## CURATOR NAME: CLASS

## GENE NAME: EUGENIA_DRAFT_1

## DNA MASTER NOTES:

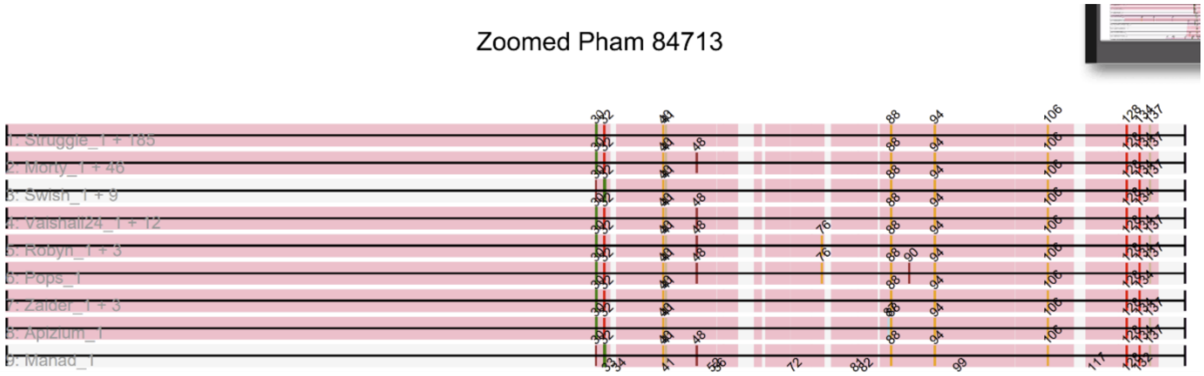
## START POSITION EVALUATION (IN ORDER):

1. Starterator

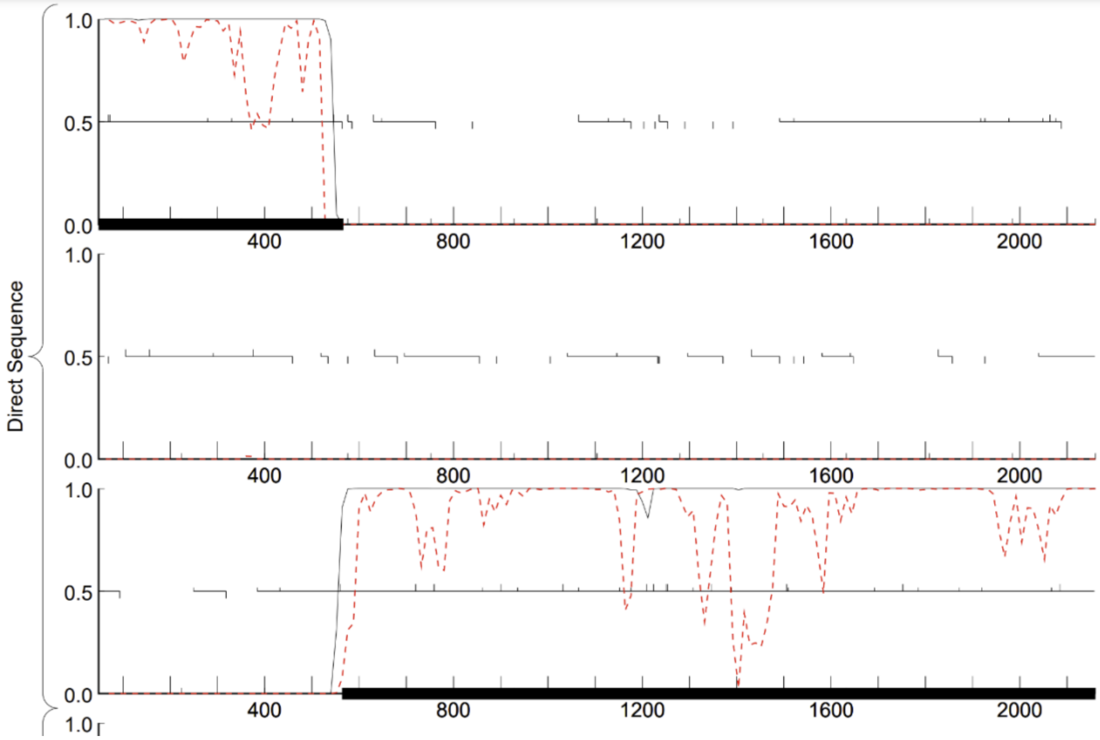
The start number called the most often in the published annotations is 30, it was called in 239 of the 310 non-draft genes in the pham.

Start 30:

- Found in 267 of 350 ( 76.3% ) of genes in pham
- Manual Annotations of this start: 239 of 310
- Called 95.9% of time when present



## 2. GeneMark coding potential



## 3. Glimmer and GeneMark agreement

YES

Glimmer Score: 13.24

Z-score: 1.724

Final Score: -6.387

| | | |
|---|---|---|
| 1.724 | -6.387 | TRUE |

## 4. Longest open reading frame (ORF) without excessive gap

YES

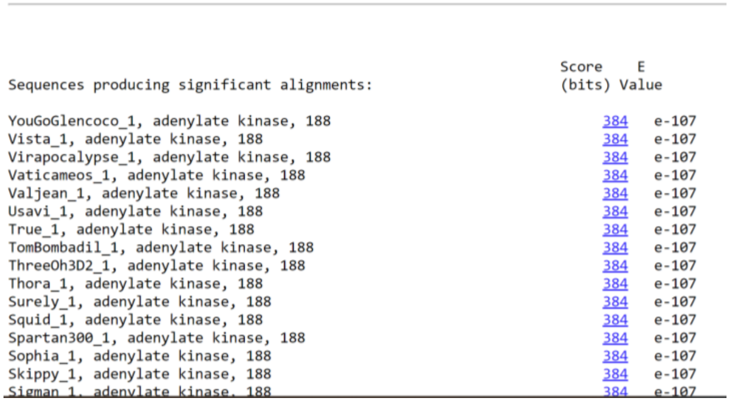
| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 1 | 567 | 567 | | 6 | 1.724 | -6.387 | TRUE | ATG | Select | ☑ |
| Forward | 10 | 567 | 558 | | 15 | 1.724 | -6.245 | | ATG | | ☐ |

## 5. If functional prediction, start include full length protein. If no functional prediction is present, write “Hypothetical protein”.

N/A

## 6. Start site similar to others in GenBank and PhagesDB

YES



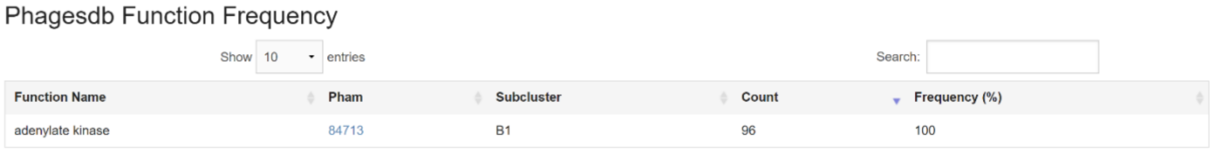
```
                                                                  Score    E
Sequences producing significant alignments:                       (bits) Value

YouGoGlencoco_1, adenylate kinase, 188                                 384   e-107
Vista_1, adenylate kinase, 188                                         384   e-107
Virapocalypse_1, adenylate kinase, 188                                 384   e-107
Vaticameos_1, adenylate kinase, 188                                    384   e-107
Valjean_1, adenylate kinase, 188                                       384   e-107
Usavi_1, adenylate kinase, 188                                         384   e-107
True_1, adenylate kinase, 188                                          384   e-107
TomBombadil_1, adenylate kinase, 188                                   384   e-107
ThreeOh3D2_1, adenylate kinase, 188                                    384   e-107
Thora_1, adenylate kinase, 188                                         384   e-107
Surely_1, adenylate kinase, 188                                        384   e-107
Squid_1, adenylate kinase, 188                                         384   e-107
Spartan300_1, adenylate kinase, 188                                    384   e-107
Sophia_1, adenylate kinase, 188                                        384   e-107
Skippy_1, adenylate kinase, 188                                        384   e-107
Sigman_1, adenylate kinase, 188                                        384   e-107
```

## 7. Associated ribosome binding site (RBS)

N/A

## 8. Function (F)

Adenylate kinase

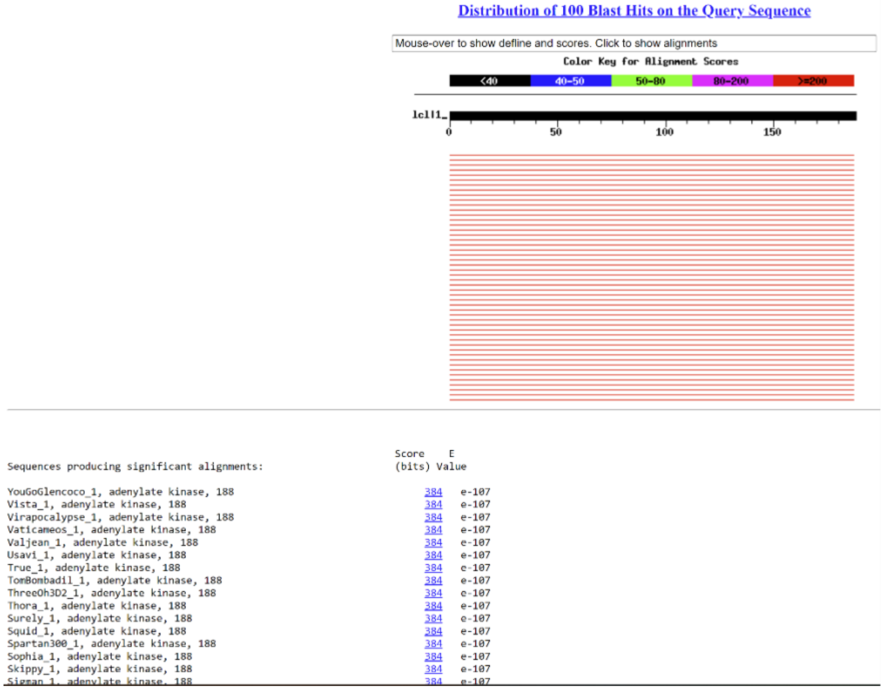


Phagesdb Function Frequency

| Function Name | Pham | Subcluster | Count | Frequency (%) |
|---|---|---|---|---|
| adenylate kinase | 84713 | B1 | 96 | 100 |

# 9. Supporting Information for Function (SIF)

## 9. a. SIF-BLAST



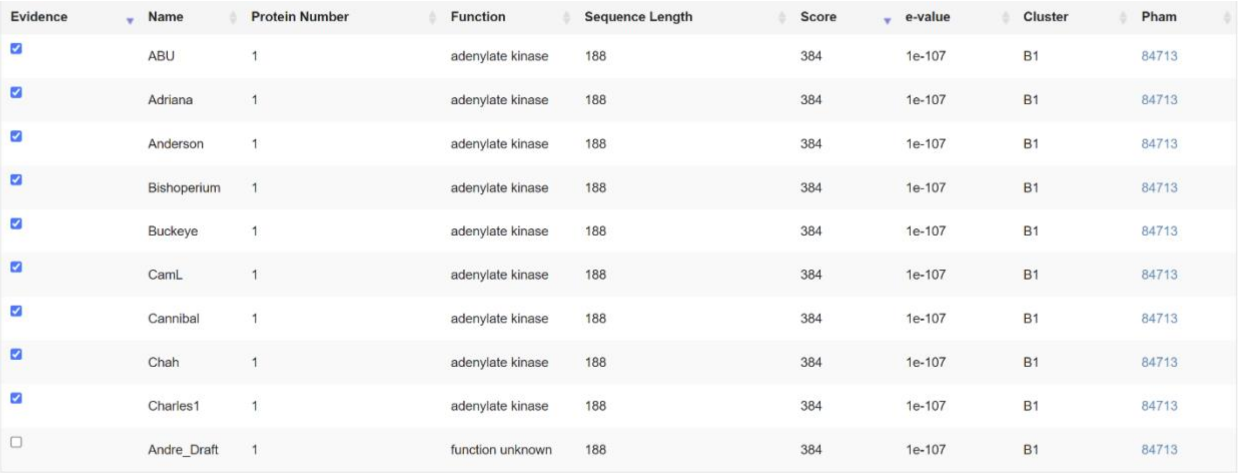
```
                                                              Score    E
Sequences producing significant alignments:                   (bits) Value

YouGoGlencoco_1, adenylate kinase, 188                           384   e-107
Vista_1, adenylate kinase, 188                                   384   e-107
Virapocalypse_1, adenylate kinase, 188                           384   e-107
Vaticameos_1, adenylate kinase, 188                              384   e-107
Valjean_1, adenylate kinase, 188                                 384   e-107
Usavi_1, adenylate kinase, 188                                   384   e-107
True_1, adenylate kinase, 188                                    384   e-107
TomBombadil_1, adenylate kinase, 188                             384   e-107
ThreeOh3D2_1, adenylate kinase, 188                              384   e-107
Thora_1, adenylate kinase, 188                                   384   e-107
Surely_1, adenylate kinase, 188                                  384   e-107
Squid_1, adenylate kinase, 188                                   384   e-107
Spartan300_1, adenylate kinase, 188                              384   e-107
Sophia_1, adenylate kinase, 188                                  384   e-107
Skippy_1, adenylate kinase, 188                                  384   e-107
Sigman_1, adenylate kinase, 188                                  384   e-107
```

## 9. b. SIF-HHPred

| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | 6LN3_A | Adenylate kinase; adenylate kinase, ATP, AMP, Structural Genomics, RIKEN Structural Genomics/Proteomics Initiative, RSGI, TRANSFERASE; HET: AMP, ATP, EPE; 2.0A {Aeropyrum pernix (strain ATCC 700893 / DSM 11879 / JCM 9820 / NBRC 100138 / K1)} | 99.8 | 97.8723 | 1 | 200 | 4 | 188 | 5.2e-17 |

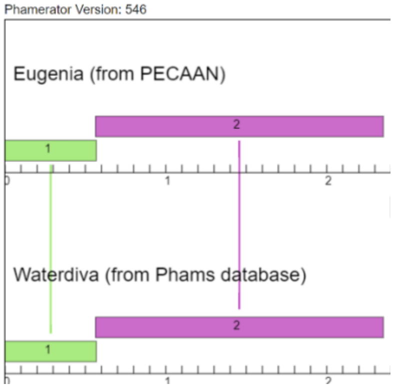
## 9. c. SIF-Synteny

Aligns with ABU, Adriana, Buckeye, CamL, Chah, Charles1, etc.



| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | ABU | 1 | adenylate kinase | 188 | 384 | 1e-107 | B1 | 84713 |
| ☑ | Adriana | 1 | adenylate kinase | 188 | 384 | 1e-107 | B1 | 84713 |
| ☑ | Anderson | 1 | adenylate kinase | 188 | 384 | 1e-107 | B1 | 84713 |
| ☑ | Bishoperium | 1 | adenylate kinase | 188 | 384 | 1e-107 | B1 | 84713 |
| ☑ | Buckeye | 1 | adenylate kinase | 188 | 384 | 1e-107 | B1 | 84713 |
| ☑ | CamL | 1 | adenylate kinase | 188 | 384 | 1e-107 | B1 | 84713 |
| ☑ | Cannibal | 1 | adenylate kinase | 188 | 384 | 1e-107 | B1 | 84713 |
| ☑ | Chah | 1 | adenylate kinase | 188 | 384 | 1e-107 | B1 | 84713 |
| ☑ | Charles1 | 1 | adenylate kinase | 188 | 384 | 1e-107 | B1 | 84713 |
| ☐ | Andre_Draft | 1 | function unknown | 188 | 384 | 1e-107 | B1 | 84713 |

Aligns with Waterdiva (final), Usavi, Slatt, etc.



10. Any other important information.

## CURATOR NAME: CLASS

## GENE NAME: EUGENIA_DRAFT_2

## DNA MASTER NOTES:

## START POSITION EVALUATION (IN ORDER):

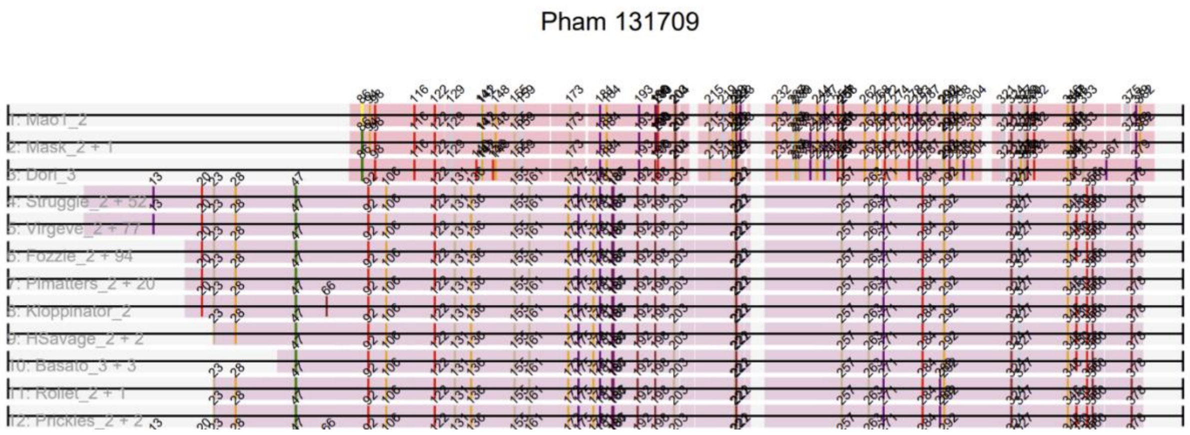
# 1. Starterator-tells us start codon

The start number called the most often in the published annotations is 47, it was called in 256 of the 451 non-draft genes in the pham.

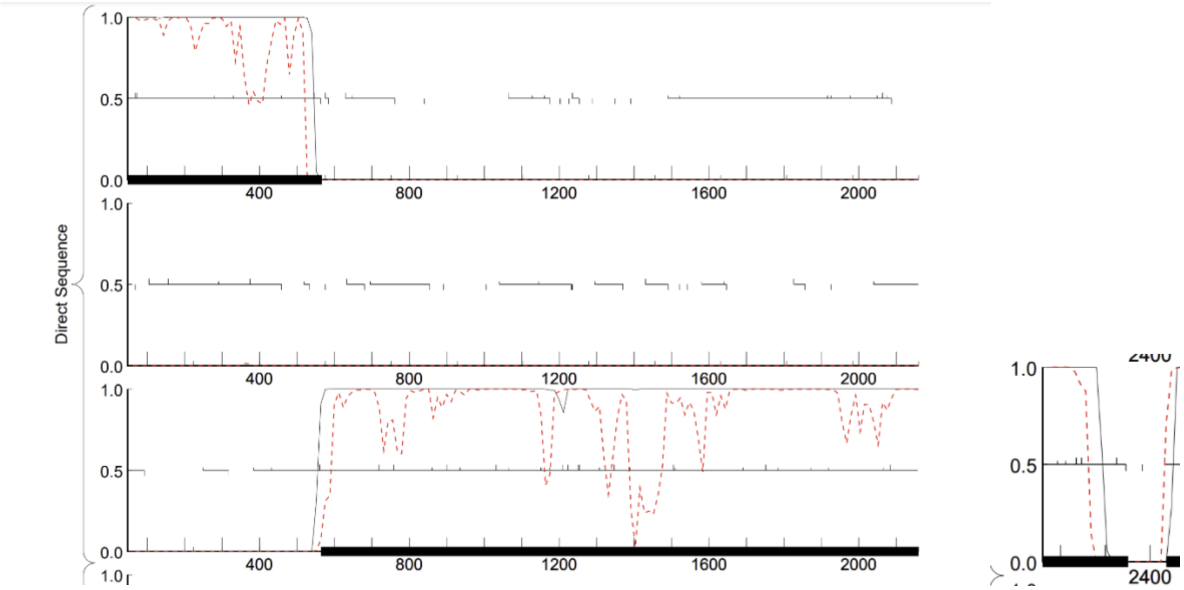
Start 47: (47 from upstream and downstream?)(Human reviewed of a final annotation)

- Found in 285 of 515 ( 55.3% ) of genes in pham
- Manual Annotations of this start: 256 of 451
- Called 100.0% of time when present

---



## 2. GeneMark coding potential



Not the longest ORF. Start of the gene around nucleotide position around 564.

## 3. Glimmer and GeneMark agreement

YES

Glimmer Score: 11.89

Z-score: 2.979

Final Score: -3.723

Phage: Eugenia Cluster: B1

Glimmer Start: 564

Glimmer Score: 11.89

GeneMark Start: 564

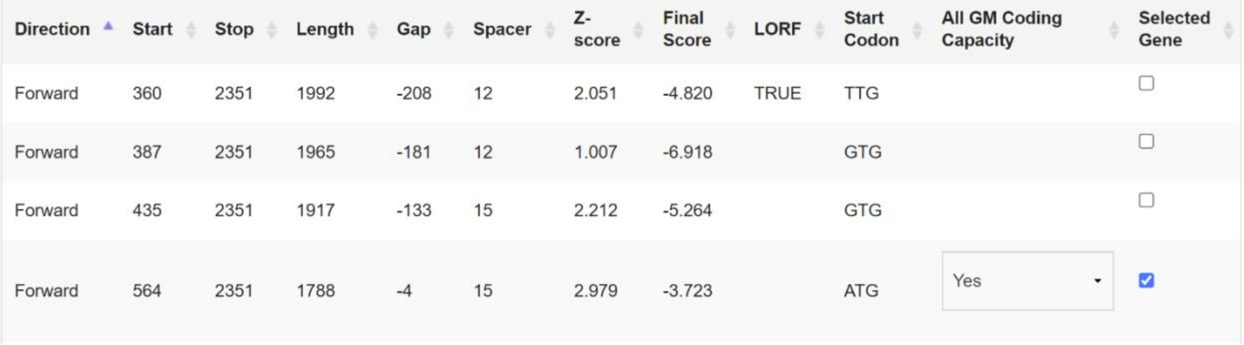
Pham Starterator: 131709

suggested start (SS)

PhagesDB: 131709

## 4. Longest open reading frame (ORF) without excessive gap

Not the longest open reading frame, but it has the smallest gap

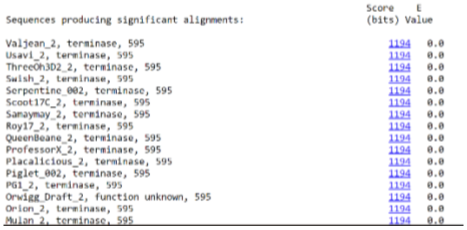


| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 360 | 2351 | 1992 | -208 | 12 | 2.051 | -4.820 | TRUE | TTG |  | ☐ |
| Forward | 387 | 2351 | 1965 | -181 | 12 | 1.007 | -6.918 |  | GTG |  | ☐ |
| Forward | 435 | 2351 | 1917 | -133 | 15 | 2.212 | -5.264 |  | GTG |  | ☐ |
| Forward | 564 | 2351 | 1788 | -4 | 15 | 2.979 | -3.723 |  | ATG | Yes | ☑ |

## 5. If functional prediction, start include full length protein. If no functional prediction is present, write “Hypothetical protein”.

## 6. Start site similar to others in GenBank and PhagesDB

YES



```
                                                                 Score     E
Sequences producing significant alignments:                      (bits) Value

Valjean_2, terminase, 595                                           1194   0.0
Usavi_2, terminase, 595                                             1194   0.0
ThreeOh3D2_2, terminase, 595                                        1194   0.0
Swish_2, terminase, 595                                             1194   0.0
Serpentine_002, terminase, 595                                      1194   0.0
Scoot17C_2, terminase, 595                                          1194   0.0
Samaymay_2, terminase, 595                                          1194   0.0
Roy17_2, terminase, 595                                             1194   0.0
QueenBeane_2, terminase, 595                                        1194   0.0
ProfessorX_2, terminase, 595                                        1194   0.0
Placalicious_2, terminase, 595                                      1194   0.0
Piglet_002, terminase, 595                                          1194   0.0
PG1_2, terminase, 595                                               1194   0.0
Orwigg_Draft_2, function unknown, 595                               1194   0.0
Orion_2, terminase, 595                                             1194   0.0
Mulan_2, terminase, 595                                             1194   0.0
```

## 7. Associated ribosome binding site (RBS)

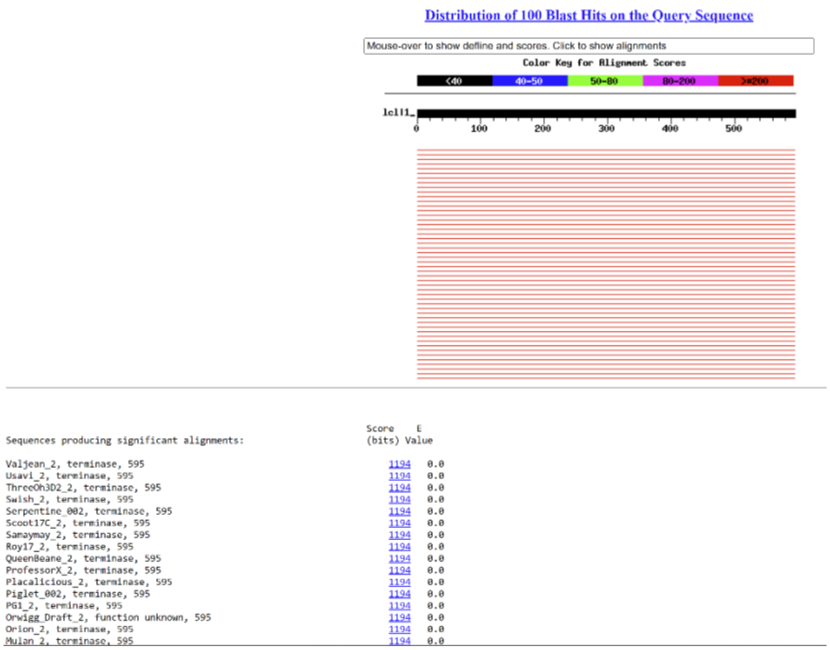
N/A

## 8. Function (F)

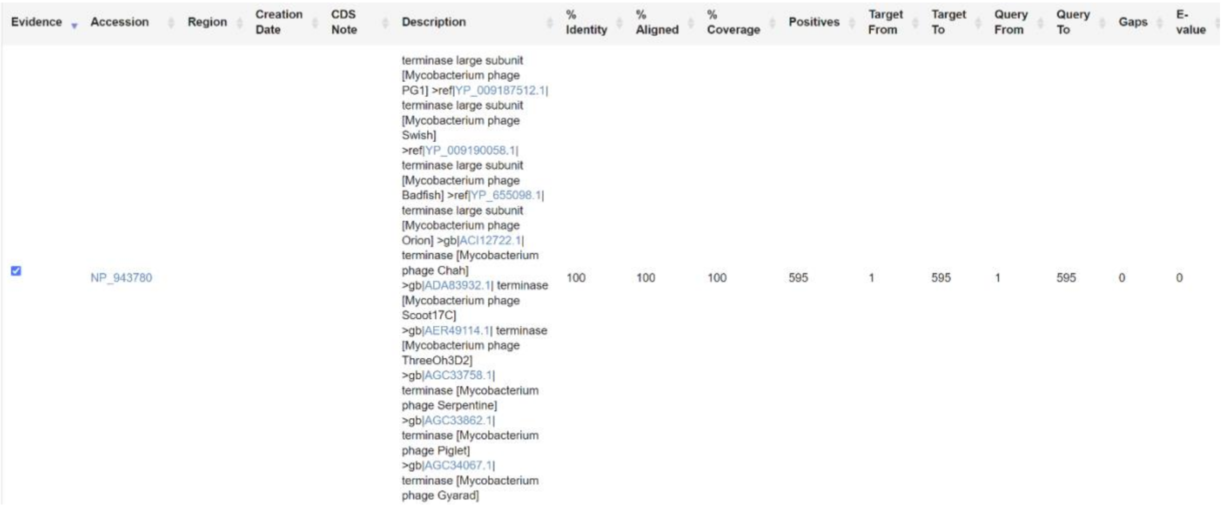
Terminase

## 9. Supporting Information for Function (SIF)

## 9. a. SIF-BLAST

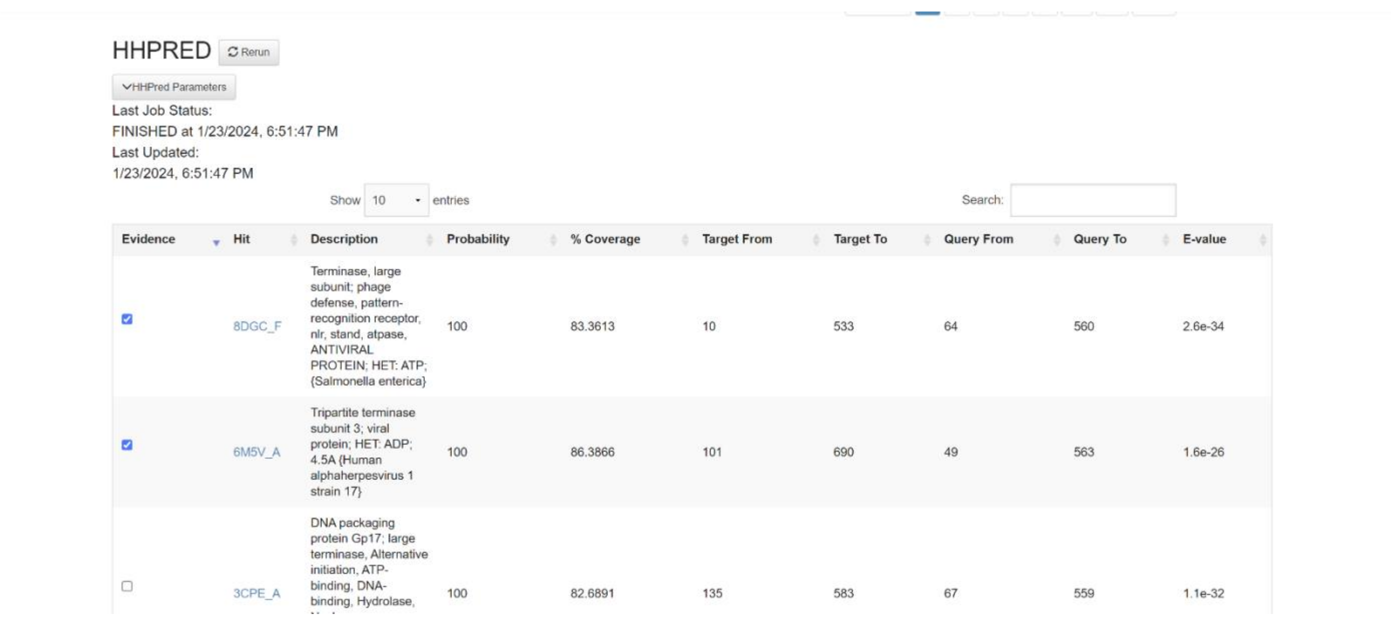


```
                                                                 Score     E
Sequences producing significant alignments:                      (bits) Value

Valjean_2, terminase, 595                                           1194   0.0
Usavi_2, terminase, 595                                             1194   0.0
ThreeOh3D2_2, terminase, 595                                        1194   0.0
Swish_2, terminase, 595                                             1194   0.0
Serpentine_002, terminase, 595                                      1194   0.0
Scoot17C_2, terminase, 595                                          1194   0.0
Samaymay_2, terminase, 595                                          1194   0.0
Roy17_2, terminase, 595                                             1194   0.0
QueenBeane_2, terminase, 595                                        1194   0.0
ProfessorX_2, terminase, 595                                        1194   0.0
Placalicious_2, terminase, 595                                      1194   0.0
Piglet_002, terminase, 595                                          1194   0.0
PG1_2, terminase, 595                                               1194   0.0
Orwigg_Draft_2, function unknown, 595                               1194   0.0
Orion_2, terminase, 595                                             1194   0.0
Mulan_2, terminase, 595                                             1194   0.0
```



| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | NP_943780 | | | | terminase large subunit [Mycobacterium phage PG1] >ref\|YP_009187512.1\| terminase large subunit [Mycobacterium phage Swish] >ref\|YP_009190058.1\| terminase large subunit [Mycobacterium phage Badfish] >ref\|YP_655098.1\| terminase large subunit [Mycobacterium phage Orion] >gb\|ACI12722.1\| terminase [Mycobacterium phage Chah] >gb\|ADA83932.1\| terminase [Mycobacterium phage Scoot17C] >gb\|AER49114.1\| terminase [Mycobacterium phage ThreeOh3D2] >gb\|AGC33758.1\| terminase [Mycobacterium phage Serpentine] >gb\|AGC33862.1\| terminase [Mycobacterium phage Piglet] >gb\|AGC34067.1\| terminase [Mycobacterium phage Gyarad] | 100 | 100 | 100 | 595 | 1 | 595 | 1 | 595 | 0 | 0 |

## 9. b. SIF-HHPred



Last Job Status:
FINISHED at 1/23/2024, 6:51:47 PM
Last Updated:
1/23/2024, 6:51:47 PM

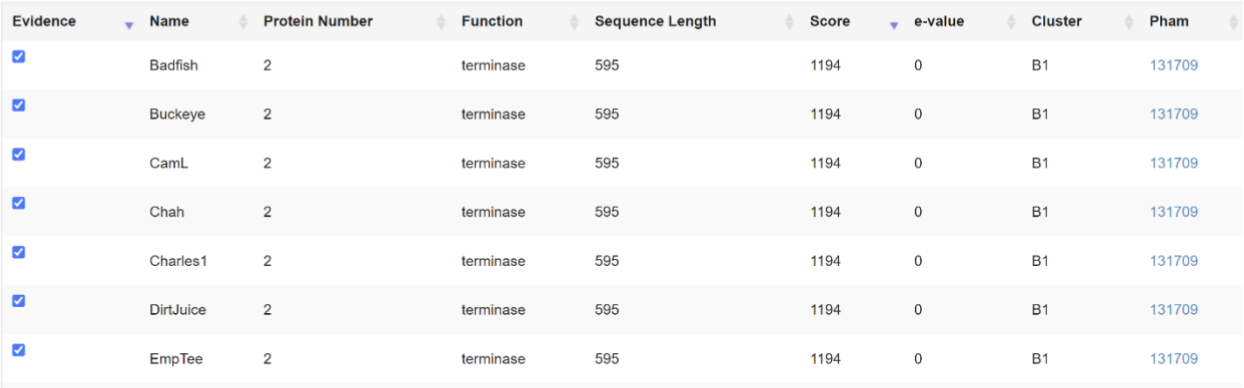
Show 10 entries

Search:

| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | 8DGC_F | Terminase, large subunit; phage defense, pattern-recognition receptor, nlr, stand, atpase, ANTIVIRAL PROTEIN; HET: ATP; {Salmonella enterica} | 100 | 83.3613 | 10 | 533 | 64 | 560 | 2.6e-34 |
| ☑ | 6M5V_A | Tripartite terminase subunit 3; viral protein; HET: ADP; 4.5A {Human alphaherpesvirus 1 strain 17} | 100 | 86.3866 | 101 | 690 | 49 | 563 | 1.6e-26 |
| ☐ | 3CPE_A | DNA packaging protein Gp17; large terminase, Alternative initiation, ATP-binding, DNA-binding, Hydrolase, | 100 | 82.6891 | 135 | 583 | 67 | 559 | 1.1e-32 |

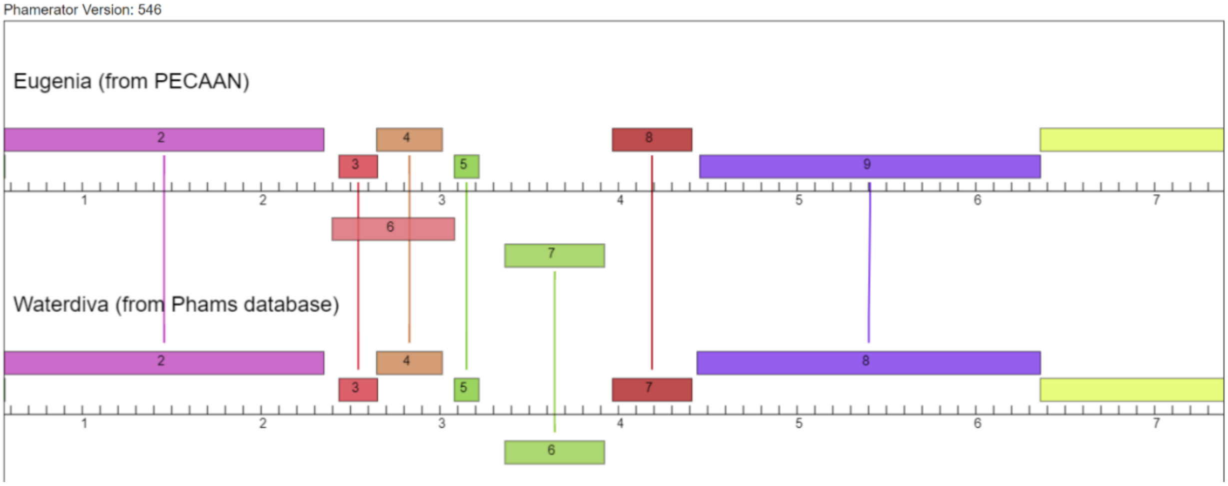
## 9. c. SIF-Synteny

Aligns with Buckeye, Badfish, CamL, Chah, Charles1, DirtJuice, EmpTee, etc.



| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | Badfish | 2 | terminase | 595 | 1194 | 0 | B1 | 131709 |
| ☑ | Buckeye | 2 | terminase | 595 | 1194 | 0 | B1 | 131709 |
| ☑ | CamL | 2 | terminase | 595 | 1194 | 0 | B1 | 131709 |
| ☑ | Chah | 2 | terminase | 595 | 1194 | 0 | B1 | 131709 |
| ☑ | Charles1 | 2 | terminase | 595 | 1194 | 0 | B1 | 131709 |
| ☑ | DirtJuice | 2 | terminase | 595 | 1194 | 0 | B1 | 131709 |
| ☑ | EmpTee | 2 | terminase | 595 | 1194 | 0 | B1 | 131709 |

Waterdiva, Usavi, etc.



## 10. Any other important information.

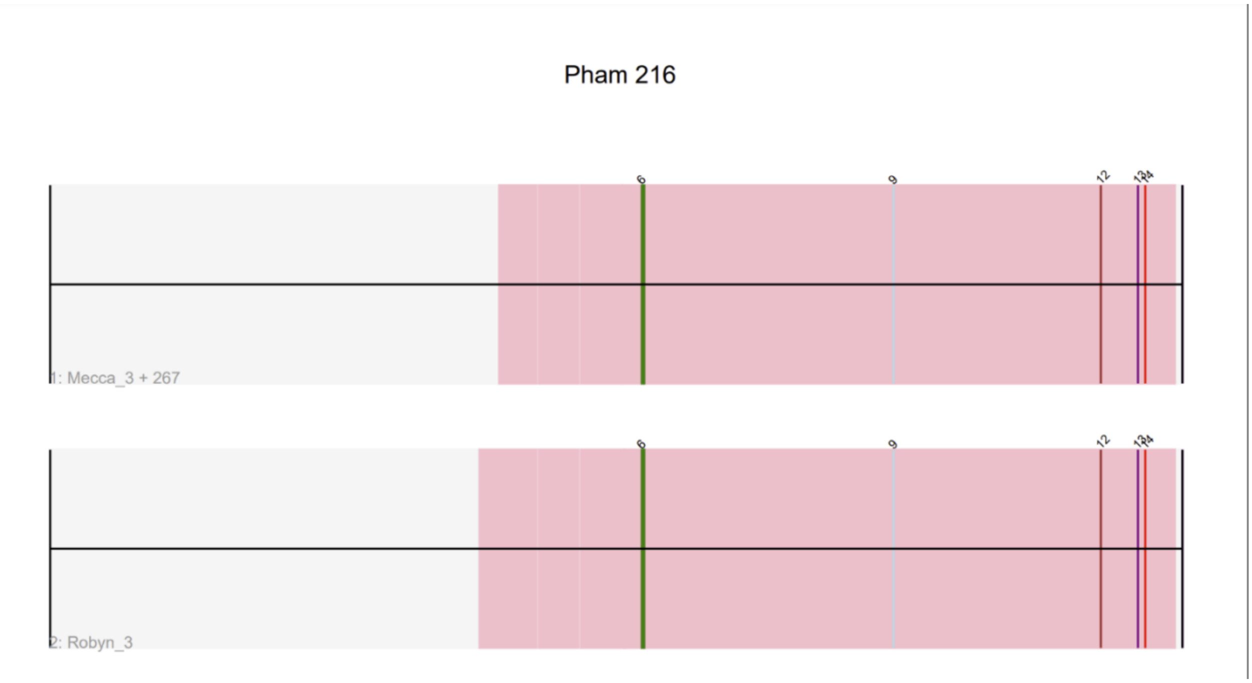
## CURATOR NAME: CLASS

## GENE NAME: EUGENIA_DRAFT_3

## DNA MASTER NOTES:

## START POSITION EVALUATION (IN ORDER):

### 1. Starterator



Eugenia’s gene is part of Track 1.

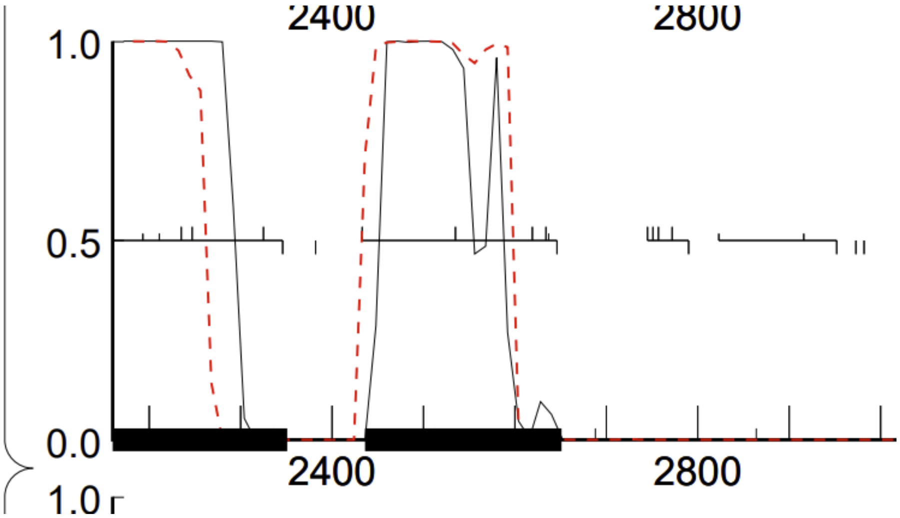
Start 6:

- Found in 269 of 274 ( 98.2% ) of genes in pham
- Manual Annotations of this start: 246 of 247
- Called 100.0% of time when present

**Note that Starterator calls Eugenia_4 instead of 3. Likely due to overlap.

## 2. GeneMark coding potential

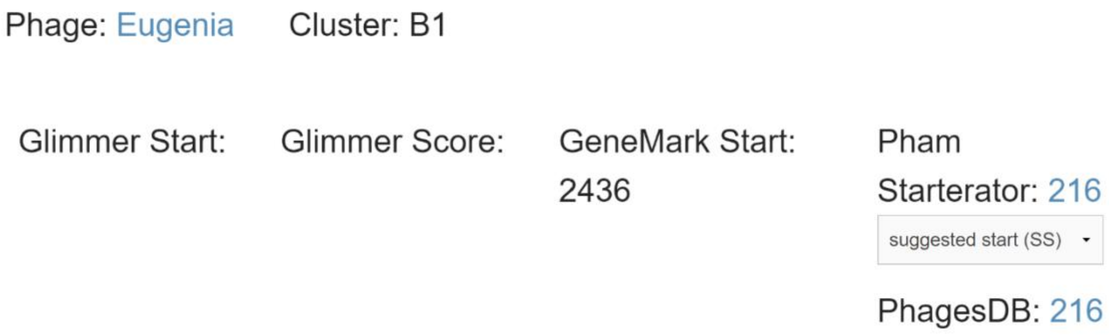
GeneMark Start: 2436



## 3. Glimmer and GeneMark agreement

No Glimmer score, so NO agreement

GeneMark Start: 2436



Phage: Eugenia Cluster: B1

Glimmer Start:

Glimmer Score:

GeneMark Start: 2436

Pham Starterator: 216

suggested start (SS)

PhagesDB: 216

## 4. Longest open reading frame (ORF) without excessive gap

YES

| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 2436 | 2651 | 216 | 84 | 13 | 3.229 | -2.666 | TRUE | ATG | Yes | ☑ |
| Forward | 2538 | 2651 | 114 | 186 | 10 | 1.666 | -5.454 |  | ATG |  | ☐ |

Z-score: 3.229

Final Score: -2.666

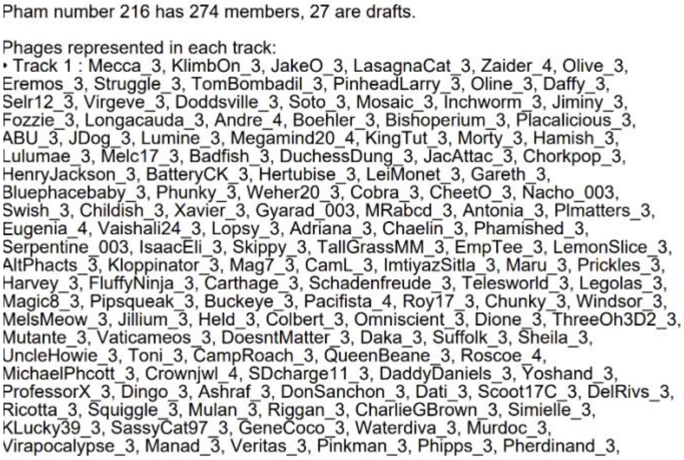
Spacer: 13

## 5. If functional prediction, start include full length protein. If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

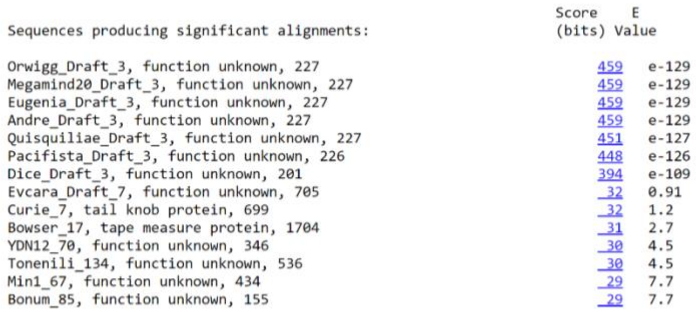
## 6. Start site similar to others in GenBank and PhagesDB

YES



Pham number 216 has 274 members, 27 are drafts.

Phages represented in each track:
- Track 1 : Mecca_3, KlimbOn_3, JakeO_3, LasagnaCat_3, Zaider_4, Olive_3, Eremos_3, Struggle_3, TomBombadil_3, PinheadLarry_3, Oline_3, Daffy_3, Selr12_3, Virgeve_3, Doddsville_3, Soto_3, Mosaic_3, Inchworm_3, Jiminy_3, Fozzie_3, Longacauda_3, Andre_4, Boehler_3, Bishoperium_3, Placalicious_3, ABU_3, JDog_3, Lumine_3, Megamind20_4, KingTut_3, Morty_3, Hamish_3, Lulumae_3, Melc17_3, Badfish_3, DuchessDung_3, JacAttac_3, Chorkpop_3, HenryJackson_3, BatteryCK_3, Hertubise_3, LeiMonet_3, Gareth_3, Bluephacebaby_3, Phunky_3, Weher20_3, Cobra_3, CheetO_3, Nacho_003, Swish_3, Childish_3, Xavier_3, Gyarad_003, MRabcd_3, Antonia_3, Plmatters_3, Eugenia_4, Vaishali24_3, Lopsy_3, Adriana_3, Chaelin_3, Phamished_3, Serpentine_003, IsaacEli_3, Skippy_3, TallGrassMM_3, EmpTee_3, LemonSlice_3, AltPhacts_3, Kloppinator_3, Mag7_3, CamL_3, ImtiyazSitla_3, Maru_3, Prickles_3, Harvey_3, FluffyNinja_3, Carthage_3, Schadenfreude_3, Telesworld_3, Legolas_3, Magic8_3, Pipsqueak_3, Buckeye_3, Pacifista_4, Roy17_3, Chunky_3, Windsor_3, MelsMeow_3, Jillium_3, Held_3, Colbert_3, Omniscient_3, Dione_3, ThreeOh3D2_3, Mutante_3, Vaticameos_3, DoesntMatter_3, Daka_3, Suffolk_3, Sheila_3, UncleHowie_3, Toni_3, CampRoach_3, QueenBeane_3, Roscoe_4, MichaelPhcott_3, Crownjwl_4, SDcharge11_3, DaddyDaniels_3, Yoshand_3, ProfessorX_3, Dingo_3, Ashraf_3, DonSanchon_3, Dati_3, Scoot17C_3, DelRivs_3, Ricotta_3, Squiggle_3, Mulan_3, Riggan_3, CharlieGBrown_3, Simielle_3, KLucky39_3, SassyCat97_3, GeneCoco_3, Waterdiva_3, Murdoc_3, Virapocalypse_3, Manad_3, Veritas_3, Pinkman_3, Phipps_3, Pherdinand_3,



```
                                                          Score    E
Sequences producing significant alignments:               (bits) Value

Orwigg_Draft_3, function unknown, 227                       459   e-129
Megamind20_Draft_3, function unknown, 227                   459   e-129
Eugenia_Draft_3, function unknown, 227                      459   e-129
Andre_Draft_3, function unknown, 227                        459   e-129
Quisquiliae_Draft_3, function unknown, 227                  451   e-127
Pacifista_Draft_3, function unknown, 226                    448   e-126
Dice_Draft_3, function unknown, 201                         394   e-109
Evcara_Draft_7, function unknown, 705                        32   0.91
Curie_7, tail knob protein, 699                              32   1.2
Bowser_17, tape measure protein, 1704                        31   2.7
YDN12_70, function unknown, 346                              30   4.5
Tonenili_134, function unknown, 536                          30   4.5
Min1_67, function unknown, 434                               29   7.7
Bonum_85, function unknown, 155                              29   7.7
```

## 7. Associated ribosome binding site (RBS)

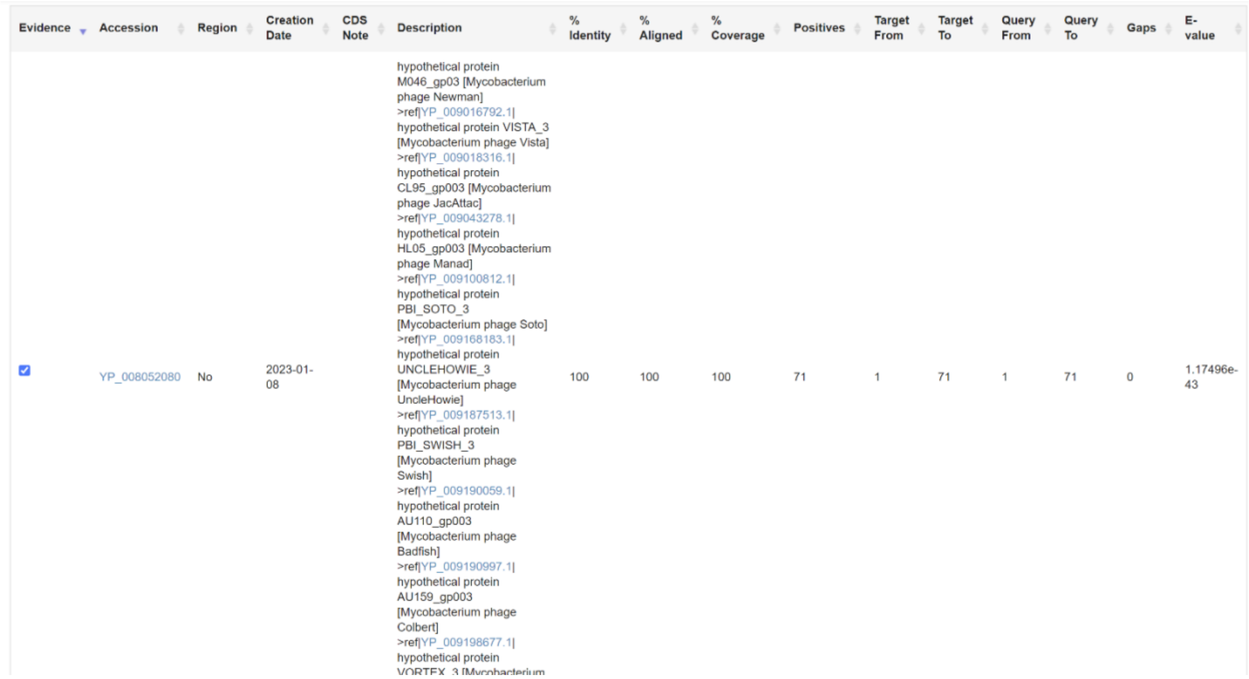
N/A

## 8. Function (F)

Hypothetical protein—no known function

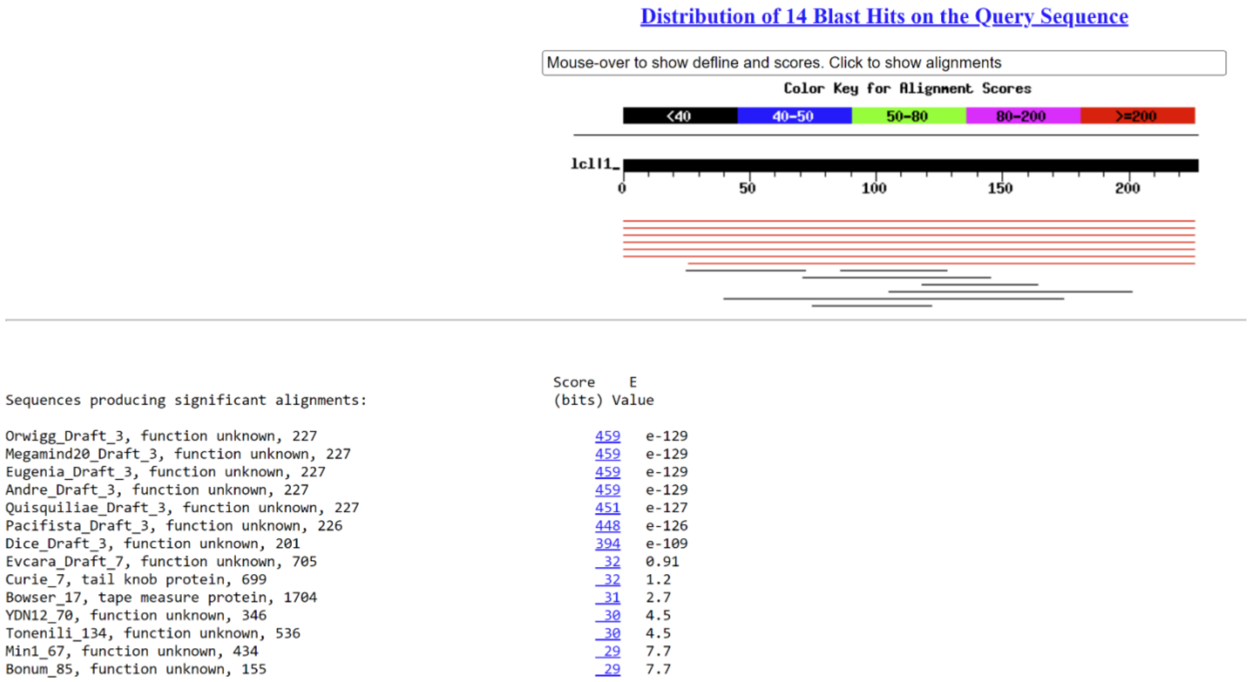
# 9. Supporting Information for Function (SIF)

## 9. a. SIF-BLAST



| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | YP_008052080 | No | 2023-01-08 | | hypothetical protein M046_gp03 [Mycobacterium phage Newman] >ref\|YP_009016792.1\| hypothetical protein VISTA_3 [Mycobacterium phage Vista] >ref\|YP_009018316.1\| hypothetical protein CL95_gp003 [Mycobacterium phage JacAttac] >ref\|YP_009043278.1\| hypothetical protein HL05_gp003 [Mycobacterium phage Manad] >ref\|YP_009100812.1\| hypothetical protein PBI_SOTO_3 [Mycobacterium phage Soto] >ref\|YP_009168183.1\| hypothetical protein UNCLEHOWIE_3 [Mycobacterium phage UncleHowie] >ref\|YP_009187513.1\| hypothetical protein PBI_SWISH_3 [Mycobacterium phage Swish] >ref\|YP_009190059.1\| hypothetical protein AU110_gp003 [Mycobacterium phage Badfish] >ref\|YP_009190997.1\| hypothetical protein AU159_gp003 [Mycobacterium phage Colbert] >ref\|YP_009198677.1\| hypothetical protein VORTEX_3 [Mycobacterium | 100 | 100 | 100 | 71 | 1 | 71 | 1 | 71 | 0 | 1.17496e-43 |

| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | Dingo | 3 | function unknown | 71 | 151 | 5e-37 | B1 | 216 |
| ☑ | DirtJuice | 3 | function unknown | 71 | 151 | 5e-37 | B1 | 216 |
| ☑ | DoesntMatter | 3 | function unknown | 71 | 151 | 5e-37 | B1 | 216 |



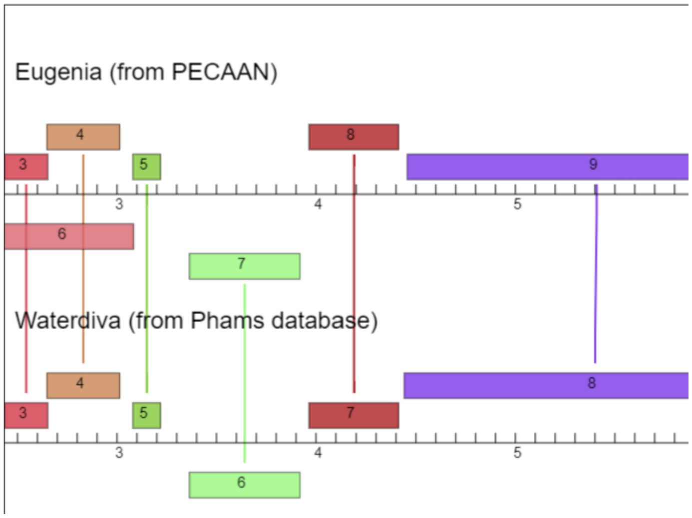
```
                                                                   Score    E
Sequences producing significant alignments:                        (bits) Value

Orwigg_Draft_3, function unknown, 227                                 459   e-129
Megamind20_Draft_3, function unknown, 227                             459   e-129
Eugenia_Draft_3, function unknown, 227                                459   e-129
Andre_Draft_3, function unknown, 227                                  459   e-129
Quisquiliae_Draft_3, function unknown, 227                            451   e-127
Pacifista_Draft_3, function unknown, 226                              448   e-126
Dice_Draft_3, function unknown, 201                                   394   e-109
Evcara_Draft_7, function unknown, 705                                  32   0.91
Curie_7, tail knob protein, 699                                        32   1.2
Bowser_17, tape measure protein, 1704                                  31   2.7
YDN12_70, function unknown, 346                                        30   4.5
Tonenili_134, function unknown, 536                                    30   4.5
Min1_67, function unknown, 434                                         29   7.7
Bonum_85, function unknown, 155                                        29   7.7
```

## 9. b. SIF-HHPred

| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | PF04475.16 | DUF555 ; Protein of unknown function (DUF555) | 49.5 | 26.7606 | 79 | 98 | 45 | 64 | 39 |

## 9. c. SIF-Synteny

Phamerator Version: 547



Additional synteny with Dingo, DirtJuice, and DoesntMatter.

| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | Dingo | 3 | function unknown | 71 | 151 | 5e-37 | B1 | 216 |
| ☑ | DirtJuice | 3 | function unknown | 71 | 151 | 5e-37 | B1 | 216 |
| ☑ | DoesntMatter | 3 | function unknown | 71 | 151 | 5e-37 | B1 | 216 |

## 10. Any other important information.

Appears to have overlap with Genes 4-6.

## CURATOR NAME: CLASS

## GENE NAME: EUGENIA_DRAFT_4

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

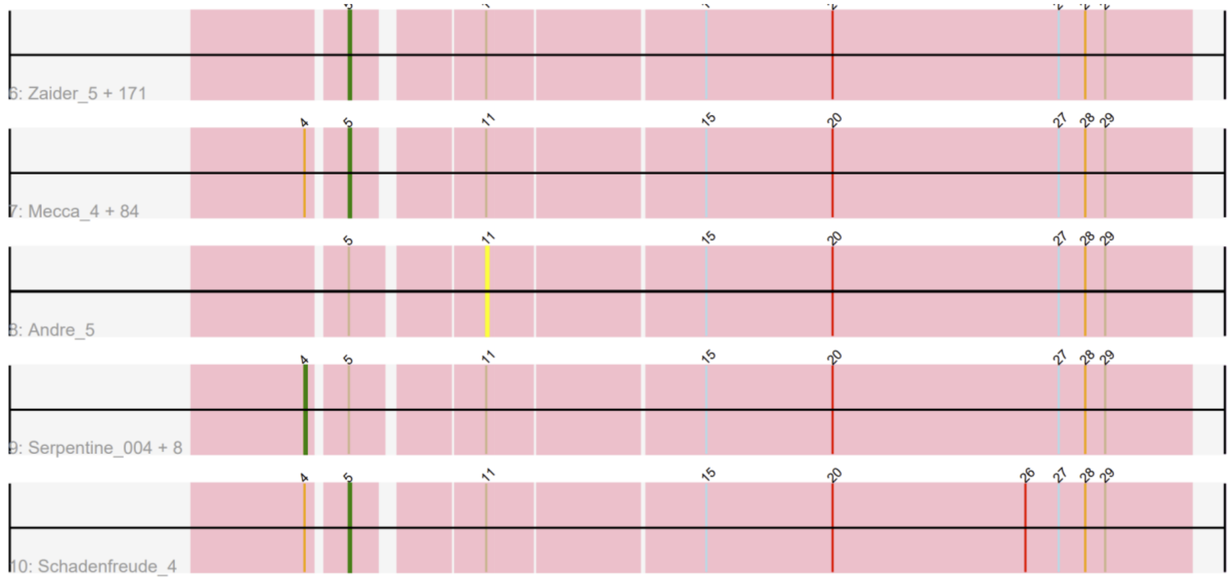
### 1. Starterator

The start number called the most often in the published annotations is 5, it was called in 239 of the 255 non-draft genes in the pham.

Start 5:

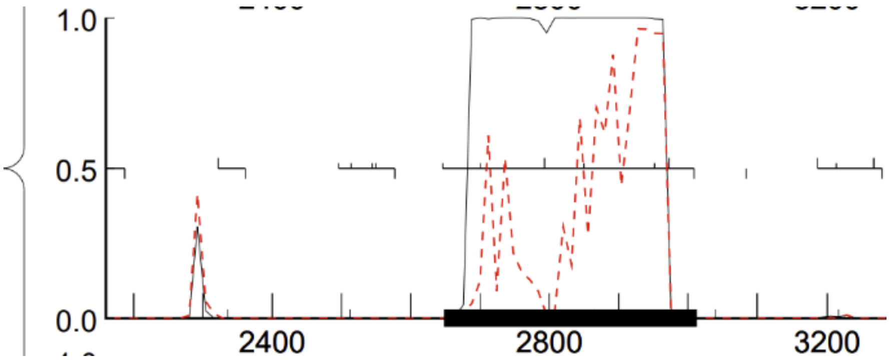
- Found in 269 of 279 ( 96.4% ) of genes in pham
- Manual Annotations of this start: 239 of 255
- Called 95.9% of time when present

*Eugenia is in Track 6. Note that Starterator calls Eugenia_5, not 4. This is likely due to overlap.



## 2. GeneMark coding potential

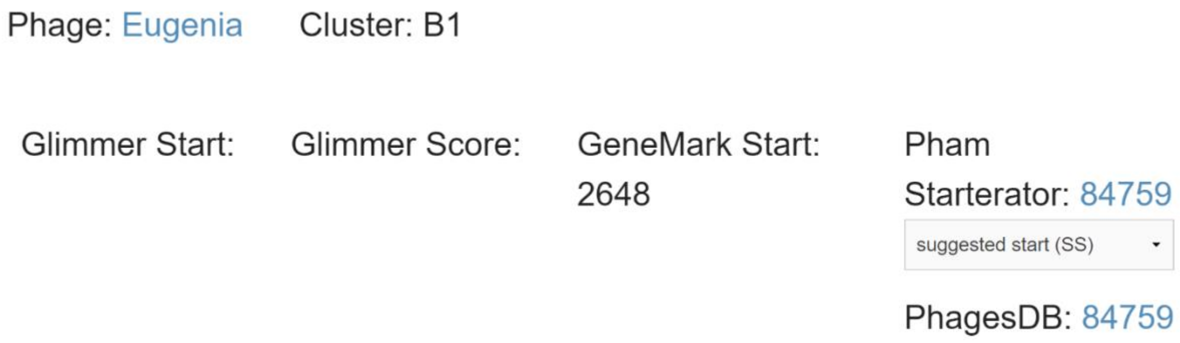
GeneMark Start: 2648



## 3. Glimmer and GeneMark agreement

No Glimmer score, so NO agreement

GeneMark Start: 2648



Phage: Eugenia Cluster: B1

Glimmer Start:

Glimmer Score:

GeneMark Start: 2648

Pham

Starterator: 84759

suggested start (SS) ▾

PhagesDB: 84759

## 4. Longest open reading frame (ORF) without excessive gap

YES

| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 2648 | 3013 | 366 | -4 | 14 | 2.634 | -4.161 | TRUE | GTG | Select | ☑ |
| Forward | 2702 | 3013 | 312 | 50 | 14 | 1.357 | -6.726 |  | GTG |  | ☐ |

Length: 366

Gap: -4

Spacer: 14

Z-Score: 2.634

Final Score: -4.161

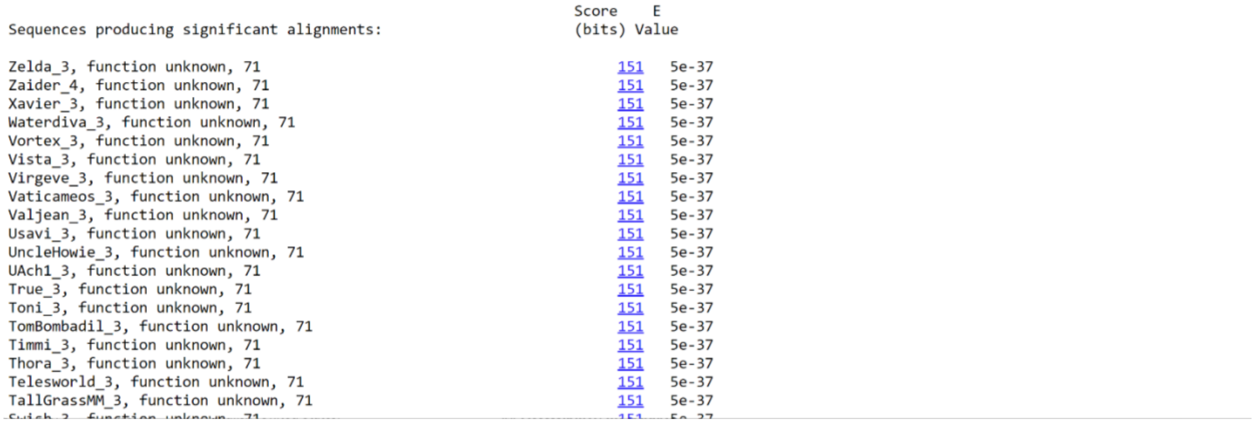
## 5. If functional prediction, start include full length protein. If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

## 6. Start site similar to others in GenBank and PhagesDB

YES

- Track 6 : Zaider_5, LasagnaCat_4, Jillium_4, KlimbOn_4, PinheadLarry_4, Olive_4, Virgeve_4, Daffy_4, Oline_4, Jiminy_4, Inchworm_4, Mosaic_4, Soto_4, Soile_4, Placalicious_4, Gophee_4, Boehler_4, Doddsville_4, Brilliant_4, ABU_4, Megamind20_5, Hamish_4, Morty_4, KingTut_4, Melc17_4, Lulumae_4, HenryJackson_4, Chorkpop_4, JacAttac_4, Hetaeria_4, Antonia_4, Hertubise_4, Durga_4, BatteryCK_4, Chaelin_4, CheetO_4, Weher20_4, MitKao_4, Bluephacebaby_4, Gareth_4, Xavier_4, Childish_4, Swish_4, Badfish_4, Eugenia_5, Adriana_4, Skippy_4, IsaacEli_4, Phamished_4, LemonSlice_4, EmpTee_4, TallGrassMM_4, Duggie_4, ImtiyazSitla_4, Mag7_4, Prickles_4, Legolas_4, Telesworld_4, Roy17_4, Burr_4, MelsMeow_4, Buckeye_4, Magic8_4, Windsor_4, Colbert_4, Held_4, DoesntMatter_4, Mutante_4, Etaye_4, Veritas_4, LeeLot_4, CampRoach_4, Toni_4, UncleHowie_4, MichaelPhcott_4, Dingo_4, Scoot17C_4, Dati_4, DonSanchon_4, Ricotta_4, CharlieGBrown_4, PhrodoBaggins_4, Manad_4, Murdoc_4, Waterdiva_4, GeneCoco_4, Vaticameos_4, LeiMonet_4, Ashraf_4, Pinkman_4, Daka_4, PhrankReynolds_4, Phareon_4, Vista_4, Maskar_4, DelRivs_4, Kailash_4, TyrionL_4, MrPhizzler_4, Trypo_4, Katniss_4, FugateOSU_4, Matalotodo_4, Solosis_4, Charles1_4, PhenghisKhan_4, Orwigg_5, PhatCats2014_4, JDog_4, Usavi_4, Anderson_4, Serendipity_4, Cher_4, PhatLouie_4, Craff_4, Schueller_4, Potter_4, Olak_4, Fang_4, Sigman_4, Kwadwo_4, Swiphy_4, Mikota_4, Kwksand96_4, JakeO_4, Bishoperium_4, Zonia_4, Banjo_4, Sophia_4, AbsoluteMadLad_4, Zelda_4, Wallhey_4, Giraffe_4, LostAndPhound_4, Surely_4, Labeouficaum_4, True_4, Timmi_4, Nyala_4, DuchessDung_4, Phleuron_4, Kikipoo_4, Oosterbaan_4, Mesh1_4, Haimas_4, HighStump_4, Emiris_4, Megatron_4, Spartan300_4, Vortex_4, Cannibal_4, Phergie_4, Orfeu_4, Squiggle_4, Aelin_4, DaddyDaniels_4, Mana_4, OliverWalter_4, DirtJuice_4, Hocus_4, UAch1_4, Rimu_4, Derpp_4, TomBombadil_4, Newman_4, Valjean_4, Hartsy_4, Thora_4, Mulan_4, Kahve_4, Longacauda_4, BlueHusk_4



```
                                                                Score    E
Sequences producing significant alignments:                     (bits) Value

Zelda_3, function unknown, 71                                      151   5e-37
Zaider_4, function unknown, 71                                     151   5e-37
Xavier_3, function unknown, 71                                     151   5e-37
Waterdiva_3, function unknown, 71                                  151   5e-37
Vortex_3, function unknown, 71                                     151   5e-37
Vista_3, function unknown, 71                                      151   5e-37
Virgeve_3, function unknown, 71                                    151   5e-37
Vaticameos_3, function unknown, 71                                 151   5e-37
Valjean_3, function unknown, 71                                    151   5e-37
Usavi_3, function unknown, 71                                      151   5e-37
UncleHowie_3, function unknown, 71                                 151   5e-37
UAch1_3, function unknown, 71                                      151   5e-37
True_3, function unknown, 71                                       151   5e-37
Toni_3, function unknown, 71                                       151   5e-37
TomBombadil_3, function unknown, 71                                151   5e-37
Timmi_3, function unknown, 71                                      151   5e-37
Thora_3, function unknown, 71                                      151   5e-37
Telesworld_3, function unknown, 71                                 151   5e-37
TallGrassMM_3, function unknown, 71                                151   5e-37
```

## 7. Associated ribosome binding site (RBS)

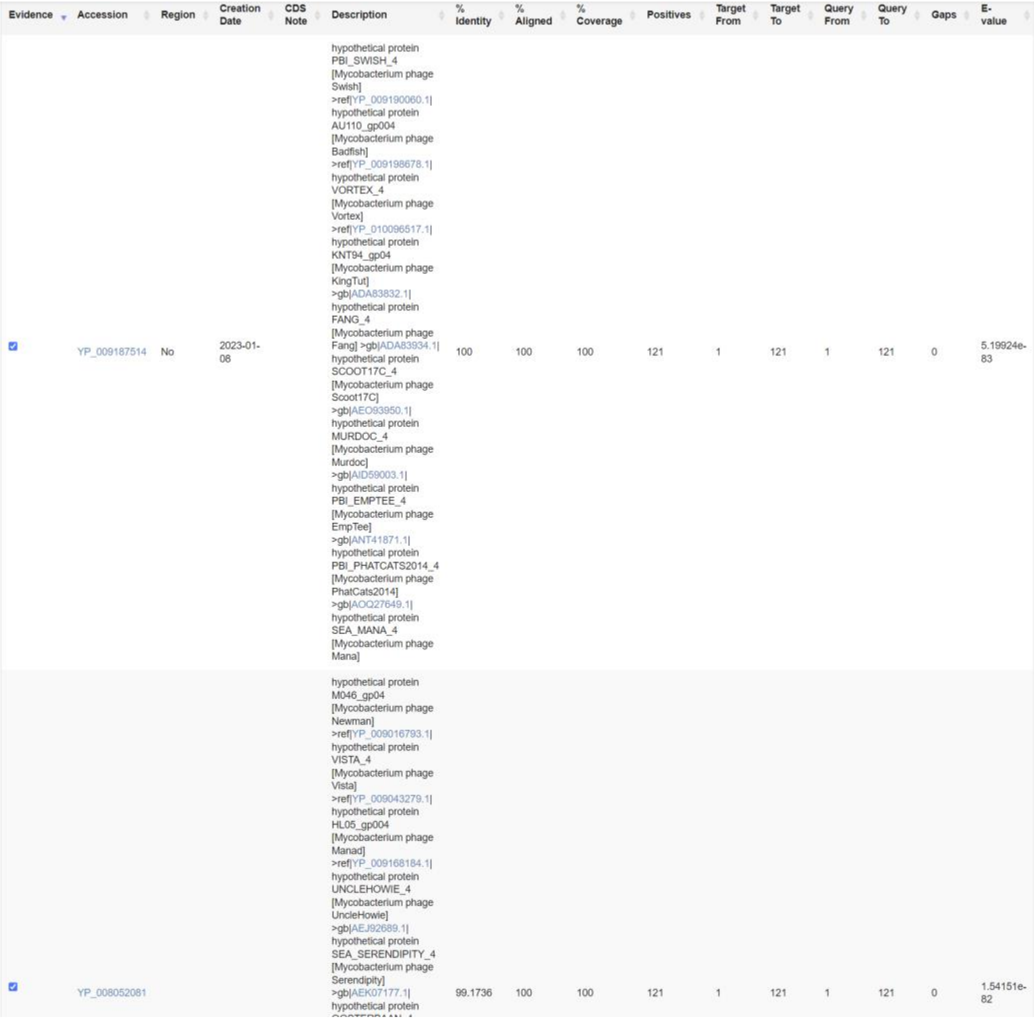
N/A

## 8. Function (F)

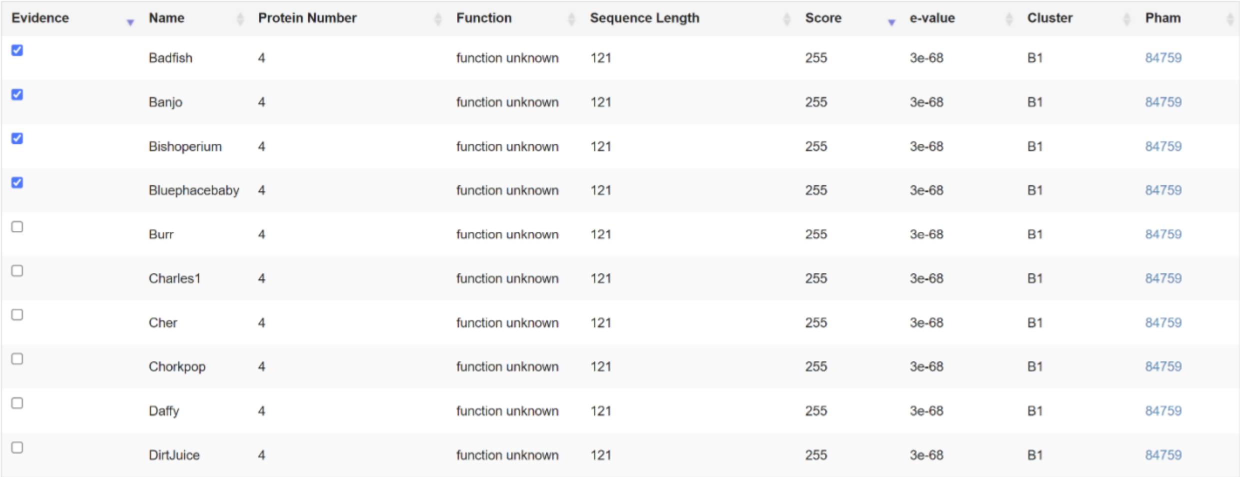
Hypothetical protein—Unknown function

# 9. Supporting Information for Function (SIF)

## 9. a. SIF-BLAST



| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | YP_009187514 | No | 2023-01-08 |  | hypothetical protein PBI_SWISH_4 [Mycobacterium phage Swish] >ref\|YP_009190060.1\| hypothetical protein AU110_gp004 [Mycobacterium phage Badfish] >ref\|YP_009198678.1\| hypothetical protein VORTEX_4 [Mycobacterium phage Vortex] >ref\|YP_010096517.1\| hypothetical protein KNT94_gp04 [Mycobacterium phage KingTut] >gb\|ADA83832.1\| hypothetical protein FANG_4 [Mycobacterium phage Fang] >gb\|ADA83934.1\| hypothetical protein SCOOT17C_4 [Mycobacterium phage Scoot17C] >gb\|AEO93950.1\| hypothetical protein MURDOC_4 [Mycobacterium phage Murdoc] >gb\|AID59003.1\| hypothetical protein PBI_EMPTEE_4 [Mycobacterium phage EmpTee] >gb\|ANT41871.1\| hypothetical protein PBI_PHATCATS2014_4 [Mycobacterium phage PhatCats2014] >gb\|AOQ27649.1\| hypothetical protein SEA_MANA_4 [Mycobacterium phage Mana] | 100 | 100 | 100 | 121 | 1 | 121 | 1 | 121 | 0 | 5.19924e-83 |
| ☑ | YP_008052081 |  |  |  | hypothetical protein M046_gp04 [Mycobacterium phage Newman] >ref\|YP_009016793.1\| hypothetical protein VISTA_4 [Mycobacterium phage Vista] >ref\|YP_009043279.1\| hypothetical protein HL05_gp004 [Mycobacterium phage Manad] >ref\|YP_009168184.1\| hypothetical protein UNCLEHOWIE_4 [Mycobacterium phage UncleHowie] >gb\|AEJ92689.1\| hypothetical protein SEA_SERENDIPITY_4 [Mycobacterium phage Serendipity] >gb\|AEK07177.1\| hypothetical protein | 99.1736 | 100 | 100 | 121 | 1 | 121 | 1 | 121 | 0 | 1.54151e-82 |



| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | Badfish | 4 | function unknown | 121 | 255 | 3e-68 | B1 | 84759 |
| ☑ | Banjo | 4 | function unknown | 121 | 255 | 3e-68 | B1 | 84759 |
| ☑ | Bishoperium | 4 | function unknown | 121 | 255 | 3e-68 | B1 | 84759 |
| ☑ | Bluephacebaby | 4 | function unknown | 121 | 255 | 3e-68 | B1 | 84759 |
| ☐ | Burr | 4 | function unknown | 121 | 255 | 3e-68 | B1 | 84759 |
| ☐ | Charles1 | 4 | function unknown | 121 | 255 | 3e-68 | B1 | 84759 |
| ☐ | Cher | 4 | function unknown | 121 | 255 | 3e-68 | B1 | 84759 |
| ☐ | Chorkpop | 4 | function unknown | 121 | 255 | 3e-68 | B1 | 84759 |
| ☐ | Daffy | 4 | function unknown | 121 | 255 | 3e-68 | B1 | 84759 |
| ☐ | DirtJuice | 4 | function unknown | 121 | 255 | 3e-68 | B1 | 84759 |

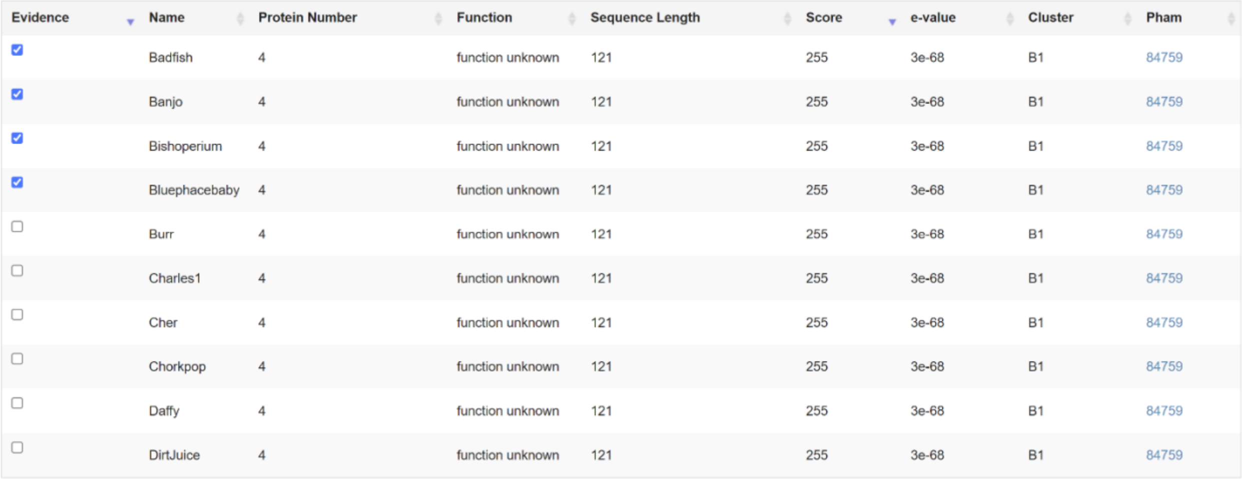
```
                                                                  Score    E
Sequences producing significant alignments:                       (bits) Value

Zelda_3, function unknown, 71                                         151   5e-37
Zaider_4, function unknown, 71                                        151   5e-37
Xavier_3, function unknown, 71                                        151   5e-37
Waterdiva_3, function unknown, 71                                     151   5e-37
Vortex_3, function unknown, 71                                        151   5e-37
Vista_3, function unknown, 71                                         151   5e-37
Virgeve_3, function unknown, 71                                       151   5e-37
Vaticameos_3, function unknown, 71                                    151   5e-37
Valjean_3, function unknown, 71                                       151   5e-37
Usavi_3, function unknown, 71                                         151   5e-37
UncleHowie_3, function unknown, 71                                    151   5e-37
UAch1_3, function unknown, 71                                         151   5e-37
True_3, function unknown, 71                                          151   5e-37
Toni_3, function unknown, 71                                          151   5e-37
TomBombadil_3, function unknown, 71                                   151   5e-37
Timmi_3, function unknown, 71                                         151   5e-37
Thora_3, function unknown, 71                                         151   5e-37
Telesworld_3, function unknown, 71                                    151   5e-37
TallGrassMM_3, function unknown, 71                                   151   5e-37
```

## 9. b. SIF-HHPred

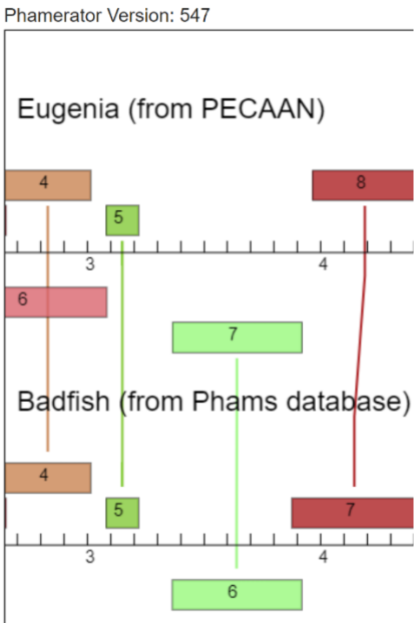
N/A

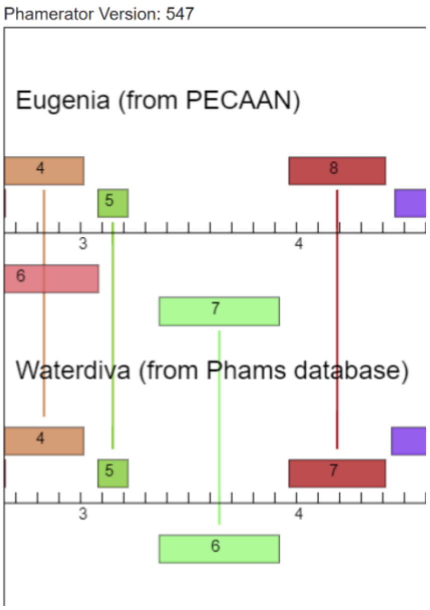
## 9. c. SIF-Synteny

Aligned with Badfish, Banjo, Bishoperium, Bluephacebaby, Waterdiva, etc.



| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | Badfish | 4 | function unknown | 121 | 255 | 3e-68 | B1 | 84759 |
| ☑ | Banjo | 4 | function unknown | 121 | 255 | 3e-68 | B1 | 84759 |
| ☑ | Bishoperium | 4 | function unknown | 121 | 255 | 3e-68 | B1 | 84759 |
| ☑ | Bluephacebaby | 4 | function unknown | 121 | 255 | 3e-68 | B1 | 84759 |
| ☐ | Burr | 4 | function unknown | 121 | 255 | 3e-68 | B1 | 84759 |
| ☐ | Charles1 | 4 | function unknown | 121 | 255 | 3e-68 | B1 | 84759 |
| ☐ | Cher | 4 | function unknown | 121 | 255 | 3e-68 | B1 | 84759 |
| ☐ | Chorkpop | 4 | function unknown | 121 | 255 | 3e-68 | B1 | 84759 |
| ☐ | Daffy | 4 | function unknown | 121 | 255 | 3e-68 | B1 | 84759 |
| ☐ | DirtJuice | 4 | function unknown | 121 | 255 | 3e-68 | B1 | 84759 |





# 10. Any other important information.

Some overlap with Gene 3 near the start.

## CURATOR NAME: ISABELLA LIMA

## GENE NAME: EUGENIA_DRAFT_5

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

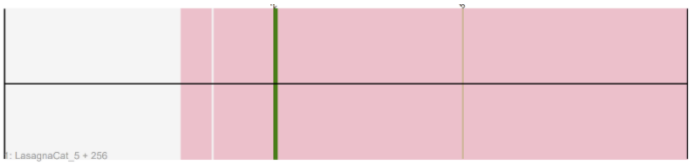
### 1. Starterator

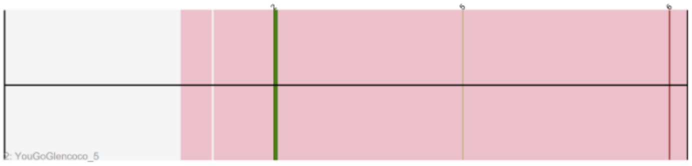
Start Number: 2

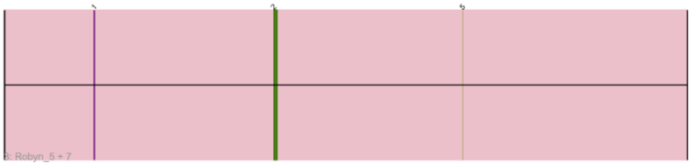
The start number called the most often in the published annotations is 2, it was called in 243 of the 245 non-draft genes in the pham.

Start 2:

- Found in 266 of 268 ( 99.3% ) of genes in pham
- Manual Annotations of this start: 243 of 245
- Called 100.0% of time when present

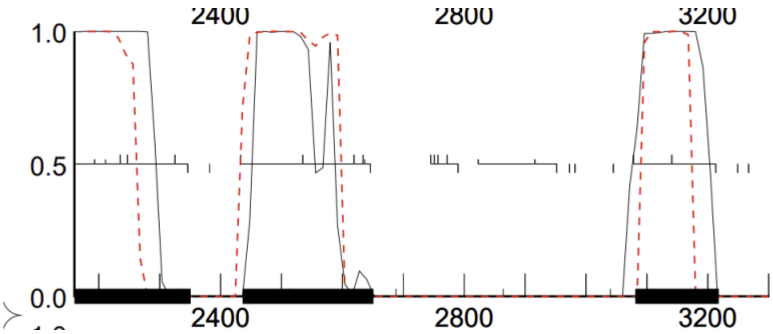






## 2. GeneMark coding potential

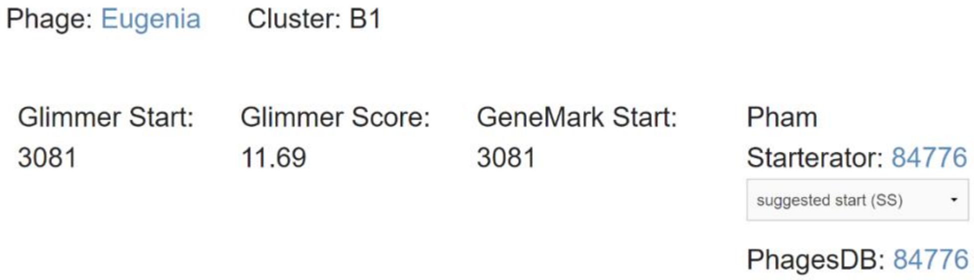
GeneMark: 3081



## 3. Glimmer and GeneMark agreement

YES

Glimmer Start: 3081
Glimmer Score: 11.69
GeneMark Start: 3081



Phage: Eugenia Cluster: B1

| Glimmer Start: | Glimmer Score: | GeneMark Start: | Pham |
|---|---|---|---|
| 3081 | 11.69 | 3081 | Starterator: 84776 |

PhagesDB: 84776

## 4. Longest open reading frame (ORF) without excessive gap

YES

| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 3081 | 3218 | 138 | 67 | 10 | 1.896 | -4.991 | TRUE | ATG | Select | ☑ |
| Forward | 3144 | 3218 | 75 | 130 | 7 | 1.373 | -6.871 |  | ATG |  | ☐ |

Length: 138

Gap: 67

Spacer: 10

Z-score: 1.896

Final Score: -4.991

## 6. Start site similar to others in GenBank and PhagesDB

YES

Phages represented in each track:
• Track 1 : LasagnaCat_5, Nicole21_5, Zaider_6, KlimbOn_5, Mecca_5,
PinheadLarry_5, Derpp_5, LemonSlice_5, Eremos_5, Olive_5, Virgeve_5, Selr12_5,
Daffy_5, Oline_5, TomBombadil_5, Inchworm_5, Jiminy_5, Mosaic_5, Soto_5,
Serpentine_005, Fozzie_5, Andre_6, Placalicious_5, Phunky_5, Antonia_5,
Boehler_5, LeiMonet_5, ABU_5, Brilliant_5, Megamind20_6, Hamish_5, KingTut_5,
Morty_5, Melc17_5, Lulumae_5, Hetaeria_5, Suffolk_5, HenryJackson_5, JacAttac_5,
Chorkpop_5, Chaelin_5, Hertubise_5, BatteryCK_5, CheetO_5, Weher20_5,
MitKao_5, Gareth_5, Bluephacebaby_5, MRabcd_5, Xavier_5, RedMaple_5,
Swish_5, Lulwa_5, Badfish_5, FluffyNinja_5, Eugenia_6, Durga_5, Plmatters_5,
Spartan300_5, Vaishali24_5, Adriana_5, IsaacEli_5, Skippy_5, Yoshand_5,
Struggle_5, Phamished_5, Soile_5, EmpTee_5, TallGrassMM_5, CamL_5,
ImtiyazSitla_5, Mag7_5, AltPhacts_5, Kloppinator_5, Schadenfreude_5, Harvey_5,
Prickles_5, Maru_5, Telesworld_5, Legolas_5, Roy17_5, Cannibal_5, Buckeye_5,
Magic8_5, Pops_5, Jillium_5, Windsor_5, FugateOSU_5, ThreeOh3D2_5,
Omniscient_5, Dione_5, Held_5, Colbert_5, DoesntMatter_5, Vaticameos_5,
Mutante_5, Sheila_5, Veritas_5, QueenBeane_5, CampRoach_5, SDcharge11_5,
UncleHowie_5, Toni_5, Podrick_5, MichaelPhcott_5, Fringe_5, Dingo_5, Newman_5,
DaddyDaniels_5, Scoot17C_5, Dati_5, DonSanchon_5, Daka_5, Ricotta_5, Dice_6,
CharlieGBrown_5, Riggan_5, Squiggle_5, UAch1_5, PhrodoBaggins_5,
Virapocalypse_5, Manad_5, Murdoc_5, GeneCoco_5, Waterdiva_5, Pherdinand_5,
Apizium_5, Pinkman_5, Phipps_5, ProfessorX_5, PhrankReynolds_5, Gophee_5,
Pacifista_6, BlackStallion_5, Maskar_5, Vista_5, Phareon_5, DelRivs_5,
MrPhizzler_5, TyrionL_5, Trypo_5, Kailash_5, Lasso_5, ShiVal_5, SassyCat97_5,
Charles1_5, Katniss_5, Solosis_5, Matalotodo_5, Schueller_5, MiniBoss_5,
Orwigg_6, Childish_5, PhatCats2014_5, Cosmolli16_5, Giraffe_5, JangoPhett_5,
Lopsy_5, Mesh1_5, Ashraf_5, Serendipity_5, JDog_5, Etaye_5, Basato_6, JakeO_5,
Cher_5, PhatLouie_5, Craff_5, Tomlarah_5, Olak_5, Potter_5, Sigman_5, Fang_5,
Kwadwo_5, Mikota_5, Swiphy_5, Grand2040_5, Morgushi_5, Kwksand96_5,
Cornobble_5, Horchata_5, Banjo_5, Zonia_5, Doddsville_5, Beaglebox_5,
Piglet_005, LeeLot_5, Samaymay_5, Squid_5, Sophia_5, AbsoluteMadLad_5,
Anderson_5, Scrick_5, Haleema_5, Bishoperium_5, Wallhey_5, MelsMeow_5,
Hocus_5, LostAndPhound_5, Kimbrough_5, Surely_5, Labeouficaum_5, Timmi_5,
True_5, FriarPreacher_5, Nyala_5, Keitherie_5, Vivaldi_5, Altwerkus_5,
DuchessDung_5, Phleuron_5, Kikipoo_5, Quisquiliae_6, Oosterbaan_5,
KingVeVeVe_5, OSmaximus_5, Haimas_5, Emiris_5, HighStump_5, Lumine_5,
Numberten_5, Megatron_5, Cobra_5, Mcshane_5, Lego3393_5, Duggie_5, Vortex_5,
Phergie_5, Tooj_5, PG1_5, Chunky_5, Burr_5, Orfeu_5, Iridoclysis_5, Aelin_5,

## 7. Associated ribosome binding site (RBS)

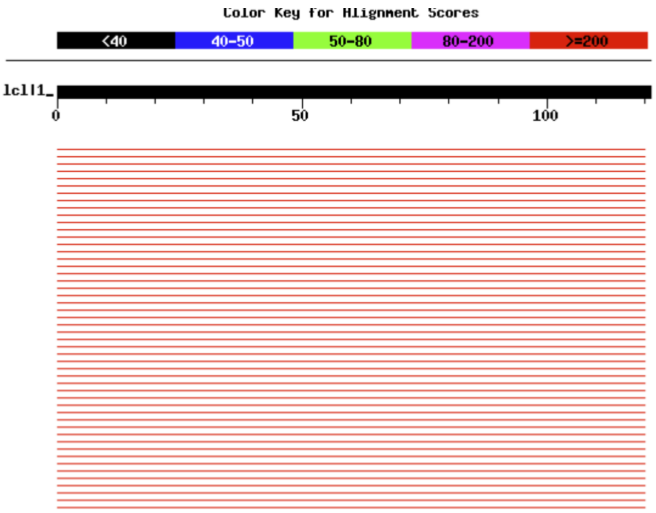
N/A

## 8. Function (F)

Hypothetical protein

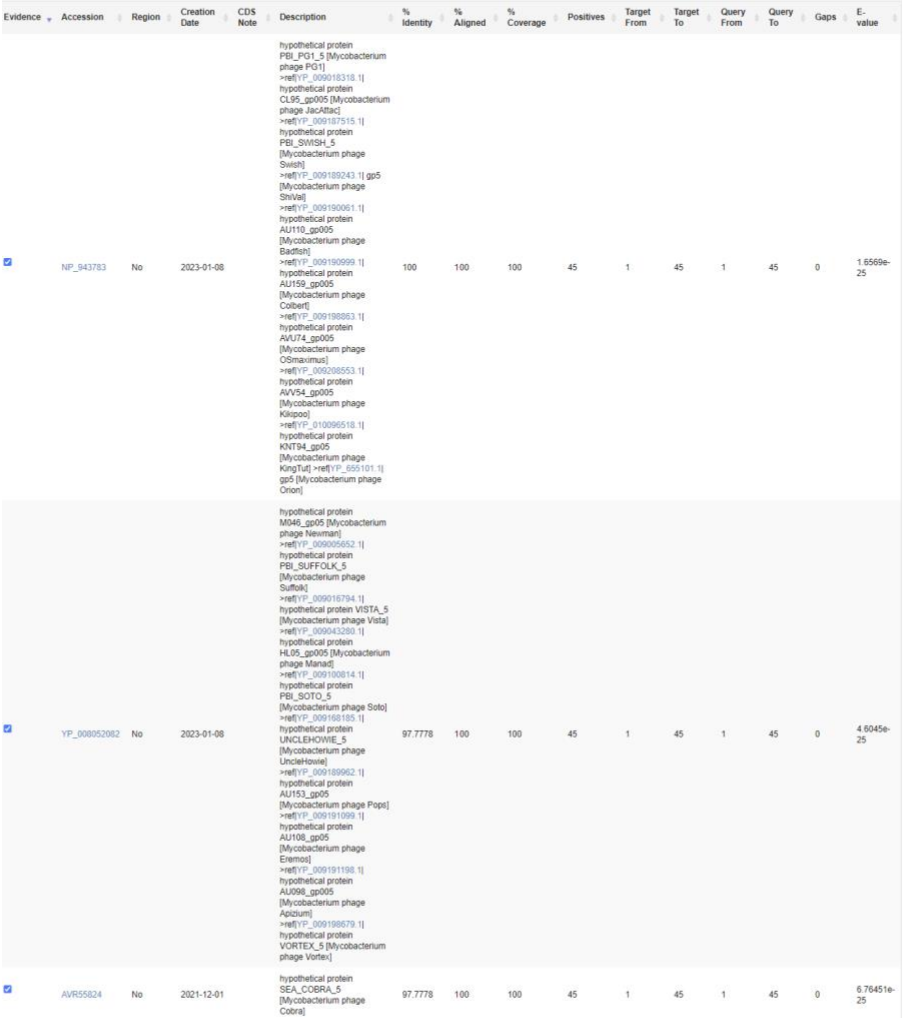
## 9. Supporting Information for Function (SIF)

## 9. a. SIF-PhageDB BLAST



```
                                                                 Score    E
Sequences producing significant alignments:                      (bits) Value

Zelda_4, function unknown, 121                                       255   3e-68
Xavier_4, function unknown, 121                                      255   3e-68
Vortex_4, function unknown, 121                                      255   3e-68
Toni_4, function unknown, 121                                        255   3e-68
Telesworld_4, function unknown, 121                                  255   3e-68
Swish_4, function unknown, 121                                       255   3e-68
Swiphy_Draft_4, function unknown, 121                                255   3e-68
Squiggle_4, function unknown, 121                                    255   3e-68
Scoot17C_4, function unknown, 121                                    255   3e-68
Roy17_4, function unknown, 121                                       255   3e-68
Placalicious_4, function unknown, 121                                255   3e-68
PhatLouie_4, function unknown, 121                                   255   3e-68
PhatCats2014_4, function unknown, 121                                255   3e-68
Phareon_4, function unknown, 121                                     255   3e-68
Orwigg_Draft_5, function unknown, 121                                255   3e-68
Nyala_4, function unknown, 121                                       255   3e-68
Murdoc_4, function unknown, 121                                      255   3e-68
Mulan_4, function unknown, 121                                       255   3e-68
Mosaic_4, function unknown, 121                                      255   3e-68
MitKao_4, function unknown, 121                                      255   3e-68
```

## 9.b. SIF- NCBI BLAST



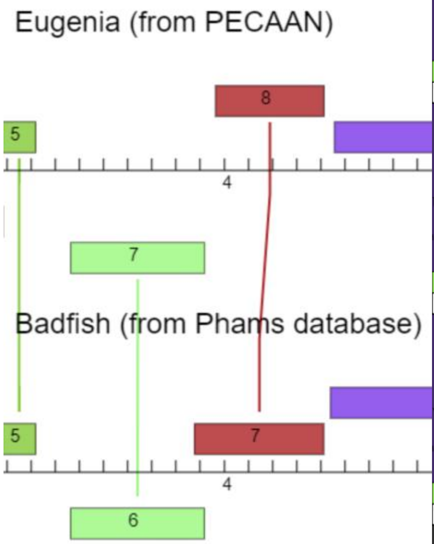
| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | NP_943783 | No | 2023-01-08 |  | hypothetical protein PBI_PG1_5 [Mycobacterium phage PG1] >ref\|YP_009018318.1\| hypothetical protein CL95_gp005 [Mycobacterium phage JacAttac] >ref\|YP_009187515.1\| hypothetical protein PBI_SWISH_5 [Mycobacterium phage Swish] >ref\|YP_009189243.1\| gp5 [Mycobacterium phage ShiVal] >ref\|YP_009190061.1\| hypothetical protein AU110_gp005 [Mycobacterium phage Badfish] >ref\|YP_009190999.1\| hypothetical protein AU159_gp005 [Mycobacterium phage Colbert] >ref\|YP_009198863.1\| hypothetical protein AVU74_gp005 [Mycobacterium phage OSmaximus] >ref\|YP_009208553.1\| hypothetical protein AVV54_gp005 [Mycobacterium phage Kikipoo] >ref\|YP_010096518.1\| hypothetical protein KNT94_gp05 [Mycobacterium phage KingTut] >ref\|YP_655101.1\| gp5 [Mycobacterium phage Orion] | 100 | 100 | 100 | 45 | 1 | 45 | 1 | 45 | 0 | 1.6569e-25 |
| ☑ | YP_008052082 | No | 2023-01-08 |  | hypothetical protein M046_gp05 [Mycobacterium phage Newman] >ref\|YP_009005652.1\| hypothetical protein PBI_SUFFOLK_5 [Mycobacterium phage Suffolk] >ref\|YP_009016794.1\| hypothetical protein VISTA_5 [Mycobacterium phage Vista] >ref\|YP_009043280.1\| hypothetical protein HL05_gp005 [Mycobacterium phage Manad] >ref\|YP_009100814.1\| hypothetical protein PBI_SOTO_5 [Mycobacterium phage Soto] >ref\|YP_009168185.1\| hypothetical protein UNCLEHOWIE_5 [Mycobacterium phage UncleHowie] >ref\|YP_009189962.1\| hypothetical protein AU153_gp05 [Mycobacterium phage Pops] >ref\|YP_009191099.1\| hypothetical protein AU108_gp05 [Mycobacterium phage Eremos] >ref\|YP_009191198.1\| hypothetical protein AU098_gp005 [Mycobacterium phage Apizium] >ref\|YP_009198679.1\| hypothetical protein VORTEX_5 [Mycobacterium phage Vortex] | 97.7778 | 100 | 100 | 45 | 1 | 45 | 1 | 45 | 0 | 4.6045e-25 |
| ☑ | AVR55824 | No | 2021-12-01 |  | hypothetical protein SEA_COBRA_5 [Mycobacterium phage Cobra] | 97.7778 | 100 | 100 | 45 | 1 | 45 | 1 | 45 | 0 | 6.76451e-25 |

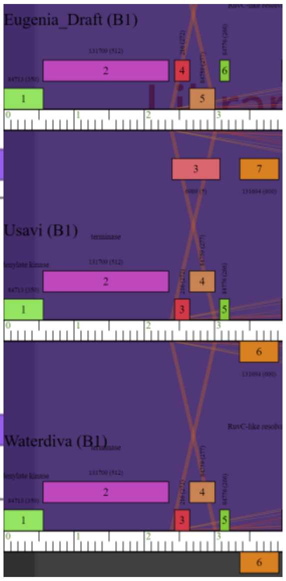
## 9. c. SIF-HHPred

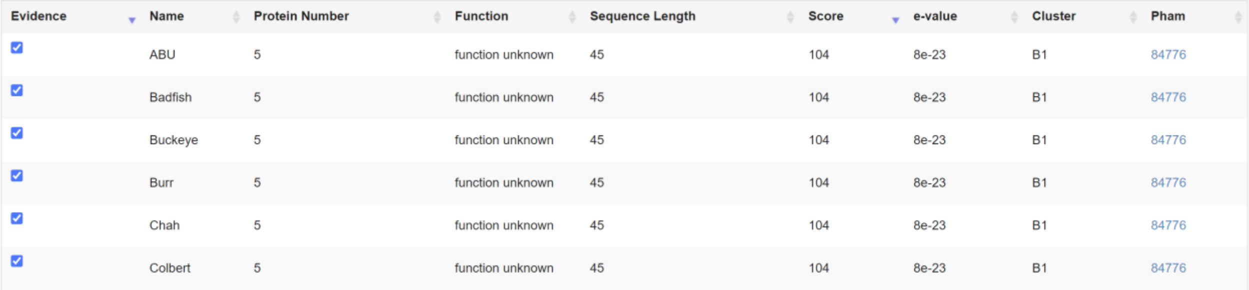
N/A

## 9. d. SIF-Synteny

Aligns with ABU, Badfish, Buckeye, Burr, Chah, Colbert, Waterdiva, etc.







| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | ABU | 5 | function unknown | 45 | 104 | 8e-23 | B1 | 84776 |
| ☑ | Badfish | 5 | function unknown | 45 | 104 | 8e-23 | B1 | 84776 |
| ☑ | Buckeye | 5 | function unknown | 45 | 104 | 8e-23 | B1 | 84776 |
| ☑ | Burr | 5 | function unknown | 45 | 104 | 8e-23 | B1 | 84776 |
| ☑ | Chah | 5 | function unknown | 45 | 104 | 8e-23 | B1 | 84776 |
| ☑ | Colbert | 5 | function unknown | 45 | 104 | 8e-23 | B1 | 84776 |

# 10. Any other important information.

N/A

## CURATOR NAME: ISABELLA LIMA

## GENE NAME: EUGENIA_DRAFT_6

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

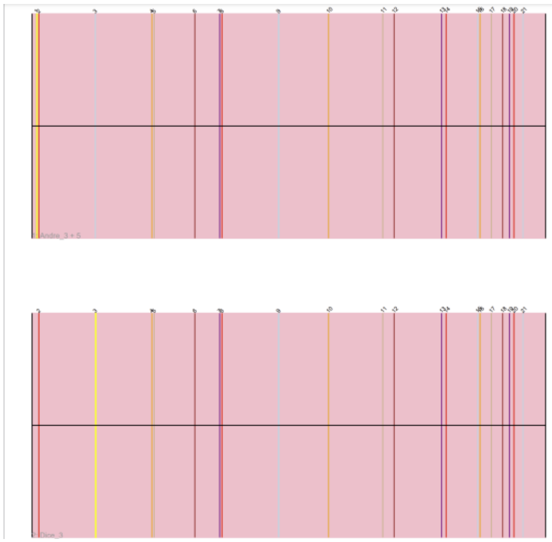
### 1. Starterator

This pham is comprised of all draft annotations. There are no annotations to summarize

Start 1:

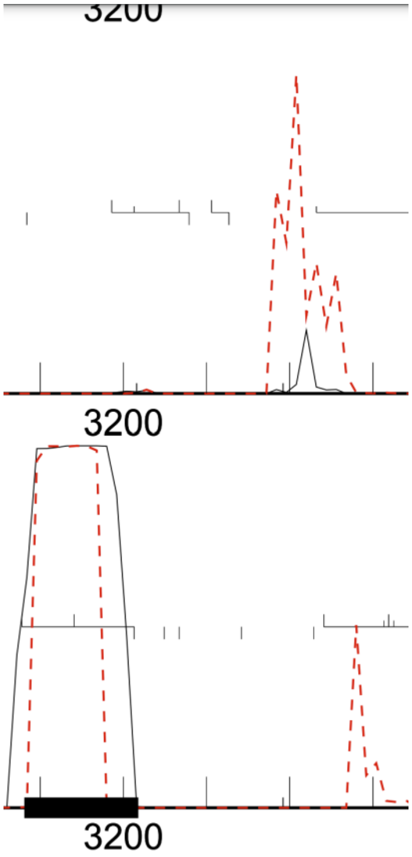
- Found in 6 of 7 ( 85.7% ) of genes in pham
- No Manual Annotations of this start.
- Called 100.0% of time when present

Eugenia is in Track 1. Note that Starterator calls Eugenia_3, likely due to overlap or discrepancy.



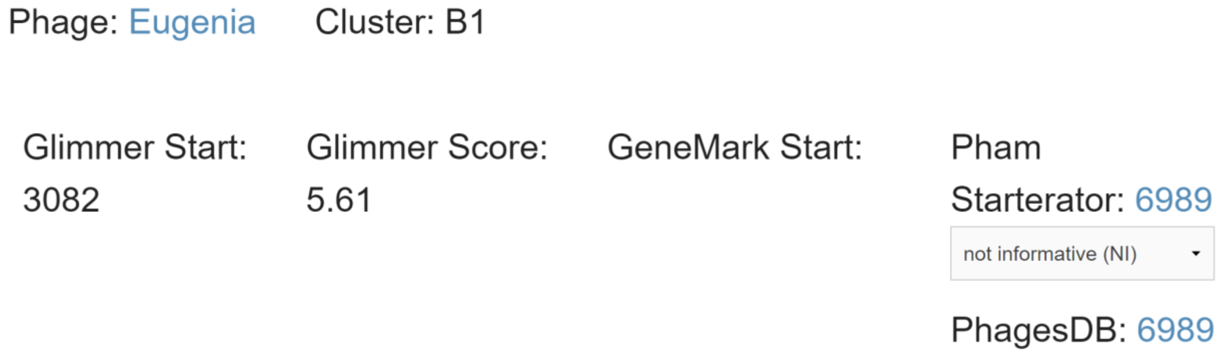
### 2. GeneMark coding potential

NO GeneMark coding potential



### 3. Glimmer and GeneMark agreement

No GeneMark Start, so NO agreement


Phage: Eugenia Cluster: B1

| Glimmer Start: | Glimmer Score: | GeneMark Start: | Pham |
|---|---|---|---|
| 3082 | 5.61 | | Starterator: 6989<br>not informative (NI)<br>PhagesDB: 6989 |

## 4. Longest open reading frame (ORF) without excessive gap

YES

| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 3082 | 2399 | 684 | 282 | 9 | 1.007 | -6.857 | TRUE | ATG | Yes | ☑ |
| Reverse | 3079 | 2399 | 681 | 285 | 12 | 1.007 | -6.918 |  | GTG |  | ☐ |

Start: 3082 (Note that this gene is in reverse)

Gap: 282

Z-score: 1.007

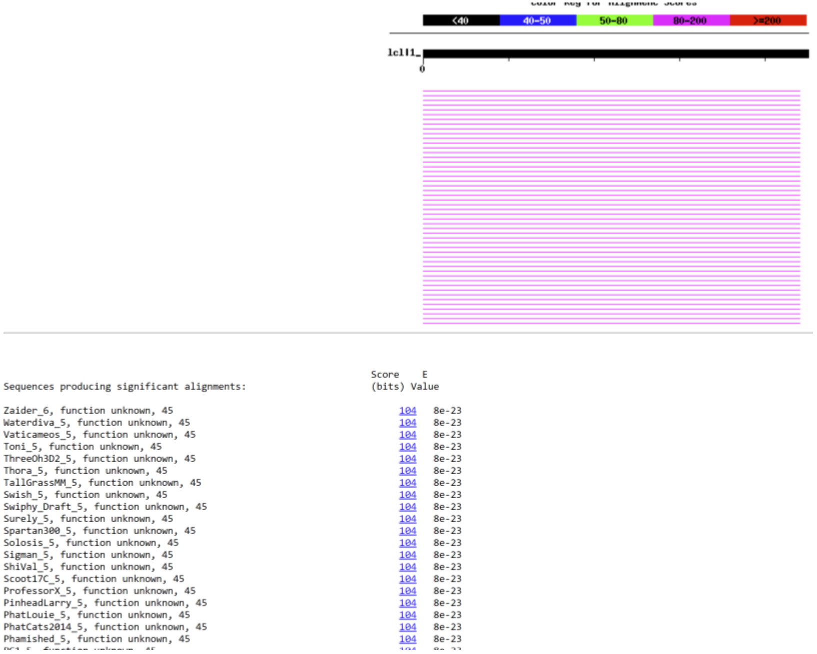
Final Score: -6.857

## 5. Function. If no functional prediction is present, write “Hypothetical protein”.

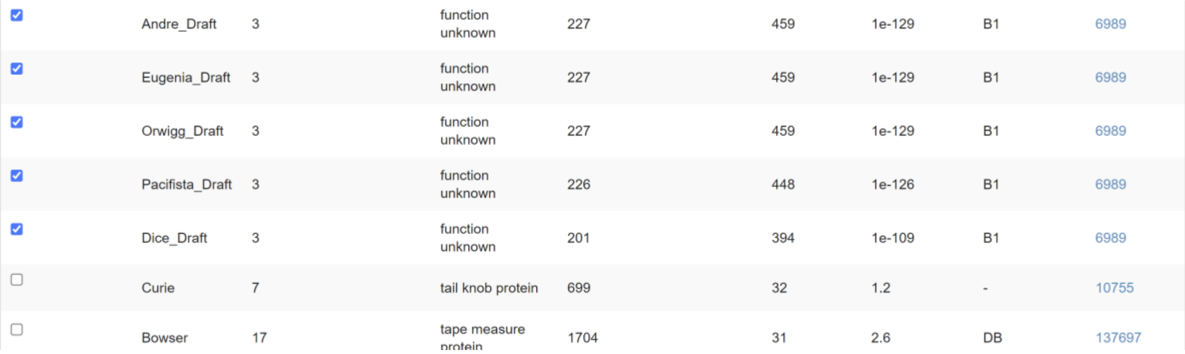
Hypothetical Protein—No Function.

## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST

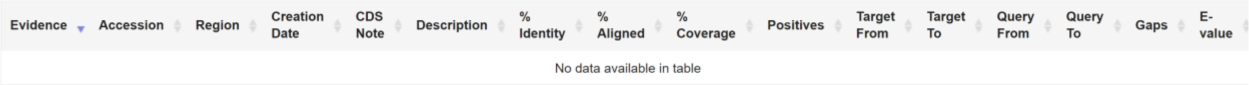


```
                                                             Score    E
Sequences producing significant alignments:                 (bits) Value

Zaider_6, function unknown, 45                                 104   8e-23
Waterdiva_5, function unknown, 45                              104   8e-23
Vaticameos_5, function unknown, 45                             104   8e-23
Toni_5, function unknown, 45                                   104   8e-23
ThreeOh3D2_5, function unknown, 45                             104   8e-23
Thora_5, function unknown, 45                                  104   8e-23
TallGrassMM_5, function unknown, 45                            104   8e-23
Swish_5, function unknown, 45                                  104   8e-23
Swiphy_Draft_5, function unknown, 45                           104   8e-23
Surely_5, function unknown, 45                                 104   8e-23
Spartan300_5, function unknown, 45                             104   8e-23
Solosis_5, function unknown, 45                                104   8e-23
Sigman_5, function unknown, 45                                 104   8e-23
ShiVal_5, function unknown, 45                                 104   8e-23
Scoot17C_5, function unknown, 45                               104   8e-23
ProfessorX_5, function unknown, 45                             104   8e-23
PinheadLarry_5, function unknown, 45                           104   8e-23
PhatLouie_5, function unknown, 45                              104   8e-23
PhatCats2014_5, function unknown, 45                           104   8e-23
Phamished_5, function unknown, 45                              104   8e-23
```



| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ☑ | Andre_Draft | 3 | function unknown | 227 | 459 | 1e-129 | B1 | 6989 |
| ☑ | Eugenia_Draft | 3 | function unknown | 227 | 459 | 1e-129 | B1 | 6989 |
| ☑ | Orwigg_Draft | 3 | function unknown | 227 | 459 | 1e-129 | B1 | 6989 |
| ☑ | Pacifista_Draft | 3 | function unknown | 226 | 448 | 1e-126 | B1 | 6989 |
| ☑ | Dice_Draft | 3 | function unknown | 201 | 394 | 1e-109 | B1 | 6989 |
| ☐ | Curie | 7 | tail knob protein | 699 | 32 | 1.2 | - | 10755 |
| ☐ | Bowser | 17 | tape measure protein | 1704 | 31 | 2.6 | DB | 137697 |

## b. SIF: NCBI BLAST

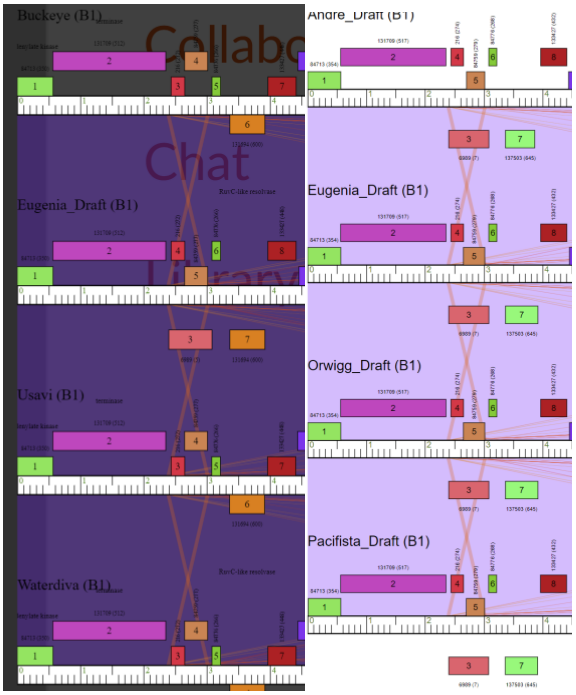


| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| No data available in table | | | | | | | | | | | | | | | |

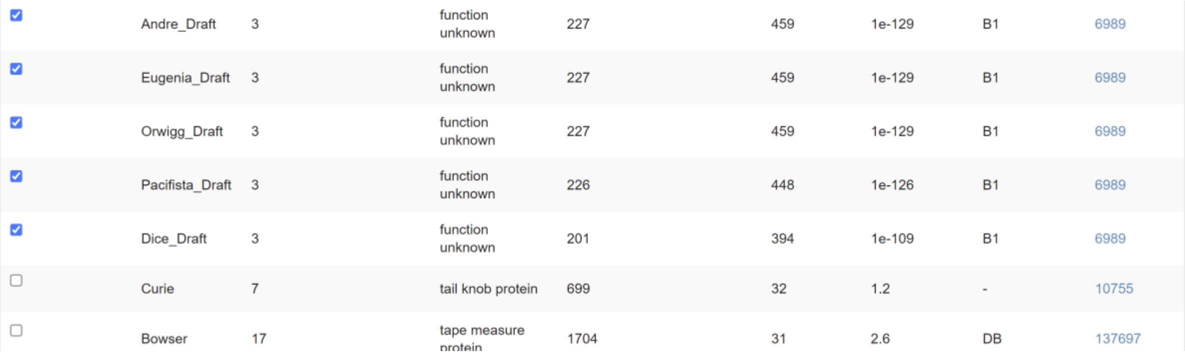
## c. SIF: HHPred

N/A

## d. SIF: Synteny-Phamerator (three genomes)



Note that there is NO synteny between the completed phages, but synteny between drafts of phages.



| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ☑ | Andre_Draft | 3 | function unknown | 227 | 459 | 1e-129 | B1 | 6989 |
| ☑ | Eugenia_Draft | 3 | function unknown | 227 | 459 | 1e-129 | B1 | 6989 |
| ☑ | Orwigg_Draft | 3 | function unknown | 227 | 459 | 1e-129 | B1 | 6989 |
| ☑ | Pacifista_Draft | 3 | function unknown | 226 | 448 | 1e-126 | B1 | 6989 |
| ☑ | Dice_Draft | 3 | function unknown | 201 | 394 | 1e-109 | B1 | 6989 |
| ☐ | Curie | 7 | tail knob protein | 699 | 32 | 1.2 | - | 10755 |
| ☐ | Bowser | 17 | tape measure protein | 1704 | 31 | 2.6 | DB | 137697 |

# 7. Any other important information.

Most matched phages were compromised of all drafts. Gene 6 will likely be deleted from the genome.

**CURATOR NAME: ISABELLA LIMA**

**GENE NAME: EUGENIA_DRAFT_6**

**DNA MASTER NOTES:**

**START POSITION EVALUATION (IN ORDER)**:

1. Starterator

2. GeneMark coding potential

3. Glimmer and GeneMark agreement

4. Longest open reading frame (ORF) without excessive gap

5. If functional prediction, start include full length protein. If no functional prediction is present, write “Hypothetical protein”.

6. Start site similar to others in GenBank and PhagesDB

7. Associated ribosome binding site (RBS)

8. Function (F)

## 9. Supporting Information for Function (SIF)

### 9. a. SIF-BLAST

### 9. b. SIF-HHPred

### 9. c. SIF-Synteny

## 10. Any other important information.

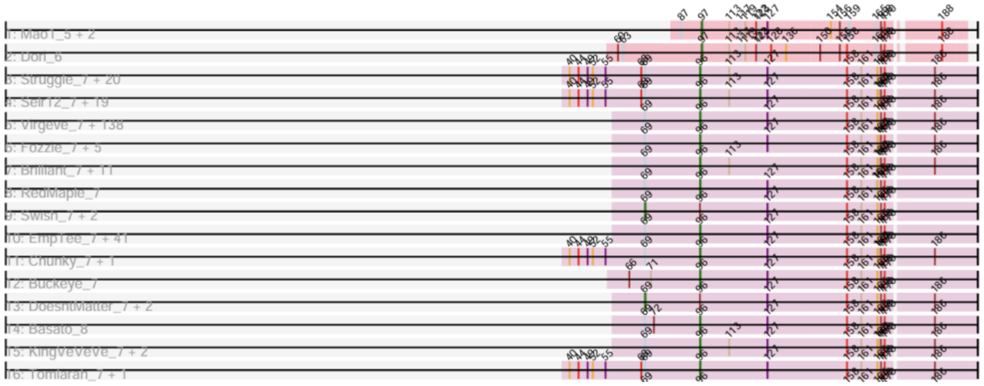
## CURATOR NAME: OLIVIA SIDOTI

## GENE NAME: EUGENIA GENE 8

## DNA MASTER NOTES: left it blank for now

## START POSITION EVALUATION (IN ORDER):

### 1. Starterator



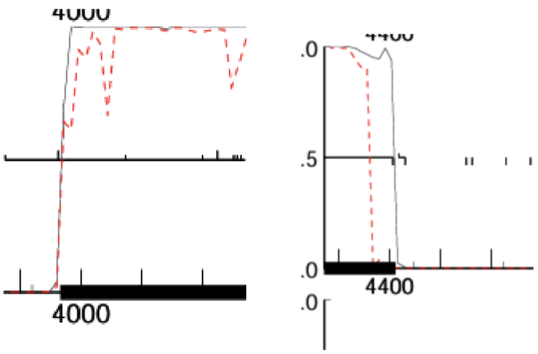
Pham number 133427 has 432 members, 47 are drafts.

The start number called the most often in the published annotations is 96, it was called in 237 of the 385 non-draft genes in the pham

Start 96:

- Found in 274 of 432 ( 63.4% ) of genes in pham
- Manual Annotations of this start: 237 of 385
- Called 95.3% of time when present

## 2. GeneMark coding potential



The ORF extends further out to the left around 3900, but the coding potential does not start until around 4000.

## 3. Glimmer and GeneMark agreement


Phage: Eugenia Cluster: B1

Glimmer Start: 3966

Glimmer Score: 14.25

GeneMark Start: 3966

Pham Starterator: 133427

PhagesDB: 133427

Yes, Glimmer and GeneMark are in agreement that the start codon is at nucleotide position 3966.

Z-Score: 2.138 (close to 2)

Final Score: -5.157

## 4. Longest open reading frame (ORF) without excessive gap

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 3966 | 4412 | 447 | 46 | 14 | 2.138 | -5.157 | ATG | Yes | ☑ |

No, this gene does not have the LORF without an excessive gap.

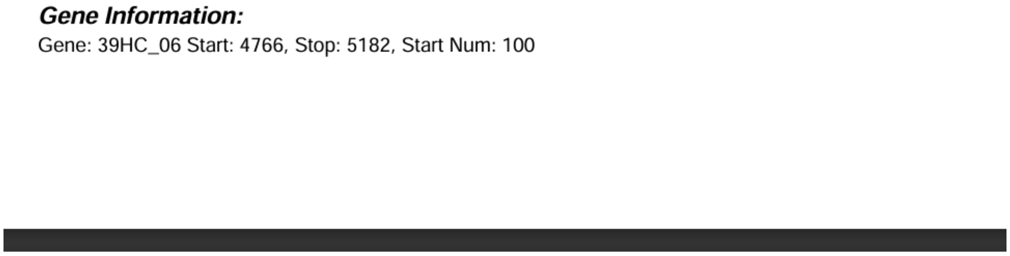
Gap: 46

5. If functional prediction, start include full length protein. If no functional prediction is present, write “Hypothetical protein”.

Hypothetical Protein

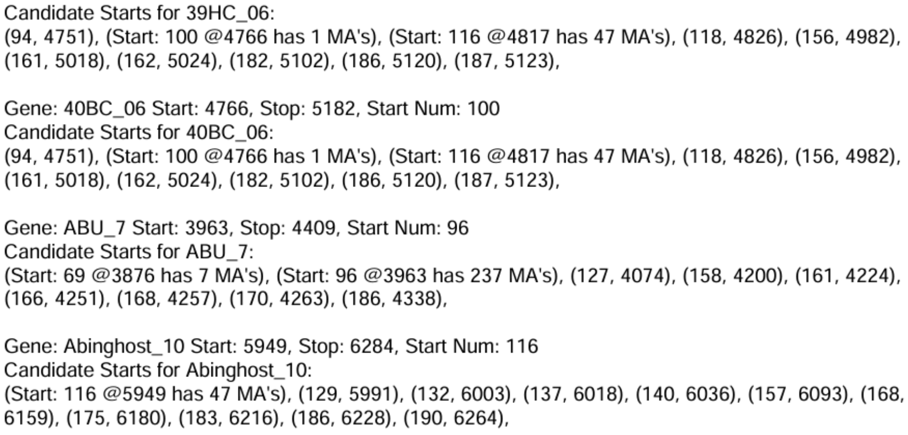
## 6. Start site similar to others in GenBank and PhagesDB

Yes



***Gene Information:***

Gene: 39HC_06 Start: 4766, Stop: 5182, Start Num: 100

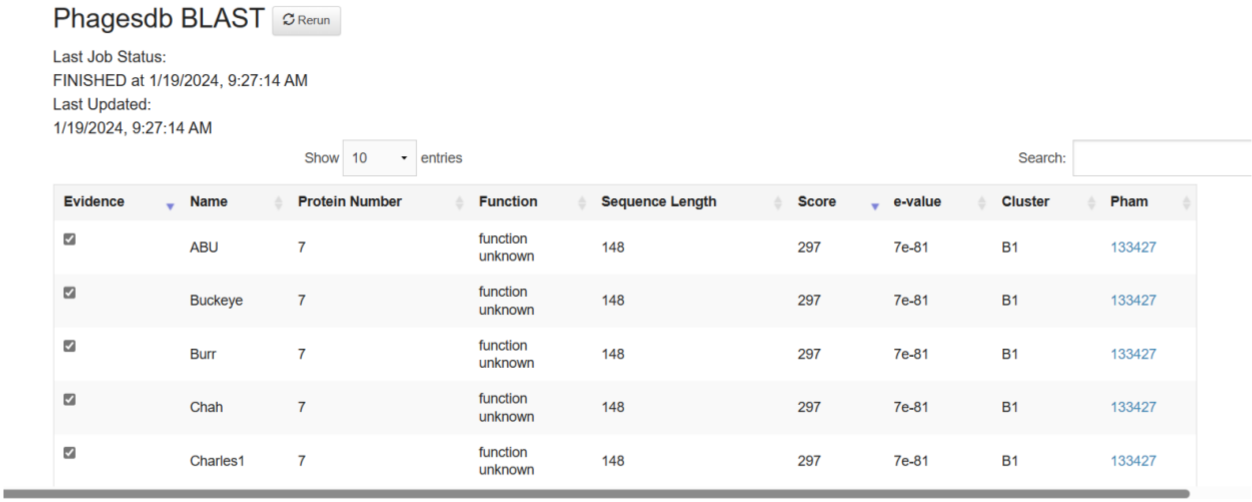


Candidate Starts for 39HC_06:
(94, 4751), (Start: 100 @4766 has 1 MA's), (Start: 116 @4817 has 47 MA's), (118, 4826), (156, 4982), (161, 5018), (162, 5024), (182, 5102), (186, 5120), (187, 5123),

Gene: 40BC_06 Start: 4766, Stop: 5182, Start Num: 100
Candidate Starts for 40BC_06:
(94, 4751), (Start: 100 @4766 has 1 MA's), (Start: 116 @4817 has 47 MA's), (118, 4826), (156, 4982), (161, 5018), (162, 5024), (182, 5102), (186, 5120), (187, 5123),

Gene: ABU_7 Start: 3963, Stop: 4409, Start Num: 96
Candidate Starts for ABU_7:
(Start: 69 @3876 has 7 MA's), (Start: 96 @3963 has 237 MA's), (127, 4074), (158, 4200), (161, 4224), (166, 4251), (168, 4257), (170, 4263), (186, 4338),

Gene: Abinghost_10 Start: 5949, Stop: 6284, Start Num: 116
Candidate Starts for Abinghost_10:
(Start: 116 @5949 has 47 MA's), (129, 5991), (132, 6003), (137, 6018), (140, 6036), (157, 6093), (168, 6159), (175, 6180), (183, 6216), (186, 6228), (190, 6264),



Phagesdb BLAST

Last Job Status:
FINISHED at 1/19/2024, 9:27:14 AM
Last Updated:
1/19/2024, 9:27:14 AM

| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | ABU | 7 | function unknown | 148 | 297 | 7e-81 | B1 | 133427 |
| ☑ | Buckeye | 7 | function unknown | 148 | 297 | 7e-81 | B1 | 133427 |
| ☑ | Burr | 7 | function unknown | 148 | 297 | 7e-81 | B1 | 133427 |
| ☑ | Chah | 7 | function unknown | 148 | 297 | 7e-81 | B1 | 133427 |
| ☑ | Charles1 | 7 | function unknown | 148 | 297 | 7e-81 | B1 | 133427 |

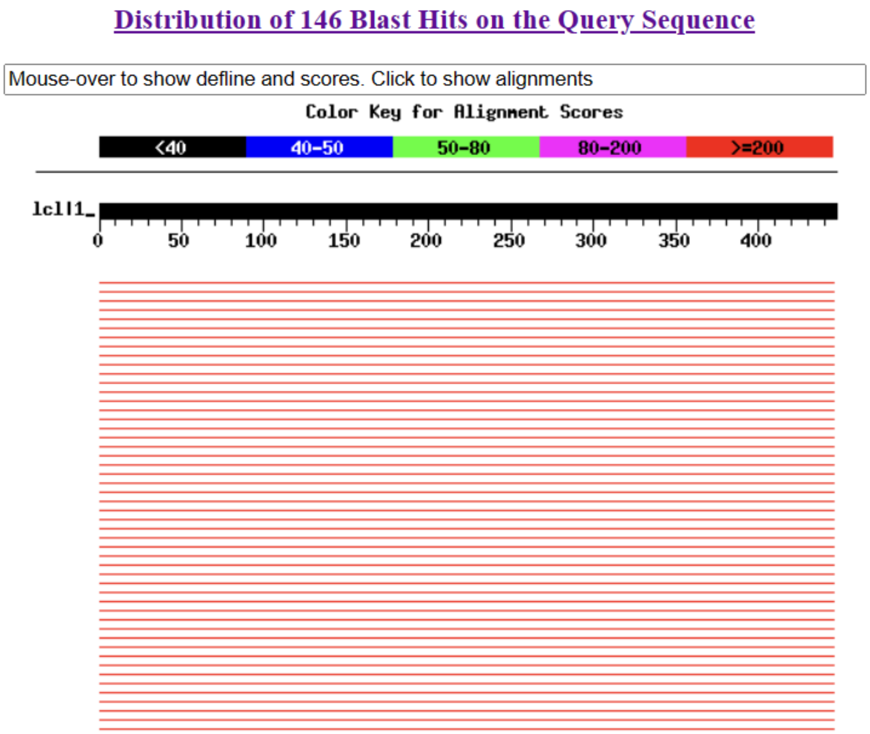
## 8. Function (F)

Unknown Function

## 9. Supporting Information for Function (SIF)

### 9. a. SIF-BLAST

Nucleotide Blast: The nucleotide sequence of Eugenia gene 8 had an alignment score with other genes at >=200. This indicates that this exact gene is present in other mycobacterium phage.

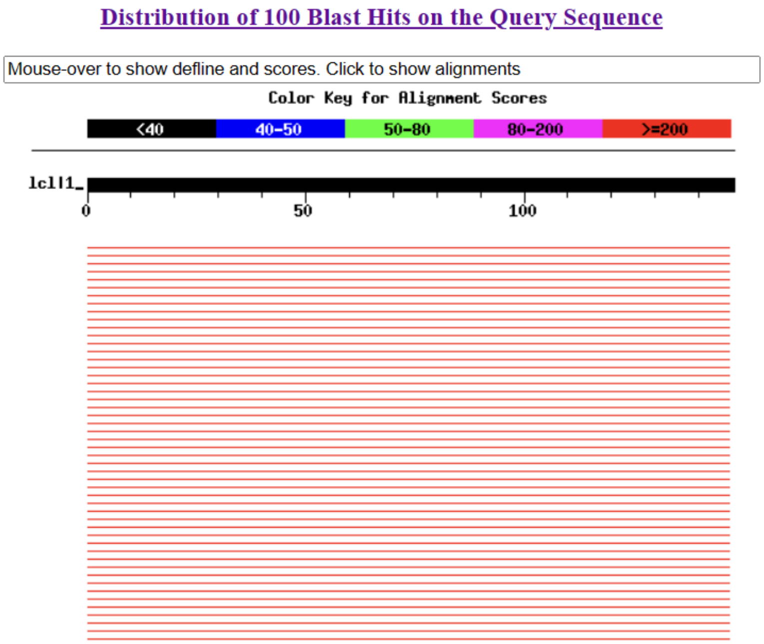


Distribution of 146 Blast Hits on the Query Sequence

Mouse-over to show defline and scores. Click to show alignments

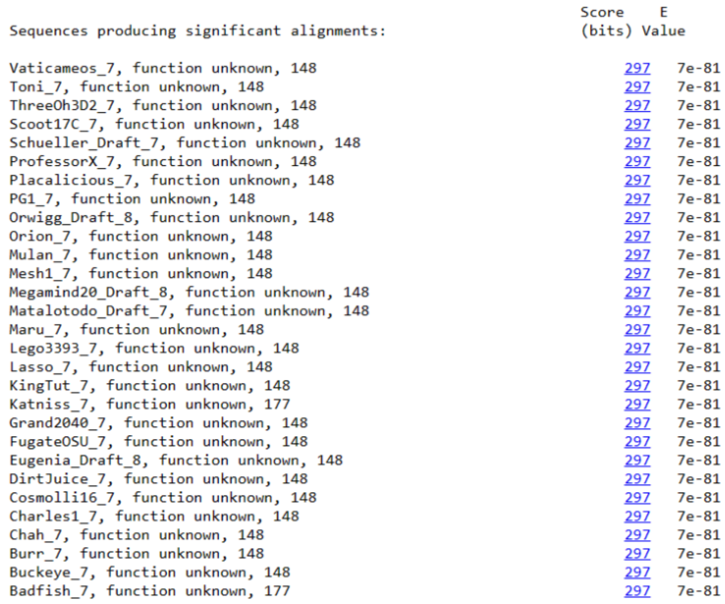
```
                                                                  Score     E
Sequences producing significant alignments:                       (bits)  Value

Mycobacterium phage Toni complete sequence, 68784 bp, circularly...   886    0.0
Mycobacterium phage Scoot17C                                          886    0.0
Mycobacterium phage Mulan complete sequence, 68521 bp, circularl...   886    0.0
Mycobacterium phage Matalotodo complete sequence, 69323 bp, circ...   886    0.0
Mycobacterium phage KingTut complete sequence, 64827 bp, circula...   886    0.0
Mycobacterium phage Katniss complete sequence, 68447 bp, circula...   886    0.0
Mycobacterium phage FugateOSU complete sequence, 69476 bp, circu...   886    0.0
Mycobacterium phage Eugenia complete sequence, 69139 bp, circula...   886    0.0
Mycobacterium phage Buckeye complete sequence, 69174 bp, circula...   886    0.0
Mycobacterium phage Badfish complete sequence, 69030 bp, circula...   886    0.0
Mycobacterium phage Swish complete sequence, 68735 bp, circularl...   878    0.0
Mycobacterium phage Phatcats2014 complete sequence, exactly 6900...   878    0.0
Mycobacterium phage EmpTee complete sequence, 68428 bp, circular...   878    0.0
Mycobacterium phage RedMaple complete sequence, 68985 bp, circul...   862    0.0
Mycobacterium phage Burr complete sequence, 68721 bp, circularly...   862    0.0
Mycobacterium phage Maru complete sequence, 68684 bp, circularly...   854    0.0
Mycobacterium phage Grand2040 complete sequence, 68591 bp, circu...   854    0.0
Mycobacterium phage Cosmolli16 complete sequence, 68978 bp, circ...   854    0.0
Mycobacterium phage Emiris complete sequence, 69027 bp, circular...   846    0.0
Mycobacterium phage MRabcd complete sequence, 68791 bp, circular...   839    0.0
Mycobacterium phage Mana complete sequence, 68479 bp, circularly...   839    0.0
Mycobacterium phage Vaticameos complete sequence, 66887 bp, circ...   831    0.0
Mycobacterium phage ThreeOh3D2                                        831    0.0
Mycobacterium phage Roy17 complete sequence, 68056 bp, circularl...   831    0.0
Mycobacterium phage ProfessorX complete sequence, 68086 bp, circ...   831    0.0
```

Protein Blast: The amino acid sequence aligned with other genome's gene in the range of >=200. All of the sequences that produced a significant alignment had no known function assigned to the gene.


Distribution of 100 Blast Hits on the Query Sequence

Mouse-over to show defline and scores. Click to show alignments

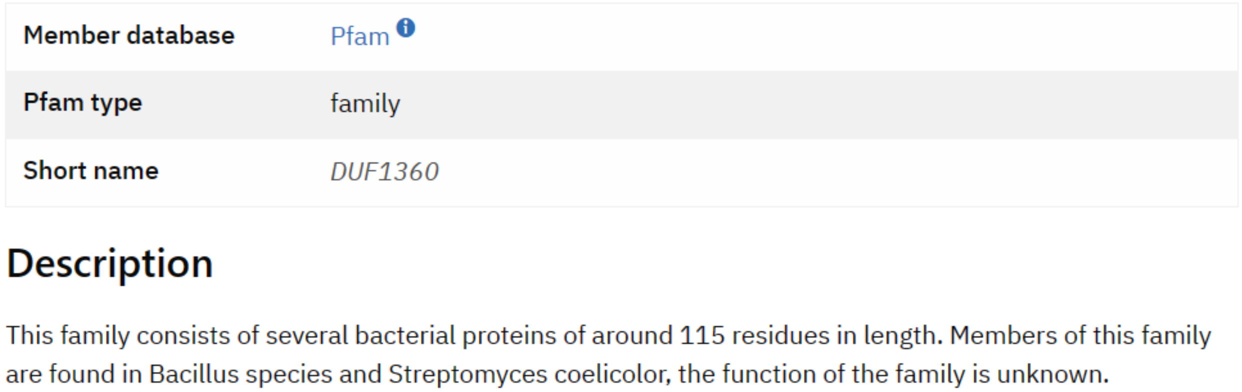
Color Key for Alignment Scores



```
                                                                   Score     E
Sequences producing significant alignments:                        (bits)  Value

Vaticameos_7, function unknown, 148                                  297    7e-81
Toni_7, function unknown, 148                                        297    7e-81
ThreeOh3D2_7, function unknown, 148                                  297    7e-81
Scoot17C_7, function unknown, 148                                    297    7e-81
Schueller_Draft_7, function unknown, 148                             297    7e-81
ProfessorX_7, function unknown, 148                                  297    7e-81
Placalicious_7, function unknown, 148                                297    7e-81
PG1_7, function unknown, 148                                         297    7e-81
Orwigg_Draft_8, function unknown, 148                                297    7e-81
Orion_7, function unknown, 148                                       297    7e-81
Mulan_7, function unknown, 148                                       297    7e-81
Mesh1_7, function unknown, 148                                       297    7e-81
Megamind20_Draft_8, function unknown, 148                            297    7e-81
Matalotodo_Draft_7, function unknown, 148                            297    7e-81
Maru_7, function unknown, 148                                        297    7e-81
Lego3393_7, function unknown, 148                                    297    7e-81
Lasso_7, function unknown, 148                                       297    7e-81
KingTut_7, function unknown, 148                                     297    7e-81
Katniss_7, function unknown, 177                                     297    7e-81
Grand2040_7, function unknown, 148                                   297    7e-81
FugateOSU_7, function unknown, 148                                   297    7e-81
Eugenia_Draft_8, function unknown, 148                               297    7e-81
DirtJuice_7, function unknown, 148                                   297    7e-81
Cosmolli16_7, function unknown, 148                                  297    7e-81
Charles1_7, function unknown, 148                                    297    7e-81
Chah_7, function unknown, 148                                        297    7e-81
Burr_7, function unknown, 148                                        297    7e-81
Buckeye_7, function unknown, 148                                     297    7e-81
Badfish_7, function unknown, 177                                     297    7e-81
```

## 9. b. SIF-HHPred

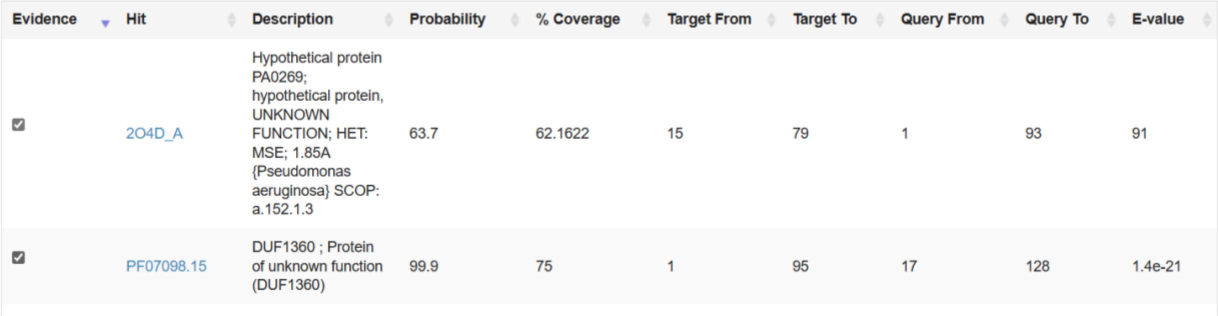
HHPred gave results for protein structures that are similar to that of the Eugenia gene-8.



| | |
|---|---|
| **Member database** | Pfam |
| **Pfam type** | family |
| **Short name** | *DUF1360* |

### Description

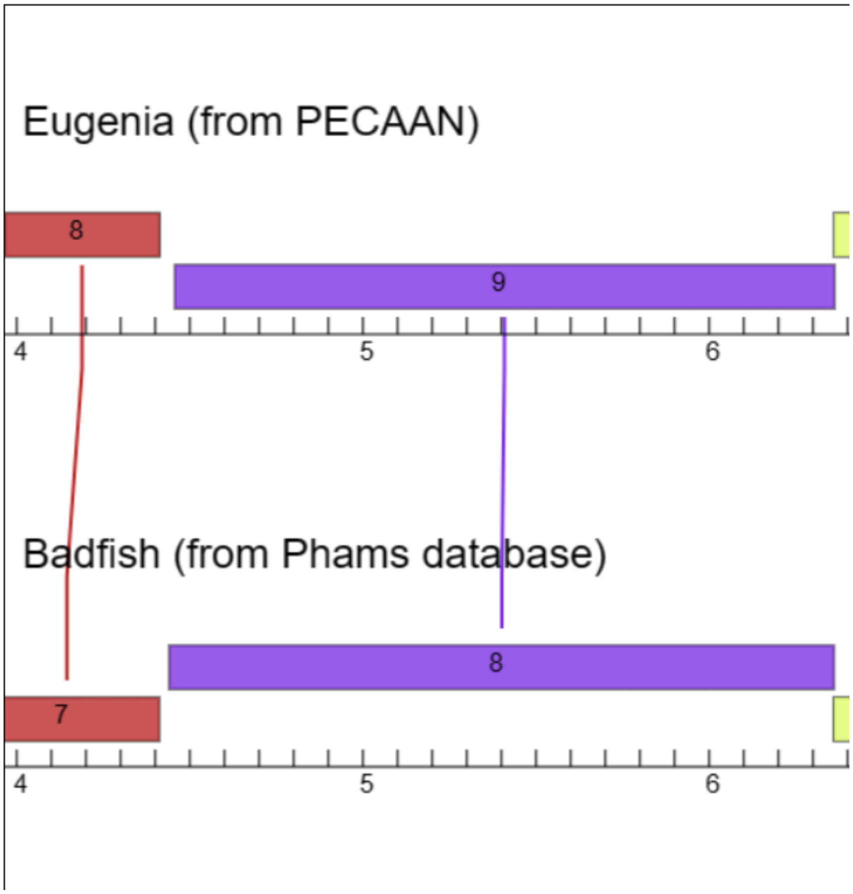
This family consists of several bacterial proteins of around 115 residues in length. Members of this family are found in Bacillus species and Streptomyces coelicolor, the function of the family is unknown.

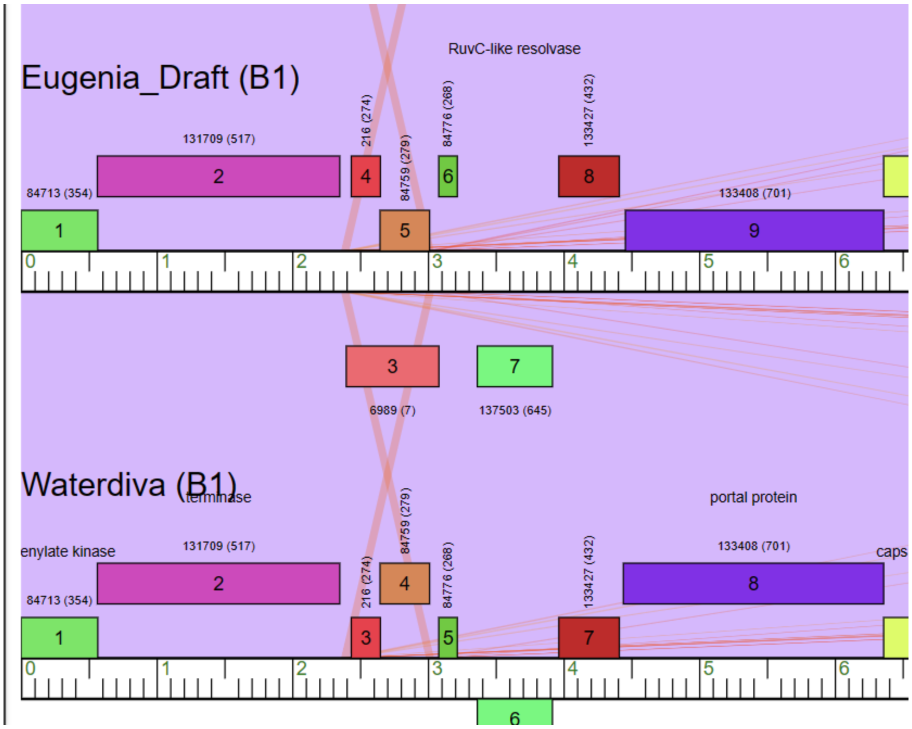


| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | 2O4D_A | Hypothetical protein PA0269; hypothetical protein, UNKNOWN FUNCTION; HET: MSE; 1.85A {Pseudomonas aeruginosa} SCOP: a.152.1.3 | 63.7 | 62.1622 | 15 | 79 | 1 | 93 | 91 |
| ☑ | PF07098.15 | DUF1360 ; Protein of unknown function (DUF1360) | 99.9 | 75 | 1 | 95 | 17 | 128 | 1.4e-21 |

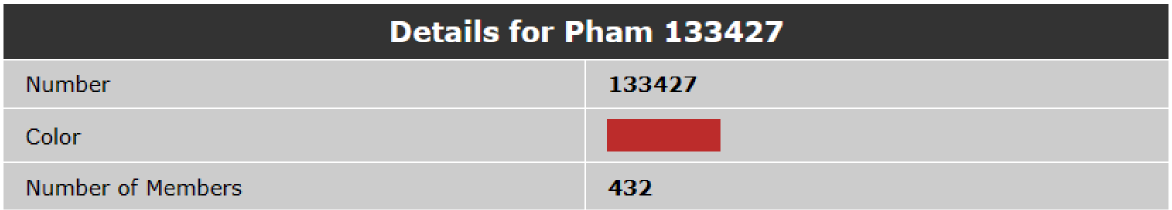
## 9. c. SIF-Synteny

Phamerator Version: 547





10. Any other important information.



| Details for Pham 133427 | |
|---|---|
| Number | **133427** |
| Color | 🟥 |
| Number of Members | **432** |

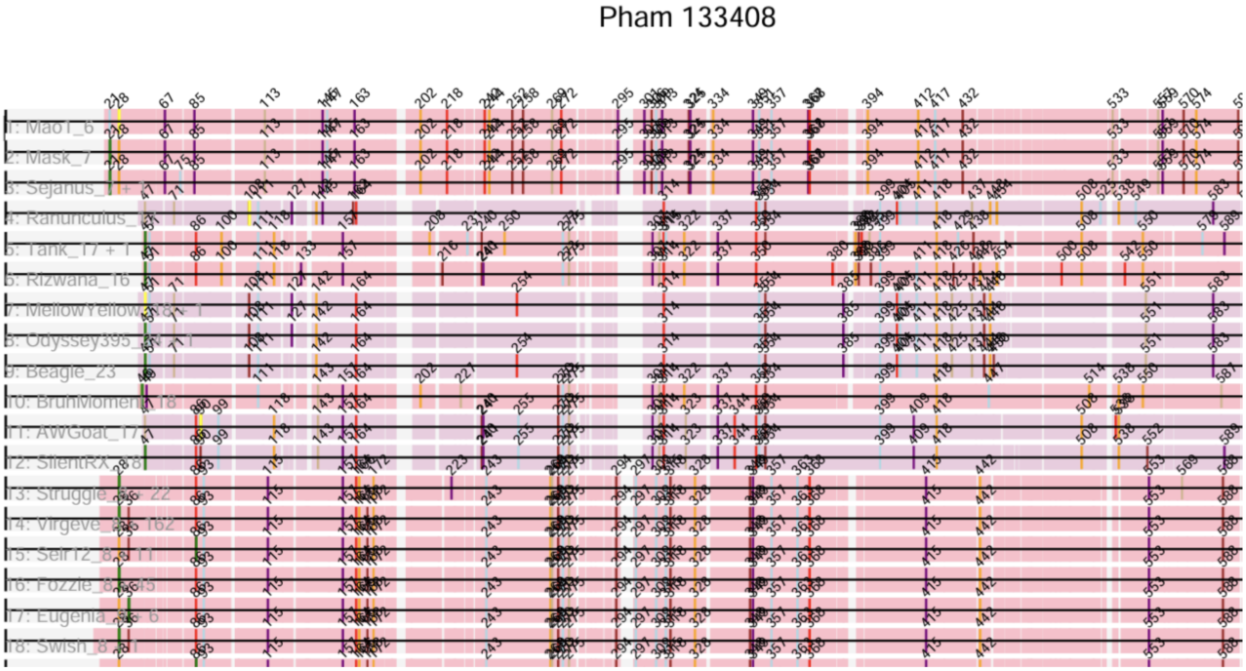
## CURATOR NAME:

## GENE NAME:

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

1. Starterator



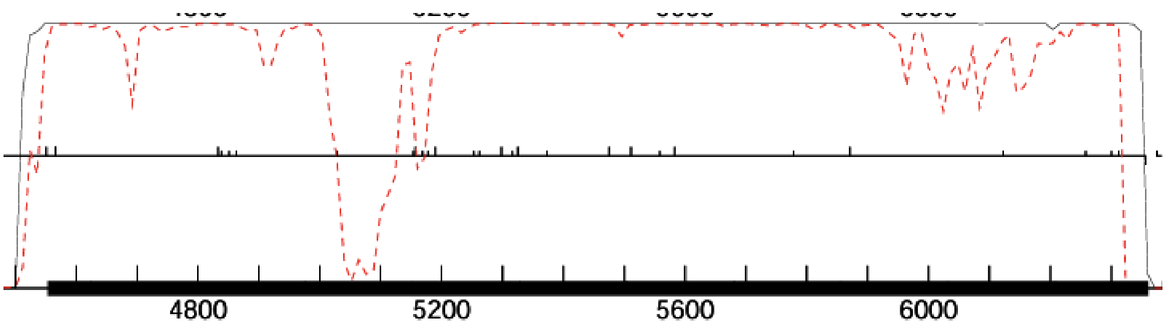
Pham number 133408 has 701 members, 100 are drafts.

The start number called the most often in the published annotations is 28, it was called in 371 of the 601 non-draft genes in the pham.

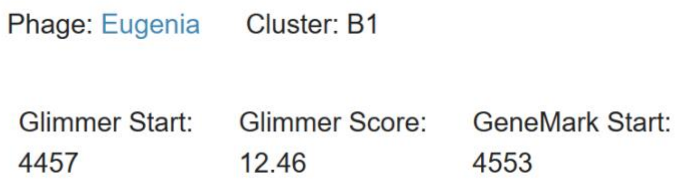
Start 28:

- Found in 447 of 701 ( 63.8% ) of genes in pham
- Manual Annotations of this start: 371 of 601
- Called 89.5% of time when present

## 2. GeneMark coding potential



## 3. Glimmer and GeneMark agreement


Phage: Eugenia    Cluster: B1

Glimmer Start: 4457
Glimmer Score: 12.46
GeneMark Start: 4553

## 4. Longest open reading frame (ORF) without excessive gap

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 4457 | 6361 | 1905 | 44 | 13 | 1.612 | -5.914 | TTG | Yes | ☑ |

No

## 5. Function. If no functional prediction is present, write “Hypothetical protein”.

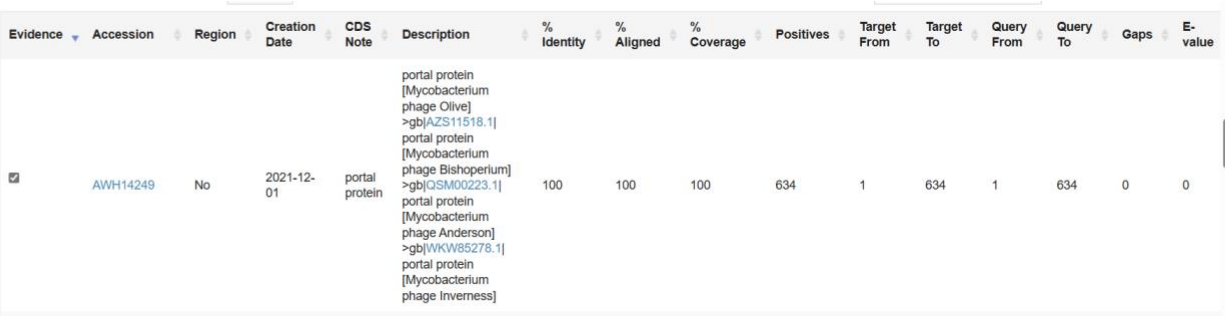
Portal Protein

## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST

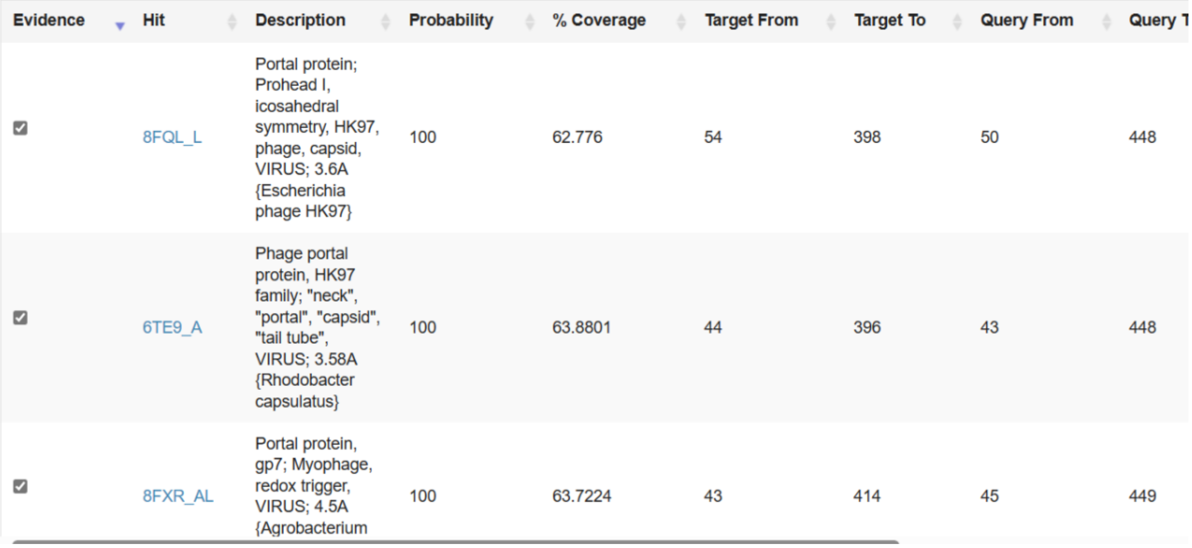
```
                                                                Score    E
Sequences producing significant alignments:                     (bits) Value

Orwigg_Draft_9, function unknown, 634                             1243    0.0
Olive_8, portal protein, 634                                      1243    0.0
Megamind20_Draft_9, function unknown, 634                         1243    0.0
Inverness_8, portal protein, 634                                  1243    0.0
Eugenia_Draft_9, function unknown, 634                            1243    0.0
Bishoperium_8, portal protein, 634                                1243    0.0
Anderson_8, portal protein, 634                                   1243    0.0
Zonia_8, portal protein, 639                                      1242    0.0
Zelda_8, portal protein, 639                                      1242    0.0
Zaider_9, portal protein, 639                                     1242    0.0
YouGoGlencoco_8, portal protein, 639                              1242    0.0
Yoshand_8, portal protein, 639                                    1242    0.0
Waterdiva_8, portal protein, 639                                  1242    0.0
Vortex_8, portal protein, 639                                     1242    0.0
Vivaldi_8, portal protein, 639                                    1242    0.0
Vista_8, portal protein, 639                                      1242    0.0
Virapocalypse_8, portal protein, 639                              1242    0.0
Vaticameos_8, portal protein, 639                                 1242    0.0
Usavi_8, portal protein, 639                                      1242    0.0
UncleHowie_8, portal protein, 639                                 1242    0.0
UAch1_8, portal protein, 639                                      1242    0.0
True_8, portal protein, 639                                       1242    0.0
Toni_8, portal protein, 639                                       1242    0.0
Tomlarah_8, portal protein, 639                                   1242    0.0
TomBombadil_8, portal protein, 639                                1242    0.0
ThreeOh3D2_8, portal protein, 639                                 1242    0.0
Thora_8, portal protein, 639                                      1242    0.0
Swish_8, portal protein, 639                                      1242    0.0
Surely_8, portal protein, 639                                     1242    0.0
Squiggle_8, portal protein, 639                                   1242    0.0
Squid_8, portal protein, 639                                      1242    0.0
Spartan300_8, portal protein, 639                                 1242    0.0
Soile_8, portal protein, 639                                      1242    0.0
```

b. SIF: NCBI BLAST

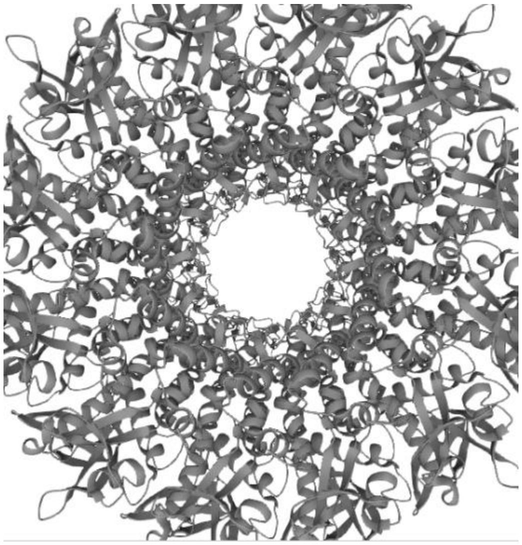


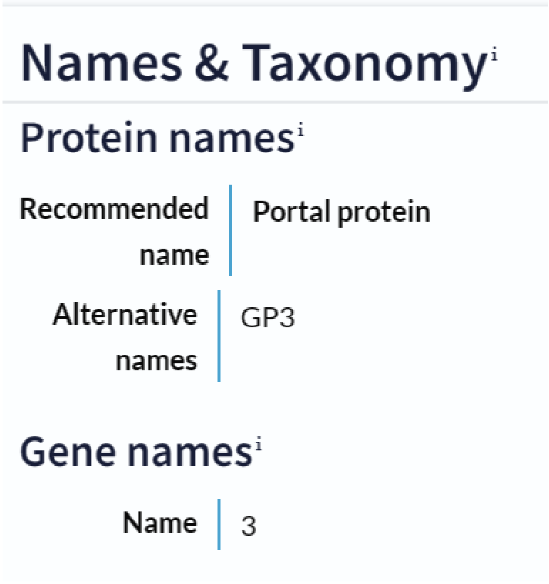
| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | AWH14249 | No | 2021-12-01 | portal protein | portal protein [Mycobacterium phage Olive] >gb\|AZS11518.1\| portal protein [Mycobacterium phage Bishoperium] >gb\|QSM00223.1\| portal protein [Mycobacterium phage Anderson] >gb\|WKW85278.1\| portal protein [Mycobacterium phage Inverness] | 100 | 100 | 100 | 634 | 1 | 634 | 1 | 634 | 0 | 0 |

## c. SIF: HHPred



| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query T |
|---|---|---|---|---|---|---|---|---|
| ☑ | 8FQL_L | Portal protein; Prohead I, icosahedral symmetry, HK97, phage, capsid, VIRUS; 3.6A {Escherichia phage HK97} | 100 | 62.776 | 54 | 398 | 50 | 448 |
| ☑ | 6TE9_A | Phage portal protein, HK97 family; "neck", "portal", "capsid", "tail tube", VIRUS; 3.58A {Rhodobacter capsulatus} | 100 | 63.8801 | 44 | 396 | 43 | 448 |
| ☑ | 8FXR_AL | Portal protein, gp7; Myophage, redox trigger, VIRUS; 4.5A {Agrobacterium | 100 | 63.7224 | 43 | 414 | 45 | 449 |





Names & Taxonomy

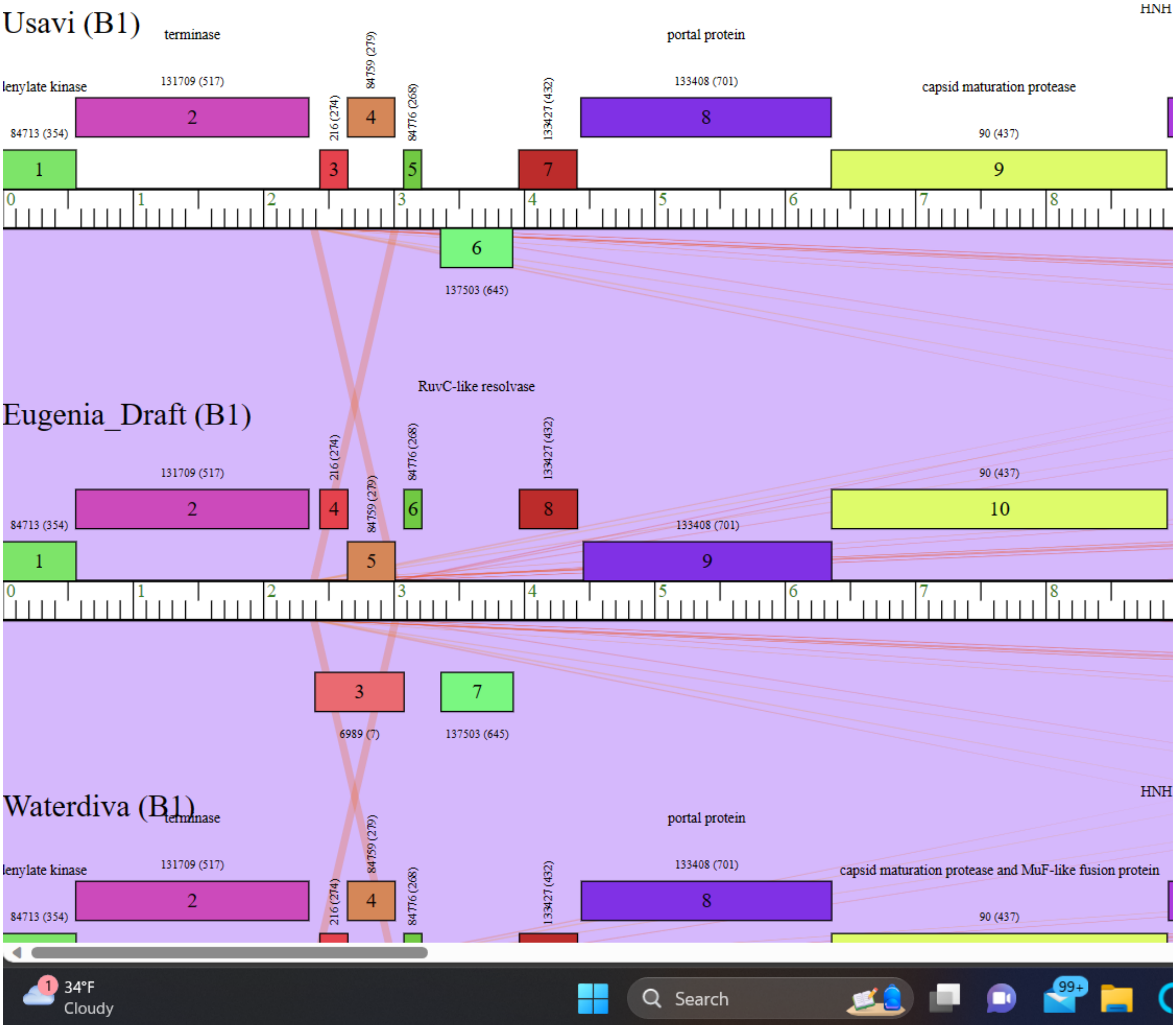
Protein names

| | |
|---|---|
| Recommended name | Portal protein |
| Alternative names | GP3 |

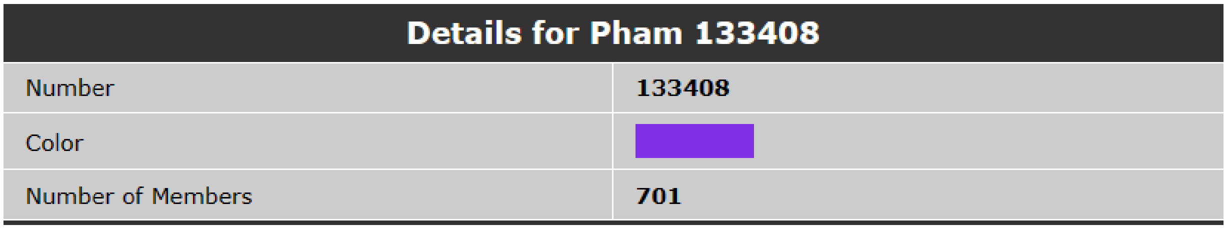
Gene names

| | |
|---|---|
| Name | 3 |

d. SIF: Synteny-Phamerator (three genomes) usavi waterdiva eugenia



7. Any other important information.



| Details for Pham 133408 | |
|---|---|
| Number | 133408 |
| Color | 🟪 |
| Number of Members | 701 |

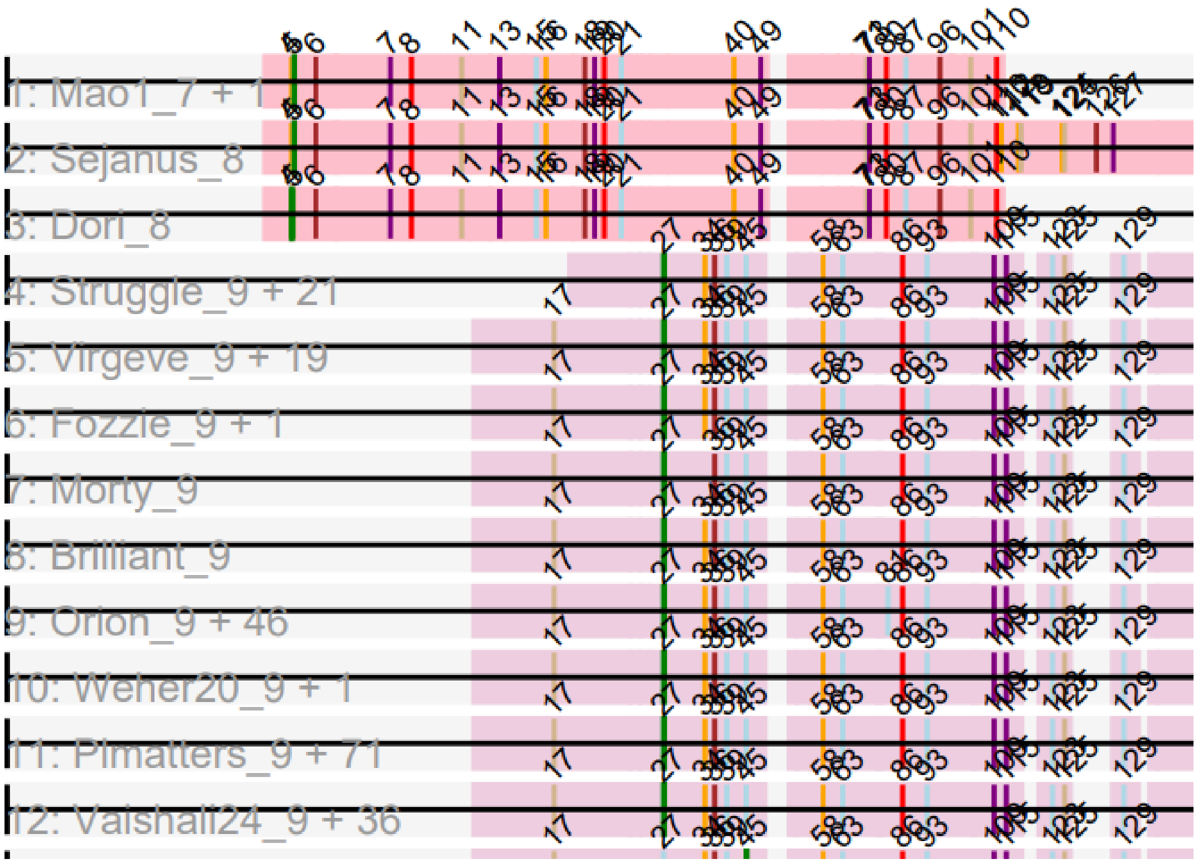
**CURATOR NAME: JOHN MOTTER**

**GENE NAME: EUGENIA GENE 10**

**DNA MASTER NOTES: N/A**

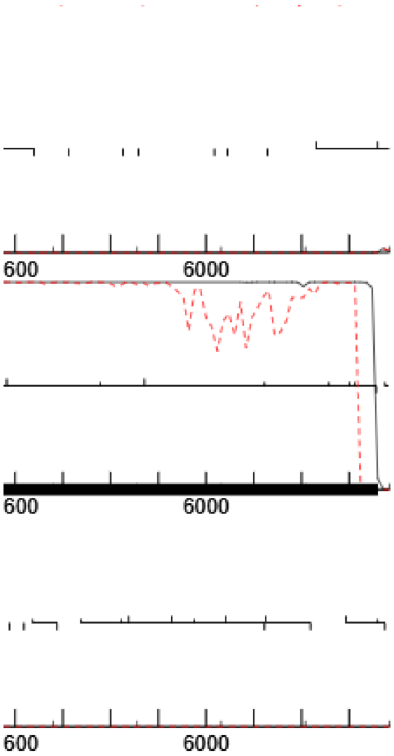
## START POSITION EVALUATION (IN ORDER):

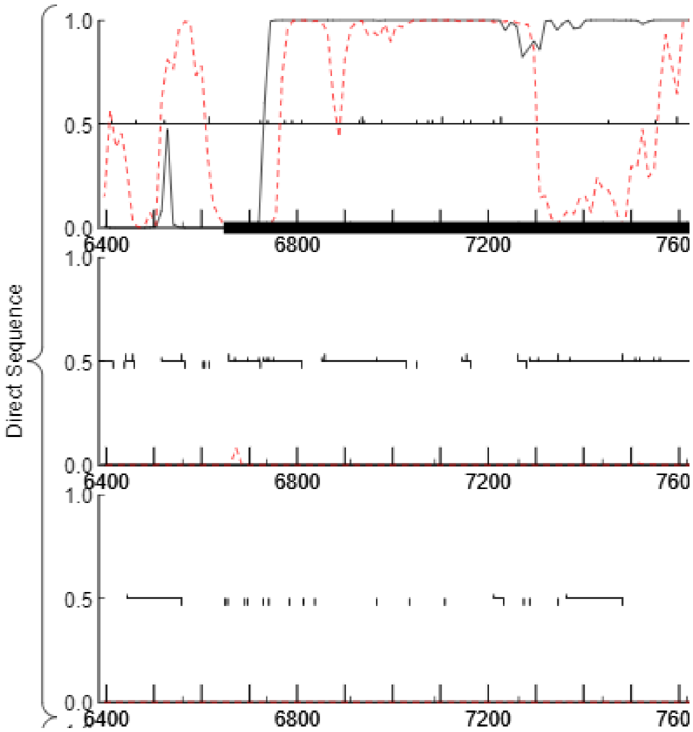
1. Starterator

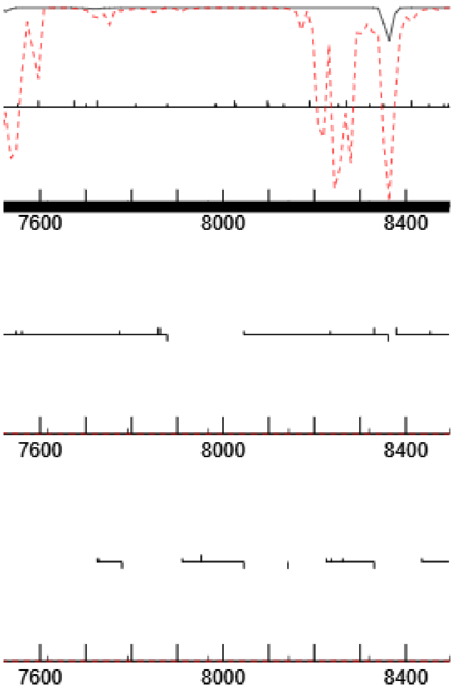


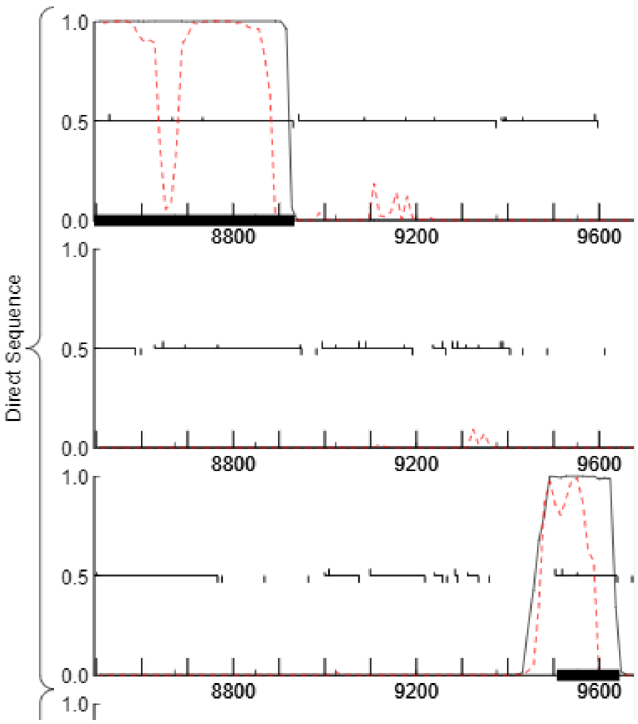
The start number called the most often in the published annotations is 27, it was called in 375 of the 386 non-draft genes in the pham.

## 2. GeneMark coding potential

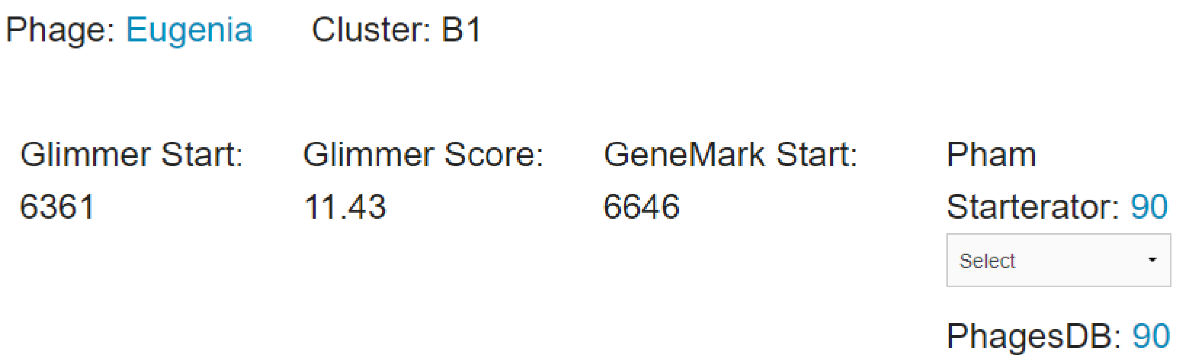








## 3. Glimmer and GeneMark agreement

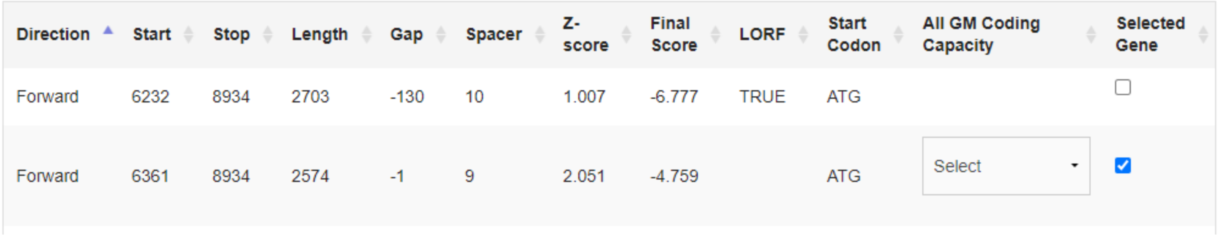

Phage: Eugenia Cluster: B1

Glimmer Start: 6361
Glimmer Score: 11.43
GeneMark Start: 6646
Pham Starterator: 90

PhagesDB: 90

Glimmer and GeneMark do not agree, and due to the presence of a large gap in the GeneMark start, it is most likely that the actual start is at 6361. GeneMark does show coding potential and an alternative start site at this position.

## 4. Longest open reading frame (ORF) without excessive gap



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 6232 | 8934 | 2703 | -130 | 10 | 1.007 | -6.777 | TRUE | ATG | | ☐ |
| Forward | 6361 | 8934 | 2574 | -1 | 9 | 2.051 | -4.759 | | ATG | Select | ☑ |

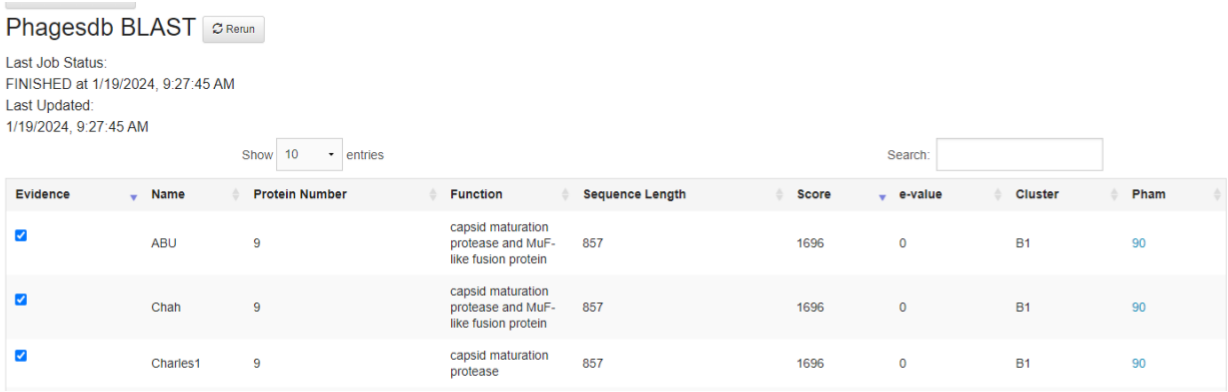
## 5. Function. Please follow this [Official SEAPHAGE Function List]()

If no functional prediction is present, write "Hypothetical protein".

Capsid maturation protease

## 6. Supporting Information for Function (SIF)

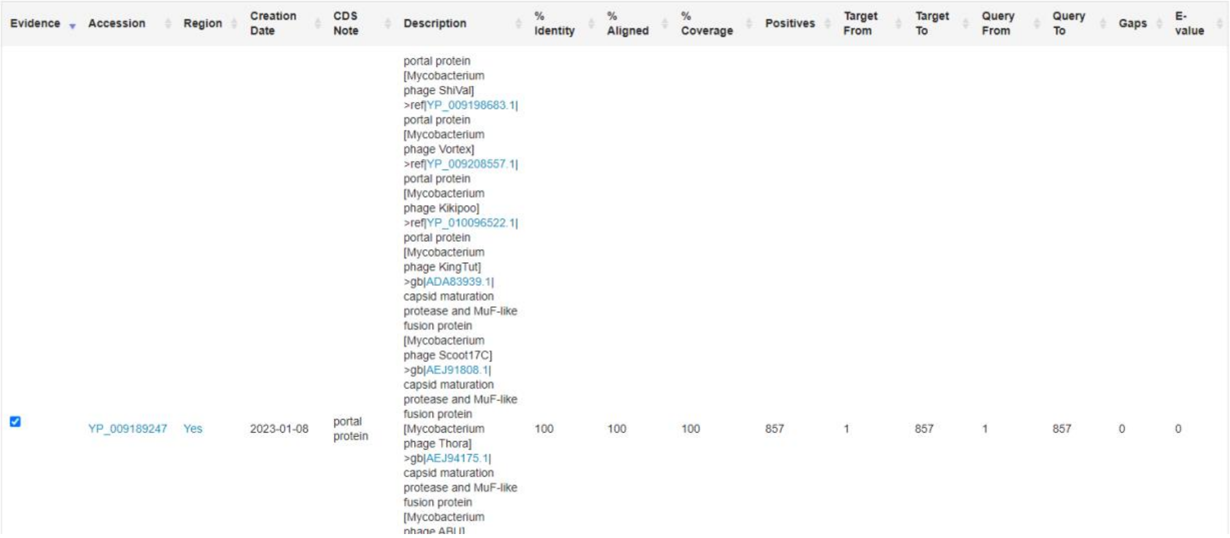
### a. SIF: PhageDb BLAST


Phagesdb BLAST

Last Job Status:
FINISHED at 1/19/2024, 9:27:45 AM
Last Updated:
1/19/2024, 9:27:45 AM

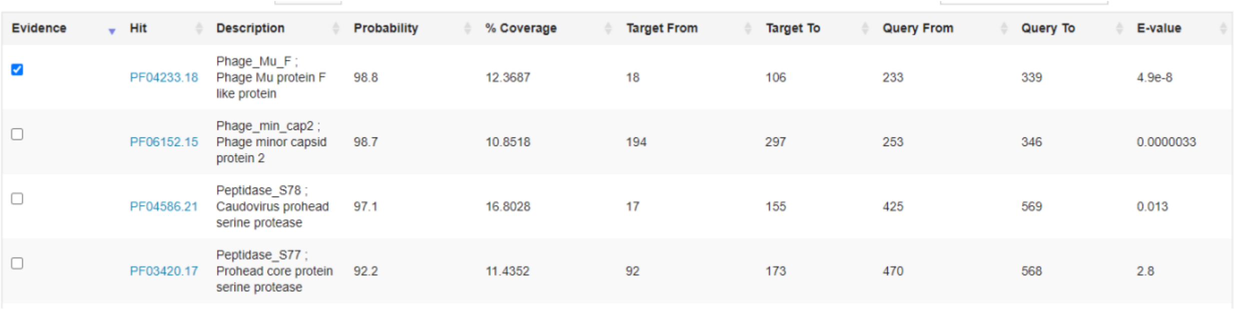
| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | ABU | 9 | capsid maturation protease and MuF-like fusion protein | 857 | 1696 | 0 | B1 | 90 |
| ☑ | Chah | 9 | capsid maturation protease and MuF-like fusion protein | 857 | 1696 | 0 | B1 | 90 |
| ☑ | Charles1 | 9 | capsid maturation protease | 857 | 1696 | 0 | B1 | 90 |

## b. SIF: NCBI BLAST



| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | YP_009189247 | Yes | 2023-01-08 | portal protein | portal protein [Mycobacterium phage ShiVal] >ref\|YP_009198683.1\| portal protein [Mycobacterium phage Vortex] >ref\|YP_009208557.1\| portal protein [Mycobacterium phage Kikipoo] >ref\|YP_010096522.1\| portal protein [Mycobacterium phage KingTut] >gb\|ADA83939.1\| capsid maturation protease and MuF-like fusion protein [Mycobacterium phage Scoot17C] >gb\|AEJ91808.1\| capsid maturation protease and MuF-like fusion protein [Mycobacterium phage Thora] >gb\|AEJ94175.1\| capsid maturation protease and MuF-like fusion protein [Mycobacterium phage ABU | 100 | 100 | 100 | 857 | 1 | 857 | 1 | 857 | 0 | 0 |

## c. SIF: HHPred

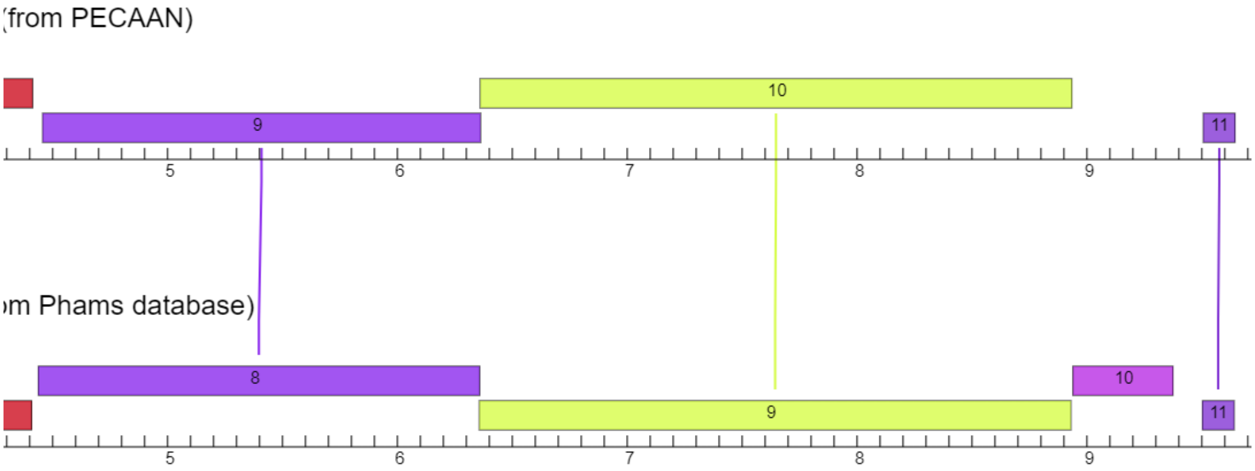


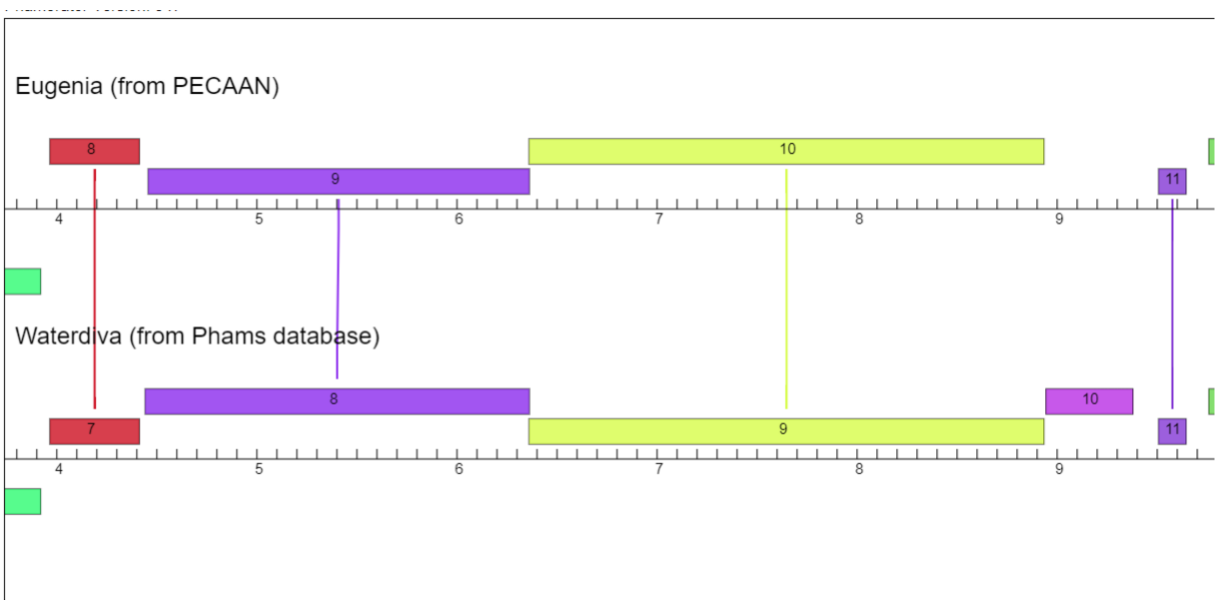
| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | PF04233.18 | Phage_Mu_F ; Phage Mu protein F like protein | 98.8 | 12.3687 | 18 | 106 | 233 | 339 | 4.9e-8 |
| ☐ | PF06152.15 | Phage_min_cap2 ; Phage minor capsid protein 2 | 98.7 | 10.8518 | 194 | 297 | 253 | 346 | 0.0000033 |
| ☐ | PF04586.21 | Peptidase_S78 ; Caudovirus prohead serine protease | 97.1 | 16.8028 | 17 | 155 | 425 | 569 | 0.013 |
| ☐ | PF03420.17 | Peptidase_S77 ; Prohead core protein serine protease | 92.2 | 11.4352 | 92 | 173 | 470 | 568 | 2.8 |

This HHPred hit has a high E value and the same predicted gene function as our gene, however it has a low %coverage. It is possible that this only matches the Mu_F like protease portion of the gene and not the major capsid protein portion. When giving the gene function, however, including the Mu_F like protease domain is not included in the approved naming conventions.

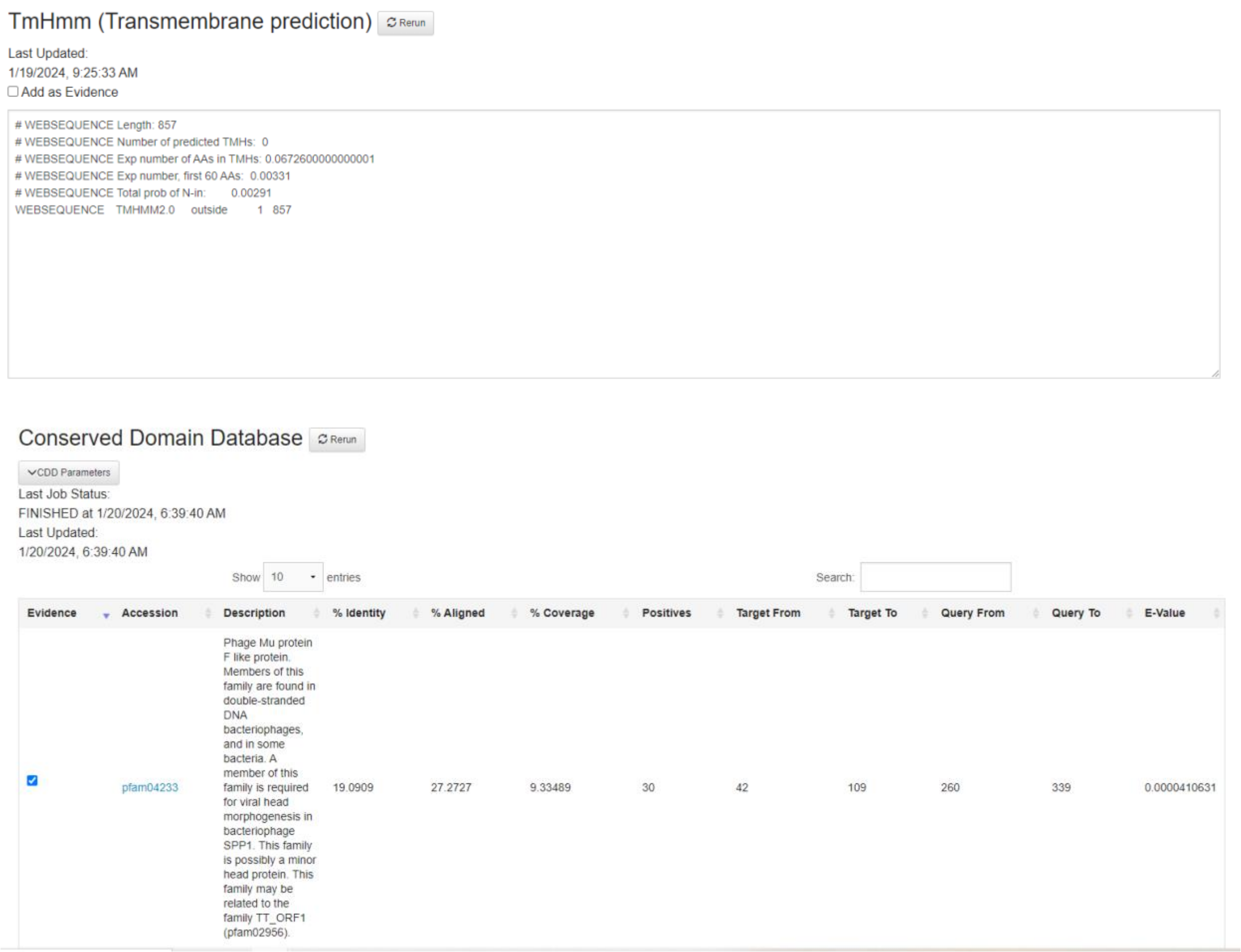
### d. SIF: Synteny-Phamerator (three genomes)

Has synteny with ABU, Chah, Charles1, ect.





### 7. Any other important information.



## TmHmm (Transmembrane prediction)

Last Updated:
1/19/2024, 9:25:33 AM
☐ Add as Evidence

```
# WEBSEQUENCE Length: 857
# WEBSEQUENCE Number of predicted TMHs:  0
# WEBSEQUENCE Exp number of AAs in TMHs: 0.0672600000000001
# WEBSEQUENCE Exp number, first 60 AAs:  0.00331
# WEBSEQUENCE Total prob of N-in:        0.00291
WEBSEQUENCE    TMHMM2.0     outside        1   857
```

## Conserved Domain Database

Last Job Status:
FINISHED at 1/20/2024, 6:39:40 AM
Last Updated:
1/20/2024, 6:39:40 AM

Show 10 entries

Search:

| Evidence | Accession | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | E-Value |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | pfam04233 | Phage Mu protein F like protein. Members of this family are found in double-stranded DNA bacteriophages, and in some bacteria. A member of this family is required for viral head morphogenesis in bacteriophage SPP1. This family is possibly a minor head protein. This family may be related to the family TT_ORF1 (pfam02956). | 19.0909 | 27.2727 | 9.33489 | 30 | 42 | 109 | 260 | 339 | 0.0000410631 |

**CURATOR NAME: SAMUEL RUTHERFORD**

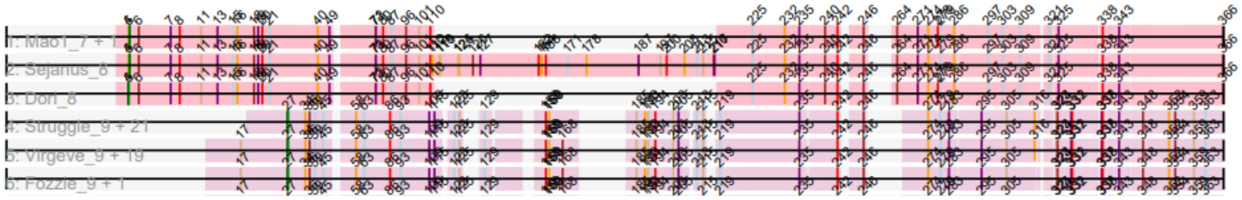
**GENE NAME: EUGENIAGENE10-RUTHERFORD**

**DNA MASTER NOTES: N/A**

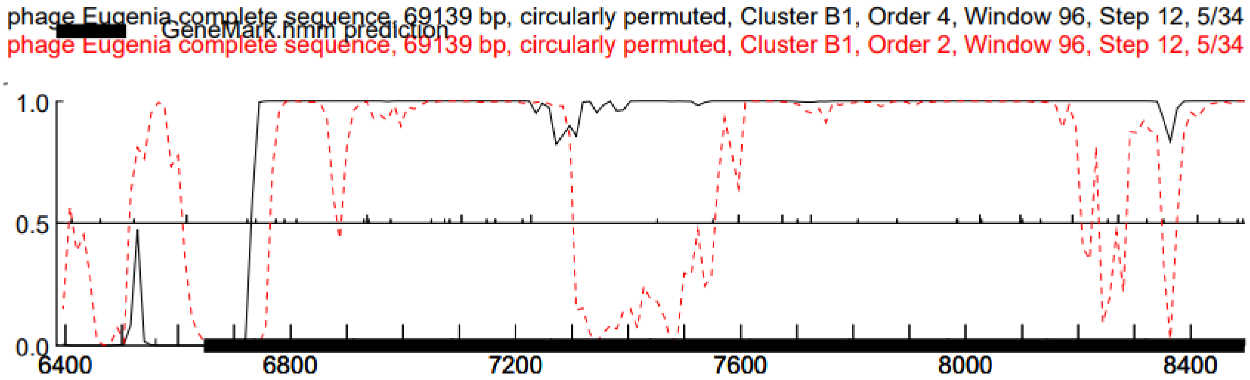
**START POSITION EVALUATION (IN ORDER):**

1. Starterator

The start number called the most often in the published annotations is 27, it was called in 375 of the 386 non-draft genes in the pham.



2. GeneMark coding potential



## 3. Glimmer and GeneMark agreement

No

Glimmer Start:

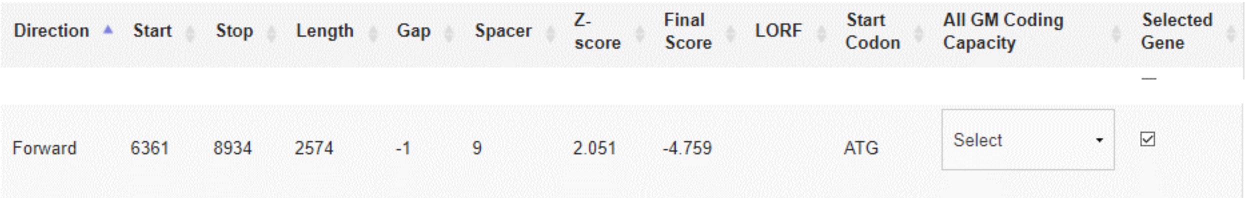
6361

Glimmer Score:

11.43

GeneMark Start:

6646



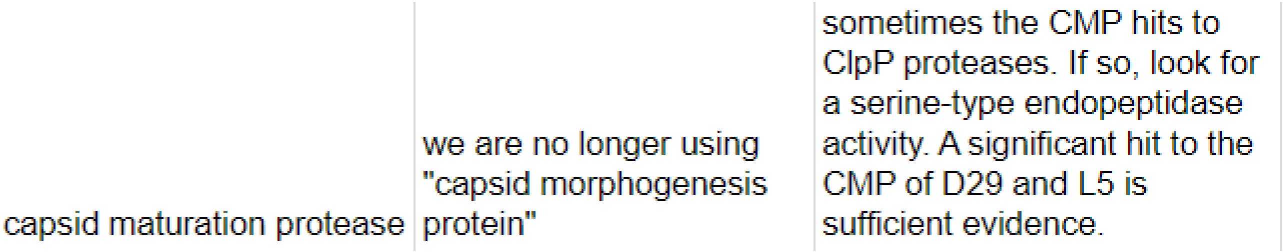
| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 6361 | 8934 | 2574 | -1 | 9 | 2.051 | -4.759 | | ATG | Select | ☑ |

## 4. Longest open reading frame (ORF) without excessive gap

No.

