Case Study: *Peaches gene 1*

Phage Genome: *Peaches*

Genome Coordinates: *457-750*

1. Collect Evidence: insert screenshots relevant to THIS REGION of the genome as indicated below. of GeneMark, BLAST results, Aragorn, Phamerator, HHPred, choose start window, frames window, etc. Make sure that your evidence is readable and complete.

A: *screenshot of BLASTN results on phagesdb.org*

B: *screenshot of phagesdb.org pham page with functions*

C: *Autoannotation results for this region (what do the DNA Master notes say are the Glimmer and GeneMark results?)*

D: *screenshot of the DNA Master frames window*

E: *screenshot of GeneMarkS output. If no coding potential in region, add GeneMark-host output(s)*

F: *screenshot of regional Phamerator map, with relevant phages*

G: *(if necessary) screenshot of DNA Master “Choose Start” window*

H: *screenshot(s) of DNA Master BLAST results, including alignment tab(s)*

I: *(if necessary)* *screenshot(s) of Starterator report*

J: *screenshots of HHPred results*

K*: (if necessary) screenshot of Aragorn results*

L*: (if necessary) screenshot of tRNAscanSE results*

M*: (if necessary) screenshot of Phamerator map of entire genome*

N: *other*

2A. Is this a gene? \_\_457-633rev No; 538-765 Yes\_\_

Rationale: Write a one-sentence explanation for each piece evidence above and how it supports your answer to 1, or write NI for not informative or N/A for not applicable

A: This region in the Subcluster A4 phages is highly conserved, across dozens of genomes. If there is a gene here, it should be in all of them (unless there is a point mutation).

B: For 457-633rev, there are only 14 members of this pham, and only 8 are not drafts. Based on A, there should be a lot more genes here, so this is suspicious.

C: Only one of the gene prediction programs chose 457-633rev gene. None chose a gene oriented in the forwards direction.

D: 457-633rev does not overlap with any of the other predicted ORFs, which is good. I don’t really like the rev ORF currently selected, and I can see there are alternate ORFs that need to be investigated in the same region.

E: There is a little tiny blip of coding potential in the rev ORF in question. There is also a tiny blip of coding potential in the top forwards frame in approximately the same region at 538-765.

F: There is some discrepancy in the annotations of the A4s in this area. Some of the draft annotations have no predicted gene in this region, Peaches has the reverse gene, and the final genomes have a gene oriented in the forwards direction.

G: This refers to the forwards ORF at 538-765 because I do not think the reverse prediction is a gene. There is only one possible start; the second would not yield a long enough ORF.

H: BLAST results show that the forwards gene at 538-765 has been added to the annotations of numerous A4 phages.

I: n/a, as there is only one possible start for the forwards ORF.

J: HHPred results of the forwards ORF at 538-765 shows that there is strong support for a HNH endonuclease here. The rev ORF does not have support for anything in HHPred.

K: n/a

L: n/a

M: n/a

2B. Rank the evidence from most-to-least compelling. include a one -to -two sentence rationale to support your order.

J. HHPred alignment to an HNH crystal structure with a 99% probability over ~90% of the sequence is the most compelling.

H. Presumably the same HHPred alignment applies to these other phage genes, and that is why these gene have been annotated in this way, and there is BLAST data for my gene.

F. The comparison between nucleotide sequences is “purple” throughout all these genomes, and the final annotations have the HNH included in the annotation.

A. all the A4s are identical in this region, such that if one of them has a gene, they should all have it.

If your answer to 2 is “no”, you are done with this case study.

If not:

3. What is the start coordinate for your gene? \_\_\_\_*538*\_\_\_\_

Refer to evidence in part 1.

Rationale: Write a one-sentence explanation for each piece evidence you’ve included and how it supports your answer to 3 or write NI for not informative or N/A for not applicable

A: n/a

B: n/a

C: n/a

D: There are only two possible start codons for this ORF, and the downstream one does not yield a long enough gene to be a gene.

E: n/a

F: it appears that the 538 start would add a gene that is approximately the same size and at the same position as the other A4s with finished annotations.

G: There are only two possible start codons for this ORF, and the downstream one does not yield a long enough gene to be a gene. The RBS scores for both starts are terrible.

H: My sequence aligns 1:1 with BellusTerra and a number of other A4 phages.

I: n/a

J: the HHPred alignment extends upstream of my first amino acid in my sequence, but I don’t have a start choice that could make my sequence even longer.

