1) The .predict file with the ORF locations from Glimmer.

Contents of predict file:

>Halanaerobium sp. MDAL1, whole genome shotgun sequence
orf00001 171 350 +3 11.68
orf00003 343 1626 +1 8.96
orf00004 1629 4733 +3 6.58
orf00005 5786 4971 -3 8.13

2) Code I ran to produce the output for number 1 (Glimmer code).

long-orfs -n -t 1.15 hprev_genome.fasta hprev.longorfs
extract -t hprev_genome.fasta hprev.longorfs > hprev.train
build-icm -r hprev.icm < hprev.train
glimmer3 -o50 -g110 -t30 halan.fasta hprev.icm halan

3) The DNA sequence of the first ORF in FASTA format.

>orf00001 171 350 len=177
ATGGGGGCAGTAATTGAAAGTAATTTAATTTCGGCTCAGAGATTGTTAAGTGATGCAGAA
ACAGATTTAACTGCTGCAAAATATGCCGTGCAGTTAAAAAAGACAGAAGTTTTGGCTGCA
GTAGAAAATATATATAAGAGCTTTACTGCAGGAGTATTAGGAGGTAATAGTAATGAA

4. Every predicted CDS of the halan.fasta file, based entirely on the Glimmer result.

orf00001 171 350
orf00003 343 1626
orf00004 1629 4733
orf00005 5786 4971

5. All possible mRNA molecules based on the FGENESB prediction.

mRNA molecules from FGENESB prediction:
Locations:
3..350
343..1626
1629..4733
4971..5786

>3..350

GGAATCAAAGAAAGTGAAATTAACTTAGATAATCTTGAGTCTGAAATAAGAATTGAACTTTCAAGCTT
ATTAAGAGAACTGGAATTAACTAAACTTAATTTAGAAACAGCAGCTAAGAATTTAAAAAGAGCTAAATTA
GAATATCAGAGCACAAAAAATAGATATCAAATGGGGGCAGTAATTGAAAGTAATTTAATTTCGGCTCAGA
GATTGTTAAGTGATGCAGAAACAGATTTAACTGCTGCAAAATATGCCGTGCAGTTAAAAAAGACAGAAGT
TTTGGCTGCAGTAGAAAATATATATAAGAGCTTTACTGCAGGAGTATTAGGAGGTAATAGTAATGAA

>343..1626
ATGAATAAAAAAACAAAAATGGCCTTAACTATCCTGCTGATAATAGCTATTGGTGCTGGA
GCTTTAATCTTTATCAGAGAGCTTAAAAATAGGGAGCCTCAGGTAGCTAAAGAGGAAGAT
TTGGGAGCAGCGGTGGAAACAGCTGAGGTTCAGAAAGGTGATTTTGAAATAATTTATAAT
TATAGTGGTACTGCAGAATATGCAGGCAAAAGAAAAATTTCTTCCCAAATCGGGGGAGAG
ATAATAAATATTTATGTTAGAGAAAGTCAAAAAGTAGAAAAAGGAGATCTTCTGGCTAGA
ATTGATGATCAAGAGCTAAAAAATAATCTCAGTTCGGCAGAGACTGCTGTTAGAGAAGCA
GAAATTGCTTTGAAGAAAGCTGAATTAGCTAAAGATATATCAAGAAATAATTTAGCGGAG
AGTAAAGCAGCTATAAAAGAAGCAGAGAGCAATTATTCTCAGTGGCAGAGTGATTATGAG
CGAGATAAAAAACTTTATCAAAAAAATGCCATTGCAAAAGCTAAGTTTGAACAGACTAAA
ACTCAGTTTCAAAAAGCTGCAGCTCAACTTGAAAGAGTACAGGCAACTCTGTCCAGTGCA
AAAAAATCAGTAGAAATTGCTGGCTTAGATGTTGAAACTACAGTCGAAAGGTTGAAAAAG
TCAAGAAATGAGCTTGAAAATGCCAGACTAAAATTTAAGGATACAGAAATTAGATCTCCA
ATTAGTGCTGAAATTGTCAACGAATTTGCAGAAGTAGGAGAAGTTACAGCAGCTGGTCAG
CCTCTTTTTGAAATAGCAAAAAGCGACAGGGTTGAAATAAAAATACAGGTGGGGATGAGT
GATCTCAATCAATTAAAGATTGGCACTAAAGCTTTAATTTCTTCTCCTGCTCTTGAGCAA
AAAGAATTTAAGGCAGTGATTTCTAAGATCGGCTCAACTGCCGACTCTAAAAGCAGAACT
ACTGAAGTAACTTTAAAATTAAAAGAAAATATTAATCTAAAAGATGGGGCCTTTGTTTCT
GCGGCTTTAATAGCGGAAGGGCTAACCGATGTCTTGATTGTTCCAGAGAAAGCAATTTTT
AACTATCAAGCAGCTTCCCATGTTTATTTAATAAAAGACGGTAGAGCAGTGAGACAAAAA
ATTGAAACAACAGTTACTAATGGTTATCAGACTGTTGTCACCTCTTTTCTCTCTGAAGGG
GATCAGATAGCAGTGACTAATCTCAATGATCTGCAGGATAAGACTAAGGTCTATTTATCT
GAGCAGGAAAATGGAGATGAT