## 5. Function. Please follow this Official SEAPHAGE Function List

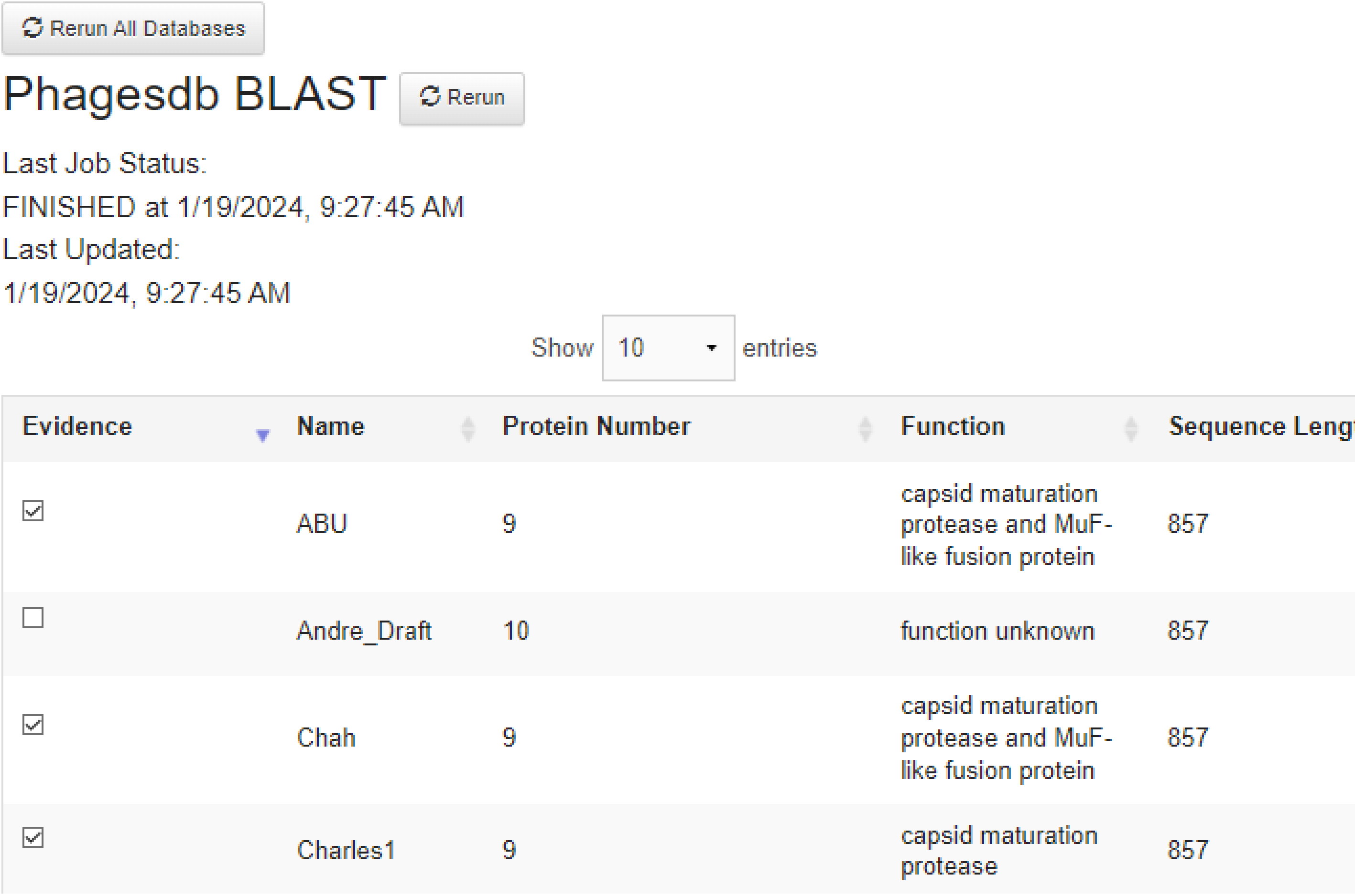
If no functional prediction is present, write “Hypothetical protein”.



| | | |
|---|---|---|
| capsid maturation protease | we are no longer using "capsid morphogenesis protein" | sometimes the CMP hits to ClpP proteases. If so, look for a serine-type endopeptidase activity. A significant hit to the CMP of D29 and L5 is sufficient evidence. |

## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST

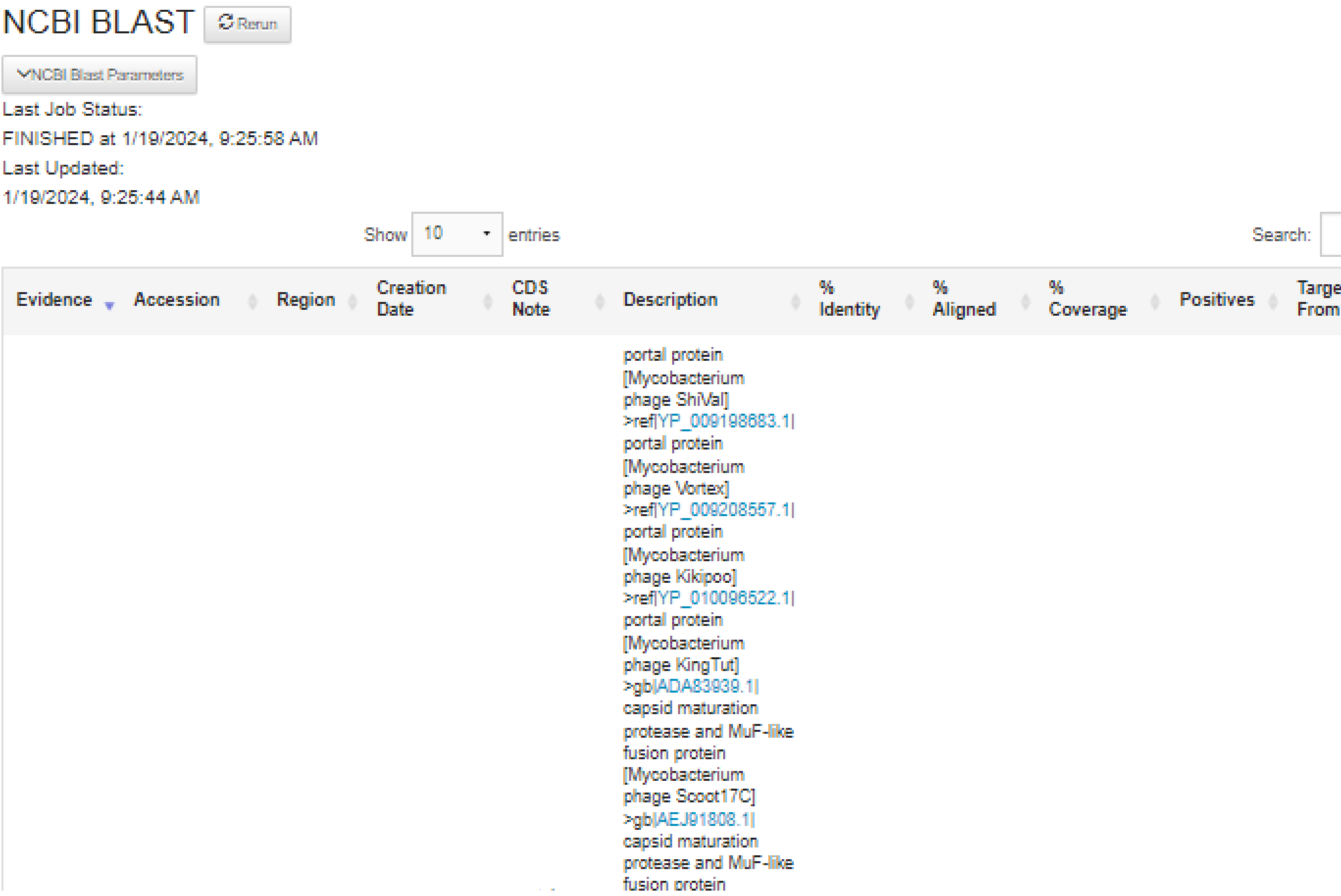


Phagesdb BLAST

Last Job Status:
FINISHED at 1/19/2024, 9:27:45 AM
Last Updated:
1/19/2024, 9:27:45 AM

| Evidence | Name | Protein Number | Function | Sequence Leng |
|---|---|---|---|---|
| ☑ | ABU | 9 | capsid maturation protease and MuF-like fusion protein | 857 |
| ☐ | Andre_Draft | 10 | function unknown | 857 |
| ☑ | Chah | 9 | capsid maturation protease and MuF-like fusion protein | 857 |
| ☑ | Charles1 | 9 | capsid maturation protease | 857 |

b. SIF: NCBI BLAST



NCBI BLAST ⟳Rerun

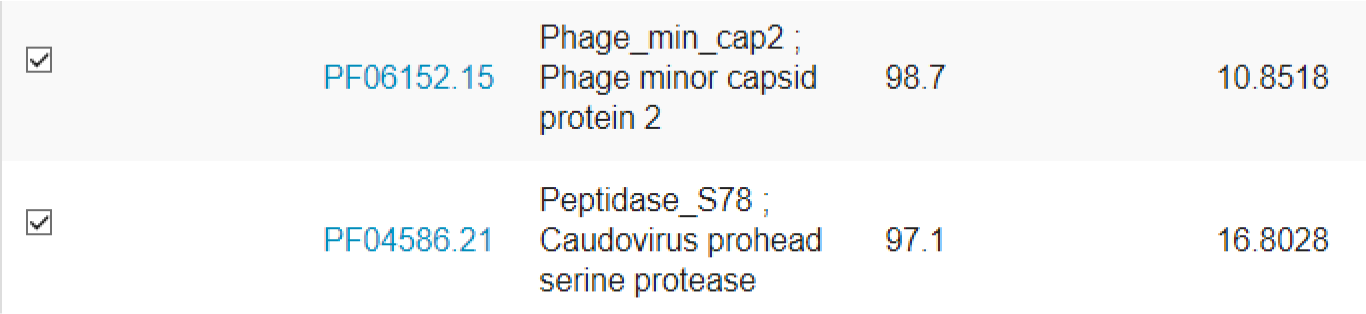
˅NCBI Blast Parameters

Last Job Status:
FINISHED at 1/19/2024, 9:25:58 AM
Last Updated:
1/19/2024, 9:25:44 AM

Show 10 entries Search:

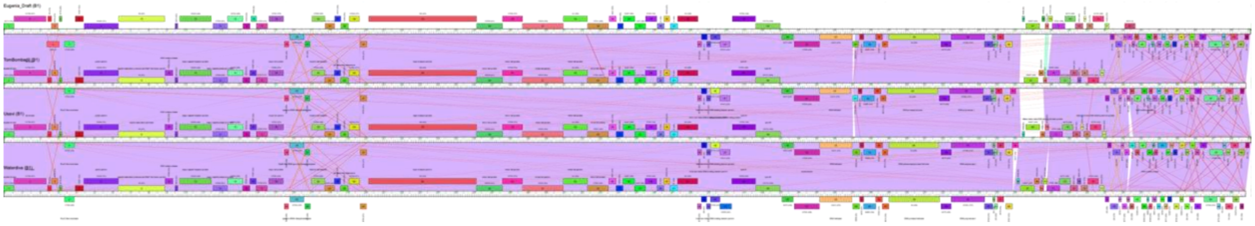
| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Targe From |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | portal protein [Mycobacterium phage ShiVal] >ref\|YP_009198683.1\| portal protein [Mycobacterium phage Vortex] >ref\|YP_009208557.1\| portal protein [Mycobacterium phage Kikipoo] >ref\|YP_010096522.1\| portal protein [Mycobacterium phage KingTut] >gb\|ADA83939.1\| capsid maturation protease and MuF-like fusion protein [Mycobacterium phage Scoot17C] >gb\|AEJ91808.1\| capsid maturation protease and MuF-like fusion protein | | | | | |

c. SIF: HHPred



| | | | | |
|---|---|---|---|---|
| ☑ | PF06152.15 | Phage_min_cap2 ; Phage minor capsid protein 2 | 98.7 | 10.8518 |
| ☑ | PF04586.21 | Peptidase_S78 ; Caudovirus prohead serine protease | 97.1 | 16.8028 |

d. SIF: Synteny-Phamerator (three genomes)



## 7. Any other important information.

No.

## CURATOR NAME: JOHN MOTTER

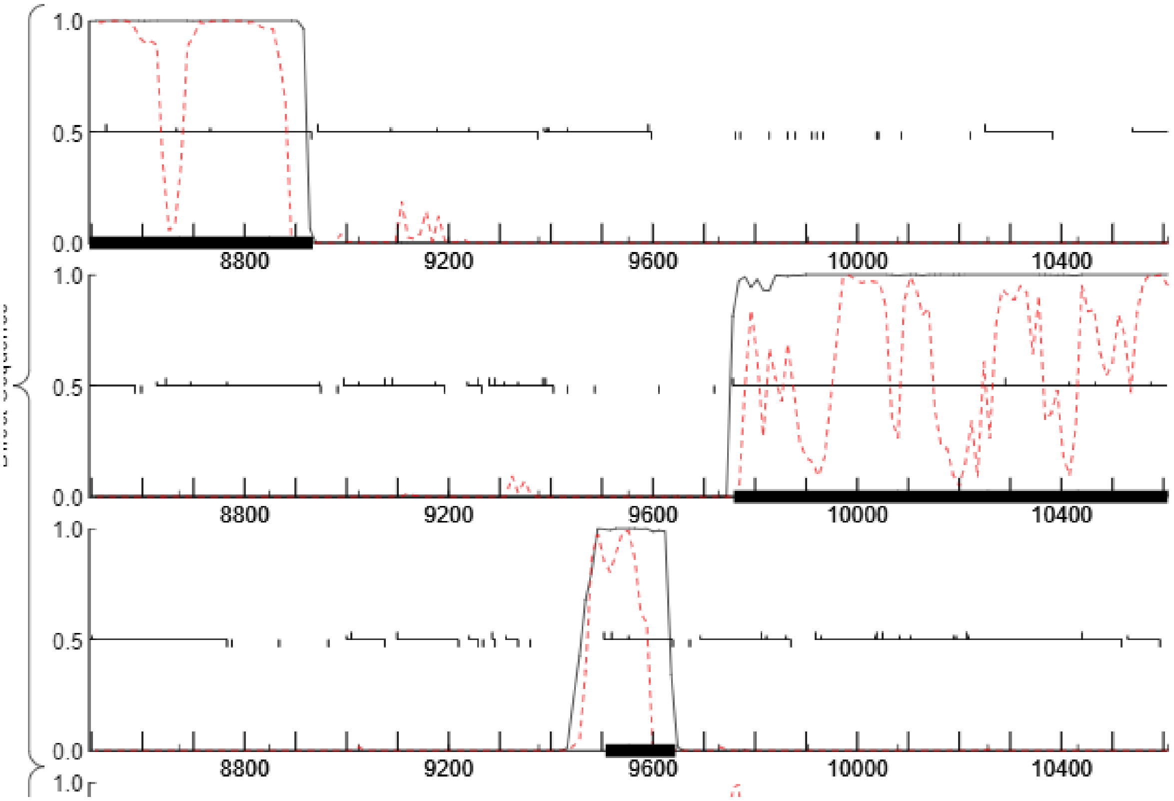
## GENE NAME: EUGENIA GENE 11

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

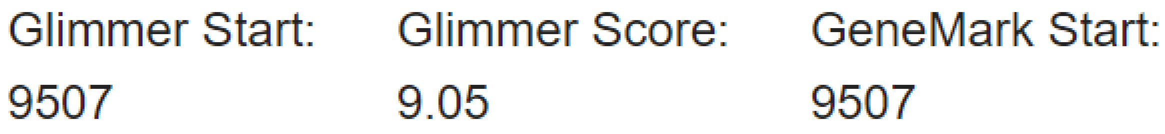
1. Starterator

## 2. GeneMark coding potential



3. Glimmer and GeneMark agreement

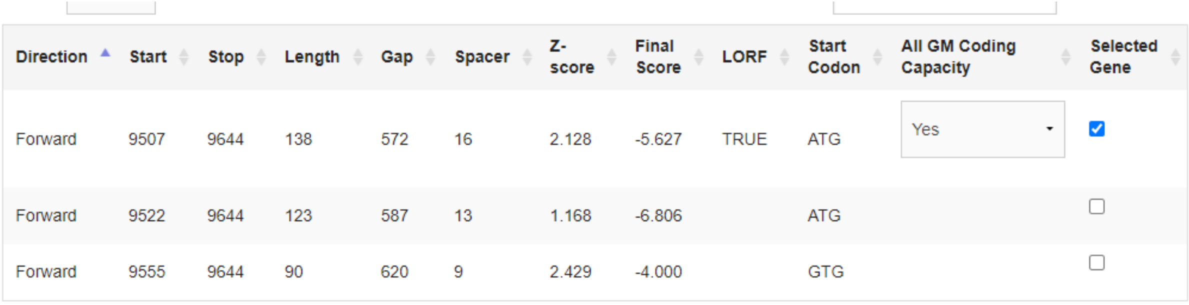

Phage: Eugenia Cluster: B1


Glimmer Start: 9507

Glimmer Score: 9.05

GeneMark Start: 9507

4. Longest open reading frame (ORF) without excessive gap



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 9507 | 9644 | 138 | 572 | 16 | 2.128 | -5.627 | TRUE | ATG | Yes | ☑ |
| Forward | 9522 | 9644 | 123 | 587 | 13 | 1.168 | -6.806 | | ATG | | ☐ |
| Forward | 9555 | 9644 | 90 | 620 | 9 | 2.429 | -4.000 | | GTG | | ☐ |

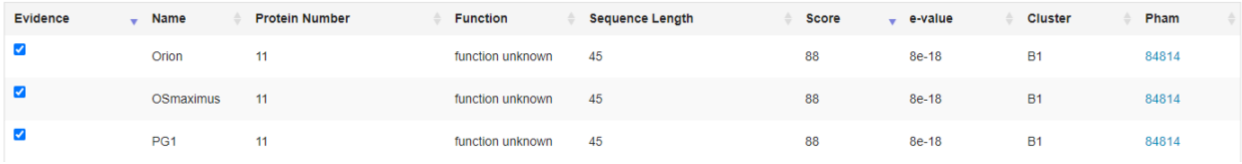
5. Function. Please follow this [Official SEAPHAGE Function List]

If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

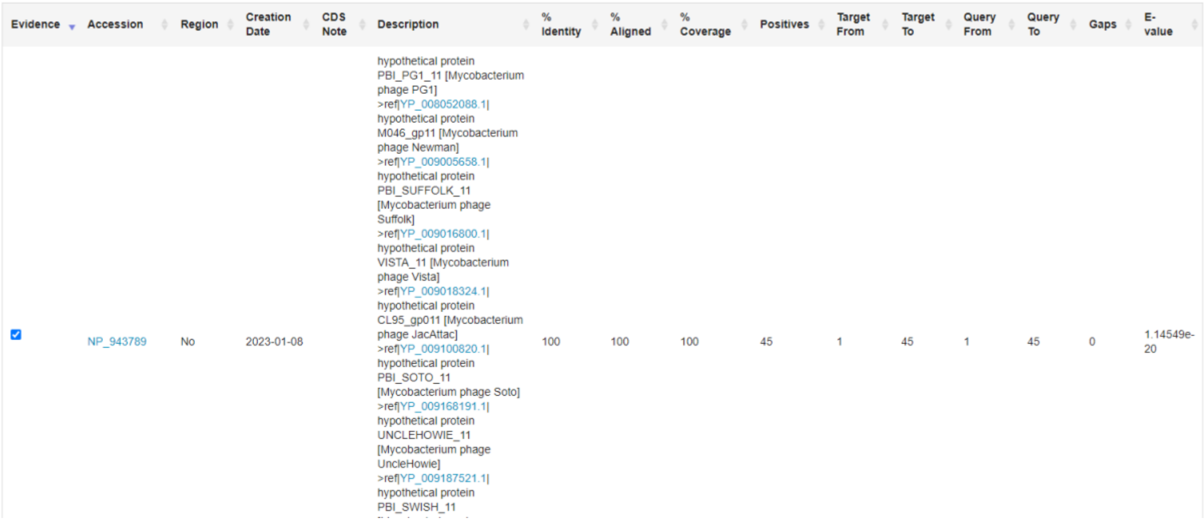
6. Supporting Information for Function (SIF)

a. SIF: PhageDb BLAST



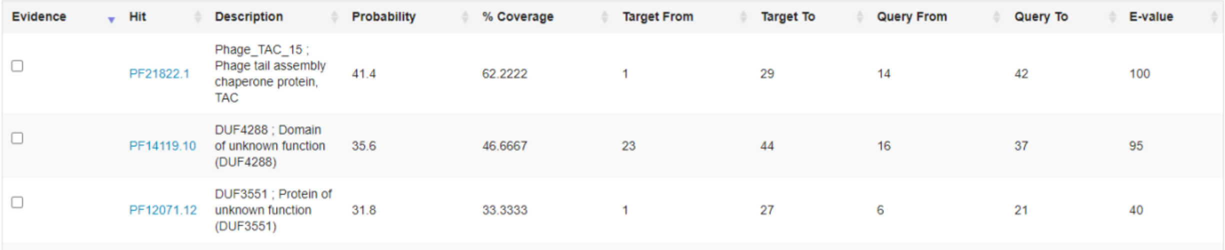
| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | Orion | 11 | function unknown | 45 | 88 | 8e-18 | B1 | 84814 |
| ☑ | OSmaximus | 11 | function unknown | 45 | 88 | 8e-18 | B1 | 84814 |
| ☑ | PG1 | 11 | function unknown | 45 | 88 | 8e-18 | B1 | 84814 |

## b. SIF: NCBI BLAST



<table>
<tr><th>Evidence</th><th>Accession</th><th>Region</th><th>Creation Date</th><th>CDS Note</th><th>Description</th><th>% Identity</th><th>% Aligned</th><th>% Coverage</th><th>Positives</th><th>Target From</th><th>Target To</th><th>Query From</th><th>Query To</th><th>Gaps</th><th>E-value</th></tr>
<tr><td>☑</td><td>NP_943789</td><td>No</td><td>2023-01-08</td><td></td><td>hypothetical protein PBI_PG1_11 [Mycobacterium phage PG1] >ref|YP_008052088.1| hypothetical protein M046_gp11 [Mycobacterium phage Newman] >ref|YP_009005658.1| hypothetical protein PBI_SUFFOLK_11 [Mycobacterium phage Suffolk] >ref|YP_009016800.1| hypothetical protein VISTA_11 [Mycobacterium phage Vista] >ref|YP_009018324.1| hypothetical protein CL95_gp011 [Mycobacterium phage JacAttac] >ref|YP_009100820.1| hypothetical protein PBI_SOTO_11 [Mycobacterium phage Soto] >ref|YP_009168191.1| hypothetical protein UNCLEHOWIE_11 [Mycobacterium phage UncleHowie] >ref|YP_009187521.1| hypothetical protein PBI_SWISH_11</td><td>100</td><td>100</td><td>100</td><td>45</td><td>1</td><td>45</td><td>1</td><td>45</td><td>0</td><td>1.14549e-20</td></tr>
</table>

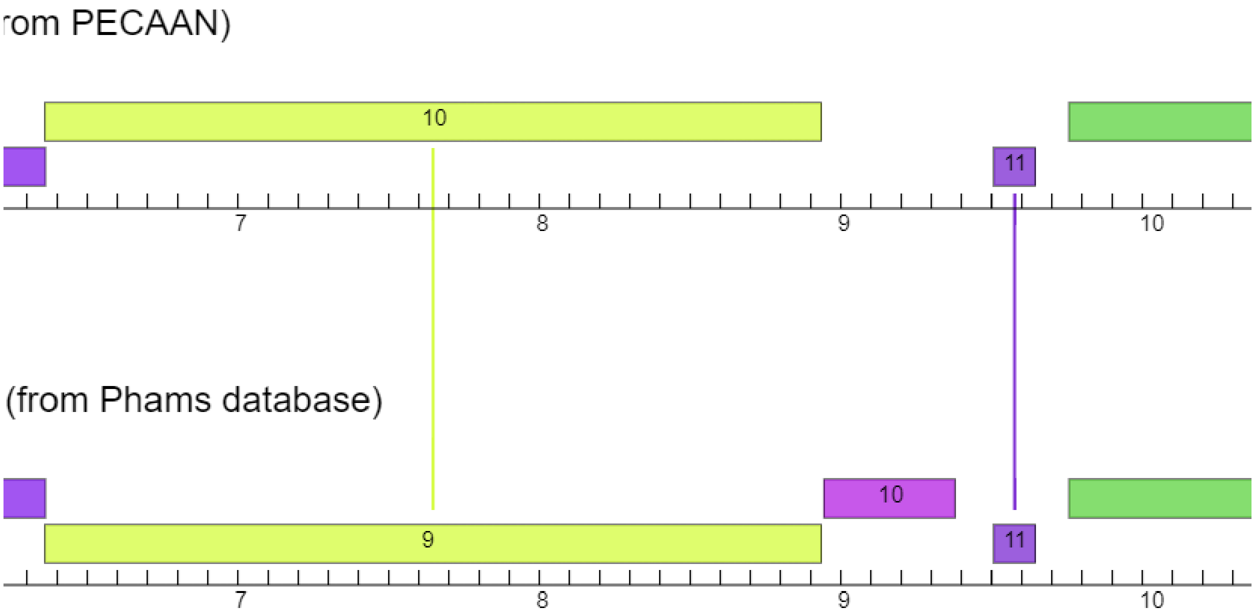
## c. SIF: HHPred

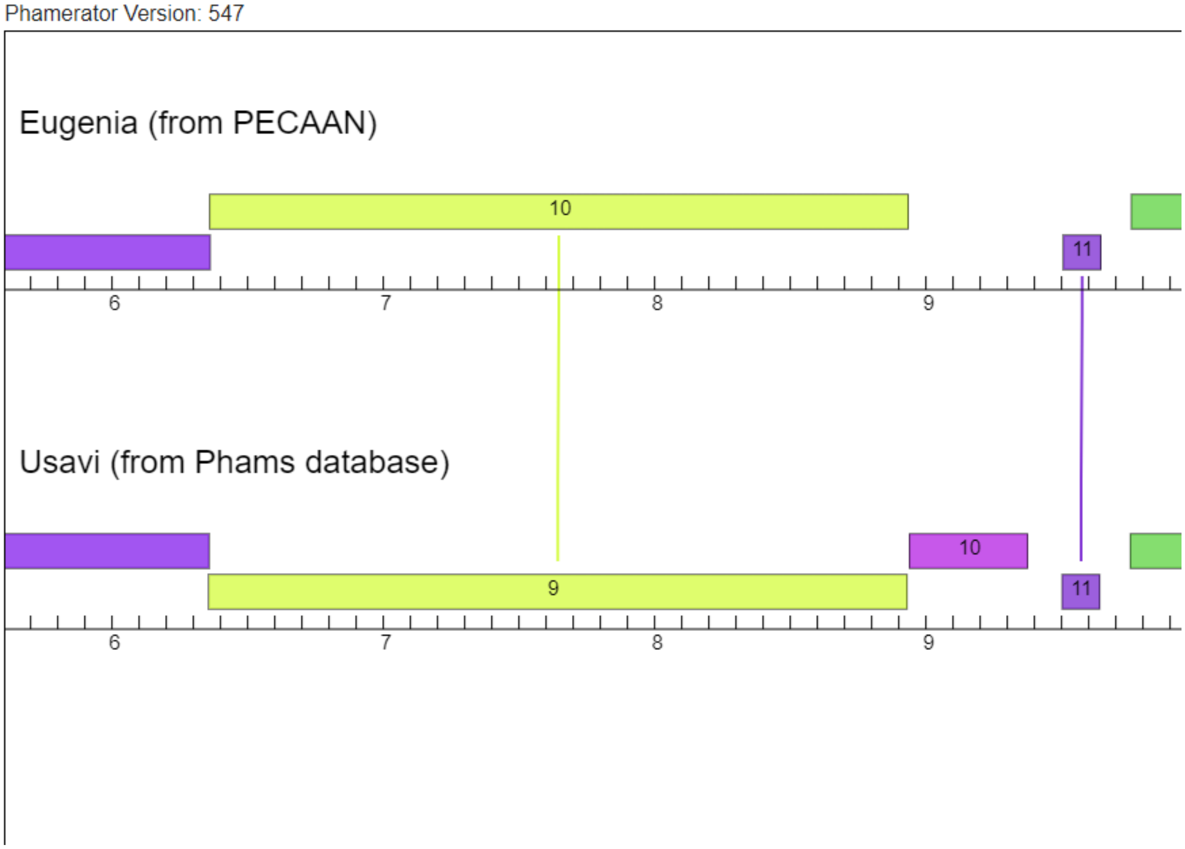


| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☐ | PF21822.1 | Phage_TAC_15 ; Phage tail assembly chaperone protein, TAC | 41.4 | 62.2222 | 1 | 29 | 14 | 42 | 100 |
| ☐ | PF14119.10 | DUF4288 ; Domain of unknown function (DUF4288) | 35.6 | 46.6667 | 23 | 44 | 16 | 37 | 95 |
| ☐ | PF12071.12 | DUF3551 ; Protein of unknown function (DUF3551) | 31.8 | 33.3333 | 1 | 27 | 6 | 21 | 40 |

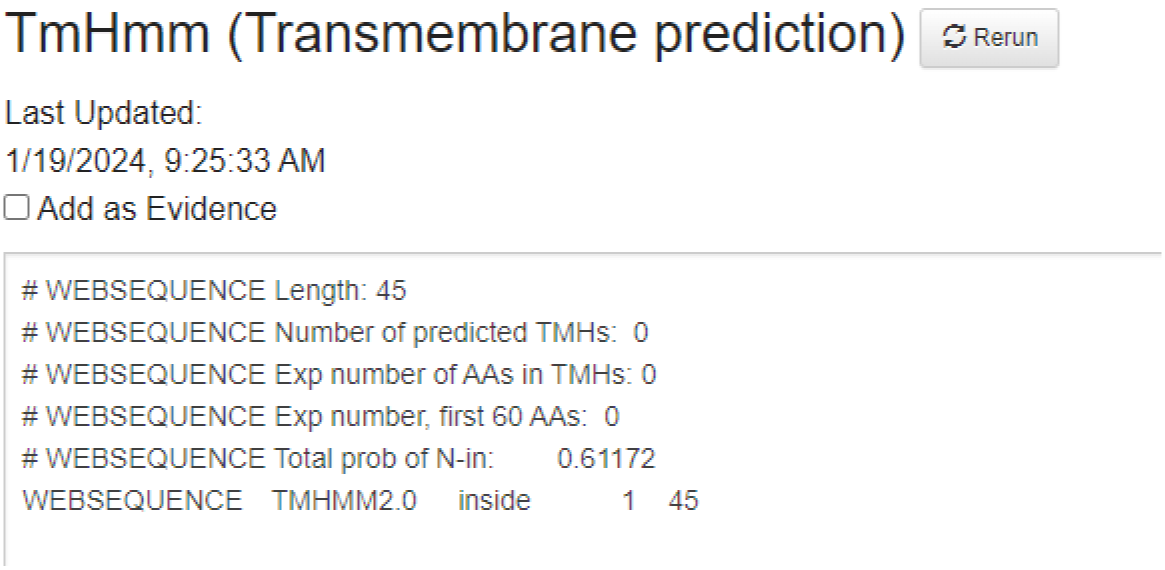
## d. SIF: Synteny-Phamerator (three genomes)

Synteny with Orion, OSmaximus, PG1, ect.


rom PECAAN)
10
11
7
8
9
10
(from Phams database)
10
9
11
7
8
9
10


Phamerator Version: 547
Eugenia (from PECAAN)
10
11
6
7
8
9
Usavi (from Phams database)
10
9
11
6
7
8
9

7. Any other important information.



TmHmm (Transmembrane prediction)

Last Updated:
1/19/2024, 9:25:33 AM

☐ Add as Evidence

```
# WEBSEQUENCE Length: 45
# WEBSEQUENCE Number of predicted TMHs:  0
# WEBSEQUENCE Exp number of AAs in TMHs: 0
# WEBSEQUENCE Exp number, first 60 AAs:  0
# WEBSEQUENCE Total prob of N-in:        0.61172
WEBSEQUENCE    TMHMM2.0     inside          1    45
```

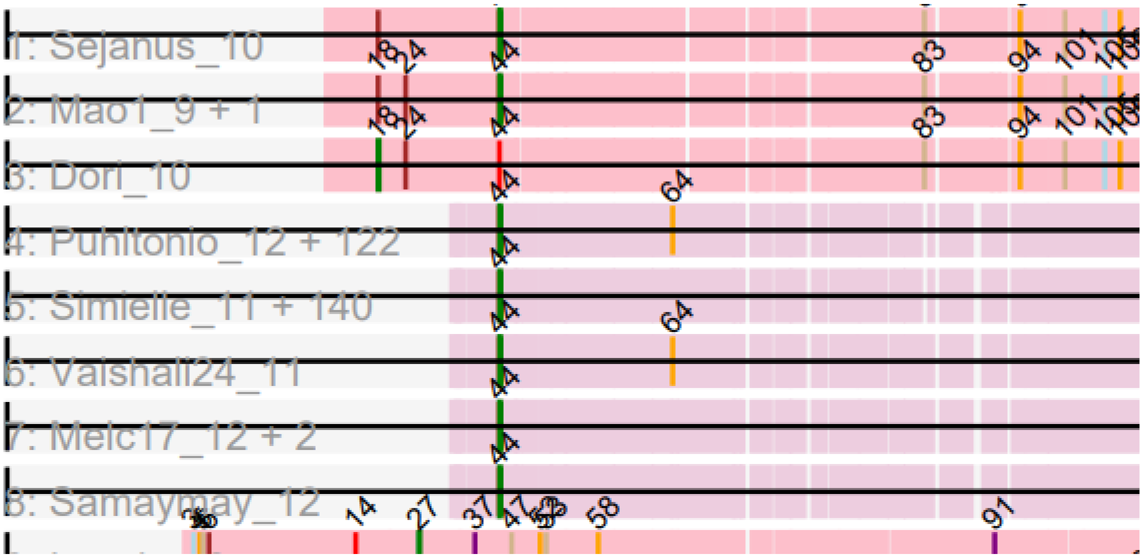
### CURATOR NAME: JOHN MOTTER

### GENE NAME: EUGENIA GENE 12

### DNA MASTER NOTES: N/A

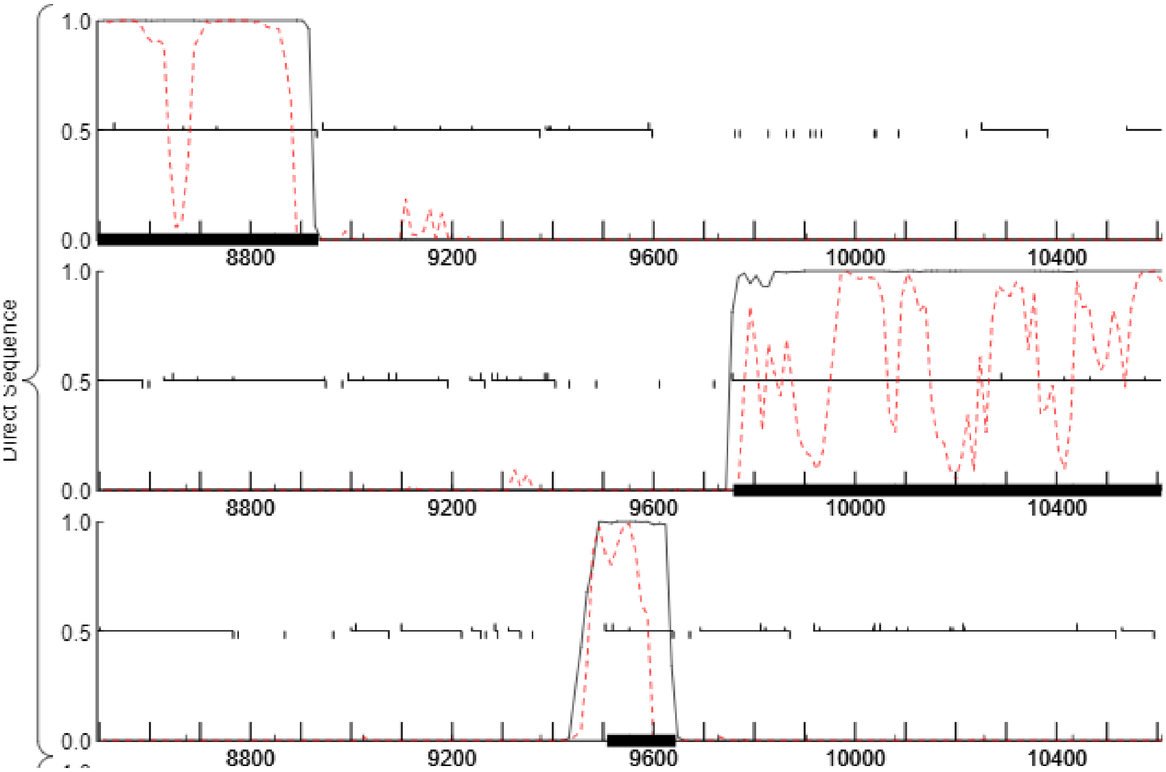
### START POSITION EVALUATION (IN ORDER):

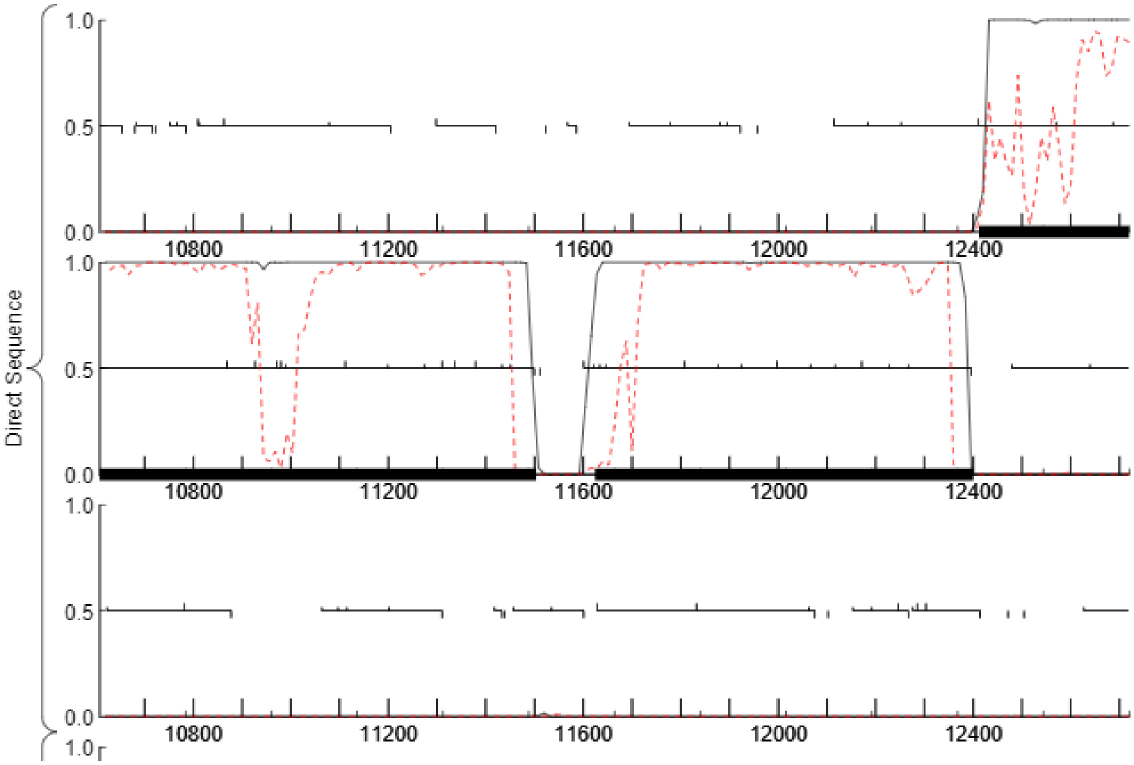
1. Starterator



The start number called the most often in the published annotations is 44, it was called in 293 of the 394 non-draft genes in the pham

2. GeneMark coding potential

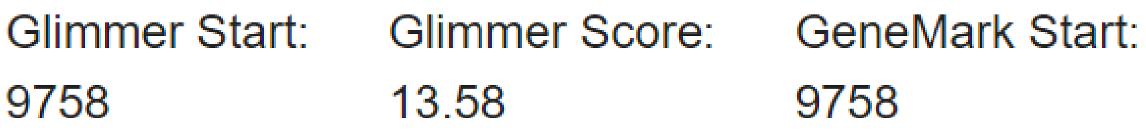

Direct Sequence
1.0
0.5
0.0
8800
9200
9600
10000
10400


Direct Sequence
1.0
0.5
0.0
10800
11200
11600
12000
12400

3. Glimmer and GeneMark agreement

Phage: Eugenia

Cluster: B1

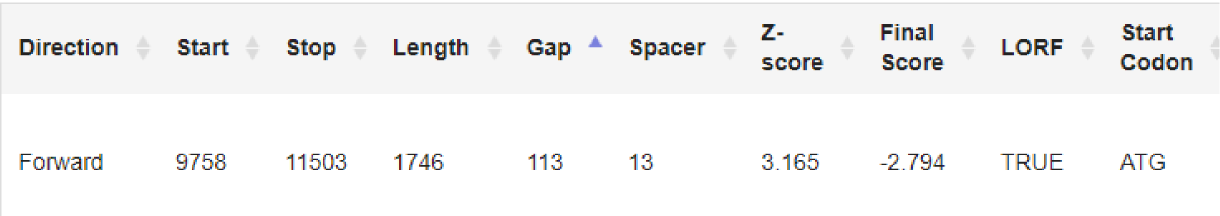


Glimmer Start: 9758

Glimmer Score: 13.58

GeneMark Start: 9758

4. Longest open reading frame (ORF) without excessive gap



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon |
|---|---|---|---|---|---|---|---|---|---|
| Forward | 9758 | 11503 | 1746 | 113 | 13 | 3.165 | -2.794 | TRUE | ATG |

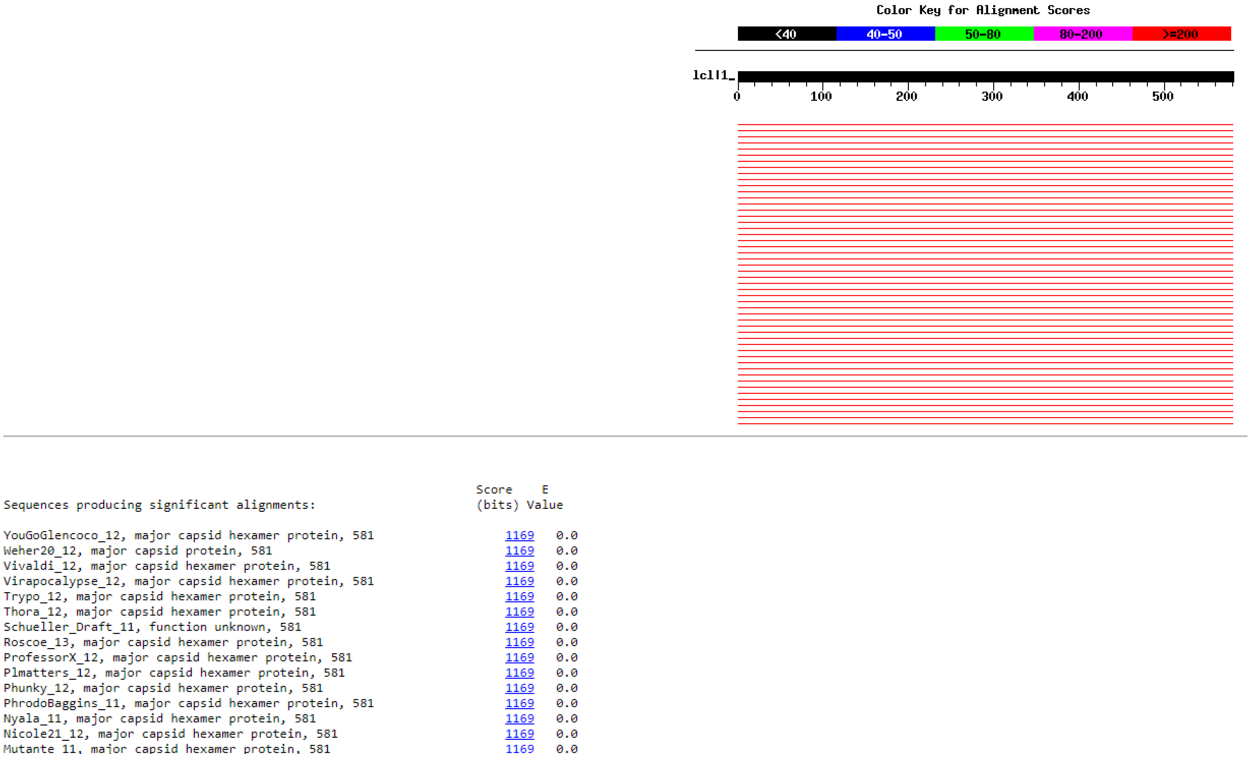
5. Function. Please follow this [Official SEAPHAGE Function List]

If no functional prediction is present, write “Hypothetical protein”.

Major Capsid Hexamer Protein

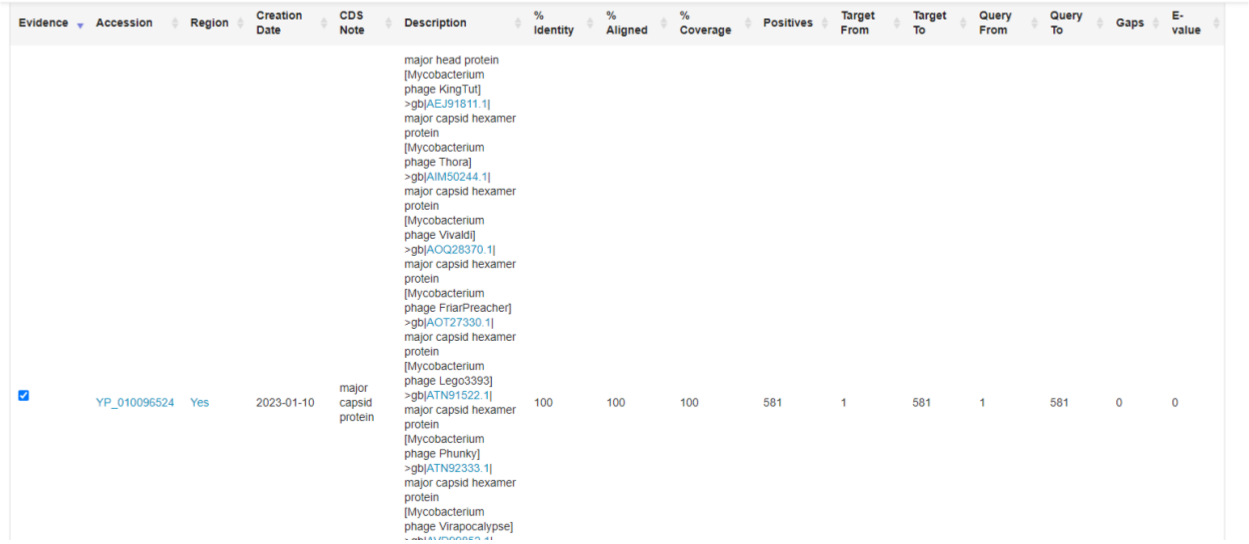
6. Supporting Information for Function (SIF)

## a. SIF: PhageDb BLAST



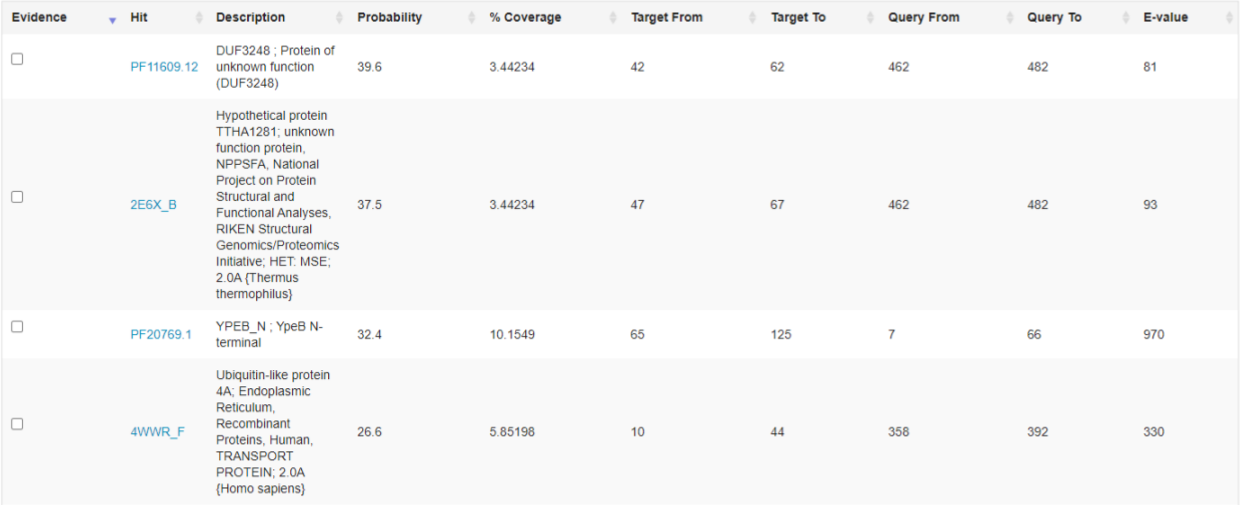
```
                                                          Score     E
Sequences producing significant alignments:               (bits) Value

YouGoGlencoco_12, major capsid hexamer protein, 581          1169   0.0
Weher20_12, major capsid protein, 581                        1169   0.0
Vivaldi_12, major capsid hexamer protein, 581                1169   0.0
Virapocalypse_12, major capsid hexamer protein, 581          1169   0.0
Trypo_12, major capsid hexamer protein, 581                  1169   0.0
Thora_12, major capsid hexamer protein, 581                  1169   0.0
Schueller_Draft_11, function unknown, 581                    1169   0.0
Roscoe_13, major capsid hexamer protein, 581                 1169   0.0
ProfessorX_12, major capsid hexamer protein, 581             1169   0.0
Plmatters_12, major capsid hexamer protein, 581              1169   0.0
Phunky_12, major capsid hexamer protein, 581                 1169   0.0
PhrodoBaggins_11, major capsid hexamer protein, 581          1169   0.0
Nyala_11, major capsid hexamer protein, 581                  1169   0.0
Nicole21_12, major capsid hexamer protein, 581               1169   0.0
Mutante 11, major capsid hexamer protein, 581                1169   0.0
```

## b. SIF: NCBI BLAST



| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | YP_010096524 | Yes | 2023-01-10 | major capsid protein | major head protein [Mycobacterium phage KingTut] >gb\|AEJ91811.1\| major capsid hexamer protein [Mycobacterium phage Thora] >gb\|AIM50244.1\| major capsid hexamer protein [Mycobacterium phage Vivaldi] >gb\|AOQ28370.1\| major capsid hexamer protein [Mycobacterium phage FriarPreacher] >gb\|AOT27330.1\| major capsid hexamer protein [Mycobacterium phage Lego3393] >gb\|ATN91522.1\| major capsid hexamer protein [Mycobacterium phage Phunky] >gb\|ATN92333.1\| major capsid hexamer protein [Mycobacterium phage Virapocalypse] | 100 | 100 | 100 | 581 | 1 | 581 | 1 | 581 | 0 | 0 |

c. SIF: HHPred

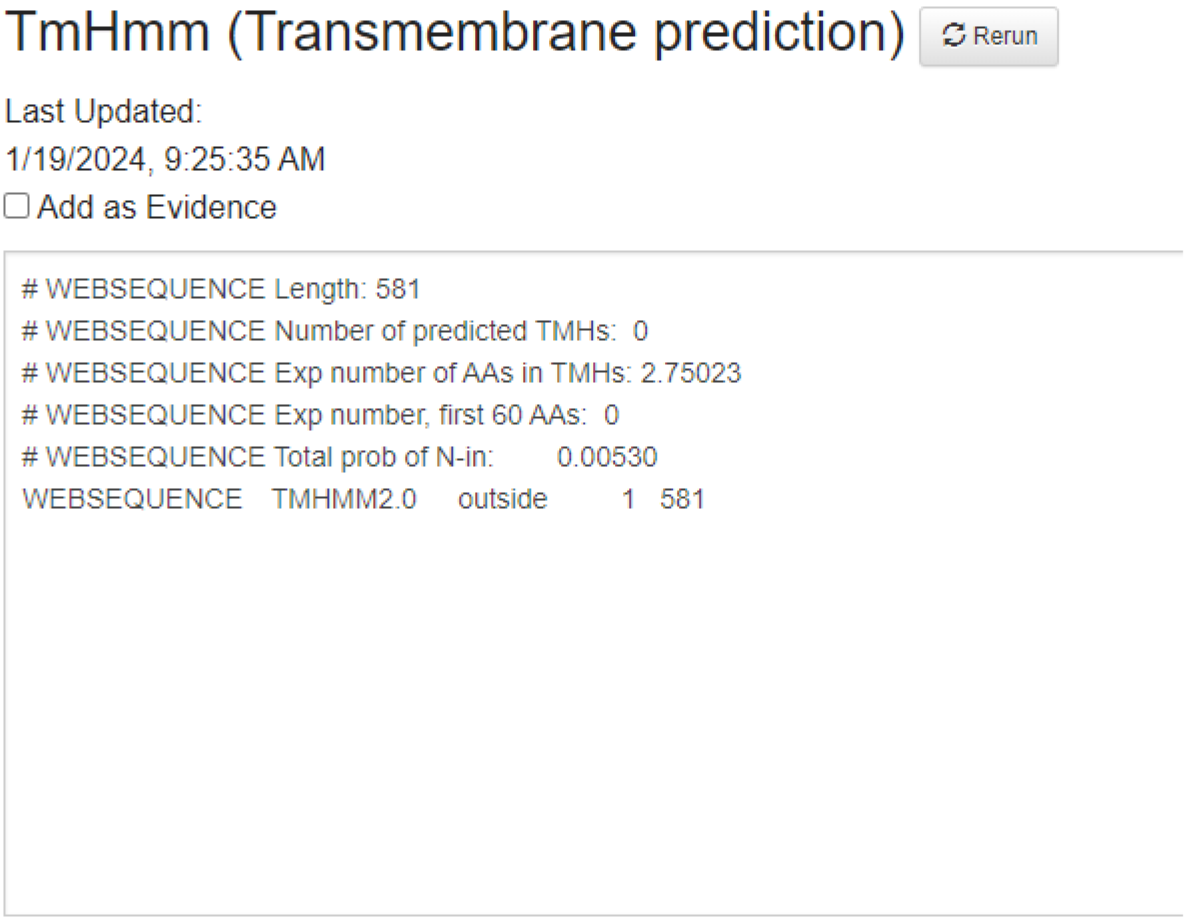


| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☐ | PF11609.12 | DUF3248 ; Protein of unknown function (DUF3248) | 39.6 | 3.44234 | 42 | 62 | 462 | 482 | 81 |
| ☐ | 2E6X_B | Hypothetical protein TTHA1281; unknown function protein, NPPSFA, National Project on Protein Structural and Functional Analyses, RIKEN Structural Genomics/Proteomics Initiative; HET: MSE; 2.0A {Thermus thermophilus} | 37.5 | 3.44234 | 47 | 67 | 462 | 482 | 93 |
| ☐ | PF20769.1 | YPEB_N ; YpeB N-terminal | 32.4 | 10.1549 | 65 | 125 | 7 | 66 | 970 |
| ☐ | 4WWR_F | Ubiquitin-like protein 4A; Endoplasmic Reticulum, Recombinant Proteins, Human, TRANSPORT PROTEIN; 2.0A {Homo sapiens} | 26.6 | 5.85198 | 10 | 44 | 358 | 392 | 330 |

d. SIF: Synteny-Phamerator (three genomes)

Has synteny with Childish, Craff, FriarPreacher, ect.

7. Any other important information.



TmHmm (Transmembrane prediction)

Last Updated:
1/19/2024, 9:25:35 AM

☐ Add as Evidence

```
# WEBSEQUENCE Length: 581
# WEBSEQUENCE Number of predicted TMHs:  0
# WEBSEQUENCE Exp number of AAs in TMHs: 2.75023
# WEBSEQUENCE Exp number, first 60 AAs:  0
# WEBSEQUENCE Total prob of N-in:        0.00530
WEBSEQUENCE    TMHMM2.0    outside       1   581
```

## CURATOR NAME: ISABELLA LIMA

## GENE NAME: EUGENIA_DRAFT_13

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

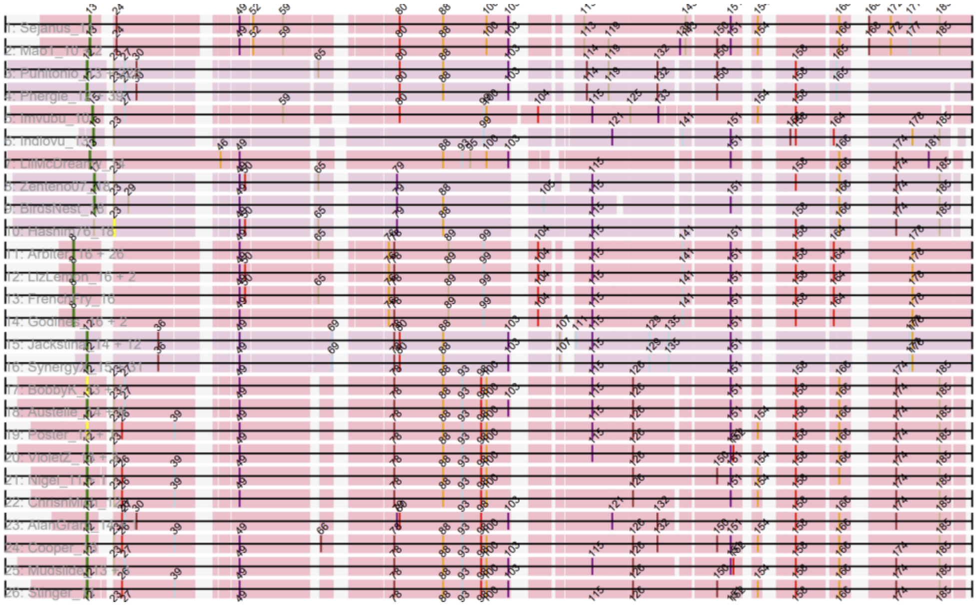
### 1. Starterator

Eugenia_13 is in Track 3 with Start 12.

The start number called the most often in the published annotations is 12, it was called in 332 of the 489 non-draft genes in the pham.

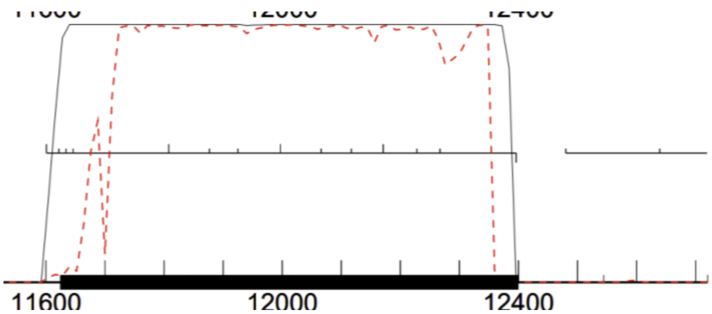
Start 12:

- Found in 376 of 561 ( 67.0% ) of genes in pham
- Manual Annotations of this start: 332 of 489
- Called 99.5% of time when present



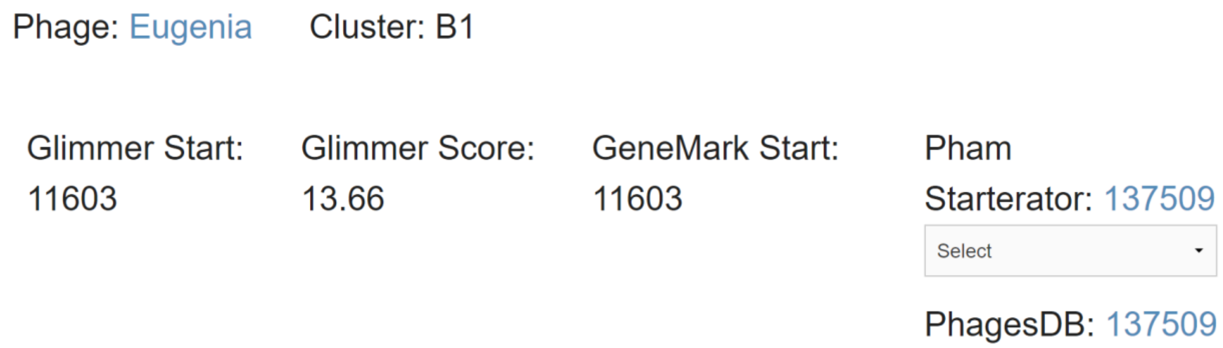
## 2. GeneMark coding potential

GeneMark Start: 11603



## 3. Glimmer and GeneMark agreement

YES, there is agreement



Phage: Eugenia Cluster: B1

Glimmer Start: 11603

Glimmer Score: 13.66

GeneMark Start: 11603

Pham Starterator: 137509

Select

PhagesDB: 137509

Glimmer: 11603

GeneMark: 11603

## 4. Longest open reading frame (ORF) without excessive gap

YES

| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 11603 | 12400 | 798 | 99 | 11 | 2.746 | -3.347 | TRUE | ATG | Select | ☑ |
| Forward | 11624 | 12400 | 777 | 120 | 11 | 0.897 | -7.061 | | GTG | | ☐ |

Length: 798

Gap: 99

Spacer: 11

Z-Score: 2.746

Final Score: -3.347

## 6. Start site similar to others in GenBank and PhagesDB

YES

- Track 3 : Puhltonio_13, Soto_13, HSavage_13, TallGrassMM_13, Zaider_14, Simielle_12, HenryJackson_12, Sheila_13, Gyarad_0013, Cornobble_12, Fringe_13, Numberten_13, Vaishali24_12, Carthage_12, Phamished_13, Doddsville_13, OliverWalter_13, Dione_12, FluffyNinja_13, Yoshand_13, Etaye_13, Virgeve_12, Toni_12, Mcshane_12, Vivaldi_13, Plmatters_13, Newman_13, Nacho_0013, JakeO_12, CheetO_13, Spartan300_13, Piglet_0012, Squid_13, CamL_13, Daffy_13, Roscoe_14, Antonia_12, Mosaic_12, DelRivs_12, Serpentine_0013, KlimbOn_13, BlueHusk_13, Hamish_13, Suffolk_13, Mikota_13, Rimu_12, Vista_13, ThreeOh3D2_13, Sophia_12, Grand2040_12, DonSanchon_12, PhatLouie_13, Badfish_13, Solosis_12, Mutante_12, Keitherie_13, Crownjwl_13, Aelin_12, Serendipity_13, DaddyDaniels_13, Inchworm_13, Placalicious_12, Dati_13, Haimas_13, DirtJuice_13, Oline_12, Jillium_12, Altwerkus_12, Nicole21_13, Surely_13, Chorkpop_13, Hetaeria_13, Slatt_13, Burr_12, KLucky39_13, Selr12_13, MrPhizzler_12, UncleHowie_13, Valjean_13, Virapocalypse_13, Pipsqueak_13, Pherdinand_13, Childish_12, MiniBoss_13, Soile_13, Ricotta_13, Mulan_12, Prickles_13, Duggie_12, Iridoclysis_13, Schueller_12, Gophee_13, Horchata_13, Morty_13, Maru_12, Mana_12, Magic8_13, Morgushi_13, MitKao_13, Olak_12, Boehler_13, PinheadLarry_13, JangoPhett_13, SDcharge11_13, BlackStallion_13, Durga_13, Riggan_13, Bishoperium_12, UAch1_13, Mesh1_13, Mecca_12, AbsoluteMadLad_12, Fang_13, TomBombadil_13, Quisquiliae_13, OSmaximus_13, Phipps_13, Craff_13, Kwadwo_12, Melc17_13, Lulwa_13, Manad_12, PhrodoBaggins_12, ProfessorX_13, FugateOSU_12, Gareth_12, Kwksand96_12, JDog_12, Beaglebox_12, Robyn_12, FriarPreacher_13, Cannibal_13, Bluephacebaby_12, Scoot17C_13, Lasso_13, LostAndPhound_12, Lopsy_12, Schadenfreude_12, Katniss_13, Tooj_13, Haleema_12, Freya_12, Pacifista_14, Chaelin_12, AltPhacts_13, Orion_13, Xavier_12, SassyCat97_12, Emiris_13, Charles1_13, PG1_13, Vortex_13, Cobra_13, Megatron_13, QueenBeane_13, Telesworld_12, Omniscient_13, LeeLot_13, Podrick_13, Giraffe_13, Murdoc_13, KingTut_12, IsaacEli_13, Vaticameos_13, Potter_12, Kikipoo_13, Pops_13, Chunky_13, Held_13, Lego3393_13, Kimbrough_13, CampRoach_12, Dingo_13, Oosterbaan_13, LeiMonet_13, Windsor_12, Roliet_13, Anderson_13, Harvey_13, Legolas_12, Tomlarah_13, Timmi_12, Matalotodo_12, Nyala_12, Phareon_13, LemonSlice_13, Chah_13, Inverness_13, Hertubise_13, Fozzie_12, Jiminy_12, Zelda_13, Banjo_12, EmpTee_13, Samaymay_13, Eugenia_13, Adriana_13, Orwigg_13, MRabcd_12, Thora_13, HighStump_12, LasagnaCat_13, Frankicide_13,

## 7. Associated ribosome binding site (RBS)

N/A

## 8. Function (F). Please follow this [Official SEAPHAGE Function List.](#) If no functional prediction is present, write "Hypothetical protein."

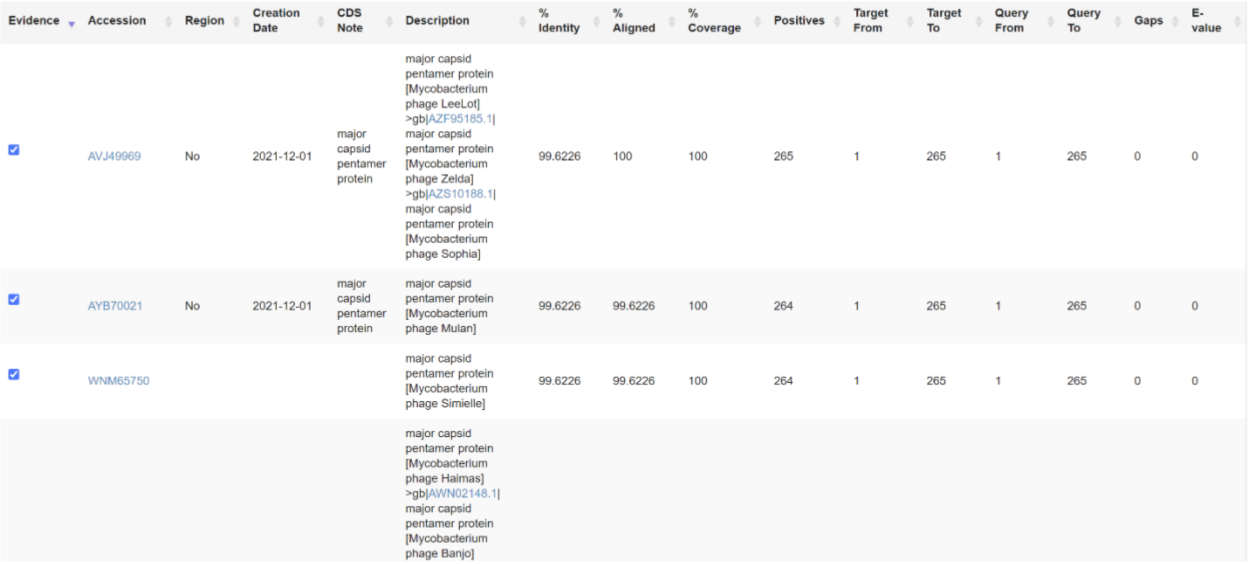
Major capsid pentamer protein: Rank 1

Phagesdb Function Frequency

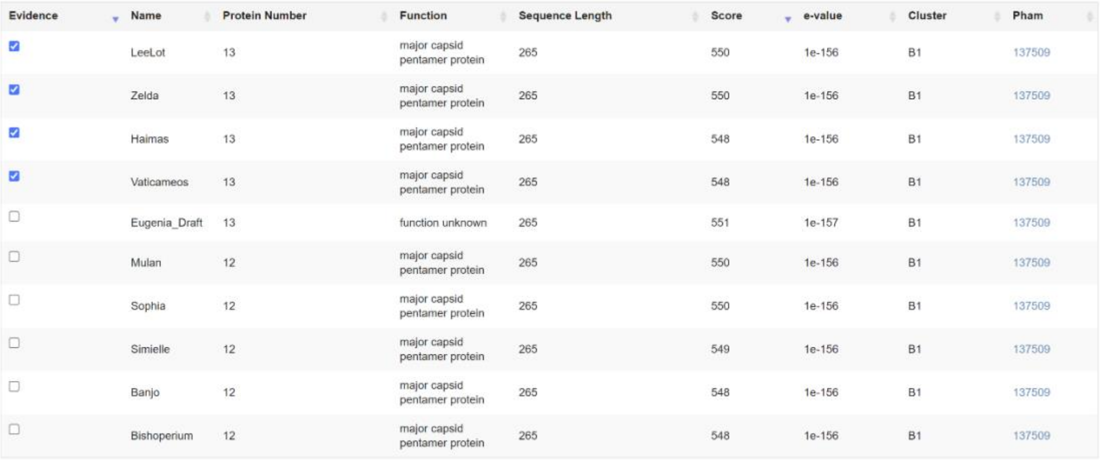
| Function Name | Pham | Subcluster | Count | Frequency (%) |
|---|---|---|---|---|
| major capsid pentamer protein | 137509 | B1 | 87 | 100 |

# 9. Supporting Information for Function (SIF)

## 9. a. SIF-BLAST



| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | AVJ49969 | No | 2021-12-01 | major capsid pentamer protein | major capsid pentamer protein [Mycobacterium phage LeeLot] >gb\|AZF95185.1\| major capsid pentamer protein [Mycobacterium phage Zelda] >gb\|AZS10188.1\| major capsid pentamer protein [Mycobacterium phage Sophia] | 99.6226 | 100 | 100 | 265 | 1 | 265 | 1 | 265 | 0 | 0 |
| ☑ | AYB70021 | No | 2021-12-01 | major capsid pentamer protein | major capsid pentamer protein [Mycobacterium phage Mulan] | 99.6226 | 99.6226 | 100 | 264 | 1 | 265 | 1 | 265 | 0 | 0 |
| ☑ | WNM65750 | | | | major capsid pentamer protein [Mycobacterium phage Simielle] | 99.6226 | 99.6226 | 100 | 264 | 1 | 265 | 1 | 265 | 0 | 0 |
| | | | | | major capsid pentamer protein [Mycobacterium phage Haimas] >gb\|AWN02148.1\| major capsid pentamer protein [Mycobacterium phage Banjo] | | | | | | | | | | |



| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | LeeLot | 13 | major capsid pentamer protein | 265 | 550 | 1e-156 | B1 | 137509 |
| ☑ | Zelda | 13 | major capsid pentamer protein | 265 | 550 | 1e-156 | B1 | 137509 |
| ☑ | Haimas | 13 | major capsid pentamer protein | 265 | 548 | 1e-156 | B1 | 137509 |
| ☑ | Vaticameos | 13 | major capsid pentamer protein | 265 | 548 | 1e-156 | B1 | 137509 |
| ☐ | Eugenia_Draft | 13 | function unknown | 265 | 551 | 1e-157 | B1 | 137509 |
| ☐ | Mulan | 12 | major capsid pentamer protein | 265 | 550 | 1e-156 | B1 | 137509 |
| ☐ | Sophia | 12 | major capsid pentamer protein | 265 | 550 | 1e-156 | B1 | 137509 |
| ☐ | Simielle | 12 | major capsid pentamer protein | 265 | 549 | 1e-156 | B1 | 137509 |
| ☐ | Banjo | 12 | major capsid pentamer protein | 265 | 548 | 1e-156 | B1 | 137509 |
| ☐ | Bishoperium | 12 | major capsid pentamer protein | 265 | 548 | 1e-156 | B1 | 137509 |

```
                                                                Score    E
Sequences producing significant alignments:                     (bits) Value

Eugenia_Draft_13, function unknown, 265                              551   e-157
Zelda_13, major capsid pentamer protein, 265                         550   e-156
Sophia_12, major capsid pentamer protein, 265                        550   e-156
LeeLot_13, major capsid pentamer protein, 265                        550   e-156
Mulan_12, major capsid pentamer protein, 265                         550   e-156
Simielle_12, major capsid pentamer protein, 265                      549   e-156
Vaticameos_13, major capsid pentamer protein, 265                    548   e-156
Orwigg_Draft_13, function unknown, 265                               548   e-156
Mecca_12, major capsid pentamer protein, 265                         548   e-156
Haimas_13, major capsid pentamer protein, 265                        548   e-156
Burr_12, major capsid pentamer protein, 265                          548   e-156
Bluephacebaby_12, major capsid pentamer protein, 265                 548   e-156
Bishoperium_12, major capsid pentamer protein, 265                   548   e-156
Banjo_12, major capsid pentamer protein, 265                         548   e-156
Weher20_13, major capsid pentamer protein, 265                       547   e-156
Vista_13, major capsid pentamer protein, 265                         547   e-156
Virapocalypse_13, major capsid pentamer protein, 265                 547   e-156
Toni_12, major capsid pentamer protein, 265                          547   e-156
TomBombadil_13, major capsid pentamer protein, 265                   547   e-156
Squid_13, major capsid pentamer protein, 265                         547   e-156
```

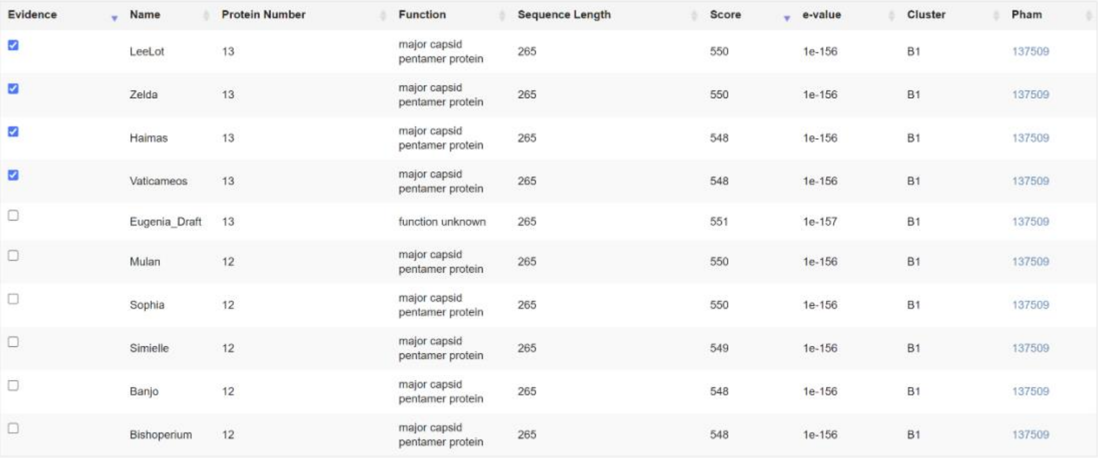
## 9. b. SIF-HHPred

N/A

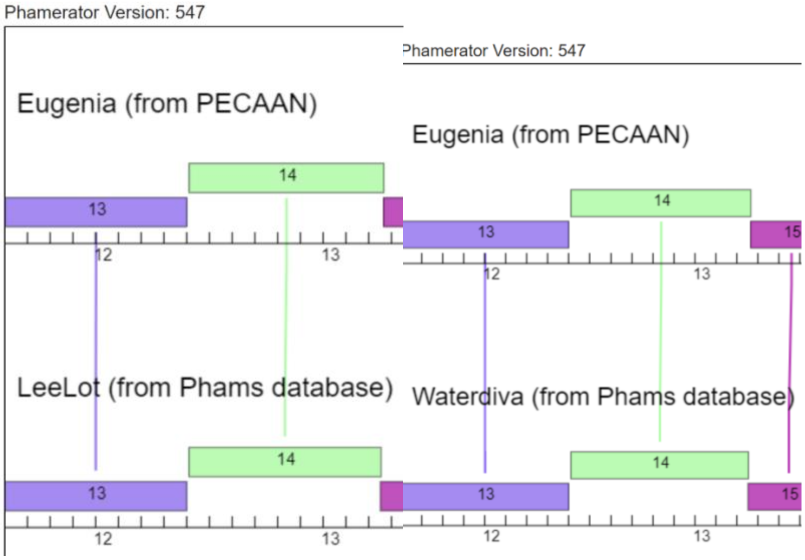
## 9. c. SIF-Synteny—Three Genomes

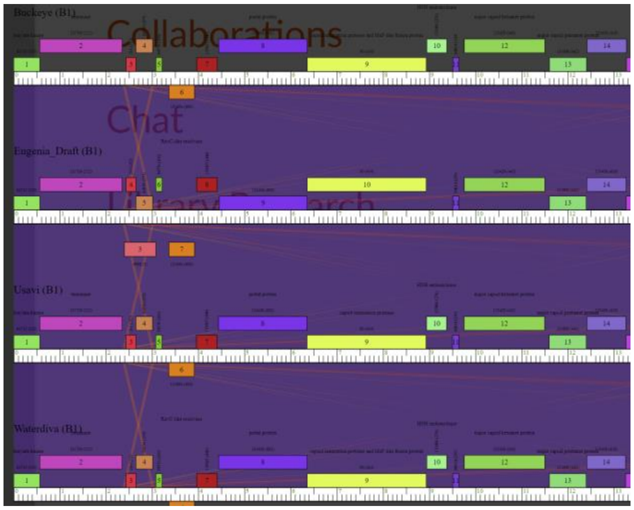
Aligned with LeeLot, Zelda, Haimas, Vaticameos, Waterdiva, Usavi, Buckeye, etc.

N/A



| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | LeeLot | 13 | major capsid pentamer protein | 265 | 550 | 1e-156 | B1 | 137509 |
| ☑ | Zelda | 13 | major capsid pentamer protein | 265 | 550 | 1e-156 | B1 | 137509 |
| ☑ | Haimas | 13 | major capsid pentamer protein | 265 | 548 | 1e-156 | B1 | 137509 |
| ☑ | Vaticameos | 13 | major capsid pentamer protein | 265 | 548 | 1e-156 | B1 | 137509 |
| ☐ | Eugenia_Draft | 13 | function unknown | 265 | 551 | 1e-157 | B1 | 137509 |
| ☐ | Mulan | 12 | major capsid pentamer protein | 265 | 550 | 1e-156 | B1 | 137509 |
| ☐ | Sophia | 12 | major capsid pentamer protein | 265 | 550 | 1e-156 | B1 | 137509 |
| ☐ | Simielle | 12 | major capsid pentamer protein | 265 | 549 | 1e-156 | B1 | 137509 |
| ☐ | Banjo | 12 | major capsid pentamer protein | 265 | 548 | 1e-156 | B1 | 137509 |
| ☐ | Bishoperium | 12 | major capsid pentamer protein | 265 | 548 | 1e-156 | B1 | 137509 |





## 10. Any other important information.

N/A

## CURATOR NAME: ISABELLA LIMA

## GENE NAME: EUGENIA_DRAFT_14

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

### 1. Starterator

Start 8 Track 2

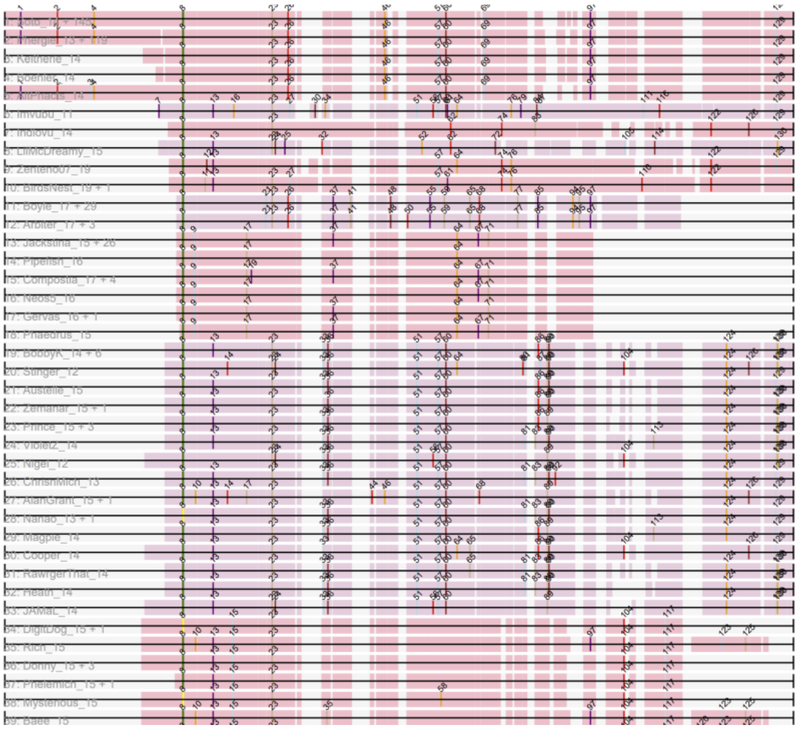
The start number called the most often in the published annotations is 8, it was called in 358 of the 358 non-draft genes in the pham.

Start 8:

• Found in 399 of 399 ( 100.0% ) of genes in pham

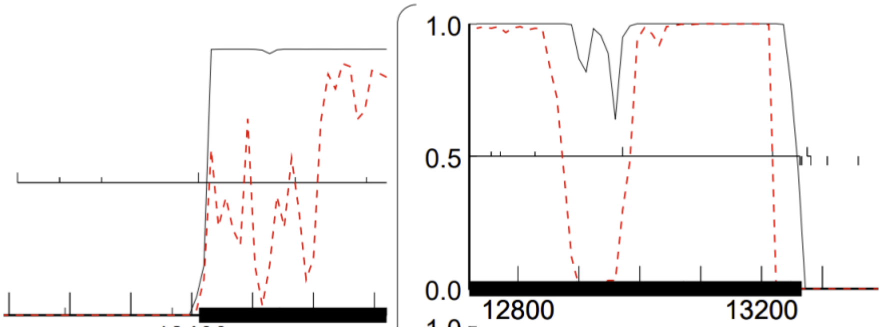
• Manual Annotations of this start: 358 of 358

• Called 100.0% of time when present



## 2. GeneMark coding potential

GeneMark: 12412

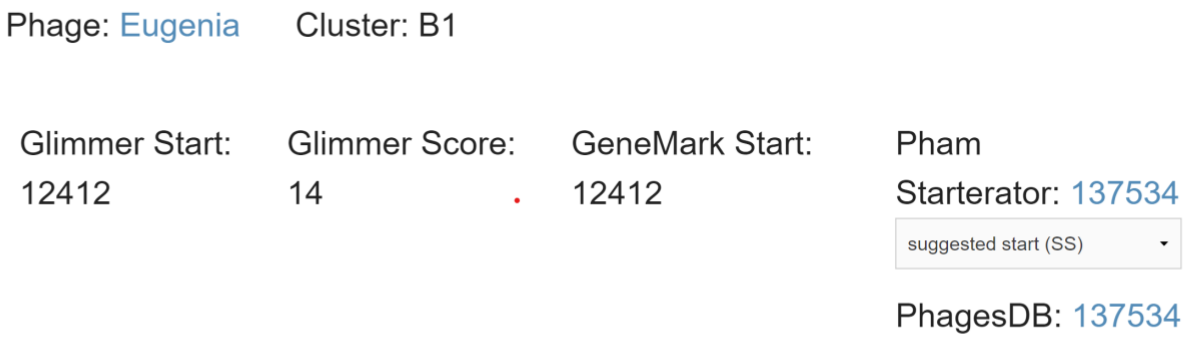


## 3. Glimmer and GeneMark agreement

YES, there is agreement

Glimmer: 12412

GeneMark: 12412



Phage: Eugenia Cluster: B1

Glimmer Start: 12412

Glimmer Score: 14

GeneMark Start: 12412

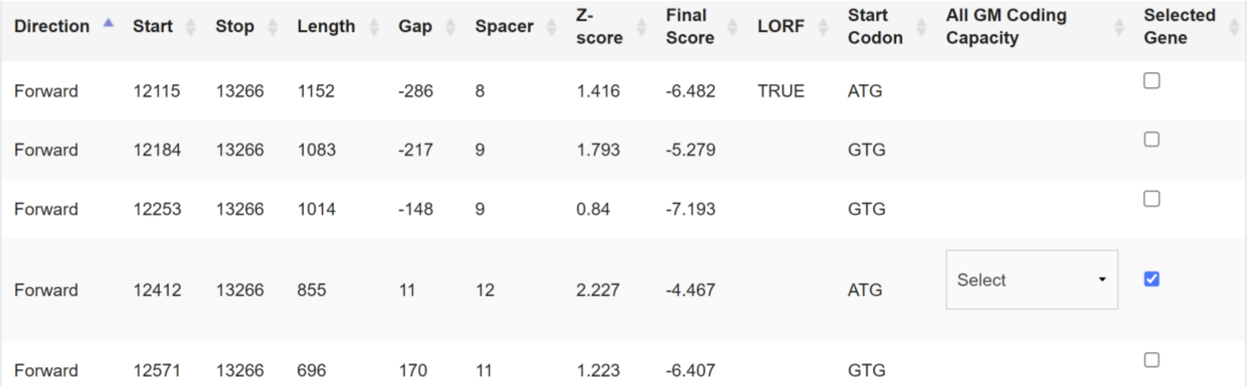
Pham

Starterator: 137534

PhagesDB: 137534

## 4. Longest open reading frame (ORF) without excessive gap

YES



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 12115 | 13266 | 1152 | -286 | 8 | 1.416 | -6.482 | TRUE | ATG | | ☐ |
| Forward | 12184 | 13266 | 1083 | -217 | 9 | 1.793 | -5.279 | | GTG | | ☐ |
| Forward | 12253 | 13266 | 1014 | -148 | 9 | 0.84 | -7.193 | | GTG | | ☐ |
| Forward | 12412 | 13266 | 855 | 11 | 12 | 2.227 | -4.467 | | ATG | Select | ☑ |
| Forward | 12571 | 13266 | 696 | 170 | 11 | 1.223 | -6.407 | | GTG | | ☐ |

Start: 12412

Gap: 11

Z-score: 2.227

Final Score: -4.467

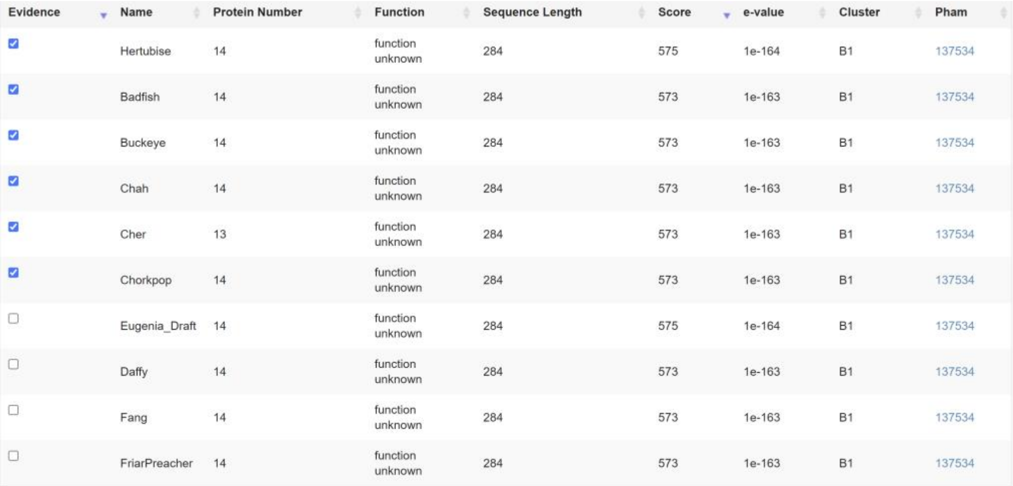
## 5. Function. Please follow this Official SEAPHAGE Function List

**If no functional prediction is present, write "Hypothetical protein".**

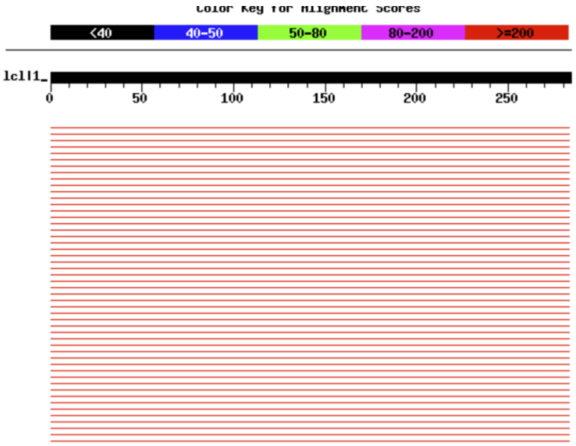
Hypothetical protein

## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST

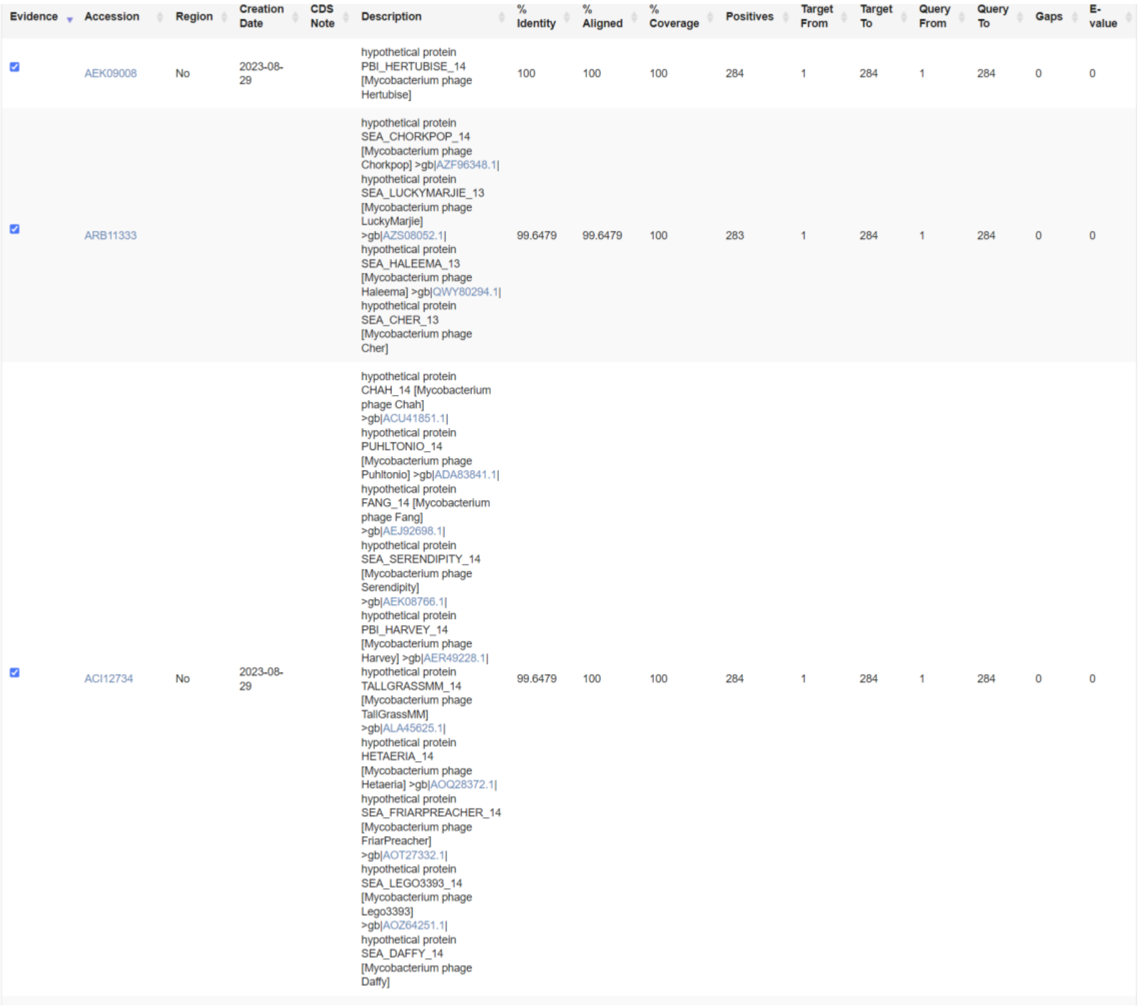


| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | Hertubise | 14 | function unknown | 284 | 575 | 1e-164 | B1 | 137534 |
| ☑ | Badfish | 14 | function unknown | 284 | 573 | 1e-163 | B1 | 137534 |
| ☑ | Buckeye | 14 | function unknown | 284 | 573 | 1e-163 | B1 | 137534 |
| ☑ | Chah | 14 | function unknown | 284 | 573 | 1e-163 | B1 | 137534 |
| ☑ | Cher | 13 | function unknown | 284 | 573 | 1e-163 | B1 | 137534 |
| ☑ | Chorkpop | 14 | function unknown | 284 | 573 | 1e-163 | B1 | 137534 |
| ☐ | Eugenia_Draft | 14 | function unknown | 284 | 575 | 1e-164 | B1 | 137534 |
| ☐ | Daffy | 14 | function unknown | 284 | 573 | 1e-163 | B1 | 137534 |
| ☐ | Fang | 14 | function unknown | 284 | 573 | 1e-163 | B1 | 137534 |
| ☐ | FriarPreacher | 14 | function unknown | 284 | 573 | 1e-163 | B1 | 137534 |



```
                                                                 Score    E
Sequences producing significant alignments:                      (bits) Value

Hertubise_14, function unknown, 284                                   575   e-164
Eugenia_Draft_14, function unknown, 284                               575   e-164
UAch1_14, function unknown, 284                                       573   e-163
TallGrassMM_14, function unknown, 284                                 573   e-163
Serendipity_14, function unknown, 284                                 573   e-163
Selr12_Draft_14, function unknown, 284                                573   e-163
RedMaple_13, function unknown, 284                                    573   e-163
Puhltonio_14, function unknown, 284                                   573   e-163
Lego3393_14, function unknown, 284                                    573   e-163
Kloppinator_14, function unknown, 284                                 573   e-163
Katniss_14, function unknown, 284                                     573   e-163
Hetaeria_14, function unknown, 284                                    573   e-163
Harvey_14, function unknown, 284                                      573   e-163
FriarPreacher_14, function unknown, 284                               573   e-163
Fang_14, function unknown, 284                                        573   e-163
Daffy_14, function unknown, 284                                       573   e-163
Chah_14, function unknown, 284                                        573   e-163
Ricotta_Draft_14, function unknown, 284                               573   e-163
LuckyMarjie_13, function unknown, 284                                 573   e-163
Haleema_13, function unknown, 284                                     573   e-163
```

## b. SIF: NCBI BLAST



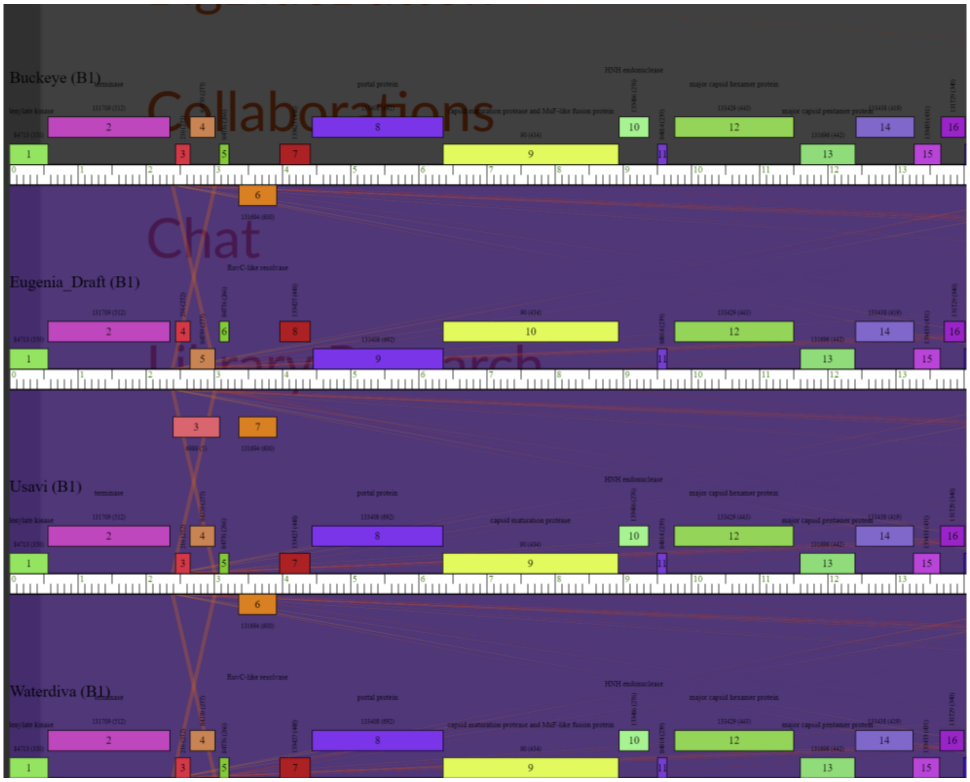
| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | AEK09008 | No | 2023-08-29 | | hypothetical protein PBI_HERTUBISE_14 [Mycobacterium phage Hertubise] | 100 | 100 | 100 | 284 | 1 | 284 | 1 | 284 | 0 | 0 |
| ☑ | ARB11333 | | | | hypothetical protein SEA_CHORKPOP_14 [Mycobacterium phage Chorkpop] >gb\|AZF96348.1\| hypothetical protein SEA_LUCKYMARJIE_13 [Mycobacterium phage LuckyMarjie] >gb\|AZS08052.1\| hypothetical protein SEA_HALEEMA_13 [Mycobacterium phage Haleema] >gb\|QWY80294.1\| hypothetical protein SEA_CHER_13 [Mycobacterium phage Cher] | 99.6479 | 99.6479 | 100 | 283 | 1 | 284 | 1 | 284 | 0 | 0 |
| ☑ | ACI12734 | No | 2023-08-29 | | hypothetical protein CHAH_14 [Mycobacterium phage Chah] >gb\|ACU41851.1\| hypothetical protein PUHLTONIO_14 [Mycobacterium phage Puhltonio] >gb\|ADA83841.1\| hypothetical protein FANG_14 [Mycobacterium phage Fang] >gb\|AEJ92698.1\| hypothetical protein SEA_SERENDIPITY_14 [Mycobacterium phage Serendipity] >gb\|AEK08766.1\| hypothetical protein PBI_HARVEY_14 [Mycobacterium phage Harvey] >gb\|AER49228.1\| hypothetical protein TALLGRASSMM_14 [Mycobacterium phage TallGrassMM] >gb\|ALA45625.1\| hypothetical protein HETAERIA_14 [Mycobacterium phage Hetaeria] >gb\|AOQ28372.1\| hypothetical protein SEA_FRIARPREACHER_14 [Mycobacterium phage FriarPreacher] >gb\|AOT27332.1\| hypothetical protein SEA_LEGO3393_14 [Mycobacterium phage Lego3393] >gb\|AOZ64251.1\| hypothetical protein SEA_DAFFY_14 [Mycobacterium phage Daffy] | 99.6479 | 100 | 100 | 284 | 1 | 284 | 1 | 284 | 0 | 0 |

## c. SIF: HHPred

N/A

## d. SIF: Synteny-Phamerator (three genomes)

Aligns with Buckeye, Waterdiva, Usavi, Hertubise, Chorkpop, etc.



## 7. Any other important information.

N/A

## CURATOR NAME: ISABELLA LIMA

## GENE NAME: EUGENIA_DRAFT_15

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

### 1. Starterator

Eugenia is in track 2 with start site 42.

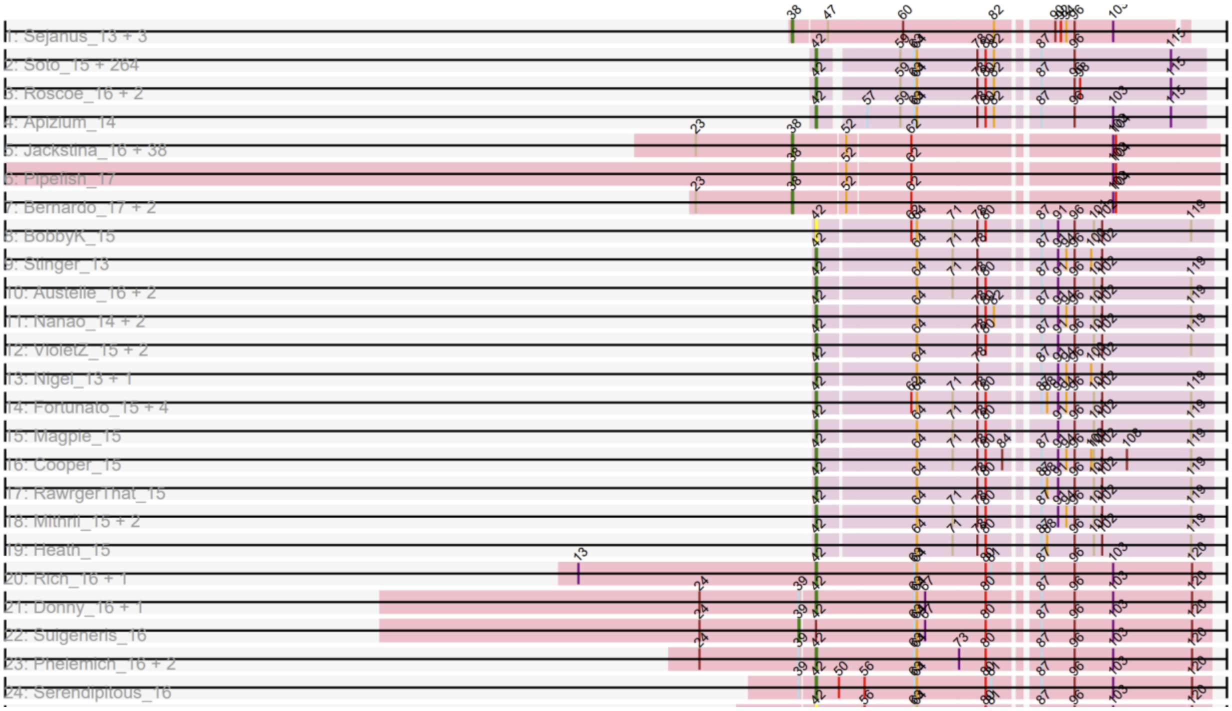
The start number called the most often in the published annotations is 42, it was called in 273 of the 388 non-draft genes in the pham.

Start 42:

- Found in 304 of 429 ( 70.9% ) of genes in pham

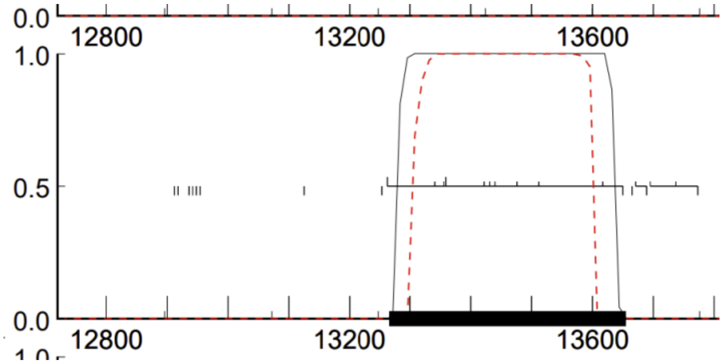
- Manual Annotations of this start: 273 of 388

- Called 99.7% of time when present



## 2. GeneMark coding potential

GeneMark: 13266

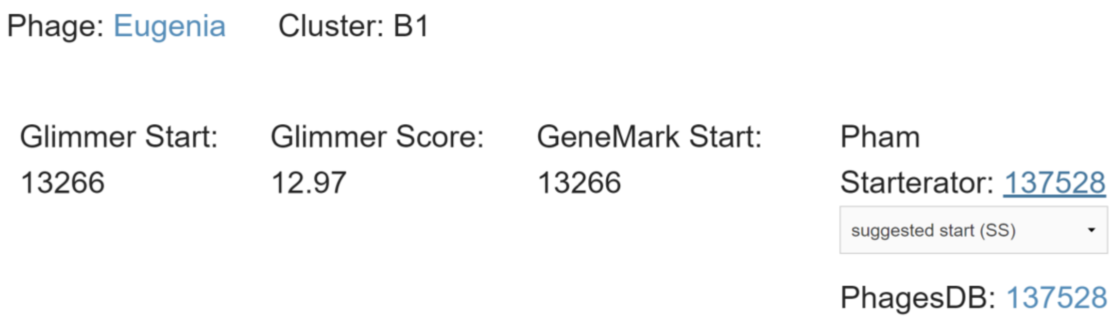


## 4. Glimmer and GeneMark agreement

YES

Glimmer: 13266

GeneMark: 13266


Phage: Eugenia Cluster: B1

Glimmer Start: 13266

Glimmer Score: 12.97

GeneMark Start: 13266

Pham Starterator: 137528

PhagesDB: 137528

## 4. Longest open reading frame (ORF) without excessive gap

YES

| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 13266 | 13655 | 390 | -1 | 12 | 2.691 | -3.535 | TRUE | ATG | Select | ☑ |
| Forward | 13344 | 13655 | 312 | 77 | 8 | 2.012 | -5.285 |  | GTG |  | ☐ |

Length: 390

Gap: -1

Z-score: 2.691

Final Score: -3.535

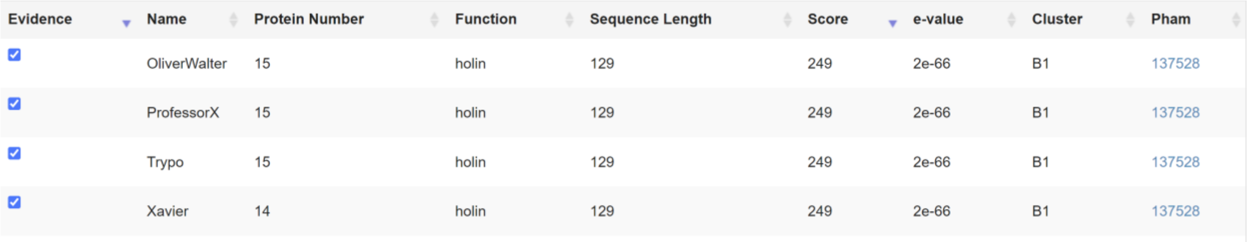
## 5. Function. Please follow this Official SEAPHAGE Function List

**If no functional prediction is present, write "Hypothetical protein".**

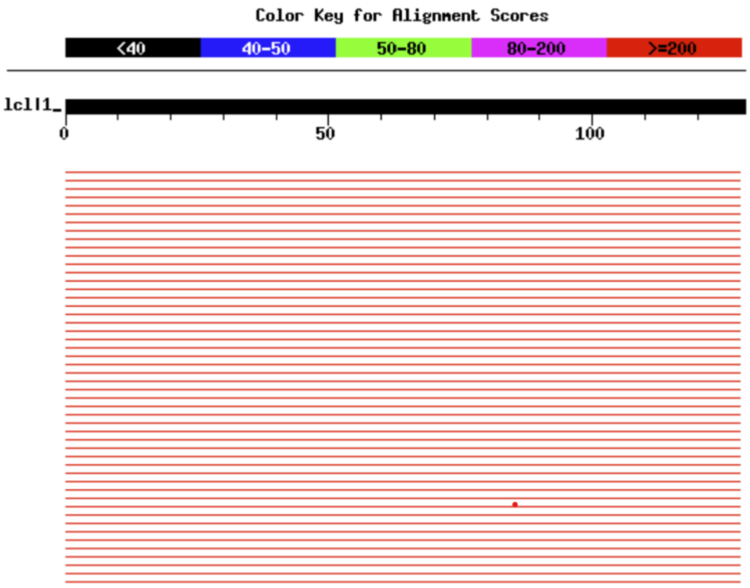
Holin (Rank 2)

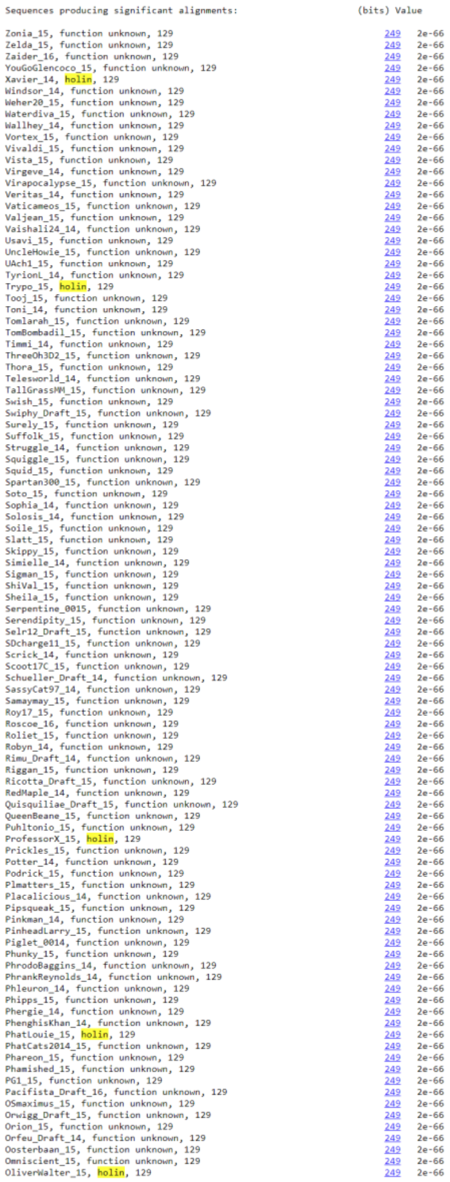
## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST



| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | OliverWalter | 15 | holin | 129 | 249 | 2e-66 | B1 | 137528 |
| ☑ | ProfessorX | 15 | holin | 129 | 249 | 2e-66 | B1 | 137528 |
| ☑ | Trypo | 15 | holin | 129 | 249 | 2e-66 | B1 | 137528 |
| ☑ | Xavier | 14 | holin | 129 | 249 | 2e-66 | B1 | 137528 |



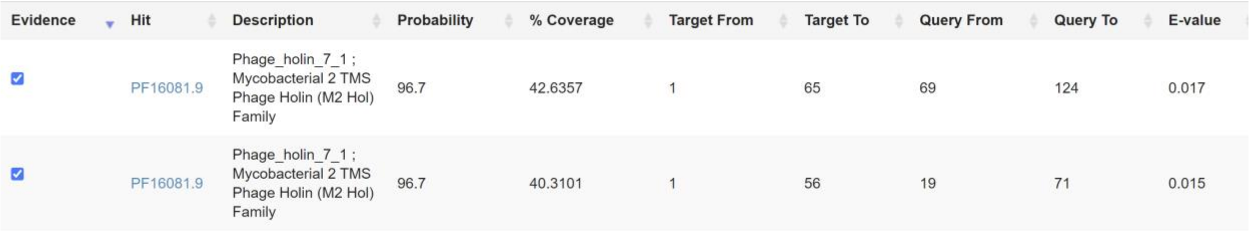


```
Sequences producing significant alignments:                      (bits) Value

Zonia_15, function unknown, 129                                        249   2e-66
Zelda_15, function unknown, 129                                        249   2e-66
Zaider_16, function unknown, 129                                       249   2e-66
YouGoGlencoco_15, function unknown, 129                                249   2e-66
Xavier_14, holin, 129                                                  249   2e-66
Windsor_14, function unknown, 129                                      249   2e-66
Weher20_15, function unknown, 129                                      249   2e-66
Waterdiva_15, function unknown, 129                                    249   2e-66
Wallhey_14, function unknown, 129                                      249   2e-66
Vortex_15, function unknown, 129                                       249   2e-66
Vivaldi_15, function unknown, 129                                      249   2e-66
Vista_15, function unknown, 129                                        249   2e-66
Virgeve_14, function unknown, 129                                      249   2e-66
Virapocalypse_15, function unknown, 129                                249   2e-66
Veritas_14, function unknown, 129                                      249   2e-66
Vaticameos_15, function unknown, 129                                   249   2e-66
Valjean_15, function unknown, 129                                      249   2e-66
Vaishali24_14, function unknown, 129                                   249   2e-66
Usavi_15, function unknown, 129                                        249   2e-66
UncleHowie_15, function unknown, 129                                   249   2e-66
UAch1_15, function unknown, 129                                        249   2e-66
TyrionL_14, function unknown, 129                                      249   2e-66
Trypo_15, holin, 129                                                   249   2e-66
Tooj_15, function unknown, 129                                         249   2e-66
Toni_14, function unknown, 129                                         249   2e-66
Tomlarah_15, function unknown, 129                                     249   2e-66
TomBombadil_15, function unknown, 129                                  249   2e-66
Timmi_14, function unknown, 129                                        249   2e-66
ThreeOh3D2_15, function unknown, 129                                   249   2e-66
Thora_15, function unknown, 129                                        249   2e-66
Telesworld_14, function unknown, 129                                   249   2e-66
TallGrassMM_15, function unknown, 129                                  249   2e-66
Swish_15, function unknown, 129                                        249   2e-66
Swiphy_Draft_15, function unknown, 129                                 249   2e-66
Surely_15, function unknown, 129                                       249   2e-66
Suffolk_15, function unknown, 129                                      249   2e-66
Struggle_14, function unknown, 129                                     249   2e-66
Squiggle_15, function unknown, 129                                     249   2e-66
Squid_15, function unknown, 129                                        249   2e-66
Spartan300_15, function unknown, 129                                   249   2e-66
Soto_15, function unknown, 129                                         249   2e-66
Sophia_14, function unknown, 129                                       249   2e-66
Solosis_14, function unknown, 129                                      249   2e-66
Soile_15, function unknown, 129                                        249   2e-66
Slatt_15, function unknown, 129                                        249   2e-66
Skippy_15, function unknown, 129                                       249   2e-66
Simielle_14, function unknown, 129                                     249   2e-66
Sigman_15, function unknown, 129                                       249   2e-66
ShiVal_15, function unknown, 129                                       249   2e-66
Sheila_15, function unknown, 129                                       249   2e-66
Serpentine_0015, function unknown, 129                                 249   2e-66
Serendipity_15, function unknown, 129                                  249   2e-66
Selr12_Draft_15, function unknown, 129                                 249   2e-66
SDcharge11_15, function unknown, 129                                   249   2e-66
Scrick_14, function unknown, 129                                       249   2e-66
Scoot17C_15, function unknown, 129                                     249   2e-66
Schueller_Draft_14, function unknown, 129                              249   2e-66
SassyCat97_14, function unknown, 129                                   249   2e-66
Samaymay_15, function unknown, 129                                     249   2e-66
Roy17_15, function unknown, 129                                        249   2e-66
Roscoe_16, function unknown, 129                                       249   2e-66
Roliet_15, function unknown, 129                                       249   2e-66
Robyn_14, function unknown, 129                                        249   2e-66
Rimu_Draft_14, function unknown, 129                                   249   2e-66
Riggan_15, function unknown, 129                                       249   2e-66
Ricotta_Draft_15, function unknown, 129                                249   2e-66
RedMaple_14, function unknown, 129                                     249   2e-66
Quisquiliae_Draft_15, function unknown, 129                            249   2e-66
QueenBeane_15, function unknown, 129                                   249   2e-66
Puhltonio_15, function unknown, 129                                    249   2e-66
ProfessorX_15, holin, 129                                              249   2e-66
Prickles_15, function unknown, 129                                     249   2e-66
Potter_14, function unknown, 129                                       249   2e-66
Podrick_15, function unknown, 129                                      249   2e-66
Plmatters_15, function unknown, 129                                    249   2e-66
Placalicious_14, function unknown, 129                                 249   2e-66
Pipsqueak_15, function unknown, 129                                    249   2e-66
Pinkman_14, function unknown, 129                                      249   2e-66
PinheadLarry_15, function unknown, 129                                 249   2e-66
Piglet_0014, function unknown, 129                                     249   2e-66
Phunky_15, function unknown, 129                                       249   2e-66
PhrodoBaggins_14, function unknown, 129                                249   2e-66
PhrankReynolds_14, function unknown, 129                               249   2e-66
Phleuron_14, function unknown, 129                                     249   2e-66
Phipps_15, function unknown, 129                                       249   2e-66
Phergie_14, function unknown, 129                                      249   2e-66
PhenghisKhan_14, function unknown, 129                                 249   2e-66
PhatLouie_15, holin, 129                                               249   2e-66
PhatCats2014_15, function unknown, 129                                 249   2e-66
Phareon_15, function unknown, 129                                      249   2e-66
Phamished_15, function unknown, 129                                    249   2e-66
PG1_15, function unknown, 129                                          249   2e-66
Pacifista_Draft_16, function unknown, 129                              249   2e-66
OSmaximus_15, function unknown, 129                                    249   2e-66
Orwigg_Draft_15, function unknown, 129                                 249   2e-66
Orion_15, function unknown, 129                                        249   2e-66
Orfeu_Draft_14, function unknown, 129                                  249   2e-66
Oosterbaan_15, function unknown, 129                                   249   2e-66
Omniscient_15, function unknown, 129                                   249   2e-66
OliverWalter_15, holin, 129                                            249   2e-66
```

## b. SIF: NCBI BLAST

N/A

## c. SIF: HHPred



| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | PF16081.9 | Phage_holin_7_1 ; Mycobacterial 2 TMS Phage Holin (M2 Hol) Family | 96.7 | 42.6357 | 1 | 65 | 69 | 124 | 0.017 |
| ☑ | PF16081.9 | Phage_holin_7_1 ; Mycobacterial 2 TMS Phage Holin (M2 Hol) Family | 96.7 | 40.3101 | 1 | 56 | 19 | 71 | 0.015 |

## d. SIF: Synteny-Phamerator (three genomes)

Buckeye, Usavi, Waterdiva, OliverWalter, ProfessorX, etc.

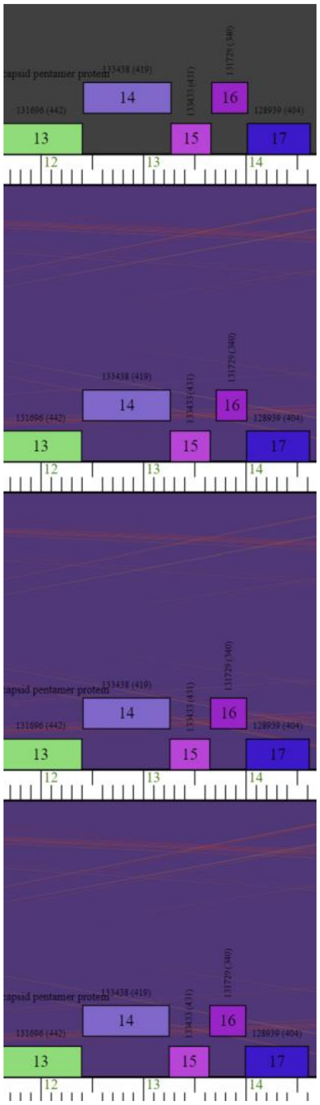


*Figure 1. (Top to bottom) Buckeye, Eugenia, Usavi, and Waterdiva.*

## 7. Any other important information.

Many phages have this gene labelled as "unknown function," while only some have their function listed as "holin." Dr. Phuntumart said this is likely because these phages with unknown Gene 15 functions haven't been updated yet.

## CURATOR NAME: ISABELLA LIMA

## GENE NAME: EUGENIA_DRAFT_16

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

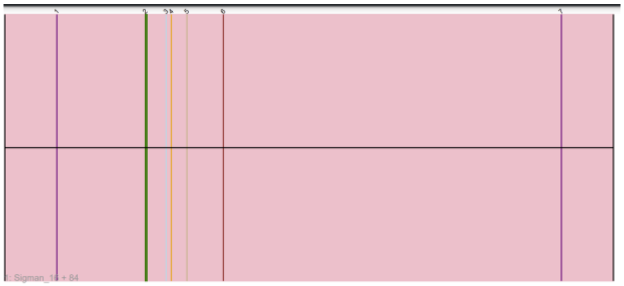
### 1. Starterator

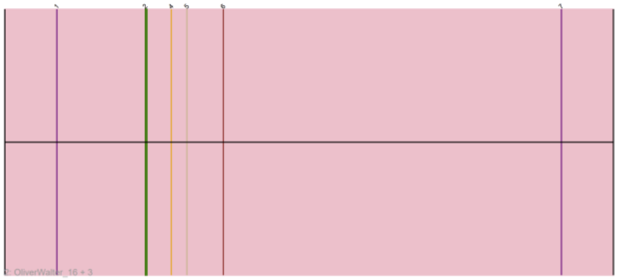
Eugenia is in track 1 with start 2.

The start number called the most often in the published annotations is 2, it was called in 66 of the 66 non-draft genes in the pham.

Start 2:

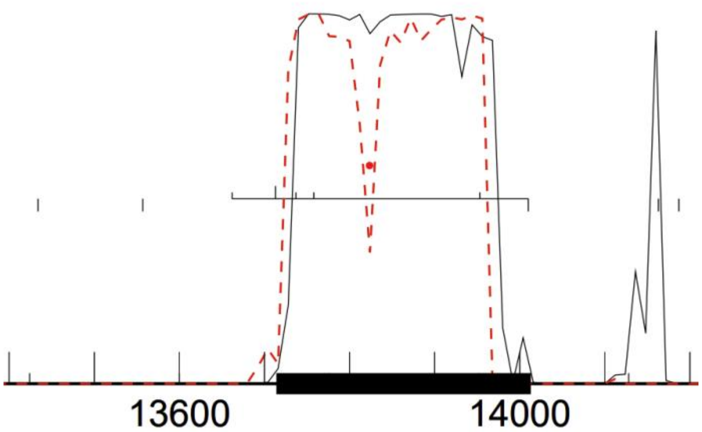
- Found in 89 of 89 ( 100.0% ) of genes in pham
- Manual Annotations of this start: 66 of 66
- Called 100.0% of time when present





## 2. GeneMark coding potential

GeneMark: 13714

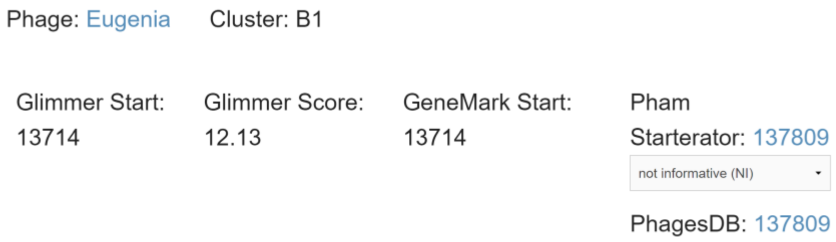


## 3. Glimmer and GeneMark agreement

YES

Glimmer: 13714

GeneMark: 13714



Phage: Eugenia    Cluster: B1

Glimmer Start: 13714

Glimmer Score: 12.13

GeneMark Start: 13714

Pham

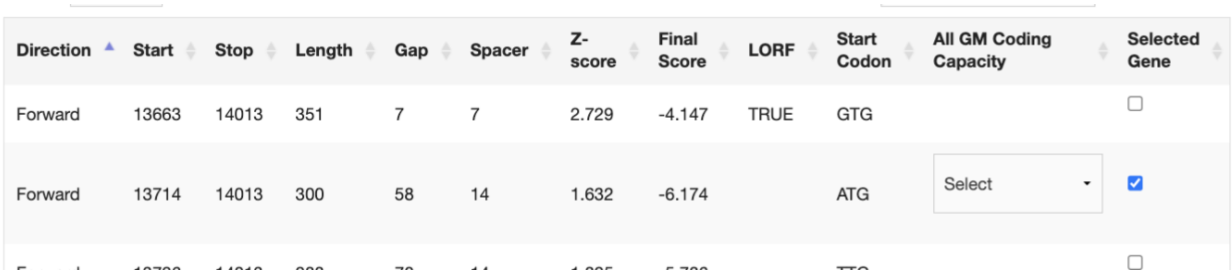
Starterator: 137809

not informative (NI)

PhagesDB: 137809

## 4. Longest open reading frame (ORF) without excessive gap

No



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 13663 | 14013 | 351 | 7 | 7 | 2.729 | -4.147 | TRUE | GTG |  | ☐ |
| Forward | 13714 | 14013 | 300 | 58 | 14 | 1.632 | -6.174 |  | ATG | Select | ☑ |

## 5. Function. Please follow this [Official SEAPHAGE Function List]()

## If no functional prediction is present, write “Hypothetical protein”.

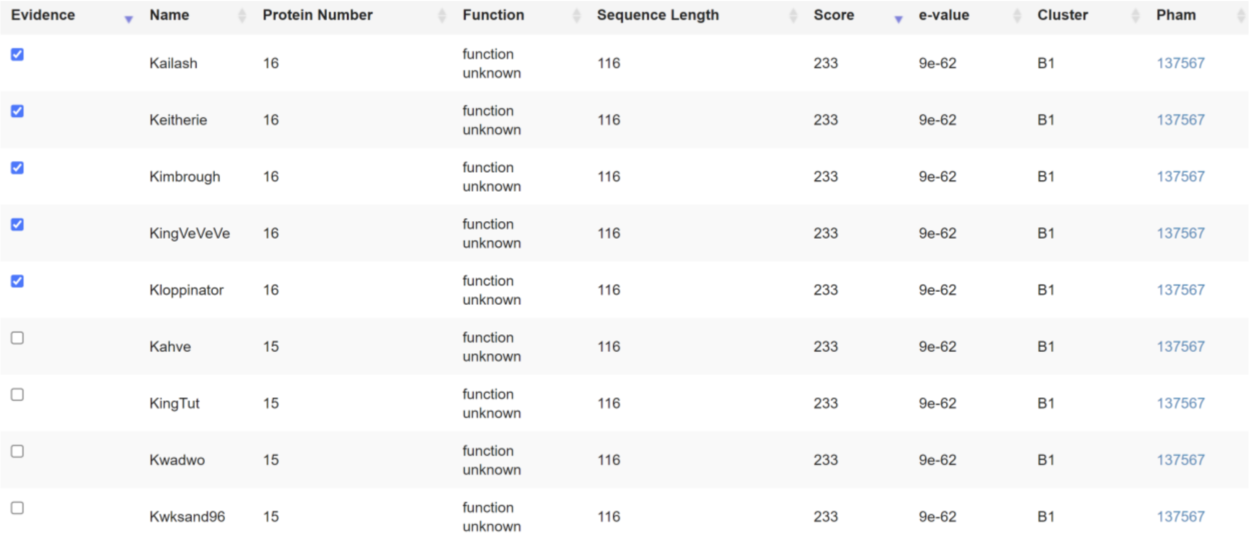
Membrane protein

TmHmm: 2

## 6. Supporting Information for Function (SIF)

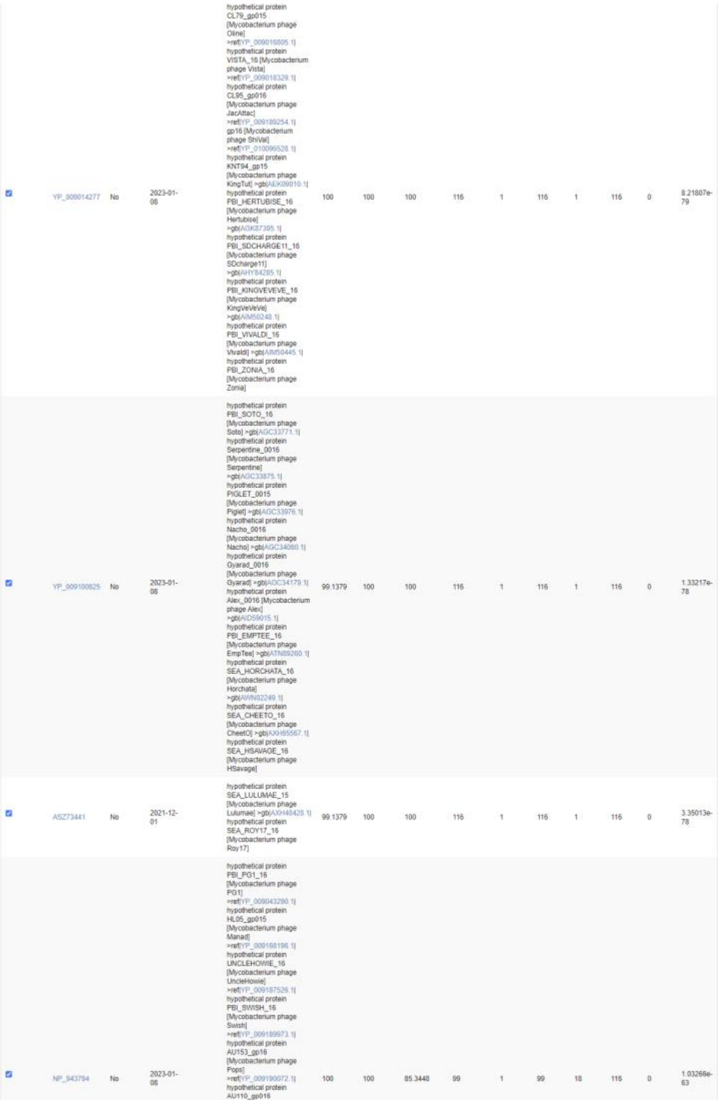
### a. SIF: PhageDb BLAST

```
                                                                  Score    E
Sequences producing significant alignments:                       (bits) Value

Zonia_16, function unknown, 116                                     194   8e-50
Zelda_16, function unknown, 116                                     194   8e-50
Zaider_17, function unknown, 116                                    194   8e-50
YouGoGlencoco_16, function unknown, 116                             194   8e-50
Yoshand_16, function unknown, 99                                    194   8e-50
Xavier_15, function unknown, 116                                    194   8e-50
Windsor_15, function unknown, 116                                   194   8e-50
Weher20_16, function unknown, 116                                   194   8e-50
Waterdiva_16, function unknown, 116                                 194   8e-50
Wallhey_15, function unknown, 116                                   194   8e-50
Vortex_16, function unknown, 99                                     194   8e-50
Vivaldi_16, function unknown, 116                                   194   8e-50
Vista_16, function unknown, 116                                     194   8e-50
Virgeve_15, function unknown, 99                                    194   8e-50
Virapocalypse_16, function unknown, 116                             194   8e-50
Veritas_15, function unknown, 116                                   194   8e-50
Vaticameos_16, function unknown, 116                                194   8e-50
Valjean_16, function unknown, 116                                   194   8e-50
Vaishali24_15, function unknown, 99                                 194   8e-50
Usavi_16, function unknown, 116                                     194   8e-50
UncleHowie_16, function unknown, 99                                 194   8e-50
UAch1_16, function unknown, 116                                     194   8e-50
TyrionL_15, function unknown, 116                                   194   8e-50
Trypo_16, function unknown, 116                                     194   8e-50
True_15, function unknown, 116                                      194   8e-50
Tooj_16, function unknown, 116                                      194   8e-50
Toni_15, function unknown, 116                                      194   8e-50
Tomlarah_16, function unknown, 116                                  194   8e-50
TomBombadil_16, function unknown, 116                               194   8e-50
Timmi_15, function unknown, 116                                     194   8e-50
ThreeOh3D2_16, function unknown, 99                                 194   8e-50
Thora_16, function unknown, 99                                      194   8e-50
Telesworld_15, function unknown, 116                                194   8e-50
TallGrassMM_16, function unknown, 99                                194   8e-50
Swish_16, function unknown, 99                                      194   8e-50
Swiphy_Draft_16, function unknown, 99                               194   8e-50
Surely_16, function unknown, 116                                    194   8e-50
Struggle_15, function unknown, 116                                  194   8e-50
Squiggle_16, function unknown, 116                                  194   8e-50
Squid_16, function unknown, 116                                     194   8e-50
Sophia_15, function unknown, 116                                    194   8e-50
Solosis_15, function unknown, 116                                   194   8e-50
Soile_16, function unknown, 116                                     194   8e-50
Slatt_16, function unknown, 116                                     194   8e-50
```



| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | Kailash | 16 | function unknown | 116 | 233 | 9e-62 | B1 | 137567 |
| ☑ | Keitherie | 16 | function unknown | 116 | 233 | 9e-62 | B1 | 137567 |
| ☑ | Kimbrough | 16 | function unknown | 116 | 233 | 9e-62 | B1 | 137567 |
| ☑ | KingVeVeVe | 16 | function unknown | 116 | 233 | 9e-62 | B1 | 137567 |
| ☑ | Kloppinator | 16 | function unknown | 116 | 233 | 9e-62 | B1 | 137567 |
| ☐ | Kahve | 15 | function unknown | 116 | 233 | 9e-62 | B1 | 137567 |
| ☐ | KingTut | 15 | function unknown | 116 | 233 | 9e-62 | B1 | 137567 |
| ☐ | Kwadwo | 15 | function unknown | 116 | 233 | 9e-62 | B1 | 137567 |
| ☐ | Kwksand96 | 15 | function unknown | 116 | 233 | 9e-62 | B1 | 137567 |

## b. SIF: NCBI BLAST



| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | YP_009014277 | No | 2023-01-08 | hypothetical protein CL79_gp015 [Mycobacterium phage Oline] >ref\|YP_009016805.1\| hypothetical protein VISTA_16 [Mycobacterium phage Vista] >ref\|YP_009018329.1\| hypothetical protein CL95_gp016 [Mycobacterium phage JacAttac] >ref\|YP_009189254.1\| gp16 [Mycobacterium phage ShiVal] >ref\|YP_010096528.1\| hypothetical protein KNT94_gp15 [Mycobacterium phage KingTut] >gb\|AEK09010.1\| hypothetical protein PBI_HERTUBISE_16 [Mycobacterium phage Hertubise] >gb\|AGK87395.1\| hypothetical protein PBI_SDCHARGE11_16 [Mycobacterium phage SDcharge11] >gb\|AHY84285.1\| hypothetical protein PBI_KINGVEVEVE_16 [Mycobacterium phage KingVeVeVe] >gb\|AIM50248.1\| hypothetical protein PBI_VIVALDI_16 [Mycobacterium phage Vivaldi] >gb\|AIM50445.1\| hypothetical protein PBI_ZONIA_16 [Mycobacterium phage Zonia] | 100 | 100 | 100 | 116 | 1 | 116 | 1 | 116 | 0 | 8.21807e-79 |
| ☑ | YP_009100825 | No | 2023-01-08 | hypothetical protein PBI_SOTO_16 [Mycobacterium phage Soto] >gb\|AGC33771.1\| hypothetical protein Serpentine_0016 [Mycobacterium phage Serpentine] >gb\|AGC33875.1\| hypothetical protein PIGLET_0015 [Mycobacterium phage Piglet] >gb\|AGC33976.1\| hypothetical protein Nacho_0016 [Mycobacterium phage Nacho] >gb\|AGC34080.1\| hypothetical protein Gyarad_0016 [Mycobacterium phage Gyarad] >gb\|AGC34179.1\| hypothetical protein Alex_0016 [Mycobacterium phage Alex] >gb\|AID59015.1\| hypothetical protein PBI_EMPTEE_16 [Mycobacterium phage EmpTee] >gb\|ATN89260.1\| hypothetical protein SEA_HORCHATA_16 [Mycobacterium phage Horchata] >gb\|AWN02249.1\| hypothetical protein SEA_CHEETO_16 [Mycobacterium phage CheetO] >gb\|AXH65567.1\| hypothetical protein SEA_HSAVAGE_16 [Mycobacterium phage HSavage] | 99.1379 | 100 | 100 | 116 | 1 | 116 | 1 | 116 | 0 | 1.33217e-78 |
| ☑ | ASZ73441 | No | 2021-12-01 | hypothetical protein SEA_LULUMAE_15 [Mycobacterium phage Lulumae] >gb\|AXH48428.1\| hypothetical protein SEA_ROY17_16 [Mycobacterium phage Roy17] | 99.1379 | 100 | 100 | 116 | 1 | 116 | 1 | 116 | 0 | 3.35013e-78 |
| ☑ | NP_943794 | No | 2023-01-08 | hypothetical protein PBI_PG1_16 [Mycobacterium phage PG1] >ref\|YP_009043290.1\| hypothetical protein HL05_gp015 [Mycobacterium phage Manad] >ref\|YP_009168196.1\| hypothetical protein UNCLEHOWIE_16 [Mycobacterium phage UncleHowie] >ref\|YP_009187526.1\| hypothetical protein PBI_SWISH_16 [Mycobacterium phage Swish] >ref\|YP_009189973.1\| hypothetical protein AU153_gp16 [Mycobacterium phage Pops] >ref\|YP_009190072.1\| hypothetical protein AU110_gp016 | 100 | 100 | 85.3448 | 99 | 1 | 99 | 18 | 116 | 0 | 1.03266e-63 |

## c. SIF: HHPred

N/A

## d. SIF: Synteny-Phamerator (three genomes)

Buckeye, Usavi, Waterdiva, Kailash, etc.

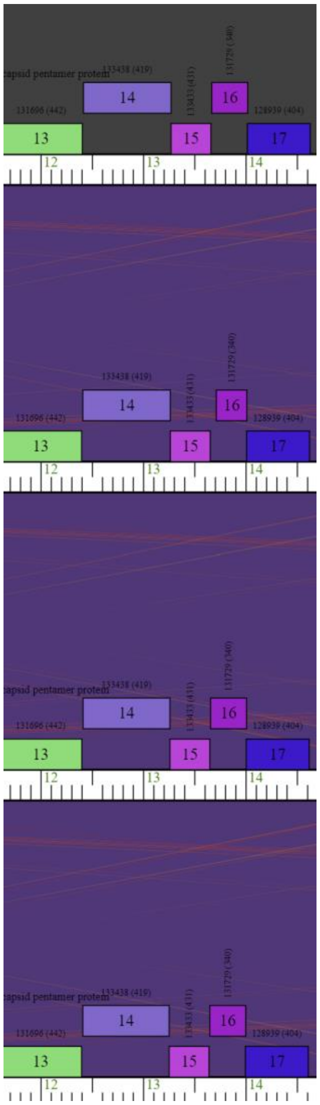


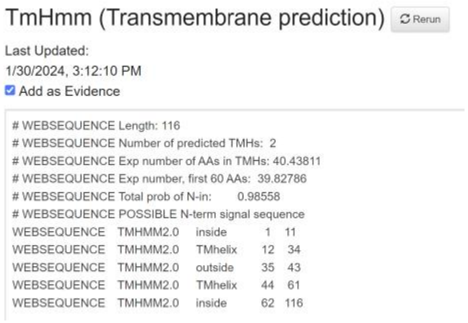
*Figure 1. (Top to bottom) Buckeye, Eugenia, Usavi, and Waterdiva.*

## 7. Any other important information.

Genemark and Glimmers’ call for start sight disagree with Pecaan’s call. However, due to PhageDB BLAST, Pecaan is possibly more correct.

While initially a hypothetical protein, the TmHmm prediction shows that Gene 16 is possibly a membrane protein. This means many other phages in this cluster’s Gene 16 could be membrane proteins as well instead of hypothetical.

TmHmm: 2



TmHmm (Transmembrane prediction)

Last Updated:
1/30/2024, 3:12:10 PM

☑ Add as Evidence

```
# WEBSEQUENCE Length: 116
# WEBSEQUENCE Number of predicted TMHs:  2
# WEBSEQUENCE Exp number of AAs in TMHs: 40.43811
# WEBSEQUENCE Exp number, first 60 AAs:  39.82786
# WEBSEQUENCE Total prob of N-in:        0.98558
# WEBSEQUENCE POSSIBLE N-term signal sequence
WEBSEQUENCE    TMHMM2.0    inside        1    11
WEBSEQUENCE    TMHMM2.0    TMhelix      12    34
WEBSEQUENCE    TMHMM2.0    outside      35    43
WEBSEQUENCE    TMHMM2.0    TMhelix      44    61
WEBSEQUENCE    TMHMM2.0    inside       62   116
```

## CURATOR NAME: ISABELLA LIMA

## GENE NAME: EUGENIA_DRAFT_17

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

# 1. Starterator

Eugenia is in Track 4 with Start 5.

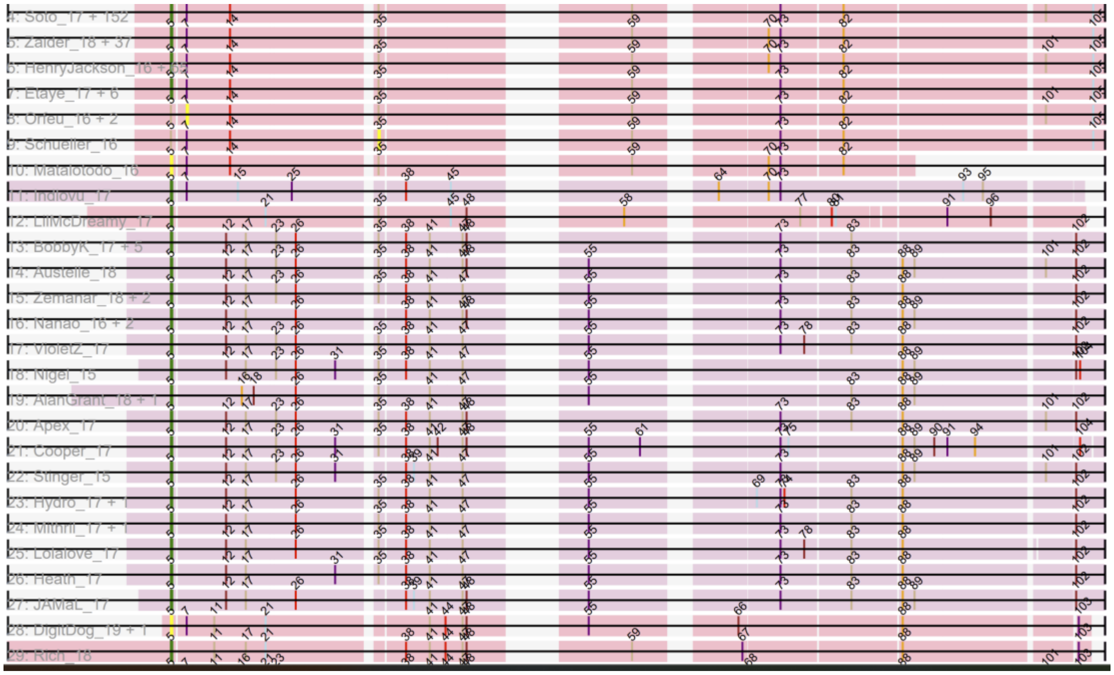
The start number called the most often in the published annotations is 5, it was called in 313 of the 374 non-draft genes in the pham.

Start 5:

- Found in 351 of 413 ( 85.0% ) of genes in pham

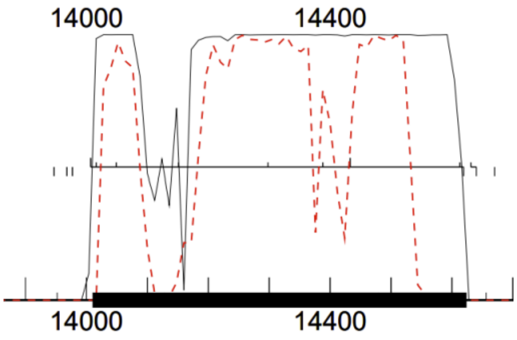
- Manual Annotations of this start: 313 of 374

- Called 97.7% of time when present



## 2. GeneMark coding potential

GeneMark: 14010

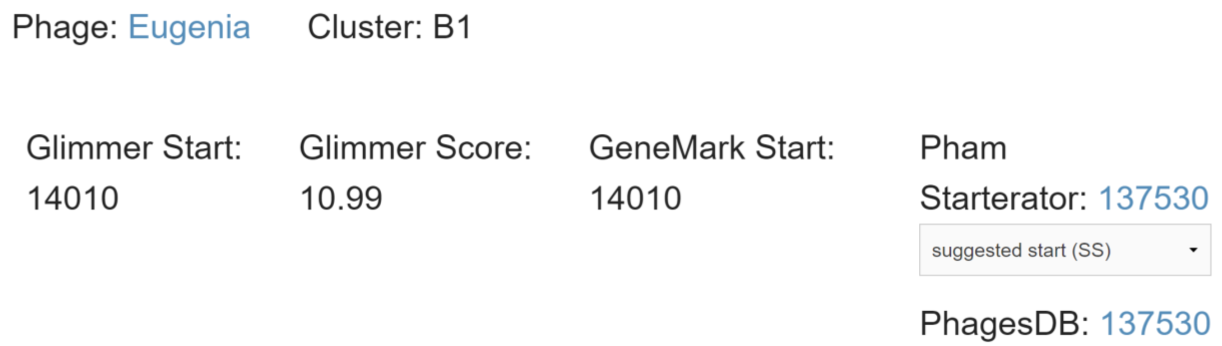


## 3. Glimmer and GeneMark agreement

YES

Glimmer: 14010

GeneMark: 14010



Phage: Eugenia Cluster: B1

Glimmer Start: 14010

Glimmer Score: 10.99

GeneMark Start: 14010

Pham Starterator: 137530

suggested start (SS)

PhagesDB: 137530

## 4. Longest open reading frame (ORF) without excessive gap

YES

| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 14010 | 14624 | 615 | -4 | 11 | 2.16 | -4.523 | TRUE | ATG | Select | ☑ |
| Forward | 14019 | 14624 | 606 | 5 | 8 | 2.138 | -5.032 |  | GTG |  | ☐ |

Start: 14010

Length: 615

Gap: -4

Z-score: 2.16

Final Score: -4.523

## 5. Function. Please follow this Official SEAPHAGE Function List

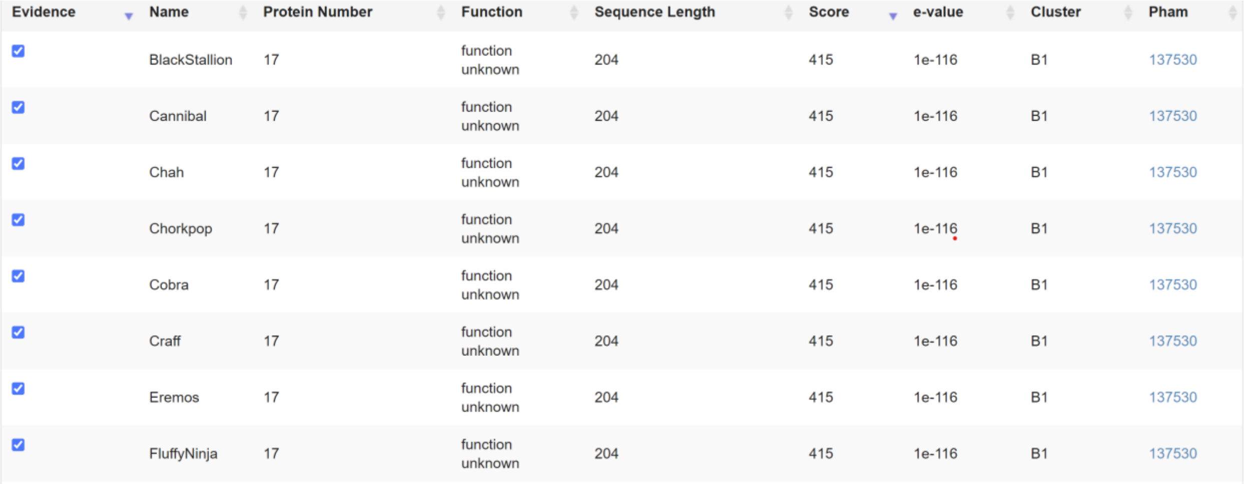
**If no functional prediction is present, write "Hypothetical protein".**

Membrane protein

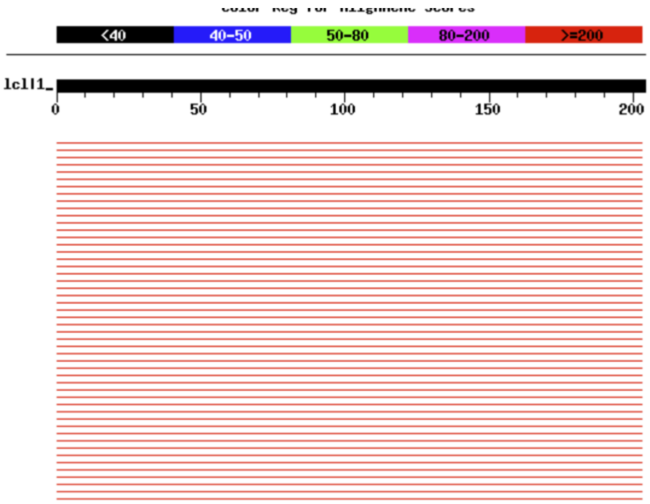
TmHmm: 2

## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST

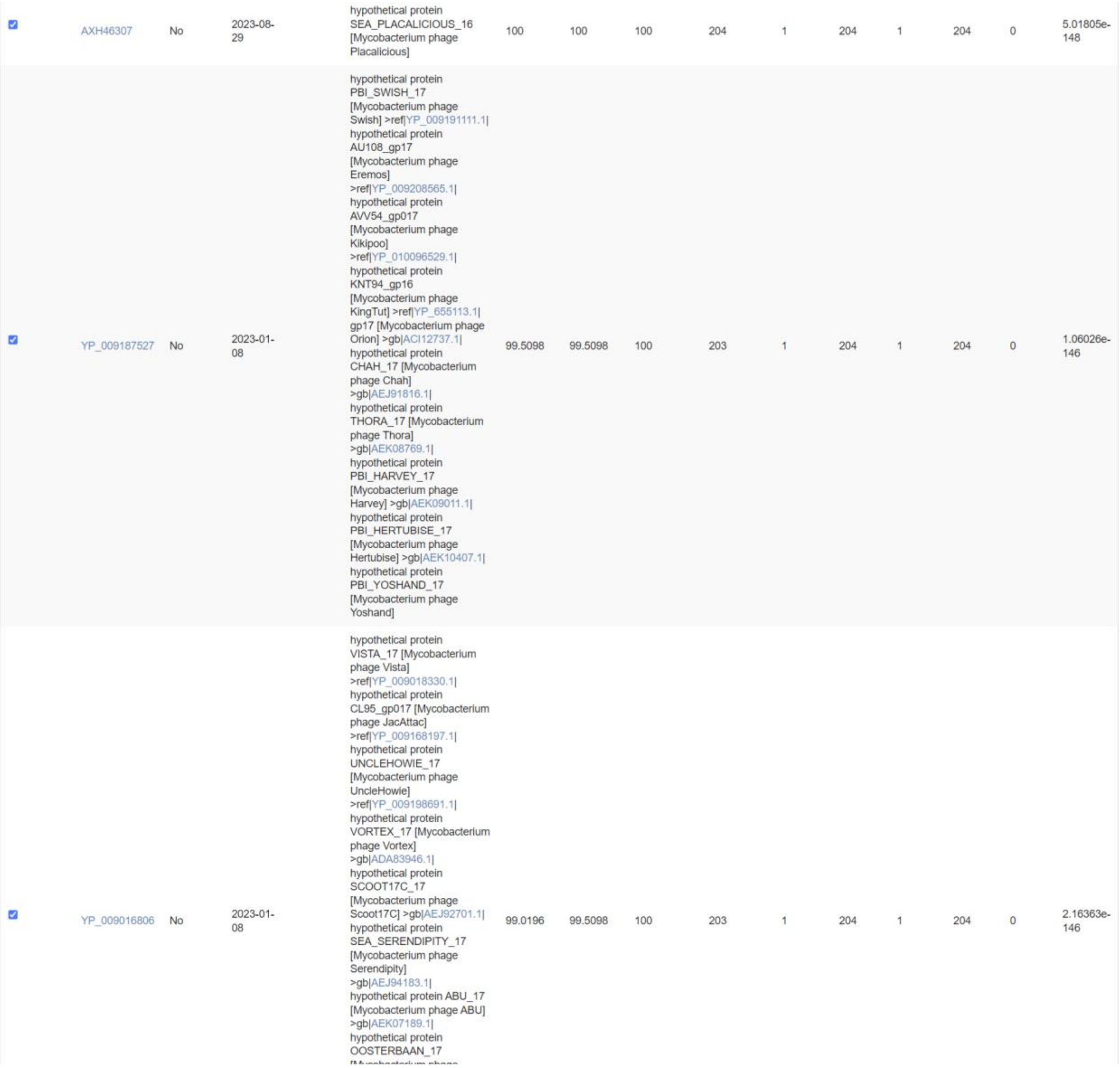


| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | BlackStallion | 17 | function unknown | 204 | 415 | 1e-116 | B1 | 137530 |
| ☑ | Cannibal | 17 | function unknown | 204 | 415 | 1e-116 | B1 | 137530 |
| ☑ | Chah | 17 | function unknown | 204 | 415 | 1e-116 | B1 | 137530 |
| ☑ | Chorkpop | 17 | function unknown | 204 | 415 | 1e-116 | B1 | 137530 |
| ☑ | Cobra | 17 | function unknown | 204 | 415 | 1e-116 | B1 | 137530 |
| ☑ | Craff | 17 | function unknown | 204 | 415 | 1e-116 | B1 | 137530 |
| ☑ | Eremos | 17 | function unknown | 204 | 415 | 1e-116 | B1 | 137530 |
| ☑ | FluffyNinja | 17 | function unknown | 204 | 415 | 1e-116 | B1 | 137530 |



```
                                                                 Score    E
Sequences producing significant alignments:                      (bits) Value

Placalicious_16, function unknown, 204                                418   e-117
Eugenia_Draft_17, function unknown, 204                               418   e-117
Yoshand_17, function unknown, 204                                     415   e-116
Xavier_16, function unknown, 204                                      415   e-116
Vivaldi_17, function unknown, 204                                     415   e-116
Virapocalypse_17, function unknown, 204                               415   e-116
Vaticameos_17, function unknown, 204                                  415   e-116
True_16, function unknown, 204                                        415   e-116
Tomlarah_17, function unknown, 204                                    415   e-116
Thora_17, function unknown, 204                                       415   e-116
Telesworld_16, function unknown, 204                                  415   e-116
TallGrassMM_17, function unknown, 204                                 415   e-116
Swish_17, function unknown, 204                                       415   e-116
Selr12_Draft_17, function unknown, 204                                415   e-116
Samaymay_17, function unknown, 204                                    415   e-116
RedMaple_16, function unknown, 204                                    415   e-116
QueenBeane_17, function unknown, 204                                  415   e-116
Plmatters_17, function unknown, 204                                   415   e-116
Phunky_17, function unknown, 204                                      415   e-116
Pherdinand_17, function unknown, 204                                  415   e-116
```

## b. SIF: NCBI BLAST



| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | AXH46307 | No | 2023-08-29 | hypothetical protein SEA_PLACALICIOUS_16 [Mycobacterium phage Placalicious] | 100 | 100 | 100 | 204 | 1 | 204 | 1 | 204 | 0 | 5.01805e-148 |
| ☑ | YP_009187527 | No | 2023-01-08 | hypothetical protein PBI_SWISH_17 [Mycobacterium phage Swish] >ref\|YP_009191111.1\| hypothetical protein AU108_gp17 [Mycobacterium phage Eremos] >ref\|YP_009208565.1\| hypothetical protein AVV54_gp017 [Mycobacterium phage Kikipoo] >ref\|YP_010096529.1\| hypothetical protein KNT94_gp16 [Mycobacterium phage KingTut] >ref\|YP_655113.1\| gp17 [Mycobacterium phage Orion] >gb\|ACI12737.1\| hypothetical protein CHAH_17 [Mycobacterium phage Chah] >gb\|AEJ91816.1\| hypothetical protein THORA_17 [Mycobacterium phage Thora] >gb\|AEK08769.1\| hypothetical protein PBI_HARVEY_17 [Mycobacterium phage Harvey] >gb\|AEK09011.1\| hypothetical protein PBI_HERTUBISE_17 [Mycobacterium phage Hertubise] >gb\|AEK10407.1\| hypothetical protein PBI_YOSHAND_17 [Mycobacterium phage Yoshand] | 99.5098 | 99.5098 | 100 | 203 | 1 | 204 | 1 | 204 | 0 | 1.06026e-146 |
| ☑ | YP_009016806 | No | 2023-01-08 | hypothetical protein VISTA_17 [Mycobacterium phage Vista] >ref\|YP_009018330.1\| hypothetical protein CL95_gp017 [Mycobacterium phage JacAttac] >ref\|YP_009168197.1\| hypothetical protein UNCLEHOWIE_17 [Mycobacterium phage UncleHowie] >ref\|YP_009198691.1\| hypothetical protein VORTEX_17 [Mycobacterium phage Vortex] >gb\|ADA83946.1\| hypothetical protein SCOOT17C_17 [Mycobacterium phage Scoot17C] >gb\|AEJ92701.1\| hypothetical protein SEA_SERENDIPITY_17 [Mycobacterium phage Serendipity] >gb\|AEJ94183.1\| hypothetical protein ABU_17 [Mycobacterium phage ABU] >gb\|AEK07189.1\| hypothetical protein OOSTERBAAN_17 | 99.0196 | 99.5098 | 100 | 203 | 1 | 204 | 1 | 204 | 0 | 2.16363e-146 |

## c. SIF: HHPred

N/A

## d. SIF: Synteny-Phamerator (three genomes)

Aligns with Buckeye, Usavi, Waterdiva, BlackStallion, FluffyNinja, etc.

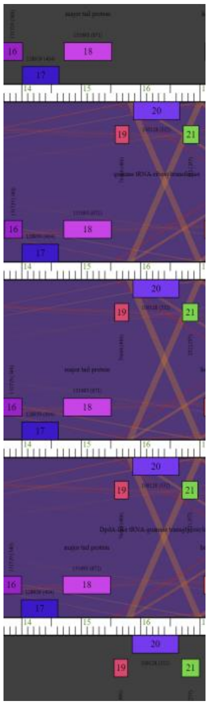
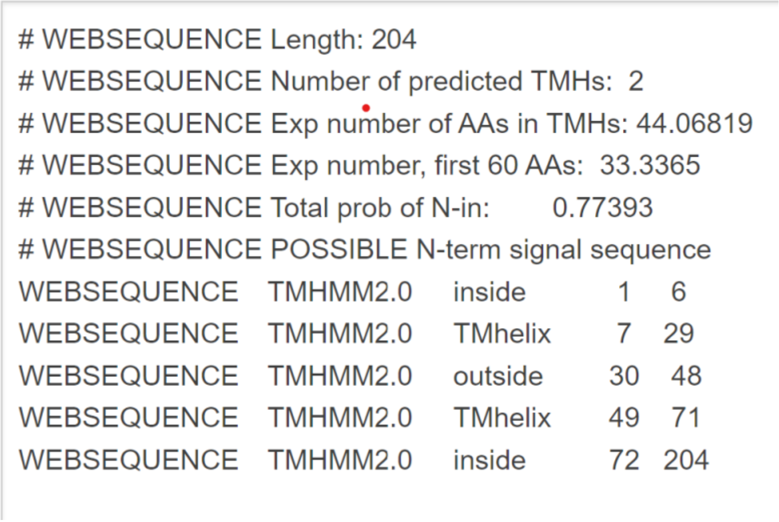


*Figure 1. (Top to bottom) Buckeye, Eugenia, Usavi, and Waterdiva.*

# 7. Any other important information.

TmHmm: 2

While initially a hypothetical protein, the TmHmm prediction shows that Gene 17 is possibly a membrane protein. This means many other phages in this cluster's Gene 17 could be membrane proteins as well instead of hypothetical.



```
# WEBSEQUENCE Length: 204
# WEBSEQUENCE Number of predicted TMHs:  2
# WEBSEQUENCE Exp number of AAs in TMHs: 44.06819
# WEBSEQUENCE Exp number, first 60 AAs:  33.3365
# WEBSEQUENCE Total prob of N-in:        0.77393
# WEBSEQUENCE POSSIBLE N-term signal sequence
WEBSEQUENCE    TMHMM2.0     inside          1     6
WEBSEQUENCE    TMHMM2.0     TMhelix         7    29
WEBSEQUENCE    TMHMM2.0     outside        30    48
WEBSEQUENCE    TMHMM2.0     TMhelix        49    71
WEBSEQUENCE    TMHMM2.0     inside         72   204
```

## CURATOR NAME: ISABELLA LIMA

## GENE NAME: EUGENIA_DRAFT_18

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

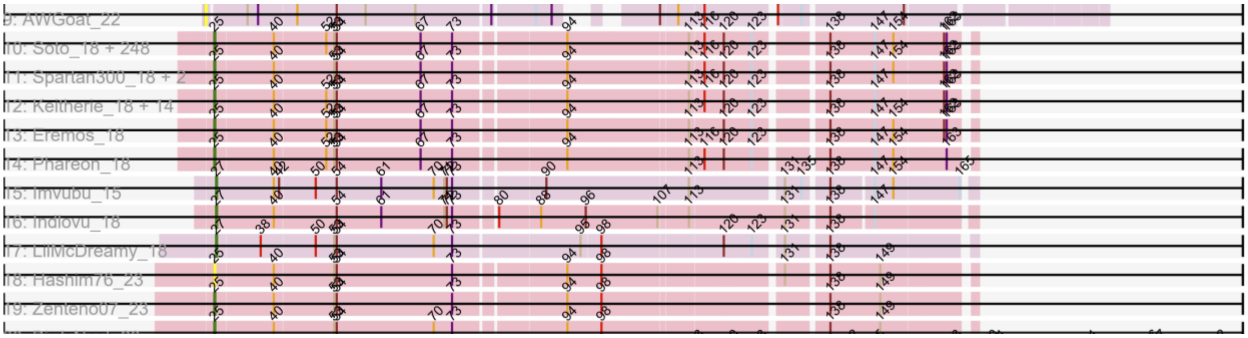
### 1. Starterator

Eugenia is in Track 10 with Start 25

The start number called the most often in the published annotations is 25, it was called in 263 of the 585 non-draft genes in the pham.

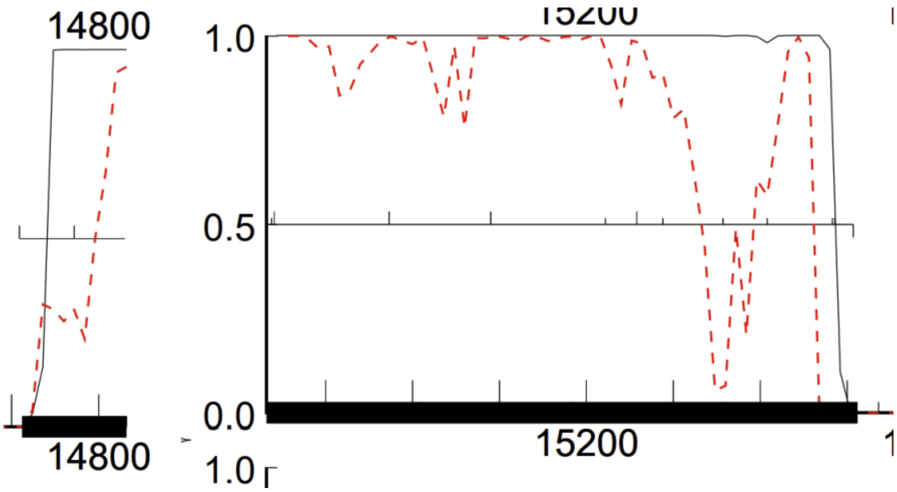
Start 25:

- Found in 288 of 678 ( 42.5% ) of genes in pham
- Manual Annotations of this start: 263 of 585
- Called 100.0% of time when present



## 2. GeneMark coding potential

GeneMark: 14712

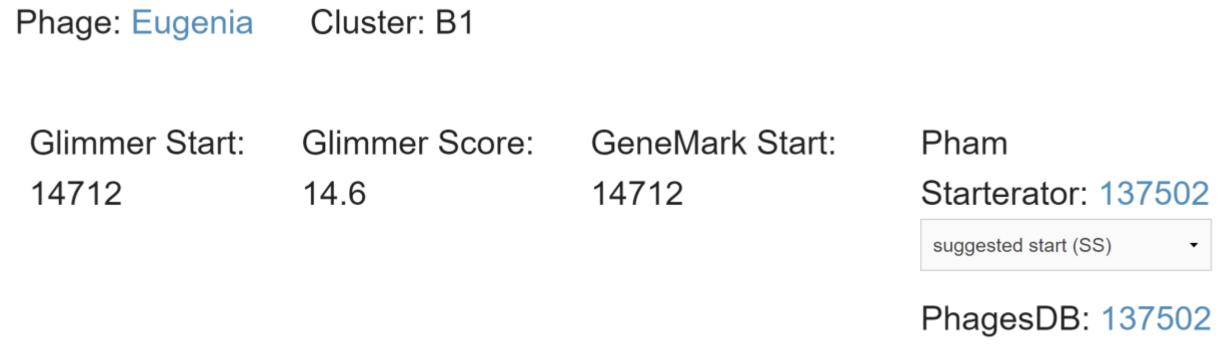


## 3. Glimmer and GeneMark agreement

YES

Glimmer: 14712

GeneMark: 14712



Phage: Eugenia Cluster: B1

Glimmer Start: 14712

Glimmer Score: 14.6

GeneMark Start: 14712

Pham Starterator: 137502

PhagesDB: 137502

## 4. Longest open reading frame (ORF) without excessive gap

| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 14712 | 15512 | 801 | 87 | 11 | 2.502 | -3.836 | TRUE | ATG | Select | ☑ |
| Forward | 14775 | 15512 | 738 | 150 | 8 | 1.687 | -5.938 |  | ATG |  | ☐ |

Start: 14712

Length: 801

Gap: 87

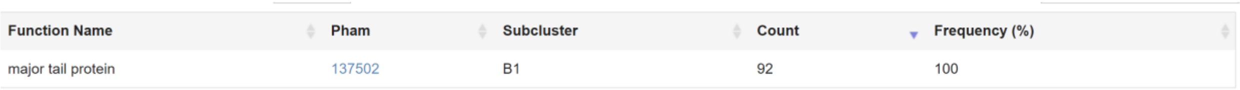
Z-score: 2.502

Final Score: -3.836

## 5. Function. Please follow this Official SEAPHAGE Function List

## If no functional prediction is present, write “Hypothetical protein”.

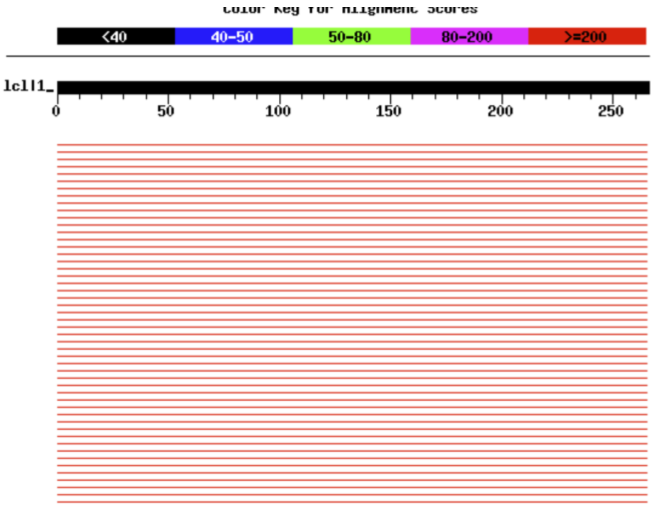
Major tail protein (Rank 1)

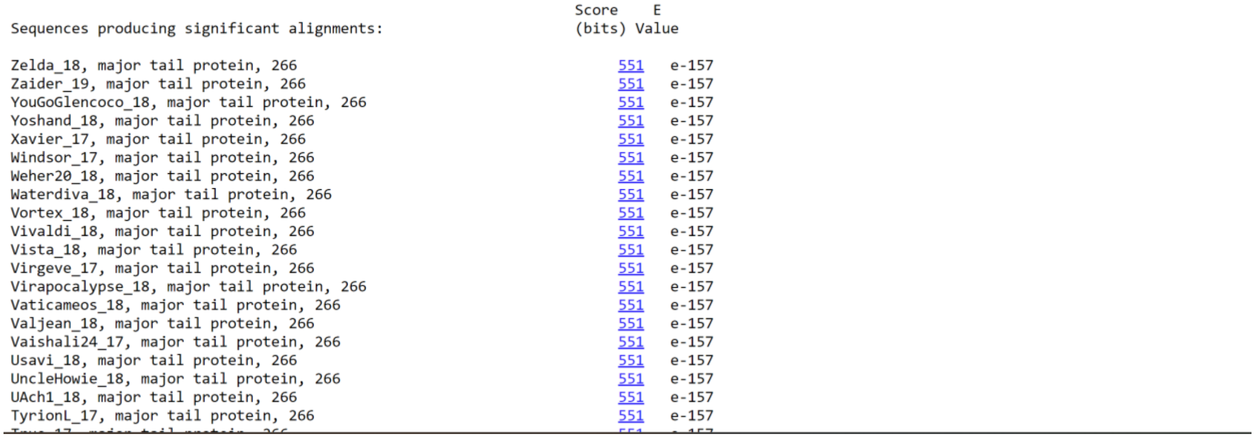


| Function Name | Pham | Subcluster | Count | Frequency (%) |
|---|---|---|---|---|
| major tail protein | 137502 | B1 | 92 | 100 |

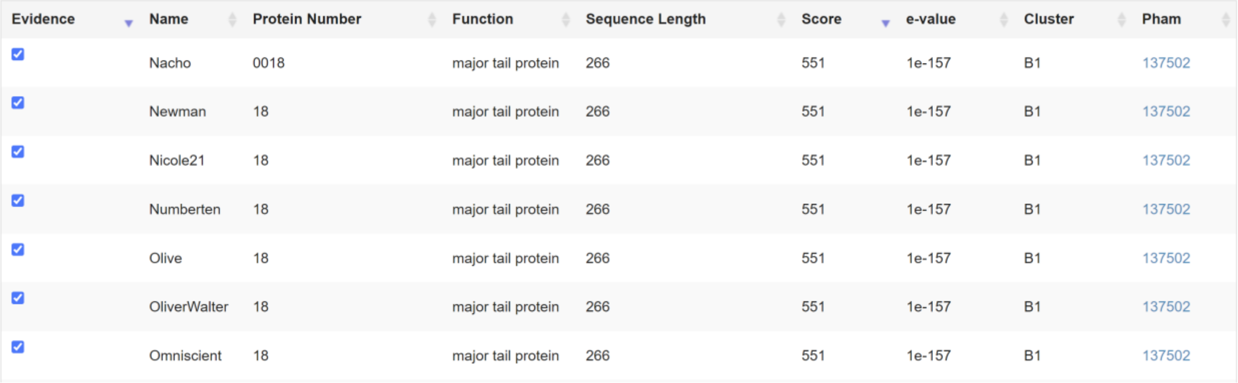
## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST



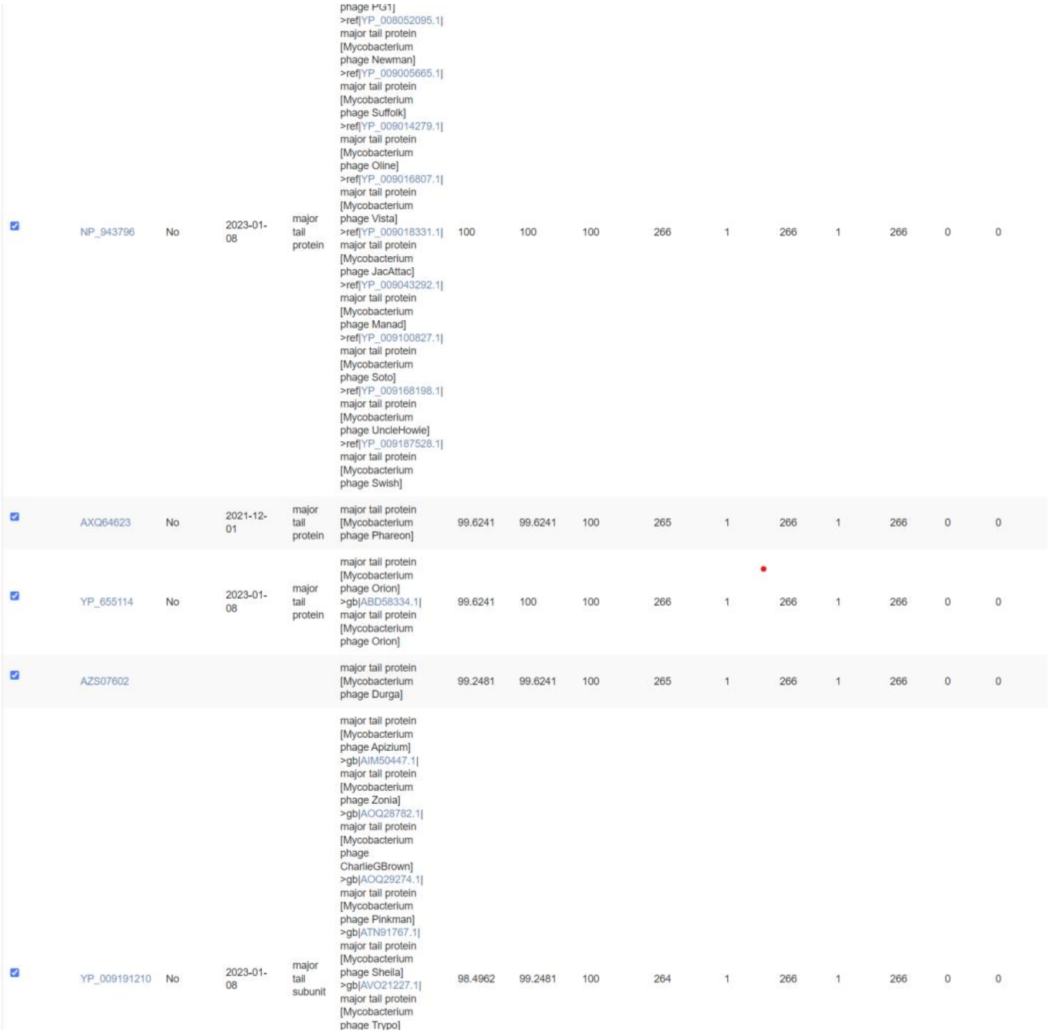


```
                                                                 Score    E
Sequences producing significant alignments:                      (bits) Value

Zelda_18, major tail protein, 266                                    551   e-157
Zaider_19, major tail protein, 266                                   551   e-157
YouGoGlencoco_18, major tail protein, 266                            551   e-157
Yoshand_18, major tail protein, 266                                  551   e-157
Xavier_17, major tail protein, 266                                   551   e-157
Windsor_17, major tail protein, 266                                  551   e-157
Weher20_18, major tail protein, 266                                  551   e-157
Waterdiva_18, major tail protein, 266                                551   e-157
Vortex_18, major tail protein, 266                                   551   e-157
Vivaldi_18, major tail protein, 266                                  551   e-157
Vista_18, major tail protein, 266                                    551   e-157
Virgeve_17, major tail protein, 266                                  551   e-157
Virapocalypse_18, major tail protein, 266                            551   e-157
Vaticameos_18, major tail protein, 266                               551   e-157
Valjean_18, major tail protein, 266                                  551   e-157
Vaishali24_17, major tail protein, 266                               551   e-157
Usavi_18, major tail protein, 266                                    551   e-157
UncleHowie_18, major tail protein, 266                               551   e-157
UAch1_18, major tail protein, 266                                    551   e-157
TyrionL_17, major tail protein, 266                                  551   e-157
```



| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | Nacho | 0018 | major tail protein | 266 | 551 | 1e-157 | B1 | 137502 |
| ☑ | Newman | 18 | major tail protein | 266 | 551 | 1e-157 | B1 | 137502 |
| ☑ | Nicole21 | 18 | major tail protein | 266 | 551 | 1e-157 | B1 | 137502 |
| ☑ | Numberten | 18 | major tail protein | 266 | 551 | 1e-157 | B1 | 137502 |
| ☑ | Olive | 18 | major tail protein | 266 | 551 | 1e-157 | B1 | 137502 |
| ☑ | OliverWalter | 18 | major tail protein | 266 | 551 | 1e-157 | B1 | 137502 |
| ☑ | Omniscient | 18 | major tail protein | 266 | 551 | 1e-157 | B1 | 137502 |

## b. SIF: NCBI BLAST



| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | NP_943796 | No | 2023-01-08 | major tail protein | phage PG1]<br>>ref\|YP_008052095.1\| major tail protein [Mycobacterium phage Newman]<br>>ref\|YP_009005665.1\| major tail protein [Mycobacterium phage Suffolk]<br>>ref\|YP_009014279.1\| major tail protein [Mycobacterium phage Oline]<br>>ref\|YP_009016807.1\| major tail protein [Mycobacterium phage Vista]<br>>ref\|YP_009018331.1\| major tail protein [Mycobacterium phage JacAttac]<br>>ref\|YP_009043292.1\| major tail protein [Mycobacterium phage Manad]<br>>ref\|YP_009100827.1\| major tail protein [Mycobacterium phage Soto]<br>>ref\|YP_009168198.1\| major tail protein [Mycobacterium phage UncleHowie]<br>>ref\|YP_009187528.1\| major tail protein [Mycobacterium phage Swish] | 100 | 100 | 100 | 266 | 1 | 266 | 1 | 266 | 0 | 0 |
| ☑ | AXQ64623 | No | 2021-12-01 | major tail protein | major tail protein [Mycobacterium phage Phareon] | 99.6241 | 99.6241 | 100 | 265 | 1 | 266 | 1 | 266 | 0 | 0 |
| ☑ | YP_655114 | No | 2023-01-08 | major tail protein | major tail protein [Mycobacterium phage Orion]<br>>gb\|ABD58334.1\| major tail protein [Mycobacterium phage Orion] | 99.6241 | 100 | 100 | 266 | 1 | 266 | 1 | 266 | 0 | 0 |
| ☑ | AZS07602 | | | | major tail protein [Mycobacterium phage Durga] | 99.2481 | 99.6241 | 100 | 265 | 1 | 266 | 1 | 266 | 0 | 0 |
| ☑ | YP_009191210 | No | 2023-01-08 | major tail subunit | major tail protein [Mycobacterium phage Apizium]<br>>gb\|AIM50447.1\| major tail protein [Mycobacterium phage Zonia]<br>>gb\|AOQ28782.1\| major tail protein [Mycobacterium phage CharlieGBrown]<br>>gb\|AOQ29274.1\| major tail protein [Mycobacterium phage Pinkman]<br>>gb\|ATN91767.1\| major tail protein [Mycobacterium phage Sheila]<br>>gb\|AVO21227.1\| major tail protein [Mycobacterium phage Trypo] | 98.4962 | 99.2481 | 100 | 264 | 1 | 266 | 1 | 266 | 0 | 0 |

## c. SIF: HHPred

N/A

## d. SIF: Synteny-Phamerator (three genomes)

Aligns with Buckeye, Usavi, Waterdiva, Nacho, Olive, etc.

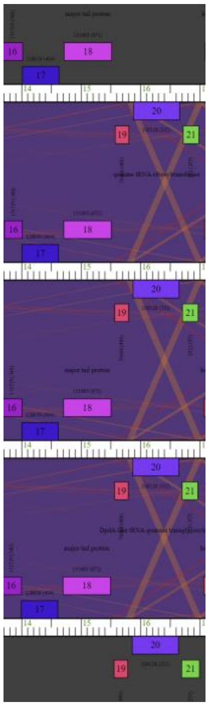


*Figure 1. (Top to bottom) Buckeye, Eugenia, Usavi, and Waterdiva.*

# 7. Any other important information.

## CURATOR NAME: ISABELLA LIMA

## GENE NAME: EUGENIA_DRAFT_19

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

# 1. Starterator

Eugenia is in Track 1, with Start 15

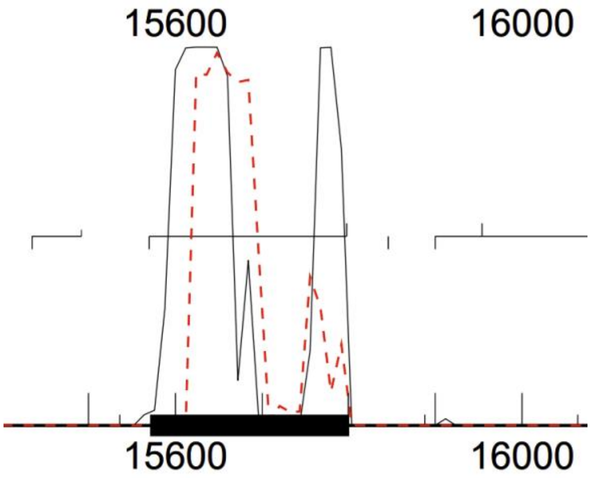
The start number called the most often in the published annotations is 15, it was called in 419 of the 436 non-draft genes in the pham.

Start 15:

- Found in 479 of 488 ( 98.2% ) of genes in pham
- Manual Annotations of this start: 419 of 436
- Called 97.5% of time when present

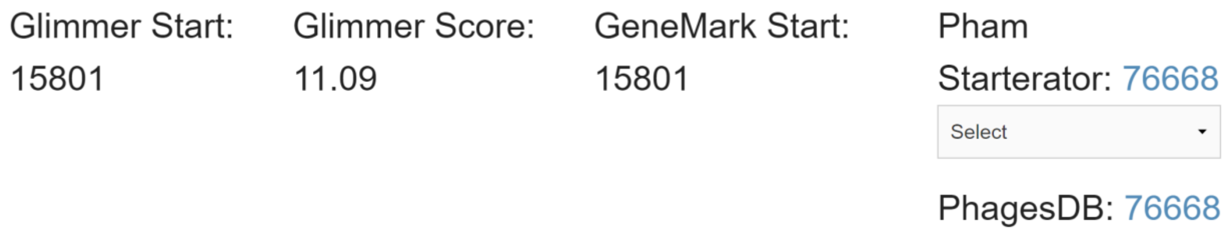
## 2. GeneMark coding potential

GeneMark: 15801 (Reverse)



## 3. Glimmer and GeneMark agreement

YES



| Glimmer Start: | Glimmer Score: | GeneMark Start: | Pham |
|---|---|---|---|
| 15801 | 11.09 | 15801 | Starterator: 76668 |

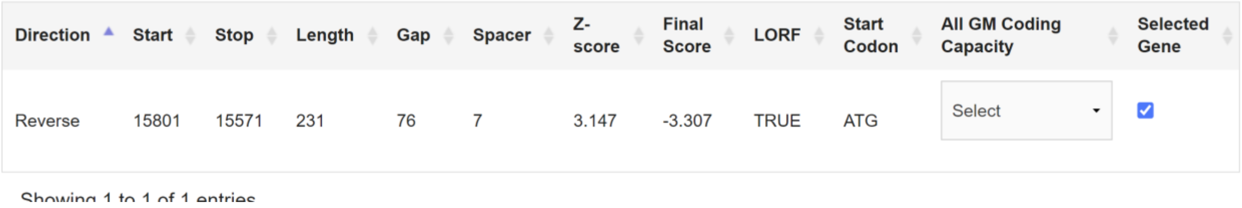
PhagesDB: 76668

Glimmer: 15801

GeneMark: 15801

## 4. Longest open reading frame (ORF) without excessive gap

YES



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 15801 | 15571 | 231 | 76 | 7 | 3.147 | -3.307 | TRUE | ATG | Select | ☑ |

Showing 1 to 1 of 1 entries

Start: 15801

Length: 231

Gap: 76

Z-score: 3.147

Final core: -3.307

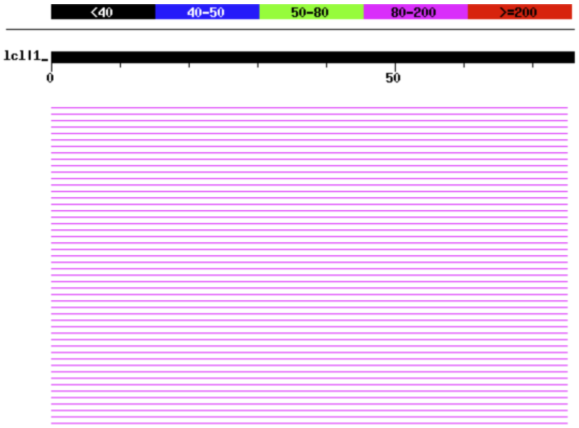
## 5. Function. Please follow this Official SEAPHAGE Function List

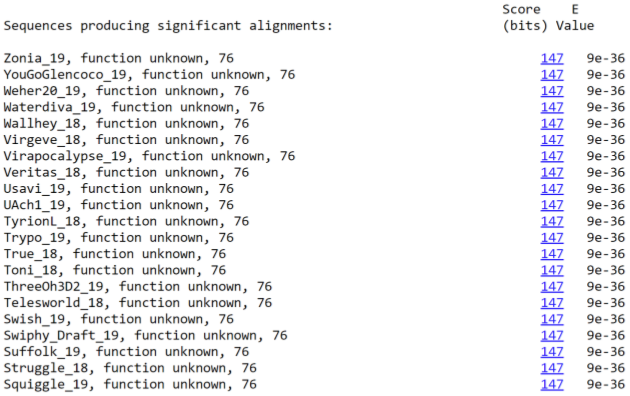
## If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

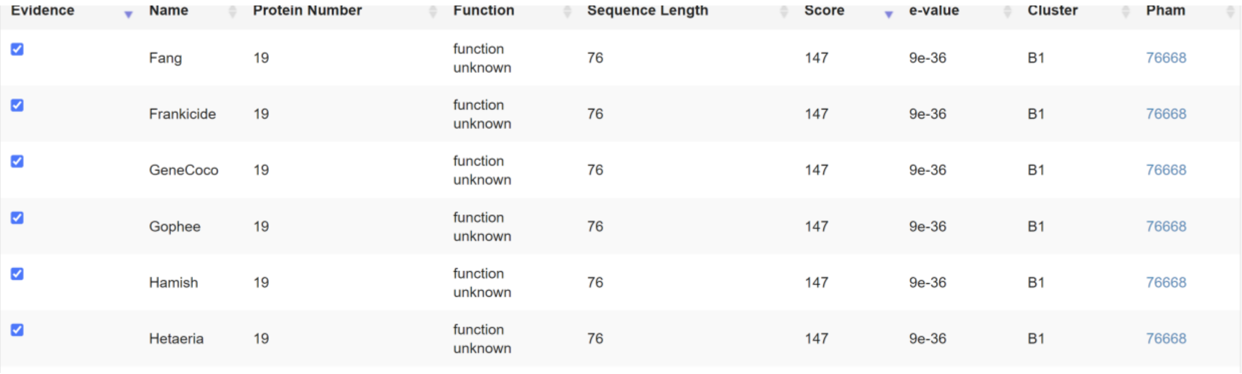
## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST



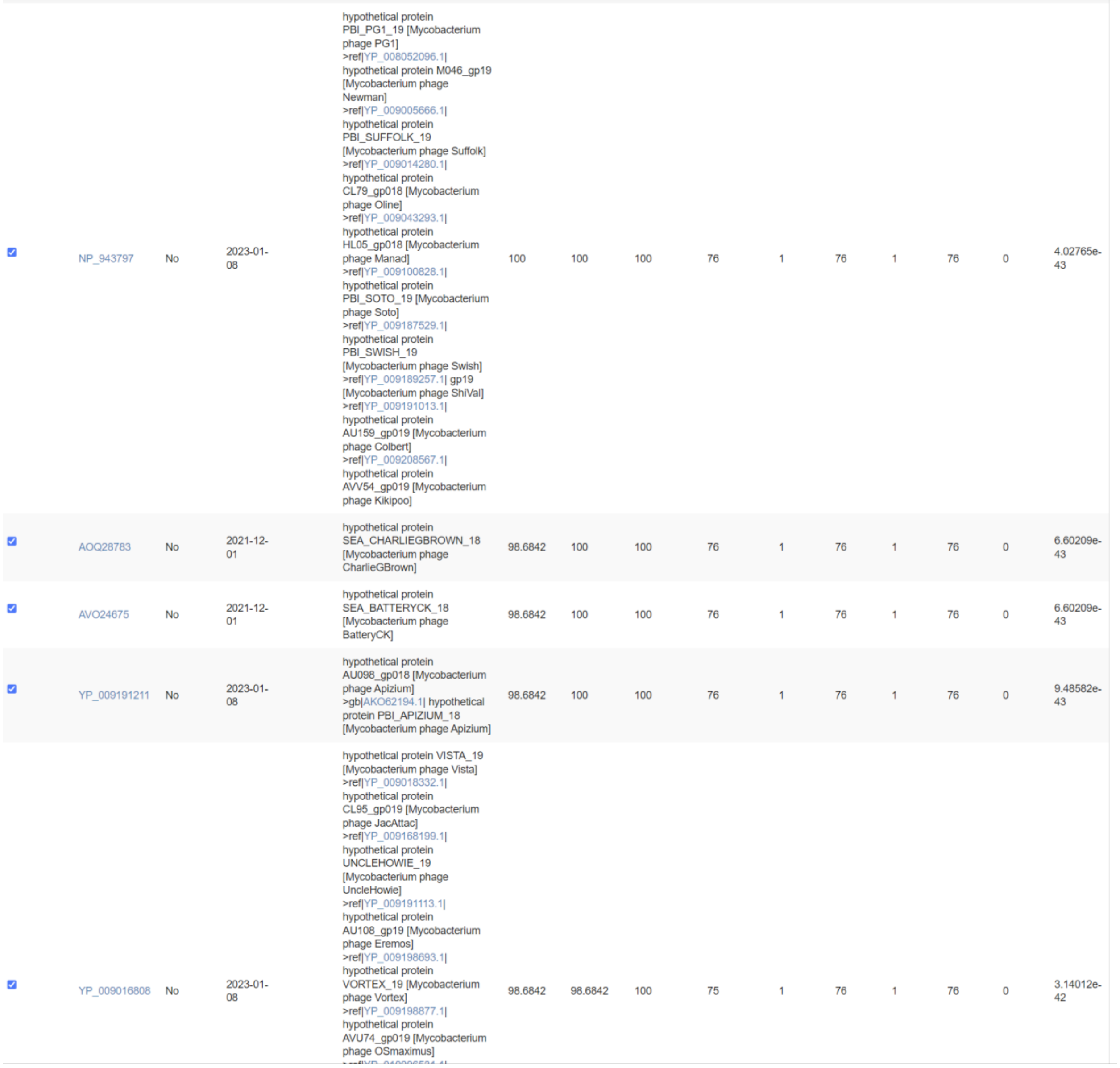


```
                                                                Score     E
Sequences producing significant alignments:                     (bits) Value

Zonia_19, function unknown, 76                                      147    9e-36
YouGoGlencoco_19, function unknown, 76                              147    9e-36
Weher20_19, function unknown, 76                                    147    9e-36
Waterdiva_19, function unknown, 76                                  147    9e-36
Wallhey_18, function unknown, 76                                    147    9e-36
Virgeve_18, function unknown, 76                                    147    9e-36
Virapocalypse_19, function unknown, 76                              147    9e-36
Veritas_18, function unknown, 76                                    147    9e-36
Usavi_19, function unknown, 76                                      147    9e-36
UAch1_19, function unknown, 76                                      147    9e-36
TyrionL_18, function unknown, 76                                    147    9e-36
Trypo_19, function unknown, 76                                      147    9e-36
True_18, function unknown, 76                                       147    9e-36
Toni_18, function unknown, 76                                       147    9e-36
ThreeOh3D2_19, function unknown, 76                                 147    9e-36
Telesworld_18, function unknown, 76                                 147    9e-36
Swish_19, function unknown, 76                                      147    9e-36
Swiphy_Draft_19, function unknown, 76                               147    9e-36
Suffolk_19, function unknown, 76                                    147    9e-36
Struggle_18, function unknown, 76                                   147    9e-36
Squiggle_19, function unknown, 76                                   147    9e-36
```



| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | Fang | 19 | function unknown | 76 | 147 | 9e-36 | B1 | 76668 |
| ☑ | Frankicide | 19 | function unknown | 76 | 147 | 9e-36 | B1 | 76668 |
| ☑ | GeneCoco | 19 | function unknown | 76 | 147 | 9e-36 | B1 | 76668 |
| ☑ | Gophee | 19 | function unknown | 76 | 147 | 9e-36 | B1 | 76668 |
| ☑ | Hamish | 19 | function unknown | 76 | 147 | 9e-36 | B1 | 76668 |
| ☑ | Hetaeria | 19 | function unknown | 76 | 147 | 9e-36 | B1 | 76668 |

## b. SIF: NCBI BLAST



| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | NP_943797 | No | 2023-01-08 | hypothetical protein PBI_PG1_19 [Mycobacterium phage PG1] >ref\|YP_008052096.1\| hypothetical protein M046_gp19 [Mycobacterium phage Newman] >ref\|YP_009005666.1\| hypothetical protein PBI_SUFFOLK_19 [Mycobacterium phage Suffolk] >ref\|YP_009014280.1\| hypothetical protein CL79_gp018 [Mycobacterium phage Oline] >ref\|YP_009043293.1\| hypothetical protein HL05_gp018 [Mycobacterium phage Manad] >ref\|YP_009100828.1\| hypothetical protein PBI_SOTO_19 [Mycobacterium phage Soto] >ref\|YP_009187529.1\| hypothetical protein PBI_SWISH_19 [Mycobacterium phage Swish] >ref\|YP_009189257.1\| gp19 [Mycobacterium phage ShiVal] >ref\|YP_009191013.1\| hypothetical protein AU159_gp019 [Mycobacterium phage Colbert] >ref\|YP_009208567.1\| hypothetical protein AVV54_gp019 [Mycobacterium phage Kikipoo] | 100 | 100 | 100 | 76 | 1 | 76 | 1 | 76 | 0 | 4.02765e-43 |
| ☑ | AOQ28783 | No | 2021-12-01 | hypothetical protein SEA_CHARLIEGBROWN_18 [Mycobacterium phage CharlieGBrown] | 98.6842 | 100 | 100 | 76 | 1 | 76 | 1 | 76 | 0 | 6.60209e-43 |
| ☑ | AVO24675 | No | 2021-12-01 | hypothetical protein SEA_BATTERYCK_18 [Mycobacterium phage BatteryCK] | 98.6842 | 100 | 100 | 76 | 1 | 76 | 1 | 76 | 0 | 6.60209e-43 |
| ☑ | YP_009191211 | No | 2023-01-08 | hypothetical protein AU098_gp018 [Mycobacterium phage Apizium] >gb\|AKO62194.1\| hypothetical protein PBI_APIZIUM_18 [Mycobacterium phage Apizium] | 98.6842 | 100 | 100 | 76 | 1 | 76 | 1 | 76 | 0 | 9.48582e-43 |
| ☑ | YP_009016808 | No | 2023-01-08 | hypothetical protein VISTA_19 [Mycobacterium phage Vista] >ref\|YP_009018332.1\| hypothetical protein CL95_gp019 [Mycobacterium phage JacAttac] >ref\|YP_009168199.1\| hypothetical protein UNCLEHOWIE_19 [Mycobacterium phage UncleHowie] >ref\|YP_009191113.1\| hypothetical protein AU108_gp19 [Mycobacterium phage Eremos] >ref\|YP_009198693.1\| hypothetical protein VORTEX_19 [Mycobacterium phage Vortex] >ref\|YP_009198877.1\| hypothetical protein AVU74_gp019 [Mycobacterium phage OSmaximus] | 98.6842 | 98.6842 | 100 | 75 | 1 | 76 | 1 | 76 | 0 | 3.14012e-42 |

## c. SIF: HHPred

N/A

## d. SIF: Synteny-Phamerator (three genomes)

Aligns with Buckeye, Waterdiva, Usavi, Fang, Hamish, etc.

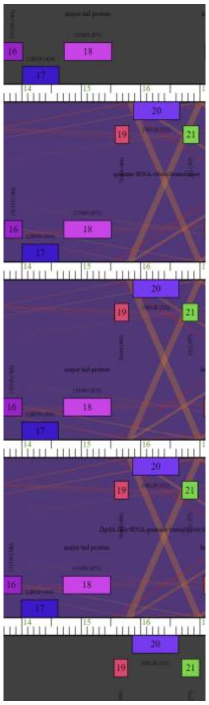


*Figure 1. (Top to bottom) Buckeye, Eugenia, Usavi, and Waterdiva.*

# 7. Any other important information.

Reverse.

## CURATOR NAME: ISABELLA LIMA

## GENE NAME: EUGENIA_DRAFT_20

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

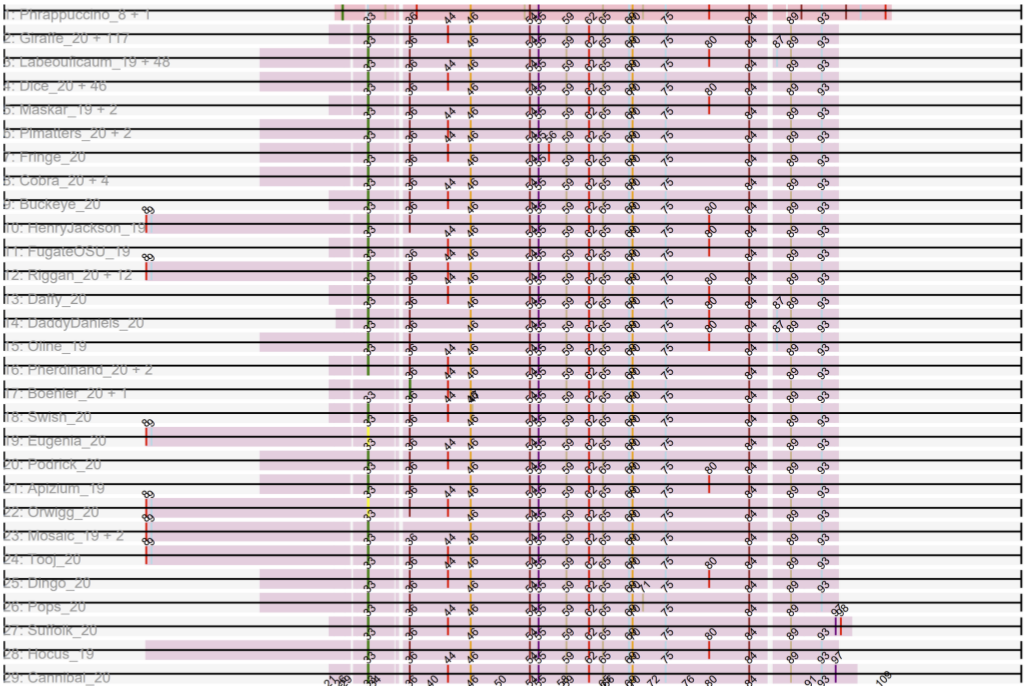
### 1. Starterator

Eugenia is in Track 19 with Start 33.

The start number called the most often in the published annotations is 33, it was called in 304 of the 310 non-draft genes in the pham.

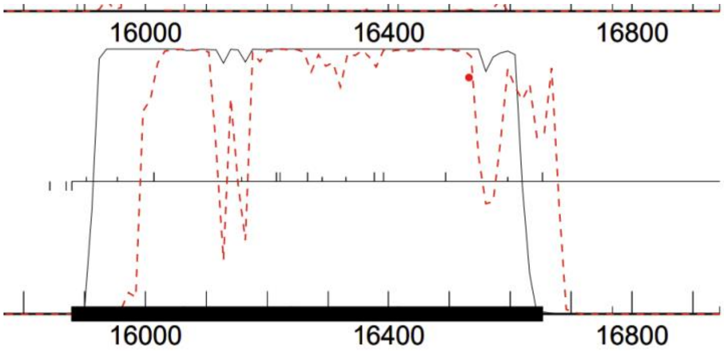
Start 33:

- Found in 334 of 337 ( 99.1% ) of genes in pham
- Manual Annotations of this start: 304 of 310
- Called 99.1% of time when present



## 2. GeneMark coding potential

GeneMark: 16654

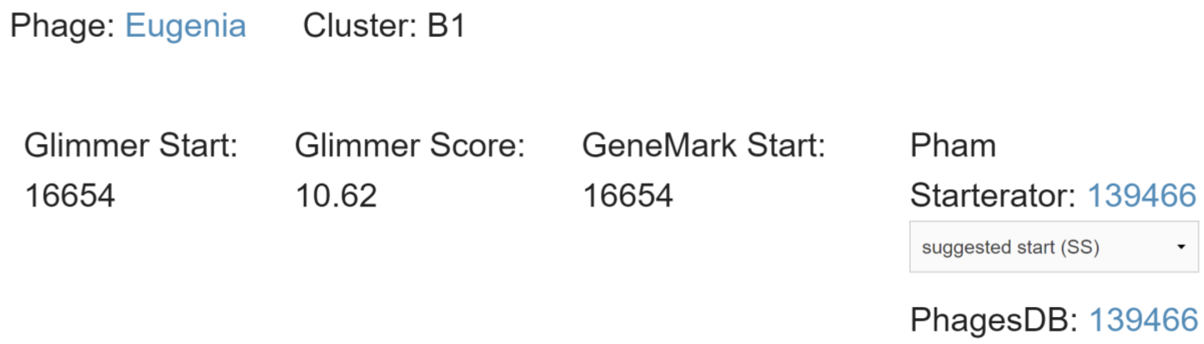


## 3. Glimmer and GeneMark agreement

YES

Glimmer: 16654

GeneMark: 16654



Phage: Eugenia Cluster: B1

Glimmer Start: 16654

Glimmer Score: 10.62

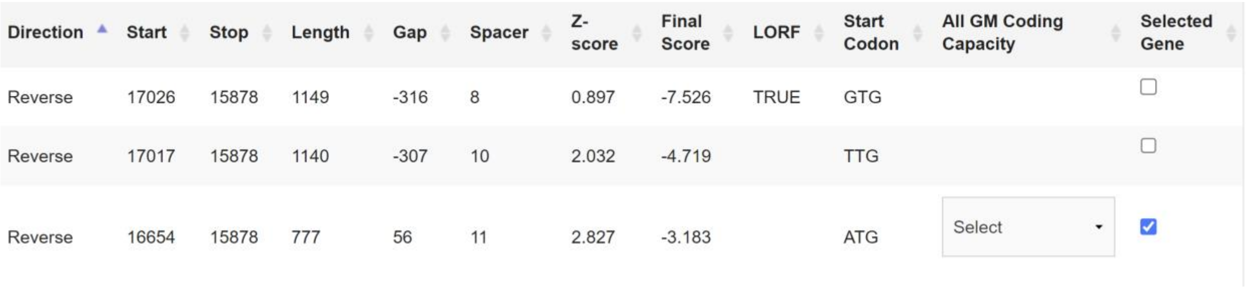
GeneMark Start: 16654

Pham Starterator: 139466

PhagesDB: 139466

## 4. Longest open reading frame (ORF) without excessive gap

NO, but has smallest gap.



