**CapnMurica Cover Sheet May 15, 2014**

**Overall comments:**

1. We did not use the BLAST function in DNAMaster for any of the information that is in our Notes section of the dnam5 file—we used the protein BLAST on NCBI’s website so that we would also see the Conserved Domains that were identified, and also so we would detect relationships to sequences not in phagesdb (e.g. most of our closest non-Arthro hits are to Rhodococcus phages: see below).
2. There were many small ORFs on the right arm that did not pull up any hits or a few insignificant hits when BLASTed. There were also several “large” gaps on the right arm where we added some ORFs that were not called by Glimmer or GeneMark but we saw significant coding potential and they had strong SD scores in DNAMaster.
3. We used both the heuristic and Arthrobacter predictions when we looked at the coding potential output from GeneMark. The heuristic model output seemed a better fit with the ORFs.
4. We did not call any “reverse” genes in this genome. There were some initially predicted with the auto-annotation in DNAMaster, but we ended up deleting all of them after more thorough analysis.
5. The most similar genomes were Arthrobacter phages Gordon (very similar) and Circum (more distant), in phagesdb. Otherwise, the most similar phage matches for the majority of the ORFs we BLASTed were Rhodococcus phage ReqiPepy6 and RequiPoco6. There were very few ORFs whose top NCBI match was from an explicit phage genome (they may be prophages), but even if there were other hypothetical matches we only mentioned the top phage hit.

**Specific Comments:**

1. We predicted a programmed frameshift for ORFs 5 and 6. This was the first time that one of our annotations has had a frameshift—we are fairly confident that we called as instructed in the DNAMaster Annotation Guide, but please double check this section.
2. There is a 142 bp gap between the 3’ end of gp34 and the 5’ end of gp35. There are no alternative start codons to extend gp35 and no potential ORFs that we could add ourselves. (the region of gp34-gp38 had relatively large gaps between all the ORFs (40-400 bp), but we could not identify any ORFs that we believed could fill in these gaps.
3. Also large gaps between ORFS 42-44 (75-130 bp), but we couldn’t extend the called ORFs or find evidence of ORFs that we could insert in these regions.
4. ORFs 59 and 60 may be joined by a programmed frameshift at position 44951, but we chose not to annotate them that way because of the non-traditional location in the genome, because we had already called one frameshift in a more traditional region, and because neither the unfused nor fused ORFs generated any BLAST hits.
5. ORF 76—we called the start at 51866, but Glimmer and GeneMark called start at 51910, which had a substantially lower SD score and would have excluded the beginning of a conserved AmsA domain.
6. ORF 77—we called the start at 52343, but Glimmer and GeneMark called start at 52355, which had a substantially lower SD score and still gave an overlap of 10 nt with upstream ORF.
7. 189 bp gap between 3’ end of ORF 78 and ORF 79, but we saw no alternative start codons to extend gp35 and no potential ORFs that we could add ourselves.
8. ORF 84 has 55 bp overlap with preceding gene in the same orientation. This was the only way we could include these 2 ORFs with significant BLAST hits and coding potential.
9. ORF 86—we called start at 57825, but Glimmer and GeneMark called start at 57819 (ATG, SD: 378), but we favored the GTG with the higher SD score as it has no effect on coding potential and only results in a 3 nt gap.
10. We added 3 ORFs to the genome (45, 64 and 75). They were not called by Glimmer or GeneMark, but they filled in large gaps and these ORFs had significant coding potential and good SD scores, although none of them had significant BLAST hits.