**Badger Annotation Cover Sheet, Claire Rinehart, WKU. 5/17/15.**

Gene 1 Additional Notes: Starterator is indeterminate at calling the start. The start at 609 (Z-value=1.915) gives a protein shorter than the call at 543 (Z-value=1.525). Therefore the BLAST alignments at the beginning of Wile gp1 are at Subject = 23 and Query = 1. Both Wile gp1 and Badger gp1 have conserved domains and align at 98% identity and 100% positives. These conserved domains have been translated into Region and Site notations in Wile:

 Region 1..49

 /region\_name="HNHc"

 /note="HNH nucleases; HNH endonuclease signature which is

 found in viral, prokaryotic, and eukaryotic proteins. The

 alignment includes members of the large group of homing

 endonucleases, yeast intron 1 protein, MutS, as well as

 bacterial colicins, pyocins, and...; cd00085"

 /db\_xref="CDD:238038"

 Site order(22,24..26,36..37,41..42,45,49)

 /site\_type="active"

 /db\_xref="CDD:238038"

In order to apply these notations to Badger, we would need to extend the start site from 609 to 543. I am unsure if aligning the CDD carries sufficient weight to move the start site.

Gene 24 Additional Notes: Frameshift at 15437.

Gene 25 Additional Notes: Large overlap with the previous frameshifted gene. Since frameshift does not happen every translation, there should be sufficient time to express this gene, even with the overlap.

Gene 54 Additional Notes: This gene was added to conform with Wile. This may be a primase that is frame-shifted with the next downstream gene.

Gene 73 Additional Notes: Start site was extended to the suggested Starterator site. This start is consistent with it's close neighbor, Wile.

Gene 74 Additional Notes: I am torn between the start at 46530, which has a Z-value of 0.27 at GTG and an 11 bp overlap, and the Starterator suggested site at 46542, that has a Z-value of 1.95 at TTG and an overlap of 23 bp. It would seem easier to restart translation close to the stop site and a GTG is more common. Wile\_74 (100% match) used 46530 but Backyardigan\_73 (100% match), Peaches\_76, and several othes at 97% positive use the 46542 start. I changed it to 46542.

Gene 77 Additional Notes: I am torn between the Starterator and Glimmer/GeneMark call for the start at 47660 (Z-value 2.994, ATG, 23 bp gap) and the longer overlapping start at 47684 (Z-value 1.482, TTG, 4 bp overlap). I will leave the start at the Glimmer / Starterator called site.

Gene 81 Additional Notes: The called start site (Z=0.940, GTG) is very poor and does not benefit from an overlap. But, moving the start site to 48581 would miss almost all of the GeneMarkSmeg coding potential.