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 17026 | 15878 | 1149 | -316 | 8 | 0.897 | -7.526 | TRUE | GTG | | ☐ |
| Reverse | 17017 | 15878 | 1140 | -307 | 10 | 2.032 | -4.719 | | TTG | | ☐ |
| Reverse | 16654 | 15878 | 777 | 56 | 11 | 2.827 | -3.183 | | ATG | Select | ☑ |

Start: 16654

Length: 777

Gap: 56

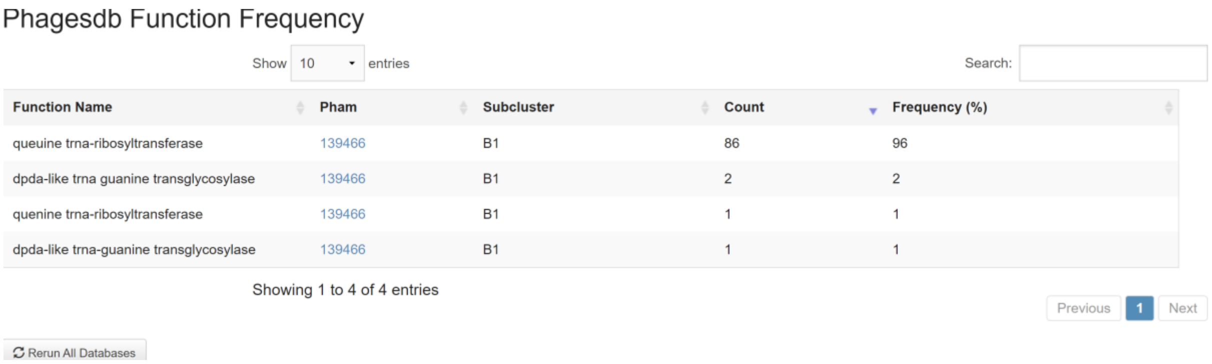
Z-score: 2.827

Final Score: -3.183

## 5. Function. Please follow this Official SEAPHAGE Function List

**If no functional prediction is present, write "Hypothetical protein".**

dpdA-like tRNA-guanine transglycosylase



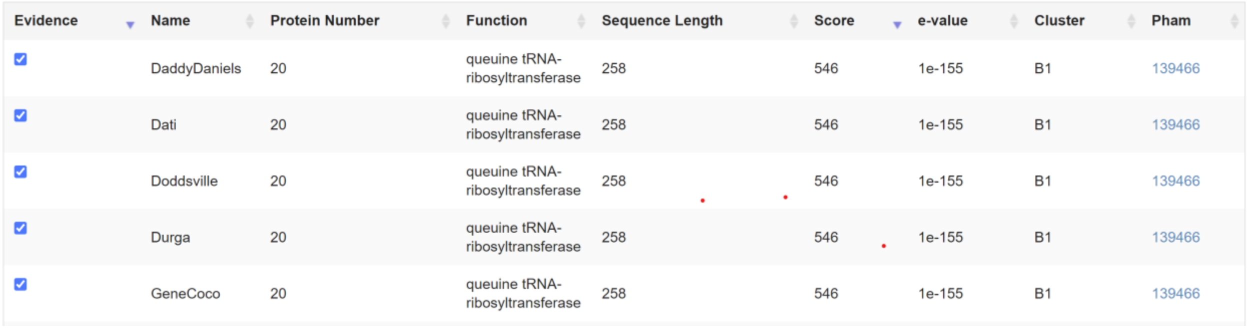
Phagesdb Function Frequency

| Function Name | Pham | Subcluster | Count | Frequency (%) |
|---|---|---|---|---|
| queuine trna-ribosyltransferase | 139466 | B1 | 86 | 96 |
| dpda-like trna guanine transglycosylase | 139466 | B1 | 2 | 2 |
| quenine trna-ribosyltransferase | 139466 | B1 | 1 | 1 |
| dpda-like trna-guanine transglycosylase | 139466 | B1 | 1 | 1 |

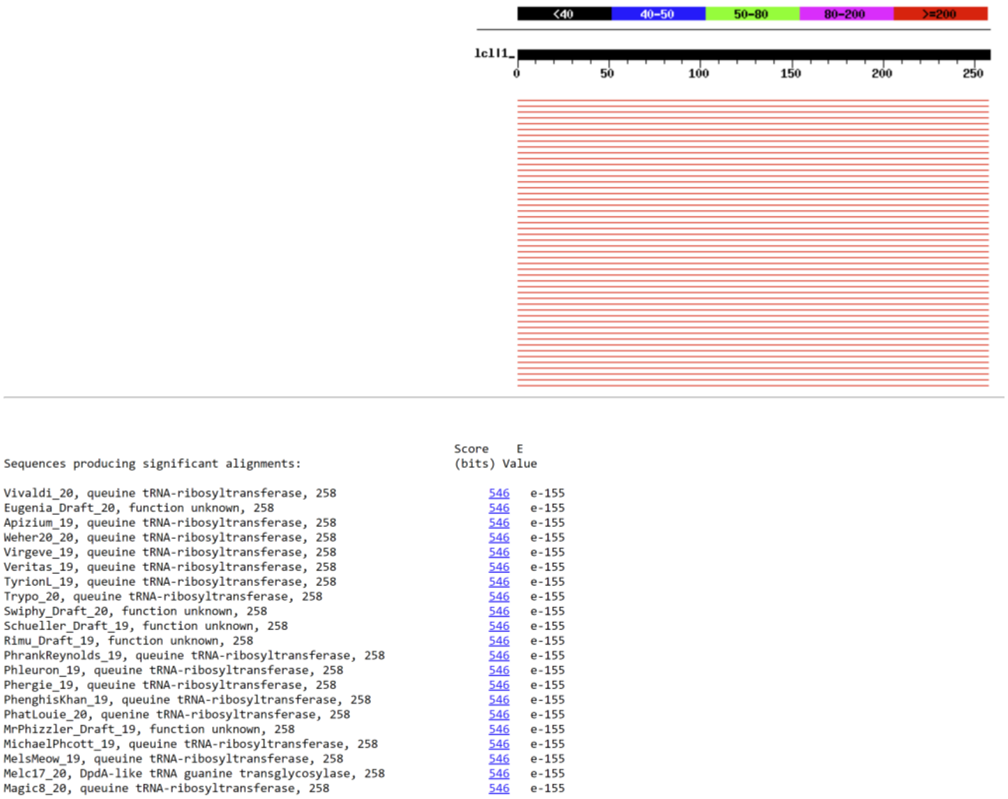
Showing 1 to 4 of 4 entries

## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST

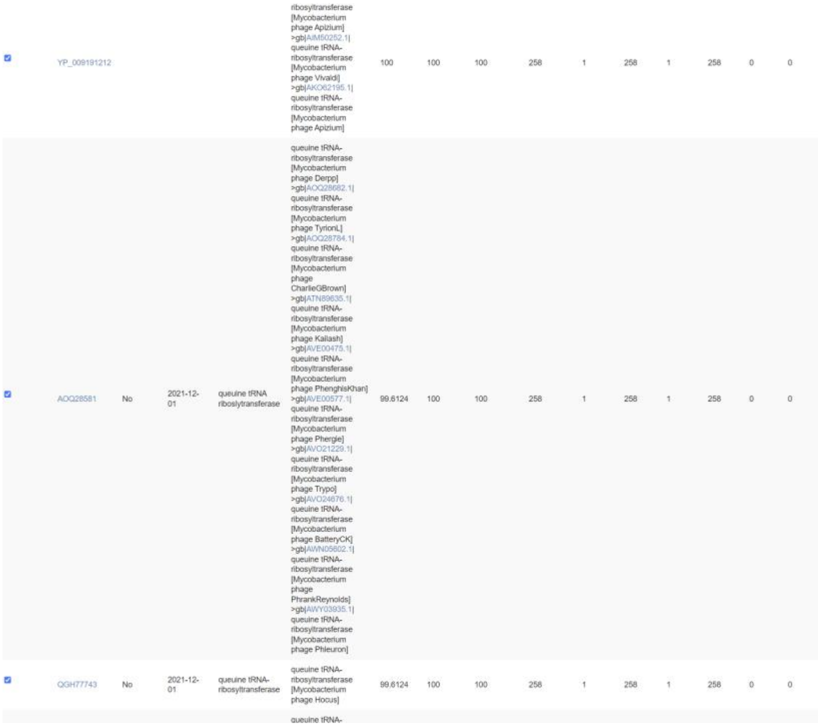


| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | DaddyDaniels | 20 | queuine tRNA-ribosyltransferase | 258 | 546 | 1e-155 | B1 | 139466 |
| ☑ | Dati | 20 | queuine tRNA-ribosyltransferase | 258 | 546 | 1e-155 | B1 | 139466 |
| ☑ | Doddsville | 20 | queuine tRNA-ribosyltransferase | 258 | 546 | 1e-155 | B1 | 139466 |
| ☑ | Durga | 20 | queuine tRNA-ribosyltransferase | 258 | 546 | 1e-155 | B1 | 139466 |
| ☑ | GeneCoco | 20 | queuine tRNA-ribosyltransferase | 258 | 546 | 1e-155 | B1 | 139466 |

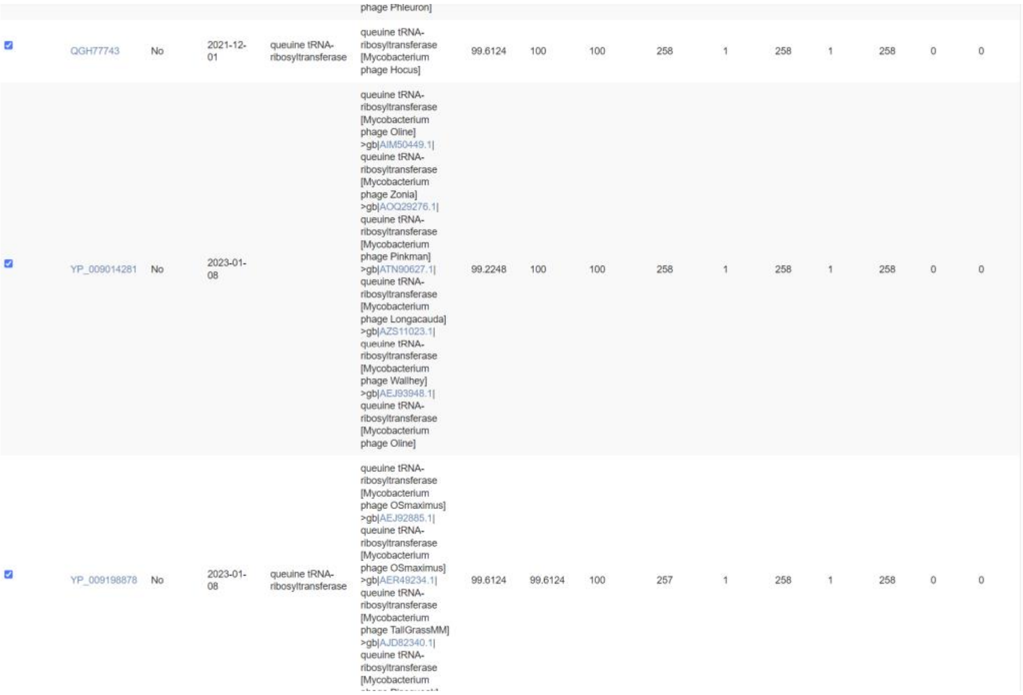


```
                                                                     Score    E
Sequences producing significant alignments:                          (bits) Value

Vivaldi_20, queuine tRNA-ribosyltransferase, 258                        546   e-155
Eugenia_Draft_20, function unknown, 258                                 546   e-155
Apizium_19, queuine tRNA-ribosyltransferase, 258                        546   e-155
Weher20_20, queuine tRNA-ribosyltransferase, 258                        546   e-155
Virgeve_19, queuine tRNA-ribosyltransferase, 258                        546   e-155
Veritas_19, queuine tRNA-ribosyltransferase, 258                        546   e-155
TyrionL_19, queuine tRNA-ribosyltransferase, 258                        546   e-155
Trypo_20, queuine tRNA-ribosyltransferase, 258                          546   e-155
Swiphy_Draft_20, function unknown, 258                                  546   e-155
Schueller_Draft_19, function unknown, 258                               546   e-155
Rimu_Draft_19, function unknown, 258                                    546   e-155
PhrankReynolds_19, queuine tRNA-ribosyltransferase, 258                 546   e-155
Phleuron_19, queuine tRNA-ribosyltransferase, 258                       546   e-155
Phergie_19, queuine tRNA-ribosyltransferase, 258                        546   e-155
PhenghisKhan_19, queuine tRNA-ribosyltransferase, 258                   546   e-155
PhatLouie_20, quenine tRNA-ribosyltransferase, 258                      546   e-155
MrPhizzler_Draft_19, function unknown, 258                              546   e-155
MichaelPhcott_19, queuine tRNA-ribosyltransferase, 258                  546   e-155
MelsMeow_19, queuine tRNA-ribosyltransferase, 258                       546   e-155
Melc17_20, DpdA-like tRNA guanine transglycosylase, 258                 546   e-155
Magic8_20, queuine tRNA-ribosyltransferase, 258                         546   e-155
```

## b. SIF: NCBI BLAST

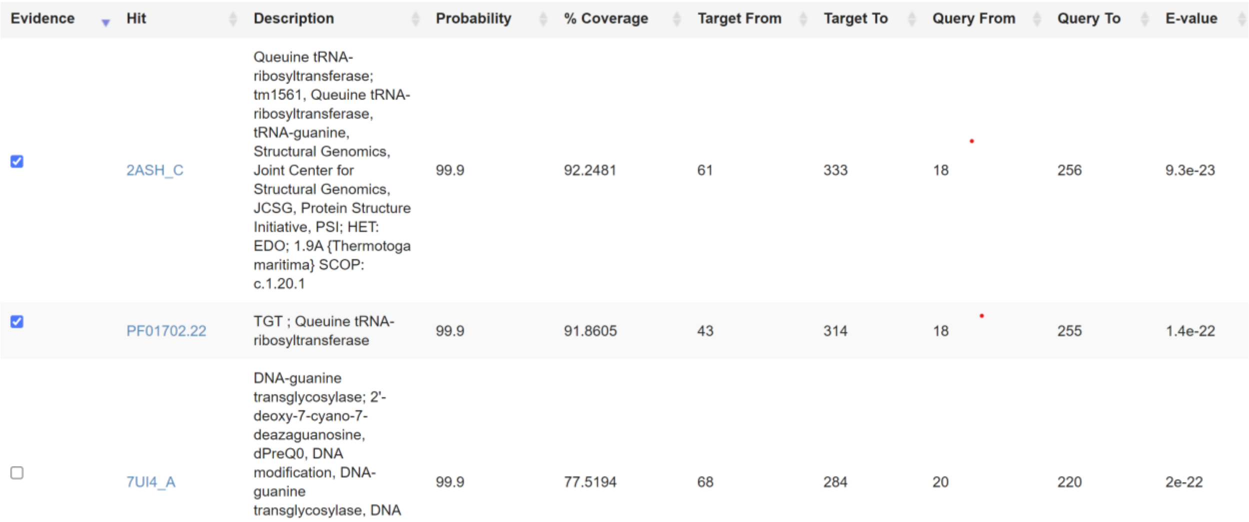


| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | YP_009191212 | | | | ribosyltransferase [Mycobacterium phage Apizium] >gb\|AIM50252.1\| queuine tRNA-ribosyltransferase [Mycobacterium phage Vivaldi] >gb\|AKO62195.1\| queuine tRNA-ribosyltransferase [Mycobacterium phage Apizium] | 100 | 100 | 100 | 258 | 1 | 258 | 1 | 258 | 0 | 0 |
| ☑ | AOQ28581 | No | 2021-12-01 | queuine tRNA ribosylytransferase | queuine tRNA-ribosyltransferase [Mycobacterium phage Derpp] >gb\|AOQ28682.1\| queuine tRNA-ribosyltransferase [Mycobacterium phage TyrionL] >gb\|AOQ28784.1\| queuine tRNA-ribosyltransferase [Mycobacterium phage CharlieGBrown] >gb\|ATN89635.1\| queuine tRNA-ribosyltransferase [Mycobacterium phage Kailash] >gb\|AVE00475.1\| queuine tRNA-ribosyltransferase [Mycobacterium phage PhenghisKhan] >gb\|AVE00577.1\| queuine tRNA-ribosyltransferase [Mycobacterium phage Phergie] >gb\|AVO21229.1\| queuine tRNA-ribosyltransferase [Mycobacterium phage Trypo] >gb\|AVO24676.1\| queuine tRNA-ribosyltransferase [Mycobacterium phage BatteryCK] >gb\|AWN05602.1\| queuine tRNA-ribosyltransferase [Mycobacterium phage PhrankReynolds] >gb\|AWY03935.1\| queuine tRNA-ribosyltransferase [Mycobacterium phage Phleuron] | 99.6124 | 100 | 100 | 258 | 1 | 258 | 1 | 258 | 0 | 0 |
| ☑ | QGH77743 | No | 2021-12-01 | queuine tRNA-ribosyltransferase | queuine tRNA-ribosyltransferase [Mycobacterium phage Hocus] | 99.6124 | 100 | 100 | 258 | 1 | 258 | 1 | 258 | 0 | 0 |
| | | | | | queuine tRNA- | | | | | | | | | | |



| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | phage Phleuron] | | | | | | | | | | |
| ☑ | QGH77743 | No | 2021-12-01 | queuine tRNA-ribosyltransferase | queuine tRNA-ribosyltransferase [Mycobacterium phage Hocus] | 99.6124 | 100 | 100 | 258 | 1 | 258 | 1 | 258 | 0 | 0 |
| ☑ | YP_009014281 | No | 2023-01-08 | | queuine tRNA-ribosyltransferase [Mycobacterium phage Oline] >gb\|AIM50449.1\| queuine tRNA-ribosyltransferase [Mycobacterium phage Zonia] >gb\|AOQ29276.1\| queuine tRNA-ribosyltransferase [Mycobacterium phage Pinkman] >gb\|ATN90627.1\| queuine tRNA-ribosyltransferase [Mycobacterium phage Longacauda] >gb\|AZS11023.1\| queuine tRNA-ribosyltransferase [Mycobacterium phage Wallhey] >gb\|AEJ93948.1\| queuine tRNA-ribosyltransferase [Mycobacterium phage Oline] | 99.2248 | 100 | 100 | 258 | 1 | 258 | 1 | 258 | 0 | 0 |
| ☑ | YP_009198878 | No | 2023-01-08 | queuine tRNA-ribosyltransferase | queuine tRNA-ribosyltransferase [Mycobacterium phage OSmaximus] >gb\|AEJ92885.1\| queuine tRNA-ribosyltransferase [Mycobacterium phage OSmaximus] >gb\|AER49234.1\| queuine tRNA-ribosyltransferase [Mycobacterium phage TallGrassMM] >gb\|AJD82340.1\| queuine tRNA-ribosyltransferase [Mycobacterium | 99.6124 | 99.6124 | 100 | 257 | 1 | 258 | 1 | 258 | 0 | 0 |

## c. SIF: HHPred



| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | 2ASH_C | Queuine tRNA-ribosyltransferase; tm1561, Queuine tRNA-ribosyltransferase, tRNA-guanine, Structural Genomics, Joint Center for Structural Genomics, JCSG, Protein Structure Initiative, PSI; HET: EDO; 1.9A {Thermotoga maritima} SCOP: c.1.20.1 | 99.9 | 92.2481 | 61 | 333 | 18 | 256 | 9.3e-23 |
| ☑ | PF01702.22 | TGT ; Queuine tRNA-ribosyltransferase | 99.9 | 91.8605 | 43 | 314 | 18 | 255 | 1.4e-22 |
| ☐ | 7UI4_A | DNA-guanine transglycosylase; 2'-deoxy-7-cyano-7-deazaguanosine, dPreQ0, DNA modification, DNA-guanine transglycosylase, DNA | 99.9 | 77.5194 | 68 | 284 | 20 | 220 | 2e-22 |

d. SIF: Synteny-Phamerator (three genomes)

Aligns with Waterdiva, Usavi, Buckeye, DaddyDaniels, GeneCoco, etc.

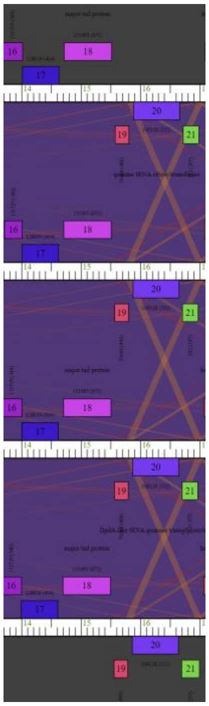


*Figure 1. (Top to bottom) Buckeye, Eugenia, Usavi, and Waterdiva.*

## 7. Any other important information.

Reverse.

## CURATOR NAME: ISABELLA LIMA

## GENE NAME: EUGENIA_DRAFT_21

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

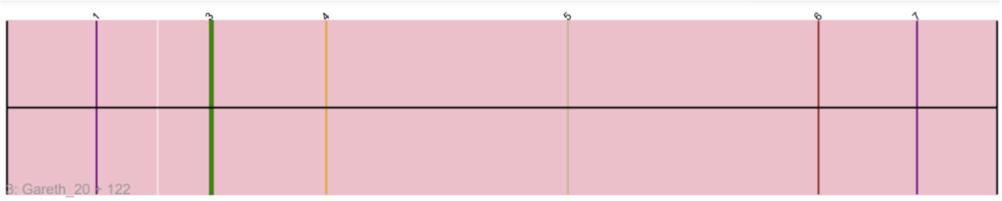
### 1. Starterator

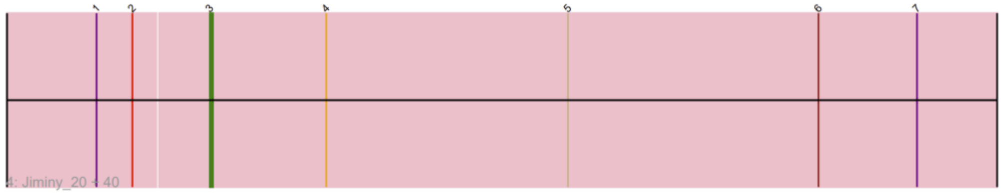
Eugenia is in Track 3 with start site 3.

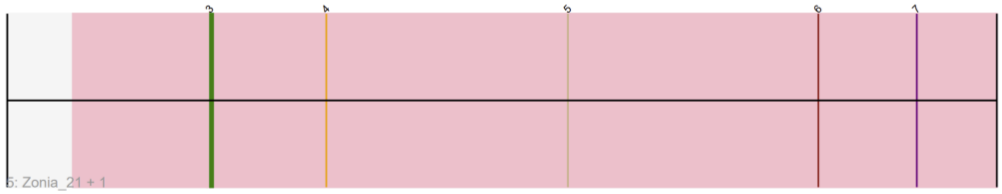
The start number called the most often in the published annotations is 3, it was called in 148 of the 243 non-draft genes in the pham.

Start 3:

- Found in 261 of 261 ( 100.0% ) of genes in pham
- Manual Annotations of this start: 148 of 243
- Called 63.6% of time when present

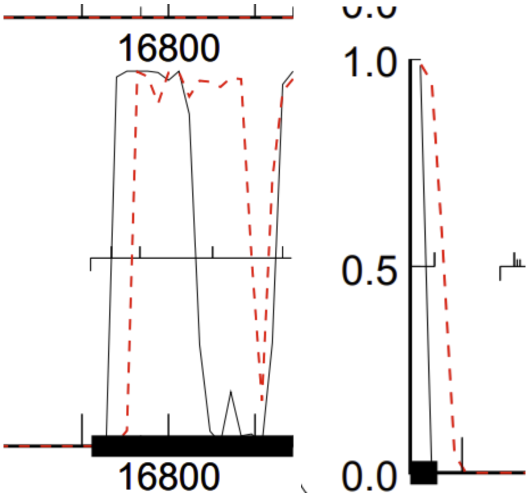






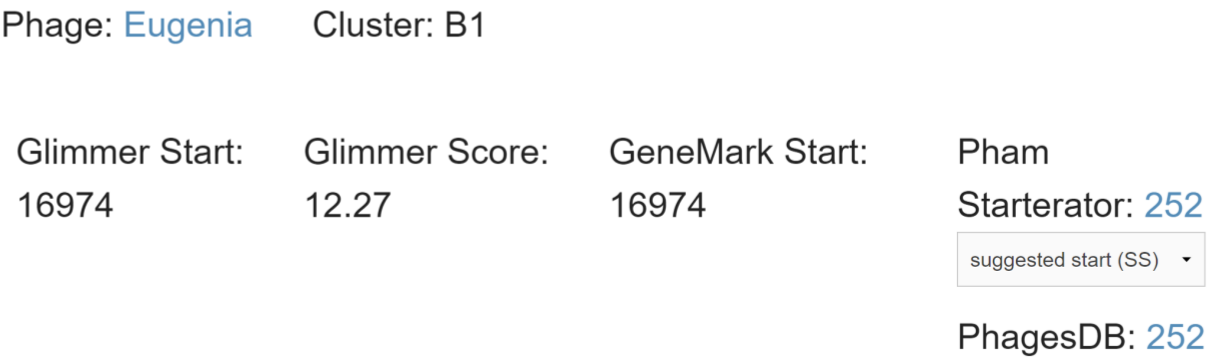
## 2. GeneMark coding potential

GeneMark: 16974



## 3. Glimmer and GeneMark agreement

YES



Phage: Eugenia Cluster: B1

Glimmer Start: 16974

Glimmer Score: 12.27

GeneMark Start: 16974

Pham Starterator: 252

suggested start (SS) ▾

PhagesDB: 252

Glimmer: 16974

GeneMark: 16974

**Note that Glimmer and GeneMark do not agree with the longest ORF without excessive gap.

## 4. Longest open reading frame (ORF) without excessive gap

No.

| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 17010 | 16711 | 300 | 73 | 14 | 1.821 | -5.794 | TRUE | TTG |  | ☐ |
| Reverse | 16974 | 16711 | 264 | 109 | 7 | 3.147 | -3.307 |  | ATG | Select ▾ | ☑ |

## 5. Function. Please follow this [Official SEAPHAGE Function List]

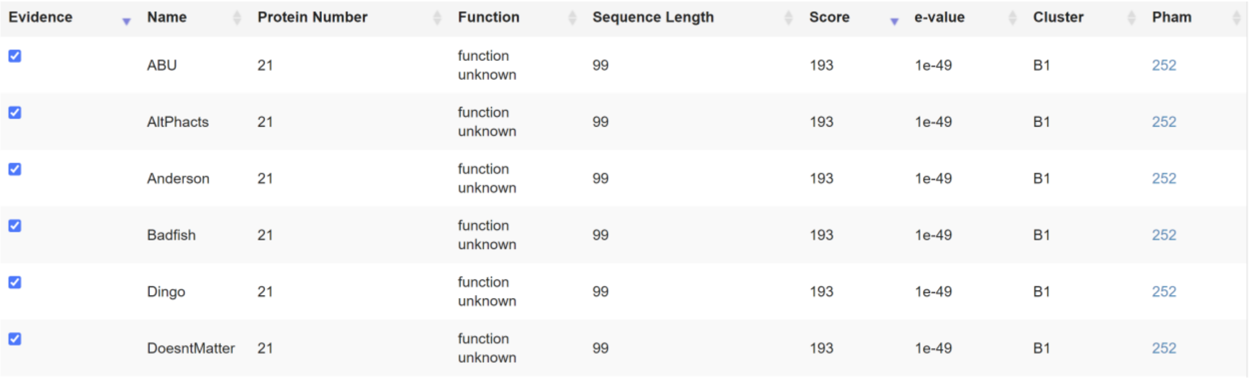
## If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein—function unknown

## 6. Supporting Information for Function (SIF)

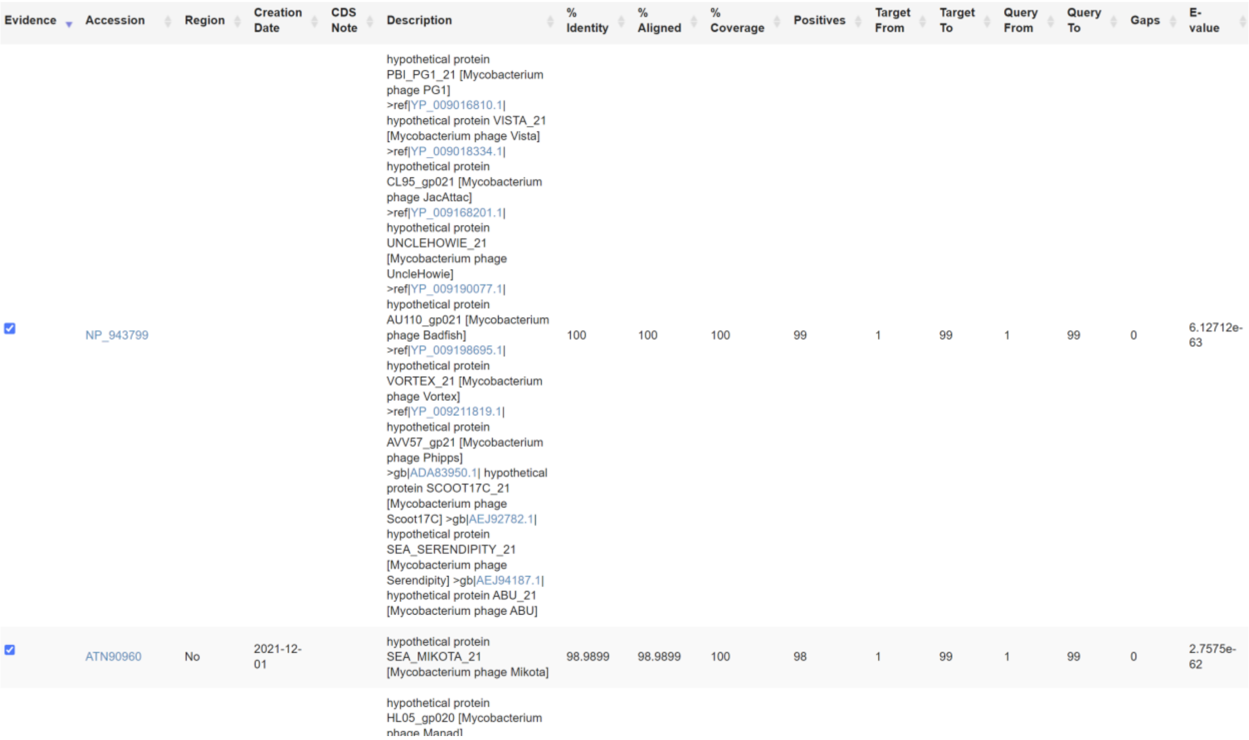
### a. SIF: PhageDb BLAST

```
                                                                 Score    E
Sequences producing significant alignments:                      (bits) Value

Zelda_21, function unknown, 87                                      169   3e-42
Zaider_22, function unknown, 87                                     169   3e-42
Yoshand_21, function unknown, 87                                    169   3e-42
Vortex_21, function unknown, 99                                     169   3e-42
Vista_21, function unknown, 99                                      169   3e-42
Virapocalypse_21, function unknown, 99                              169   3e-42
Valjean_21, function unknown, 87                                    169   3e-42
Vaishali24_20, function unknown, 87                                 169   3e-42
UncleHowie_21, function unknown, 99                                 169   3e-42
True_20, function unknown, 87                                       169   3e-42
TomBombadil_21, function unknown, 87                                169   3e-42
Timmi_20, function unknown, 87                                      169   3e-42
Thora_21, function unknown, 87                                      169   3e-42
TallGrassMM_21, function unknown, 99                                169   3e-42
Surely_21, function unknown, 87                                     169   3e-42
Squid_21, function unknown, 99                                      169   3e-42
Spartan300_21, function unknown, 87                                 169   3e-42
Sophia_20, function unknown, 87                                     169   3e-42
Solosis_20, function unknown, 87                                    169   3e-42
Slatt_21, function unknown, 87                                      169   3e-42
```



| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | ABU | 21 | function unknown | 99 | 193 | 1e-49 | B1 | 252 |
| ☑ | AltPhacts | 21 | function unknown | 99 | 193 | 1e-49 | B1 | 252 |
| ☑ | Anderson | 21 | function unknown | 99 | 193 | 1e-49 | B1 | 252 |
| ☑ | Badfish | 21 | function unknown | 99 | 193 | 1e-49 | B1 | 252 |
| ☑ | Dingo | 21 | function unknown | 99 | 193 | 1e-49 | B1 | 252 |
| ☑ | DoesntMatter | 21 | function unknown | 99 | 193 | 1e-49 | B1 | 252 |

## b. SIF: NCBI BLAST



| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | NP_943799 | | | | hypothetical protein PBI_PG1_21 [Mycobacterium phage PG1] >ref\|YP_009016810.1\| hypothetical protein VISTA_21 [Mycobacterium phage Vista] >ref\|YP_009018334.1\| hypothetical protein CL95_gp021 [Mycobacterium phage JacAttac] >ref\|YP_009168201.1\| hypothetical protein UNCLEHOWIE_21 [Mycobacterium phage UncleHowie] >ref\|YP_009190077.1\| hypothetical protein AU110_gp021 [Mycobacterium phage Badfish] >ref\|YP_009198695.1\| hypothetical protein VORTEX_21 [Mycobacterium phage Vortex] >ref\|YP_009211819.1\| hypothetical protein AVV57_gp21 [Mycobacterium phage Phipps] >gb\|ADA83950.1\| hypothetical protein SCOOT17C_21 [Mycobacterium phage Scoot17C] >gb\|AEJ92782.1\| hypothetical protein SEA_SERENDIPITY_21 [Mycobacterium phage Serendipity] >gb\|AEJ94187.1\| hypothetical protein ABU_21 [Mycobacterium phage ABU] | 100 | 100 | 100 | 99 | 1 | 99 | 1 | 99 | 0 | 6.12712e-63 |
| ☑ | ATN90960 | No | 2021-12-01 | | hypothetical protein SEA_MIKOTA_21 [Mycobacterium phage Mikota] | 98.9899 | 98.9899 | 100 | 98 | 1 | 99 | 1 | 99 | 0 | 2.7575e-62 |
| | | | | | hypothetical protein HL05_gp020 [Mycobacterium phage Manad] | | | | | | | | | | |

## c. SIF: HHPred

N/A

## d. SIF: Synteny-Phamerator (three genomes)

Aligns with Buckeye, Usavi, Waterdiva, ABU, Dingo, etc.

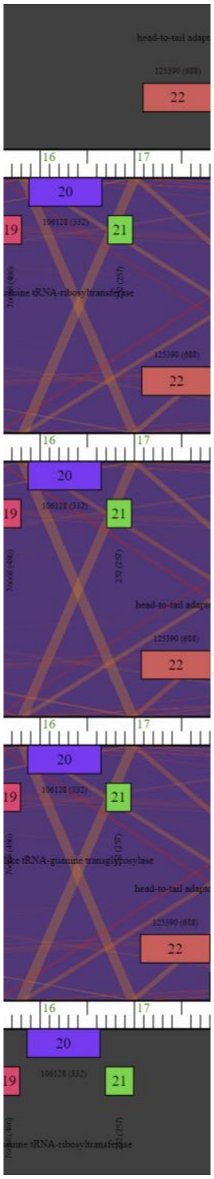


*Figure 1. (Top to bottom) Buckeye, Eugenia, Usavi, and Waterdiva.*

## 7. Any other important information.

Note that Glimmer and GeneMark do not agree with the longest ORF without excessive gap.

## CURATOR NAME: ISABELLA LIMA

## GENE NAME: EUGENIA_DRAFT_22

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

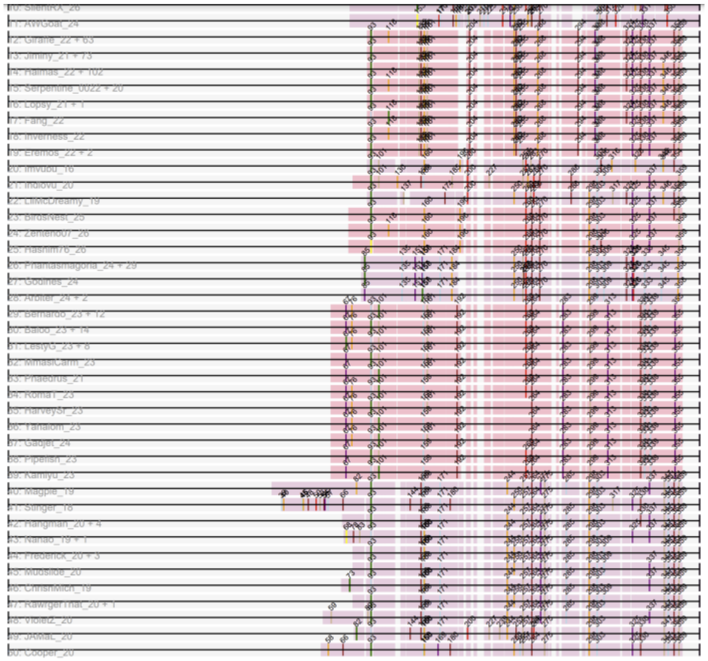
# 1. Starterator

Eugenia is in Track 14 with start 93.

The start number called the most often in the published annotations is 93, it was called in 305 of the 597 non-draft genes in the pham.

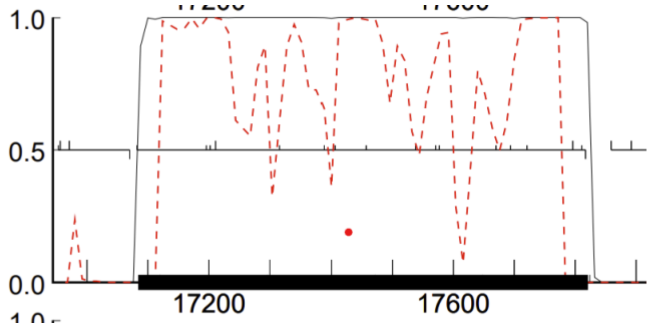
Start 93:

- Found in 385 of 700 ( 55.0% ) of genes in pham
- Manual Annotations of this start: 305 of 597
- Called 89.9% of time when present



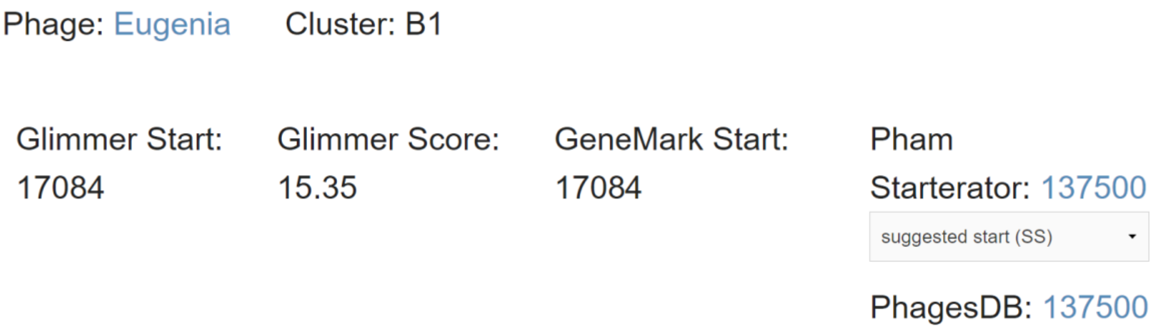
## 2. GeneMark coding potential

GeneMark: 17084



## 3. Glimmer and GeneMark agreement

Yes



Phage: Eugenia Cluster: B1

| Glimmer Start: | Glimmer Score: | GeneMark Start: | Pham |
|---|---|---|---|
| 17084 | 15.35 | 17084 | Starterator: 137500 |

PhagesDB: 137500

Glimmer: 17084

GeneMark: 17084

## 4. Longest open reading frame (ORF) without excessive gap

YES

| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 17084 | 17821 | 738 | 73 | 14 | 1.878 | -5.680 | TRUE | GTG | Select | ☑ |

Start: 17804

Length: 738

Gap: 73

Z-score: 1.878

Final score: -5.680

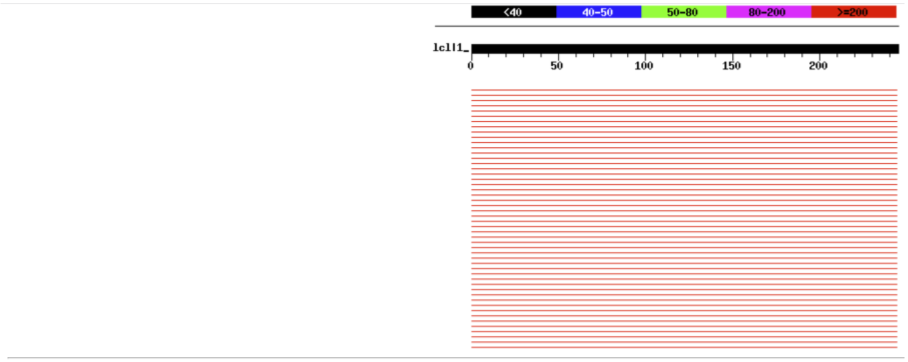
## 5. Function. Please follow this [Official SEAPHAGE Function List]

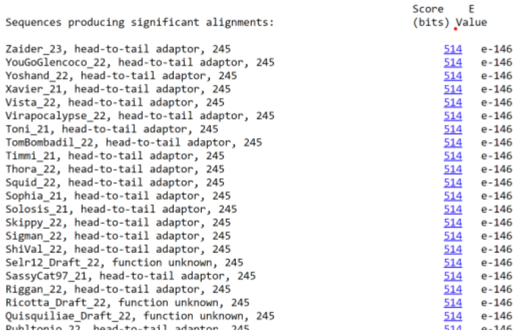
## If no functional prediction is present, write "Hypothetical protein".

Head-to-tail adaptor (Rank 3)

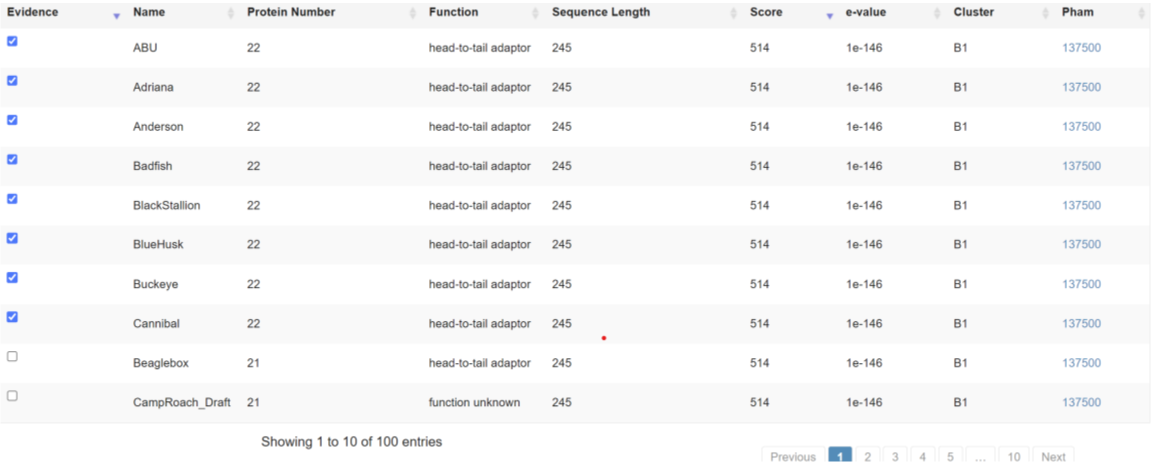
## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST





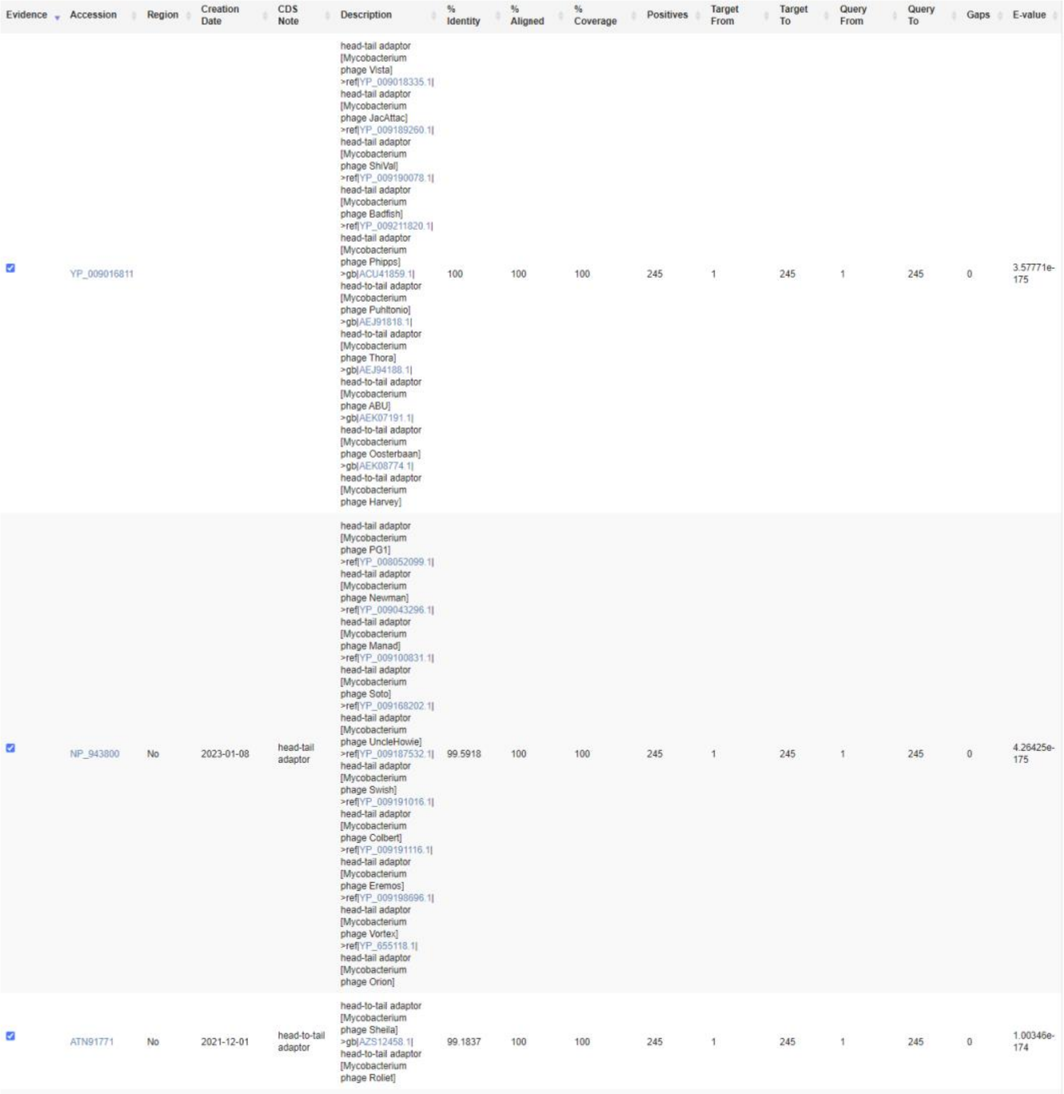
```
                                                              Score    E
Sequences producing significant alignments:                   (bits) Value

Zaider_23, head-to-tail adaptor, 245                             514   e-146
YouGoGlencoco_22, head-to-tail adaptor, 245                      514   e-146
Yoshand_22, head-to-tail adaptor, 245                            514   e-146
Xavier_21, head-to-tail adaptor, 245                             514   e-146
Vista_22, head-to-tail adaptor, 245                              514   e-146
Virapocalypse_22, head-to-tail adaptor, 245                      514   e-146
Toni_21, head-to-tail adaptor, 245                               514   e-146
TomBombadil_22, head-to-tail adaptor, 245                        514   e-146
Timmi_21, head-to-tail adaptor, 245                              514   e-146
Thora_22, head-to-tail adaptor, 245                              514   e-146
Squid_22, head-to-tail adaptor, 245                              514   e-146
Sophia_21, head-to-tail adaptor, 245                             514   e-146
Solosis_21, head-to-tail adaptor, 245                            514   e-146
Skippy_22, head-to-tail adaptor, 245                             514   e-146
Sigman_22, head-to-tail adaptor, 245                             514   e-146
ShiVal_22, head-to-tail adaptor, 245                             514   e-146
Selr12_Draft_22, function unknown, 245                           514   e-146
SassyCat97_21, head-to-tail adaptor, 245                         514   e-146
Riggan_22, head-to-tail adaptor, 245                             514   e-146
Ricotta_Draft_22, function unknown, 245                          514   e-146
Quisquiliae_Draft_22, function unknown, 245                      514   e-146
```



| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | ABU | 22 | head-to-tail adaptor | 245 | 514 | 1e-146 | B1 | 137500 |
| ☑ | Adriana | 22 | head-to-tail adaptor | 245 | 514 | 1e-146 | B1 | 137500 |
| ☑ | Anderson | 22 | head-to-tail adaptor | 245 | 514 | 1e-146 | B1 | 137500 |
| ☑ | Badfish | 22 | head-to-tail adaptor | 245 | 514 | 1e-146 | B1 | 137500 |
| ☑ | BlackStallion | 22 | head-to-tail adaptor | 245 | 514 | 1e-146 | B1 | 137500 |
| ☑ | BlueHusk | 22 | head-to-tail adaptor | 245 | 514 | 1e-146 | B1 | 137500 |
| ☑ | Buckeye | 22 | head-to-tail adaptor | 245 | 514 | 1e-146 | B1 | 137500 |
| ☑ | Cannibal | 22 | head-to-tail adaptor | 245 | 514 | 1e-146 | B1 | 137500 |
| ☐ | Beaglebox | 21 | head-to-tail adaptor | 245 | 514 | 1e-146 | B1 | 137500 |
| ☐ | CampRoach_Draft | 21 | function unknown | 245 | 514 | 1e-146 | B1 | 137500 |

Showing 1 to 10 of 100 entries

## b. SIF: NCBI BLAST



| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | YP_009016811 | | | | head-tail adaptor [Mycobacterium phage Vista] >ref\|YP_009018335.1\| head-tail adaptor [Mycobacterium phage JacAttac] >ref\|YP_009189260.1\| head-tail adaptor [Mycobacterium phage ShiVal] >ref\|YP_009190078.1\| head-tail adaptor [Mycobacterium phage Badfish] >ref\|YP_009211820.1\| head-tail adaptor [Mycobacterium phage Phipps] >gb\|ACU41859.1\| head-to-tail adaptor [Mycobacterium phage Puhltonio] >gb\|AEJ91818.1\| head-to-tail adaptor [Mycobacterium phage Thora] >gb\|AEJ94188.1\| head-to-tail adaptor [Mycobacterium phage ABU] >gb\|AEK07191.1\| head-to-tail adaptor [Mycobacterium phage Oosterbaan] >gb\|AEK08774.1\| head-to-tail adaptor [Mycobacterium phage Harvey] | 100 | 100 | 100 | 245 | 1 | 245 | 1 | 245 | 0 | 3.57771e-175 |
| ☑ | NP_943800 | No | 2023-01-08 | head-tail adaptor | head-tail adaptor [Mycobacterium phage PG1] >ref\|YP_008052099.1\| head-tail adaptor [Mycobacterium phage Newman] >ref\|YP_009043296.1\| head-tail adaptor [Mycobacterium phage Manad] >ref\|YP_009100831.1\| head-tail adaptor [Mycobacterium phage Soto] >ref\|YP_009168202.1\| head-tail adaptor [Mycobacterium phage UncleHowie] >ref\|YP_009187532.1\| head-tail adaptor [Mycobacterium phage Swish] >ref\|YP_009191016.1\| head-tail adaptor [Mycobacterium phage Colbert] >ref\|YP_009191116.1\| head-tail adaptor [Mycobacterium phage Eremos] >ref\|YP_009198696.1\| head-tail adaptor [Mycobacterium phage Vortex] >ref\|YP_655118.1\| head-tail adaptor [Mycobacterium phage Orion] | 99.5918 | 100 | 100 | 245 | 1 | 245 | 1 | 245 | 0 | 4.26425e-175 |
| ☑ | ATN91771 | No | 2021-12-01 | head-to-tail adaptor | head-to-tail adaptor [Mycobacterium phage Sheila] >gb\|AZS12458.1\| head-to-tail adaptor [Mycobacterium phage Roliet] | 99.1837 | 100 | 100 | 245 | 1 | 245 | 1 | 245 | 0 | 1.00346e-174 |

## c. SIF: HHPred

N/A

### d. SIF: Synteny-Phamerator (three genomes)

Aligns with Buckeye, Waterdiva, Usavi, BlackStallion, Cannibal, etc.

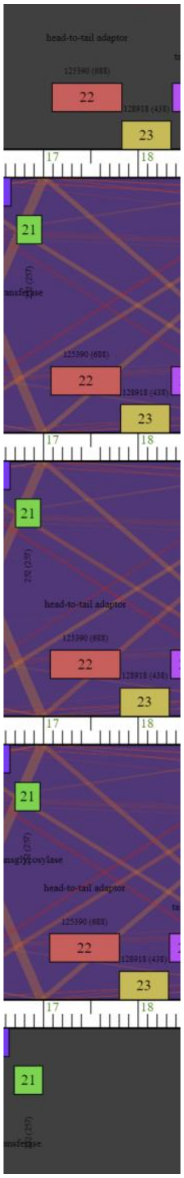


*Figure 1. (Top to bottom) Buckeye, Eugenia, Usavi, and Waterdiva.*

## 7. Any other important information.

N/A

## CURATOR NAME: ISABELLA LIMA

## GENE NAME: EUGENIA_DRAFT_23

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

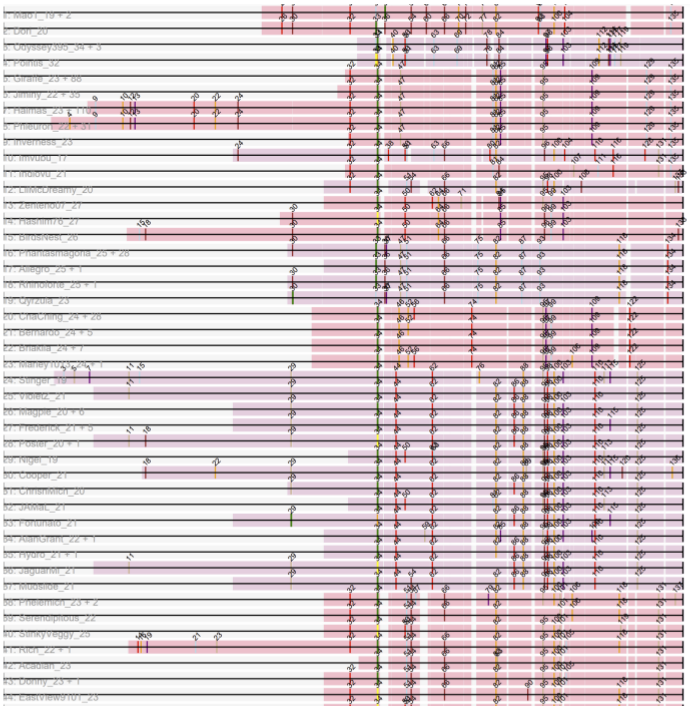
## 1. Starterator

Eugenia is in Track 7 with start 34.

The start number called the most often in the published annotations is 34, it was called in 337 of the 396 non-draft genes in the pham.

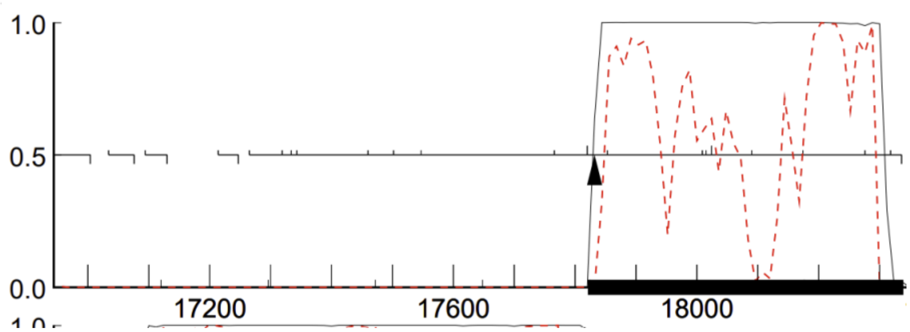
Start 34:

- Found in 384 of 453 ( 84.8% ) of genes in pham
- Manual Annotations of this start: 337 of 396
- Called 98.7% of time when present

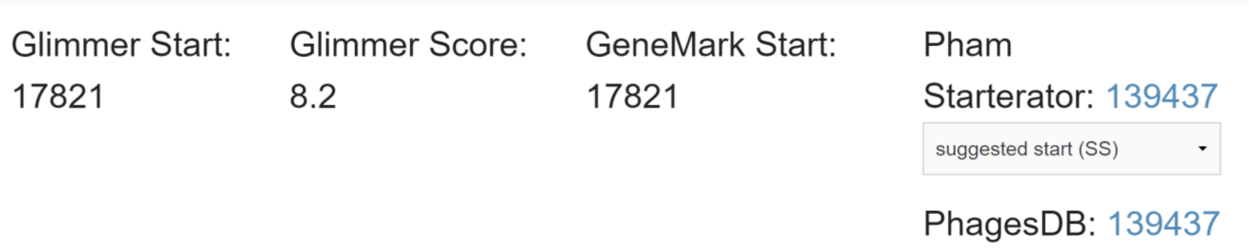


## 2. GeneMark coding potential

GeneMark: 17821



## 3. Glimmer and GeneMark agreement



| Glimmer Start: | Glimmer Score: | GeneMark Start: | Pham Starterator: |
|---|---|---|---|
| 17821 | 8.2 | 17821 | 139437 |

PhagesDB: 139437

## 4. Longest open reading frame (ORF) without excessive gap

YES

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 17767 | 18339 | 573 | -55 | 12 | 1.32 | -6.290 | GTG | | ☐ |
| Forward | 17821 | 18339 | 519 | -1 | 9 | 2.139 | -4.583 | ATG | Select | ☑ |
| Forward | 17854 | 18339 | 486 | 32 | 18 | 1.185 | -8.025 | GTG | | ☐ |

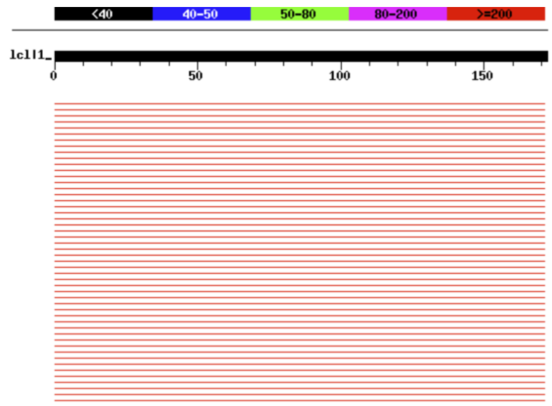
## 5. Function. Please follow this Official SEAPHAGE Function List

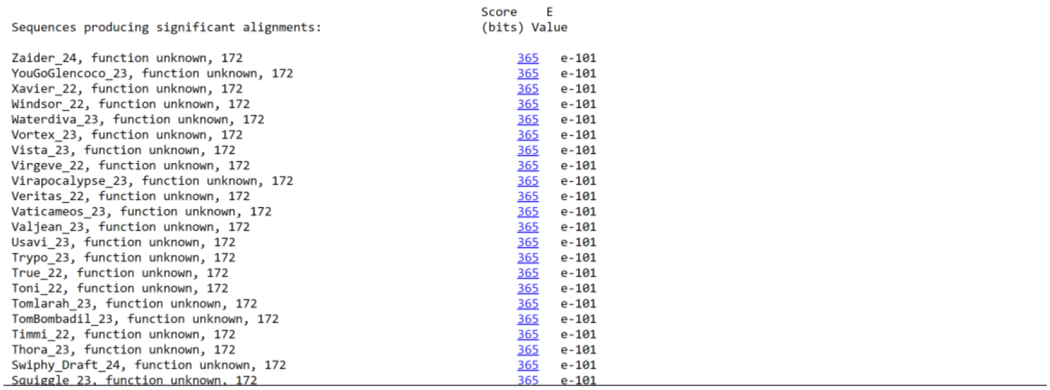
## If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

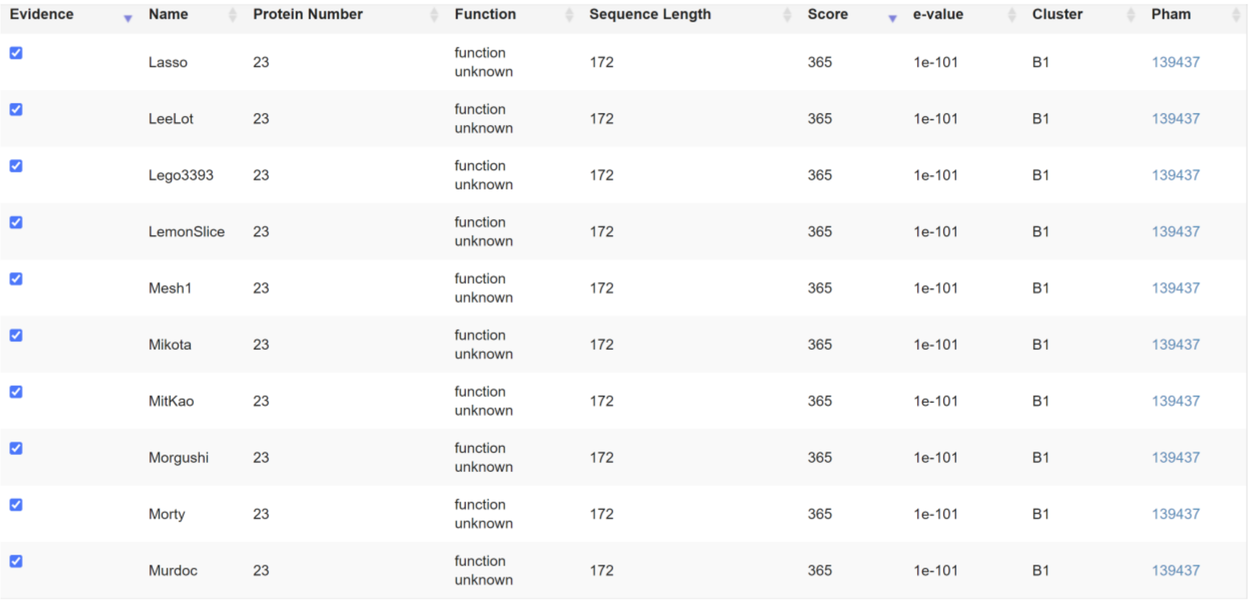
# 6. Supporting Information for Function (SIF)

## a. SIF: PhageDb BLAST



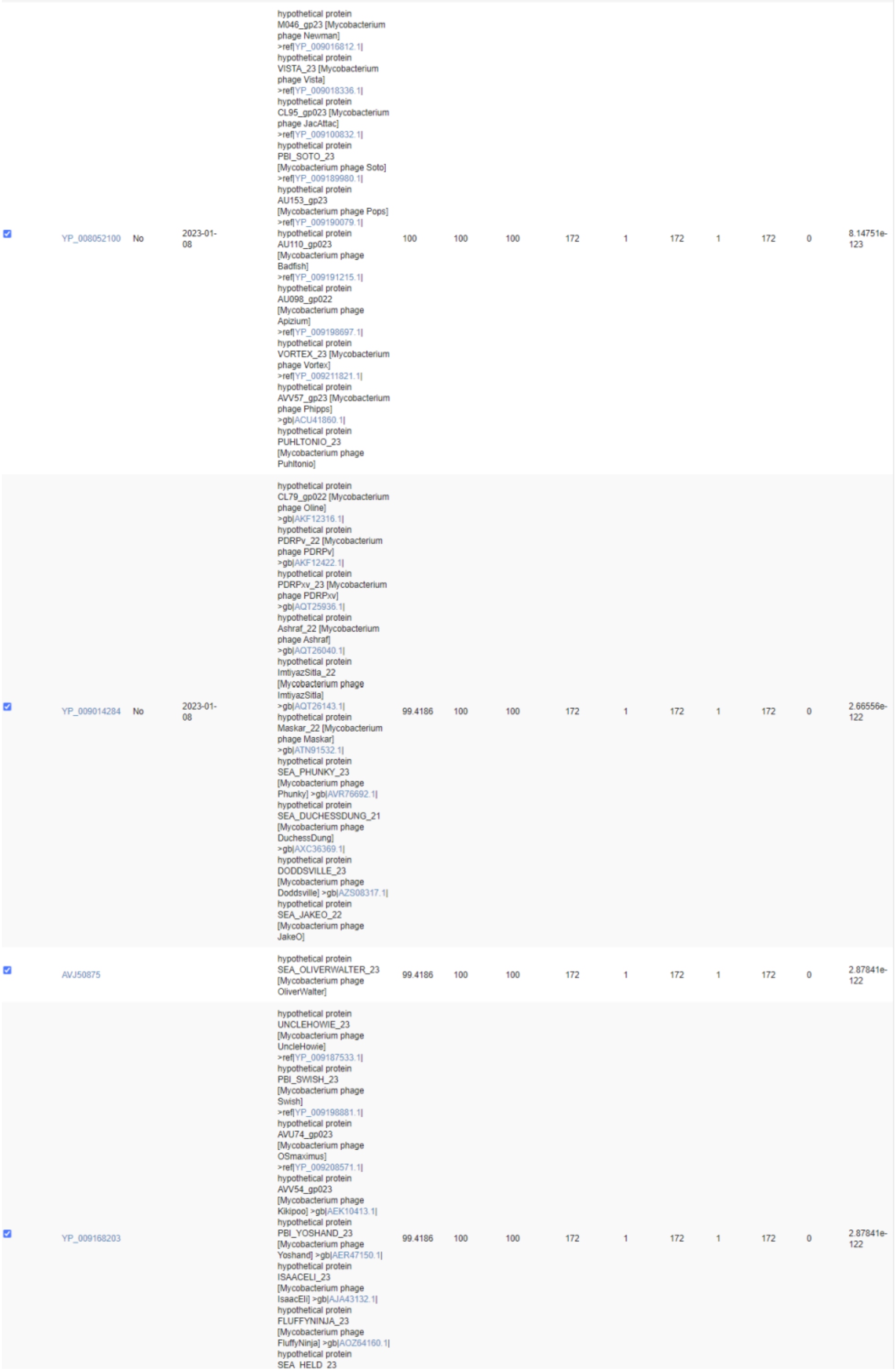


```
                                                                 Score    E
Sequences producing significant alignments:                      (bits) Value

Zaider_24, function unknown, 172                                     365   e-101
YouGoGlencoco_23, function unknown, 172                              365   e-101
Xavier_22, function unknown, 172                                     365   e-101
Windsor_22, function unknown, 172                                    365   e-101
Waterdiva_23, function unknown, 172                                  365   e-101
Vortex_23, function unknown, 172                                     365   e-101
Vista_23, function unknown, 172                                      365   e-101
Virgeve_22, function unknown, 172                                    365   e-101
Virapocalypse_23, function unknown, 172                              365   e-101
Veritas_22, function unknown, 172                                    365   e-101
Vaticameos_23, function unknown, 172                                 365   e-101
Valjean_23, function unknown, 172                                    365   e-101
Usavi_23, function unknown, 172                                      365   e-101
Trypo_23, function unknown, 172                                      365   e-101
True_22, function unknown, 172                                       365   e-101
Toni_22, function unknown, 172                                       365   e-101
Tomlarah_23, function unknown, 172                                   365   e-101
TomBombadil_23, function unknown, 172                                365   e-101
Timmi_22, function unknown, 172                                      365   e-101
Thora_23, function unknown, 172                                      365   e-101
Swiphy_Draft_24, function unknown, 172                               365   e-101
Squiggle_23, function unknown, 172                                   365   e-101
```



| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | Lasso | 23 | function unknown | 172 | 365 | 1e-101 | B1 | 139437 |
| ☑ | LeeLot | 23 | function unknown | 172 | 365 | 1e-101 | B1 | 139437 |
| ☑ | Lego3393 | 23 | function unknown | 172 | 365 | 1e-101 | B1 | 139437 |
| ☑ | LemonSlice | 23 | function unknown | 172 | 365 | 1e-101 | B1 | 139437 |
| ☑ | Mesh1 | 23 | function unknown | 172 | 365 | 1e-101 | B1 | 139437 |
| ☑ | Mikota | 23 | function unknown | 172 | 365 | 1e-101 | B1 | 139437 |
| ☑ | MitKao | 23 | function unknown | 172 | 365 | 1e-101 | B1 | 139437 |
| ☑ | Morgushi | 23 | function unknown | 172 | 365 | 1e-101 | B1 | 139437 |
| ☑ | Morty | 23 | function unknown | 172 | 365 | 1e-101 | B1 | 139437 |
| ☑ | Murdoc | 23 | function unknown | 172 | 365 | 1e-101 | B1 | 139437 |

# b. SIF: NCBI BLAST



| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | YP_008052100 | No | 2023-01-08 | hypothetical protein M046_gp23 [Mycobacterium phage Newman] >ref\|YP_009016812.1\| hypothetical protein VISTA_23 [Mycobacterium phage Vista] >ref\|YP_009018336.1\| hypothetical protein CL95_gp023 [Mycobacterium phage JacAttac] >ref\|YP_009100832.1\| hypothetical protein PBI_SOTO_23 [Mycobacterium phage Soto] >ref\|YP_009189980.1\| hypothetical protein AU153_gp23 [Mycobacterium phage Pops] >ref\|YP_009190079.1\| hypothetical protein AU110_gp023 [Mycobacterium phage Badfish] >ref\|YP_009191215.1\| hypothetical protein AU098_gp022 [Mycobacterium phage Apizium] >ref\|YP_009198697.1\| hypothetical protein VORTEX_23 [Mycobacterium phage Vortex] >ref\|YP_009211821.1\| hypothetical protein AVV57_gp23 [Mycobacterium phage Phipps] >gb\|ACU41860.1\| hypothetical protein PUHLTONIO_23 [Mycobacterium phage Puhltonio] | 100 | 100 | 100 | 172 | 1 | 172 | 1 | 172 | 0 | 8.14751e-123 |
| ☑ | YP_009014284 | No | 2023-01-08 | hypothetical protein CL79_gp022 [Mycobacterium phage Oline] >gb\|AKF12316.1\| hypothetical protein PDRPv_22 [Mycobacterium phage PDRPv] >gb\|AKF12422.1\| hypothetical protein PDRPxv_23 [Mycobacterium phage PDRPxv] >gb\|AQT25936.1\| hypothetical protein Ashraf_22 [Mycobacterium phage Ashraf] >gb\|AQT26040.1\| hypothetical protein ImtiyazSitla_22 [Mycobacterium phage ImtiyazSitla] >gb\|AQT26143.1\| hypothetical protein Maskar_22 [Mycobacterium phage Maskar] >gb\|ATN91532.1\| hypothetical protein SEA_PHUNKY_23 [Mycobacterium phage Phunky] >gb\|AVR76692.1\| hypothetical protein SEA_DUCHESSDUNG_21 [Mycobacterium phage DuchessDung] >gb\|AXC36369.1\| hypothetical protein DODDSVILLE_23 [Mycobacterium phage Doddsville] >gb\|AZS08317.1\| hypothetical protein SEA_JAKEO_22 [Mycobacterium phage JakeO] | 99.4186 | 100 | 100 | 172 | 1 | 172 | 1 | 172 | 0 | 2.66556e-122 |
| ☑ | AVJ50875 | | | hypothetical protein SEA_OLIVERWALTER_23 [Mycobacterium phage OliverWalter] | 99.4186 | 100 | 100 | 172 | 1 | 172 | 1 | 172 | 0 | 2.87841e-122 |
| ☑ | YP_009168203 | | | hypothetical protein UNCLEHOWIE_23 [Mycobacterium phage UncleHowie] >ref\|YP_009187533.1\| hypothetical protein PBI_SWISH_23 [Mycobacterium phage Swish] >ref\|YP_009198881.1\| hypothetical protein AVU74_gp023 [Mycobacterium phage OSmaximus] >ref\|YP_009208571.1\| hypothetical protein AVV54_gp023 [Mycobacterium phage Kikipoo] >gb\|AEK10413.1\| hypothetical protein PBI_YOSHAND_23 [Mycobacterium phage Yoshand] >gb\|AER47150.1\| hypothetical protein ISAACELI_23 [Mycobacterium phage IsaacEli] >gb\|AJA43132.1\| hypothetical protein FLUFFYNINJA_23 [Mycobacterium phage FluffyNinja] >gb\|AOZ64160.1\| hypothetical protein SEA_HELD_23 | 99.4186 | 100 | 100 | 172 | 1 | 172 | 1 | 172 | 0 | 2.87841e-122 |

### c. SIF: HHPred

N/A

### d. SIF: Synteny-Phamerator (three genomes)

Aligns with Buckeye, Waterdiva, Usavi, Lasso, Morty, etc.

*Figure 1. (Top to bottom) Buckeye, Eugenia, Usavi, and Waterdiva.*

## 7. Any other important information.

N/A

## CURATOR NAME: ISABELLA LIMA

## GENE NAME: EUGENIA_DRAFT_24

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

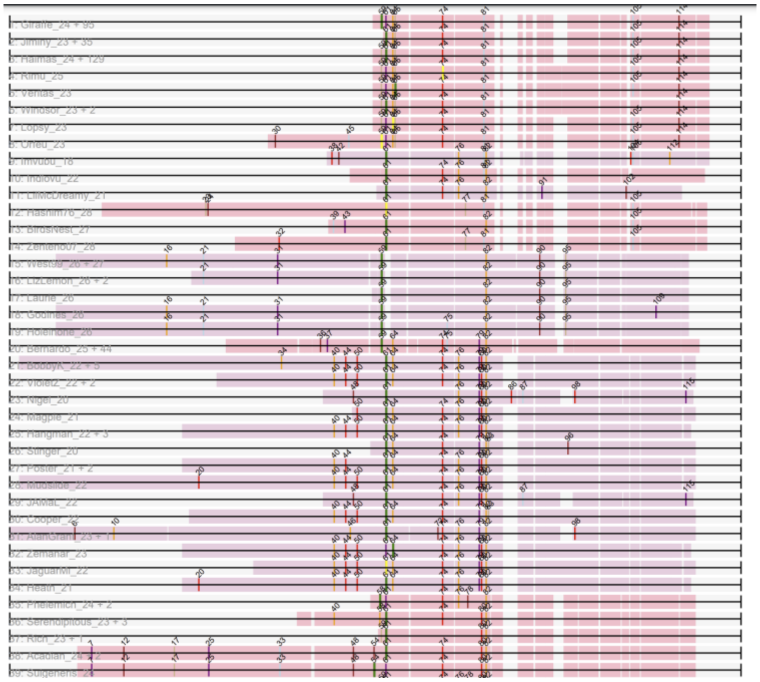
### 1. Starterator

Eugenia is in Track 3 with start 61.

The start number called the most often in the published annotations is 61, it was called in 191 of the 447 non-draft genes in the pham.

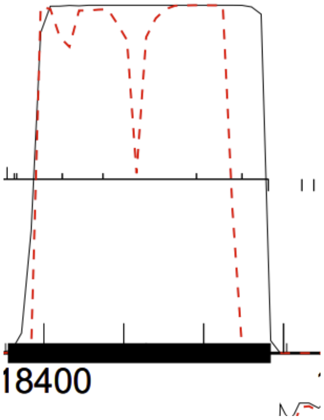
Start 61:

- Found in 328 of 514 ( 63.8% ) of genes in pham
- Manual Annotations of this start: 191 of 447
- Called 67.4% of time when present



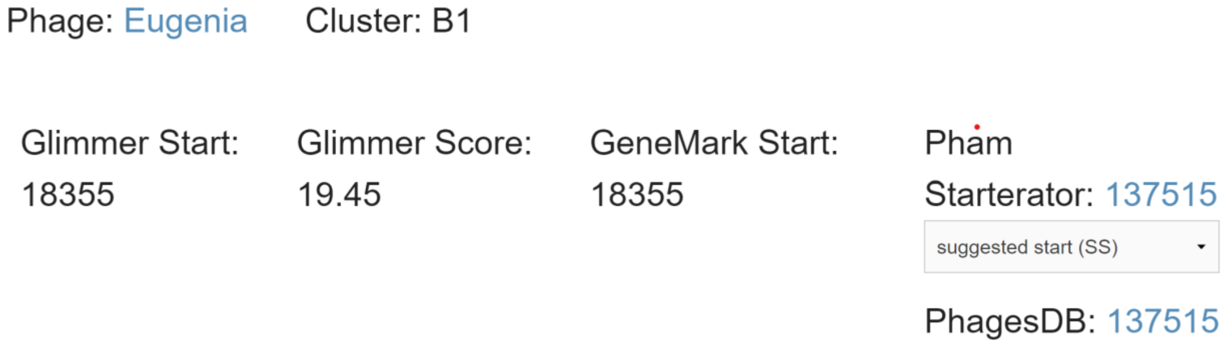
## 2. GeneMark coding potential

GeneMark: 18355



## 3. Glimmer and GeneMark agreement

YES



Phage: Eugenia Cluster: B1

Glimmer Start: 18355

Glimmer Score: 19.45

GeneMark Start: 18355

Pham

Starterator: 137515

suggested start (SS)

PhagesDB: 137515

Glimmer: 18355

GeneMark: 18355

**Please note that Glimmer and GeneMark do NOT agree with the longest ORF with shortest gap’s start site.

## 4. Longest open reading frame (ORF) without excessive gap

No.

| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 18349 | 18684 | 336 | 9 | 5 | 2.471 | -5.142 | TRUE | ATG | | ☐ |
| Forward | 18355 | 18684 | 330 | 15 | 11 | 2.471 | -3.899 | | ATG | Select | ☑ |

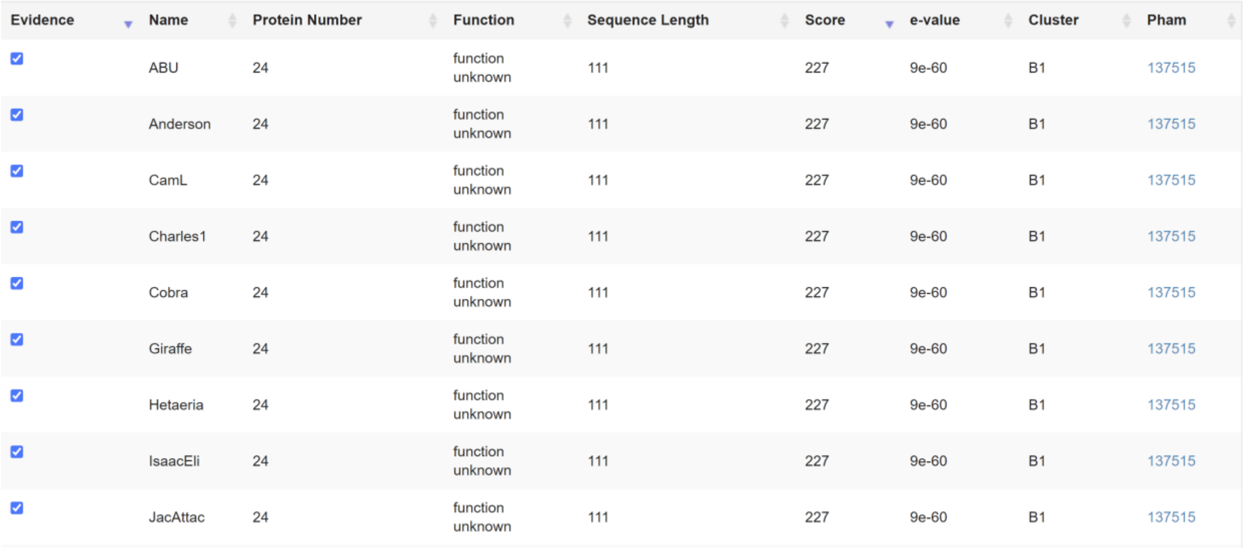
## 5. Function. Please follow this [Official SEAPHAGE Function List]

## If no functional prediction is present, write “Hypothetical protein”.

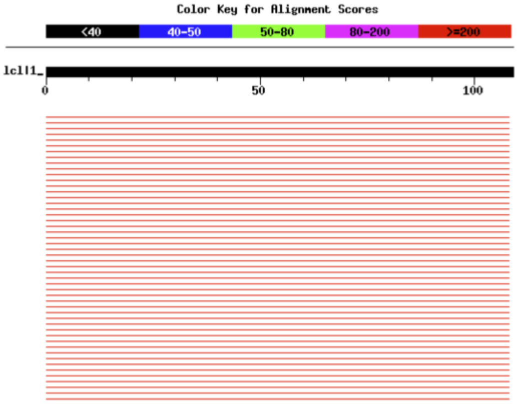
Hypothetical protein

## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST



| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | ABU | 24 | function unknown | 111 | 227 | 9e-60 | B1 | 137515 |
| ☑ | Anderson | 24 | function unknown | 111 | 227 | 9e-60 | B1 | 137515 |
| ☑ | CamL | 24 | function unknown | 111 | 227 | 9e-60 | B1 | 137515 |
| ☑ | Charles1 | 24 | function unknown | 111 | 227 | 9e-60 | B1 | 137515 |
| ☑ | Cobra | 24 | function unknown | 111 | 227 | 9e-60 | B1 | 137515 |
| ☑ | Giraffe | 24 | function unknown | 111 | 227 | 9e-60 | B1 | 137515 |
| ☑ | Hetaeria | 24 | function unknown | 111 | 227 | 9e-60 | B1 | 137515 |
| ☑ | IsaacEli | 24 | function unknown | 111 | 227 | 9e-60 | B1 | 137515 |
| ☑ | JacAttac | 24 | function unknown | 111 | 227 | 9e-60 | B1 | 137515 |



```
                                                                  Score    E
Sequences producing significant alignments:                       (bits) Value

Zelda_24, function unknown, 109                                        223   1e-58
Zaider_25, function unknown, 109                                       223   1e-58
YouGoGlencoco_24, function unknown, 111                                223   1e-58
Xavier_23, function unknown, 109                                       223   1e-58
Waterdiva_24, function unknown, 111                                    223   1e-58
Vivaldi_24, function unknown, 109                                      223   1e-58
Vista_24, function unknown, 111                                        223   1e-58
Virapocalypse_24, function unknown, 109                                223   1e-58
Vaticameos_24, function unknown, 111                                   223   1e-58
UncleHowie_24, function unknown, 111                                   223   1e-58
UAch1_24, function unknown, 109                                        223   1e-58
True_23, function unknown, 109                                         223   1e-58
Tomlarah_24, function unknown, 109                                     223   1e-58
TomBombadil_24, function unknown, 111                                  223   1e-58
Timmi_23, function unknown, 111                                        223   1e-58
Thora_24, function unknown, 111                                        223   1e-58
Telesworld_23, function unknown, 111                                   223   1e-58
TallGrassMM_24, function unknown, 111                                  223   1e-58
Surely_24, function unknown, 111                                       223   1e-58
Squid_24, function unknown, 111                                        223   1e-58
```

## b. SIF: NCBI BLAST

| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | NP_943802 | No | 2023-01-08 |  | hypothetical protein PBI_PG1_24 [Mycobacterium phage PG1] >ref\|YP_009016813.1\| hypothetical protein VISTA_24 [Mycobacterium phage Vista] >ref\|YP_009018337.1\| hypothetical protein CL95_gp024 [Mycobacterium phage JacAttac] >ref\|YP_009168204.1\| hypothetical protein UNCLEHOWIE_24 [Mycobacterium phage UncleHowie] >ref\|YP_009208572.1\| hypothetical protein AVV54_gp024 [Mycobacterium phage Kikipoo] >ref\|YP_009211822.1\| hypothetical protein AVV57_gp24 [Mycobacterium phage Phipps] >ref\|YP_655120.1\| gp24 [Mycobacterium phage Orion] >gb\|ACI12745.1\| hypothetical protein CHAH_25 [Mycobacterium phage Chah] >gb\|ACU41861.1\| hypothetical protein PUHLTONIO_24 [Mycobacterium phage Puhltonio] >gb\|ADA83953.1\| hypothetical protein SCOOT17C_24 [Mycobacterium phage Scoot17C] | 100 | 100 | 100 | 111 | 1 | 111 | 1 | 111 | 0 | 3.25515e-74 |
| ☑ | YP_009005671 | No | 2023-01-08 |  | hypothetical protein PBI_SUFFOLK_24 [Mycobacterium phage Suffolk] >ref\|YP_009189262.1\| gp24 [Mycobacterium phage ShiVal] >ref\|YP_009191018.1\| hypothetical protein AU159_gp024 [Mycobacterium phage Colbert] >ref\|YP_009198698.1\| hypothetical protein VORTEX_24 [Mycobacterium phage Vortex] >gb\|ADA83851.1\| hypothetical protein FANG_24 [Mycobacterium phage Fang] >gb\|AEJ92705.1\| hypothetical protein SEA_SERENDIPITY_24 [Mycobacterium phage Serendipity] >gb\|AEJ95210.1\| hypothetical protein KLUCKY39_24 [Mycobacterium phage KLucky39] >gb\|AEK08776.1\| hypothetical protein PBI_HARVEY_24 [Mycobacterium phage Harvey] >gb\|AEK10414.1\| hypothetical protein PBI_YOSHAND_24 [Mycobacterium phage Yoshand] >gb\|AEO94063.1\| hypothetical protein MORGUSHI_24 [Mycobacterium phage Morgushi] | 99.0991 | 99.0991 | 100 | 110 | 1 | 111 | 1 | 111 | 0 | 1.43302e-73 |
| ☑ | YP_009043298 | No | 2023-01-08 |  | hypothetical protein HL05_gp023 [Mycobacterium phage Manad] >gb\|AHZ95283.1\| hypothetical protein PBI_MANAD_23 [Mycobacterium phage Manad] | 99.0991 | 99.0991 | 100 | 110 | 1 | 111 | 1 | 111 | 0 | 3.64359e-73 |
| ☑ | YP_009190080 | No | 2023-01-08 |  | hypothetical protein AU110_gp024 [Mycobacterium phage Badfish] >ref\|YP_009191118.1\| hypothetical protein AU108_gp24 [Mycobacterium phage Eremos] >gb\|AEK09018.1\| hypothetical protein PBI_HERTUBISE_24 [Mycobacterium phage Hertubise] >gb\|AIM50256.1\| hypothetical protein PBI_VIVALDI_24 [Mycobacterium phage Vivaldi] >gb\|AJA43790.1\| hypothetical protein SIGMAN_24 [Mycobacterium phage Sigman] >gb\|AKO62300.1\| hypothetical protein PBI_PHAMISHED_24 [Mycobacterium phage Phamished] >gb\|AMS01486.1\| hypothetical protein PBI_POTTER_23 [Mycobacterium phage Potter] >gb\|AOQ27668.1\| hypothetical protein SEA_MANA_23 [Mycobacterium phage Mana] >gb\|AOQ28381.1\| hypothetical protein SEA_FRIARPREACHER_24 [Mycobacterium phage FriarPreacher] >gb\|AOT23496.1\| hypothetical protein SEA_IRIDOCLYSIS_24 [Mycobacterium phage Iridoclysis] | 100 | 100 | 98.1982 | 109 | 1 | 109 | 3 | 111 | 0 | 1.71507e-72 |
|  |  |  |  |  | hypothetical protein M046_gp24 [Mycobacterium |  |  |  |  |  |  |  |  |  |  |

## c. SIF: HHPred

N/A

### d. SIF: Synteny-Phamerator (three genomes)

Aligns with Buckeye, Waterdiva, Usavi, Giraffe, Cobra, etc.

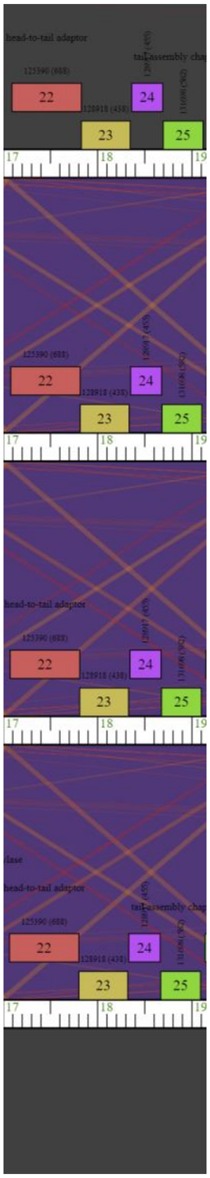


*Figure 1. (Top to bottom) Buckeye, Eugenia, Usavi, and Waterdiva.*

## 7. Any other important information.

Glimmer and GeneMark do NOT agree with the longest ORF with shortest gap’s start site.

## CURATOR NAME: ISABELLA LIMA

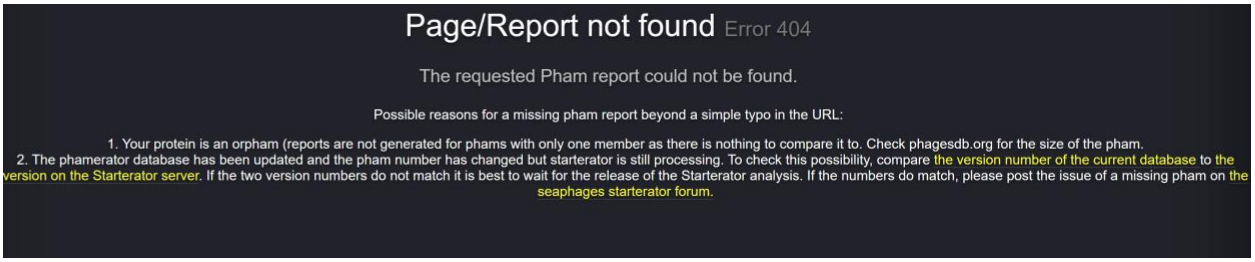
## GENE NAME: EUGENIA_DRAFT_25

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

### 1. Starterator

Waiting on New Release



Page/Report not found Error 404

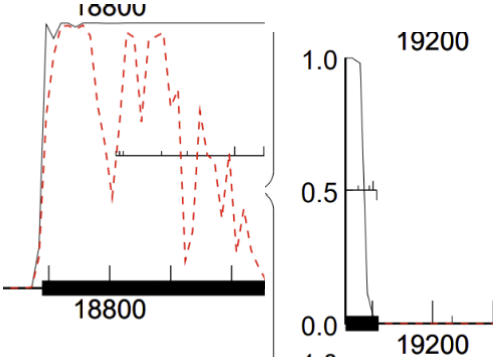
The requested Pham report could not be found.

Possible reasons for a missing pham report beyond a simple typo in the URL:

1. Your protein is an orpham (reports are not generated for phams with only one member as there is nothing to compare it to. Check phagesdb.org for the size of the pham.
2. The phamerator database has been updated and the pham number has changed but starterator is still processing. To check this possibility, compare the version number of the current database to the version on the Starterator server. If the two version numbers do not match it is best to wait for the release of the Starterator analysis. If the numbers do match, please post the issue of a missing pham on the seaphages starterator forum.

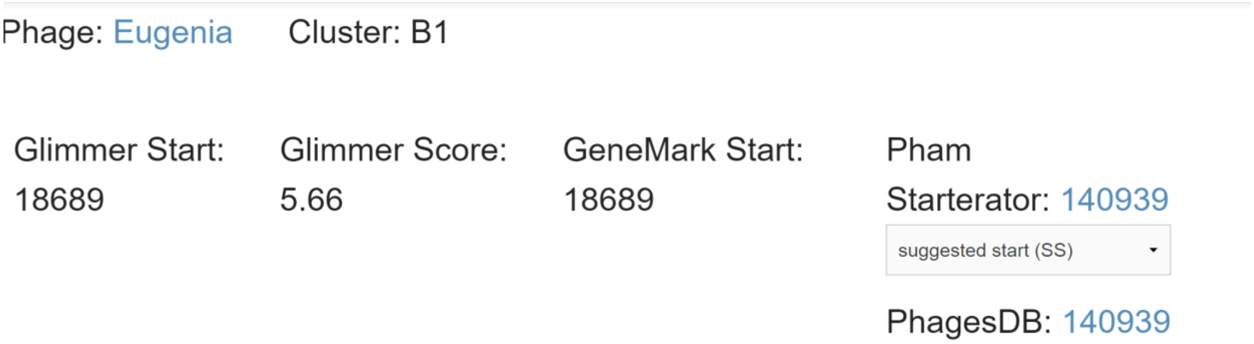
### 2. GeneMark coding potential

GeneMark: 18689



## 3. Glimmer and GeneMark agreement

YES



Phage: Eugenia Cluster: B1

Glimmer Start: 18689

Glimmer Score: 5.66

GeneMark Start: 18689

Pham Starterator: 140939

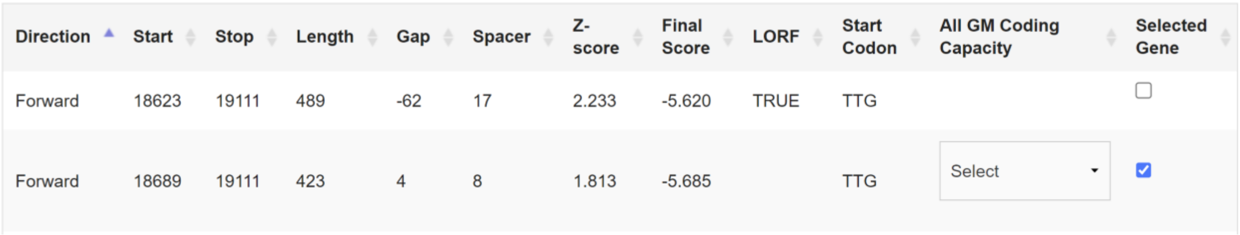
PhagesDB: 140939

Glimmer: 18689

GeneMark: 18689

## 4. Longest open reading frame (ORF) without excessive gap

YES



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 18623 | 19111 | 489 | -62 | 17 | 2.233 | -5.620 | TRUE | TTG |  | ☐ |
| Forward | 18689 | 19111 | 423 | 4 | 8 | 1.813 | -5.685 |  | TTG | Select | ☑ |

Start: 18689

Length: 423

Gap: 4

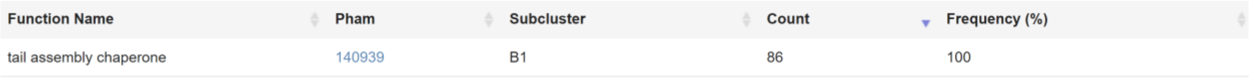
Z-score: 1.813

Final Score: -5.685

## 5. Function. Please follow this [Official SEAPHAGE Function List]()

## If no functional prediction is present, write “Hypothetical protein”.

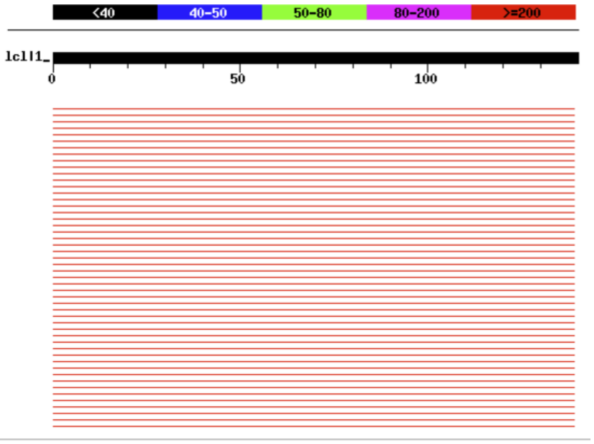
Tail Assembly Chaperone (Rank 3)

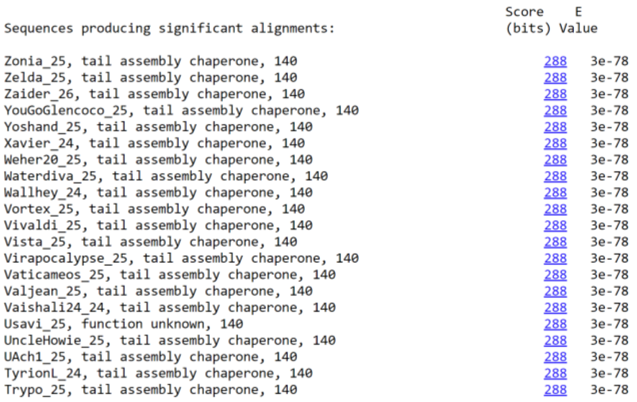


| Function Name | Pham | Subcluster | Count | Frequency (%) |
|---|---|---|---|---|
| tail assembly chaperone | 140939 | B1 | 86 | 100 |

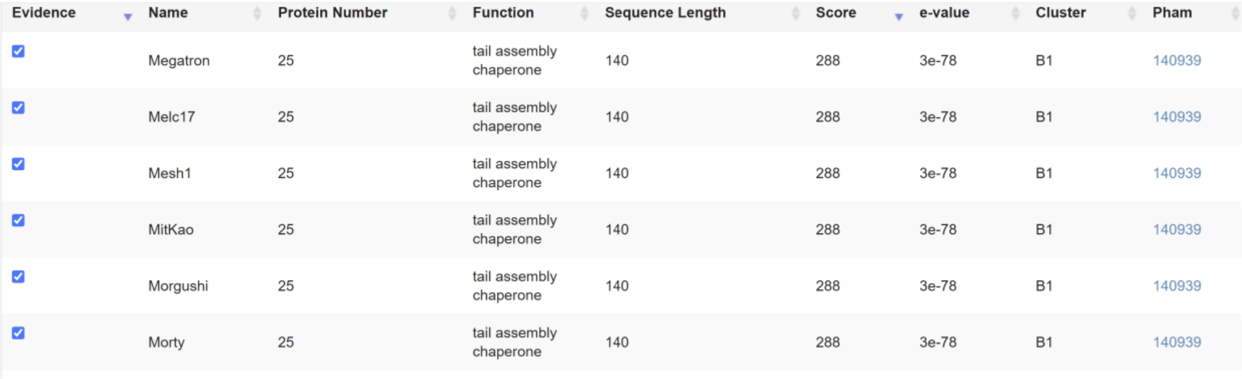
# 6. Supporting Information for Function (SIF)

## a. SIF: PhageDb BLAST





```
                                                                  Score    E
Sequences producing significant alignments:                      (bits) Value

Zonia_25, tail assembly chaperone, 140                               288   3e-78
Zelda_25, tail assembly chaperone, 140                               288   3e-78
Zaider_26, tail assembly chaperone, 140                              288   3e-78
YouGoGlencoco_25, tail assembly chaperone, 140                       288   3e-78
Yoshand_25, tail assembly chaperone, 140                             288   3e-78
Xavier_24, tail assembly chaperone, 140                              288   3e-78
Weher20_25, tail assembly chaperone, 140                             288   3e-78
Waterdiva_25, tail assembly chaperone, 140                           288   3e-78
Wallhey_24, tail assembly chaperone, 140                             288   3e-78
Vortex_25, tail assembly chaperone, 140                              288   3e-78
Vivaldi_25, tail assembly chaperone, 140                             288   3e-78
Vista_25, tail assembly chaperone, 140                               288   3e-78
Virapocalypse_25, tail assembly chaperone, 140                       288   3e-78
Vaticameos_25, tail assembly chaperone, 140                          288   3e-78
Valjean_25, tail assembly chaperone, 140                             288   3e-78
Vaishali24_24, tail assembly chaperone, 140                          288   3e-78
Usavi_25, function unknown, 140                                      288   3e-78
UncleHowie_25, tail assembly chaperone, 140                          288   3e-78
UAch1_25, tail assembly chaperone, 140                               288   3e-78
TyrionL_24, tail assembly chaperone, 140                             288   3e-78
Trypo_25, tail assembly chaperone, 140                               288   3e-78
```



| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | Megatron | 25 | tail assembly chaperone | 140 | 288 | 3e-78 | B1 | 140939 |
| ☑ | Melc17 | 25 | tail assembly chaperone | 140 | 288 | 3e-78 | B1 | 140939 |
| ☑ | Mesh1 | 25 | tail assembly chaperone | 140 | 288 | 3e-78 | B1 | 140939 |
| ☑ | MitKao | 25 | tail assembly chaperone | 140 | 288 | 3e-78 | B1 | 140939 |
| ☑ | Morgushi | 25 | tail assembly chaperone | 140 | 288 | 3e-78 | B1 | 140939 |
| ☑ | Morty | 25 | tail assembly chaperone | 140 | 288 | 3e-78 | B1 | 140939 |

## b. SIF: NCBI BLAST

| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | NP_943803 | No | 2023-01-08 | tail assembly chaperone | tail assembly chaperone [Mycobacterium phage PG1] >ref\|YP_008052102.1\| tail assembly chaperone [Mycobacterium phage Newman] >ref\|YP_009014286.1\| tail assembly chaperone [Mycobacterium phage Oline] >ref\|YP_009016814.1\| tail assembly chaperone [Mycobacterium phage Vista] >ref\|YP_009018338.1\| tail assembly chaperone [Mycobacterium phage JacAttac] >ref\|YP_009043299.1\| tail assembly chaperone [Mycobacterium phage Manad] >ref\|YP_009100834.1\| tail assembly chaperone [Mycobacterium phage Soto] >ref\|YP_009168205.1\| tail assembly chaperone [Mycobacterium phage UncleHowie] >ref\|YP_009187535.1\| tail assembly chaperone [Mycobacterium phage Swish] >ref\|YP_009189263.1\| tail assembly chaperone [Mycobacterium phage ShiVal] | 100 | 100 | 100 | 140 | 1 | 140 | 1 | 140 | 0 | 4.38876e-96 |
| ☑ | YP_009005672 | No | 2023-01-08 | tail assembly chaperone | tail assembly chaperone [Mycobacterium phage Suffolk] >ref\|YP_009191019.1\| tail assembly chaperone [Mycobacterium phage Colbert] >gb\|AEJ95211.1\| tail assembly chaperone [Mycobacterium phage KLucky39] >gb\|AGC33779.1\| tail assembly chaperone [Mycobacterium phage Serpentine] >gb\|AGC33884.1\| tail assembly chaperone [Mycobacterium phage Piglet] >gb\|AGC33985.1\| tail assembly chaperone [Mycobacterium phage Nacho] >gb\|AGC34089.1\| tail assembly chaperone [Mycobacterium phage Gyarad] >gb\|AGC34188.1\| hypothetical protein Alex_0025 [Mycobacterium phage Alex] >gb\|AHY84294.1\| tail assembly chaperone [Mycobacterium phage KingVeVeVe] >gb\|ALH46129.1\| tail assembly chaperone [Mycobacterium phage Squid] | 99.2857 | 100 | 100 | 140 | 1 | 140 | 1 | 140 | 0 | 1.05608e-95 |
| ☑ | QBI99564 | No | 2021-12-01 | tail assembly chaperone | tail assembly chaperone [Mycobacterium phage Robyn] | 99.2857 | 99.2857 | 100 | 139 | 1 | 140 | 1 | 140 | 0 | 1.73063e-95 |
| ☑ | UVK61520 | | | | tail assembly chaperone [Mycobacterium phage Soile] >gb\|WNM65864.1\| hypothetical protein SEA_DELRIVS__24 [Mycobacterium phage DelRivs] | 99.2857 | 99.2857 | 100 | 139 | 1 | 140 | 1 | 140 | 0 | 1.97427e-95 |
| ☑ | AVJ49981 | No | 2021-12-01 | tail assembly chaperone | tail assembly chaperone [Mycobacterium phage LeeLot] >gb\|AZS10200.1\| tail assembly chaperone | 99.2857 | 100 | 100 | 140 | 1 | 140 | 1 | 140 | 0 | 2.10866e-95 |

## c. SIF: HHPred

N/A

### d. SIF: Synteny-Phamerator (three genomes)

Aligns with Buckeye, Waterdiva, Usavi, Megatron, MitKao, etc.

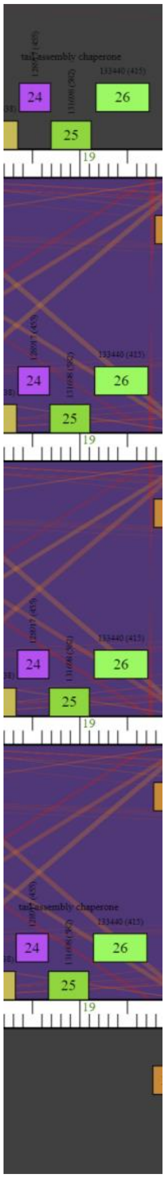


*Figure 1. (Top to bottom) Buckeye, Eugenia, Usavi, and Waterdiva.*

7. Any other important information.

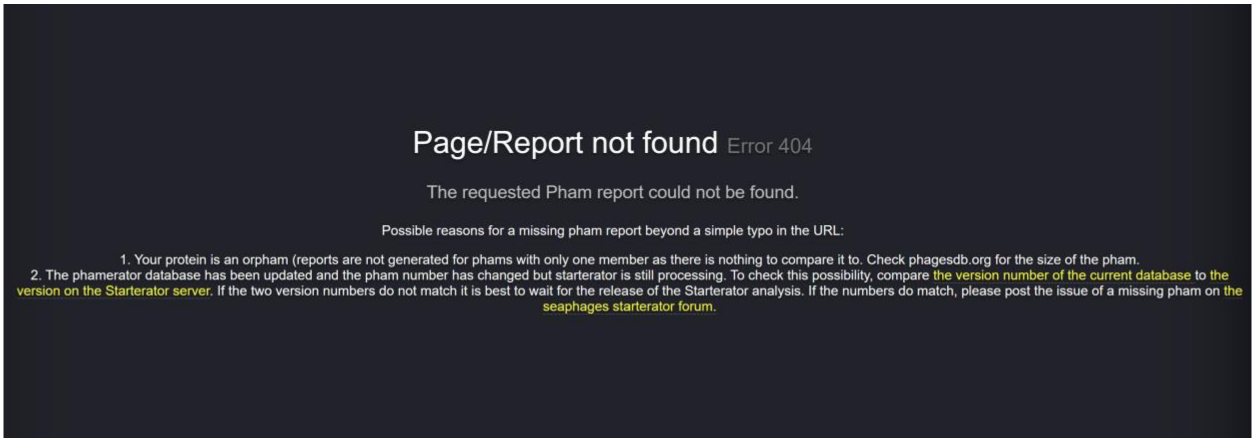
## CURATOR NAME: ISABELLA LIMA

## GENE NAME: EUGENIA_DRAFT_26

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

## 1. Starterator



Page/Report not found Error 404

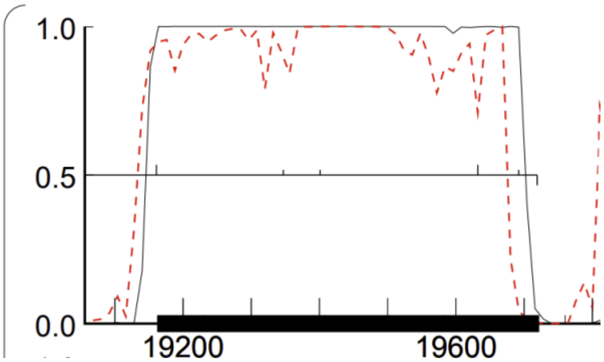
The requested Pham report could not be found.

Possible reasons for a missing pham report beyond a simple typo in the URL:

1. Your protein is an orpham (reports are not generated for phams with only one member as there is nothing to compare it to. Check phagesdb.org for the size of the pham.
2. The phamerator database has been updated and the pham number has changed but starterator is still processing. To check this possibility, compare the version number of the current database to the version on the Starterator server. If the two version numbers do not match it is best to wait for the release of the Starterator analysis. If the numbers do match, please post the issue of a missing pham on the seaphages starterator forum.

## 2. GeneMark coding potential

GeneMark: 19162



## 3. Glimmer and GeneMark agreement

YES

Phage: Eugenia Cluster: B1

| Glimmer Start: | Glimmer Score: | GeneMark Start: | Pham |
|---|---|---|---|
| 19162 | 16.54 | 19162 | Starterator: 142200 |

PhagesDB: 142200

Glimmer: 19162

GeneMark: 19162

## 4. Longest open reading frame (ORF) without excessive gap

YES

| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 18991 | 19722 | 732 | -121 | 9 | 1.207 | -6.455 | TRUE | GTG |  | ☐ |
| Forward | 19162 | 19722 | 561 | 50 | 11 | 2.904 | -3.030 |  | ATG | Select | ☑ |

Start: 19162

Length: 561

Gap: 50

Z-score: 2.904

Final Score: -3.030

## 5. Function. Please follow this [Official SEAPHAGE Function List]

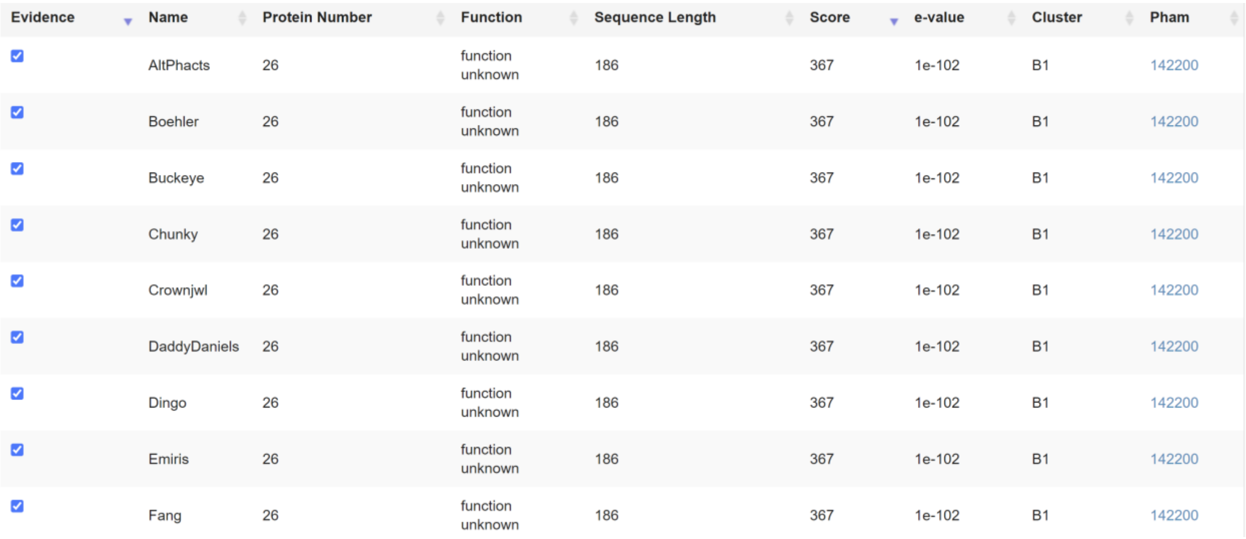
## If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

# 6. Supporting Information for Function (SIF)

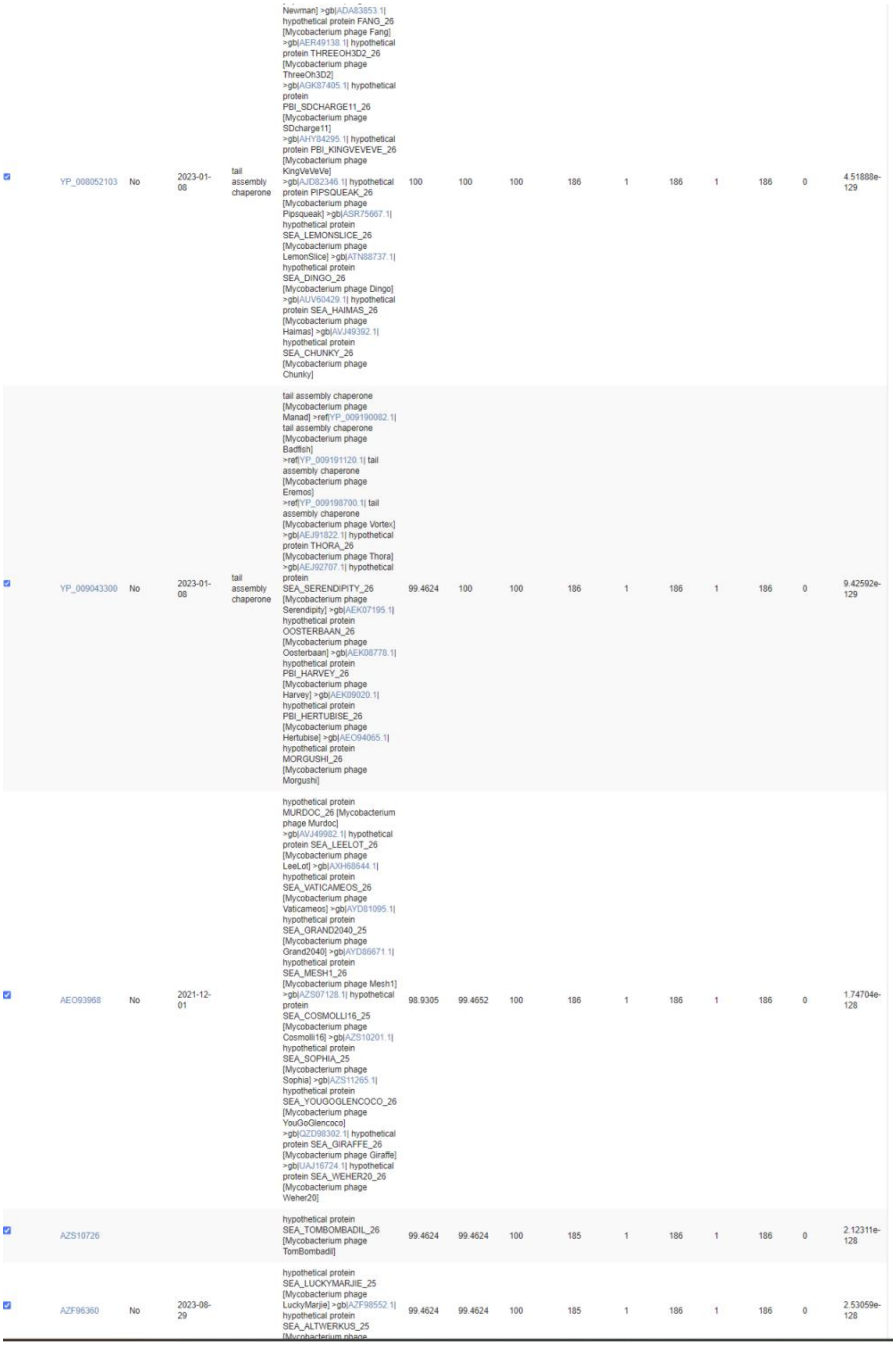
## a. SIF: PhageDb BLAST

```
                                                               Score    E
Sequences producing significant alignments:                    (bits) Value

ThreeOh3D2_26, function unknown, 186                              367   e-102
SDcharge11_26, function unknown, 186                              367   e-102
Roscoe_27, function unknown, 186                                  367   e-102
Placalicious_25, function unknown, 186                            367   e-102
Pipsqueak_26, function unknown, 186                               367   e-102
Newman_26, function unknown, 186                                  367   e-102
LemonSlice_26, function unknown, 186                              367   e-102
KingVeVeVe_26, function unknown, 186                              367   e-102
Hamish_26, function unknown, 186                                  367   e-102
Haimas_26, function unknown, 186                                  367   e-102
FugateOSU_25, function unknown, 186                               367   e-102
Freya_25, function unknown, 186                                   367   e-102
Fang_26, function unknown, 186                                    367   e-102
Eugenia_Draft_26, function unknown, 186                           367   e-102
Emiris_26, function unknown, 186                                  367   e-102
Dingo_26, function unknown, 186                                   367   e-102
DaddyDaniels_26, function unknown, 186                            367   e-102
Crownjwl_26, function unknown, 186                                367   e-102
Chunky_26, function unknown, 186                                  367   e-102
Buckeye_26, function unknown, 186                                 367   e-102
Boehler_26, function unknown, 186                                 367   e-102
AltPhacts 26, function unknown, 186                               367   e-102
```



| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | AltPhacts | 26 | function unknown | 186 | 367 | 1e-102 | B1 | 142200 |
| ☑ | Boehler | 26 | function unknown | 186 | 367 | 1e-102 | B1 | 142200 |
| ☑ | Buckeye | 26 | function unknown | 186 | 367 | 1e-102 | B1 | 142200 |
| ☑ | Chunky | 26 | function unknown | 186 | 367 | 1e-102 | B1 | 142200 |
| ☑ | Crownjwl | 26 | function unknown | 186 | 367 | 1e-102 | B1 | 142200 |
| ☑ | DaddyDaniels | 26 | function unknown | 186 | 367 | 1e-102 | B1 | 142200 |
| ☑ | Dingo | 26 | function unknown | 186 | 367 | 1e-102 | B1 | 142200 |
| ☑ | Emiris | 26 | function unknown | 186 | 367 | 1e-102 | B1 | 142200 |
| ☑ | Fang | 26 | function unknown | 186 | 367 | 1e-102 | B1 | 142200 |

## b. SIF: NCBI BLAST



| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | YP_008052103 | No | 2023-01-08 | tail assembly chaperone | Newman] >gb\|ADA83853.1\| hypothetical protein FANG_26 [Mycobacterium phage Fang] >gb\|AER49138.1\| hypothetical protein THREEOH3D2_26 [Mycobacterium phage ThreeOh3D2] >gb\|AGK87405.1\| hypothetical protein PBI_SDCHARGE11_26 [Mycobacterium phage SDcharge11] >gb\|AHY84295.1\| hypothetical protein PBI_KINGVEVEVE_26 [Mycobacterium phage KingVeVeVe] >gb\|AJD82346.1\| hypothetical protein PIPSQUEAK_26 [Mycobacterium phage Pipsqueak] >gb\|ASR75667.1\| hypothetical protein SEA_LEMONSLICE_26 [Mycobacterium phage LemonSlice] >gb\|ATN88737.1\| hypothetical protein SEA_DINGO_26 [Mycobacterium phage Dingo] >gb\|AUV60429.1\| hypothetical protein SEA_HAIMAS_26 [Mycobacterium phage Haimas] >gb\|AVJ49392.1\| hypothetical protein SEA_CHUNKY_26 [Mycobacterium phage Chunky] | 100 | 100 | 100 | 186 | 1 | 186 | 1 | 186 | 0 | 4.51888e-129 |
| ☑ | YP_009043300 | No | 2023-01-08 | tail assembly chaperone | tail assembly chaperone [Mycobacterium phage Manad] >ref\|YP_009190082.1\| tail assembly chaperone [Mycobacterium phage Badfish] >ref\|YP_009191120.1\| tail assembly chaperone [Mycobacterium phage Eremos] >ref\|YP_009198700.1\| tail assembly chaperone [Mycobacterium phage Vortex] >gb\|AEJ91822.1\| hypothetical protein THORA_26 [Mycobacterium phage Thora] >gb\|AEJ92707.1\| hypothetical protein SEA_SERENDIPITY_26 [Mycobacterium phage Serendipity] >gb\|AEK07195.1\| hypothetical protein OOSTERBAAN_26 [Mycobacterium phage Oosterbaan] >gb\|AEK08778.1\| hypothetical protein PBI_HARVEY_26 [Mycobacterium phage Harvey] >gb\|AEK09020.1\| hypothetical protein PBI_HERTUBISE_26 [Mycobacterium phage Hertubise] >gb\|AEO94065.1\| hypothetical protein MORGUSHI_26 [Mycobacterium phage Morgushi] | 99.4624 | 100 | 100 | 186 | 1 | 186 | 1 | 186 | 0 | 9.42592e-129 |
| ☑ | AEO93968 | No | 2021-12-01 | | hypothetical protein MURDOC_26 [Mycobacterium phage Murdoc] >gb\|AVJ49982.1\| hypothetical protein SEA_LEELOT_26 [Mycobacterium phage LeeLot] >gb\|AXH68644.1\| hypothetical protein SEA_VATICAMEOS_26 [Mycobacterium phage Vaticameos] >gb\|AYD81095.1\| hypothetical protein SEA_GRAND2040_25 [Mycobacterium phage Grand2040] >gb\|AYD86671.1\| hypothetical protein SEA_MESH1_26 [Mycobacterium phage Mesh1] >gb\|AZS07128.1\| hypothetical protein SEA_COSMOLLI16_25 [Mycobacterium phage Cosmolli16] >gb\|AZS10201.1\| hypothetical protein SEA_SOPHIA_25 [Mycobacterium phage Sophia] >gb\|AZS11265.1\| hypothetical protein SEA_YOUGOGLENCOCO_26 [Mycobacterium phage YouGoGlencoco] >gb\|QZD98302.1\| hypothetical protein SEA_GIRAFFE_26 [Mycobacterium phage Giraffe] >gb\|UAJ16724.1\| hypothetical protein SEA_WEHER20_26 [Mycobacterium phage Weher20] | 98.9305 | 99.4652 | 100 | 186 | 1 | 186 | 1 | 186 | 0 | 1.74704e-128 |
| ☑ | AZS10726 | | | | hypothetical protein SEA_TOMBOMBADIL_26 [Mycobacterium phage TomBombadil] | 99.4624 | 99.4624 | 100 | 185 | 1 | 186 | 1 | 186 | 0 | 2.12311e-128 |
| ☑ | AZF96360 | No | 2023-08-29 | | hypothetical protein SEA_LUCKYMARJIE_25 [Mycobacterium phage LuckyMarjie] >gb\|AZF98552.1\| hypothetical protein SEA_ALTWERKUS_25 [Mycobacterium phage | 99.4624 | 99.4624 | 100 | 185 | 1 | 186 | 1 | 186 | 0 | 2.53059e-128 |

## c. SIF: HHPred

N/A

### d. SIF: Synteny-Phamerator (three genomes)

Aligns with Buckeye, Usavi, Waterdiva, AltPhacts, and Chunky.

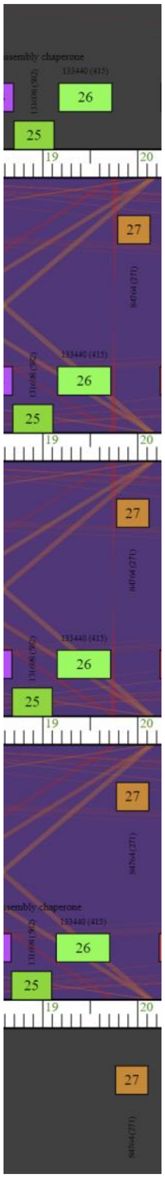


*Figure 1. (Top to bottom) Buckeye, Eugenia, Usavi, and Waterdiva.*

## 7. Any other important information.

0 TmHmms

## CURATOR NAME: ISABELLA LIMA

## GENE NAME: EUGENIA_DRAFT_27

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

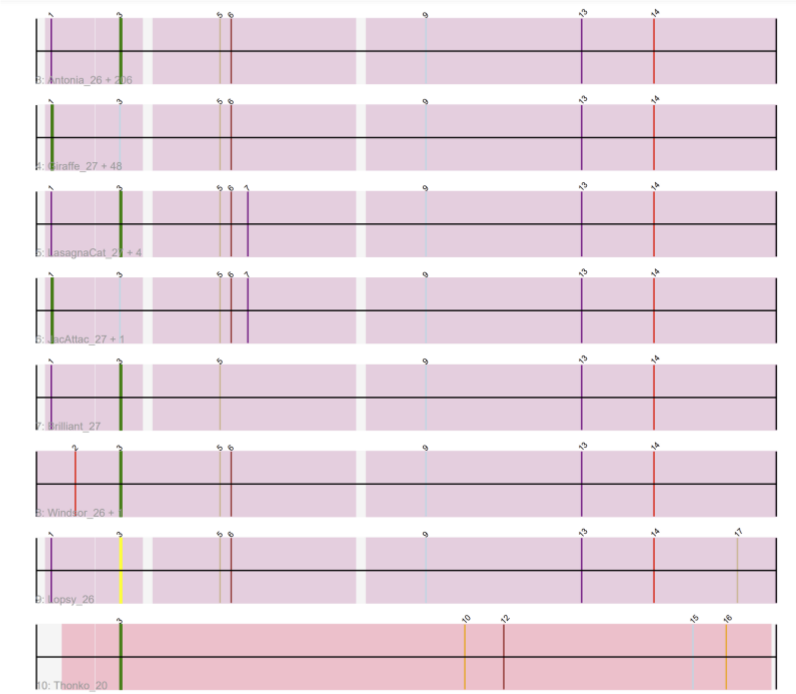
### 1. Starterator

Eugenia is in Track 3 with Start 3.

The start number called the most often in the published annotations is 3, it was called in 196 of the 250 non-draft genes in the pham.

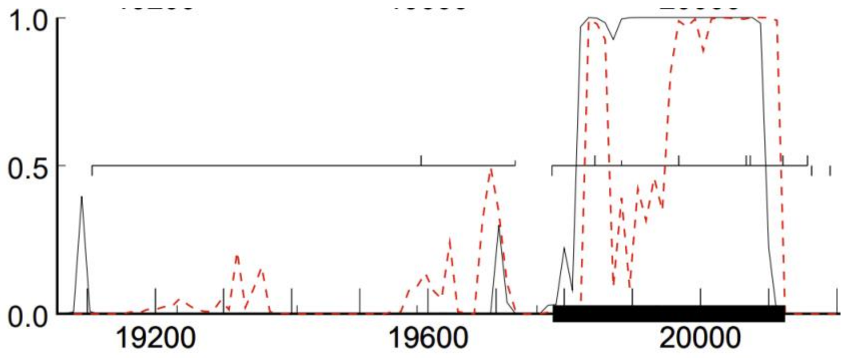
Start 3:

- Found in 268 of 274 ( 97.8% ) of genes in pham
- Manual Annotations of this start: 196 of 250
- Called 81.0% of time when present



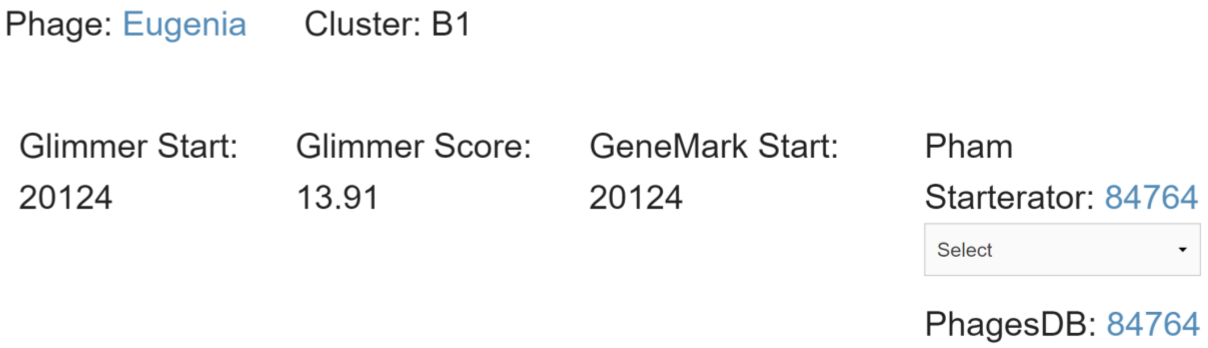
## 2. GeneMark coding potential

GeneMark: 20124



## 3. Glimmer and GeneMark agreement

YES



Phage: Eugenia Cluster: B1

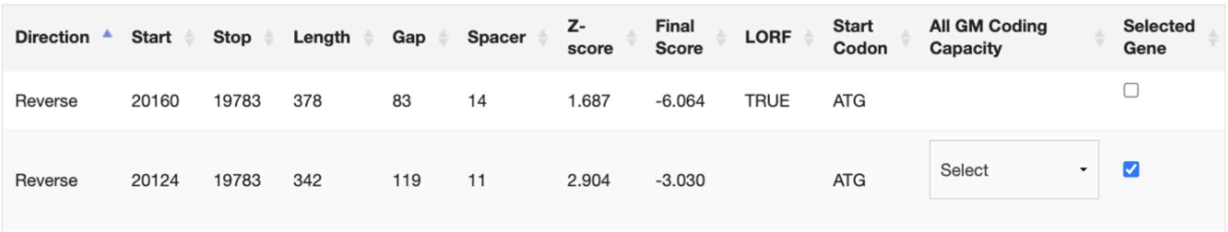
| Glimmer Start: | Glimmer Score: | GeneMark Start: | Pham |
|---|---|---|---|
| 20124 | 13.91 | 20124 | Starterator: 84764<br>Select<br>PhagesDB: 84764 |

Glimmer: 20124

GeneMark: 20124

## 4. Longest open reading frame (ORF) without excessive gap

No.



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 20160 | 19783 | 378 | 83 | 14 | 1.687 | -6.064 | TRUE | ATG |  | ☐ |
| Reverse | 20124 | 19783 | 342 | 119 | 11 | 2.904 | -3.030 |  | ATG | Select | ☑ |

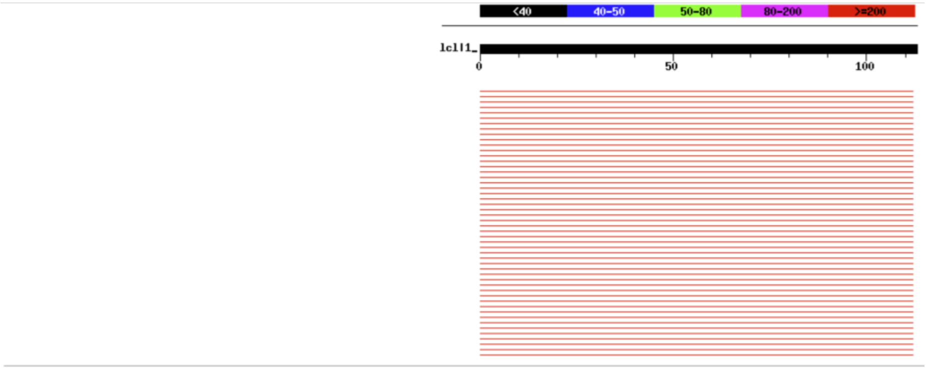
## 5. Function. Please follow this [Official SEAPHAGE Function List]

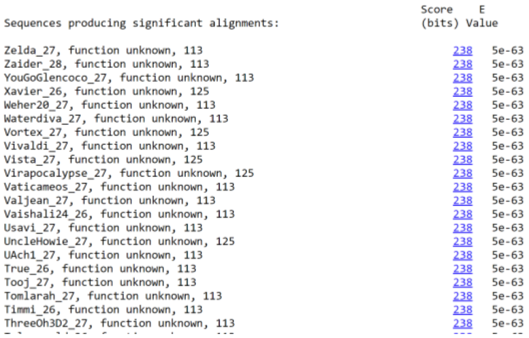
## If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

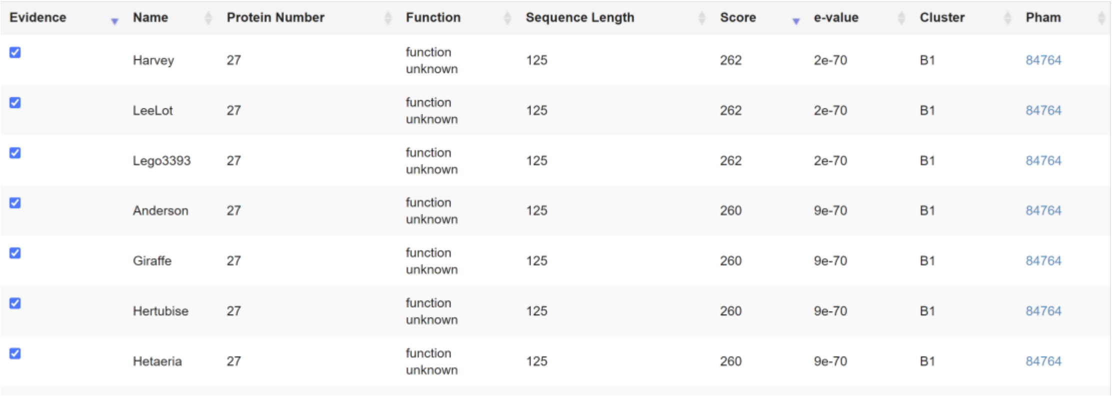
## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST



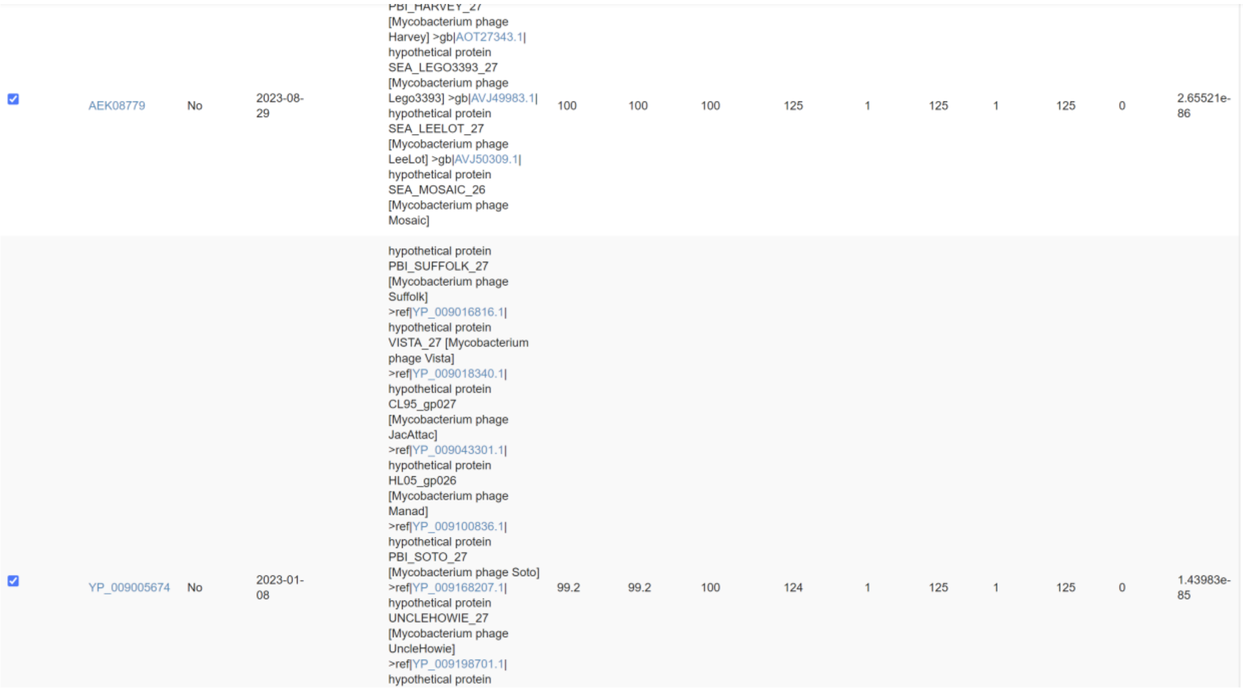


```
                                                            Score    E
Sequences producing significant alignments:                 (bits) Value

Zelda_27, function unknown, 113                               238    5e-63
Zaider_28, function unknown, 113                              238    5e-63
YouGoGlencoco_27, function unknown, 113                       238    5e-63
Xavier_26, function unknown, 125                              238    5e-63
Weher20_27, function unknown, 113                             238    5e-63
Waterdiva_27, function unknown, 113                           238    5e-63
Vortex_27, function unknown, 125                              238    5e-63
Vivaldi_27, function unknown, 113                             238    5e-63
Vista_27, function unknown, 125                               238    5e-63
Virapocalypse_27, function unknown, 125                       238    5e-63
Vaticameos_27, function unknown, 113                          238    5e-63
Valjean_27, function unknown, 113                             238    5e-63
Vaishali24_26, function unknown, 113                          238    5e-63
Usavi_27, function unknown, 113                               238    5e-63
UncleHowie_27, function unknown, 125                          238    5e-63
UAch1_27, function unknown, 113                               238    5e-63
True_26, function unknown, 113                                238    5e-63
Tooj_27, function unknown, 113                                238    5e-63
Tomlarah_27, function unknown, 113                            238    5e-63
Timmi_26, function unknown, 113                               238    5e-63
ThreeOh3D2_27, function unknown, 113                          238    5e-63
```



| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | Harvey | 27 | function unknown | 125 | 262 | 2e-70 | B1 | 84764 |
| ☑ | LeeLot | 27 | function unknown | 125 | 262 | 2e-70 | B1 | 84764 |
| ☑ | Lego3393 | 27 | function unknown | 125 | 262 | 2e-70 | B1 | 84764 |
| ☑ | Anderson | 27 | function unknown | 125 | 260 | 9e-70 | B1 | 84764 |
| ☑ | Giraffe | 27 | function unknown | 125 | 260 | 9e-70 | B1 | 84764 |
| ☑ | Hertubise | 27 | function unknown | 125 | 260 | 9e-70 | B1 | 84764 |
| ☑ | Hetaeria | 27 | function unknown | 125 | 260 | 9e-70 | B1 | 84764 |

## b. SIF: NCBI BLAST



| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | AEK08779 | No | 2023-08-29 | PBI_HARVEY_27 [Mycobacterium phage Harvey] >gb\|AOT27343.1\| hypothetical protein SEA_LEGO3393_27 [Mycobacterium phage Lego3393] >gb\|AVJ49983.1\| hypothetical protein SEA_LEELOT_27 [Mycobacterium phage LeeLot] >gb\|AVJ50309.1\| hypothetical protein SEA_MOSAIC_26 [Mycobacterium phage Mosaic] | 100 | 100 | 100 | 125 | 1 | 125 | 1 | 125 | 0 | 2.65521e-86 |
| ☑ | YP_009005674 | No | 2023-01-08 | hypothetical protein PBI_SUFFOLK_27 [Mycobacterium phage Suffolk] >ref\|YP_009016816.1\| hypothetical protein VISTA_27 [Mycobacterium phage Vista] >ref\|YP_009018340.1\| hypothetical protein CL95_gp027 [Mycobacterium phage JacAttac] >ref\|YP_009043301.1\| hypothetical protein HL05_gp026 [Mycobacterium phage Manad] >ref\|YP_009100836.1\| hypothetical protein PBI_SOTO_27 [Mycobacterium phage Soto] >ref\|YP_009168207.1\| hypothetical protein UNCLEHOWIE_27 [Mycobacterium phage UncleHowie] >ref\|YP_009198701.1\| hypothetical protein | 99.2 | 99.2 | 100 | 124 | 1 | 125 | 1 | 125 | 0 | 1.43983e-85 |

## c. SIF: HHPred

N/A

## d. SIF: Synteny-Phamerator (three genomes)

Aligns with Buckeye, Waterdiva, Usavi, Harvey, and LeeLot

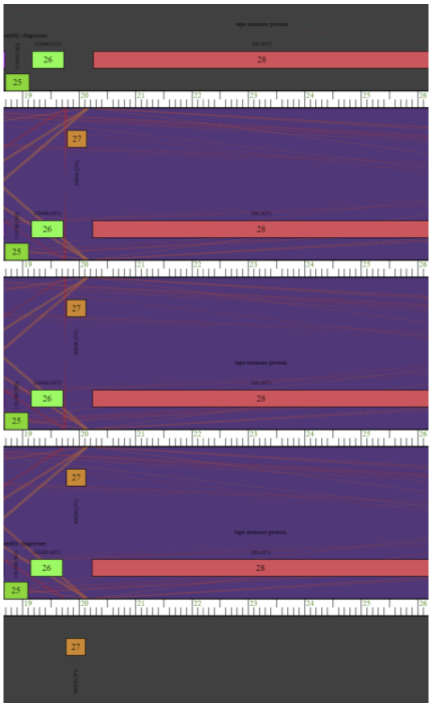


*Figure 1. (Top to bottom) Buckeye, Eugenia, Usavi, and Waterdiva.*

## 7. Any other important information.

Glimmer and GeneMark start site doesn't agree with the longest ORF with shortest gap's start site.

## CURATOR NAME: ISABELLA LIMA

## GENE NAME: EUGENIA_DRAFT_28

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

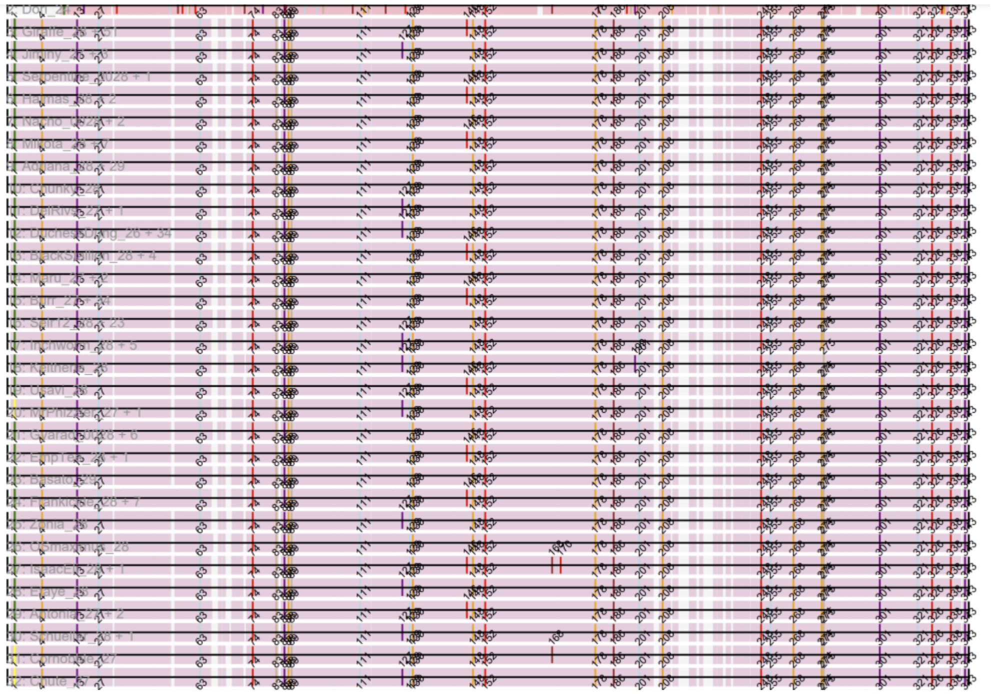
### 1. Starterator

Eugenia is in Track 15 with Start site 1.

The start number called the most often in the published annotations is 1, it was called in 295 of the 376 non-draft genes in the pham.

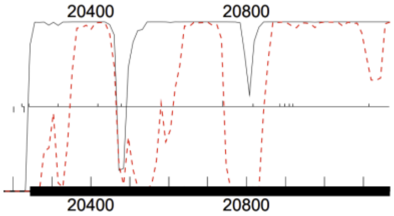
Start 1:

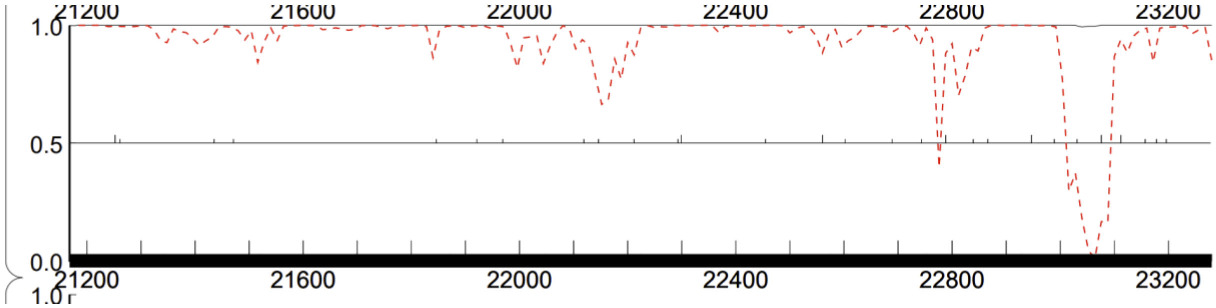
- Found in 339 of 422 ( 80.3% ) of genes in pham
- Manual Annotations of this start: 295 of 376
- Called 99.7% of time when present

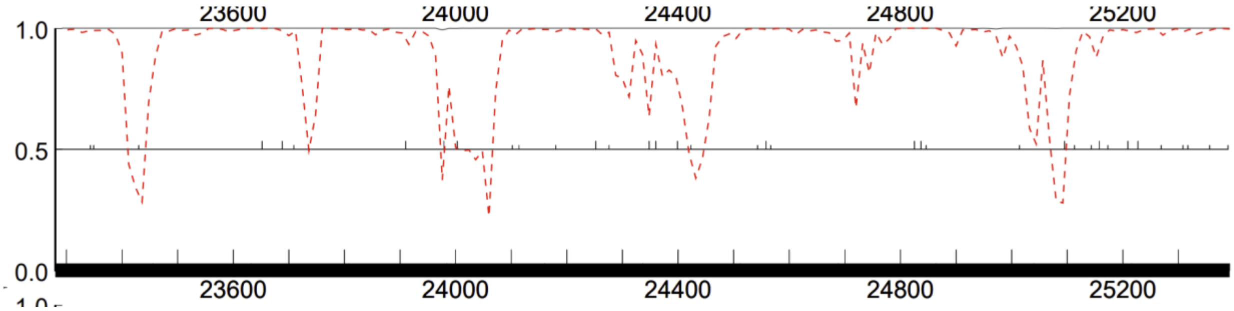


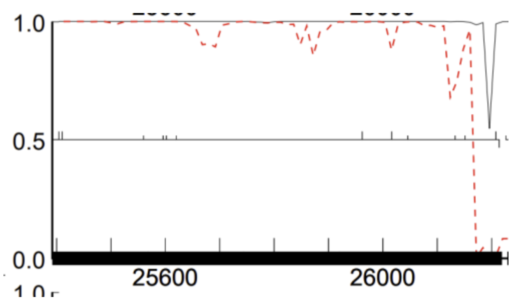
## 2. GeneMark coding potential

GeneMark: 20244



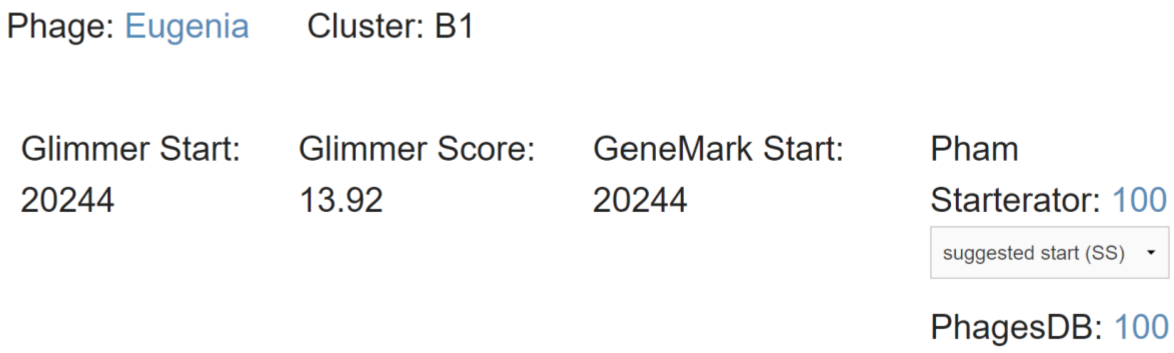






## 3. Glimmer and GeneMark agreement

YES



Phage: Eugenia Cluster: B1

Glimmer Start: 20244

Glimmer Score: 13.92

GeneMark Start: 20244

Pham Starterator: 100

PhagesDB: 100

Glimmer: 20244

GeneMark: 20244

## 4. Longest open reading frame (ORF) without excessive gap

YES

| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 20244 | 26219 | 5976 | 83 | 14 | 0.942 | -7.561 | TRUE | GTG | Select | ☑ |
| Forward | 20319 | 26219 | 5901 | 158 | 10 | 0.777 | -7.239 |  | GTG |  | ☐ |

Start: 20244

Length: 5976

Gap: 83

Z-score: 0.942

Final score: -7.561

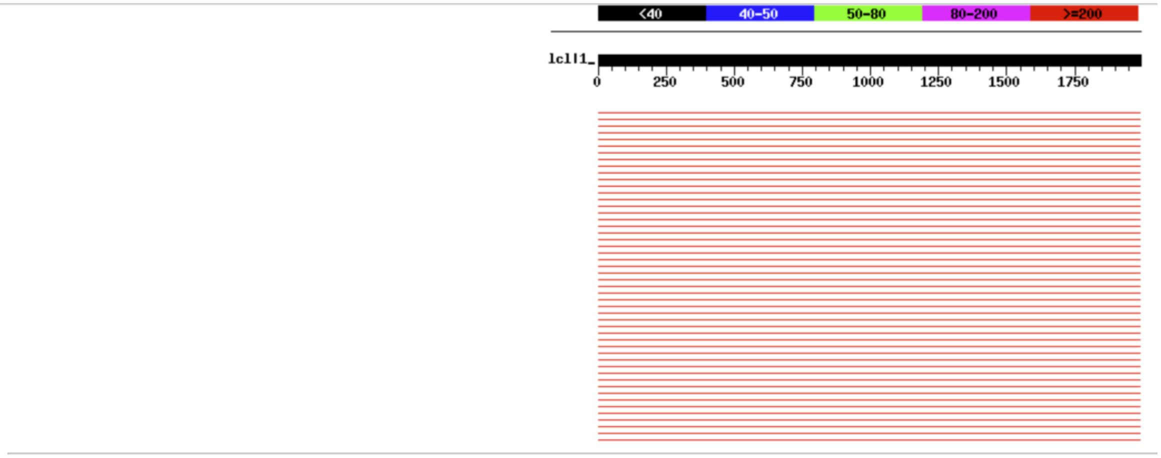
## 5. Function. Please follow this Official SEAPHAGE Function List

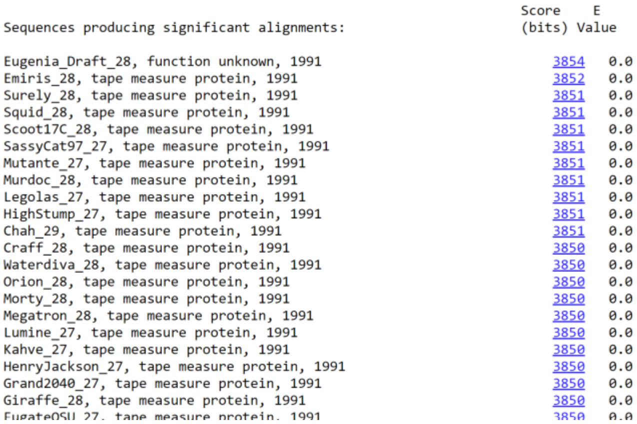
**If no functional prediction is present, write “Hypothetical protein”.**

Tape Measure Protein (Rank 1)

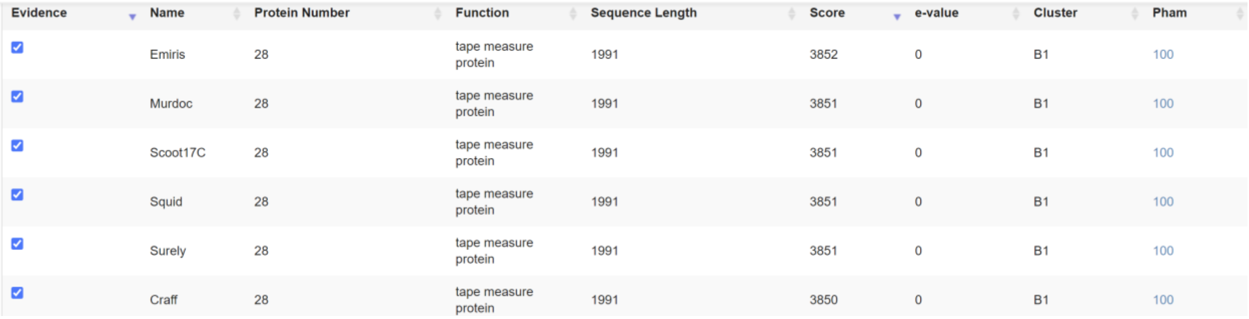
## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST



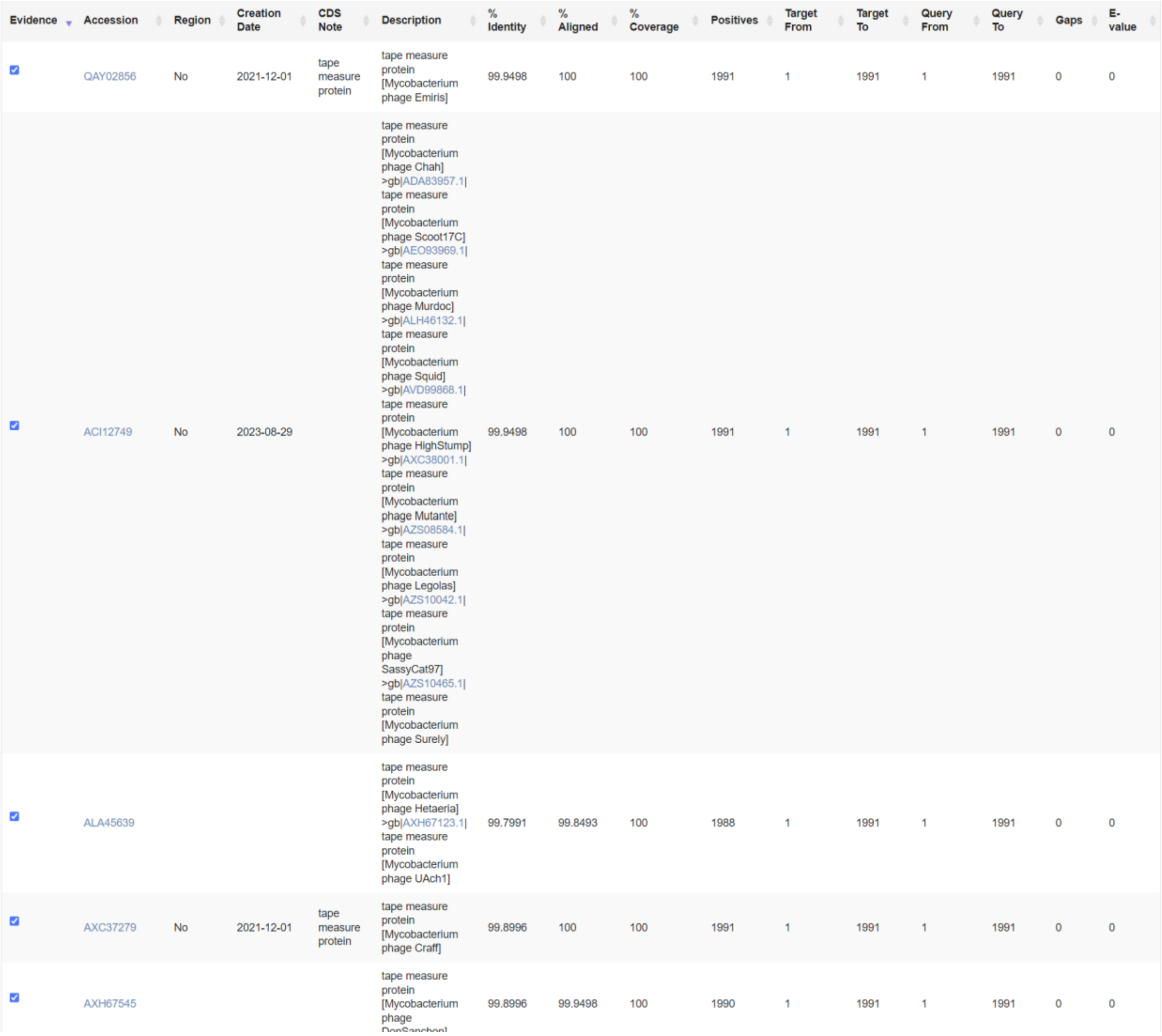


```
                                                                 Score    E
Sequences producing significant alignments:                      (bits) Value

Eugenia_Draft_28, function unknown, 1991                             3854   0.0
Emiris_28, tape measure protein, 1991                                3852   0.0
Surely_28, tape measure protein, 1991                                3851   0.0
Squid_28, tape measure protein, 1991                                 3851   0.0
Scoot17C_28, tape measure protein, 1991                              3851   0.0
SassyCat97_27, tape measure protein, 1991                            3851   0.0
Mutante_27, tape measure protein, 1991                               3851   0.0
Murdoc_28, tape measure protein, 1991                                3851   0.0
Legolas_27, tape measure protein, 1991                               3851   0.0
HighStump_27, tape measure protein, 1991                             3851   0.0
Chah_29, tape measure protein, 1991                                  3851   0.0
Craff_28, tape measure protein, 1991                                 3850   0.0
Waterdiva_28, tape measure protein, 1991                             3850   0.0
Orion_28, tape measure protein, 1991                                 3850   0.0
Morty_28, tape measure protein, 1991                                 3850   0.0
Megatron_28, tape measure protein, 1991                              3850   0.0
Lumine_27, tape measure protein, 1991                                3850   0.0
Kahve_27, tape measure protein, 1991                                 3850   0.0
HenryJackson_27, tape measure protein, 1991                          3850   0.0
Grand2040_27, tape measure protein, 1991                             3850   0.0
Giraffe_28, tape measure protein, 1991                               3850   0.0
FugateOSU_27, tape measure protein, 1991                             3850   0.0
```



| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | Emiris | 28 | tape measure protein | 1991 | 3852 | 0 | B1 | 100 |
| ☑ | Murdoc | 28 | tape measure protein | 1991 | 3851 | 0 | B1 | 100 |
| ☑ | Scoot17C | 28 | tape measure protein | 1991 | 3851 | 0 | B1 | 100 |
| ☑ | Squid | 28 | tape measure protein | 1991 | 3851 | 0 | B1 | 100 |
| ☑ | Surely | 28 | tape measure protein | 1991 | 3851 | 0 | B1 | 100 |
| ☑ | Craff | 28 | tape measure protein | 1991 | 3850 | 0 | B1 | 100 |

## b. SIF: NCBI BLAST



| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | QAY02856 | No | 2021-12-01 | tape measure protein | tape measure protein [Mycobacterium phage Emiris] | 99.9498 | 100 | 100 | 1991 | 1 | 1991 | 1 | 1991 | 0 | 0 |
| ☑ | ACI12749 | No | 2023-08-29 |  | tape measure protein [Mycobacterium phage Chah] >gb\|ADA83957.1\| tape measure protein [Mycobacterium phage Scoot17C] >gb\|AEO93969.1\| tape measure protein [Mycobacterium phage Murdoc] >gb\|ALH46132.1\| tape measure protein [Mycobacterium phage Squid] >gb\|AVD99868.1\| tape measure protein [Mycobacterium phage HighStump] >gb\|AXC38001.1\| tape measure protein [Mycobacterium phage Mutante] >gb\|AZS08584.1\| tape measure protein [Mycobacterium phage Legolas] >gb\|AZS10042.1\| tape measure protein [Mycobacterium phage SassyCat97] >gb\|AZS10465.1\| tape measure protein [Mycobacterium phage Surely] | 99.9498 | 100 | 100 | 1991 | 1 | 1991 | 1 | 1991 | 0 | 0 |
| ☑ | ALA45639 |  |  |  | tape measure protein [Mycobacterium phage Hetaeria] >gb\|AXH67123.1\| tape measure protein [Mycobacterium phage UAch1] | 99.7991 | 99.8493 | 100 | 1988 | 1 | 1991 | 1 | 1991 | 0 | 0 |
| ☑ | AXC37279 | No | 2021-12-01 | tape measure protein | tape measure protein [Mycobacterium phage Craff] | 99.8996 | 100 | 100 | 1991 | 1 | 1991 | 1 | 1991 | 0 | 0 |
| ☑ | AXH67545 |  |  |  | tape measure protein [Mycobacterium phage | 99.8996 | 99.9498 | 100 | 1990 | 1 | 1991 | 1 | 1991 | 0 | 0 |

## c. SIF: HHPred

N/A

### d. SIF: Synteny-Phamerator (three genomes)

Aligns with Buckeye, Usavi, Waterdiva, Emiris, Murdoc, etc.

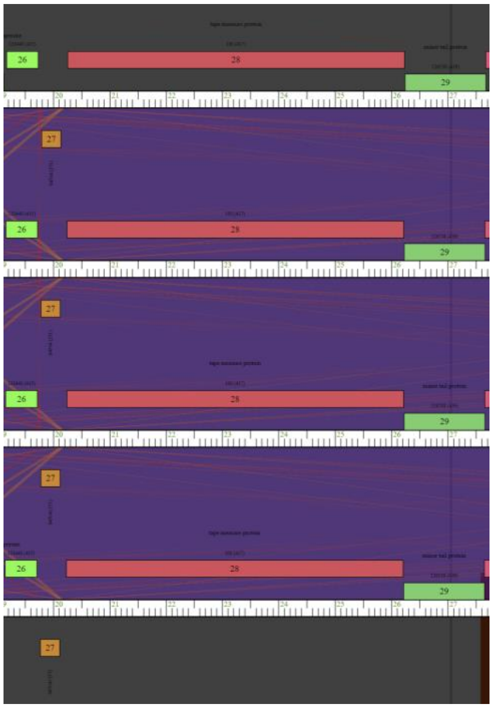


*Figure 1. (Top to bottom) Buckeye, Eugenia, Usavi, and Waterdiva.*

7. Any other important information.

## CURATOR NAME: ISABELLA LIMA

## GENE NAME: EUGENIA_DRAFT_29

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

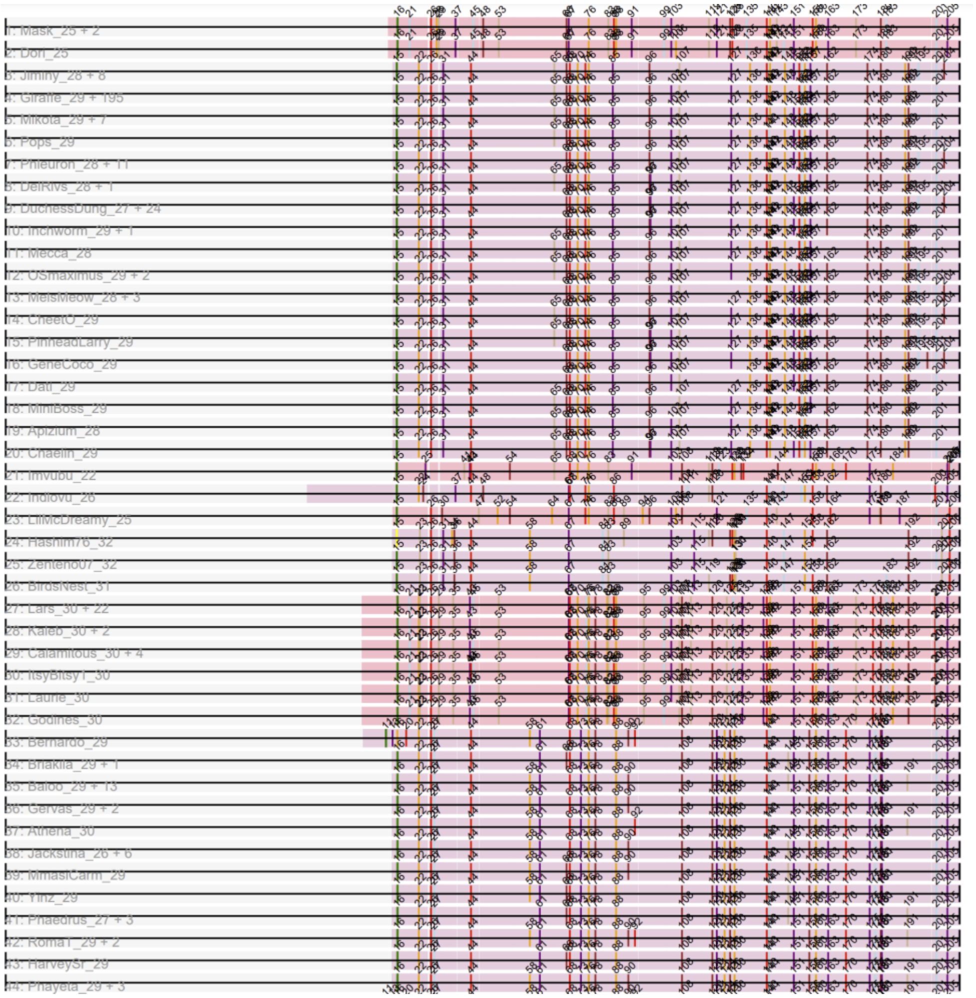
## 1. Starterator

Eugenia is in Track 4 with Start site.

The start number called the most often in the published annotations is 15, it was called in 285 of the 392 non-draft genes in the pham.

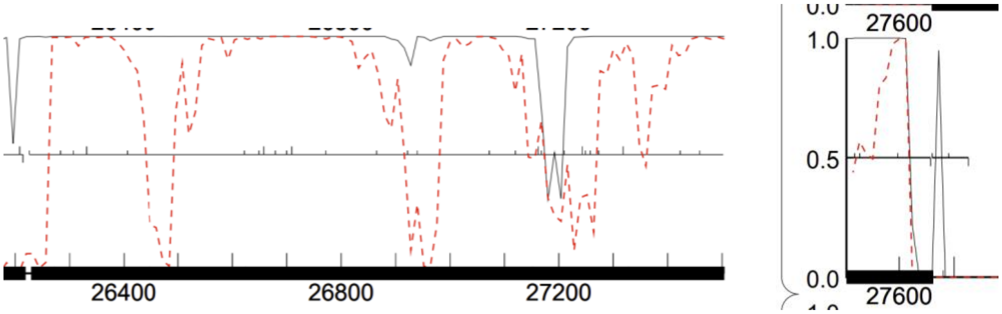
Start 15:

- Found in 326 of 444 ( 73.4% ) of genes in pham
- Manual Annotations of this start: 285 of 392
- Called 100.0% of time when present



# 2. GeneMark coding potential

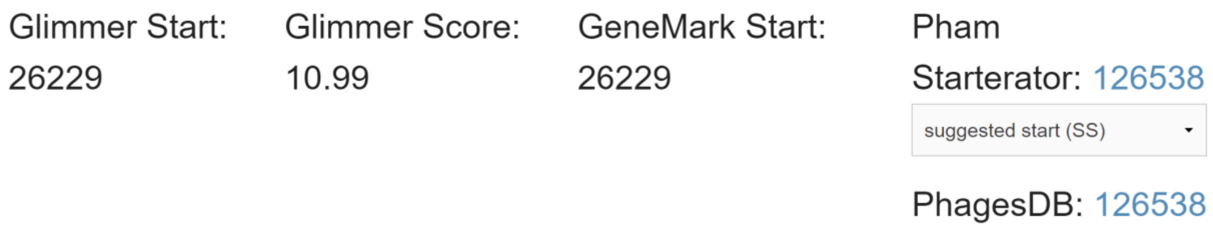
GeneMark: 26229



## 3. Glimmer and GeneMark agreement

YES


Phage: Eugenia Cluster: B1


Glimmer Start: 26229

Glimmer Score: 10.99

GeneMark Start: 26229

Pham Starterator: 126538

suggested start (SS)

PhagesDB: 126538

Glimmer: 26229

GeneMark: 26229

## 4. Longest open reading frame (ORF) without excessive gap

YES

| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 26229 | 27662 | 1434 | 9 | 10 | 2.482 | -3.814 | TRUE | GTG | Select | ☑ |
| Forward | 26286 | 27662 | 1377 | 66 | 8 | 2.138 | -5.032 | | GTG | | ☐ |

Start: 26229

Length: 1434

Gap: 9

Z-score: 2.482

Final Score: -3.814

## 5. Function. Please follow this [Official SEAPHAGE Function List]()

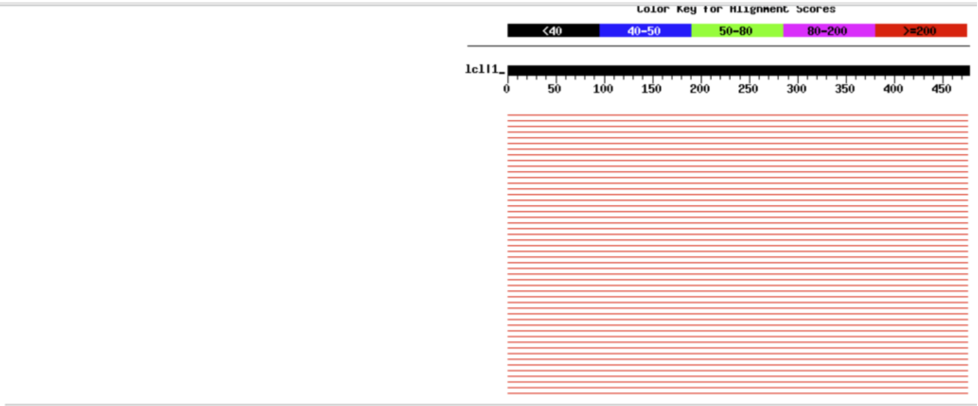
## If no functional prediction is present, write "Hypothetical protein".

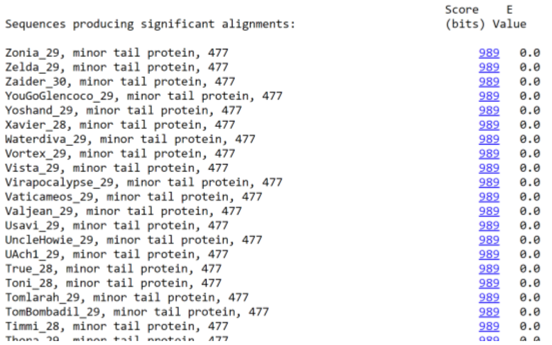
Minor tail protein (Rank 1)

| Function Name | Pham | Subcluster | Count | Frequency (%) |
|---|---|---|---|---|
| minor tail protein | 126538 | B1 | 93 | 100 |

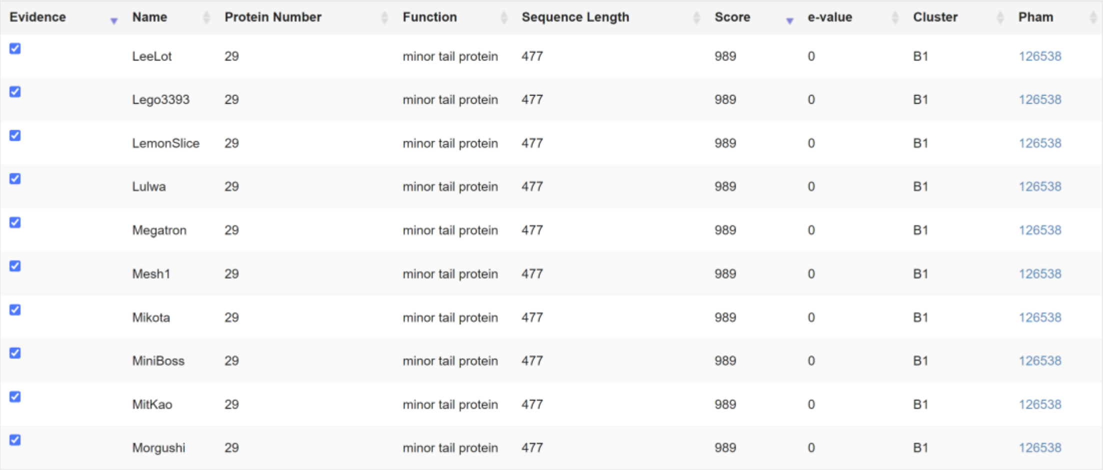
## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST



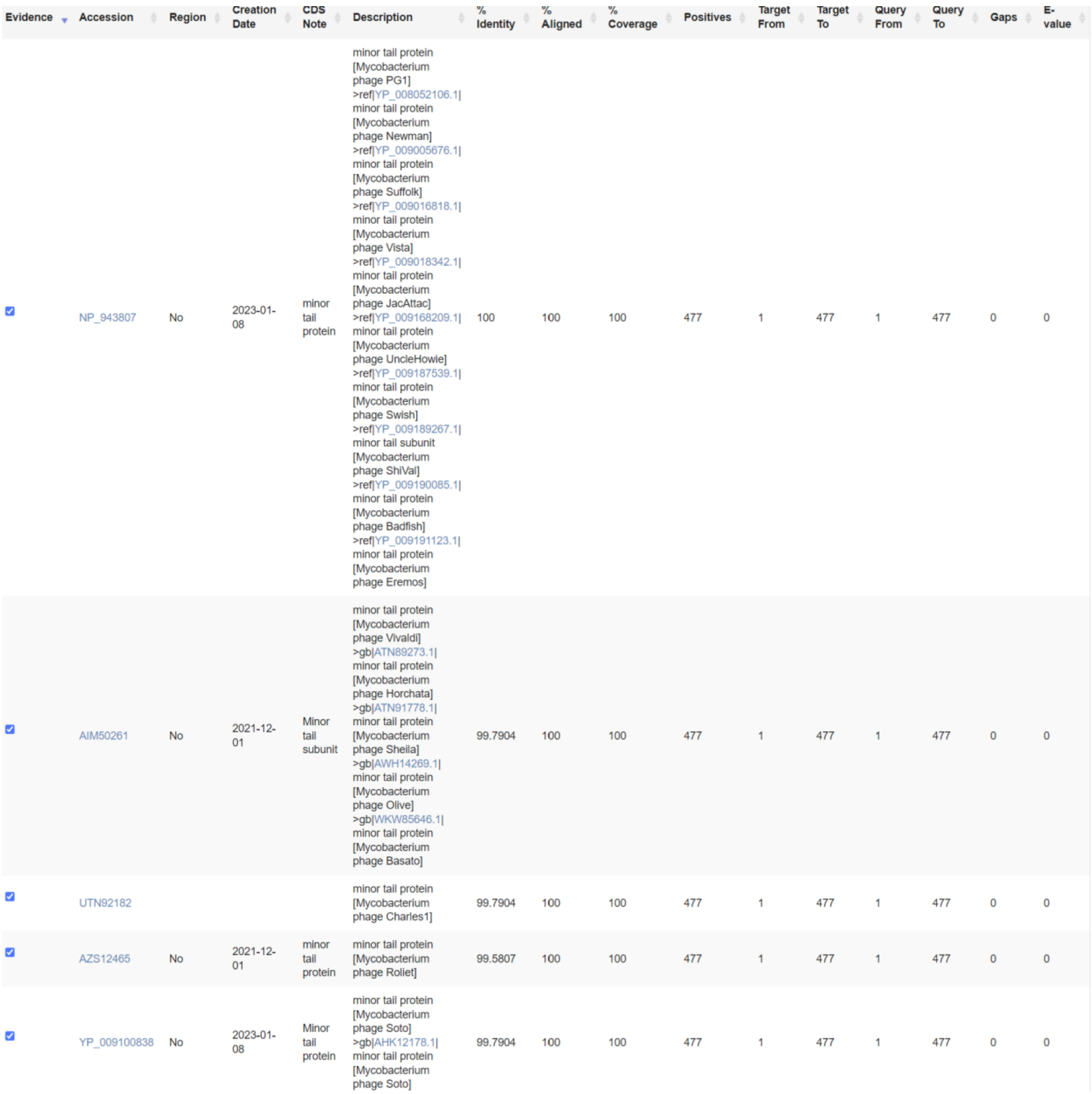


```
                                                                 Score    E
Sequences producing significant alignments:                      (bits) Value

Zonia_29, minor tail protein, 477                                   989   0.0
Zelda_29, minor tail protein, 477                                   989   0.0
Zaider_30, minor tail protein, 477                                  989   0.0
YouGoGlencoco_29, minor tail protein, 477                           989   0.0
Yoshand_29, minor tail protein, 477                                 989   0.0
Xavier_28, minor tail protein, 477                                  989   0.0
Waterdiva_29, minor tail protein, 477                               989   0.0
Vortex_29, minor tail protein, 477                                  989   0.0
Vista_29, minor tail protein, 477                                   989   0.0
Virapocalypse_29, minor tail protein, 477                           989   0.0
Vaticameos_29, minor tail protein, 477                              989   0.0
Valjean_29, minor tail protein, 477                                 989   0.0
Usavi_29, minor tail protein, 477                                   989   0.0
UncleHowie_29, minor tail protein, 477                              989   0.0
UAch1_29, minor tail protein, 477                                   989   0.0
True_28, minor tail protein, 477                                    989   0.0
Toni_28, minor tail protein, 477                                    989   0.0
Tomlarah_29, minor tail protein, 477                                989   0.0
TomBombadil_29, minor tail protein, 477                             989   0.0
Timmi_28, minor tail protein, 477                                   989   0.0
```



| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | LeeLot | 29 | minor tail protein | 477 | 989 | 0 | B1 | 126538 |
| ☑ | Lego3393 | 29 | minor tail protein | 477 | 989 | 0 | B1 | 126538 |
| ☑ | LemonSlice | 29 | minor tail protein | 477 | 989 | 0 | B1 | 126538 |
| ☑ | Lulwa | 29 | minor tail protein | 477 | 989 | 0 | B1 | 126538 |
| ☑ | Megatron | 29 | minor tail protein | 477 | 989 | 0 | B1 | 126538 |
| ☑ | Mesh1 | 29 | minor tail protein | 477 | 989 | 0 | B1 | 126538 |
| ☑ | Mikota | 29 | minor tail protein | 477 | 989 | 0 | B1 | 126538 |
| ☑ | MiniBoss | 29 | minor tail protein | 477 | 989 | 0 | B1 | 126538 |
| ☑ | MitKao | 29 | minor tail protein | 477 | 989 | 0 | B1 | 126538 |
| ☑ | Morgushi | 29 | minor tail protein | 477 | 989 | 0 | B1 | 126538 |

## b. SIF: NCBI BLAST



| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | NP_943807 | No | 2023-01-08 | minor tail protein | minor tail protein [Mycobacterium phage PG1] >ref\|YP_008052106.1\| minor tail protein [Mycobacterium phage Newman] >ref\|YP_009005676.1\| minor tail protein [Mycobacterium phage Suffolk] >ref\|YP_009016818.1\| minor tail protein [Mycobacterium phage Vista] >ref\|YP_009018342.1\| minor tail protein [Mycobacterium phage JacAttac] >ref\|YP_009168209.1\| minor tail protein [Mycobacterium phage UncleHowie] >ref\|YP_009187539.1\| minor tail protein [Mycobacterium phage Swish] >ref\|YP_009189267.1\| minor tail subunit [Mycobacterium phage ShiVal] >ref\|YP_009190085.1\| minor tail protein [Mycobacterium phage Badfish] >ref\|YP_009191123.1\| minor tail protein [Mycobacterium phage Eremos] | 100 | 100 | 100 | 477 | 1 | 477 | 1 | 477 | 0 | 0 |
| ☑ | AIM50261 | No | 2021-12-01 | Minor tail subunit | minor tail protein [Mycobacterium phage Vivaldi] >gb\|ATN89273.1\| minor tail protein [Mycobacterium phage Horchata] >gb\|ATN91778.1\| minor tail protein [Mycobacterium phage Sheila] >gb\|AWH14269.1\| minor tail protein [Mycobacterium phage Olive] >gb\|WKW85646.1\| minor tail protein [Mycobacterium phage Basato] | 99.7904 | 100 | 100 | 477 | 1 | 477 | 1 | 477 | 0 | 0 |
| ☑ | UTN92182 | | | | minor tail protein [Mycobacterium phage Charles1] | 99.7904 | 100 | 100 | 477 | 1 | 477 | 1 | 477 | 0 | 0 |
| ☑ | AZS12465 | No | 2021-12-01 | minor tail protein | minor tail protein [Mycobacterium phage Roliet] | 99.5807 | 100 | 100 | 477 | 1 | 477 | 1 | 477 | 0 | 0 |
| ☑ | YP_009100838 | No | 2023-01-08 | Minor tail protein | minor tail protein [Mycobacterium phage Soto] >gb\|AHK12178.1\| minor tail protein [Mycobacterium phage Soto] | 99.7904 | 100 | 100 | 477 | 1 | 477 | 1 | 477 | 0 | 0 |

## c. SIF: HHPred

N/A

### d. SIF: Synteny-Phamerator (three genomes)

Aligns with Buckeye, Usavi, Waterdiva, LeeLot, MiniBoss, etc.

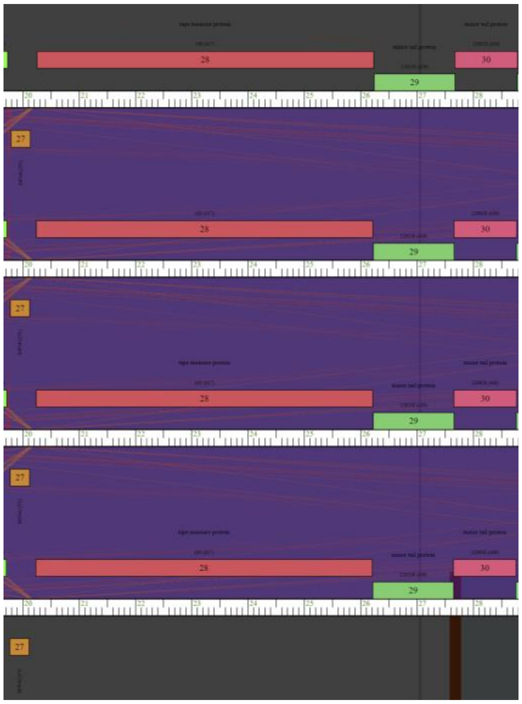


*Figure 1. (Top to bottom) Buckeye, Eugenia, Usavi, and Waterdiva.*

## 7. Any other important information.

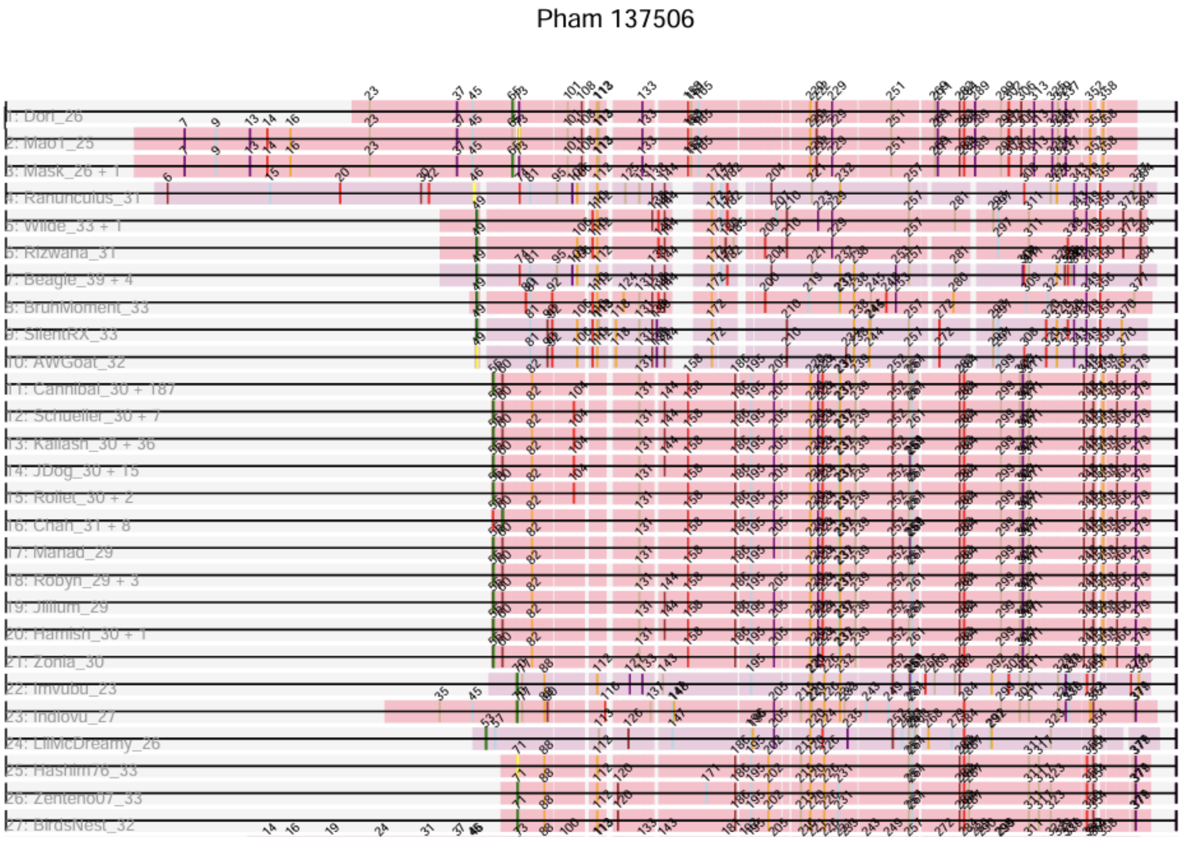
## CURATOR NAME: OLIVIA SIDOTI

## GENE NAME: EUGENIA GENE-30

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

1. Starterator



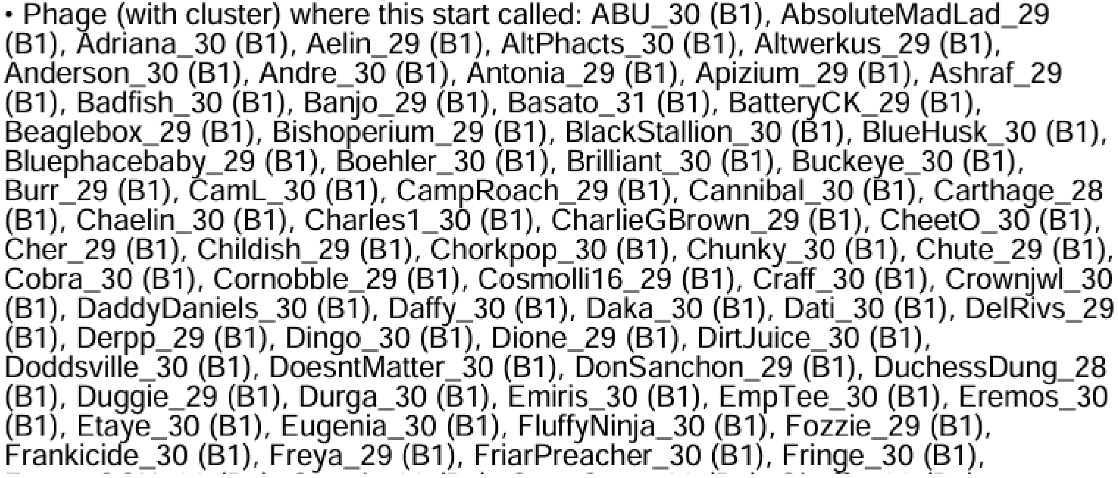
The start number called the most often in the published annotations is 56, it was called in 238 of the 498 non-draft genes in the pham.

Start 56:

- Found in 270 of 577 ( 46.8% ) of genes in pham

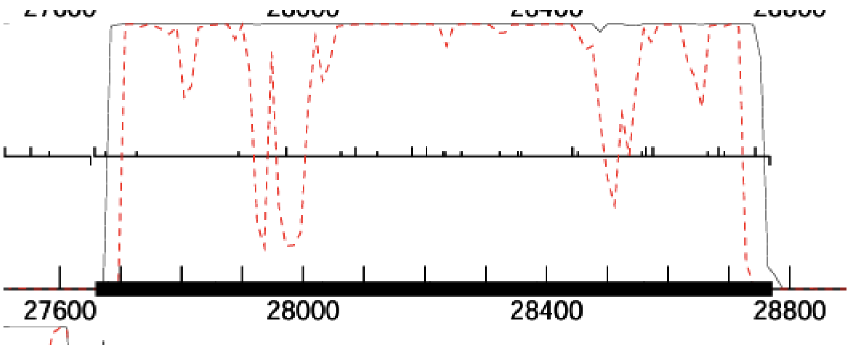
- Manual Annotations of this start: 238 of 498

- Called 96.7% of time when present



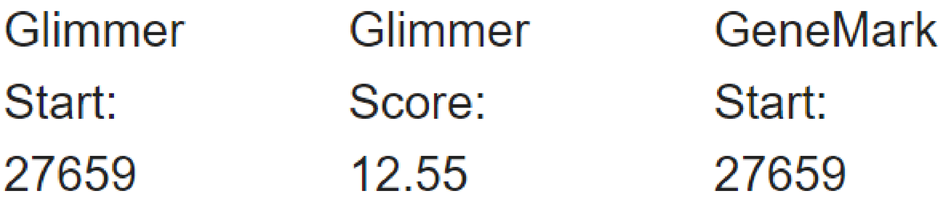
- Phage (with cluster) where this start called: ABU_30 (B1), AbsoluteMadLad_29 (B1), Adriana_30 (B1), Aelin_29 (B1), AltPhacts_30 (B1), Altwerkus_29 (B1), Anderson_30 (B1), Andre_30 (B1), Antonia_29 (B1), Apizium_29 (B1), Ashraf_29 (B1), Badfish_30 (B1), Banjo_29 (B1), Basato_31 (B1), BatteryCK_29 (B1), Beaglebox_29 (B1), Bishoperium_29 (B1), BlackStallion_30 (B1), BlueHusk_30 (B1), Bluephacebaby_29 (B1), Boehler_30 (B1), Brilliant_30 (B1), Buckeye_30 (B1), Burr_29 (B1), CamL_30 (B1), CampRoach_29 (B1), Cannibal_30 (B1), Carthage_28 (B1), Chaelin_30 (B1), Charles1_30 (B1), CharlieGBrown_29 (B1), CheetO_30 (B1), Cher_29 (B1), Childish_29 (B1), Chorkpop_30 (B1), Chunky_30 (B1), Chute_29 (B1), Cobra_30 (B1), Cornobble_29 (B1), Cosmolli16_29 (B1), Craff_30 (B1), Crownjwl_30 (B1), DaddyDaniels_30 (B1), Daffy_30 (B1), Daka_30 (B1), Dati_30 (B1), DelRivs_29 (B1), Derpp_29 (B1), Dingo_30 (B1), Dione_29 (B1), DirtJuice_30 (B1), Doddsville_30 (B1), DoesntMatter_30 (B1), DonSanchon_29 (B1), DuchessDung_28 (B1), Duggie_29 (B1), Durga_30 (B1), Emiris_30 (B1), EmpTee_30 (B1), Eremos_30 (B1), Etaye_30 (B1), Eugenia_30 (B1), FluffyNinja_30 (B1), Fozzie_29 (B1), Frankicide_30 (B1), Freya_29 (B1), FriarPreacher_30 (B1), Fringe_30 (B1),

2. GeneMark coding potential



3. Glimmer and GeneMark agreement

Phage: Eugenia Cluster: B1



| Glimmer Start: | Glimmer Score: | GeneMark Start: |
|---|---|---|
| 27659 | 12.55 | 27659 |

Yes, Glimmer and GeneMark are in agreement that the start site is at base pair 27659.

4. Longest open reading frame (ORF) without excessive gap

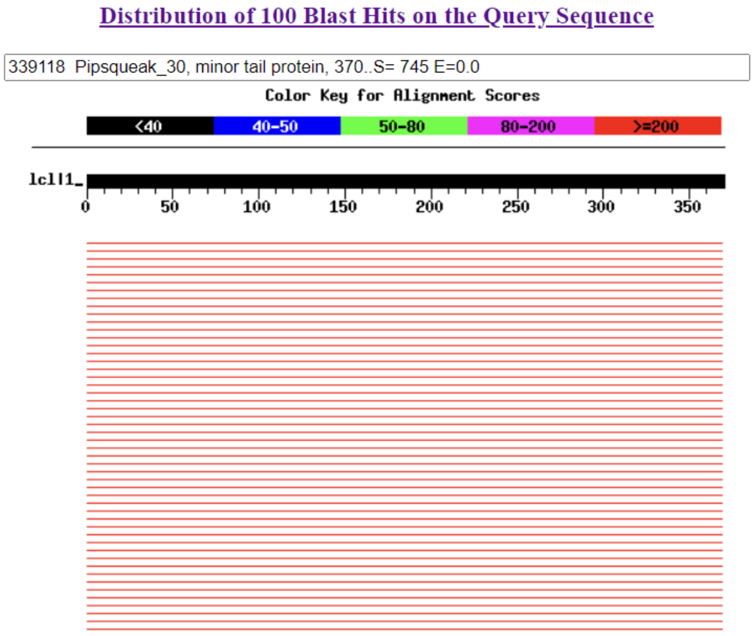
| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 27659 | 28771 | 1113 | -4 | 12 | 2.277 | -4.367 | TRUE | ATG | Select | ☑ |

5. Function. If no functional prediction is present, write “Hypothetical protein”.

Minor Tail Protein

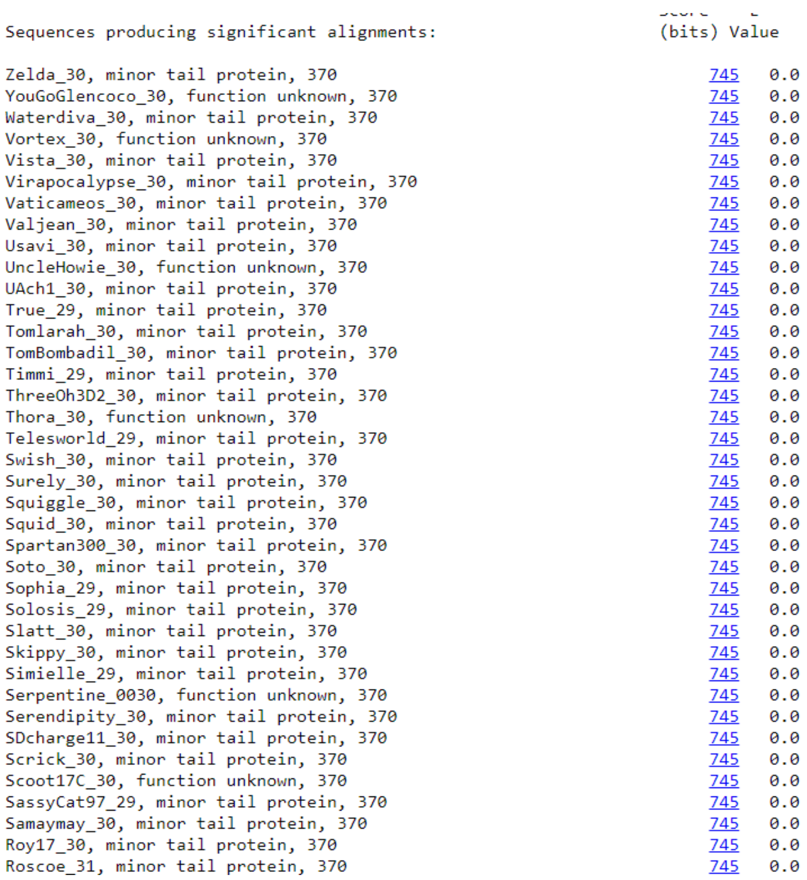
6. Supporting Information for Function (SIF)

a. SIF: PhageDb BLAST



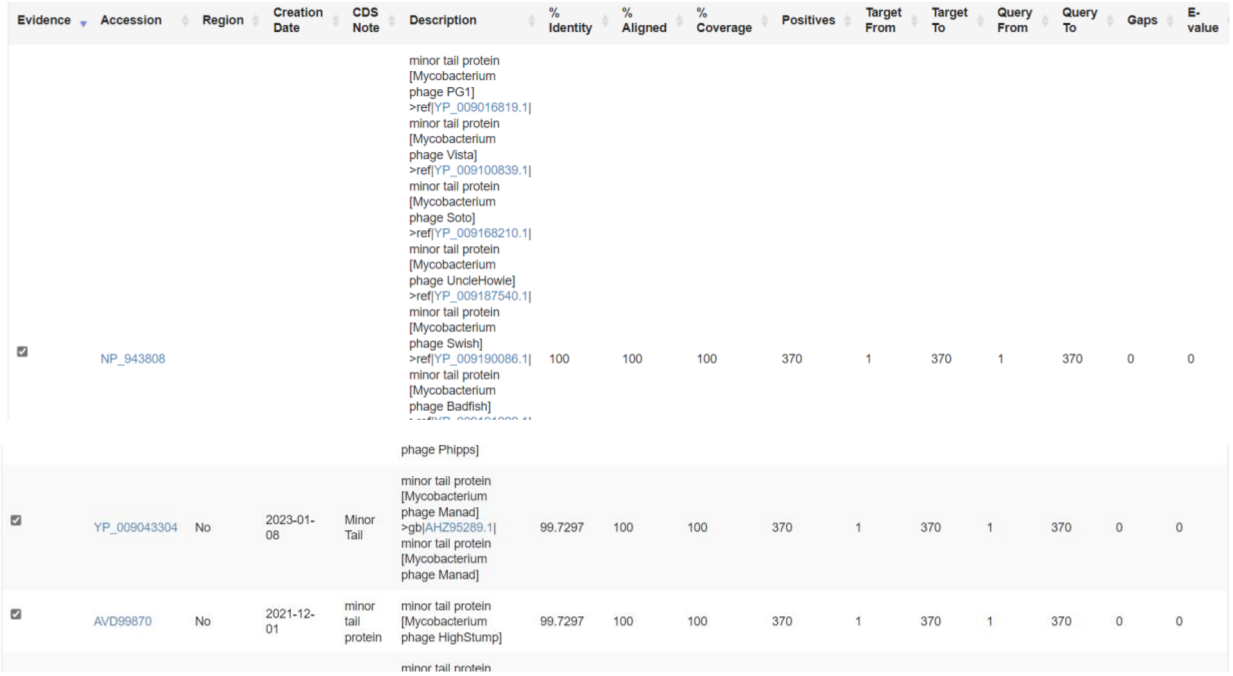
Distribution of 100 Blast Hits on the Query Sequence

339118 Pipsqueak_30, minor tail protein, 370..S= 745 E=0.0



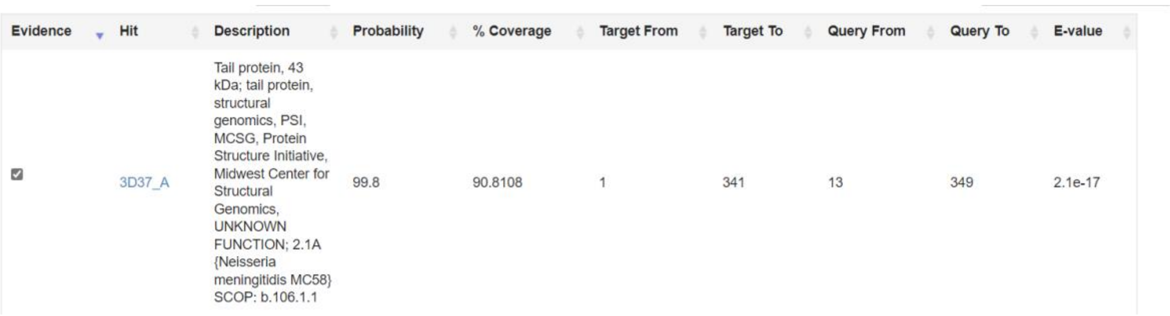
```
Sequences producing significant alignments:                       (bits) Value

Zelda_30, minor tail protein, 370                                     745     0.0
YouGoGlencoco_30, function unknown, 370                               745     0.0
Waterdiva_30, minor tail protein, 370                                 745     0.0
Vortex_30, function unknown, 370                                      745     0.0
Vista_30, minor tail protein, 370                                     745     0.0
Virapocalypse_30, minor tail protein, 370                             745     0.0
Vaticameos_30, minor tail protein, 370                                745     0.0
Valjean_30, minor tail protein, 370                                   745     0.0
Usavi_30, minor tail protein, 370                                     745     0.0
UncleHowie_30, function unknown, 370                                  745     0.0
UAch1_30, minor tail protein, 370                                     745     0.0
True_29, minor tail protein, 370                                      745     0.0
Tomlarah_30, minor tail protein, 370                                  745     0.0
TomBombadil_30, minor tail protein, 370                               745     0.0
Timmi_29, minor tail protein, 370                                     745     0.0
ThreeOh3D2_30, minor tail protein, 370                                745     0.0
Thora_30, function unknown, 370                                       745     0.0
Telesworld_29, minor tail protein, 370                                745     0.0
Swish_30, minor tail protein, 370                                     745     0.0
Surely_30, minor tail protein, 370                                    745     0.0
Squiggle_30, minor tail protein, 370                                  745     0.0
Squid_30, minor tail protein, 370                                     745     0.0
Spartan300_30, minor tail protein, 370                                745     0.0
Soto_30, minor tail protein, 370                                      745     0.0
Sophia_29, minor tail protein, 370                                    745     0.0
Solosis_29, minor tail protein, 370                                   745     0.0
Slatt_30, minor tail protein, 370                                     745     0.0
Skippy_30, minor tail protein, 370                                    745     0.0
Simielle_29, minor tail protein, 370                                  745     0.0
Serpentine_0030, function unknown, 370                                745     0.0
Serendipity_30, minor tail protein, 370                               745     0.0
SDcharge11_30, minor tail protein, 370                                745     0.0
Scrick_30, minor tail protein, 370                                    745     0.0
Scoot17C_30, function unknown, 370                                    745     0.0
SassyCat97_29, minor tail protein, 370                                745     0.0
Samaymay_30, minor tail protein, 370                                  745     0.0
Roy17_30, minor tail protein, 370                                     745     0.0
Roscoe_31, minor tail protein, 370                                    745     0.0
```

## b. SIF: NCBI BLAST



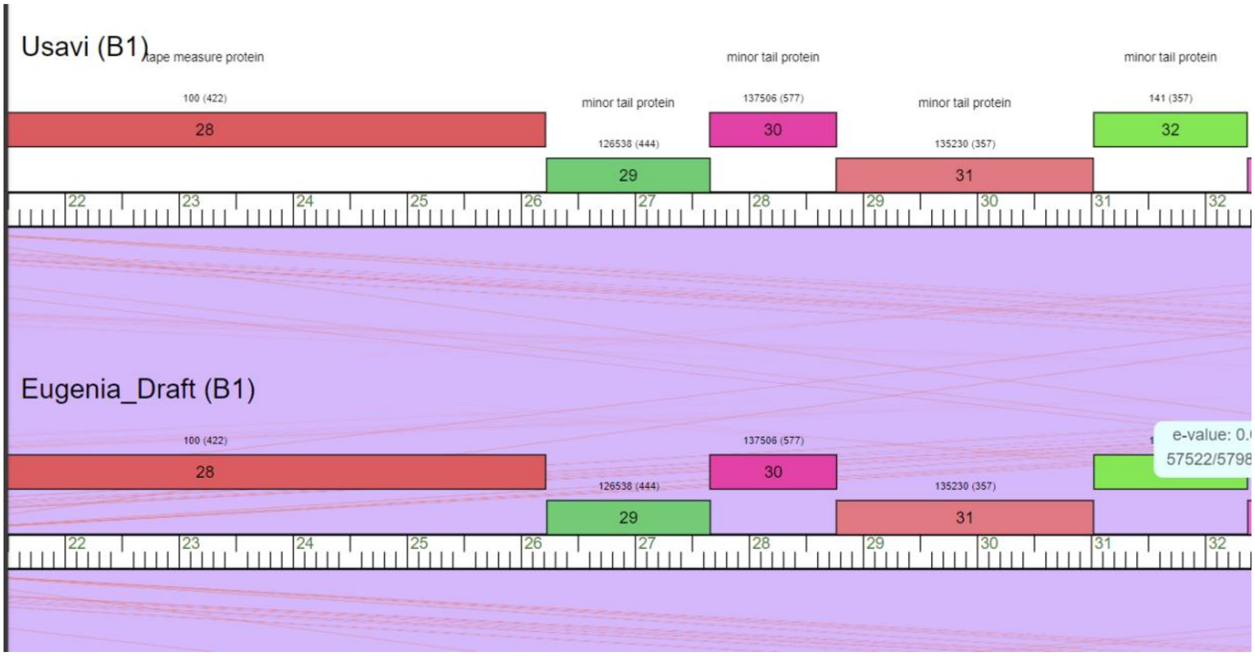
| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | NP_943808 | | | | minor tail protein [Mycobacterium phage PG1] >ref\|YP_009016819.1\| minor tail protein [Mycobacterium phage Vista] >ref\|YP_009100839.1\| minor tail protein [Mycobacterium phage Soto] >ref\|YP_009168210.1\| minor tail protein [Mycobacterium phage UncleHowie] >ref\|YP_009187540.1\| minor tail protein [Mycobacterium phage Swish] >ref\|YP_009190086.1\| minor tail protein [Mycobacterium phage Badfish] | 100 | 100 | 100 | 370 | 1 | 370 | 1 | 370 | 0 | 0 |
| | | | | | phage Phipps] | | | | | | | | | | |
| ☑ | YP_009043304 | No | 2023-01-08 | Minor Tail | minor tail protein [Mycobacterium phage Manad] >gb\|AHZ95289.1\| minor tail protein [Mycobacterium phage Manad] | 99.7297 | 100 | 100 | 370 | 1 | 370 | 1 | 370 | 0 | 0 |
| ☑ | AVD99870 | No | 2021-12-01 | minor tail protein | minor tail protein [Mycobacterium phage HighStump] | 99.7297 | 100 | 100 | 370 | 1 | 370 | 1 | 370 | 0 | 0 |
| | | | | | minor tail protein | | | | | | | | | | |

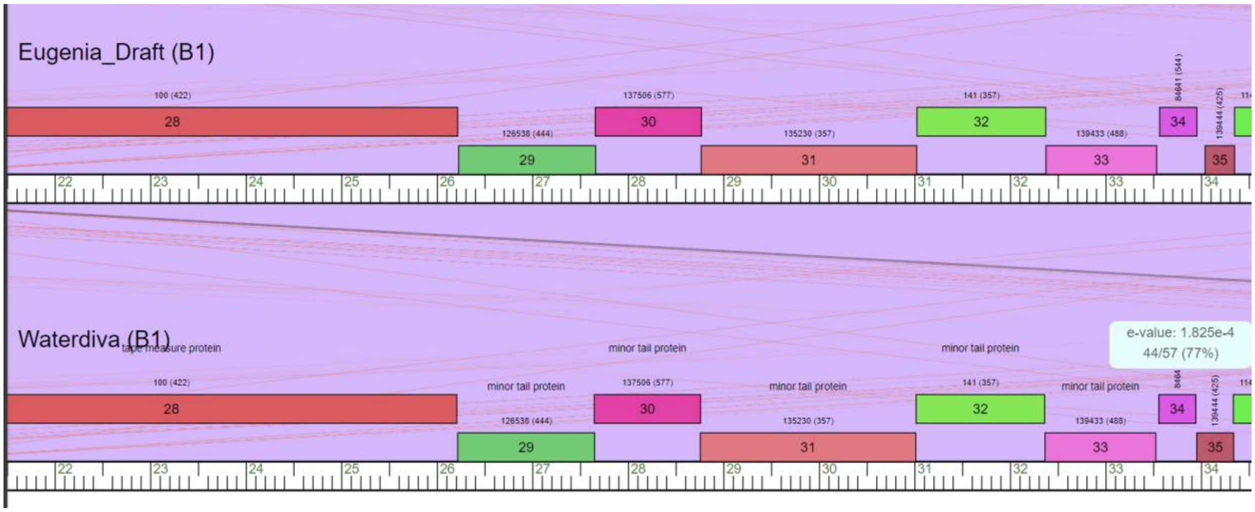
## c. SIF: HHPred



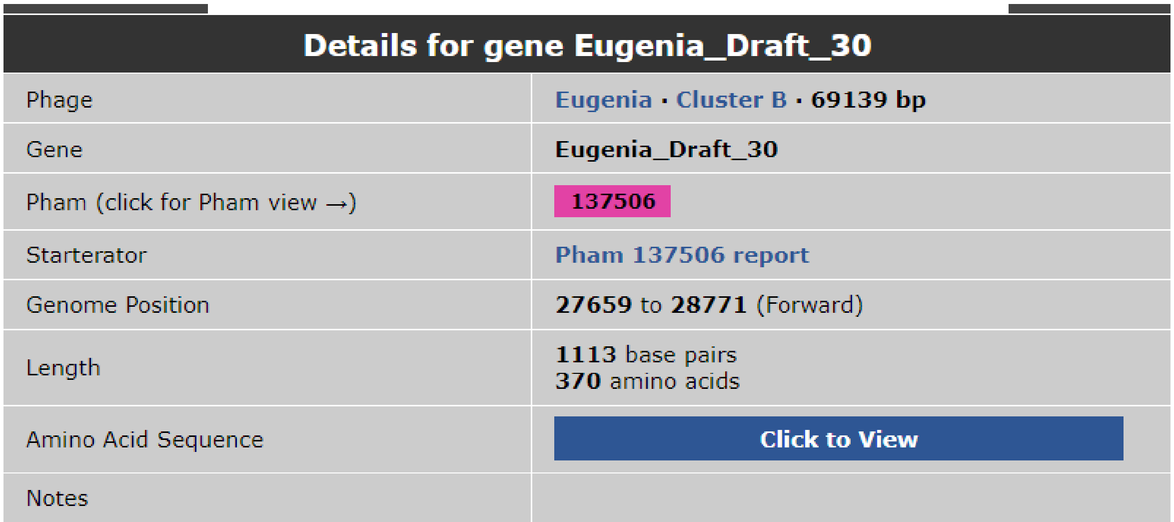
| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | 3D37_A | Tail protein, 43 kDa; tail protein, structural genomics, PSI, MCSG, Protein Structure Initiative, Midwest Center for Structural Genomics, UNKNOWN FUNCTION; 2.1A {Neisseria meningitidis MC58} SCOP: b.106.1.1 | 99.8 | 90.8108 | 1 | 341 | 13 | 349 | 2.1e-17 |

## d. SIF: Synteny-Phamerator (three genomes)





7. Any other important information.



### Details for gene Eugenia_Draft_30

| | |
|---|---|
| Phage | Eugenia · Cluster B · **69139 bp** |
| Gene | **Eugenia_Draft_30** |
| Pham (click for Pham view →) | **137506** |
| Starterator | Pham 137506 report |
| Genome Position | **27659** to **28771** (Forward) |
| Length | **1113** base pairs<br>**370** amino acids |
| Amino Acid Sequence | Click to View |
| Notes | |

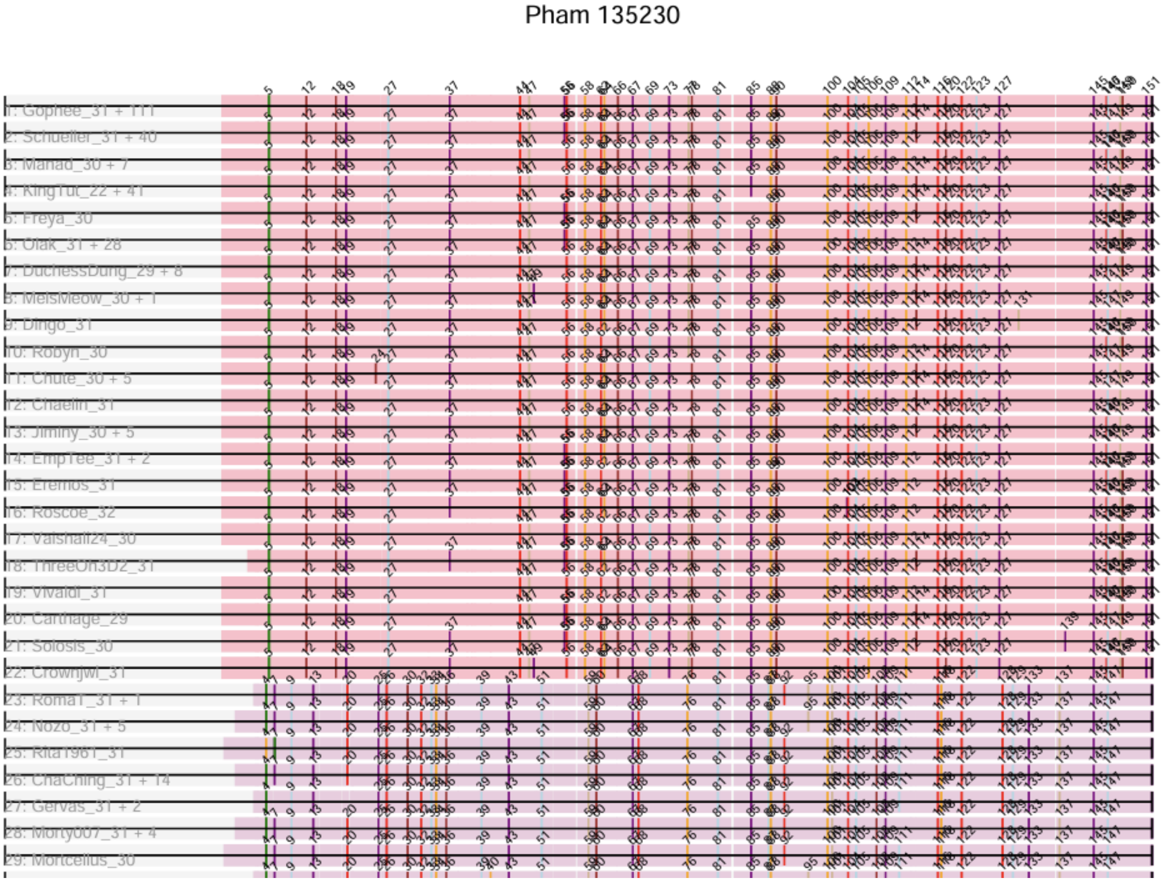
## SSCURATOR NAME: OLIVIA SIDOTI

## GENE NAME: EUGENIA GENE-31

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

1. Starterator

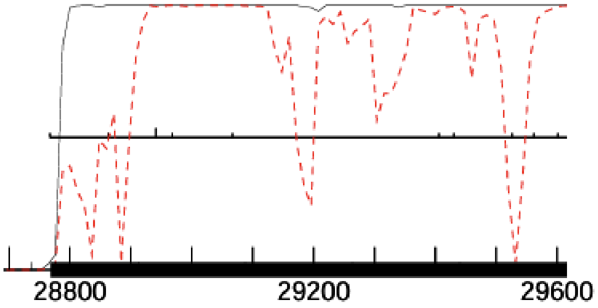


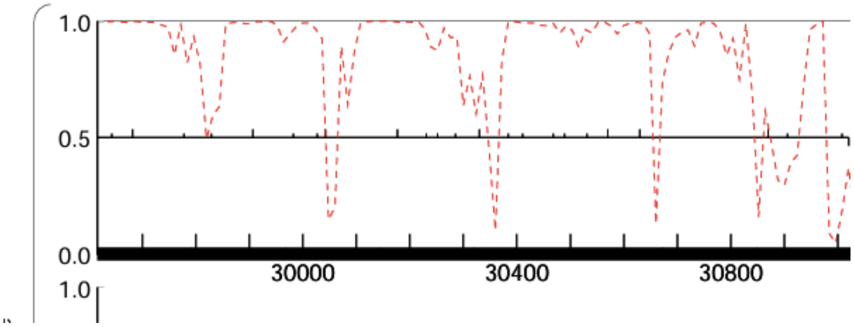
Pham number 135230 has 357 members, 37 are drafts.