K: n/a

L: n/a

3.B Rank the evidence from most-to-least compelling, include a one-to-two sentence rationale to support your order.

J: any shorter and I would miss out on crystal structure alignment

D. only one of the two possible start choices yields a lengthy enough product.

G. same reason as D.

H. my sequence aligns 1:1 with a number of other A4 genes. This is good because the underlying sequences appear to be very similar.

4A. What is the function of this gene? (choose from approved list): (HNH endonuclease*)*

4B: Rationale for evidence:

Rationale: Write a one-sentence explanation for each piece evidence you’ve included and how it supports your answer to 4A or write NI for not informative or N/A for not applicable

A: n/a

B: n/a

C: n/a

D: n/a

E: n/a

F: the final annotations in Phamerator show this gene as “HNH endonuclease”

G: n/a

H: the identical phage protein sequences in GenBank are labeled “HNH endonuclease”

I: n/a

J: the HHPred alignment shows that ~90% of my protein query sequence aligns to a crystal structure of an HNH endonuclease with a probability of 99% (number three in the HHPred list).

K: n/a

L:n/a

4C: Rank the evidence from most-to-least compelling. Include a one-to-two sentence rationale.

J. The the HHPred alignment shows that ~90% of my protein query sequence aligns to a crystal structure of an HNH endonuclease with a probability of 99% (number three in the HHPred list). Crystal structures and their associated functional assignments are among the most reliable entries in databases. The alignment runs for most of the query sequence, and only omits the first 30 and last 10 residues of the template crystal structure.

H. Presumably similar HHPred evidence exists for all these phage protein sequences in GenBank, as they are identical in amino acid sequence to the one I am exploring.

F. Due to the evidence in J, phage genes have been assigned the HNH endonuclease function which results in their presence in GenBank and on Phamerator maps.

