The “**Additional Notes**” section found in some gene’s notes were conducted on Phagesdb.org protein blast, in order to look for alignment to and functions of previously QC annotated phages.

**Starterator notes:**

I do not have the newest VirtualBox, and I believe that is why I was experiencing some problems with Starterator. When I try to run the report, Starterator states “Startstaerator has encountered an error. Please try again.” I attempted to run the report for each individual gene, and the report successfully ran for the following genes; **6**, **7**, **11**, **12**, **13**, **14**, **15**, **21**, **32**, **34**, **36**, **37**, **47**, **52**, **53**, **57**, **59**, **61**, **62**, **65**, **66**, **67**, **69**, **70**, **75**, **76**, **79**, **80**, **82**, **83**, **86**, **87**, **93**, **98**, **103**, **104**, **105**, **106**, **107**, **115**. Therefore, these genes are reported properly in the notes. For the genes that did not work, I used the Starterator index found here (<http://phages.wustl.edu/starterator/>). I used phamerator to find the Pham number, and clicked on the Pham report in the index. It does not suggest a start site for each individual phage, but it present the most annotated start site for that Pham. Therefore, for the genes that didn’t work in Starterator report, I reported “Has most annotated start site” or “Does not have the most annotated start site”, using the index. Is there any way I can fix this error or update the software? My email is ConboyA@chc.edu. Thanks so much!

**Gene 6** - SugarP\_Isomerase domain from NCBI Blast, but HHPred (and phagesdb blast) shows immunity protein, although alignment is off (3 to 4).

**Gene 9** - Start site changed from 7317 to 7314 to align 1-to-1 with previously QC annotated PrincessTrina\_10. Start site change slightly worsens SD score, but gives better alignment to blast results.

**Gene 19** - Has good HHPred hit in terms of probability, but a not so good e-value. Alignment is also off to other phage homologue genes.

**Gene 20** - type III restriction-modification system StyLTI enzyme res from BlastP multidomains

**Gene 24** - Although 1:1 alignment to PrincessTrina, I couldn't find any other evidence to call this a tail protein.

**Gene 25** - Called this gene a minor tail protein due to HHPred results, and alignment to PrincessTrina\_27, Jawnski\_21, and a few other phages who all list protein as minor tail.

**Gene 29** - Considered calling this gene a Holin because of alignments to multiple previously QC annotated phages including BarretLemon\_27, Martha\_27, Sonny\_27, and TaeYoung\_27. These are all part of the same Pham 8364, as is Colucci, even though they are in different clusters. Additionally, their e-values are good (E-15 to E-29) However, I couldn't find other evidence supporting Holin function.

**Gene 31** - The 1:1 alignments listed as a Holin were originally called a Holin in PrincessTrina\_32 because it was adjacent to a lysin gene, and was given modest HHPred matches to a holin. I have listed that result in the notes of gene 31.

**Gene 32** - Best blast results were to Camplyobacter bacteria strains. Start site was changed from 28361 to 28400 to match GeneMark. This gave a much longer ORF, and better blast alignment results. It also yielded a better SD score. However, it does make the overlap of the next gene longer (16 bp), and goes against Starterator. As for gene function, BlastP multidomain showed a good result; protein listed as “acyl carrier protein, Synthase II”.

**Gene 33** - Start site was changed from 28509 to 28557 to match GeneMark. This gave a longer ORF, and better blast alignment.

**Gene 37** - was deleted because it was a forward gene amongst reverse genes. Additionally, it was a very small ORF and mostly overlapped gene 38. It also did not have any good blast results. Lastly, the gene was not called by GeneMark.

**Gene 42** - HHPred gives a pretty good result, but with a slightly higher e-value than we'd like. Hesitated to call a function based on the HHPred result because of the organism it was found in and because of the evalue.

**Gene 49 -** Start site changed from 38339 to 39303 to increase SD score, decrease gap, increase ORF, better e-values on blast results.

**Gene 51** - Start site changed from 38995 to 38917, although this start site was not suggested by Glimmer or GeneMark. However, this new start site yields a 1:1 alignment with many previously QC annotated phages including PrincessTrina, ArV1, HumptyDumpty and EdgarPoe. This site also reduced the gap by a lot, gave the best SD score, the longest ORF, and gave better blast results. Although this start site change does increase alignment to many phage genomes, the original start site (38995) offers an interesting 1:1 alignment with Arthrobacter bacteria species "hypothetical protein [Arthrobacter sp. Soil763] - score = 100 bits(249); e-value = 2e-24; method = compositional matrix adjust; identities = 56/129(43%); positives = 72/129(55%); gaps = 2/129(1%).

**Gene 52** - was deleted because it was a reverse gene amongst forward genes. It did not have any significant blast results on BlastP, PhagesDB Protein Blast, HHPred, etc.

**Gene 57** - was deleted because it was a very small ORF (40 amino acids) that showed no significant blast results, alignments to any other phages via phagesdb.org protein blast, and was also not called by GeneMark at all.

**Gene 59** - was deleted because it was a very small ORF (30 aa) and did not have any significant blastp, hhpred or phagesdb results.

**Gene 67** - changed ss from 44903 to 44921; better alignment, agrees with starterator, but shorter ORF.

**Gene 70** - goes against Starterator (but agrees with glimmer and genemark).

**Gene 74** - No function was called, but BlastP has a good match to a superfamily over the interval 49-128.

**Gene 80** - No function was listed, but the BlastP superfamily covers a large portion of this short gene, with a good e-value. (Periplasmic glucose/galactose-binding protein (GGBP) involved in chemotaxis).

**Gene 85** - Start site was changed from 53346 to 53289. This change made alignments to blast result much better (from 20:1 to 1:1), gives a slightly better SD score, includes all coding potential, and gives a larger ORF. However, it does give a rather large overlap of the previous gene (34 bp). Function had

**Gene 93** - was deleted because it not called by GeneMark, had no significant hits on blast, HHpred, phamerator and starterator. Additionally, it did not match any ORF from other phages (blasted on PhagesDB.org).

**Gene 111** - Agrees with Glimmer because of better alignments. BlastP yields a modest hit and e-value to an HNH endonuclease in Mycobacterium species, but very strong hit to Peptidase in HHPred. Listed as a peptidase because of the HHPred hit.

**Gene 113** - matches a superfamily to a flagellar assembly protein H.

**Gene 114** - agrees with Glimmer. Even though it is a smaller ORF, it has a much better SD score, and a better alignment to the HTH DNA binding protein found in PrincessTrina.