The start number called the most often in the published annotations is 5, it was called in 275 of the 320 non-draft genes in the pham

Start 5: • Found in 312 of 357 ( 87.4% ) of genes in pham • Manual Annotations of this start: 275 of 320 • Called 99.7% of time when present

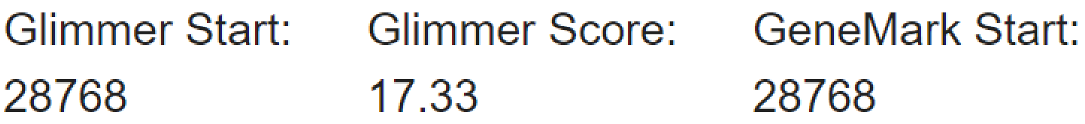
## 2. GeneMark coding potential





## 3. Glimmer and GeneMark agreement

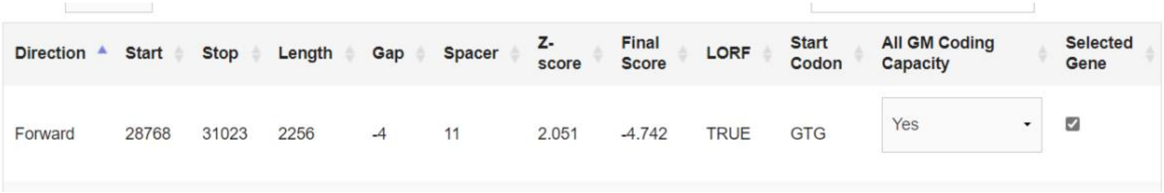
Phage: Eugenia Cluster: B1



| Glimmer Start: | Glimmer Score: | GeneMark Start: |
|---|---|---|
| 28768 | 17.33 | 28768 |

Yes, Glimmer and GeneMark are in agreement that the start is at base pair 28768.

## 4. Longest open reading frame (ORF) without excessive gap



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 28768 | 31023 | 2256 | -4 | 11 | 2.051 | -4.742 | TRUE | GTG | Yes | ☑ |

Yes, this gene is the longest open reading frame.

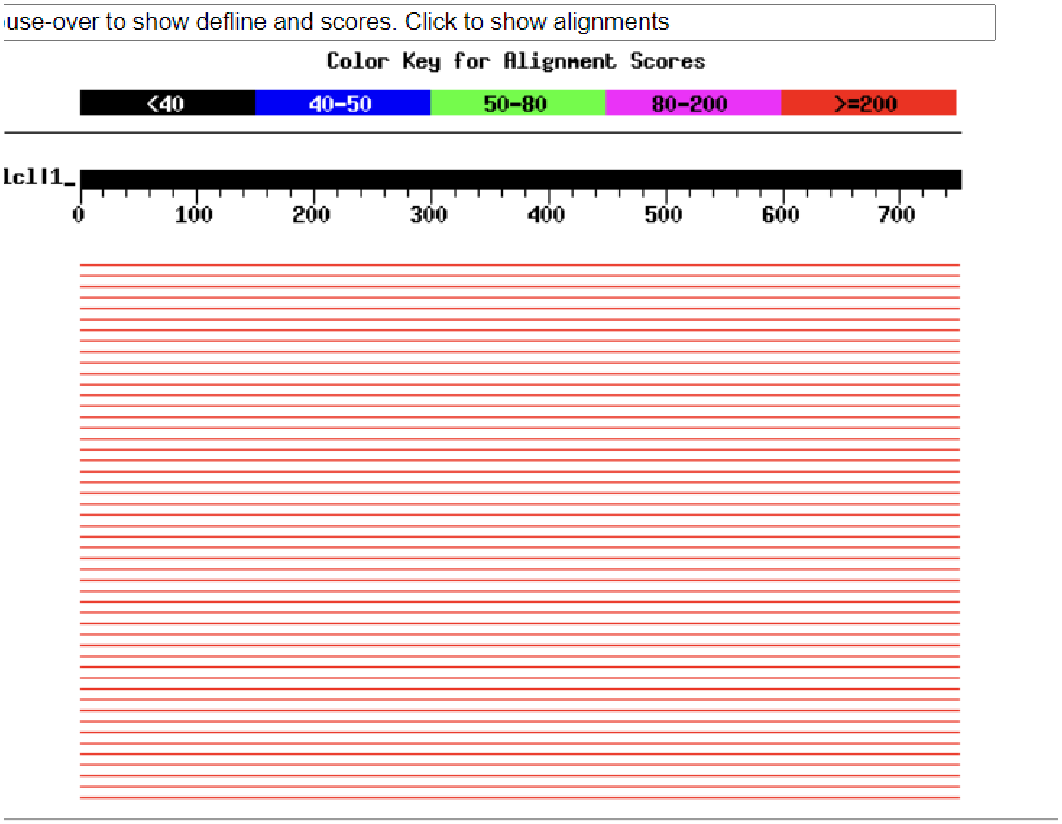
## 5. Function. If no functional prediction is present, write "Hypothetical protein".

Minor Tail Protein

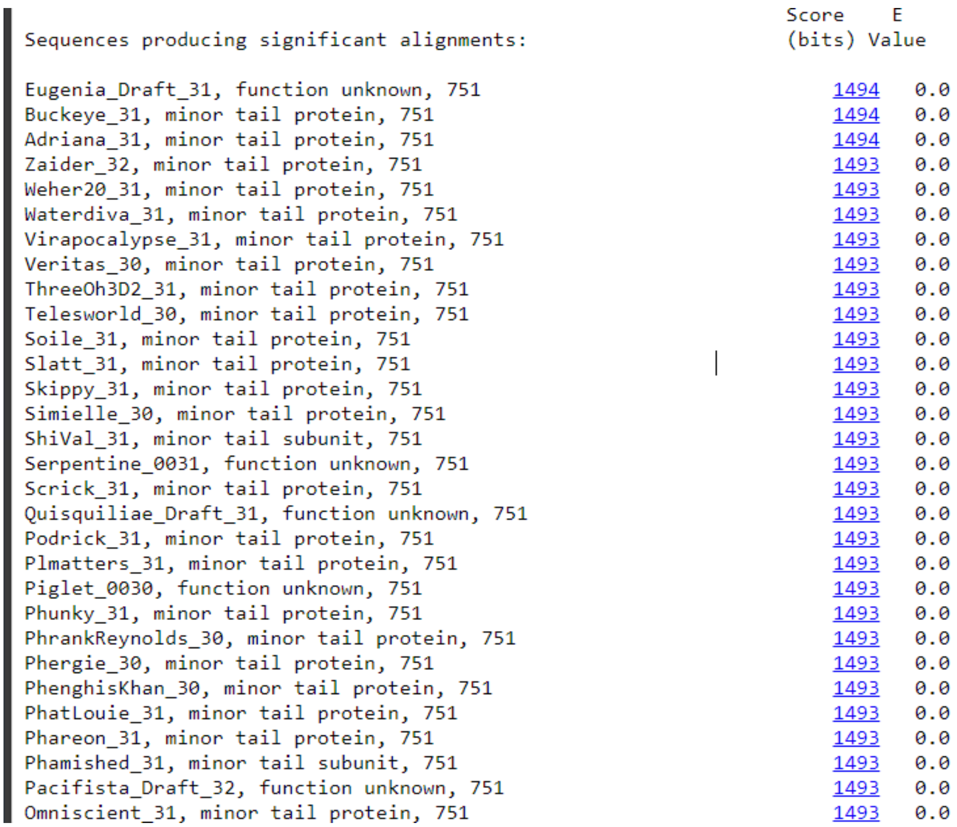
## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST

## Distribution of 100 Blast Hits on the Query Sequence

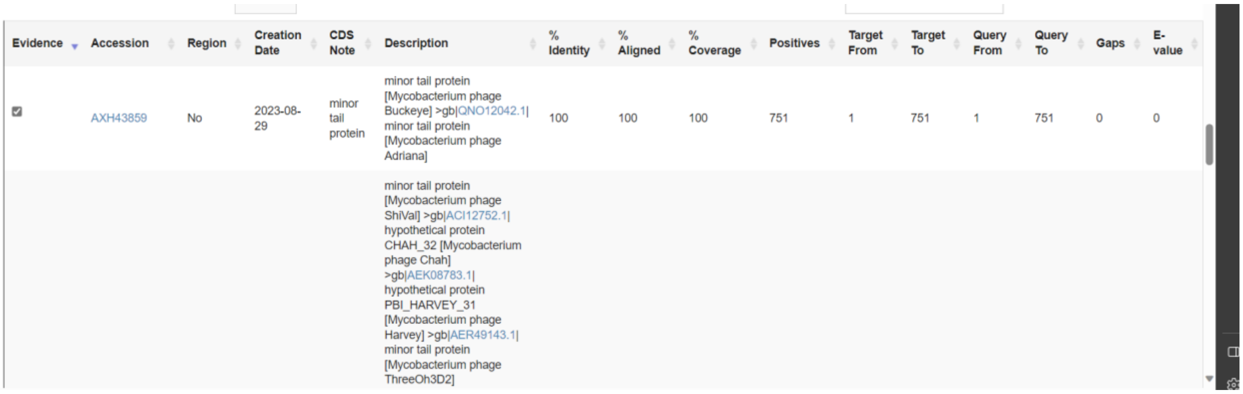


use-over to show defline and scores. Click to show alignments



```
                                                                  Score     E
Sequences producing significant alignments:                       (bits)  Value

Eugenia_Draft_31, function unknown, 751                              1494   0.0
Buckeye_31, minor tail protein, 751                                  1494   0.0
Adriana_31, minor tail protein, 751                                  1494   0.0
Zaider_32, minor tail protein, 751                                   1493   0.0
Weher20_31, minor tail protein, 751                                  1493   0.0
Waterdiva_31, minor tail protein, 751                                1493   0.0
Virapocalypse_31, minor tail protein, 751                            1493   0.0
Veritas_30, minor tail protein, 751                                  1493   0.0
ThreeOh3D2_31, minor tail protein, 751                               1493   0.0
Telesworld_30, minor tail protein, 751                               1493   0.0
Soile_31, minor tail protein, 751                                    1493   0.0
Slatt_31, minor tail protein, 751                                    1493   0.0
Skippy_31, minor tail protein, 751                                   1493   0.0
Simielle_30, minor tail protein, 751                                 1493   0.0
ShiVal_31, minor tail subunit, 751                                   1493   0.0
Serpentine_0031, function unknown, 751                               1493   0.0
Scrick_31, minor tail protein, 751                                   1493   0.0
Quisquiliae_Draft_31, function unknown, 751                          1493   0.0
Podrick_31, minor tail protein, 751                                  1493   0.0
Plmatters_31, minor tail protein, 751                                1493   0.0
Piglet_0030, function unknown, 751                                   1493   0.0
Phunky_31, minor tail protein, 751                                   1493   0.0
PhrankReynolds_30, minor tail protein, 751                           1493   0.0
Phergie_30, minor tail protein, 751                                  1493   0.0
PhenghisKhan_30, minor tail protein, 751                             1493   0.0
PhatLouie_31, minor tail protein, 751                                1493   0.0
Phareon_31, minor tail protein, 751                                  1493   0.0
Phamished_31, minor tail subunit, 751                                1493   0.0
Pacifista_Draft_32, function unknown, 751                            1493   0.0
Omniscient_31, minor tail protein, 751                               1493   0.0
```

### b. SIF: NCBI BLAST

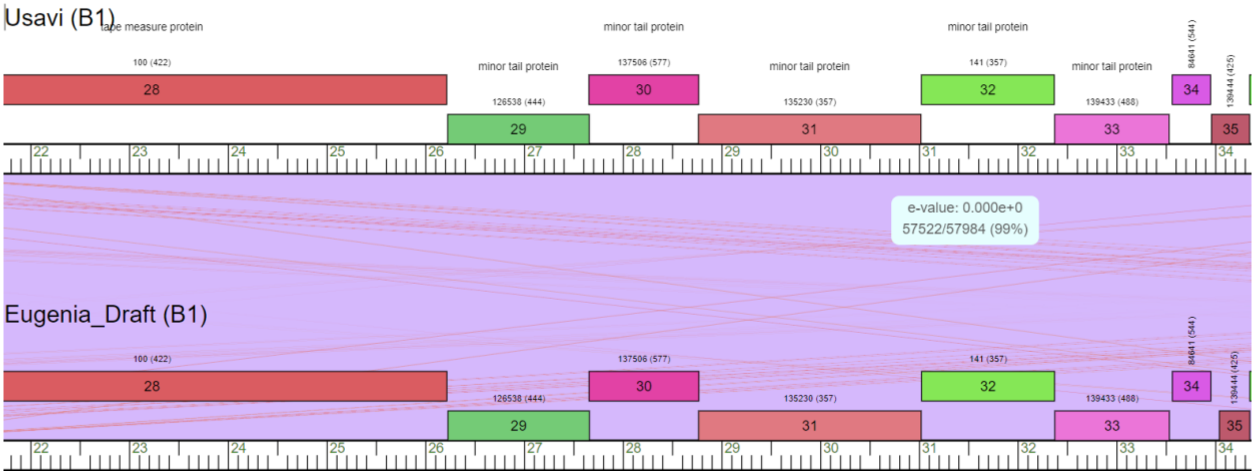


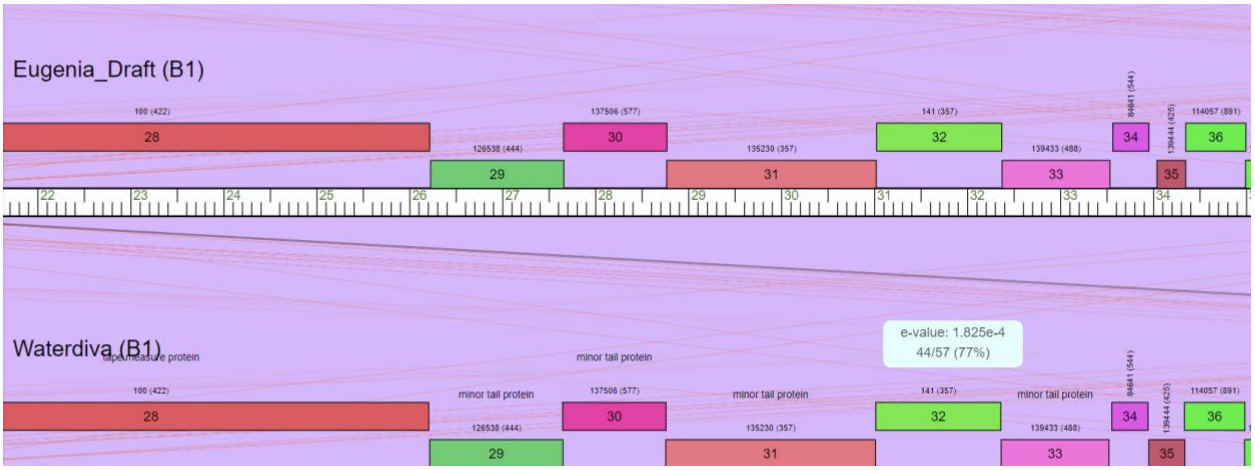
| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | AXH43859 | No | 2023-08-29 | minor tail protein | minor tail protein [Mycobacterium phage Buckeye] >gb\|QNO12042.1\| minor tail protein [Mycobacterium phage Adriana] | 100 | 100 | 100 | 751 | 1 | 751 | 1 | 751 | 0 | 0 |
| | | | | | minor tail protein [Mycobacterium phage ShiVal] >gb\|ACI12752.1\| hypothetical protein CHAH_32 [Mycobacterium phage Chah] >gb\|AEK08783.1\| hypothetical protein PBI_HARVEY_31 [Mycobacterium phage Harvey] >gb\|AER49143.1\| minor tail protein [Mycobacterium phage ThreeOh3D2] | | | | | | | | | | |

### c. SIF: HHPred

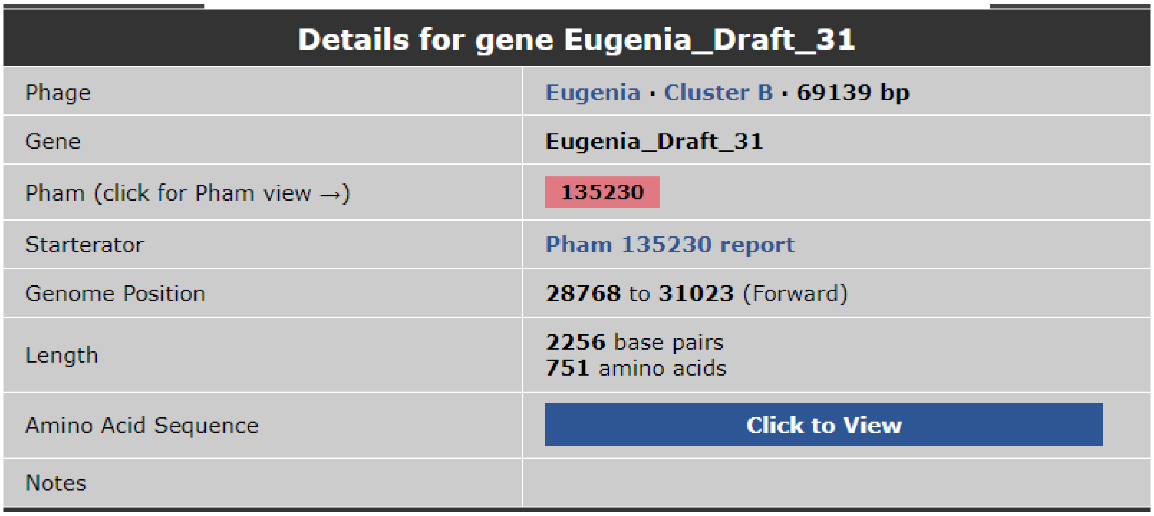
No Evidence from HHPred

### d. SIF: Synteny-Phamerator (three genomes)





7. Any other important information.



## Details for gene Eugenia_Draft_31

| | |
|---|---|
| Phage | Eugenia · Cluster B · **69139 bp** |
| Gene | **Eugenia_Draft_31** |
| Pham (click for Pham view →) | **135230** |
| Starterator | **Pham 135230 report** |
| Genome Position | **28768** to **31023** (Forward) |
| Length | **2256** base pairs<br>**751** amino acids |
| Amino Acid Sequence | **Click to View** |
| Notes | |

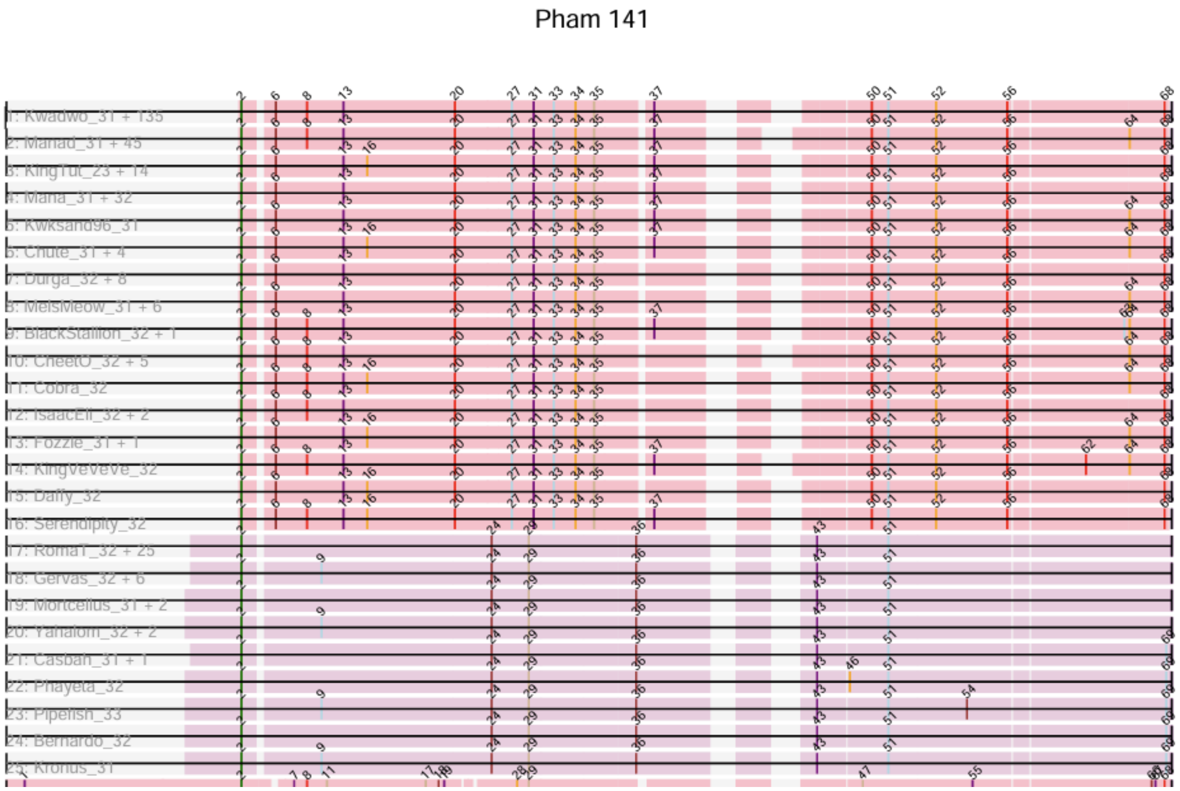
## CURATOR NAME: OLIVIA SIDOTI

## GENE NAME: EUGENIA GENE-32

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

1. Starterator



Pham number 141 has 357 members, 37 are drafts.

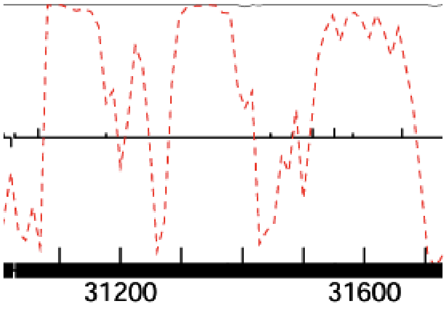
The start number called the most often in the published annotations is 2, it was called in 320 of the 320 non-draft genes in the pham.

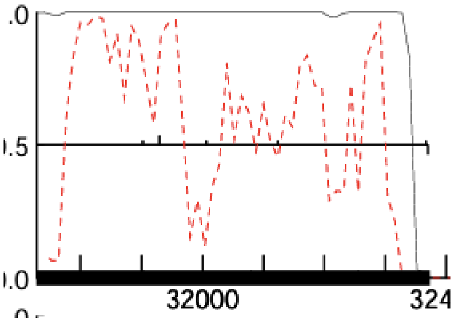
Start 2:

- Found in 357 of 357 ( 100.0% ) of genes in pham

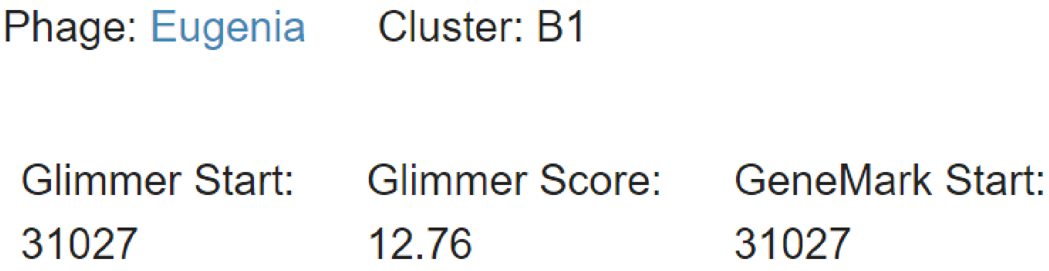
- Manual Annotations of this start: 320 of 320
- Called 100.0% of time when present

## 2. GeneMark coding potential





## 3. Glimmer and GeneMark agreement



Phage: Eugenia Cluster: B1

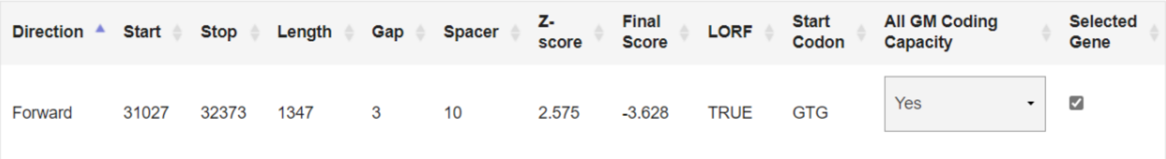
Glimmer Start: 31027

Glimmer Score: 12.76

GeneMark Start: 31027

Yes, Glimmer and GeneMark are in agreement that the start is at base pair 31027.

## 4. Longest open reading frame (ORF) without excessive gap



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 31027 | 32373 | 1347 | 3 | 10 | 2.575 | -3.628 | TRUE | GTG | Yes | ☑ |

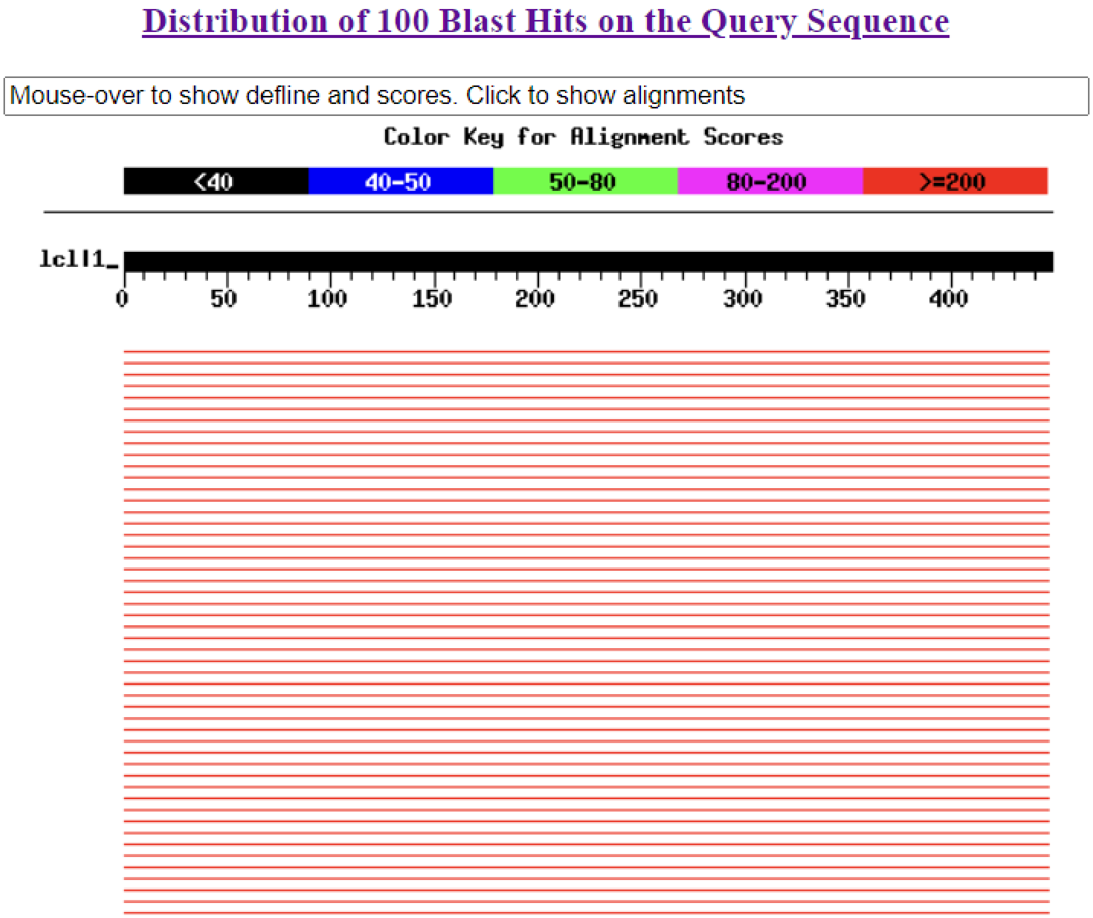
Yes, this gene has the longest open reading frame.

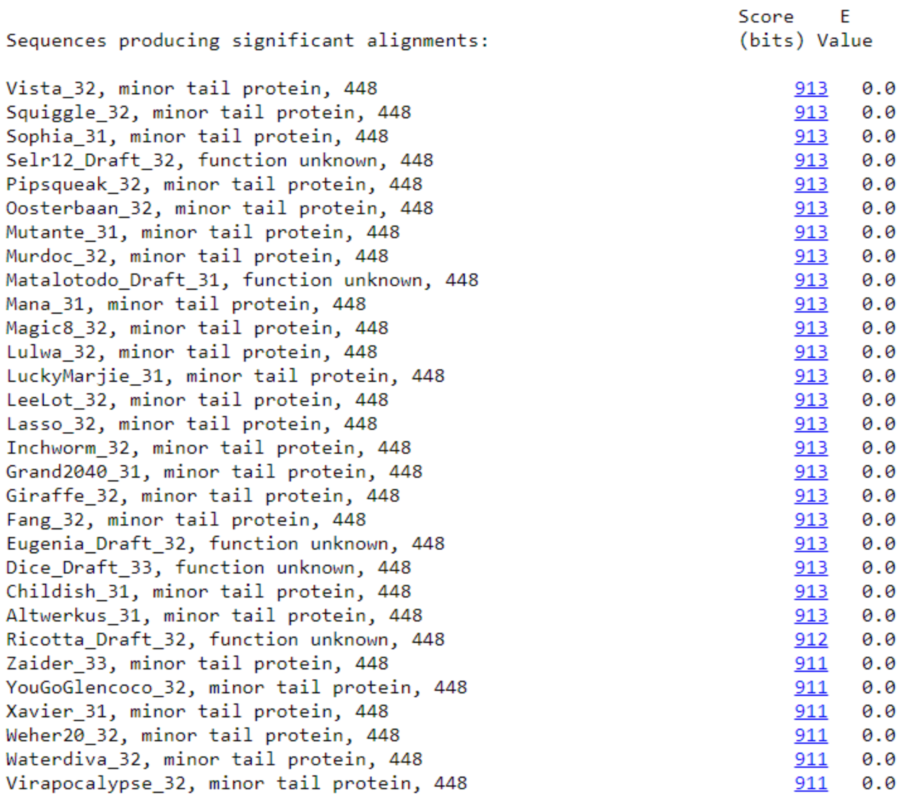
## 5. Function. If no functional prediction is present, write "Hypothetical protein".

Minor Tail Protein

6. Supporting Information for Function (SIF)

a. SIF: PhageDb BLAST

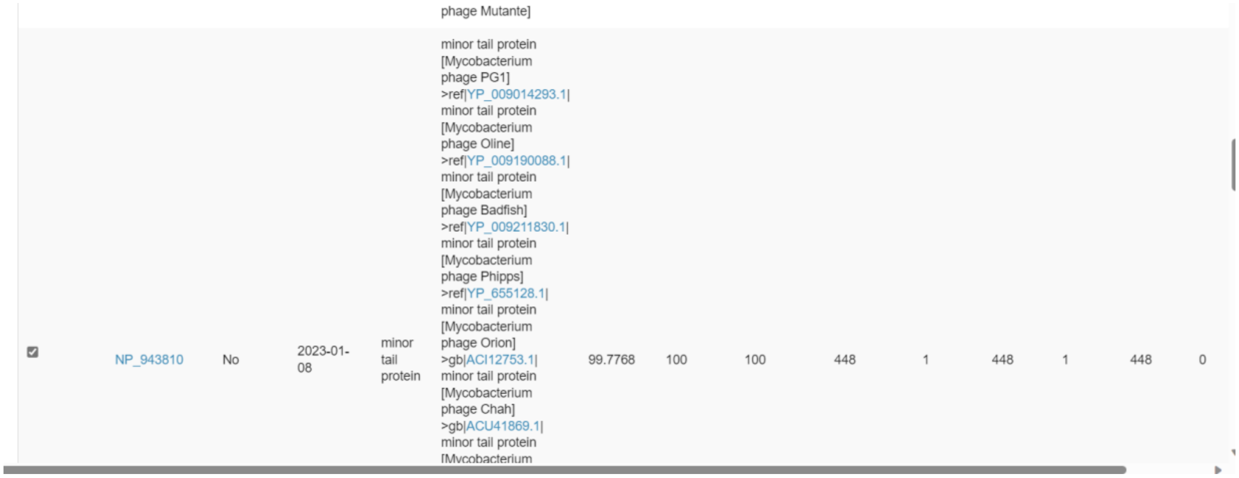




```
                                                                  Score     E
Sequences producing significant alignments:                       (bits)  Value

Vista_32, minor tail protein, 448                                      913   0.0
Squiggle_32, minor tail protein, 448                                   913   0.0
Sophia_31, minor tail protein, 448                                     913   0.0
Selr12_Draft_32, function unknown, 448                                 913   0.0
Pipsqueak_32, minor tail protein, 448                                  913   0.0
Oosterbaan_32, minor tail protein, 448                                 913   0.0
Mutante_31, minor tail protein, 448                                    913   0.0
Murdoc_32, minor tail protein, 448                                     913   0.0
Matalotodo_Draft_31, function unknown, 448                             913   0.0
Mana_31, minor tail protein, 448                                       913   0.0
Magic8_32, minor tail protein, 448                                     913   0.0
Lulwa_32, minor tail protein, 448                                      913   0.0
LuckyMarjie_31, minor tail protein, 448                                913   0.0
LeeLot_32, minor tail protein, 448                                     913   0.0
Lasso_32, minor tail protein, 448                                      913   0.0
Inchworm_32, minor tail protein, 448                                   913   0.0
Grand2040_31, minor tail protein, 448                                  913   0.0
Giraffe_32, minor tail protein, 448                                    913   0.0
Fang_32, minor tail protein, 448                                       913   0.0
Eugenia_Draft_32, function unknown, 448                                913   0.0
Dice_Draft_33, function unknown, 448                                   913   0.0
Childish_31, minor tail protein, 448                                   913   0.0
Altwerkus_31, minor tail protein, 448                                  913   0.0
Ricotta_Draft_32, function unknown, 448                                912   0.0
Zaider_33, minor tail protein, 448                                     911   0.0
YouGoGlencoco_32, minor tail protein, 448                              911   0.0
Xavier_31, minor tail protein, 448                                     911   0.0
Weher20_32, minor tail protein, 448                                    911   0.0
Waterdiva_32, minor tail protein, 448                                  911   0.0
Virapocalypse_32, minor tail protein, 448                              911   0.0
```

## b. SIF: NCBI BLAST

| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gap |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | YP_009016821 | | | | minor tail protein [Mycobacterium phage Vista] >gb\|ADA83859.1\| minor tail protein [Mycobacterium phage Fang] >gb\|AEK07200.1\| minor tail protein [Mycobacterium phage Oosterbaan] >gb\|AEO93973.1\| minor tail protein [Mycobacterium phage Murdoc] >gb\|AIM49770.1\| minor tail protein [Mycobacterium phage Lasso] >gb\|AJD82352.1\| minor tail protein [Mycobacterium phage Pipsqueak] | 100 | 100 | 100 | 448 | 1 | 448 | 1 | 448 | 0 |

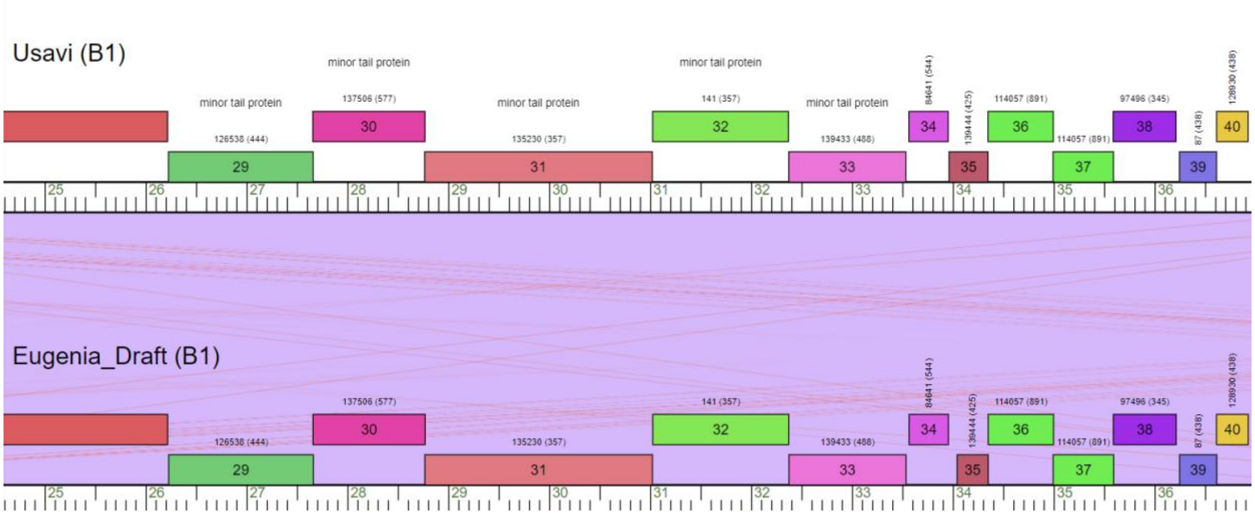


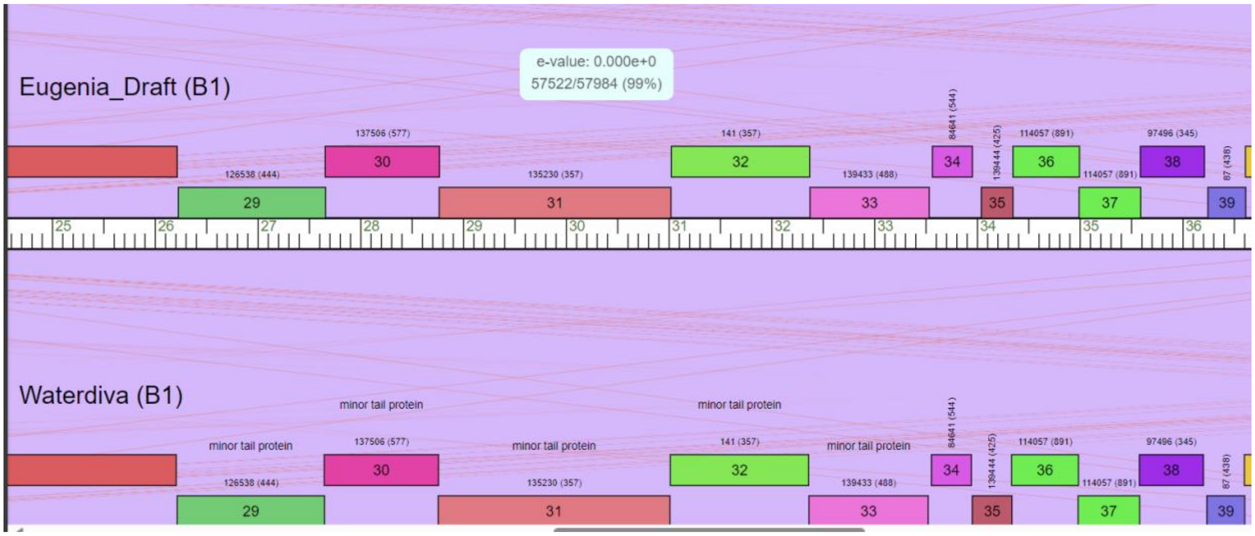
| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | phage Mutante] | | | | | | | | | |
| ☑ | NP_943810 | No | 2023-01-08 | minor tail protein | minor tail protein [Mycobacterium phage PG1] >ref\|YP_009014293.1\| minor tail protein [Mycobacterium phage Oline] >ref\|YP_009190088.1\| minor tail protein [Mycobacterium phage Badfish] >ref\|YP_009211830.1\| minor tail protein [Mycobacterium phage Phipps] >ref\|YP_655128.1\| minor tail protein [Mycobacterium phage Orion] >gb\|ACI12753.1\| minor tail protein [Mycobacterium phage Chah] >gb\|ACU41869.1\| minor tail protein [Mycobacterium | 99.7768 | 100 | 100 | 448 | 1 | 448 | 1 | 448 | 0 |

## c. SIF: HHPred

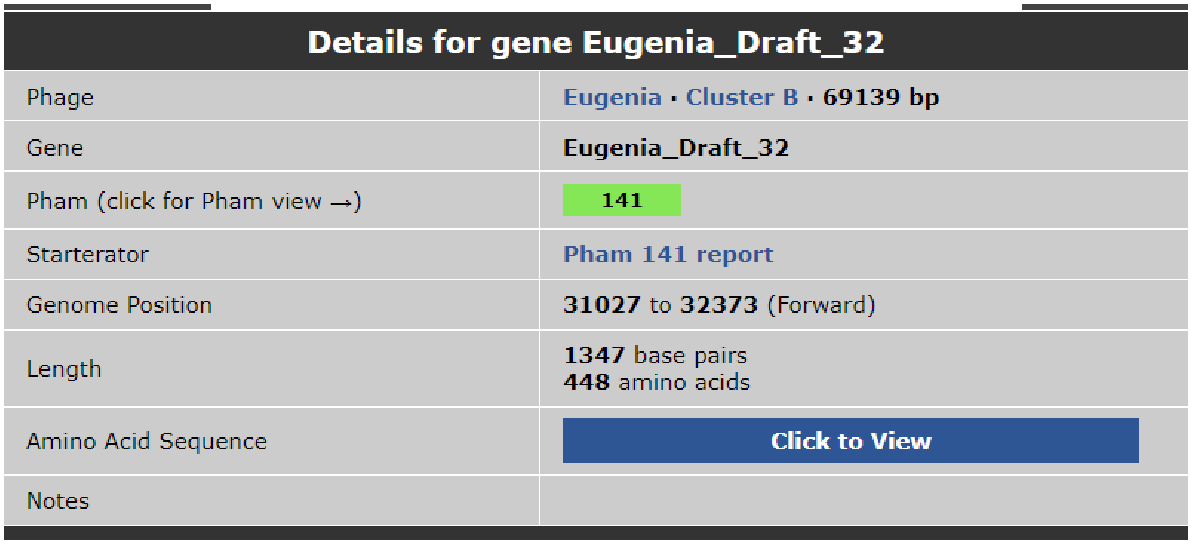
No evidence from HHPred

## d. SIF: Synteny-Phamerator (three genomes)





7. Any other important information.



| Details for gene Eugenia_Draft_32 | |
|---|---|
| Phage | Eugenia · Cluster B · **69139** bp |
| Gene | **Eugenia_Draft_32** |
| Pham (click for Pham view →) | **141** |
| Starterator | **Pham 141 report** |
| Genome Position | **31027** to **32373** (Forward) |
| Length | **1347** base pairs<br>**448** amino acids |
| Amino Acid Sequence | **Click to View** |
| Notes | |

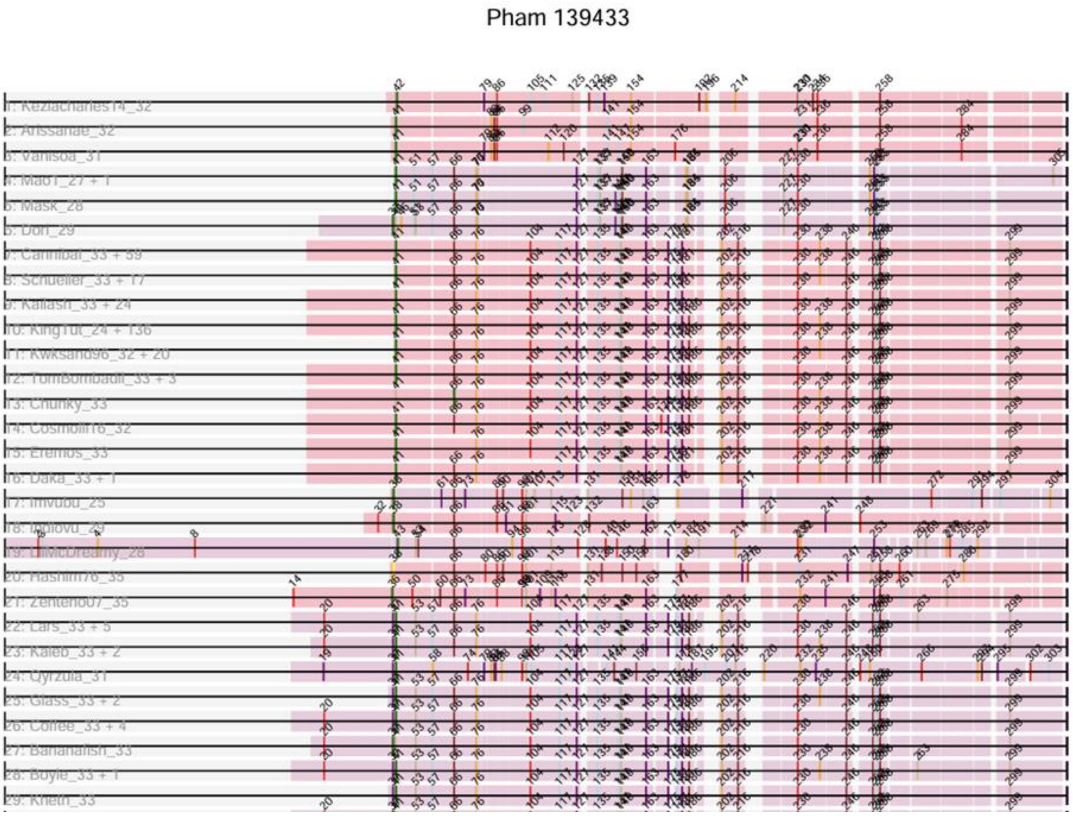
## CURATOR NAME: OLIVIA SIDOTI

## GENE NAME:EUGENIA GENE 33

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

### 1. Starterator



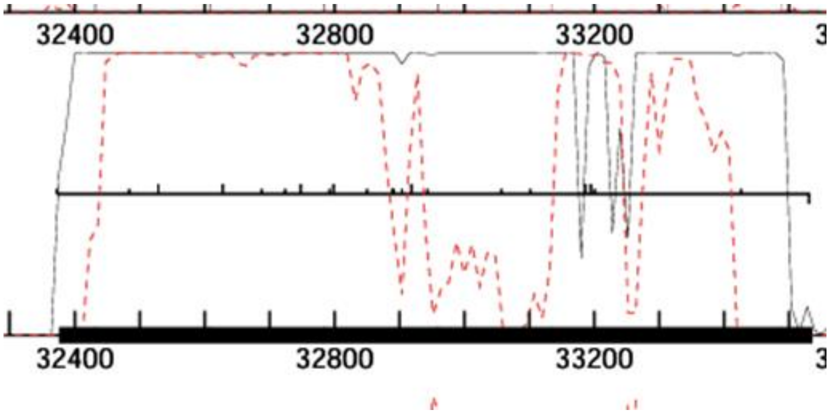
Pham number 139433 has 488 members, 55 are drafts.

The start number called the most often in the published annotations is 41, it was called in 281 of the 433 non-draft genes in the pham.

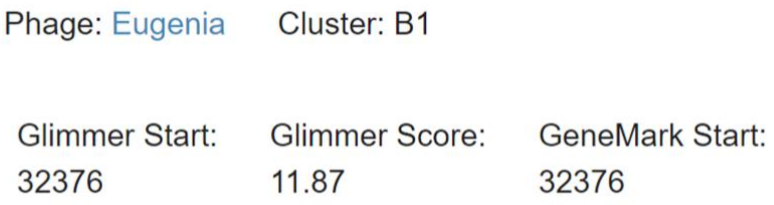
Start 41:

- Found in 323 of 488 ( 66.2% ) of genes in pham
- Manual Annotations of this start: 281 of 433
- Called 95.0% of time when present

## 2. GeneMark coding potential

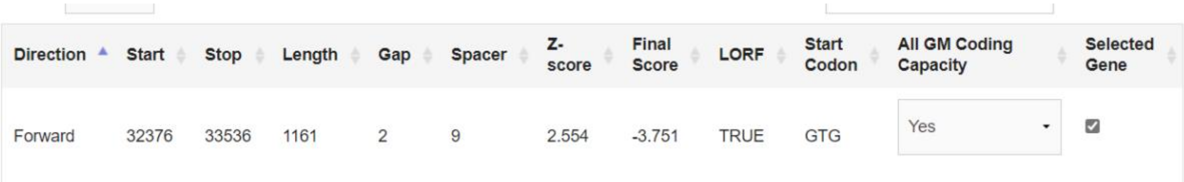


## 3. Glimmer and GeneMark agreement


Phage: Eugenia Cluster: B1

| Glimmer Start: | Glimmer Score: | GeneMark Start: |
|---|---|---|
| 32376 | 11.87 | 32376 |

## 4. Longest open reading frame (ORF) without excessive gap



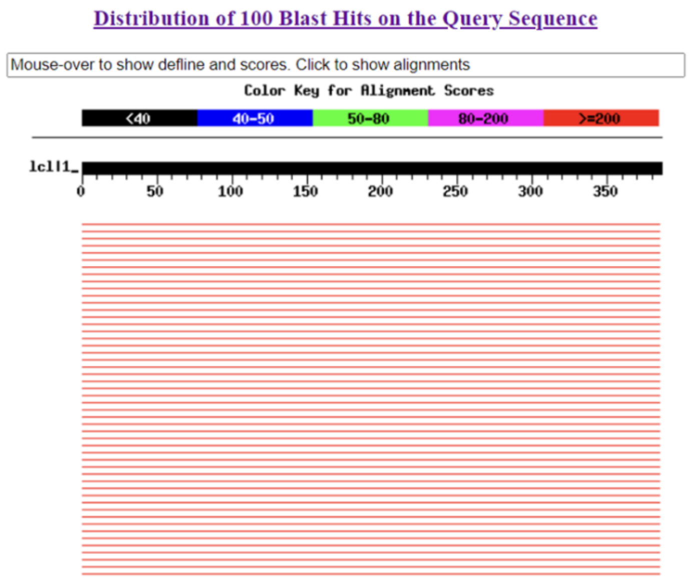
| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 32376 | 33536 | 1161 | 2 | 9 | 2.554 | -3.751 | TRUE | GTG | Yes | ☑ |

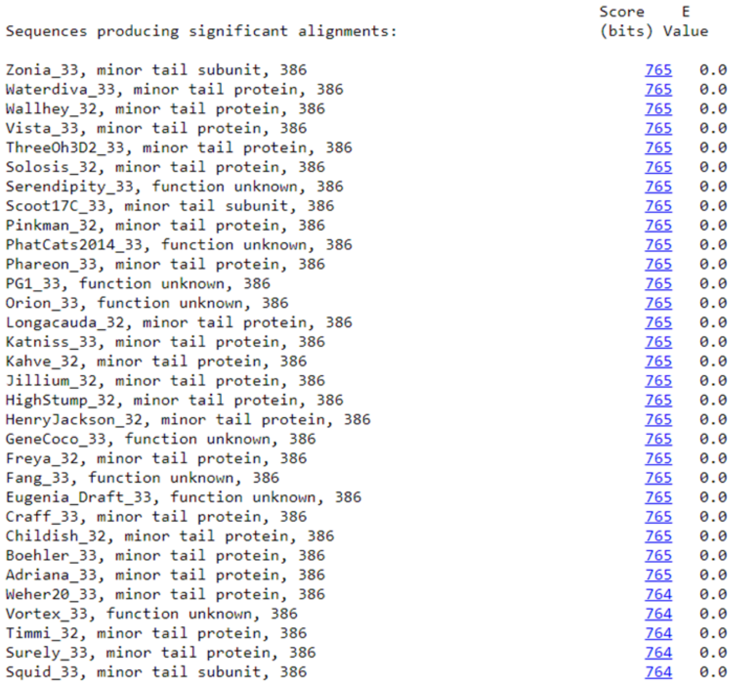
5. Function. If no functional prediction is present, write “Hypothetical protein”.

Minor tail protein

6. Supporting Information for Function (SIF)

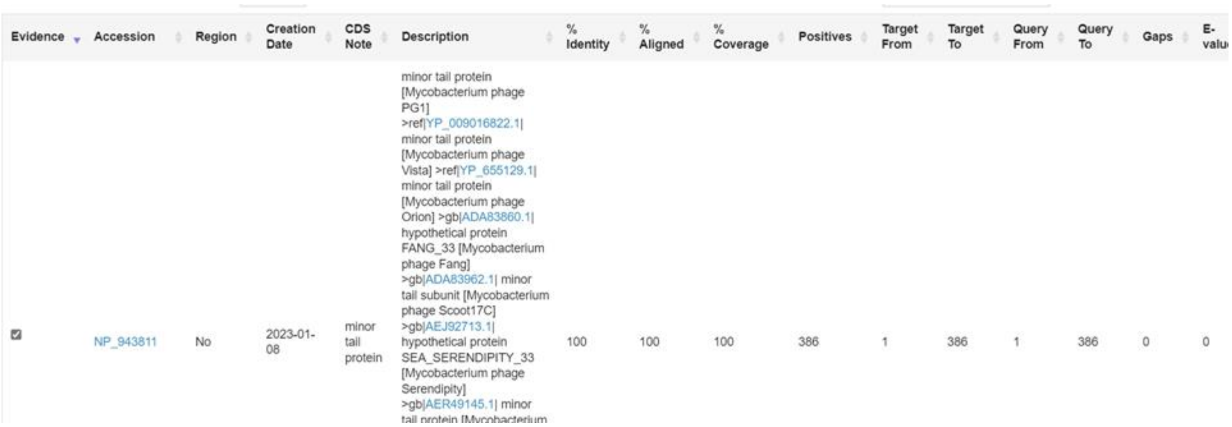
a. SIF: PhageDb BLAST



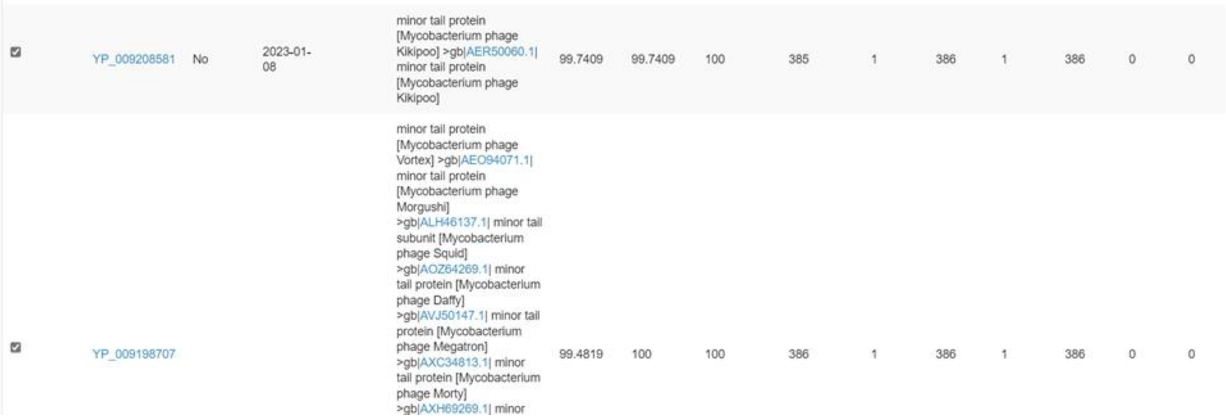


```
                                                                 Score     E
Sequences producing significant alignments:                      (bits)  Value

Zonia_33, minor tail subunit, 386                                    765    0.0
Waterdiva_33, minor tail protein, 386                                765    0.0
Wallhey_32, minor tail protein, 386                                  765    0.0
Vista_33, minor tail protein, 386                                    765    0.0
ThreeOh3D2_33, minor tail protein, 386                               765    0.0
Solosis_32, minor tail protein, 386                                  765    0.0
Serendipity_33, function unknown, 386                                765    0.0
Scoot17C_33, minor tail subunit, 386                                 765    0.0
Pinkman_32, minor tail protein, 386                                  765    0.0
PhatCats2014_33, function unknown, 386                               765    0.0
Phareon_33, minor tail protein, 386                                  765    0.0
PG1_33, function unknown, 386                                        765    0.0
Orion_33, function unknown, 386                                      765    0.0
Longacauda_32, minor tail protein, 386                               765    0.0
Katniss_33, minor tail protein, 386                                  765    0.0
Kahve_32, minor tail protein, 386                                    765    0.0
Jillium_32, minor tail protein, 386                                  765    0.0
HighStump_32, minor tail protein, 386                                765    0.0
HenryJackson_32, minor tail protein, 386                             765    0.0
GeneCoco_33, function unknown, 386                                   765    0.0
Freya_32, minor tail protein, 386                                    765    0.0
Fang_33, function unknown, 386                                       765    0.0
Eugenia_Draft_33, function unknown, 386                              765    0.0
Craff_33, minor tail protein, 386                                    765    0.0
Childish_32, minor tail protein, 386                                 765    0.0
Boehler_33, minor tail protein, 386                                  765    0.0
Adriana_33, minor tail protein, 386                                  765    0.0
Weher20_33, minor tail protein, 386                                  764    0.0
Vortex_33, function unknown, 386                                     764    0.0
Timmi_32, minor tail protein, 386                                    764    0.0
Surely_33, minor tail protein, 386                                   764    0.0
Squid_33, minor tail subunit, 386                                    764    0.0
```

## b. SIF: NCBI BLAST



| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-valu |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | NP_943811 | No | 2023-01-08 | minor tail protein | minor tail protein [Mycobacterium phage PG1] >ref\|YP_009016822.1\| minor tail protein [Mycobacterium phage Vista] >ref\|YP_655129.1\| minor tail protein [Mycobacterium phage Orion] >gb\|ADA83860.1\| hypothetical protein FANG_33 [Mycobacterium phage Fang] >gb\|ADA83962.1\| minor tail subunit [Mycobacterium phage Scoot17C] >gb\|AEJ92713.1\| hypothetical protein SEA_SERENDIPITY_33 [Mycobacterium phage Serendipity] >gb\|AER49145.1\| minor tail protein [Mycobacterium | 100 | 100 | 100 | 386 | 1 | 386 | 1 | 386 | 0 | 0 |

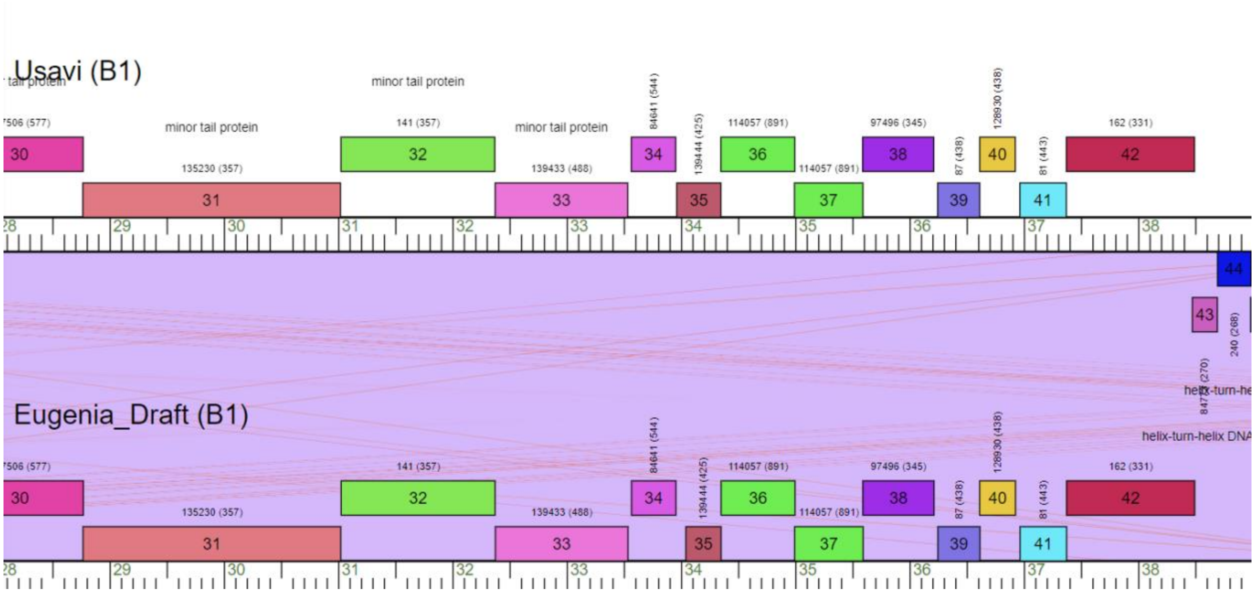


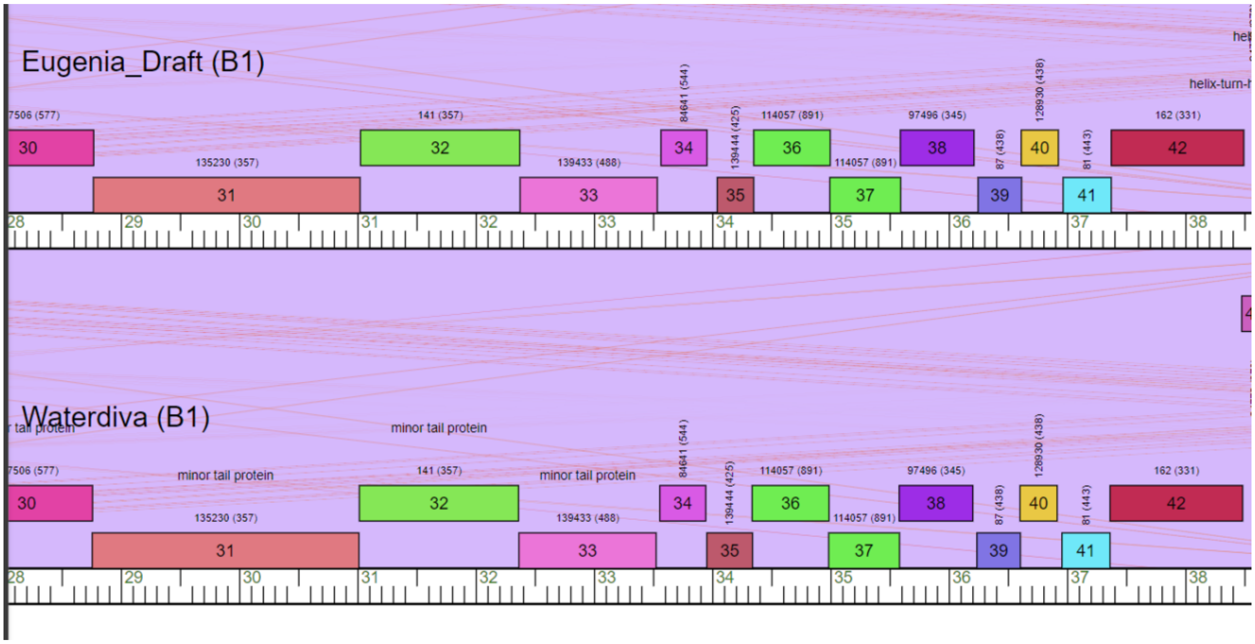
| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | YP_009208581 | No | 2023-01-08 | minor tail protein [Mycobacterium phage Kikipoo] >gb\|AER50060.1\| minor tail protein [Mycobacterium phage Kikipoo] | 99.7409 | 99.7409 | 100 | 385 | 1 | 386 | 1 | 386 | 0 | 0 |
| ☑ | YP_009198707 | | | minor tail protein [Mycobacterium phage Vortex] >gb\|AEO94071.1\| minor tail protein [Mycobacterium phage Morgushi] >gb\|ALH46137.1\| minor tail subunit [Mycobacterium phage Squid] >gb\|AOZ64269.1\| minor tail protein [Mycobacterium phage Daffy] >gb\|AVJ50147.1\| minor tail protein [Mycobacterium phage Megatron] >gb\|AXC34813.1\| minor tail protein [Mycobacterium phage Morty] >gb\|AXH69269.1\| minor | 99.4819 | 100 | 100 | 386 | 1 | 386 | 1 | 386 | 0 | 0 |

## c. SIF: HHPred

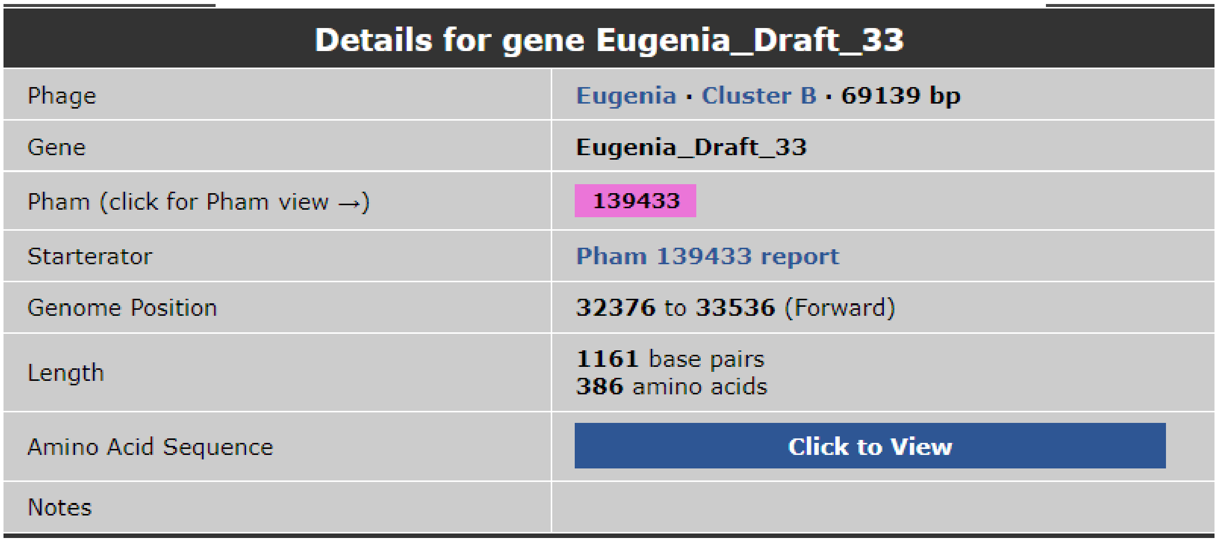
No evidence from HHPred

## d. SIF: Synteny-Phamerator (three genomes)





## 7. Any other important information:



| **Details for gene Eugenia_Draft_33** | |
|---|---|
| Phage | Eugenia · Cluster B · **69139** bp |
| Gene | **Eugenia_Draft_33** |
| Pham (click for Pham view →) | **139433** |
| Starterator | Pham 139433 report |
| Genome Position | **32376** to **33536** (Forward) |
| Length | **1161** base pairs<br>**386** amino acids |
| Amino Acid Sequence | Click to View |
| Notes | |

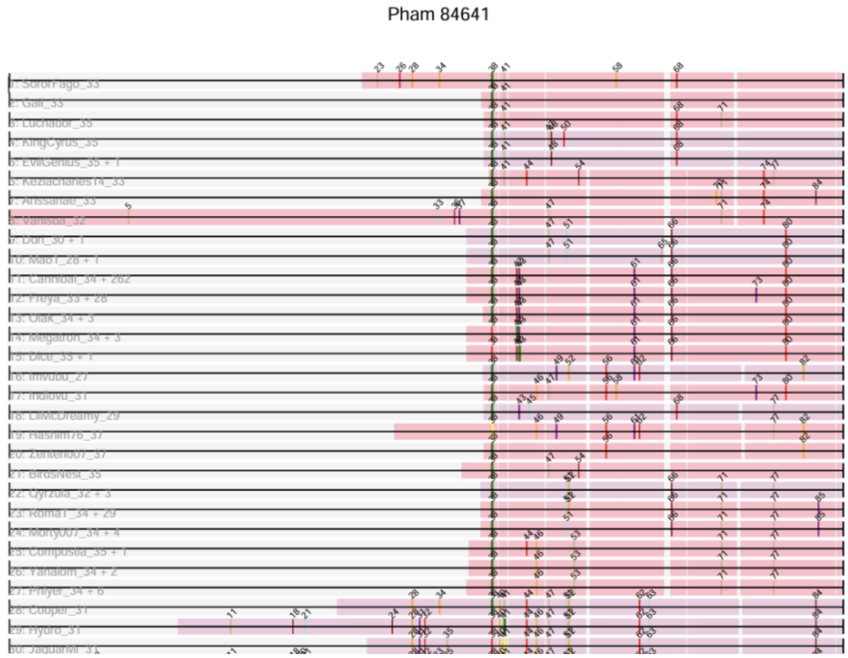
## CURATOR NAME: OLIVIA SIDOTI

## GENE NAME:EUGENIA GENE 34

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

### 1. Starterator



Pham number 84641 has 544 members, 62 are drafts.

The start number called the most often in the published annotations is 38, it was called in 447 of the 482 non-draft genes in the pham.

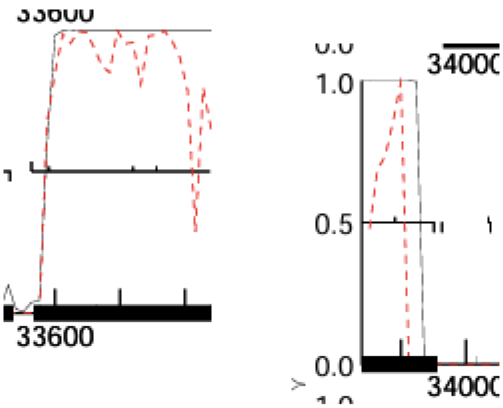
Start 38:

- Found in 512 of 544 ( 94.1% ) of genes in pham

- Manual Annotations of this start: 447 of 482

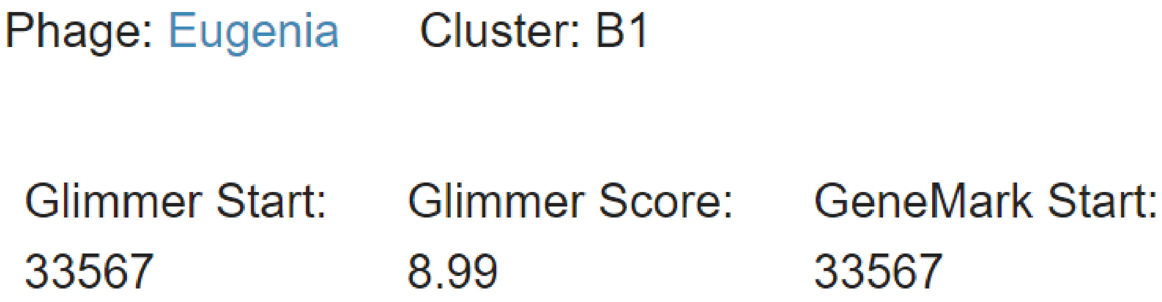
- Called 96.3% of time when present

## 2. GeneMark coding potential



Start: 33,567 Stop:33,956

## 3. Glimmer and GeneMark agreement

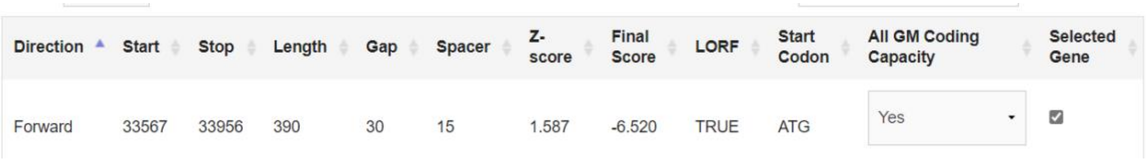


Phage: Eugenia Cluster: B1

| Glimmer Start: | Glimmer Score: | GeneMark Start: |
|---|---|---|
| 33567 | 8.99 | 33567 |

Yes, Glimmer and Genemark are in agreement that the start is at nucleotide 33567.

## 4. Longest open reading frame (ORF) without excessive gap



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 33567 | 33956 | 390 | 30 | 15 | 1.587 | -6.520 | TRUE | ATG | Yes | ☑ |

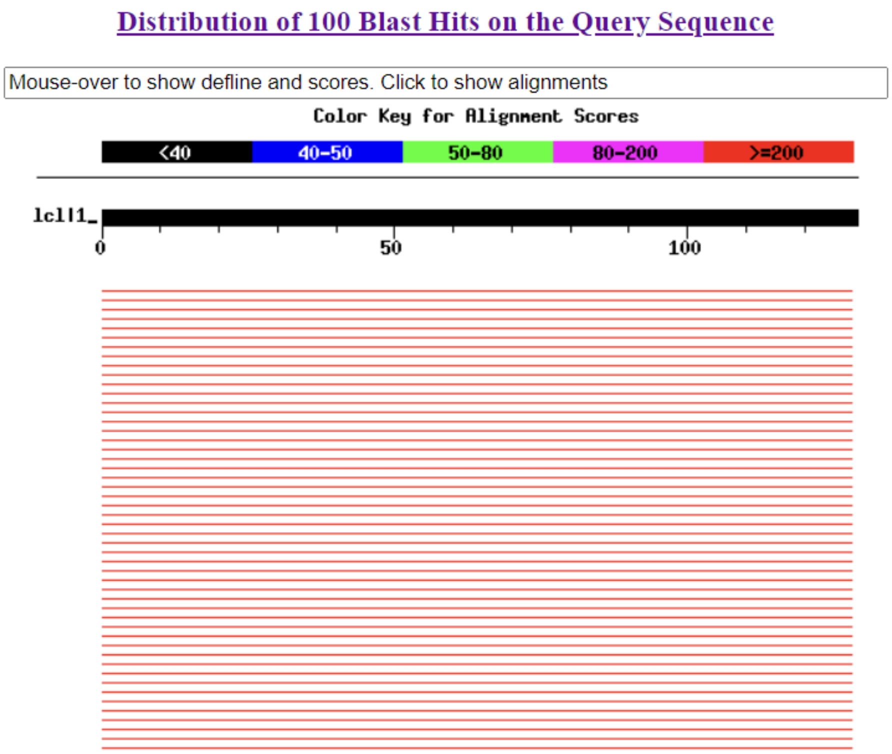
Yes, this gene is the longest open reading frame.

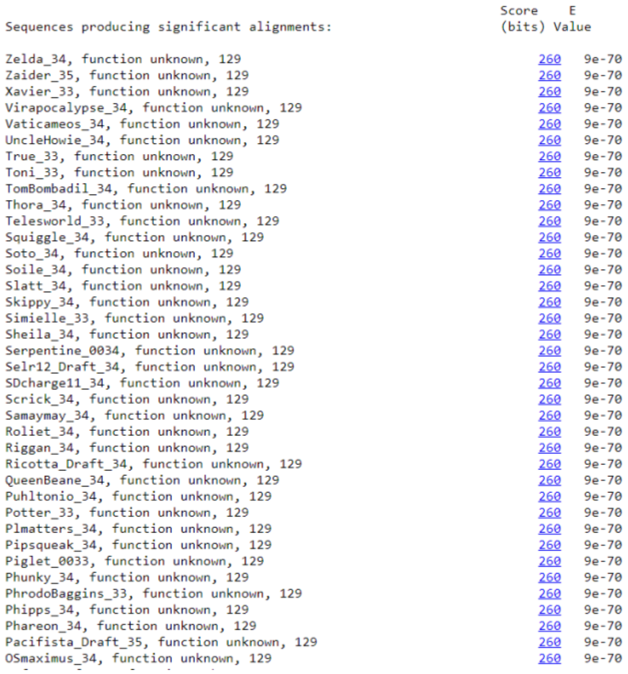
## 5. Function. If no functional prediction is present, write “Hypothetical protein”.

Hypothetical Protein

## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST





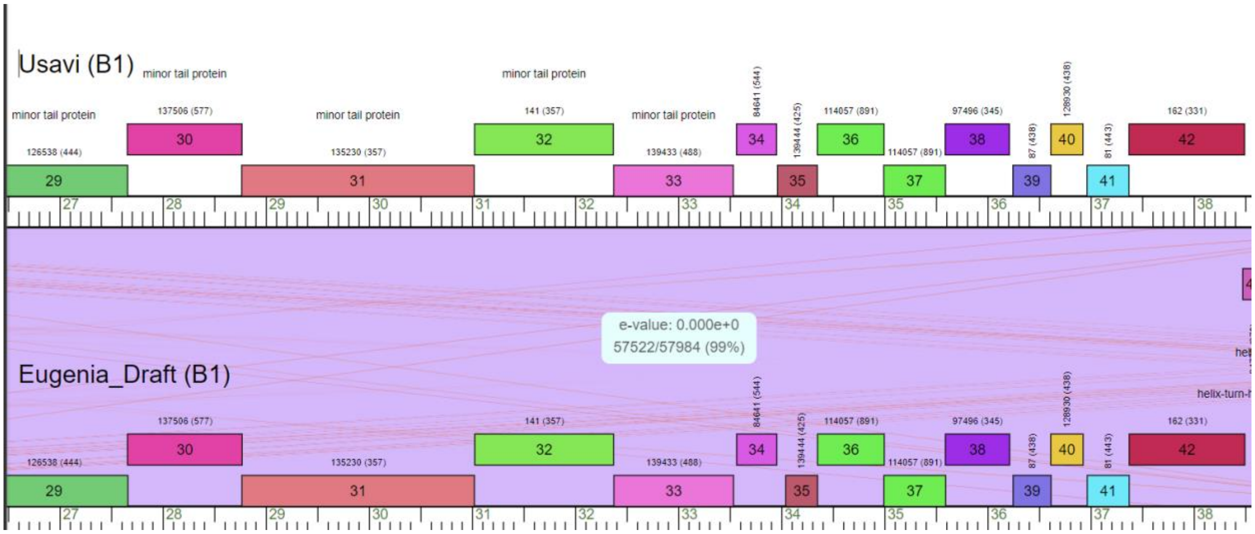
```
                                                                Score    E
Sequences producing significant alignments:                     (bits) Value

Zelda_34, function unknown, 129                                      260   9e-70
Zaider_35, function unknown, 129                                     260   9e-70
Xavier_33, function unknown, 129                                     260   9e-70
Virapocalypse_34, function unknown, 129                              260   9e-70
Vaticameos_34, function unknown, 129                                 260   9e-70
UncleHowie_34, function unknown, 129                                 260   9e-70
True_33, function unknown, 129                                       260   9e-70
Toni_33, function unknown, 129                                       260   9e-70
TomBombadil_34, function unknown, 129                                260   9e-70
Thora_34, function unknown, 129                                      260   9e-70
Telesworld_33, function unknown, 129                                 260   9e-70
Squiggle_34, function unknown, 129                                   260   9e-70
Soto_34, function unknown, 129                                       260   9e-70
Soile_34, function unknown, 129                                      260   9e-70
Slatt_34, function unknown, 129                                      260   9e-70
Skippy_34, function unknown, 129                                     260   9e-70
Simielle_33, function unknown, 129                                   260   9e-70
Sheila_34, function unknown, 129                                     260   9e-70
Serpentine_0034, function unknown, 129                               260   9e-70
Selr12_Draft_34, function unknown, 129                               260   9e-70
SDcharge11_34, function unknown, 129                                 260   9e-70
Scrick_34, function unknown, 129                                     260   9e-70
Samaymay_34, function unknown, 129                                   260   9e-70
Roliet_34, function unknown, 129                                     260   9e-70
Riggan_34, function unknown, 129                                     260   9e-70
Ricotta_Draft_34, function unknown, 129                              260   9e-70
QueenBeane_34, function unknown, 129                                 260   9e-70
Puhltonio_34, function unknown, 129                                  260   9e-70
Potter_33, function unknown, 129                                     260   9e-70
Plmatters_34, function unknown, 129                                  260   9e-70
Pipsqueak_34, function unknown, 129                                  260   9e-70
Piglet_0033, function unknown, 129                                   260   9e-70
Phunky_34, function unknown, 129                                     260   9e-70
PhrodoBaggins_33, function unknown, 129                              260   9e-70
Phipps_34, function unknown, 129                                     260   9e-70
Phareon_34, function unknown, 129                                    260   9e-70
Pacifista_Draft_35, function unknown, 129                            260   9e-70
OSmaximus_34, function unknown, 129                                  260   9e-70
```

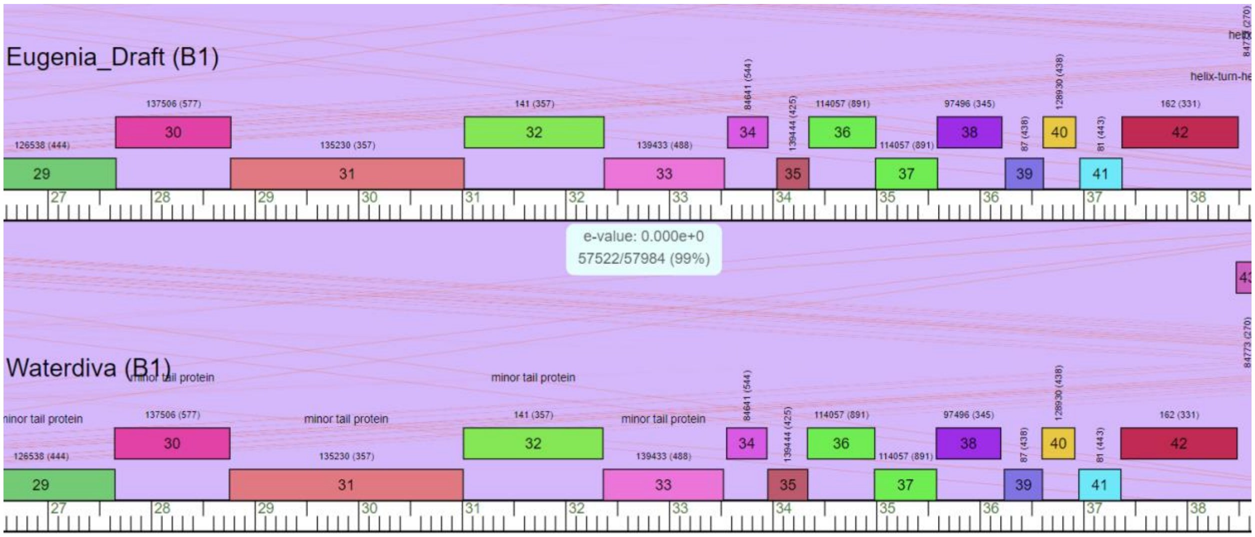
### b. SIF: NCBI BLAST

### c. SIF: HHPred

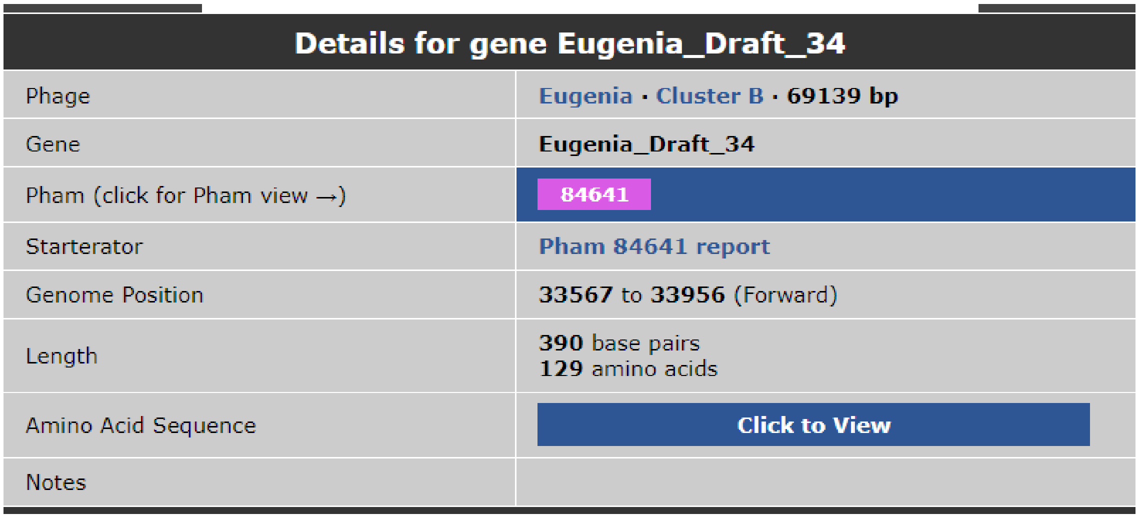
No evidence from HHPred

### d. SIF: Synteny-Phamerator (three genomes)





7. Any other important information.



### Details for gene Eugenia_Draft_34

| | |
|---|---|
| Phage | Eugenia · Cluster B · **69139 bp** |
| Gene | **Eugenia_Draft_34** |
| Pham (click for Pham view →) | 84641 |
| Starterator | Pham 84641 report |
| Genome Position | **33567** to **33956** (Forward) |
| Length | **390** base pairs<br>**129** amino acids |
| Amino Acid Sequence | Click to View |
| Notes | |

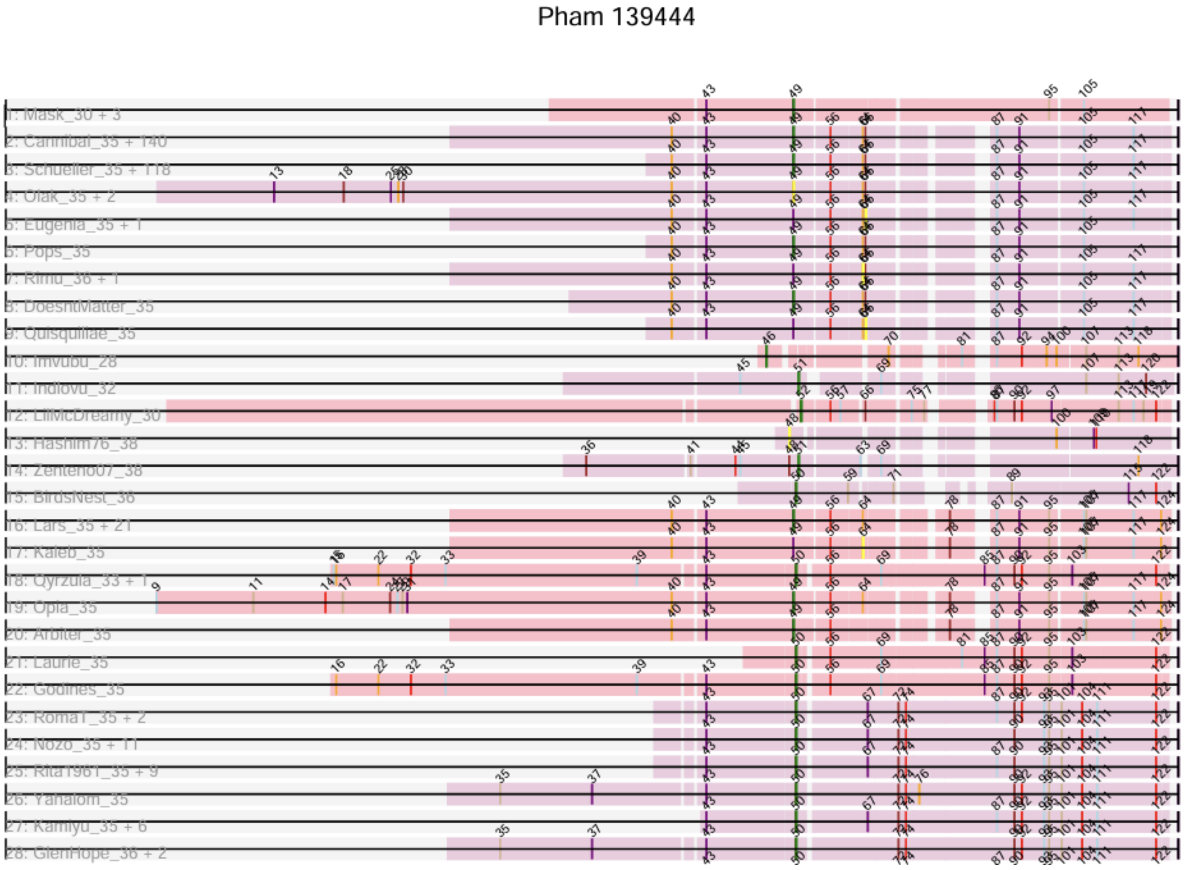
## **CURATOR NAME: OLIVIA SIDOTI**

## **GENE NAME: EUGENIA GENE 35**

## **DNA MASTER NOTES: N/A**

## **START POSITION EVALUATION (IN ORDER):**

### 1. Starterator



Pham number 139444 has 425 members, 46 are drafts.

The start number called the most often in the published annotations is 49, it was called in 281 of the 379 non-draft genes in the pham

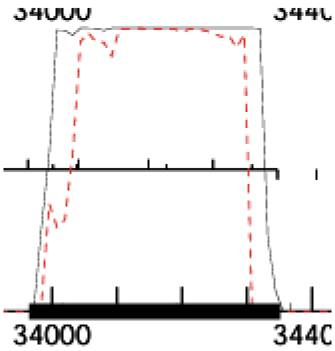
Start 49:

- Found in 307 of 425 ( 72.2% ) of genes in pham

- Manual Annotations of this start: 281 of 379

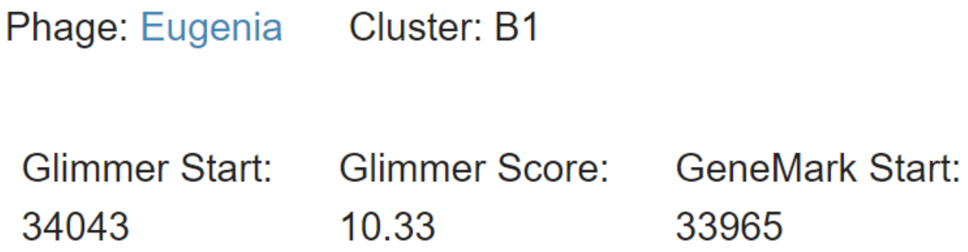
- Called 98.0% of time when present

## 2. GeneMark coding potential



Start:33,965 Stop:34,351

## 3. Glimmer and GeneMark agreement



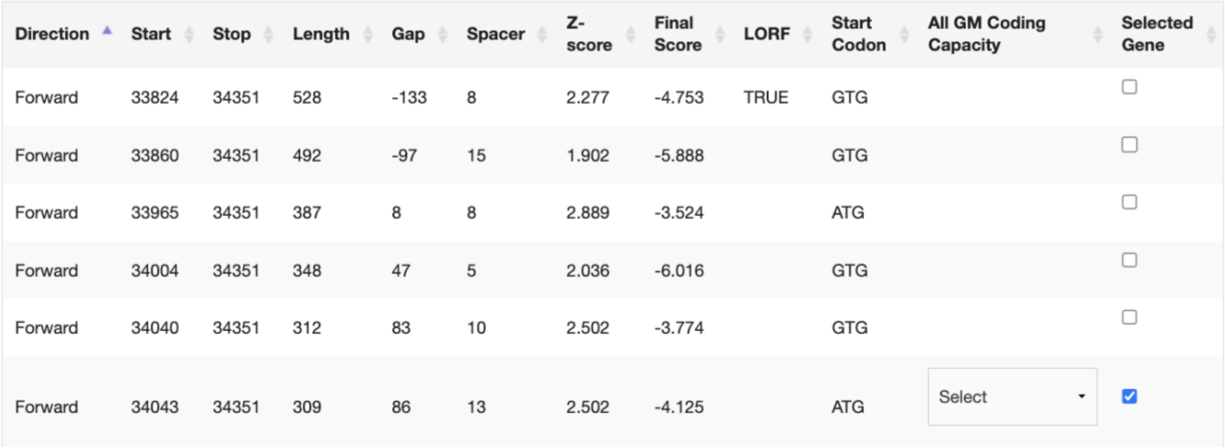
Phage: Eugenia Cluster: B1

| Glimmer Start: | Glimmer Score: | GeneMark Start: |
|---|---|---|
| 34043 | 10.33 | 33965 |

No, Glimmer and Gene Mark are not in agreement about where the start of the gene is.

## 4. Longest open reading frame (ORF) without excessive gap

No. Glimmer and GeneMark have different start site.



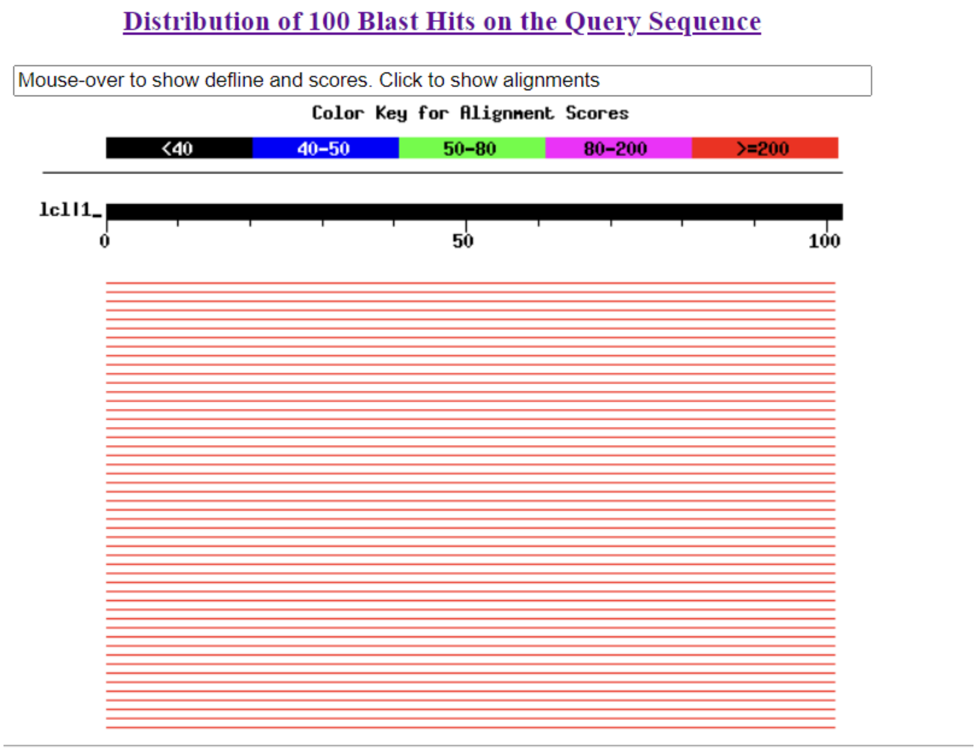
| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 33824 | 34351 | 528 | -133 | 8 | 2.277 | -4.753 | TRUE | GTG |  | ☐ |
| Forward | 33860 | 34351 | 492 | -97 | 15 | 1.902 | -5.888 |  | GTG |  | ☐ |
| Forward | 33965 | 34351 | 387 | 8 | 8 | 2.889 | -3.524 |  | ATG |  | ☐ |
| Forward | 34004 | 34351 | 348 | 47 | 5 | 2.036 | -6.016 |  | GTG |  | ☐ |
| Forward | 34040 | 34351 | 312 | 83 | 10 | 2.502 | -3.774 |  | GTG |  | ☐ |
| Forward | 34043 | 34351 | 309 | 86 | 13 | 2.502 | -4.125 |  | ATG | Select | ☑ |

5. Function. If no functional prediction is present, write “Hypothetical protein”.

Hypothetical Protein

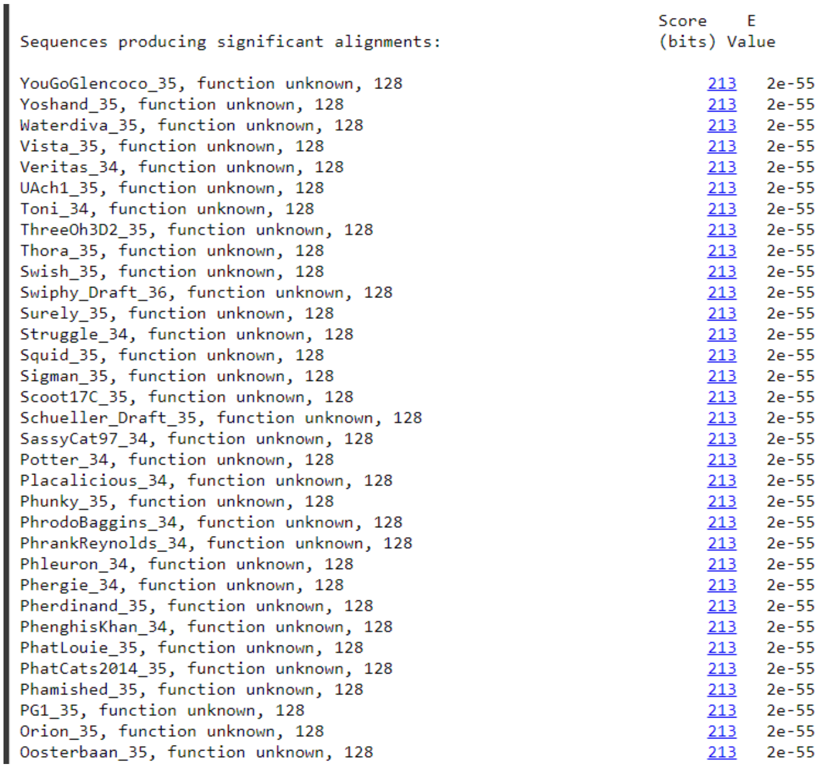
6. Supporting Information for Function (SIF)

a. SIF: PhageDb BLAST



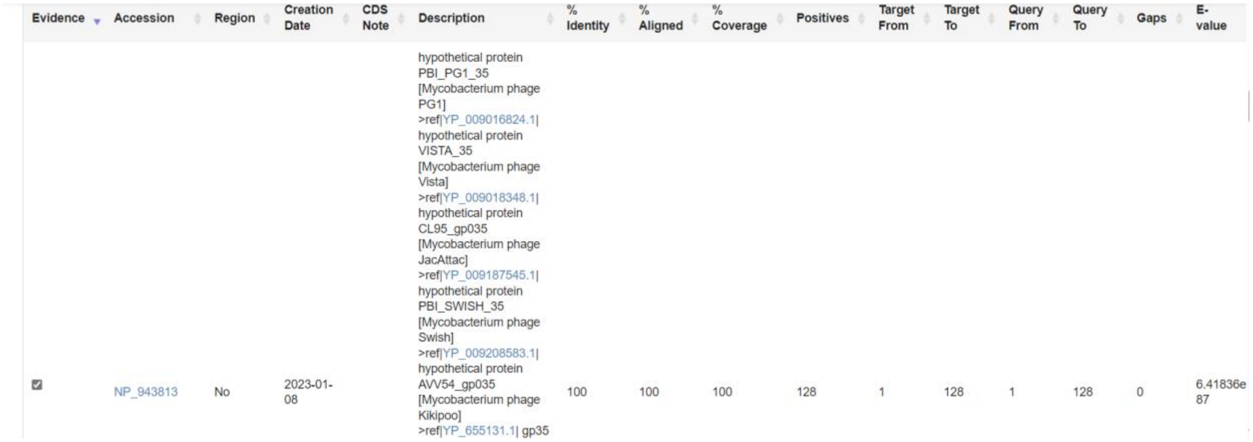
Distribution of 100 Blast Hits on the Query Sequence

Mouse-over to show defline and scores. Click to show alignments



```
                                                                  Score    E
Sequences producing significant alignments:                       (bits) Value

YouGoGlencoco_35, function unknown, 128                              213    2e-55
Yoshand_35, function unknown, 128                                    213    2e-55
Waterdiva_35, function unknown, 128                                  213    2e-55
Vista_35, function unknown, 128                                      213    2e-55
Veritas_34, function unknown, 128                                    213    2e-55
UAch1_35, function unknown, 128                                      213    2e-55
Toni_34, function unknown, 128                                       213    2e-55
ThreeOh3D2_35, function unknown, 128                                 213    2e-55
Thora_35, function unknown, 128                                      213    2e-55
Swish_35, function unknown, 128                                      213    2e-55
Swiphy_Draft_36, function unknown, 128                               213    2e-55
Surely_35, function unknown, 128                                     213    2e-55
Struggle_34, function unknown, 128                                   213    2e-55
Squid_35, function unknown, 128                                      213    2e-55
Sigman_35, function unknown, 128                                     213    2e-55
Scoot17C_35, function unknown, 128                                   213    2e-55
Schueller_Draft_35, function unknown, 128                            213    2e-55
SassyCat97_34, function unknown, 128                                 213    2e-55
Potter_34, function unknown, 128                                     213    2e-55
Placalicious_34, function unknown, 128                               213    2e-55
Phunky_35, function unknown, 128                                     213    2e-55
PhrodoBaggins_34, function unknown, 128                              213    2e-55
PhrankReynolds_34, function unknown, 128                             213    2e-55
Phleuron_34, function unknown, 128                                   213    2e-55
Phergie_34, function unknown, 128                                    213    2e-55
Pherdinand_35, function unknown, 128                                 213    2e-55
PhenghisKhan_34, function unknown, 128                               213    2e-55
PhatLouie_35, function unknown, 128                                  213    2e-55
PhatCats2014_35, function unknown, 128                               213    2e-55
Phamished_35, function unknown, 128                                  213    2e-55
PG1_35, function unknown, 128                                        213    2e-55
Orion_35, function unknown, 128                                      213    2e-55
Oosterbaan_35, function unknown, 128                                 213    2e-55
```

b. SIF: NCBI BLAST



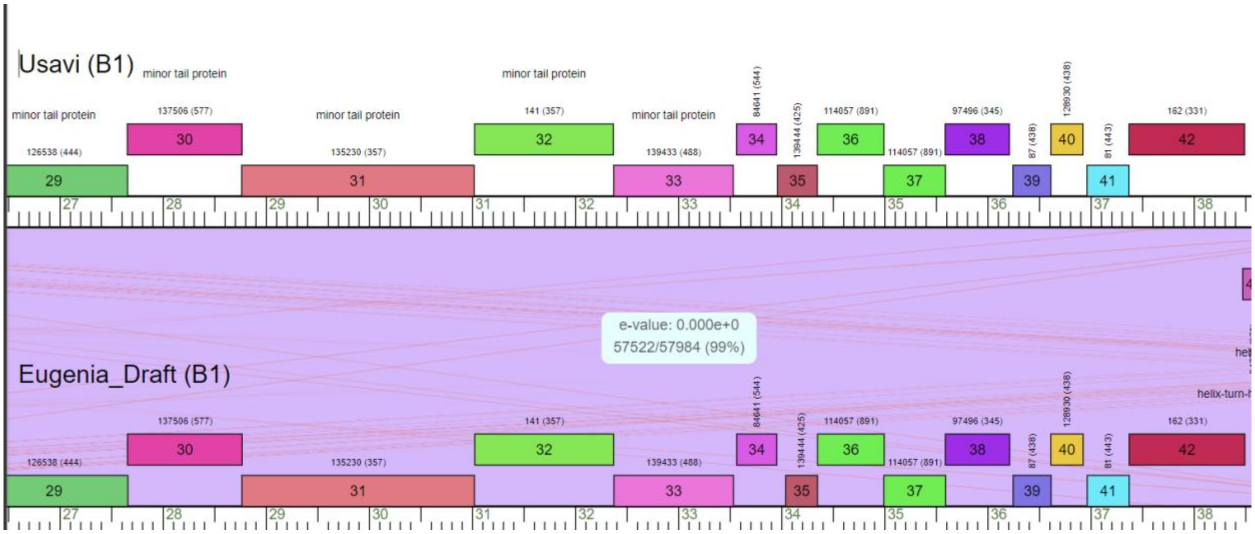
| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | NP_943813 | No | 2023-01-08 |  | hypothetical protein PBI_PG1_35 [Mycobacterium phage PG1] >ref\|YP_009016824.1\| hypothetical protein VISTA_35 [Mycobacterium phage Vista] >ref\|YP_009018348.1\| hypothetical protein CL95_gp035 [Mycobacterium phage JacAttac] >ref\|YP_009187545.1\| hypothetical protein PBI_SWISH_35 [Mycobacterium phage Swish] >ref\|YP_009208583.1\| hypothetical protein AVV54_gp035 [Mycobacterium phage Kikipoo] >ref\|YP_655131.1\| gp35 | 100 | 100 | 100 | 128 | 1 | 128 | 1 | 128 | 0 | 6.41836e 87 |

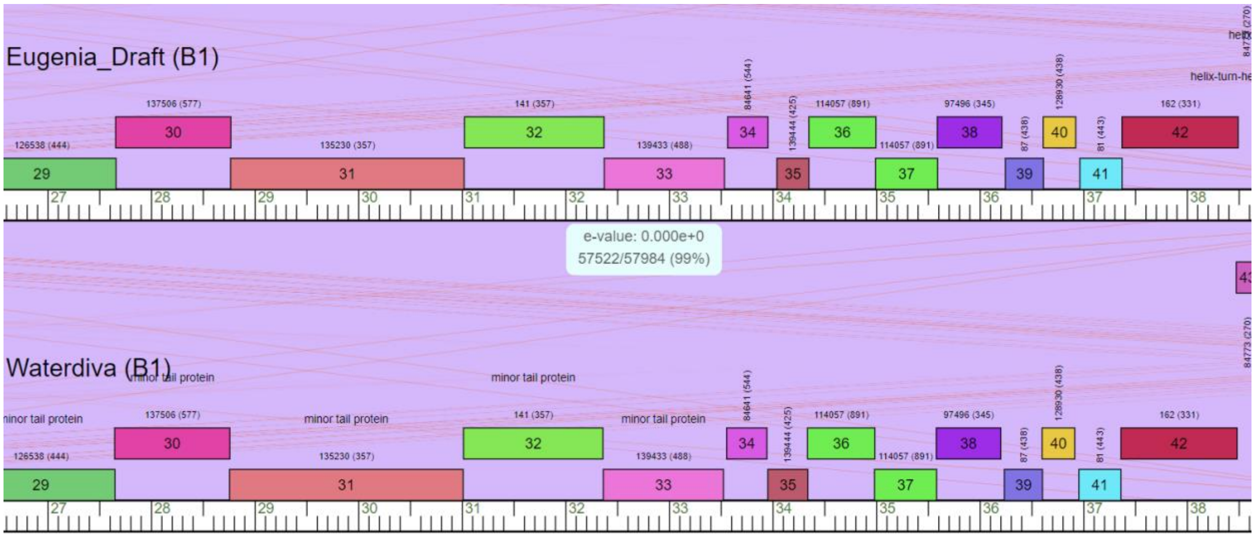
| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | AVJ49401 | No | 2021-12-01 |  | hypothetical protein SEA_CHUNKY_35 [Mycobacterium phage Chunky] >gb\|AZS07137.1\| hypothetical protein SEA_COSMOLLI16_34 [Mycobacterium phage Cosmolli16] >gb\|QWY82883.1\| hypothetical protein SEA_FREYA_34 [Mycobacterium phage Freya] | 99.2188 | 100 | 100 | 128 | 1 | 128 | 1 | 128 | 0 | 1.00666e-86 |

c. SIF: HHPred

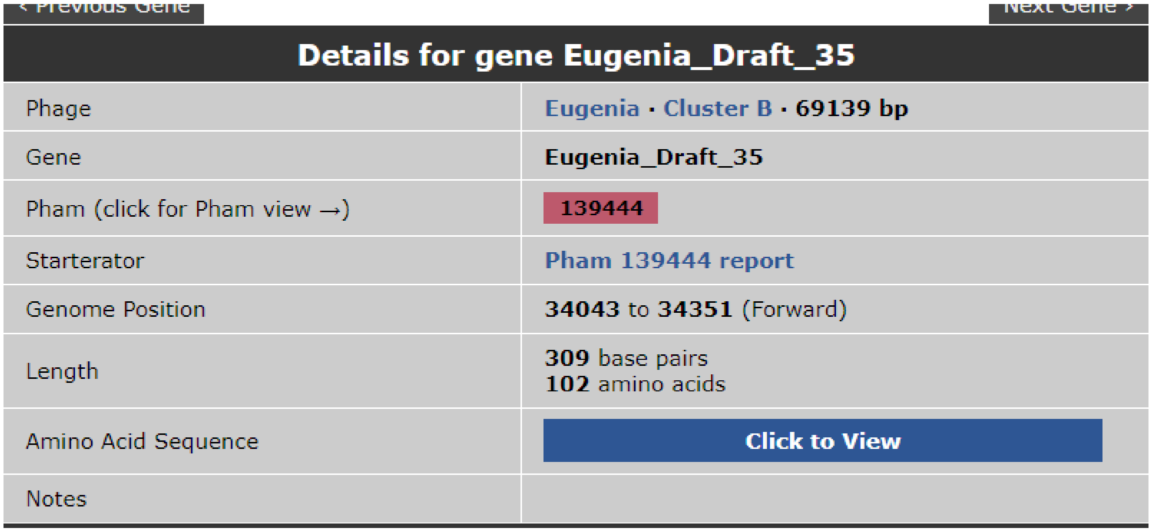
No evidence from HHPred

d. SIF: Synteny-Phamerator (three genomes)





7. Any other important information.



## Details for gene Eugenia_Draft_35

| | |
|---|---|
| Phage | Eugenia · Cluster B · **69139 bp** |
| Gene | **Eugenia_Draft_35** |
| Pham (click for Pham view →) | 139444 |
| Starterator | Pham 139444 report |
| Genome Position | **34043** to **34351** (Forward) |
| Length | **309** base pairs<br>**102** amino acids |
| Amino Acid Sequence | Click to View |
| Notes | |

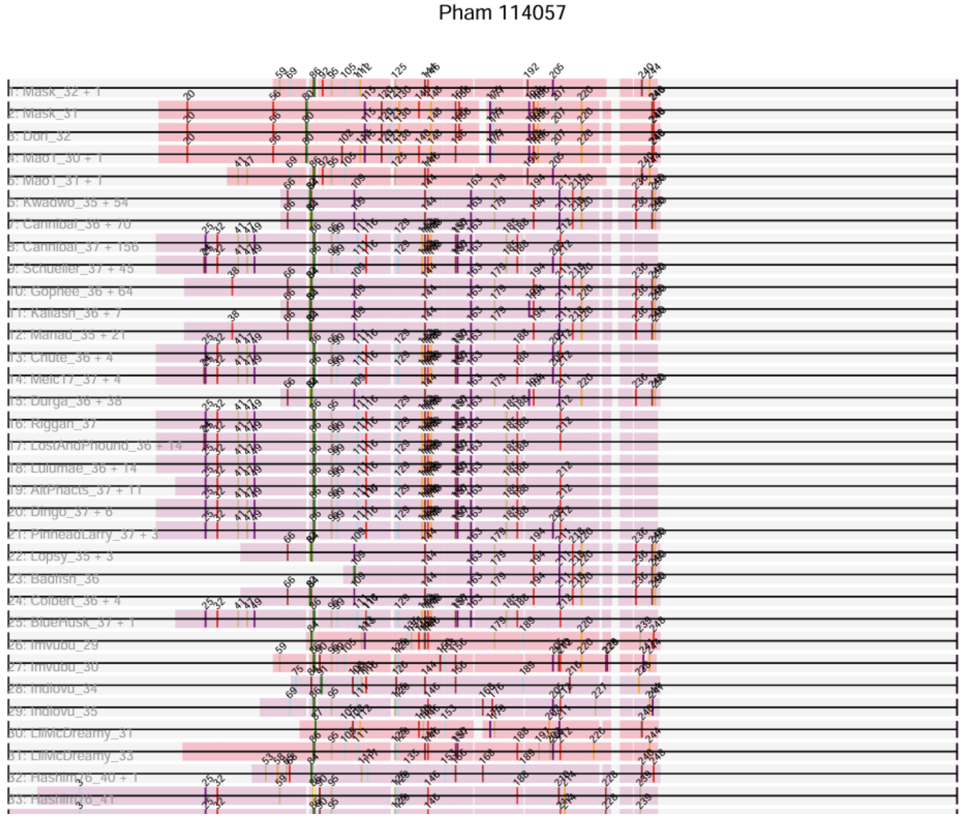
## CURATOR NAME: OLIVIA SIDOTI

## GENE NAME:

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

### 1. Starterator



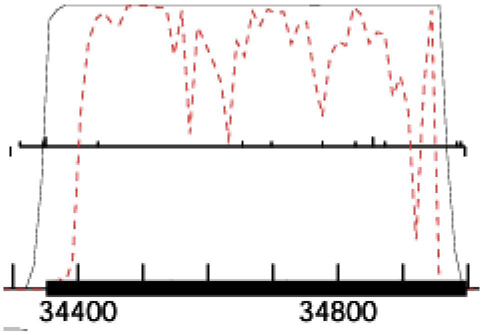
Pham number 114057 has 891 members, 106 are drafts.

The start number called the most often in the published annotations is 86, it was called in 388 of the 785 non-draft genes in the pham.

Start 86:

- Found in 439 of 891 ( 49.3% ) of genes in pham
- Manual Annotations of this start: 388 of 785
- Called 99.8% of time when present

## 2. GeneMark coding potential

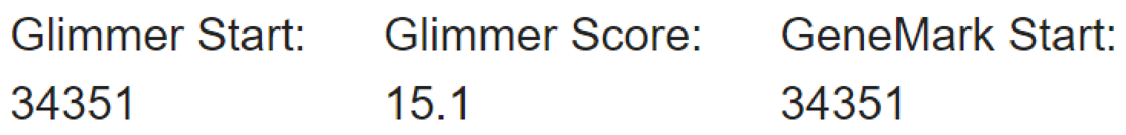


Start: 34,351

Stop: 34,998

## 3. Glimmer and GeneMark agreement

Phage: Eugenia Cluster: B1



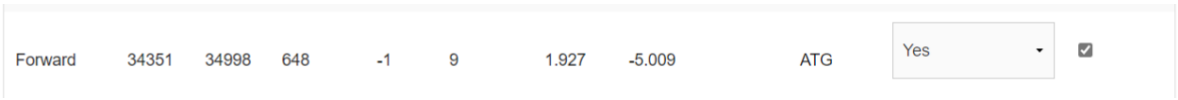
Glimmer Start: 34351

Glimmer Score: 15.1

GeneMark Start: 34351

Yes, Glimmer and Gene Mark are in

4. Longest open reading frame (ORF) without excessive gap



| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 34351 | 34998 | 648 | -1 | 9 | 1.927 | -5.009 | ATG | Yes | ☑ |

No, this gene does not have the longest open reading frame.

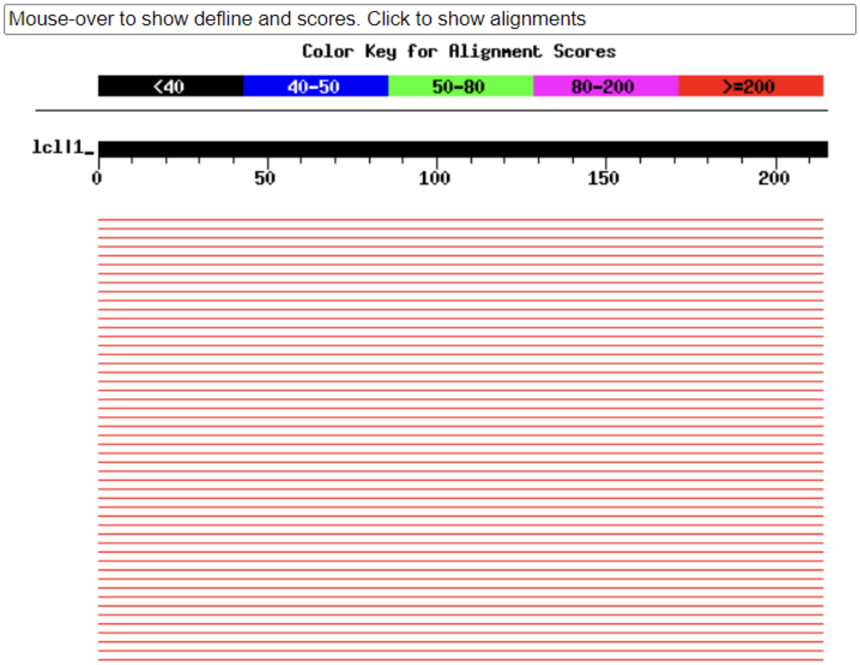
5. Function. If no functional prediction is present, write "Hypothetical protein".

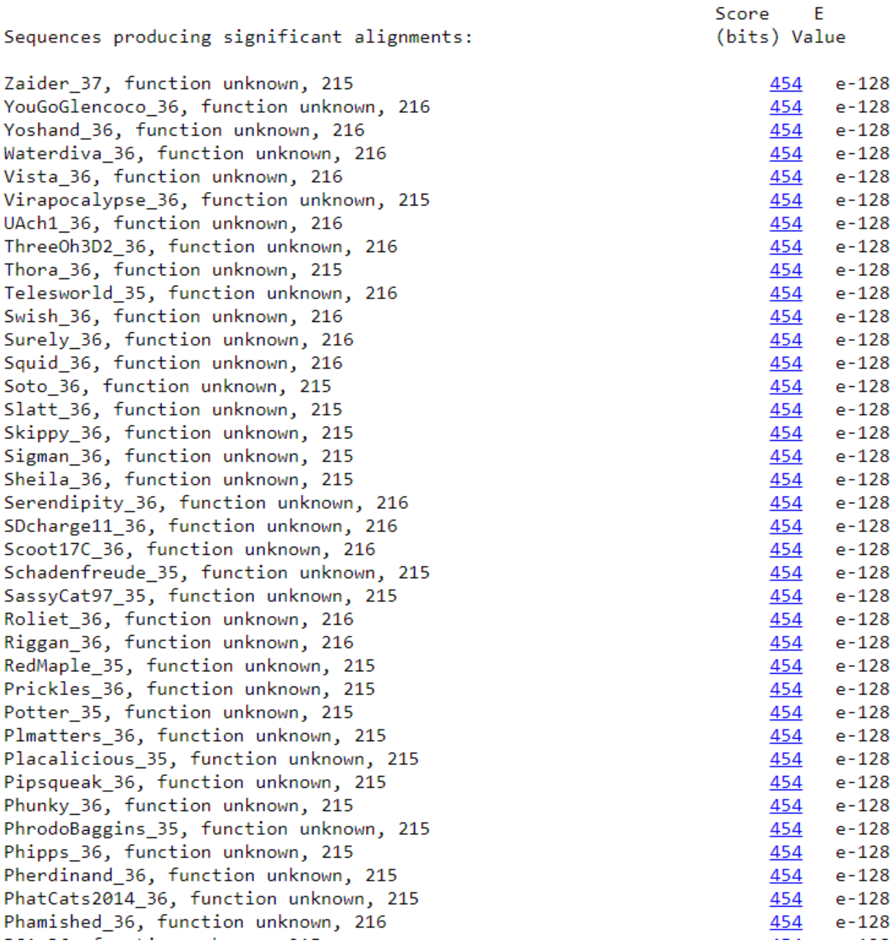
Hypothetical Protein.

6. Supporting Information for Function (SIF)

a. SIF: PhageDb BLAST

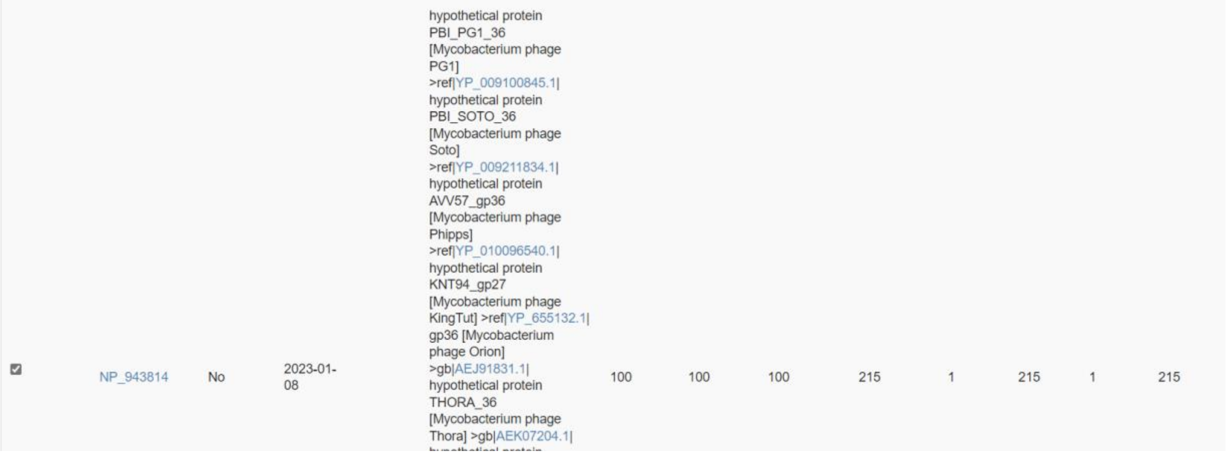
## Distribution of 100 Blast Hits on the Query Sequence



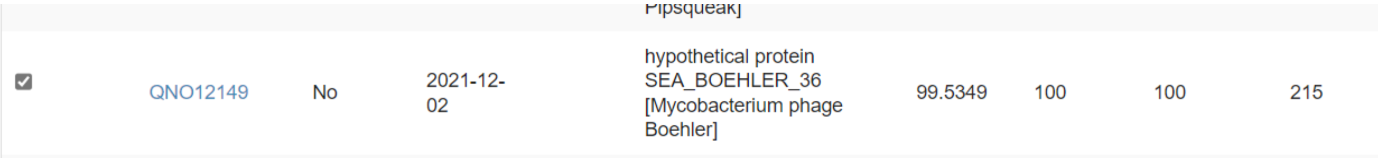


```
                                                                  Score    E
Sequences producing significant alignments:                      (bits) Value

Zaider_37, function unknown, 215                                      454   e-128
YouGoGlencoco_36, function unknown, 216                               454   e-128
Yoshand_36, function unknown, 216                                     454   e-128
Waterdiva_36, function unknown, 216                                   454   e-128
Vista_36, function unknown, 216                                       454   e-128
Virapocalypse_36, function unknown, 215                               454   e-128
UAch1_36, function unknown, 216                                       454   e-128
ThreeOh3D2_36, function unknown, 216                                  454   e-128
Thora_36, function unknown, 215                                       454   e-128
Telesworld_35, function unknown, 216                                  454   e-128
Swish_36, function unknown, 216                                       454   e-128
Surely_36, function unknown, 216                                      454   e-128
Squid_36, function unknown, 216                                       454   e-128
Soto_36, function unknown, 215                                        454   e-128
Slatt_36, function unknown, 215                                       454   e-128
Skippy_36, function unknown, 215                                      454   e-128
Sigman_36, function unknown, 215                                      454   e-128
Sheila_36, function unknown, 215                                      454   e-128
Serendipity_36, function unknown, 216                                 454   e-128
SDcharge11_36, function unknown, 216                                  454   e-128
Scoot17C_36, function unknown, 216                                    454   e-128
Schadenfreude_35, function unknown, 215                               454   e-128
SassyCat97_35, function unknown, 215                                  454   e-128
Roliet_36, function unknown, 216                                      454   e-128
Riggan_36, function unknown, 216                                      454   e-128
RedMaple_35, function unknown, 215                                    454   e-128
Prickles_36, function unknown, 215                                    454   e-128
Potter_35, function unknown, 215                                      454   e-128
Plmatters_36, function unknown, 215                                   454   e-128
Placalicious_35, function unknown, 215                                454   e-128
Pipsqueak_36, function unknown, 215                                   454   e-128
Phunky_36, function unknown, 215                                      454   e-128
PhrodoBaggins_35, function unknown, 215                               454   e-128
Phipps_36, function unknown, 215                                      454   e-128
Pherdinand_36, function unknown, 215                                  454   e-128
PhatCats2014_36, function unknown, 215                                454   e-128
Phamished_36, function unknown, 216                                   454   e-128
```

## b. SIF: NCBI BLAST



| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | NP_943814 | No | 2023-01-08 | hypothetical protein PBI_PG1_36 [Mycobacterium phage PG1] >ref\|YP_009100845.1\| hypothetical protein PBI_SOTO_36 [Mycobacterium phage Soto] >ref\|YP_009211834.1\| hypothetical protein AVV57_gp36 [Mycobacterium phage Phipps] >ref\|YP_010096540.1\| hypothetical protein KNT94_gp27 [Mycobacterium phage KingTut] >ref\|YP_655132.1\| gp36 [Mycobacterium phage Orion] >gb\|AEJ91831.1\| hypothetical protein THORA_36 [Mycobacterium phage Thora] >gb\|AEK07204.1\| | 100 | 100 | 100 | 215 | 1 | 215 | 1 | 215 |

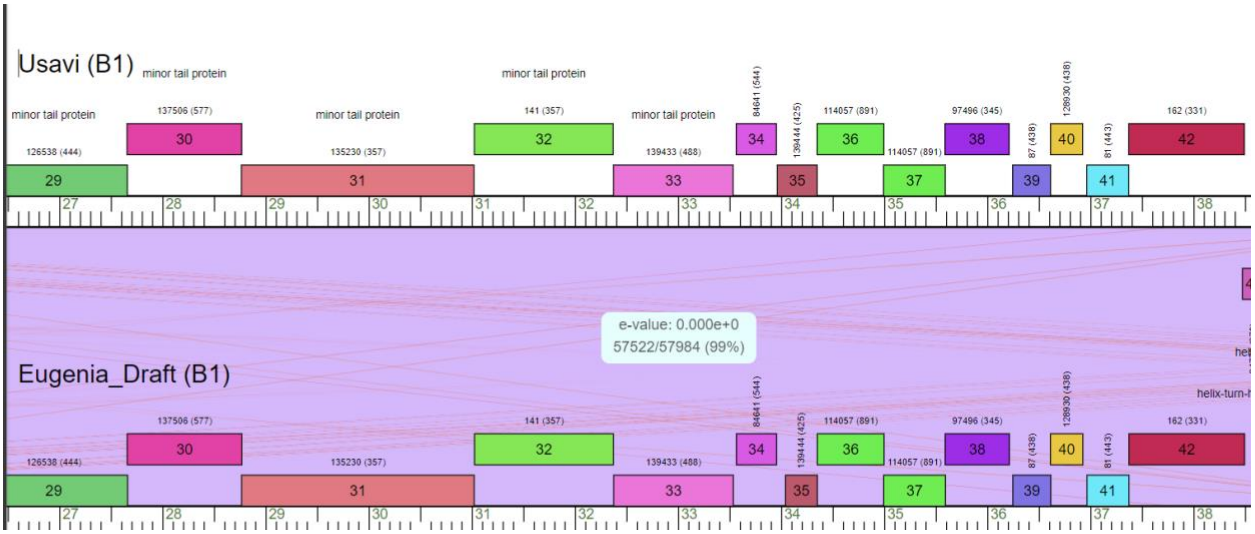


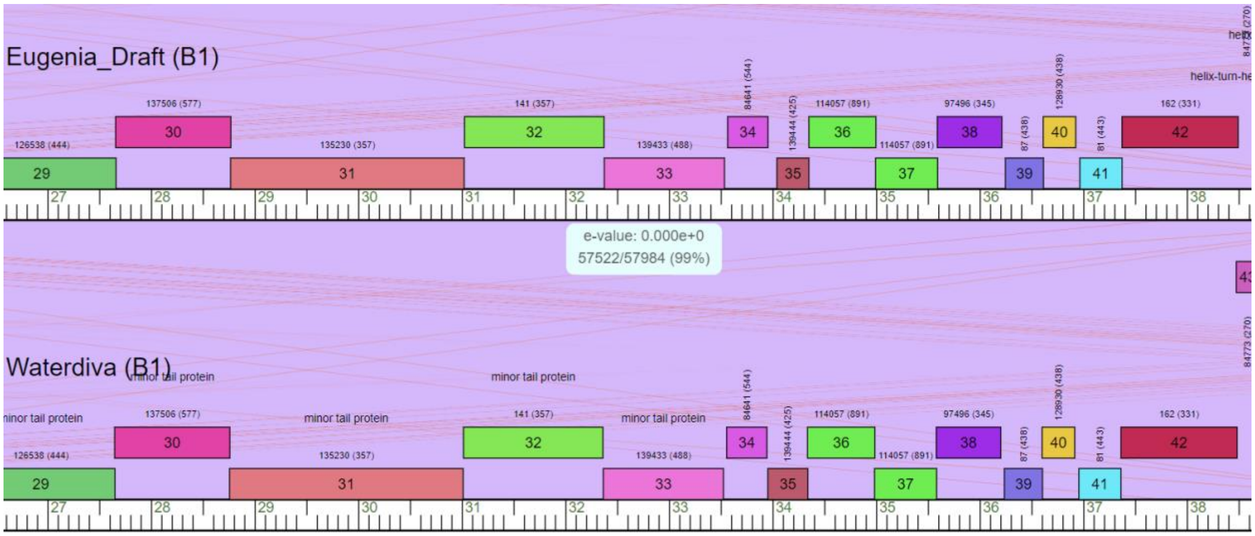
| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | Pipsqueak] | | | | |
| ☑ | QNO12149 | No | 2021-12-02 | hypothetical protein SEA_BOEHLER_36 [Mycobacterium phage Boehler] | 99.5349 | 100 | 100 | 215 |

## c. SIF: HHPred

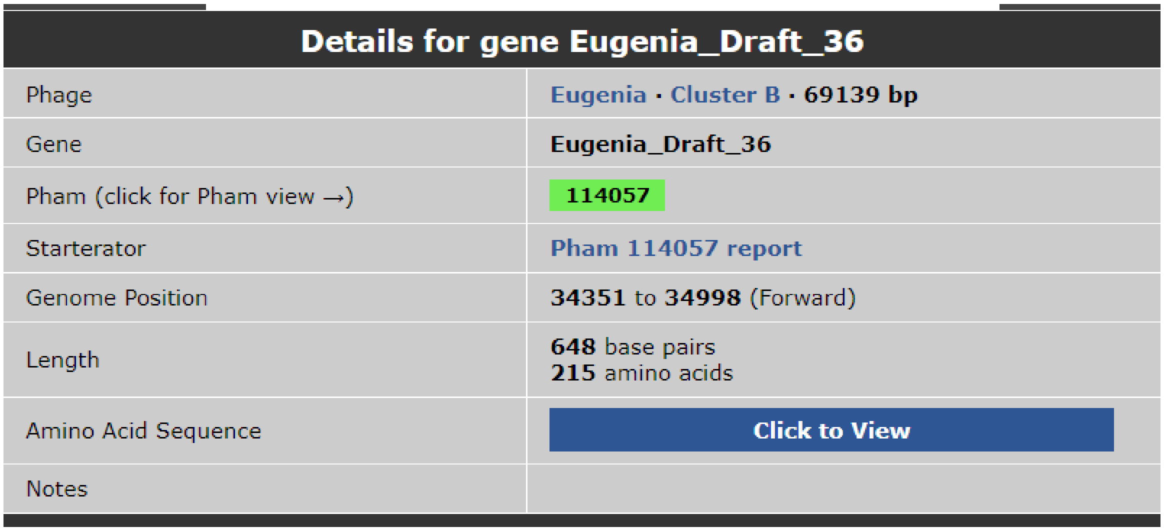
No evidence from HHPred.

## d. SIF: Synteny-Phamerator (three genomes)





7. Any other important information.



**Details for gene Eugenia_Draft_36**

| | |
|---|---|
| Phage | Eugenia · Cluster B · **69139 bp** |
| Gene | **Eugenia_Draft_36** |
| Pham (click for Pham view →) | **114057** |
| Starterator | Pham 114057 report |
| Genome Position | **34351** to **34998** (Forward) |
| Length | **648** base pairs<br>**215** amino acids |
| Amino Acid Sequence | Click to View |
| Notes | |

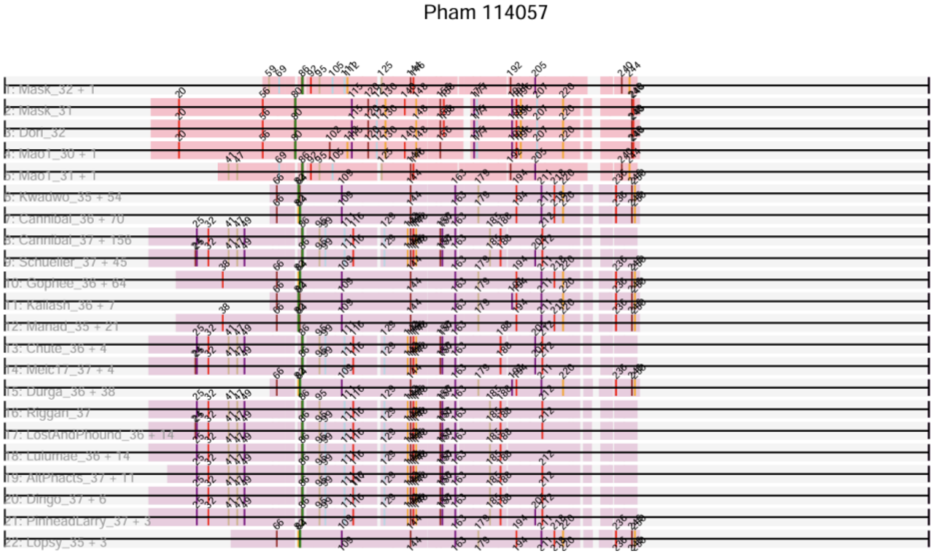
**CURATOR NAME: OLIVIA SIDOTI**

**GENE NAME: EUGENIA GENE 37**

**DNA MASTER NOTES: N/A**

**START POSITION EVALUATION (IN ORDER)**:

1. Starterator



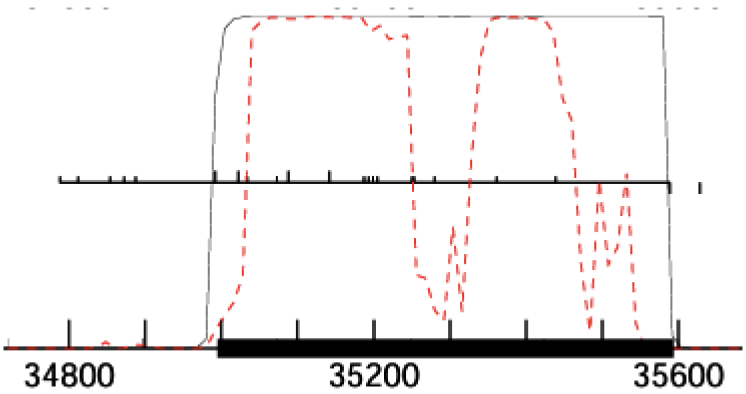
Pham number 114057 has 891 members, 106 are drafts.

The start number called the most often in the published annotations is 86, it was called in 388 of the 785 non-draft genes in the pham.

Start 86:

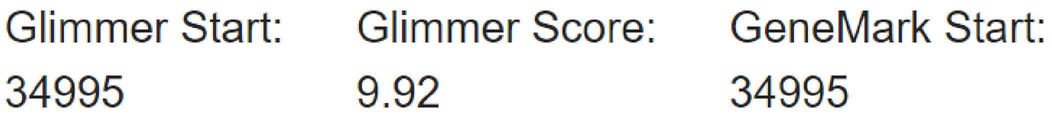
- Found in 439 of 891 ( 49.3% ) of genes in pham
- Manual Annotations of this start: 388 of 785
- Called 99.8% of time when present

## 2. GeneMark coding potential



## 3. Glimmer and GeneMark agreement


Phage: Eugenia Cluster: B1

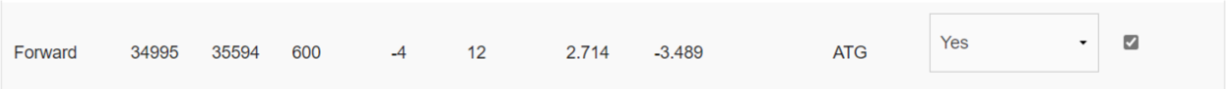

Glimmer Start: 34995

Glimmer Score: 9.92

GeneMark Start: 34995

Yes, Glimmer and GeneMak are in agreement that the start is at position 34995.

## 4. Longest open reading frame (ORF) without excessive gap



| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 34995 | 35594 | 600 | -4 | 12 | 2.714 | -3.489 | ATG | Yes | ☑ |

No, this gene does not have the longest open reading frame.

Gap: -4

Spacer: 12

Z-Score: 2.714

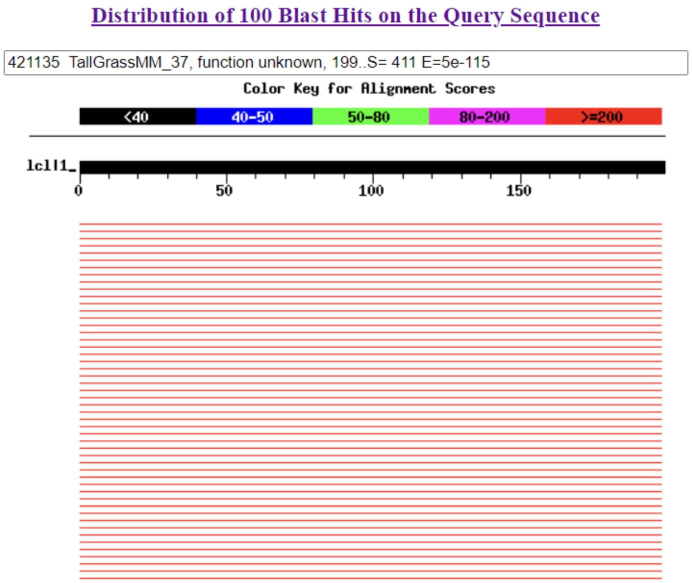
Final Score: -3.488

## 5. Function. If no functional prediction is present, write “Hypothetical protein”.

Hypothetical Protein

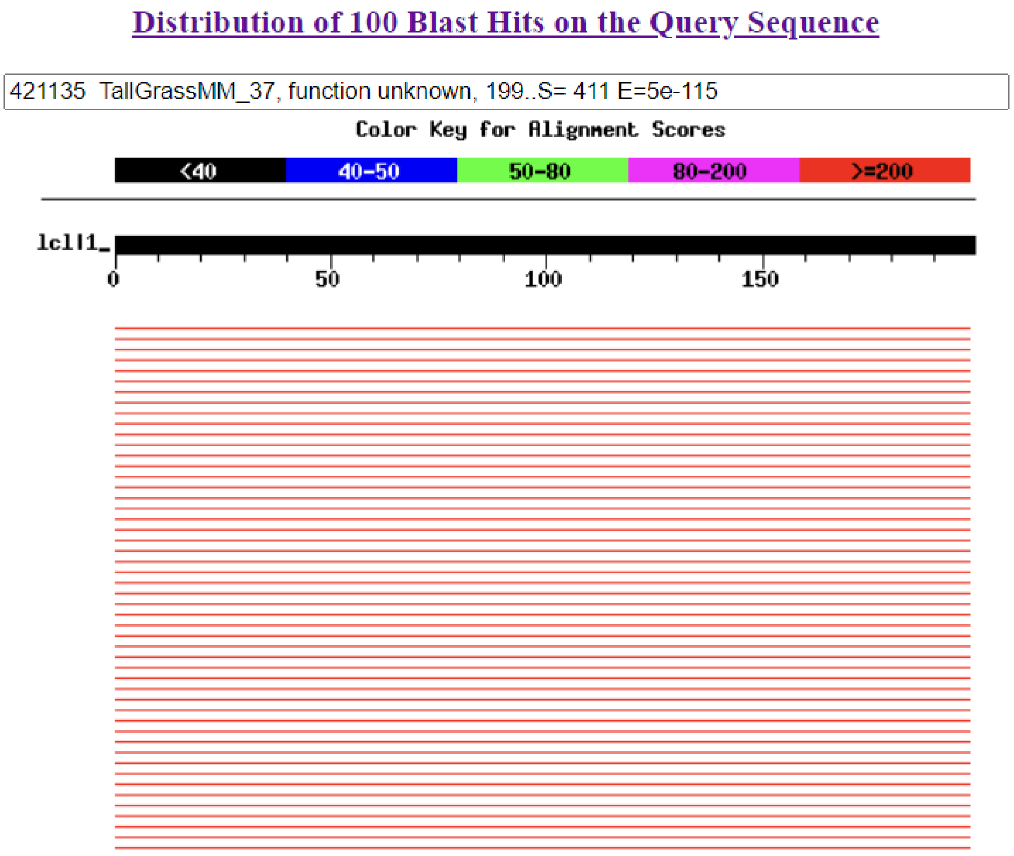
## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST



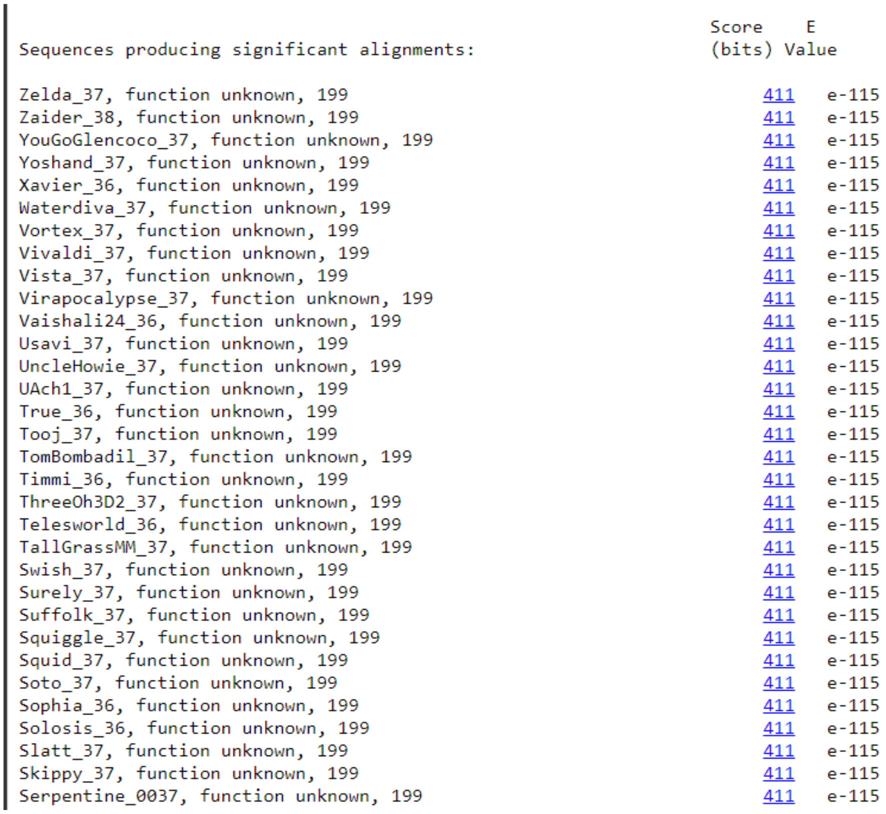
Distribution of 100 Blast Hits on the Query Sequence

421135 TallGrassMM_37, function unknown, 199..S= 411 E=5e-115



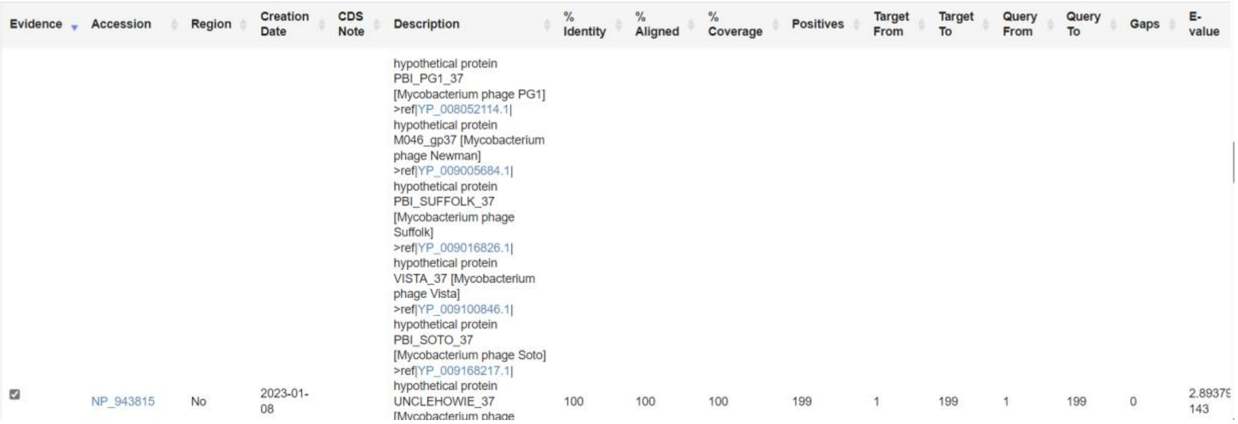
## Distribution of 100 Blast Hits on the Query Sequence

421135 TallGrassMM_37, function unknown, 199..S= 411 E=5e-115



```
                                                                  Score    E
Sequences producing significant alignments:                       (bits) Value

Zelda_37, function unknown, 199                                      411   e-115
Zaider_38, function unknown, 199                                     411   e-115
YouGoGlencoco_37, function unknown, 199                              411   e-115
Yoshand_37, function unknown, 199                                    411   e-115
Xavier_36, function unknown, 199                                     411   e-115
Waterdiva_37, function unknown, 199                                  411   e-115
Vortex_37, function unknown, 199                                     411   e-115
Vivaldi_37, function unknown, 199                                    411   e-115
Vista_37, function unknown, 199                                      411   e-115
Virapocalypse_37, function unknown, 199                              411   e-115
Vaishali24_36, function unknown, 199                                 411   e-115
Usavi_37, function unknown, 199                                      411   e-115
UncleHowie_37, function unknown, 199                                 411   e-115
UAch1_37, function unknown, 199                                      411   e-115
True_36, function unknown, 199                                       411   e-115
Tooj_37, function unknown, 199                                       411   e-115
TomBombadil_37, function unknown, 199                                411   e-115
Timmi_36, function unknown, 199                                      411   e-115
ThreeOh3D2_37, function unknown, 199                                 411   e-115
Telesworld_36, function unknown, 199                                 411   e-115
TallGrassMM_37, function unknown, 199                                411   e-115
Swish_37, function unknown, 199                                      411   e-115
Surely_37, function unknown, 199                                     411   e-115
Suffolk_37, function unknown, 199                                    411   e-115
Squiggle_37, function unknown, 199                                   411   e-115
Squid_37, function unknown, 199                                      411   e-115
Soto_37, function unknown, 199                                       411   e-115
Sophia_36, function unknown, 199                                     411   e-115
Solosis_36, function unknown, 199                                    411   e-115
Slatt_37, function unknown, 199                                      411   e-115
Skippy_37, function unknown, 199                                     411   e-115
Serpentine_0037, function unknown, 199                               411   e-115
```

## b. SIF: NCBI BLAST



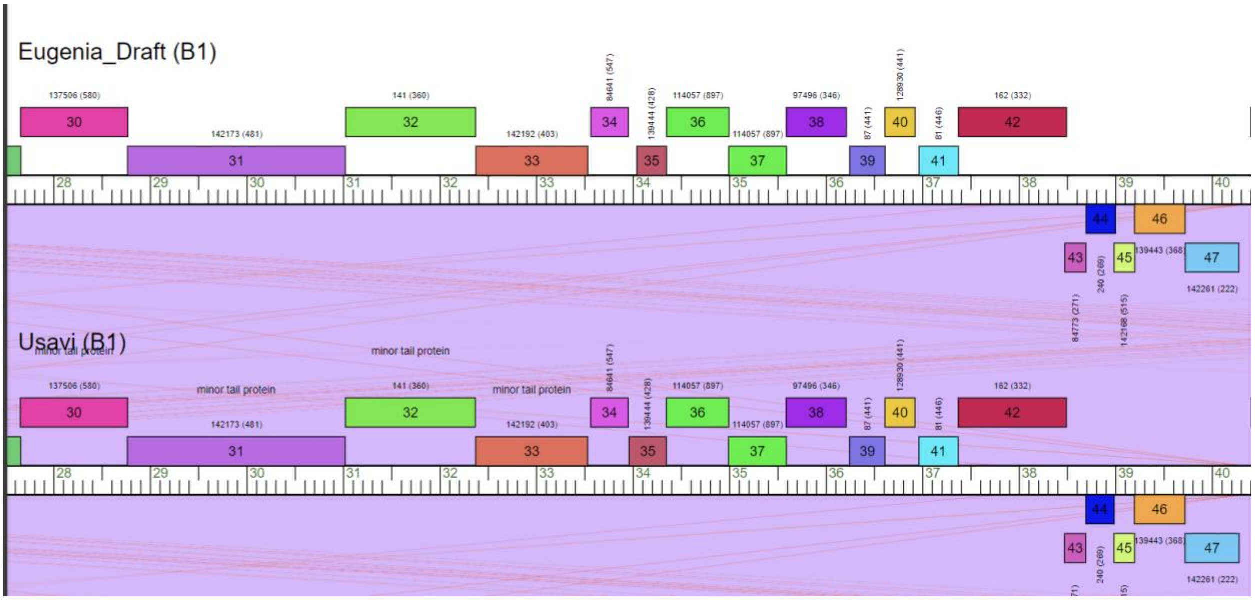
| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | NP_943815 | No | 2023-01-08 |  | hypothetical protein PBI_PG1_37 [Mycobacterium phage PG1] >ref\|YP_008052114.1\| hypothetical protein M046_gp37 [Mycobacterium phage Newman] >ref\|YP_009005684.1\| hypothetical protein PBI_SUFFOLK_37 [Mycobacterium phage Suffolk] >ref\|YP_009016826.1\| hypothetical protein VISTA_37 [Mycobacterium phage Vista] >ref\|YP_009100846.1\| hypothetical protein PBI_SOTO_37 [Mycobacterium phage Soto] >ref\|YP_009168217.1\| hypothetical protein UNCLEHOWIE_37 [Mycobacterium phage | 100 | 100 | 100 | 199 | 1 | 199 | 1 | 199 | 0 | 2.8937 143 |

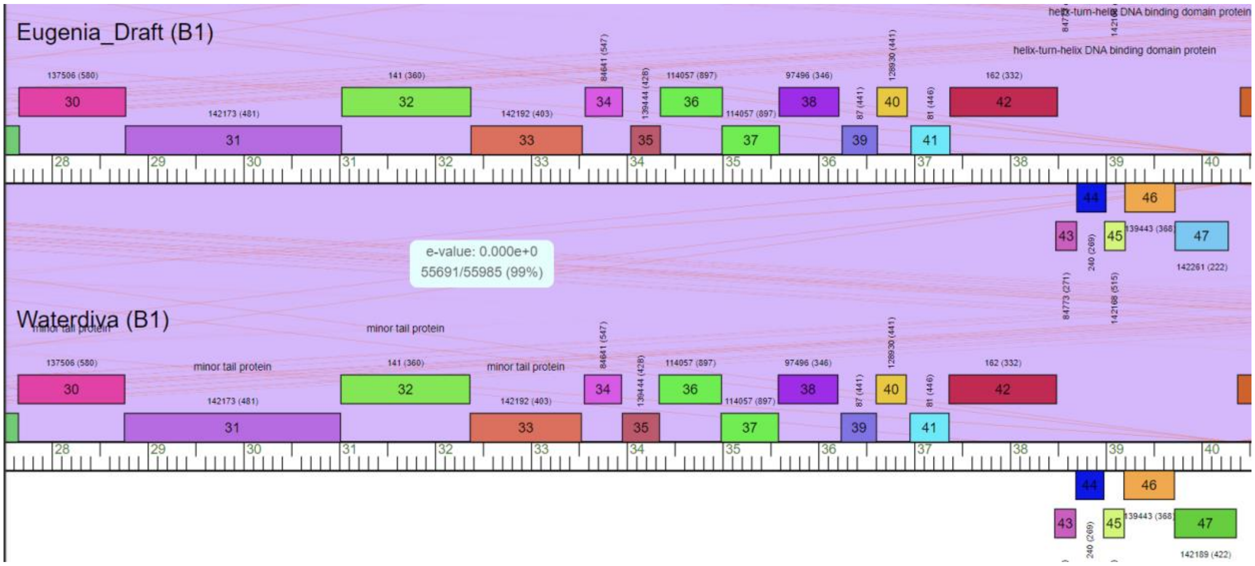
| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | YP_009043311 | No | 2023-01-08 | hypothetical protein HL05_gp036 [Mycobacterium phage Manad] >gb\|AHZ95296.1\| hypothetical protein PBI_MANAD_36 [Mycobacterium phage Manad] >gb\|AOQ28394.1\| hypothetical protein SEA_FRIARPREACHER_37 [Mycobacterium phage FriarPreacher] | 99.4975 | 100 | 100 | 199 | 1 | 199 | 1 | 199 | 0 | 7.27434<br>143 |
| ☑ | AZF96896 | No | 2021-12-01 | hypothetical protein SEA_JILLIUM_36 [Mycobacterium phage Jillium] | 99.4975 | 100 | 100 | 199 | 1 | 199 | 1 | 199 | 0 | 7.6007€<br>143 |

## c. SIF: HHPred

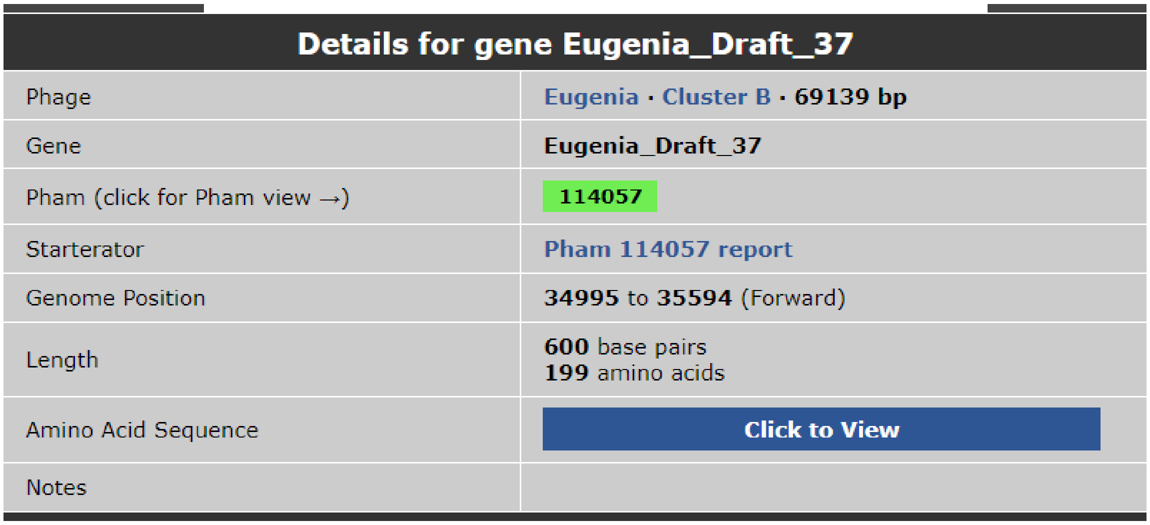
No Evidence from HHPred

## d. SIF: Synteny-Phamerator (three genomes)





## 7. Any other important information.



**Details for gene Eugenia_Draft_37**

| | |
|---|---|
| Phage | Eugenia · Cluster B · 69139 bp |
| Gene | Eugenia_Draft_37 |
| Pham (click for Pham view →) | 114057 |
| Starterator | Pham 114057 report |
| Genome Position | 34995 to 35594 (Forward) |
| Length | 600 base pairs<br>199 amino acids |
| Amino Acid Sequence | Click to View |
| Notes | |

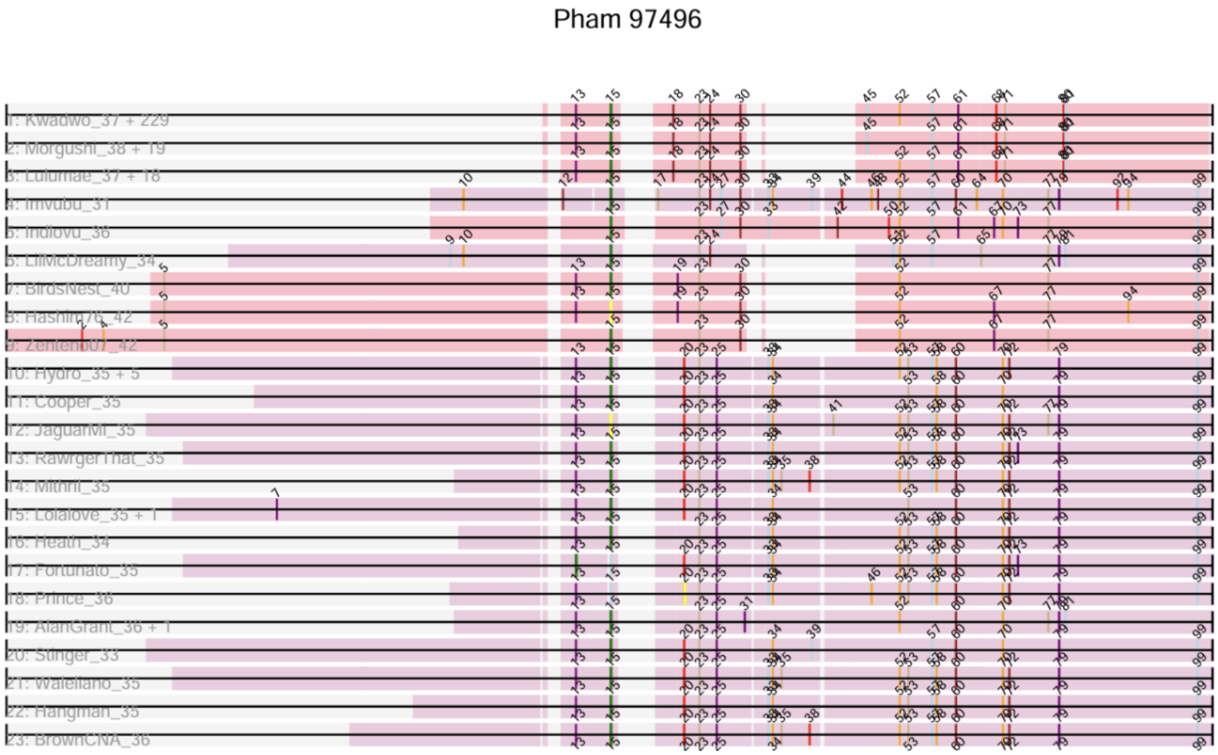
## **CURATOR NAME: OLIVIA SIDOTI**

## **GENE NAME: EUGENIA GENE 38**

## **DNA MASTER NOTES: N/A**

## **START POSITION EVALUATION (IN ORDER):**

### 1. Starterator



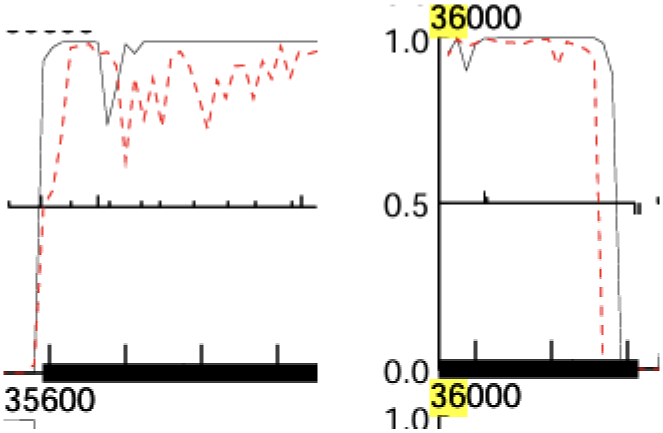
Pham number 97496 has 345 members, 43 are drafts.

The start number called the most often in the published annotations is 15, it was called in 299 of the 302 non-draft genes in the pham.

Start 15:

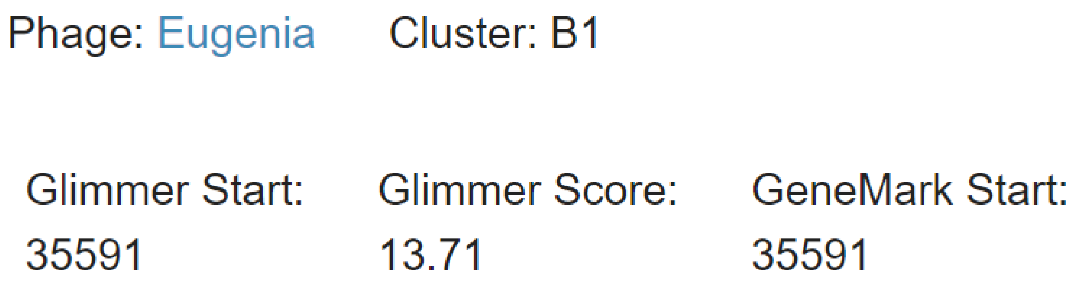
- Found in 343 of 345 ( 99.4% ) of genes in pham
- Manual Annotations of this start: 299 of 302
- Called 97.7% of time when present

## 2. GeneMark coding potential



Start: 35,591 Stop: 36,214

## 3. Glimmer and GeneMark agreement


Phage: Eugenia Cluster: B1

| Glimmer Start: | Glimmer Score: | GeneMark Start: |
|---|---|---|
| 35591 | 13.71 | 35591 |

## 4. Longest open reading frame (ORF) without excessive gap

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 35591 | 36214 | 624 | -4 | 12 | 2.203 | -4.516 | ATG | Yes | ☑ |

No, this gene is not the longest open reading frame.

Gap: -4

Spacer: 12

Z-Score: 2.203

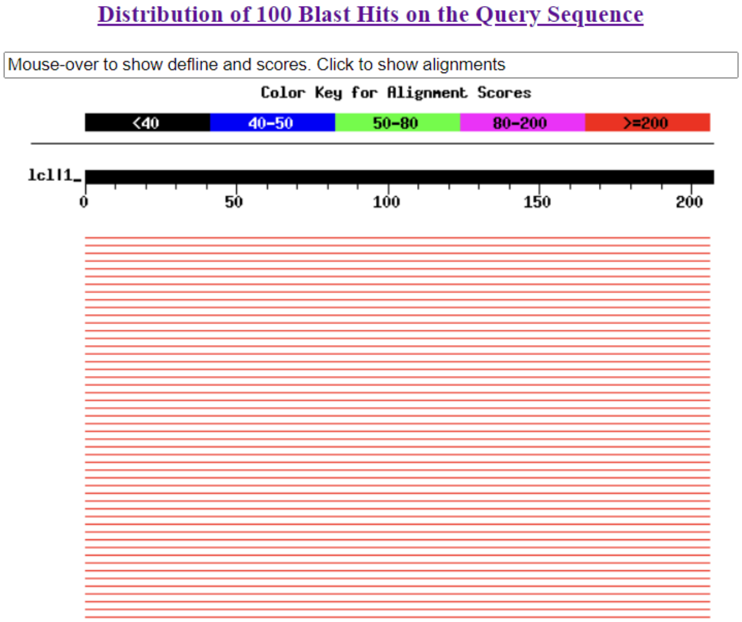
Final Score: -4.516

## 5. Function. If no functional prediction is present, write “Hypothetical protein”.

Hypothetical Protein

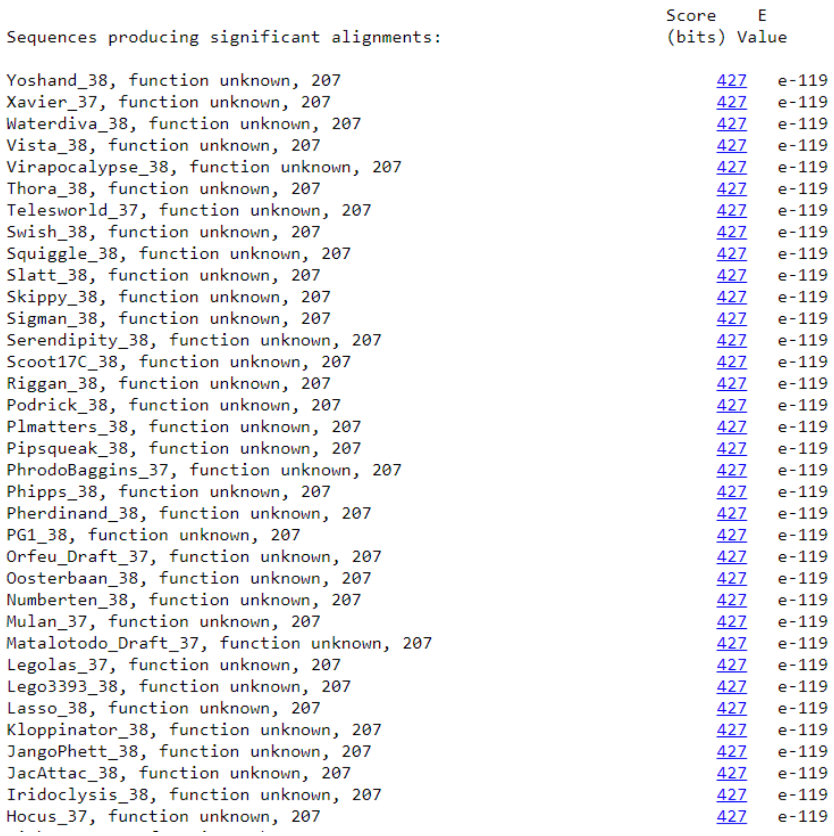
## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST



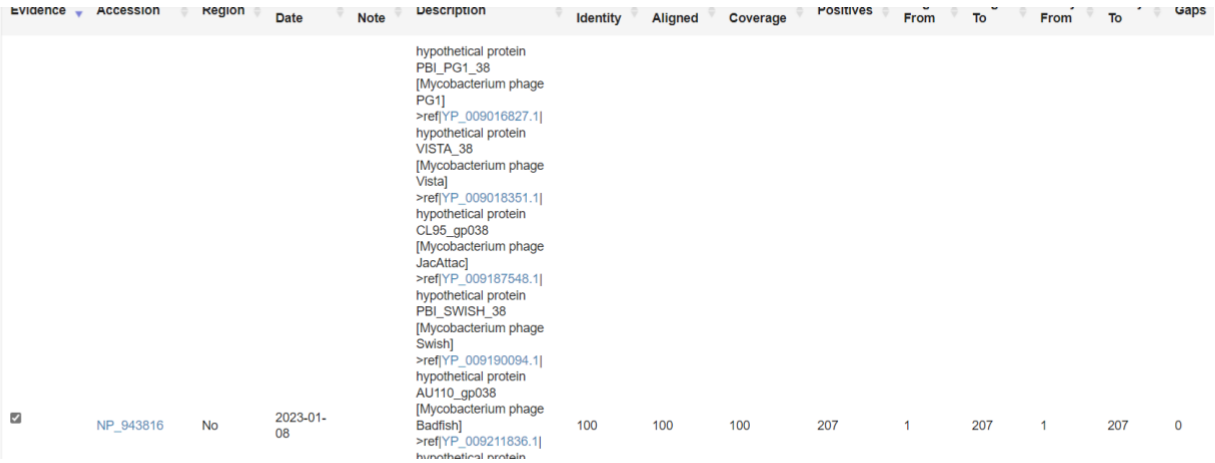
Distribution of 100 Blast Hits on the Query Sequence

Mouse-over to show defline and scores. Click to show alignments



```
                                                                 Score    E
Sequences producing significant alignments:                      (bits)  Value

Yoshand_38, function unknown, 207                                    427   e-119
Xavier_37, function unknown, 207                                     427   e-119
Waterdiva_38, function unknown, 207                                  427   e-119
Vista_38, function unknown, 207                                      427   e-119
Virapocalypse_38, function unknown, 207                              427   e-119
Thora_38, function unknown, 207                                      427   e-119
Telesworld_37, function unknown, 207                                 427   e-119
Swish_38, function unknown, 207                                      427   e-119
Squiggle_38, function unknown, 207                                   427   e-119
Slatt_38, function unknown, 207                                      427   e-119
Skippy_38, function unknown, 207                                     427   e-119
Sigman_38, function unknown, 207                                     427   e-119
Serendipity_38, function unknown, 207                                427   e-119
Scoot17C_38, function unknown, 207                                   427   e-119
Riggan_38, function unknown, 207                                     427   e-119
Podrick_38, function unknown, 207                                    427   e-119
Plmatters_38, function unknown, 207                                  427   e-119
Pipsqueak_38, function unknown, 207                                  427   e-119
PhrodoBaggins_37, function unknown, 207                              427   e-119
Phipps_38, function unknown, 207                                     427   e-119
Pherdinand_38, function unknown, 207                                 427   e-119
PG1_38, function unknown, 207                                        427   e-119
Orfeu_Draft_37, function unknown, 207                                427   e-119
Oosterbaan_38, function unknown, 207                                 427   e-119
Numberten_38, function unknown, 207                                  427   e-119
Mulan_37, function unknown, 207                                      427   e-119
Matalotodo_Draft_37, function unknown, 207                           427   e-119
Legolas_37, function unknown, 207                                    427   e-119
Lego3393_38, function unknown, 207                                   427   e-119
Lasso_38, function unknown, 207                                      427   e-119
Kloppinator_38, function unknown, 207                                427   e-119
JangoPhett_38, function unknown, 207                                 427   e-119
JacAttac_38, function unknown, 207                                   427   e-119
Iridoclysis_38, function unknown, 207                                427   e-119
Hocus_37, function unknown, 207                                      427   e-119
```

## b. SIF: NCBI BLAST



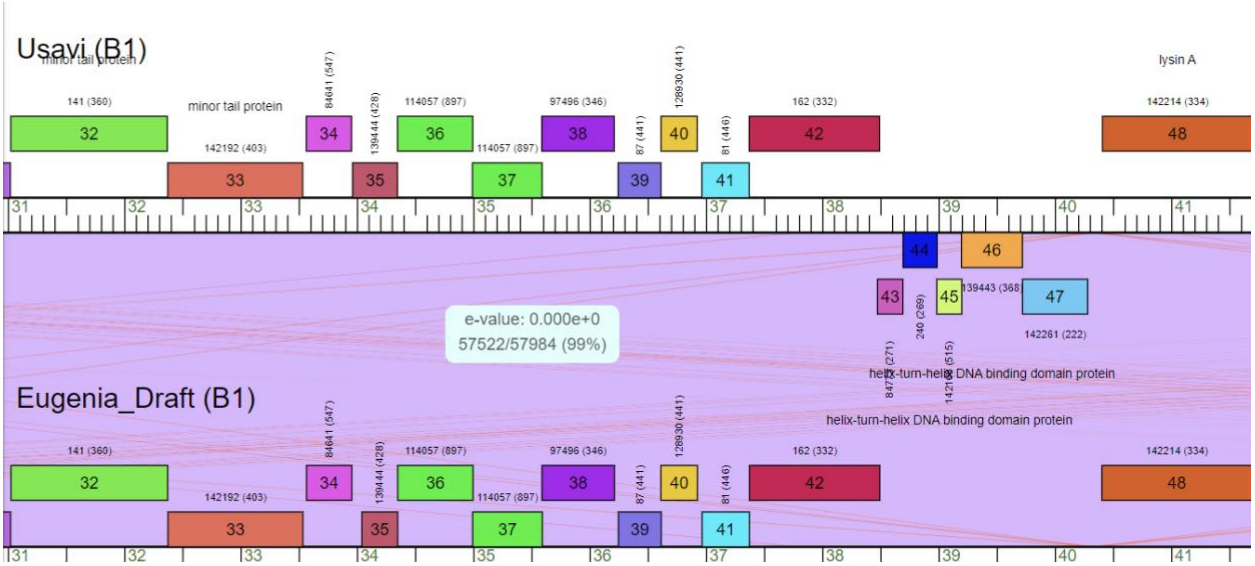
| Evidence | Accession | Region | Date | Note | Description | Identity | Aligned | Coverage | Positives | From | To | From | To | Gaps |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | NP_943816 | No | 2023-01-08 |  | hypothetical protein PBI_PG1_38 [Mycobacterium phage PG1] >ref\|YP_009016827.1\| hypothetical protein VISTA_38 [Mycobacterium phage Vista] >ref\|YP_009018351.1\| hypothetical protein CL95_gp038 [Mycobacterium phage JacAttac] >ref\|YP_009187548.1\| hypothetical protein PBI_SWISH_38 [Mycobacterium phage Swish] >ref\|YP_009190094.1\| hypothetical protein AU110_gp038 [Mycobacterium phage Badfish] >ref\|YP_009211836.1\| | 100 | 100 | 100 | 207 | 1 | 207 | 1 | 207 | 0 |

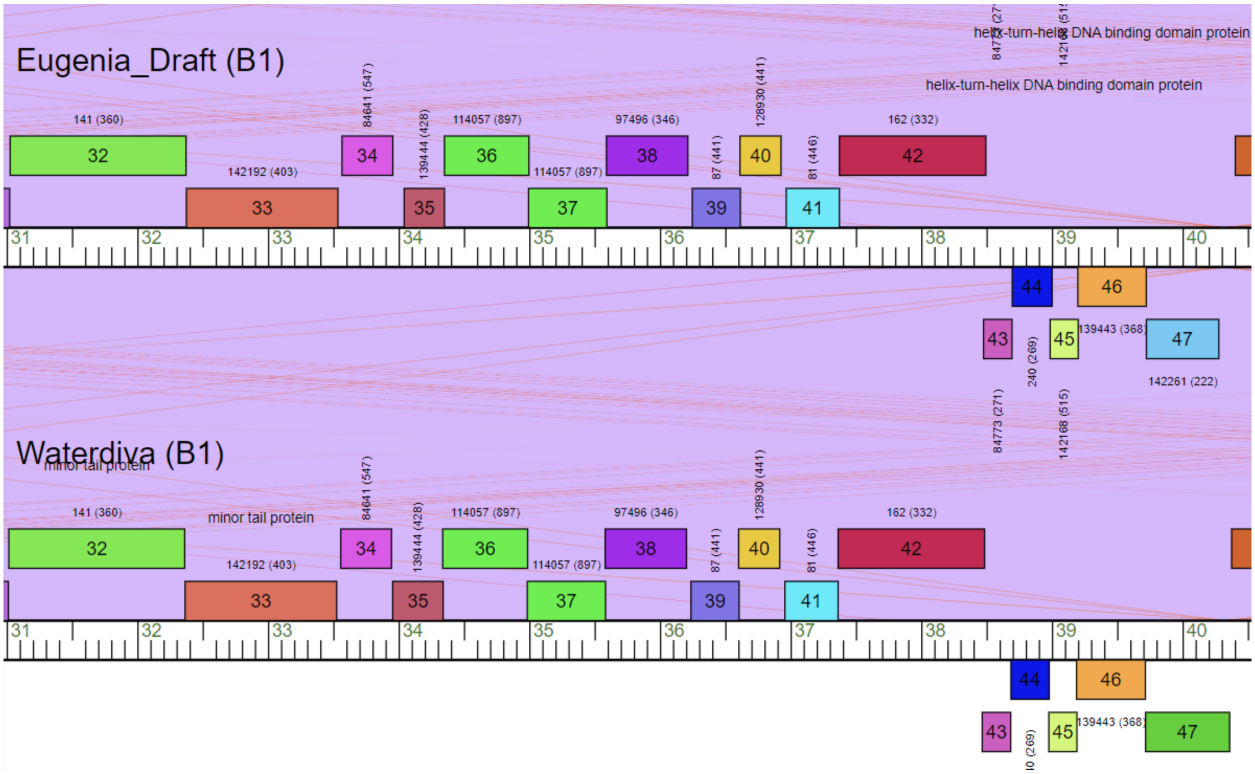
| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | [Mycobacterium phage<br>Serendipity] | | | | | | | | | | |
| ☑ | ALA45649 | No | 2021-12-01 | hypothetical protein HETAERIA_38 [Mycobacterium phage Hetaeria] >gb\|AXH67133.1\| hypothetical protein SEA_UACH1_38 [Mycobacterium phage UAch1] | 99.0338 | 100 | 100 | 207 | 1 | 207 | 1\| | 207 | 0 | 1.24347<br>149 |
| | | | | hypothetical protein | | | | | | | | | | |

## c. SIF: HHPred

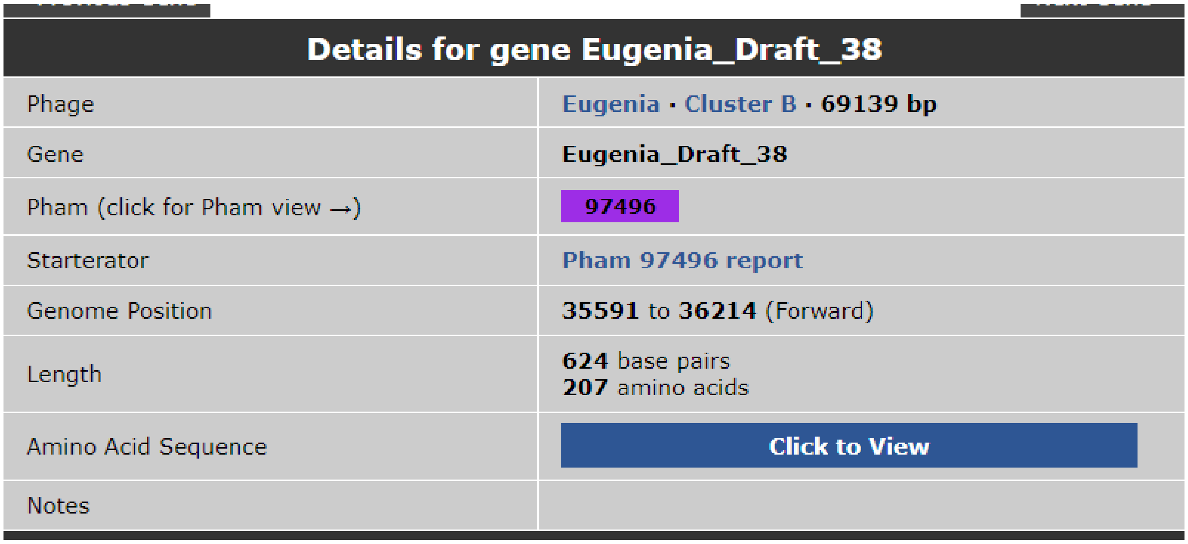
No Evidence from HHPred

## d. SIF: Synteny-Phamerator (three genomes)





## 7. Any other important information.



| Details for gene Eugenia_Draft_38 | |
|---|---|
| Phage | Eugenia · Cluster B · **69139 bp** |
| Gene | **Eugenia_Draft_38** |
| Pham (click for Pham view →) | **97496** |
| Starterator | Pham 97496 report |
| Genome Position | **35591** to **36214** (Forward) |
| Length | **624** base pairs<br>**207** amino acids |
| Amino Acid Sequence | Click to View |
| Notes | |

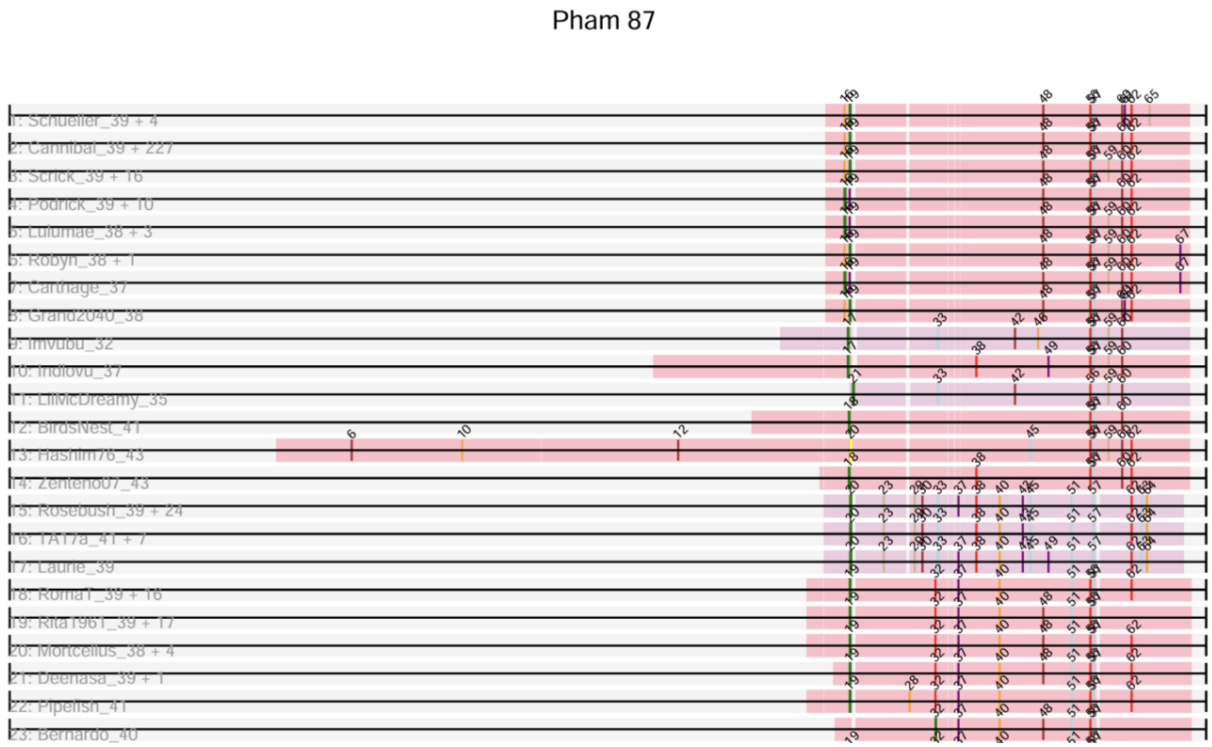
### **CURATOR NAME: OLIVIA SIDOTI**

### **GENE NAME: EUGENIA GENE 39**

### **DNA MASTER NOTES: N/A**

### **START POSITION EVALUATION (IN ORDER):**

1. Starterator



Pham number 87 has 438 members, 50 are drafts.

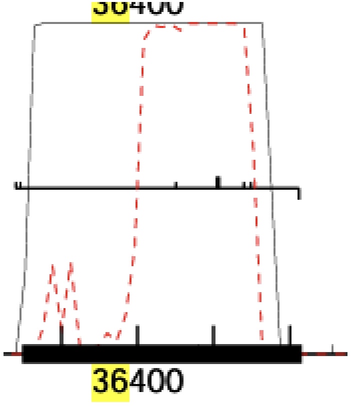
The start number called the most often in the published annotations is 19, it was called in 274 of the 388 non-draft genes in the pham.

Start 19:

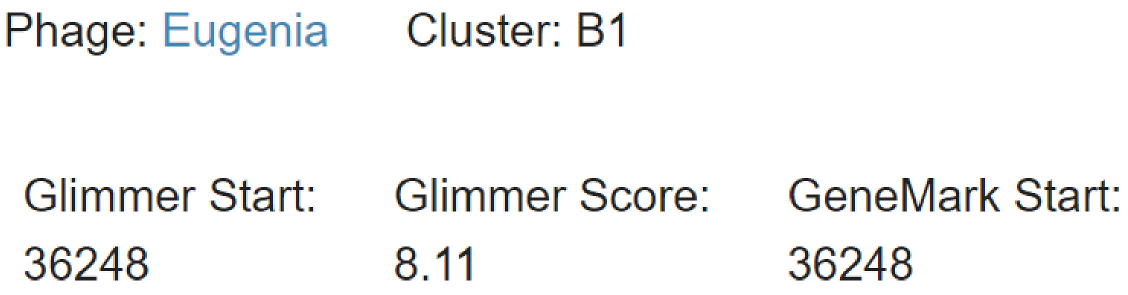
- Found in 316 of 438 ( 72.1% ) of genes in pham
- Manual Annotations of this start: 274 of 388
- Called 94.9% of time when present

## 2. GeneMark coding potential

Start: 36,248
Stop: 36,616



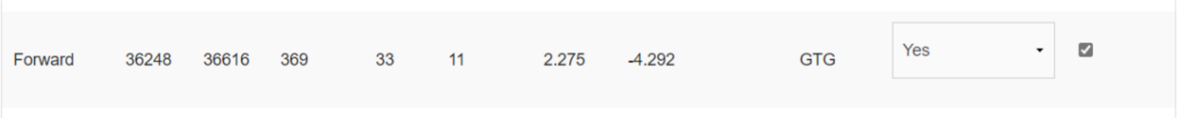
## 3. Glimmer and GeneMark agreement


Phage: Eugenia Cluster: B1

| Glimmer Start: | Glimmer Score: | GeneMark Start: |
|---|---|---|
| 36248 | 8.11 | 36248 |

Yes, Glimmer and Genemark are in agreement that the start is at position 36248.

## 4. Longest open reading frame (ORF) without excessive gap



| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 36248 | 36616 | 369 | 33 | 11 | 2.275 | -4.292 | GTG | Yes | ☑ |

No, this gene is not the longest open reading frame.

Gap: 33

Spacer: 11

Z-Score: 2.275

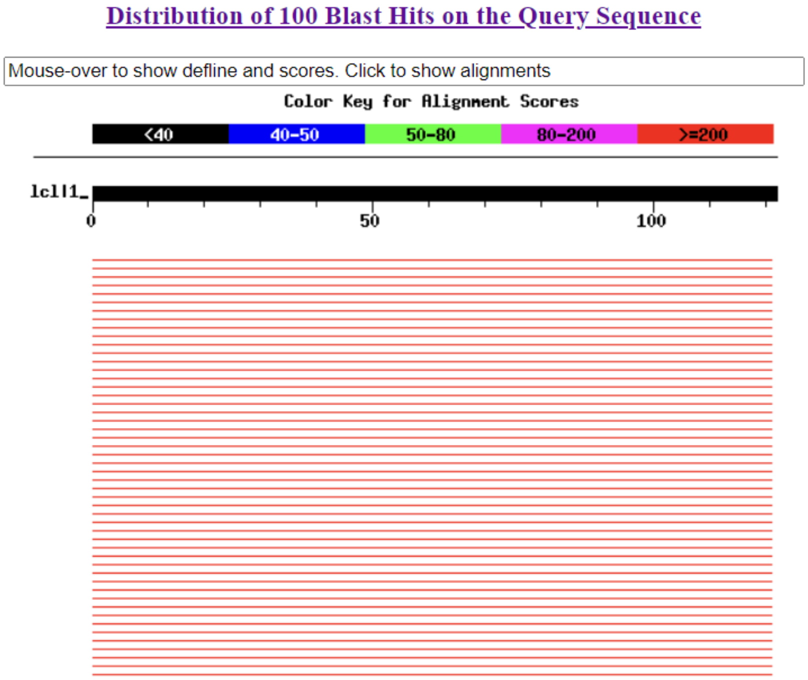
Final Score: -4.292

## 5. Function. If no functional prediction is present, write “Hypothetical protein”.

Hypothetical Protein

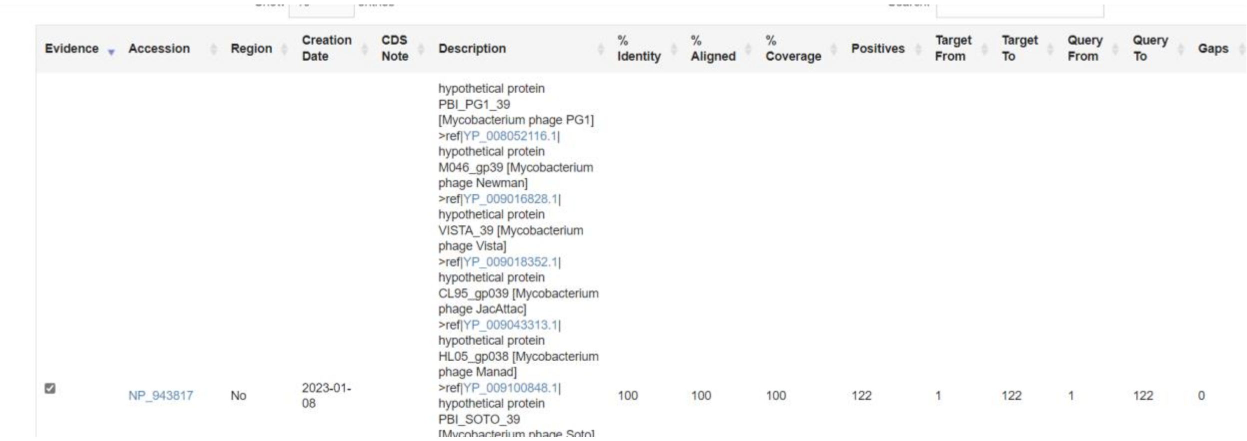
## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST

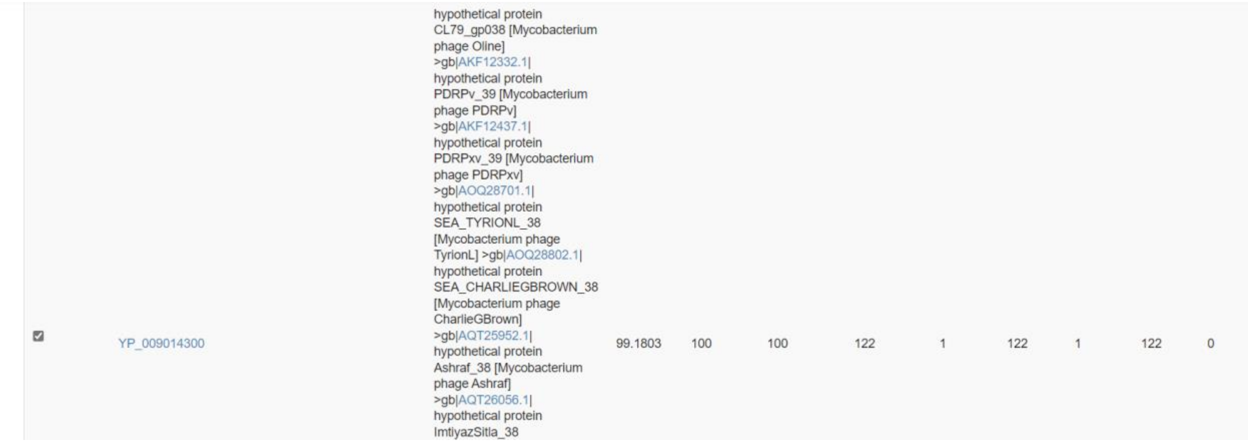


```
                                                                 Score     E
Sequences producing significant alignments:                      (bits)  Value

Zonia_39, function unknown, 122                                       247    6e-66
Zelda_39, function unknown, 122                                       247    6e-66
Zaider_40, function unknown, 122                                      247    6e-66
YouGoGlencoco_39, function unknown, 122                               247    6e-66
Yoshand_39, function unknown, 122                                     247    6e-66
Xavier_38, function unknown, 122                                      247    6e-66
Waterdiva_39, function unknown, 122                                   247    6e-66
Wallhey_38, function unknown, 122                                     247    6e-66
Vortex_39, function unknown, 122                                      247    6e-66
Vivaldi_39, function unknown, 122                                     247    6e-66
Vista_39, function unknown, 122                                       247    6e-66
Virapocalypse_39, function unknown, 122                               247    6e-66
Vaishali24_38, function unknown, 122                                  247    6e-66
Usavi_39, function unknown, 122                                       247    6e-66
UncleHowie_39, function unknown, 122                                  247    6e-66
UAch1_39, function unknown, 122                                       247    6e-66
True_38, function unknown, 122                                        247    6e-66
Tooj_39, function unknown, 122                                        247    6e-66
Toni_38, function unknown, 122                                        247    6e-66
Tomlarah_39, function unknown, 122                                    247    6e-66
TomBombadil_39, function unknown, 122                                 247    6e-66
Timmi_38, function unknown, 122                                       247    6e-66
ThreeOh3D2_39, function unknown, 122                                  247    6e-66
Thora_39, function unknown, 122                                       247    6e-66
Telesworld_38, function unknown, 122                                  247    6e-66
TallGrassMM_39, function unknown, 122                                 247    6e-66
Swish_39, function unknown, 122                                       247    6e-66
Surely_39, function unknown, 122                                      247    6e-66
Squiggle_39, function unknown, 122                                    247    6e-66
Squid_39, function unknown, 122                                       247    6e-66
Spartan300_39, function unknown, 122                                  247    6e-66
Soto_39, function unknown, 122                                        247    6e-66
Sophia_38, function unknown, 122                                      247    6e-66
Solosis_38, function unknown, 122                                     247    6e-66
Slatt_39, function unknown, 122                                       247    6e-66
Skippy_39, function unknown, 122                                      247    6e-66
Simielle_38, function unknown, 122                                    247    6e-66
Sigman_39, function unknown, 122                                      247    6e-66
```

## b. SIF: NCBI BLAST



| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | NP_943817 | No | 2023-01-08 | | hypothetical protein PBI_PG1_39 [Mycobacterium phage PG1] >ref\|YP_008052116.1\| hypothetical protein M046_gp39 [Mycobacterium phage Newman] >ref\|YP_009016828.1\| hypothetical protein VISTA_39 [Mycobacterium phage Vista] >ref\|YP_009018352.1\| hypothetical protein CL95_gp039 [Mycobacterium phage JacAttac] >ref\|YP_009043313.1\| hypothetical protein HL05_gp038 [Mycobacterium phage Manad] >ref\|YP_009100848.1\| hypothetical protein PBI_SOTO_39 [Mycobacterium phage Soto] | 100 | 100 | 100 | 122 | 1 | 122 | 1 | 122 | 0 |



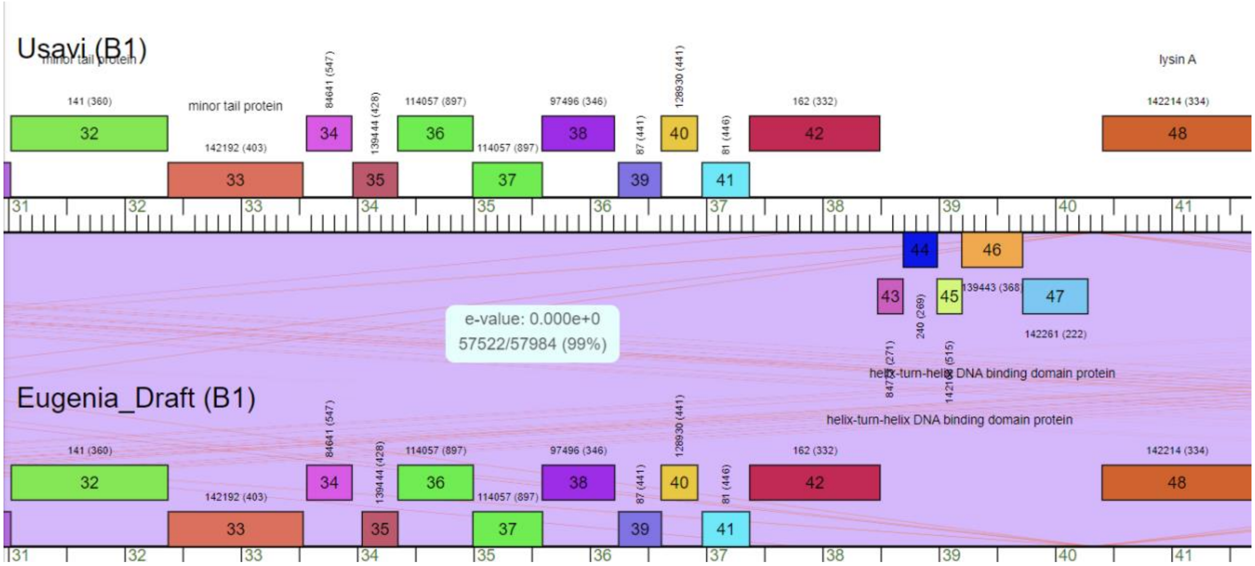
| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | YP_009014300 | | | | hypothetical protein CL79_gp038 [Mycobacterium phage Oline] >gb\|AKF12332.1\| hypothetical protein PDRPv_39 [Mycobacterium phage PDRPv] >gb\|AKF12437.1\| hypothetical protein PDRPxv_39 [Mycobacterium phage PDRPxv] >gb\|AOQ28701.1\| hypothetical protein SEA_TYRIONL_38 [Mycobacterium phage TyrionL] >gb\|AOQ28802.1\| hypothetical protein SEA_CHARLIEGBROWN_38 [Mycobacterium phage CharlieGBrown] >gb\|AQT25952.1\| hypothetical protein Ashraf_38 [Mycobacterium phage Ashraf] >gb\|AQT26056.1\| hypothetical protein ImtiyazSitla_38 | 99.1803 | 100 | 100 | 122 | 1 | 122 | 1 | 122 | 0 |

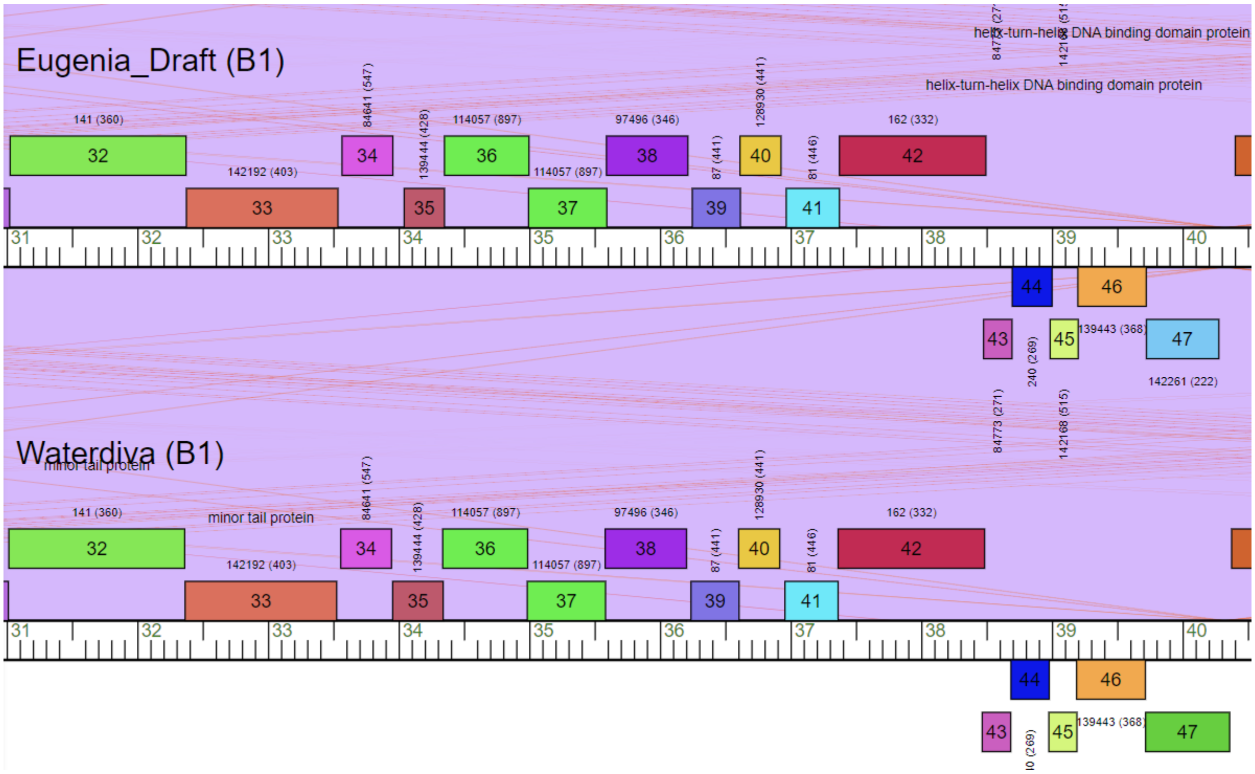
| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | PhenghisKhan] | | | | | | | | | |
| ☑ | YP_009005686 | No | 2023-01-08 | | hypothetical protein PBI_SUFFOLK_39 [Mycobacterium phage Suffolk] >gb\|AHB79675.1\| hypothetical protein PBI_SUFFOLK_39 [Mycobacterium phage Suffolk] | 99.1803 | 100 | 100 | 122 | 1 | 122 | 1 | 122 | 0 |

## c. SIF: HHPred

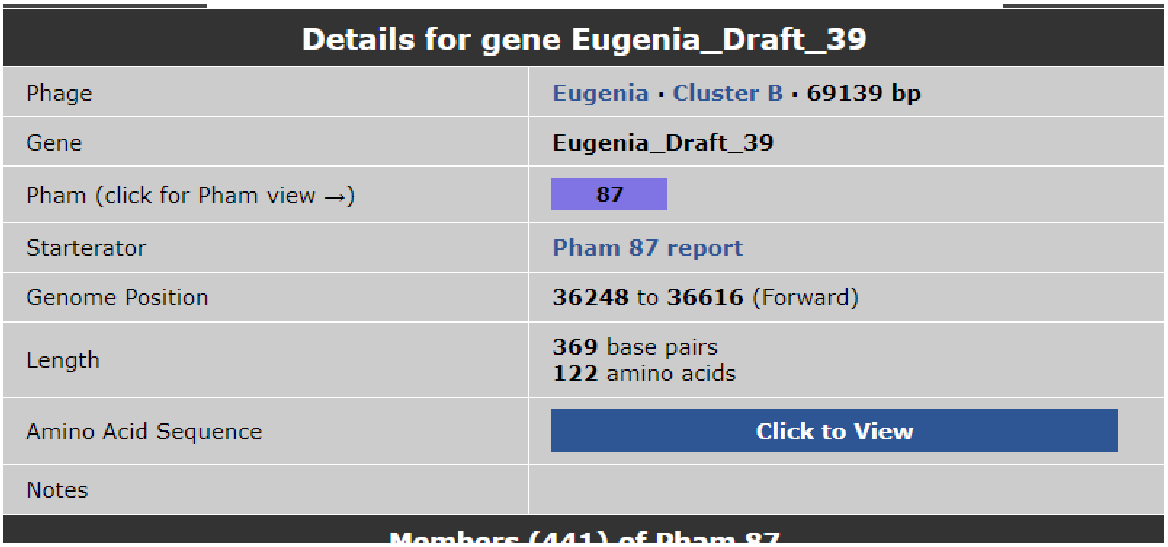
No Evidence from HHPred

## d. SIF: Synteny-Phamerator (three genomes)





7. Any other important information.



**Details for gene Eugenia_Draft_39**

| | |
|---|---|
| Phage | Eugenia · Cluster B · **69139 bp** |
| Gene | **Eugenia_Draft_39** |
| Pham (click for Pham view →) | **87** |
| Starterator | Pham 87 report |
| Genome Position | **36248** to **36616** (Forward) |
| Length | **369** base pairs<br>**122** amino acids |
| Amino Acid Sequence | Click to View |
| Notes | |

**Members (441) of Pham 87**

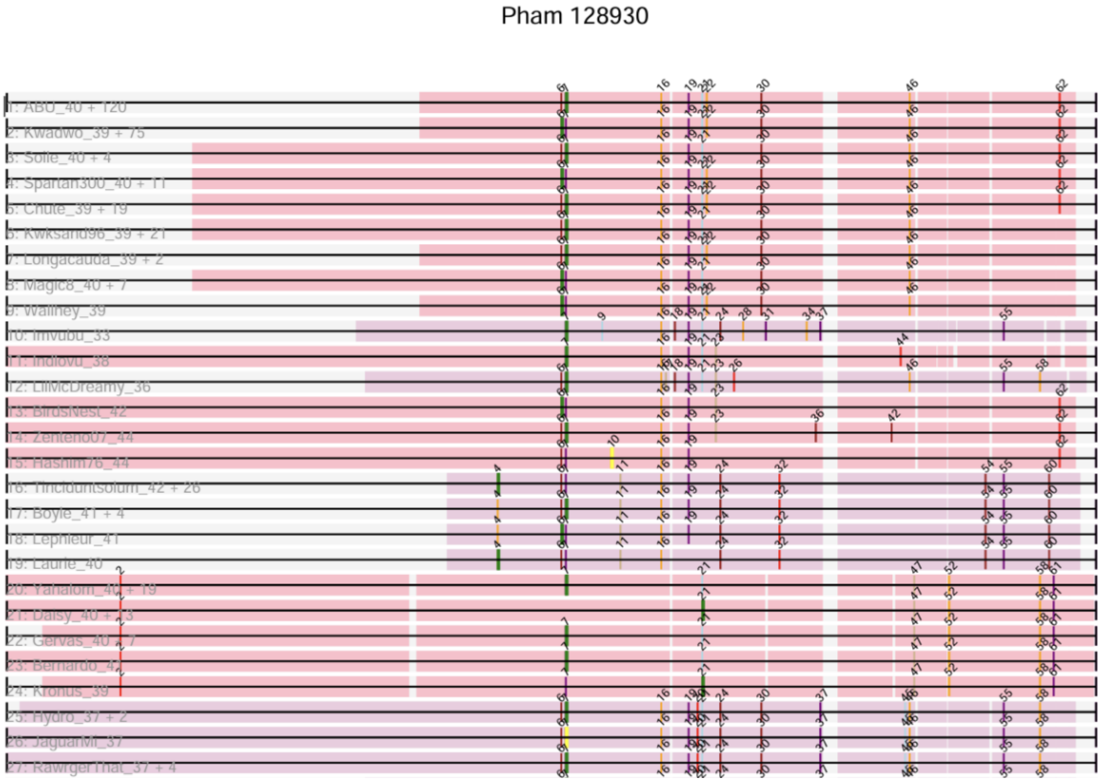
**Curator name: Olivia Sidoti**

**Gene Name: Eugenia Gene 40**

**DNA Master Notes: N/A**

**Start Position Evaluation (In order)**:

1. Starterator



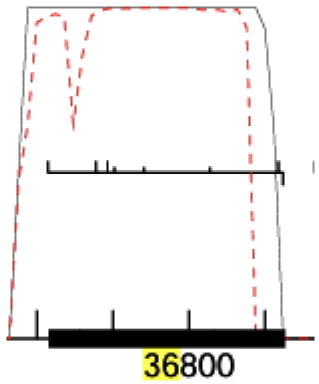
Pham number 128930 has 438 members, 51 are drafts.

The start number called the most often in the published annotations is 7, it was called in 220 of the 387 non-draft genes in the pham.

Start 7:

- Found in 392 of 438 ( 89.5% ) of genes in pham
- Manual Annotations of this start: 220 of 387
- Called 64.3% of time when present

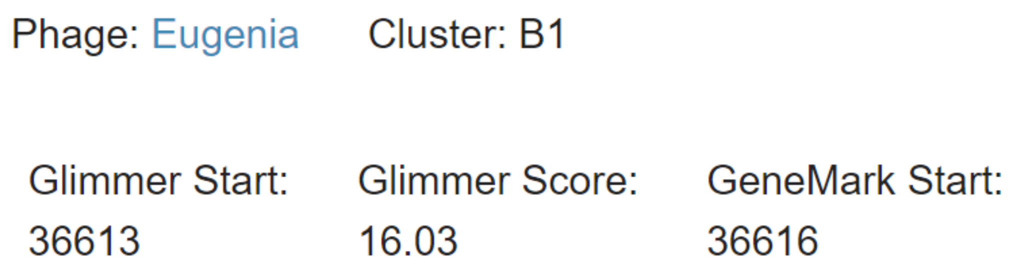
2. GeneMark coding potential



Start: 36,613

Stop: 36,927

3. Glimmer and GeneMark agreement

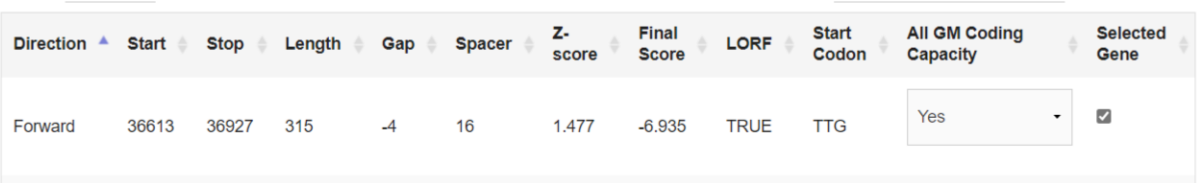


Phage: Eugenia Cluster: B1

| Glimmer Start: | Glimmer Score: | GeneMark Start: |
|---|---|---|
| 36613 | 16.03 | 36616 |

No, Glimmer and GeneMark are not in agreement about the start position.