5. How does this function influence your start choice or gene/not a gene decisions?

a. confirms

b. required adjustment of start or gene prediction

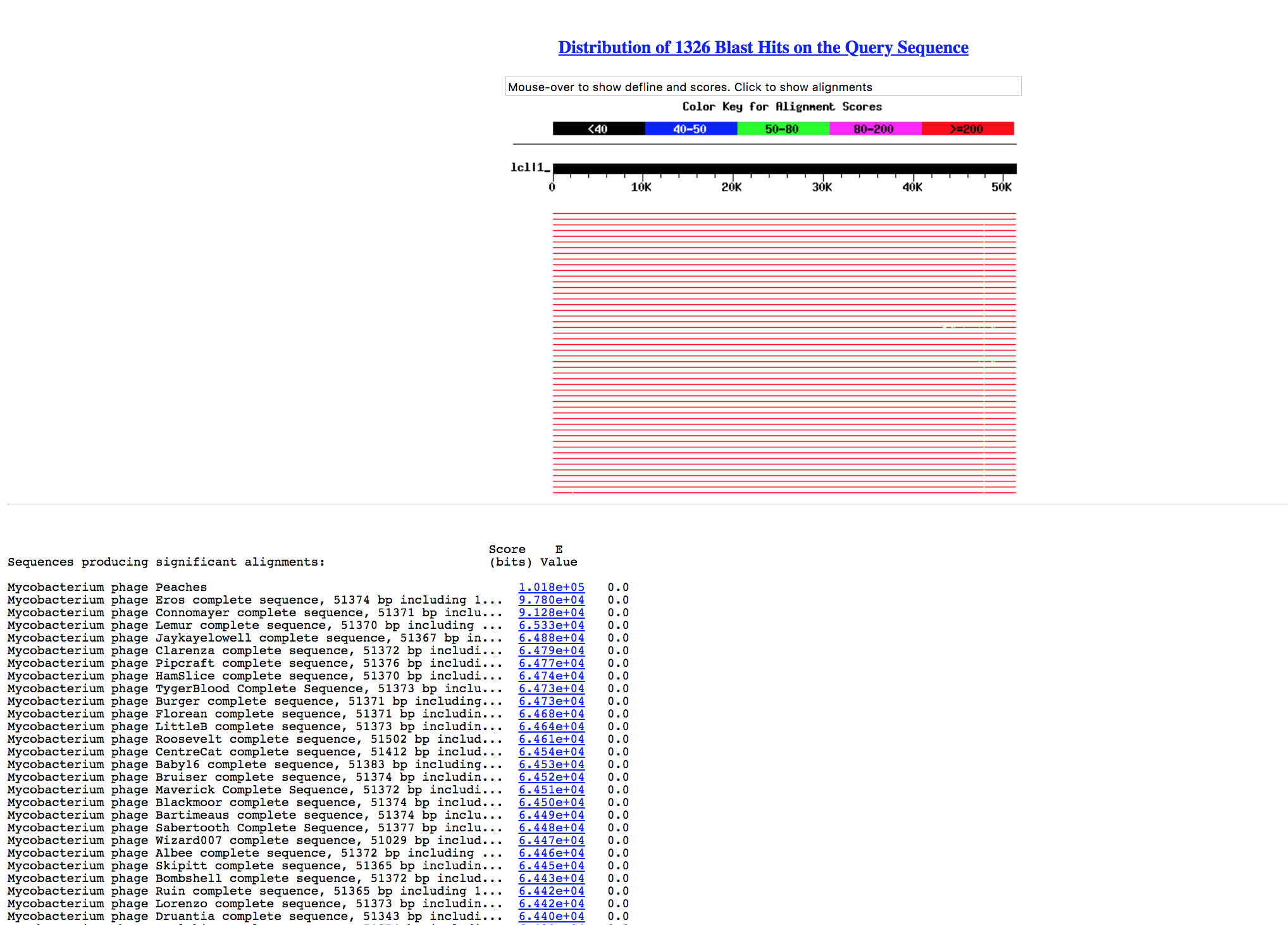
c. not informative

A: confirms

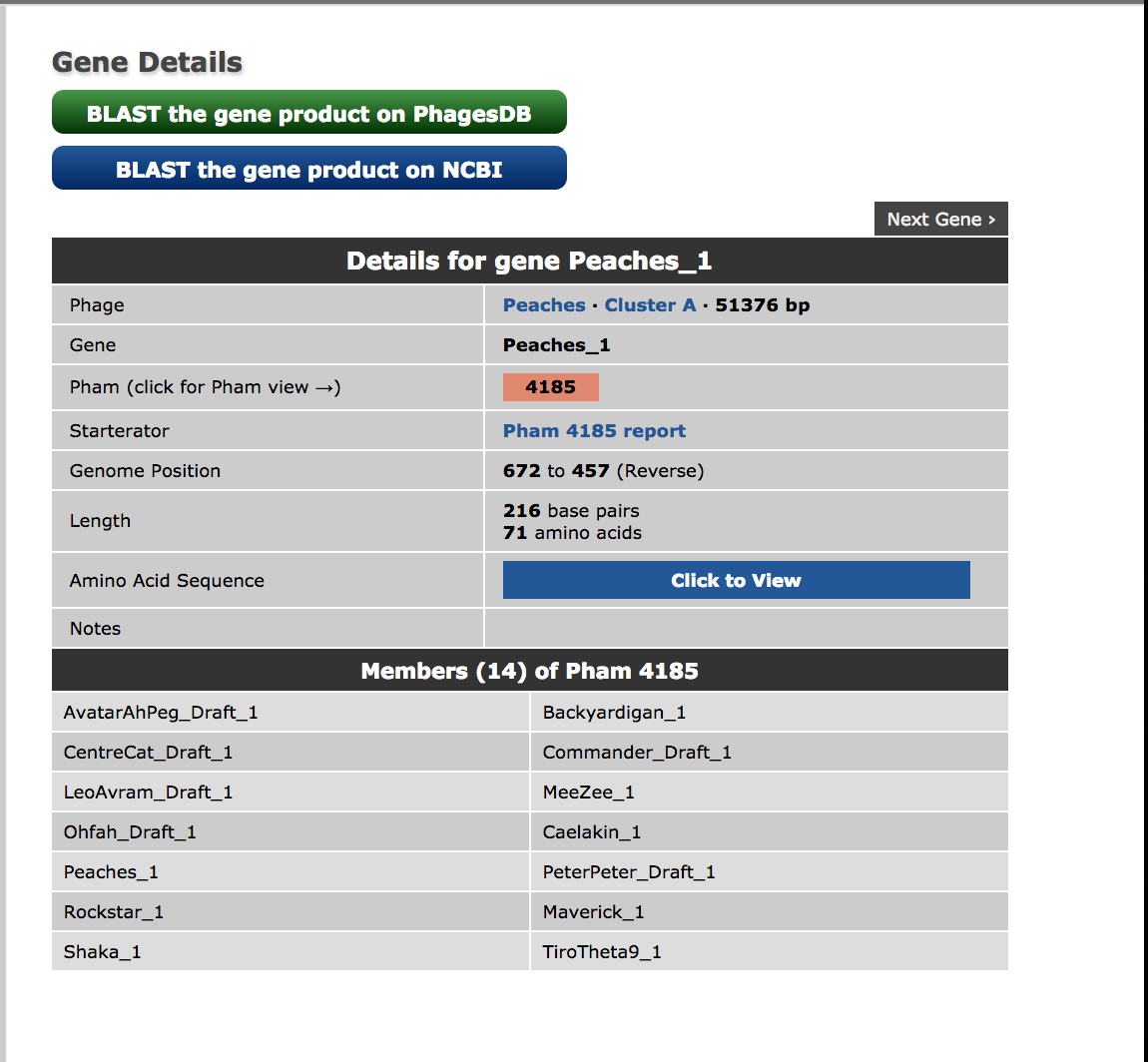
6. Does this function make sense in the context of the genome as a whole? Refer to the list of common phage functions, their most common genome locations, and the number of times they usually appear in a genome.

