LadyBird Annotation Cover Sheet May 26, 2015

Michael Kart and Charles Hauser

Amanda C. Callejon, Darius Douglas, Dylan J. Fox Matisen Harper, Nicolas V. Hyde, Jose F. Lerma, Samantha C. Parsons, Crisel Suarez and Margaret A. Walsh

Annotation Concerns:

1. gp1:
   1. Removed gp1 from annotation. This leaves a large gap at start of genome (1-570). No homology to any sequence in NCBI based on blastx. DNAM\_1: Original Glimmer call @bp 264 has strength 15.70 SSC: (REV)CP: SD: SCS: GAP: Blast: LO: ST: F: FS
2. gp24, 25
   1. Translational frameshift fuses these 2 genes giving rise to a full-length tail assembly chaperone
3. Inserted gp34
   1. Large gap (763nt) present between gp33-gp34. Some homology to Mtb var Erdman in this region. However we annotated based on blastx homology to a hypothetical protein present in EagleEye. Not predicted by Glimmer or GeneMark.
4. gp78
   1. Both GeneMark and Glimmer predictions are believed to be wrong - needed to be called at the first start in order to catch 1:1blast alignment. Ran blastx at NCBI with filtering low complexity = aligned 22:22; ran blastx without filtering low complexity, and this returned 1:1 alignment
5. gp90
   1. Removed this predicted gene. Short predicted gene, no coding potential on GeneMark-Smeg output. NCBI’s BLAST was carried out to check gap, and returned no putative conserved domains.
6. gp97
   1. Removed this predicted gene. Short predicted gene. NCBI’s BLAST was carried out to check gap, and returned no putative conserved domains. Blastx for 51579-51499 (GeneMark called ORF) matched same hypothetical protein 81:9. Blastx for 51603-51499 (second ORF) matched same hypothetical protein 105:1.
7. gp98
   1. Removed this predicted gene. Not called by GeneMark. NCBI’s BLASTp was carried out to check overlap with DNAM\_99; NCBI returned bad alignments to hypothetical proteins
8. tRNA(Gln)
   1. 1 tRNA identified, tRNA-Gln (CTG); 4416-4493.