4. Longest open reading frame (ORF) without excessive gap



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 36613 | 36927 | 315 | -4 | 16 | 1.477 | -6.935 | TRUE | TTG | Yes | ☑ |

Yes, this gene is the longest open reading frame.

Gap: -4

Spacer: -16

Z-Score: 1.477

Final Score: -6.

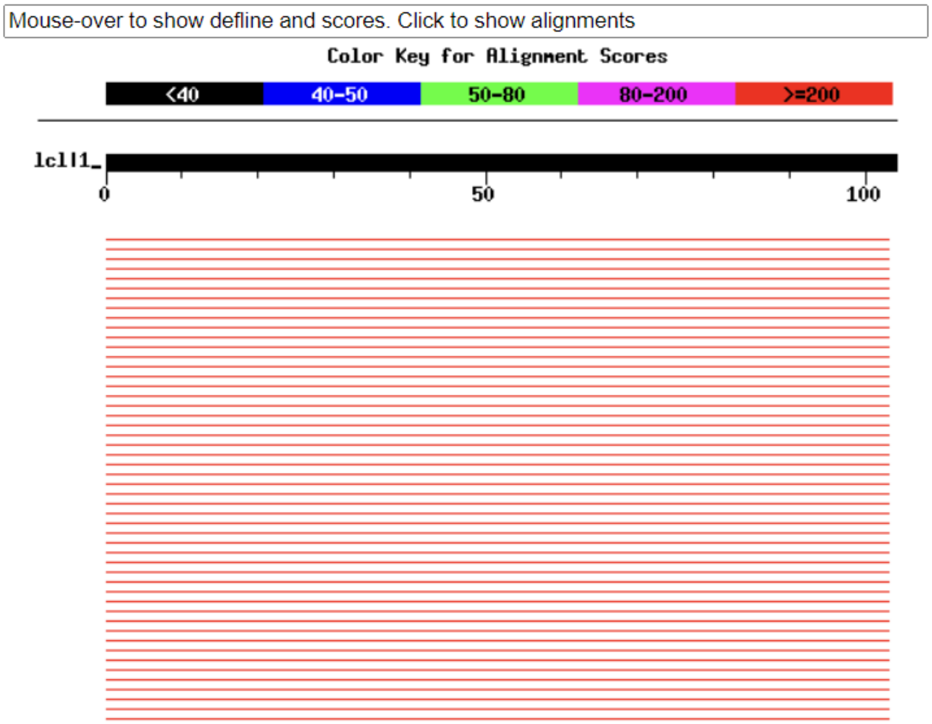
5. Function. If no functional prediction is present, write “Hypothetical protein”.

Hypothetical Protein

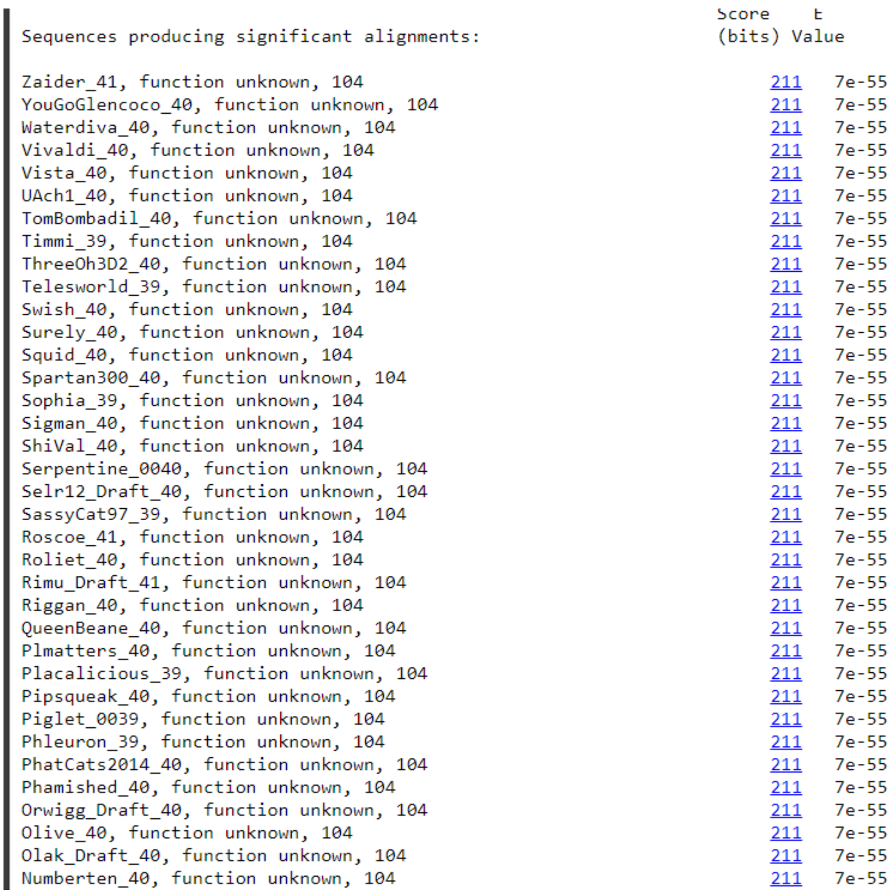
6. Supporting Information for Function (SIF)

a. SIF: PhageDb BLAST

## Distribution of 100 Blast Hits on the Query Sequence

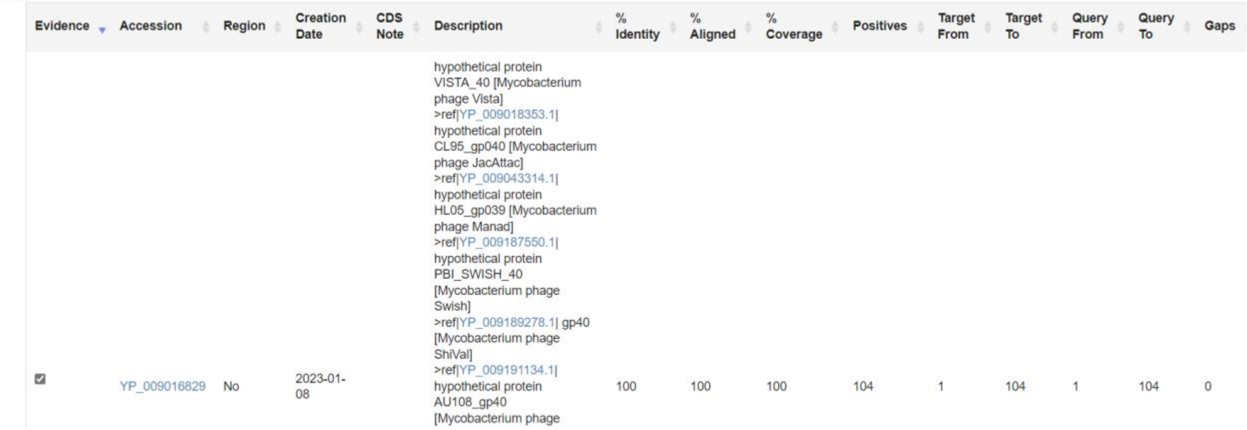


Mouse-over to show defline and scores. Click to show alignments



```
                                                                 Score     E
Sequences producing significant alignments:                      (bits) Value

Zaider_41, function unknown, 104                                      211    7e-55
YouGoGlencoco_40, function unknown, 104                               211    7e-55
Waterdiva_40, function unknown, 104                                   211    7e-55
Vivaldi_40, function unknown, 104                                     211    7e-55
Vista_40, function unknown, 104                                       211    7e-55
UAch1_40, function unknown, 104                                       211    7e-55
TomBombadil_40, function unknown, 104                                 211    7e-55
Timmi_39, function unknown, 104                                       211    7e-55
ThreeOh3D2_40, function unknown, 104                                  211    7e-55
Telesworld_39, function unknown, 104                                  211    7e-55
Swish_40, function unknown, 104                                       211    7e-55
Surely_40, function unknown, 104                                      211    7e-55
Squid_40, function unknown, 104                                       211    7e-55
Spartan300_40, function unknown, 104                                  211    7e-55
Sophia_39, function unknown, 104                                      211    7e-55
Sigman_40, function unknown, 104                                      211    7e-55
ShiVal_40, function unknown, 104                                      211    7e-55
Serpentine_0040, function unknown, 104                                211    7e-55
Selr12_Draft_40, function unknown, 104                                211    7e-55
SassyCat97_39, function unknown, 104                                  211    7e-55
Roscoe_41, function unknown, 104                                      211    7e-55
Roliet_40, function unknown, 104                                      211    7e-55
Rimu_Draft_41, function unknown, 104                                  211    7e-55
Riggan_40, function unknown, 104                                      211    7e-55
QueenBeane_40, function unknown, 104                                  211    7e-55
Plmatters_40, function unknown, 104                                   211    7e-55
Placalicious_39, function unknown, 104                                211    7e-55
Pipsqueak_40, function unknown, 104                                   211    7e-55
Piglet_0039, function unknown, 104                                    211    7e-55
Phleuron_39, function unknown, 104                                    211    7e-55
PhatCats2014_40, function unknown, 104                                211    7e-55
Phamished_40, function unknown, 104                                   211    7e-55
Orwigg_Draft_40, function unknown, 104                                211    7e-55
Olive_40, function unknown, 104                                       211    7e-55
Olak_Draft_40, function unknown, 104                                  211    7e-55
Numberten_40, function unknown, 104                                   211    7e-55
```

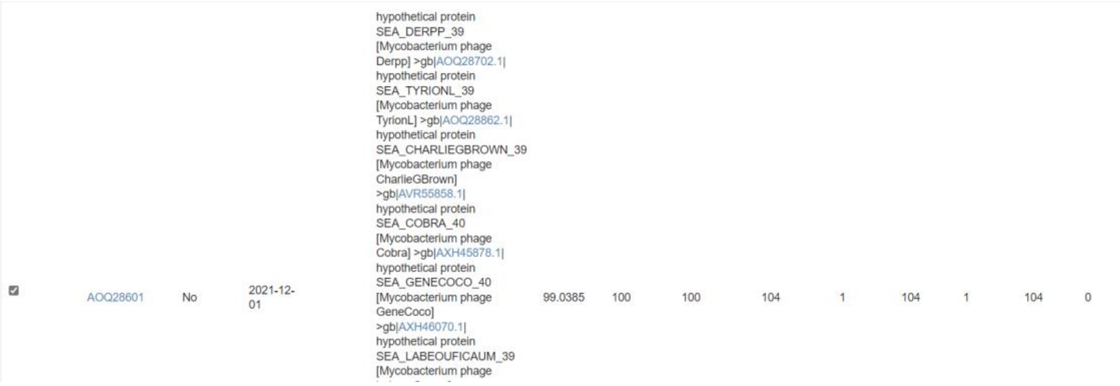
## b. SIF: NCBI BLAST



| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | YP_009016829 | No | 2023-01-08 | | hypothetical protein VISTA_40 [Mycobacterium phage Vista] >ref\|YP_009018353.1\| hypothetical protein CL95_gp040 [Mycobacterium phage JacAttac] >ref\|YP_009043314.1\| hypothetical protein HL05_gp039 [Mycobacterium phage Manad] >ref\|YP_009187550.1\| hypothetical protein PBI_SWISH_40 [Mycobacterium phage Swish] >ref\|YP_009189278.1\| gp40 [Mycobacterium phage ShiVal] >ref\|YP_009191134.1\| hypothetical protein AU108_gp40 [Mycobacterium phage | 100 | 100 | 100 | 104 | 1 | 104 | 1 | 104 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | ASZ73429 | No | 2021-12-01 | hypothetical protein SEA_LULUMAE_39 [Mycobacterium phage Lulumae] >gb\|AXH48452.1\| hypothetical protein SEA_ROY17_40 [Mycobacterium phage Roy17] | 99.0385 | 100 | 100 | 104 | 1 | 104 | 1 | 104 | 0 |

hypothetical protein

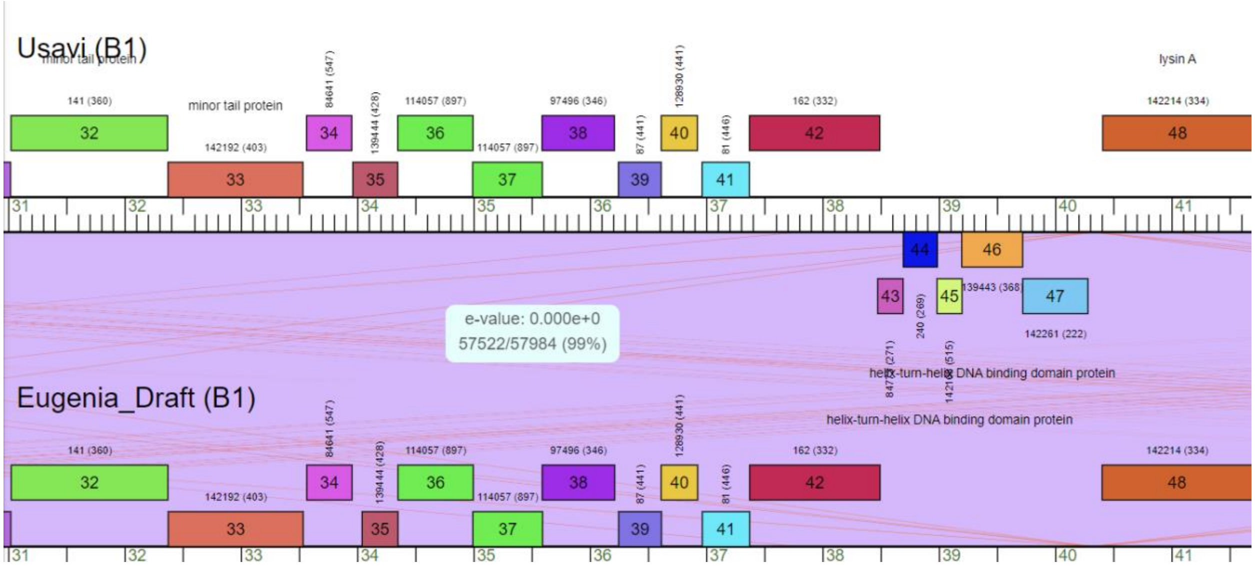


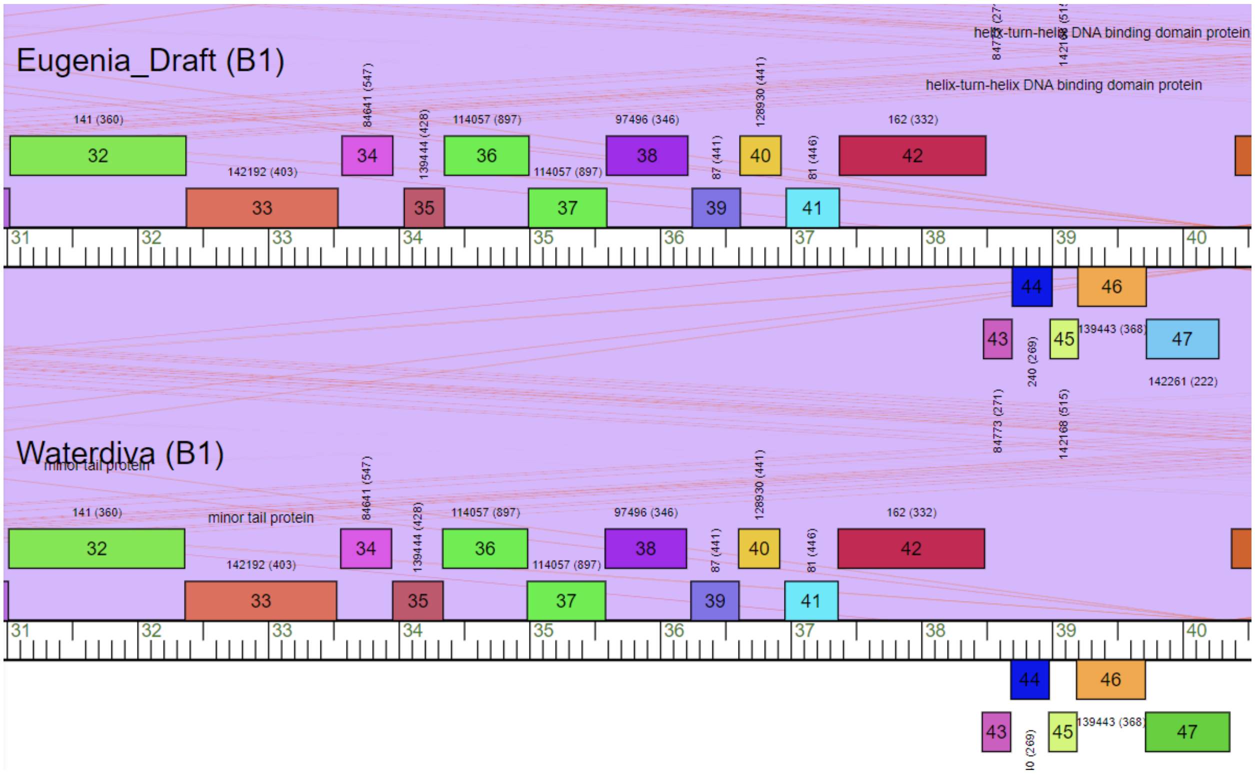
| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | AOQ28601 | No | 2021-12-01 | hypothetical protein SEA_DERPP_39 [Mycobacterium phage Derpp] >gb\|AOQ28702.1\| hypothetical protein SEA_TYRIONL_39 [Mycobacterium phage TyrionL] >gb\|AOQ28862.1\| hypothetical protein SEA_CHARLIEGBROWN_39 [Mycobacterium phage CharlieGBrown] >gb\|AVR55858.1\| hypothetical protein SEA_COBRA_40 [Mycobacterium phage Cobra] >gb\|AXH45878.1\| hypothetical protein SEA_GENECOCO_40 [Mycobacterium phage GeneCoco] >gb\|AXH46070.1\| hypothetical protein SEA_LABEOUFICAUM_39 [Mycobacterium phage | 99.0385 | 100 | 100 | 104 | 1 | 104 | 1 | 104 | 0 |

c. SIF: HHPred

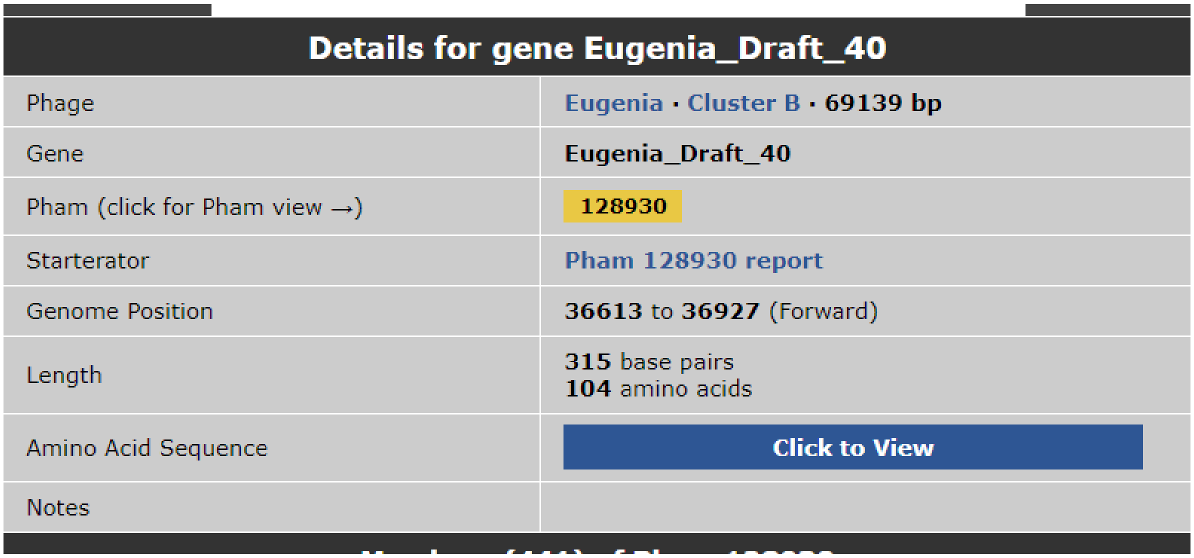
No Evidence from HHPred

d. SIF: Synteny-Phamerator (three genomes)





7. Any other important information.



| Details for gene Eugenia_Draft_40 | |
|---|---|
| Phage | Eugenia · Cluster B · 69139 bp |
| Gene | Eugenia_Draft_40 |
| Pham (click for Pham view →) | 128930 |
| Starterator | Pham 128930 report |
| Genome Position | 36613 to 36927 (Forward) |
| Length | 315 base pairs<br>104 amino acids |
| Amino Acid Sequence | Click to View |
| Notes | |

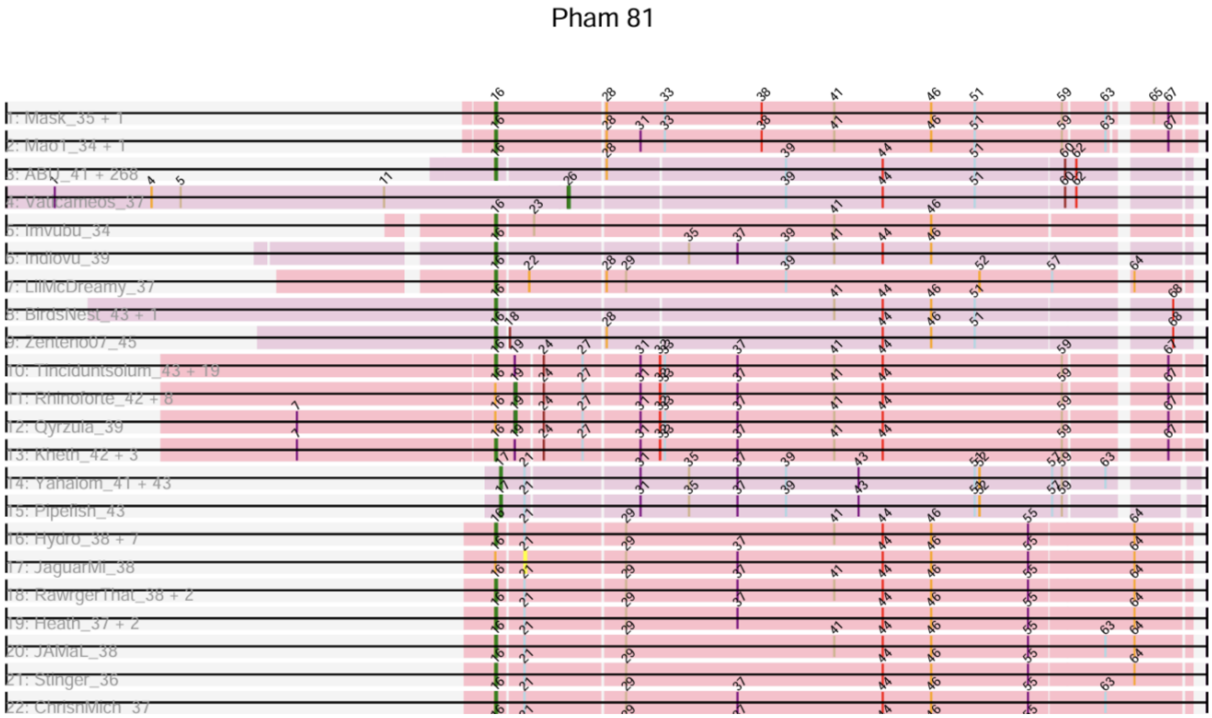
## CURATOR NAME: OLIVIA SIDOTI

## GENE NAME: EUGENIA GENE 41

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

1. Starterator



Pham number 81 has 443 members, 51 are drafts.

The start number called the most often in the published annotations is 16, it was called in 335 of the 392 non-draft genes in the pham

Start 16:

- Found in 397 of 443 ( 89.6% ) of genes in pham

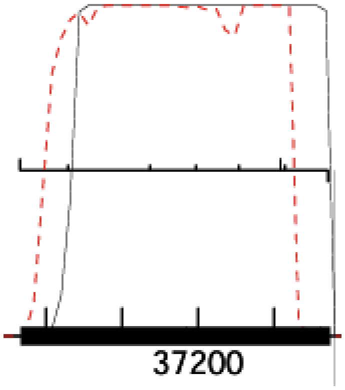
- Manual Annotations of this start: 335 of 392

- Called 95.2% of time when present

## 2. GeneMark coding potential

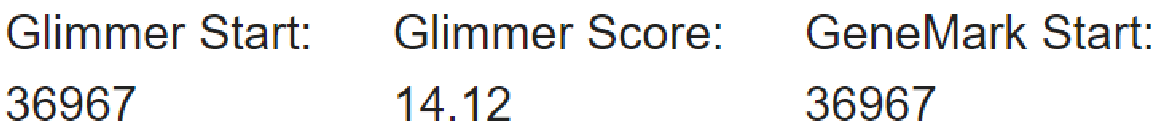
## Start: 36,967

## Stop: 37,374



## 3. Glimmer and GeneMark agreement

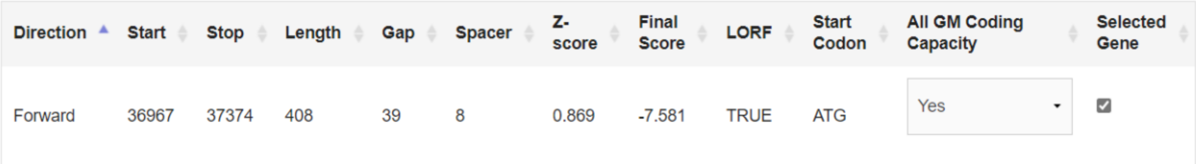
Phage: Eugenia      Cluster: B1



| Glimmer Start: | Glimmer Score: | GeneMark Start: |
|---|---|---|
| 36967 | 14.12 | 36967 |

Yes, Glimmer and GeneMark agree that the start position is at 36967.

## 4. Longest open reading frame (ORF) without excessive gap



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 36967 | 37374 | 408 | 39 | 8 | 0.869 | -7.581 | TRUE | ATG | Yes | ☑ |

Yes, this gene is the longest open reading frame.

Gap: 39

Spacer: 8

Z-Score: 0.869

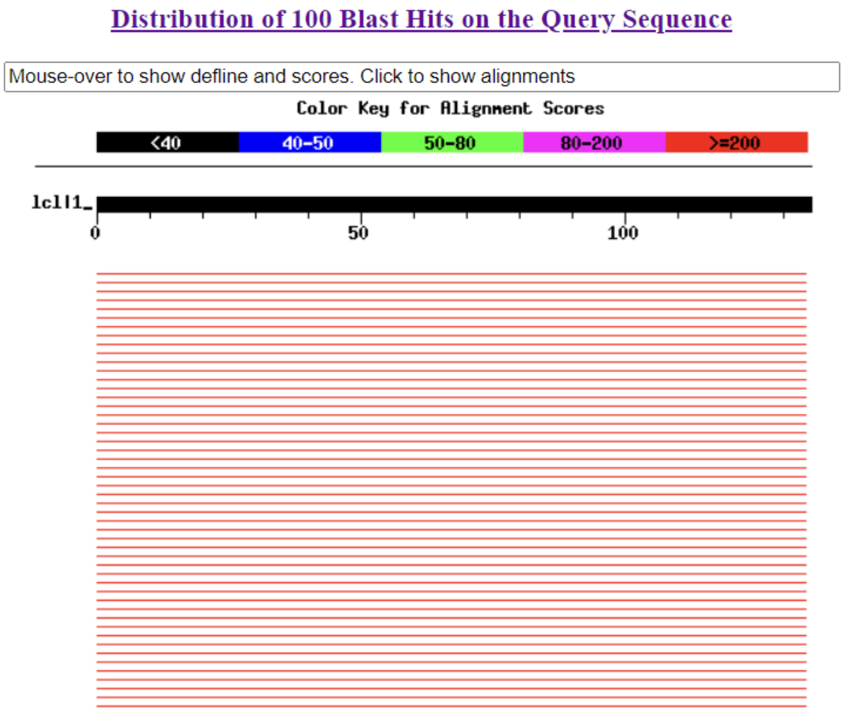
Final Score: -7.581

## 5. Function. If no functional prediction is present, write “Hypothetical protein”.

Hypothetical Protein

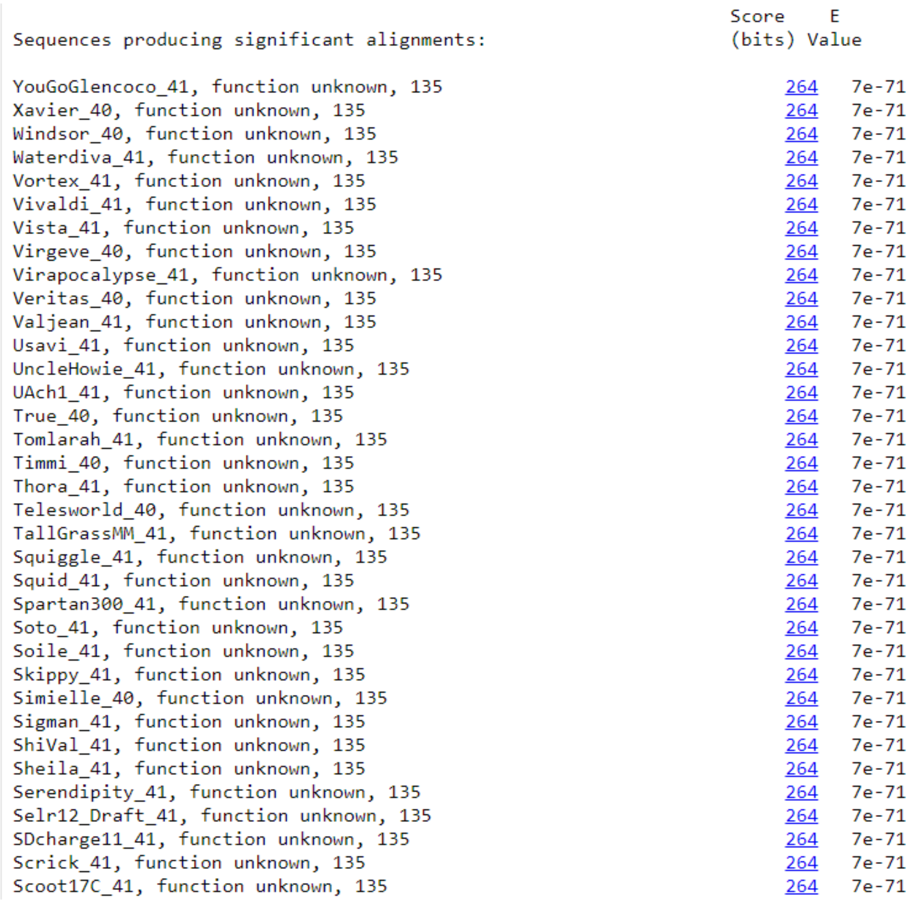
## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST



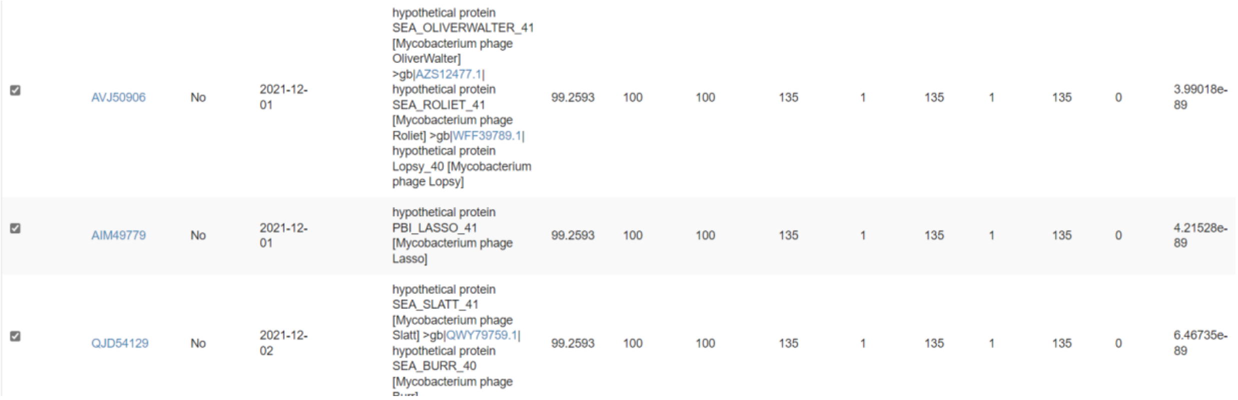
Distribution of 100 Blast Hits on the Query Sequence

Mouse-over to show defline and scores. Click to show alignments



```
                                                                 Score     E
Sequences producing significant alignments:                      (bits)  Value

YouGoGlencoco_41, function unknown, 135                             264    7e-71
Xavier_40, function unknown, 135                                    264    7e-71
Windsor_40, function unknown, 135                                   264    7e-71
Waterdiva_41, function unknown, 135                                 264    7e-71
Vortex_41, function unknown, 135                                    264    7e-71
Vivaldi_41, function unknown, 135                                   264    7e-71
Vista_41, function unknown, 135                                     264    7e-71
Virgeve_40, function unknown, 135                                   264    7e-71
Virapocalypse_41, function unknown, 135                             264    7e-71
Veritas_40, function unknown, 135                                   264    7e-71
Valjean_41, function unknown, 135                                   264    7e-71
Usavi_41, function unknown, 135                                     264    7e-71
UncleHowie_41, function unknown, 135                                264    7e-71
UAch1_41, function unknown, 135                                     264    7e-71
True_40, function unknown, 135                                      264    7e-71
Tomlarah_41, function unknown, 135                                  264    7e-71
Timmi_40, function unknown, 135                                     264    7e-71
Thora_41, function unknown, 135                                     264    7e-71
Telesworld_40, function unknown, 135                                264    7e-71
TallGrassMM_41, function unknown, 135                               264    7e-71
Squiggle_41, function unknown, 135                                  264    7e-71
Squid_41, function unknown, 135                                     264    7e-71
Spartan300_41, function unknown, 135                                264    7e-71
Soto_41, function unknown, 135                                      264    7e-71
Soile_41, function unknown, 135                                     264    7e-71
Skippy_41, function unknown, 135                                    264    7e-71
Simielle_40, function unknown, 135                                  264    7e-71
Sigman_41, function unknown, 135                                    264    7e-71
ShiVal_41, function unknown, 135                                    264    7e-71
Sheila_41, function unknown, 135                                    264    7e-71
Serendipity_41, function unknown, 135                               264    7e-71
Selr12_Draft_41, function unknown, 135                              264    7e-71
SDcharge11_41, function unknown, 135                                264    7e-71
Scrick_41, function unknown, 135                                    264    7e-71
Scoot17C_41, function unknown, 135                                  264    7e-71
```

## b. SIF: NCBI BLAST

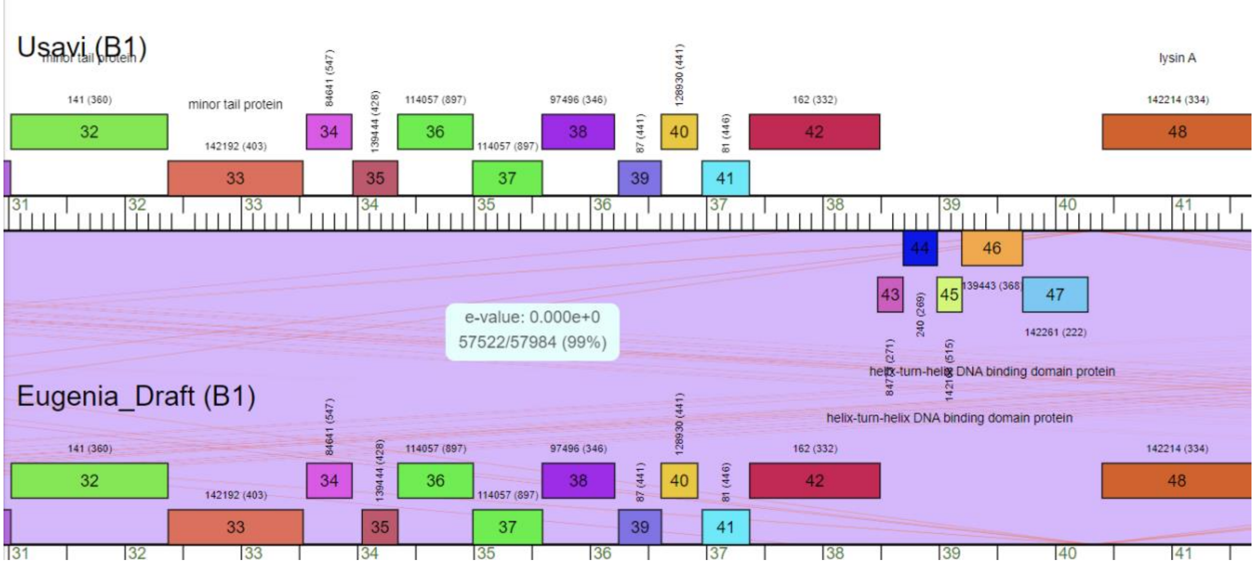


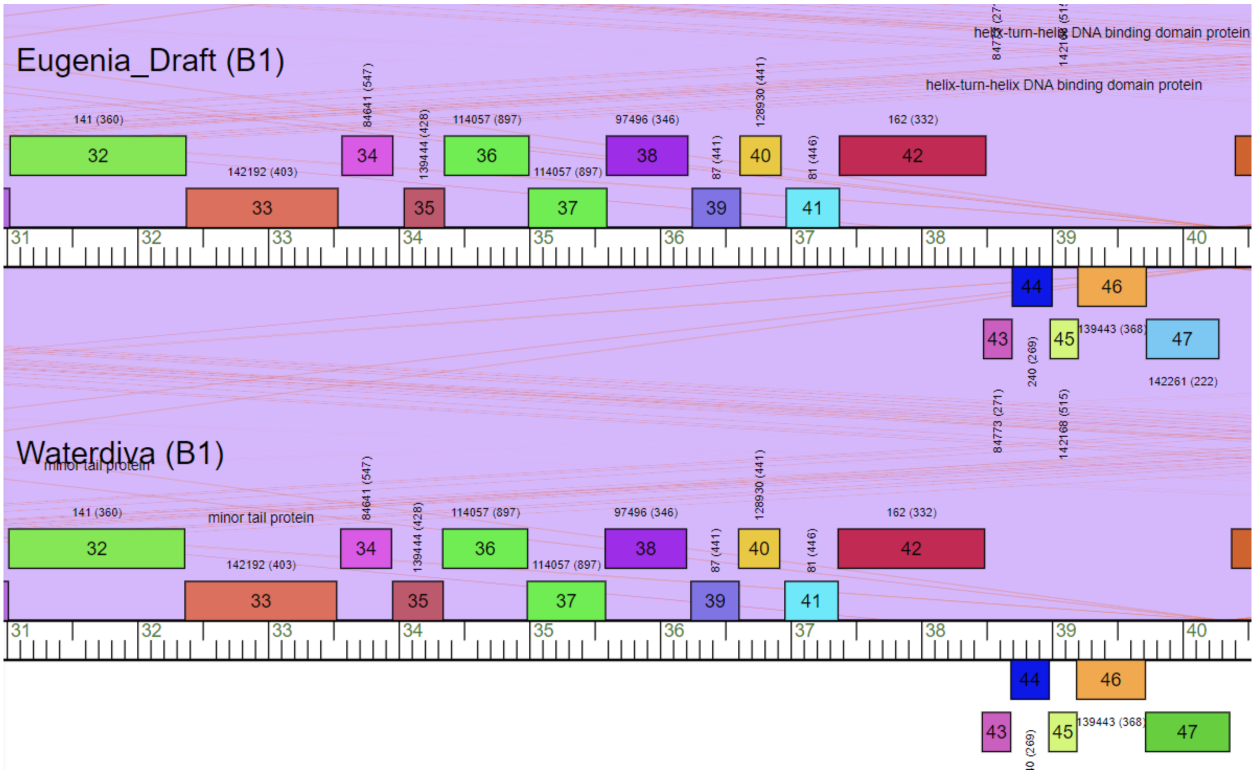
| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | AVJ50906 | No | 2021-12-01 | hypothetical protein SEA_OLIVERWALTER_41 [Mycobacterium phage OliverWalter] >gb\|AZS12477.1\| hypothetical protein SEA_ROLIET_41 [Mycobacterium phage Roliet] >gb\|WFF39789.1\| hypothetical protein Lopsy_40 [Mycobacterium phage Lopsy] | 99.2593 | 100 | 100 | 135 | 1 | 135 | 1 | 135 | 0 | 3.99018e-89 |
| ☑ | AIM49779 | No | 2021-12-01 | hypothetical protein PBI_LASSO_41 [Mycobacterium phage Lasso] | 99.2593 | 100 | 100 | 135 | 1 | 135 | 1 | 135 | 0 | 4.21528e-89 |
| ☑ | QJD54129 | No | 2021-12-02 | hypothetical protein SEA_SLATT_41 [Mycobacterium phage Slatt] >gb\|QWY79759.1\| hypothetical protein SEA_BURR_40 [Mycobacterium phage | 99.2593 | 100 | 100 | 135 | 1 | 135 | 1 | 135 | 0 | 6.46735e-89 |

### c. SIF: HHPred

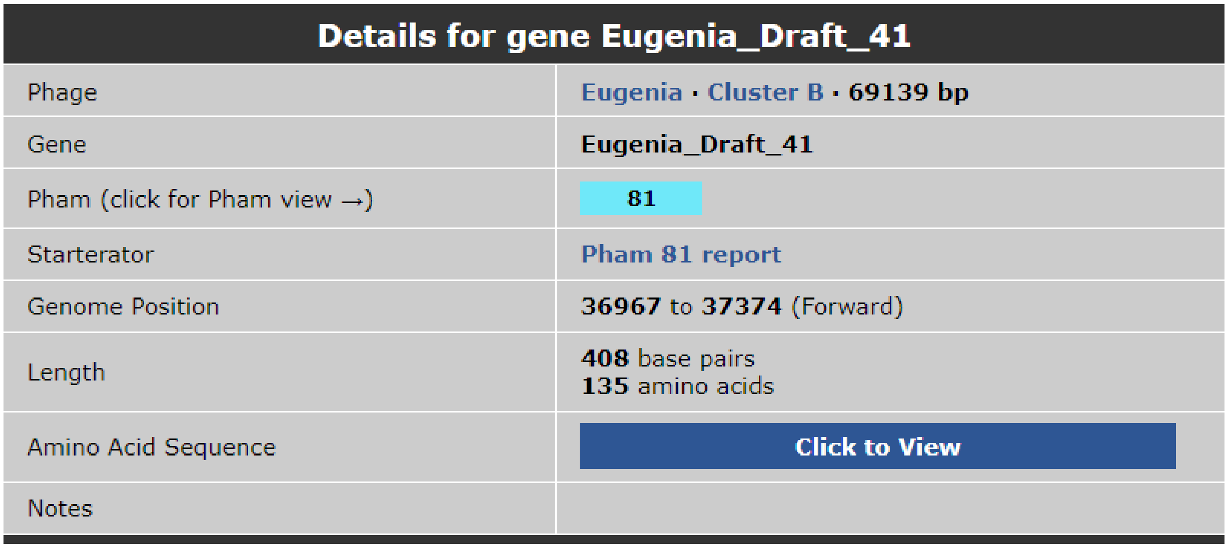
No Evidence from HHPred

### d. SIF: Synteny-Phamerator (three genomes)





# 7. Any other important information.



| Details for gene Eugenia_Draft_41 | |
|---|---|
| Phage | Eugenia • Cluster B • 69139 bp |
| Gene | Eugenia_Draft_41 |
| Pham (click for Pham view →) | 81 |
| Starterator | Pham 81 report |
| Genome Position | 36967 to 37374 (Forward) |
| Length | 408 base pairs<br>135 amino acids |
| Amino Acid Sequence | Click to View |
| Notes | |

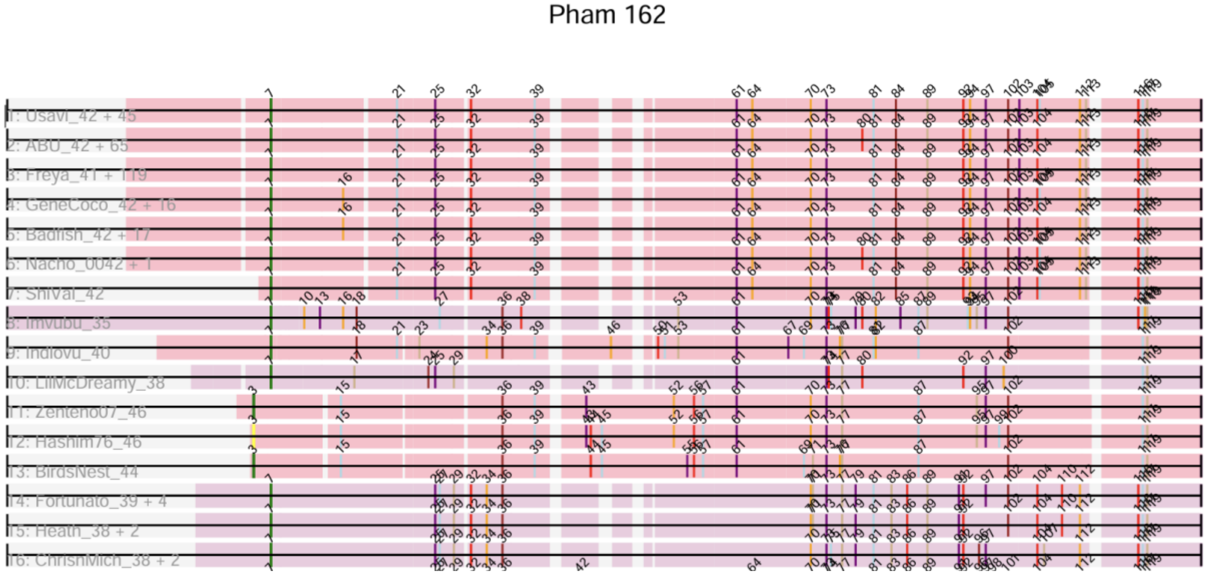
**CURATOR NAME: OLIVIA SIDOTI**

**GENE NAME: EUGENIA GENE 41**

**DNA MASTER NOTES: N/A**

**START POSITION EVALUATION (IN ORDER)**:

1. Starterator



Pham number 162 has 331 members, 43 are drafts.

The start number called the most often in the published annotations is 7, it was called in 282 of the 288 non-draft genes in the pham.

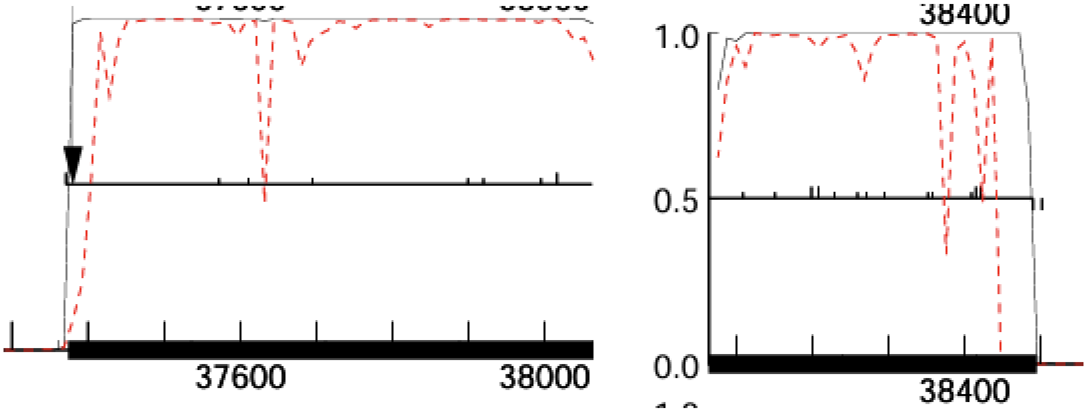
Start 7:

- Found in 318 of 331 ( 96.1% ) of genes in pham
- Manual Annotations of this start: 282 of 288
- Called 100.0% of time when present

2. GeneMark coding potential

Start: 37,374

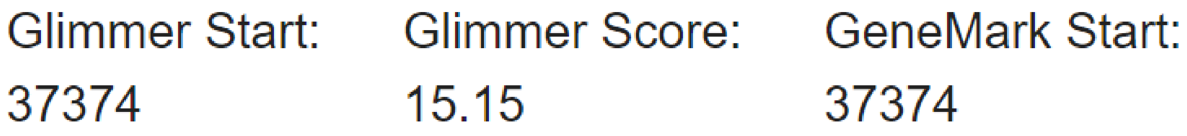
Stop: 38,495



Black Triangle:

3. Glimmer and GeneMark agreement

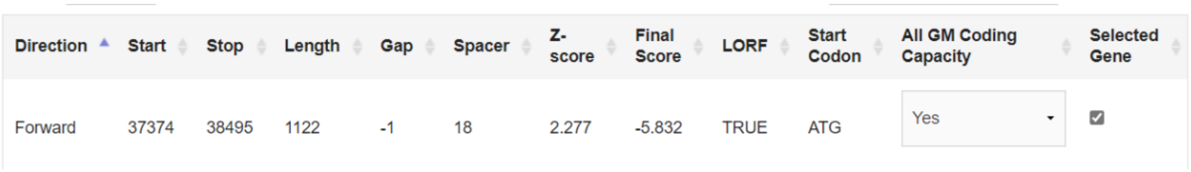
Phage: Eugenia Cluster: B1



| Glimmer Start: | Glimmer Score: | GeneMark Start: |
|---|---|---|
| 37374 | 15.15 | 37374 |

Yes, Glimmer and GeneMark are in agreement that the start position is at 37374.

4. Longest open reading frame (ORF) without excessive gap



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Forward | 37374 | 38495 | 1122 | -1 | 18 | 2.277 | -5.832 | TRUE | ATG | Yes | ☑ |

Yes, this gene is the longest open reading frame.

Gap: -1

Spacer: 18

Z-Score: 2.277

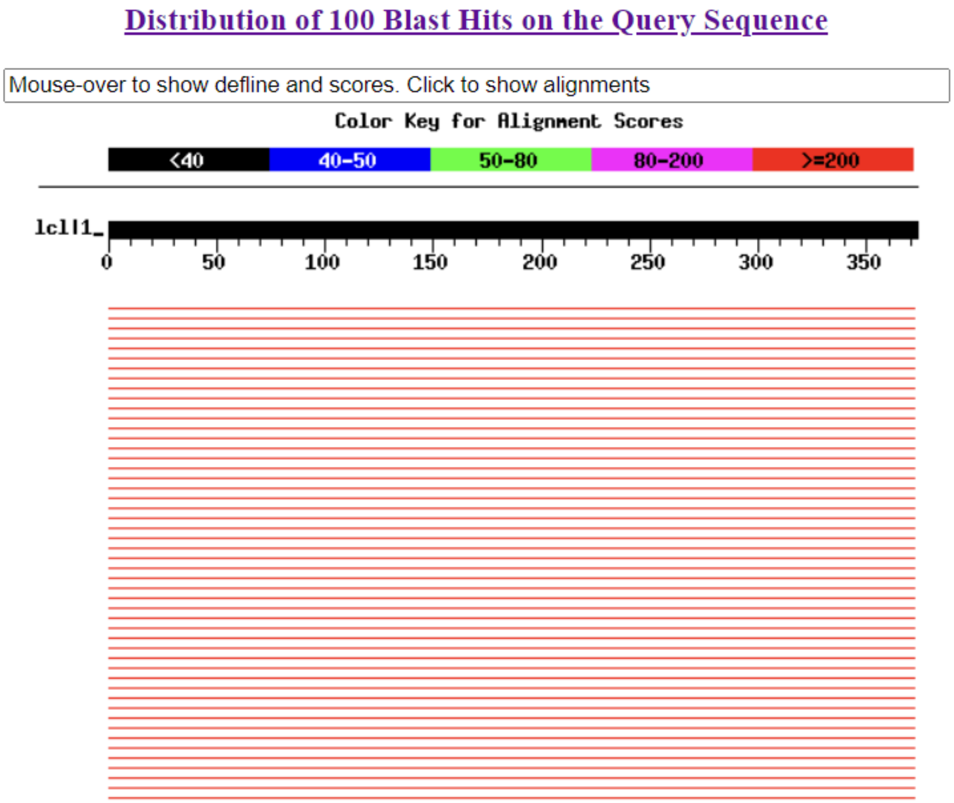
Final Score: -5.832

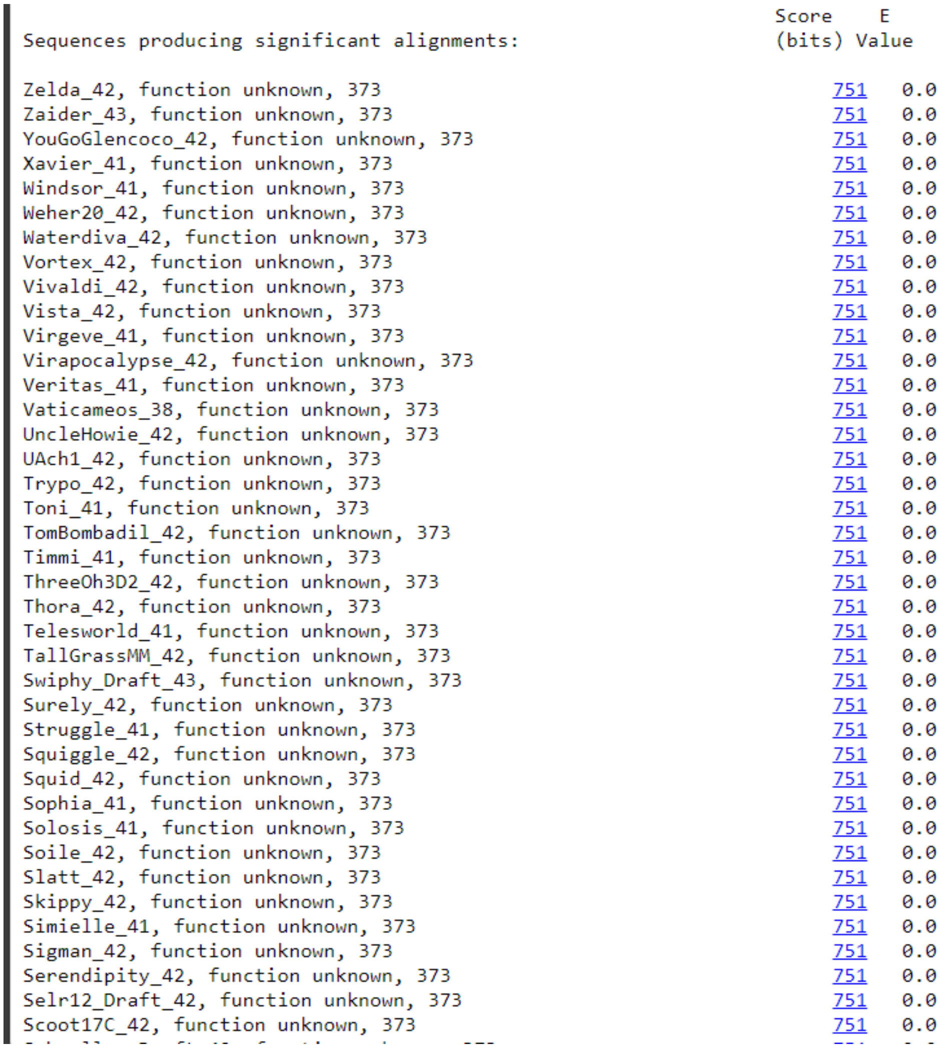
5. Function. If no functional prediction is present, write "Hypothetical protein".

Hypothetical Protein

6. Supporting Information for Function (SIF)

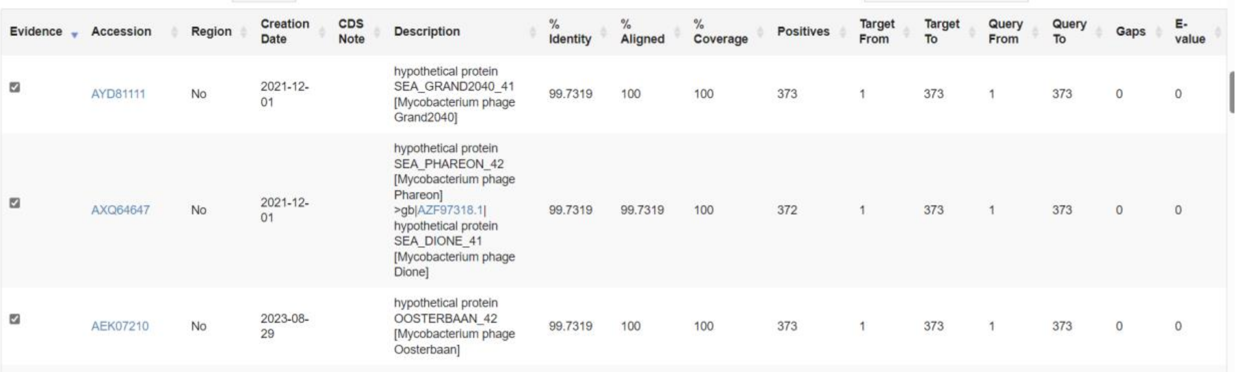
a. SIF: PhageDb BLAST





```
                                                                 Score     E
Sequences producing significant alignments:                      (bits)  Value

Zelda_42, function unknown, 373                                      751    0.0
Zaider_43, function unknown, 373                                     751    0.0
YouGoGlencoco_42, function unknown, 373                              751    0.0
Xavier_41, function unknown, 373                                     751    0.0
Windsor_41, function unknown, 373                                    751    0.0
Weher20_42, function unknown, 373                                    751    0.0
Waterdiva_42, function unknown, 373                                  751    0.0
Vortex_42, function unknown, 373                                     751    0.0
Vivaldi_42, function unknown, 373                                    751    0.0
Vista_42, function unknown, 373                                      751    0.0
Virgeve_41, function unknown, 373                                    751    0.0
Virapocalypse_42, function unknown, 373                              751    0.0
Veritas_41, function unknown, 373                                    751    0.0
Vaticameos_38, function unknown, 373                                 751    0.0
UncleHowie_42, function unknown, 373                                 751    0.0
UAch1_42, function unknown, 373                                      751    0.0
Trypo_42, function unknown, 373                                      751    0.0
Toni_41, function unknown, 373                                       751    0.0
TomBombadil_42, function unknown, 373                                751    0.0
Timmi_41, function unknown, 373                                      751    0.0
ThreeOh3D2_42, function unknown, 373                                 751    0.0
Thora_42, function unknown, 373                                      751    0.0
Telesworld_41, function unknown, 373                                 751    0.0
TallGrassMM_42, function unknown, 373                                751    0.0
Swiphy_Draft_43, function unknown, 373                               751    0.0
Surely_42, function unknown, 373                                     751    0.0
Struggle_41, function unknown, 373                                   751    0.0
Squiggle_42, function unknown, 373                                   751    0.0
Squid_42, function unknown, 373                                      751    0.0
Sophia_41, function unknown, 373                                     751    0.0
Solosis_41, function unknown, 373                                    751    0.0
Soile_42, function unknown, 373                                      751    0.0
Slatt_42, function unknown, 373                                      751    0.0
Skippy_42, function unknown, 373                                     751    0.0
Simielle_41, function unknown, 373                                   751    0.0
Sigman_42, function unknown, 373                                     751    0.0
Serendipity_42, function unknown, 373                                751    0.0
Selr12_Draft_42, function unknown, 373                               751    0.0
Scoot17C_42, function unknown, 373                                   751    0.0
```

b. SIF: NCBI BLAST

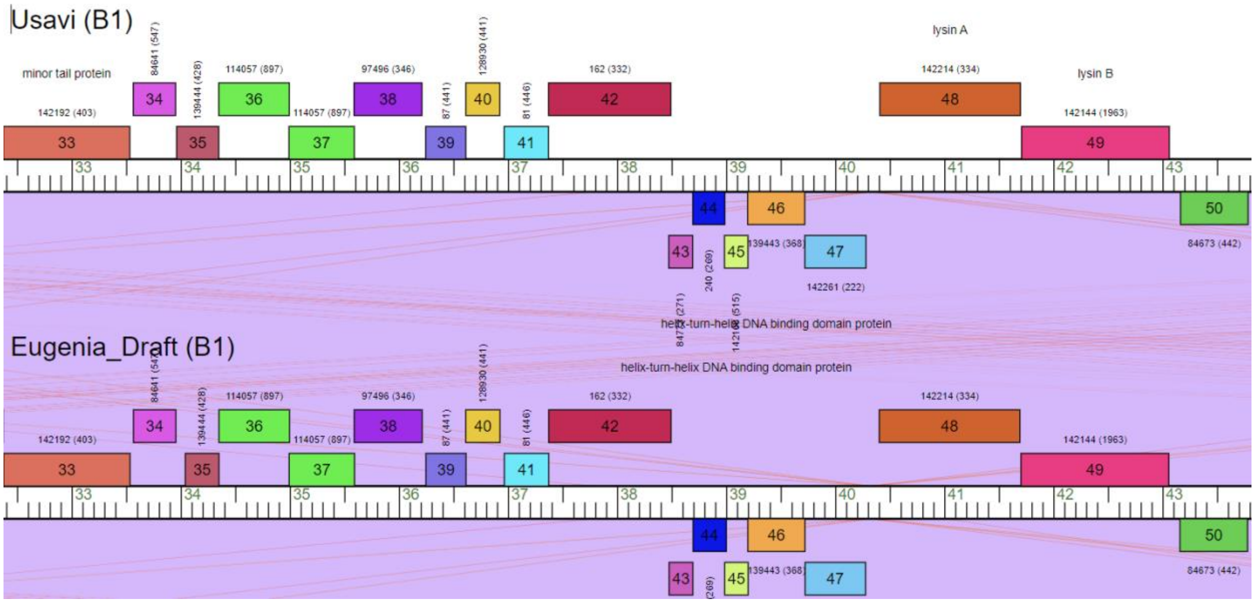


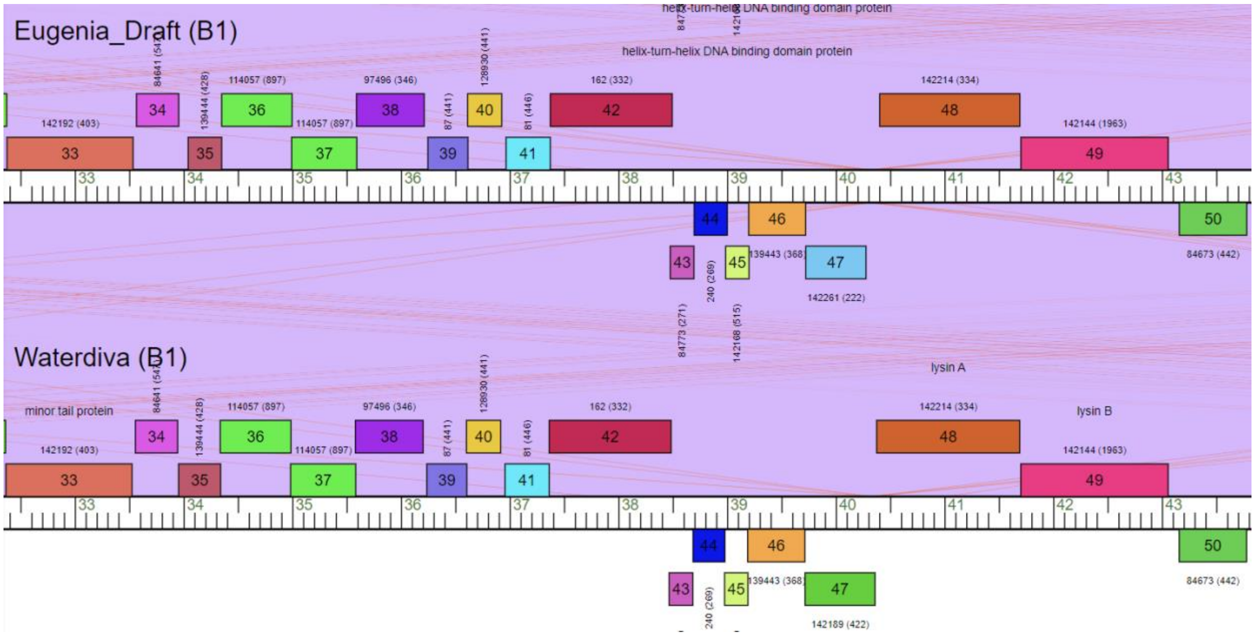
| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | AYD81111 | No | 2021-12-01 | | hypothetical protein SEA_GRAND2040_41 [Mycobacterium phage Grand2040] | 99.7319 | 100 | 100 | 373 | 1 | 373 | 1 | 373 | 0 | 0 |
| ☑ | AXQ64647 | No | 2021-12-01 | | hypothetical protein SEA_PHAREON_42 [Mycobacterium phage Phareon] >gb\|AZF97318.1\| hypothetical protein SEA_DIONE_41 [Mycobacterium phage Dione] | 99.7319 | 99.7319 | 100 | 372 | 1 | 373 | 1 | 373 | 0 | 0 |
| ☑ | AEK07210 | No | 2023-08-29 | | hypothetical protein OOSTERBAAN_42 [Mycobacterium phage Oosterbaan] | 99.7319 | 100 | 100 | 373 | 1 | 373 | 1 | 373 | 0 | 0 |

c. SIF: HHPred

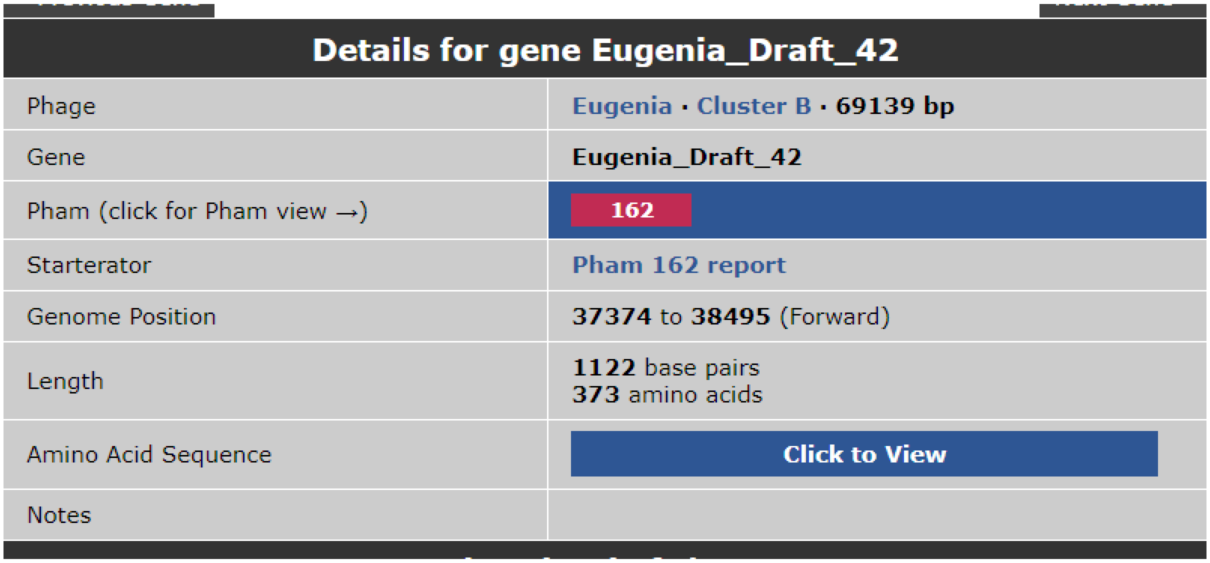
No Evidence from HHPred

d. SIF: Synteny-Phamerator (three genomes)





7. Any other important information.



| **Details for gene Eugenia_Draft_42** | |
|---|---|
| Phage | Eugenia · Cluster B · **69139 bp** |
| Gene | **Eugenia_Draft_42** |
| Pham (click for Pham view →) | **162** |
| Starterator | **Pham 162 report** |
| Genome Position | **37374** to **38495** (Forward) |
| Length | **1122** base pairs<br>**373** amino acids |
| Amino Acid Sequence | **Click to View** |
| Notes | |

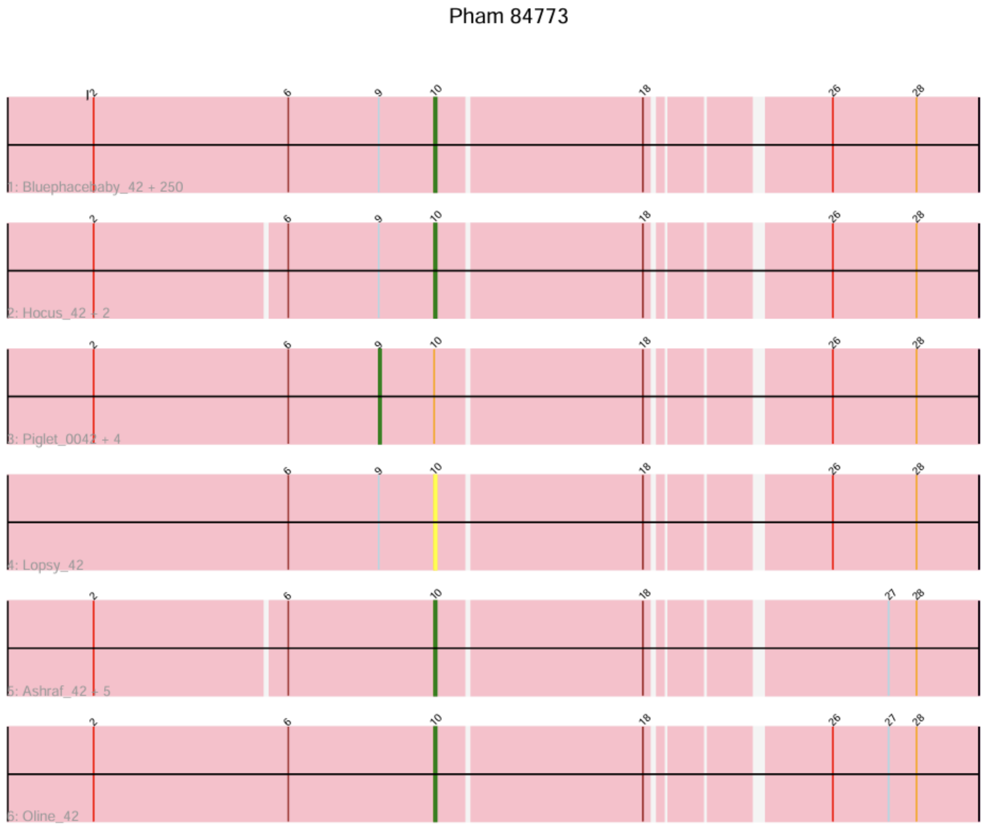
**CURATOR NAME: OLIVIA SIDOTI**

**GENE NAME: EUGENIA GENE 43**

**DNA MASTER NOTES: N/A**

**START POSITION EVALUATION (IN ORDER):**

1. Starterator



Pham number 84773 has 270 members, 24 are drafts.

The start number called the most often in the published annotations is 10, it was called in 238 of the 246 non-draft genes in the pham.

Start 10:

- Found in 267 of 270 ( 98.9% ) of genes in pham

- Manual Annotations of this start: 238 of 246

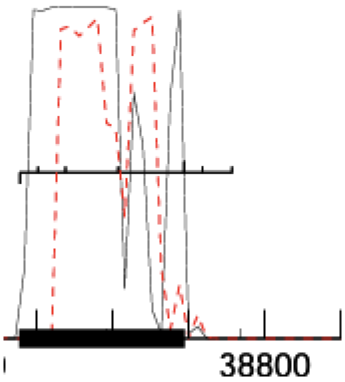
- Called 98.1% of time when present

2. GeneMark coding potential

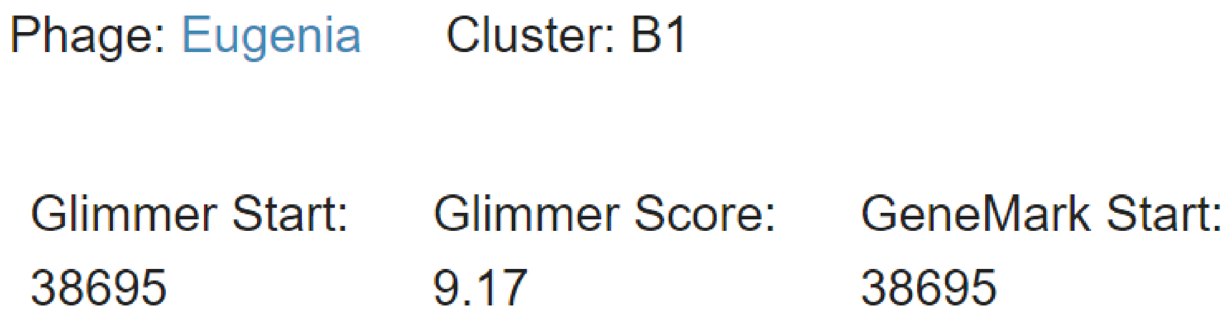
Start: 38,695

Stop: 38,477

Reverse



3. Glimmer and GeneMark agreement



Phage: Eugenia Cluster: B1

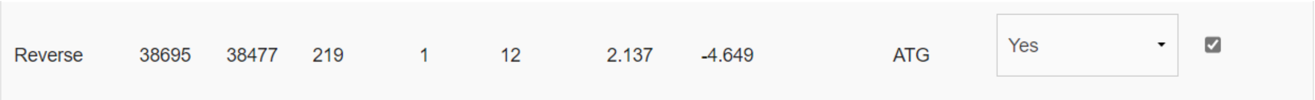
Glimmer Start: 38695

Glimmer Score: 9.17

GeneMark Start: 38695

Yes, Glimmer and GeneMark are in agreement that the start position is at

4. Longest open reading frame (ORF) without excessive gap



| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 38695 | 38477 | 219 | 1 | 12 | 2.137 | -4.649 | ATG | Yes ▾ | ☑ |

No, this gene is not the longest open reading frame.

Gap: 1

Spacer: 12

Z-Score: 2.137

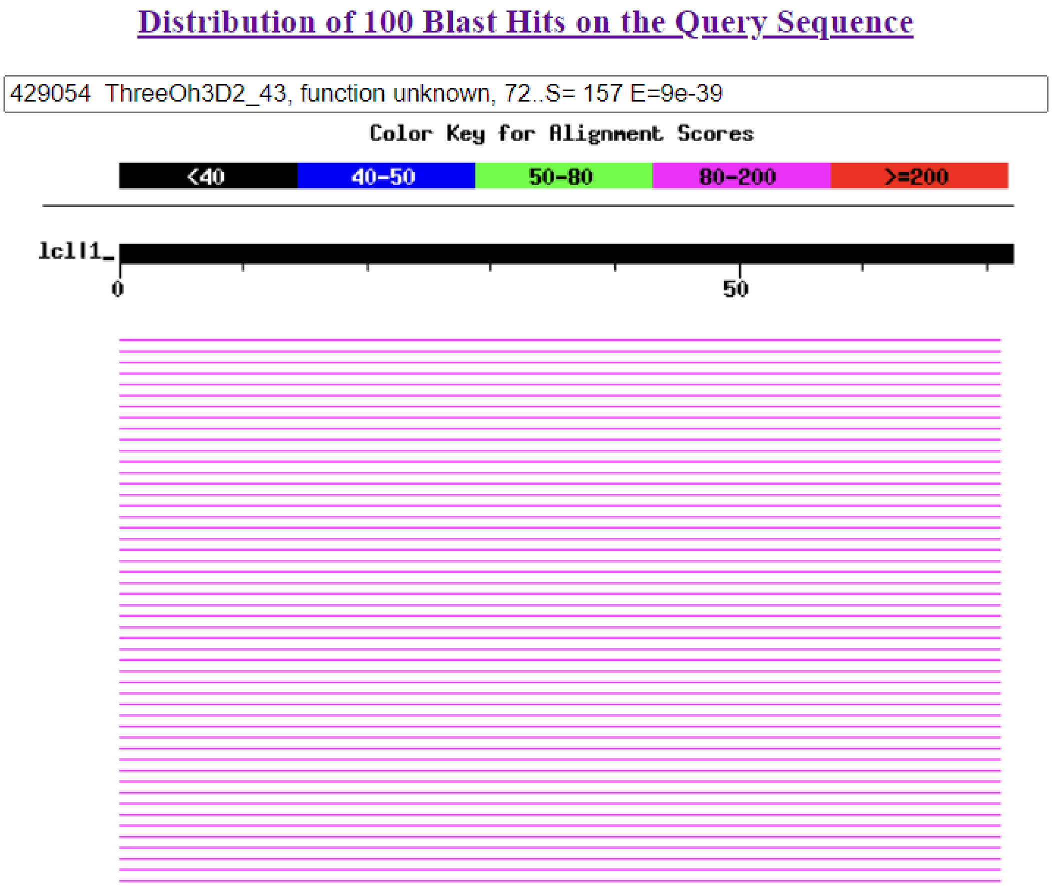
Final Score: -4.649

5. Function. If no functional prediction is present, write “Hypothetical protein”.

Hypothetical Protein

6. Supporting Information for Function (SIF)

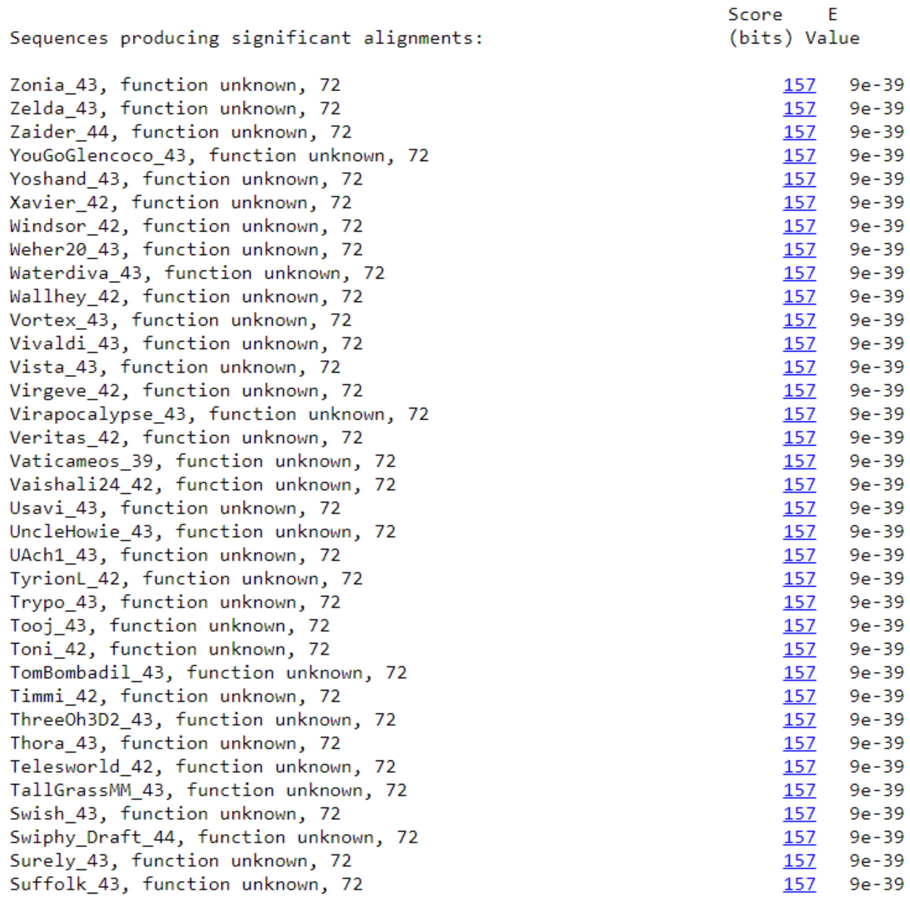
a. SIF: PhageDb BLAST



Distribution of 100 Blast Hits on the Query Sequence

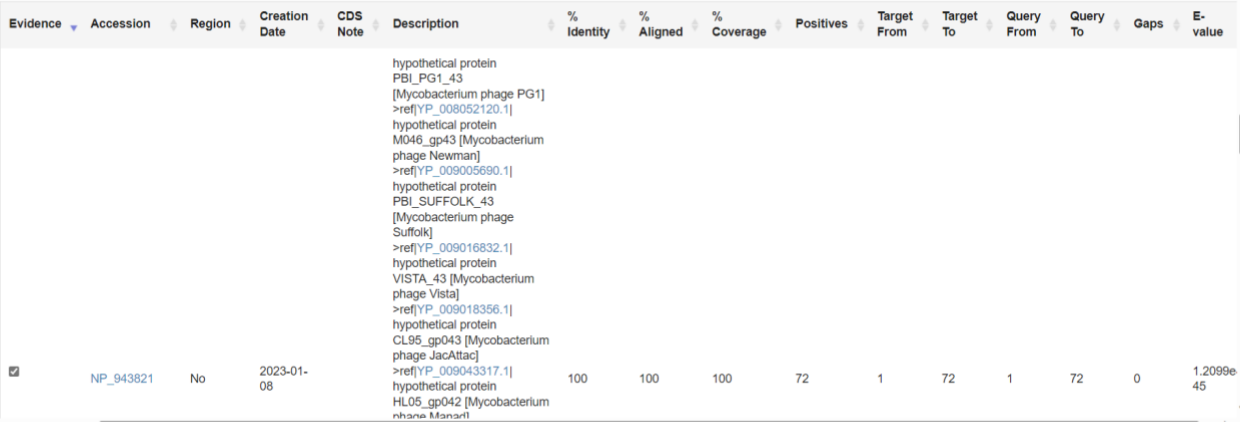
429054 ThreeOh3D2_43, function unknown, 72..S= 157 E=9e-39

80-200 % alignment



```
                                                                 Score     E
Sequences producing significant alignments:                      (bits)  Value

Zonia_43, function unknown, 72                                        157    9e-39
Zelda_43, function unknown, 72                                        157    9e-39
Zaider_44, function unknown, 72                                       157    9e-39
YouGoGlencoco_43, function unknown, 72                                157    9e-39
Yoshand_43, function unknown, 72                                      157    9e-39
Xavier_42, function unknown, 72                                       157    9e-39
Windsor_42, function unknown, 72                                      157    9e-39
Weher20_43, function unknown, 72                                      157    9e-39
Waterdiva_43, function unknown, 72                                    157    9e-39
Wallhey_42, function unknown, 72                                      157    9e-39
Vortex_43, function unknown, 72                                       157    9e-39
Vivaldi_43, function unknown, 72                                      157    9e-39
Vista_43, function unknown, 72                                        157    9e-39
Virgeve_42, function unknown, 72                                      157    9e-39
Virapocalypse_43, function unknown, 72                                157    9e-39
Veritas_42, function unknown, 72                                      157    9e-39
Vaticameos_39, function unknown, 72                                   157    9e-39
Vaishali24_42, function unknown, 72                                   157    9e-39
Usavi_43, function unknown, 72                                        157    9e-39
UncleHowie_43, function unknown, 72                                   157    9e-39
UAch1_43, function unknown, 72                                        157    9e-39
TyrionL_42, function unknown, 72                                      157    9e-39
Trypo_43, function unknown, 72                                        157    9e-39
Tooj_43, function unknown, 72                                         157    9e-39
Toni_42, function unknown, 72                                         157    9e-39
TomBombadil_43, function unknown, 72                                  157    9e-39
Timmi_42, function unknown, 72                                        157    9e-39
ThreeOh3D2_43, function unknown, 72                                   157    9e-39
Thora_43, function unknown, 72                                        157    9e-39
Telesworld_42, function unknown, 72                                   157    9e-39
TallGrassMM_43, function unknown, 72                                  157    9e-39
Swish_43, function unknown, 72                                        157    9e-39
Swiphy_Draft_44, function unknown, 72                                 157    9e-39
Surely_43, function unknown, 72                                       157    9e-39
Suffolk_43, function unknown, 72                                      157    9e-39
```

b. SIF: NCBI BLAST

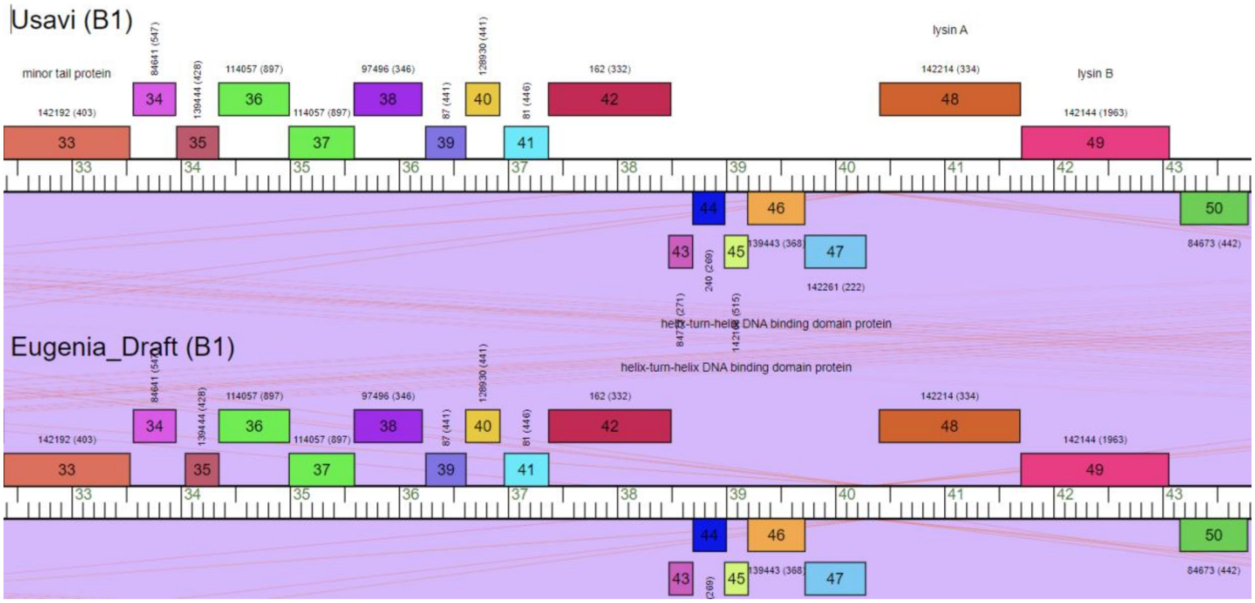


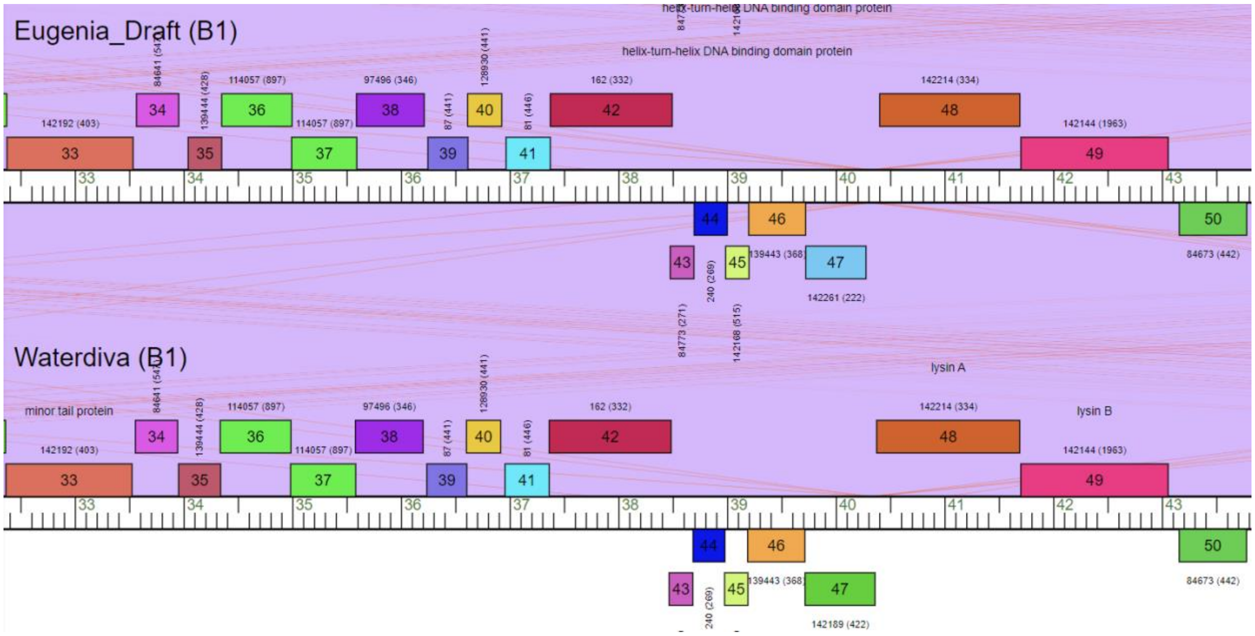
| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | NP_943821 | No | 2023-01-08 |  | hypothetical protein PBI_PG1_43 [Mycobacterium phage PG1] >ref\|YP_008052120.1\| hypothetical protein M046_gp43 [Mycobacterium phage Newman] >ref\|YP_009005690.1\| hypothetical protein PBI_SUFFOLK_43 [Mycobacterium phage Suffolk] >ref\|YP_009016832.1\| hypothetical protein VISTA_43 [Mycobacterium phage Vista] >ref\|YP_009018356.1\| hypothetical protein CL95_gp043 [Mycobacterium phage JacAttac] >ref\|YP_009043317.1\| hypothetical protein HL05_gp042 [Mycobacterium phage Manad] | 100 | 100 | 100 | 72 | 1 | 72 | 1 | 72 | 0 | 1.2099e-45 |

## c. SIF: HHPred

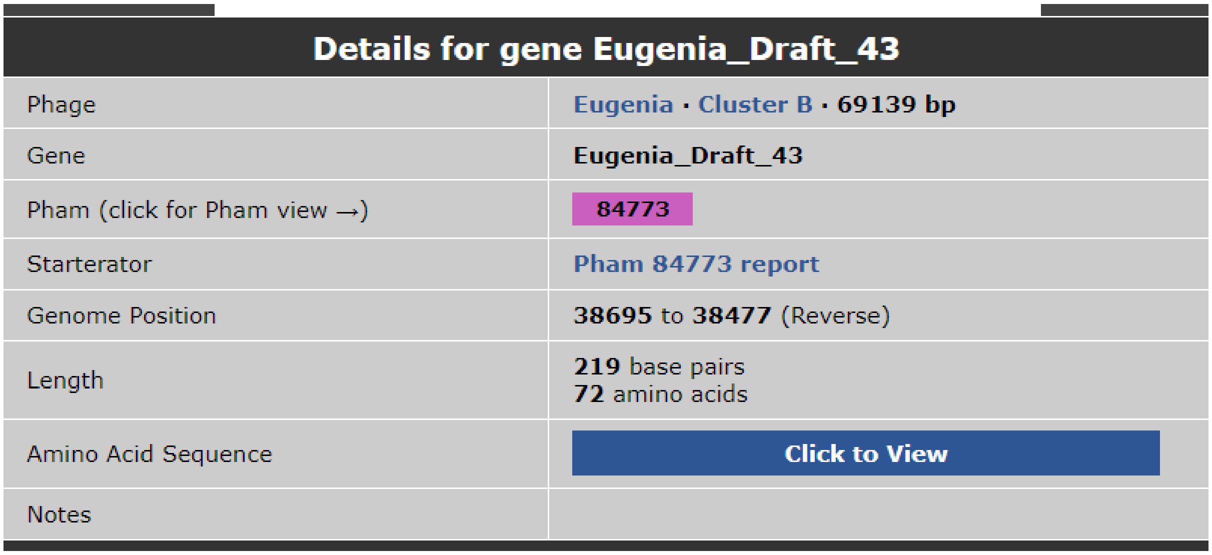
No Evidence from HHPred

## d. SIF: Synteny-Phamerator (three genomes)





7. Any other important information.



| Details for gene Eugenia_Draft_43 | |
|---|---|
| Phage | Eugenia • Cluster B • 69139 bp |
| Gene | **Eugenia_Draft_43** |
| Pham (click for Pham view →) | **84773** |
| Starterator | Pham 84773 report |
| Genome Position | **38695** to **38477** (Reverse) |
| Length | **219** base pairs<br>**72** amino acids |
| Amino Acid Sequence | Click to View |
| Notes | |

**CURATOR NAME: OLIVIA SIDOTI**

**GENE NAME: EUGENIA GENE 44**

**DNA MASTER NOTES: N/A**

**START POSITION EVALUATION (IN ORDER):**

1. Starterator

Pham number 84773 has 270 members, 24 are drafts

The start number called the most often in the published annotations is 10, it was called in 238 of the 246 non-draft genes in the pham.

Start 10:

- Found in 267 of 270 ( 98.9% ) of genes in pham

- Manual Annotations of this start: 238 of 246

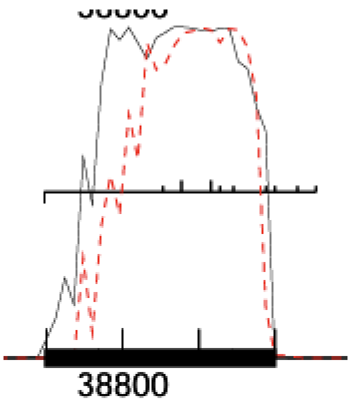
- Called 98.1% of time when present

2. GeneMark coding potential

Start: 39,002

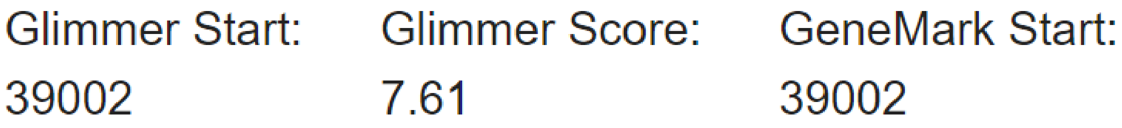
Stop: 38,697

(Reverse)



3. Glimmer and GeneMark agreement

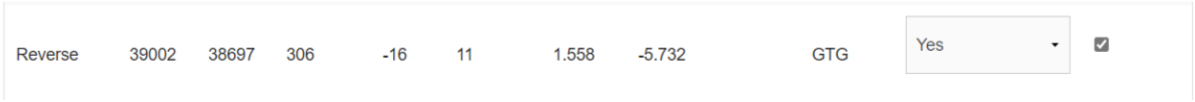
Phage: Eugenia Cluster: B1



| Glimmer Start: | Glimmer Score: | GeneMark Start: |
|---|---|---|
| 39002 | 7.61 | 39002 |

Yes, Glimmer and GeneMark are in agreement that the start position is at 39002

## 4. Longest open reading frame (ORF) without excessive gap



| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 39002 | 38697 | 306 | -16 | 11 | 1.558 | -5.732 | GTG | Yes | ☑ |

No, this gene is not the longest open reading frame.

Gap: -16

Spacer: 11

Z-Score: 1.558

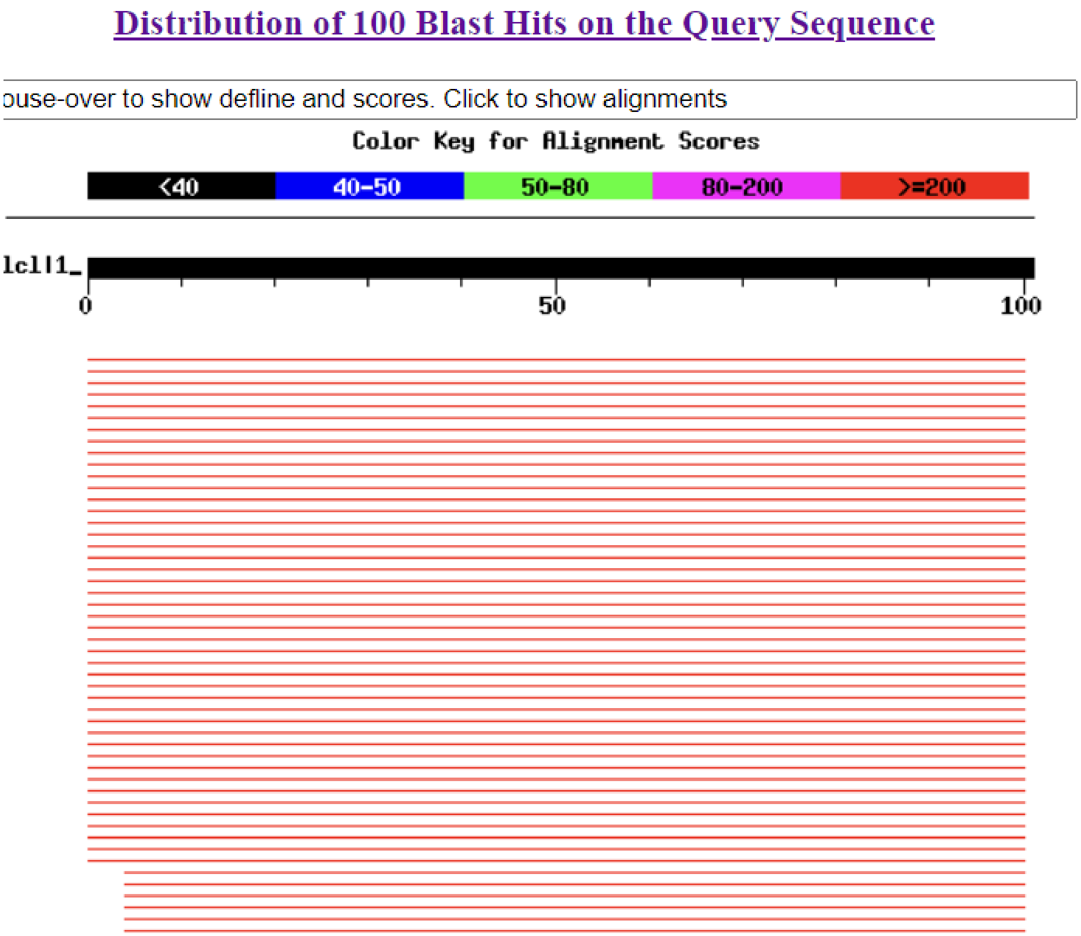
Final Score: -5.732

## 5. Function. If no functional prediction is present, write “Hypothetical protein”.

Membrane protein

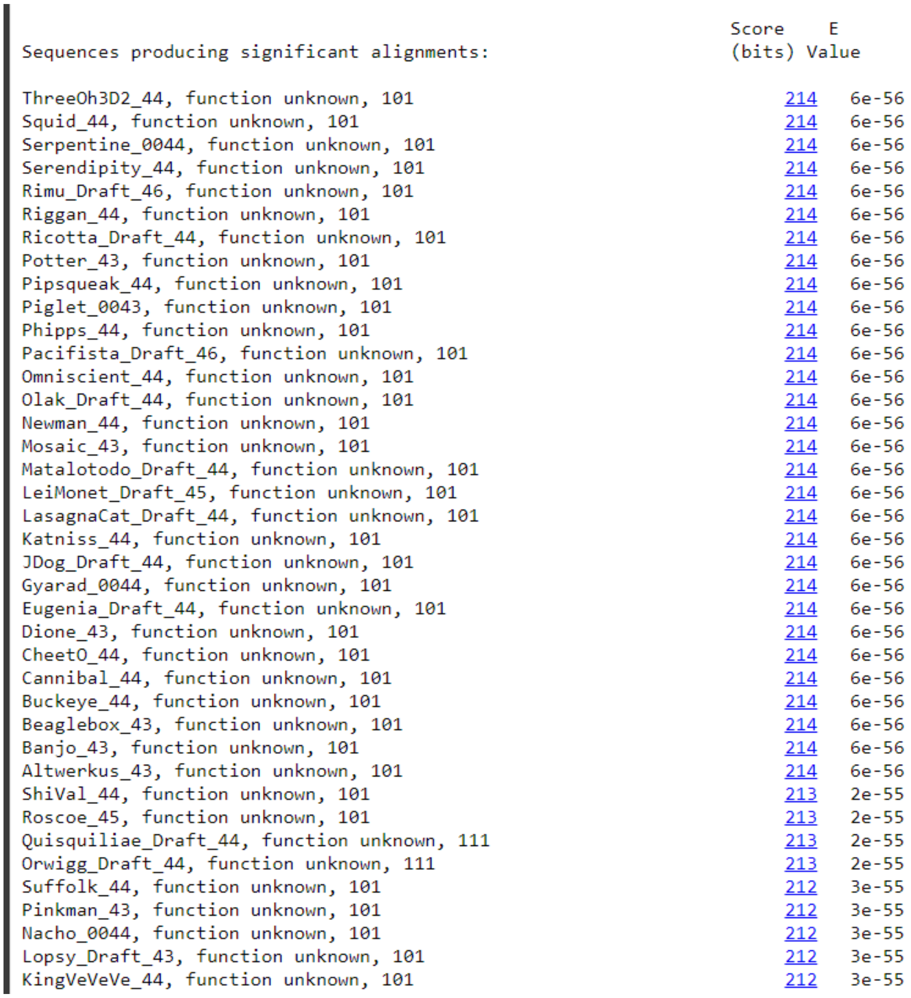
## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST



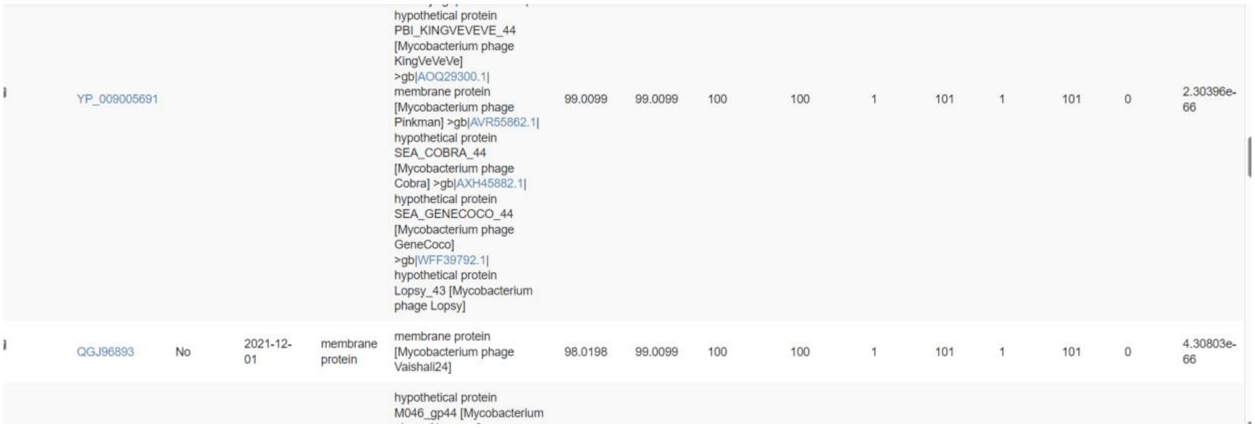
Distribution of 100 Blast Hits on the Query Sequence

ouse-over to show defline and scores. Click to show alignments



```
                                                                   Score     E
Sequences producing significant alignments:                        (bits)  Value

ThreeOh3D2_44, function unknown, 101                                    214    6e-56
Squid_44, function unknown, 101                                         214    6e-56
Serpentine_0044, function unknown, 101                                  214    6e-56
Serendipity_44, function unknown, 101                                   214    6e-56
Rimu_Draft_46, function unknown, 101                                    214    6e-56
Riggan_44, function unknown, 101                                        214    6e-56
Ricotta_Draft_44, function unknown, 101                                 214    6e-56
Potter_43, function unknown, 101                                        214    6e-56
Pipsqueak_44, function unknown, 101                                     214    6e-56
Piglet_0043, function unknown, 101                                      214    6e-56
Phipps_44, function unknown, 101                                        214    6e-56
Pacifista_Draft_46, function unknown, 101                               214    6e-56
Omniscient_44, function unknown, 101                                    214    6e-56
Olak_Draft_44, function unknown, 101                                    214    6e-56
Newman_44, function unknown, 101                                        214    6e-56
Mosaic_43, function unknown, 101                                        214    6e-56
Matalotodo_Draft_44, function unknown, 101                              214    6e-56
LeiMonet_Draft_45, function unknown, 101                                214    6e-56
LasagnaCat_Draft_44, function unknown, 101                              214    6e-56
Katniss_44, function unknown, 101                                       214    6e-56
JDog_Draft_44, function unknown, 101                                    214    6e-56
Gyarad_0044, function unknown, 101                                      214    6e-56
Eugenia_Draft_44, function unknown, 101                                 214    6e-56
Dione_43, function unknown, 101                                         214    6e-56
CheetO_44, function unknown, 101                                        214    6e-56
Cannibal_44, function unknown, 101                                      214    6e-56
Buckeye_44, function unknown, 101                                       214    6e-56
Beaglebox_43, function unknown, 101                                     214    6e-56
Banjo_43, function unknown, 101                                         214    6e-56
Altwerkus_43, function unknown, 101                                     214    6e-56
ShiVal_44, function unknown, 101                                        213    2e-55
Roscoe_45, function unknown, 101                                        213    2e-55
Quisquiliae_Draft_44, function unknown, 111                             213    2e-55
Orwigg_Draft_44, function unknown, 111                                  213    2e-55
Suffolk_44, function unknown, 101                                       212    3e-55
Pinkman_43, function unknown, 101                                       212    3e-55
Nacho_0044, function unknown, 101                                       212    3e-55
Lopsy_Draft_43, function unknown, 101                                   212    3e-55
KingVeVeVe_44, function unknown, 101                                    212    3e-55
```

b. SIF: NCBI BLAST

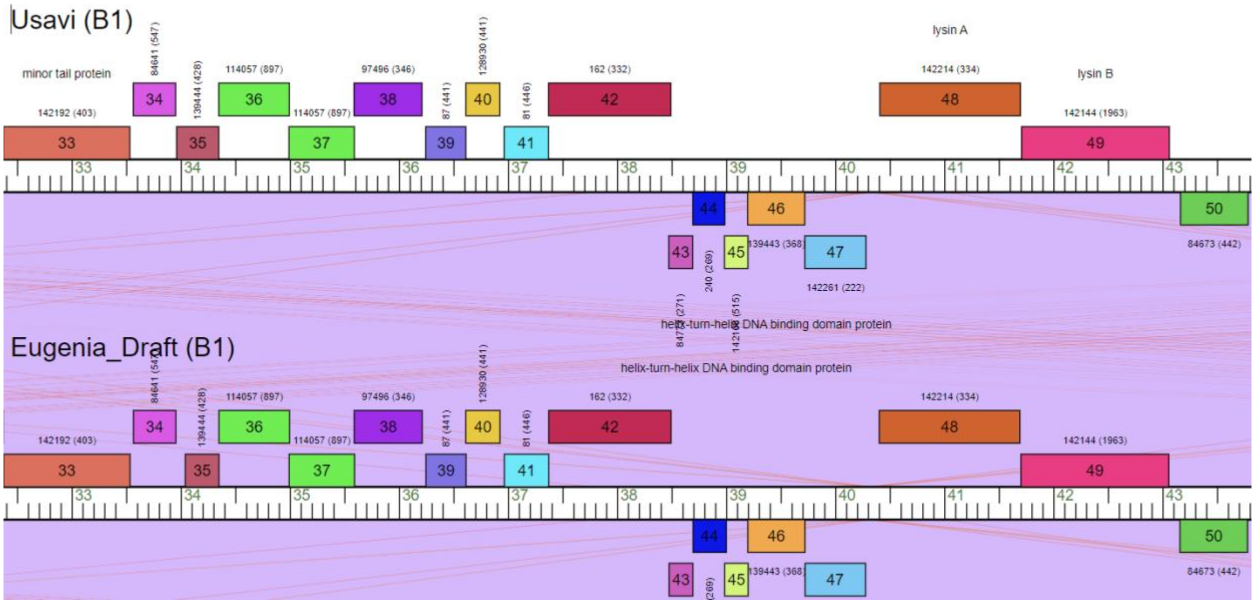


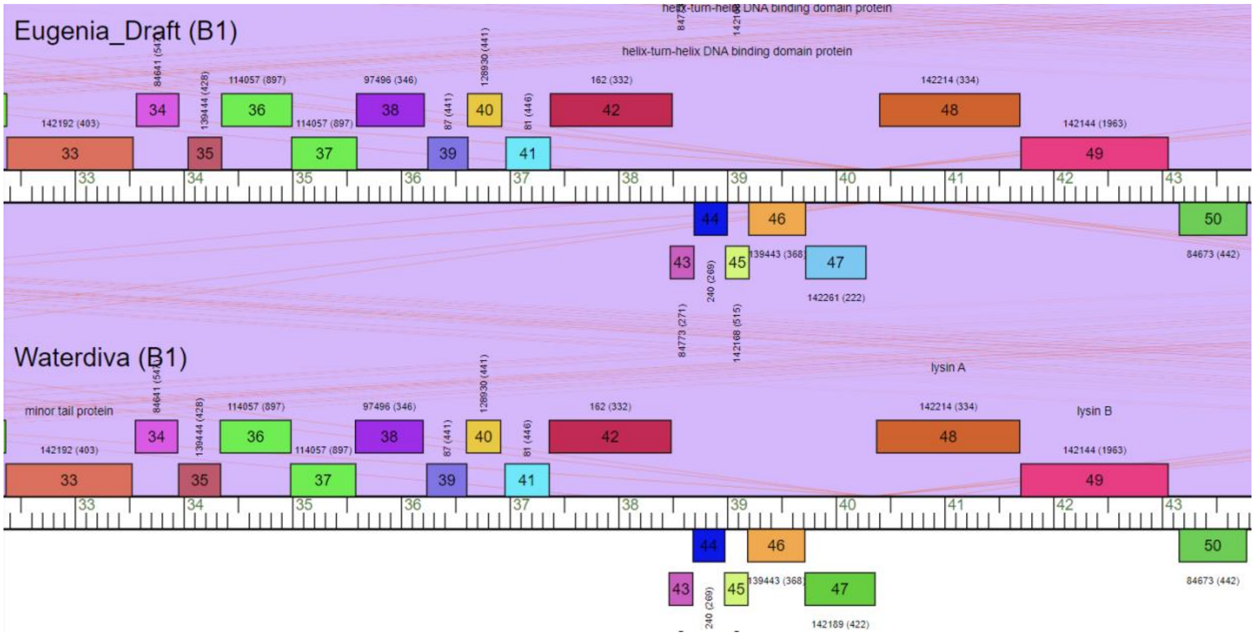
| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| YP_009005691 | | | | hypothetical protein PBI_KINGVEVEVE_44 [Mycobacterium phage KingVeVeVe] >gb\|AOQ29300.1\| membrane protein [Mycobacterium phage Pinkman] >gb\|AVR55862.1\| hypothetical protein SEA_COBRA_44 [Mycobacterium phage Cobra] >gb\|AXH45882.1\| hypothetical protein SEA_GENECOCO_44 [Mycobacterium phage GeneCoco] >gb\|WFF39792.1\| hypothetical protein Lopsy_43 [Mycobacterium phage Lopsy] | 99.0099 | 99.0099 | 100 | 100 | 1 | 101 | 1 | 101 | 0 | 2.30396e-66 |
| QGJ96893 | No | 2021-12-01 | membrane protein | membrane protein [Mycobacterium phage Vaishali24] | 98.0198 | 99.0099 | 100 | 100 | 1 | 101 | 1 | 101 | 0 | 4.30803e-66 |
| | | | | hypothetical protein M046_gp44 [Mycobacterium | | | | | | | | | | |

c. SIF: HHPred

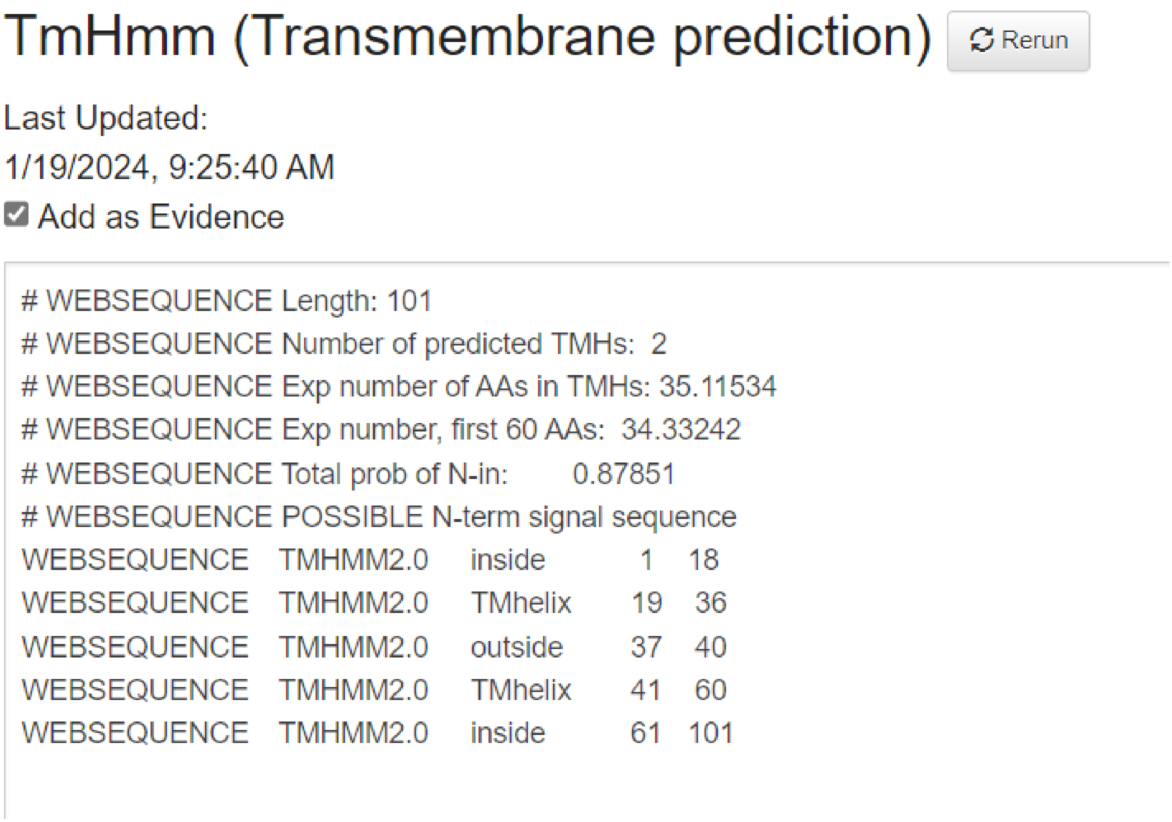
No Evidence from HHPred

d. SIF: Synteny-Phamerator (three genomes)





7. Any other important information.

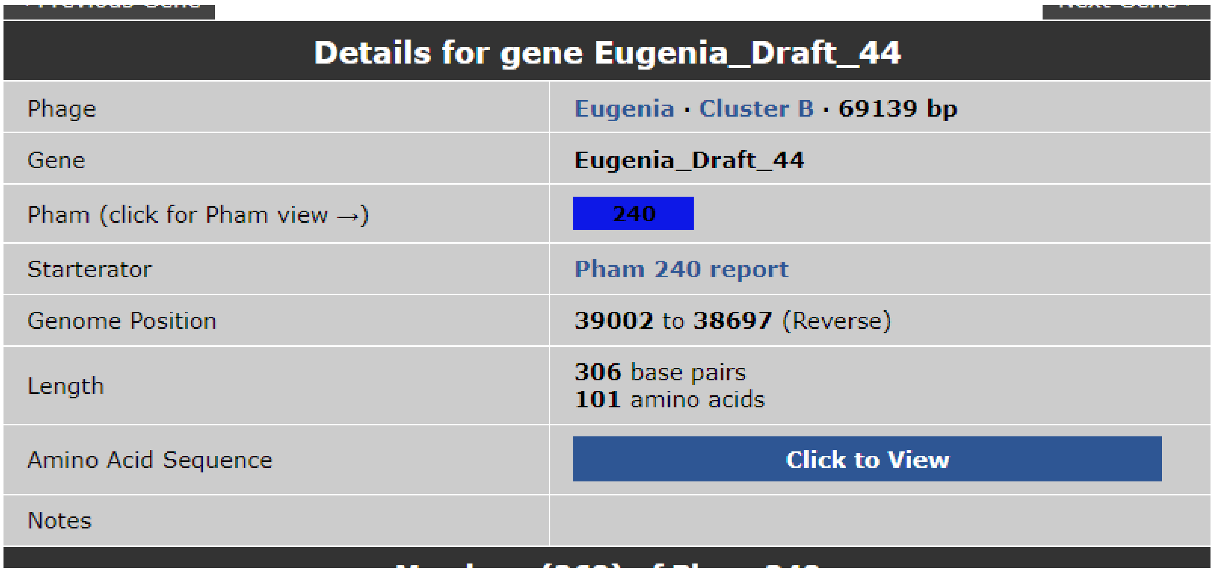


TmHmm (Transmembrane prediction)

Last Updated:
1/19/2024, 9:25:40 AM

☑ Add as Evidence

```
# WEBSEQUENCE Length: 101
# WEBSEQUENCE Number of predicted TMHs:  2
# WEBSEQUENCE Exp number of AAs in TMHs: 35.11534
# WEBSEQUENCE Exp number, first 60 AAs:  34.33242
# WEBSEQUENCE Total prob of N-in:        0.87851
# WEBSEQUENCE POSSIBLE N-term signal sequence
WEBSEQUENCE   TMHMM2.0    inside       1    18
WEBSEQUENCE   TMHMM2.0    TMhelix     19    36
WEBSEQUENCE   TMHMM2.0    outside     37    40
WEBSEQUENCE   TMHMM2.0    TMhelix     41    60
WEBSEQUENCE   TMHMM2.0    inside      61   101
```



**Details for gene Eugenia_Draft_44**

| | |
|---|---|
| Phage | Eugenia · Cluster B · **69139 bp** |
| Gene | **Eugenia_Draft_44** |
| Pham (click for Pham view →) | **240** |
| Starterator | Pham 240 report |
| Genome Position | **39002** to **38697** (Reverse) |
| Length | **306** base pairs<br>**101** amino acids |
| Amino Acid Sequence | Click to View |
| Notes | |

**CURATOR NAME: OLIVIA SIDOTI**

**GENE NAME: EUGENIA GENE 45**

**DNA MASTER NOTES: N/A**

**START POSITION EVALUATION (IN ORDER)**:

1. Starterator

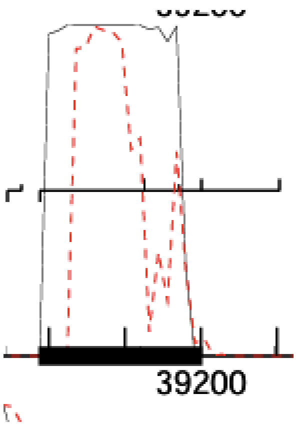
N/A

2. GeneMark coding potential

Start: 39,202

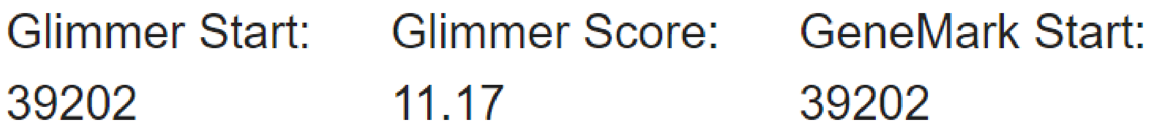
Stop: 38,987

Reverse



## 3. Glimmer and GeneMark agreement

Phage: Eugenia Cluster: B1



Glimmer Start: 39202

Glimmer Score: 11.17

GeneMark Start: 39202

Yes, Glimmer and GeneMark are in agreement that the start position is at 39202

## 4. Longest open reading frame (ORF) without excessive gap

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 39202 | 38987 | 216 | -4 | 10 | 2.241 | -4.299 | ATG | Yes | ☑ |

No, this gene is not the longest open reading frame.

Gap: -4

Spacer: 10

Z-Score: 2.241

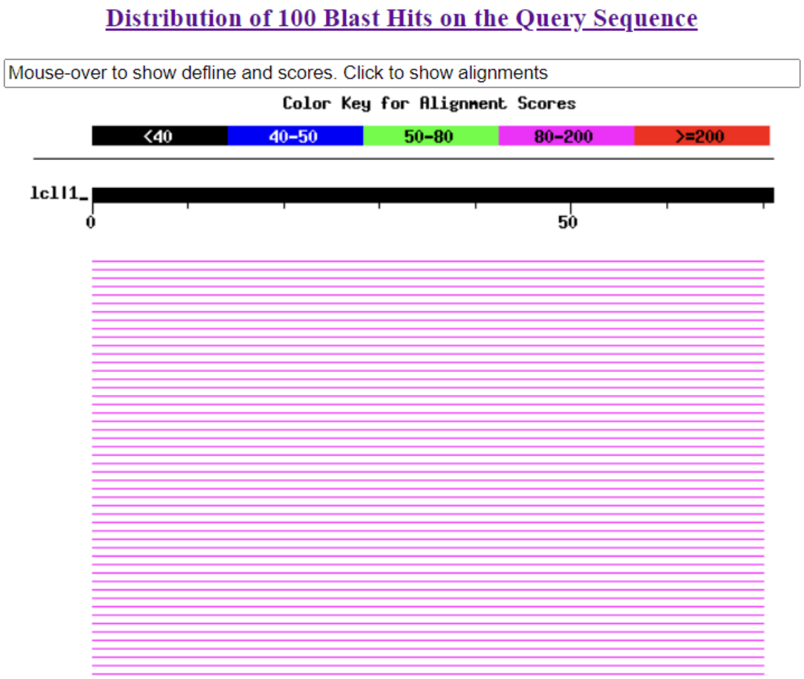
Final Score: -4.299

## 5. Function. If no functional prediction is present, write “Hypothetical protein”.

Helix-turn-helix binding protein

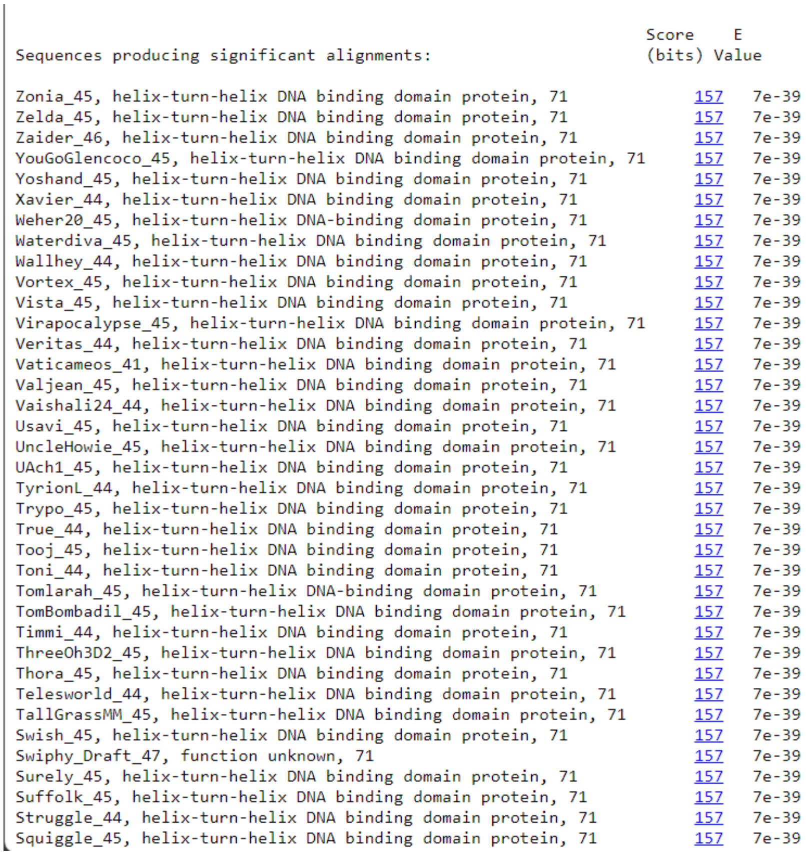
## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST



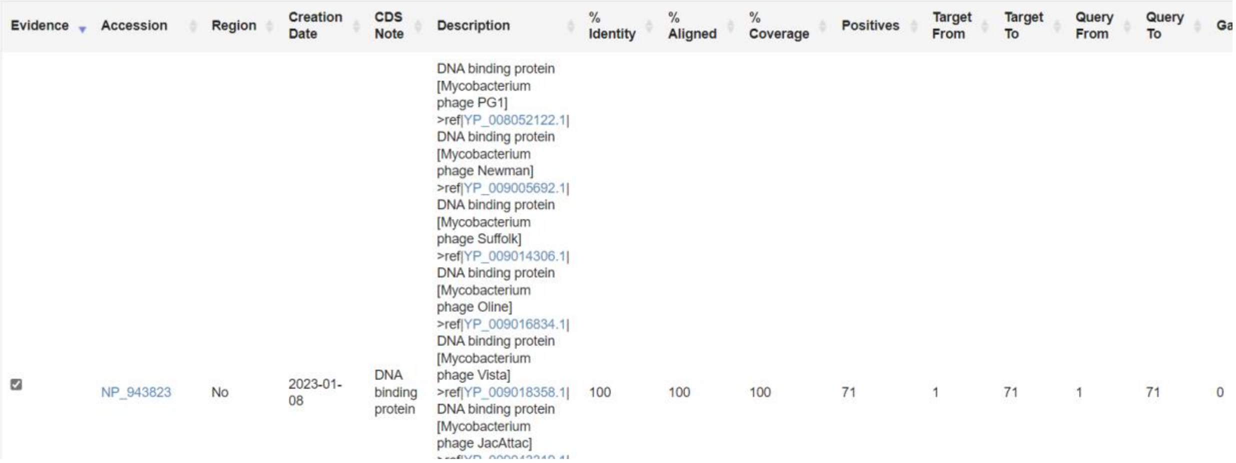
Distribution of 100 Blast Hits on the Query Sequence

Mouse-over to show defline and scores. Click to show alignments

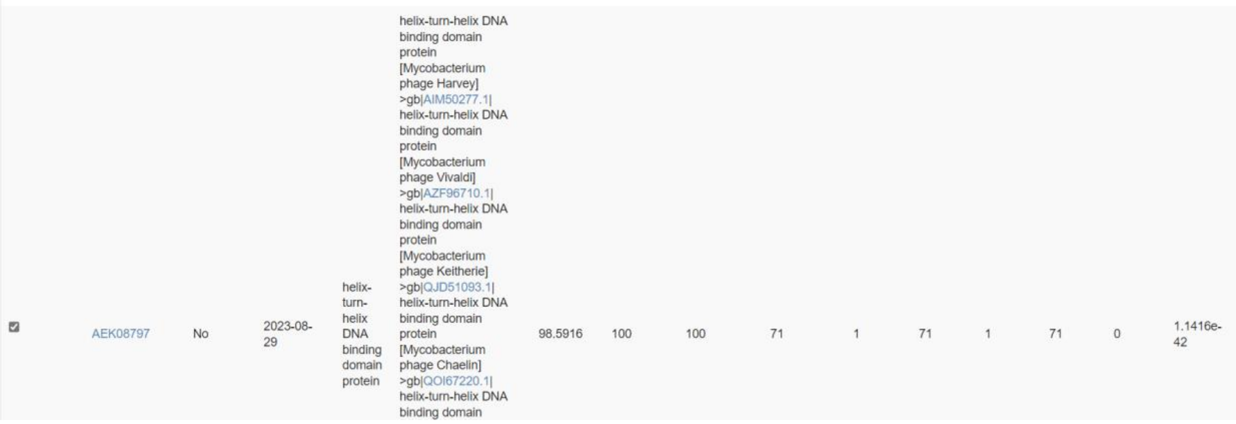


```
                                                                   Score     E
Sequences producing significant alignments:                       (bits)  Value

Zonia_45, helix-turn-helix DNA binding domain protein, 71              157    7e-39
Zelda_45, helix-turn-helix DNA binding domain protein, 71              157    7e-39
Zaider_46, helix-turn-helix DNA binding domain protein, 71             157    7e-39
YouGoGlencoco_45, helix-turn-helix DNA binding domain protein, 71      157    7e-39
Yoshand_45, helix-turn-helix DNA binding domain protein, 71            157    7e-39
Xavier_44, helix-turn-helix DNA binding domain protein, 71             157    7e-39
Weher20_45, helix-turn-helix DNA-binding domain protein, 71            157    7e-39
Waterdiva_45, helix-turn-helix DNA binding domain protein, 71          157    7e-39
Wallhey_44, helix-turn-helix DNA binding domain protein, 71            157    7e-39
Vortex_45, helix-turn-helix DNA binding domain protein, 71             157    7e-39
Vista_45, helix-turn-helix DNA binding domain protein, 71              157    7e-39
Virapocalypse_45, helix-turn-helix DNA binding domain protein, 71      157    7e-39
Veritas_44, helix-turn-helix DNA binding domain protein, 71            157    7e-39
Vaticameos_41, helix-turn-helix DNA binding domain protein, 71         157    7e-39
Valjean_45, helix-turn-helix DNA binding domain protein, 71            157    7e-39
Vaishali24_44, helix-turn-helix DNA binding domain protein, 71         157    7e-39
Usavi_45, helix-turn-helix DNA binding domain protein, 71              157    7e-39
UncleHowie_45, helix-turn-helix DNA binding domain protein, 71         157    7e-39
UAch1_45, helix-turn-helix DNA binding domain protein, 71              157    7e-39
TyrionL_44, helix-turn-helix DNA binding domain protein, 71            157    7e-39
Trypo_45, helix-turn-helix DNA binding domain protein, 71              157    7e-39
True_44, helix-turn-helix DNA binding domain protein, 71               157    7e-39
Tooj_45, helix-turn-helix DNA binding domain protein, 71               157    7e-39
Toni_44, helix-turn-helix DNA binding domain protein, 71               157    7e-39
Tomlarah_45, helix-turn-helix DNA-binding domain protein, 71           157    7e-39
TomBombadil_45, helix-turn-helix DNA binding domain protein, 71        157    7e-39
Timmi_44, helix-turn-helix DNA binding domain protein, 71              157    7e-39
ThreeOh3D2_45, helix-turn-helix DNA binding domain protein, 71         157    7e-39
Thora_45, helix-turn-helix DNA binding domain protein, 71              157    7e-39
Telesworld_44, helix-turn-helix DNA binding domain protein, 71         157    7e-39
TallGrassMM_45, helix-turn-helix DNA binding domain protein, 71        157    7e-39
Swish_45, helix-turn-helix DNA binding domain protein, 71              157    7e-39
Swiphy_Draft_47, function unknown, 71                                  157    7e-39
Surely_45, helix-turn-helix DNA binding domain protein, 71             157    7e-39
Suffolk_45, helix-turn-helix DNA binding domain protein, 71            157    7e-39
Struggle_44, helix-turn-helix DNA binding domain protein, 71           157    7e-39
Squiggle_45, helix-turn-helix DNA binding domain protein, 71           157    7e-39
```

## b. SIF: NCBI BLAST

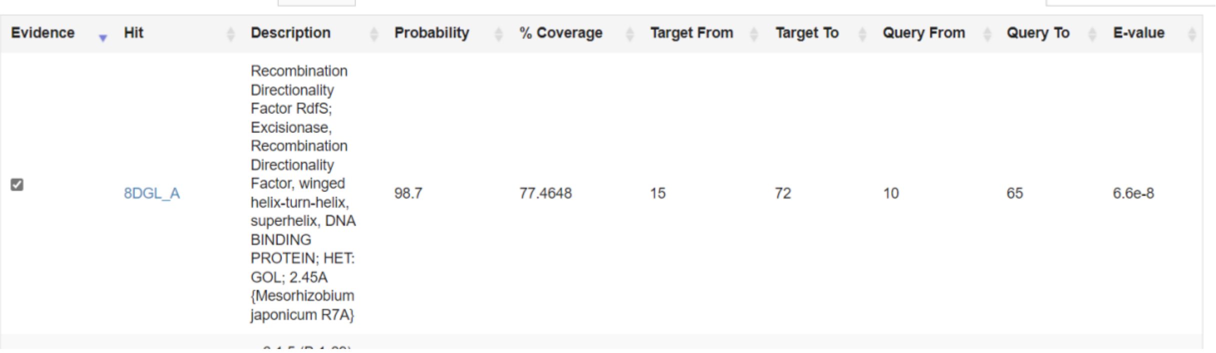


| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Ga |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | NP_943823 | No | 2023-01-08 | DNA binding protein | DNA binding protein [Mycobacterium phage PG1] >ref\|YP_008052122.1\| DNA binding protein [Mycobacterium phage Newman] >ref\|YP_009005692.1\| DNA binding protein [Mycobacterium phage Suffolk] >ref\|YP_009014306.1\| DNA binding protein [Mycobacterium phage Oline] >ref\|YP_009016834.1\| DNA binding protein [Mycobacterium phage Vista] >ref\|YP_009018358.1\| DNA binding protein [Mycobacterium phage JacAttac] | 100 | 100 | 100 | 71 | 1 | 71 | 1 | 71 | 0 |



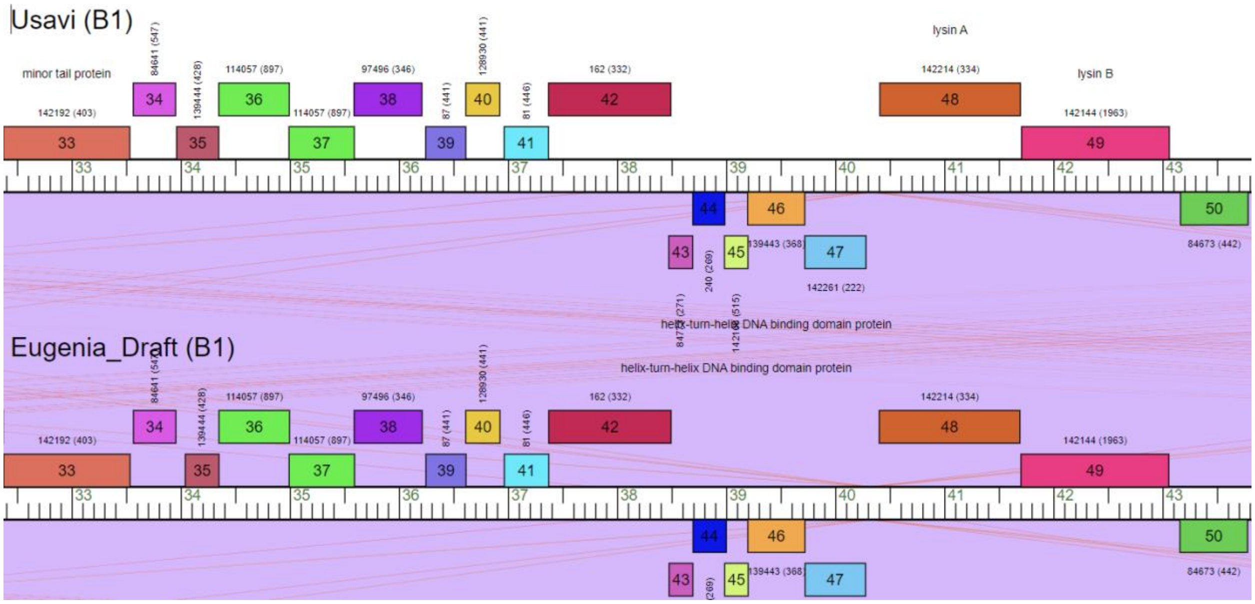
| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | AEK08797 | No | 2023-08-29 | helix-turn-helix DNA binding domain protein | helix-turn-helix DNA binding domain protein [Mycobacterium phage Harvey] >gb\|AIM50277.1\| helix-turn-helix DNA binding domain protein [Mycobacterium phage Vivaldi] >gb\|AZF96710.1\| helix-turn-helix DNA binding domain protein [Mycobacterium phage Keitherie] >gb\|QJD51093.1\| helix-turn-helix DNA binding domain protein [Mycobacterium phage Chaelin] >gb\|QOI67220.1\| helix-turn-helix DNA binding domain | 98.5916 | 100 | 100 | 71 | 1 | 71 | 1 | 71 | 0 | 1.1416e-42 |

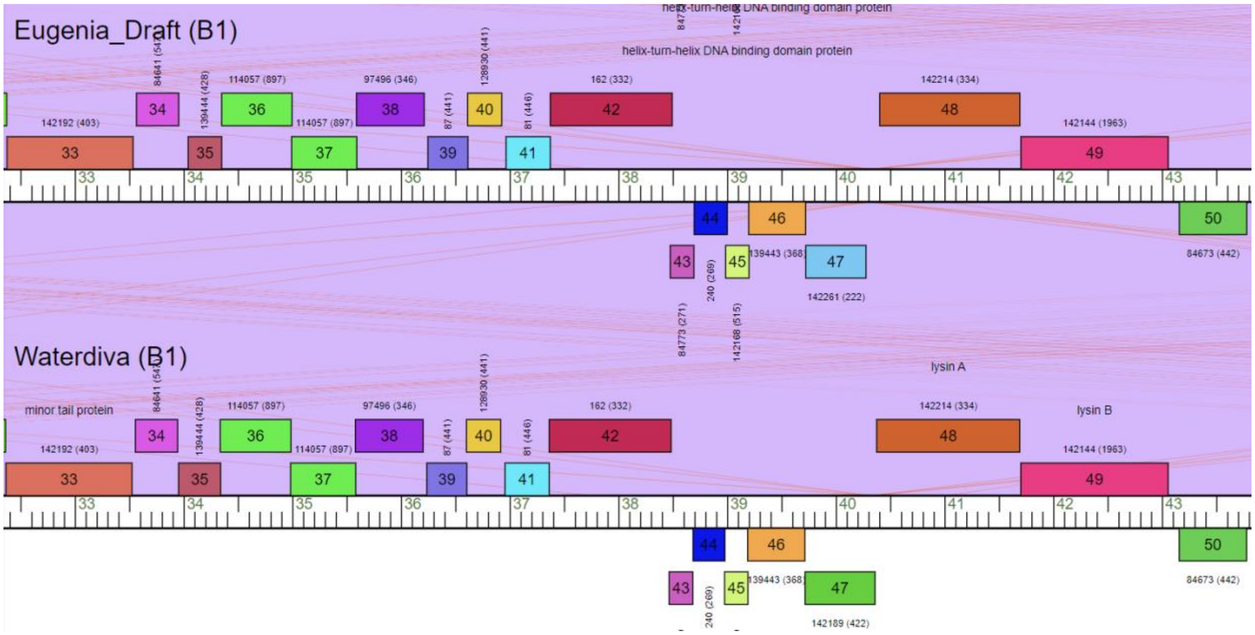
## c. SIF: HHPred



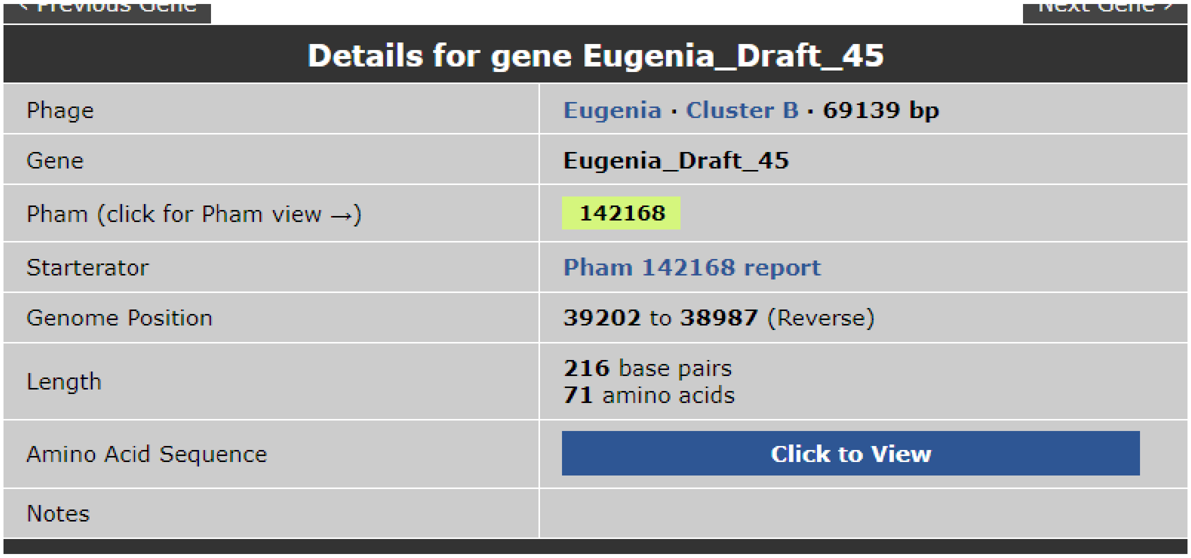
| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | 8DGL_A | Recombination Directionality Factor RdfS; Excisionase, Recombination Directionality Factor, winged helix-turn-helix, superhelix, DNA BINDING PROTEIN; HET: GOL; 2.45A {Mesorhizobium japonicum R7A} | 98.7 | 77.4648 | 15 | 72 | 10 | 65 | 6.6e-8 |

## d. SIF: Synteny-Phamerator (three genomes)





7. Any other important information.



## Details for gene Eugenia_Draft_45

| | |
|---|---|
| Phage | Eugenia · Cluster B · **69139 bp** |
| Gene | **Eugenia_Draft_45** |
| Pham (click for Pham view →) | **142168** |
| Starterator | **Pham 142168 report** |
| Genome Position | **39202** to **38987** (Reverse) |
| Length | **216** base pairs<br>**71** amino acids |
| Amino Acid Sequence | **Click to View** |
| Notes | |

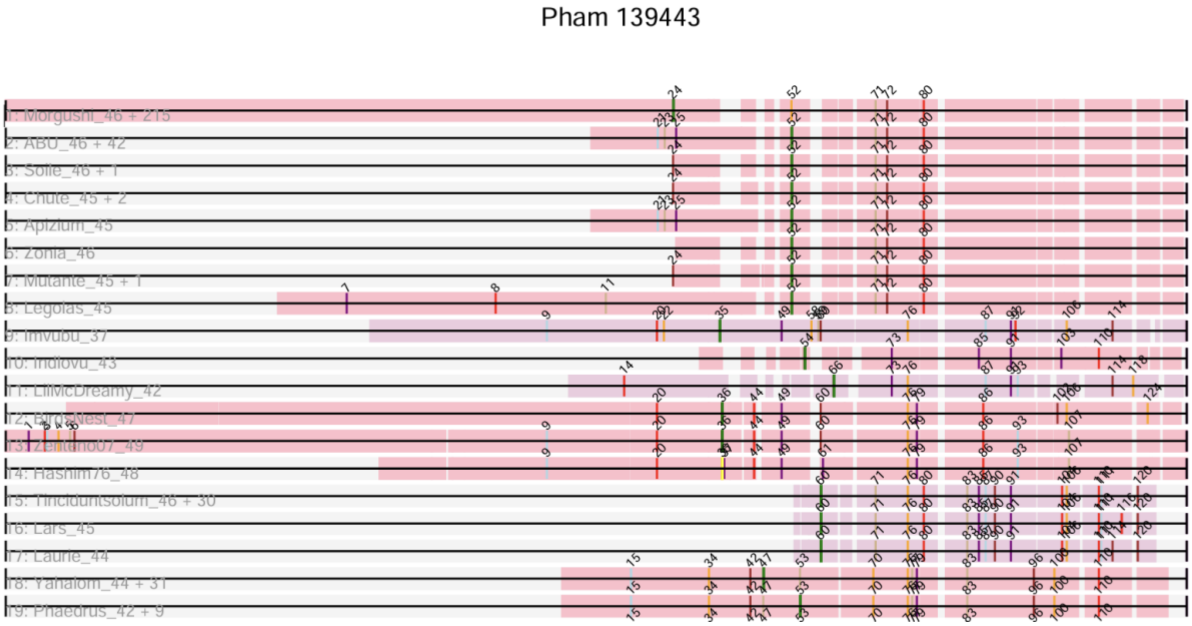
## CURATOR NAME: OLIVIA SIDOTI

## GENE NAME: EUGENIA GENE 46

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

1. Starterator



Pham number 139443 has 426 members, 49 are drafts.

The start number called the most often in the published annotations is 24, it was called in 198 of the 377 non-draft genes in the pham.

Start 24:

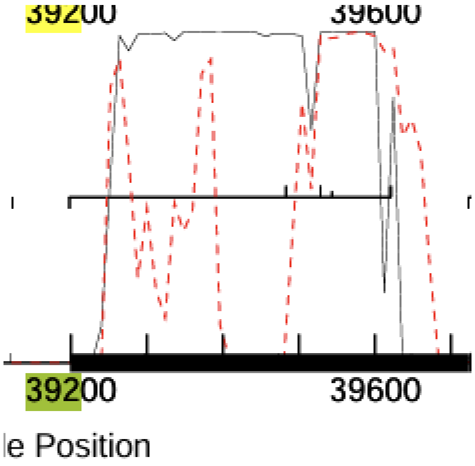
- Found in 223 of 426 ( 52.3% ) of genes in pham
- Manual Annotations of this start: 198 of 377
- Called 96.9% of time when present

2. GeneMark coding potential

Start: 39,723

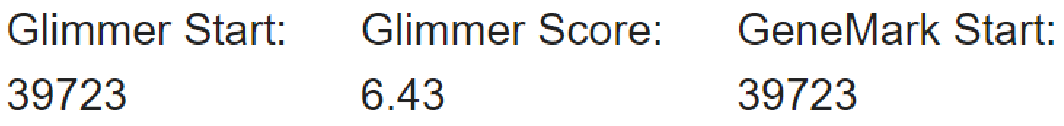
Stop: 39,199

Reverse



3. Glimmer and GeneMark agreement

Phage: Eugenia Cluster: B1



Glimmer Start: 39723

Glimmer Score: 6.43

GeneMark Start: 39723

Yes, Glimmer and GeneMark are in agreement that the start position is at 39723.

## 4. Longest open reading frame (ORF) without excessive gap

| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 39723 | 39199 | 525 | 0 | 10 | 1.793 | -5.199 | TRUE | TTG | Yes | ☑ |

Yes, this gene is the longest open reading frame.

Gap: 0

Spacer: 10

Z-Score: 1.793

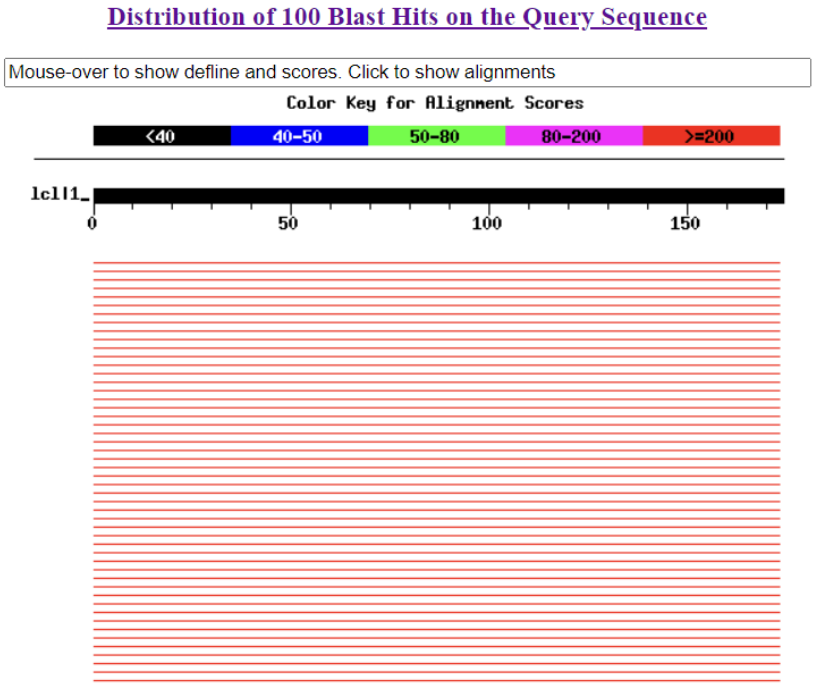
Final Score: -5.199

## 5. Function. If no functional prediction is present, write “Hypothetical protein”.

helix-turn-helix DNA binding domain

# 6. Supporting Information for Function (SIF)

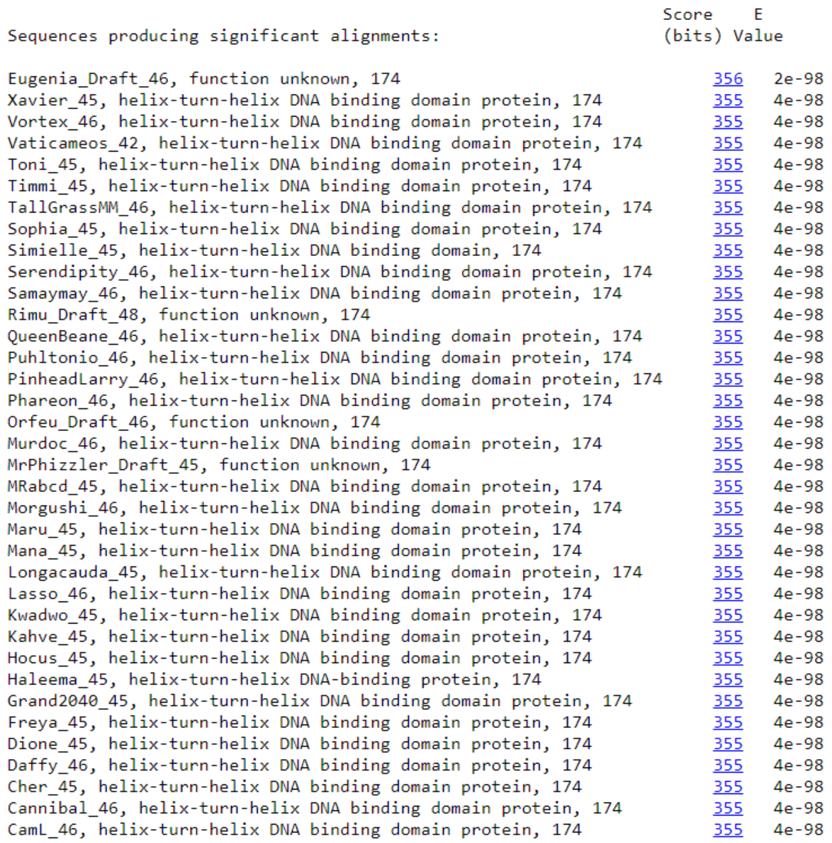
## a. SIF: PhageDb BLAST



Distribution of 100 Blast Hits on the Query Sequence

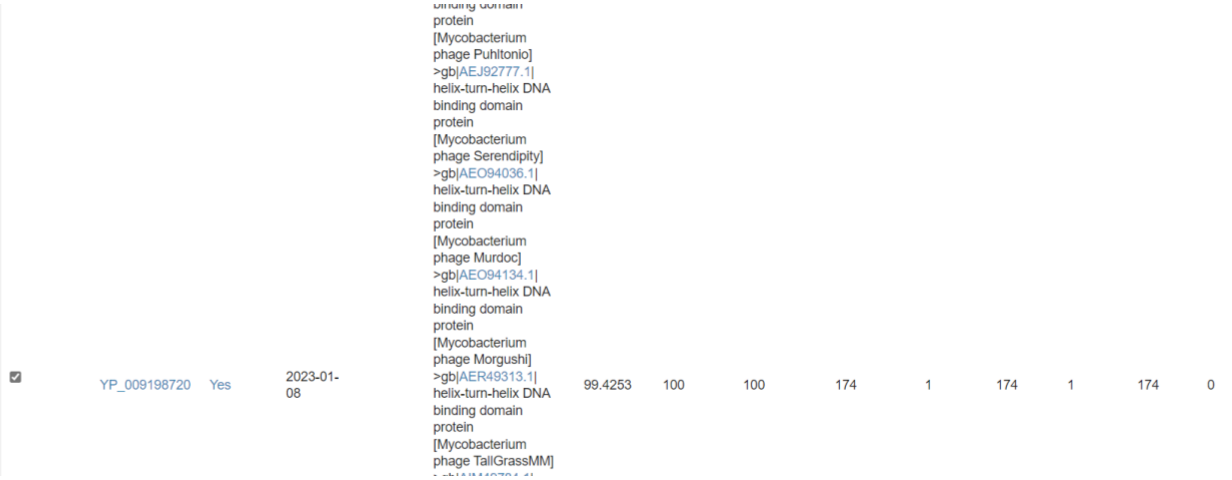
Mouse-over to show defline and scores. Click to show alignments

Color Key for Alignment Scores

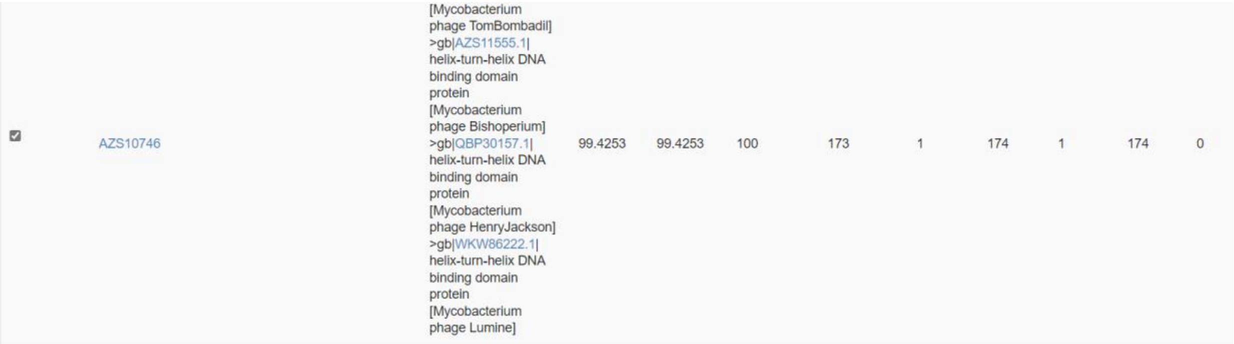


```
                                                                 Score    E
Sequences producing significant alignments:                      (bits) Value

Eugenia_Draft_46, function unknown, 174                              356   2e-98
Xavier_45, helix-turn-helix DNA binding domain protein, 174          355   4e-98
Vortex_46, helix-turn-helix DNA binding domain protein, 174          355   4e-98
Vaticameos_42, helix-turn-helix DNA binding domain protein, 174      355   4e-98
Toni_45, helix-turn-helix DNA binding domain protein, 174            355   4e-98
Timmi_45, helix-turn-helix DNA binding domain protein, 174           355   4e-98
TallGrassMM_46, helix-turn-helix DNA binding domain protein, 174     355   4e-98
Sophia_45, helix-turn-helix DNA binding domain protein, 174          355   4e-98
Simielle_45, helix-turn-helix DNA binding domain, 174                355   4e-98
Serendipity_46, helix-turn-helix DNA binding domain protein, 174     355   4e-98
Samaymay_46, helix-turn-helix DNA binding domain protein, 174        355   4e-98
Rimu_Draft_48, function unknown, 174                                 355   4e-98
QueenBeane_46, helix-turn-helix DNA binding domain protein, 174      355   4e-98
Puhltonio_46, helix-turn-helix DNA binding domain protein, 174       355   4e-98
PinheadLarry_46, helix-turn-helix DNA binding domain protein, 174    355   4e-98
Phareon_46, helix-turn-helix DNA binding domain protein, 174         355   4e-98
Orfeu_Draft_46, function unknown, 174                                355   4e-98
Murdoc_46, helix-turn-helix DNA binding domain protein, 174          355   4e-98
MrPhizzler_Draft_45, function unknown, 174                           355   4e-98
MRabcd_45, helix-turn-helix DNA binding domain protein, 174          355   4e-98
Morgushi_46, helix-turn-helix DNA binding domain protein, 174        355   4e-98
Maru_45, helix-turn-helix DNA binding domain protein, 174            355   4e-98
Mana_45, helix-turn-helix DNA binding domain protein, 174            355   4e-98
Longacauda_45, helix-turn-helix DNA binding domain protein, 174      355   4e-98
Lasso_46, helix-turn-helix DNA binding domain protein, 174           355   4e-98
Kwadwo_45, helix-turn-helix DNA binding domain protein, 174          355   4e-98
Kahve_45, helix-turn-helix DNA binding domain protein, 174           355   4e-98
Hocus_45, helix-turn-helix DNA binding domain protein, 174           355   4e-98
Haleema_45, helix-turn-helix DNA-binding protein, 174                355   4e-98
Grand2040_45, helix-turn-helix DNA binding domain protein, 174       355   4e-98
Freya_45, helix-turn-helix DNA binding domain protein, 174           355   4e-98
Dione_45, helix-turn-helix DNA binding domain protein, 174           355   4e-98
Daffy_46, helix-turn-helix DNA binding domain protein, 174           355   4e-98
Cher_45, helix-turn-helix DNA binding domain protein, 174            355   4e-98
Cannibal_46, helix-turn-helix DNA binding domain protein, 174        355   4e-98
CamL_46, helix-turn-helix DNA binding domain protein, 174            355   4e-98
```

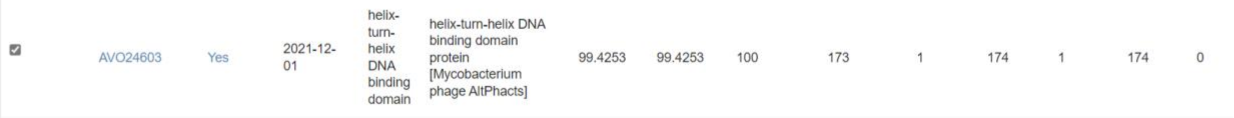
b. SIF: NCBI BLAST



| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | YP_009198720 | Yes | 2023-01-08 | | binding domain protein [Mycobacterium phage Puhltonio] >gb\|AEJ92777.1\| helix-turn-helix DNA binding domain protein [Mycobacterium phage Serendipity] >gb\|AEO94036.1\| helix-turn-helix DNA binding domain protein [Mycobacterium phage Murdoc] >gb\|AEO94134.1\| helix-turn-helix DNA binding domain protein [Mycobacterium phage Morgushi] >gb\|AER49313.1\| helix-turn-helix DNA binding domain protein [Mycobacterium phage TallGrassMM] | 99.4253 | 100 | 100 | 174 | 1 | 174 | 1 | 174 | 0 |



| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | AZS10746 | | | | [Mycobacterium phage TomBombadil] >gb\|AZS11555.1\| helix-turn-helix DNA binding domain protein [Mycobacterium phage Bishoperium] >gb\|QBP30157.1\| helix-turn-helix DNA binding domain protein [Mycobacterium phage HenryJackson] >gb\|WKW86222.1\| helix-turn-helix DNA binding domain protein [Mycobacterium phage Lumine] | 99.4253 | 99.4253 | 100 | 173 | 1 | 174 | 1 | 174 | 0 |

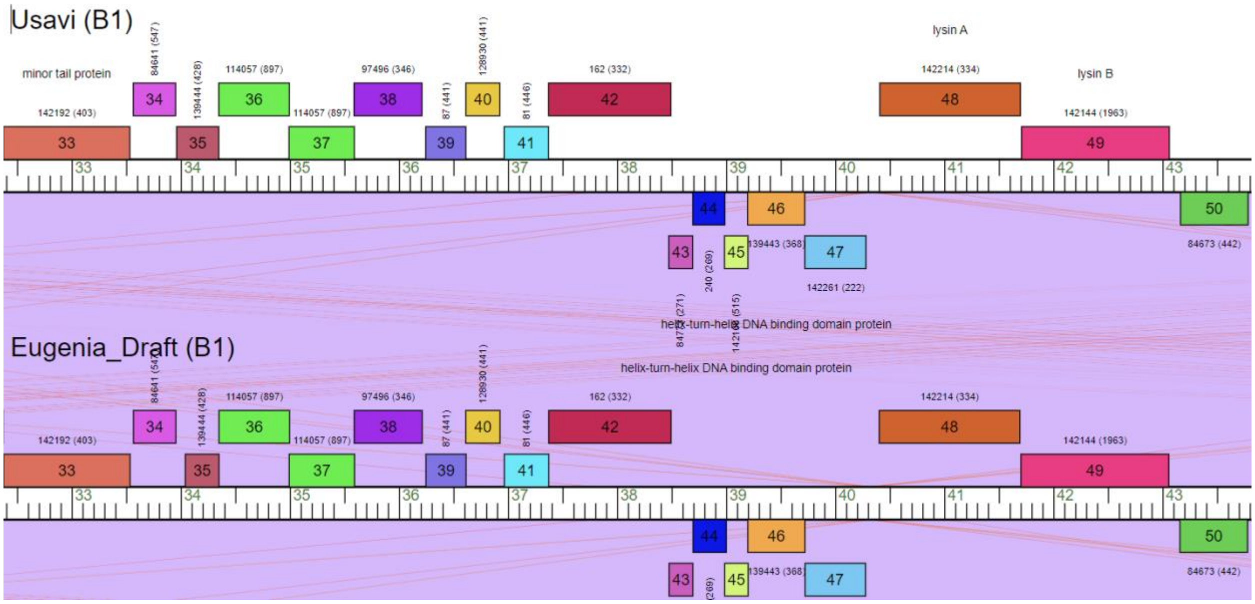


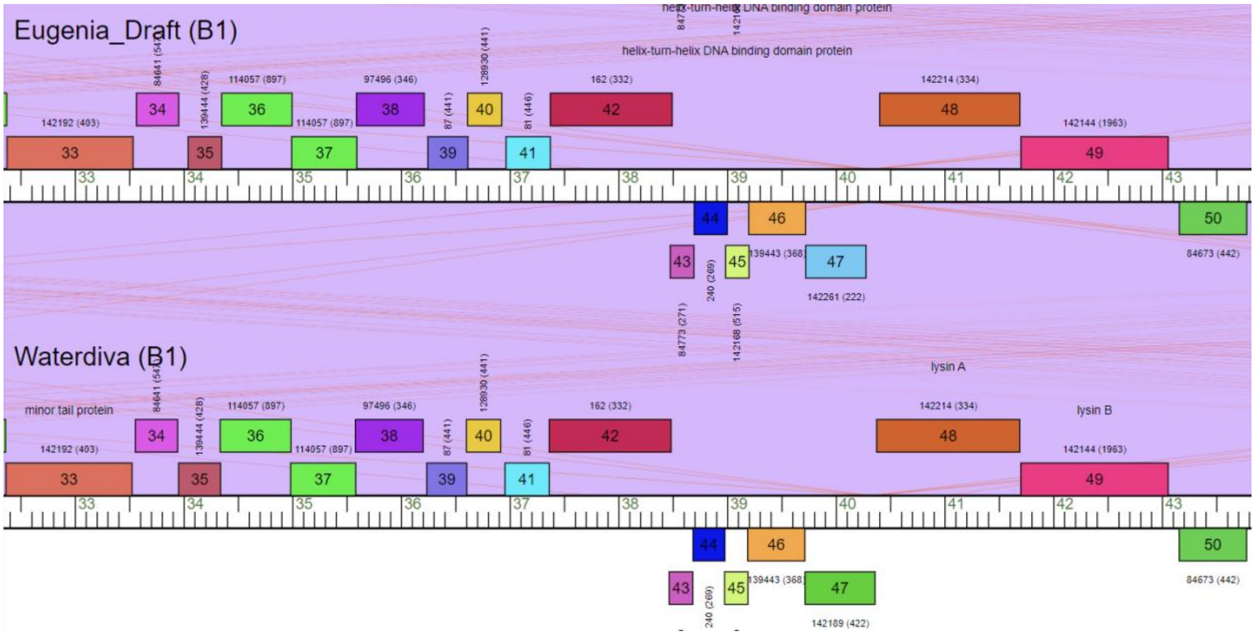
| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | AVO24603 | Yes | 2021-12-01 | helix-turn-helix DNA binding domain | helix-turn-helix DNA binding domain protein [Mycobacterium phage AltPhacts] | 99.4253 | 99.4253 | 100 | 173 | 1 | 174 | 1 | 174 | 0 |

### c. SIF: HHPred

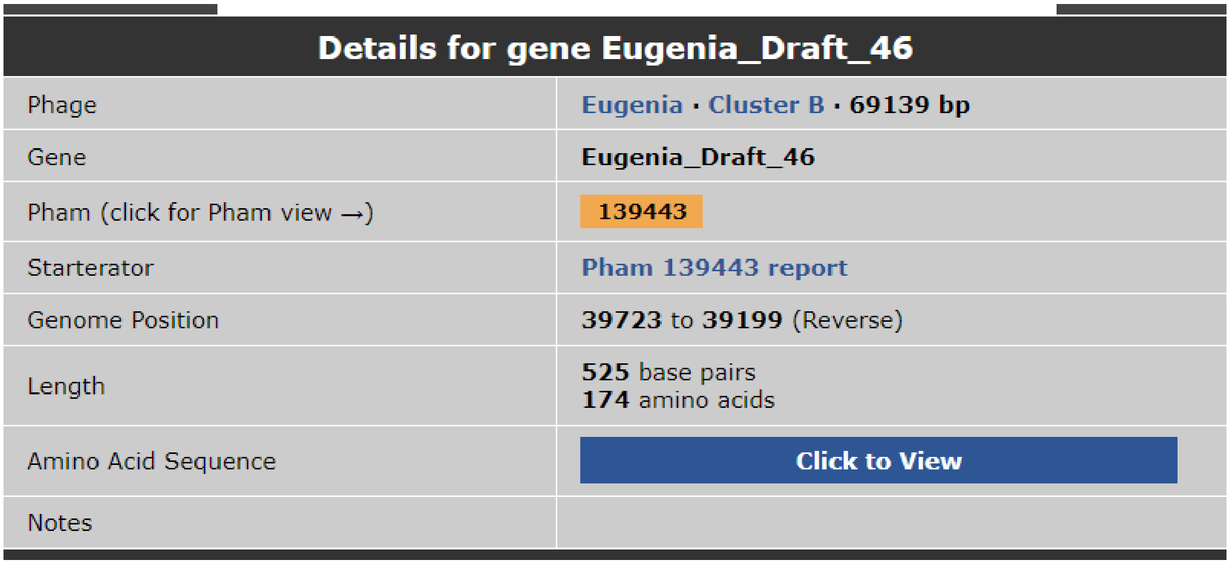
No Evidence from HHPred

### d. SIF: Synteny-Phamerator (three genomes)





7. Any other important information.



## Details for gene Eugenia_Draft_46

| | |
|---|---|
| Phage | Eugenia · Cluster B · **69139 bp** |
| Gene | **Eugenia_Draft_46** |
| Pham (click for Pham view →) | **139443** |
| Starterator | **Pham 139443 report** |
| Genome Position | **39723** to **39199** (Reverse) |
| Length | **525** base pairs<br>**174** amino acids |
| Amino Acid Sequence | **Click to View** |
| Notes | |

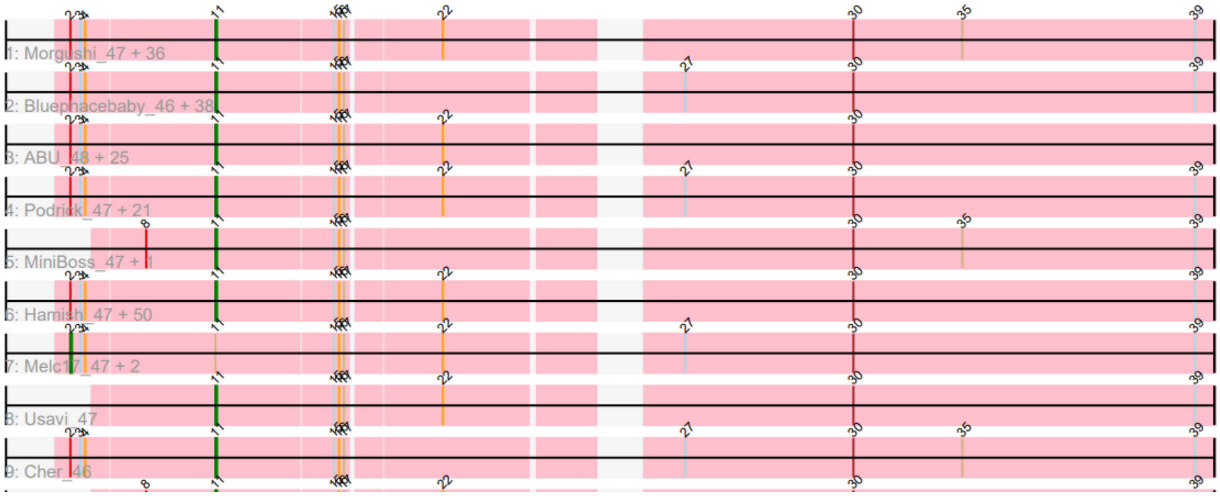
## CURATOR NAME: BELLA N

## GENE NAME: EUGENIA-47

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

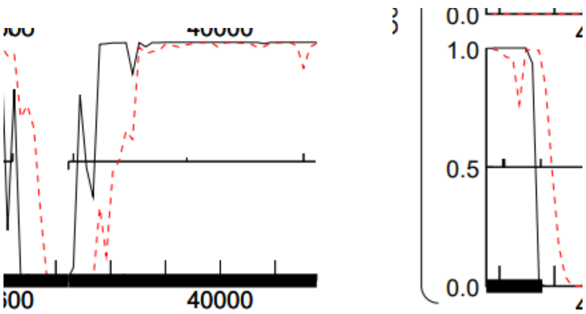
1. Starterator



The start number called the most often in the published annotations is 11, it was called in 187 of the 207 non-draft genes in the pham.
Start 11:

- Found in 215 of 223 ( 96.4% ) of genes in pham
- Manual Annotations of this start: 187 of 207
- Called 93.5% of time when present

2. GeneMark coding potential



The black bar at the bottom represents the whole ORF of gene 47, it is about 500 nucleotides long.

3. Glimmer and GeneMark agreement

There is agreement between Glimmer and GeneMark

Glimmer Start: 40278

Glimmer Score: 13.55

GeneMark Start: 40278

4. Longest open reading frame (ORF) without excessive gap

The selected gene is not the LORF with a length of 555 nucleotides. It has a Z-score of 2.852, a spacer of 12, and a final score of -3.212. It is a reverse gene.

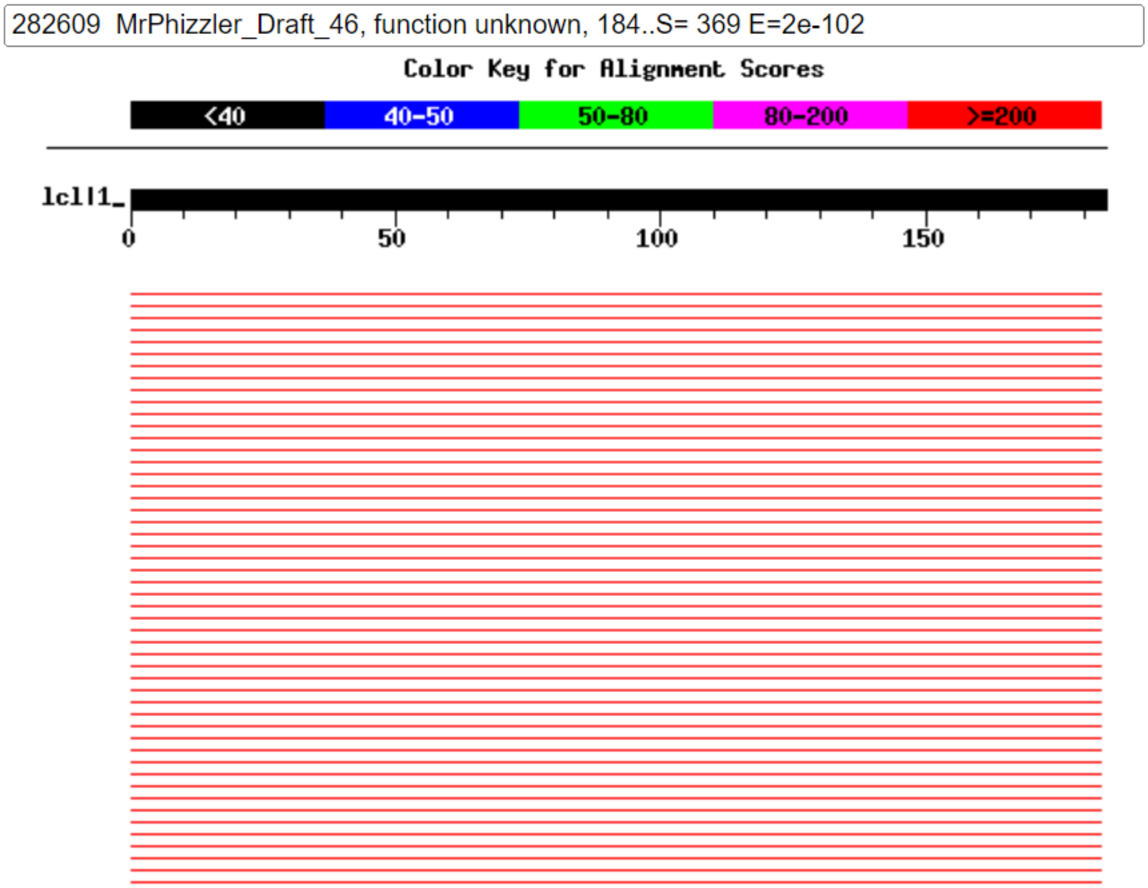
5. Function. If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

6. Supporting Information for Function (SIF)

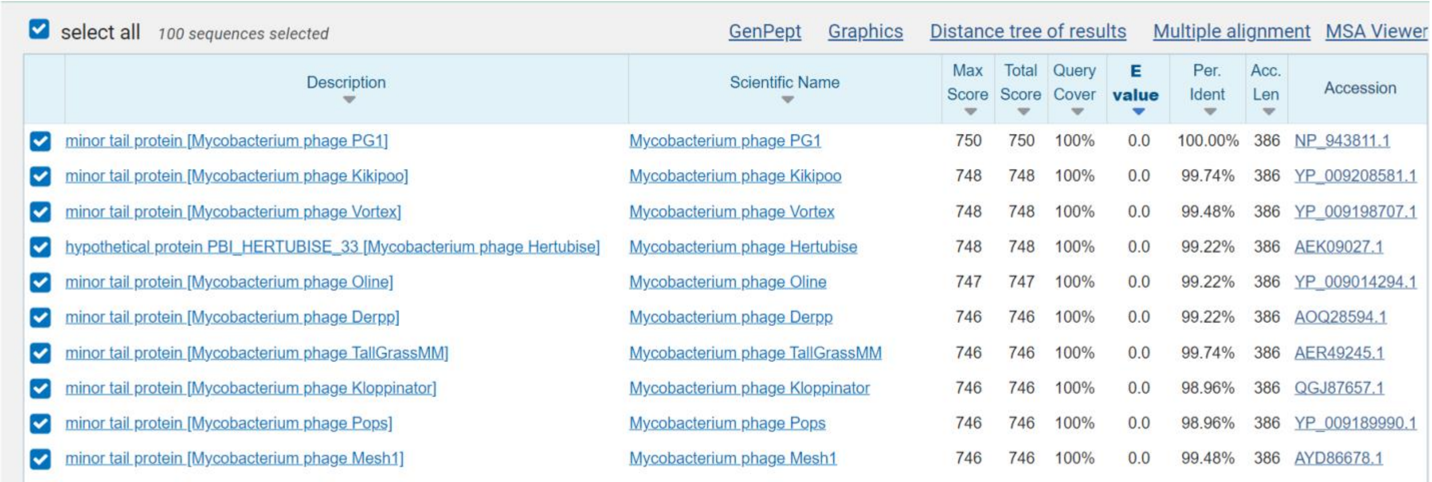
   a. SIF: PhageDb BLAST

# Distribution of 100 Blast Hits on the Query Sequence



```
                                                                  Score    E
Sequences producing significant alignments:                      (bits) Value

Xavier_46, function unknown, 212                                     369   e-102
Vortex_47, function unknown, 212                                     369   e-102
Toni_46, function unknown, 184                                       369   e-102
TomBombadil_47, helix-turn-helix DNA-binding protein, 184            369   e-102
Timmi_46, function unknown, 184                                      369   e-102
TallGrassMM_47, function unknown, 212                                369   e-102
Sophia_46, function unknown, 184                                     369   e-102
Simielle_46, function unknown, 184                                   369   e-102
Serendipity_47, function unknown, 184                                369   e-102
Roy17_47, function unknown, 212                                      369   e-102
Rimu_Draft_49, function unknown, 184                                 369   e-102
Ricotta_Draft_47, function unknown, 184                              369   e-102
Puhltonio_47, function unknown, 212                                  369   e-102
PinheadLarry_47, function unknown, 184                               369   e-102
```

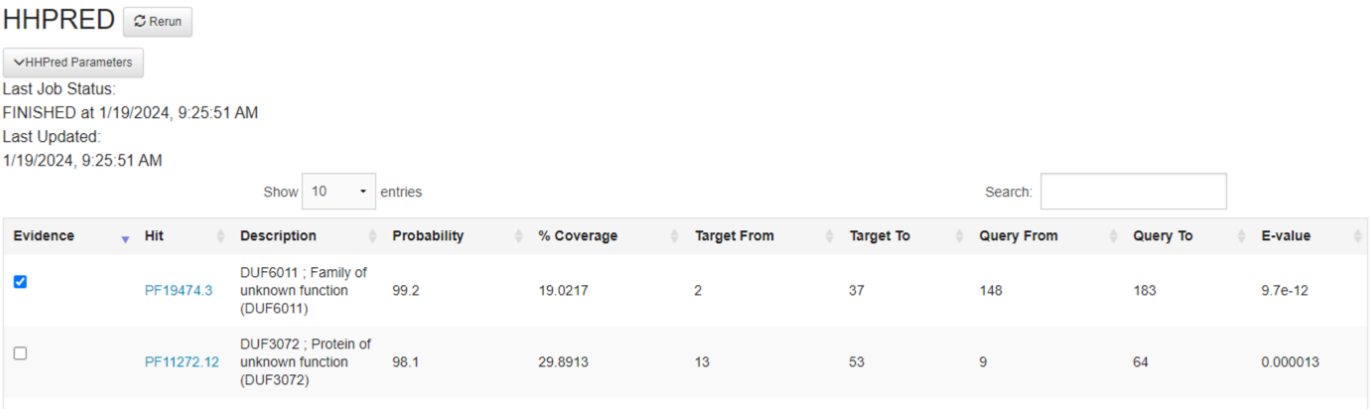
## b. SIF: NCBI BLAST



*100 sequences selected*

| | Description | Scientific Name | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | minor tail protein [Mycobacterium phage PG1] | Mycobacterium phage PG1 | 750 | 750 | 100% | 0.0 | 100.00% | 386 | NP_943811.1 |
| ☑ | minor tail protein [Mycobacterium phage Kikipoo] | Mycobacterium phage Kikipoo | 748 | 748 | 100% | 0.0 | 99.74% | 386 | YP_009208581.1 |
| ☑ | minor tail protein [Mycobacterium phage Vortex] | Mycobacterium phage Vortex | 748 | 748 | 100% | 0.0 | 99.48% | 386 | YP_009198707.1 |
| ☑ | hypothetical protein PBI_HERTUBISE_33 [Mycobacterium phage Hertubise] | Mycobacterium phage Hertubise | 748 | 748 | 100% | 0.0 | 99.22% | 386 | AEK09027.1 |
| ☑ | minor tail protein [Mycobacterium phage Oline] | Mycobacterium phage Oline | 747 | 747 | 100% | 0.0 | 99.22% | 386 | YP_009014294.1 |
| ☑ | minor tail protein [Mycobacterium phage Derpp] | Mycobacterium phage Derpp | 746 | 746 | 100% | 0.0 | 99.22% | 386 | AOQ28594.1 |
| ☑ | minor tail protein [Mycobacterium phage TallGrassMM] | Mycobacterium phage TallGrassMM | 746 | 746 | 100% | 0.0 | 99.74% | 386 | AER49245.1 |
| ☑ | minor tail protein [Mycobacterium phage Kloppinator] | Mycobacterium phage Kloppinator | 746 | 746 | 100% | 0.0 | 98.96% | 386 | QGJ87657.1 |
| ☑ | minor tail protein [Mycobacterium phage Pops] | Mycobacterium phage Pops | 746 | 746 | 100% | 0.0 | 98.96% | 386 | YP_009189990.1 |
| ☑ | minor tail protein [Mycobacterium phage Mesh1] | Mycobacterium phage Mesh1 | 746 | 746 | 100% | 0.0 | 99.48% | 386 | AYD86678.1 |

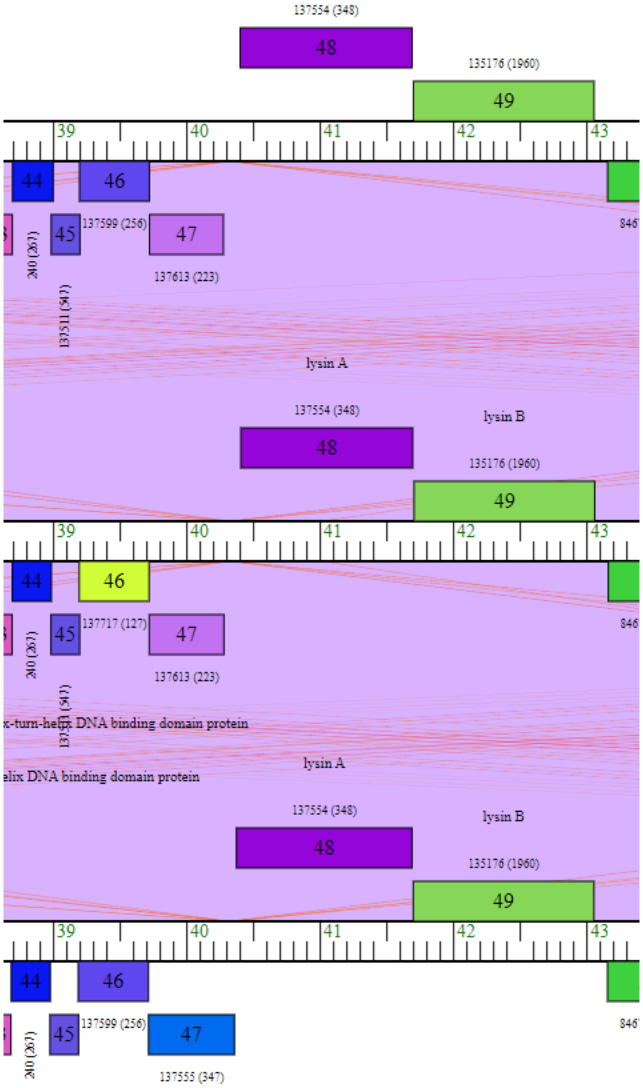
## c. SIF: HHPred



Last Job Status:
FINISHED at 1/19/2024, 9:25:51 AM
Last Updated:
1/19/2024, 9:25:51 AM

| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | PF19474.3 | DUF6011 ; Family of unknown function (DUF6011) | 99.2 | 19.0217 | 2 | 37 | 148 | 183 | 9.7e-12 |
| ☐ | PF11272.12 | DUF3072 ; Protein of unknown function (DUF3072) | 98.1 | 29.8913 | 13 | 53 | 9 | 64 | 0.000013 |

## d. SIF: Synteny-Phamerator (three genomes)



7. Any other important information.

No TmHmm

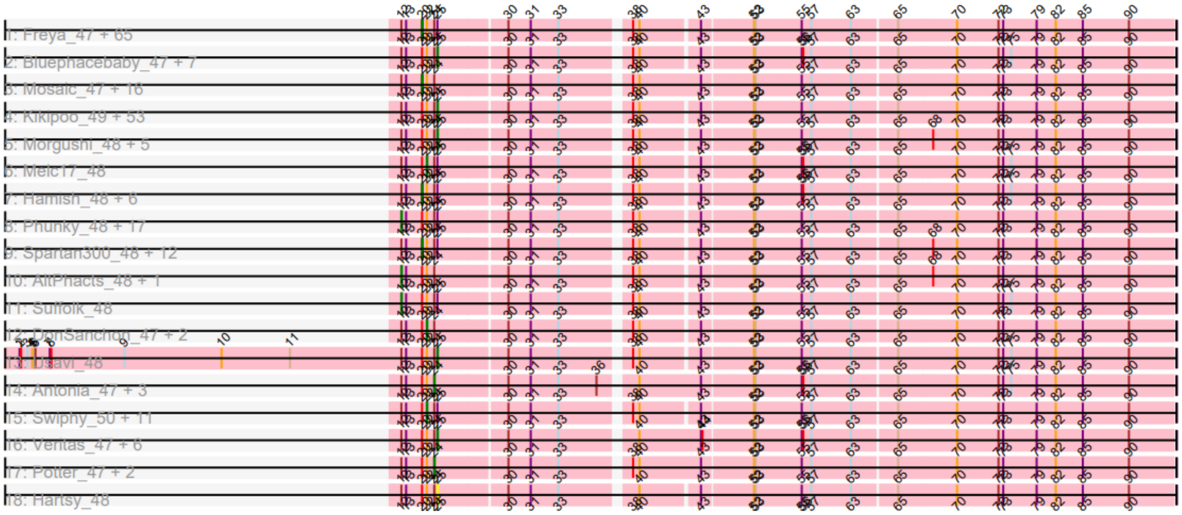
**CURATOR NAME: BELLA N**

**GENE NAME: EUGENIA-48**

**DNA MASTER NOTES: N/A**

**START POSITION EVALUATION (IN ORDER)**:

1. Starterator

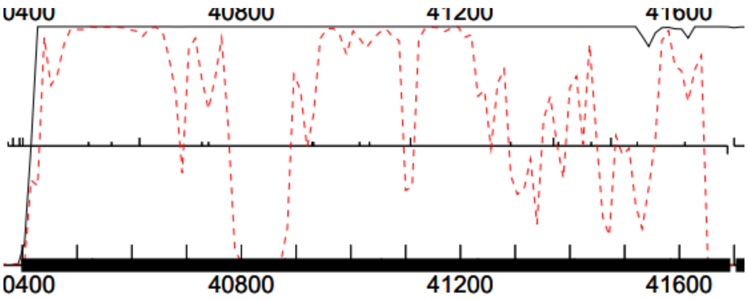


The start number called the most often in the published annotations is 20, it was called in 130 of the 305 non-draft genes in the pham.

Start 20:

- Found in 269 of 333 ( 80.8% ) of genes in pham
- Manual Annotations of this start: 130 of 305
- Called 49.1% of time when present

2. GeneMark coding potential



The black bar at the bottom represents the whole ORF it may seem like it continues on, but there is a break around 41600.

3. Glimmer and GeneMark agreement

There is not agreement between Glimmer and GeneMark

Glimmer Start: 40404

Glimmer Score: 15.84

GeneMark Start: 40386

4. Longest open reading frame (ORF) without excessive gap

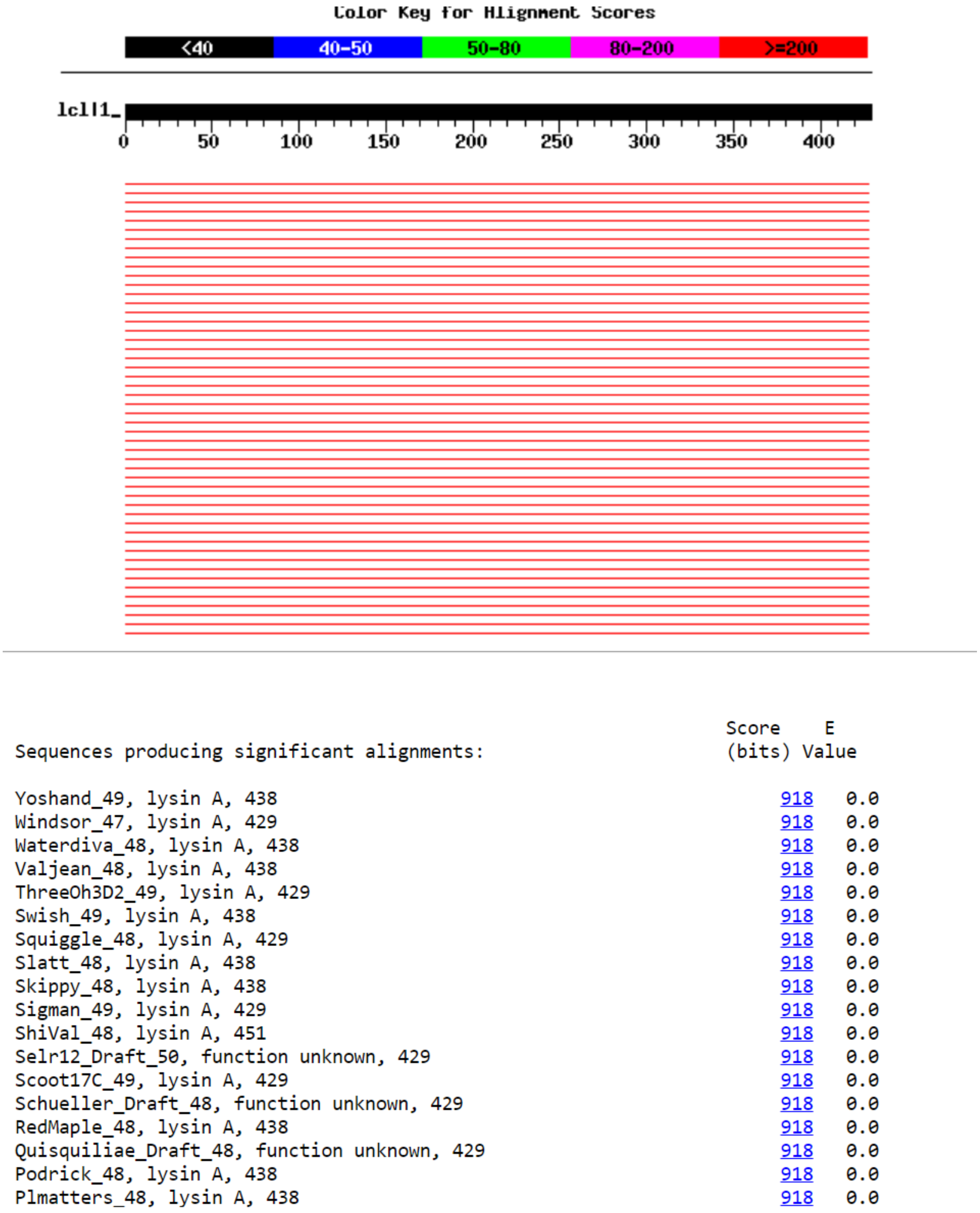
The selected gene is not the LORF with a length of 1290 nucleotides. It has a Z-score of 2.453, a spacer of 11, and a final score of -3.935.

5. Function. If no functional prediction is present, write “Hypothetical protein”.

Lysin A

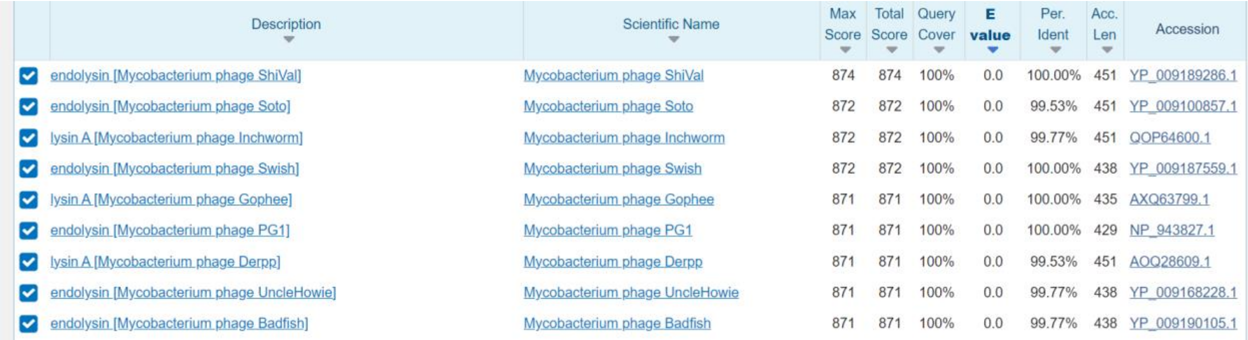
6. Supporting Information for Function (SIF)

a. SIF: PhageDb BLAST



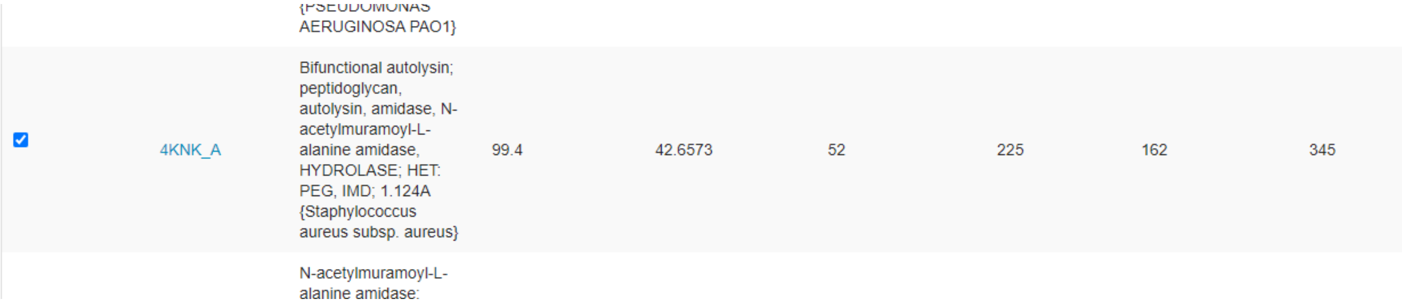
```
                                                                   Score     E
Sequences producing significant alignments:                        (bits)  Value

Yoshand_49, lysin A, 438                                              918    0.0
Windsor_47, lysin A, 429                                              918    0.0
Waterdiva_48, lysin A, 438                                            918    0.0
Valjean_48, lysin A, 438                                              918    0.0
ThreeOh3D2_49, lysin A, 429                                           918    0.0
Swish_49, lysin A, 438                                                918    0.0
Squiggle_48, lysin A, 429                                             918    0.0
Slatt_48, lysin A, 438                                                918    0.0
Skippy_48, lysin A, 438                                               918    0.0
Sigman_49, lysin A, 429                                               918    0.0
ShiVal_48, lysin A, 451                                               918    0.0
Selr12_Draft_50, function unknown, 429                                918    0.0
Scoot17C_49, lysin A, 429                                             918    0.0
Schueller_Draft_48, function unknown, 429                             918    0.0
RedMaple_48, lysin A, 438                                             918    0.0
Quisquiliae_Draft_48, function unknown, 429                           918    0.0
Podrick_48, lysin A, 438                                              918    0.0
Plmatters_48, lysin A, 438                                            918    0.0
```

b. SIF: NCBI BLAST



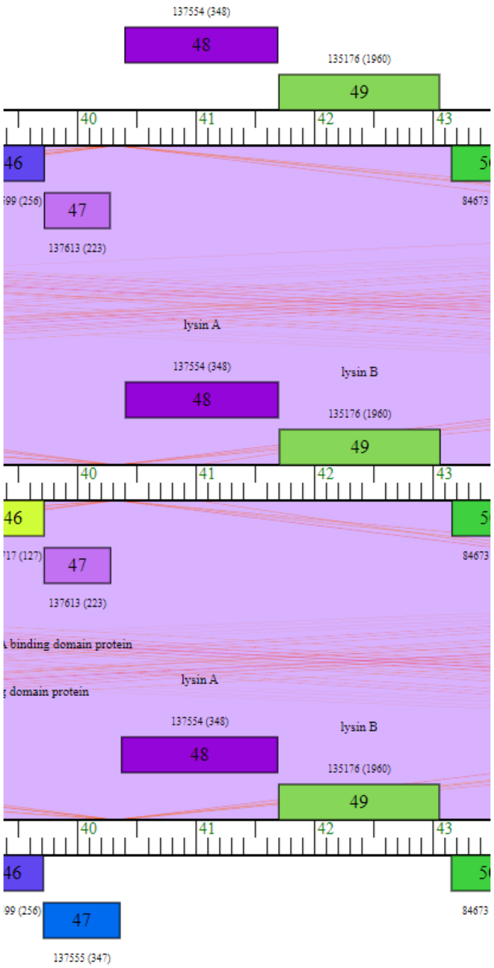
| | Description | Scientific Name | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | endolysin [Mycobacterium phage ShiVal] | Mycobacterium phage ShiVal | 874 | 874 | 100% | 0.0 | 100.00% | 451 | YP_009189286.1 |
| ☑ | endolysin [Mycobacterium phage Soto] | Mycobacterium phage Soto | 872 | 872 | 100% | 0.0 | 99.53% | 451 | YP_009100857.1 |
| ☑ | lysin A [Mycobacterium phage Inchworm] | Mycobacterium phage Inchworm | 872 | 872 | 100% | 0.0 | 99.77% | 451 | QOP64600.1 |
| ☑ | endolysin [Mycobacterium phage Swish] | Mycobacterium phage Swish | 872 | 872 | 100% | 0.0 | 100.00% | 438 | YP_009187559.1 |
| ☑ | lysin A [Mycobacterium phage Gophee] | Mycobacterium phage Gophee | 871 | 871 | 100% | 0.0 | 100.00% | 435 | AXQ63799.1 |
| ☑ | endolysin [Mycobacterium phage PG1] | Mycobacterium phage PG1 | 871 | 871 | 100% | 0.0 | 100.00% | 429 | NP_943827.1 |
| ☑ | lysin A [Mycobacterium phage Derpp] | Mycobacterium phage Derpp | 871 | 871 | 100% | 0.0 | 99.53% | 451 | AOQ28609.1 |
| ☑ | endolysin [Mycobacterium phage UncleHowie] | Mycobacterium phage UncleHowie | 871 | 871 | 100% | 0.0 | 99.77% | 438 | YP_009168228.1 |
| ☑ | endolysin [Mycobacterium phage Badfish] | Mycobacterium phage Badfish | 871 | 871 | 100% | 0.0 | 99.77% | 438 | YP_009190105.1 |

## c. SIF: HHPred



| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | {PSEUDOMONAS AERUGINOSA PAO1} | | | | | | |
| ☑ | 4KNK_A | Bifunctional autolysin; peptidoglycan, autolysin, amidase, N-acetylmuramoyl-L-alanine amidase, HYDROLASE; HET: PEG, IMD; 1.124A {Staphylococcus aureus subsp. aureus} | 99.4 | 42.6573 | 52 | 225 | 162 | 345 |
| | | N-acetylmuramoyl-L-alanine amidase: | | | | | | |

## d. SIF: Synteny-Phamerator (three genomes)



7. Any other important information.

Synteny with ABU, Adriana, Anderson, and others.

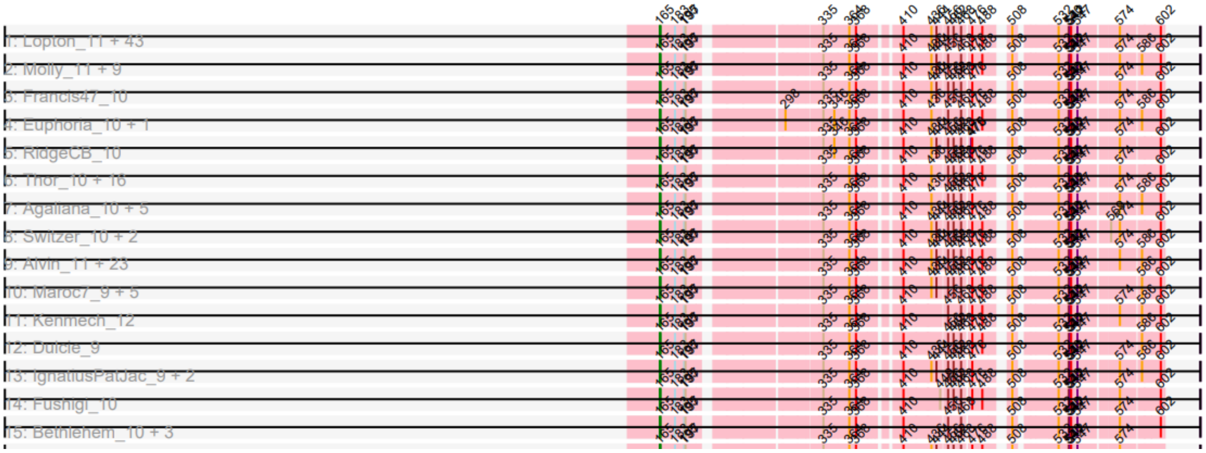
**CURATOR NAME: BELLA N**

**GENE NAME: EUGENIA-49**

**DNA MASTER NOTES: N/A**

**START POSITION EVALUATION (IN ORDER)**:

1. Starterator

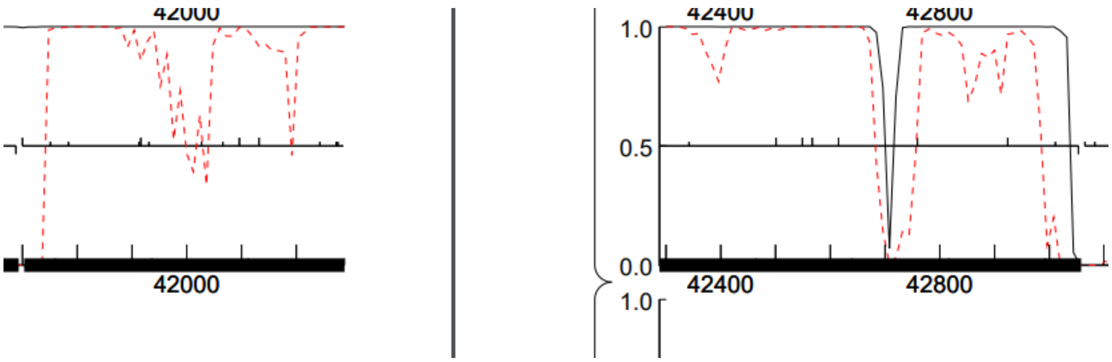


The start number called the most often in the published annotations is 80, it was called in 246 of the 1825 non-draft genes in the pham.

Start 80:

- Found in 270 of 1980 ( 13.6% ) of genes in pham
- Manual Annotations of this start: 246 of 1825
- Called 100.0% of time when present

2. GeneMark coding potential



The black bar at the bottom represents the whole ORF it is about 1400 nucleotides.

3. Glimmer and GeneMark agreement

There is agreement between Glimmer and GeneMark

Glimmer Start: 41703

Glimmer Score: 16.83

GeneMark Start: 41703

4. Longest open reading frame (ORF) without excessive gap

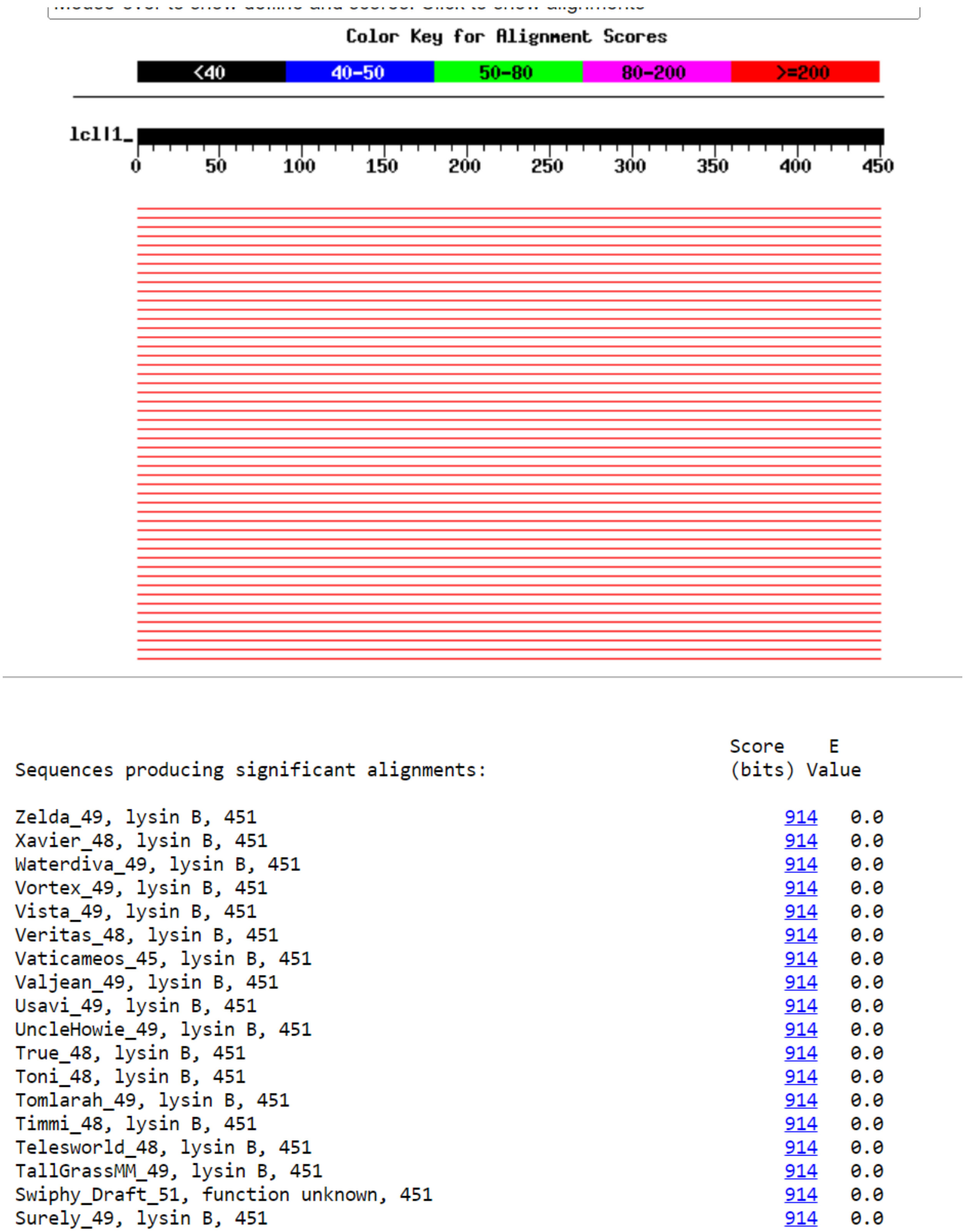
The selected gene is the LORF with a length of 1356 nucleotides. It has a Z-score of 2.227, a spacer of 10, and a final score of -4.326.