yes, in many phages there is an HNH endonuclease gene that is located near the terminase gene in the genome. This particular HNH endonuclease aids the terminase in DNA packaging.

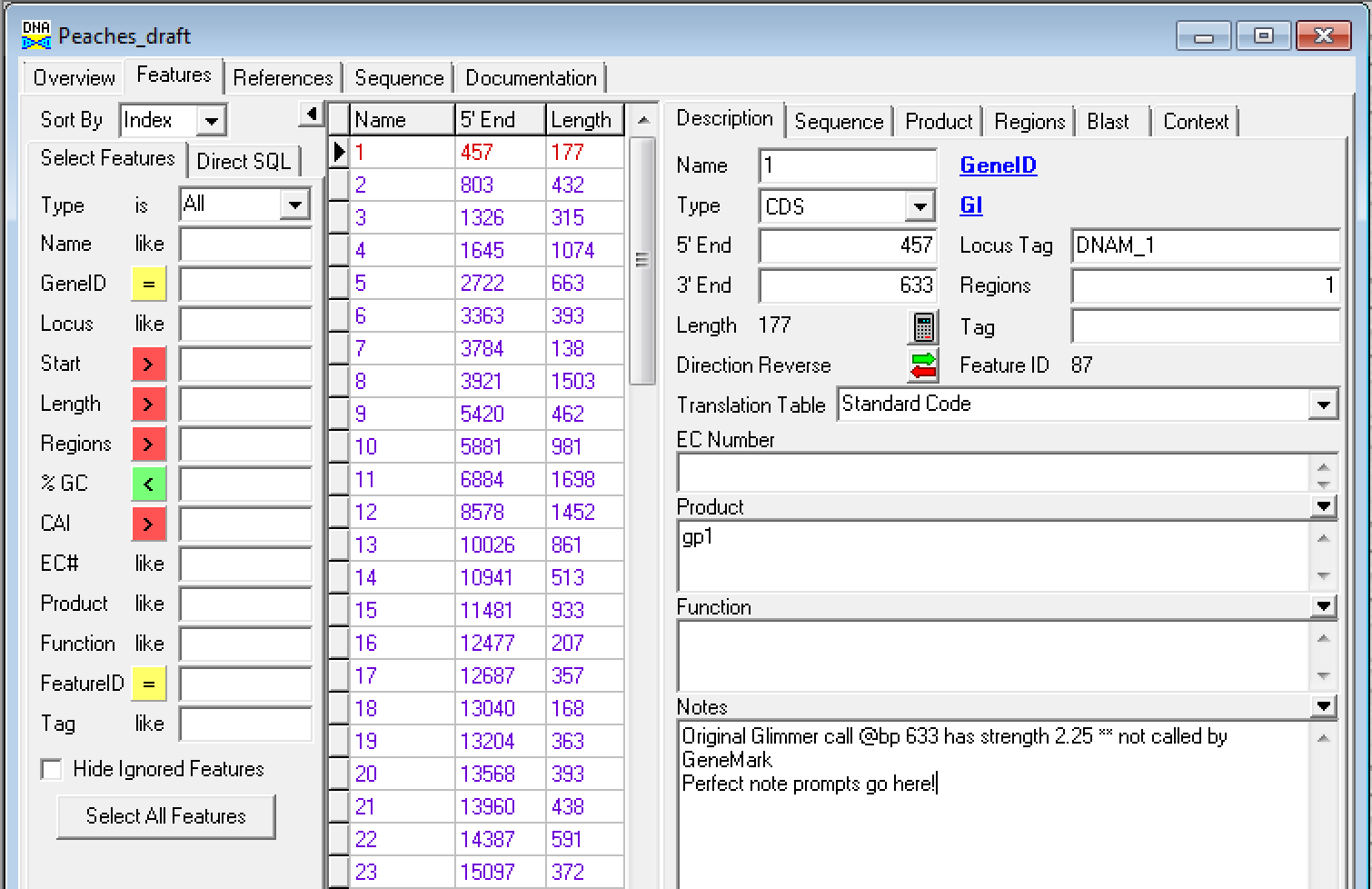
A.