>1629..4733
TAATTAATCATCTCCATTTTCCTGCTCAGATAAATAGACCTTAGTCTTATCCTGCAGATC
ATTGAGATTAGTCACTGCTATCTGATCCCCTTCAGAGAGAAAAGAGGTGACAACAGTCTG
ATAACCATTAGTAACTGTTGTTTCAATTTTTTGTCTCACTGCTCTACCGTCTTTTATTAA
ATAAACATGGGAAGCTGCTTGATAGTTAAAAATTGCTTTCTCTGGAACAATCAAGACATC
GGTTAGCCCTTCCGCTATTAAAGCCGCAGAAACAAAGGCCCCATCTTTTAGATTAATATT
TTCTTTTAATTTTAAAGTTACTTCAGTAGTTCTGCTTTTAGAGTCGGCAGTTGAGCCGAT
CTTAGAAATCACTGCCTTAAATTCTTTTTGCTCAAGAGCAGGAGAAGAAATTAAAGCTTT
AGTGCCAATCTTTAATTGATTGAGATCACTCATCCCCACCTGTATTTTTATTTCAACCCT
GTCGCTTTTTGCTATTTCAAAAAGAGGCTGACCAGCTGCTGTAACTTCTCCTACTTCTGC
AAATTCGTTGACAATTTCAGCACTAATTGGAGATCTAATTTCTGTATCCTTAAATTTTAG
TCTGGCATTTTCAAGCTCATTTCTTGACTTTTTCAACCTTTCGACTGTAGTTTCAACATC
TAAGCCAGCAATTTCTACTGATTTTTTTGCACTGGACAGAGTTGCCTGTACTCTTTCAAG
TTGAGCTGCAGCTTTTTGAAACTGAGTTTTAGTCTGTTCAAACTTAGCTTTTGCAATGGC

ATTTTTTTGATAAAGTTTTTTATCTCGCTCATAATCACTCTGCCACTGAGAATAATTGCT
CTCTGCTTCTTTTATAGCTGCTTTACTCTCCGCTAAATTATTTCTTGATATATCTTTAGC
TAATTCAGCTTTCTTCAAAGCAATTTCTGCTTCTCTAACAGCAGTCTCTGCCGAACTGAG
ATTATTTTTTAGCTCTTGATCATCAATTCTAGCCAGAAGATCTCCTTTTTCTACTTTTTG
ACTTTCTCTAACATAAATATTTATTATCTCTCCCCCGATTTGGGAAGAAATTTTTCTTTT
GCCTGCATATTCTGCAGTACCACTATAATTATAAATTATTTCAAAATCACCTTTCTGAAC
CTCAGCTGTTTCCACCGCTGCTCCCAAATCTTCCTCTTTAGCTACCTGAGGCTCCCTATT
TTTAAGCTCTCTGATAAAGATTAAAGCTCCAGCACCAATAGCTATTATCAGCAGGATAGT
TAAGGCCATTTTTGTTTTTTTATTCATTACTATTACCTCCTAATACTCCTGCAGTAAAGC
TCTTATATATATTTTCTACTGCAGCCAAAACTTCTGTCTTTTTTAACTGCACGGCATATT
TTGCAGCAGTTAAATCTGTTTCTGCATCACTTAACAATCTCTGAGCCGAAATTAAATTAC
TTTCAATTACTGCCCCCATTTGATATCTATTTTTTGTGCTCTGATATTCTAATTTAGCTC
TTTTTAAATTCTTAGCTGCTGTTTCTAAATTAAGTTTAGTTAATTCCAGTTCTCTTAATA
AGCTTGAAAGTTCAATTCTTATTTCAGACTCAAGATTATCTAAGTTAATTTCACTTTCTT
TGATTCCCGATTTTTTTAAACTAAAATAAATTTTTTAAACTAAAATAAAAGTTATAAACA
AATGATATCTTTGGCTTATTTTATCGCATTAAAGTATTTGAATCAAGCTATTTAACTGAT
TAAAAAAATTAATTTGCCTTTTGATGTTAAATAATTAATGGCTGCATCTTATTGGATCCT
ATATAATGTTGATGAGAAATATTAATTAGAAAGGGTGATTTAAATGAAAAAATTTGAACT
GAAAAATGGAAATAAAATGCCTGCACTGGGATTAGGGACCTCAGGTTTACGAGGTAAAGA
ATGTACTCAAGTAGTAAAAGAAGCTCTCGAGCTGGGCTACCGACAGGTAGACACTGCTGA
CATGTATGGAAATCACAGAGCGATTGCTGAAGCATTAAATGAATCTGATGTAAGGCGTGA
AGATTTGTTTATTACTTCTAAAATCCAGAGTGAAGATTTAGAATATAGACAGCTAAAAAA
GACTGCCTCTCGCCTCTTAGATGAACTTGATCTAAAATATTTTGACCTGCTTTTAATTCA
CTGGCCCAGTCCAGAAGTTCCGGTTGAAGAATCTTTAAAAGCAATGAAAGAATTAAAAGA
AGCTGGTAAAGCTAAAAATATCGGAGTCAGCAATTTTACTATTCCACTTCTCAAAAAAGC
CTTAGCTGCTTATCCTGATTTAATAACTGTTAATCAGGTAGAATTTCACCCGACTCTTTA
TCAAAAAGAACTTTTAGACTTTGCTTTCAAAAATGATATTATTCTAACTGCTTATGCTCC
GCTGGCCCAGGGAGAAGTATTTGAAAATAGCGTCTTAAAATCACTGGGAGAAAAATACGA
TAAATCTCCTGCACAGCTGGCTTTAAGGTGGCTGGTTGAGAAAAATATTGCAGTTATTCC
TAAAGCAAGTTCTAAAGCTCATCTTAAAAATAACTTAGAGATCTTCGACTGGGACTTCCC
AATTGATGCAGCTCGAGAAATGGAGCTATTGGATCAAAATAACCGCTTAATTGATCCCGG
TTACCCAAATTTTGATTAAAATTATTAAACCCAGCCTTTAATCGGGCCGGGTTTTGCTAT
CTCAATCTCACATTAATAATGACGCTCAATAAAATTTTTTGACGCCCCTATCATACCAAA
ATCATTACTTTTAACTGCTTCCCTAACTGAAAATCCCCTGTTCACAGATCAGAATTTGAT
TGCTCCATCTGATTACTGTTGAAAGGGGATAAATCAAAAGCGATATTTTAATTTCTAGCA
GCTCCAGTTATAAA