5. Function. If no functional prediction is present, write “Hypothetical protein”.

Lysin B

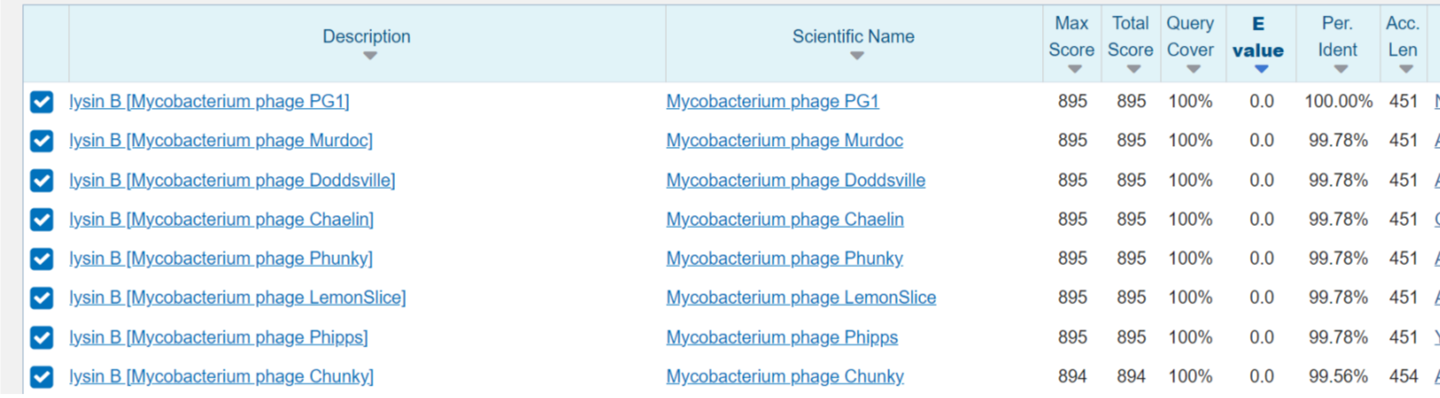
6. Supporting Information for Function (SIF)

a. SIF: PhageDb BLAST



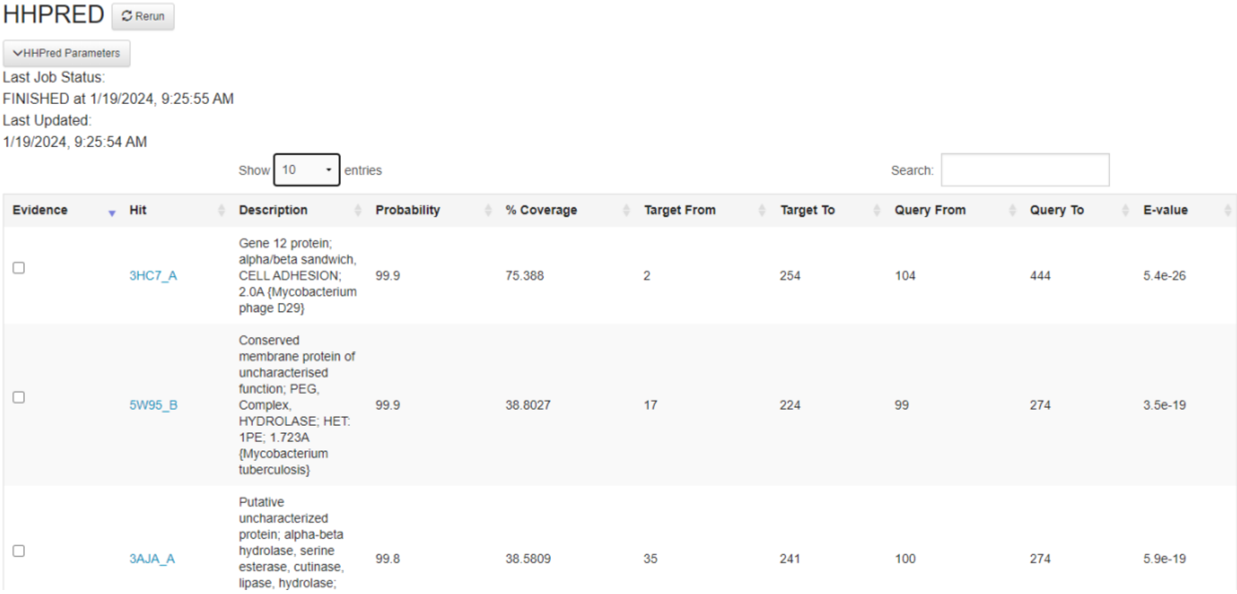
```
                                                                 Score     E
Sequences producing significant alignments:                      (bits)  Value

Zelda_49, lysin B, 451                                              914    0.0
Xavier_48, lysin B, 451                                             914    0.0
Waterdiva_49, lysin B, 451                                          914    0.0
Vortex_49, lysin B, 451                                             914    0.0
Vista_49, lysin B, 451                                              914    0.0
Veritas_48, lysin B, 451                                            914    0.0
Vaticameos_45, lysin B, 451                                         914    0.0
Valjean_49, lysin B, 451                                            914    0.0
Usavi_49, lysin B, 451                                              914    0.0
UncleHowie_49, lysin B, 451                                         914    0.0
True_48, lysin B, 451                                               914    0.0
Toni_48, lysin B, 451                                               914    0.0
Tomlarah_49, lysin B, 451                                           914    0.0
Timmi_48, lysin B, 451                                              914    0.0
Telesworld_48, lysin B, 451                                         914    0.0
TallGrassMM_49, lysin B, 451                                        914    0.0
Swiphy_Draft_51, function unknown, 451                              914    0.0
Surely_49, lysin B, 451                                             914    0.0
```

## b. SIF: NCBI BLAST



| | Description | Scientific Name | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len |
|---|---|---|---|---|---|---|---|---|
| ☑ | lysin B [Mycobacterium phage PG1] | Mycobacterium phage PG1 | 895 | 895 | 100% | 0.0 | 100.00% | 451 |
| ☑ | lysin B [Mycobacterium phage Murdoc] | Mycobacterium phage Murdoc | 895 | 895 | 100% | 0.0 | 99.78% | 451 |
| ☑ | lysin B [Mycobacterium phage Doddsville] | Mycobacterium phage Doddsville | 895 | 895 | 100% | 0.0 | 99.78% | 451 |
| ☑ | lysin B [Mycobacterium phage Chaelin] | Mycobacterium phage Chaelin | 895 | 895 | 100% | 0.0 | 99.78% | 451 |
| ☑ | lysin B [Mycobacterium phage Phunky] | Mycobacterium phage Phunky | 895 | 895 | 100% | 0.0 | 99.78% | 451 |
| ☑ | lysin B [Mycobacterium phage LemonSlice] | Mycobacterium phage LemonSlice | 895 | 895 | 100% | 0.0 | 99.78% | 451 |
| ☑ | lysin B [Mycobacterium phage Phipps] | Mycobacterium phage Phipps | 895 | 895 | 100% | 0.0 | 99.78% | 451 |
| ☑ | lysin B [Mycobacterium phage Chunky] | Mycobacterium phage Chunky | 894 | 894 | 100% | 0.0 | 99.56% | 454 |

## c. SIF: HHPred

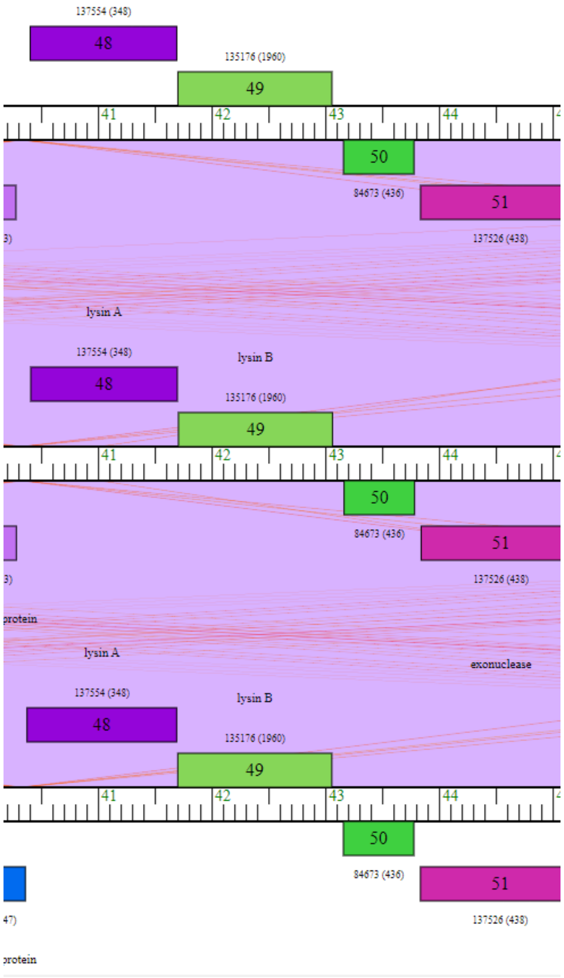


Last Job Status:
FINISHED at 1/19/2024, 9:25:55 AM
Last Updated:
1/19/2024, 9:25:54 AM

| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☐ | 3HC7_A | Gene 12 protein; alpha/beta sandwich, CELL ADHESION; 2.0A {Mycobacterium phage D29} | 99.9 | 75.388 | 2 | 254 | 104 | 444 | 5.4e-26 |
| ☐ | 5W95_B | Conserved membrane protein of uncharacterised function; PEG, Complex, HYDROLASE; HET: 1PE; 1.723A {Mycobacterium tuberculosis} | 99.9 | 38.8027 | 17 | 224 | 99 | 274 | 3.5e-19 |
| ☐ | 3AJA_A | Putative uncharacterized protein; alpha-beta hydrolase, serine esterase, cutinase, lipase, hydrolase; | 99.8 | 38.5809 | 35 | 241 | 100 | 274 | 5.9e-19 |

No HHPRED evidence was selected

## d. SIF: Synteny-Phamerator (three genomes)



7. Any other important information.

Synteny with Fang, Fozzie, Frankincide, and others

3 suspected TmHmms

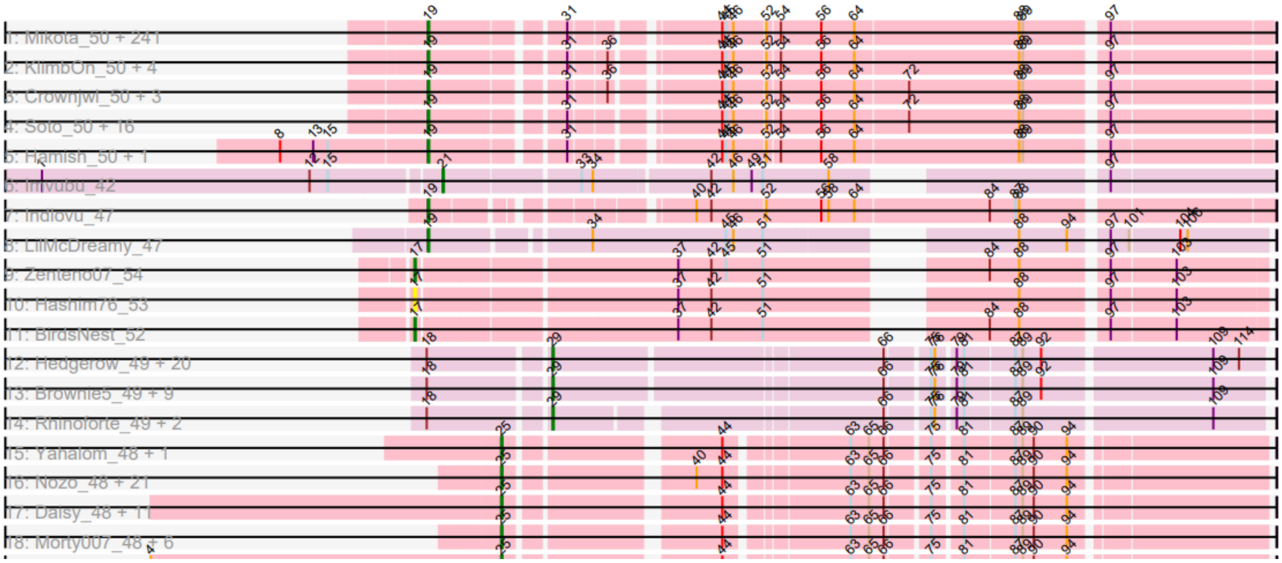
## CURATOR NAME: BELLA N

## GENE NAME: EUGENIA-50

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

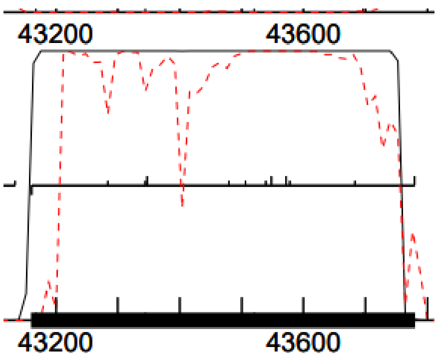
1. Starterator



The start number called the most often in the published annotations is 19, it was called in 284 of the 388 non-draft genes in the pham.
Start 19:

- Found in 321 of 439 ( 73.1% ) of genes in pham
- Manual Annotations of this start: 284 of 388
- Called 100.0% of time when present

2. GeneMark coding potential



The black bar at the bottom represents the whole ORF it is about 600 nucleotides and this is a reverse gene.

3. Glimmer and GeneMark agreement

There is agreement between Glimmer and GeneMark

Glimmer Start: 43780

Glimmer Score: 16.59

GeneMark Start: 43780

4. Longest open reading frame (ORF) without excessive gap

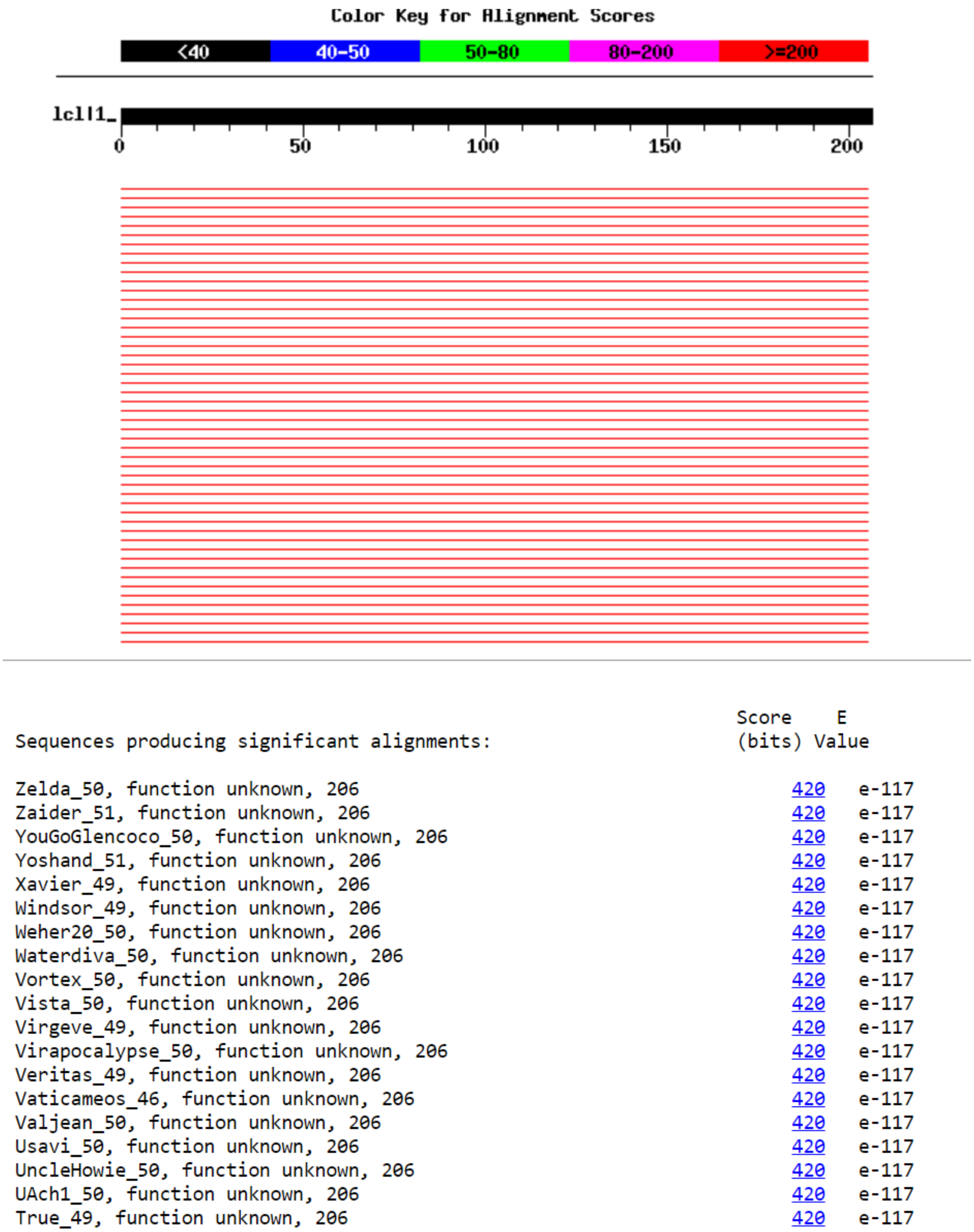
The selected gene is the LORF with a length of 621 nucleotides. It has a Z-score of 3.408, a spacer of 13, and a final score of -2.305. This is a reverse gene.

5. Function. If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

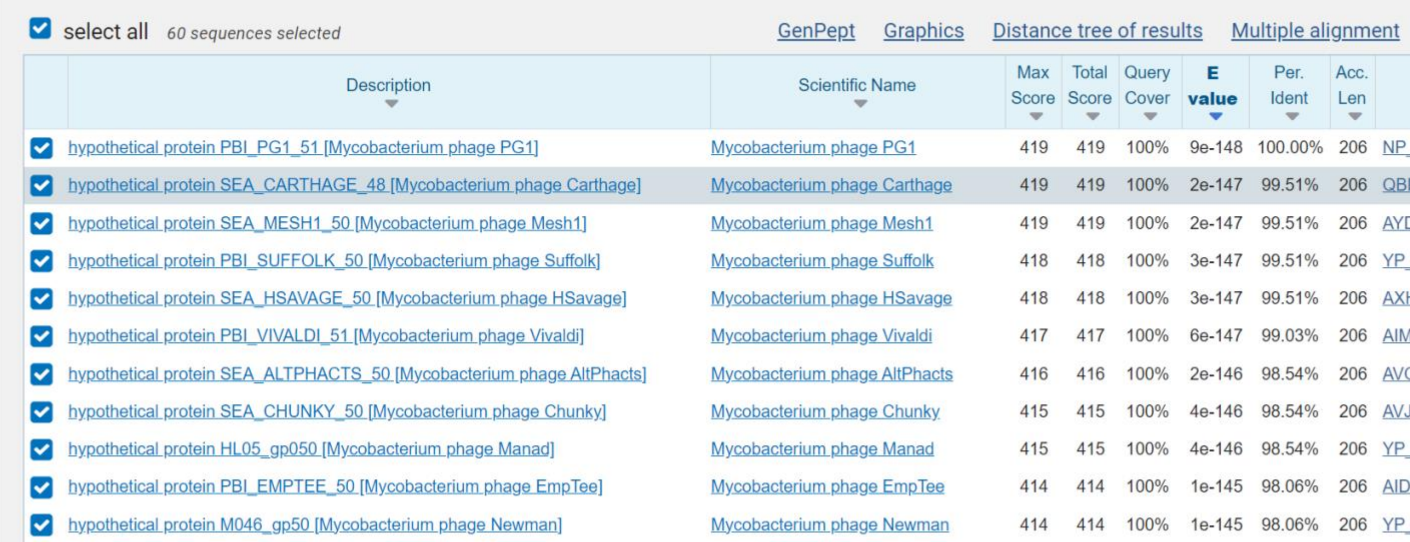
6. Supporting Information for Function (SIF)

a. SIF: PhageDb BLAST



```
                                                                Score    E
Sequences producing significant alignments:                      (bits) Value

Zelda_50, function unknown, 206                                     420    e-117
Zaider_51, function unknown, 206                                    420    e-117
YouGoGlencoco_50, function unknown, 206                             420    e-117
Yoshand_51, function unknown, 206                                   420    e-117
Xavier_49, function unknown, 206                                    420    e-117
Windsor_49, function unknown, 206                                   420    e-117
Weher20_50, function unknown, 206                                   420    e-117
Waterdiva_50, function unknown, 206                                 420    e-117
Vortex_50, function unknown, 206                                    420    e-117
Vista_50, function unknown, 206                                     420    e-117
Virgeve_49, function unknown, 206                                   420    e-117
Virapocalypse_50, function unknown, 206                             420    e-117
Veritas_49, function unknown, 206                                   420    e-117
Vaticameos_46, function unknown, 206                                420    e-117
Valjean_50, function unknown, 206                                   420    e-117
Usavi_50, function unknown, 206                                     420    e-117
UncleHowie_50, function unknown, 206                                420    e-117
UAch1_50, function unknown, 206                                     420    e-117
True_49, function unknown, 206                                      420    e-117
```

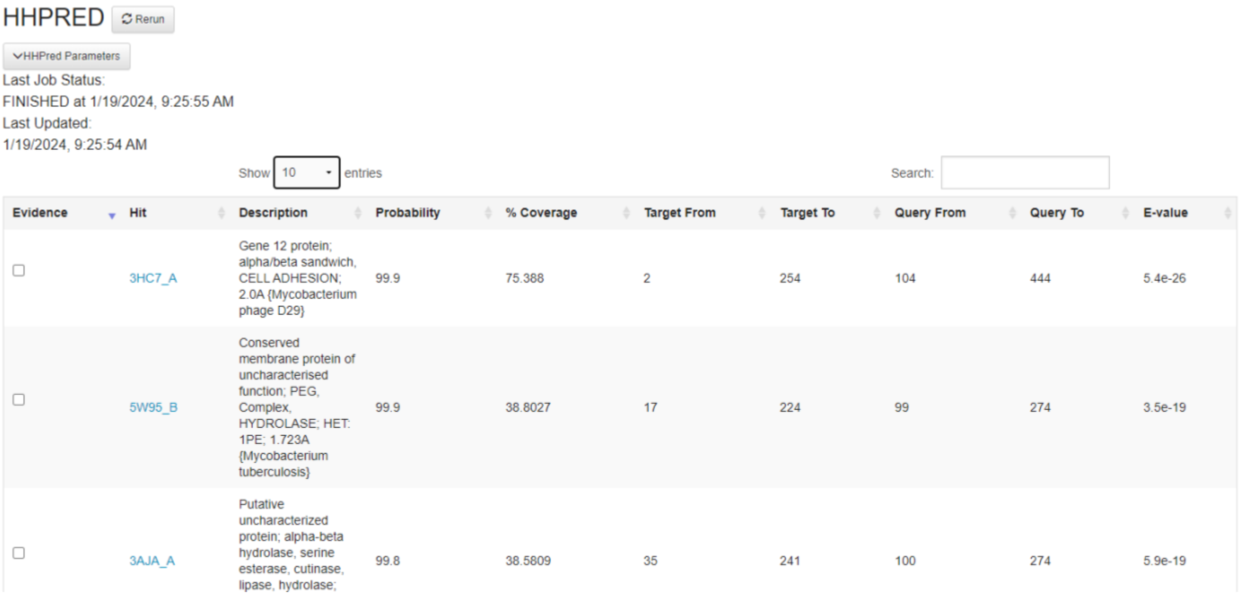
## b. SIF: NCBI BLAST



☑ select all 60 sequences selected

| ☐ | Description | Scientific Name | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | hypothetical protein PBI_PG1_51 [Mycobacterium phage PG1] | Mycobacterium phage PG1 | 419 | 419 | 100% | 9e-148 | 100.00% | 206 | NP_ |
| ☑ | hypothetical protein SEA_CARTHAGE_48 [Mycobacterium phage Carthage] | Mycobacterium phage Carthage | 419 | 419 | 100% | 2e-147 | 99.51% | 206 | QB |
| ☑ | hypothetical protein SEA_MESH1_50 [Mycobacterium phage Mesh1] | Mycobacterium phage Mesh1 | 419 | 419 | 100% | 2e-147 | 99.51% | 206 | AY |
| ☑ | hypothetical protein PBI_SUFFOLK_50 [Mycobacterium phage Suffolk] | Mycobacterium phage Suffolk | 418 | 418 | 100% | 3e-147 | 99.51% | 206 | YP_ |
| ☑ | hypothetical protein SEA_HSAVAGE_50 [Mycobacterium phage HSavage] | Mycobacterium phage HSavage | 418 | 418 | 100% | 3e-147 | 99.51% | 206 | AX |
| ☑ | hypothetical protein PBI_VIVALDI_51 [Mycobacterium phage Vivaldi] | Mycobacterium phage Vivaldi | 417 | 417 | 100% | 6e-147 | 99.03% | 206 | AIM |
| ☑ | hypothetical protein SEA_ALTPHACTS_50 [Mycobacterium phage AltPhacts] | Mycobacterium phage AltPhacts | 416 | 416 | 100% | 2e-146 | 98.54% | 206 | AV |
| ☑ | hypothetical protein SEA_CHUNKY_50 [Mycobacterium phage Chunky] | Mycobacterium phage Chunky | 415 | 415 | 100% | 4e-146 | 98.54% | 206 | AVJ |
| ☑ | hypothetical protein HL05_gp050 [Mycobacterium phage Manad] | Mycobacterium phage Manad | 415 | 415 | 100% | 4e-146 | 98.54% | 206 | YP_ |
| ☑ | hypothetical protein PBI_EMPTEE_50 [Mycobacterium phage EmpTee] | Mycobacterium phage EmpTee | 414 | 414 | 100% | 1e-145 | 98.06% | 206 | AID |
| ☑ | hypothetical protein M046_gp50 [Mycobacterium phage Newman] | Mycobacterium phage Newman | 414 | 414 | 100% | 1e-145 | 98.06% | 206 | YP_ |

## c. SIF: HHPred



Last Job Status:
FINISHED at 1/19/2024, 9:25:55 AM
Last Updated:
1/19/2024, 9:25:54 AM

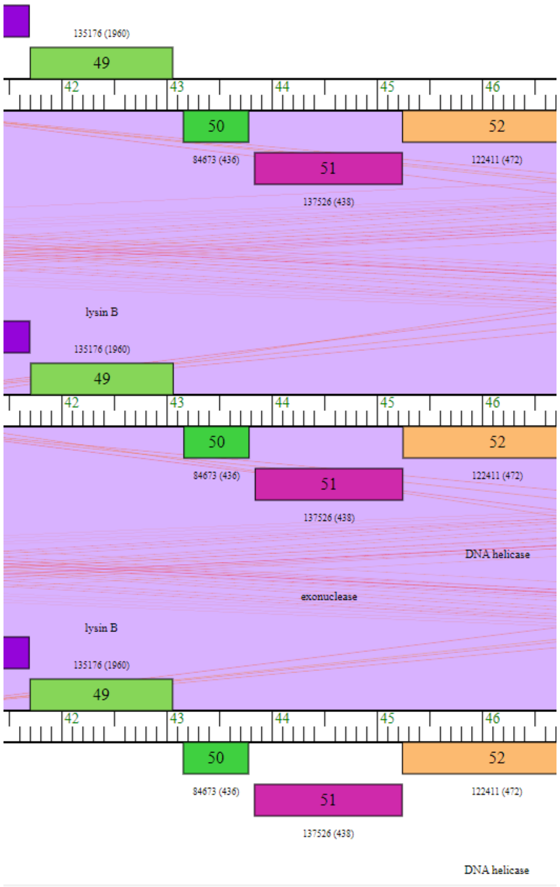
Show 10 entries

Search:

| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☐ | 3HC7_A | Gene 12 protein; alpha/beta sandwich, CELL ADHESION; 2.0A {Mycobacterium phage D29} | 99.9 | 75.388 | 2 | 254 | 104 | 444 | 5.4e-26 |
| ☐ | 5W95_B | Conserved membrane protein of uncharacterised function; PEG, Complex, HYDROLASE; HET: 1PE; 1.723A {Mycobacterium tuberculosis} | 99.9 | 38.8027 | 17 | 224 | 99 | 274 | 3.5e-19 |
| ☐ | 3AJA_A | Putative uncharacterized protein; alpha-beta hydrolase, serine esterase, cutinase, lipase, hydrolase; | 99.8 | 38.5809 | 35 | 241 | 100 | 274 | 5.9e-19 |

No HHPRED evidence was selected

## d. SIF: Synteny-Phamerator (three genomes)



7. Any other important information.

Synteny with Mulan, Murdoc, Mutante and others.

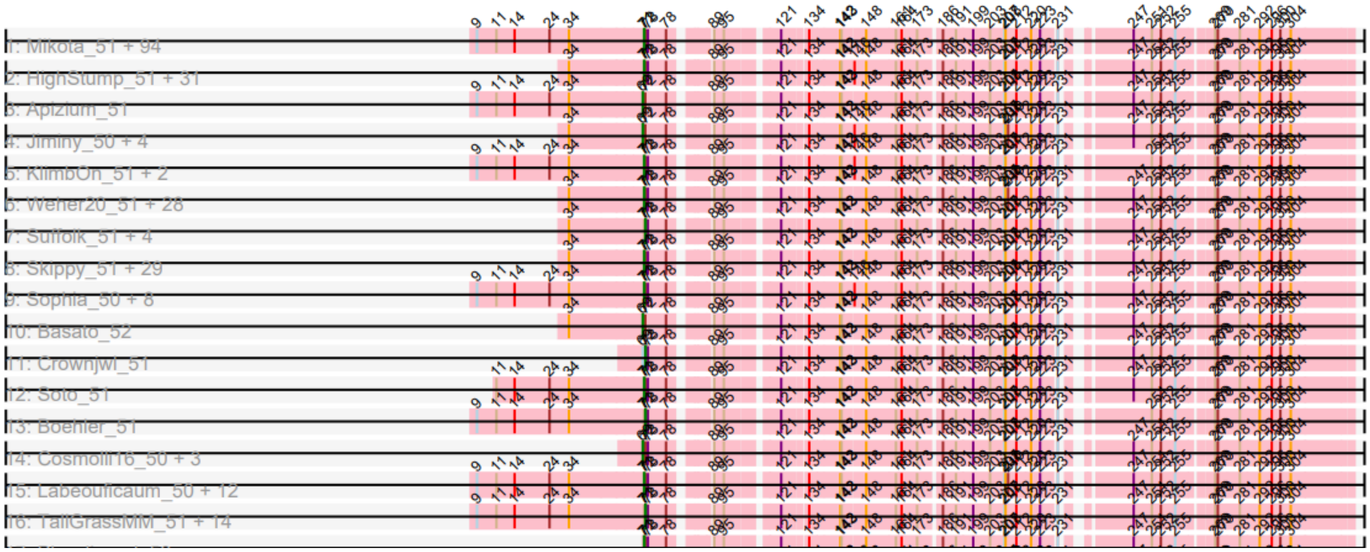
**CURATOR NAME: BELLA N**

**GENE NAME: EUGENIA-51**

**DNA MASTER NOTES: N/A**

**START POSITION EVALUATION (IN ORDER)**:

1. Starterator

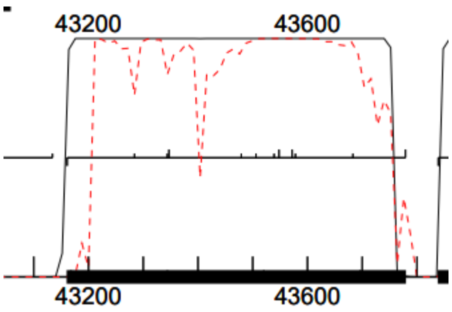


The start number called the most often in the published annotations is 71, it was called in 201 of the 389 non-draft genes in the pham.

Start 71:

- Found in 248 of 441 ( 56.2% ) of genes in pham
- Manual Annotations of this start: 201 of 389
- Called 87.5% of time when present

2. GeneMark coding potential



The black bar at the bottom represents the whole ORF it is about 1500 nucleotides and this is a reverse gene.

3. Glimmer and GeneMark agreement

There is agreement between Glimmer and GeneMark

Glimmer Start: 45241

Glimmer Score: 11.94

GeneMark Start: 45241

4. Longest open reading frame (ORF) without excessive gap

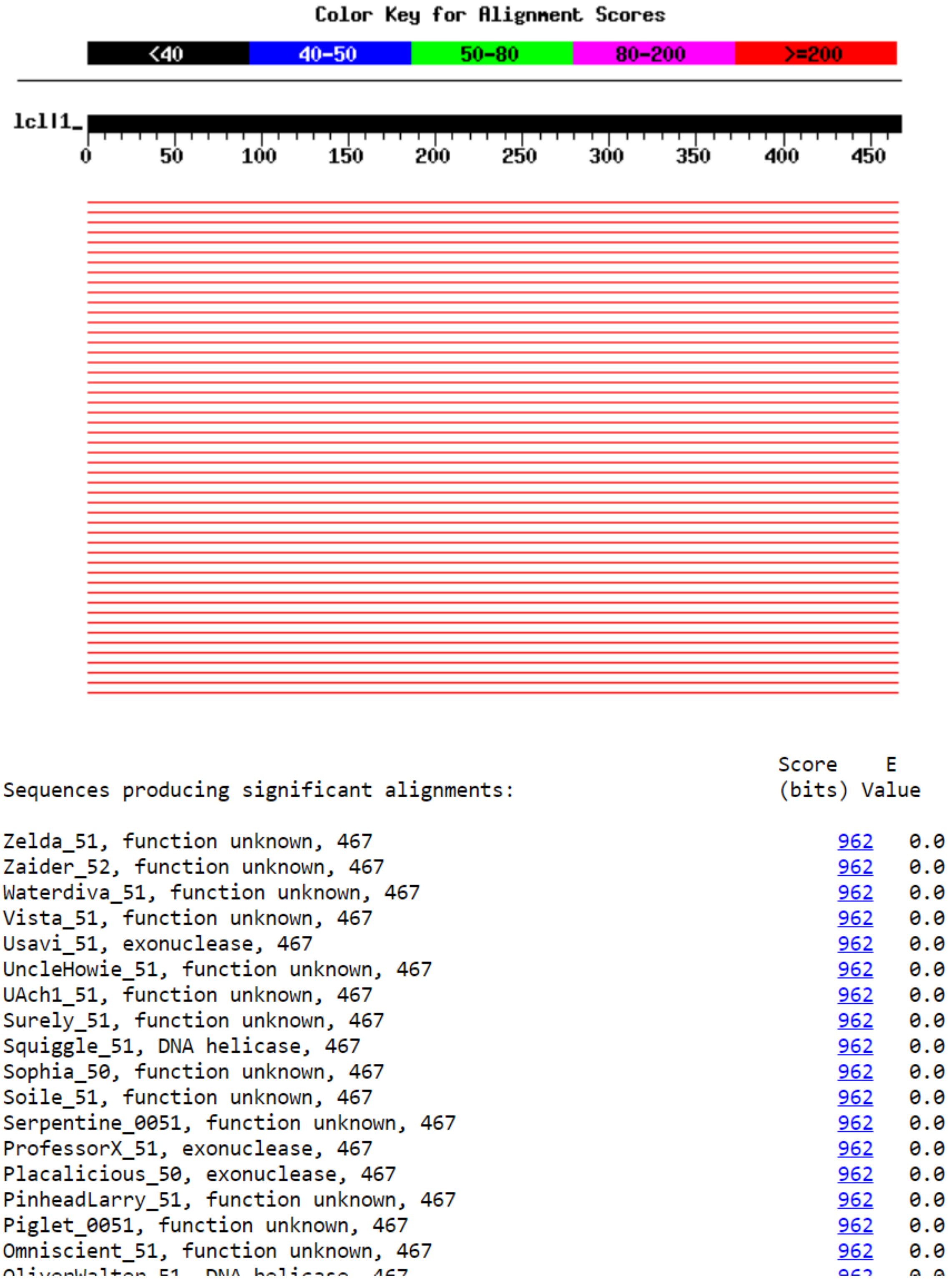
The selected gene is not the LORF with a length of 1569 nucleotides. It has a Z-score of 2.852, a spacer of 16, and a final score of -4.173. This is a reverse gene.

5. Function. If no functional prediction is present, write “Hypothetical protein”.

Exonuclease

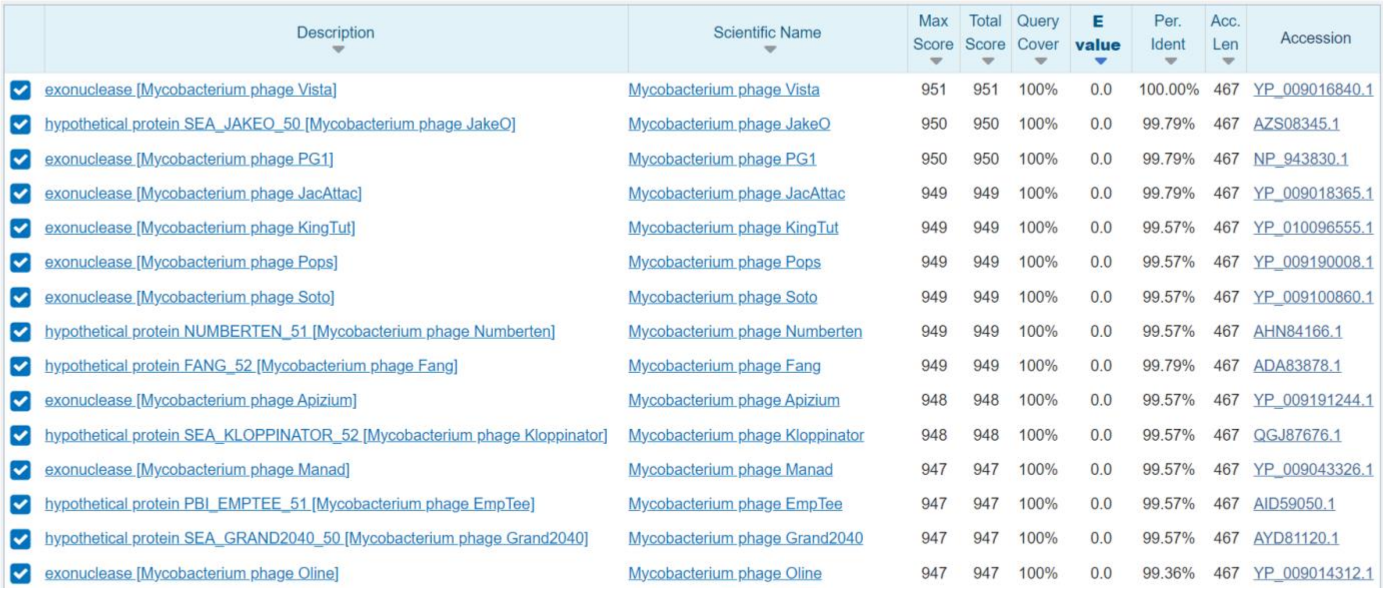
6. Supporting Information for Function (SIF)

a. SIF: PhageDb BLAST



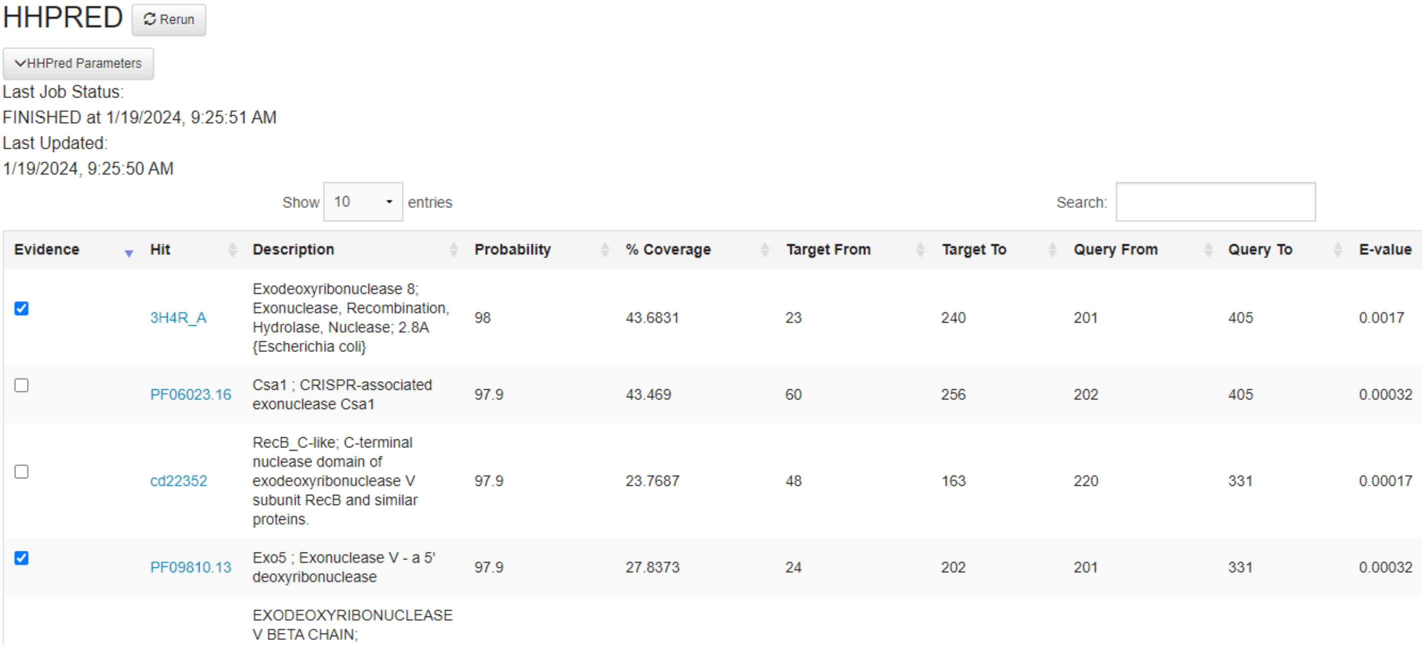
```
                                                                   Score     E
Sequences producing significant alignments:                       (bits)  Value

Zelda_51, function unknown, 467                                       962     0.0
Zaider_52, function unknown, 467                                      962     0.0
Waterdiva_51, function unknown, 467                                   962     0.0
Vista_51, function unknown, 467                                       962     0.0
Usavi_51, exonuclease, 467                                            962     0.0
UncleHowie_51, function unknown, 467                                  962     0.0
UAch1_51, function unknown, 467                                       962     0.0
Surely_51, function unknown, 467                                      962     0.0
Squiggle_51, DNA helicase, 467                                        962     0.0
Sophia_50, function unknown, 467                                      962     0.0
Soile_51, function unknown, 467                                       962     0.0
Serpentine_0051, function unknown, 467                                962     0.0
ProfessorX_51, exonuclease, 467                                       962     0.0
Placalicious_50, exonuclease, 467                                     962     0.0
PinheadLarry_51, function unknown, 467                                962     0.0
Piglet_0051, function unknown, 467                                    962     0.0
Omniscient_51, function unknown, 467                                  962     0.0
OliverWalter_51, DNA helicase, 467                                    962     0.0
```

b. SIF: NCBI BLAST



| | Description | Scientific Name | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | exonuclease [Mycobacterium phage Vista] | Mycobacterium phage Vista | 951 | 951 | 100% | 0.0 | 100.00% | 467 | YP_009016840.1 |
| ☑ | hypothetical protein SEA_JAKEO_50 [Mycobacterium phage JakeO] | Mycobacterium phage JakeO | 950 | 950 | 100% | 0.0 | 99.79% | 467 | AZS08345.1 |
| ☑ | exonuclease [Mycobacterium phage PG1] | Mycobacterium phage PG1 | 950 | 950 | 100% | 0.0 | 99.79% | 467 | NP_943830.1 |
| ☑ | exonuclease [Mycobacterium phage JacAttac] | Mycobacterium phage JacAttac | 949 | 949 | 100% | 0.0 | 99.79% | 467 | YP_009018365.1 |
| ☑ | exonuclease [Mycobacterium phage KingTut] | Mycobacterium phage KingTut | 949 | 949 | 100% | 0.0 | 99.57% | 467 | YP_010096555.1 |
| ☑ | exonuclease [Mycobacterium phage Pops] | Mycobacterium phage Pops | 949 | 949 | 100% | 0.0 | 99.57% | 467 | YP_009190008.1 |
| ☑ | exonuclease [Mycobacterium phage Soto] | Mycobacterium phage Soto | 949 | 949 | 100% | 0.0 | 99.57% | 467 | YP_009100860.1 |
| ☑ | hypothetical protein NUMBERTEN_51 [Mycobacterium phage Numberten] | Mycobacterium phage Numberten | 949 | 949 | 100% | 0.0 | 99.57% | 467 | AHN84166.1 |
| ☑ | hypothetical protein FANG_52 [Mycobacterium phage Fang] | Mycobacterium phage Fang | 949 | 949 | 100% | 0.0 | 99.79% | 467 | ADA83878.1 |
| ☑ | exonuclease [Mycobacterium phage Apizium] | Mycobacterium phage Apizium | 948 | 948 | 100% | 0.0 | 99.57% | 467 | YP_009191244.1 |
| ☑ | hypothetical protein SEA_KLOPPINATOR_52 [Mycobacterium phage Kloppinator] | Mycobacterium phage Kloppinator | 948 | 948 | 100% | 0.0 | 99.57% | 467 | QGJ87676.1 |
| ☑ | exonuclease [Mycobacterium phage Manad] | Mycobacterium phage Manad | 947 | 947 | 100% | 0.0 | 99.57% | 467 | YP_009043326.1 |
| ☑ | hypothetical protein PBI_EMPTEE_51 [Mycobacterium phage EmpTee] | Mycobacterium phage EmpTee | 947 | 947 | 100% | 0.0 | 99.57% | 467 | AID59050.1 |
| ☑ | hypothetical protein SEA_GRAND2040_50 [Mycobacterium phage Grand2040] | Mycobacterium phage Grand2040 | 947 | 947 | 100% | 0.0 | 99.57% | 467 | AYD81120.1 |
| ☑ | exonuclease [Mycobacterium phage Oline] | Mycobacterium phage Oline | 947 | 947 | 100% | 0.0 | 99.36% | 467 | YP_009014312.1 |

## c. SIF: HHPred

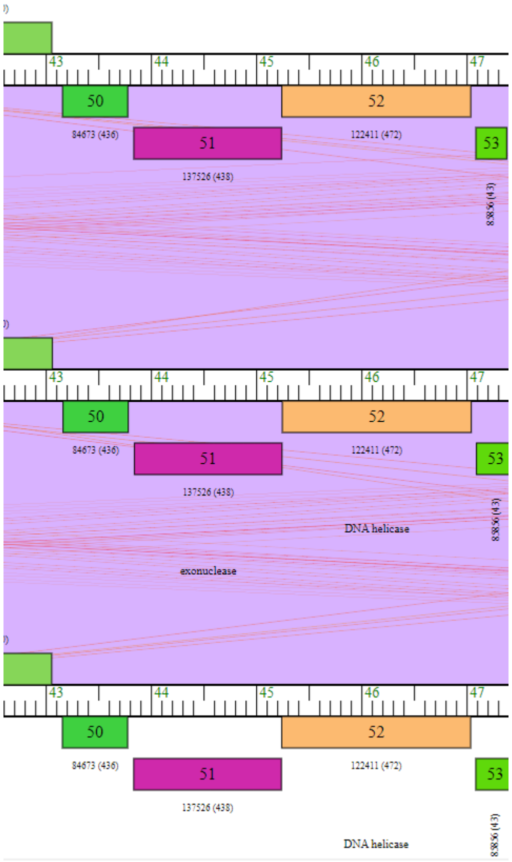


HHPRED

Last Job Status:
FINISHED at 1/19/2024, 9:25:51 AM
Last Updated:
1/19/2024, 9:25:50 AM

| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | 3H4R_A | Exodeoxyribonuclease 8; Exonuclease, Recombination, Hydrolase, Nuclease; 2.8A {Escherichia coli} | 98 | 43.6831 | 23 | 240 | 201 | 405 | 0.0017 |
| ☐ | PF06023.16 | Csa1 ; CRISPR-associated exonuclease Csa1 | 97.9 | 43.469 | 60 | 256 | 202 | 405 | 0.00032 |
| ☐ | cd22352 | RecB_C-like; C-terminal nuclease domain of exodeoxyribonuclease V subunit RecB and similar proteins. | 97.9 | 23.7687 | 48 | 163 | 220 | 331 | 0.00017 |
| ☑ | PF09810.13 | Exo5 ; Exonuclease V - a 5' deoxyribonuclease | 97.9 | 27.8373 | 24 | 202 | 201 | 331 | 0.00032 |
| | | EXODEOXYRIBONUCLEASE V BETA CHAIN; | | | | | | | |

## d. SIF: Synteny-Phamerator (three genomes)



7. Any other important information.

Synteny with Charles1, Frankicide, FugateOSU, Placalicious and others

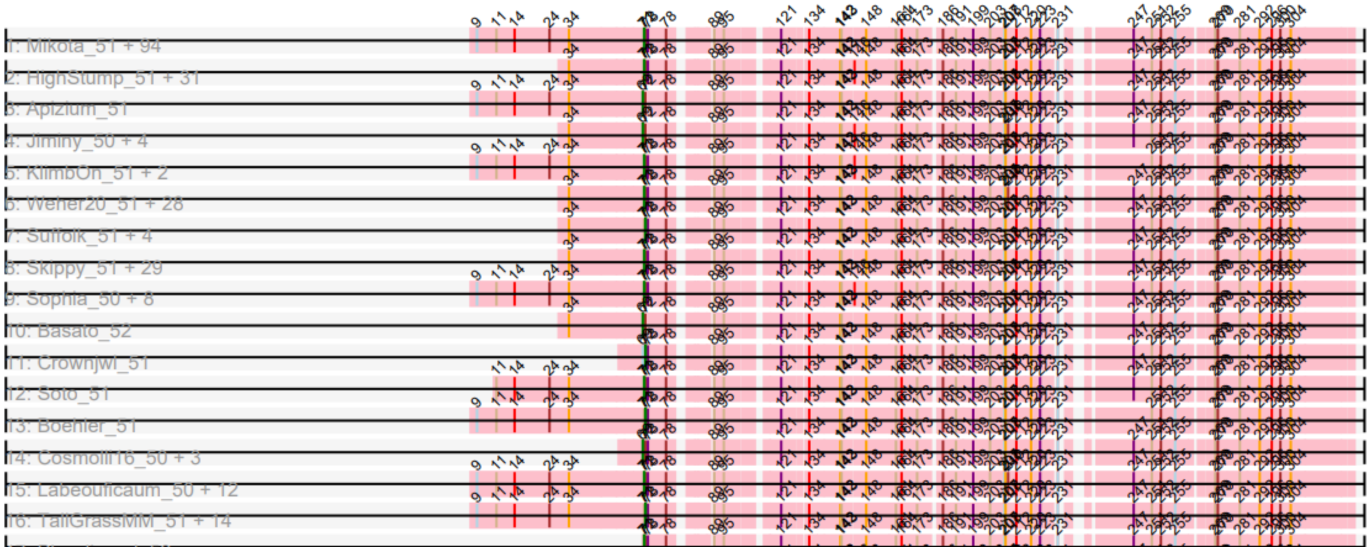
**CURATOR NAME: BELLA N**

**GENE NAME: EUGENIA-52**

**DNA MASTER NOTES: N/A**

**START POSITION EVALUATION (IN ORDER)**:

1. Starterator

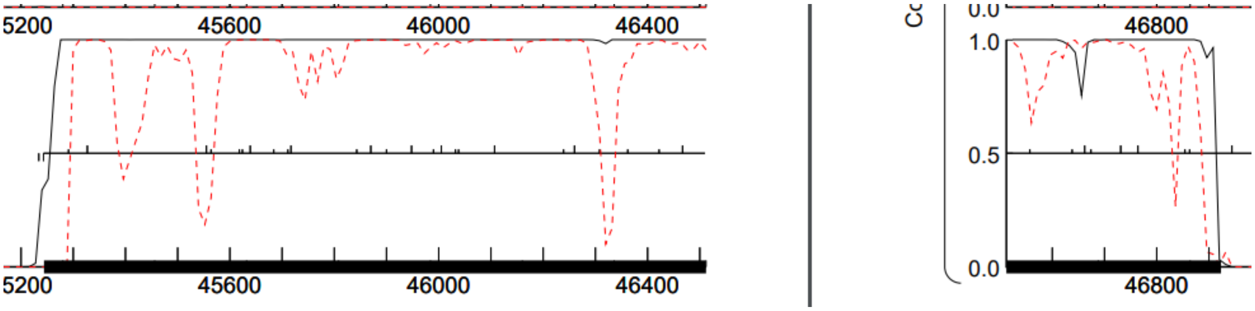


The start number called the most often in the published annotations is 71, it was called in 201 of the 389 non-draft genes in the pham.

Start 71:

- Found in 248 of 441 ( 56.2% ) of genes in pham
- Manual Annotations of this start: 201 of 389
- Called 87.5% of time when present

2. GeneMark coding potential



The black bar at the bottom represents the whole ORF it is about 1800 nucleotides and this is a reverse gene.

3. Glimmer and GeneMark agreement

There is not agreement between Glimmer and GeneMark

Glimmer Start: 47040

Glimmer Score: 12.38

GeneMark Start: 46923

4. Longest open reading frame (ORF) without excessive gap

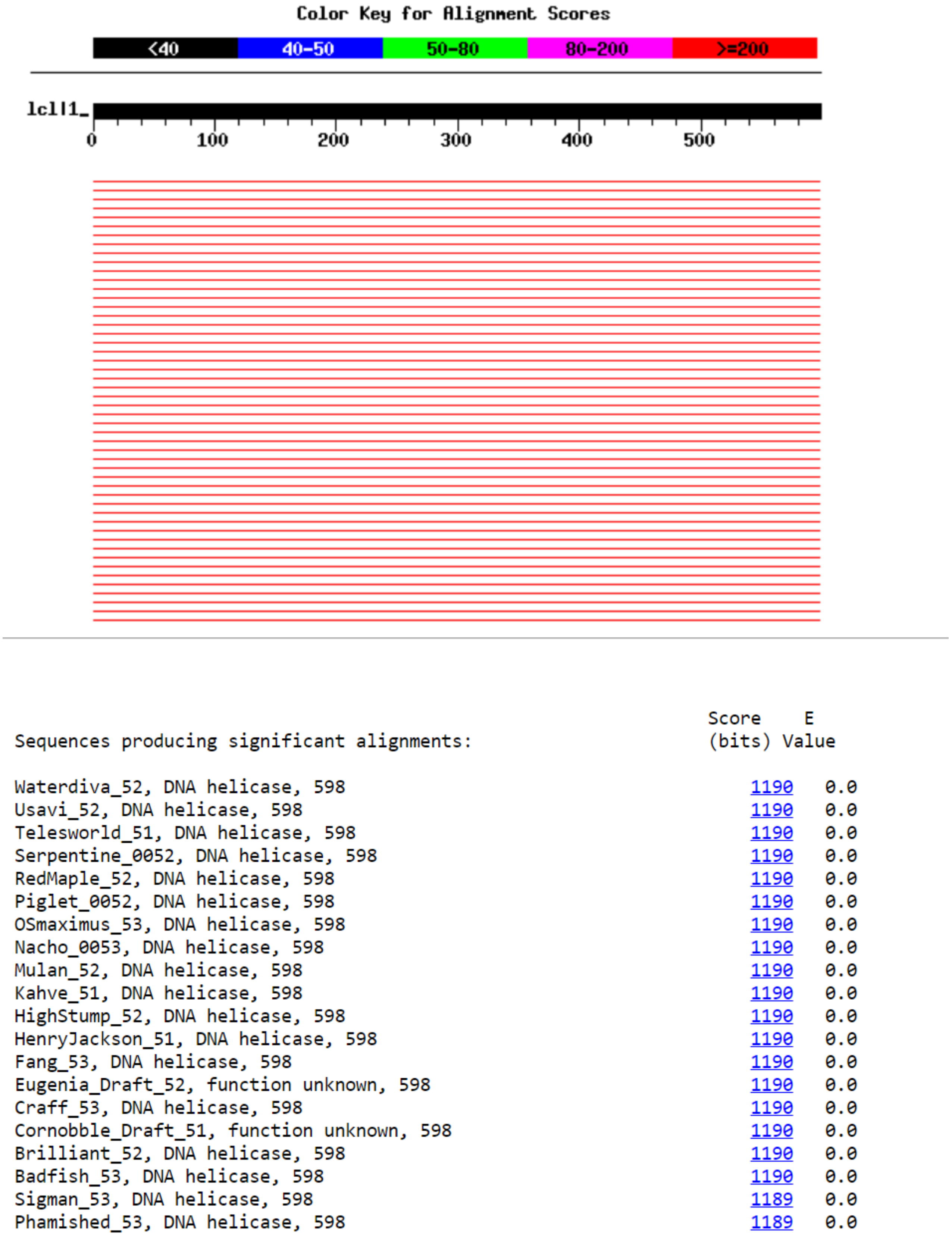
The selected gene is the LORF with a length of 1797 nucleotides. It has a Z-score of 2.222, a spacer of 17, and a final score of -5.642. This is a reverse gene.

5. Function. If no functional prediction is present, write “Hypothetical protein”.

DNA helicase

6. Supporting Information for Function (SIF)

a. SIF: PhageDb BLAST

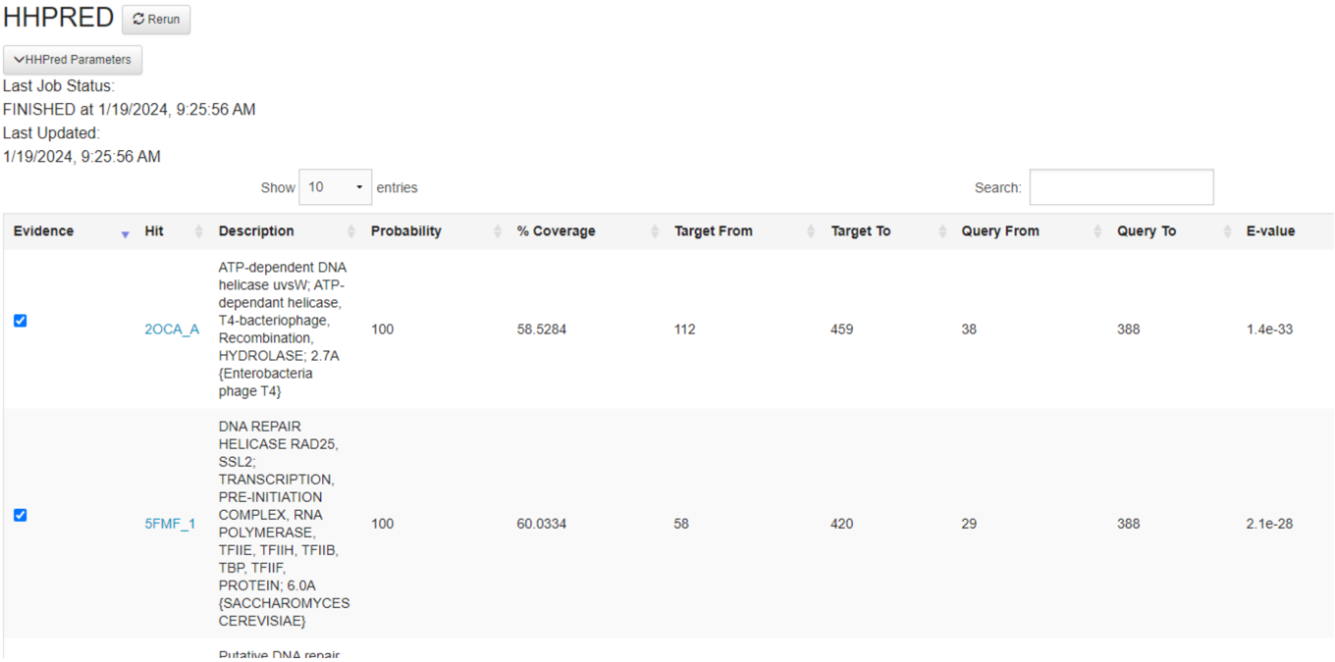


```
                                                                  Score     E
Sequences producing significant alignments:                       (bits)  Value

Waterdiva_52, DNA helicase, 598                                      1190    0.0
Usavi_52, DNA helicase, 598                                          1190    0.0
Telesworld_51, DNA helicase, 598                                     1190    0.0
Serpentine_0052, DNA helicase, 598                                   1190    0.0
RedMaple_52, DNA helicase, 598                                       1190    0.0
Piglet_0052, DNA helicase, 598                                       1190    0.0
OSmaximus_53, DNA helicase, 598                                      1190    0.0
Nacho_0053, DNA helicase, 598                                        1190    0.0
Mulan_52, DNA helicase, 598                                          1190    0.0
Kahve_51, DNA helicase, 598                                          1190    0.0
HighStump_52, DNA helicase, 598                                      1190    0.0
HenryJackson_51, DNA helicase, 598                                   1190    0.0
Fang_53, DNA helicase, 598                                           1190    0.0
Eugenia_Draft_52, function unknown, 598                              1190    0.0
Craff_53, DNA helicase, 598                                          1190    0.0
Cornobble_Draft_51, function unknown, 598                            1190    0.0
Brilliant_52, DNA helicase, 598                                      1190    0.0
Badfish_53, DNA helicase, 598                                        1190    0.0
Sigman_53, DNA helicase, 598                                         1189    0.0
Phamished_53, DNA helicase, 598                                      1189    0.0
```

## b. SIF: NCBI BLAST

| | Description | Scientific Name | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Acces |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | DNA helicase [Mycobacterium phage Badfish] | Mycobacterium phage Badfish | 1221 | 1221 | 100% | 0.0 | 100.00% | 598 | YP_0091 |
| ☑ | DNA helicase [Mycobacterium phage Lego3393] | Mycobacterium phage Lego3393 | 1220 | 1220 | 99% | 0.0 | 100.00% | 598 | AOT2740 |
| ☑ | DNA helicase [Mycobacterium phage JacAttac] | Mycobacterium phage JacAttac | 1220 | 1220 | 100% | 0.0 | 99.83% | 598 | YP_0090 |
| ☑ | DNA helicase [Mycobacterium phage Kikipoo] | Mycobacterium phage Kikipoo | 1219 | 1219 | 100% | 0.0 | 99.83% | 598 | YP_0092 |
| ☑ | DNA helicase [Mycobacterium phage Swish] | Mycobacterium phage Swish | 1218 | 1218 | 100% | 0.0 | 99.67% | 598 | YP_0091 |
| ☑ | DNA helicase [Mycobacterium phage Adriana] | Mycobacterium phage Adriana | 1215 | 1215 | 100% | 0.0 | 99.50% | 598 | QNO1206 |
| ☑ | DNA helicase [Mycobacterium phage Melc17] | Mycobacterium phage Melc17 | 1198 | 1198 | 100% | 0.0 | 98.17% | 602 | WAB1000 |
| ☑ | DNA helicase [Mycobacterium phage Oline] | Mycobacterium phage Oline | 1184 | 1184 | 100% | 0.0 | 97.18% | 598 | YP_0090 |
| ☑ | DNA helicase [Mycobacterium phage Anderson] | Mycobacterium phage Anderson | 1183 | 1183 | 100% | 0.0 | 97.01% | 598 | QSM0026 |
| ☑ | DNA helicase [Mycobacterium phage Crownjwl] | Mycobacterium phage Crownjwl | 1183 | 1183 | 100% | 0.0 | 97.35% | 599 | AYD8333 |
| ☑ | DNA helicase [Mycobacterium phage BlueHusk] | Mycobacterium phage BlueHusk | 1182 | 1182 | 100% | 0.0 | 97.01% | 598 | WNM666 |
| ☑ | DNA helicase [Mycobacterium phage Kailash] | Mycobacterium phage Kailash | 1182 | 1182 | 100% | 0.0 | 97.01% | 598 | ATN8966 |
| ☑ | DNA helicase [Mycobacterium phage True] | Mycobacterium phage True | 1181 | 1181 | 100% | 0.0 | 96.84% | 598 | QWY8177 |
| ☑ | DNA helicase [Mycobacterium phage Duggie] | Mycobacterium phage Duggie | 1181 | 1181 | 100% | 0.0 | 97.01% | 598 | QGJ8750 |
| ☑ | DNA helicase [Mycobacterium phage Robyn] | Mycobacterium phage Robyn | 1180 | 1180 | 100% | 0.0 | 96.36% | 601 | QBI99591 |
| ☑ | DNA helicase [Mycobacterium phage AltPhacts] | Mycobacterium phage AltPhacts | 1179 | 1179 | 100% | 0.0 | 97.19% | 600 | AVO2460 |

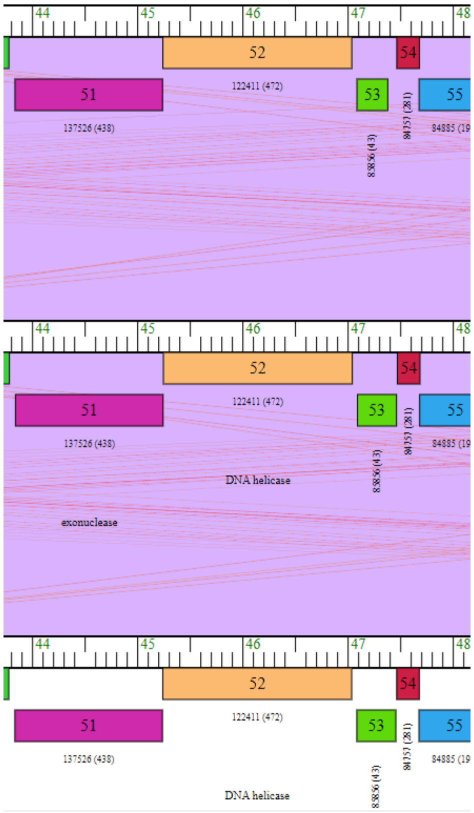
## c. SIF: HHPred



Last Job Status:
FINISHED at 1/19/2024, 9:25:56 AM
Last Updated:
1/19/2024, 9:25:56 AM

| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | 2OCA_A | ATP-dependent DNA helicase uvsW; ATP-dependant helicase, T4-bacteriophage, Recombination, HYDROLASE; 2.7A {Enterobacteria phage T4} | 100 | 58.5284 | 112 | 459 | 38 | 388 | 1.4e-33 |
| ☑ | 5FMF_1 | DNA REPAIR HELICASE RAD25, SSL2; TRANSCRIPTION, PRE-INITIATION COMPLEX, RNA POLYMERASE, TFIIE, TFIIH, TFIIB, TBP, TFIIF, PROTEIN; 6.0A {SACCHAROMYCES CEREVISIAE} | 100 | 60.0334 | 58 | 420 | 29 | 388 | 2.1e-28 |
| | | Putative DNA repair | | | | | | | |

## d. SIF: Synteny-Phamerator (three genomes)



7. Any other important information.

Synteny with Badfish, Brilliant, Craff and more.

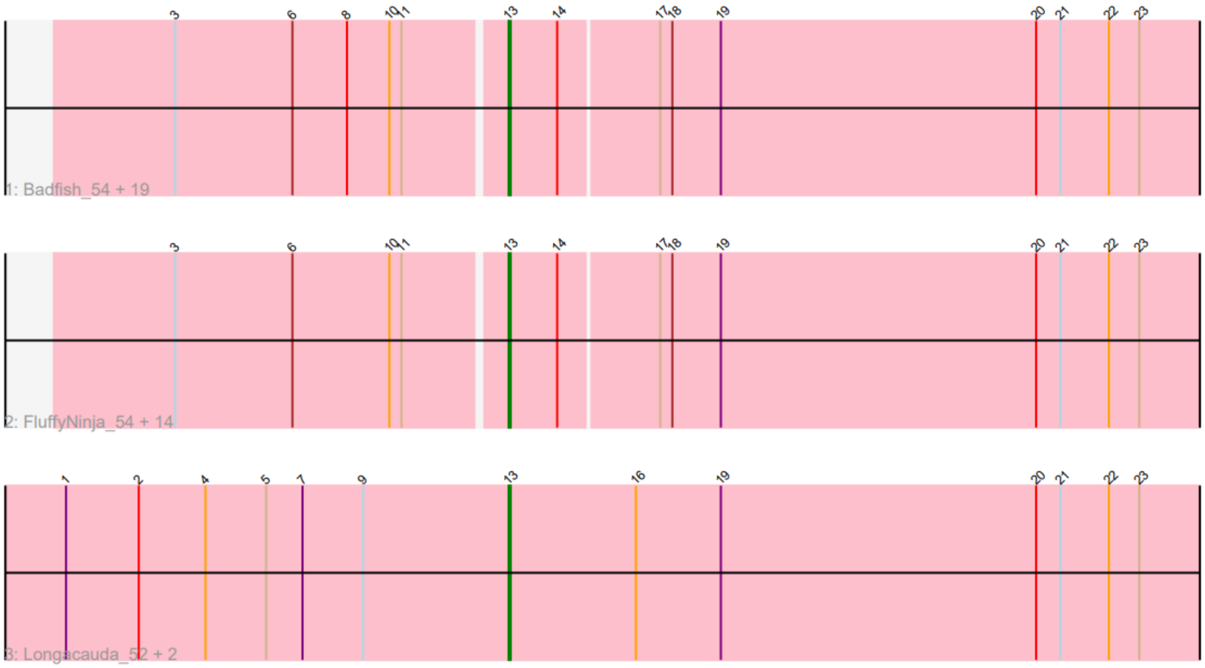
## CURATOR NAME: BELLA N

## GENE NAME: EUGENIA-53

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

1. Starterator

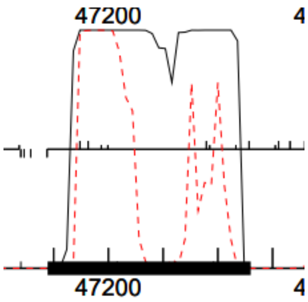


The start number called the most often in the published annotations is 13, it was called in 40 of the 41 non-draft genes in the pham.
Start 13:

- Found in 42 of 43 ( 97.7% ) of genes in pham
- Manual Annotations of this start: 40 of 41

- Called 95.2% of time when present

2. GeneMark coding potential



The black bar at the bottom represents the whole ORF it is about 400 nucleotides and this is a reverse gene.

3. Glimmer and GeneMark agreement

There is not agreement between Glimmer and GeneMark

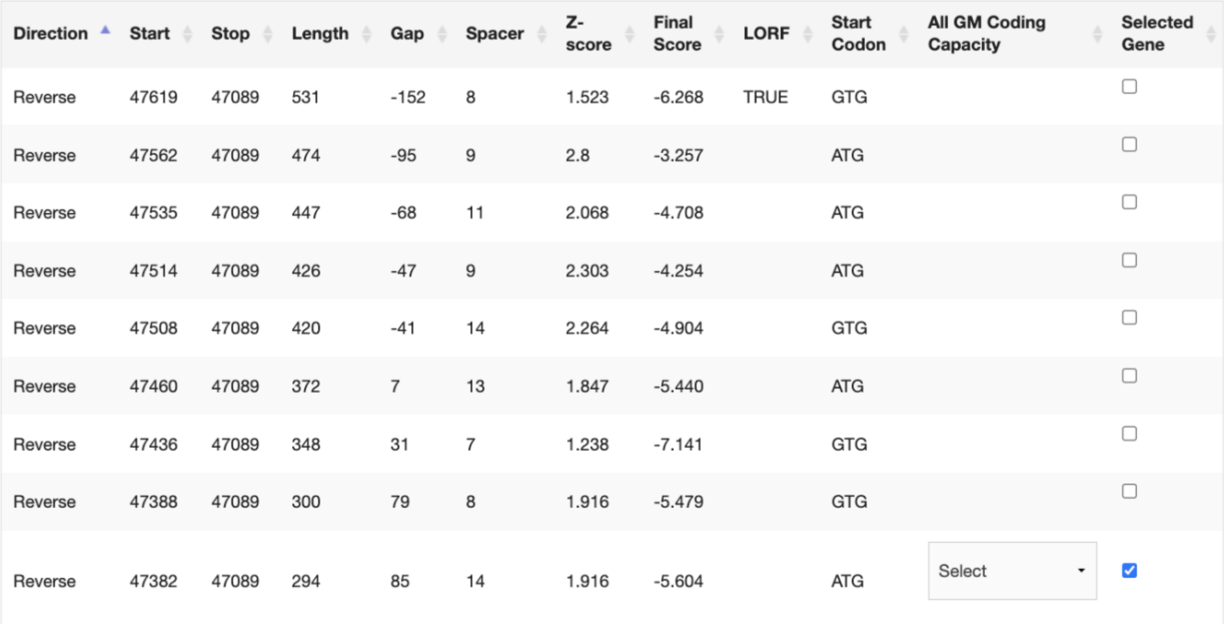
Glimmer Start: 47382

Glimmer Score: 4.75

GeneMark Start: 47460

4. Longest open reading frame (ORF) without excessive gap

No.



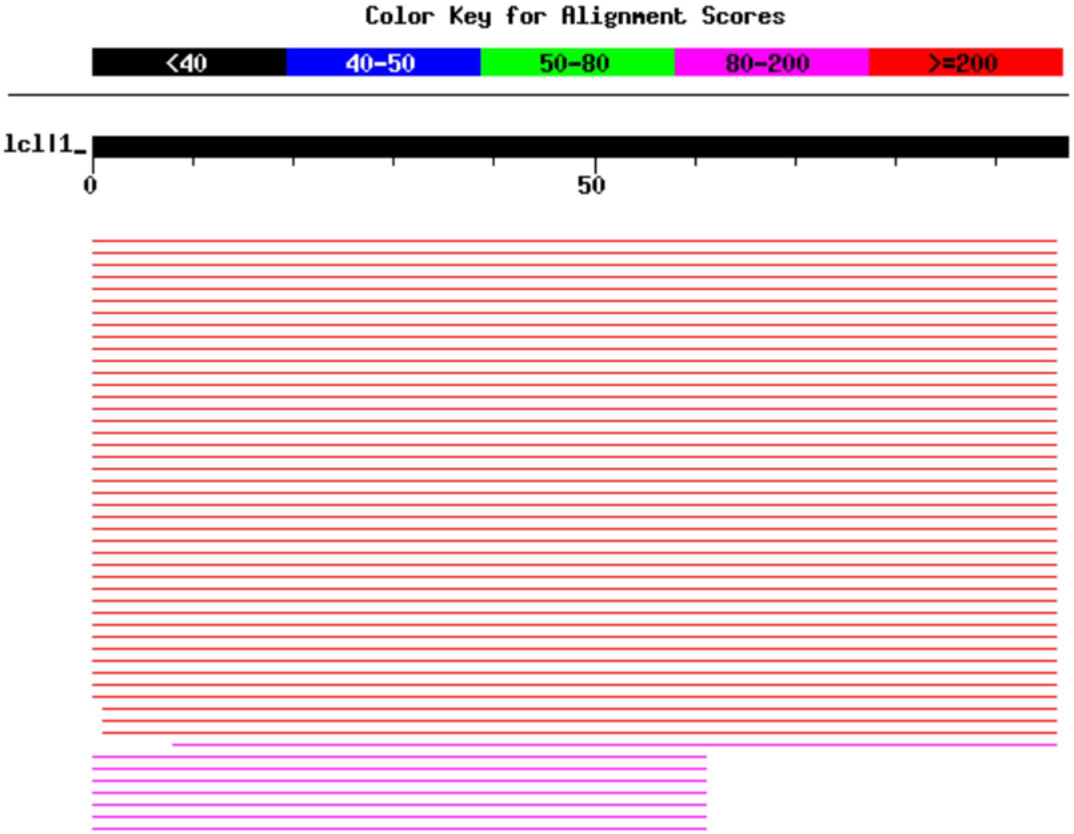
| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 47619 | 47089 | 531 | -152 | 8 | 1.523 | -6.268 | TRUE | GTG | | ☐ |
| Reverse | 47562 | 47089 | 474 | -95 | 9 | 2.8 | -3.257 | | ATG | | ☐ |
| Reverse | 47535 | 47089 | 447 | -68 | 11 | 2.068 | -4.708 | | ATG | | ☐ |
| Reverse | 47514 | 47089 | 426 | -47 | 9 | 2.303 | -4.254 | | ATG | | ☐ |
| Reverse | 47508 | 47089 | 420 | -41 | 14 | 2.264 | -4.904 | | GTG | | ☐ |
| Reverse | 47460 | 47089 | 372 | 7 | 13 | 1.847 | -5.440 | | ATG | | ☐ |
| Reverse | 47436 | 47089 | 348 | 31 | 7 | 1.238 | -7.141 | | GTG | | ☐ |
| Reverse | 47388 | 47089 | 300 | 79 | 8 | 1.916 | -5.479 | | GTG | | ☐ |
| Reverse | 47382 | 47089 | 294 | 85 | 14 | 1.916 | -5.604 | | ATG | Select | ☑ |

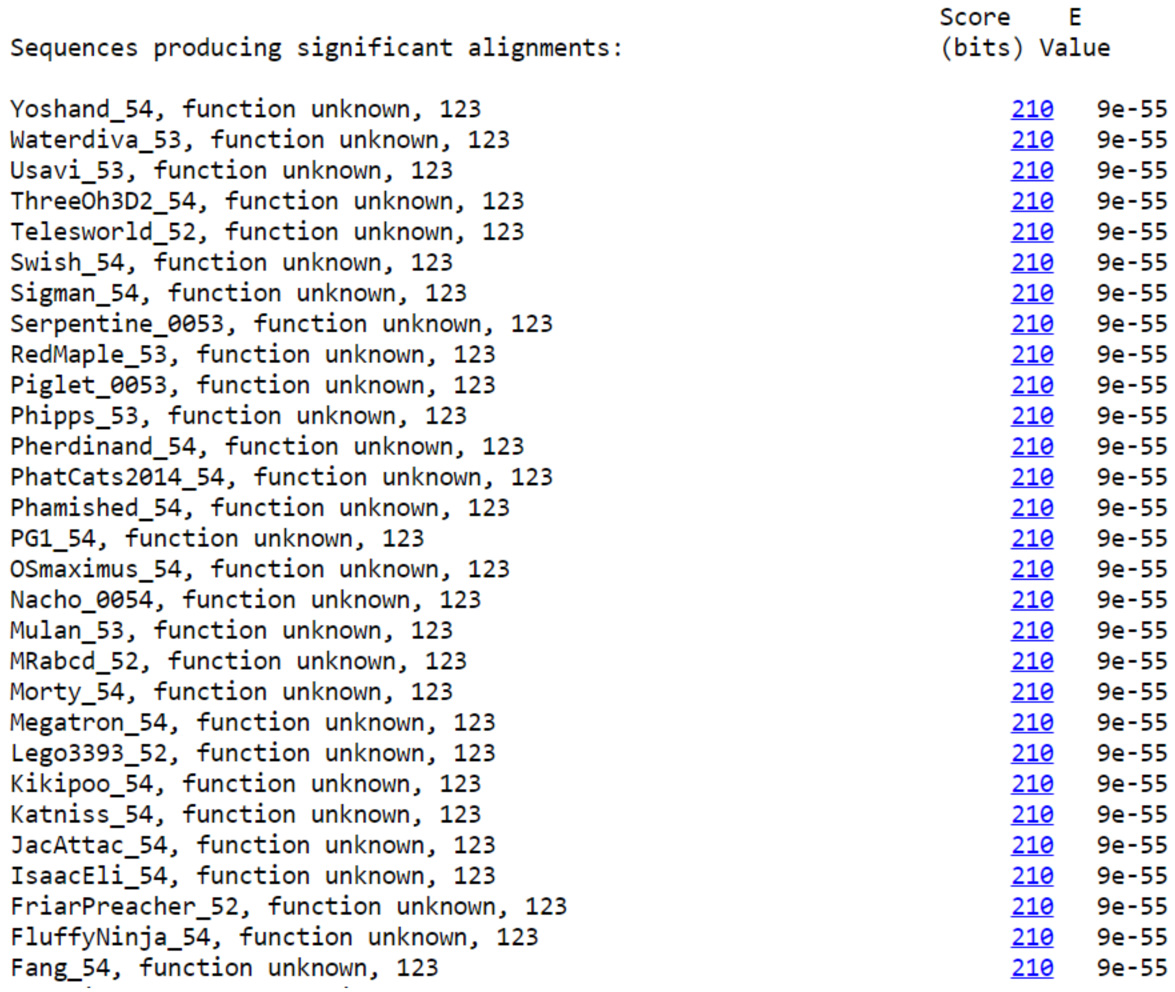
5. Function. If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

6. Supporting Information for Function (SIF)

   a. SIF: PhageDb BLAST


Color Key for Alignment Scores
<40
40-50
50-80
80-200
>=200
lcl|1_
0
50

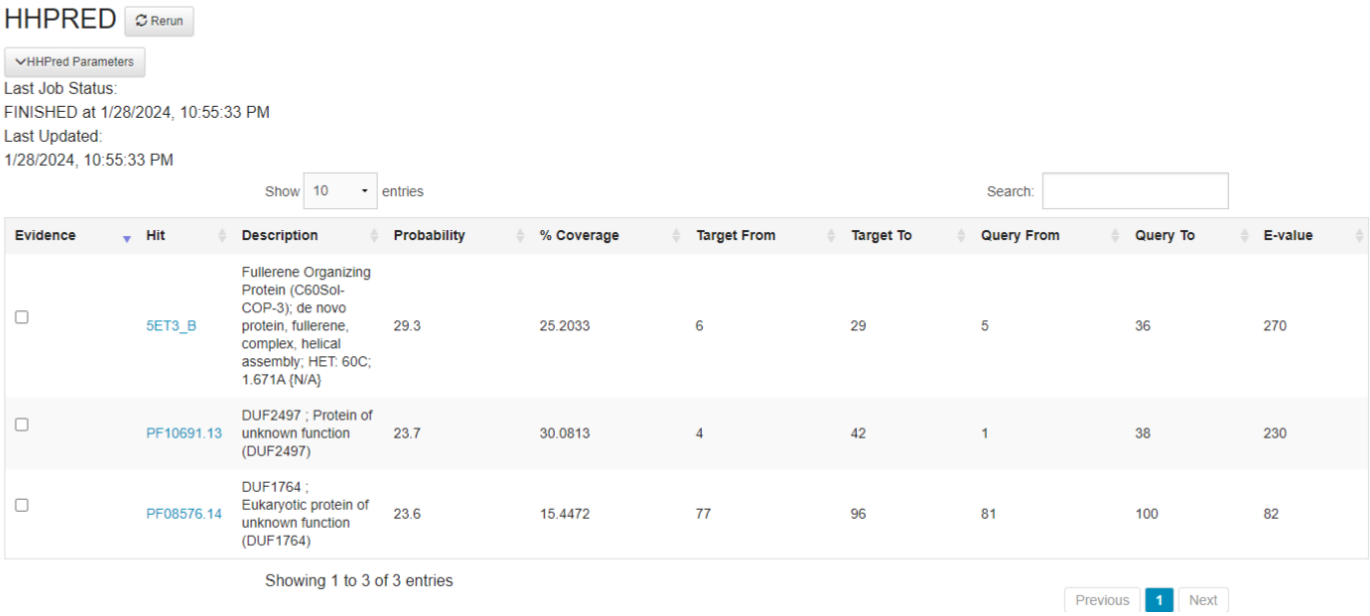


```
                                                                Score     E
Sequences producing significant alignments:                     (bits) Value

Yoshand_54, function unknown, 123                                 210    9e-55
Waterdiva_53, function unknown, 123                               210    9e-55
Usavi_53, function unknown, 123                                   210    9e-55
ThreeOh3D2_54, function unknown, 123                              210    9e-55
Telesworld_52, function unknown, 123                              210    9e-55
Swish_54, function unknown, 123                                   210    9e-55
Sigman_54, function unknown, 123                                  210    9e-55
Serpentine_0053, function unknown, 123                            210    9e-55
RedMaple_53, function unknown, 123                                210    9e-55
Piglet_0053, function unknown, 123                                210    9e-55
Phipps_53, function unknown, 123                                  210    9e-55
Pherdinand_54, function unknown, 123                              210    9e-55
PhatCats2014_54, function unknown, 123                            210    9e-55
Phamished_54, function unknown, 123                               210    9e-55
PG1_54, function unknown, 123                                     210    9e-55
OSmaximus_54, function unknown, 123                               210    9e-55
Nacho_0054, function unknown, 123                                 210    9e-55
Mulan_53, function unknown, 123                                   210    9e-55
MRabcd_52, function unknown, 123                                  210    9e-55
Morty_54, function unknown, 123                                   210    9e-55
Megatron_54, function unknown, 123                                210    9e-55
Lego3393_52, function unknown, 123                                210    9e-55
Kikipoo_54, function unknown, 123                                 210    9e-55
Katniss_54, function unknown, 123                                 210    9e-55
JacAttac_54, function unknown, 123                                210    9e-55
IsaacEli_54, function unknown, 123                                210    9e-55
FriarPreacher_52, function unknown, 123                           210    9e-55
FluffyNinja_54, function unknown, 123                             210    9e-55
Fang_54, function unknown, 123                                    210    9e-55
```

## b. SIF: NCBI BLAST

| | Description | Scientific Name | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Ac |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | hypothetical protein SEA_LEGO3393_52 [Mycobacterium phage Lego3393] | Mycobacterium phage Lego3393 | 202 | 202 | 100% | 6e-65 | 100.00% | 123 | AOT27 |
| ☑ | hypothetical protein PBI_PG1_54 [Mycobacterium phage PG1] | Mycobacterium phage PG1 | 201 | 201 | 100% | 1e-64 | 100.00% | 123 | NP_94 |
| ☑ | gp54 [Mycobacterium phage Orion] | Mycobacterium phage Orion | 199 | 199 | 100% | 6e-64 | 98.97% | 123 | YP_65 |
| ☑ | hypothetical protein SEA_LONGACAUDA_52 [Mycobacterium phage Longacau… | Mycobacterium phage Longacauda | 194 | 194 | 98% | 8e-62 | 96.88% | 124 | ATN90 |
| ☑ | hypothetical protein SEA_FRINGE_54 [Mycobacterium phage Fringe] | Mycobacterium phage Fringe | 186 | 186 | 91% | 8e-59 | 98.88% | 89 | AZS07 |
| ☑ | hypothetical protein SEA_GRAND2040_53 [Mycobacterium phage Grand2040] | Mycobacterium phage Grand2040 | 128 | 128 | 64% | 6e-36 | 95.24% | 89 | AYD81 |
| ☑ | hypothetical protein SEA_TRUE_52 [Mycobacterium phage True] | Mycobacterium phage True | 133 | 133 | 63% | 8e-36 | 98.39% | 279 | QWY8 |
| ☑ | hypothetical protein VORTEX_53 [Mycobacterium phage Vortex] | Mycobacterium phage Vortex | 127 | 127 | 67% | 8e-36 | 90.77% | 97 | YP_00 |
| ☑ | hypothetical protein SEA_WEHER20_53 [Mycobacterium phage Weher20] | Mycobacterium phage Weher20 | 133 | 133 | 63% | 9e-36 | 98.39% | 279 | UAJ16 |
| ☑ | hypothetical protein SEA_FOZZIE_52 [Mycobacterium phage Fozzie] | Mycobacterium phage Fozzie | 127 | 127 | 67% | 2e-35 | 90.77% | 97 | WNM7 |
| ☑ | hypothetical protein SEA_LONGACAUDA_54 [Mycobacterium phage Longacau… | Mycobacterium phage Longacauda | 126 | 126 | 64% | 2e-35 | 93.65% | 89 | ATN90 |
| ☑ | hypothetical protein OOSTERBAAN_53 [Mycobacterium phage Oosterbaan] | Mycobacterium phage Oosterbaan | 126 | 126 | 67% | 4e-35 | 89.23% | 97 | AEK07 |
| ☑ | hypothetical protein THORA_53 [Mycobacterium phage Thora] | Mycobacterium phage Thora | 126 | 126 | 67% | 4e-35 | 89.23% | 97 | AEJ91 |

### c. SIF: HHPred



HHPRED

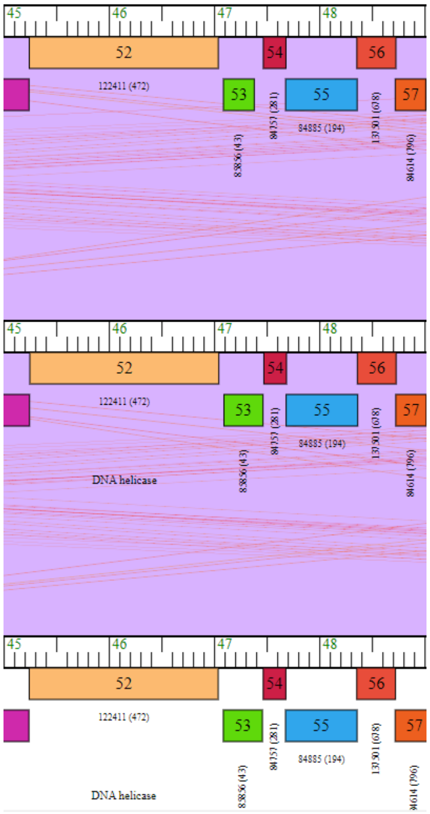
Last Job Status:
FINISHED at 1/28/2024, 10:55:33 PM
Last Updated:
1/28/2024, 10:55:33 PM

| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☐ | 5ET3_B | Fullerene Organizing Protein (C60Sol-COP-3); de novo protein, fullerene, complex, helical assembly; HET: 60C; 1.671A {N/A} | 29.3 | 25.2033 | 6 | 29 | 5 | 36 | 270 |
| ☐ | PF10691.13 | DUF2497 ; Protein of unknown function (DUF2497) | 23.7 | 30.0813 | 4 | 42 | 1 | 38 | 230 |
| ☐ | PF08576.14 | DUF1764 ; Eukaryotic protein of unknown function (DUF1764) | 23.6 | 15.4472 | 77 | 96 | 81 | 100 | 82 |

Showing 1 to 3 of 3 entries

No HHPRED evidence selected

### d. SIF: Synteny-Phamerator (three genomes)



7. Any other important information.

Synteny with ABU, Adriana, Badfish, and more.

No TmHmm

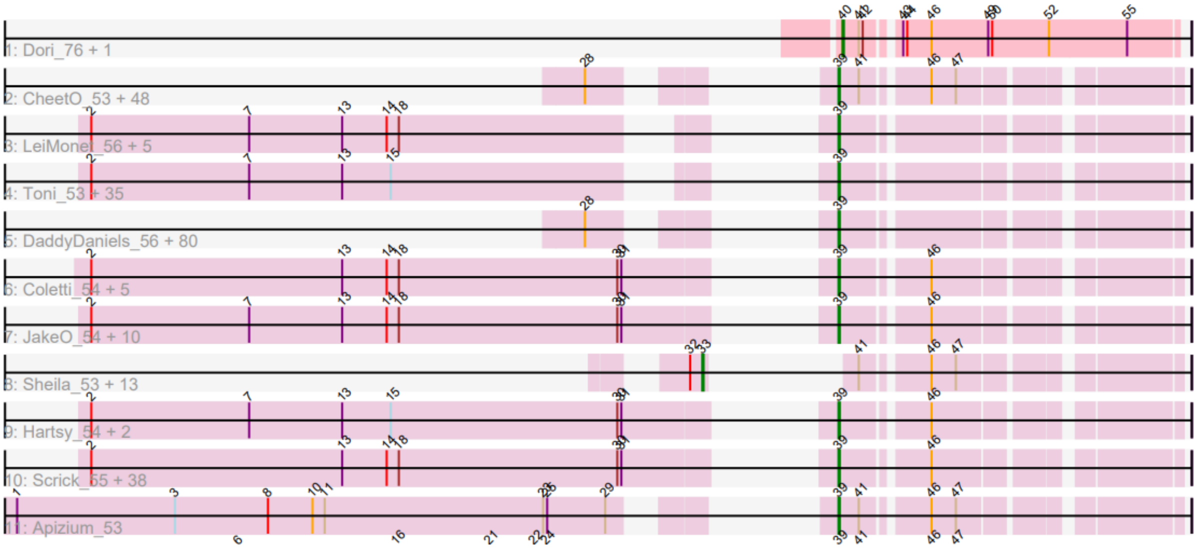
## CURATOR NAME: BELLA N

## GENE NAME: EUGENIA-54

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

1. Starterator

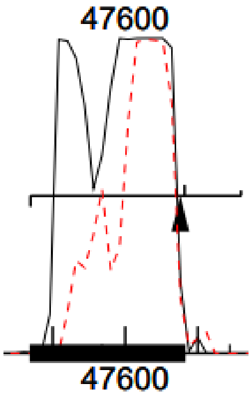


The start number called the most often in the published annotations is 39, it was called in 231 of the 259 non-draft genes in the pham.

Start 39:

- Found in 255 of 283 ( 90.1% ) of genes in pham
- Manual Annotations of this start: 231 of 259
- Called 99.6% of time when present

2. GeneMark coding potential



The black bar at the bottom represents the whole ORF, it is about 200 nucleotides long. The triangle present would indicate a shift, but since its function is not a tail assembly chaperone, it can be ignored.

3. Glimmer and GeneMark agreement

There is agreement between Glimmer and GeneMark

Glimmer Start: 47683

Glimmer Score: 13.71

GeneMark Start: 47683

4. Longest open reading frame (ORF) without excessive gap

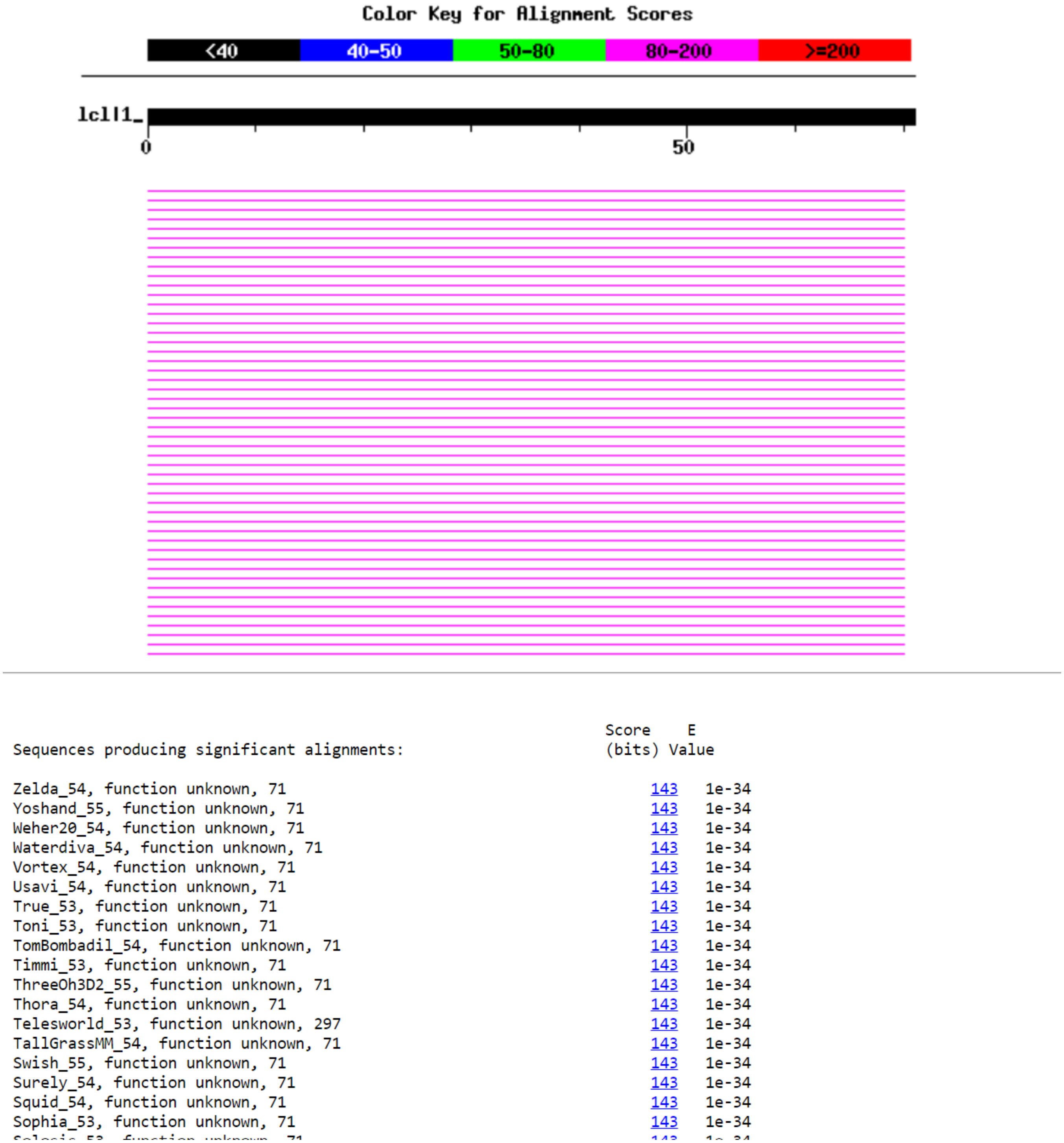
The selected gene is not the LORF with a length of 216 nucleotides. It has a Z-score of 3.229, a spacer of 11, and a final score of -2.377. It is a reverse gene.

5. Function. If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

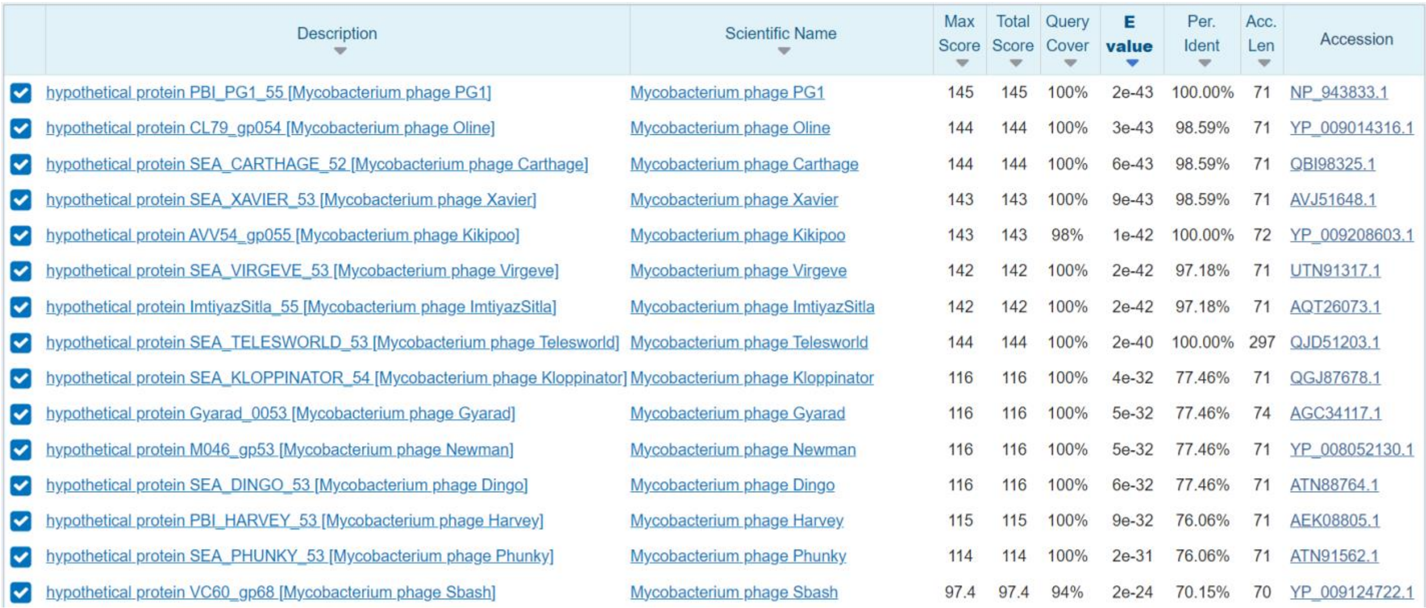
# 6. Supporting Information for Function (SIF)

## a. SIF: PhageDb BLAST



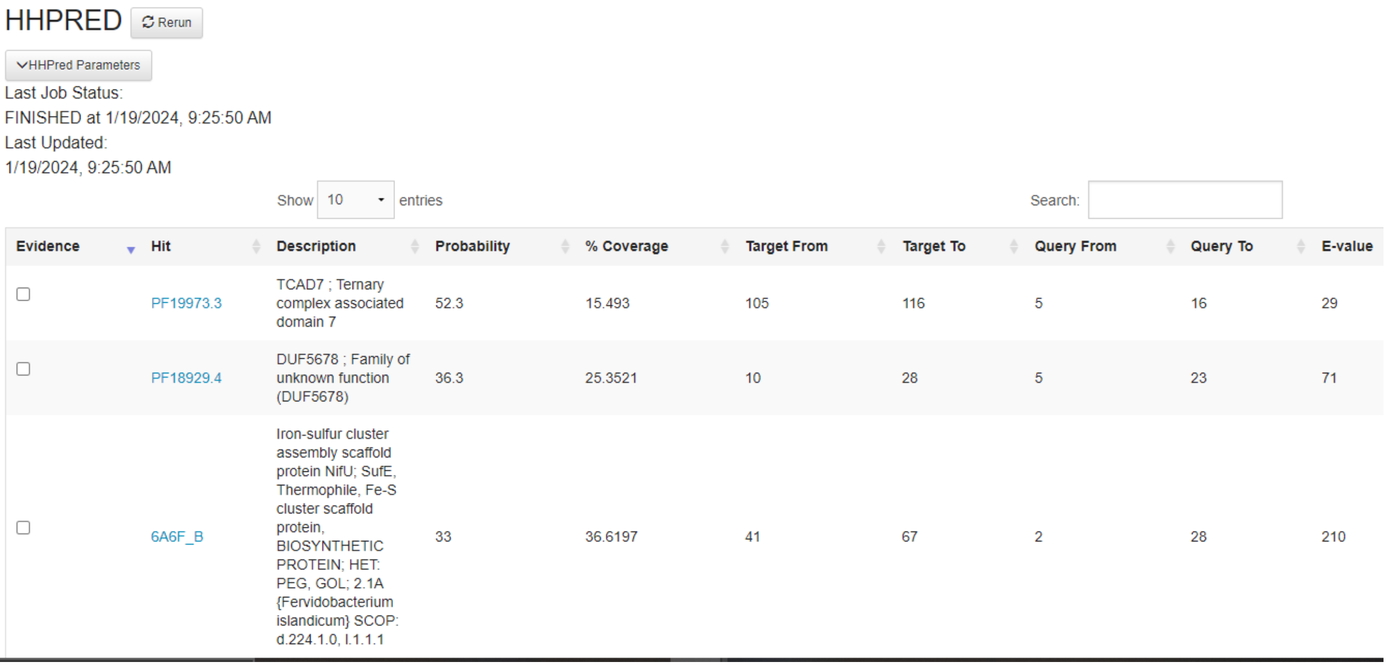
```
                                                                 Score     E
Sequences producing significant alignments:                      (bits)  Value

Zelda_54, function unknown, 71                                      143    1e-34
Yoshand_55, function unknown, 71                                    143    1e-34
Weher20_54, function unknown, 71                                    143    1e-34
Waterdiva_54, function unknown, 71                                  143    1e-34
Vortex_54, function unknown, 71                                     143    1e-34
Usavi_54, function unknown, 71                                      143    1e-34
True_53, function unknown, 71                                       143    1e-34
Toni_53, function unknown, 71                                       143    1e-34
TomBombadil_54, function unknown, 71                                143    1e-34
Timmi_53, function unknown, 71                                      143    1e-34
ThreeOh3D2_55, function unknown, 71                                 143    1e-34
Thora_54, function unknown, 71                                      143    1e-34
Telesworld_53, function unknown, 297                                143    1e-34
TallGrassMM_54, function unknown, 71                                143    1e-34
Swish_55, function unknown, 71                                      143    1e-34
Surely_54, function unknown, 71                                     143    1e-34
Squid_54, function unknown, 71                                      143    1e-34
Sophia_53, function unknown, 71                                     143    1e-34
```

## b. SIF: NCBI BLAST



| | Description | Scientific Name | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | hypothetical protein PBI_PG1_55 [Mycobacterium phage PG1] | Mycobacterium phage PG1 | 145 | 145 | 100% | 2e-43 | 100.00% | 71 | NP_943833.1 |
| ☑ | hypothetical protein CL79_gp054 [Mycobacterium phage Oline] | Mycobacterium phage Oline | 144 | 144 | 100% | 3e-43 | 98.59% | 71 | YP_009014316.1 |
| ☑ | hypothetical protein SEA_CARTHAGE_52 [Mycobacterium phage Carthage] | Mycobacterium phage Carthage | 144 | 144 | 100% | 6e-43 | 98.59% | 71 | QBI98325.1 |
| ☑ | hypothetical protein SEA_XAVIER_53 [Mycobacterium phage Xavier] | Mycobacterium phage Xavier | 143 | 143 | 100% | 9e-43 | 98.59% | 71 | AVJ51648.1 |
| ☑ | hypothetical protein AVV54_gp055 [Mycobacterium phage Kikipoo] | Mycobacterium phage Kikipoo | 143 | 143 | 98% | 1e-42 | 100.00% | 72 | YP_009208603.1 |
| ☑ | hypothetical protein SEA_VIRGEVE_53 [Mycobacterium phage Virgeve] | Mycobacterium phage Virgeve | 142 | 142 | 100% | 2e-42 | 97.18% | 71 | UTN91317.1 |
| ☑ | hypothetical protein ImtiyazSitla_55 [Mycobacterium phage ImtiyazSitla] | Mycobacterium phage ImtiyazSitla | 142 | 142 | 100% | 2e-42 | 97.18% | 71 | AQT26073.1 |
| ☑ | hypothetical protein SEA_TELESWORLD_53 [Mycobacterium phage Telesworld] | Mycobacterium phage Telesworld | 144 | 144 | 100% | 2e-40 | 100.00% | 297 | QJD51203.1 |
| ☑ | hypothetical protein SEA_KLOPPINATOR_54 [Mycobacterium phage Kloppinator] | Mycobacterium phage Kloppinator | 116 | 116 | 100% | 4e-32 | 77.46% | 71 | QGJ87678.1 |
| ☑ | hypothetical protein Gyarad_0053 [Mycobacterium phage Gyarad] | Mycobacterium phage Gyarad | 116 | 116 | 100% | 5e-32 | 77.46% | 74 | AGC34117.1 |
| ☑ | hypothetical protein M046_gp53 [Mycobacterium phage Newman] | Mycobacterium phage Newman | 116 | 116 | 100% | 5e-32 | 77.46% | 71 | YP_008052130.1 |
| ☑ | hypothetical protein SEA_DINGO_53 [Mycobacterium phage Dingo] | Mycobacterium phage Dingo | 116 | 116 | 100% | 6e-32 | 77.46% | 71 | ATN88764.1 |
| ☑ | hypothetical protein PBI_HARVEY_53 [Mycobacterium phage Harvey] | Mycobacterium phage Harvey | 115 | 115 | 100% | 9e-32 | 76.06% | 71 | AEK08805.1 |
| ☑ | hypothetical protein SEA_PHUNKY_53 [Mycobacterium phage Phunky] | Mycobacterium phage Phunky | 114 | 114 | 100% | 2e-31 | 76.06% | 71 | ATN91562.1 |
| ☑ | hypothetical protein VC60_gp68 [Mycobacterium phage Sbash] | Mycobacterium phage Sbash | 97.4 | 97.4 | 94% | 2e-24 | 70.15% | 70 | YP_009124722.1 |

## c. SIF: HHPred



HHPRED

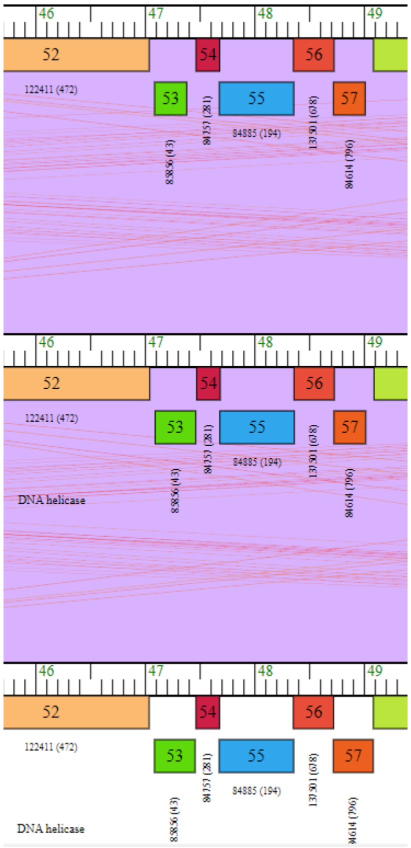
Last Job Status:
FINISHED at 1/19/2024, 9:25:50 AM
Last Updated:
1/19/2024, 9:25:50 AM

Show 10 entries Search:

| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☐ | PF19973.3 | TCAD7 ; Ternary complex associated domain 7 | 52.3 | 15.493 | 105 | 116 | 5 | 16 | 29 |
| ☐ | PF18929.4 | DUF5678 ; Family of unknown function (DUF5678) | 36.3 | 25.3521 | 10 | 28 | 5 | 23 | 71 |
| ☐ | 6A6F_B | Iron-sulfur cluster assembly scaffold protein NifU; SufE, Thermophile, Fe-S cluster scaffold protein, BIOSYNTHETIC PROTEIN; HET: PEG, GOL; 2.1A {Fervidobacterium islandicum} SCOP: d.224.1.0, l.1.1.1 | 33 | 36.6197 | 41 | 67 | 2 | 28 | 210 |

No evidence from HHPRED was used.

## d. SIF: Synteny-Phamerator (three genomes)



7. Any other important information.

No TmHmm

Synteny with DelRivs, Etaye, Fang, and more

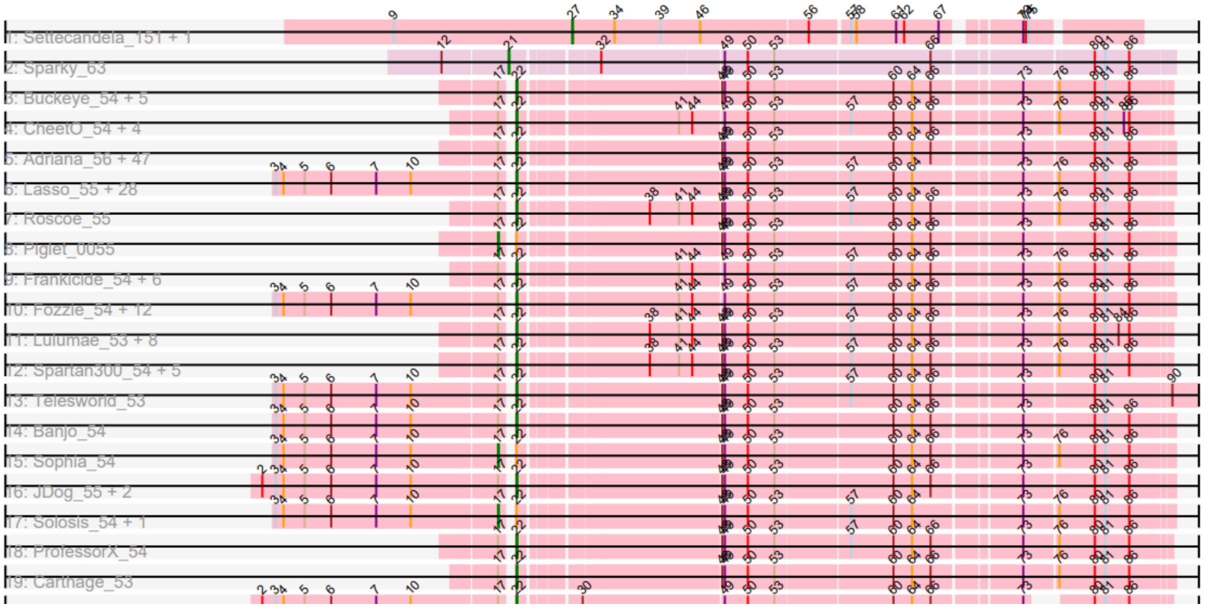
**CURATOR NAME: BELLA N**

**GENE NAME: EUGENIA-55**

**DNA MASTER NOTES: N/A**

**START POSITION EVALUATION (IN ORDER)**:

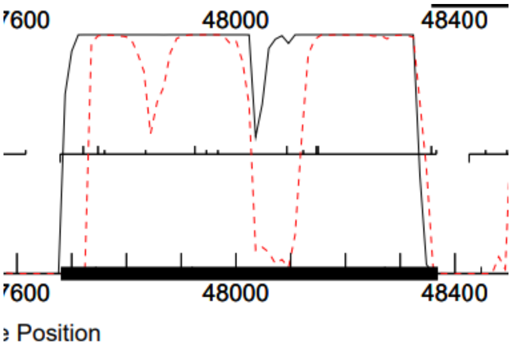
1. Starterator



The start number called the most often in the published annotations is 22, it was called in 132 of the 184 non-draft genes in the pham.
Start 22:

- Found in 142 of 194 ( 73.2% ) of genes in pham
- Manual Annotations of this start: 132 of 184
- Called 96.5% of time when present

2. GeneMark coding potential



The black bar at the bottom represents the whole ORF; the ORF is about 600 nucleotides long.

3. Glimmer and GeneMark agreement

There is not agreement between Glimmer and GeneMark

Glimmer Start: 48360

Glimmer Score: 13.35

GeneMark Start: 48369

4. Longest open reading frame (ORF) without excessive gap

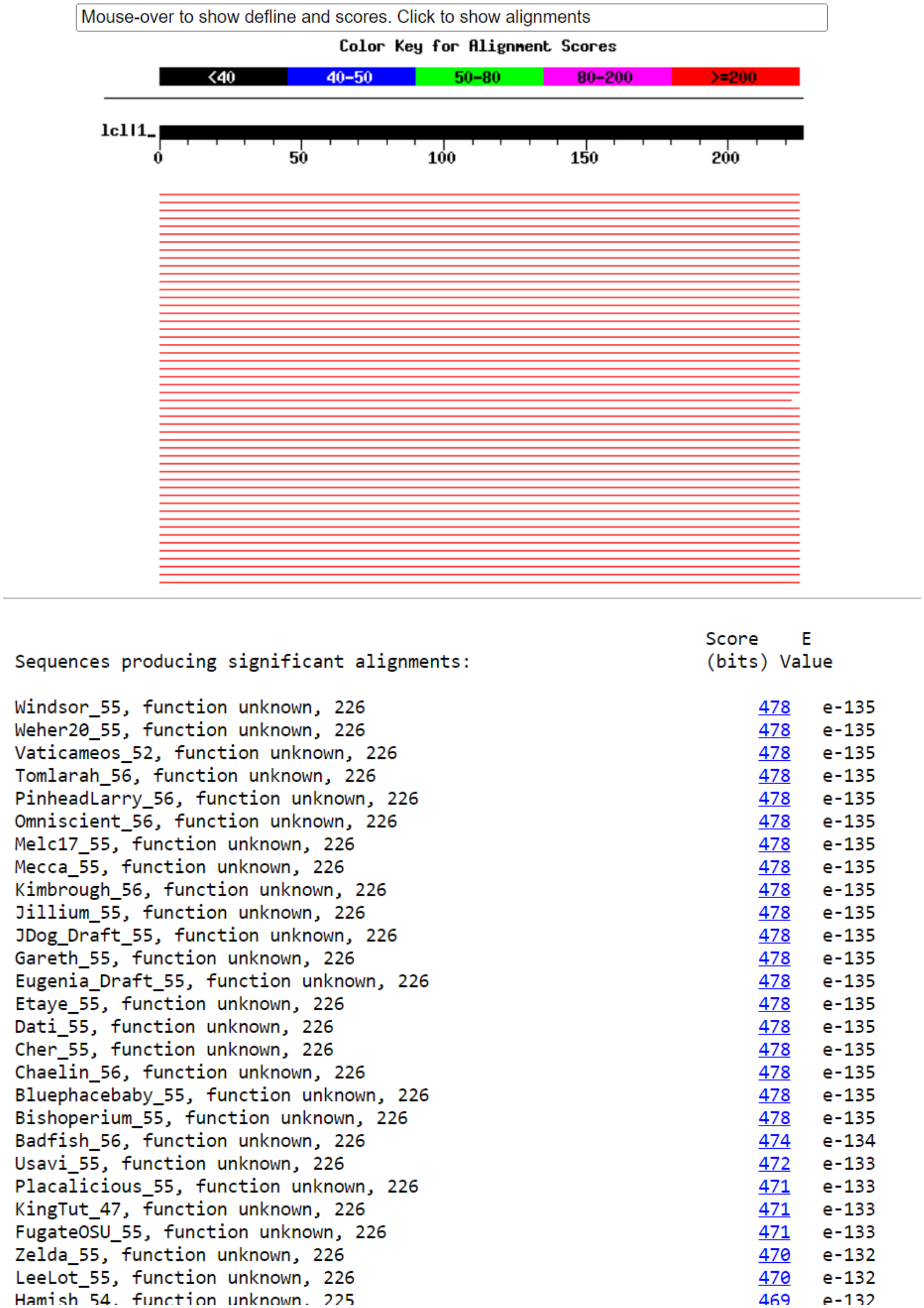
The selected gene is not the LORF with a length of 681 nucleotides. It has a Z-score of 2.081, a spacer of 12, and a final score of -4.760.

5. Function. If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

6. Supporting Information for Function (SIF)

## a. SIF: PhageDb BLAST

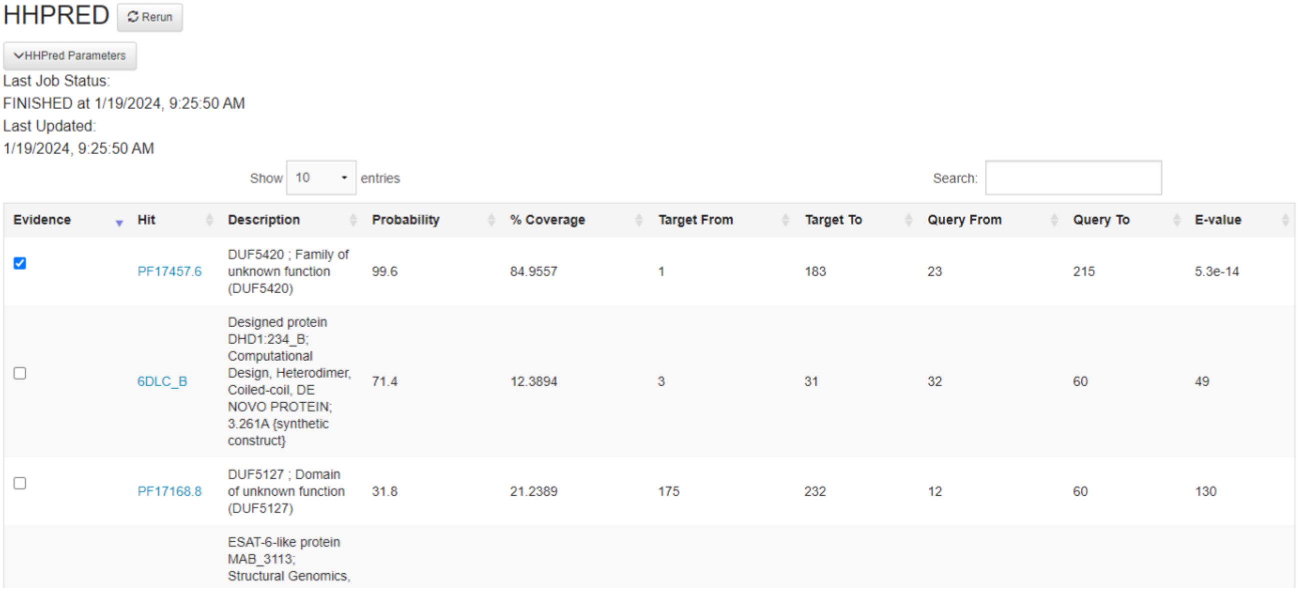


```
                                                                  Score     E
Sequences producing significant alignments:                       (bits)  Value

Windsor_55, function unknown, 226                                     478   e-135
Weher20_55, function unknown, 226                                     478   e-135
Vaticameos_52, function unknown, 226                                  478   e-135
Tomlarah_56, function unknown, 226                                    478   e-135
PinheadLarry_56, function unknown, 226                                478   e-135
Omniscient_56, function unknown, 226                                  478   e-135
Melc17_55, function unknown, 226                                      478   e-135
Mecca_55, function unknown, 226                                       478   e-135
Kimbrough_56, function unknown, 226                                   478   e-135
Jillium_55, function unknown, 226                                     478   e-135
JDog_Draft_55, function unknown, 226                                  478   e-135
Gareth_55, function unknown, 226                                      478   e-135
Eugenia_Draft_55, function unknown, 226                               478   e-135
Etaye_55, function unknown, 226                                       478   e-135
Dati_55, function unknown, 226                                        478   e-135
Cher_55, function unknown, 226                                        478   e-135
Chaelin_56, function unknown, 226                                     478   e-135
Bluephacebaby_55, function unknown, 226                               478   e-135
Bishoperium_55, function unknown, 226                                 478   e-135
Badfish_56, function unknown, 226                                     474   e-134
Usavi_55, function unknown, 226                                       472   e-133
Placalicious_55, function unknown, 226                                471   e-133
KingTut_47, function unknown, 226                                     471   e-133
FugateOSU_55, function unknown, 226                                   471   e-133
Zelda_55, function unknown, 226                                       470   e-132
LeeLot_55, function unknown, 226                                      470   e-132
Hamish_54, function unknown, 225                                      469   e-132
```

## b. SIF: NCBI BLAST

| | | | Score | Score | Cover | Value | Ident | Len | |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | hypothetical protein SEA_GARETH_55 [Mycobacterium phage Gareth] | Mycobacterium phage Gareth | 463 | 463 | 100% | 2e-164 | 100.00% | 226 | AXC37577.1 |
| ☑ | hypothetical protein AU110_gp056 [Mycobacterium phage Badfish] | Mycobacterium phage Badfish | 459 | 459 | 100% | 9e-163 | 99.12% | 226 | YP_009190112.1 |
| ☑ | hypothetical protein SEA_USAVI_55 [Mycobacterium phage Usavi] | Mycobacterium phage Usavi | 457 | 457 | 100% | 5e-162 | 98.67% | 226 | WNM64440.1 |
| ☑ | hypothetical protein KNT94_gp47 [Mycobacterium phage KingTut] | Mycobacterium phage KingTut | 457 | 457 | 100% | 6e-162 | 98.67% | 226 | YP_010096560.1 |
| ☑ | hypothetical protein SEA_LEELOT_55 [Mycobacterium phage LeeLot] | Mycobacterium phage LeeLot | 456 | 456 | 100% | 2e-161 | 98.23% | 226 | AVJ50011.1 |
| ☑ | hypothetical protein SEA_HAMISH_54 [Mycobacterium phage Hamish] | Mycobacterium phage Hamish | 455 | 455 | 98% | 3e-161 | 99.10% | 225 | AYD86425.1 |
| ☑ | hypothetical protein PIGLET_0055 [Mycobacterium phage Piglet] | Mycobacterium phage Piglet | 454 | 454 | 100% | 1e-160 | 97.35% | 229 | AGC33915.1 |
| ☑ | hypothetical protein PBI_PG1_56 [Mycobacterium phage PG1] | Mycobacterium phage PG1 | 453 | 453 | 100% | 2e-160 | 97.35% | 226 | NP_943834.1 |
| ☑ | hypothetical protein SEA_SOPHIA_54 [Mycobacterium phage Sophia] | Mycobacterium phage Sophia | 452 | 452 | 100% | 3e-160 | 97.79% | 229 | AZS10230.1 |
| ☑ | hypothetical protein SEA_DONSANCHON_53 [Mycobacterium phage DonSanc… | Mycobacterium phage DonSanchon | 451 | 451 | 98% | 8e-160 | 98.21% | 228 | AXH67571.1 |
| ☑ | hypothetical protein SEA_ADRIANA_56 [Mycobacterium phage Adriana] | Mycobacterium phage Adriana | 451 | 451 | 100% | 8e-160 | 97.79% | 226 | QNO12067.1 |
| ☑ | hypothetical protein SEA_MORTY_56 [Mycobacterium phage Morty] | Mycobacterium phage Morty | 451 | 451 | 100% | 9e-160 | 96.90% | 226 | AXC34836.1 |
| ☑ | hypothetical protein VISTA_54 [Mycobacterium phage Vista] | Mycobacterium phage Vista | 451 | 451 | 98% | 1e-159 | 98.21% | 225 | YP_009016843.1 |
| ☑ | hypothetical protein SEA_MULAN_55 [Mycobacterium phage Mulan] | Mycobacterium phage Mulan | 451 | 451 | 100% | 1e-159 | 96.90% | 226 | AYB70064.1 |
| ☑ | hypothetical protein SEA_HAIMAS_54 [Mycobacterium phage Haimas] | Mycobacterium phage Haimas | 449 | 449 | 98% | 8e-159 | 97.31% | 225 | AUV60457.1 |
| ☑ | hypothetical protein SEA_BUCKEYE_54 [Mycobacterium phage Buckeye] | Mycobacterium phage Buckeye | 445 | 445 | 98% | 3e-157 | 96.41% | 225 | AXH43882.1 |
| | hypothetical protein VORTEX_55 [Mycobacterium phage Vortex] | Mycobacterium phage Vortex | 444 | 444 | 100% | 6e-157 | 96.02% | 226 | YP_009198729.1 |

## c. SIF: HHPred

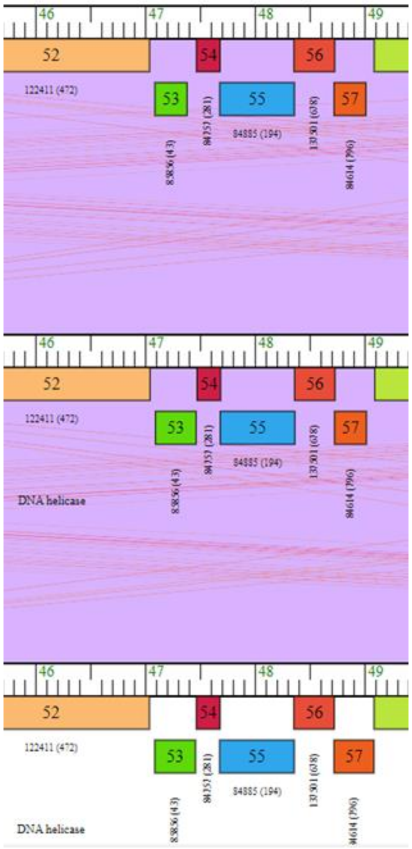


Last Job Status:
FINISHED at 1/19/2024, 9:25:50 AM
Last Updated:
1/19/2024, 9:25:50 AM

Show 10 entries Search:

| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | PF17457.6 | DUF5420 ; Family of unknown function (DUF5420) | 99.6 | 84.9557 | 1 | 183 | 23 | 215 | 5.3e-14 |
| ☐ | 6DLC_B | Designed protein DHD1:234_B; Computational Design, Heterodimer, Coiled-coil, DE NOVO PROTEIN; 3.261A {synthetic construct} | 71.4 | 12.3894 | 3 | 31 | 32 | 60 | 49 |
| ☐ | PF17168.8 | DUF5127 ; Domain of unknown function (DUF5127) | 31.8 | 21.2389 | 175 | 232 | 12 | 60 | 130 |
| | | ESAT-6-like protein MAB_3113; Structural Genomics, | | | | | | | |

## d. SIF: Synteny-Phamerator (three genomes)



7. Any other important information.

Synteny with Bishoperium, Bluephacebaby, Chaelin.

No TmHmms

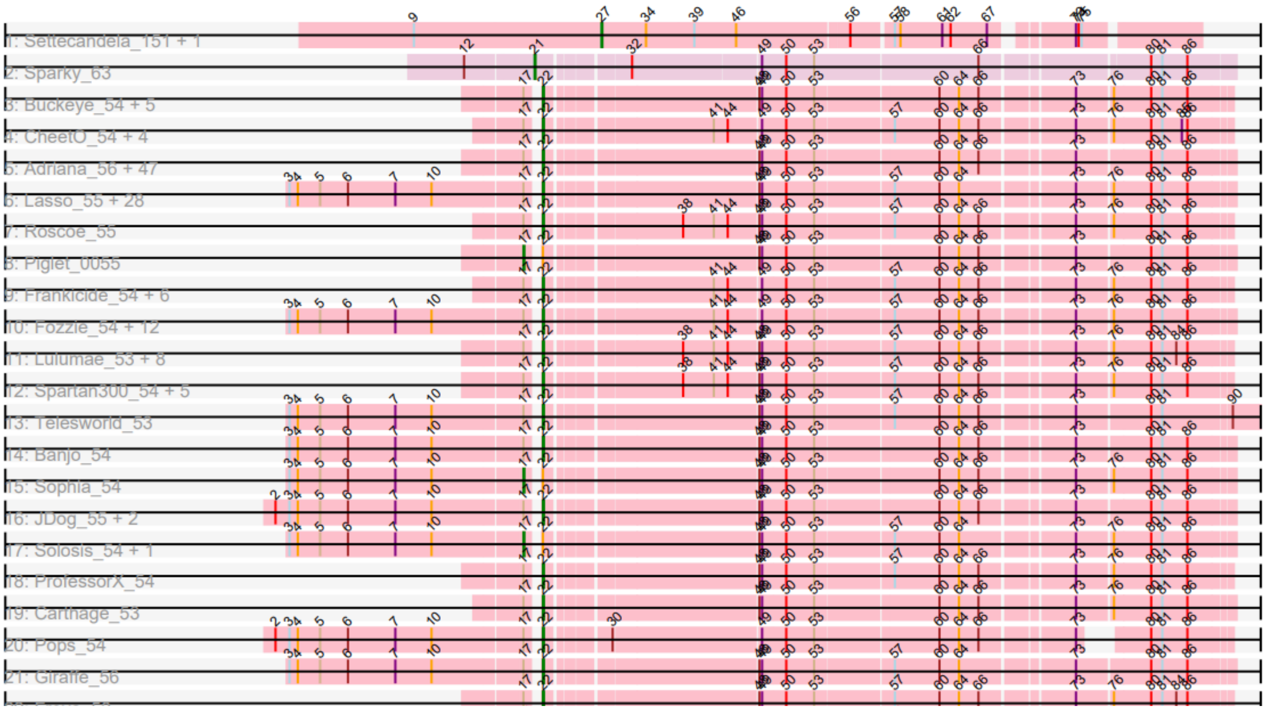
**CURATOR NAME: BELLA N**

**GENE NAME: EUGENIA-56**

**DNA MASTER NOTES: N/A**

**START POSITION EVALUATION (IN ORDER)**:

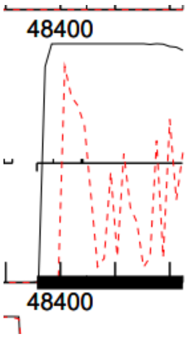
1. Starterator

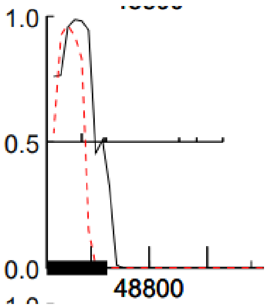


The start number called the most often in the published annotations is 22, it was called in 132 of the 184 non-draft genes in the pham.
Start 22:

- Found in 142 of 194 ( 73.2% ) of genes in pham
- Manual Annotations of this start: 132 of 184
- Called 96.5% of time when present

### 2. GeneMark coding potential





The black bar at the bottom represents the whole ORF it is about 700 nucleotides and this is a reverse gene.

### 3. Glimmer and GeneMark agreement

There is not agreement between Glimmer and GeneMark

Glimmer Start: 48360

Glimmer Score: 13.35

GeneMark Start: 48369

### 4. Longest open reading frame (ORF) without excessive gap

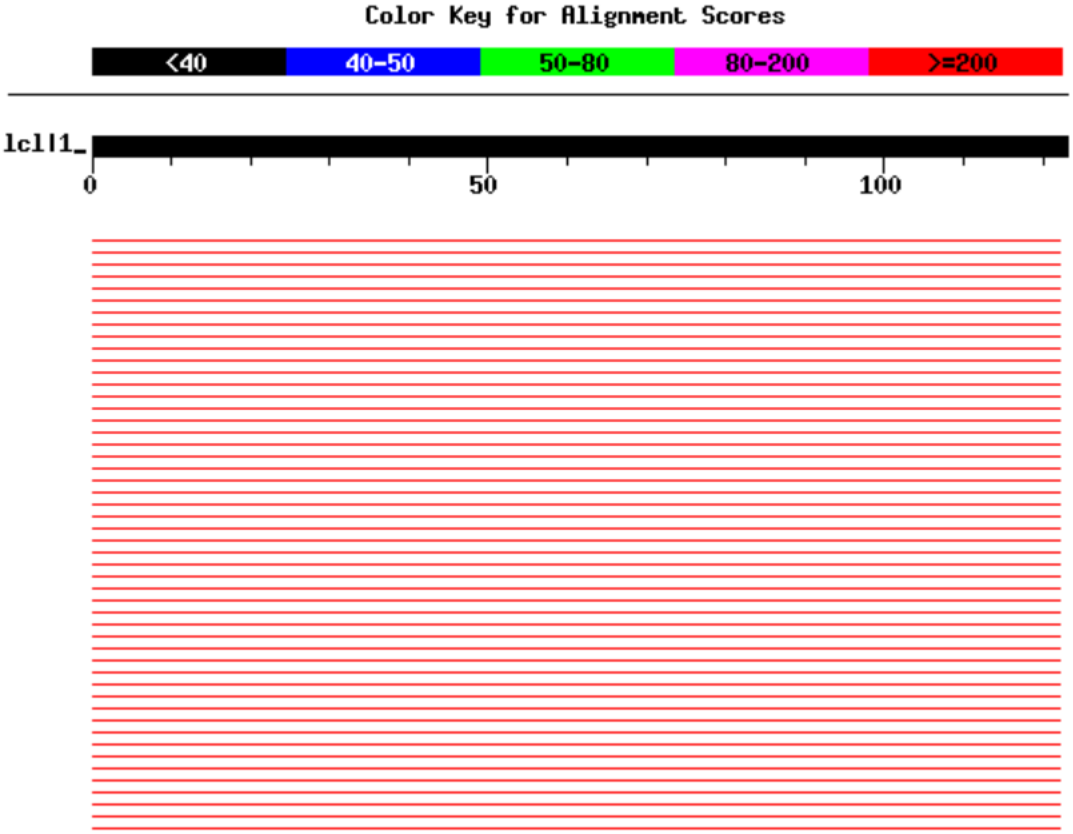
The selected gene is not  LORF with a length of 681 nucleotides. It has a Z-score of 2.081, a spacer of 12, and a final score of  -4.760.

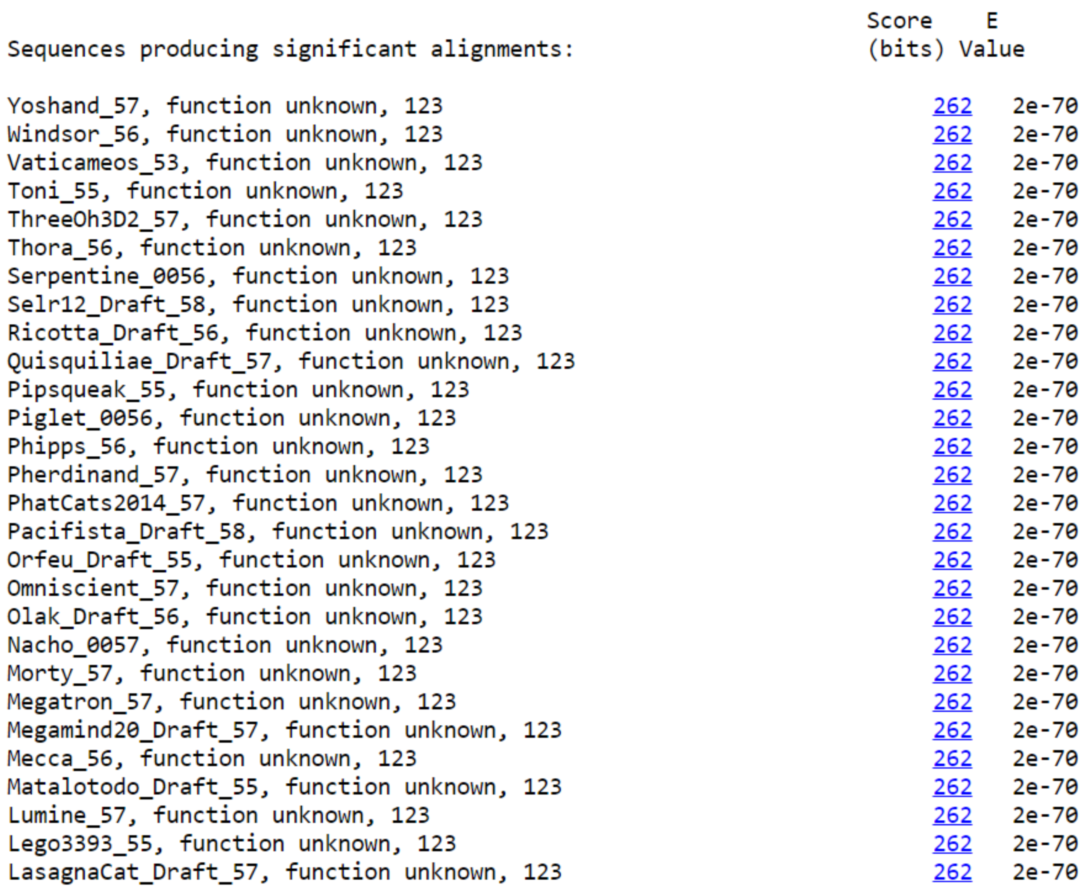
### 5. Function. If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

### 6. Supporting Information for Function (SIF)

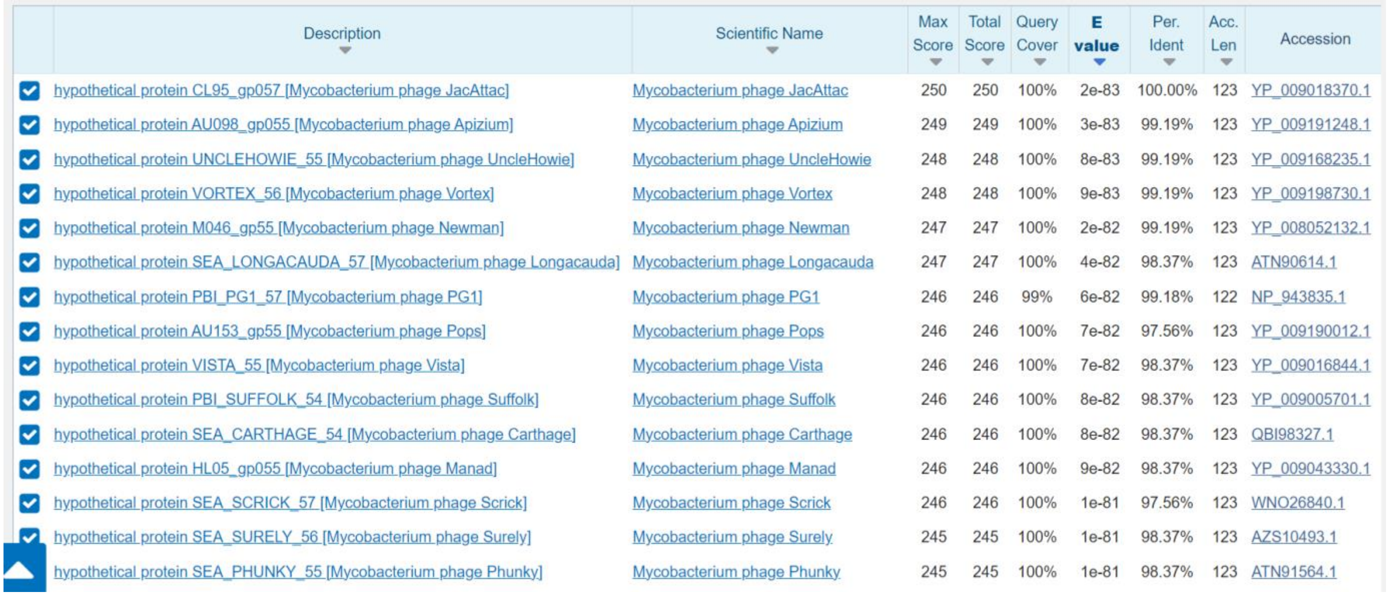
a. SIF: PhageDb BLAST


Color Key for Alignment Scores
<40
40-50
50-80
80-200
>=200
lcl|1_
0
50
100



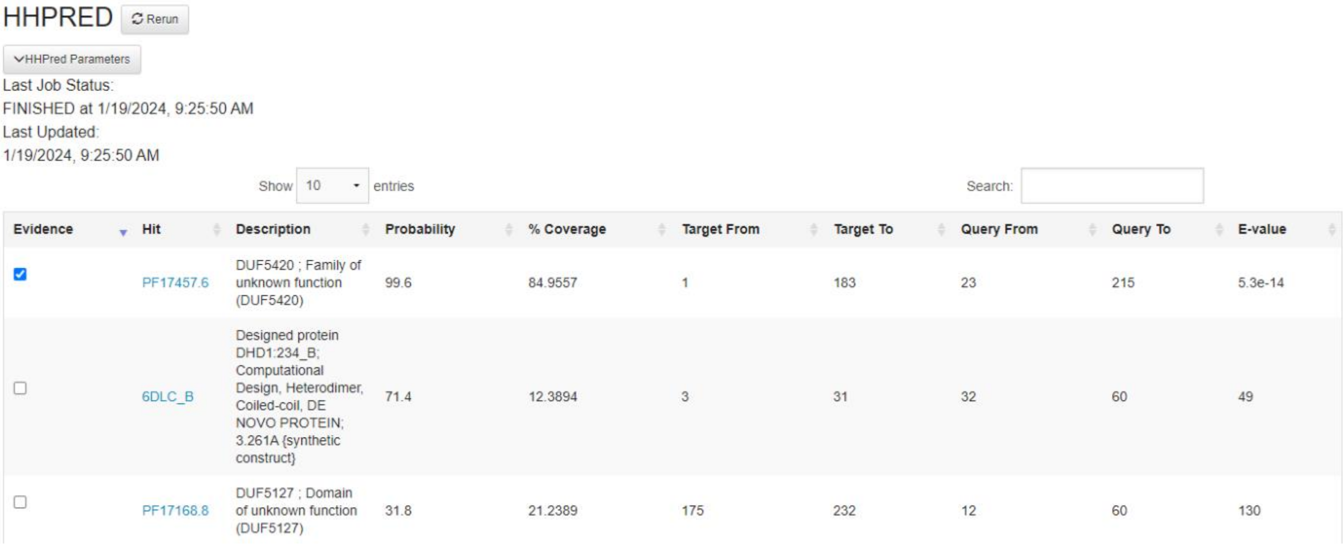
```
                                                                Score     E
Sequences producing significant alignments:                    (bits)  Value

Yoshand_57, function unknown, 123                                  262    2e-70
Windsor_56, function unknown, 123                                  262    2e-70
Vaticameos_53, function unknown, 123                               262    2e-70
Toni_55, function unknown, 123                                     262    2e-70
ThreeOh3D2_57, function unknown, 123                               262    2e-70
Thora_56, function unknown, 123                                    262    2e-70
Serpentine_0056, function unknown, 123                             262    2e-70
Selr12_Draft_58, function unknown, 123                             262    2e-70
Ricotta_Draft_56, function unknown, 123                            262    2e-70
Quisquiliae_Draft_57, function unknown, 123                        262    2e-70
Pipsqueak_55, function unknown, 123                                262    2e-70
Piglet_0056, function unknown, 123                                 262    2e-70
Phipps_56, function unknown, 123                                   262    2e-70
Pherdinand_57, function unknown, 123                               262    2e-70
PhatCats2014_57, function unknown, 123                             262    2e-70
Pacifista_Draft_58, function unknown, 123                          262    2e-70
Orfeu_Draft_55, function unknown, 123                              262    2e-70
Omniscient_57, function unknown, 123                               262    2e-70
Olak_Draft_56, function unknown, 123                               262    2e-70
Nacho_0057, function unknown, 123                                  262    2e-70
Morty_57, function unknown, 123                                    262    2e-70
Megatron_57, function unknown, 123                                 262    2e-70
Megamind20_Draft_57, function unknown, 123                         262    2e-70
Mecca_56, function unknown, 123                                    262    2e-70
Matalotodo_Draft_55, function unknown, 123                         262    2e-70
Lumine_57, function unknown, 123                                   262    2e-70
Lego3393_55, function unknown, 123                                 262    2e-70
LasagnaCat_Draft_57, function unknown, 123                         262    2e-70
```

b. SIF: NCBI BLAST



| | Description | Scientific Name | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | hypothetical protein CL95_gp057 [Mycobacterium phage JacAttac] | Mycobacterium phage JacAttac | 250 | 250 | 100% | 2e-83 | 100.00% | 123 | YP_009018370.1 |
| ☑ | hypothetical protein AU098_gp055 [Mycobacterium phage Apizium] | Mycobacterium phage Apizium | 249 | 249 | 100% | 3e-83 | 99.19% | 123 | YP_009191248.1 |
| ☑ | hypothetical protein UNCLEHOWIE_55 [Mycobacterium phage UncleHowie] | Mycobacterium phage UncleHowie | 248 | 248 | 100% | 8e-83 | 99.19% | 123 | YP_009168235.1 |
| ☑ | hypothetical protein VORTEX_56 [Mycobacterium phage Vortex] | Mycobacterium phage Vortex | 248 | 248 | 100% | 9e-83 | 99.19% | 123 | YP_009198730.1 |
| ☑ | hypothetical protein M046_gp55 [Mycobacterium phage Newman] | Mycobacterium phage Newman | 247 | 247 | 100% | 2e-82 | 99.19% | 123 | YP_008052132.1 |
| ☑ | hypothetical protein SEA_LONGACAUDA_57 [Mycobacterium phage Longacauda] | Mycobacterium phage Longacauda | 247 | 247 | 100% | 4e-82 | 98.37% | 123 | ATN90614.1 |
| ☑ | hypothetical protein PBI_PG1_57 [Mycobacterium phage PG1] | Mycobacterium phage PG1 | 246 | 246 | 99% | 6e-82 | 99.18% | 122 | NP_943835.1 |
| ☑ | hypothetical protein AU153_gp55 [Mycobacterium phage Pops] | Mycobacterium phage Pops | 246 | 246 | 100% | 7e-82 | 97.56% | 123 | YP_009190012.1 |
| ☑ | hypothetical protein VISTA_55 [Mycobacterium phage Vista] | Mycobacterium phage Vista | 246 | 246 | 100% | 7e-82 | 98.37% | 123 | YP_009016844.1 |
| ☑ | hypothetical protein PBI_SUFFOLK_54 [Mycobacterium phage Suffolk] | Mycobacterium phage Suffolk | 246 | 246 | 100% | 8e-82 | 98.37% | 123 | YP_009005701.1 |
| ☑ | hypothetical protein SEA_CARTHAGE_54 [Mycobacterium phage Carthage] | Mycobacterium phage Carthage | 246 | 246 | 100% | 8e-82 | 98.37% | 123 | QBI98327.1 |
| ☑ | hypothetical protein HL05_gp055 [Mycobacterium phage Manad] | Mycobacterium phage Manad | 246 | 246 | 100% | 9e-82 | 98.37% | 123 | YP_009043330.1 |
| ☑ | hypothetical protein SEA_SCRICK_57 [Mycobacterium phage Scrick] | Mycobacterium phage Scrick | 246 | 246 | 100% | 1e-81 | 97.56% | 123 | WNO26840.1 |
| ☑ | hypothetical protein SEA_SURELY_56 [Mycobacterium phage Surely] | Mycobacterium phage Surely | 245 | 245 | 100% | 1e-81 | 98.37% | 123 | AZS10493.1 |
| | hypothetical protein SEA_PHUNKY_55 [Mycobacterium phage Phunky] | Mycobacterium phage Phunky | 245 | 245 | 100% | 1e-81 | 98.37% | 123 | ATN91564.1 |

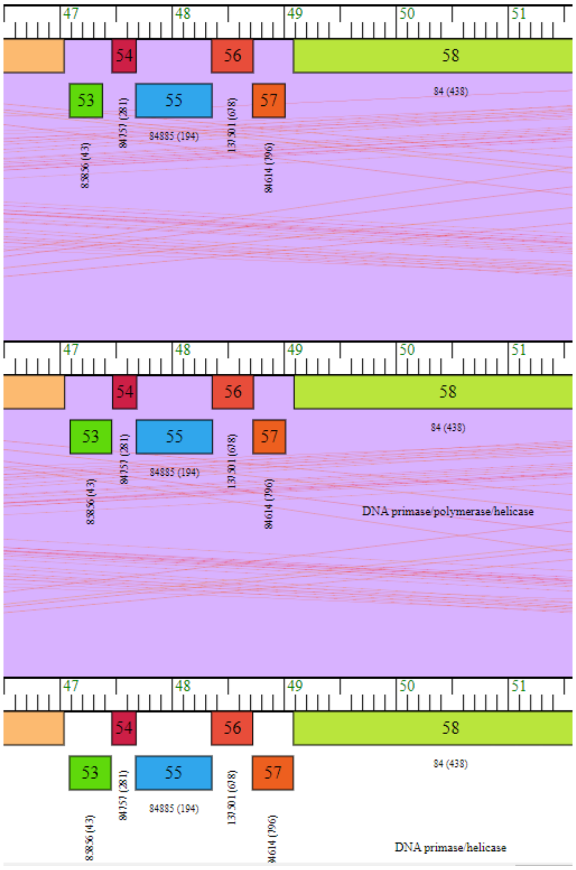
## c. SIF: HHPred



Last Job Status:
FINISHED at 1/19/2024, 9:25:50 AM
Last Updated:
1/19/2024, 9:25:50 AM

| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | PF17457.6 | DUF5420 ; Family of unknown function (DUF5420) | 99.6 | 84.9557 | 1 | 183 | 23 | 215 | 5.3e-14 |
| ☐ | 6DLC_B | Designed protein DHD1:234_B; Computational Design, Heterodimer, Coiled-coil, DE NOVO PROTEIN; 3.261A {synthetic construct} | 71.4 | 12.3894 | 3 | 31 | 32 | 60 | 49 |
| ☐ | PF17168.8 | DUF5127 ; Domain of unknown function (DUF5127) | 31.8 | 21.2389 | 175 | 232 | 12 | 60 | 130 |

## d. SIF: Synteny-Phamerator (three genomes)



7. Any other important information.

Synteny with Bishoperium, Bluephacebaby, Chaelin.

No TmHmms

### CURATOR NAME: BELLA N

### GENE NAME: EUGENIA-57

### DNA MASTER NOTES: N/A

### START POSITION EVALUATION (IN ORDER):

1. Starterator

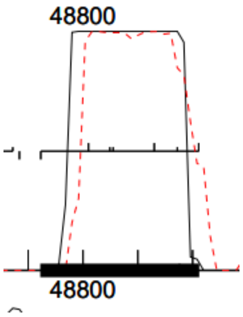
The start number called the most often in the published annotations is 44, it was called in 514 of the 741 non-draft genes in the pham.

Start 44:

- Found in 560 of 807 ( 69.4% ) of genes in pham
- Manual Annotations of this start: 514 of 741

- Called 99.6% of time when present

2. GeneMark coding potential



The black bar at the bottom represents the whole ORF it is about 300 nucleotides and this is a reverse gene.

3. Glimmer and GeneMark agreement

There is agreement between Glimmer and GeneMark

Glimmer Start: 49012

Glimmer Score: 14.89

GeneMark Start: 49012

4. Longest open reading frame (ORF) without excessive gap

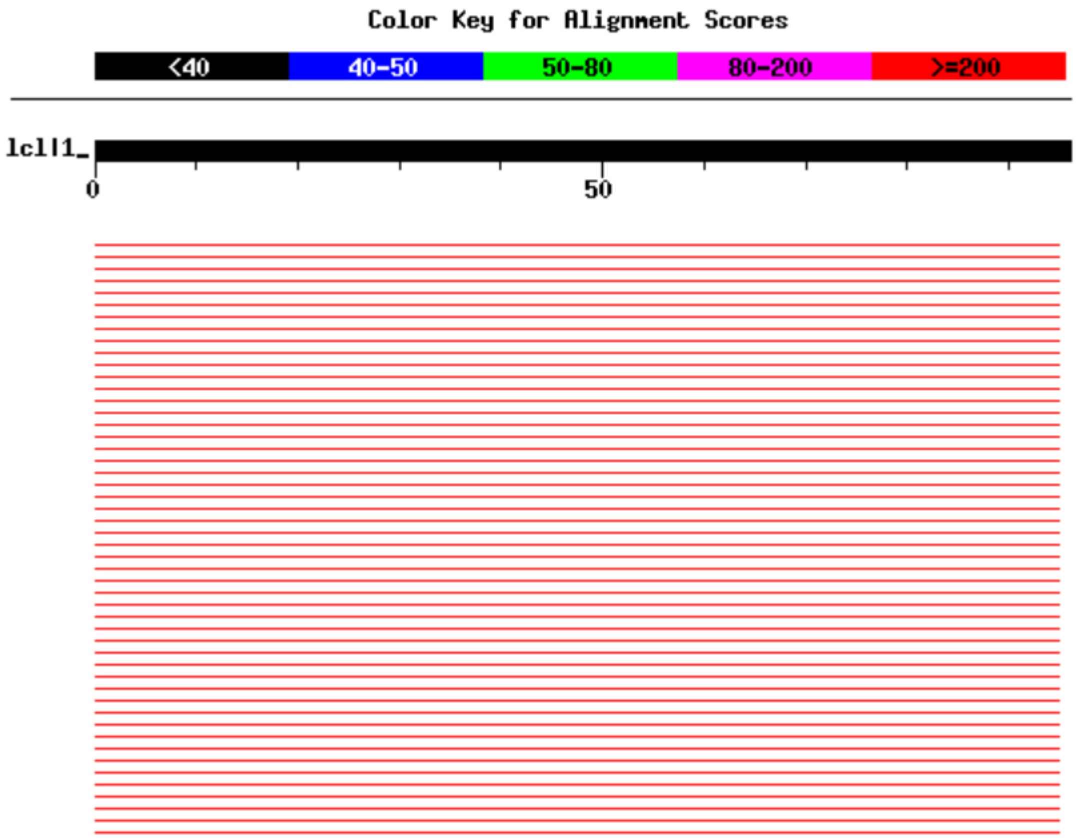
The selected gene is not the LORF with a length of 291 nucleotides. It has a Z-score of 3.408, a spacer of 11, and a final score of -2.016. This is a reverse gene.

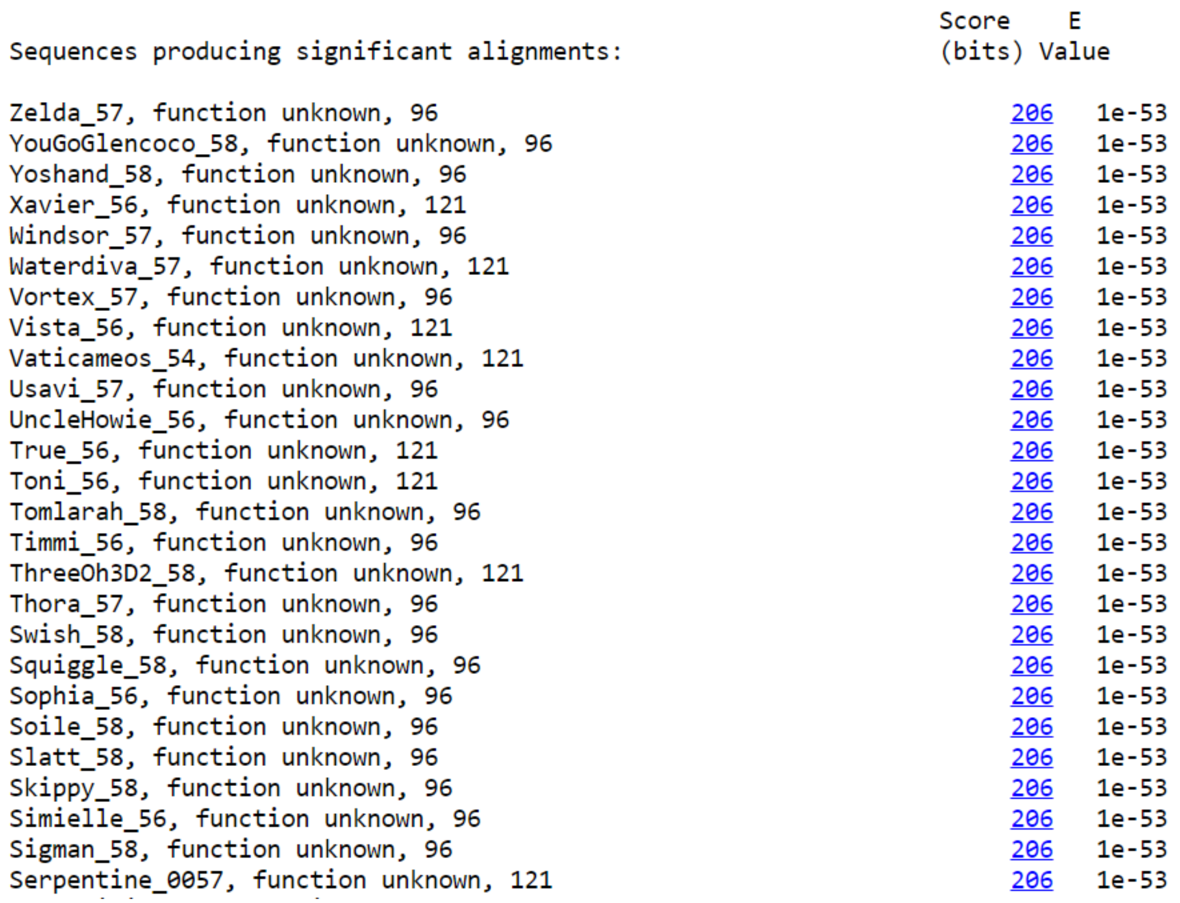
5. Function. If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST





```
                                                                  Score     E
Sequences producing significant alignments:                       (bits)  Value

Zelda_57, function unknown, 96                                      206    1e-53
YouGoGlencoco_58, function unknown, 96                              206    1e-53
Yoshand_58, function unknown, 96                                    206    1e-53
Xavier_56, function unknown, 121                                    206    1e-53
Windsor_57, function unknown, 96                                    206    1e-53
Waterdiva_57, function unknown, 121                                 206    1e-53
Vortex_57, function unknown, 96                                     206    1e-53
Vista_56, function unknown, 121                                     206    1e-53
Vaticameos_54, function unknown, 121                                206    1e-53
Usavi_57, function unknown, 96                                      206    1e-53
UncleHowie_56, function unknown, 96                                 206    1e-53
True_56, function unknown, 121                                      206    1e-53
Toni_56, function unknown, 121                                      206    1e-53
Tomlarah_58, function unknown, 96                                   206    1e-53
Timmi_56, function unknown, 96                                      206    1e-53
ThreeOh3D2_58, function unknown, 121                                206    1e-53
Thora_57, function unknown, 96                                      206    1e-53
Swish_58, function unknown, 96                                      206    1e-53
Squiggle_58, function unknown, 96                                   206    1e-53
Sophia_56, function unknown, 96                                     206    1e-53
Soile_58, function unknown, 96                                      206    1e-53
Slatt_58, function unknown, 96                                      206    1e-53
Skippy_58, function unknown, 96                                     206    1e-53
Simielle_56, function unknown, 96                                   206    1e-53
Sigman_58, function unknown, 96                                     206    1e-53
Serpentine_0057, function unknown, 121                              206    1e-53
```

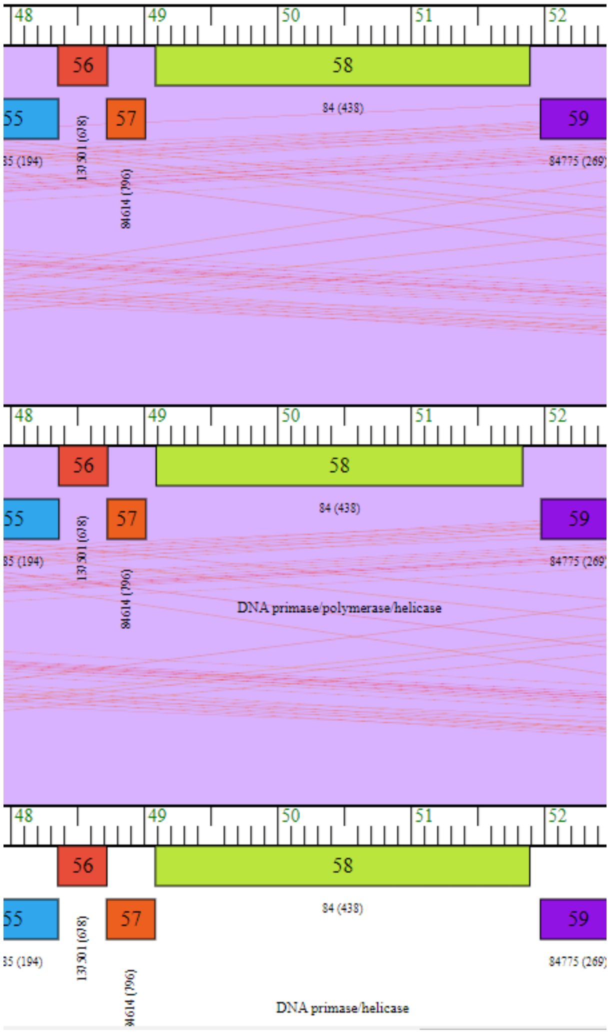
## b. SIF: NCBI BLAST

| | Description | Scientific Name | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | hypothetical protein VISTA_56 [Mycobacterium phage Vista] | Mycobacterium phage Vista | 200 | 200 | 100% | 3e-64 | 100.00% | 121 | YP_009016845.1 |
| ☑ | hypothetical protein CL95_gp058 [Mycobacterium phage JacAttac] | Mycobacterium phage JacAttac | 199 | 199 | 100% | 6e-64 | 100.00% | 121 | YP_009018371.1 |
| ☑ | hypothetical protein SEA_PODRICK_55 [Mycobacterium phage Podrick] | Mycobacterium phage Podrick | 199 | 199 | 100% | 6e-64 | 100.00% | 121 | AXQ64859.1 |
| ☑ | hypothetical protein TALLGRASSMM_57 [Mycobacterium phage TallGrassMM] | Mycobacterium phage TallGrassMM | 199 | 199 | 100% | 8e-64 | 98.96% | 121 | AER49268.1 |
| ☑ | hypothetical protein PBI_PG1_58 [Mycobacterium phage PG1] | Mycobacterium phage PG1 | 198 | 198 | 100% | 1e-63 | 100.00% | 96 | NP_943836.1 |
| ☑ | hypothetical protein PBI_PHAMISHED_58 [Mycobacterium phage Phamished] | Mycobacterium phage Phamished | 196 | 196 | 100% | 1e-62 | 97.92% | 121 | AKO62334.1 |
| ☑ | hypothetical protein SEA_IRIDOCLYSIS_58 [Mycobacterium phage Iridoclysis] | Mycobacterium phage Iridoclysis | 195 | 195 | 100% | 1e-62 | 98.96% | 96 | AOT23530.1 |
| ☑ | hypothetical protein CL79_gp057 [Mycobacterium phage Oline] | Mycobacterium phage Oline | 193 | 193 | 100% | 7e-62 | 96.88% | 96 | YP_009014319.1 |
| ☑ | hypothetical protein M046_gp56 [Mycobacterium phage Newman] | Mycobacterium phage Newman | 194 | 194 | 100% | 1e-61 | 96.88% | 121 | YP_008052133.1 |
| ☑ | hypothetical protein SEA_LULUMAE_55 [Mycobacterium phage Lulumae] | Mycobacterium phage Lulumae | 193 | 193 | 100% | 2e-61 | 96.88% | 121 | ASZ73484.1 |
| ☑ | hypothetical protein PBI_VIVALDI_58 [Mycobacterium phage Vivaldi] | Mycobacterium phage Vivaldi | 193 | 193 | 100% | 3e-61 | 95.83% | 121 | AIM50290.1 |
| ☑ | hypothetical protein PBI_SOTO_55 [Mycobacterium phage Soto] | Mycobacterium phage Soto | 192 | 192 | 100% | 3e-61 | 96.88% | 96 | YP_009100864.1 |
| ☑ | hypothetical protein HETAERIA_56 [Mycobacterium phage Hetaeria] | Mycobacterium phage Hetaeria | 192 | 192 | 100% | 6e-61 | 95.83% | 121 | ALA45667.1 |
| ☑ | hypothetical protein KLUCKY39_55 [Mycobacterium phage KLucky39] | Mycobacterium phage KLucky39 | 190 | 190 | 100% | 3e-60 | 95.83% | 121 | AEJ95241.1 |
| ☑ | hypothetical protein N850_gp031 [Mycobacterium phage Jabbawokkie] | Mycobacterium phage Jabbawokkie | 128 | 128 | 100% | 5e-36 | 64.95% | 98 | YP_008410703.1 |
| ☑ | hypothetical protein SEA_NIMBO_26 [Mycobacterium phage Nimbo] | Mycobacterium phage Nimbo | 125 | 125 | 97% | 5e-35 | 70.53% | 98 | AXQ61667.1 |
| | hypothetical protein CL78_gp029 [Mycobacterium phage Avani] | Mycobacterium phage Avani | 124 | 124 | 97% | 2e-34 | 68.42% | 98 | YP_009013124.1 |

c. SIF: HHPred

No HHPRED evidence was selected.

d. SIF: Synteny-Phamerator (three genomes)



7. Any other important information.

Synteny with Haleema, Harvey, Held and others.

No TmHmms

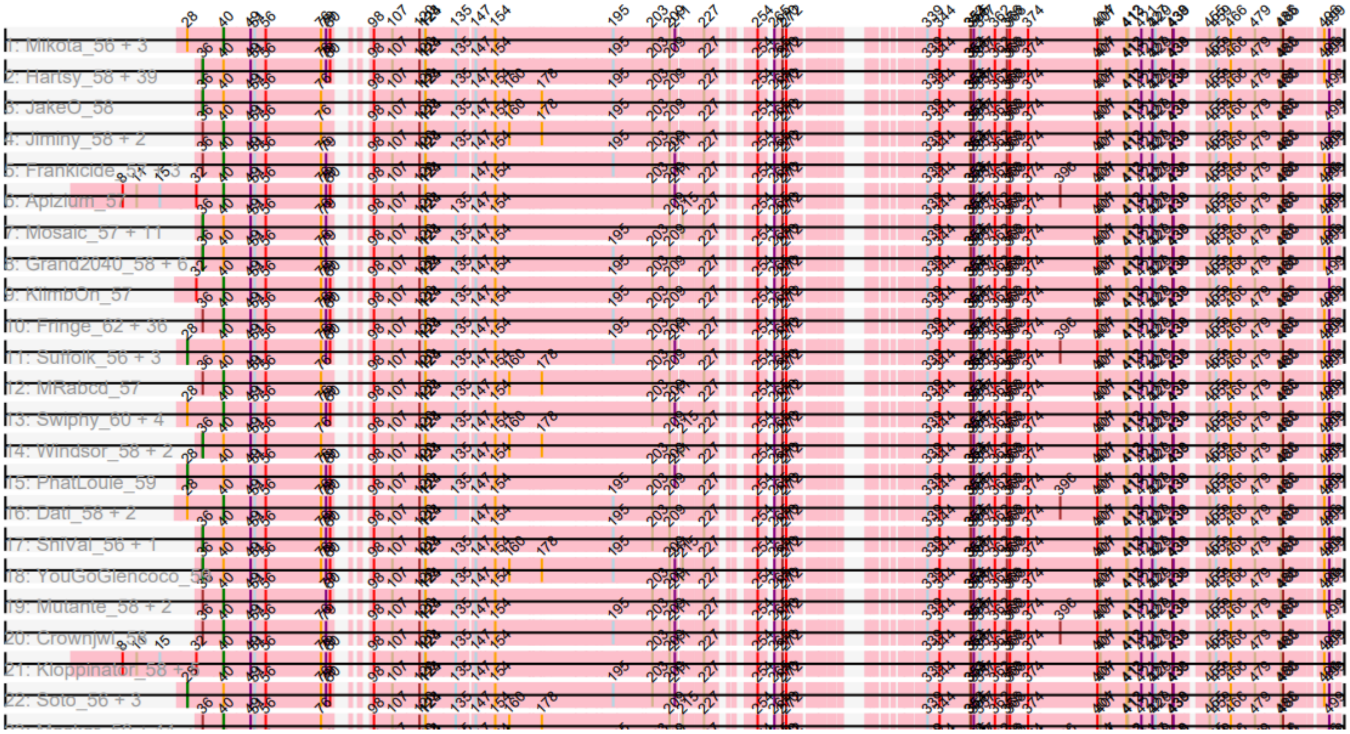
## CURATOR NAME: BELLA N

## GENE NAME: EUGENIA-58

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

1. Starterator

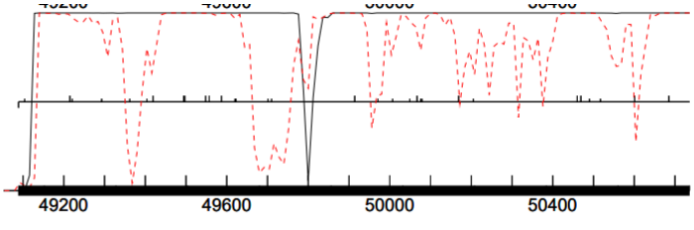


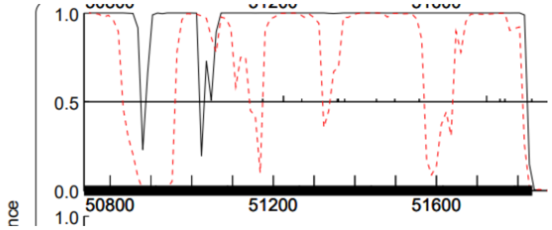
The start number called the most often in the published annotations is 40, it was called in 279 of the 390 non-draft genes in the pham.
Start 40:

- Found in 433 of 444 ( 97.5% ) of genes in pham
- Manual Annotations of this start: 279 of 390

- Called 73.0% of time when present

2. GeneMark coding potential





The black bar at the bottom represents the whole ORF it is very long at around 3000 nucleotides and this is a reverse gene.

3. Glimmer and GeneMark agreement

There is not agreement between Glimmer and GeneMark

Glimmer Start: 51895

Glimmer Score: 15.8

GeneMark Start: 51835

4. Longest open reading frame (ORF) without excessive gap

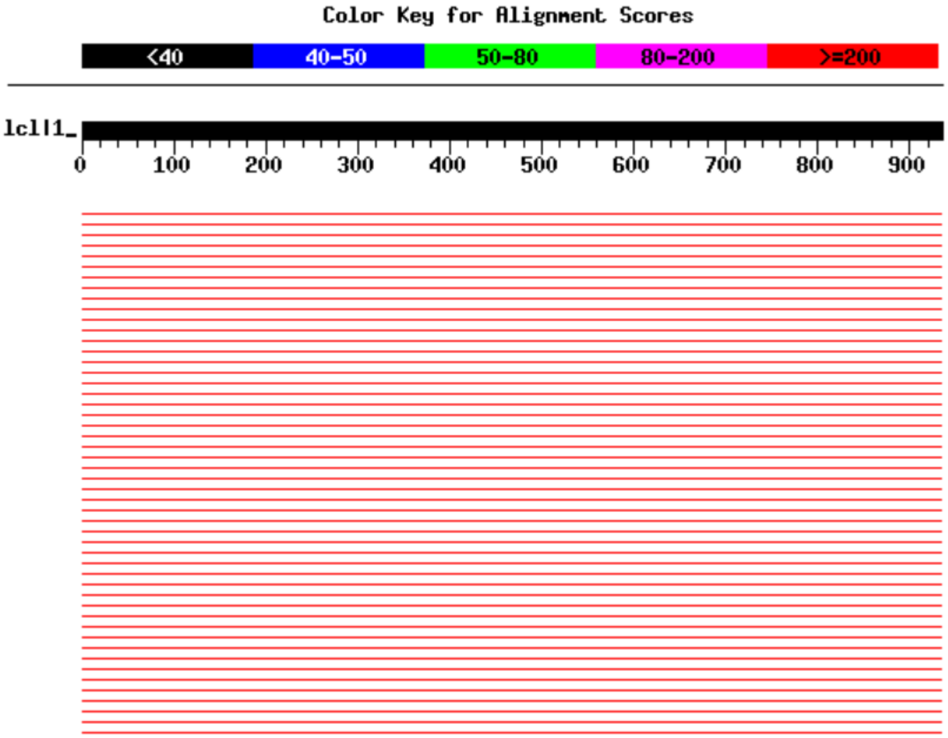
The selected gene is the LORF with a length of 2808 nucleotides. It has a Z-score of 1.658, a spacer of 10, and a final score of -2.315. This is a reverse gene.

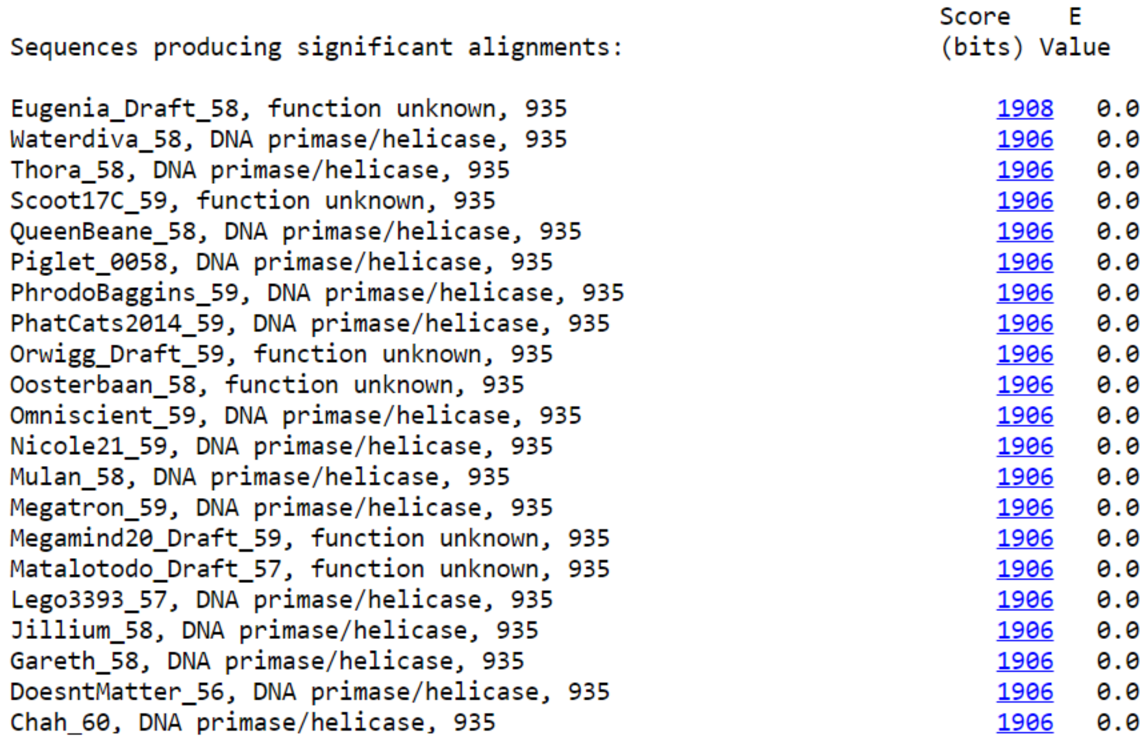
5. Function. If no functional prediction is present, write “Hypothetical protein”.

DNA primase/helicase

## 6. Supporting Information for Function (SIF)

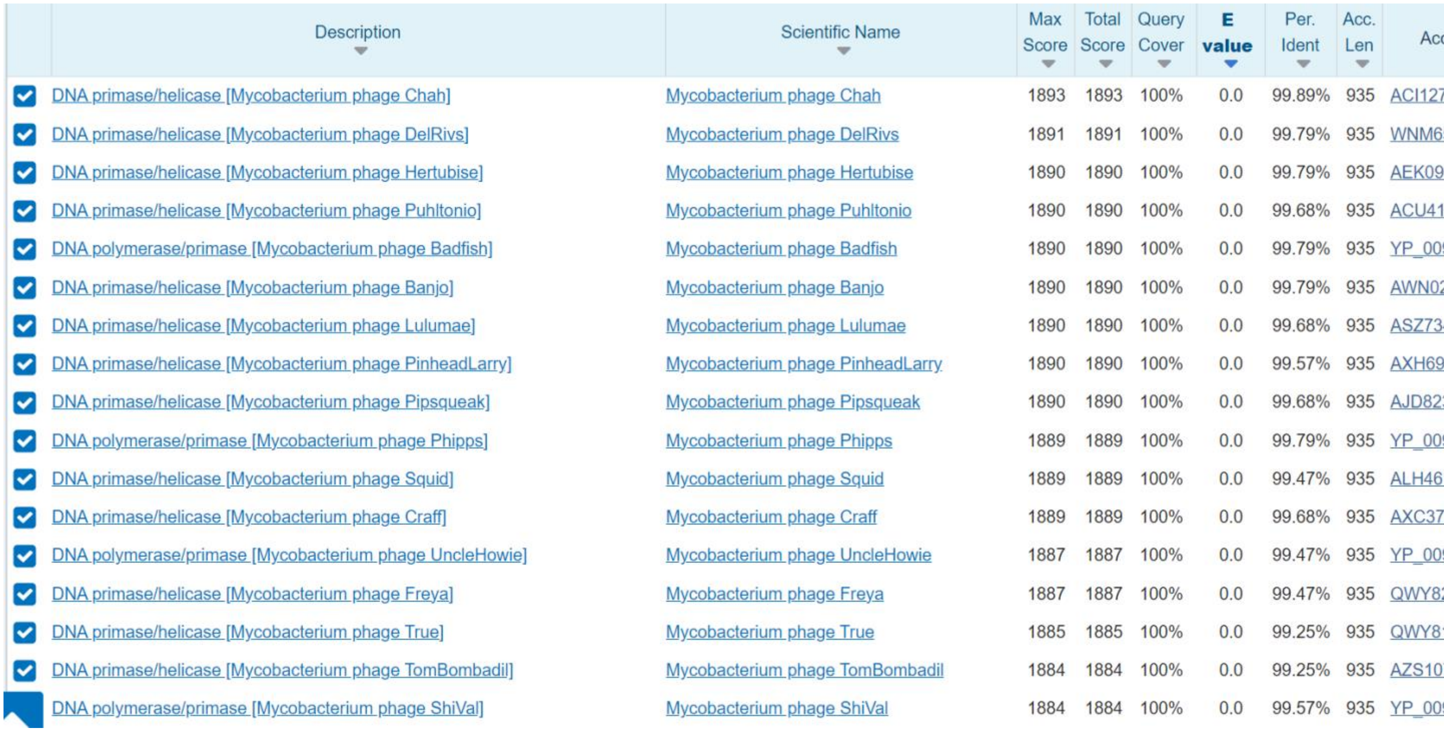
### a. SIF: PhageDb BLAST





```
                                                             Score    E
Sequences producing significant alignments:                  (bits) Value

Eugenia_Draft_58, function unknown, 935                          1908   0.0
Waterdiva_58, DNA primase/helicase, 935                          1906   0.0
Thora_58, DNA primase/helicase, 935                              1906   0.0
Scoot17C_59, function unknown, 935                               1906   0.0
QueenBeane_58, DNA primase/helicase, 935                         1906   0.0
Piglet_0058, DNA primase/helicase, 935                           1906   0.0
PhrodoBaggins_59, DNA primase/helicase, 935                      1906   0.0
PhatCats2014_59, DNA primase/helicase, 935                       1906   0.0
Orwigg_Draft_59, function unknown, 935                           1906   0.0
Oosterbaan_58, function unknown, 935                             1906   0.0
Omniscient_59, DNA primase/helicase, 935                         1906   0.0
Nicole21_59, DNA primase/helicase, 935                           1906   0.0
Mulan_58, DNA primase/helicase, 935                              1906   0.0
Megatron_59, DNA primase/helicase, 935                           1906   0.0
Megamind20_Draft_59, function unknown, 935                       1906   0.0
Matalotodo_Draft_57, function unknown, 935                       1906   0.0
Lego3393_57, DNA primase/helicase, 935                           1906   0.0
Jillium_58, DNA primase/helicase, 935                            1906   0.0
Gareth_58, DNA primase/helicase, 935                             1906   0.0
DoesntMatter_56, DNA primase/helicase, 935                       1906   0.0
Chah_60, DNA primase/helicase, 935                               1906   0.0
```

b. SIF: NCBI BLAST

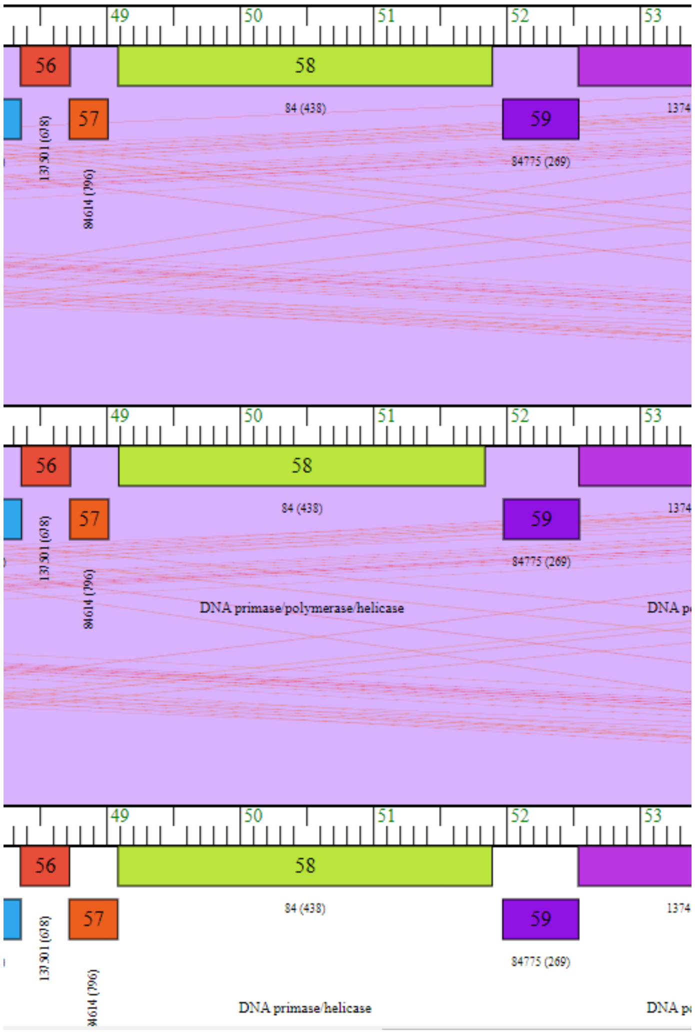


| | Description | Scientific Name | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Ac |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | DNA primase/helicase [Mycobacterium phage Chah] | Mycobacterium phage Chah | 1893 | 1893 | 100% | 0.0 | 99.89% | 935 | ACI127 |
| ☑ | DNA primase/helicase [Mycobacterium phage DelRivs] | Mycobacterium phage DelRivs | 1891 | 1891 | 100% | 0.0 | 99.79% | 935 | WNM6 |
| ☑ | DNA primase/helicase [Mycobacterium phage Hertubise] | Mycobacterium phage Hertubise | 1890 | 1890 | 100% | 0.0 | 99.79% | 935 | AEK09 |
| ☑ | DNA primase/helicase [Mycobacterium phage Puhltonio] | Mycobacterium phage Puhltonio | 1890 | 1890 | 100% | 0.0 | 99.68% | 935 | ACU41 |
| ☑ | DNA polymerase/primase [Mycobacterium phage Badfish] | Mycobacterium phage Badfish | 1890 | 1890 | 100% | 0.0 | 99.79% | 935 | YP_00 |
| ☑ | DNA primase/helicase [Mycobacterium phage Banjo] | Mycobacterium phage Banjo | 1890 | 1890 | 100% | 0.0 | 99.79% | 935 | AWN0 |
| ☑ | DNA primase/helicase [Mycobacterium phage Lulumae] | Mycobacterium phage Lulumae | 1890 | 1890 | 100% | 0.0 | 99.68% | 935 | ASZ73 |
| ☑ | DNA primase/helicase [Mycobacterium phage PinheadLarry] | Mycobacterium phage PinheadLarry | 1890 | 1890 | 100% | 0.0 | 99.57% | 935 | AXH69 |
| ☑ | DNA primase/helicase [Mycobacterium phage Pipsqueak] | Mycobacterium phage Pipsqueak | 1890 | 1890 | 100% | 0.0 | 99.68% | 935 | AJD82 |
| ☑ | DNA polymerase/primase [Mycobacterium phage Phipps] | Mycobacterium phage Phipps | 1889 | 1889 | 100% | 0.0 | 99.79% | 935 | YP_00 |
| ☑ | DNA primase/helicase [Mycobacterium phage Squid] | Mycobacterium phage Squid | 1889 | 1889 | 100% | 0.0 | 99.47% | 935 | ALH46 |
| ☑ | DNA primase/helicase [Mycobacterium phage Craff] | Mycobacterium phage Craff | 1889 | 1889 | 100% | 0.0 | 99.68% | 935 | AXC37 |
| ☑ | DNA polymerase/primase [Mycobacterium phage UncleHowie] | Mycobacterium phage UncleHowie | 1887 | 1887 | 100% | 0.0 | 99.47% | 935 | YP_00 |
| ☑ | DNA primase/helicase [Mycobacterium phage Freya] | Mycobacterium phage Freya | 1887 | 1887 | 100% | 0.0 | 99.47% | 935 | QWY8 |
| ☑ | DNA primase/helicase [Mycobacterium phage True] | Mycobacterium phage True | 1885 | 1885 | 100% | 0.0 | 99.25% | 935 | QWY8 |
| ☑ | DNA primase/helicase [Mycobacterium phage TomBombadil] | Mycobacterium phage TomBombadil | 1884 | 1884 | 100% | 0.0 | 99.25% | 935 | AZS10 |
| | DNA polymerase/primase [Mycobacterium phage ShiVal] | Mycobacterium phage ShiVal | 1884 | 1884 | 100% | 0.0 | 99.57% | 935 | YP_00 |

c. SIF: HHPred

No HHPRED evidence was selected

d. SIF: Synteny-Phamerator (three genomes)



7. Any other important information.

Synteny with Altwerkus, CamL, Chah and others

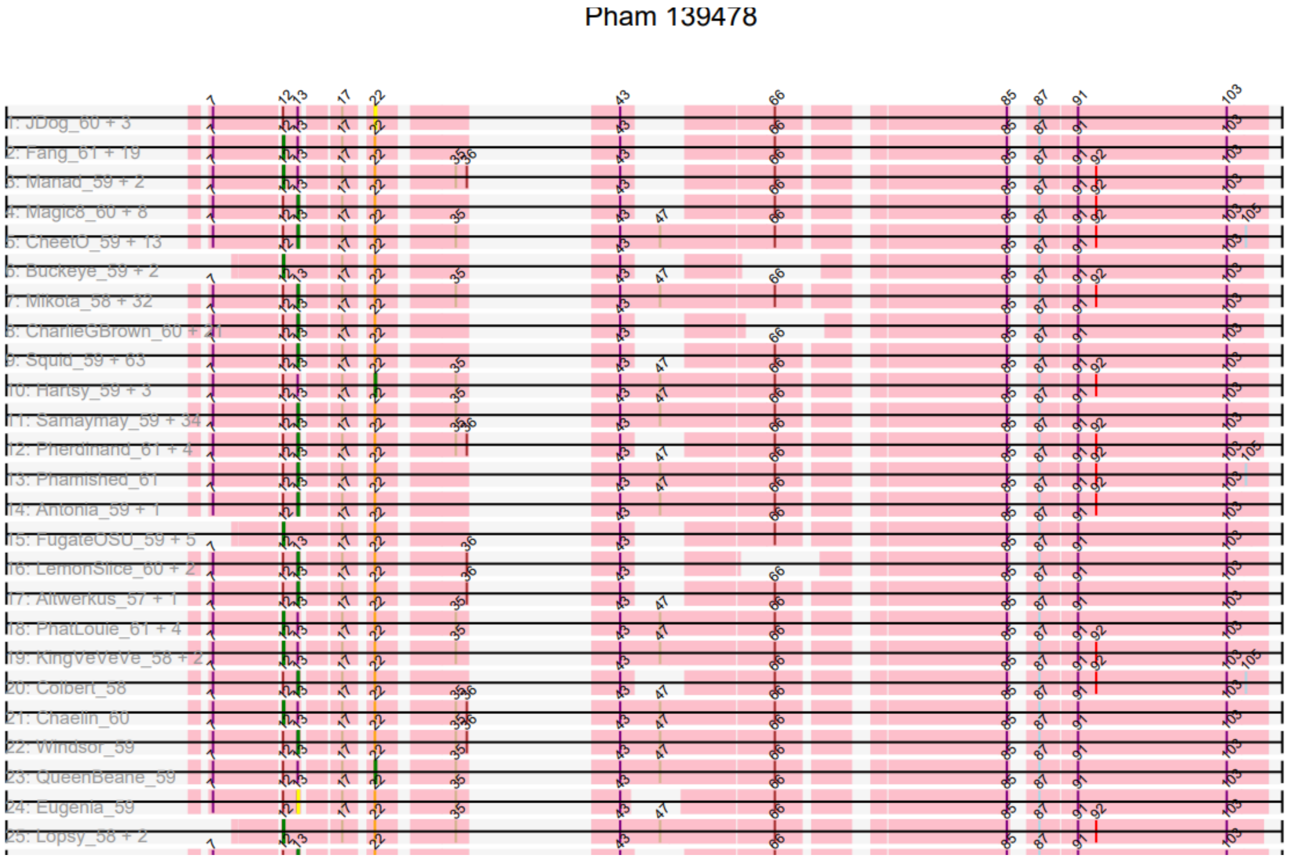
## CURATOR NAME: BELLA N

## GENE NAME: EUGENIA-59

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

1. Starterator

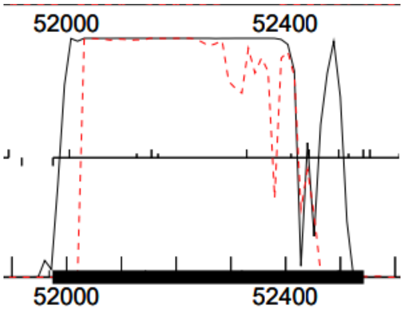


The start number called the most often in the published annotations is 13, it was called in 193 of the 275 non-draft genes in the pham.
Start 13:

- Found in 256 of 311 ( 82.3% ) of genes in pham

- Manual Annotations of this start: 193 of 275
- Called 78.9% of time when present

2. GeneMark coding potential



The black bar at the bottom represents the whole ORF it is about 600 nucleotides and this is a reverse gene.

3. Glimmer and GeneMark agreement

There is agreement between Glimmer and GeneMark

Glimmer Start: 52543

Glimmer Score: 11.59

GeneMark Start: 52543

4. Longest open reading frame (ORF) without excessive gap

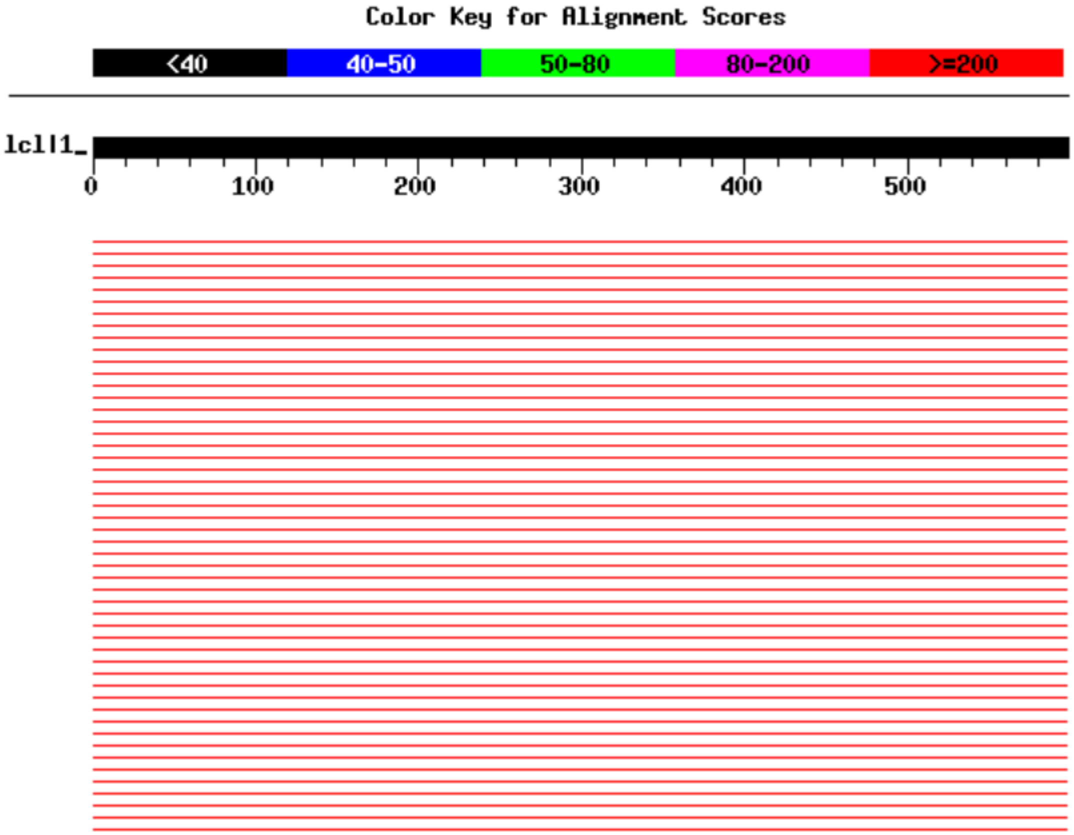
The selected gene is not the LORF with a length of 570 nucleotides. It has a Z-score of 1.6, a spacer of 10, and a final score of -5.585. This is a reverse gene.

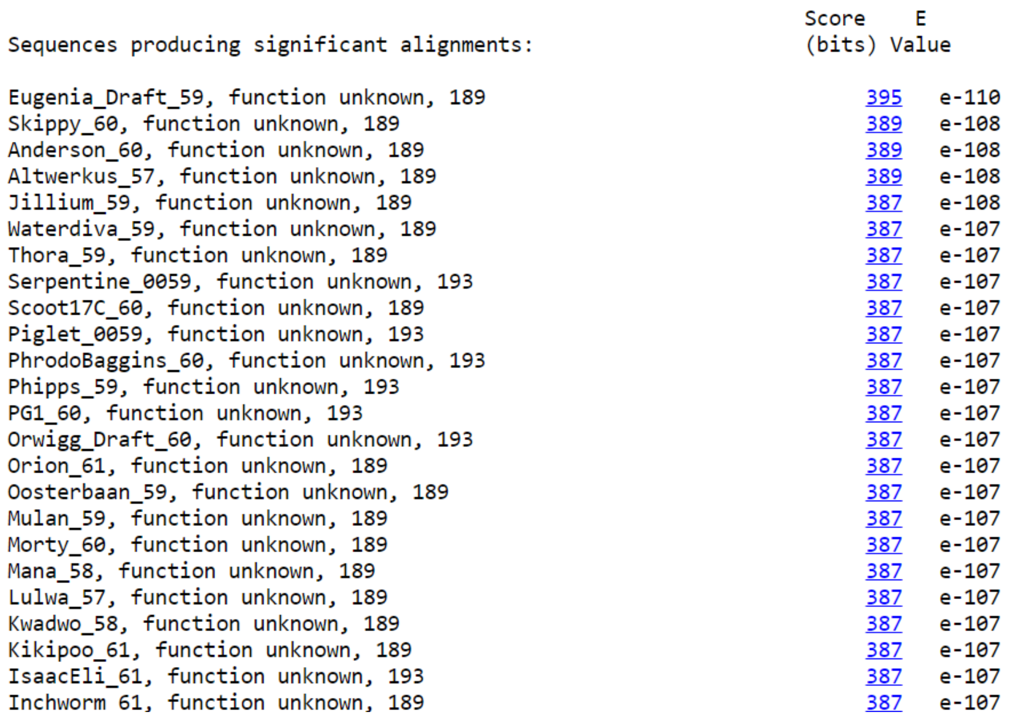
5. Function. If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

## 6. Supporting Information for Function (SIF)

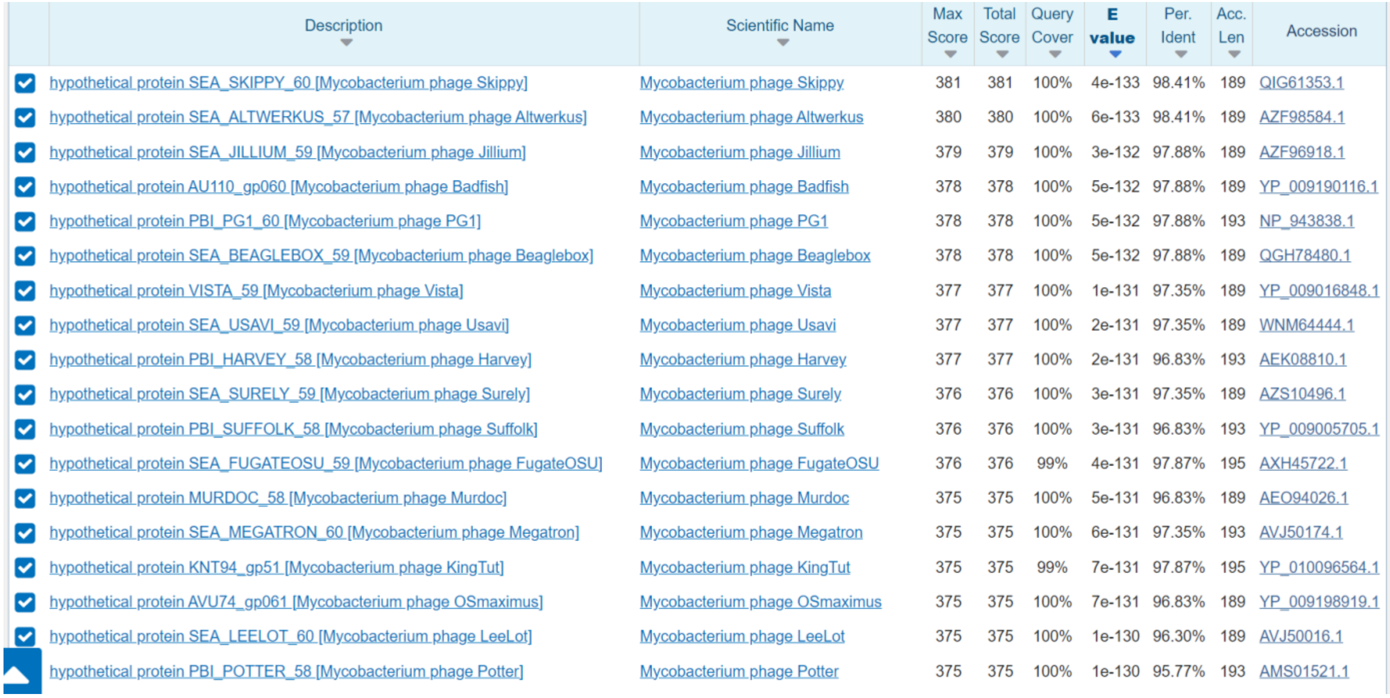
### a. SIF: PhageDb BLAST





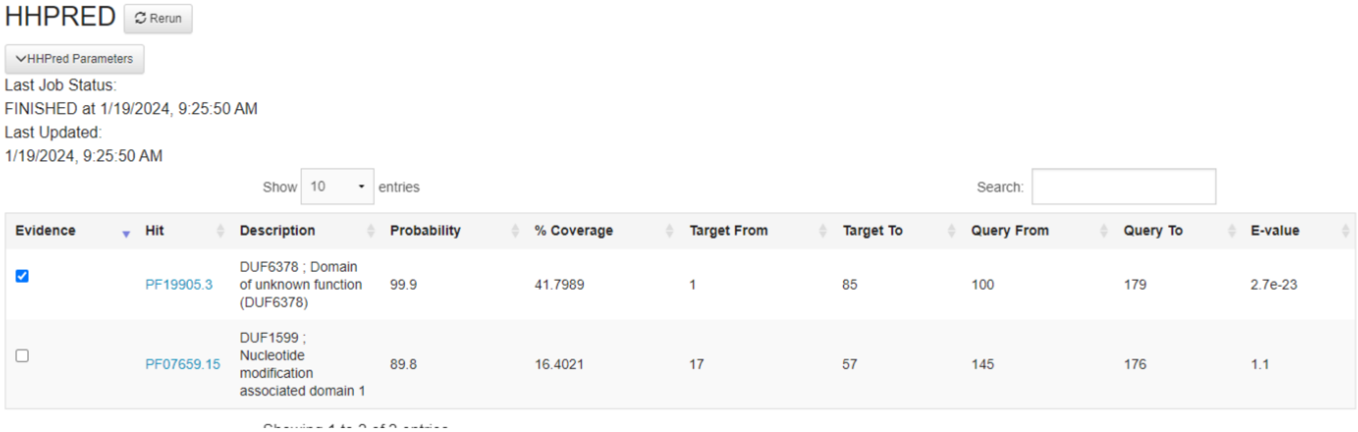
```
                                                              Score    E
Sequences producing significant alignments:                   (bits) Value

Eugenia_Draft_59, function unknown, 189                          395   e-110
Skippy_60, function unknown, 189                                 389   e-108
Anderson_60, function unknown, 189                               389   e-108
Altwerkus_57, function unknown, 189                              389   e-108
Jillium_59, function unknown, 189                                387   e-108
Waterdiva_59, function unknown, 189                              387   e-107
Thora_59, function unknown, 189                                  387   e-107
Serpentine_0059, function unknown, 193                           387   e-107
Scoot17C_60, function unknown, 189                               387   e-107
Piglet_0059, function unknown, 193                               387   e-107
PhrodoBaggins_60, function unknown, 193                          387   e-107
Phipps_59, function unknown, 193                                 387   e-107
PG1_60, function unknown, 193                                    387   e-107
Orwigg_Draft_60, function unknown, 193                           387   e-107
Orion_61, function unknown, 189                                  387   e-107
Oosterbaan_59, function unknown, 189                             387   e-107
Mulan_59, function unknown, 189                                  387   e-107
Morty_60, function unknown, 189                                  387   e-107
Mana_58, function unknown, 189                                   387   e-107
Lulwa_57, function unknown, 189                                  387   e-107
Kwadwo_58, function unknown, 189                                 387   e-107
Kikipoo_61, function unknown, 189                                387   e-107
IsaacEli_61, function unknown, 193                               387   e-107
Inchworm 61, function unknown, 189                               387   e-107
```

## b. SIF: NCBI BLAST



| | Description | Scientific Name | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | hypothetical protein SEA_SKIPPY_60 [Mycobacterium phage Skippy] | Mycobacterium phage Skippy | 381 | 381 | 100% | 4e-133 | 98.41% | 189 | QIG61353.1 |
| ☑ | hypothetical protein SEA_ALTWERKUS_57 [Mycobacterium phage Altwerkus] | Mycobacterium phage Altwerkus | 380 | 380 | 100% | 6e-133 | 98.41% | 189 | AZF98584.1 |
| ☑ | hypothetical protein SEA_JILLIUM_59 [Mycobacterium phage Jillium] | Mycobacterium phage Jillium | 379 | 379 | 100% | 3e-132 | 97.88% | 189 | AZF96918.1 |
| ☑ | hypothetical protein AU110_gp060 [Mycobacterium phage Badfish] | Mycobacterium phage Badfish | 378 | 378 | 100% | 5e-132 | 97.88% | 189 | YP_009190116.1 |
| ☑ | hypothetical protein PBI_PG1_60 [Mycobacterium phage PG1] | Mycobacterium phage PG1 | 378 | 378 | 100% | 5e-132 | 97.88% | 193 | NP_943838.1 |
| ☑ | hypothetical protein SEA_BEAGLEBOX_59 [Mycobacterium phage Beaglebox] | Mycobacterium phage Beaglebox | 378 | 378 | 100% | 5e-132 | 97.88% | 189 | QGH78480.1 |
| ☑ | hypothetical protein VISTA_59 [Mycobacterium phage Vista] | Mycobacterium phage Vista | 377 | 377 | 100% | 1e-131 | 97.35% | 189 | YP_009016848.1 |
| ☑ | hypothetical protein SEA_USAVI_59 [Mycobacterium phage Usavi] | Mycobacterium phage Usavi | 377 | 377 | 100% | 2e-131 | 97.35% | 189 | WNM64444.1 |
| ☑ | hypothetical protein PBI_HARVEY_58 [Mycobacterium phage Harvey] | Mycobacterium phage Harvey | 377 | 377 | 100% | 2e-131 | 96.83% | 193 | AEK08810.1 |
| ☑ | hypothetical protein SEA_SURELY_59 [Mycobacterium phage Surely] | Mycobacterium phage Surely | 376 | 376 | 100% | 3e-131 | 97.35% | 189 | AZS10496.1 |
| ☑ | hypothetical protein PBI_SUFFOLK_58 [Mycobacterium phage Suffolk] | Mycobacterium phage Suffolk | 376 | 376 | 100% | 3e-131 | 96.83% | 193 | YP_009005705.1 |
| ☑ | hypothetical protein SEA_FUGATEOSU_59 [Mycobacterium phage FugateOSU] | Mycobacterium phage FugateOSU | 376 | 376 | 99% | 4e-131 | 97.87% | 195 | AXH45722.1 |
| ☑ | hypothetical protein MURDOC_58 [Mycobacterium phage Murdoc] | Mycobacterium phage Murdoc | 375 | 375 | 100% | 5e-131 | 96.83% | 189 | AEO94026.1 |
| ☑ | hypothetical protein SEA_MEGATRON_60 [Mycobacterium phage Megatron] | Mycobacterium phage Megatron | 375 | 375 | 100% | 6e-131 | 97.35% | 193 | AVJ50174.1 |
| ☑ | hypothetical protein KNT94_gp51 [Mycobacterium phage KingTut] | Mycobacterium phage KingTut | 375 | 375 | 99% | 7e-131 | 97.87% | 195 | YP_010096564.1 |
| ☑ | hypothetical protein AVU74_gp061 [Mycobacterium phage OSmaximus] | Mycobacterium phage OSmaximus | 375 | 375 | 100% | 7e-131 | 96.83% | 189 | YP_009198919.1 |
| ☑ | hypothetical protein SEA_LEELOT_60 [Mycobacterium phage LeeLot] | Mycobacterium phage LeeLot | 375 | 375 | 100% | 1e-130 | 96.30% | 189 | AVJ50016.1 |
| | hypothetical protein PBI_POTTER_58 [Mycobacterium phage Potter] | Mycobacterium phage Potter | 375 | 375 | 100% | 1e-130 | 95.77% | 193 | AMS01521.1 |

## c. SIF: HHPred

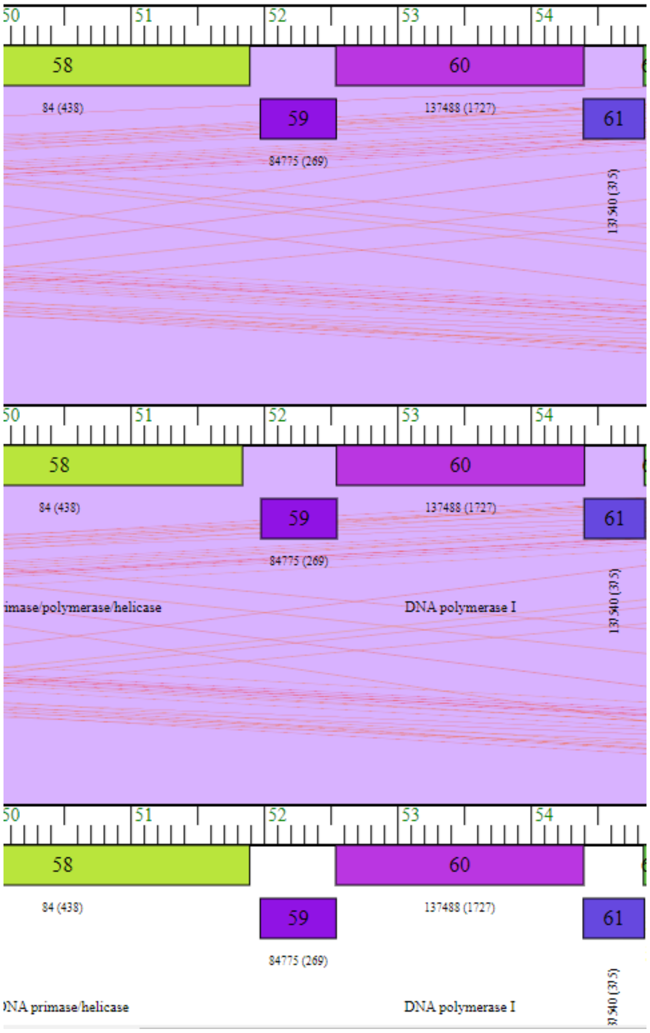


HHPRED

Last Job Status:
FINISHED at 1/19/2024, 9:25:50 AM
Last Updated:
1/19/2024, 9:25:50 AM

| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | PF19905.3 | DUF6378 ; Domain of unknown function (DUF6378) | 99.9 | 41.7989 | 1 | 85 | 100 | 179 | 2.7e-23 |
| ☐ | PF07659.15 | DUF1599 ; Nucleotide modification associated domain 1 | 89.8 | 16.4021 | 17 | 57 | 145 | 176 | 1.1 |

## d. SIF: Synteny-Phamerator (three genomes)



7. Any other important information.

Synteny with Altwerkus, Anderson, Skippy and more.