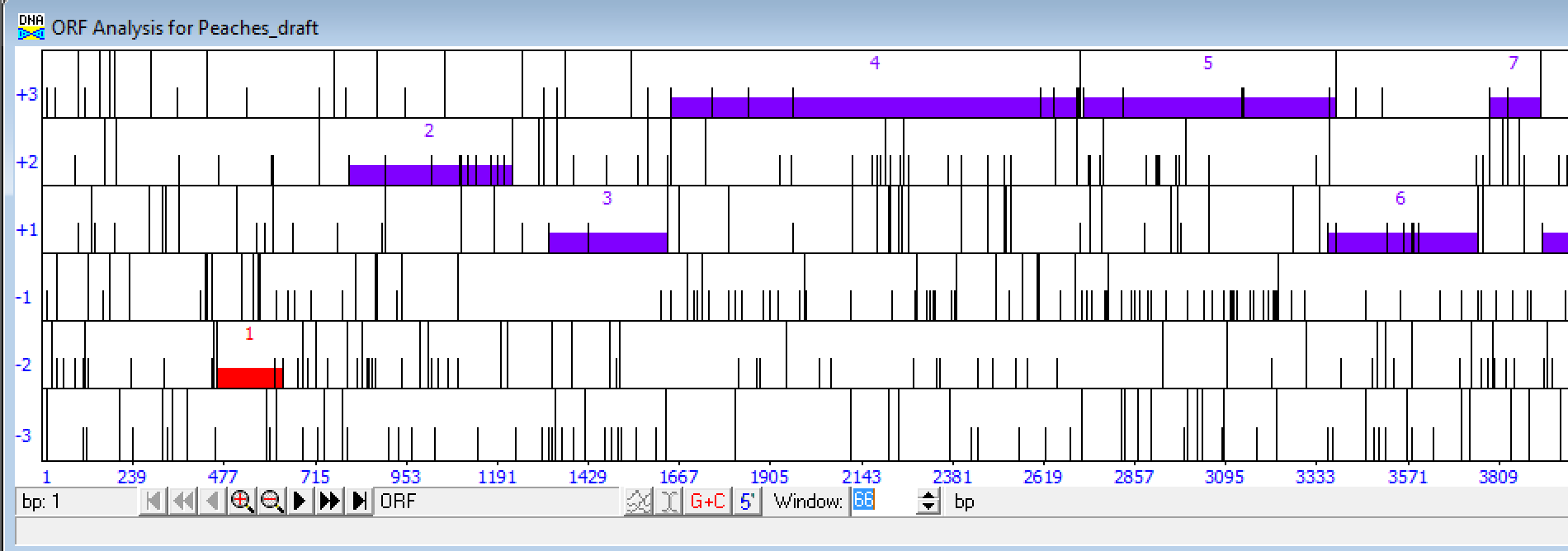
B.



