I am submitting the annotated Kersh genome. There were a couple of points I wanted to check:

1. For many of the genes, we could identify a functional domain that comprised only part of the predicted protein. In cases like this, we reported the functional domain as the gene function – it almost all cases, this is what other annotators had done as well. For example, gp28 had a GSDL family lipase domain, and that is the function we reported. Another example of this is gp40 which contains an HTH DNA binding domain.

2. For reverse transcribed genes, the gap was measured from the beginning of the gene to the end of the next reverse transcribed gene. So for example, the gap for gene 37, a reverse transcribed gene was measured between gp 37 and gp 38.

3. We are most unsure about the predictions we made for gene 46. DNA master originally called gp46 much longer than we did, but it left no space between the two genes which are transcribed in opposite directions. Starterator was of little help – gp46 has been called at all different lengths in other genomes. We picked a start with a relatively good SD score that captured all the coding potential.

4. We usually chose a start that gave a four bp overlap over one with a better SD score, as the new guidelines suggested. We used Starterator as much as possible, but it was often not informative in cases like this, perhaps because in the past, annotators choose a better SD score over a 4 bp overlap?

Thanks,

Mandy Butler

mabutler@ucsd.edu