No TmHmms

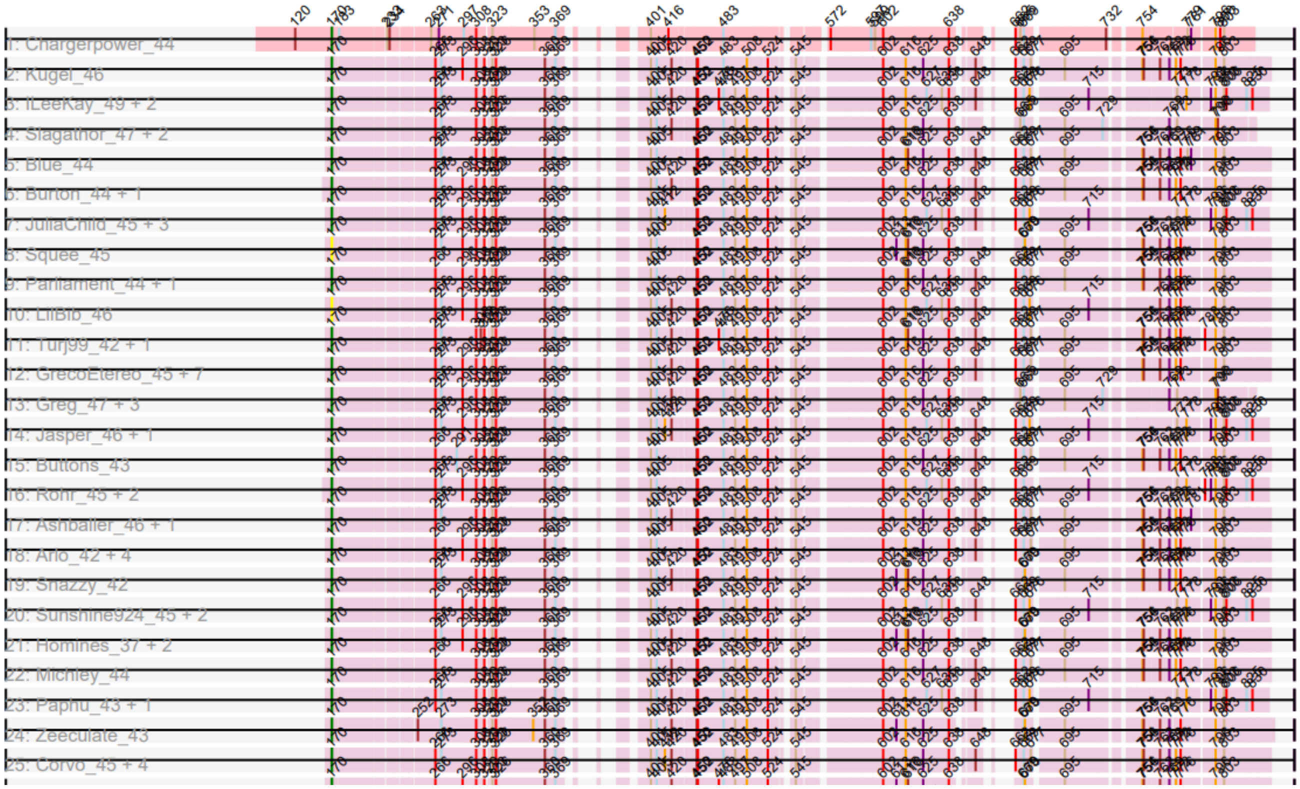
**CURATOR NAME: BELLA N**

**GENE NAME: EUGENIA-60**

**DNA MASTER NOTES: N/A**

**START POSITION EVALUATION (IN ORDER)**:

1. Starterator

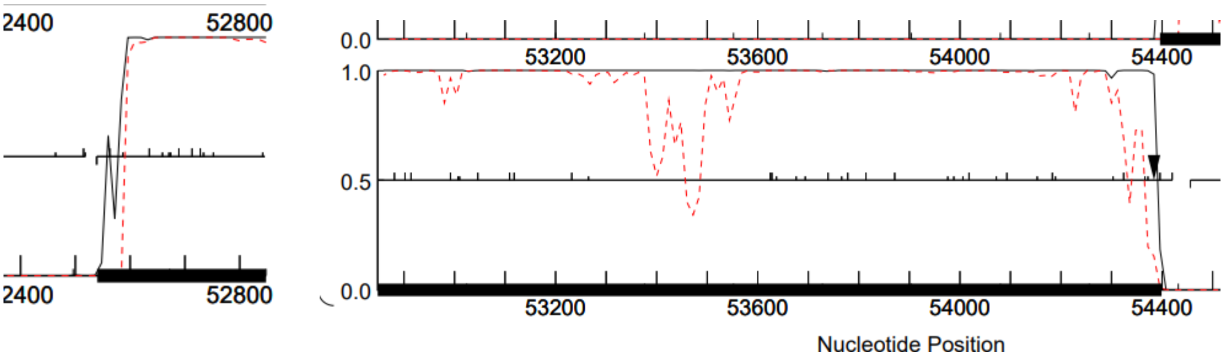


The start number called the most often in the published annotations is 170, it was called in 872 of the 1333 non-draft genes in the pham.
Start 170:

- Found in 987 of 1492 ( 66.2% ) of genes in pham
- Manual Annotations of this start: 872 of 1333

- Called 97.3% of time when present

2. GeneMark coding potential



The black bar at the bottom represents the whole ORF it is about 1800 nucleotides and this is a reverse gene. There is a triangle indicating that there my be a shift, but because this gene is not a tail assembly chaperone it can be dismissed.

3. Glimmer and GeneMark agreement

There is agreement between Glimmer and GeneMark

Glimmer Start: 54399

Glimmer Score: 12.71

GeneMark Start: 54399

4. Longest open reading frame (ORF) without excessive gap

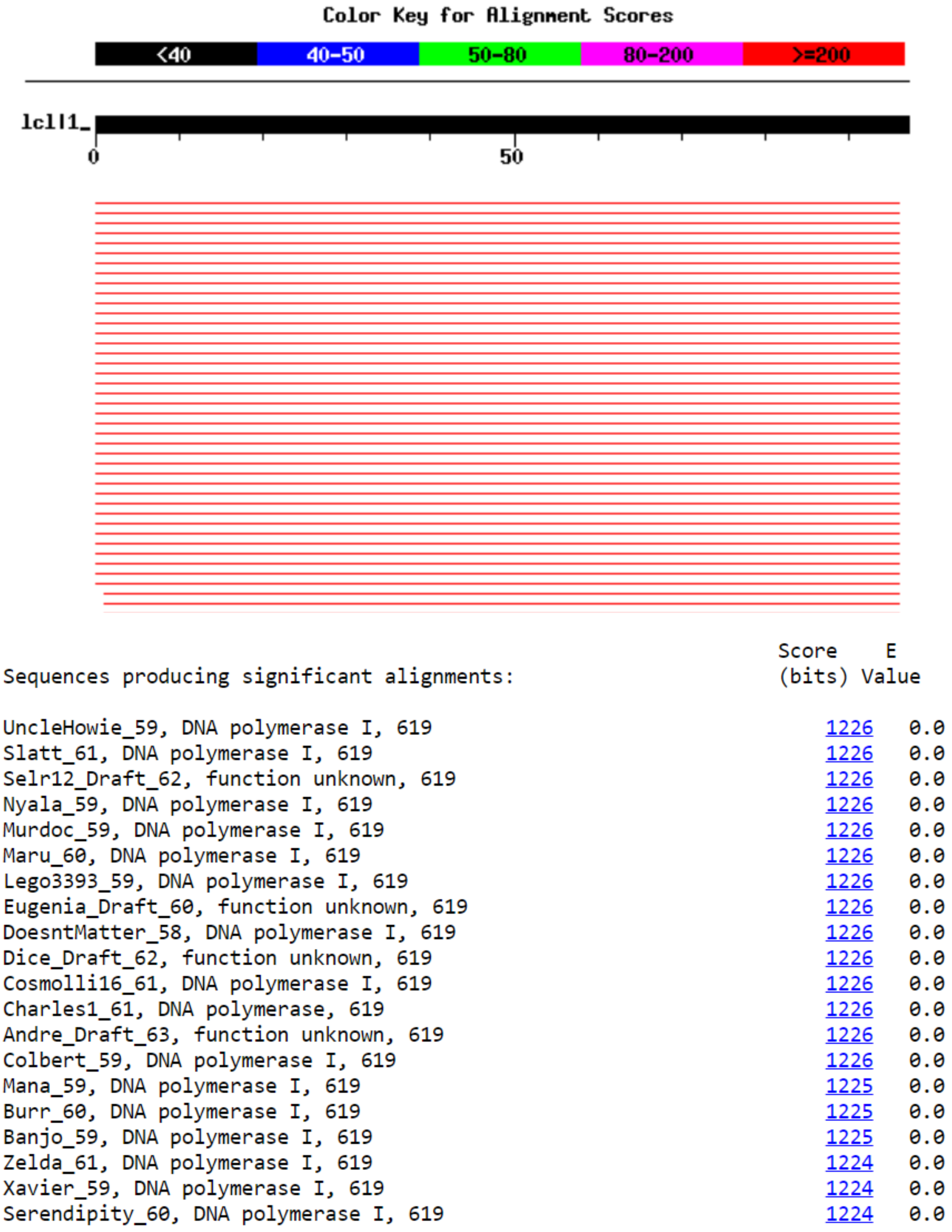
The selected gene is not the LORF with a length of 1860 nucleotides. It has a Z-score of 1.979, a spacer of 11, and a final score of  -4.888. This is a reverse gene.

5. Function. If no functional prediction is present, write “Hypothetical protein”.

DNA polymerase I

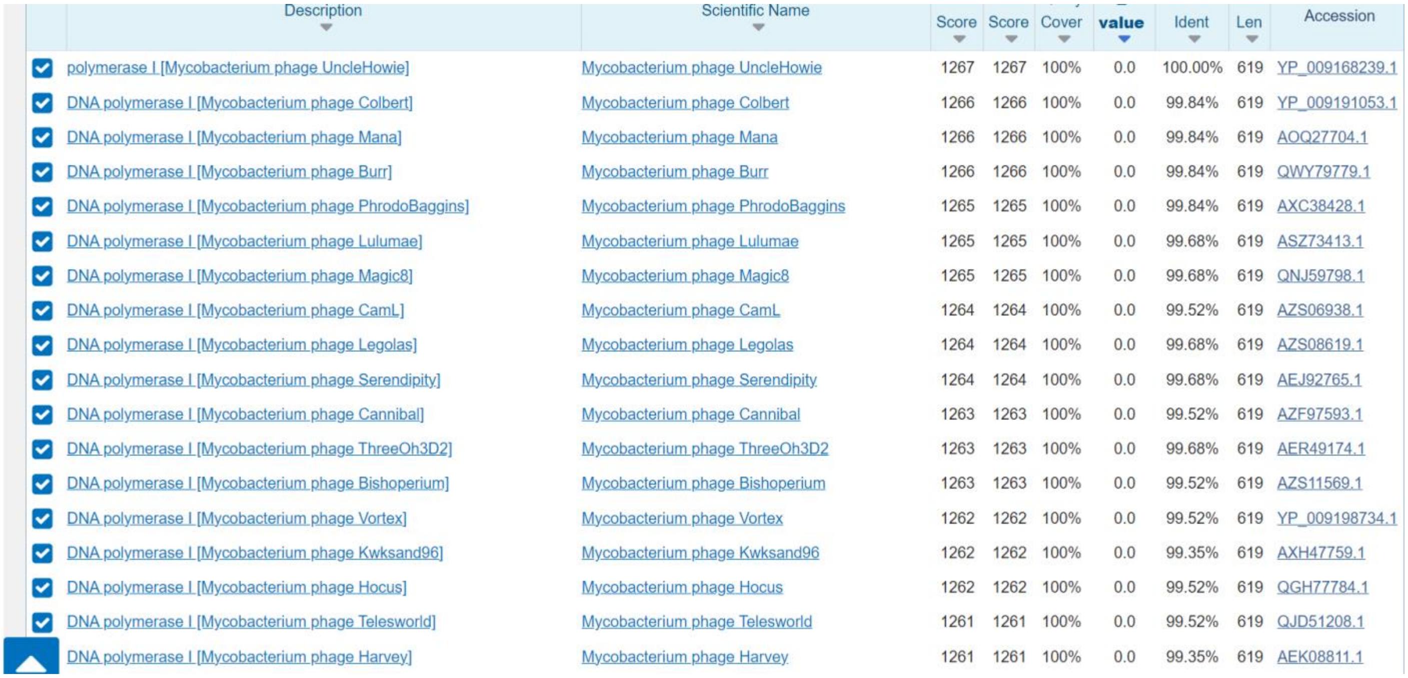
## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST



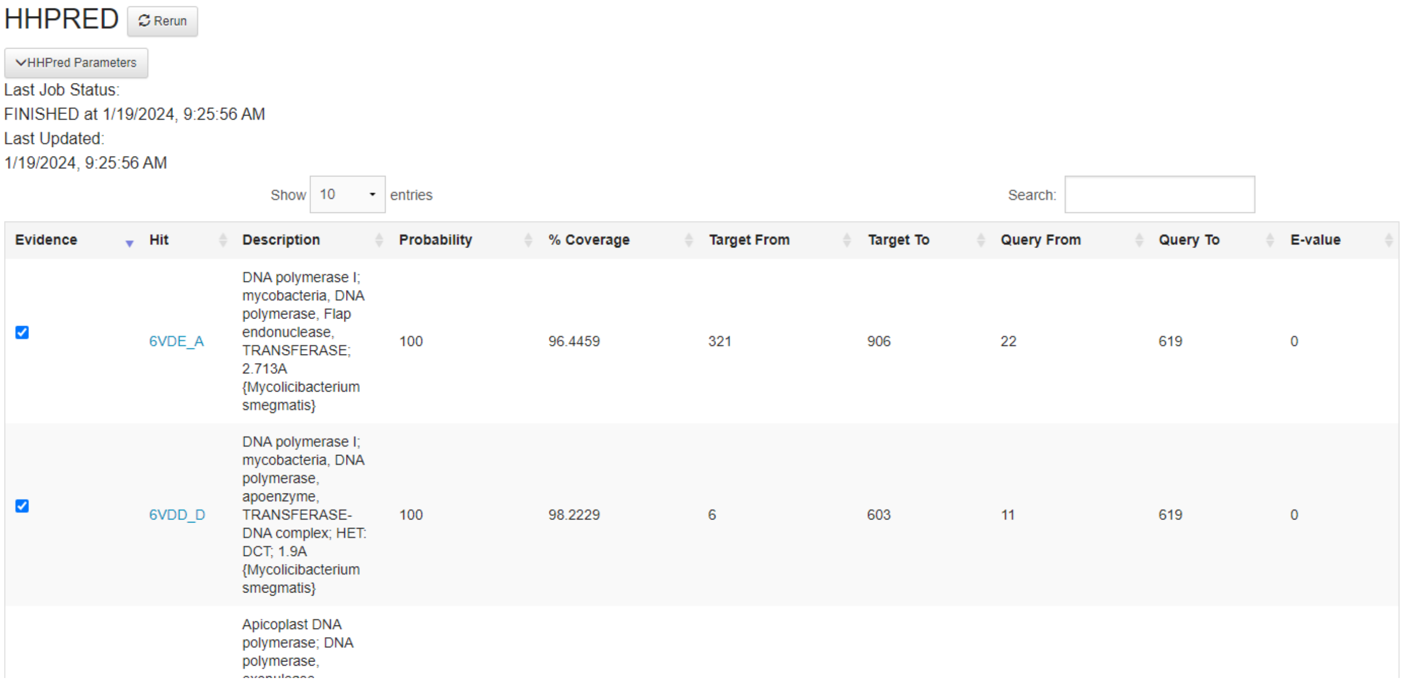
```
                                                                  Score     E
Sequences producing significant alignments:                      (bits)  Value

UncleHowie_59, DNA polymerase I, 619                               1226    0.0
Slatt_61, DNA polymerase I, 619                                    1226    0.0
Selr12_Draft_62, function unknown, 619                             1226    0.0
Nyala_59, DNA polymerase I, 619                                    1226    0.0
Murdoc_59, DNA polymerase I, 619                                   1226    0.0
Maru_60, DNA polymerase I, 619                                     1226    0.0
Lego3393_59, DNA polymerase I, 619                                 1226    0.0
Eugenia_Draft_60, function unknown, 619                            1226    0.0
DoesntMatter_58, DNA polymerase I, 619                             1226    0.0
Dice_Draft_62, function unknown, 619                               1226    0.0
Cosmolli16_61, DNA polymerase I, 619                               1226    0.0
Charles1_61, DNA polymerase, 619                                   1226    0.0
Andre_Draft_63, function unknown, 619                              1226    0.0
Colbert_59, DNA polymerase I, 619                                  1226    0.0
Mana_59, DNA polymerase I, 619                                     1225    0.0
Burr_60, DNA polymerase I, 619                                     1225    0.0
Banjo_59, DNA polymerase I, 619                                    1225    0.0
Zelda_61, DNA polymerase I, 619                                    1224    0.0
Xavier_59, DNA polymerase I, 619                                   1224    0.0
Serendipity_60, DNA polymerase I, 619                              1224    0.0
```

## b. SIF: NCBI BLAST



| | Description | Scientific Name | Score | Score | Cover | value | Ident | Len | Accession |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | polymerase I [Mycobacterium phage UncleHowie] | Mycobacterium phage UncleHowie | 1267 | 1267 | 100% | 0.0 | 100.00% | 619 | YP_009168239.1 |
| ☑ | DNA polymerase I [Mycobacterium phage Colbert] | Mycobacterium phage Colbert | 1266 | 1266 | 100% | 0.0 | 99.84% | 619 | YP_009191053.1 |
| ☑ | DNA polymerase I [Mycobacterium phage Mana] | Mycobacterium phage Mana | 1266 | 1266 | 100% | 0.0 | 99.84% | 619 | AOQ27704.1 |
| ☑ | DNA polymerase I [Mycobacterium phage Burr] | Mycobacterium phage Burr | 1266 | 1266 | 100% | 0.0 | 99.84% | 619 | QWY79779.1 |
| ☑ | DNA polymerase I [Mycobacterium phage PhrodoBaggins] | Mycobacterium phage PhrodoBaggins | 1265 | 1265 | 100% | 0.0 | 99.84% | 619 | AXC38428.1 |
| ☑ | DNA polymerase I [Mycobacterium phage Lulumae] | Mycobacterium phage Lulumae | 1265 | 1265 | 100% | 0.0 | 99.68% | 619 | ASZ73413.1 |
| ☑ | DNA polymerase I [Mycobacterium phage Magic8] | Mycobacterium phage Magic8 | 1265 | 1265 | 100% | 0.0 | 99.68% | 619 | QNJ59798.1 |
| ☑ | DNA polymerase I [Mycobacterium phage CamL] | Mycobacterium phage CamL | 1264 | 1264 | 100% | 0.0 | 99.52% | 619 | AZS06938.1 |
| ☑ | DNA polymerase I [Mycobacterium phage Legolas] | Mycobacterium phage Legolas | 1264 | 1264 | 100% | 0.0 | 99.68% | 619 | AZS08619.1 |
| ☑ | DNA polymerase I [Mycobacterium phage Serendipity] | Mycobacterium phage Serendipity | 1264 | 1264 | 100% | 0.0 | 99.68% | 619 | AEJ92765.1 |
| ☑ | DNA polymerase I [Mycobacterium phage Cannibal] | Mycobacterium phage Cannibal | 1263 | 1263 | 100% | 0.0 | 99.52% | 619 | AZF97593.1 |
| ☑ | DNA polymerase I [Mycobacterium phage ThreeOh3D2] | Mycobacterium phage ThreeOh3D2 | 1263 | 1263 | 100% | 0.0 | 99.68% | 619 | AER49174.1 |
| ☑ | DNA polymerase I [Mycobacterium phage Bishoperium] | Mycobacterium phage Bishoperium | 1263 | 1263 | 100% | 0.0 | 99.52% | 619 | AZS11569.1 |
| ☑ | DNA polymerase I [Mycobacterium phage Vortex] | Mycobacterium phage Vortex | 1262 | 1262 | 100% | 0.0 | 99.52% | 619 | YP_009198734.1 |
| ☑ | DNA polymerase I [Mycobacterium phage Kwksand96] | Mycobacterium phage Kwksand96 | 1262 | 1262 | 100% | 0.0 | 99.35% | 619 | AXH47759.1 |
| ☑ | DNA polymerase I [Mycobacterium phage Hocus] | Mycobacterium phage Hocus | 1262 | 1262 | 100% | 0.0 | 99.52% | 619 | QGH77784.1 |
| ☑ | DNA polymerase I [Mycobacterium phage Telesworld] | Mycobacterium phage Telesworld | 1261 | 1261 | 100% | 0.0 | 99.52% | 619 | QJD51208.1 |
| | DNA polymerase I [Mycobacterium phage Harvey] | Mycobacterium phage Harvey | 1261 | 1261 | 100% | 0.0 | 99.35% | 619 | AEK08811.1 |

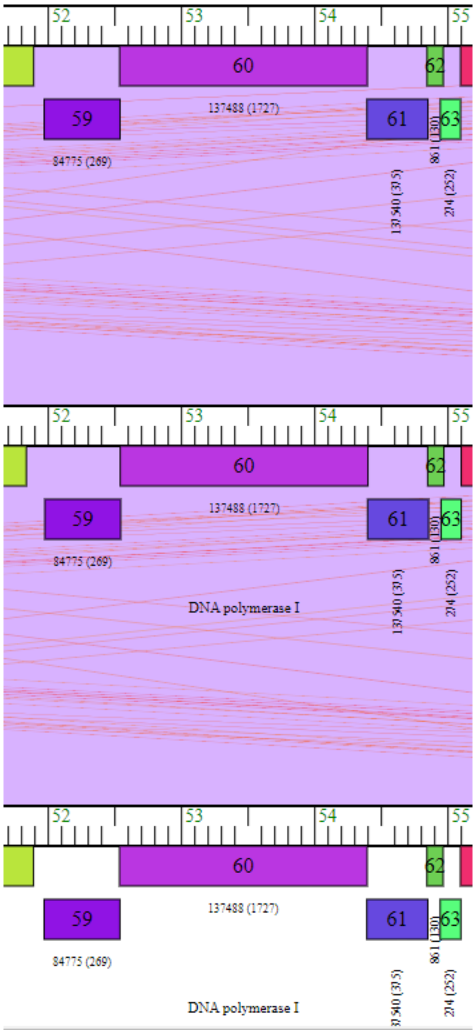
## c. SIF: HHPred



Last Job Status:
FINISHED at 1/19/2024, 9:25:56 AM
Last Updated:
1/19/2024, 9:25:56 AM

| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | 6VDE_A | DNA polymerase I; mycobacteria, DNA polymerase, Flap endonuclease, TRANSFERASE; 2.713A {Mycolicibacterium smegmatis} | 100 | 96.4459 | 321 | 906 | 22 | 619 | 0 |
| ☑ | 6VDD_D | DNA polymerase I; mycobacteria, DNA polymerase, apoenzyme, TRANSFERASE-DNA complex; HET: DCT; 1.9A {Mycolicibacterium smegmatis} | 100 | 98.2229 | 6 | 603 | 11 | 619 | 0 |
| | | Apicoplast DNA polymerase; DNA polymerase, | | | | | | | |

## d. SIF: Synteny-Phamerator (three genomes)



7. Any other important information.

Synteny with Charles1, Colbert, Cosmolli16, DoesntMatter and more.

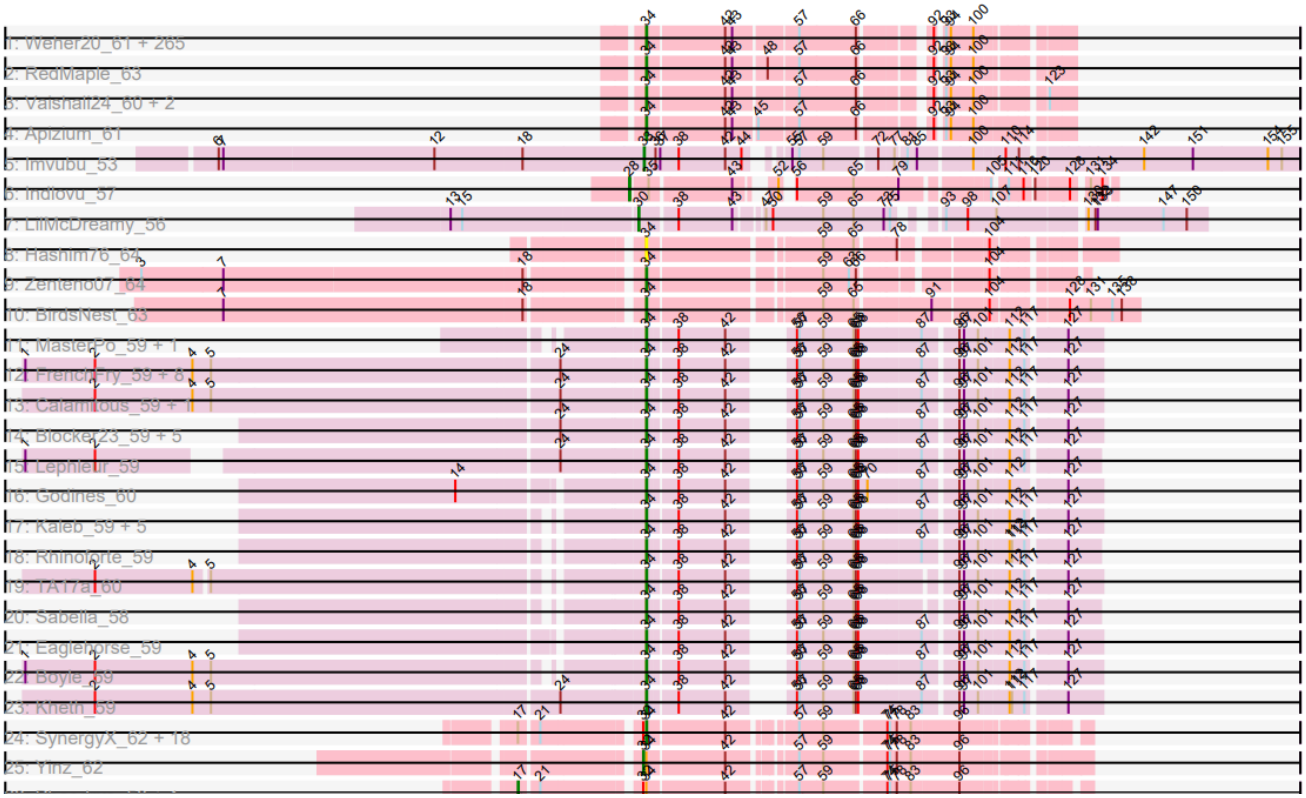
**CURATOR NAME: BELLA N**

**GENE NAME: EUGENIA-61**

**DNA MASTER NOTES: N/A**

**START POSITION EVALUATION (IN ORDER):**

1. Starterator

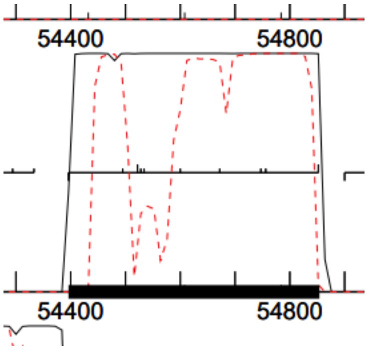


The start number called the most often in the published annotations is 34, it was called in 313 of the 348 non-draft genes in the pham.
Start 34:

- Found in 359 of 389 ( 92.3% ) of genes in pham
- Manual Annotations of this start: 313 of 348

- Called 96.4% of time when present

2. GeneMark coding potential



The black bar at the bottom represents the whole ORF it is about 500 nucleotides and this is a reverse gene.

3. Glimmer and GeneMark agreement

There is agreement between Glimmer and GeneMark

Glimmer Start: 54854

Glimmer Score: 13.13

GeneMark Start: 54854

4. Longest open reading frame (ORF) without excessive gap

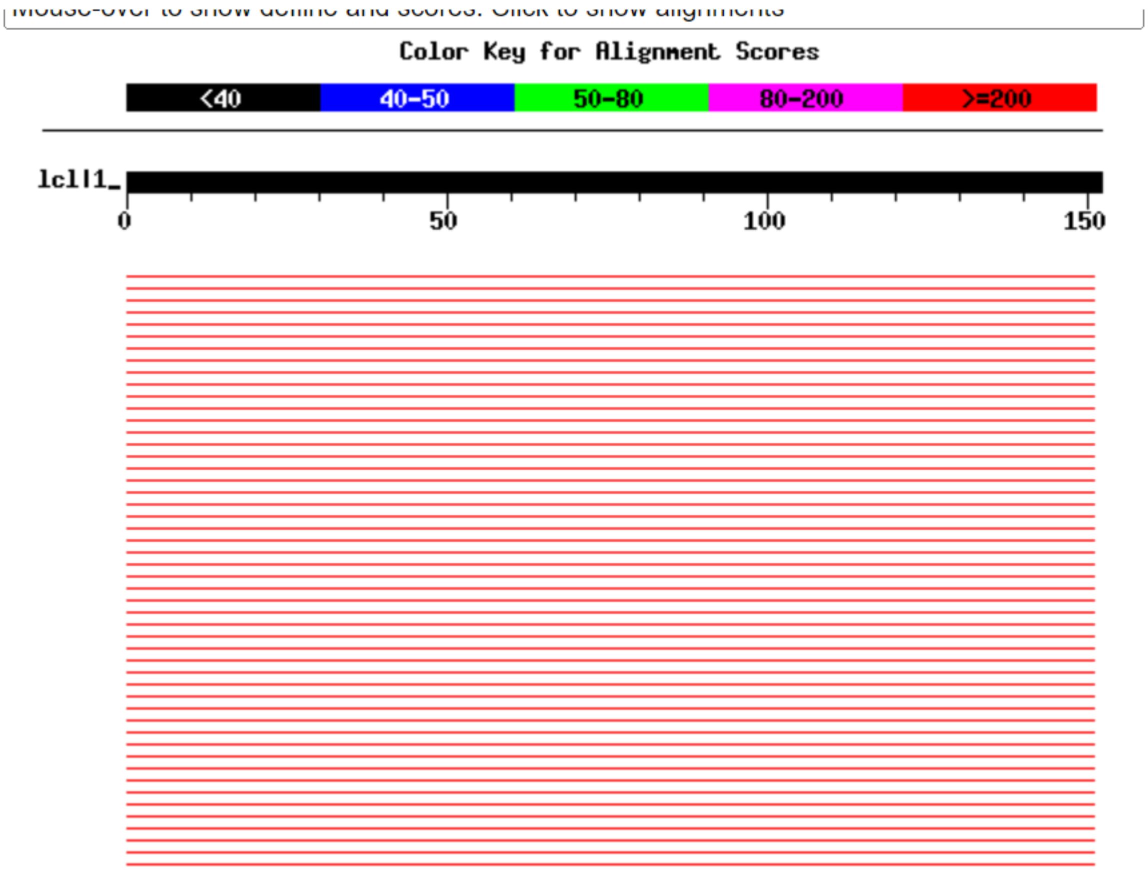
The selected gene is the LORF with a length of 459 nucleotides. It has a Z-score of 1.399, a spacer of 11, and a final score of  -6.053. This is a reverse gene.

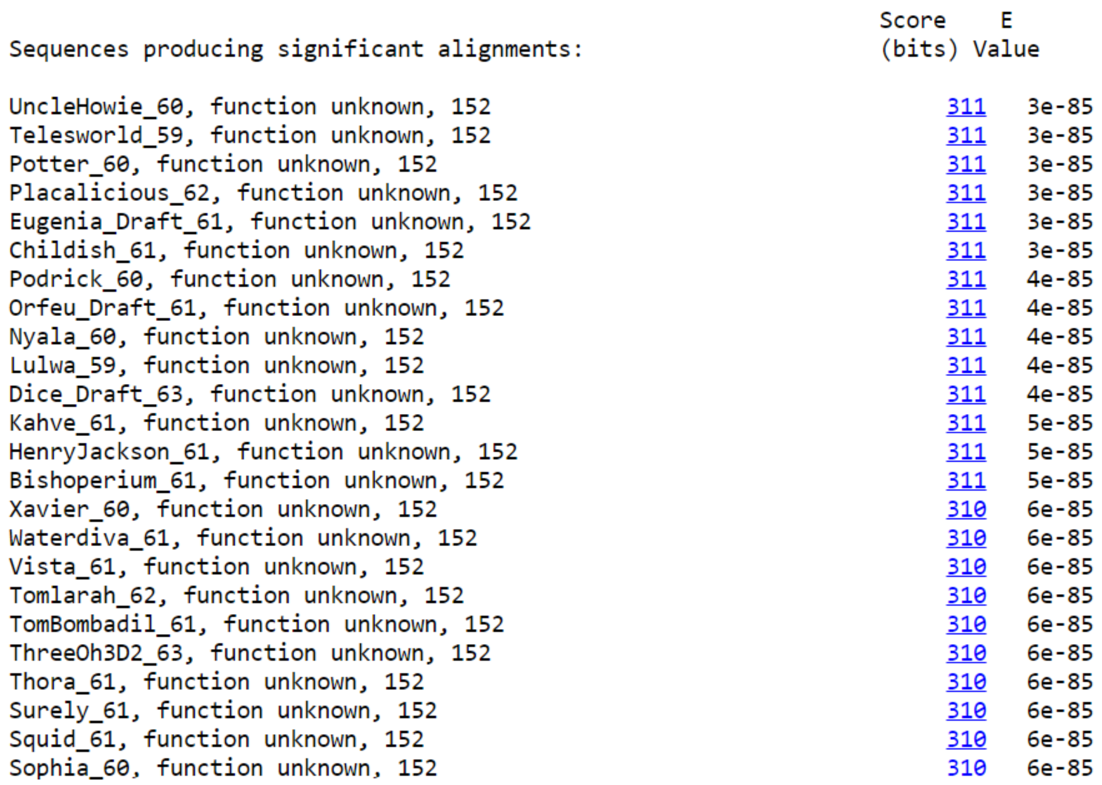
5. Function. If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

6. Supporting Information for Function (SIF)

a. SIF: PhageDb BLAST





```
                                                                  Score     E
Sequences producing significant alignments:                       (bits)  Value

UncleHowie_60, function unknown, 152                                 311    3e-85
Telesworld_59, function unknown, 152                                 311    3e-85
Potter_60, function unknown, 152                                     311    3e-85
Placalicious_62, function unknown, 152                               311    3e-85
Eugenia_Draft_61, function unknown, 152                              311    3e-85
Childish_61, function unknown, 152                                   311    3e-85
Podrick_60, function unknown, 152                                    311    4e-85
Orfeu_Draft_61, function unknown, 152                                311    4e-85
Nyala_60, function unknown, 152                                      311    4e-85
Lulwa_59, function unknown, 152                                      311    4e-85
Dice_Draft_63, function unknown, 152                                 311    4e-85
Kahve_61, function unknown, 152                                      311    5e-85
HenryJackson_61, function unknown, 152                               311    5e-85
Bishoperium_61, function unknown, 152                                311    5e-85
Xavier_60, function unknown, 152                                     310    6e-85
Waterdiva_61, function unknown, 152                                  310    6e-85
Vista_61, function unknown, 152                                      310    6e-85
Tomlarah_62, function unknown, 152                                   310    6e-85
TomBombadil_61, function unknown, 152                                310    6e-85
ThreeOh3D2_63, function unknown, 152                                 310    6e-85
Thora_61, function unknown, 152                                      310    6e-85
Surely_61, function unknown, 152                                     310    6e-85
Squid_61, function unknown, 152                                      310    6e-85
Sophia_60, function unknown, 152                                     310    6e-85
```

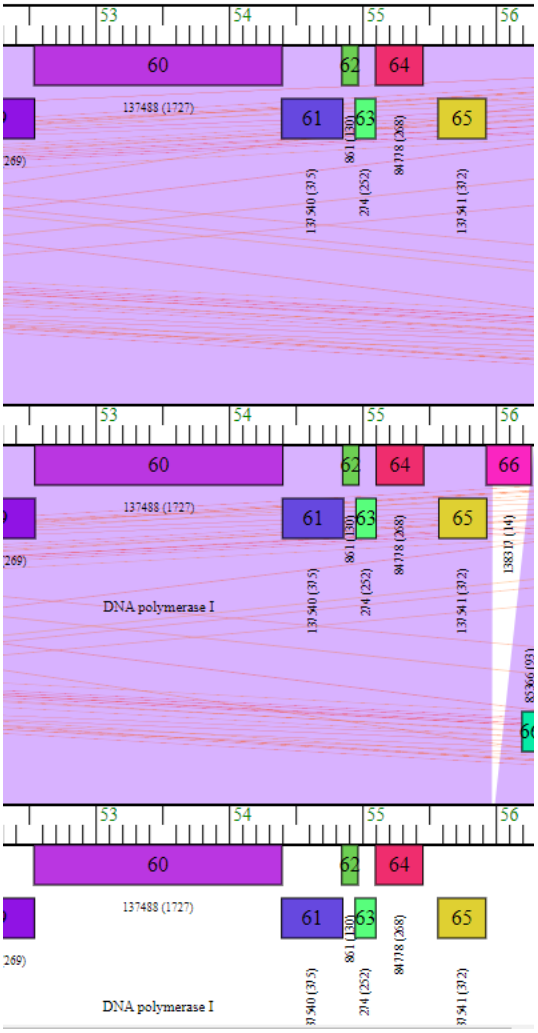
b. SIF: NCBI BLAST

| | Description | Scientific Name | Max Score | Total Score | Query Cover | E value | Per. Ident |
|---|---|---|---|---|---|---|---|
| ☑ | hypothetical protein UNCLEHOWIE_60 [Mycobacterium phage UncleHowie] | Mycobacterium phage UncleHowie | 308 | 308 | 100% | 1e-105 | 100.00% |
| ☑ | hypothetical protein SEA_PODRICK_60 [Mycobacterium phage Podrick] | Mycobacterium phage Podrick | 308 | 308 | 100% | 2e-105 | 99.34% |
| ☑ | hypothetical protein VISTA_61 [Mycobacterium phage Vista] | Mycobacterium phage Vista | 307 | 307 | 100% | 3e-105 | 99.34% |
| ☑ | hypothetical protein SEA_KAHVE_61 [Mycobacterium phage Kahve] | Mycobacterium phage Kahve | 307 | 307 | 100% | 3e-105 | 98.68% |
| ☑ | hypothetical protein AVV54_gp063 [Mycobacterium phage Kikipoo] | Mycobacterium phage Kikipoo | 307 | 307 | 100% | 5e-105 | 98.68% |
| ☑ | hypothetical protein Nacho_0063 [Mycobacterium phage Nacho] | Mycobacterium phage Nacho | 306 | 306 | 100% | 8e-105 | 98.03% |
| ☑ | hypothetical protein SEA_JILLIUM_61 [Mycobacterium phage Jillium] | Mycobacterium phage Jillium | 306 | 306 | 100% | 9e-105 | 98.68% |
| ☑ | hypothetical protein SEA_CHORKPOP_61 [Mycobacterium phage Chorkpop] | Mycobacterium phage Chorkpop | 306 | 306 | 100% | 1e-104 | 98.68% |
| ☑ | hypothetical protein SEA_TIMMI_60 [Mycobacterium phage Timmi] | Mycobacterium phage Timmi | 305 | 305 | 100% | 2e-104 | 98.68% |
| ☑ | hypothetical protein SEA_FRINGE_65 [Mycobacterium phage Fringe] | Mycobacterium phage Fringe | 305 | 305 | 100% | 2e-104 | 98.03% |
| ☑ | hypothetical protein SEA_DADDYDANIELS_63 [Mycobacterium phage DaddyDani… | Mycobacterium phage DaddyDaniels | 305 | 305 | 100% | 2e-104 | 98.03% |
| ☑ | hypothetical protein SEA_BLUEPHACEBABY_61 [Mycobacterium phage Bluephac… | Mycobacterium phage Bluephacebaby | 305 | 305 | 100% | 2e-104 | 98.68% |
| ☑ | hypothetical protein TALLGRASSMM_61 [Mycobacterium phage TallGrassMM] | Mycobacterium phage TallGrassMM | 305 | 305 | 100% | 3e-104 | 98.03% |
| ☑ | hypothetical protein SEA_MULAN_61 [Mycobacterium phage Mulan] | Mycobacterium phage Mulan | 305 | 305 | 100% | 3e-104 | 97.37% |
| ☑ | hypothetical protein PBI_EMPTEE_61 [Mycobacterium phage EmpTee] | Mycobacterium phage EmpTee | 305 | 305 | 100% | 4e-104 | 97.37% |
| ☑ | hypothetical protein SEA_HIGHSTUMP_63 [Mycobacterium phage HighStump] | Mycobacterium phage HighStump | 305 | 305 | 100% | 4e-104 | 97.37% |
| | hypothetical protein AU108_gp61 [Mycobacterium phage Eremos] | Mycobacterium phage Eremos | 304 | 304 | 100% | 6e-104 | 98.03% |
| | hypothetical protein SEA_PINHEADLARRY_62 [Mycobacterium phage PinheadLarry] | Mycobacterium phage PinheadLarry | 304 | 304 | 100% | 6e-104 | 97.37% |

c. SIF: HHPred

No HHPRED evidence selected

d. SIF: Synteny-Phamerator (three genomes)



7. Any other important information.

Synteny with Bishoperium, Childish, HenryJackson, and more.

No TmHmm

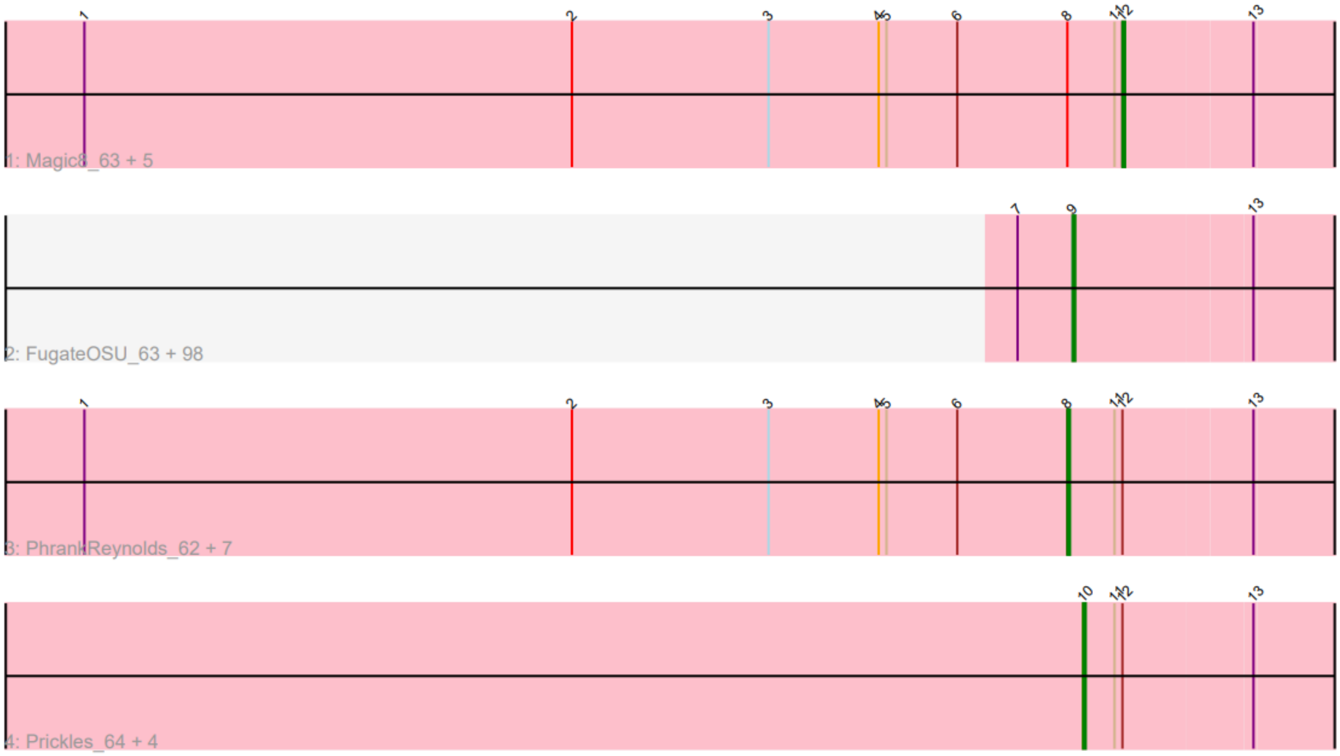
**CURATOR NAME: BELLA N**

**GENE NAME: EUGENIA-62**

**DNA MASTER NOTES: N/A**

**START POSITION EVALUATION (IN ORDER)**:

1. Starterator



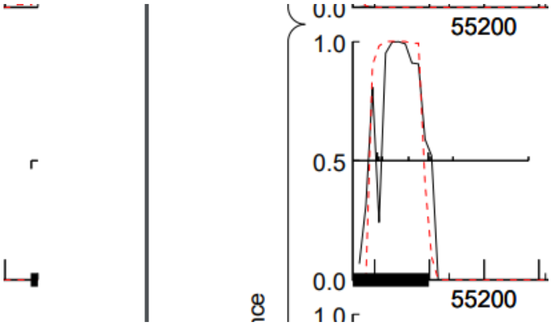
The start number called the most often in the published annotations is 9, it was called in 92 of the 118 non-draft genes in the pham.

Start 9:

- Found in 106 of 131 ( 80.9% ) of genes in pham
- Manual Annotations of this start: 92 of 118

- Called 93.4% of time when present

2. GeneMark coding potential



The black bar at the bottom represents the whole ORF it is about 150 nucleotides and this is a reverse gene.

3. Glimmer and GeneMark agreement

There is not agreement between Glimmer and GeneMark

Glimmer Start: 54969

Glimmer Score: 1.12

GeneMark Start: N/A

4. Longest open reading frame (ORF) without excessive gap

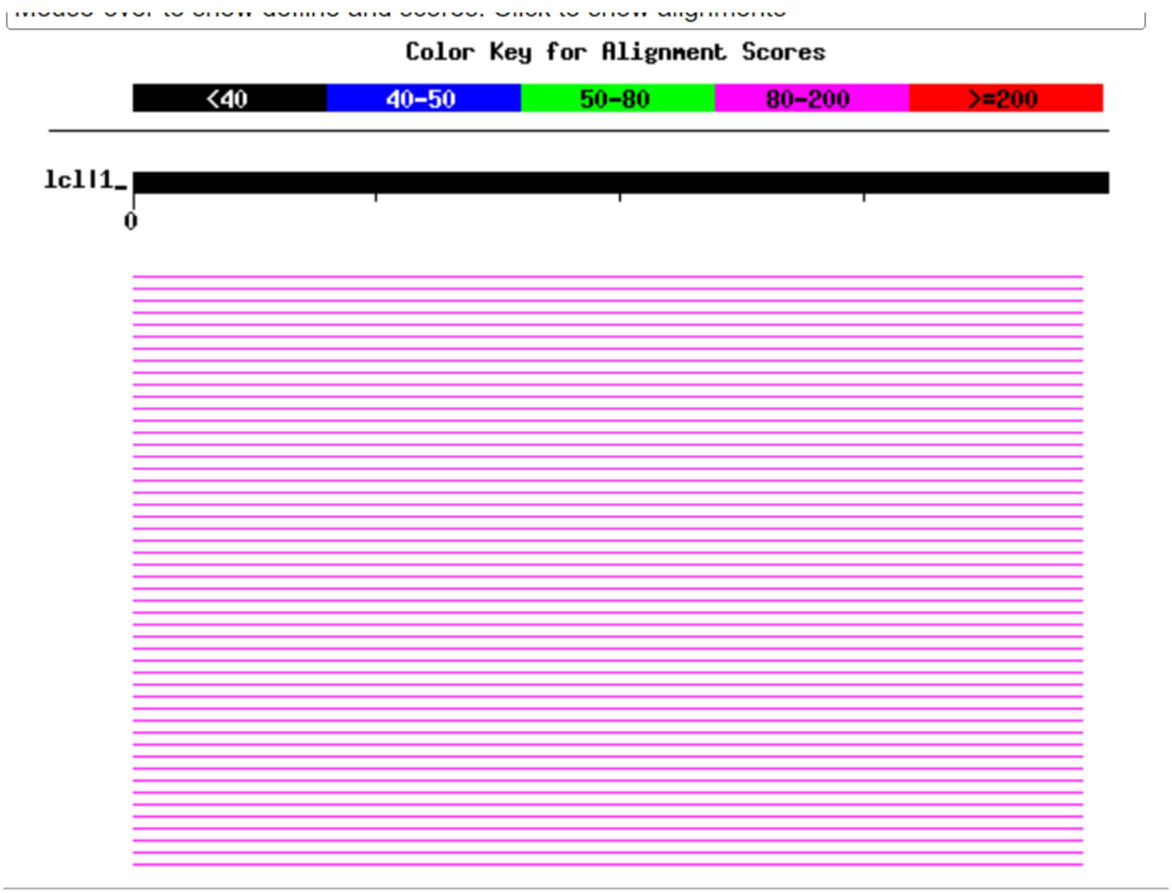
The selected gene is not the LORF with a length of 123 nucleotides. It has a Z-score of 1.895, a spacer of 5, and a final score of -6.2. This is a reverse gene.

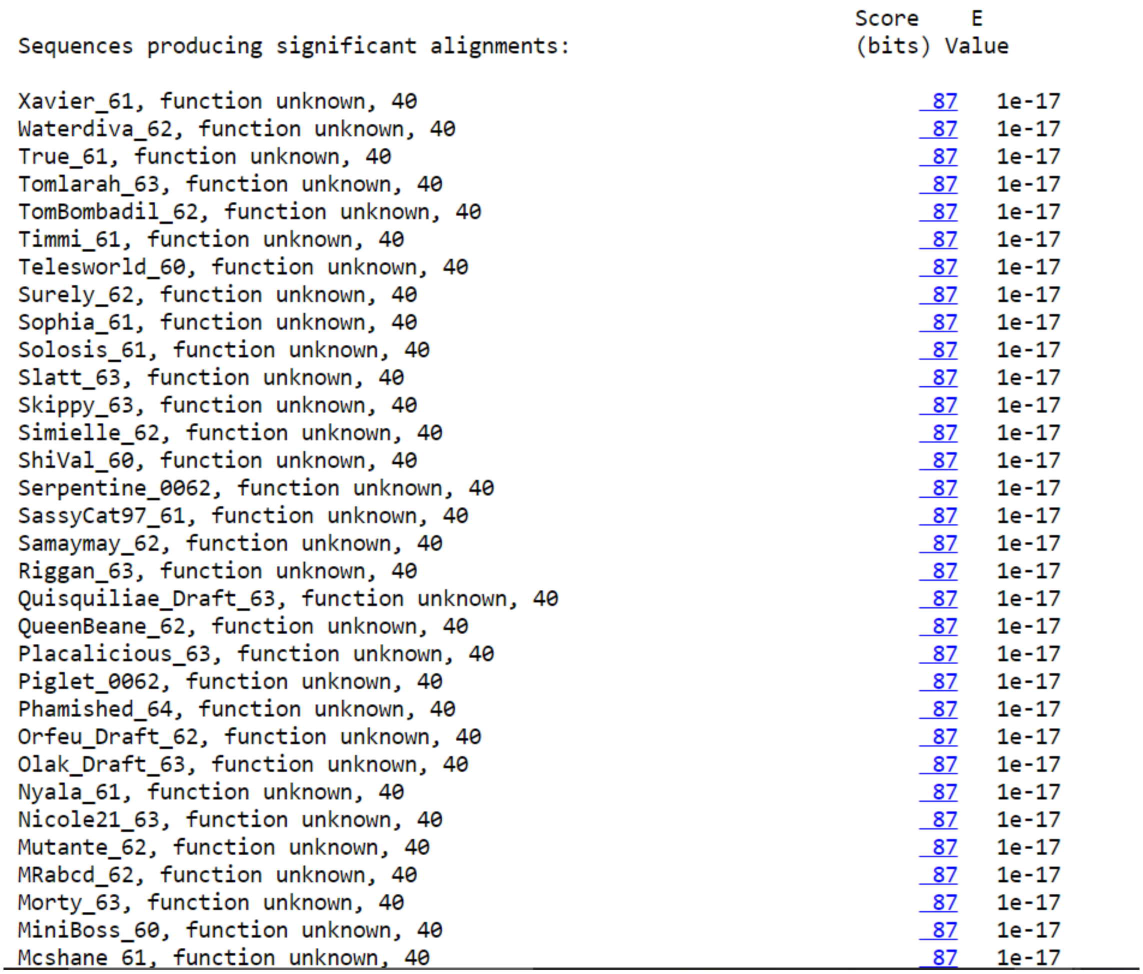
5. Function. If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

6. Supporting Information for Function (SIF)

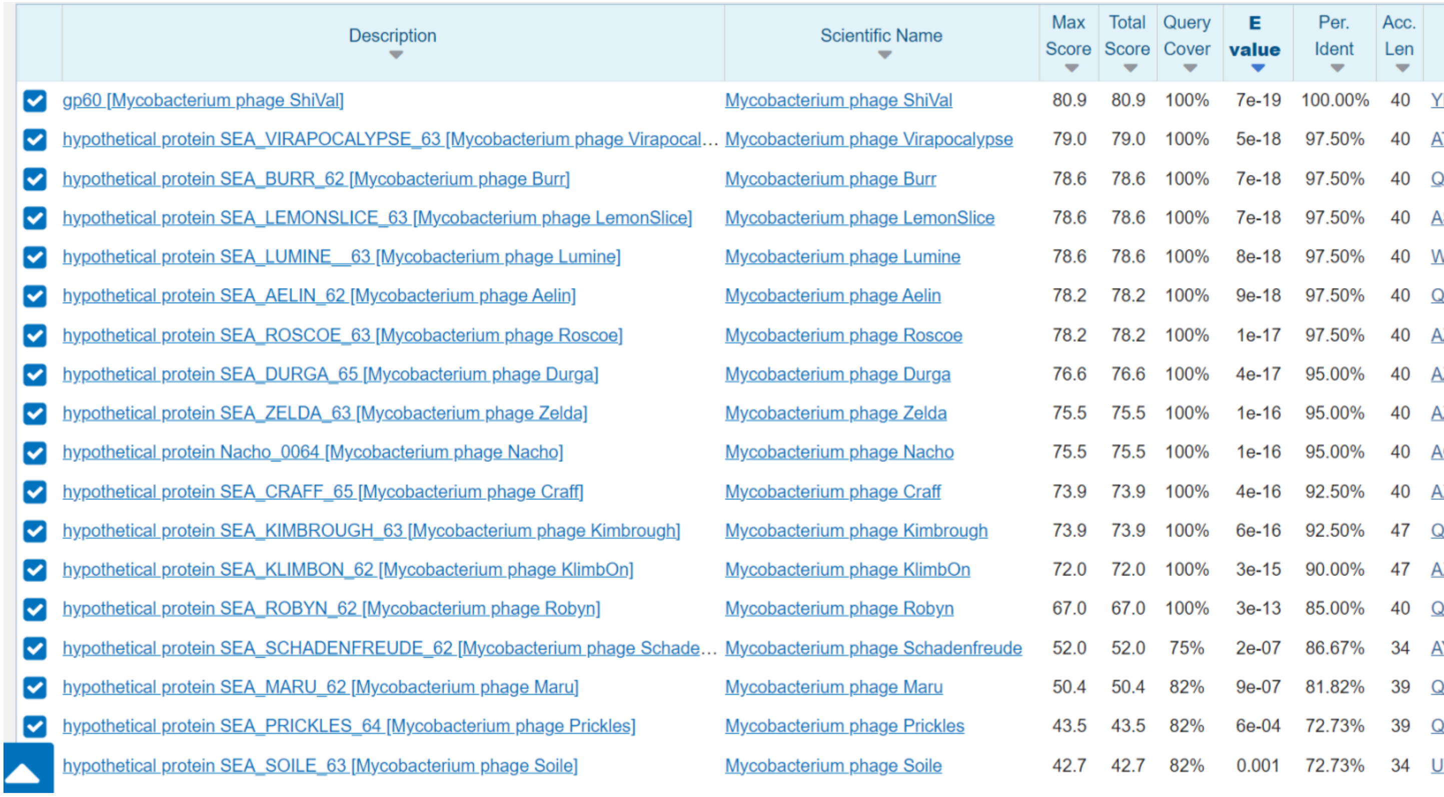
a. SIF: PhageDb BLAST





```
                                                                 Score    E
Sequences producing significant alignments:                      (bits) Value

Xavier_61, function unknown, 40                                     87    1e-17
Waterdiva_62, function unknown, 40                                  87    1e-17
True_61, function unknown, 40                                       87    1e-17
Tomlarah_63, function unknown, 40                                   87    1e-17
TomBombadil_62, function unknown, 40                                87    1e-17
Timmi_61, function unknown, 40                                      87    1e-17
Telesworld_60, function unknown, 40                                 87    1e-17
Surely_62, function unknown, 40                                     87    1e-17
Sophia_61, function unknown, 40                                     87    1e-17
Solosis_61, function unknown, 40                                    87    1e-17
Slatt_63, function unknown, 40                                      87    1e-17
Skippy_63, function unknown, 40                                     87    1e-17
Simielle_62, function unknown, 40                                   87    1e-17
ShiVal_60, function unknown, 40                                     87    1e-17
Serpentine_0062, function unknown, 40                               87    1e-17
SassyCat97_61, function unknown, 40                                 87    1e-17
Samaymay_62, function unknown, 40                                   87    1e-17
Riggan_63, function unknown, 40                                     87    1e-17
Quisquiliae_Draft_63, function unknown, 40                          87    1e-17
QueenBeane_62, function unknown, 40                                 87    1e-17
Placalicious_63, function unknown, 40                               87    1e-17
Piglet_0062, function unknown, 40                                   87    1e-17
Phamished_64, function unknown, 40                                  87    1e-17
Orfeu_Draft_62, function unknown, 40                                87    1e-17
Olak_Draft_63, function unknown, 40                                 87    1e-17
Nyala_61, function unknown, 40                                      87    1e-17
Nicole21_63, function unknown, 40                                   87    1e-17
Mutante_62, function unknown, 40                                    87    1e-17
MRabcd_62, function unknown, 40                                     87    1e-17
Morty_63, function unknown, 40                                      87    1e-17
MiniBoss_60, function unknown, 40                                   87    1e-17
Mcshane 61, function unknown, 40                                    87    1e-17
```

b. SIF: NCBI BLAST

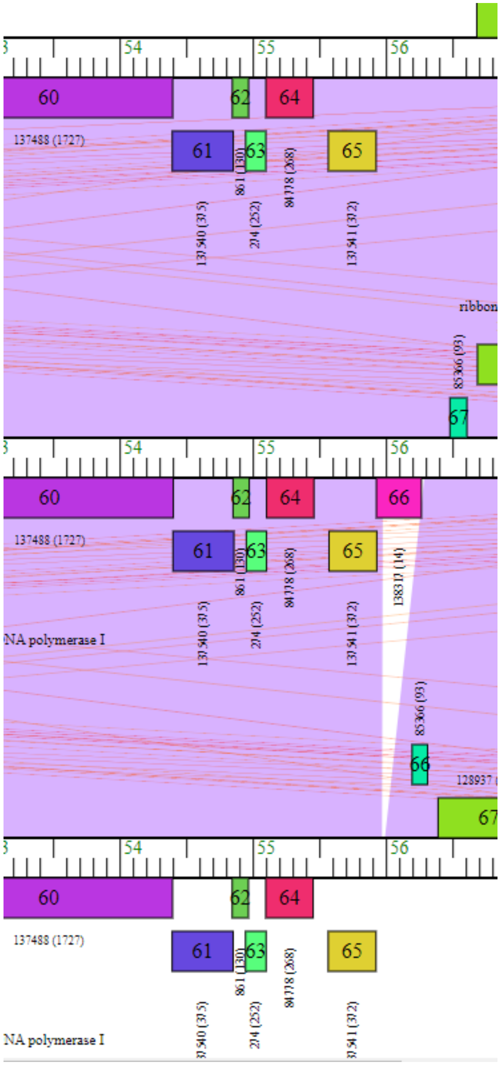


| | Description | Scientific Name | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | gp60 [Mycobacterium phage ShiVal] | Mycobacterium phage ShiVal | 80.9 | 80.9 | 100% | 7e-19 | 100.00% | 40 | Y |
| ☑ | hypothetical protein SEA_VIRAPOCALYPSE_63 [Mycobacterium phage Virapocal… | Mycobacterium phage Virapocalypse | 79.0 | 79.0 | 100% | 5e-18 | 97.50% | 40 | A |
| ☑ | hypothetical protein SEA_BURR_62 [Mycobacterium phage Burr] | Mycobacterium phage Burr | 78.6 | 78.6 | 100% | 7e-18 | 97.50% | 40 | Q |
| ☑ | hypothetical protein SEA_LEMONSLICE_63 [Mycobacterium phage LemonSlice] | Mycobacterium phage LemonSlice | 78.6 | 78.6 | 100% | 7e-18 | 97.50% | 40 | A |
| ☑ | hypothetical protein SEA_LUMINE__63 [Mycobacterium phage Lumine] | Mycobacterium phage Lumine | 78.6 | 78.6 | 100% | 8e-18 | 97.50% | 40 | W |
| ☑ | hypothetical protein SEA_AELIN_62 [Mycobacterium phage Aelin] | Mycobacterium phage Aelin | 78.2 | 78.2 | 100% | 9e-18 | 97.50% | 40 | Q |
| ☑ | hypothetical protein SEA_ROSCOE_63 [Mycobacterium phage Roscoe] | Mycobacterium phage Roscoe | 78.2 | 78.2 | 100% | 1e-17 | 97.50% | 40 | A |
| ☑ | hypothetical protein SEA_DURGA_65 [Mycobacterium phage Durga] | Mycobacterium phage Durga | 76.6 | 76.6 | 100% | 4e-17 | 95.00% | 40 | A |
| ☑ | hypothetical protein SEA_ZELDA_63 [Mycobacterium phage Zelda] | Mycobacterium phage Zelda | 75.5 | 75.5 | 100% | 1e-16 | 95.00% | 40 | A |
| ☑ | hypothetical protein Nacho_0064 [Mycobacterium phage Nacho] | Mycobacterium phage Nacho | 75.5 | 75.5 | 100% | 1e-16 | 95.00% | 40 | A |
| ☑ | hypothetical protein SEA_CRAFF_65 [Mycobacterium phage Craff] | Mycobacterium phage Craff | 73.9 | 73.9 | 100% | 4e-16 | 92.50% | 40 | A |
| ☑ | hypothetical protein SEA_KIMBROUGH_63 [Mycobacterium phage Kimbrough] | Mycobacterium phage Kimbrough | 73.9 | 73.9 | 100% | 6e-16 | 92.50% | 47 | Q |
| ☑ | hypothetical protein SEA_KLIMBON_62 [Mycobacterium phage KlimbOn] | Mycobacterium phage KlimbOn | 72.0 | 72.0 | 100% | 3e-15 | 90.00% | 47 | A |
| ☑ | hypothetical protein SEA_ROBYN_62 [Mycobacterium phage Robyn] | Mycobacterium phage Robyn | 67.0 | 67.0 | 100% | 3e-13 | 85.00% | 40 | Q |
| ☑ | hypothetical protein SEA_SCHADENFREUDE_62 [Mycobacterium phage Schade… | Mycobacterium phage Schadenfreude | 52.0 | 52.0 | 75% | 2e-07 | 86.67% | 34 | A |
| ☑ | hypothetical protein SEA_MARU_62 [Mycobacterium phage Maru] | Mycobacterium phage Maru | 50.4 | 50.4 | 82% | 9e-07 | 81.82% | 39 | Q |
| ☑ | hypothetical protein SEA_PRICKLES_64 [Mycobacterium phage Prickles] | Mycobacterium phage Prickles | 43.5 | 43.5 | 82% | 6e-04 | 72.73% | 39 | Q |
| | hypothetical protein SEA_SOILE_63 [Mycobacterium phage Soile] | Mycobacterium phage Soile | 42.7 | 42.7 | 82% | 0.001 | 72.73% | 34 | U |

c. SIF: HHPred

No HHPRED evidence selected

d. SIF: Synteny-Phamerator (three genomes)



7. Any other important information.

Synteny with Altwerkus, Anderson, BlueHusk, and more.

No TmHmm

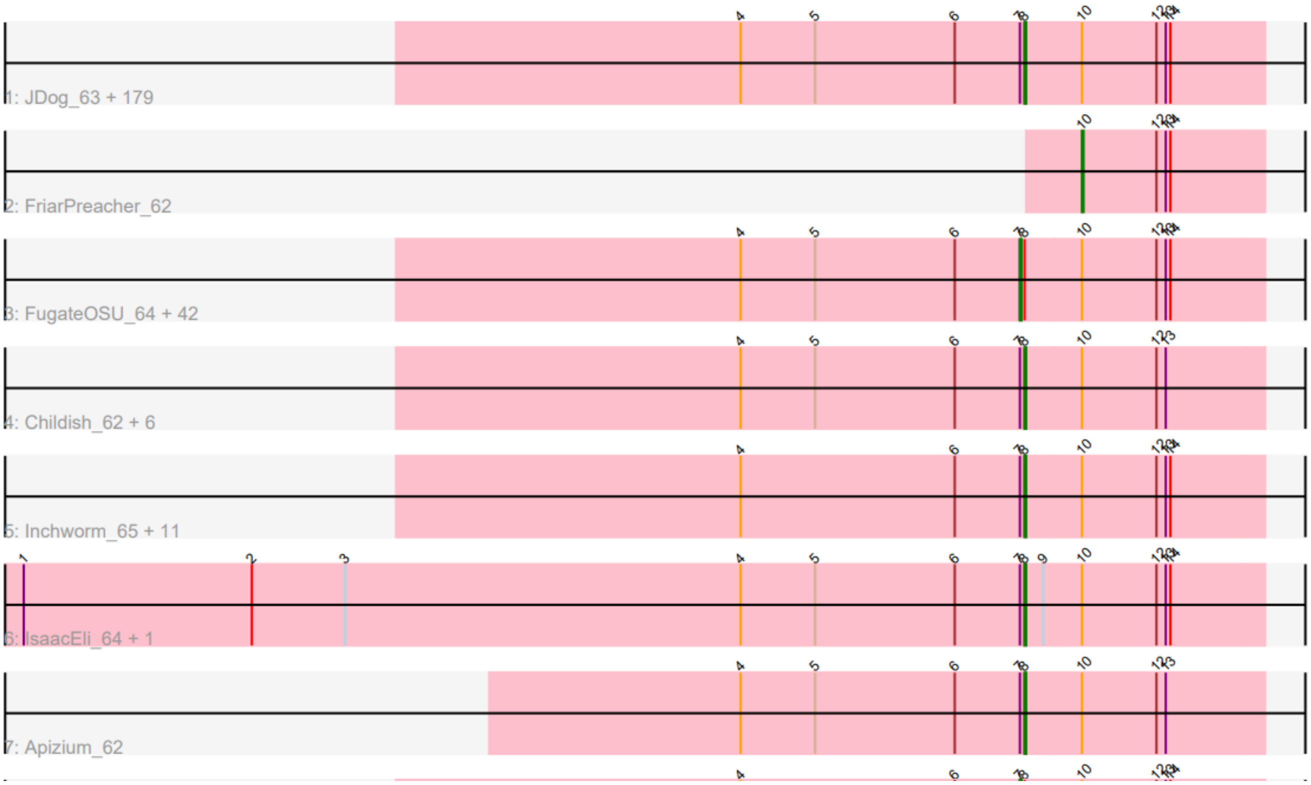
## CURATOR NAME: BELLA N

## GENE NAME: EUGENIA-63

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

1. Starterator



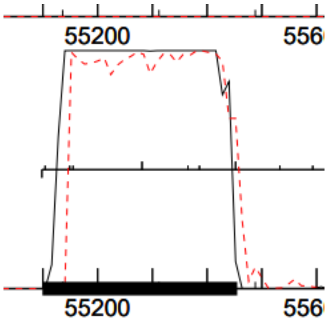
The start number called the most often in the published annotations is 8, it was called in 181 of the 230 non-draft genes in the pham.

Start 8:

- Found in 253 of 254 ( 99.6% ) of genes in pham

- Manual Annotations of this start: 181 of 230
- Called 80.6% of time when present

2. GeneMark coding potential



The black bar at the bottom represents the whole ORF it is about 300 nucleotides and this is a reverse gene.

3. Glimmer and GeneMark agreement

There is not agreement between Glimmer and GeneMark

Glimmer Start: N/A

Glimmer Score: N/A

GeneMark Start: 55099

4. Longest open reading frame (ORF) without excessive gap

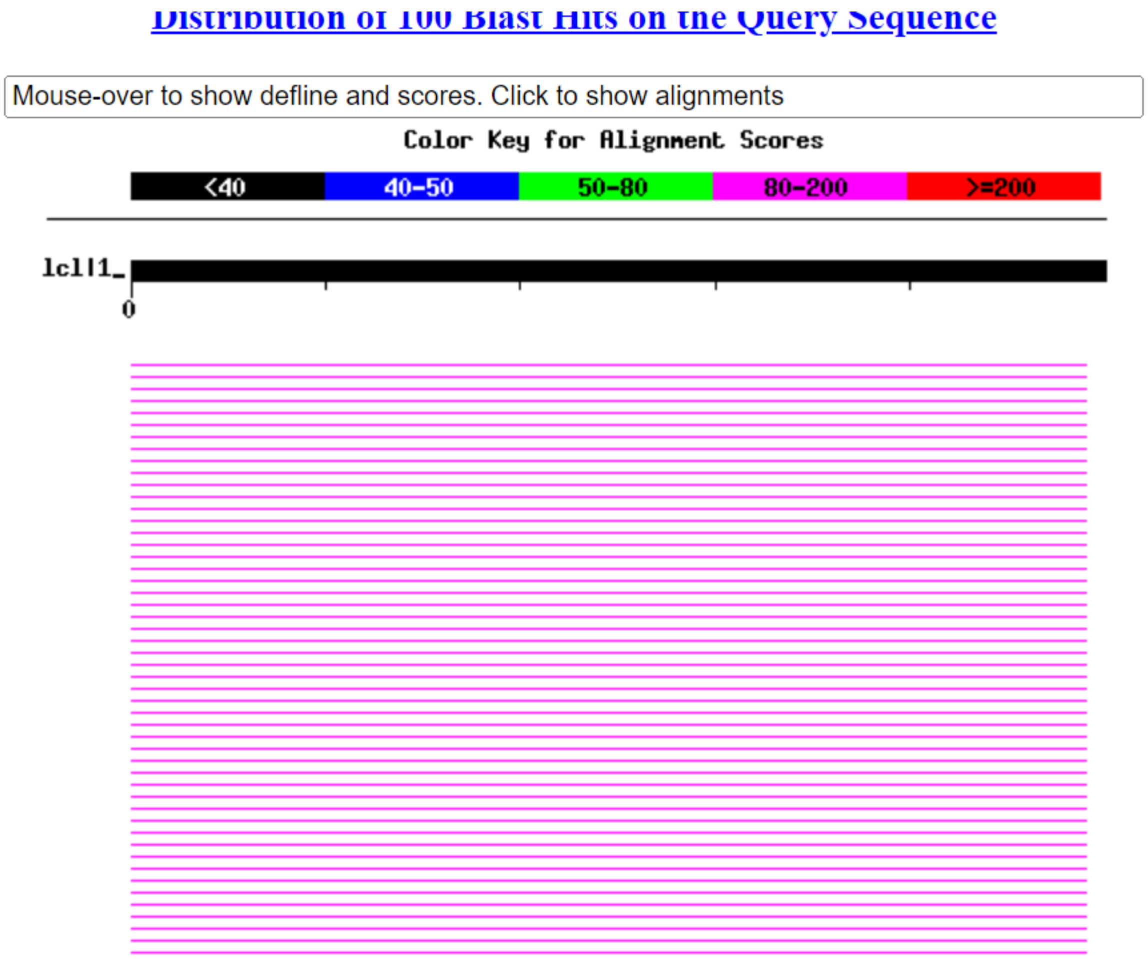
The selected gene is not the LORF with a length of 153 nucleotides. It has a Z-score of 2.554, a spacer of 9, and a final score of -3.751. This is a reverse gene.

5. Function. If no functional prediction is present, write “Hypothetical protein”.

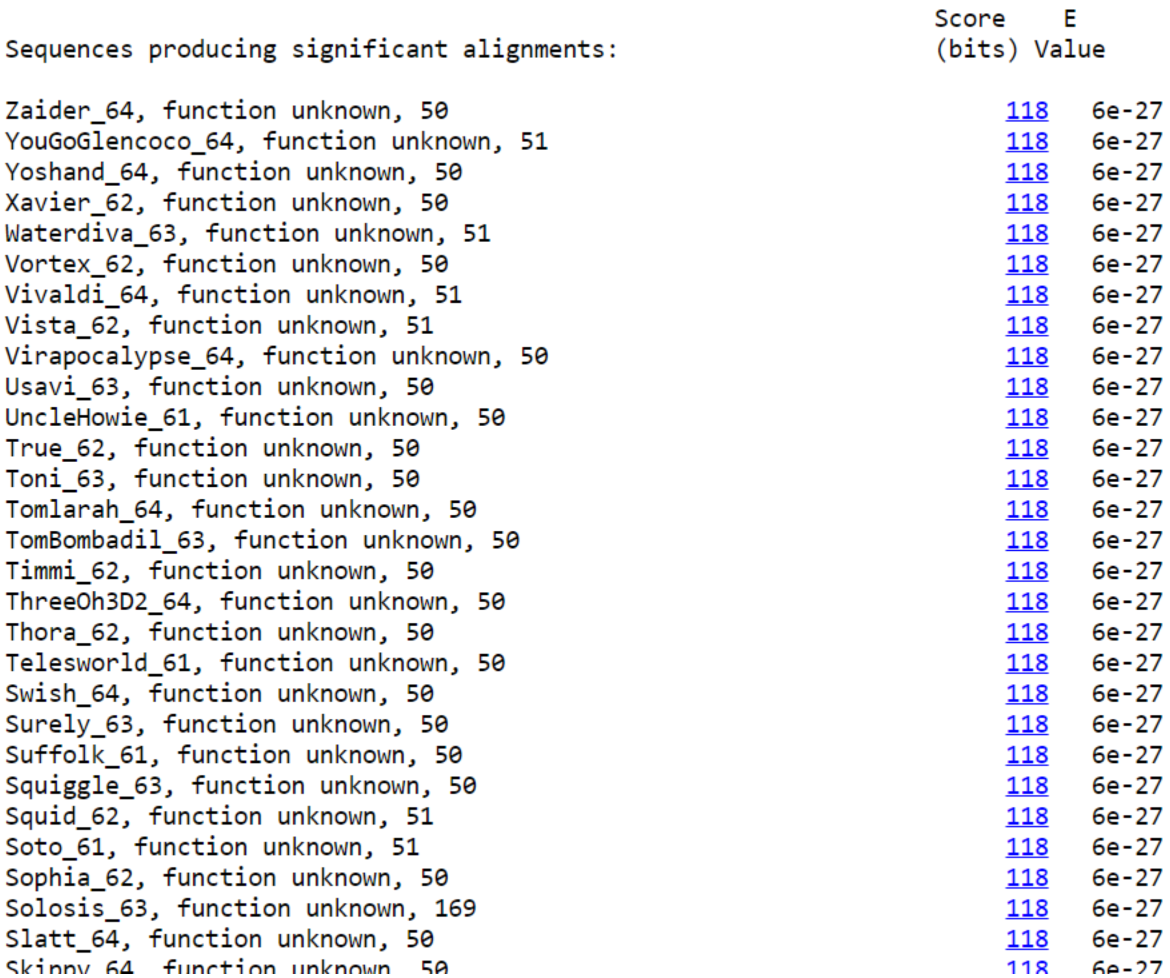
Hypothetical protein

## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST

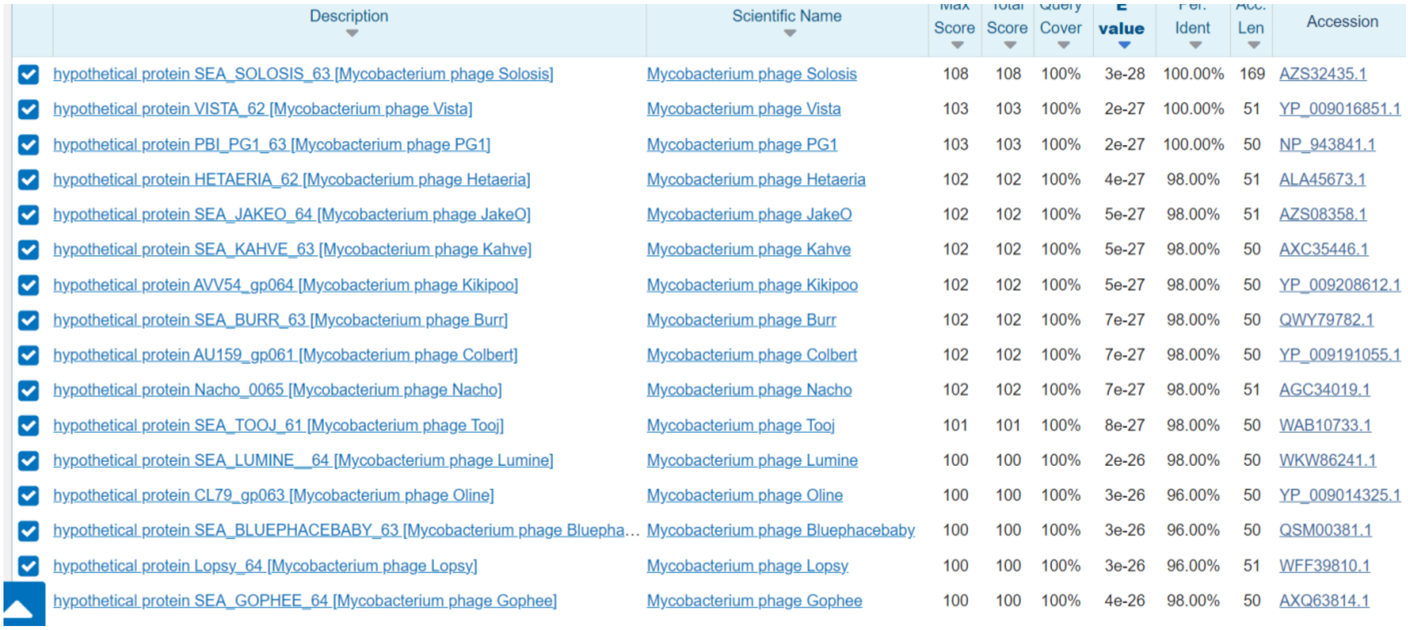

Distribution of 100 Blast Hits on the Query Sequence

Mouse-over to show defline and scores. Click to show alignments



```
                                                              Score     E
Sequences producing significant alignments:                   (bits)  Value

Zaider_64, function unknown, 50                                 118    6e-27
YouGoGlencoco_64, function unknown, 51                          118    6e-27
Yoshand_64, function unknown, 50                                118    6e-27
Xavier_62, function unknown, 50                                 118    6e-27
Waterdiva_63, function unknown, 51                              118    6e-27
Vortex_62, function unknown, 50                                 118    6e-27
Vivaldi_64, function unknown, 51                                118    6e-27
Vista_62, function unknown, 51                                  118    6e-27
Virapocalypse_64, function unknown, 50                          118    6e-27
Usavi_63, function unknown, 50                                  118    6e-27
UncleHowie_61, function unknown, 50                             118    6e-27
True_62, function unknown, 50                                   118    6e-27
Toni_63, function unknown, 50                                   118    6e-27
Tomlarah_64, function unknown, 50                               118    6e-27
TomBombadil_63, function unknown, 50                            118    6e-27
Timmi_62, function unknown, 50                                  118    6e-27
ThreeOh3D2_64, function unknown, 50                             118    6e-27
Thora_62, function unknown, 50                                  118    6e-27
Telesworld_61, function unknown, 50                             118    6e-27
Swish_64, function unknown, 50                                  118    6e-27
Surely_63, function unknown, 50                                 118    6e-27
Suffolk_61, function unknown, 50                                118    6e-27
Squiggle_63, function unknown, 50                               118    6e-27
Squid_62, function unknown, 51                                  118    6e-27
Soto_61, function unknown, 51                                   118    6e-27
Sophia_62, function unknown, 50                                 118    6e-27
Solosis_63, function unknown, 169                               118    6e-27
Slatt_64, function unknown, 50                                  118    6e-27
Skinny_64, function unknown, 50                                 118    6e-27
```

b. SIF: NCBI BLAST

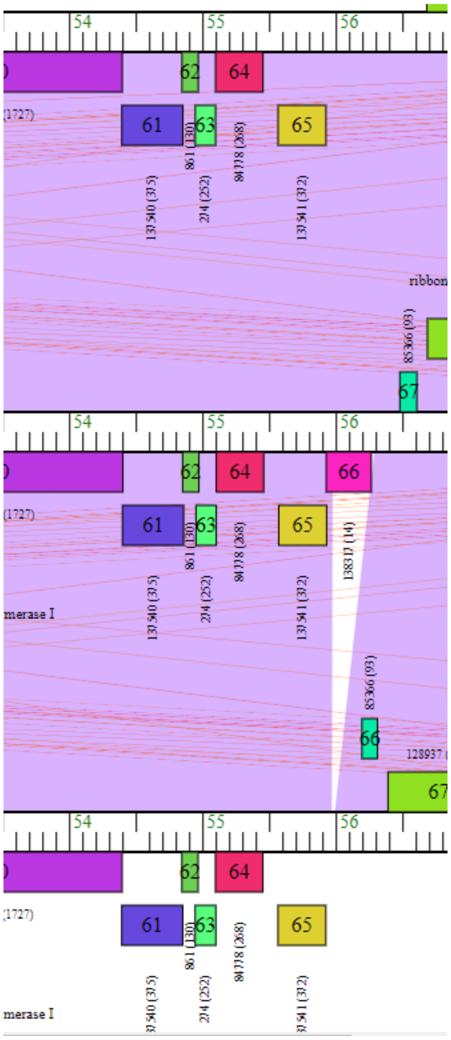


| | Description | Scientific Name | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | hypothetical protein SEA_SOLOSIS_63 [Mycobacterium phage Solosis] | Mycobacterium phage Solosis | 108 | 108 | 100% | 3e-28 | 100.00% | 169 | AZS32435.1 |
| ☑ | hypothetical protein VISTA_62 [Mycobacterium phage Vista] | Mycobacterium phage Vista | 103 | 103 | 100% | 2e-27 | 100.00% | 51 | YP_009016851.1 |
| ☑ | hypothetical protein PBI_PG1_63 [Mycobacterium phage PG1] | Mycobacterium phage PG1 | 103 | 103 | 100% | 2e-27 | 100.00% | 50 | NP_943841.1 |
| ☑ | hypothetical protein HETAERIA_62 [Mycobacterium phage Hetaeria] | Mycobacterium phage Hetaeria | 102 | 102 | 100% | 4e-27 | 98.00% | 51 | ALA45673.1 |
| ☑ | hypothetical protein SEA_JAKEO_64 [Mycobacterium phage JakeO] | Mycobacterium phage JakeO | 102 | 102 | 100% | 5e-27 | 98.00% | 51 | AZS08358.1 |
| ☑ | hypothetical protein SEA_KAHVE_63 [Mycobacterium phage Kahve] | Mycobacterium phage Kahve | 102 | 102 | 100% | 5e-27 | 98.00% | 50 | AXC35446.1 |
| ☑ | hypothetical protein AVV54_gp064 [Mycobacterium phage Kikipoo] | Mycobacterium phage Kikipoo | 102 | 102 | 100% | 5e-27 | 98.00% | 50 | YP_009208612.1 |
| ☑ | hypothetical protein SEA_BURR_63 [Mycobacterium phage Burr] | Mycobacterium phage Burr | 102 | 102 | 100% | 7e-27 | 98.00% | 50 | QWY79782.1 |
| ☑ | hypothetical protein AU159_gp061 [Mycobacterium phage Colbert] | Mycobacterium phage Colbert | 102 | 102 | 100% | 7e-27 | 98.00% | 50 | YP_009191055.1 |
| ☑ | hypothetical protein Nacho_0065 [Mycobacterium phage Nacho] | Mycobacterium phage Nacho | 102 | 102 | 100% | 7e-27 | 98.00% | 51 | AGC34019.1 |
| ☑ | hypothetical protein SEA_TOOJ_61 [Mycobacterium phage Tooj] | Mycobacterium phage Tooj | 101 | 101 | 100% | 8e-27 | 98.00% | 50 | WAB10733.1 |
| ☑ | hypothetical protein SEA_LUMINE__64 [Mycobacterium phage Lumine] | Mycobacterium phage Lumine | 100 | 100 | 100% | 2e-26 | 98.00% | 50 | WKW86241.1 |
| ☑ | hypothetical protein CL79_gp063 [Mycobacterium phage Oline] | Mycobacterium phage Oline | 100 | 100 | 100% | 3e-26 | 96.00% | 50 | YP_009014325.1 |
| ☑ | hypothetical protein SEA_BLUEPHACEBABY_63 [Mycobacterium phage Bluepha… | Mycobacterium phage Bluephacebaby | 100 | 100 | 100% | 3e-26 | 96.00% | 50 | QSM00381.1 |
| ☑ | hypothetical protein Lopsy_64 [Mycobacterium phage Lopsy] | Mycobacterium phage Lopsy | 100 | 100 | 100% | 3e-26 | 96.00% | 51 | WFF39810.1 |
| | hypothetical protein SEA_GOPHEE_64 [Mycobacterium phage Gophee] | Mycobacterium phage Gophee | 100 | 100 | 100% | 4e-26 | 98.00% | 50 | AXQ63814.1 |

c. SIF: HHPred

No HHPRED evidence selected

d. SIF: Synteny-Phamerator (three genomes)



7. Any other important information.

Synteny with Kloppinator, KLucky39, Kwadwo, and more.

No TmHmm

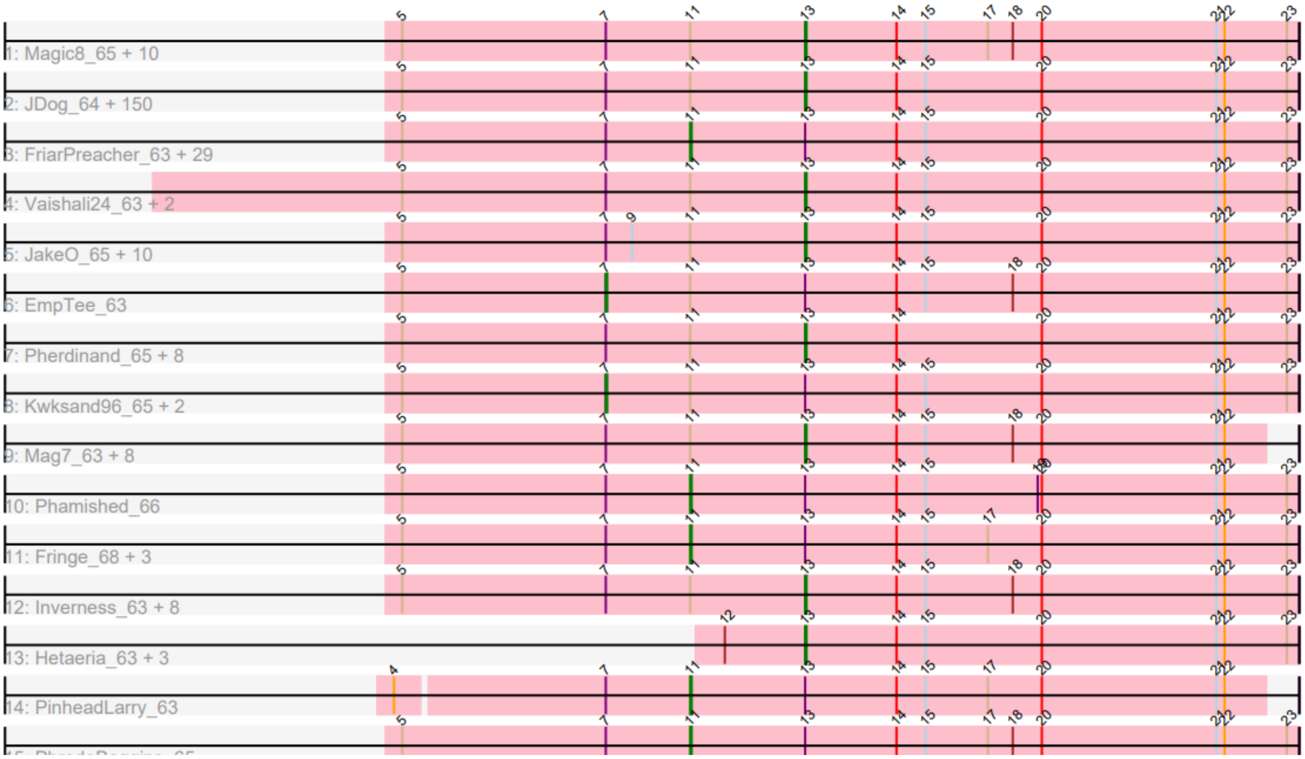
**CURATOR NAME: BELLA N**

**GENE NAME: EUGENIA-64**

**DNA MASTER NOTES: N/A**

**START POSITION EVALUATION (IN ORDER)**:

1. Starterator



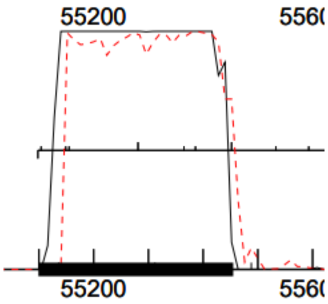
The start number called the most often in the published annotations is 13, it was called in 193 of the 245 non-draft genes in the pham.

Start 13:

- Found in 270 of 270 ( 100.0% ) of genes in pham

- Manual Annotations of this start: 193 of 245
- Called 80.4% of time when present

2. GeneMark coding potential



The black bar at the bottom represents the whole ORF it is about 500 nucleotides and this is a reverse gene.

3. Glimmer and GeneMark agreement

There is agreement between Glimmer and GeneMark

Glimmer Start: 55455

Glimmer Score: 13.47

GeneMark Start: 55455

4. Longest open reading frame (ORF) without excessive gap

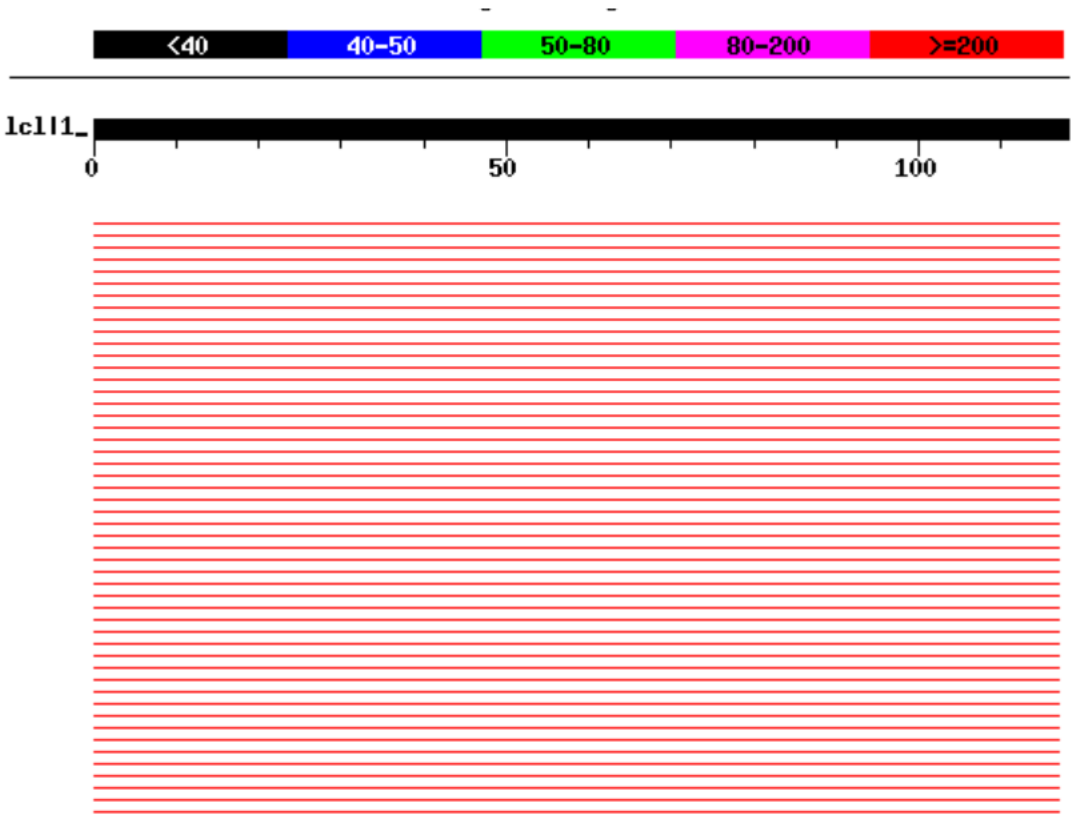
The selected gene is not the LORF with a length of 357 nucleotides. It has a Z-score of 2.634, a spacer of 14, and a final score of -4.161. This is a reverse gene.

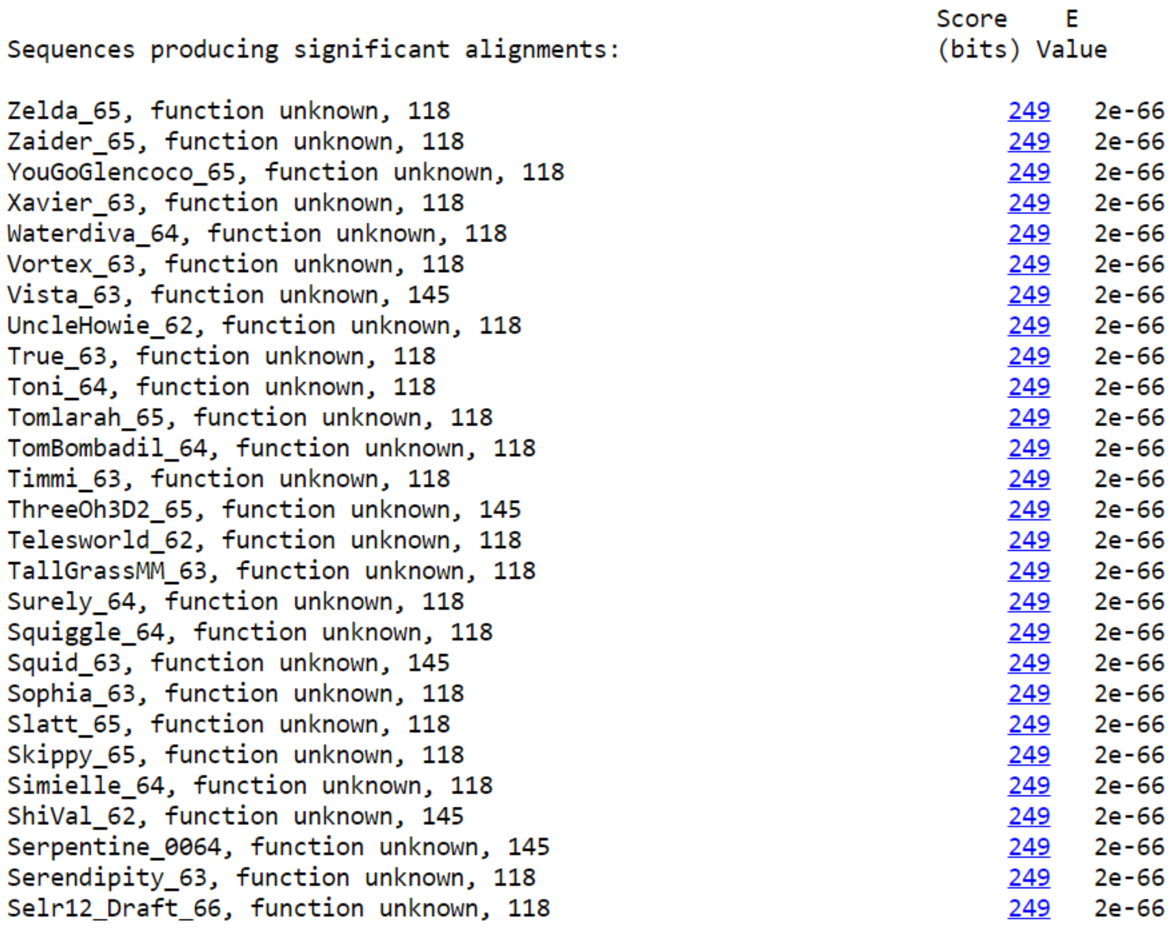
5. Function. If no functional prediction is present, write “Hypothetical protein”.

Membrane protein

## 6. Supporting Information for Function (SIF)

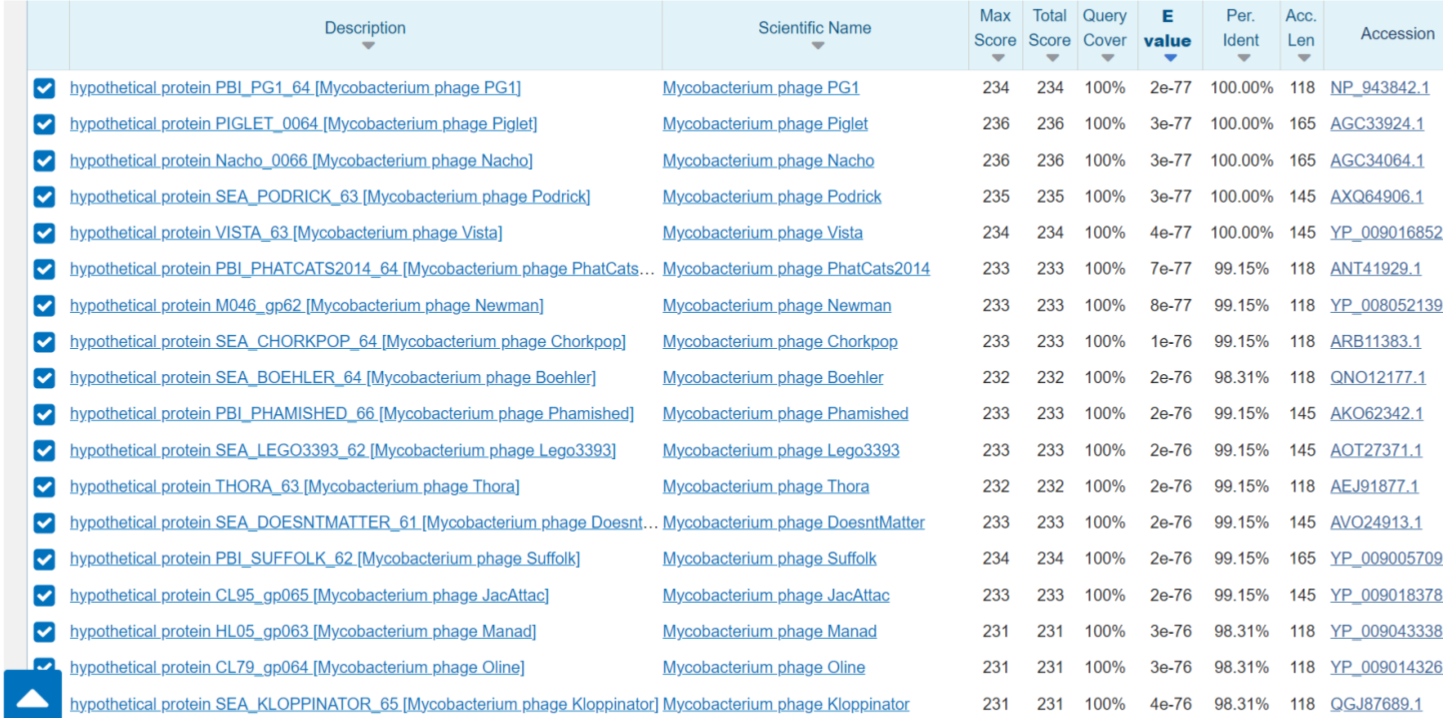
### a. SIF: PhageDb BLAST





```
                                                                Score     E
Sequences producing significant alignments:                    (bits)  Value

Zelda_65, function unknown, 118                                    249    2e-66
Zaider_65, function unknown, 118                                   249    2e-66
YouGoGlencoco_65, function unknown, 118                            249    2e-66
Xavier_63, function unknown, 118                                   249    2e-66
Waterdiva_64, function unknown, 118                                249    2e-66
Vortex_63, function unknown, 118                                   249    2e-66
Vista_63, function unknown, 145                                    249    2e-66
UncleHowie_62, function unknown, 118                               249    2e-66
True_63, function unknown, 118                                     249    2e-66
Toni_64, function unknown, 118                                     249    2e-66
Tomlarah_65, function unknown, 118                                 249    2e-66
TomBombadil_64, function unknown, 118                              249    2e-66
Timmi_63, function unknown, 118                                    249    2e-66
ThreeOh3D2_65, function unknown, 145                               249    2e-66
Telesworld_62, function unknown, 118                               249    2e-66
TallGrassMM_63, function unknown, 118                              249    2e-66
Surely_64, function unknown, 118                                   249    2e-66
Squiggle_64, function unknown, 118                                 249    2e-66
Squid_63, function unknown, 145                                    249    2e-66
Sophia_63, function unknown, 118                                   249    2e-66
Slatt_65, function unknown, 118                                    249    2e-66
Skippy_65, function unknown, 118                                   249    2e-66
Simielle_64, function unknown, 118                                 249    2e-66
ShiVal_62, function unknown, 145                                   249    2e-66
Serpentine_0064, function unknown, 145                             249    2e-66
Serendipity_63, function unknown, 118                              249    2e-66
Selr12_Draft_66, function unknown, 118                             249    2e-66
```

b. SIF: NCBI BLAST

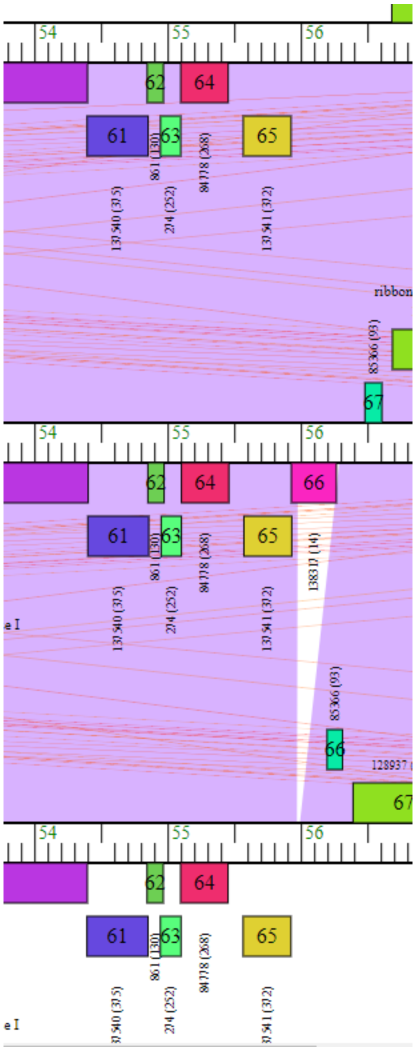


| | Description | Scientific Name | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | hypothetical protein PBI_PG1_64 [Mycobacterium phage PG1] | Mycobacterium phage PG1 | 234 | 234 | 100% | 2e-77 | 100.00% | 118 | NP_943842.1 |
| ☑ | hypothetical protein PIGLET_0064 [Mycobacterium phage Piglet] | Mycobacterium phage Piglet | 236 | 236 | 100% | 3e-77 | 100.00% | 165 | AGC33924.1 |
| ☑ | hypothetical protein Nacho_0066 [Mycobacterium phage Nacho] | Mycobacterium phage Nacho | 236 | 236 | 100% | 3e-77 | 100.00% | 165 | AGC34064.1 |
| ☑ | hypothetical protein SEA_PODRICK_63 [Mycobacterium phage Podrick] | Mycobacterium phage Podrick | 235 | 235 | 100% | 3e-77 | 100.00% | 145 | AXQ64906.1 |
| ☑ | hypothetical protein VISTA_63 [Mycobacterium phage Vista] | Mycobacterium phage Vista | 234 | 234 | 100% | 4e-77 | 100.00% | 145 | YP_009016852 |
| ☑ | hypothetical protein PBI_PHATCATS2014_64 [Mycobacterium phage PhatCats… | Mycobacterium phage PhatCats2014 | 233 | 233 | 100% | 7e-77 | 99.15% | 118 | ANT41929.1 |
| ☑ | hypothetical protein M046_gp62 [Mycobacterium phage Newman] | Mycobacterium phage Newman | 233 | 233 | 100% | 8e-77 | 99.15% | 118 | YP_008052139 |
| ☑ | hypothetical protein SEA_CHORKPOP_64 [Mycobacterium phage Chorkpop] | Mycobacterium phage Chorkpop | 233 | 233 | 100% | 1e-76 | 99.15% | 118 | ARB11383.1 |
| ☑ | hypothetical protein SEA_BOEHLER_64 [Mycobacterium phage Boehler] | Mycobacterium phage Boehler | 232 | 232 | 100% | 2e-76 | 98.31% | 118 | QNO12177.1 |
| ☑ | hypothetical protein PBI_PHAMISHED_66 [Mycobacterium phage Phamished] | Mycobacterium phage Phamished | 233 | 233 | 100% | 2e-76 | 99.15% | 145 | AKO62342.1 |
| ☑ | hypothetical protein SEA_LEGO3393_62 [Mycobacterium phage Lego3393] | Mycobacterium phage Lego3393 | 233 | 233 | 100% | 2e-76 | 99.15% | 145 | AOT27371.1 |
| ☑ | hypothetical protein THORA_63 [Mycobacterium phage Thora] | Mycobacterium phage Thora | 232 | 232 | 100% | 2e-76 | 99.15% | 118 | AEJ91877.1 |
| ☑ | hypothetical protein SEA_DOESNTMATTER_61 [Mycobacterium phage Doesnt… | Mycobacterium phage DoesntMatter | 233 | 233 | 100% | 2e-76 | 99.15% | 145 | AVO24913.1 |
| ☑ | hypothetical protein PBI_SUFFOLK_62 [Mycobacterium phage Suffolk] | Mycobacterium phage Suffolk | 234 | 234 | 100% | 2e-76 | 99.15% | 165 | YP_009005709 |
| ☑ | hypothetical protein CL95_gp065 [Mycobacterium phage JacAttac] | Mycobacterium phage JacAttac | 233 | 233 | 100% | 2e-76 | 99.15% | 145 | YP_009018378 |
| ☑ | hypothetical protein HL05_gp063 [Mycobacterium phage Manad] | Mycobacterium phage Manad | 231 | 231 | 100% | 3e-76 | 98.31% | 118 | YP_009043338 |
| ☑ | hypothetical protein CL79_gp064 [Mycobacterium phage Oline] | Mycobacterium phage Oline | 231 | 231 | 100% | 3e-76 | 98.31% | 118 | YP_009014326 |
| | hypothetical protein SEA_KLOPPINATOR_65 [Mycobacterium phage Kloppinator] | Mycobacterium phage Kloppinator | 231 | 231 | 100% | 4e-76 | 98.31% | 118 | QGJ87689.1 |

c. SIF: HHPred

No HHPRED evidence selected

d. SIF: Synteny-Phamerator (three genomes)



7. Any other important information.

Synteny with Emiris, Fang, Freya, and more.

1 suspected TmHmm

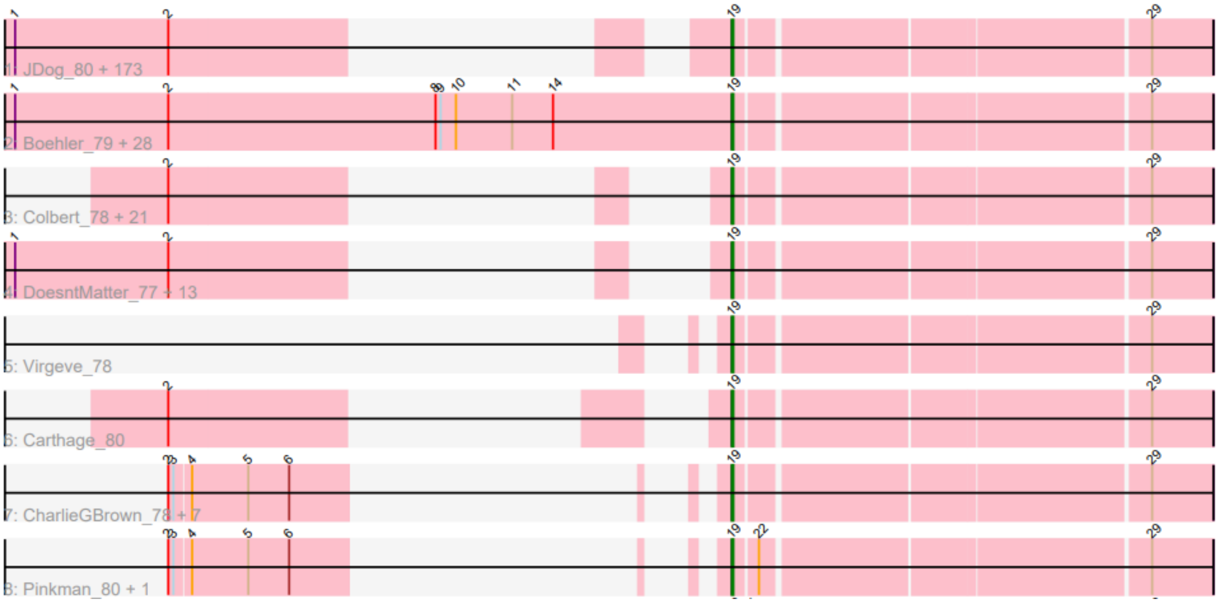
## CURATOR NAME: JOHN MOTTER

## GENE NAME: EUGENIA GENE 82

## DNA MASTER NOTES: N/A

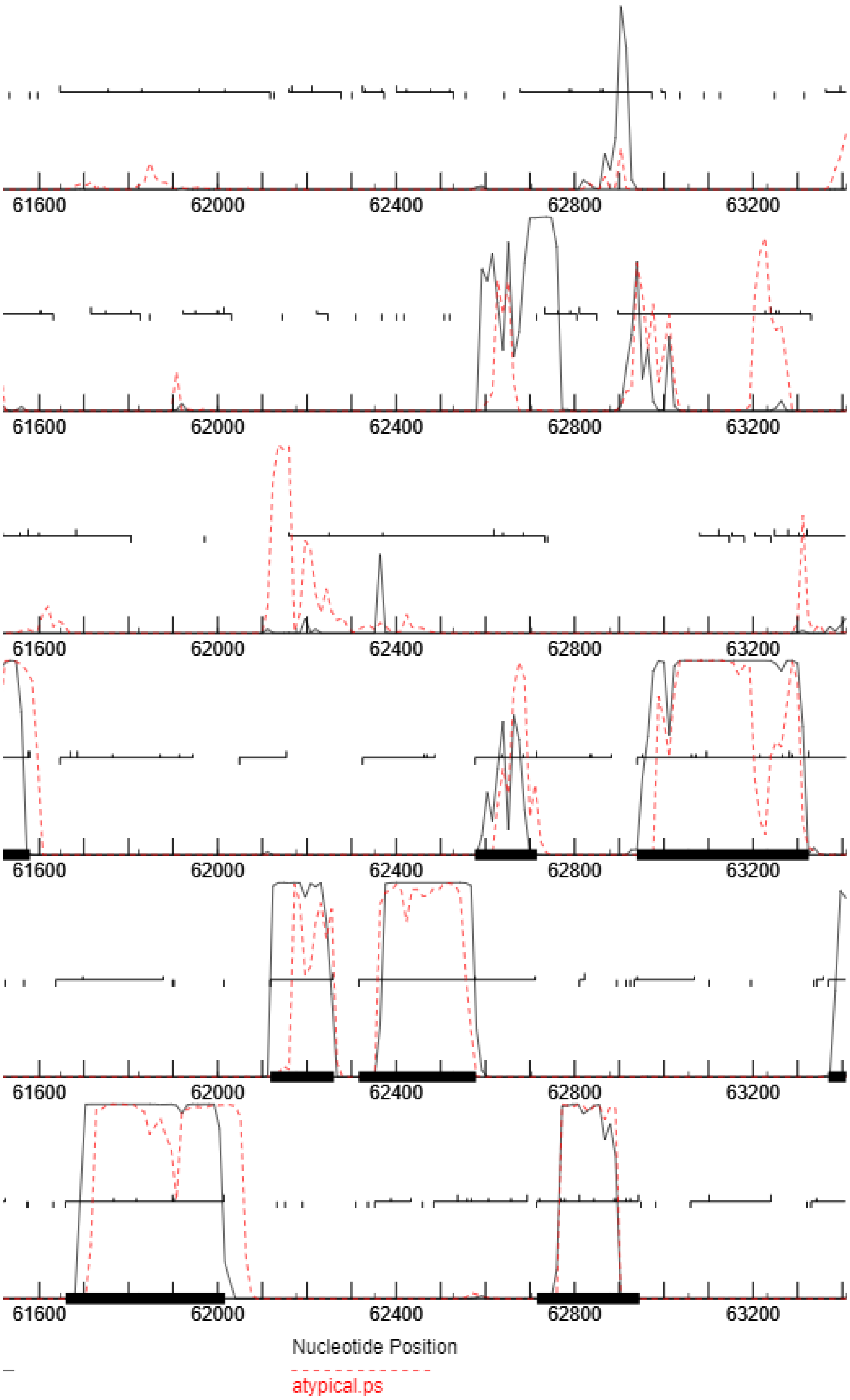
## START POSITION EVALUATION (IN ORDER):

### 1. Starterator



The start number called the most often in the published annotations is 19, it was called in 239 of the 244 non-draft genes in the pham.

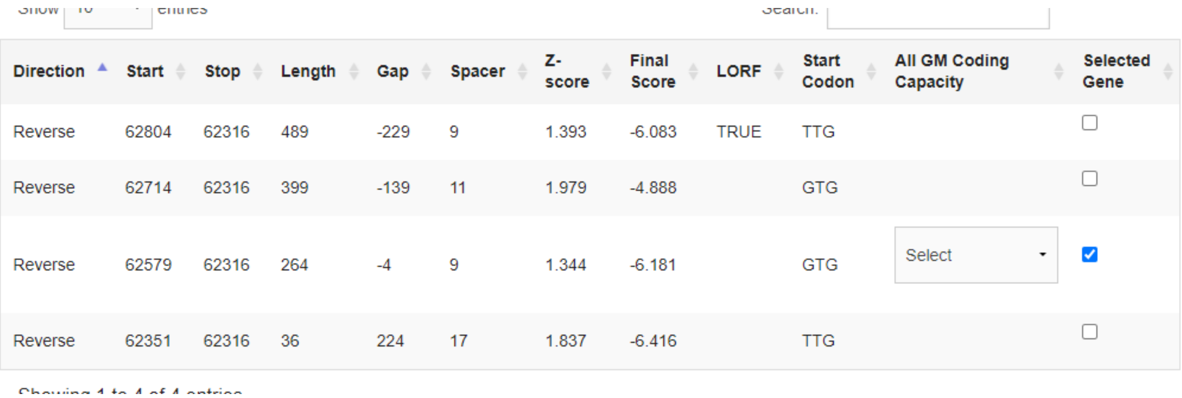
### 2. GeneMark coding potential


61600
62000
62400
62800
63200
Nucleotide Position
atypical.ps

## 3. Glimmer and GeneMark agreement

Glimmer did not call a gene at this position. GeneMark predicted that this gene would begin at 62579

## 4. Longest open reading frame (ORF) without excessive gap



Show 10 entries Search:

| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 62804 | 62316 | 489 | -229 | 9 | 1.393 | -6.083 | TRUE | TTG | | ☐ |
| Reverse | 62714 | 62316 | 399 | -139 | 11 | 1.979 | -4.888 | | GTG | | ☐ |
| Reverse | 62579 | 62316 | 264 | -4 | 9 | 1.344 | -6.181 | | GTG | Select | ☑ |
| Reverse | 62351 | 62316 | 36 | 224 | 17 | 1.837 | -6.416 | | TTG | | ☐ |

Showing 1 to 4 of 4 entries

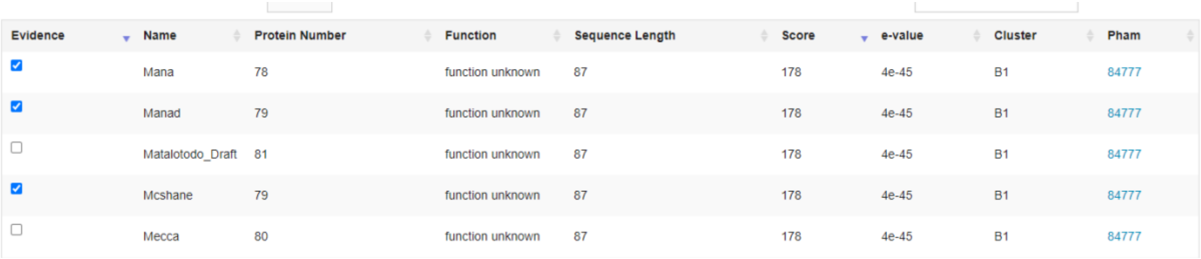
## 5. Function. Please follow this [Official SEAPHAGE Function List]()

If no functional prediction is present, write “Hypothetical protein”.

Hypothetical Protein

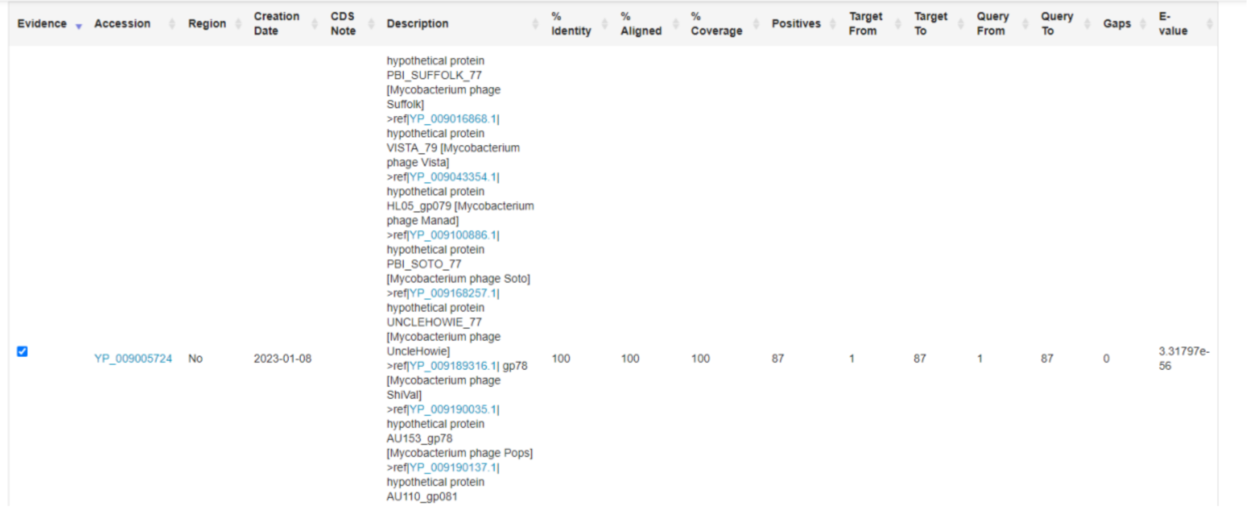
## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST



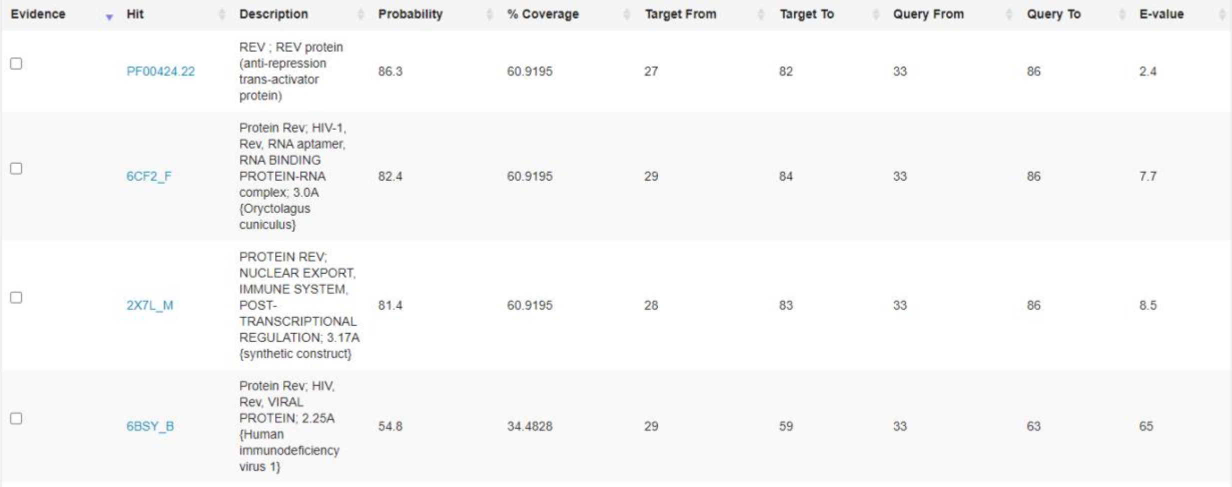
| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | Mana | 78 | function unknown | 87 | 178 | 4e-45 | B1 | 84777 |
| ☑ | Manad | 79 | function unknown | 87 | 178 | 4e-45 | B1 | 84777 |
| ☐ | Matalotodo_Draft | 81 | function unknown | 87 | 178 | 4e-45 | B1 | 84777 |
| ☑ | Mcshane | 79 | function unknown | 87 | 178 | 4e-45 | B1 | 84777 |
| ☐ | Mecca | 80 | function unknown | 87 | 178 | 4e-45 | B1 | 84777 |

### b. SIF: NCBI BLAST



| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | YP_009005724 | No | 2023-01-08 | | hypothetical protein PBI_SUFFOLK_77 [Mycobacterium phage Suffolk] >ref\|YP_009016868.1\| hypothetical protein VISTA_79 [Mycobacterium phage Vista] >ref\|YP_009043354.1\| hypothetical protein HL05_gp079 [Mycobacterium phage Manad] >ref\|YP_009100886.1\| hypothetical protein PBI_SOTO_77 [Mycobacterium phage Soto] >ref\|YP_009168257.1\| hypothetical protein UNCLEHOWIE_77 [Mycobacterium phage UncleHowie] >ref\|YP_009189316.1\| gp78 [Mycobacterium phage ShiVal] >ref\|YP_009190035.1\| hypothetical protein AU153_gp78 [Mycobacterium phage Pops] >ref\|YP_009190137.1\| hypothetical protein AU110_gp081 | 100 | 100 | 100 | 87 | 1 | 87 | 1 | 87 | 0 | 3.31797e-56 |

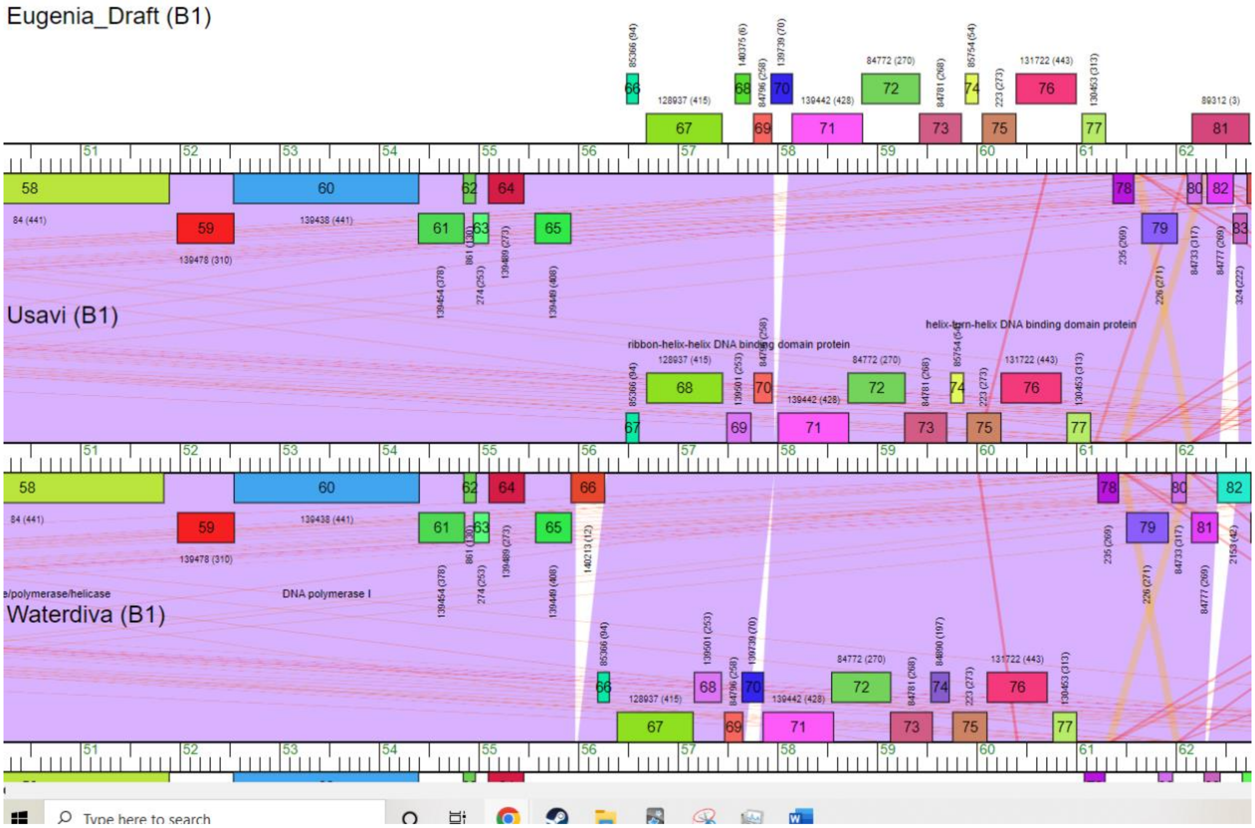
## c. SIF: HHPred



| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☐ | PF00424.22 | REV ; REV protein (anti-repression trans-activator protein) | 86.3 | 60.9195 | 27 | 82 | 33 | 86 | 2.4 |
| ☐ | 6CF2_F | Protein Rev; HIV-1, Rev, RNA aptamer, RNA BINDING PROTEIN-RNA complex; 3.0A {Oryctolagus cuniculus} | 82.4 | 60.9195 | 29 | 84 | 33 | 86 | 7.7 |
| ☐ | 2X7L_M | PROTEIN REV; NUCLEAR EXPORT, IMMUNE SYSTEM, POST-TRANSCRIPTIONAL REGULATION; 3.17A {synthetic construct} | 81.4 | 60.9195 | 28 | 83 | 33 | 86 | 8.5 |
| ☐ | 6BSY_B | Protein Rev; HIV, Rev, VIRAL PROTEIN; 2.25A {Human immunodeficiency virus 1} | 54.8 | 34.4828 | 29 | 59 | 33 | 63 | 65 |

## d. SIF: Synteny-Phamerator (three genomes)

Has synteny with Mana, Manad, Mcshane, ect.



# 7. Any other important information.

## Transmembrane prediction

Transmembrane domains

```
# WEBSEQUENCE Length: 87
# WEBSEQUENCE Number of predicted TMHs:  0
# WEBSEQUENCE Exp number of AAs in TMHs: 0.00422
# WEBSEQUENCE Exp number, first 60 AAs:  0.00034
# WEBSEQUENCE Total prob of N-in:        0.47778
WEBSEQUENCE      TMHMM2.0   outside      1    87
```

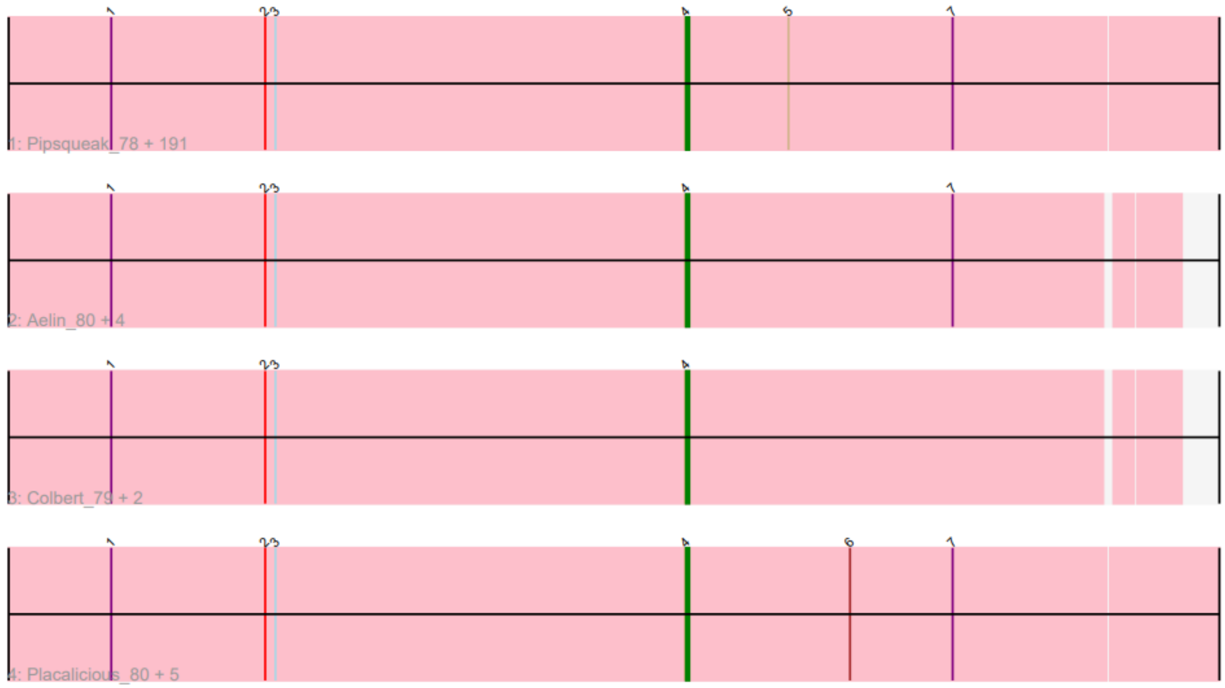
## Curator name: John Motter

## Gene Name: eugenia Gene 83

## DNA Master Notes: N/A

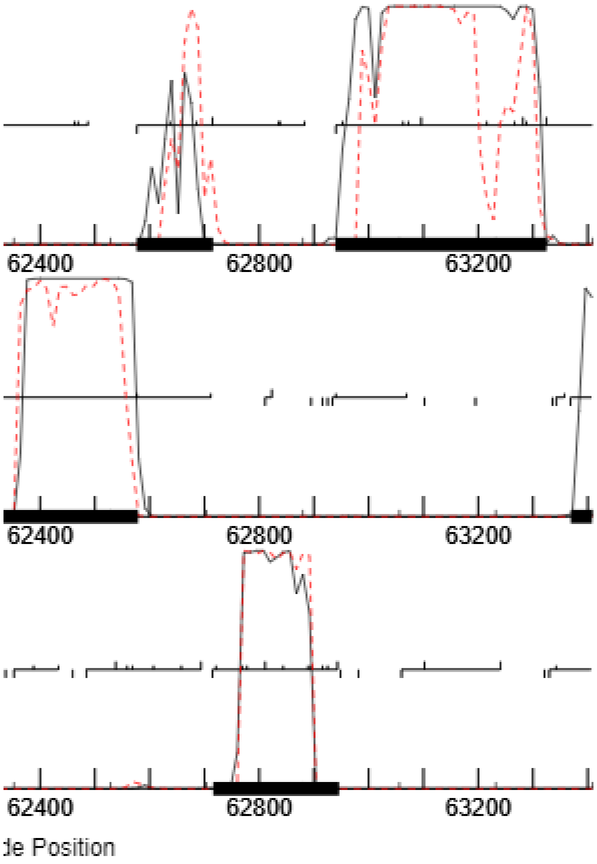
## Start Position Evaluation (In order):

### 1. Starterator



The start number called the most often in the published annotations is 4, it was called in 199 of the 199 non-draft genes in the pham.

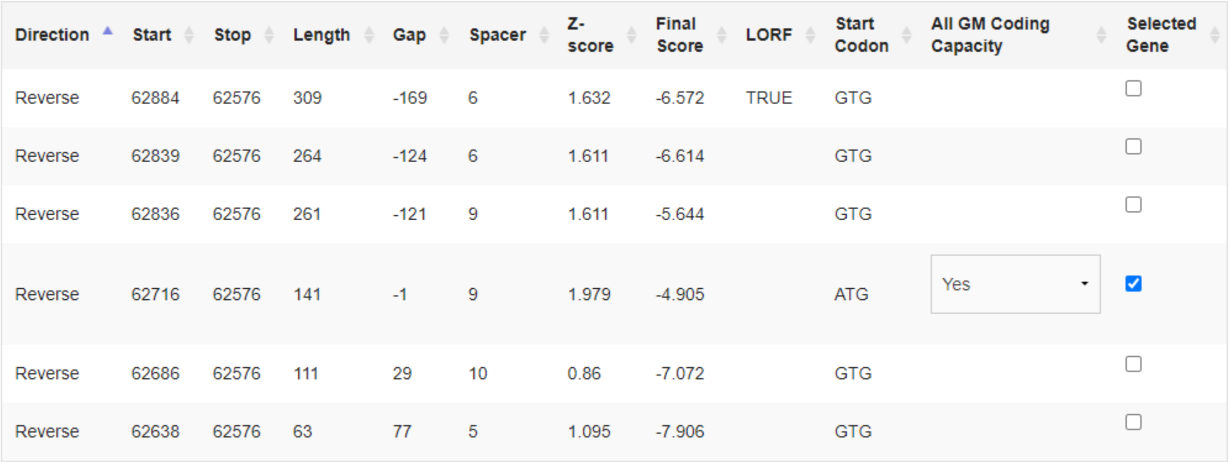
### 2. GeneMark coding potential



## 3. Glimmer and GeneMark agreement

Glimmer did not call the gene, however GeneMark called the gene at 26716

## 4. Longest open reading frame (ORF) without excessive gap



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 62884 | 62576 | 309 | -169 | 6 | 1.632 | -6.572 | TRUE | GTG | | ☐ |
| Reverse | 62839 | 62576 | 264 | -124 | 6 | 1.611 | -6.614 | | GTG | | ☐ |
| Reverse | 62836 | 62576 | 261 | -121 | 9 | 1.611 | -5.644 | | GTG | | ☐ |
| Reverse | 62716 | 62576 | 141 | -1 | 9 | 1.979 | -4.905 | | ATG | Yes | ☑ |
| Reverse | 62686 | 62576 | 111 | 29 | 10 | 0.86 | -7.072 | | GTG | | ☐ |
| Reverse | 62638 | 62576 | 63 | 77 | 5 | 1.095 | -7.906 | | GTG | | ☐ |

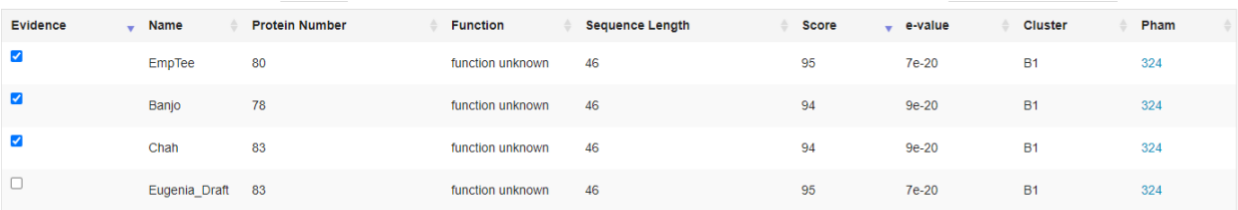
## 5. Function. Please follow this Official SEAPHAGE Function List

If no functional prediction is present, write “Hypothetical protein”.

Membrane protein

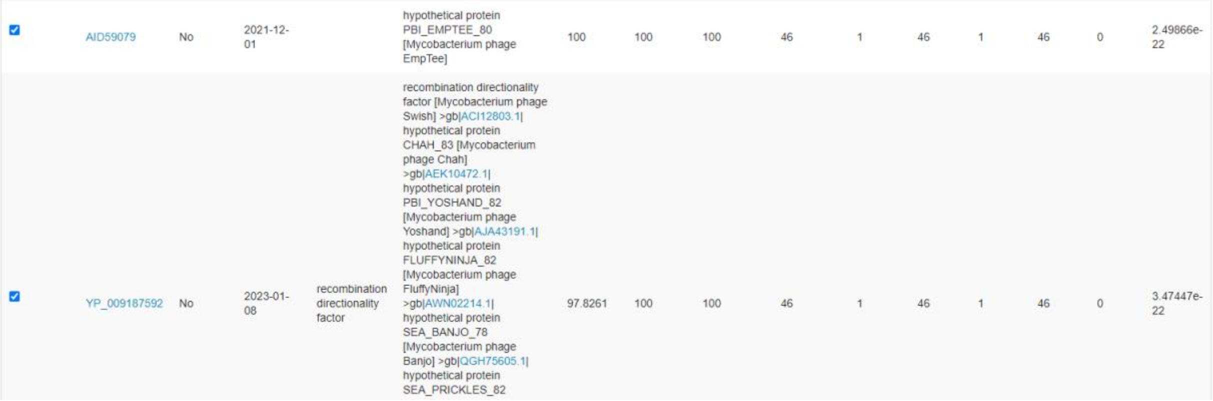
## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST



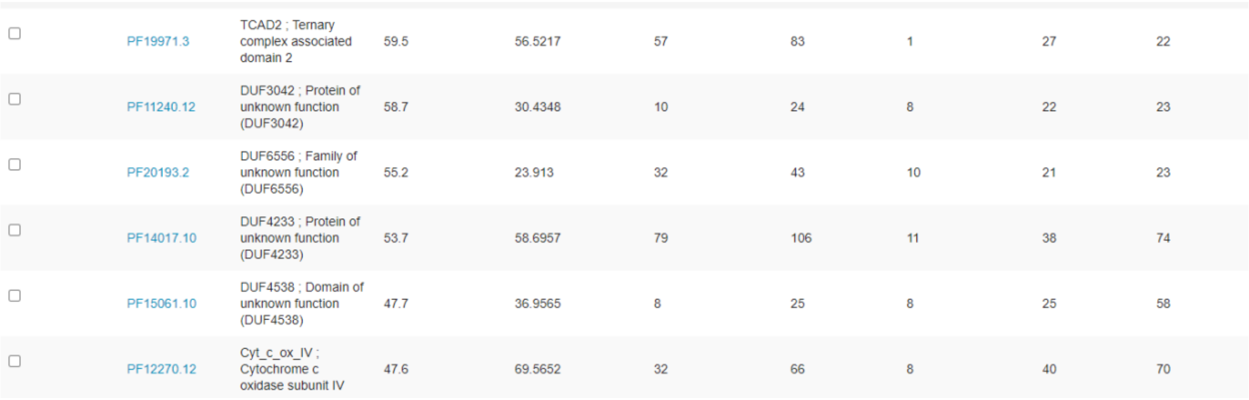
| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | EmpTee | 80 | function unknown | 46 | 95 | 7e-20 | B1 | 324 |
| ☑ | Banjo | 78 | function unknown | 46 | 94 | 9e-20 | B1 | 324 |
| ☑ | Chah | 83 | function unknown | 46 | 94 | 9e-20 | B1 | 324 |
| ☐ | Eugenia_Draft | 83 | function unknown | 46 | 95 | 7e-20 | B1 | 324 |

### b. SIF: NCBI BLAST



| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | AID59079 | No | 2021-12-01 | | hypothetical protein PBI_EMPTEE_80 [Mycobacterium phage EmpTee] | 100 | 100 | 100 | 46 | 1 | 46 | 1 | 46 | 0 | 2.49866e-22 |
| ☑ | YP_009187592 | No | 2023-01-08 | recombination directionality factor | recombination directionality factor [Mycobacterium phage Swish] >gb\|ACI12803.1\| hypothetical protein CHAH_83 [Mycobacterium phage Chah] >gb\|AEK10472.1\| hypothetical protein PBI_YOSHAND_82 [Mycobacterium phage Yoshand] >gb\|AJA43191.1\| hypothetical protein FLUFFYNINJA_82 [Mycobacterium phage FluffyNinja] >gb\|AWN02214.1\| hypothetical protein SEA_BANJO_78 [Mycobacterium phage Banjo] >gb\|QGH75605.1\| hypothetical protein SEA_PRICKLES_82 | 97.8261 | 100 | 100 | 46 | 1 | 46 | 1 | 46 | 0 | 3.47447e-22 |

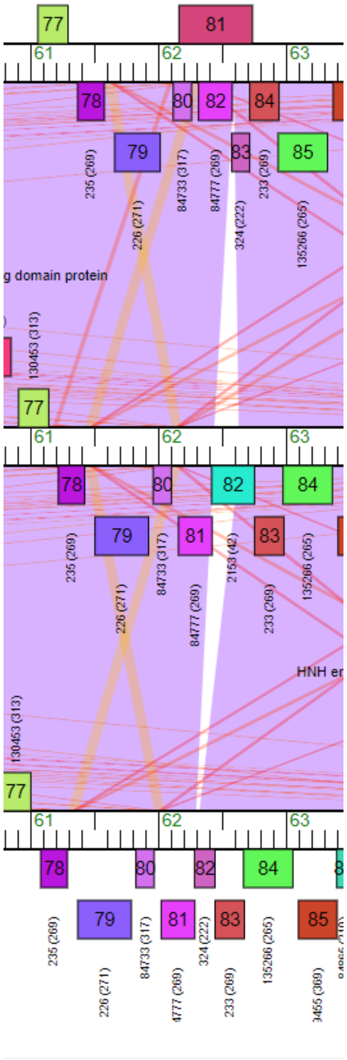
### c. SIF: HHPred



| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ☐ | PF19971.3 | TCAD2 ; Ternary complex associated domain 2 | 59.5 | 56.5217 | 57 | 83 | 1 | 27 | 22 |
| ☐ | PF11240.12 | DUF3042 ; Protein of unknown function (DUF3042) | 58.7 | 30.4348 | 10 | 24 | 8 | 22 | 23 |
| ☐ | PF20193.2 | DUF6556 ; Family of unknown function (DUF6556) | 55.2 | 23.913 | 32 | 43 | 10 | 21 | 23 |
| ☐ | PF14017.10 | DUF4233 ; Protein of unknown function (DUF4233) | 53.7 | 58.6957 | 79 | 106 | 11 | 38 | 74 |
| ☐ | PF15061.10 | DUF4538 ; Domain of unknown function (DUF4538) | 47.7 | 36.9565 | 8 | 25 | 8 | 25 | 58 |
| ☐ | PF12270.12 | Cyt_c_ox_IV ; Cytochrome c oxidase subunit IV | 47.6 | 69.5652 | 32 | 66 | 8 | 40 | 70 |

### d. SIF: Synteny-Phamerator (three genomes)

Shows synteny with EmpTee, Banjo, Chah, ect.



## 7. Any other important information.

## Transmembrane Protein

```
# WEBSEQUENCE Length: 46
# WEBSEQUENCE Number of predicted TMHs:  1
# WEBSEQUENCE Exp number of AAs in TMHs: 20.9652
# WEBSEQUENCE Exp number, first 60 AAs:  20.9652
# WEBSEQUENCE Total prob of N-in:        0.00149
# WEBSEQUENCE POSSIBLE N-term signal sequence
```

```
WEBSEQUENCE     TMHMM2.0   outside     1     9
WEBSEQUENCE     TMHMM2.0   TMhelix          10    29
WEBSEQUENCE     TMHMM2.0   inside     30    46
```

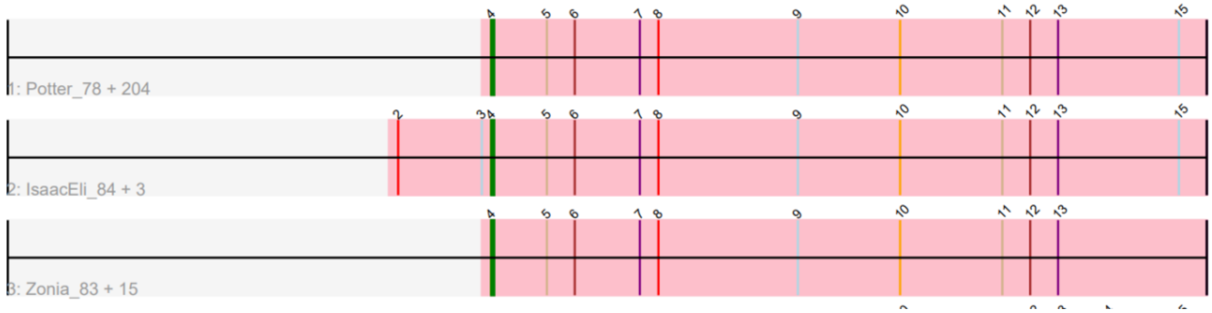
## CURATOR NAME: JOHN MOTTER

## GENE NAME: EUGENIA GENE 84

## DNA MASTER NOTES: N/A

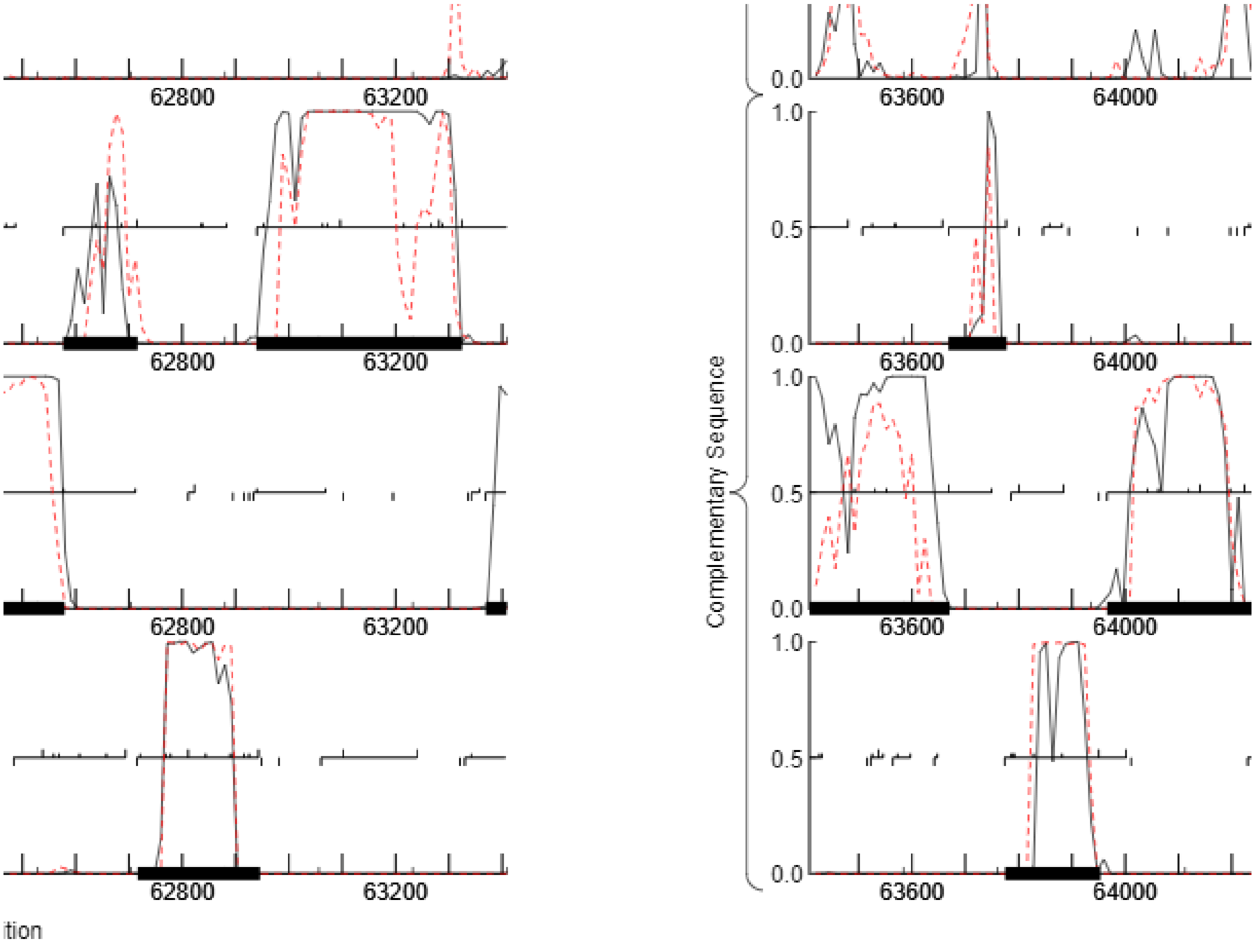
## START POSITION EVALUATION (IN ORDER):

### 1. Starterator

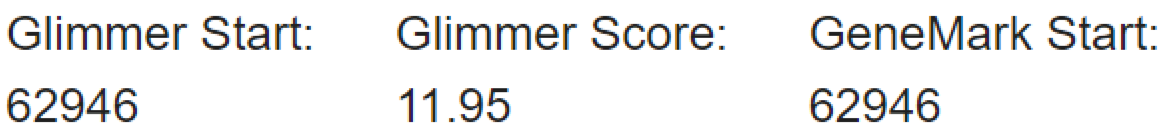


The start number called the most often in the published annotations is 4, it was called in 241 of the 245 non-draft genes in the pham.

### 2. GeneMark coding potential

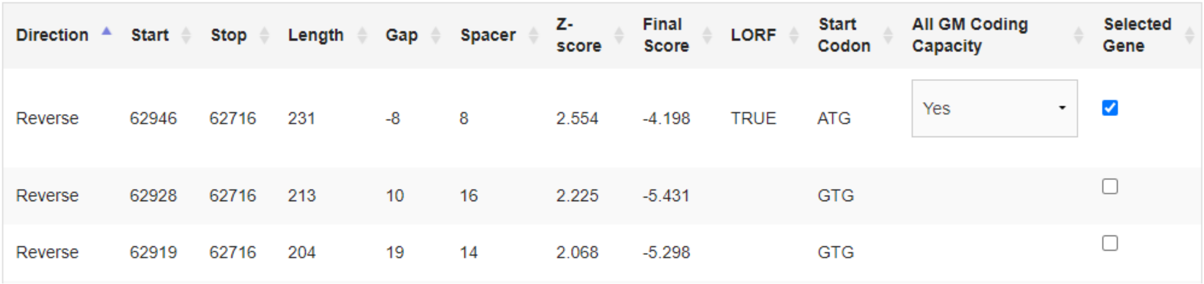


### 3. Glimmer and GeneMark agreement



| Glimmer Start: | Glimmer Score: | GeneMark Start: |
|---|---|---|
| 62946 | 11.95 | 62946 |

### 4. Longest open reading frame (ORF) without excessive gap



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 62946 | 62716 | 231 | -8 | 8 | 2.554 | -4.198 | TRUE | ATG | Yes | ☑ |
| Reverse | 62928 | 62716 | 213 | 10 | 16 | 2.225 | -5.431 |  | GTG |  | ☐ |
| Reverse | 62919 | 62716 | 204 | 19 | 14 | 2.068 | -5.298 |  | GTG |  | ☐ |

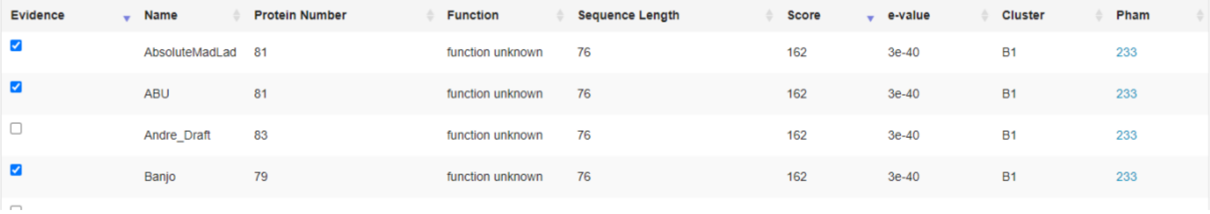
5. Function. Please follow this [Official SEAPHAGE Function List]()

If no functional prediction is present, write “Hypothetical protein”.

hypothetical protein

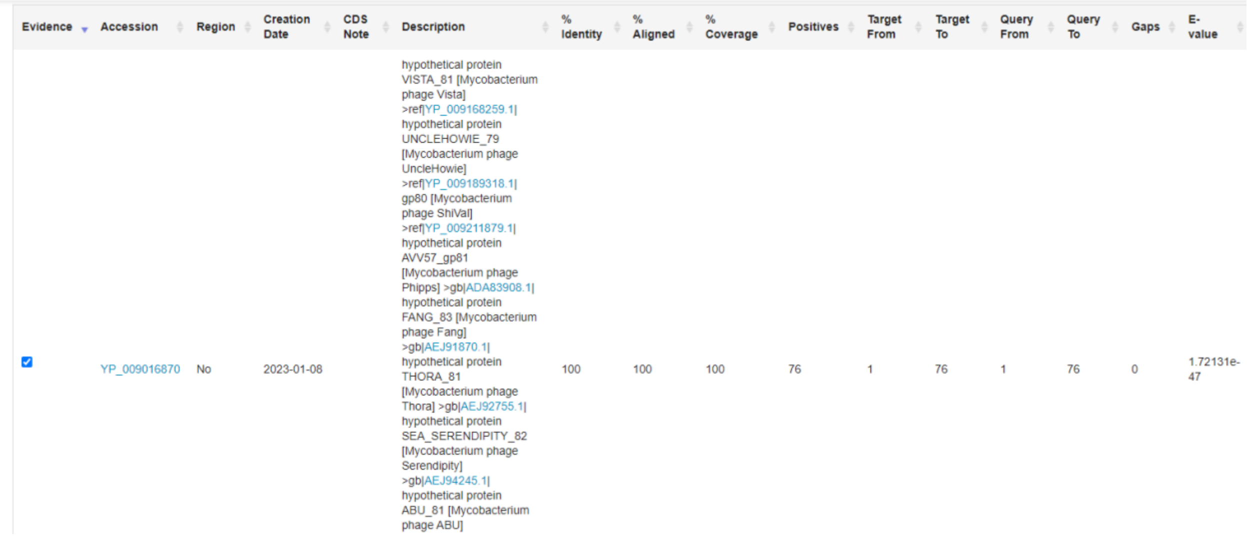
6. Supporting Information for Function (SIF)

a. SIF: PhageDb BLAST



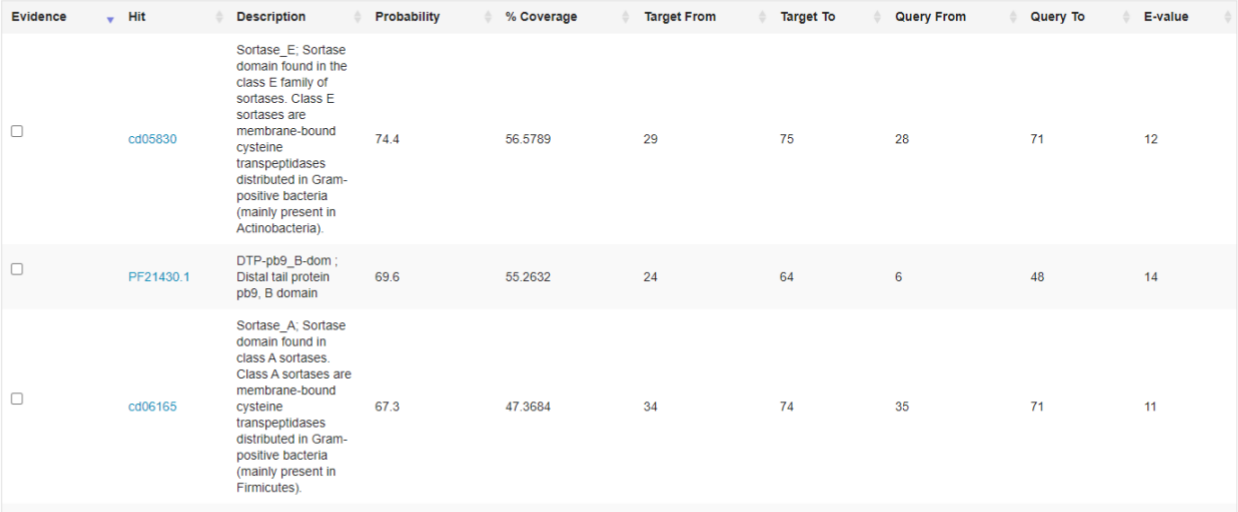
| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | AbsoluteMadLad | 81 | function unknown | 76 | 162 | 3e-40 | B1 | 233 |
| ☑ | ABU | 81 | function unknown | 76 | 162 | 3e-40 | B1 | 233 |
| ☐ | Andre_Draft | 83 | function unknown | 76 | 162 | 3e-40 | B1 | 233 |
| ☑ | Banjo | 79 | function unknown | 76 | 162 | 3e-40 | B1 | 233 |

b. SIF: NCBI BLAST



| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | YP_009016870 | No | 2023-01-08 | | hypothetical protein VISTA_81 [Mycobacterium phage Vista] >ref\|YP_009168259.1\| hypothetical protein UNCLEHOWIE_79 [Mycobacterium phage UncleHowie] >ref\|YP_009189318.1\| gp80 [Mycobacterium phage ShiVal] >ref\|YP_009211879.1\| hypothetical protein AVV57_gp81 [Mycobacterium phage Phipps] >gb\|ADA83908.1\| hypothetical protein FANG_83 [Mycobacterium phage Fang] >gb\|AEJ91870.1\| hypothetical protein THORA_81 [Mycobacterium phage Thora] >gb\|AEJ92755.1\| hypothetical protein SEA_SERENDIPITY_82 [Mycobacterium phage Serendipity] >gb\|AEJ94245.1\| hypothetical protein ABU_81 [Mycobacterium phage ABU] | 100 | 100 | 100 | 76 | 1 | 76 | 1 | 76 | 0 | 1.72131e-47 |

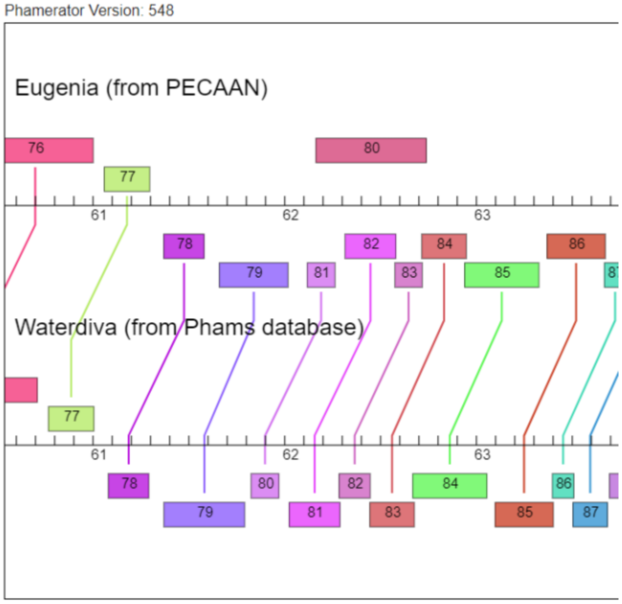
c. SIF: HHPred

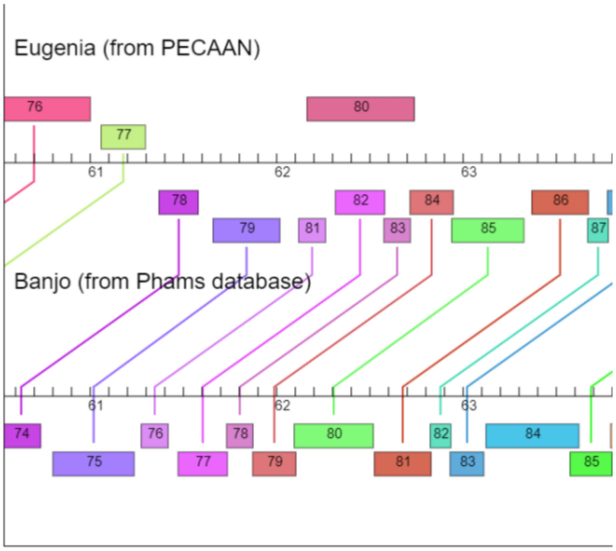


| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☐ | cd05830 | Sortase_E; Sortase domain found in the class E family of sortases. Class E sortases are membrane-bound cysteine transpeptidases distributed in Gram-positive bacteria (mainly present in Actinobacteria). | 74.4 | 56.5789 | 29 | 75 | 28 | 71 | 12 |
| ☐ | PF21430.1 | DTP-pb9_B-dom ; Distal tail protein pb9, B domain | 69.6 | 55.2632 | 24 | 64 | 6 | 48 | 14 |
| ☐ | cd06165 | Sortase_A; Sortase domain found in class A sortases. Class A sortases are membrane-bound cysteine transpeptidases distributed in Gram-positive bacteria (mainly present in Firmicutes). | 67.3 | 47.3684 | 34 | 74 | 35 | 71 | 11 |

### d. SIF: Synteny-Phamerator (three genomes)

Has synteny with AbsoluteMadLad, ABU, Banjo, etc.





## 7. Any other important information.

### Transmembrane Prediction

```
# WEBSEQUENCE Length: 76
# WEBSEQUENCE Number of predicted TMHs:  0
# WEBSEQUENCE Exp number of AAs in TMHs: 0.00488
# WEBSEQUENCE Exp number, first 60 AAs:  0.00351
# WEBSEQUENCE Total prob of N-in:        0.08565
WEBSEQUENCE      TMHMM2.0   outside      1    76
```

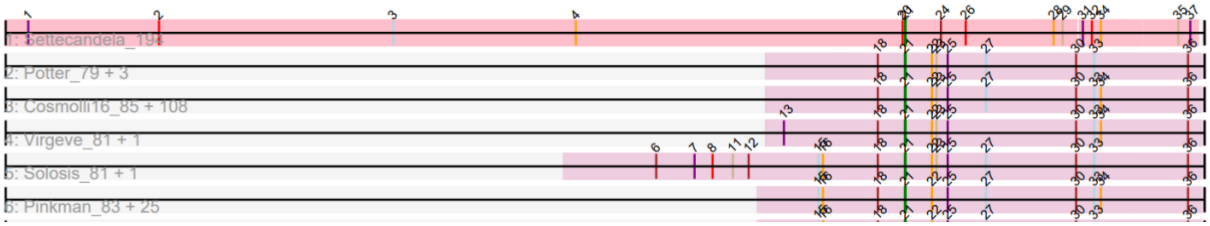
## CURATOR NAME: JOHN MOTTER

## GENE NAME: EUGENIA GENE 85

## DNA MASTER NOTES: N/A

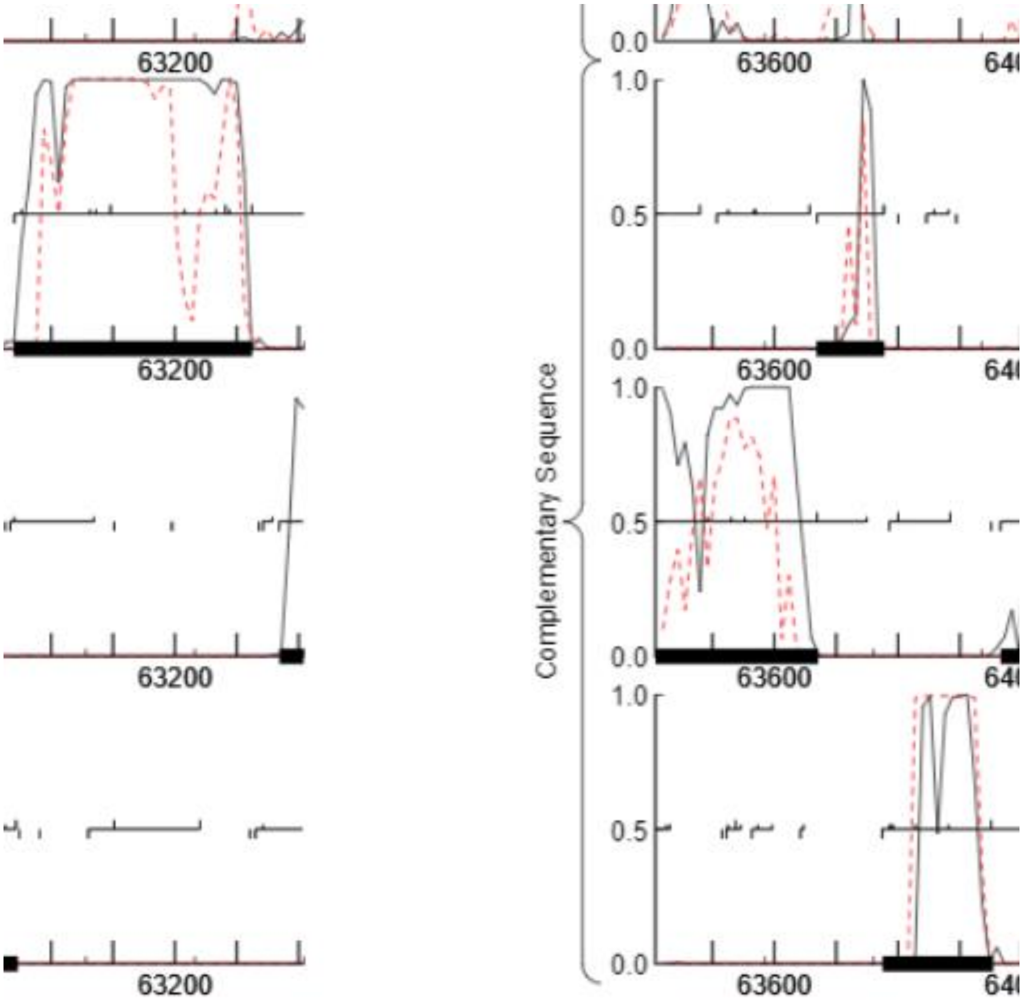
## START POSITION EVALUATION (IN ORDER):

### 1. Starterator

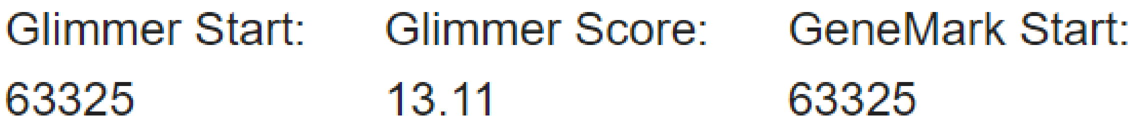


The start number called the most often in the published annotations is 21, it was called in 198 of the 240 non-draft genes in the pham.

### 2. GeneMark coding potential



3. Glimmer and GeneMark agreement



| Glimmer Start: | Glimmer Score: | GeneMark Start: |
|---|---|---|
| 63325 | 13.11 | 63325 |

4. Longest open reading frame (ORF) without excessive gap

No.

| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 63481 | 62939 | 543 | -113 | 18 | 2.442 | -5.502 | TRUE | ATG | | ☐ |
| Reverse | 63361 | 62939 | 423 | 7 | 9 | 1.354 | -6.161 | | TTG | | ☐ |
| Reverse | 63325 | 62939 | 387 | 43 | 16 | 2.517 | -4.846 | | ATG | Select | ☑ |

5. Function. Please follow this Official SEAPHAGE Function List

If no functional prediction is present, write “Hypothetical protein”.

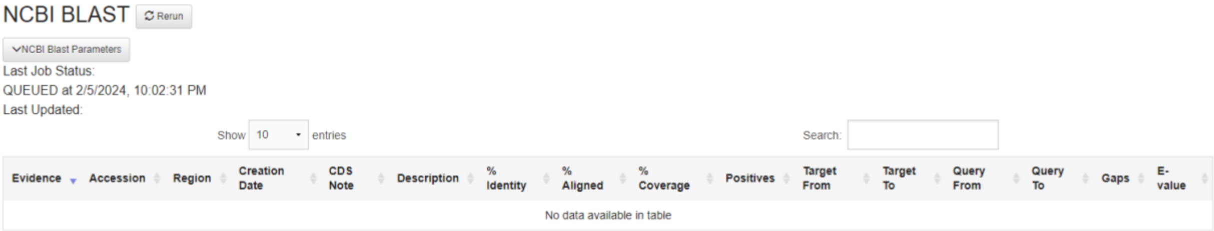
hypothetical protein

6. Supporting Information for Function (SIF)

a. SIF: PhageDb BLAST

| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | Banjo | 80 | function unknown | 140 | 279 | 2e-75 | B1 | 135266 |
| ☑ | Cobra | 83 | function unknown | 140 | 279 | 2e-75 | B1 | 135266 |
| ☑ | Craff | 85 | function unknown | 140 | 279 | 2e-75 | B1 | 135266 |

b. SIF: NCBI BLAST



NCBI BLAST

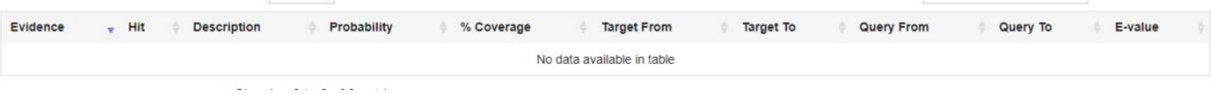
Last Job Status:
QUEUED at 2/5/2024, 10:02:31 PM
Last Updated:

Show 10 entries

Search:

| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| No data available in table | | | | | | | | | | | | | | | |

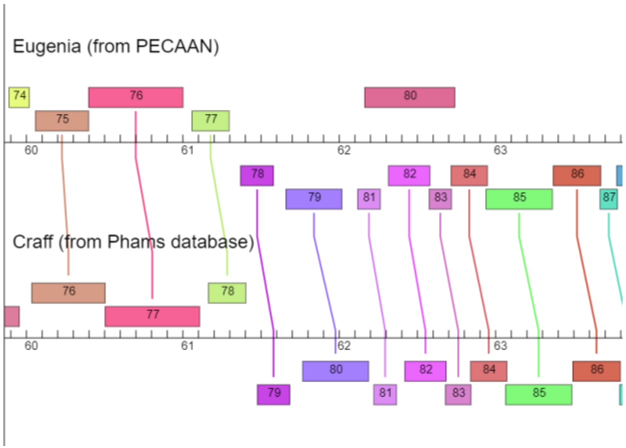
c. SIF: HHPred

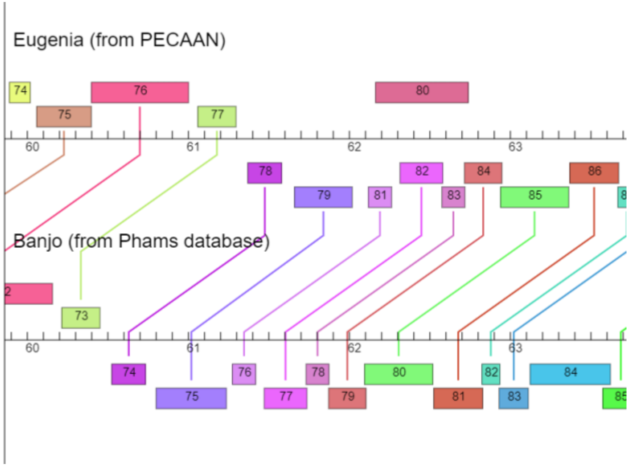


| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| No data available in table | | | | | | | | | |

d. SIF: Synteny-Phamerator (three genomes)

Shows synteny with Banjo, Cobra, Craff, etc.





## 7. Any other important information.

## Transmembrane Prediction

```
# WEBSEQUENCE Length: 140
# WEBSEQUENCE Number of predicted TMHs:  0
# WEBSEQUENCE Exp number of AAs in TMHs: 0
# WEBSEQUENCE Exp number, first 60 AAs:  0
# WEBSEQUENCE Total prob of N-in:        0.08887
WEBSEQUENCE     TMHMM2.0   outside     1   140
```

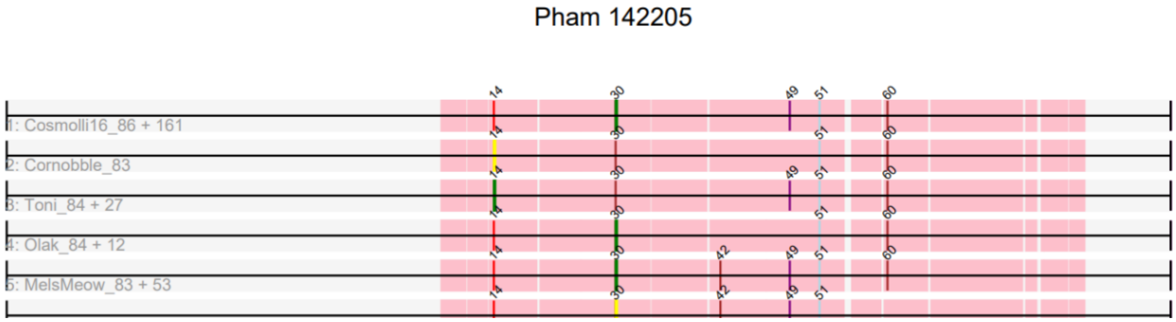
**Curator Name: John Motter**

## GENE NAME: EUGENIA GENE 86

## DNA MASTER NOTES: N/A

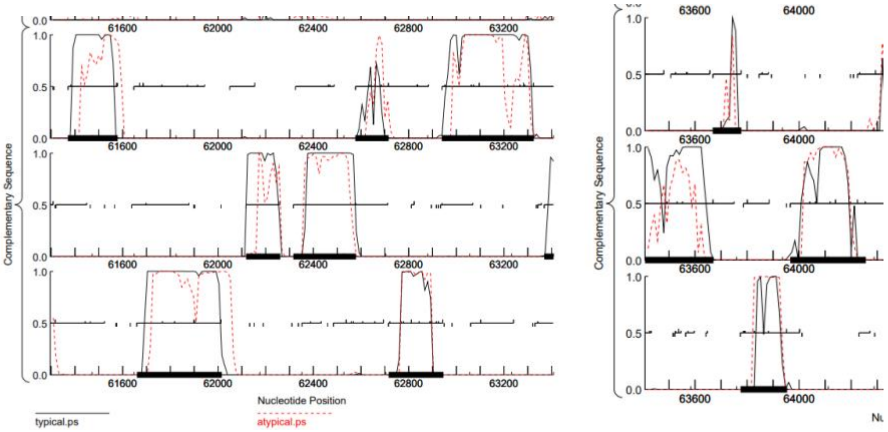
## START POSITION EVALUATION (IN ORDER):

### 1. Starterator

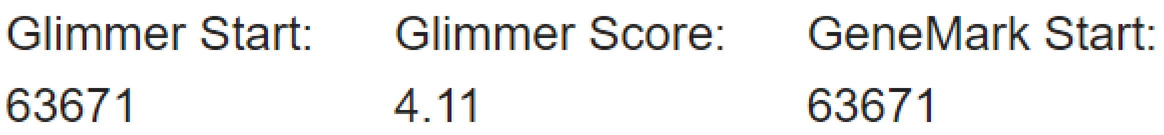


The start number called the most often in the published annotations is 30, it was called in 211 of the 334 non-draft genes in the pham.

### 2. GeneMark coding potential

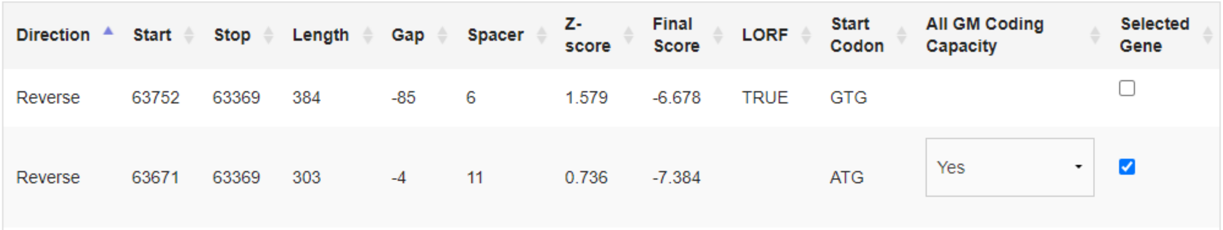


### 3. Glimmer and GeneMark agreement



| Glimmer Start: | Glimmer Score: | GeneMark Start: |
|---|---|---|
| 63671 | 4.11 | 63671 |

4. Longest open reading frame (ORF) without excessive gap



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 63752 | 63369 | 384 | -85 | 6 | 1.579 | -6.678 | TRUE | GTG | | ☐ |
| Reverse | 63671 | 63369 | 303 | -4 | 11 | 0.736 | -7.384 | | ATG | Yes | ☑ |

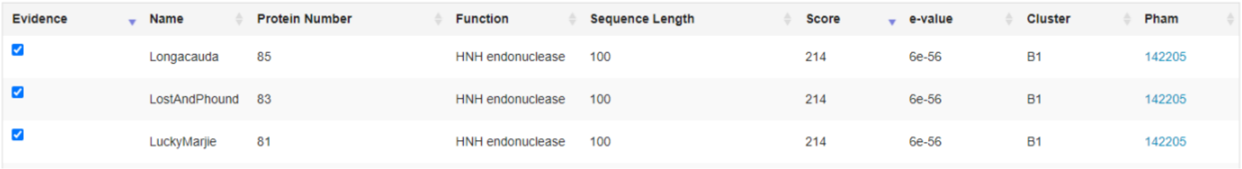
5. Function. Please follow this [Official SEAPHAGE Function List]

If no functional prediction is present, write “Hypothetical protein”.

HNH endonuclease

6. Supporting Information for Function (SIF)

a. SIF: PhageDb BLAST

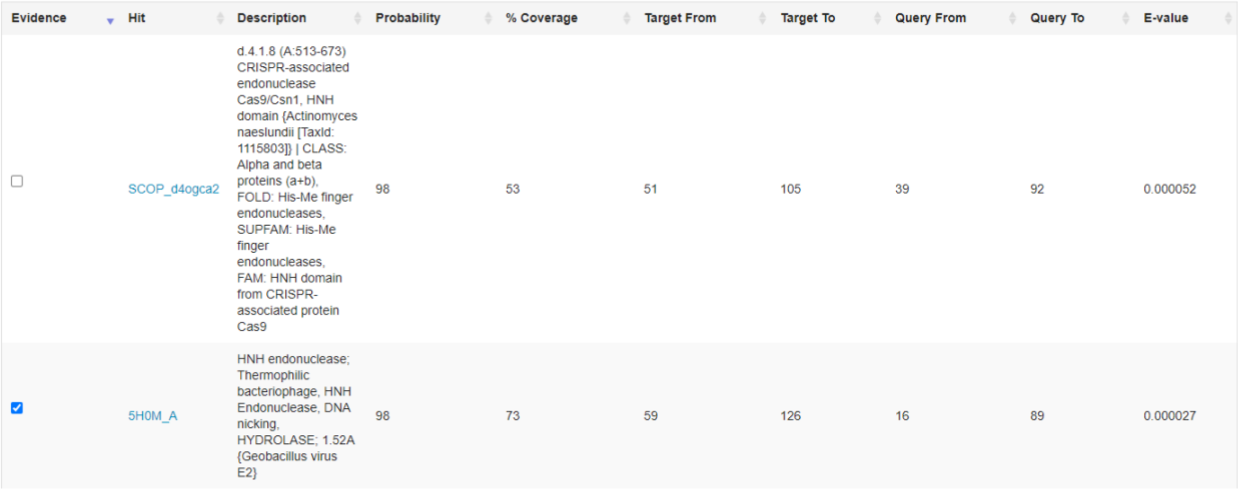


| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | Longacauda | 85 | HNH endonuclease | 100 | 214 | 6e-56 | B1 | 142205 |
| ☑ | LostAndPhound | 83 | HNH endonuclease | 100 | 214 | 6e-56 | B1 | 142205 |
| ☑ | LuckyMarjie | 81 | HNH endonuclease | 100 | 214 | 6e-56 | B1 | 142205 |

b. SIF: NCBI BLAST

| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | YP_009016872 | No | 2023-01-08 | | HNH endonuclease [Mycobacterium phage Vista] >ref\|YP_009018397.1\| HNH endonuclease [Mycobacterium phage JacAttac] >ref\|YP_009190141.1\| HNH endonuclease [Mycobacterium phage Badfish] >ref\|YP_009211881.1\| HNH endonuclease [Mycobacterium phage Phipps] >ref\|YP_010096588.1\| HNH endonuclease [Mycobacterium phage KingTut] >gb\|ADA84013.1\| hypothetical protein SCOOT17C_85 [Mycobacterium phage Scoot17C] >gb\|AEJ91868.1\| HNH endonuclease domain protein [Mycobacterium phage Thora] >gb\|AEJ94247.1\| hypothetical protein ABU_83 | 100 | 100 | 100 | 100 | 1 | 100 | 1 | 100 | 0 | 8.5402e-66 |

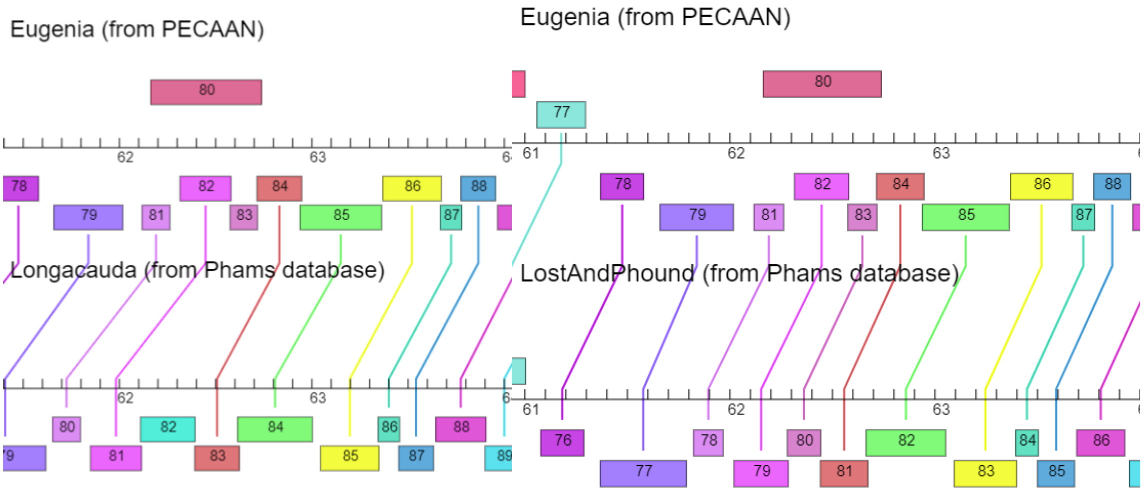
c. SIF: HHPred



| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☐ | SCOP_d4ogca2 | d.4.1.8 (A:513-673) CRISPR-associated endonuclease Cas9/Csn1, HNH domain {Actinomyces naeslundii [TaxId: 1115803]} \| CLASS: Alpha and beta proteins (a+b), FOLD: His-Me finger endonucleases, SUPFAM: His-Me finger endonucleases, FAM: HNH domain from CRISPR-associated protein Cas9 | 98 | 53 | 51 | 105 | 39 | 92 | 0.000052 |
| ☑ | 5H0M_A | HNH endonuclease; Thermophilic bacteriophage, HNH Endonuclease, DNA nicking, HYDROLASE; 1.52A {Geobacillus virus E2} | 98 | 73 | 59 | 126 | 16 | 89 | 0.000027 |

d. SIF: Synteny-Phamerator (three genomes)

Shows synteny with Longacauda, LostAndPhound, LuckeyMarjie, etc.



## 7. Any other important information.

Transmembrane prediction

```
# WEBSEQUENCE Length: 100
# WEBSEQUENCE Number of predicted TMHs:  0
# WEBSEQUENCE Exp number of AAs in TMHs: 0
# WEBSEQUENCE Exp number, first 60 AAs:  0
# WEBSEQUENCE Total prob of N-in:        0.61520
WEBSEQUENCE      TMHMM2.0   inside       1   100
```

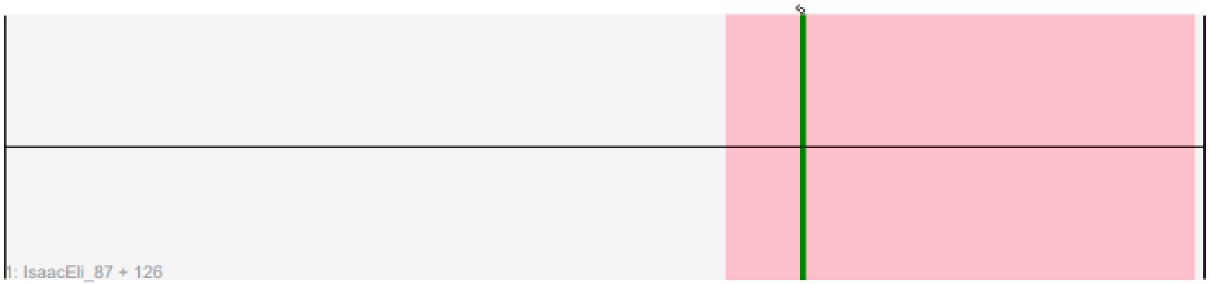
## CURATOR NAME: JOHN MOTTER

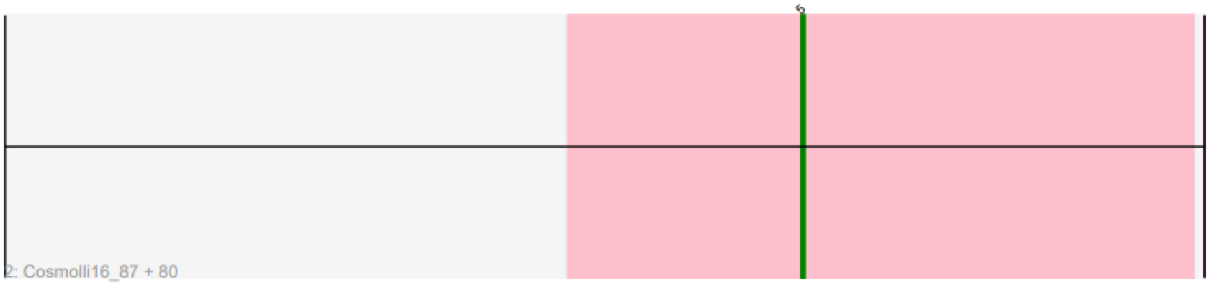
## GENE NAME: EUGENIA GENE 87

## DNA MASTER NOTES: N/A

## START POSITION EVALUATION (IN ORDER):

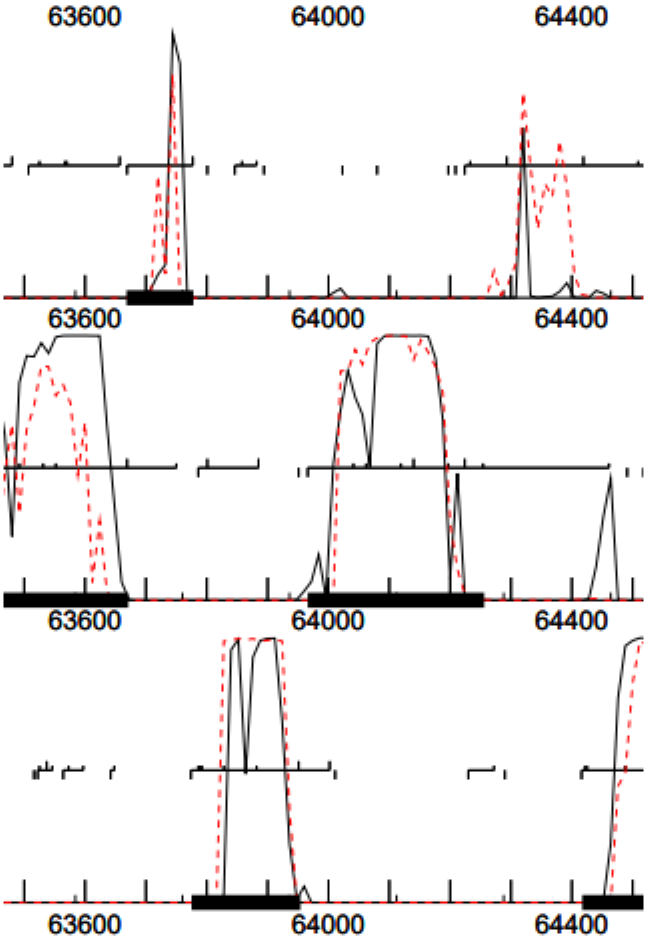
### 1. Starterator





The start number called the most often in the published annotations is 5, it was called in 191 of the 193 non-draft genes in the pham.

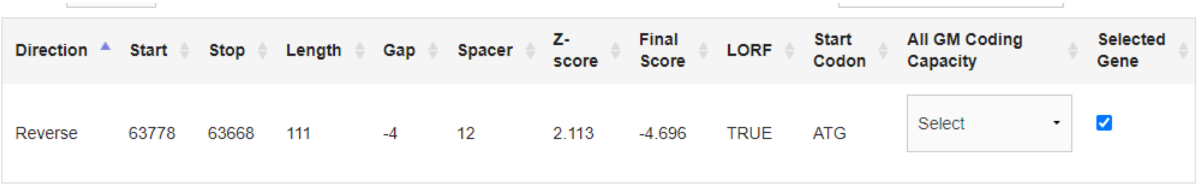
### 2. GeneMark coding potential



## 3. Glimmer and GeneMark agreement

Glimmer did not call the gene, however GeneMark called the gene at 63778

## 4. Longest open reading frame (ORF) without excessive gap



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 63778 | 63668 | 111 | -4 | 12 | 2.113 | -4.696 | TRUE | ATG | Select | ☑ |

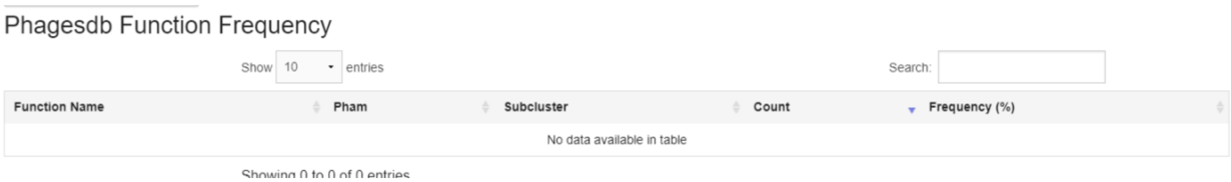
## 5. Function. Please follow this [Official SEAPHAGE Function List]()

If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

## 6. Supporting Information for Function (SIF)

## a. SIF: PhageDb BLAST

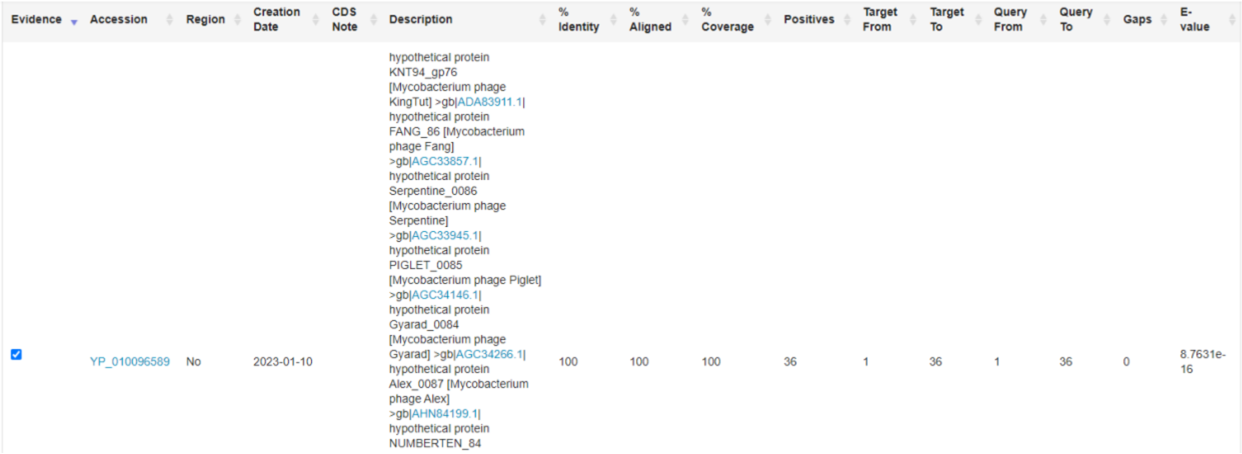


Phagesdb Function Frequency

| Function Name | Pham | Subcluster | Count | Frequency (%) |
|---|---|---|---|---|
| No data available in table | | | | |

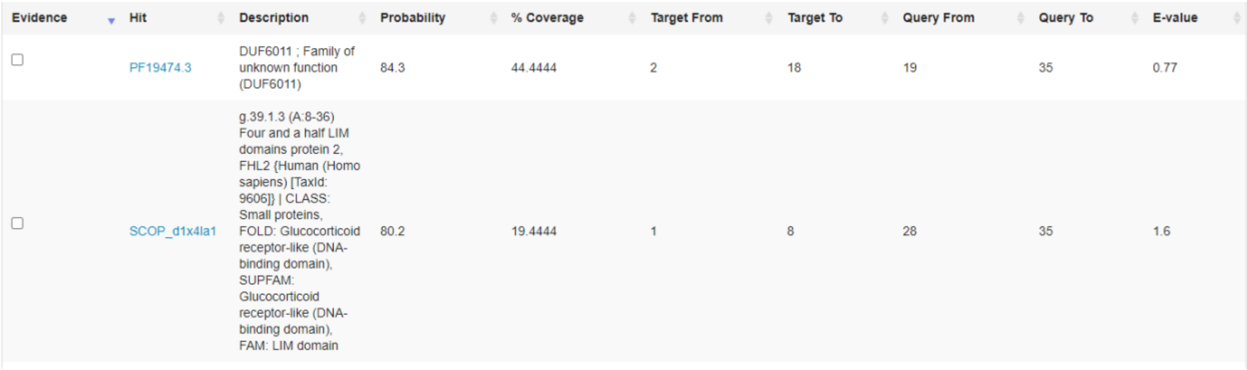
Showing 0 to 0 of 0 entries

## b. SIF: NCBI BLAST



| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | YP_010096589 | No | 2023-01-10 | | hypothetical protein KNT94_gp76 [Mycobacterium phage KingTut] >gb\|ADA83911.1\| hypothetical protein FANG_86 [Mycobacterium phage Fang] >gb\|AGC33857.1\| hypothetical protein Serpentine_0086 [Mycobacterium phage Serpentine] >gb\|AGC33945.1\| hypothetical protein PIGLET_0085 [Mycobacterium phage Piglet] >gb\|AGC34146.1\| hypothetical protein Gyarad_0084 [Mycobacterium phage Gyarad] >gb\|AGC34266.1\| hypothetical protein Alex_0087 [Mycobacterium phage Alex] >gb\|AHN84199.1\| hypothetical protein NUMBERTEN_84 | 100 | 100 | 100 | 36 | 1 | 36 | 1 | 36 | 0 | 8.7631e-16 |

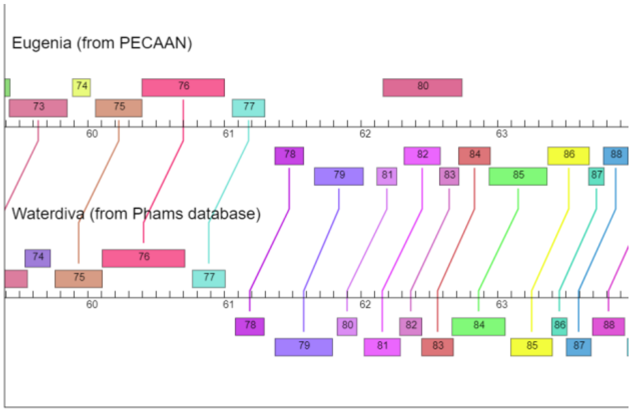
## c. SIF: HHPred

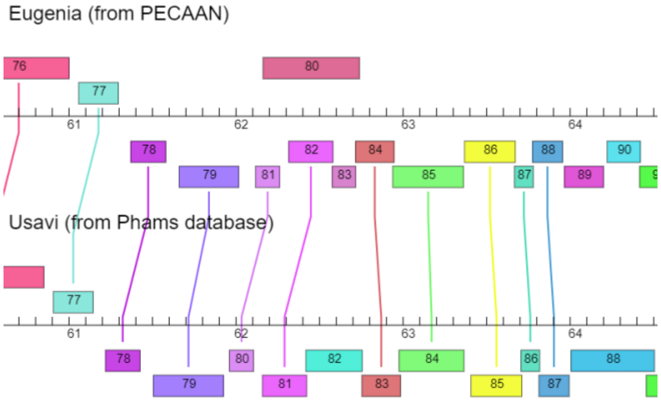


| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☐ | PF19474.3 | DUF6011 ; Family of unknown function (DUF6011) | 84.3 | 44.4444 | 2 | 18 | 19 | 35 | 0.77 |
| ☐ | SCOP_d1x4la1 | g.39.1.3 (A:8-36) Four and a half LIM domains protein 2, FHL2 {Human (Homo sapiens) [TaxId: 9606]} \| CLASS: Small proteins, FOLD: Glucocorticoid receptor-like (DNA-binding domain), SUPFAM: Glucocorticoid receptor-like (DNA-binding domain), FAM: LIM domain | 80.2 | 19.4444 | 1 | 8 | 28 | 35 | 1.6 |

## d. SIF: Synteny-Phamerator (three genomes)

Shows synteny with Adriana, Anderson, Ashraf, etc.





## 7. Any other important information.

## Transmembrane Prediction

```
# WEBSEQUENCE Length: 36
# WEBSEQUENCE Number of predicted TMHs:  0
# WEBSEQUENCE Exp number of AAs in TMHs: 0.00553
# WEBSEQUENCE Exp number, first 60 AAs:  0.00553
# WEBSEQUENCE Total prob of N-in:        0.49689
WEBSEQUENCE      TMHMM2.0   inside       1    36
```

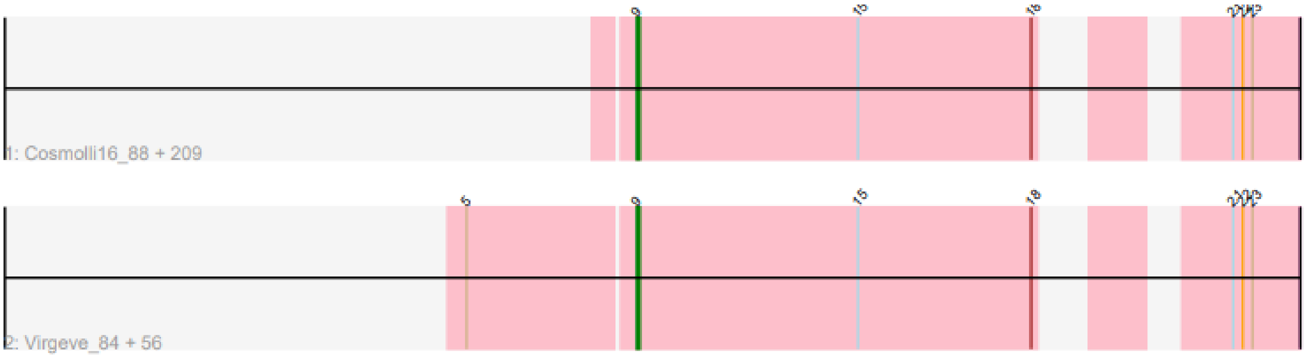
## CURATOR NAME: JOHN MOTTER

## GENE NAME: EUGENIA GENE 88

## DNA MASTER NOTES: N/A

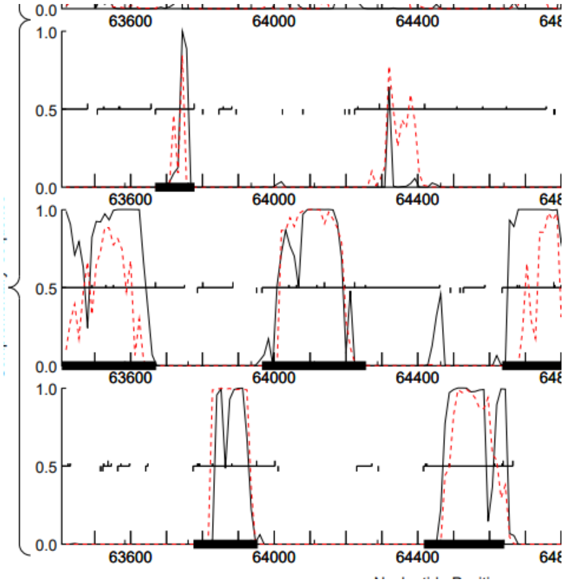
## START POSITION EVALUATION (IN ORDER):

### 1. Starterator

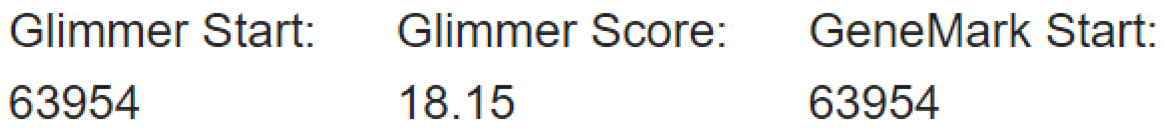


The start number called the most often in the published annotations is 9, it was called in 246 of the 247 non-draft genes in the pham.

### 2. GeneMark coding potential

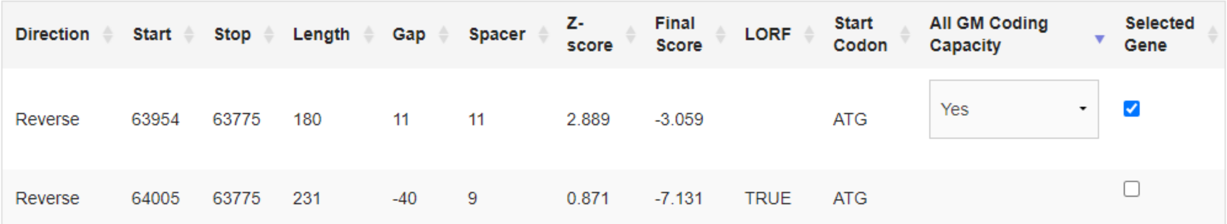


3. Glimmer and GeneMark agreement



| Glimmer Start: | Glimmer Score: | GeneMark Start: |
|---|---|---|
| 63954 | 18.15 | 63954 |

4. Longest open reading frame (ORF) without excessive gap



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 63954 | 63775 | 180 | 11 | 11 | 2.889 | -3.059 | | ATG | Yes | ☑ |
| Reverse | 64005 | 63775 | 231 | -40 | 9 | 0.871 | -7.131 | TRUE | ATG | | ☐ |

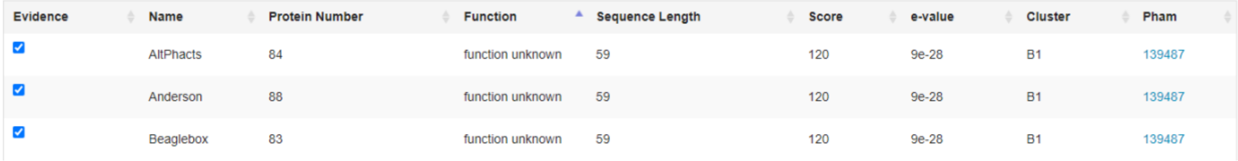
5. Function. Please follow this [Official SEAPHAGE Function List]

If no functional prediction is present, write "Hypothetical protein".