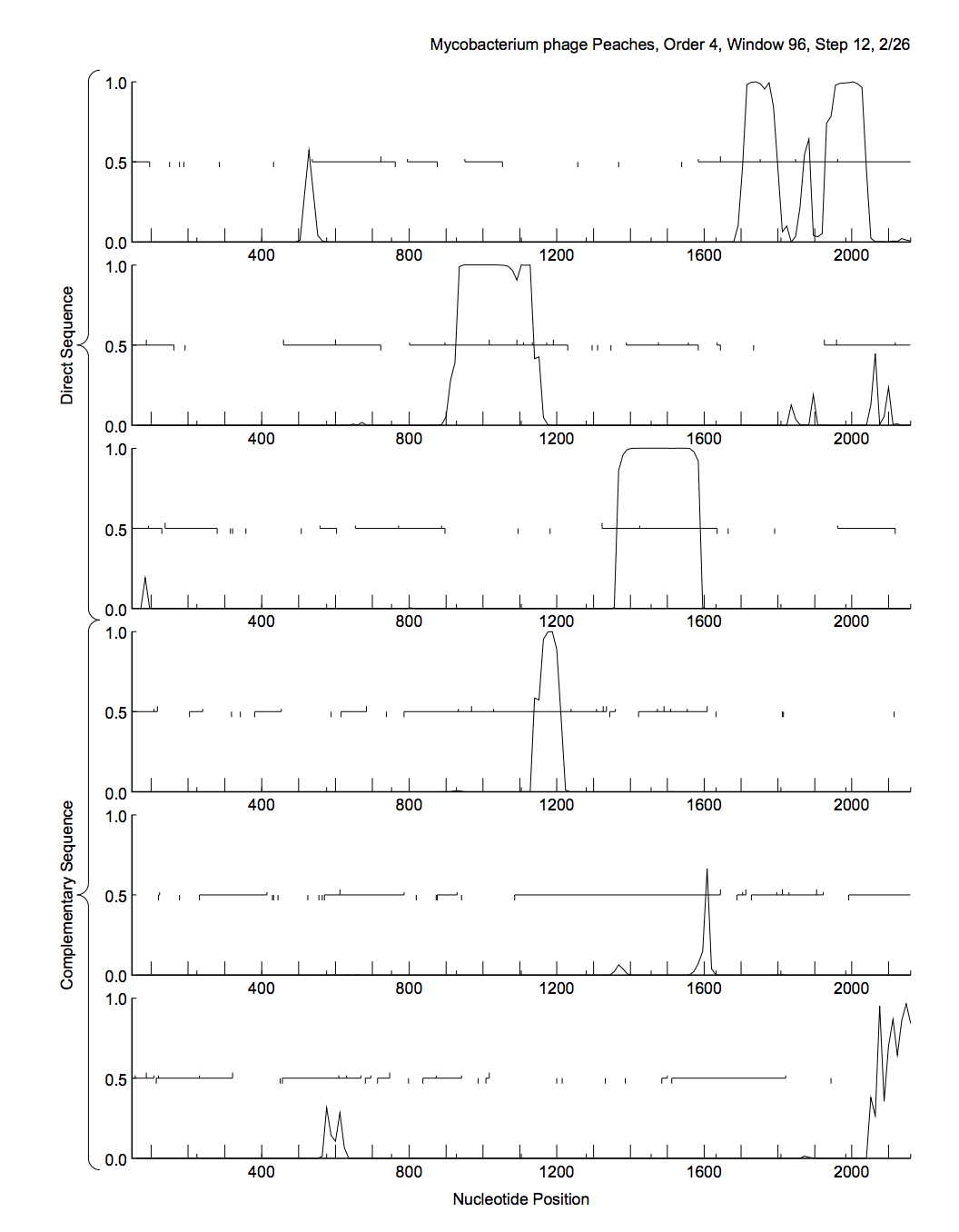
C.



