

Phage DoRead Start at 52714 bp with 8 bp overlap vs start at 52718 with 4 bp overlap?

Phage DoRead draft gene 103 (stop 52875 bp) in the +3 frame has significant coding potential (CP) only in GeneMarkS, and has an 8 bp overlap with the upstream gene which stops at 52721 bp. However, the gene in the +2 frame (stop 54067 bp) and is auto-called to start at 52907 bp has significant CP in both GeneMark_smeg & GeneMark_TB to allow it to start at 52718 bp, with a 4 bp ATGA overlap. Per Guiding Principle 12a, the ribosome is expected to prefer the operon. Keeping draft gene 103 (stop 52875 bp) means that the operon would have to go backwards to start at 52714 bp to make an 8 bp overlap rather than the ATGA overlap (start 52718 bp).

So far, only phage Modragons gp104 takes this operon as the start. Several BLASTp results in phagesDB keep the 8 pb overlap start (52714 - 52875 bp) which is found in 69.5% of genes in pham. The start at 52718 which gives the 4 bp overlap is found in 54.2% of genes in pham.

I also note that there is a string of operon genes with 1-4 bp overlaps starting from the two upstream genes all the way to the end of the genome (only the start at 52714 bp with an 8 bp overlap would break this pattern).

What is your verdict? Change the start to 52718 bp and prefer the operon (meaning deleting the gene which stops at stop 52875 bp) or keep both genes?

See below:

```

W T A P M D D A A R A R L E L R R S N A S Q P H R N R H R E R K T
G L H Q W M T L P A P A S N Y A D P T L P S R T A T G T E N A K P
L D C T N G * R C P R P E R T T P I Q R F P A A P Q P A P R T Q N R
52701 CTGGACTGCACCAATGGATGCGCTGCCCGCGCCCGCTCGAACTACGCCGATCCACGCTTCCCAGCCGCACCGCAACCGGCACCGAGAACGCCAAAAC
.....|
GACCTGACGTGGTTACCTACTGCGACGGCGCGGGCGAGCTTGATGCGGCTAGGTTGCGAAGGGTCGGCGTGGCGTGGCCGTTGCGCTTTGG
Q V A G I S S A A R A R R R S R R R D L A E W G C R L R C R S R L V
A P S C W H I V S G A G A E F * A S G V S G L R V A V P V S F A F G
S S Q V L P H R Q G R G G R V V G I W R K G A A G C G A G L V C F

```