Hypothetical Protein

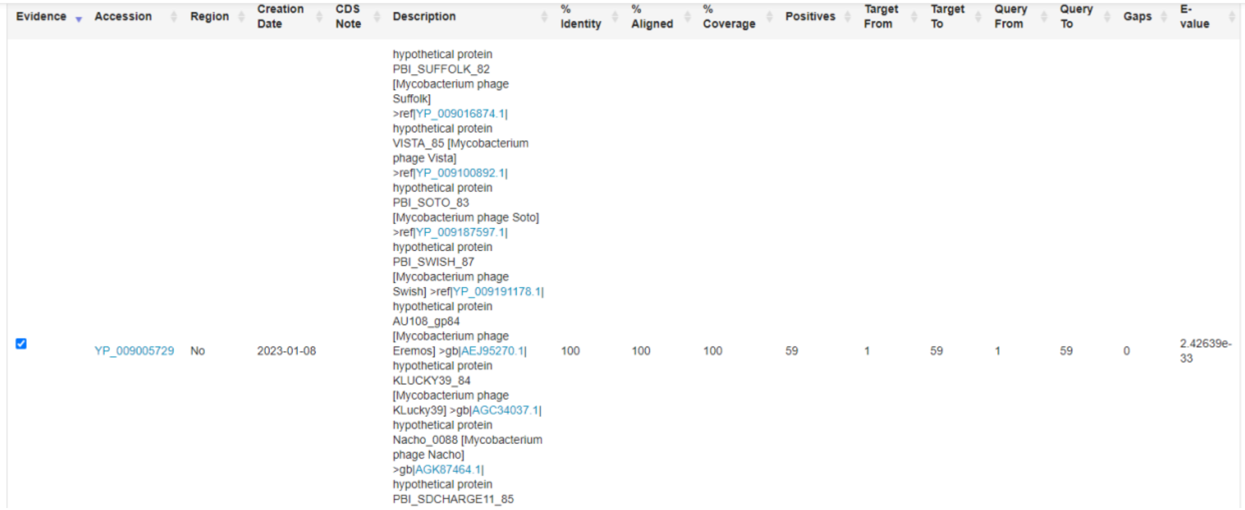
6. Supporting Information for Function (SIF)

a. SIF: PhageDb BLAST



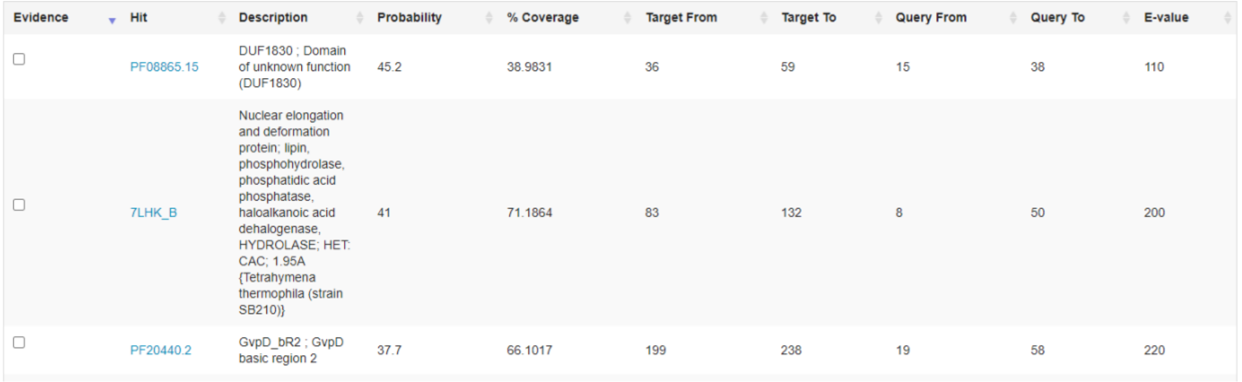
| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | AltPhacts | 84 | function unknown | 59 | 120 | 9e-28 | B1 | 139487 |
| ☑ | Anderson | 88 | function unknown | 59 | 120 | 9e-28 | B1 | 139487 |
| ☑ | Beaglebox | 83 | function unknown | 59 | 120 | 9e-28 | B1 | 139487 |

b. SIF: NCBI BLAST



| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | YP_009005729 | No | 2023-01-08 |  | hypothetical protein PBI_SUFFOLK_82 [Mycobacterium phage Suffolk] >ref\|YP_009016874.1\| hypothetical protein VISTA_85 [Mycobacterium phage Vista] >ref\|YP_009100892.1\| hypothetical protein PBI_SOTO_83 [Mycobacterium phage Soto] >ref\|YP_009187597.1\| hypothetical protein PBI_SWISH_87 [Mycobacterium phage Swish] >ref\|YP_009191178.1\| hypothetical protein AU108_gp84 [Mycobacterium phage Eremos] >gb\|AEJ95270.1\| hypothetical protein KLUCKY39_84 [Mycobacterium phage KLucky39] >gb\|AGC34037.1\| hypothetical protein Nacho_0088 [Mycobacterium phage Nacho] >gb\|AGK87464.1\| hypothetical protein PBI_SDCHARGE11_85 | 100 | 100 | 100 | 59 | 1 | 59 | 1 | 59 | 0 | 2.42639e-33 |

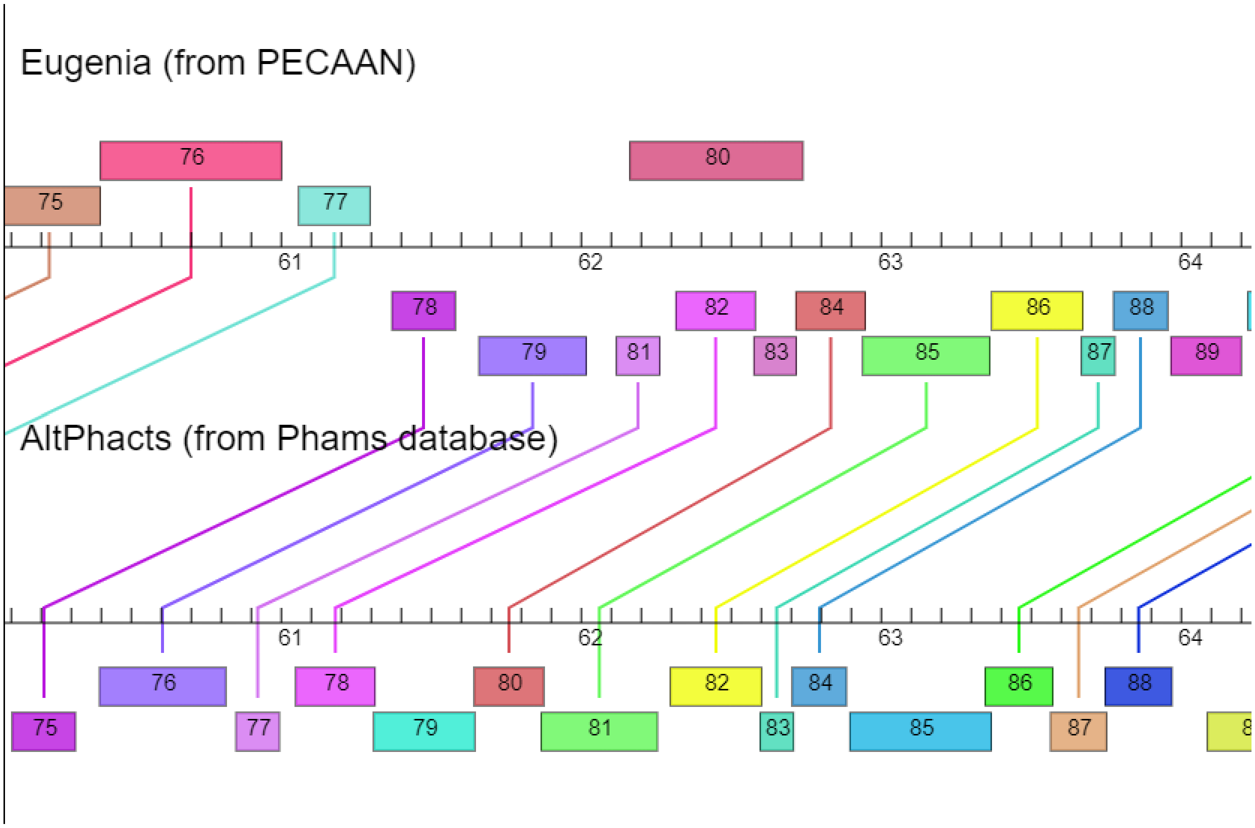
## c. SIF: HHPred

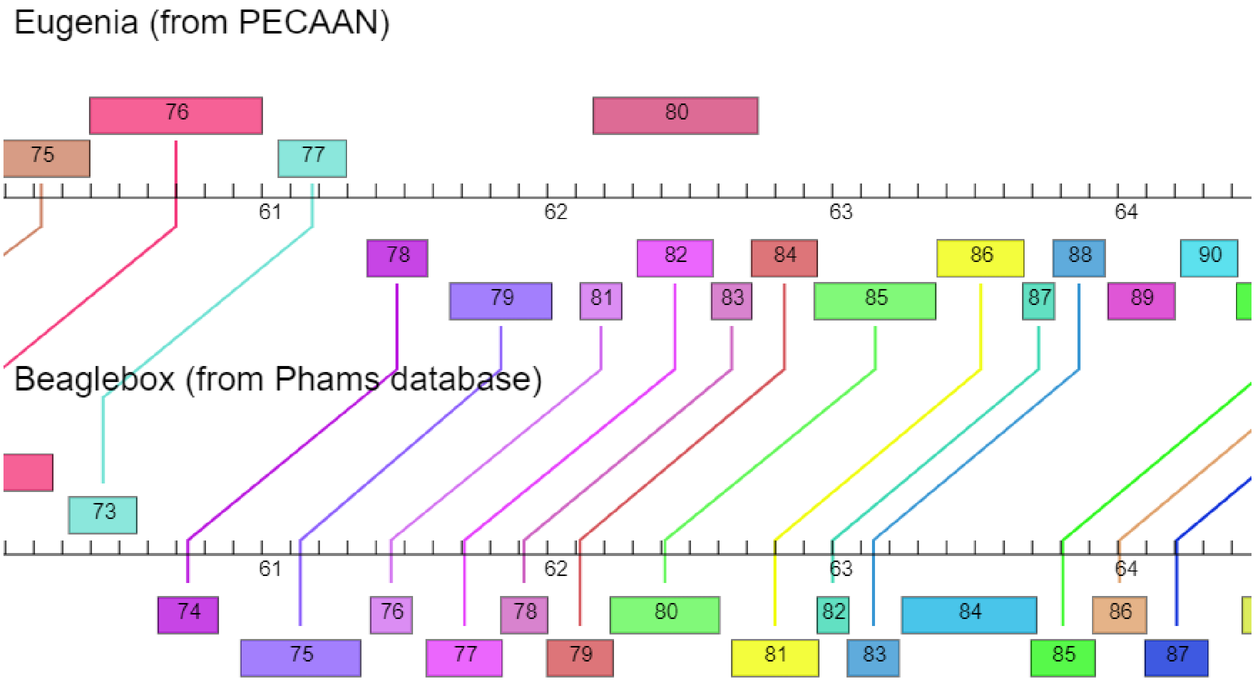


| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☐ | PF08865.15 | DUF1830 ; Domain of unknown function (DUF1830) | 45.2 | 38.9831 | 36 | 59 | 15 | 38 | 110 |
| ☐ | 7LHK_B | Nuclear elongation and deformation protein; lipin, phosphohydrolase, phosphatidic acid phosphatase, haloalkanoic acid dehalogenase, HYDROLASE; HET: CAC; 1.95A {Tetrahymena thermophila (strain SB210)} | 41 | 71.1864 | 83 | 132 | 8 | 50 | 200 |
| ☐ | PF20440.2 | GvpD_bR2 ; GvpD basic region 2 | 37.7 | 66.1017 | 199 | 238 | 19 | 58 | 220 |

## d. SIF: Synteny-Phamerator (three genomes)

Shows synteny with AltPhacts, Anderson, Beaglebox etc.


Eugenia (from PECAAN)
76
80
75
77
61
62
63
64
78
82
84
86
88
79
81
83
85
87
89
AltPhacts (from Phams database)
61
62
63
64
76
78
80
82
84
86
88
75
77
79
81
83
85
87


Eugenia (from PECAAN)
76
80
75
77
61
62
63
64
78
82
84
86
88
90
79
81
83
85
87
89
Beaglebox (from Phams database)
73
61
62
63
64
74
76
78
80
82
84
86
75
77
79
81
83
85
87

## 7. Any other important information.

Transmembrane prediction

```
# WEBSEQUENCE Length: 59

# WEBSEQUENCE Number of predicted TMHs:  0

# WEBSEQUENCE Exp number of AAs in TMHs: 0

# WEBSEQUENCE Exp number, first 60 AAs:  0

# WEBSEQUENCE Total prob of N-in:        0.35100

WEBSEQUENCE      TMHMM2.0   outside     1    59
```

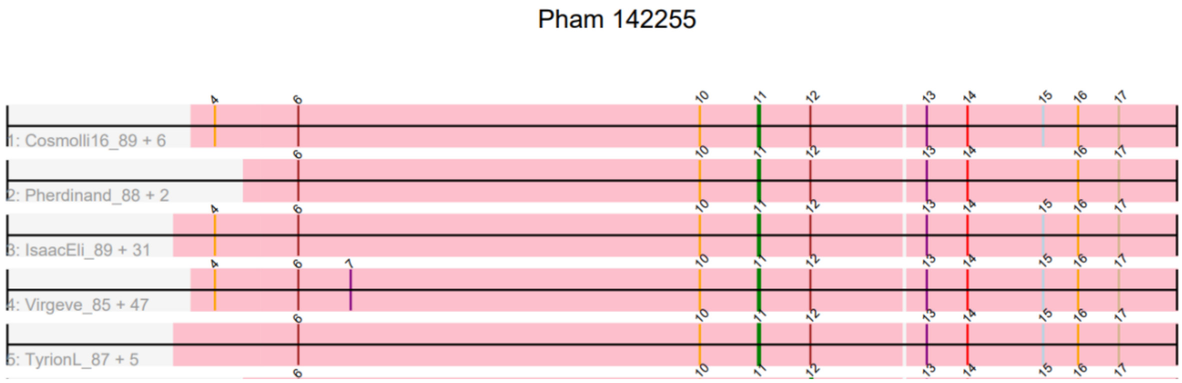
## CURATOR NAME: JOHN MOTTER

## GENE NAME: EUGENIA GENE 90

## DNA MASTER NOTES: N/A

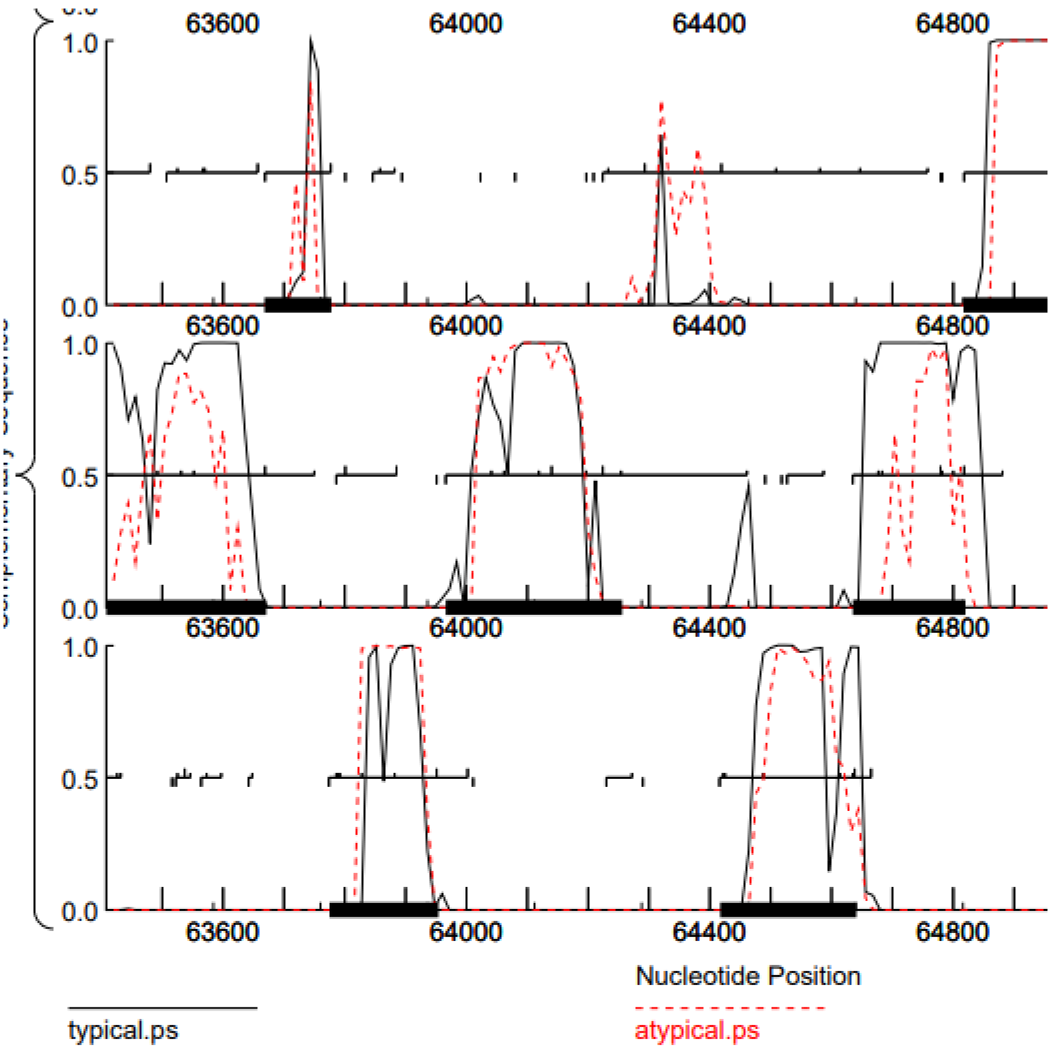
## START POSITION EVALUATION (IN ORDER):

### 1. Starterator



The start number called the most often in the published annotations is 11, it was called in 193 of the 215 non-draft genes in the pham.

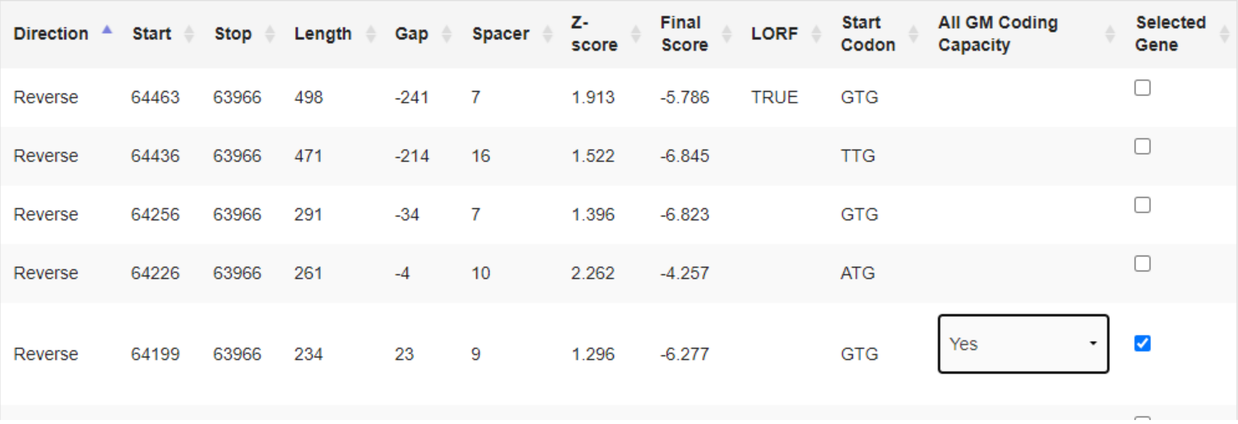
### 2. GeneMark coding potential



## 3. Glimmer and GeneMark agreement

| Glimmer Start: | Glimmer Score: | GeneMark Start: | Pham |
|---|---|---|---|
| 64199 | 11.97 | 64199 | Starterator: 142255<br>suggested start (SS)<br>PhagesDB: 142255 |

## 4. Longest open reading frame (ORF) without excessive gap



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 64463 | 63966 | 498 | -241 | 7 | 1.913 | -5.786 | TRUE | GTG | | ☐ |
| Reverse | 64436 | 63966 | 471 | -214 | 16 | 1.522 | -6.845 | | TTG | | ☐ |
| Reverse | 64256 | 63966 | 291 | -34 | 7 | 1.396 | -6.823 | | GTG | | ☐ |
| Reverse | 64226 | 63966 | 261 | -4 | 10 | 2.262 | -4.257 | | ATG | | ☐ |
| Reverse | 64199 | 63966 | 234 | 23 | 9 | 1.296 | -6.277 | | GTG | Yes | ☑ |

Although longer open reading frames without excessive gaps exist, these genes do not have synteny with similar phages and yield no results in NCBI BLAS, suggesting that the Starterator and Glimmer predicted start sites should be used.

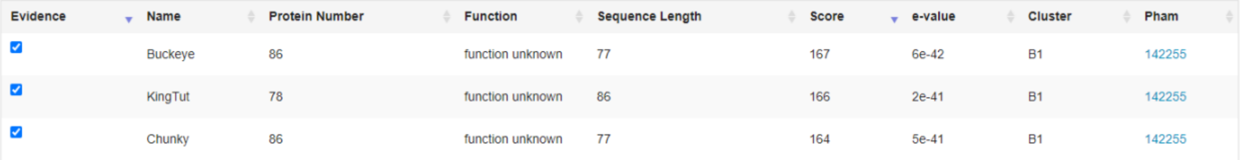
## 5. Function. Please follow this [Official SEAPHAGE Function List]()

If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

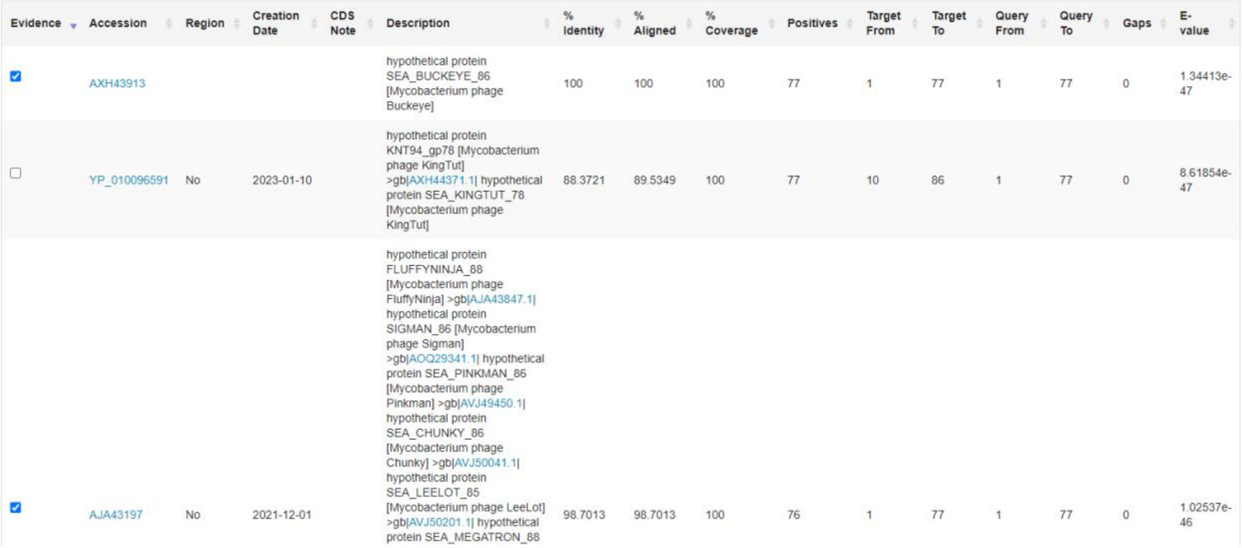
## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST



| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | Buckeye | 86 | function unknown | 77 | 167 | 6e-42 | B1 | 142255 |
| ☑ | KingTut | 78 | function unknown | 86 | 166 | 2e-41 | B1 | 142255 |
| ☑ | Chunky | 86 | function unknown | 77 | 164 | 5e-41 | B1 | 142255 |

## b. SIF: NCBI BLAST



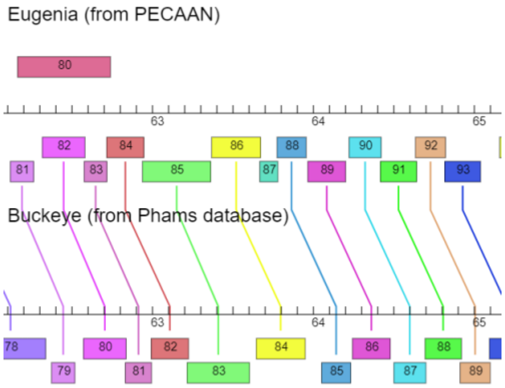
| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | AXH43913 | | | | hypothetical protein SEA_BUCKEYE_86 [Mycobacterium phage Buckeye] | 100 | 100 | 100 | 77 | 1 | 77 | 1 | 77 | 0 | 1.34413e-47 |
| ☐ | YP_010096591 | No | 2023-01-10 | | hypothetical protein KNT94_gp78 [Mycobacterium phage KingTut] >gb\|AXH44371.1\| hypothetical protein SEA_KINGTUT_78 [Mycobacterium phage KingTut] | 88.3721 | 89.5349 | 100 | 77 | 10 | 86 | 1 | 77 | 0 | 8.61854e-47 |
| ☑ | AJA43197 | No | 2021-12-01 | | hypothetical protein FLUFFYNINJA_88 [Mycobacterium phage FluffyNinja] >gb\|AJA43847.1\| hypothetical protein SIGMAN_86 [Mycobacterium phage Sigman] >gb\|AOQ29341.1\| hypothetical protein SEA_PINKMAN_86 [Mycobacterium phage Pinkman] >gb\|AVJ49450.1\| hypothetical protein SEA_CHUNKY_86 [Mycobacterium phage Chunky] >gb\|AVJ50041.1\| hypothetical protein SEA_LEELOT_85 [Mycobacterium phage LeeLot] >gb\|AVJ50201.1\| hypothetical protein SEA_MEGATRON_88 | 98.7013 | 98.7013 | 100 | 76 | 1 | 77 | 1 | 77 | 0 | 1.02537e-46 |

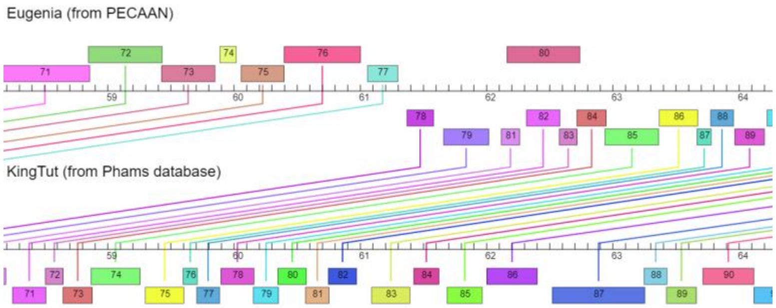
## c. SIF: HHPred

No informative results.

### d. SIF: Synteny-Phamerator (three genomes)

Shows synteny with Buckeye, KingTut, Chunky, etc.





## 7. Any other important information.

Transmembrane prediction

```
# WEBSEQUENCE Length: 77
# WEBSEQUENCE Number of predicted TMHs:  0
# WEBSEQUENCE Exp number of AAs in TMHs: 0.00022
# WEBSEQUENCE Exp number, first 60 AAs:  0.00022
# WEBSEQUENCE Total prob of N-in:        0.31052
WEBSEQUENCE      TMHMM2.0   outside      1    77
```

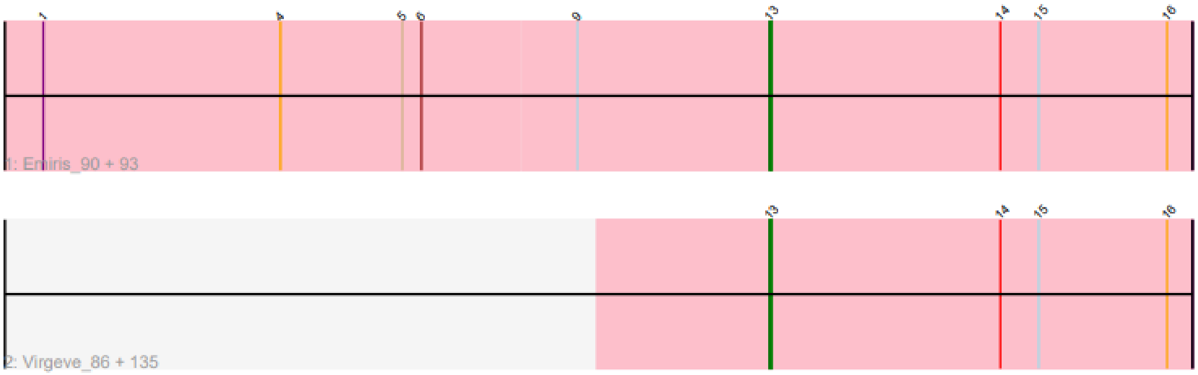
## CURATOR NAME: JOHN MOTTER

## GENE NAME: EUGENIA GENE 90

## DNA MASTER NOTES: N/A

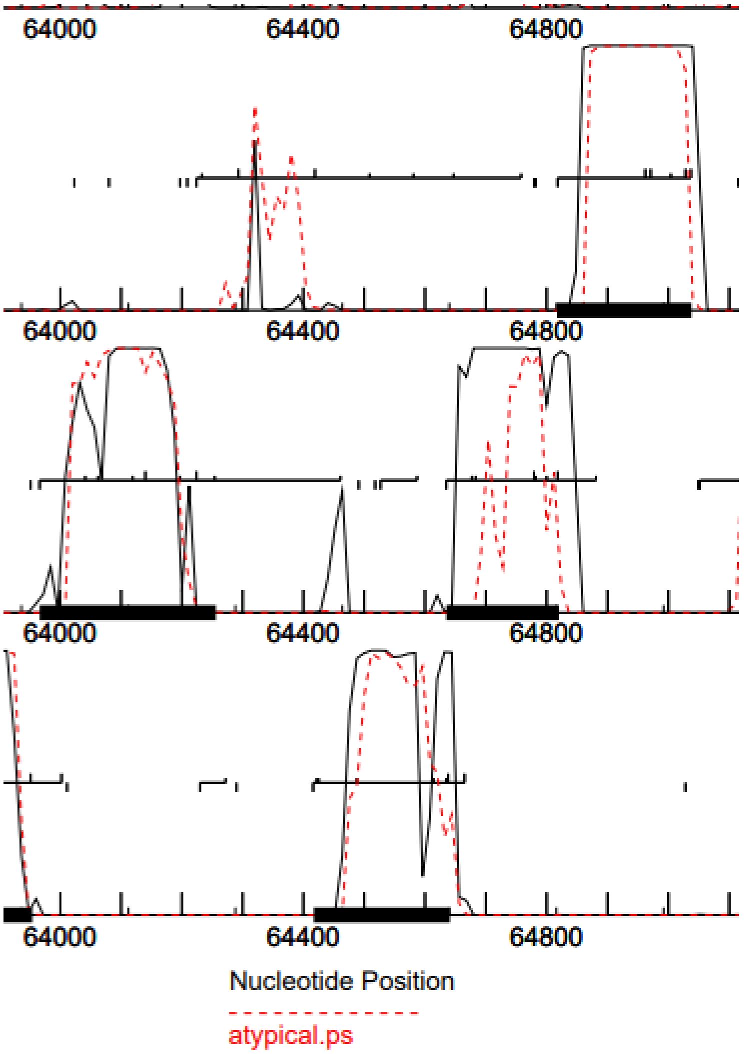
## START POSITION EVALUATION (IN ORDER):

### 1. Starterator



The start number called the most often in the published annotations is 13, it was called in 213 of the 213 non-draft genes in the pham.

### 2. GeneMark coding potential



## 3. Glimmer and GeneMark agreement

Only Glimmer called this gene, predicting that it would start at 64420.

## 4. Longest open reading frame (ORF) without excessive gap

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 64582 | 64223 | 360 | -166 | 15 | 1.558 | -6.577 | GTG | | ☐ |
| Reverse | 64510 | 64223 | 288 | -94 | 7 | 1.638 | -6.338 | GTG | | ☐ |
| Reverse | 64420 | 64223 | 198 | -4 | 6 | 2.113 | -5.605 | ATG | Yes | ☑ |
| Reverse | 64312 | 64223 | 90 | 104 | 11 | 1.362 | -6.126 | TTG | | ☐ |

5. Function. Please follow this [Official SEAPHAGE Function List]

If no functional prediction is present, write “Hypothetical protein”.

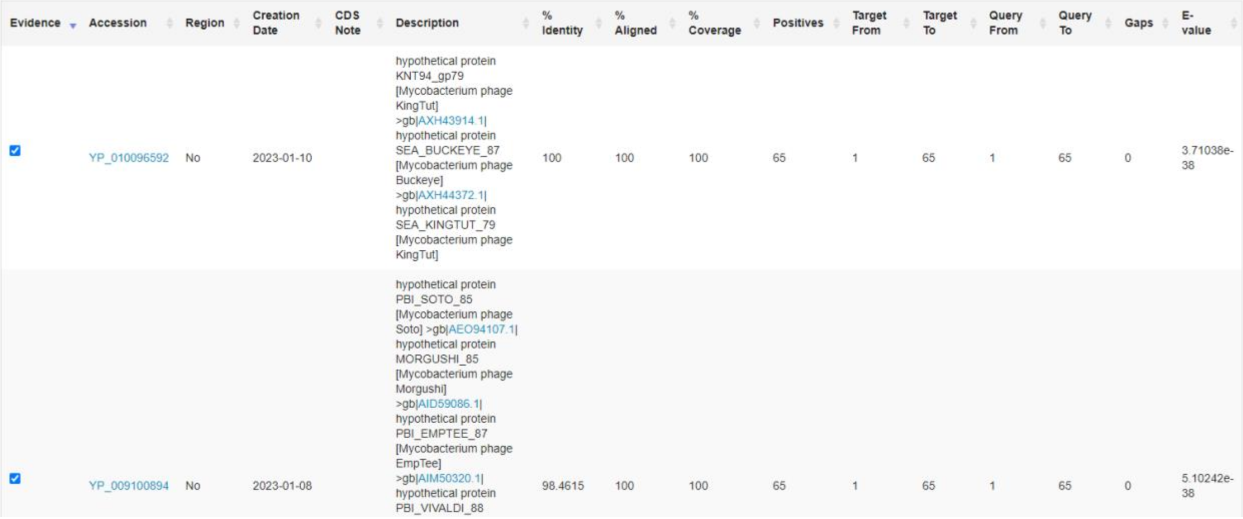
Hypothetical Protein

6. Supporting Information for Function (SIF)

a. SIF: PhageDb BLAST

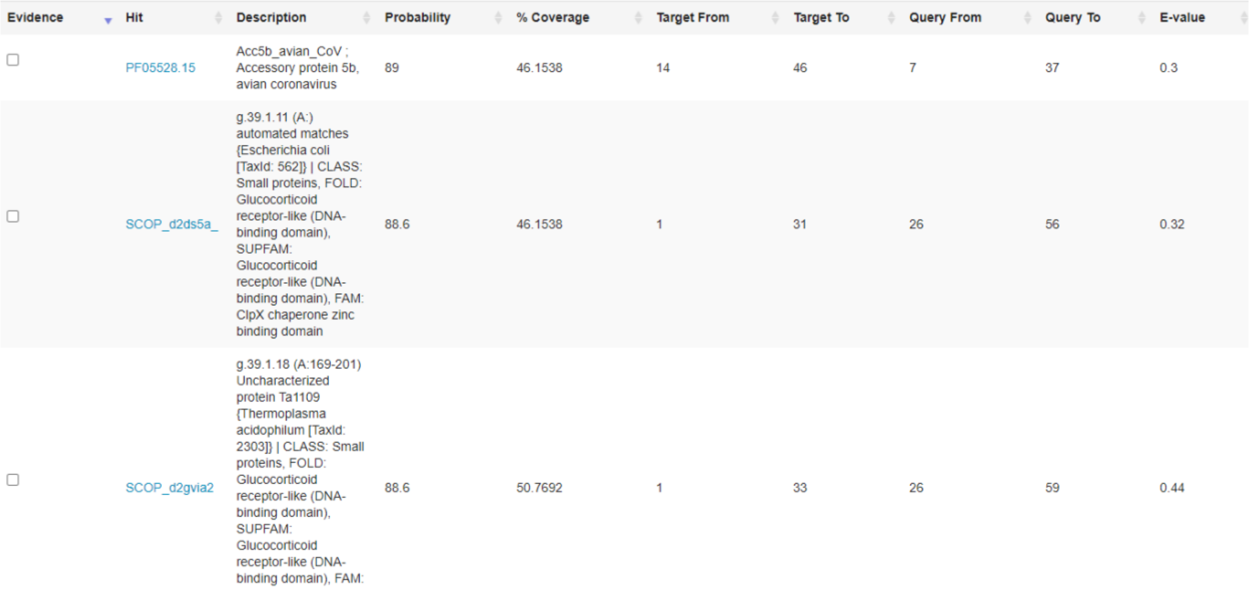
| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ☑ | Buckeye | 87 | function unknown | 65 | 139 | 3e-33 | B1 | 302 |
| ☐ | Eugenia_Draft | 90 | function unknown | 65 | 139 | 3e-33 | B1 | 302 |
| ☑ | KingTut | 79 | function unknown | 65 | 139 | 3e-33 | B1 | 302 |
| ☑ | Antonia | 87 | function unknown | 65 | 138 | 4e-33 | B1 | 302 |

b. SIF: NCBI BLAST



| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | YP_010096592 | No | 2023-01-10 | | hypothetical protein KNT94_gp79 [Mycobacterium phage KingTut] >gb\|AXH43914.1\| hypothetical protein SEA_BUCKEYE_87 [Mycobacterium phage Buckeye] >gb\|AXH44372.1\| hypothetical protein SEA_KINGTUT_79 [Mycobacterium phage KingTut] | 100 | 100 | 100 | 65 | 1 | 65 | 1 | 65 | 0 | 3.71038e-38 |
| ☑ | YP_009100894 | No | 2023-01-08 | | hypothetical protein PBI_SOTO_85 [Mycobacterium phage Soto] >gb\|AEO94107.1\| hypothetical protein MORGUSHI_85 [Mycobacterium phage Morgushi] >gb\|AID59086.1\| hypothetical protein PBI_EMPTEE_87 [Mycobacterium phage EmpTee] >gb\|AIM50320.1\| hypothetical protein PBI_VIVALDI_88 | 98.4615 | 100 | 100 | 65 | 1 | 65 | 1 | 65 | 0 | 5.10242e-38 |

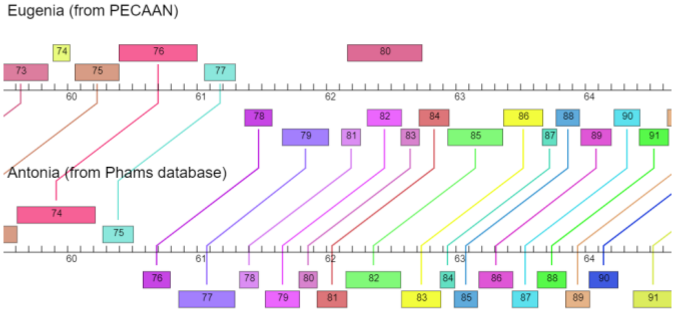
### c. SIF: HHPred

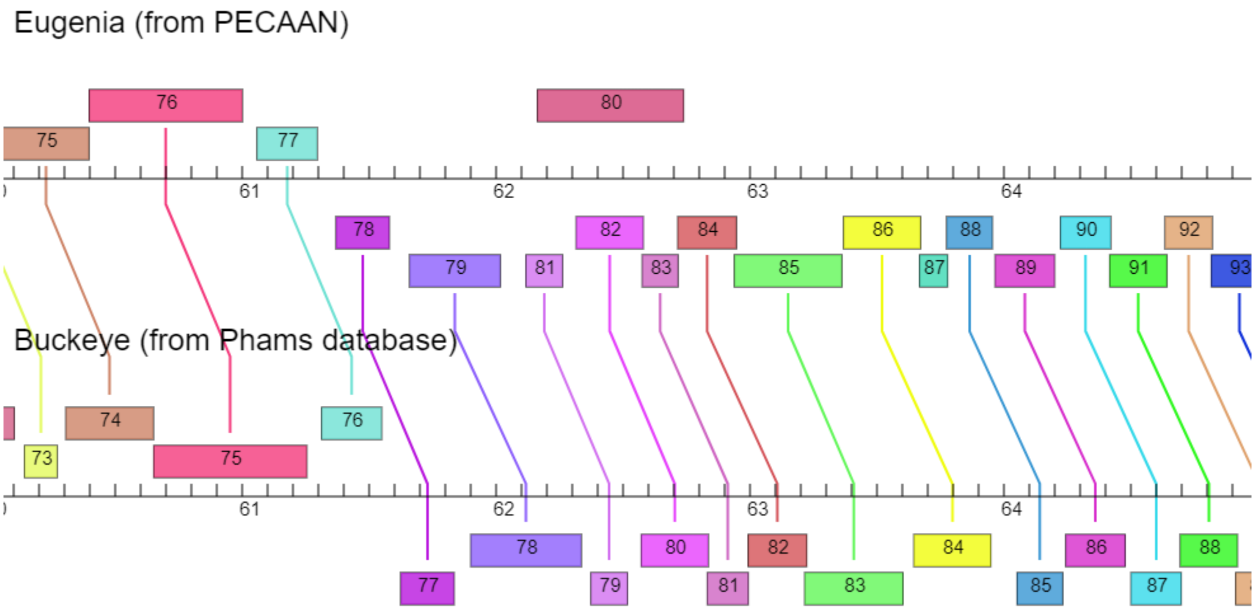


| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☐ | PF05528.15 | Acc5b_avian_CoV ; Accessory protein 5b, avian coronavirus | 89 | 46.1538 | 14 | 46 | 7 | 37 | 0.3 |
| ☐ | SCOP_d2ds5a_ | g.39.1.11 (A:) automated matches {Escherichia coli [TaxId: 562]} \| CLASS: Small proteins, FOLD: Glucocorticoid receptor-like (DNA-binding domain), SUPFAM: Glucocorticoid receptor-like (DNA-binding domain), FAM: ClpX chaperone zinc binding domain | 88.6 | 46.1538 | 1 | 31 | 26 | 56 | 0.32 |
| ☐ | SCOP_d2gvia2 | g.39.1.18 (A:169-201) Uncharacterized protein Ta1109 {Thermoplasma acidophilum [TaxId: 2303]} \| CLASS: Small proteins, FOLD: Glucocorticoid receptor-like (DNA-binding domain), SUPFAM: Glucocorticoid receptor-like (DNA-binding domain), FAM: | 88.6 | 50.7692 | 1 | 33 | 26 | 59 | 0.44 |

### d. SIF: Synteny-Phamerator (three genomes)

Shows synteny with Buckeye, KingTut, Antonia, etc.





## 7. Any other important information.

### Transmembrane Prediction

```
# WEBSEQUENCE Length: 65
# WEBSEQUENCE Number of predicted TMHs:  0
# WEBSEQUENCE Exp number of AAs in TMHs: 0.00017
# WEBSEQUENCE Exp number, first 60 AAs:  0.00016
# WEBSEQUENCE Total prob of N-in:        0.66623
WEBSEQUENCE      TMHMM2.0   inside       1    65
```

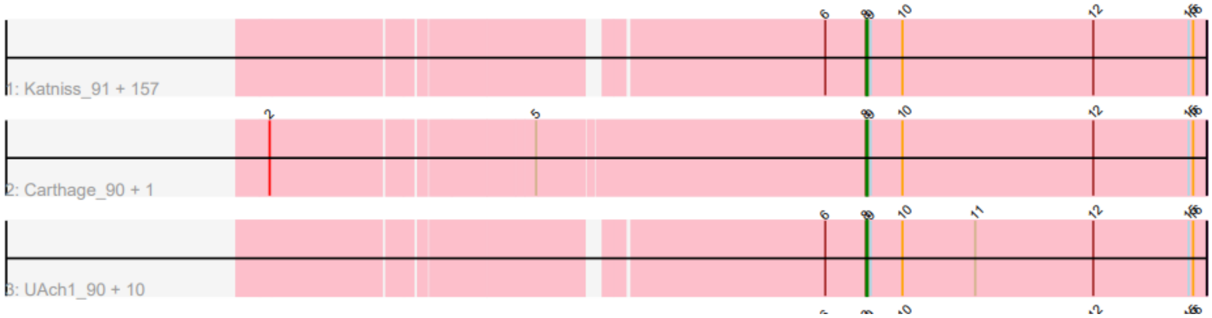
## CURATOR NAME: JOHN MOTTER

## GENE NAME: EUGENIA GENE 91

## DNA MASTER NOTES: N/A

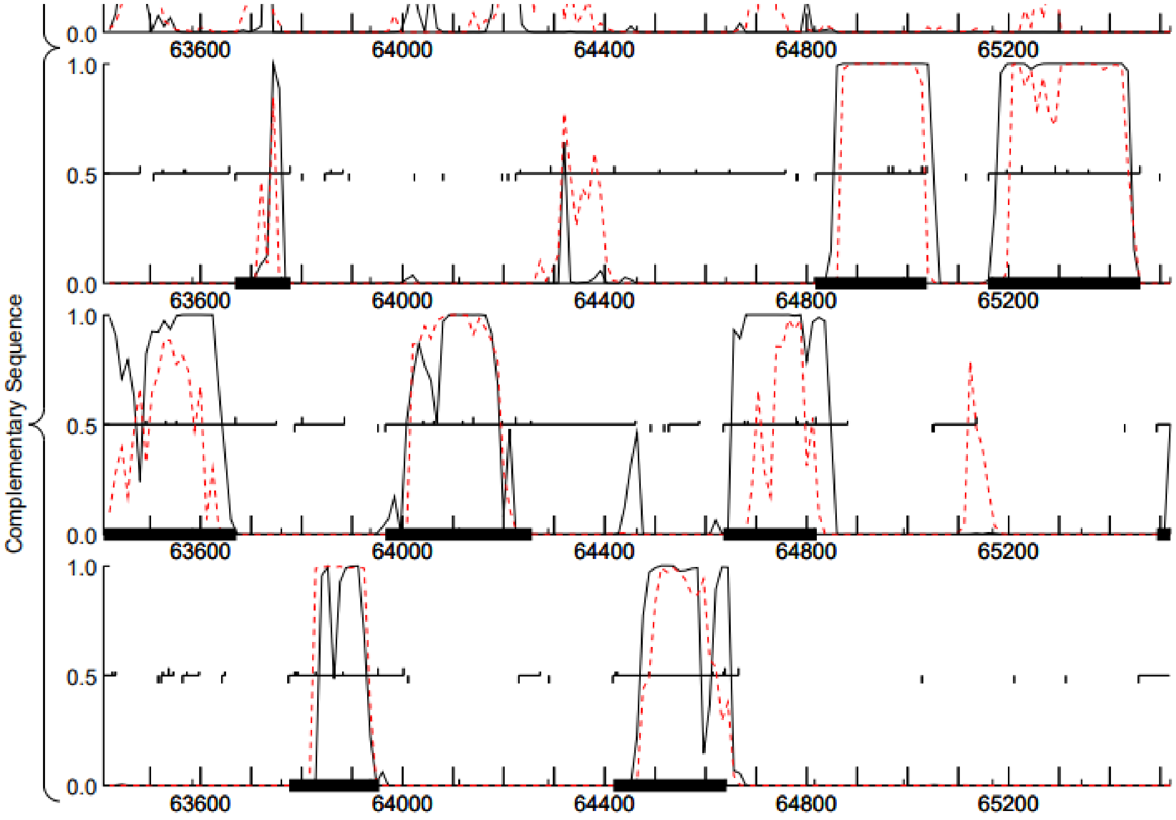
## START POSITION EVALUATION (IN ORDER):

1. Starterator

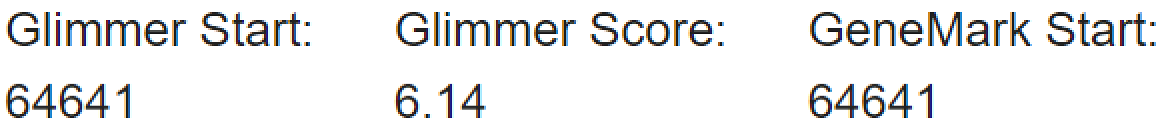


The start number called the most often in the published annotations is 8, it was called in 155 of the 243 non-draft genes in the pham.

2. GeneMark coding potential



3. Glimmer and GeneMark agreement



| Glimmer Start: | Glimmer Score: | GeneMark Start: |
|---|---|---|
| 64641 | 6.14 | 64641 |

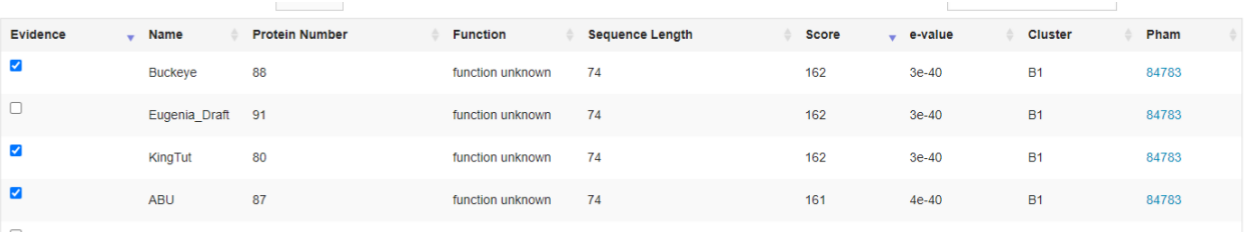
4. Longest open reading frame (ORF) without excessive gap

5. Function. Please follow this [Official SEAPHAGE Function List]

If no functional prediction is present, write “Hypothetical protein”.

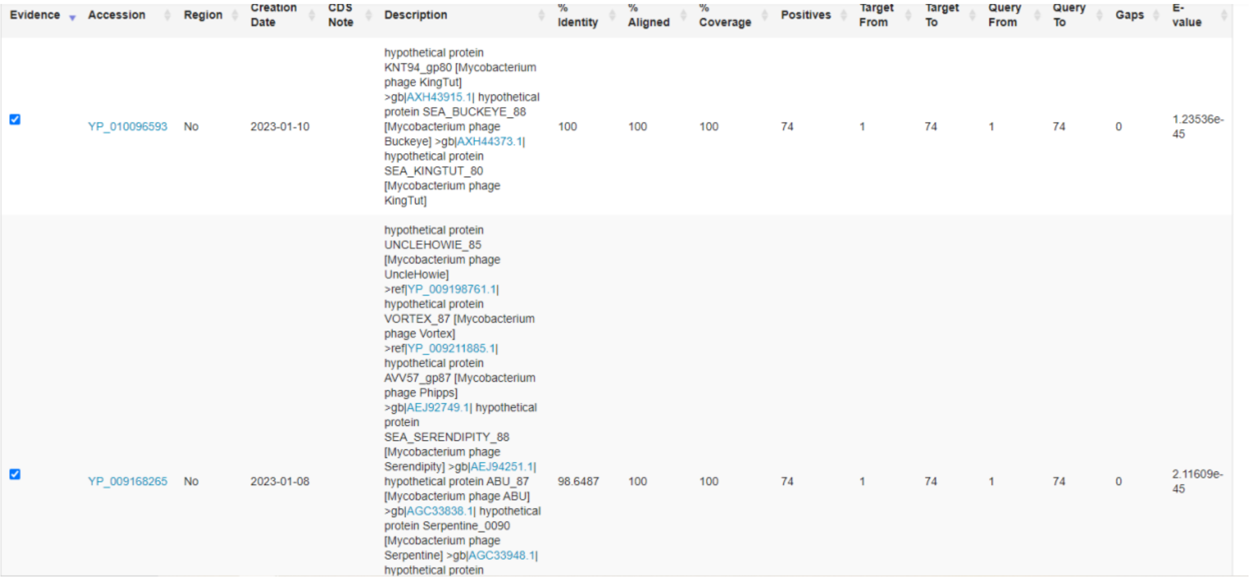
# 6. Supporting Information for Function (SIF)

## a. SIF: PhageDb BLAST



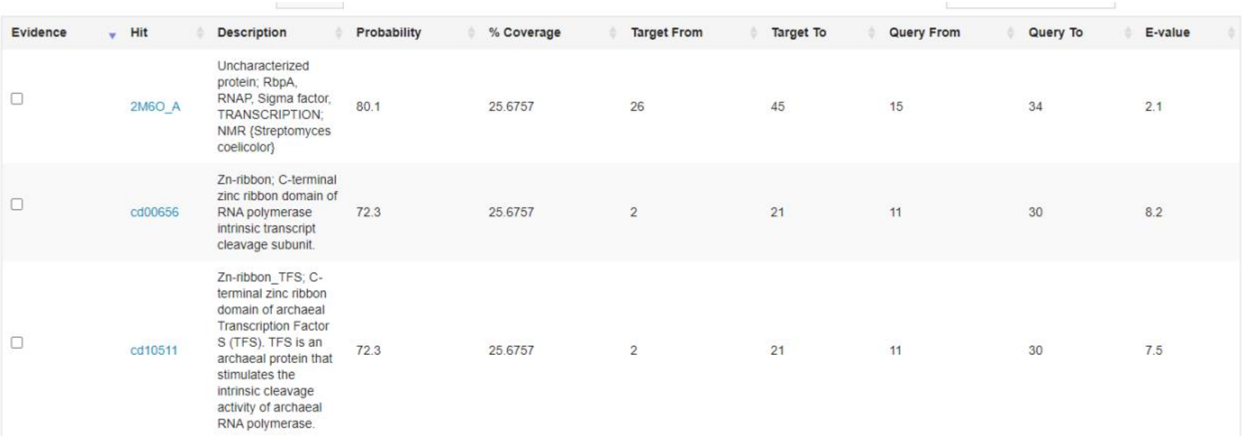
| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | Buckeye | 88 | function unknown | 74 | 162 | 3e-40 | B1 | 84783 |
| ☐ | Eugenia_Draft | 91 | function unknown | 74 | 162 | 3e-40 | B1 | 84783 |
| ☑ | KingTut | 80 | function unknown | 74 | 162 | 3e-40 | B1 | 84783 |
| ☑ | ABU | 87 | function unknown | 74 | 161 | 4e-40 | B1 | 84783 |

## b. SIF: NCBI BLAST



| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | YP_010096593 | No | 2023-01-10 | | hypothetical protein KNT94_gp80 [Mycobacterium phage KingTut] >gb\|AXH43915.1\| hypothetical protein SEA_BUCKEYE_88 [Mycobacterium phage Buckeye] >gb\|AXH44373.1\| hypothetical protein SEA_KINGTUT_80 [Mycobacterium phage KingTut] | 100 | 100 | 100 | 74 | 1 | 74 | 1 | 74 | 0 | 1.23536e-45 |
| ☑ | YP_009168265 | No | 2023-01-08 | | hypothetical protein UNCLEHOWIE_85 [Mycobacterium phage UncleHowie] >ref\|YP_009198761.1\| hypothetical protein VORTEX_87 [Mycobacterium phage Vortex] >ref\|YP_009211885.1\| hypothetical protein AVV57_gp87 [Mycobacterium phage Phipps] >gb\|AEJ92749.1\| hypothetical protein SEA_SERENDIPITY_88 [Mycobacterium phage Serendipity] >gb\|AEJ94251.1\| hypothetical protein ABU_87 [Mycobacterium phage ABU] >gb\|AGC33838.1\| hypothetical protein Serpentine_0090 [Mycobacterium phage Serpentine] >gb\|AGC33948.1\| hypothetical protein | 98.6487 | 100 | 100 | 74 | 1 | 74 | 1 | 74 | 0 | 2.11609e-45 |

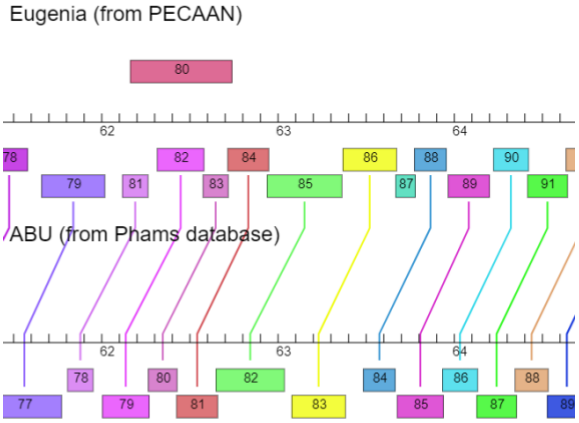
## c. SIF: HHPred

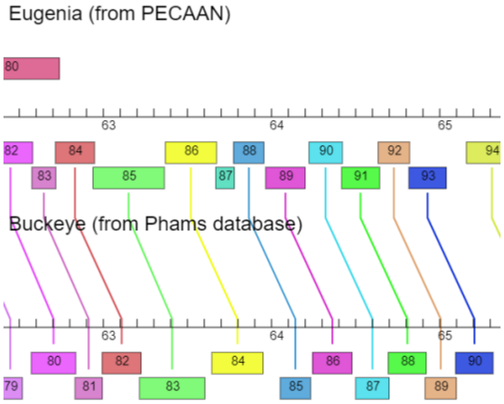


| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☐ | 2M6O_A | Uncharacterized protein; RbpA, RNAP, Sigma factor, TRANSCRIPTION; NMR {Streptomyces coelicolor} | 80.1 | 25.6757 | 26 | 45 | 15 | 34 | 2.1 |
| ☐ | cd00656 | Zn-ribbon; C-terminal zinc ribbon domain of RNA polymerase intrinsic transcript cleavage subunit. | 72.3 | 25.6757 | 2 | 21 | 11 | 30 | 8.2 |
| ☐ | cd10511 | Zn-ribbon_TFS; C-terminal zinc ribbon domain of archaeal Transcription Factor S (TFS). TFS is an archaeal protein that stimulates the intrinsic cleavage activity of archaeal RNA polymerase. | 72.3 | 25.6757 | 2 | 21 | 11 | 30 | 7.5 |

### d. SIF: Synteny-Phamerator (three genomes)

Shows synteny with Buckeye, KingTut, ABU, etc.





## 7. Any other important information.

Transmembrane prediction

```
# WEBSEQUENCE Length: 74
# WEBSEQUENCE Number of predicted TMHs:  0
# WEBSEQUENCE Exp number of AAs in TMHs: 0.00241
# WEBSEQUENCE Exp number, first 60 AAs:  0.00202
# WEBSEQUENCE Total prob of N-in:        0.21232
WEBSEQUENCE      TMHMM2.0   outside      1    74
```

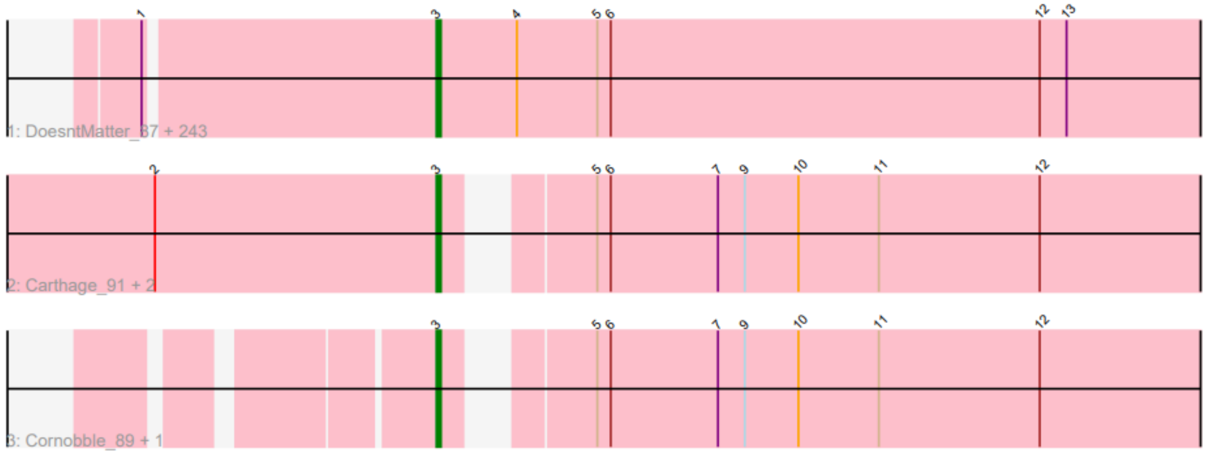
## CURATOR NAME: JOHN MOTTER

## GENE NAME: EUGENIA GENE 92

## DNA MASTER NOTES: N/A

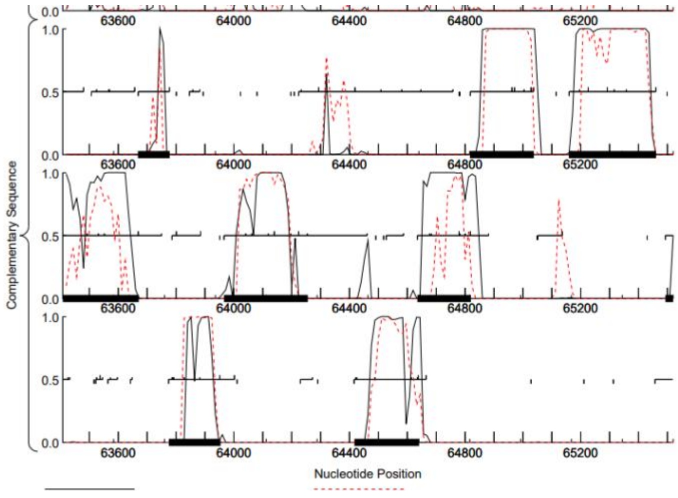
## START POSITION EVALUATION (IN ORDER):

### 1. Starterator

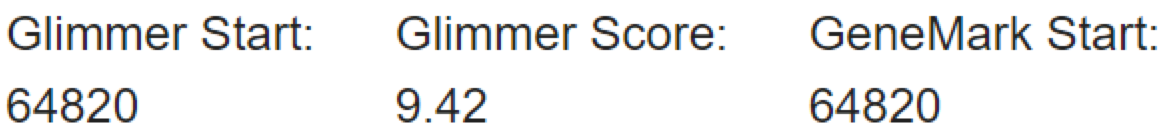


The start number called the most often in the published annotations is 3, it was called in 245 of the 246 non-draft genes in the pham.

### 2. GeneMark coding potential

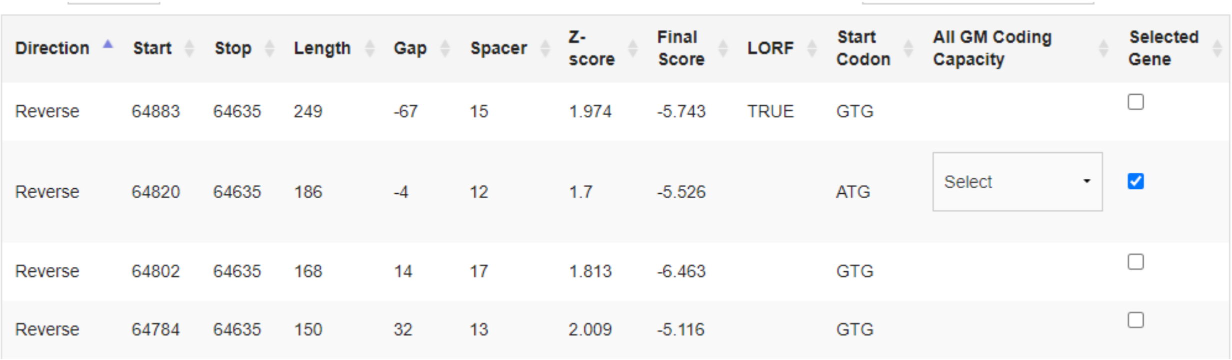


3. Glimmer and GeneMark agreement



| Glimmer Start: | Glimmer Score: | GeneMark Start: |
|---|---|---|
| 64820 | 9.42 | 64820 |

4. Longest open reading frame (ORF) without excessive gap



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 64883 | 64635 | 249 | -67 | 15 | 1.974 | -5.743 | TRUE | GTG |  | ☐ |
| Reverse | 64820 | 64635 | 186 | -4 | 12 | 1.7 | -5.526 |  | ATG | Select | ☑ |
| Reverse | 64802 | 64635 | 168 | 14 | 17 | 1.813 | -6.463 |  | GTG |  | ☐ |
| Reverse | 64784 | 64635 | 150 | 32 | 13 | 2.009 | -5.116 |  | GTG |  | ☐ |

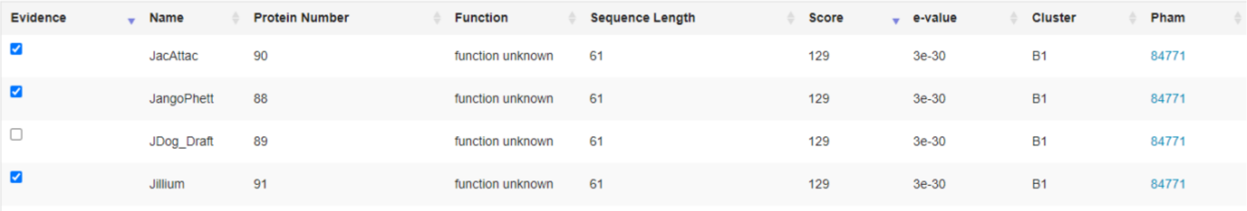
5. Function. Please follow this Official SEAPHAGE Function List

If no functional prediction is present, write “Hypothetical protein”.

Hypothetical Protein

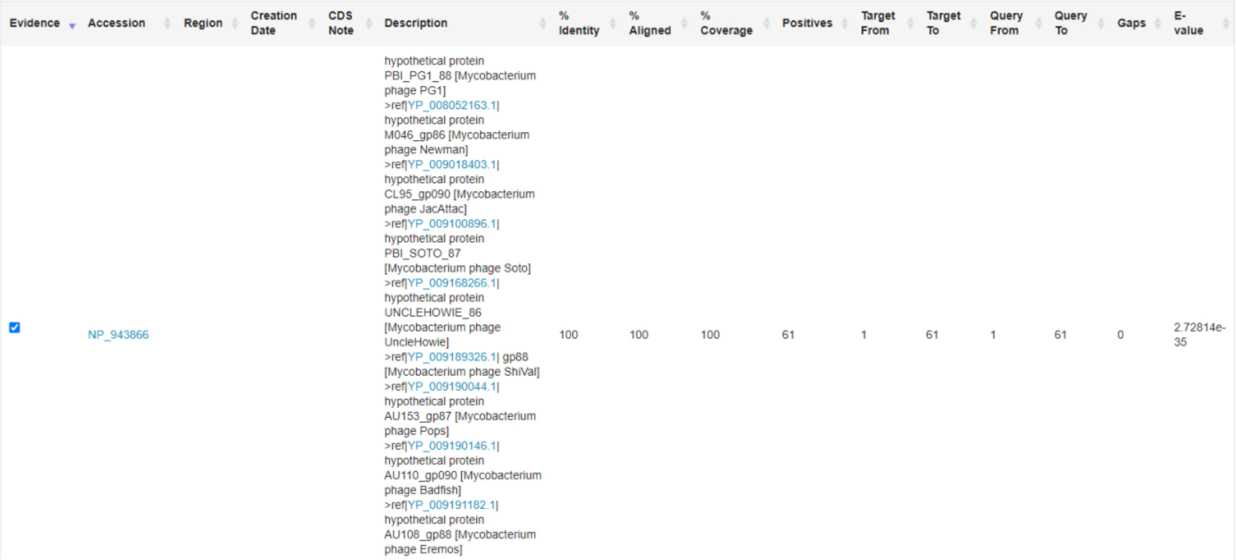
6. Supporting Information for Function (SIF)

a. SIF: PhageDb BLAST



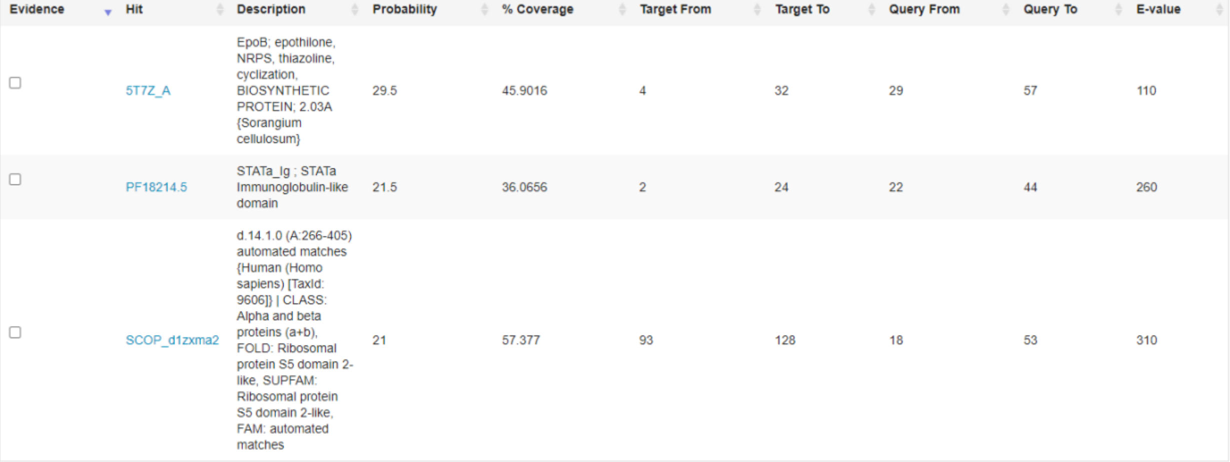
| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | JacAttac | 90 | function unknown | 61 | 129 | 3e-30 | B1 | 84771 |
| ☑ | JangoPhett | 88 | function unknown | 61 | 129 | 3e-30 | B1 | 84771 |
| ☐ | JDog_Draft | 89 | function unknown | 61 | 129 | 3e-30 | B1 | 84771 |
| ☑ | Jillium | 91 | function unknown | 61 | 129 | 3e-30 | B1 | 84771 |

## b. SIF: NCBI BLAST



| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | NP_943866 | | | | hypothetical protein PBI_PG1_88 [Mycobacterium phage PG1] >ref\|YP_008052163.1\| hypothetical protein M046_gp86 [Mycobacterium phage Newman] >ref\|YP_009018403.1\| hypothetical protein CL95_gp090 [Mycobacterium phage JacAttac] >ref\|YP_009100896.1\| hypothetical protein PBI_SOTO_87 [Mycobacterium phage Soto] >ref\|YP_009168266.1\| hypothetical protein UNCLEHOWIE_86 [Mycobacterium phage UncleHowie] >ref\|YP_009189326.1\| gp88 [Mycobacterium phage ShiVal] >ref\|YP_009190044.1\| hypothetical protein AU153_gp87 [Mycobacterium phage Pops] >ref\|YP_009190146.1\| hypothetical protein AU110_gp090 [Mycobacterium phage Badfish] >ref\|YP_009191182.1\| hypothetical protein AU108_gp88 [Mycobacterium phage Eremos] | 100 | 100 | 100 | 61 | 1 | 61 | 1 | 61 | 0 | 2.72814e-35 |

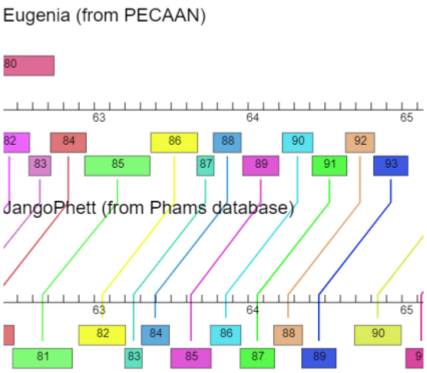
## c. SIF: HHPred

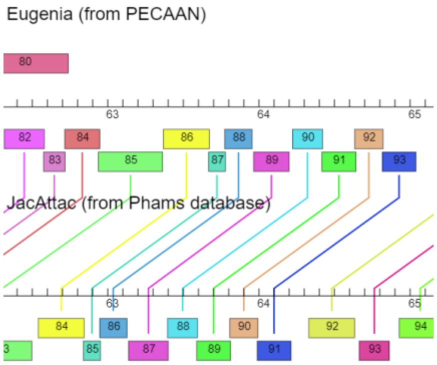


| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☐ | 5T7Z_A | EpoB; epothilone, NRPS, thiazoline, cyclization, BIOSYNTHETIC PROTEIN; 2.03A {Sorangium cellulosum} | 29.5 | 45.9016 | 4 | 32 | 29 | 57 | 110 |
| ☐ | PF18214.5 | STATa_Ig ; STATa Immunoglobulin-like domain | 21.5 | 36.0656 | 2 | 24 | 22 | 44 | 260 |
| ☐ | SCOP_d1zxma2 | d.14.1.0 (A:266-405) automated matches {Human (Homo sapiens) [TaxId: 9606]} \| CLASS: Alpha and beta proteins (a+b), FOLD: Ribosomal protein S5 domain 2-like, SUPFAM: Ribosomal protein S5 domain 2-like, FAM: automated matches | 21 | 57.377 | 93 | 128 | 18 | 53 | 310 |

### d. SIF: Synteny-Phamerator (three genomes)

Shows synteny with JangoPhett, JacAttac, Jillium, etc.





## 7. Any other important information.

Transmembrane prediction

```
# WEBSEQUENCE Length: 61
# WEBSEQUENCE Number of predicted TMHs:  0
# WEBSEQUENCE Exp number of AAs in TMHs: 0.00511
# WEBSEQUENCE Exp number, first 60 AAs:  0.00511
# WEBSEQUENCE Total prob of N-in:        0.14261
WEBSEQUENCE      TMHMM2.0   outside      1    61
```

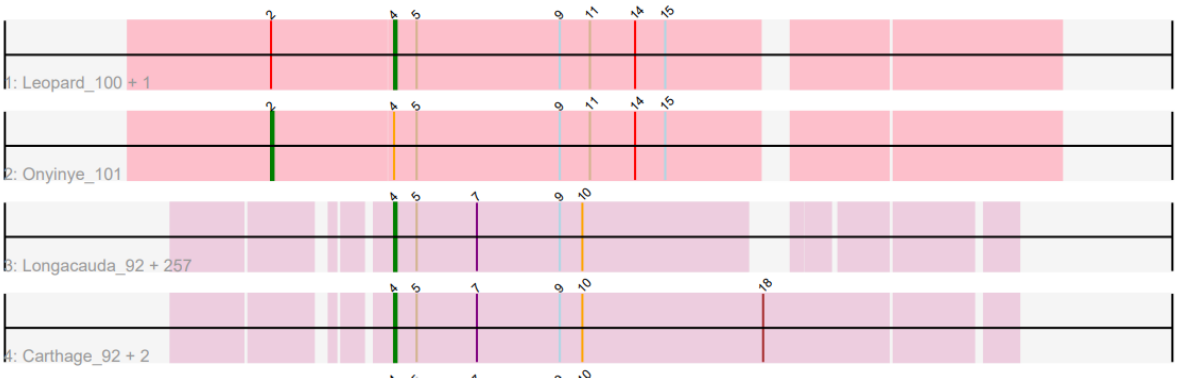
## CURATOR NAME: JOHN MOTTER

## GENE NAME: EUGENIA GENE 93

## DNA MASTER NOTES: N/A

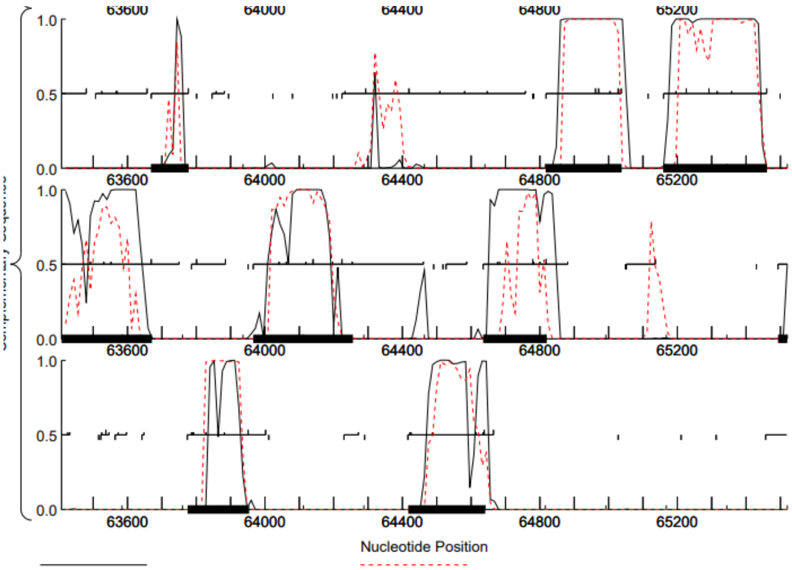
## START POSITION EVALUATION (IN ORDER):

### 1. Starterator

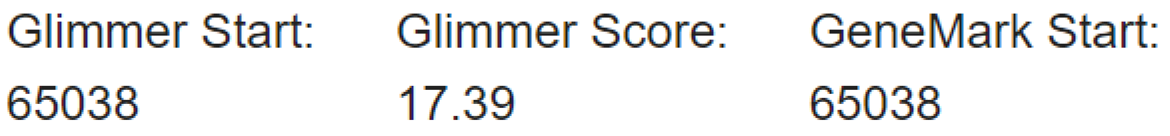


The start number called the most often in the published annotations is 4, it was called in 245 of the 251 non-draft genes in the pham.

### 2. GeneMark coding potential

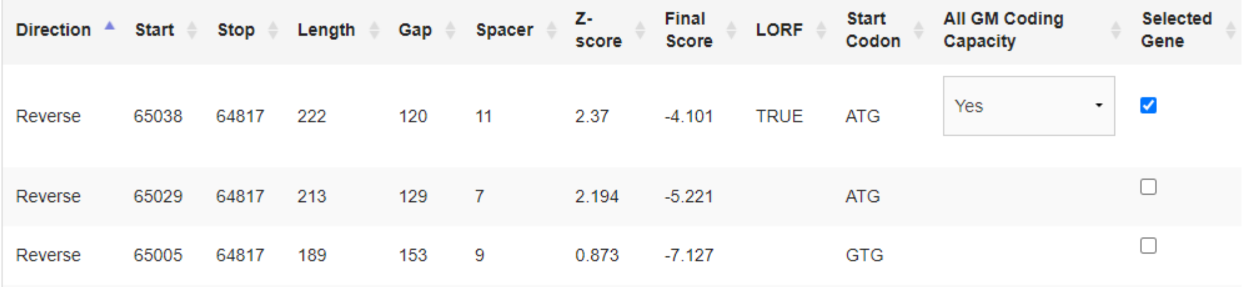


3. Glimmer and GeneMark agreement



| Glimmer Start: | Glimmer Score: | GeneMark Start: |
|---|---|---|
| 65038 | 17.39 | 65038 |

4. Longest open reading frame (ORF) without excessive gap



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 65038 | 64817 | 222 | 120 | 11 | 2.37 | -4.101 | TRUE | ATG | Yes | ☑ |
| Reverse | 65029 | 64817 | 213 | 129 | 7 | 2.194 | -5.221 |  | ATG |  | ☐ |
| Reverse | 65005 | 64817 | 189 | 153 | 9 | 0.873 | -7.127 |  | GTG |  | ☐ |

Gap is present, however a similar gap is present in phages that Eugenia shows synteny with (Buckeye, JacAttac, etc.)

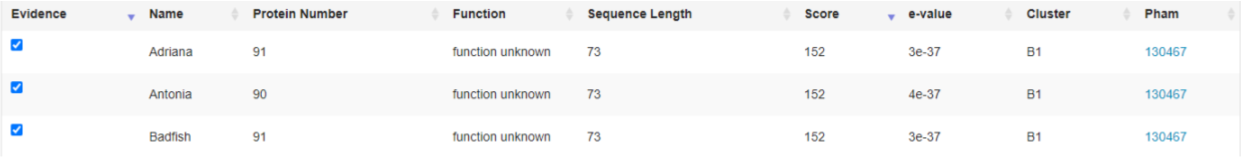
5. Function. Please follow this [Official SEAPHAGE Function List]

If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

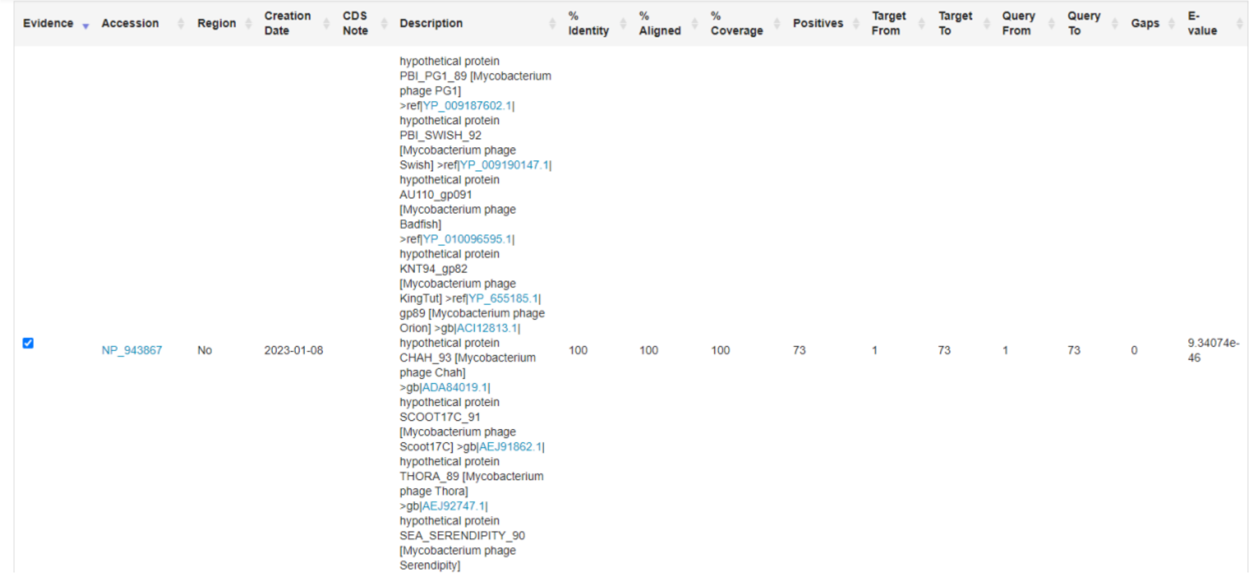
6. Supporting Information for Function (SIF)

a. SIF: PhageDb BLAST



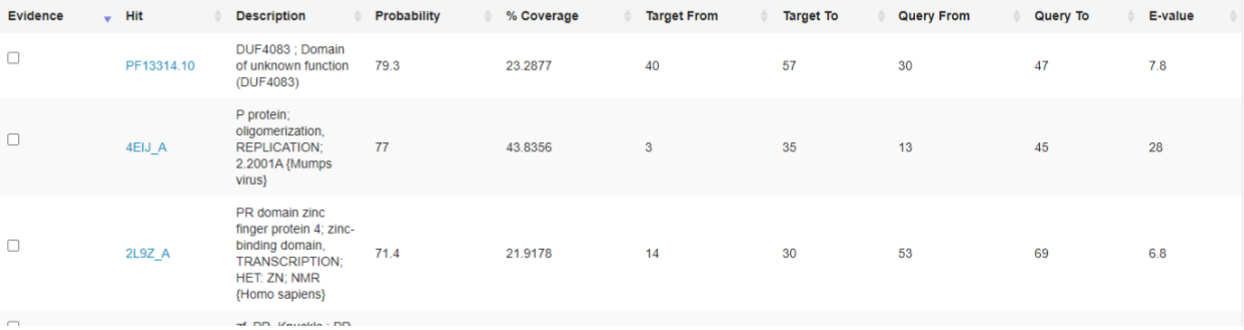
| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | Adriana | 91 | function unknown | 73 | 152 | 3e-37 | B1 | 130467 |
| ☑ | Antonia | 90 | function unknown | 73 | 152 | 4e-37 | B1 | 130467 |
| ☑ | Badfish | 91 | function unknown | 73 | 152 | 3e-37 | B1 | 130467 |

### b. SIF: NCBI BLAST



| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | NP_943867 | No | 2023-01-08 |  | hypothetical protein PBI_PG1_89 [Mycobacterium phage PG1] >ref\|YP_009187602.1\| hypothetical protein PBI_SWISH_92 [Mycobacterium phage Swish] >ref\|YP_009190147.1\| hypothetical protein AU110_gp091 [Mycobacterium phage Badfish] >ref\|YP_010096595.1\| hypothetical protein KNT94_gp82 [Mycobacterium phage KingTut] >ref\|YP_655185.1\| gp89 [Mycobacterium phage Orion] >gb\|ACI12813.1\| hypothetical protein CHAH_93 [Mycobacterium phage Chah] >gb\|ADA84019.1\| hypothetical protein SCOOT17C_91 [Mycobacterium phage Scoot17C] >gb\|AEJ91862.1\| hypothetical protein THORA_89 [Mycobacterium phage Thora] >gb\|AEJ92747.1\| hypothetical protein SEA_SERENDIPITY_90 [Mycobacterium phage Serendipity] | 100 | 100 | 100 | 73 | 1 | 73 | 1 | 73 | 0 | 9.34074e-46 |

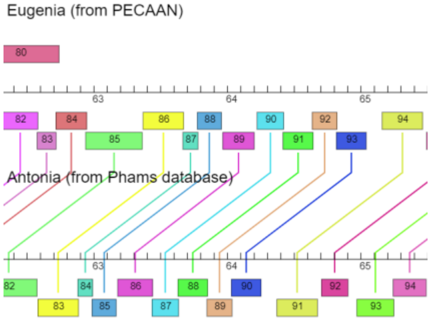
### c. SIF: HHPred

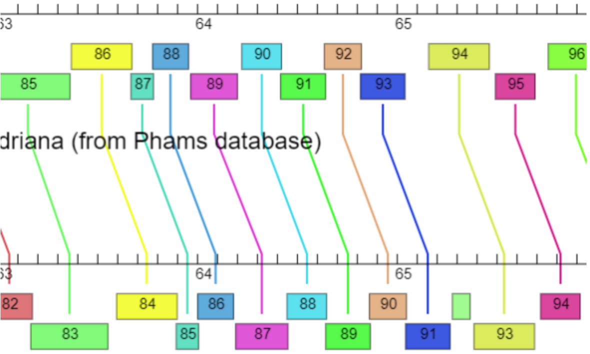


| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☐ | PF13314.10 | DUF4083 ; Domain of unknown function (DUF4083) | 79.3 | 23.2877 | 40 | 57 | 30 | 47 | 7.8 |
| ☐ | 4EIJ_A | P protein; oligomerization, REPLICATION; 2.2001A {Mumps virus} | 77 | 43.8356 | 3 | 35 | 13 | 45 | 28 |
| ☐ | 2L9Z_A | PR domain zinc finger protein 4; zinc-binding domain, TRANSCRIPTION; HET: ZN; NMR {Homo sapiens} | 71.4 | 21.9178 | 14 | 30 | 53 | 69 | 6.8 |

### d. SIF: Synteny-Phamerator (three genomes)

Shows synteny with Adriana, Antonia, Badfish, etc.





# 7. Any other important information.

Transmembrane Prediction

```
# WEBSEQUENCE Length: 73
# WEBSEQUENCE Number of predicted TMHs:  0
# WEBSEQUENCE Exp number of AAs in TMHs: 0.00111
# WEBSEQUENCE Exp number, first 60 AAs:  0.00111
# WEBSEQUENCE Total prob of N-in:        0.02768
WEBSEQUENCE      TMHMM2.0   outside      1    73
```

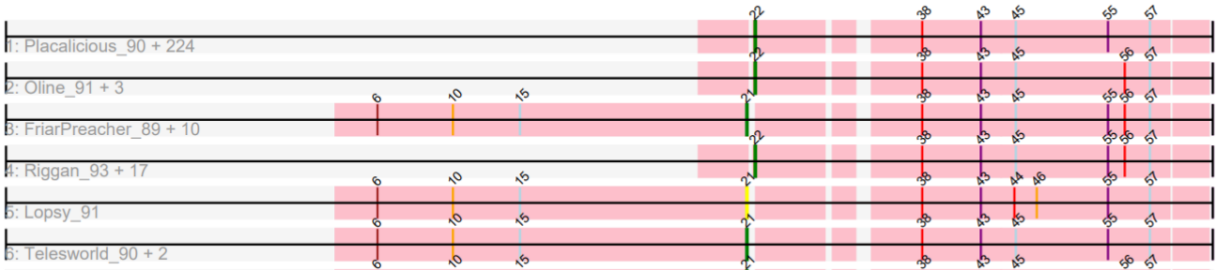
## CURATOR NAME: JOHN MOTTER

## GENE NAME: EUGENIA GENE 94

## DNA MASTER NOTES: N/A

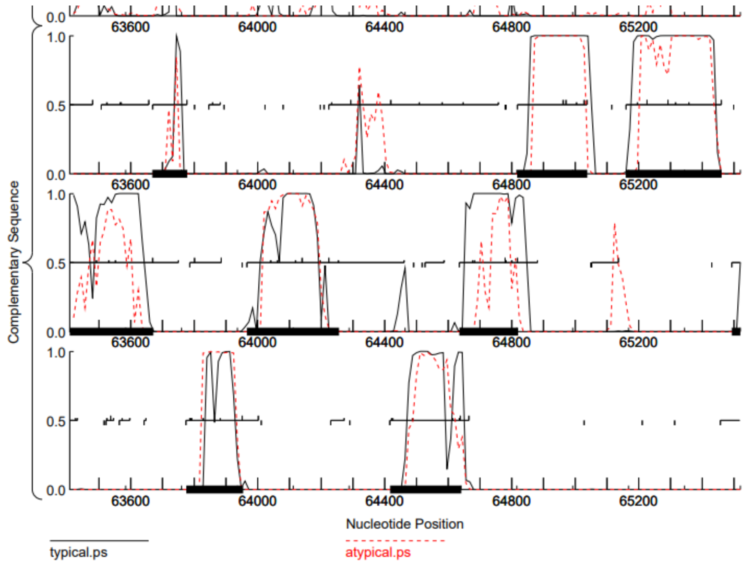
## START POSITION EVALUATION (IN ORDER):

### 1. Starterator

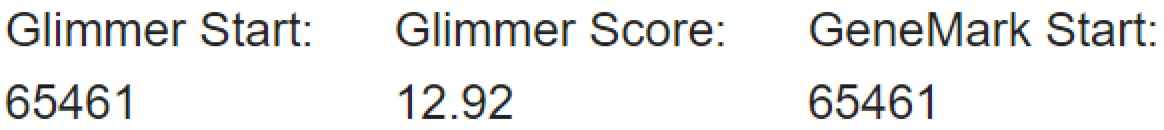


The start number called the most often in the published annotations is 22, it was called in 270 of the 292 non-draft genes in the pham.

### 2. GeneMark coding potential

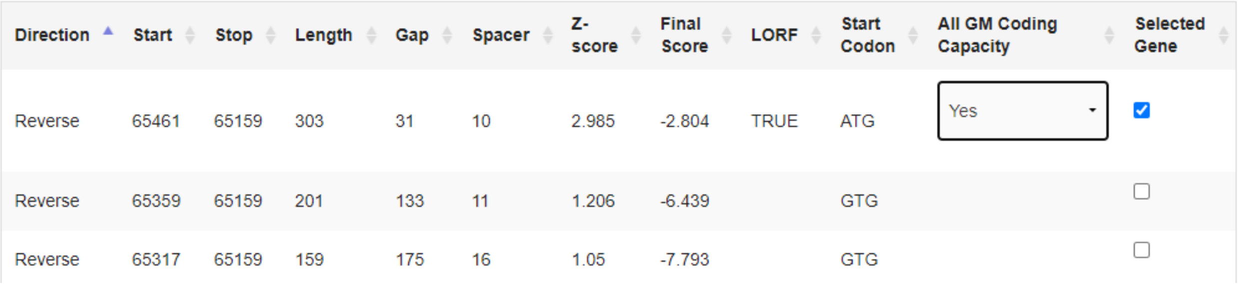


3. Glimmer and GeneMark agreement



| Glimmer Start: | Glimmer Score: | GeneMark Start: |
|---|---|---|
| 65461 | 12.92 | 65461 |

4. Longest open reading frame (ORF) without excessive gap



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 65461 | 65159 | 303 | 31 | 10 | 2.985 | -2.804 | TRUE | ATG | Yes | ☑ |
| Reverse | 65359 | 65159 | 201 | 133 | 11 | 1.206 | -6.439 | | GTG | | ☐ |
| Reverse | 65317 | 65159 | 159 | 175 | 16 | 1.05 | -7.793 | | GTG | | ☐ |

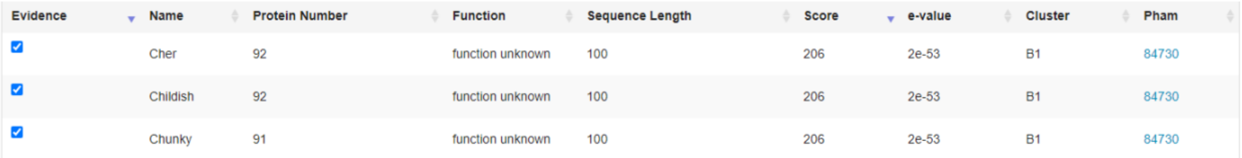
5. Function. Please follow this [Official SEAPHAGE Function List]

If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

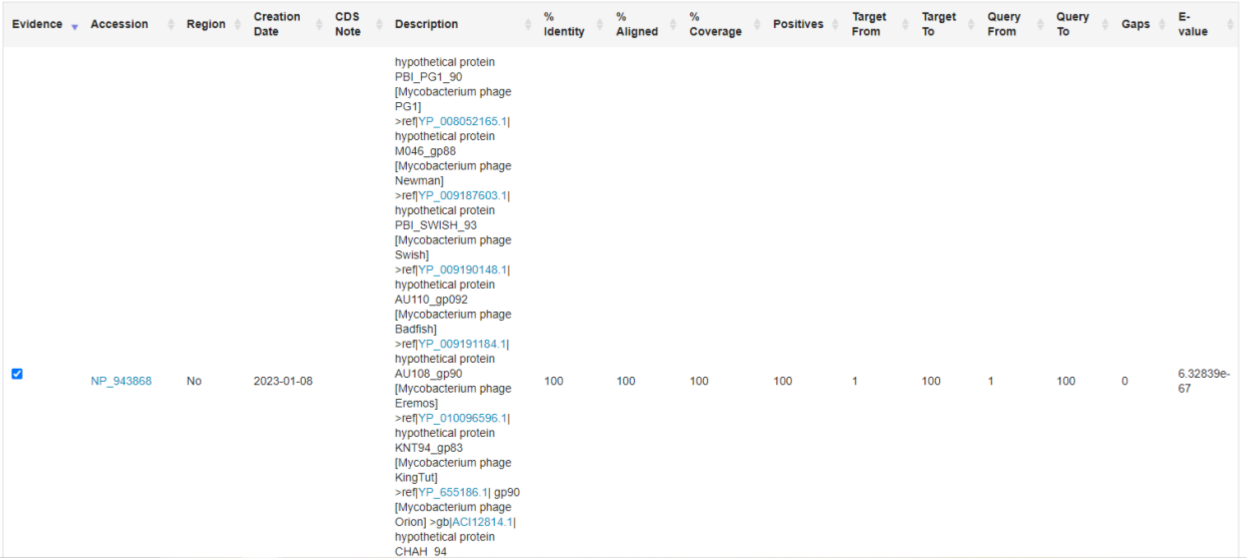
6. Supporting Information for Function (SIF)

a. SIF: PhageDb BLAST



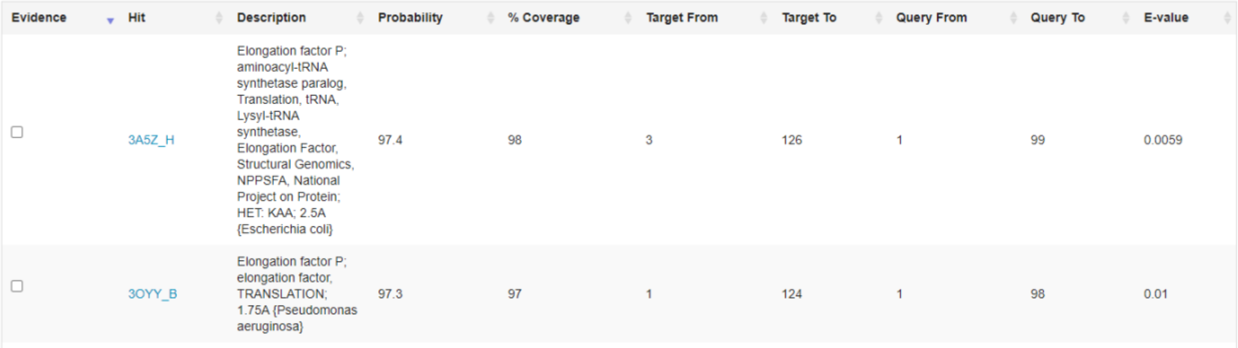
| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | Cher | 92 | function unknown | 100 | 206 | 2e-53 | B1 | 84730 |
| ☑ | Childish | 92 | function unknown | 100 | 206 | 2e-53 | B1 | 84730 |
| ☑ | Chunky | 91 | function unknown | 100 | 206 | 2e-53 | B1 | 84730 |

b. SIF: NCBI BLAST



| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | NP_943868 | No | 2023-01-08 |  | hypothetical protein PBI_PG1_90 [Mycobacterium phage PG1] >ref\|YP_008052165.1\| hypothetical protein M046_gp88 [Mycobacterium phage Newman] >ref\|YP_009187603.1\| hypothetical protein PBI_SWISH_93 [Mycobacterium phage Swish] >ref\|YP_009190148.1\| hypothetical protein AU110_gp092 [Mycobacterium phage Badfish] >ref\|YP_009191184.1\| hypothetical protein AU108_gp90 [Mycobacterium phage Eremos] >ref\|YP_010096596.1\| hypothetical protein KNT94_gp83 [Mycobacterium phage KingTut] >ref\|YP_655186.1\| gp90 [Mycobacterium phage Orion] >gb\|ACI12814.1\| hypothetical protein CHAH_94 | 100 | 100 | 100 | 100 | 1 | 100 | 1 | 100 | 0 | 6.32839e-67 |

### c. SIF: HHPred

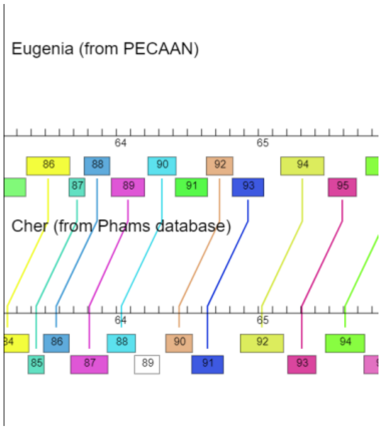


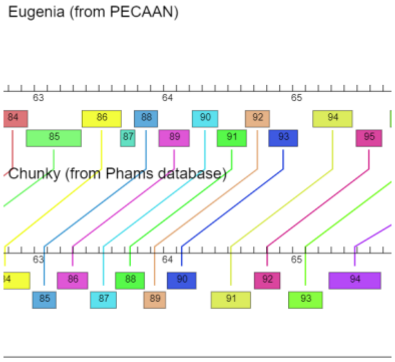
| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☐ | 3A5Z_H | Elongation factor P; aminoacyl-tRNA synthetase paralog, Translation, tRNA, Lysyl-tRNA synthetase, Elongation Factor, Structural Genomics, NPPSFA, National Project on Protein; HET: KAA; 2.5A {Escherichia coli} | 97.4 | 98 | 3 | 126 | 1 | 99 | 0.0059 |
| ☐ | 3OYY_B | Elongation factor P; elongation factor, TRANSLATION; 1.75A {Pseudomonas aeruginosa} | 97.3 | 97 | 1 | 124 | 1 | 98 | 0.01 |

Although strong matches were detected, they were not included in calling gene function or as evidence due to the low E value and the proteins being from organisms other than *Mycobacterium* phages.

### d. SIF: Synteny-Phamerator (three genomes)

Shows synteny with Cher, Childish, Chunky, etc.





## 7. Any other important information.

Transmembrane prediction

```
# WEBSEQUENCE Length: 100
# WEBSEQUENCE Number of predicted TMHs:  0
# WEBSEQUENCE Exp number of AAs in TMHs: 0.0125
# WEBSEQUENCE Exp number, first 60 AAs:  0.00302
# WEBSEQUENCE Total prob of N-in:        0.29305
WEBSEQUENCE      TMHMM2.0   outside      1   100
```

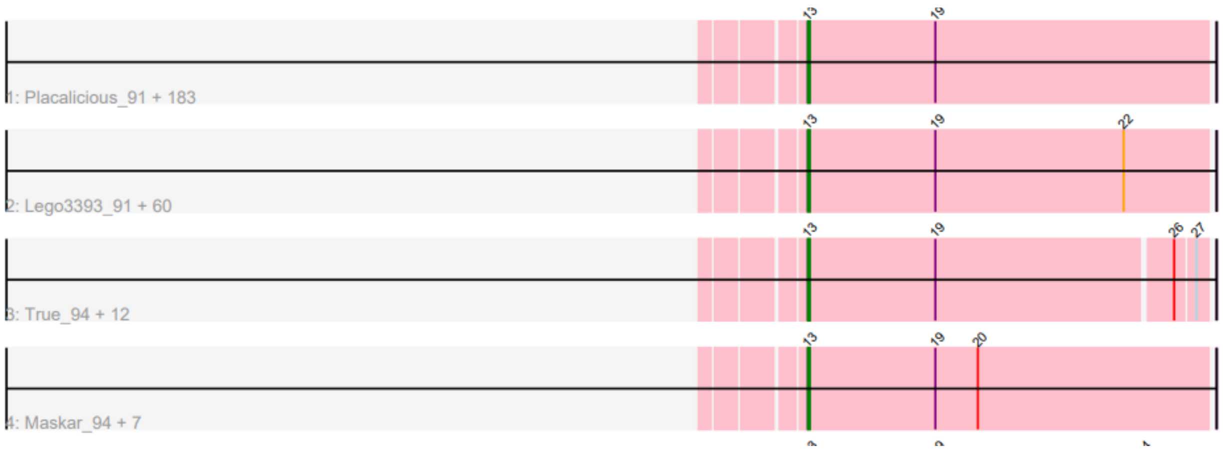
## CURATOR NAME: JOHN MOTTER

## GENE NAME: EUGENIA GENE 95

## DNA MASTER NOTES: N/A

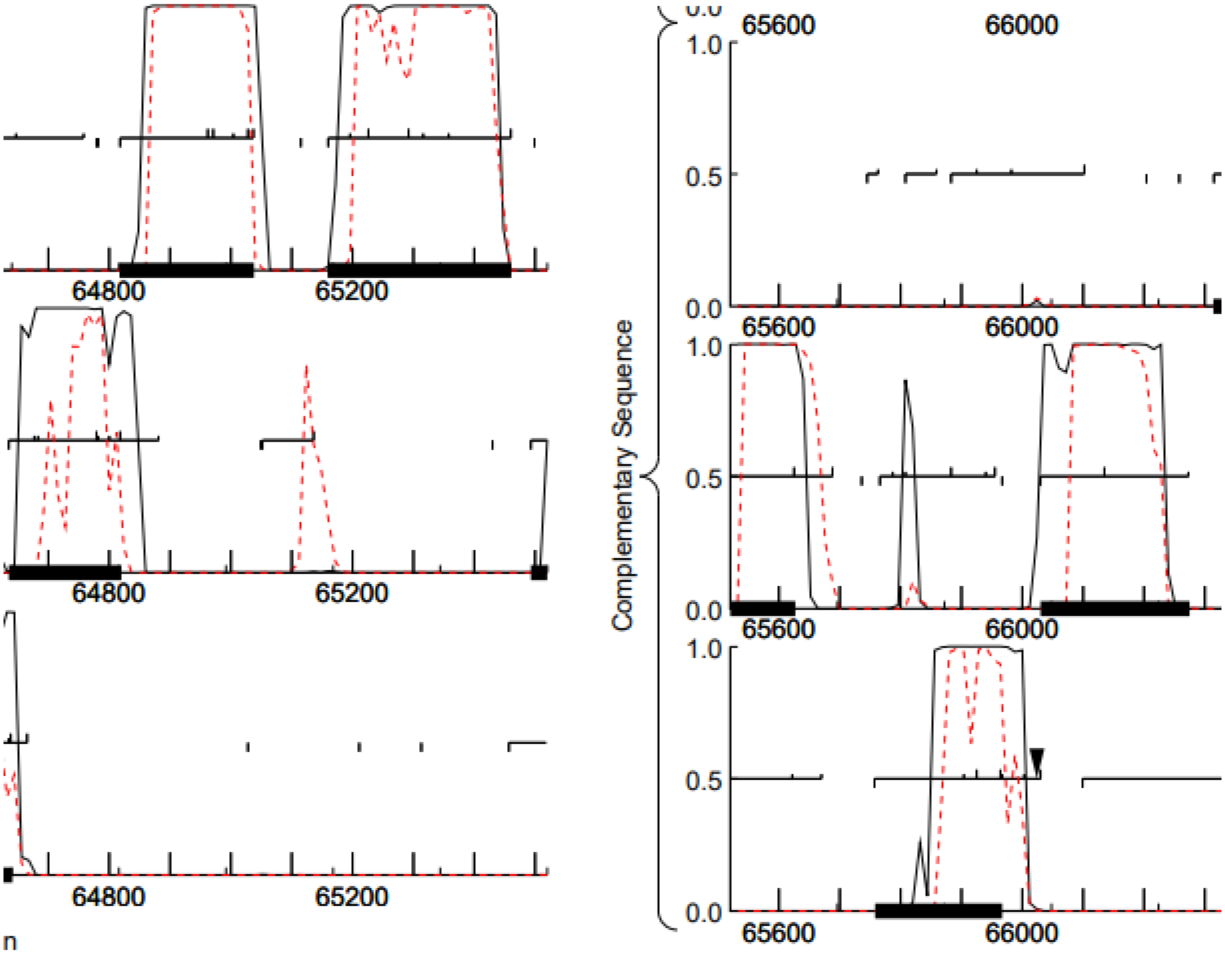
## START POSITION EVALUATION (IN ORDER):

### 1. Starterator



The start number called the most often in the published annotations is 13, it was called in 275 of the 301 non-draft genes in the pham.

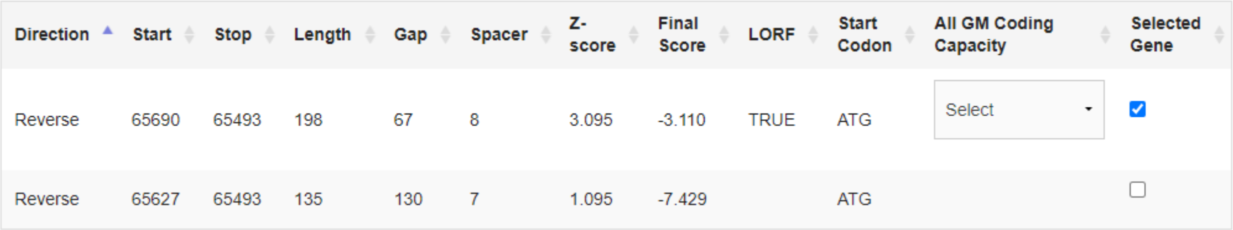
### 2. GeneMark coding potential



3. Glimmer and GeneMark agreement

| Glimmer Start: | Glimmer Score: | GeneMark Start: |
|---|---|---|
| 65690 | 10.79 | 65690 |

4. Longest open reading frame (ORF) without excessive gap



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 65690 | 65493 | 198 | 67 | 8 | 3.095 | -3.110 | TRUE | ATG | Select | ☑ |
| Reverse | 65627 | 65493 | 135 | 130 | 7 | 1.095 | -7.429 |  | ATG |  | ☐ |

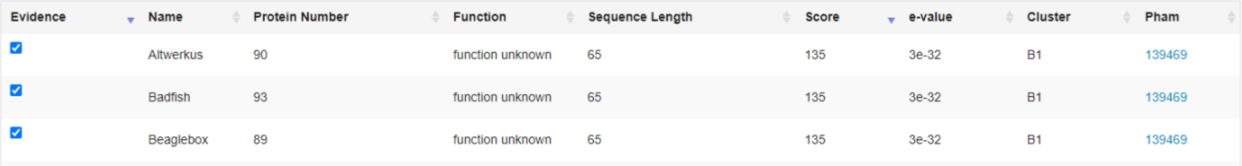
5. Function. Please follow this [Official SEAPHAGE Function List]

If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

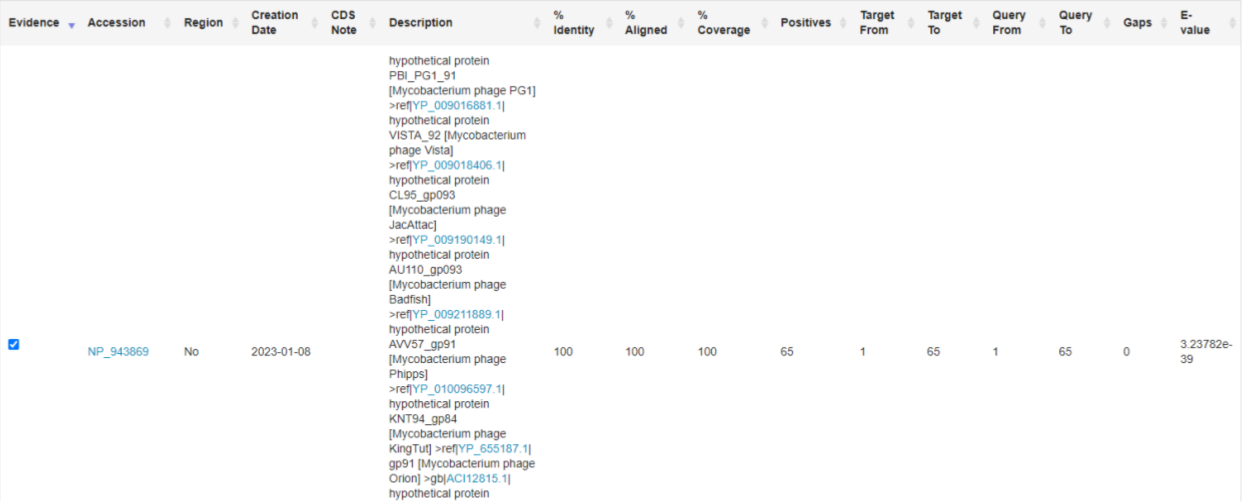
6. Supporting Information for Function (SIF)

a. SIF: PhageDb BLAST



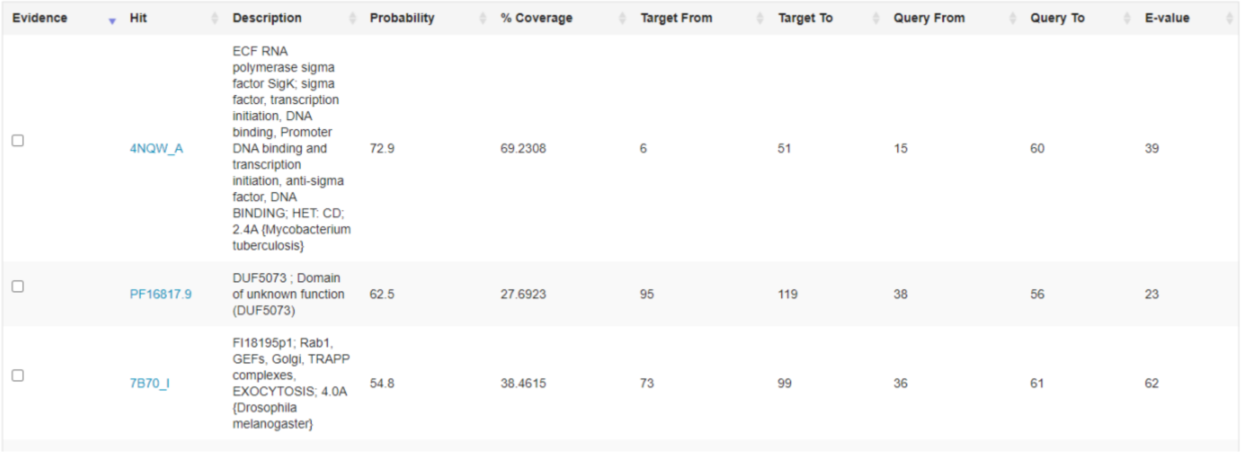
| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | Altwerkus | 90 | function unknown | 65 | 135 | 3e-32 | B1 | 139469 |
| ☑ | Badfish | 93 | function unknown | 65 | 135 | 3e-32 | B1 | 139469 |
| ☑ | Beaglebox | 89 | function unknown | 65 | 135 | 3e-32 | B1 | 139469 |

b. SIF: NCBI BLAST



| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | NP_943869 | No | 2023-01-08 |  | hypothetical protein PBI_PG1_91 [Mycobacterium phage PG1] >ref\|YP_009016881.1\| hypothetical protein VISTA_92 [Mycobacterium phage Vista] >ref\|YP_009018406.1\| hypothetical protein CL95_gp093 [Mycobacterium phage JacAttac] >ref\|YP_009190149.1\| hypothetical protein AU110_gp093 [Mycobacterium phage Badfish] >ref\|YP_009211889.1\| hypothetical protein AVV57_gp91 [Mycobacterium phage Phipps] >ref\|YP_010096597.1\| hypothetical protein KNT94_gp84 [Mycobacterium phage KingTut] >ref\|YP_655187.1\| gp91 [Mycobacterium phage Orion] >gb\|ACI12815.1\| hypothetical protein | 100 | 100 | 100 | 65 | 1 | 65 | 1 | 65 | 0 | 3.23782e-39 |

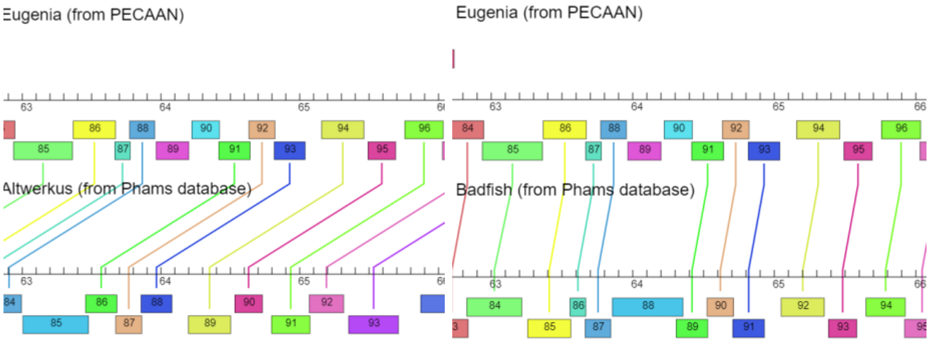
## c. SIF: HHPred



| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☐ | 4NQW_A | ECF RNA polymerase sigma factor SigK; sigma factor, transcription initiation, DNA binding, Promoter DNA binding and transcription initiation, anti-sigma factor, DNA BINDING; HET: CD; 2.4A {Mycobacterium tuberculosis} | 72.9 | 69.2308 | 6 | 51 | 15 | 60 | 39 |
| ☐ | PF16817.9 | DUF5073 ; Domain of unknown function (DUF5073) | 62.5 | 27.6923 | 95 | 119 | 38 | 56 | 23 |
| ☐ | 7B70_I | FI18195p1; Rab1, GEFs, Golgi, TRAPP complexes, EXOCYTOSIS; 4.0A {Drosophila melanogaster} | 54.8 | 38.4615 | 73 | 99 | 36 | 61 | 62 |

## d. SIF: Synteny-Phamerator (three genomes)

Shows synteny with Altwerkus, Badfish, Beaglebox, etc.



## 7. Any other important information.

Transmembrane prediction

# WEBSEQUENCE Length: 65

# WEBSEQUENCE Number of predicted TMHs: 0

# WEBSEQUENCE Exp number of AAs in TMHs: 0.00166

# WEBSEQUENCE Exp number, first 60 AAs: 0.00166

# WEBSEQUENCE Total prob of N-in: 0.27178

WEBSEQUENCE TMHMM2.0 outside 1 65

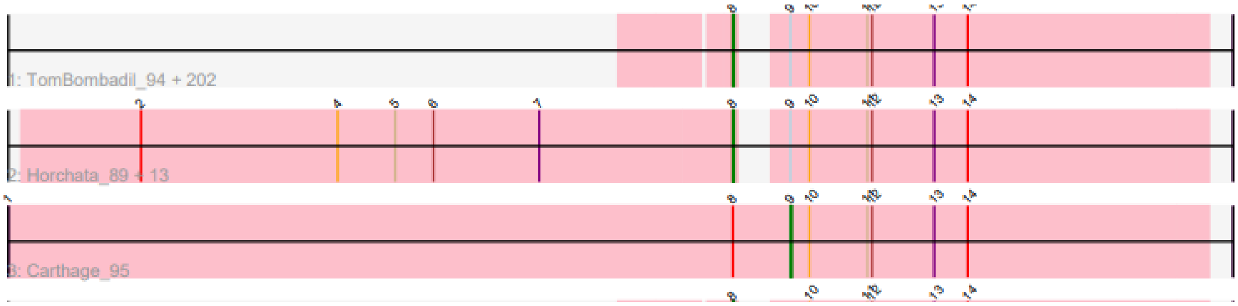
## CURATOR NAME: JOHN MOTTER

## GENE NAME: EUGENIA GENE 96

## DNA MASTER NOTES: N/A

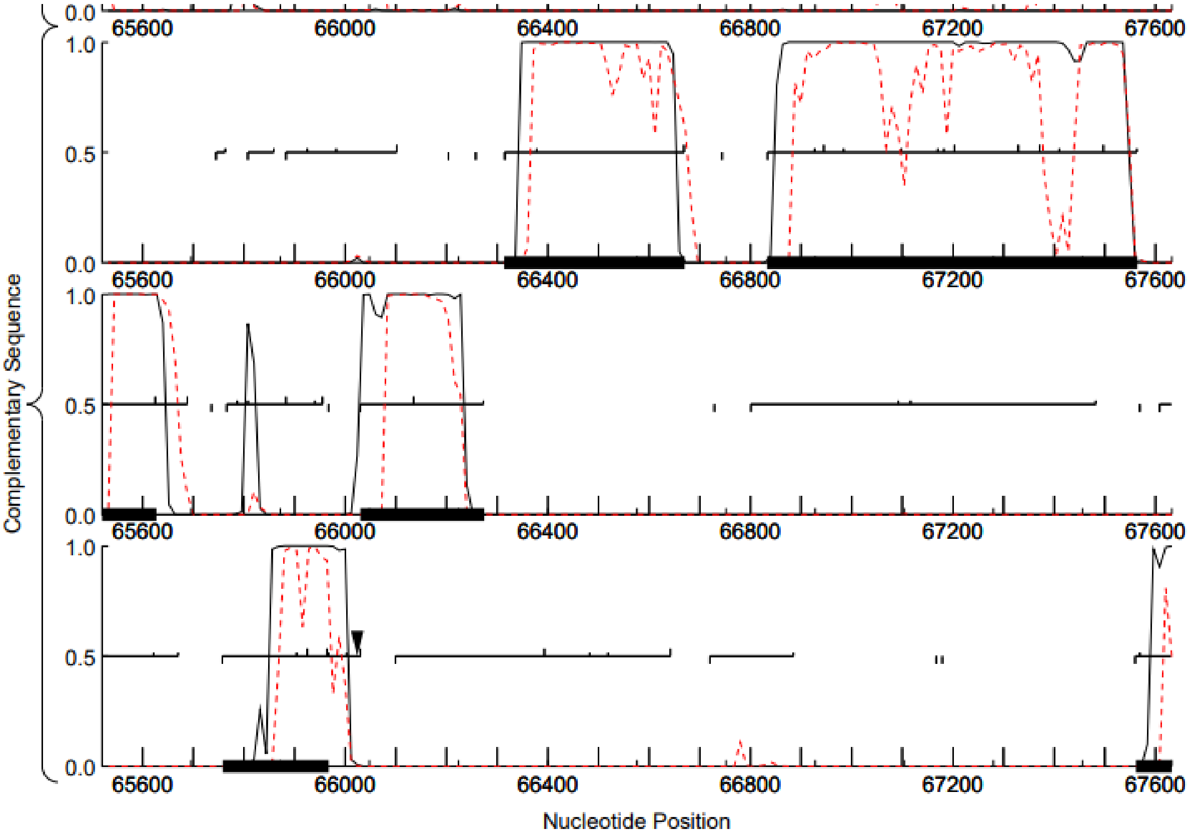
## START POSITION EVALUATION (IN ORDER):

### 1. Starterator

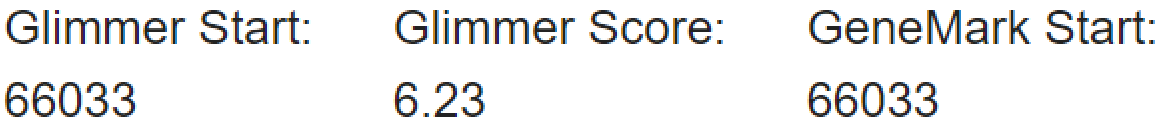


The start number called the most often in the published annotations is 8, it was called in 214 of the 223 non-draft genes in the pham.

### 2. GeneMark coding potential

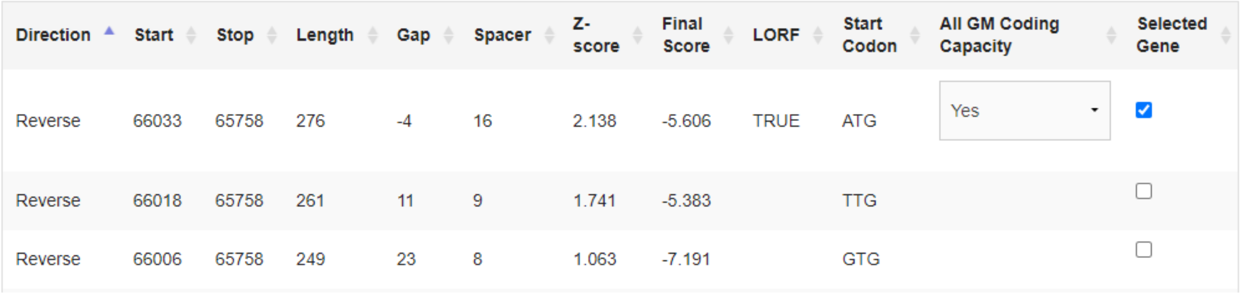


## 3. Glimmer and GeneMark agreement



| Glimmer Start: | Glimmer Score: | GeneMark Start: |
|---|---|---|
| 66033 | 6.23 | 66033 |

## 4. Longest open reading frame (ORF) without excessive gap



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 66033 | 65758 | 276 | -4 | 16 | 2.138 | -5.606 | TRUE | ATG | Yes | ☑ |
| Reverse | 66018 | 65758 | 261 | 11 | 9 | 1.741 | -5.383 | | TTG | | ☐ |
| Reverse | 66006 | 65758 | 249 | 23 | 8 | 1.063 | -7.191 | | GTG | | ☐ |

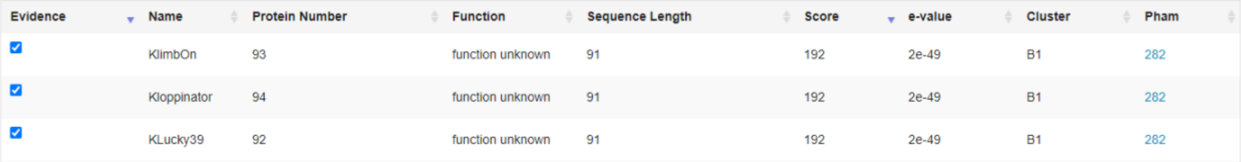
5. Function. Please follow this Official SEAPHAGE Function List

If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

6. Supporting Information for Function (SIF)

a. SIF: PhageDb BLAST

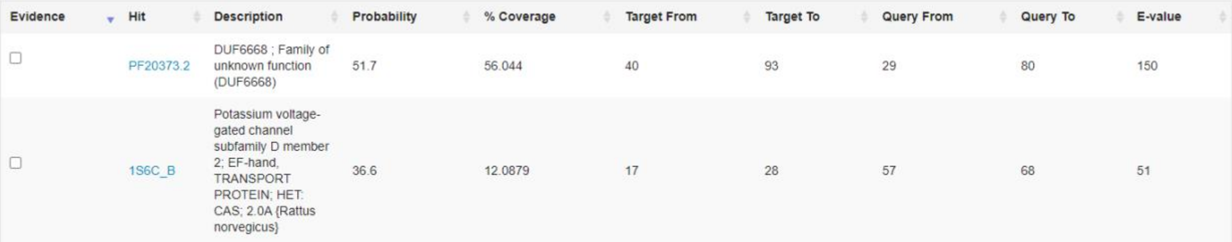


| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | KlimbOn | 93 | function unknown | 91 | 192 | 2e-49 | B1 | 282 |
| ☑ | Kloppinator | 94 | function unknown | 91 | 192 | 2e-49 | B1 | 282 |
| ☑ | KLucky39 | 92 | function unknown | 91 | 192 | 2e-49 | B1 | 282 |

b. SIF: NCBI BLAST

| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | YP_008052167 | No | 2023-01-08 | | hypothetical protein M046_gp90 [Mycobacterium phage Newman] >ref\|YP_009016882.1\| hypothetical protein VISTA_93 [Mycobacterium phage Vista] >ref\|YP_009018407.1\| hypothetical protein CL95_gp094 [Mycobacterium phage JacAttac] >ref\|YP_009168270.1\| hypothetical protein UNCLEHOWIE_90 [Mycobacterium phage UncleHowie] >ref\|YP_009189331.1\| gp93 [Mycobacterium phage ShiVal] >ref\|YP_009191086.1\| hypothetical protein AU159_gp092 [Mycobacterium phage Colbert] >ref\|YP_009191186.1\| hypothetical protein AU108_gp92 [Mycobacterium phage Eremos] >ref\|YP_009191285.1\| hypothetical protein AU098_gp092 [Mycobacterium phage Apizium] >ref\|YP_009198766.1\| hypothetical protein | 100 | 100 | 100 | 91 | 1 | 91 | 1 | 91 | 0 | 1.44108e-57 |

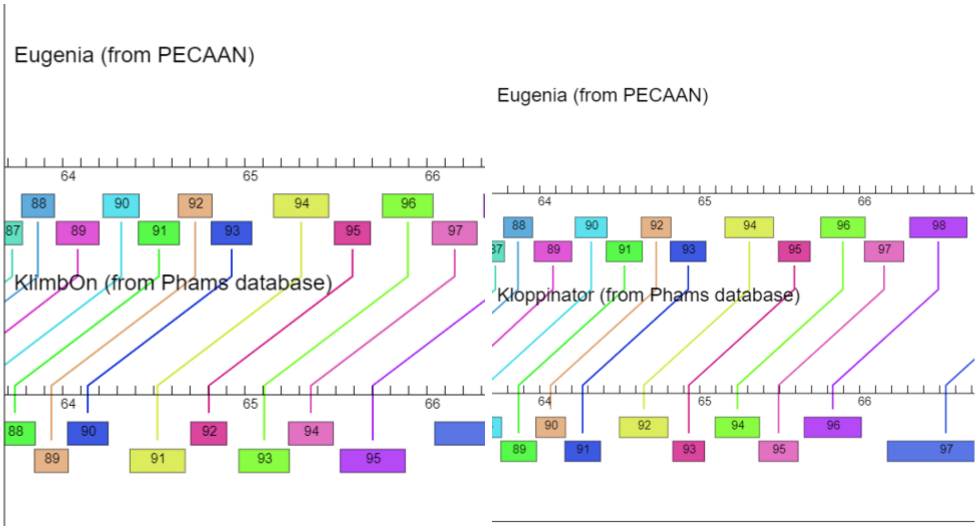
c. SIF: HHPred



| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☐ | PF20373.2 | DUF6668 ; Family of unknown function (DUF6668) | 51.7 | 56.044 | 40 | 93 | 29 | 80 | 150 |
| ☐ | 1S6C_B | Potassium voltage-gated channel subfamily D member 2; EF-hand, TRANSPORT PROTEIN; HET: CAS; 2.0A {Rattus norvegicus} | 36.6 | 12.0879 | 17 | 28 | 57 | 68 | 51 |

### d. SIF: Synteny-Phamerator (three genomes)

Shows synteny with KlimbOn, Kloppinator, KLucky39, etc.



## 7. Any other important information.

Transmembrane prediction

```
# WEBSEQUENCE Length: 91
# WEBSEQUENCE Number of predicted TMHs:  0
# WEBSEQUENCE Exp number of AAs in TMHs: 0.65872
# WEBSEQUENCE Exp number, first 60 AAs:  0.27739
# WEBSEQUENCE Total prob of N-in:        0.21755
WEBSEQUENCE      TMHMM2.0   outside      1    91
```

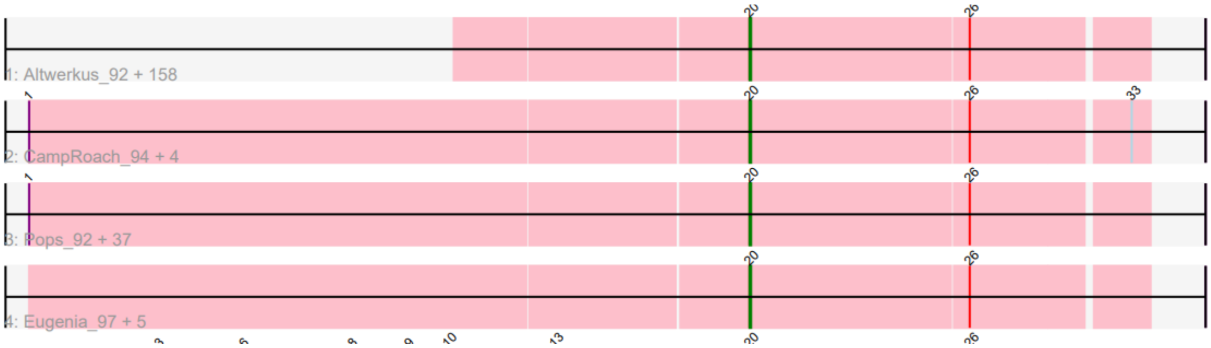
## CURATOR NAME: JOHN MOTTER

## GENE NAME: EUGENIA GENE 97

## DNA MASTER NOTES: N/A

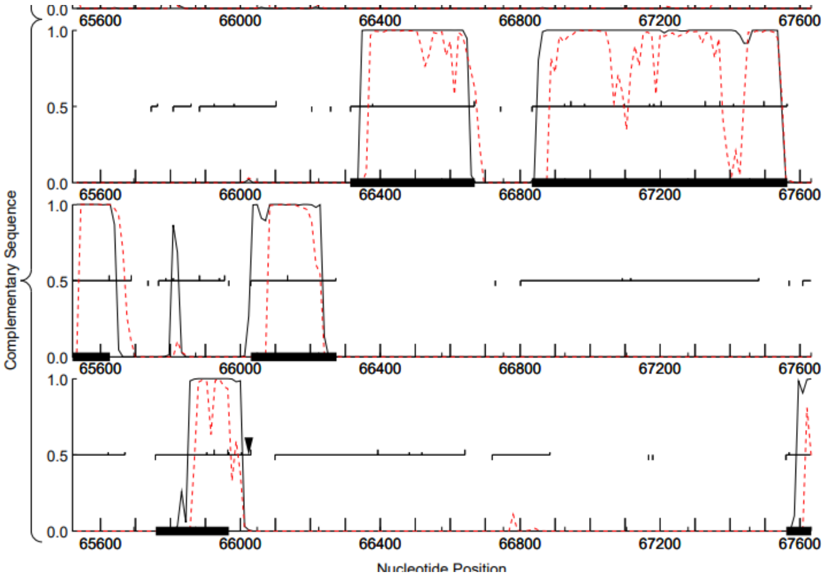
## START POSITION EVALUATION (IN ORDER):

### 1. Starterator

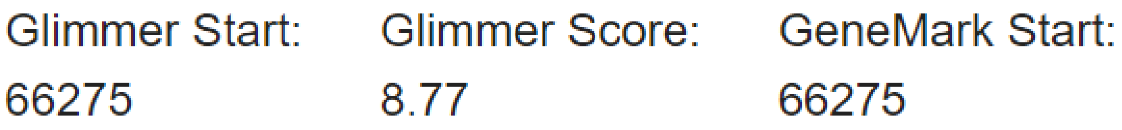


The start number called the most often in the published annotations is 20, it was called in 197 of the 204 non-draft genes in the pham.

### 2. GeneMark coding potential

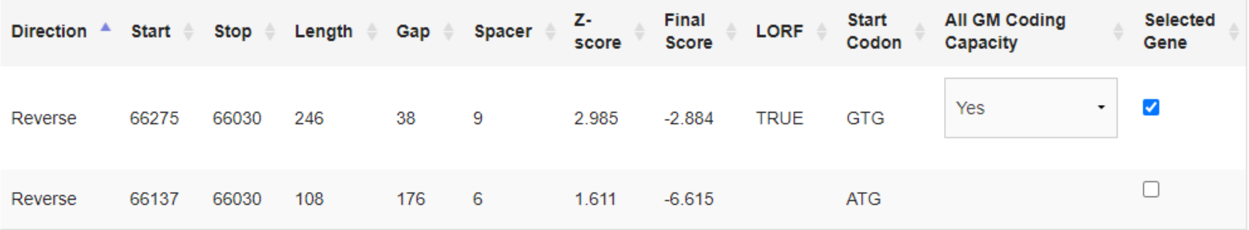


## 3. Glimmer and GeneMark agreement



| Glimmer Start: | Glimmer Score: | GeneMark Start: |
|---|---|---|
| 66275 | 8.77 | 66275 |

## 4. Longest open reading frame (ORF) without excessive gap



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 66275 | 66030 | 246 | 38 | 9 | 2.985 | -2.884 | TRUE | GTG | Yes | ☑ |
| Reverse | 66137 | 66030 | 108 | 176 | 6 | 1.611 | -6.615 |  | ATG |  | ☐ |

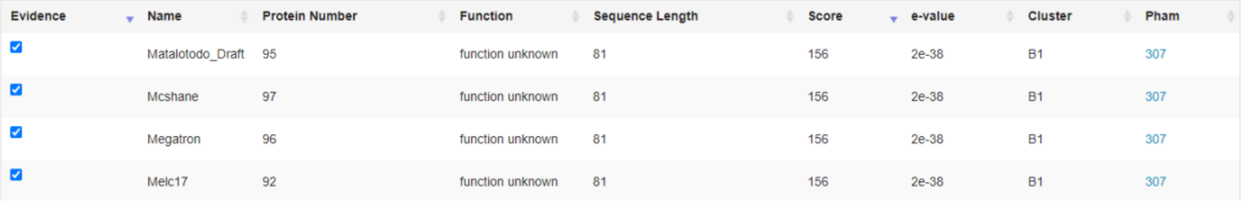
## 5. Function. Please follow this [Official SEAPHAGE Function List]

If no functional prediction is present, write "Hypothetical protein".

Hypothetical protein

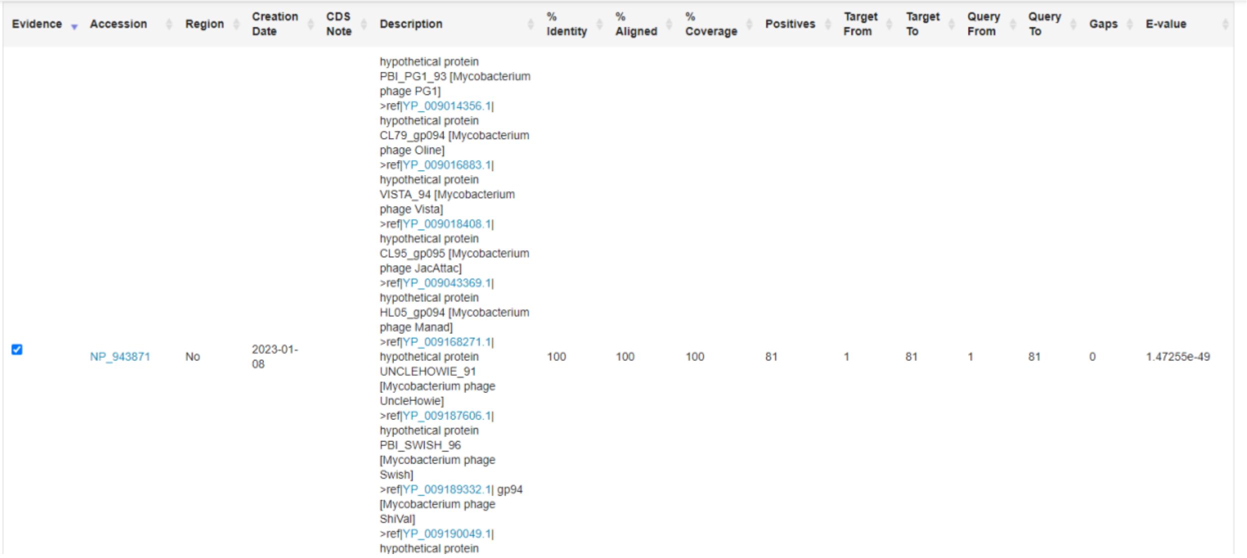
## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST



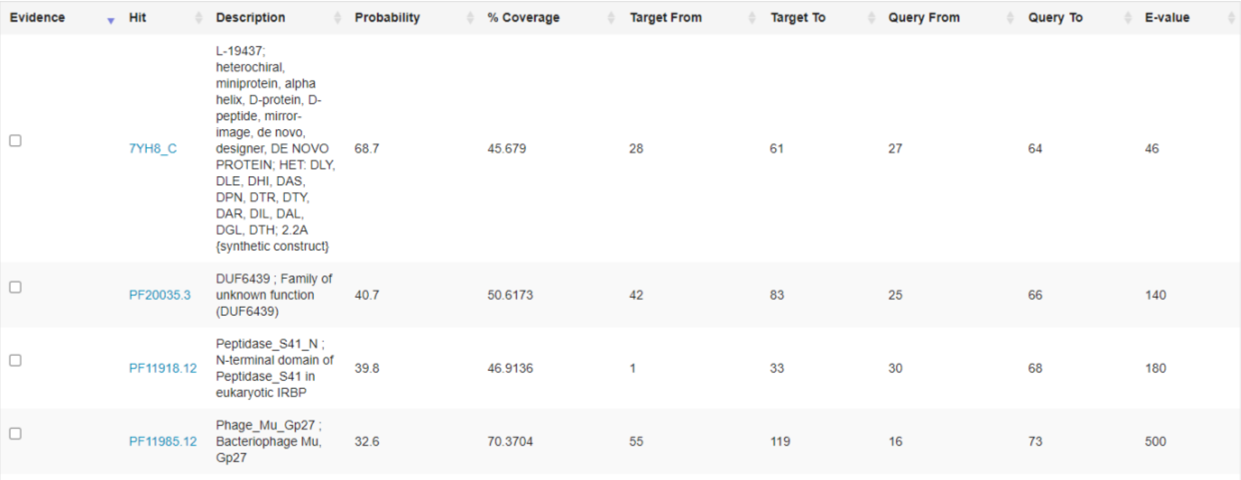
| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | Matalotodo_Draft | 95 | function unknown | 81 | 156 | 2e-38 | B1 | 307 |
| ☑ | Mcshane | 97 | function unknown | 81 | 156 | 2e-38 | B1 | 307 |
| ☑ | Megatron | 96 | function unknown | 81 | 156 | 2e-38 | B1 | 307 |
| ☑ | Melc17 | 92 | function unknown | 81 | 156 | 2e-38 | B1 | 307 |

### b. SIF: NCBI BLAST



| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | NP_943871 | No | 2023-01-08 | | hypothetical protein PBI_PG1_93 [Mycobacterium phage PG1] >ref\|YP_009014356.1\| hypothetical protein CL79_gp094 [Mycobacterium phage Oline] >ref\|YP_009016883.1\| hypothetical protein VISTA_94 [Mycobacterium phage Vista] >ref\|YP_009018408.1\| hypothetical protein CL95_gp095 [Mycobacterium phage JacAttac] >ref\|YP_009043369.1\| hypothetical protein HL05_gp094 [Mycobacterium phage Manad] >ref\|YP_009168271.1\| hypothetical protein UNCLEHOWIE_91 [Mycobacterium phage UncleHowie] >ref\|YP_009187606.1\| hypothetical protein PBI_SWISH_96 [Mycobacterium phage Swish] >ref\|YP_009189332.1\| gp94 [Mycobacterium phage ShiVal] >ref\|YP_009190049.1\| hypothetical protein | 100 | 100 | 100 | 81 | 1 | 81 | 1 | 81 | 0 | 1.47255e-49 |

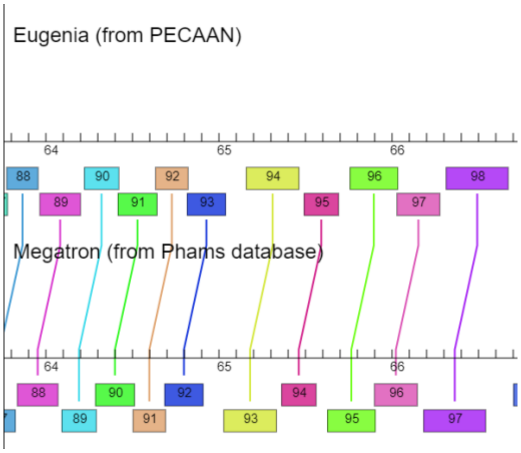
## c. SIF: HHPred

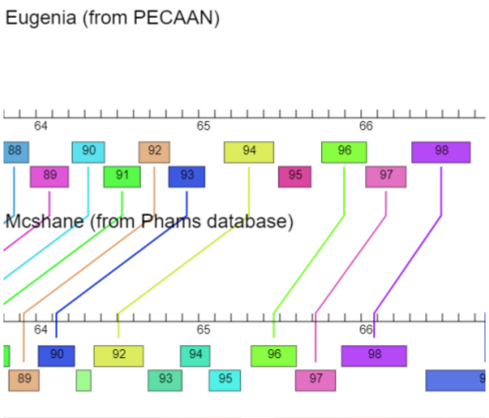


| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☐ | 7YH8_C | L-19437; heterochiral, miniprotein, alpha helix, D-protein, D-peptide, mirror-image, de novo, designer, DE NOVO PROTEIN; HET: DLY, DLE, DHI, DAS, DPN, DTR, DTY, DAR, DIL, DAL, DGL, DTH; 2.2A {synthetic construct} | 68.7 | 45.679 | 28 | 61 | 27 | 64 | 46 |
| ☐ | PF20035.3 | DUF6439 ; Family of unknown function (DUF6439) | 40.7 | 50.6173 | 42 | 83 | 25 | 66 | 140 |
| ☐ | PF11918.12 | Peptidase_S41_N ; N-terminal domain of Peptidase_S41 in eukaryotic IRBP | 39.8 | 46.9136 | 1 | 33 | 30 | 68 | 180 |
| ☐ | PF11985.12 | Phage_Mu_Gp27 ; Bacteriophage Mu, Gp27 | 32.6 | 70.3704 | 55 | 119 | 16 | 73 | 500 |

### d. SIF: Synteny-Phamerator (three genomes)

Shows synteny with Mcshane, Megatron, Melc17, etc.





## 7. Any other important information.

Transmembrane Prediction

```
# WEBSEQUENCE Length: 81
# WEBSEQUENCE Number of predicted TMHs:  0
# WEBSEQUENCE Exp number of AAs in TMHs: 0.0016
# WEBSEQUENCE Exp number, first 60 AAs:  0.00158
# WEBSEQUENCE Total prob of N-in:        0.66803
WEBSEQUENCE      TMHMM2.0   inside       1    81
```

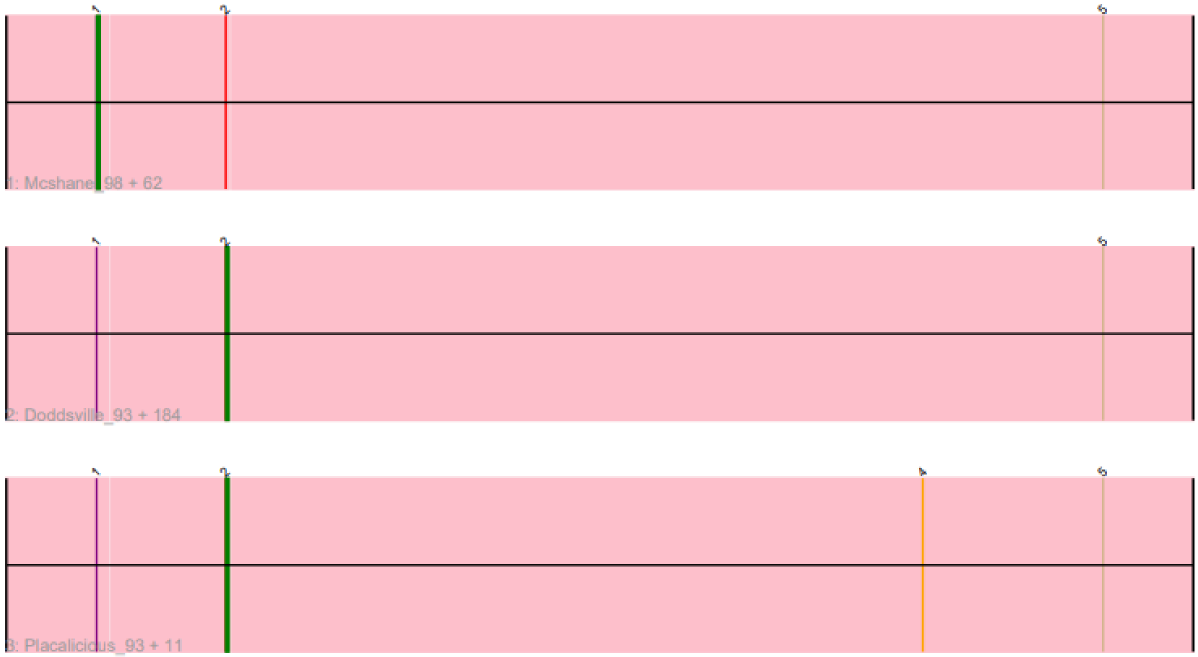
## CURATOR NAME: JOHN MOTTER

## GENE NAME: EUGENIA GENE 98

## DNA MASTER NOTES: N/A

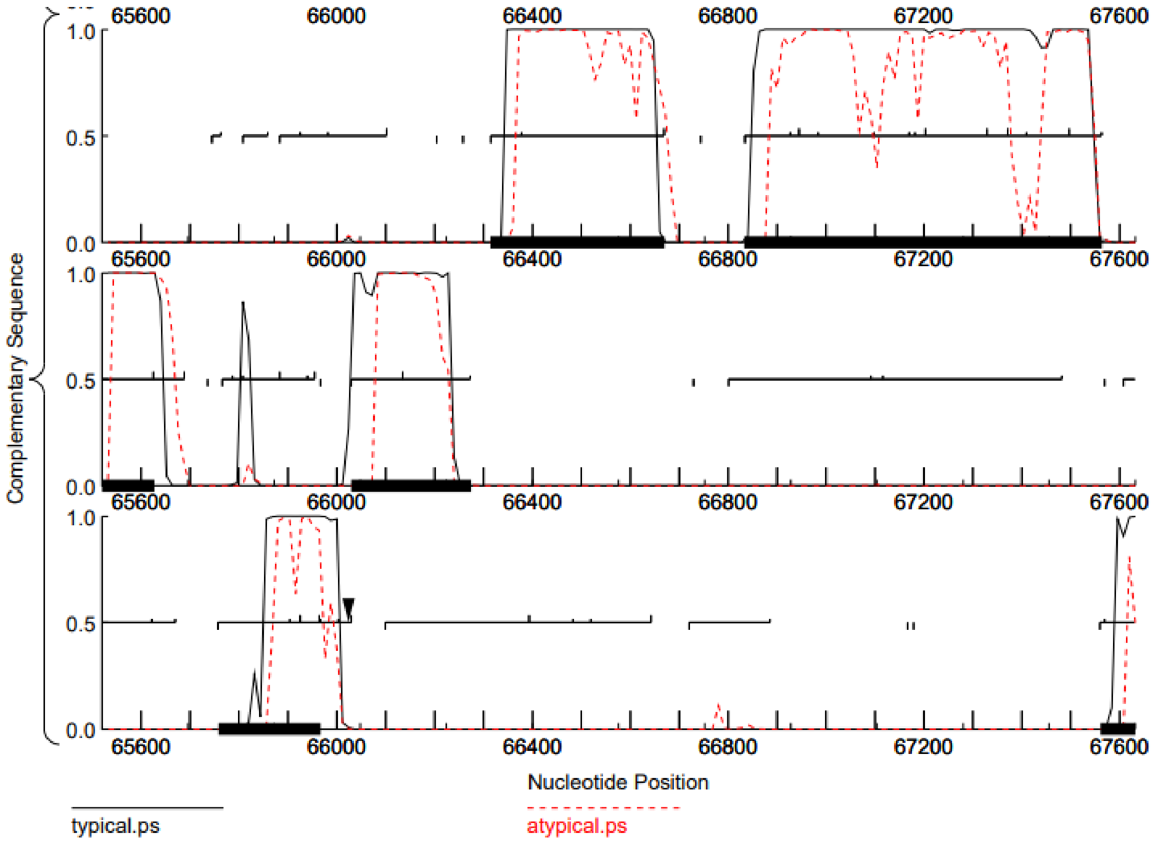
## START POSITION EVALUATION (IN ORDER):

1. Starterator

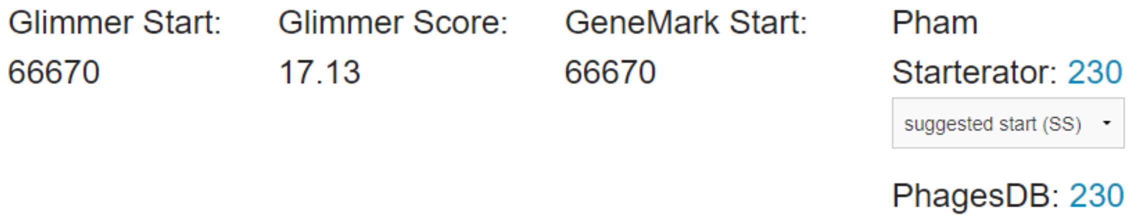


The start number called the most often in the published annotations is 2, it was called in 175 of the 246 non-draft genes in the pham.

## 2. GeneMark coding potential



## 3. Glimmer and GeneMark agreement


Glimmer Start: 66670

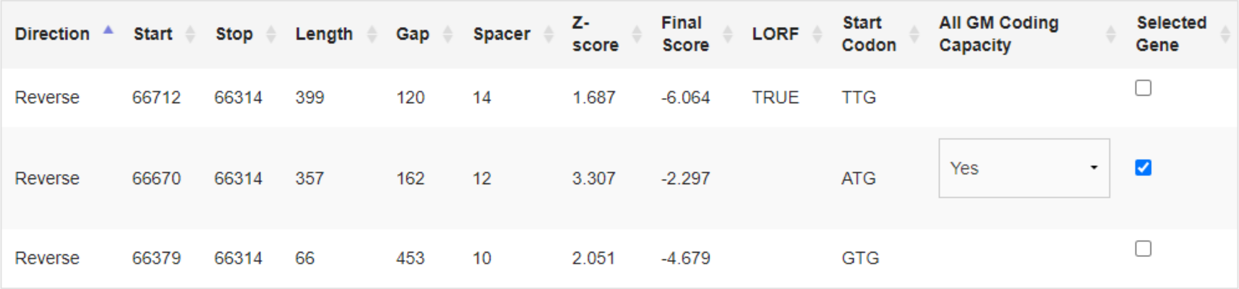
Glimmer Score: 17.13

GeneMark Start: 66670

Pham Starterator: 230

PhagesDB: 230

## 4. Longest open reading frame (ORF) without excessive gap



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 66712 | 66314 | 399 | 120 | 14 | 1.687 | -6.064 | TRUE | TTG | | ☐ |
| Reverse | 66670 | 66314 | 357 | 162 | 12 | 3.307 | -2.297 | | ATG | Yes | ☑ |
| Reverse | 66379 | 66314 | 66 | 453 | 10 | 2.051 | -4.679 | | GTG | | ☐ |

This start site was selected because it was predicted by Glimmer and GeneMark, and due to its Z-score and length. There are no longest ORFs with a short gap.

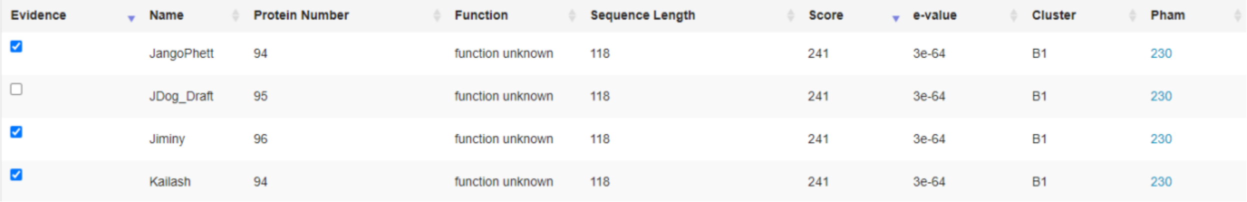
## 5. Function. Please follow this Official SEAPHAGE Function List

If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

## 6. Supporting Information for Function (SIF)

### a. SIF: PhageDb BLAST

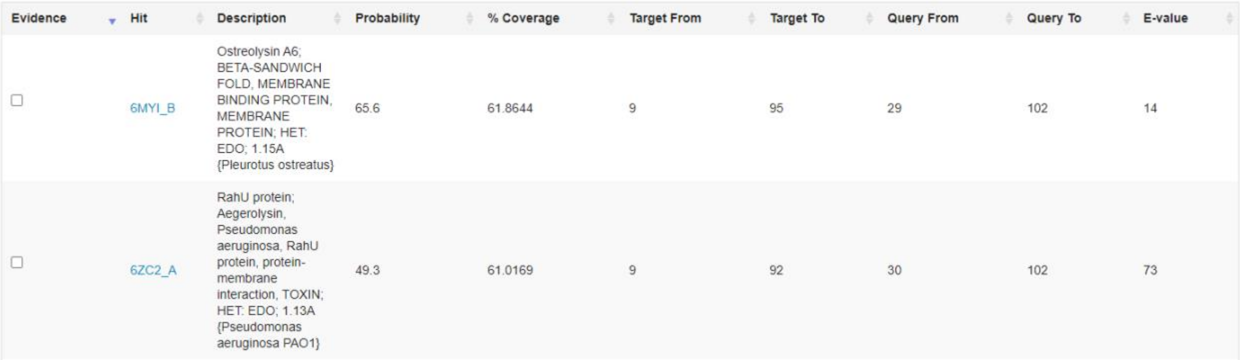


| Evidence | Name | Protein Number | Function | Sequence Length | Score | e-value | Cluster | Pham |
|---|---|---|---|---|---|---|---|---|
| ☑ | JangoPhett | 94 | function unknown | 118 | 241 | 3e-64 | B1 | 230 |
| ☐ | JDog_Draft | 95 | function unknown | 118 | 241 | 3e-64 | B1 | 230 |
| ☑ | Jiminy | 96 | function unknown | 118 | 241 | 3e-64 | B1 | 230 |
| ☑ | Kailash | 94 | function unknown | 118 | 241 | 3e-64 | B1 | 230 |

### b. SIF: NCBI BLAST

| Evidence | Accession | Region | Creation Date | CDS Note | Description | % Identity | % Aligned | % Coverage | Positives | Target From | Target To | Query From | Query To | Gaps | E-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | YP_008052168 | No | 2023-01-08 | | hypothetical protein M046_gp91 [Mycobacterium phage Newman] >ref\|YP_009014357.1\| hypothetical protein CL79_gp095 [Mycobacterium phage Oline] >ref\|YP_009043370.1\| hypothetical protein HL05_gp095 [Mycobacterium phage Manad] >ref\|YP_009191187.1\| hypothetical protein AU108_gp93 [Mycobacterium phage Eremos] >gb\|AGK87473.1\| hypothetical protein PBI_SDCHARGE11_94 [Mycobacterium phage SDcharge11] >gb\|AHY84363.1\| hypothetical protein PBI_KINGVEVEVE_94 [Mycobacterium phage KingVeVeVe] >gb\|AIM50328.1\| hypothetical protein PBI_VIVALDI_96 | 100 | 100 | 100 | 118 | 1 | 118 | 1 | 118 | 0 | 1.26374e-81 |

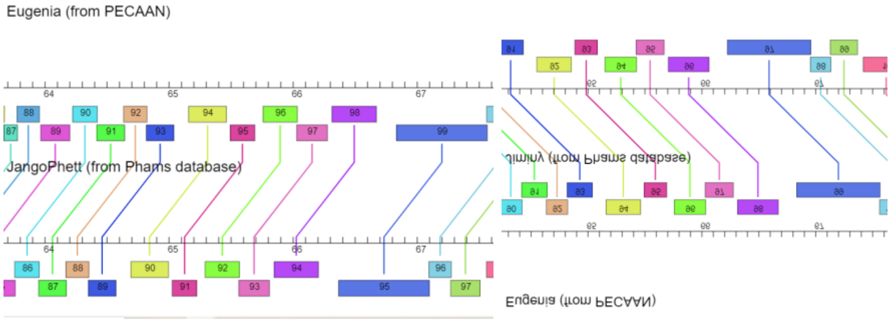
## c. SIF: HHPred



| Evidence | Hit | Description | Probability | % Coverage | Target From | Target To | Query From | Query To | E-value |
|---|---|---|---|---|---|---|---|---|---|
| ☐ | 6MYI_B | Ostreolysin A6; BETA-SANDWICH FOLD, MEMBRANE BINDING PROTEIN, MEMBRANE PROTEIN; HET: EDO; 1.15A {Pleurotus ostreatus} | 65.6 | 61.8644 | 9 | 95 | 29 | 102 | 14 |
| ☐ | 6ZC2_A | RahU protein; Aegerolysin, Pseudomonas aeruginosa, RahU protein, protein-membrane interaction, TOXIN; HET: EDO; 1.13A {Pseudomonas aeruginosa PAO1} | 49.3 | 61.0169 | 9 | 92 | 30 | 102 | 73 |

## d. SIF: Synteny-Phamerator (three genomes)

Shows synteny with JengoPhett, Jimmy, Kailash, etc.



## 7. Any other important information.

Transmembrane prediction

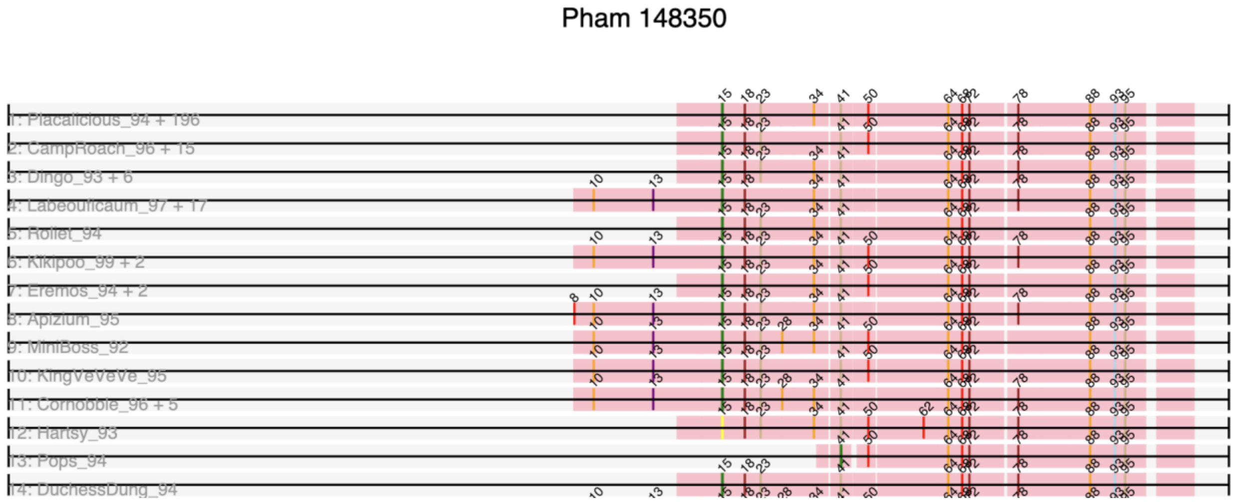
```
# WEBSEQUENCE Length: 118
# WEBSEQUENCE Number of predicted TMHs:  0
# WEBSEQUENCE Exp number of AAs in TMHs: 0.0001
# WEBSEQUENCE Exp number, first 60 AAs:  3e-05
# WEBSEQUENCE Total prob of N-in:        0.12876
WEBSEQUENCE      TMHMM2.0   outside      1   118
```

## CURATOR NAME: WEI

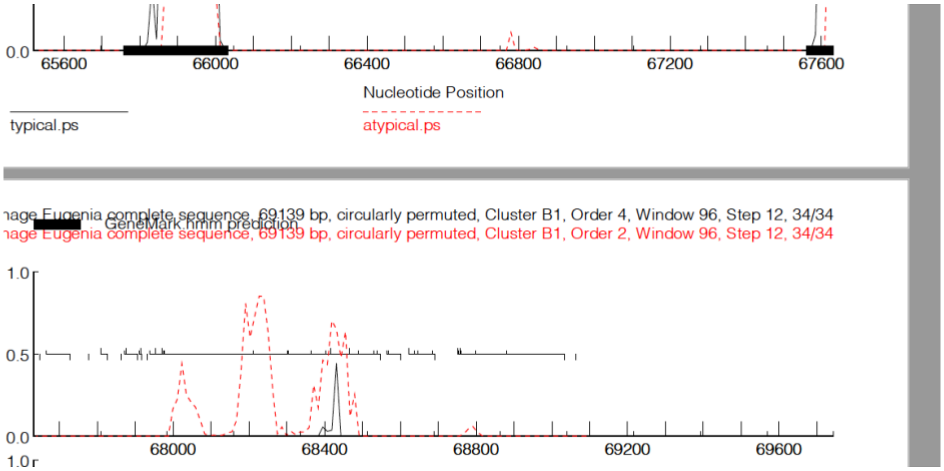
## GENE NAME: EUGENIA-DRAFT_99

## START POSITION EVALUATION (IN ORDER):

### 1. Starterator


Pham 148350

### 2. GeneMark coding potential



### 3. Glimmer and GeneMark agreement

YES

Glimmer Start:  67564
Glimmer Score:  14.65
GeneMark Start:  67564
Glimmer call @bp 67564 has strength 14.65; Genemark calls start at 67564

## 4. Longest open reading frame (ORF) without excessive gap

YES

## 5. If functional prediction, start include full length protein. If no functional prediction is present, write "Hypothetical protein".

Hypothetical protein

## 6. Start site similar to others in GenBank and PhagesDB

Charles1, Usavi, Waterdiva

## 7. Function (F)

unknown

## 9. Supporting Information for Function (SIF)

## 9. a. SIF-BLAST

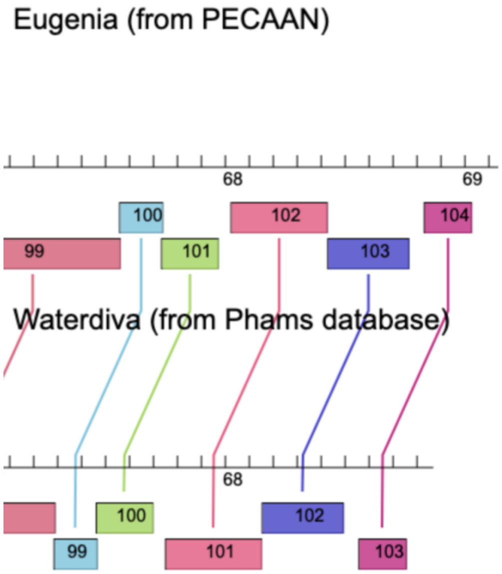
N/A

## 9. b. SIF-HHPred

N/A

9. c. SIF-Synteny

Aligns with Waterdiva, ABU, Adriana, Buckeye, CamL, Chah,

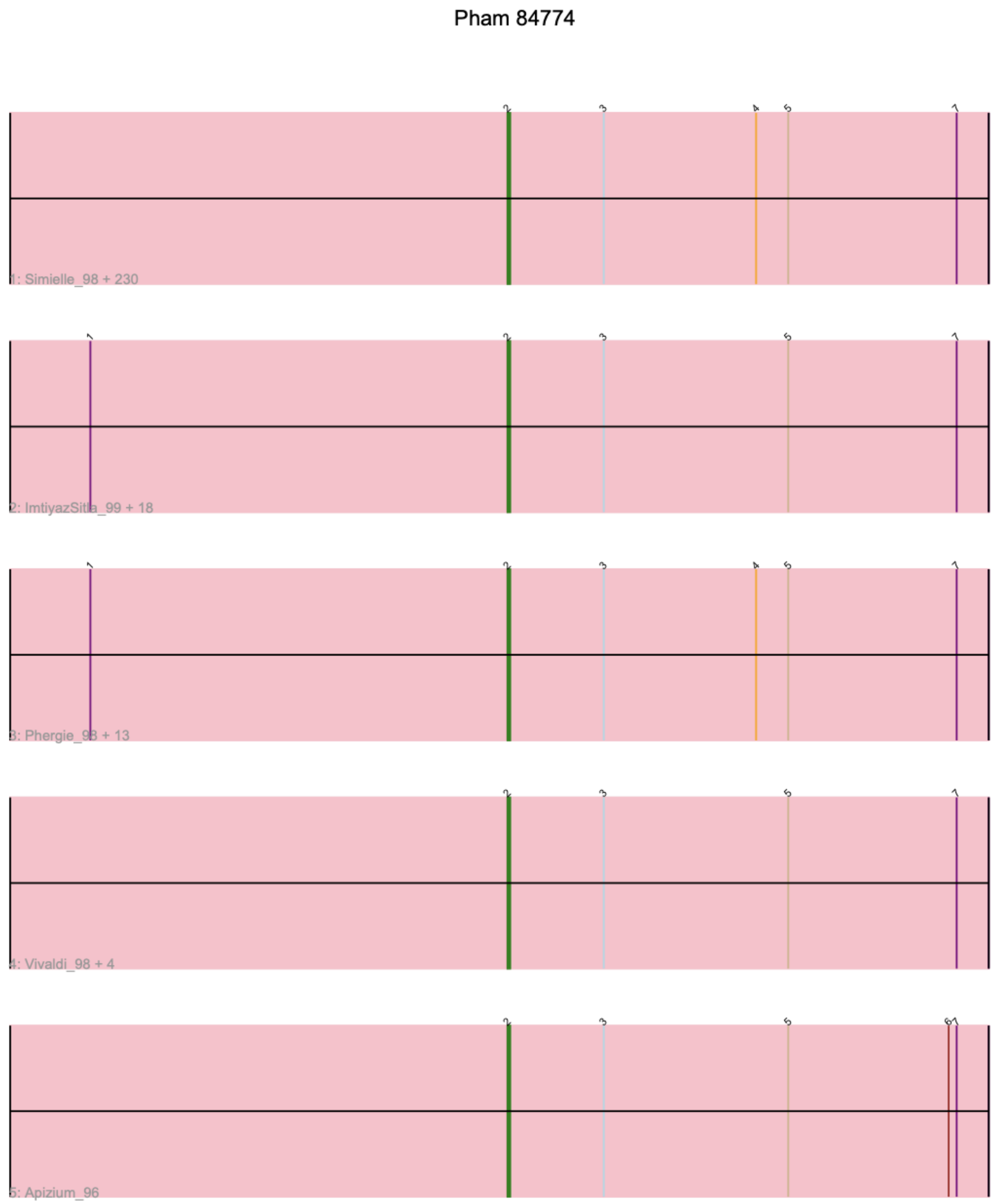


10. Any other important information.

## GENE NAME: EUGENIA-DRAFT_100

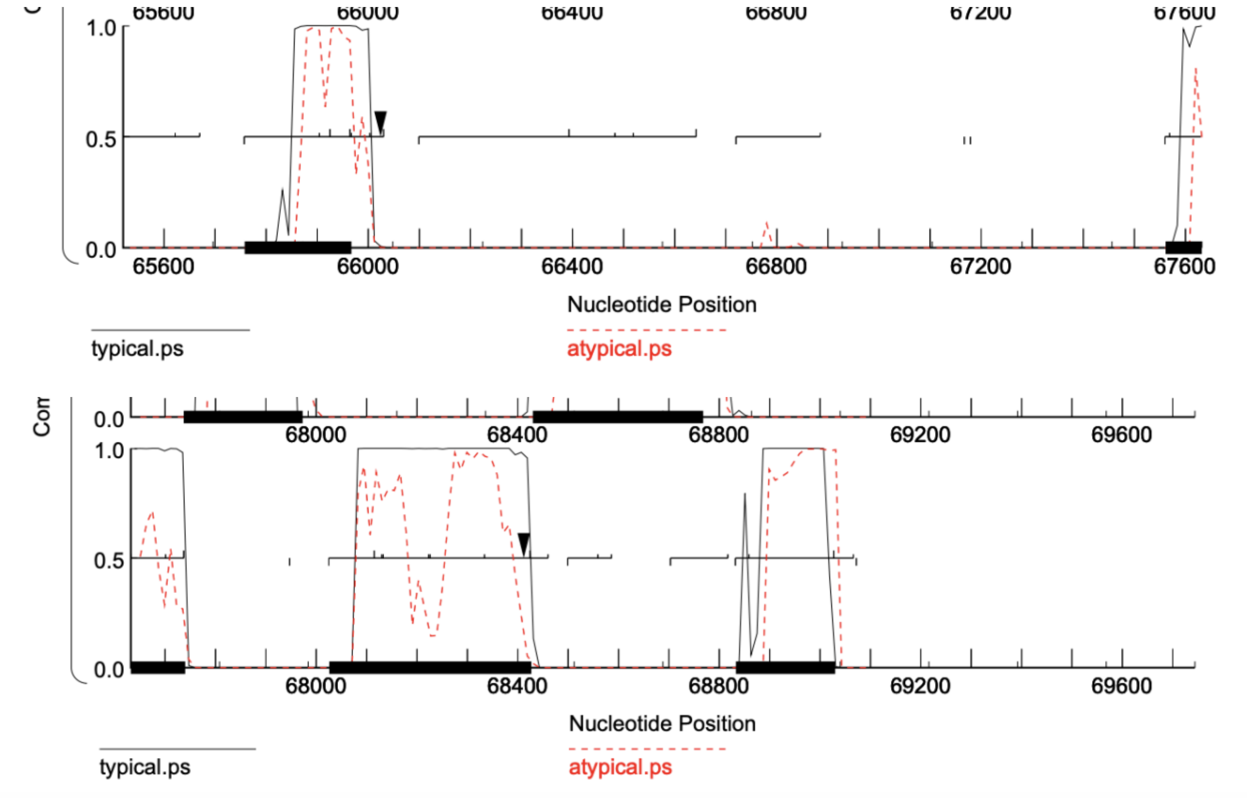
## START POSITION EVALUATION (IN ORDER):

1. Starterator



## 2. GeneMark coding potential

67,740 – 67,561



## 3. Glimmer and GeneMark agreement

YES

Glimmer Start: 67740; Glimmer Score: 8.45; GeneMark Start: 67740
Glimmer call @bp 67740 has strength 8.45; Genemark calls start at 67740

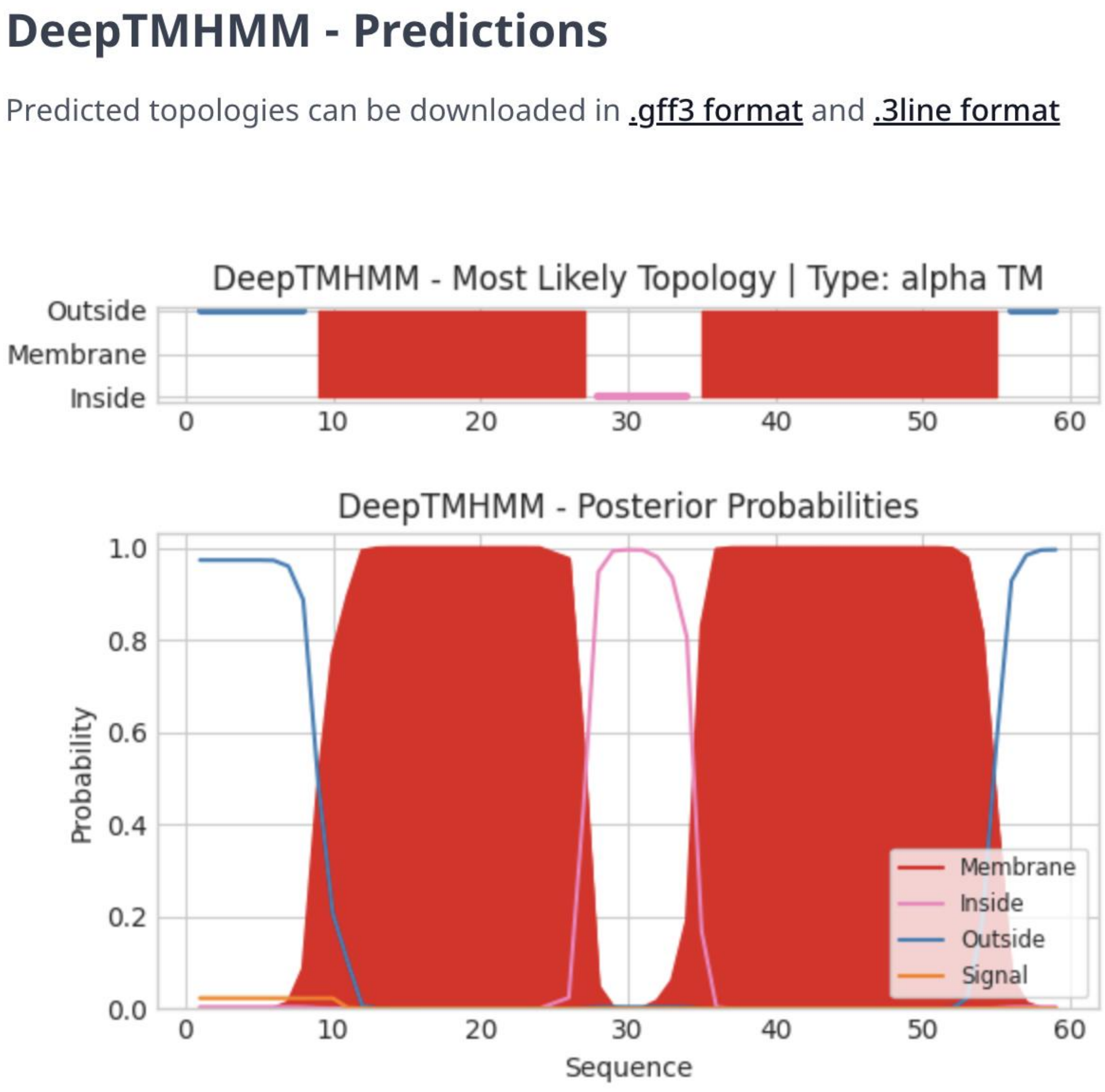
## 4. Longest open reading frame (ORF) without excessive gap

YES

| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 67740 | 67561 | 180 | -4 | 9 | 1.448 | -5.972 | TRUE | ATG | Select | ☑ |
| Reverse | 67704 | 67561 | 144 | 32 | 10 | 1.435 | -5.918 |  | GTG |  | ☐ |

5. If functional prediction, start include full length protein. If no functional prediction is present, write “Hypothetical protein”.

membrane protein; contain at least two transmembrane domains found using DeepTmHmm



## DeepTMHMM - Predictions

Predicted topologies can be downloaded in .gff3 format and .3line format

6. Start site similar to others in GenBank and PhagesDB

Lasso, Usavi, Waterdiva

## 7. Function (F)

Unknown. It may be membrane protein since contain at least two transmembrane domains found using DeepTmHmm.

## 9. Supporting Information for Function (SIF)

### 9. a. SIF-BLAST

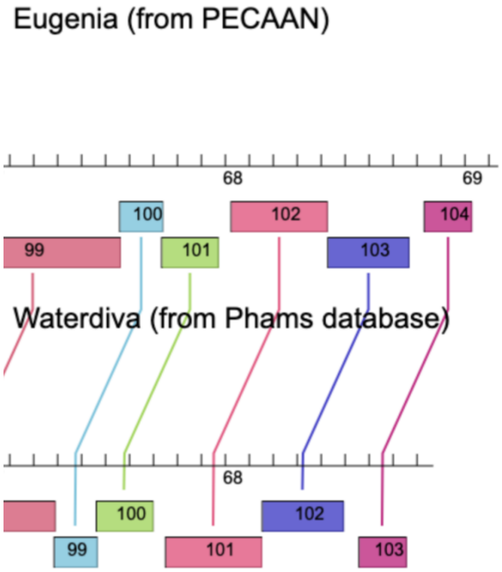
N/A

### 9. b. SIF-HHPred

N/A

### 9. c. SIF-Synteny

Aligns with Waterdiva, ABU, Adriana, Buckeye, CamL, Chah,



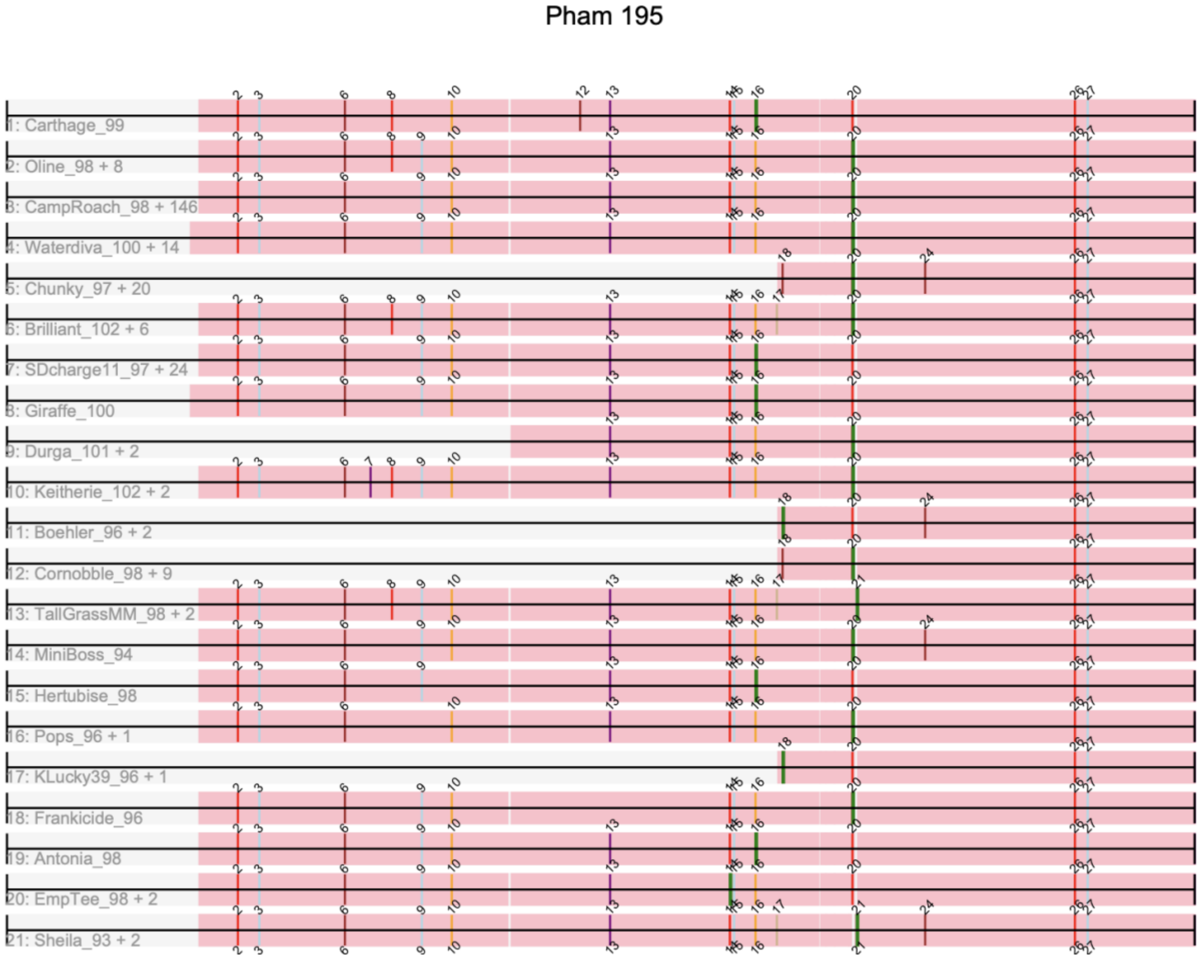
## 10. Any other important information.

## CURATOR NAME: WEI

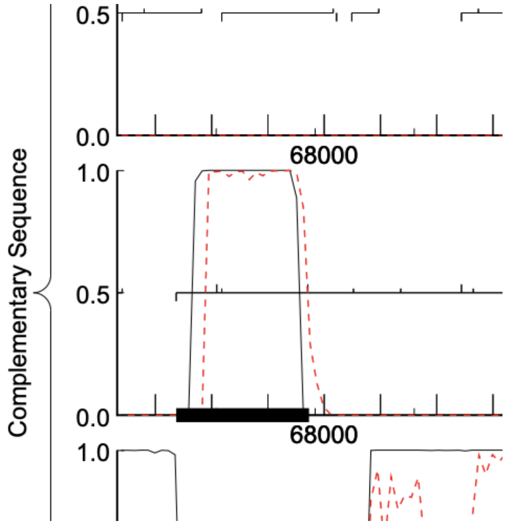
## GENE NAME: EUGENIA-DRAFT_101

## START POSITION EVALUATION (IN ORDER):

1. Starterator



2. GeneMark coding potential



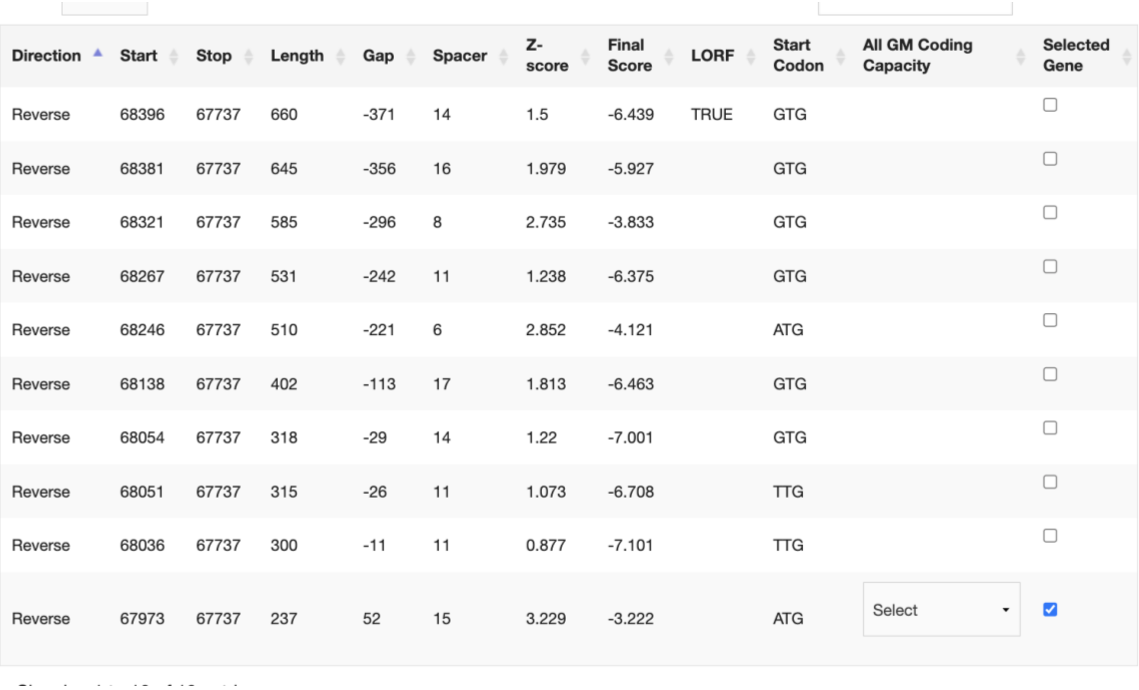
## 3. Glimmer and GeneMark agreement

YES

Glimmer Start: 67973
Glimmer Score: 17.57
GeneMark Start: 67973
Glimmer call @bp 67973 has strength 17.57; Genemark calls start at 67973

## 4. Longest open reading frame (ORF) without excessive gap

No



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 68396 | 67737 | 660 | -371 | 14 | 1.5 | -6.439 | TRUE | GTG | | ☐ |
| Reverse | 68381 | 67737 | 645 | -356 | 16 | 1.979 | -5.927 | | GTG | | ☐ |
| Reverse | 68321 | 67737 | 585 | -296 | 8 | 2.735 | -3.833 | | GTG | | ☐ |
| Reverse | 68267 | 67737 | 531 | -242 | 11 | 1.238 | -6.375 | | GTG | | ☐ |
| Reverse | 68246 | 67737 | 510 | -221 | 6 | 2.852 | -4.121 | | ATG | | ☐ |
| Reverse | 68138 | 67737 | 402 | -113 | 17 | 1.813 | -6.463 | | GTG | | ☐ |
| Reverse | 68054 | 67737 | 318 | -29 | 14 | 1.22 | -7.001 | | GTG | | ☐ |
| Reverse | 68051 | 67737 | 315 | -26 | 11 | 1.073 | -6.708 | | TTG | | ☐ |
| Reverse | 68036 | 67737 | 300 | -11 | 11 | 0.877 | -7.101 | | TTG | | ☐ |
| Reverse | 67973 | 67737 | 237 | 52 | 15 | 3.229 | -3.222 | | ATG | Select | ☑ |

5. If functional prediction, start include full length protein. If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

6. Start site similar to others in GenBank and PhagesDB

MRabcd, MrPhizzler_Draft, Mulan

7. Function (F)

unknown

9. Supporting Information for Function (SIF)

9. a. SIF-BLAST

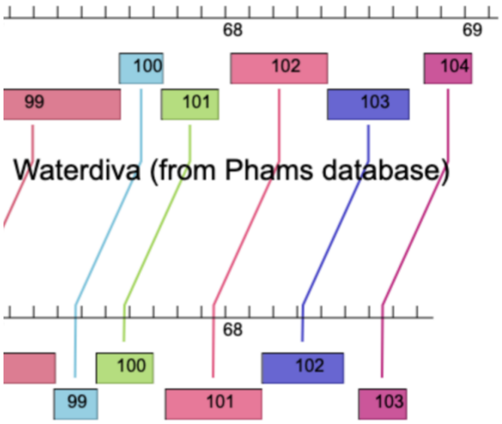
N/A

9. b. SIF-HHPred

N/A

9. c. SIF-Synteny

Aligns with Waterdiva, ABU, Adriana, Buckeye, CamL, Chah, Eugenia (from PECAAN)



Waterdiva (from Phams database)

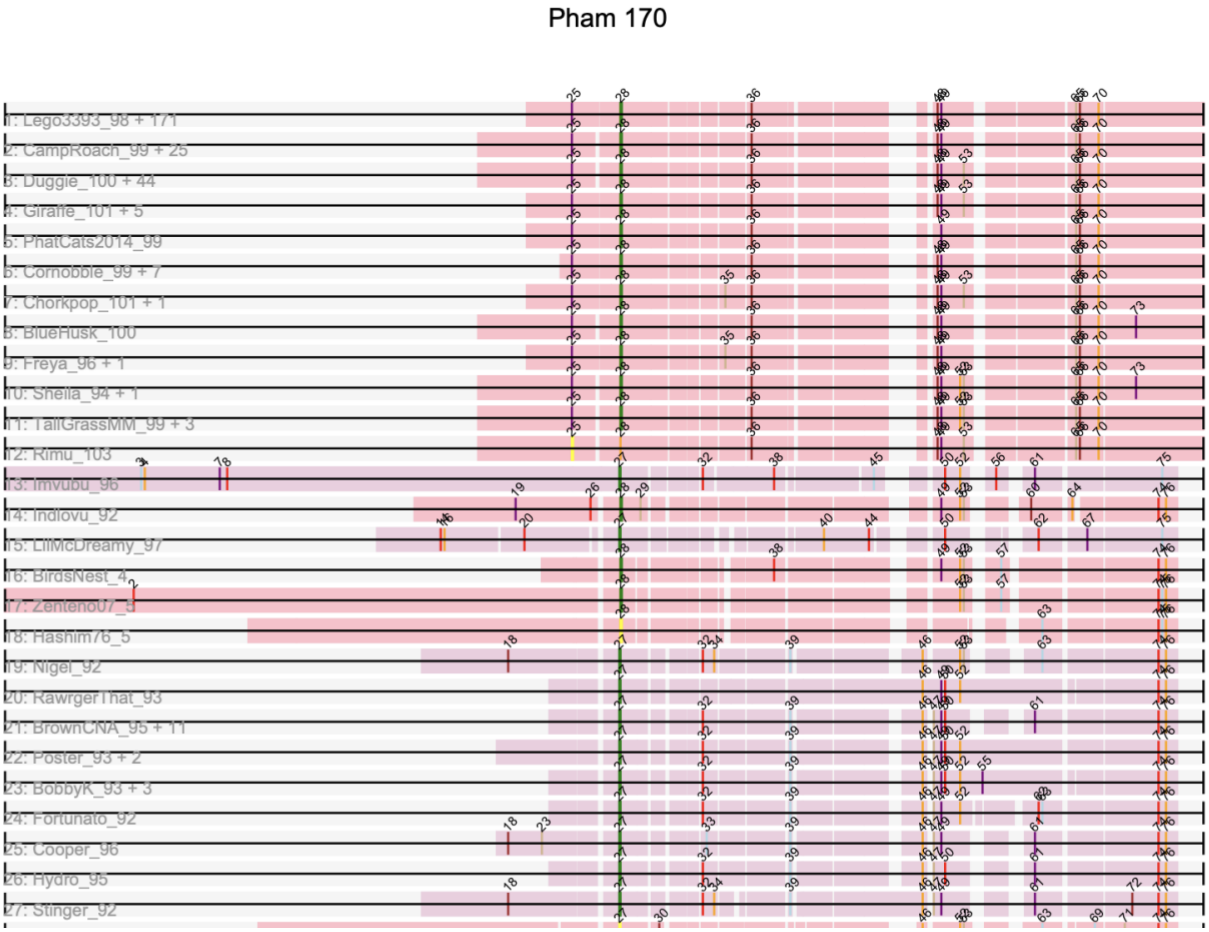
10. Any other important information.

## CURATOR NAME: WEI

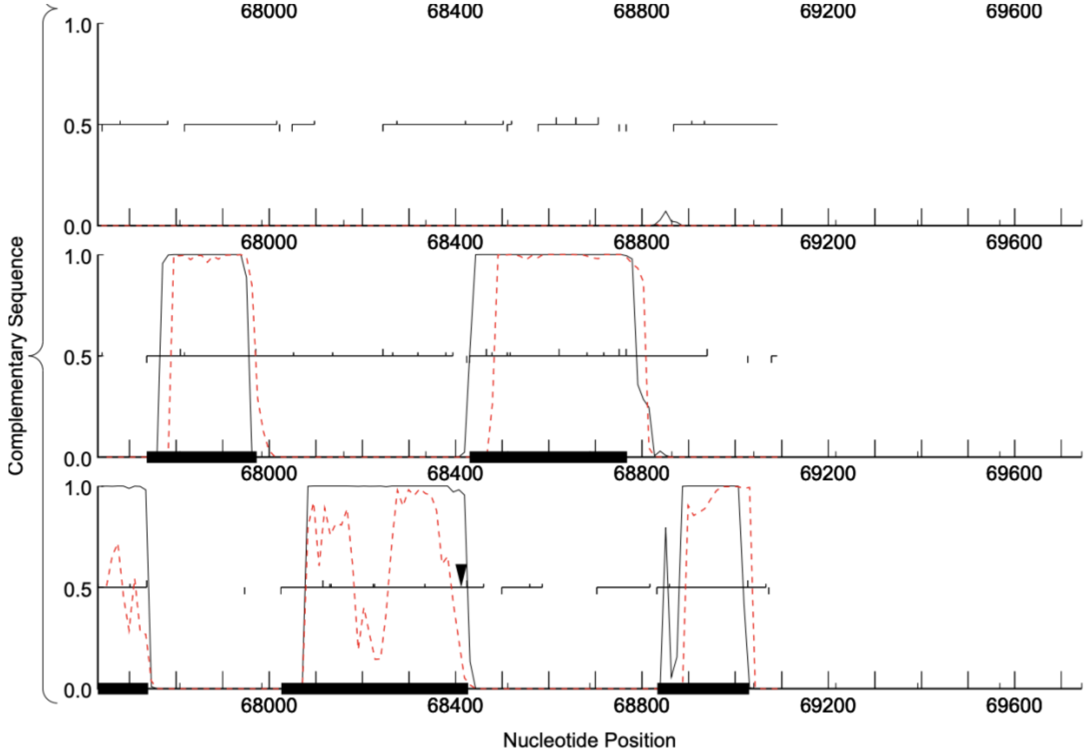
## GENE NAME: EUGENIA-DRAFT_102

## START POSITION EVALUATION (IN ORDER):

1. Starterator



2. GeneMark coding potential



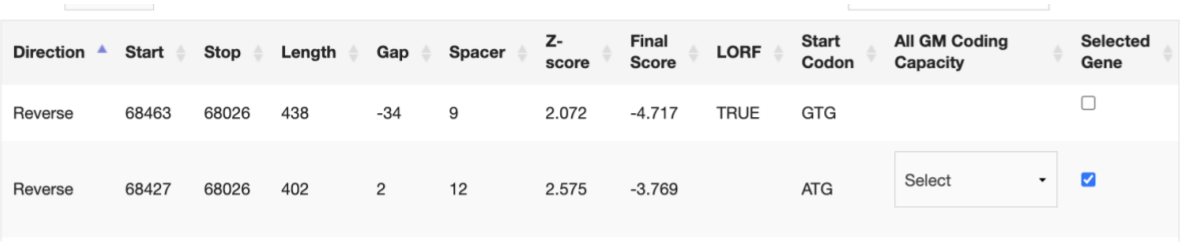
## 3. Glimmer and GeneMark agreement

YES

Glimmer Start: 68427
Glimmer Score: 11.68
GeneMark Start: 68427
Glimmer call @bp 68427 has strength 11.68; Genemark calls start at 68427

## 4. Longest open reading frame (ORF) without excessive gap

No



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 68463 | 68026 | 438 | -34 | 9 | 2.072 | -4.717 | TRUE | GTG |  | ☐ |
| Reverse | 68427 | 68026 | 402 | 2 | 12 | 2.575 | -3.769 |  | ATG | Select | ☑ |

## 5. If functional prediction, start include full length protein. If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

6. Start site similar to others in GenBank and PhagesDB
Jillium, Kahve, Katniss

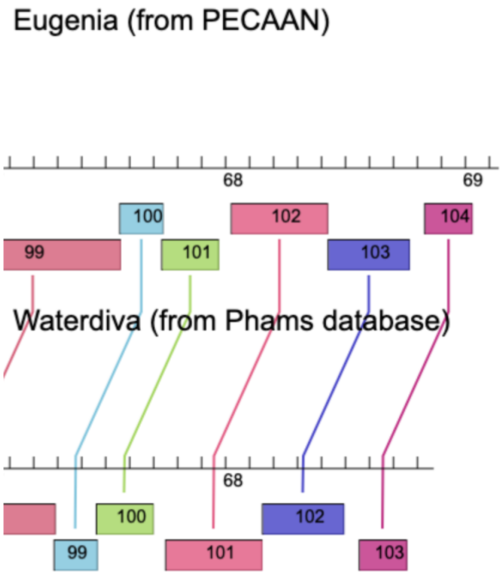
7. Function (F)
unknown

9. Supporting Information for Function (SIF)
9. a. SIF-BLAST

N/A

9. b. SIF-HHPred
N/A

9. c. SIF-Synteny
Aligns with Waterdiva, ABU, Adriana, Buckeye, CamL, Chah,



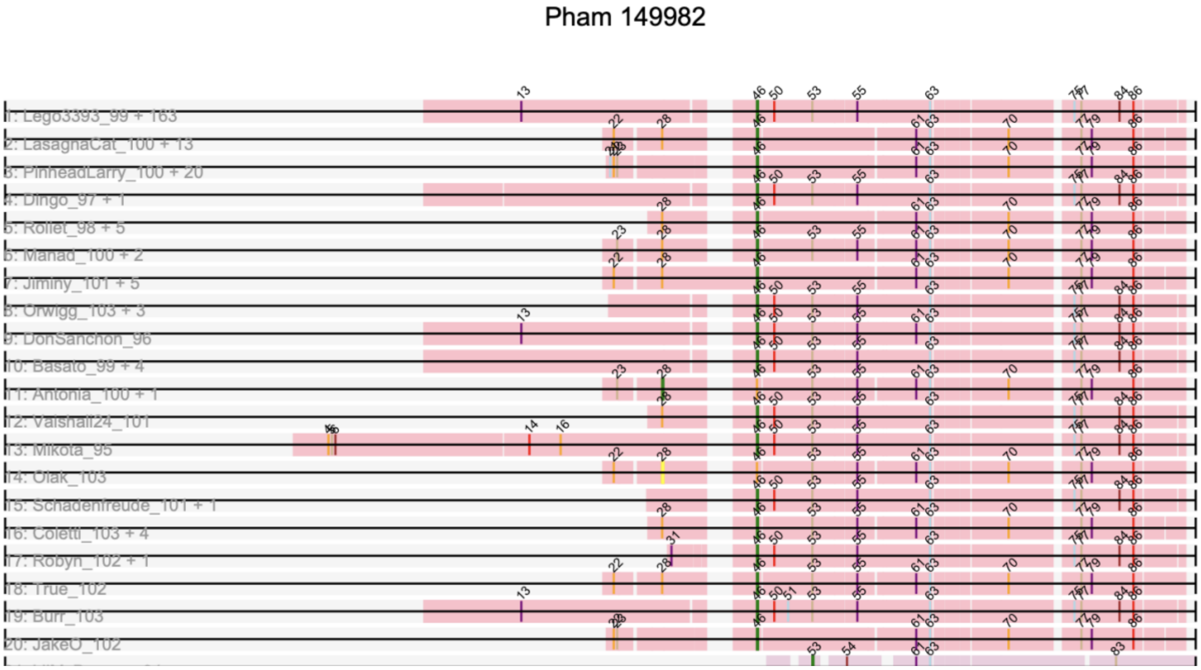
10. Any other important information.

## CURATOR NAME: WEI

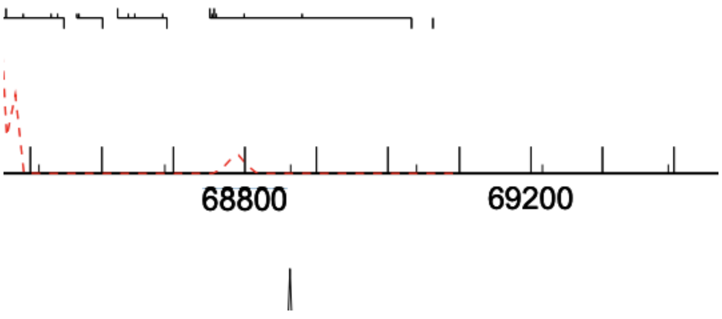
## GENE NAME: EUGENIA-DRAFT_103

## START POSITION EVALUATION (IN ORDER):

1. Starterator



2. GeneMark coding potential



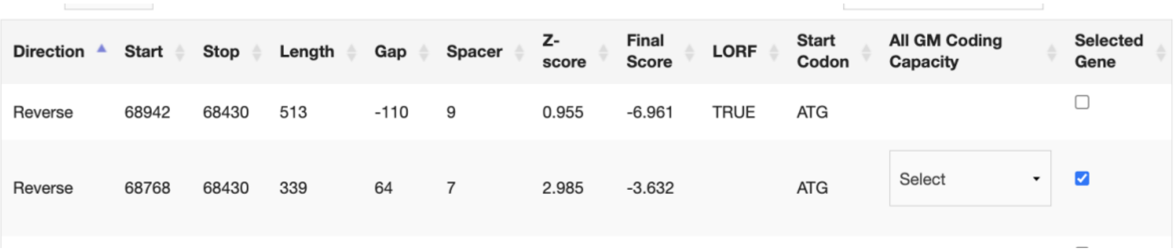
## 3. Glimmer and GeneMark agreement

YES

Glimmer Start: 68768
Glimmer Score: 16.74
GeneMark Start: 68768
Glimmer call @bp 68768 has strength 16.74; Genemark calls start at 68768

## 4. Longest open reading frame (ORF) without excessive gap

No



| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 68942 | 68430 | 513 | -110 | 9 | 0.955 | -6.961 | TRUE | ATG | | ☐ |
| Reverse | 68768 | 68430 | 339 | 64 | 7 | 2.985 | -3.632 | | ATG | Select | ☑ |

## 5. If functional prediction, start include full length protein. If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

## 6. Start site similar to others in GenBank and PhagesDB

Kikipoo, KingTut, Kloppinator

## 7. Function (F)

Unknown

## 9. Supporting Information for Function (SIF)

## 9. a. SIF-BLAST

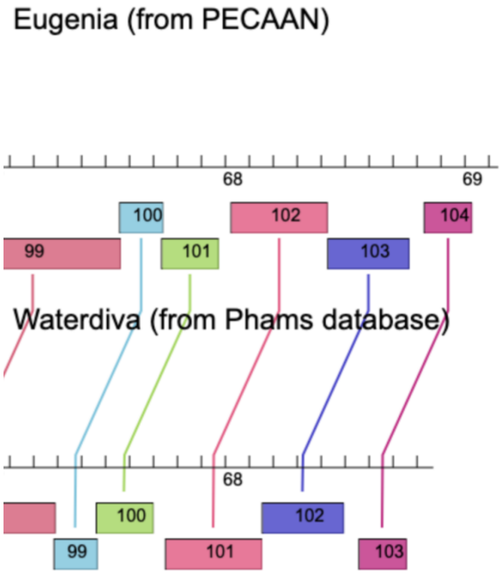
N/A

9. b. SIF-HHPred

N/A

9. c. SIF-Synteny

Aligns with Waterdiva, ABU, Adriana, Buckeye, CamL, Chah,



10. Any other important information.

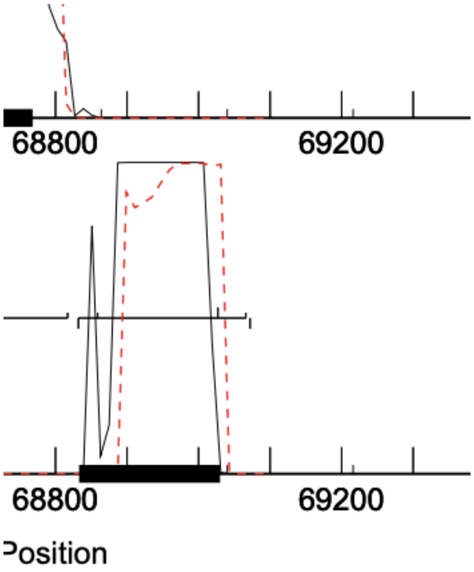
## CURATOR NAME: WEI

## GENE NAME: EUGENIA-DRAFT_104

## START POSITION EVALUATION (IN ORDER):

1. Starterator

2. GeneMark coding potential



## 3. Glimmer and GeneMark agreement

YES

Glimmer Start: 69030
Glimmer Score: 14.05
GeneMark Start: 69030
Glimmer call @bp 69030  has strength 14.05; Genemark calls start at 69030

## 4. Longest open reading frame (ORF) without excessive gap

No

| Direction | Start | Stop | Length | Gap | Spacer | Z-score | Final Score | LORF | Start Codon | All GM Coding Capacity | Selected Gene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Reverse | 69069 | 68833 | 237 | | 13 | 1.724 | -5.688 | TRUE | GTG | | ☐ |
| Reverse | 69030 | 68833 | 198 | | 10 | 2.735 | -3.306 | | ATG | Select | ☑ |

## 5. If functional prediction, start include full length protein. If no functional prediction is present, write “Hypothetical protein”.

Hypothetical protein

## 6. Start site similar to others in GenBank and PhagesDB

KingTut, Kloppinator, Kwadwo

## 7. Function (F)

unknown

## 9. Supporting Information for Function (SIF)

### 9. a. SIF-BLAST

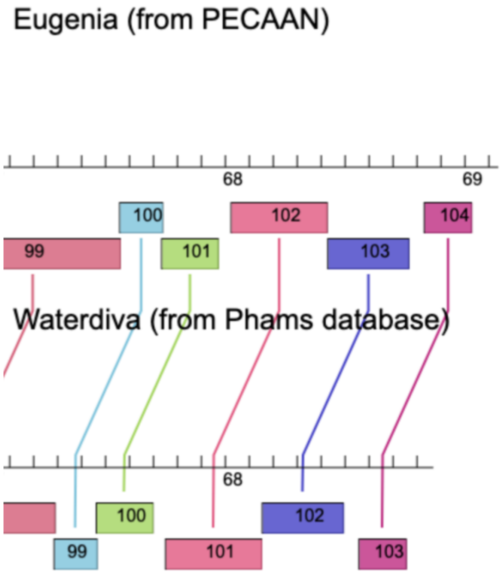
N/A

### 9. b. SIF-HHPred

N/A

### 9. c. SIF-Synteny

Aligns with Waterdiva, ABU, Adriana, Buckeye, CamL, Chah,



## 10. Any other important information.