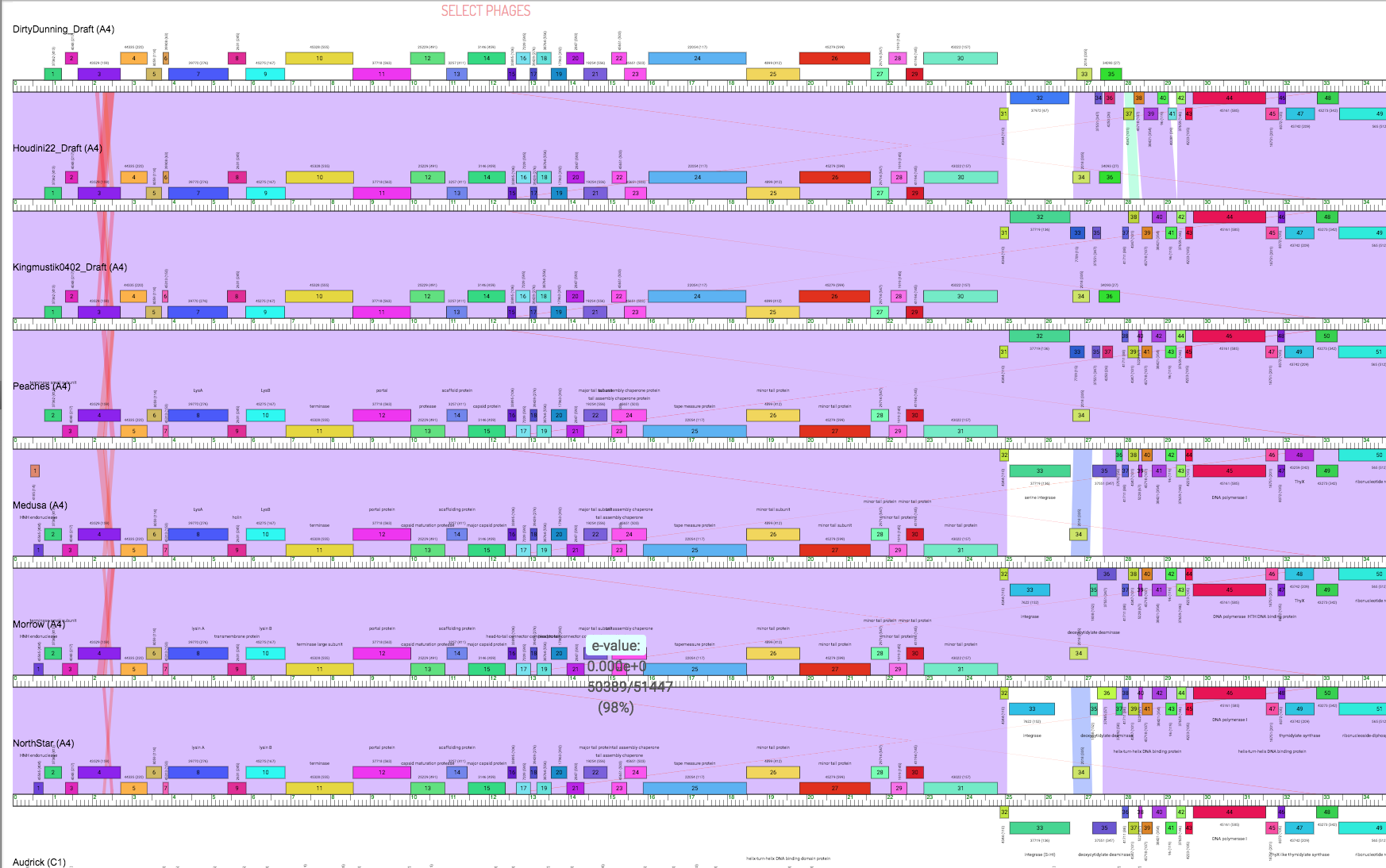
D.

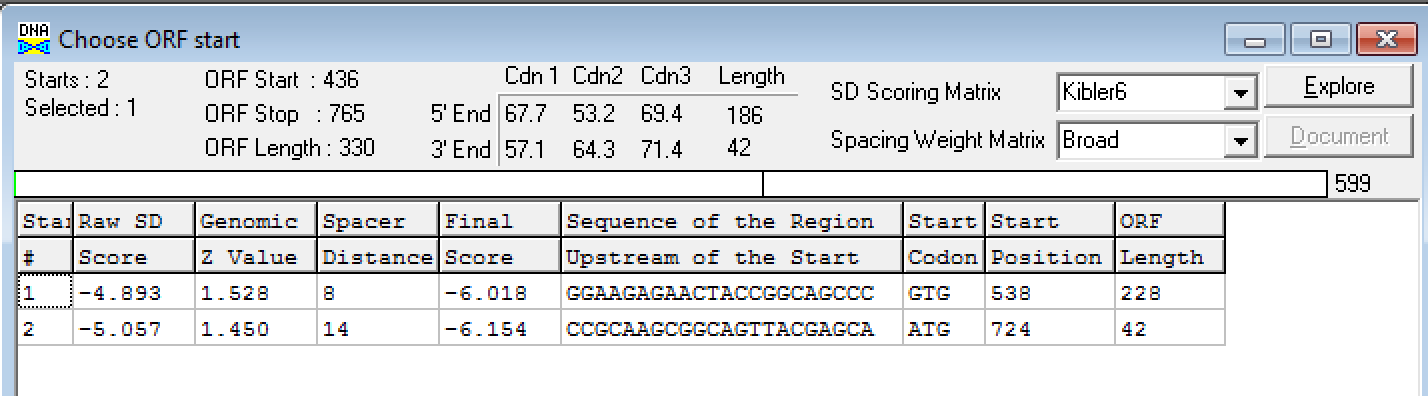


E.