>4971..5786
ATGAAAAAATTTGAACTGAAAAATGGAAATAAAATGCCTGCACTGGGATTAGGGACCTCA
GGTTTACGAGGTAAAGAATGTACTCAAGTAGTAAAAGAAGCTCTCGAGCTGGGCTACCGA
CAGGTAGACACTGCTGACATGTATGGAAATCACAGAGCGATTGCTGAAGCATTAAATGAA
TCTGATGTAAGGCGTGAAGATTTGTTTATTACTTCTAAAATCCAGAGTGAAGATTTAGAA
TATAGACAGCTAAAAAAGACTGCCTCTCGCCTCTTAGATGAACTTGATCTAAAATATTTT
GACCTGCTTTTAATTCACTGGCCCAGTCCAGAAGTTCCGGTTGAAGAATCTTTAAAAGCA

ATGAAAGAATTAAAAGAAGCTGGTAAAGCTAAAAATATCGGAGTCAGCAATTTTACTATT
CCACTTCTCAAAAAAGCCTTAGCTGCTTATCCTGATTTAATAACTGTTAATCAGGTAGAA
TTTCACCCGACTCTTTATCAAAAAGAACTTTTAGACTTTGCTTTCAAAAATGATATTATT
CTAACTGCTTATGCTCCGCTGGCCCAGGGAGAAGTATTTGAAAATAGCGTCTTAAAATCA
CTGGGAGAAAAATACGATAAATCTCCTGCACAGCTGGCTTTAAGGTGGCTGGTTGAGAAA
AATATTGCAGTTATTCCTAAAGCAAGTTCTAAAGCTCATCTTAAAAATAACTTAGAGATC
TTCGACTGGGACTTCCCAATTGATGCAGCTCGAGAAATGGAGCTATTGGATCAAAATAAC
CGCTTAATTGATCCCGGTTACCCAAATTTTGAT

6. Found locations where FGENESB and Glimmer differ in CDS prediction.

Glimmer and FGENESB only differ at the beginning of the first CDS. Glimmer predicts the first to start at position 171, whereas FGENESB predicts it starts at position 3. Both predict the first CDS ends at 350. All other positions are in agreement.

FGENESB

```
Prediction of potential genes in microbial  genomes
 Time:   Tue Jan  1 00:00:00 2005
 Seq name: Halanaerobium sp. MDAL1, whole genome shotgun sequence
 Length of sequence - 6000 bp
 Number of predicted genes - 4
 Number of transcription units - 2, operons - 1
     N      Tu/Op   Conserved  S             Start         End    Score
                    pairs(N/Pv)
     1     1 Op  1     .       +    CDS           3 -       350    298
     2     1 Op  2     .       +    CDS         343 -      1626   1063
     3     1 Op  3     .       +    CDS        1629 -      4733   1901
     4     2 Tu  1     .       -    CDS        4971 -      5786    654
```

Used two attached sequences: mouse_genomic.txt and mouse_cdna.txt from the organism Mus musculus. The cDNA is an alternately spliced transcript that was verified by NCBI on September 13, 2016 based on RNA-seq data. Ran Splign to get the mRNA coordinates and the cDNA coordinates of the genomic sequence:

1. Coordinates of the mRNA & the CDS locations based on the provided genomic sequence.

mRNA locations:
601..677
3390..3701
3847..4370

CDS Locations:
3415..3701
3847..4165