

Cover Letter Extra Notes

General Notes

* all SD (RBS) annotations were recorded using settings Kibler 6 and Karlin Medium - we did not record this in the annotation notes to avoid repetition.
* In this file I have included a list of all of the major changes we made (gene additions and deletions). I have highlighted in **bold and italics** the annotations that I think need the most attention.

Thank You!

Specific Notes

***Translational Frameshift? ORF 21 (stop 12342) and 22 (stop 13334) – tail assembly chaperones :***

***In cluster A phages there is usually a translational frameshift at this location. There are 84 cluster E genomes. 9 of these annotate a translational frameshift, ~ 35 don’t and 47 are draft genomes. Also phamerator predicts a frameshift (I am not sure whether phamerator is including phages from other clusters?). We have provisionally annotated the frameshift based on the 'slippy' sequence as compared to the sequence of Phage Tuco. I am not entirely convinced on whether this is a true translational frameshift and would like the QC team to evaluate. I have included an image below of where the frameshift would be located (based on comparison with phage Tuco). Happy to annotate to the original call if QC team disagree with our current call.***

ORF 31 SSC:28958-29143 is a **lone** gene on the reverse strand surrounded by genes on the forward strand, this gene placement is conserved in similar homology phages so we are confident that this annotation is correct

Note - Starterator identifies a possible translational frameshift between gene 41 and 42. Since this is not the translational frameshift accepted by Genbank, it has not been annotated as a translational frameshift .

Added gene annotation at ORF100 SSC: 59043-59606 (Fwd strand) –

Start Site chosen based on Blast alignments and the fact that this ORF doesn’t overlap with preceding gene

SSC: 59043-59606. CP: Yes, it captures all CP. SD: F= -4.985 Z=1.857, not best scores. SCS: This gene was added manually, GAP: 41 bp gap. Blast: 1:1 with gp 100 Mycobacterium phage 244, top hit. LO: Longest w/o too much overlap. ST: N/A. F: DNA methylase. FS: see DNA Master file

Added gene annotation @ 63385-63885

ORF 112 SSC: 63385-63885, CP: yes, captures all CP. SD: F=-4.699; Z= 2.002. (best score) SCS: this gene annotation was added manually. GAP:11 bp gap. Blast: 1:1 with gp102 phage 244 LO: No, not the longest ORF ST:N/A F: HNH endonuclease. FS: see DNA Master file

Two gene annotation on the reverse strand were deleted and two genes annotations were added instead on the forward strand.

1. ORF 120 SSC: 67113-67325; CP:Yes, captures all of it;. SD: Z= 1.807; F= -5.108 (best score) SCS: This gene annotation was added manually after two gene annotations autoannotated on the reverse strand were deleted; GAP: 8bp overlap; Blast: 1:1 alignment with gp 110 phage 244; LO:Longest w/o significant overlap; ST:N/A F: NFK. FS: Checked phagesbp. NCBI. HHpred. Phamerator

2. ORF 121 SSC: 67306-67560; CP: Yes, captures all of it; SD:Z= 1.438; F= -6.152 (not best score but longest ORF) SCS: This gene annotation was added manually after two gene annotations autoannotated on the reverse strand were deleted; GAP:20bp overlap (similar overlaps observed in closely related phage, Mindy, TeardropMSU and Tuco Blast: 1:1 alignment with gp121 Phaux; LO: yes. longest ORF, ST:N/A. F:NKF.

***ORF 129 \*(stop site 69283) – provisionally calling this as a HTH DNA binding protein, MER like based on synteny with phage Tuco, CrystalP128; NelitzaMV\_125, Dusk and Goku. Note that this has not been called functionally in many other phage. Please examine. Note - No putative conserved domains have been detected***

Added gene annotation at ORF 135 SSC: 71519-71881 reverse strand - (stop site 71519)

SSC: 71519-71881 (REV). CP: Yes, captures all CP. SD: F=-3.212, Z=2.740, best score. SCS: This gene annotation was added manually. GAP: 16 bp gap. Blast: 1:1 gp128 Goldilocks. LO: Yes, longest ORF. ST: N/A. F: NKF

ORFs with upstream gaps > 250 bp

ORF 52 stop 36631 237 bp gap – conserved.

ORF 63 – 544 bp gap – conserved in other similar genomes

Between 68905 and 69282 - 378 bp gap – strand switch

571 bp gap at end of genome

Other:

Note that between 37643 - 37660 a putative attP site was detected based on a genomic comparison of Kimchi with *M. smegmatis* using NCBI BLASTn