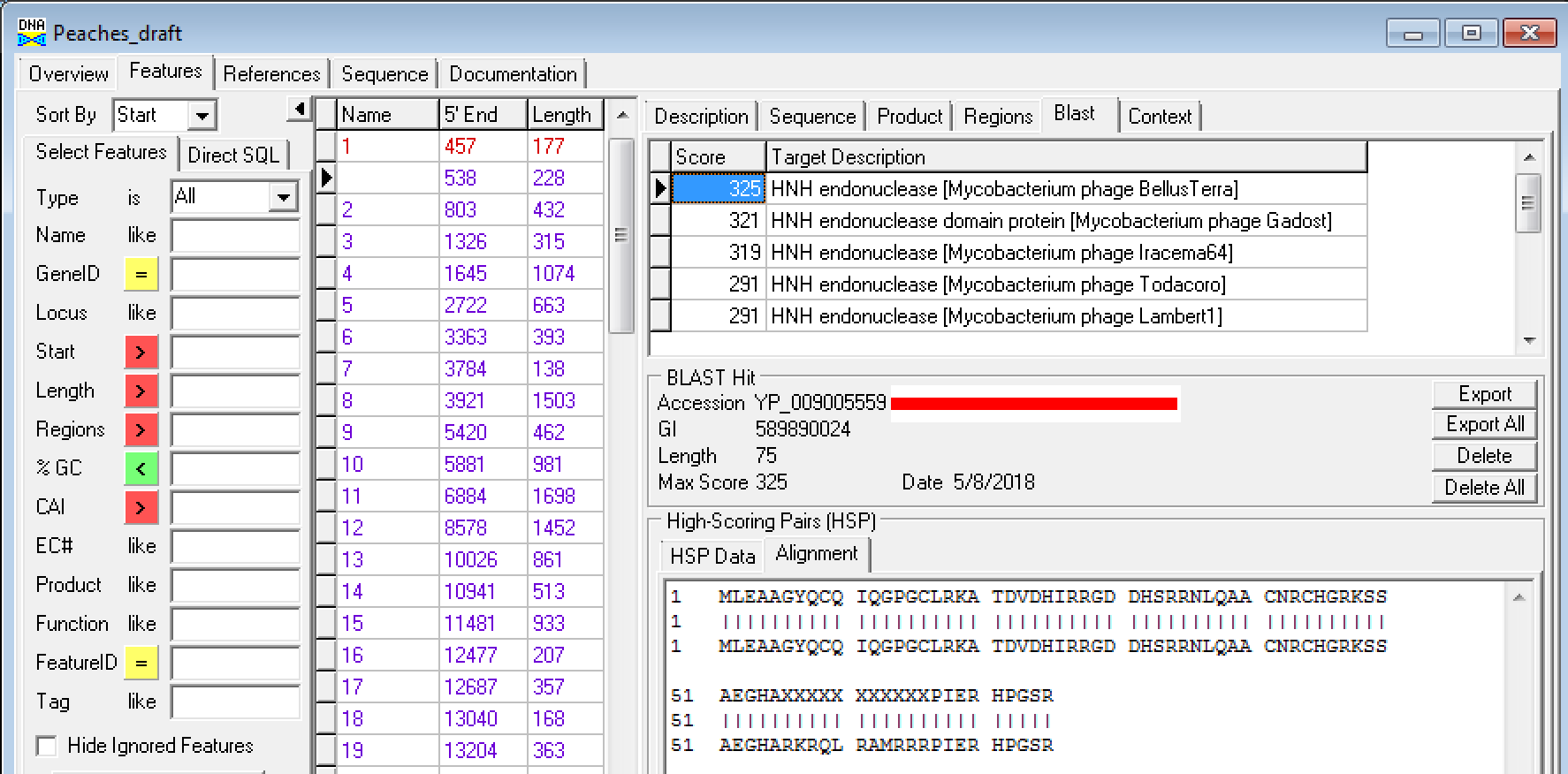


F.



G.

H.



J.