Genome Annotation Info

I tried to make a large file with all of the annotation info, Word kept crashing. I can certainly share the Google drive info with you if needed. Below are the genes we found difficult and notes are provided.

Elinal

|  |  |  |  |
| --- | --- | --- | --- |
| Gene # | Issue | Steps | Evidence |
| (547-630) | Has a -1 frame shift, coding potential, and is annotated in other genomes (most of the CT cluster), therefore we believe this gene should be kept | Keep the gene | Timeline  Description automatically generated with medium confidence |
| (816-899) | Has a -4 frameshift at its start site but lacks coding potential and is annotated in other genomes (Pons) therefore, we would like you to make the call. | Your call |  |
| (7294-7704) | -14bp start (GeneMark)  31bp gap (Glimmer) | We chose the -14bp gap | Text  Description automatically generated with medium confidence |
| (8643-8918)  and  (8643-9391) | Tail assembly chaperone – slippery sequence GGGGAAT (8910-8916) exists and we believe this is the gpG – gpGT translation frameshift | We do believe this to be the slippery sequence when comparing to other genomes – Pons, Big Chungus, etc. | Timeline  Description automatically generated |
| (18085-19053) | 3 starts available, -10, -4, and -1. We were uncertain which to pick  Glimmer has 18076 and GeneMark has 18085 | Your call | Graphical user interface, application, table  Description automatically generated |
| (22849-24279) | LOR is a -26bp overlap  Glimmer and GeneMark us the 1bp gap at 22849 | We believe the predictive tools are correct |  |
| (24276-24707) | Two start sites LOR is a -16bp overlap….GeneMArk predicts the -4bp overlap is correct at 24276 | 24276 is the correct start site | Graphical user interface, application  Description automatically generated |
| (25066-24773)  Reverse | 1bp gap is the start site (25066)  Pecaan has a strange -125 start site prediction | 25066 is the start |  |
| (?-25285)  Reverse | Multiple start sites  It looks like 25285 was chosen in Phamerator (unsure why)  -89, -77, -47, 28, 61, and 94 are predicted…94 is 25285 | We are unsure which start site. There is coding potential and Pons has an identical gene (33) | A picture containing diagram  Description automatically generated |
| (?-25643)  Reverse | Large gap. Glimmer predicts 25912 (188bp gap) and GeneMark predicts 25948 (152bp gap) | Coding potential suggests the longer gene and a start at 25948  Your call | Graphical user interface, application  Description automatically generated |
| (?-26686)  Reverse | Large gaps in predicted start sites  Glimmer and GeneMark predict 27399 | The coding potential predicts a large gap at the start site  Starterator says 3 of 10 pick this start  We picked 27399 | Diagram  Description automatically generated |
| (27986-27618)  Reverse | Long gap. GeneMark and Glimmer both choose the first site 27986 | We are confident just wanted to point this out |  |
| (31632-30580)  Reverse | Three overlapping start sites: -38, -14, and -8.  Glimmer predicts the -38, and Gene Mark predicts the -14 | The coding potential suggests the -14 start site is correct  The NCBI blast 1:1 ratio disagree with this choice | Graphical user interface, application, table  Description automatically generated  Diagram  Description automatically generated |
| (31846-31619) | GeneMark predicts 31846 as start  There is coding potential | We believe the gene should be kept | Diagram  Description automatically generated |
| (44627-44968) | Large gap  198,234, and 288bp  Glimmer chooses 44717  GeneMark  44627  The coding potential isn’t the best | We chose the LOR |  |
| (45087-45371) | Large gaps  Glimmer and GeneMArk choose the second LOR | We stayed with the predictions |  |
| (46696-46995) | Large gaps  GeneMark 46786 and Glimmer 46702 differ  The LOR is not predicted (46696) | We would like you to choose, the coding potential is weak | Chart, histogram  Description automatically generated |
| (47994-48332) | Only Glimmer predicts a start (48084)  Large gaps  There is no coding potential | Remove the gene? | Chart, box and whisker chart  Description automatically generated |