

2021-22

BIO F242
Introduction to Bioinformatics

Experiment - 3

Name : Suchismita Tripathy
ID : 2019A7PS0554P

Gene Name : HUS1 Checkpoint Clamp Component (HUS1)
transcript variant X1
Organism : Theropithecus gelada
Accession Number : XM_025380797.1

14 February, 2022

1 Aim

To predict the functional (Exons) and structural (Introns, TFBS, Promotor) regions in given gene sequence of Theropithecus gelada HUS1 Checkpoint Clamp Component (HUS1) transcript variant X1 using GENSCAN FGENESH

2 Materials Required

GENSCAN, FGENESH, gene sequence of Theropithecus gelada HUS1 Checkpoint Clamp Component (HUS1) transcript variant X1

3 Procedure

1. Open the home page of NCBI and retrieve the Theropithecus gelada HUS1 Checkpoint Clamp Component (HUS1) transcript variant X1 gene sequence in Fasta format having “Accession No” (XM_025380797.1).
2. Open the home page of GenScan (<http://hollywood.mit.edu/GENSCAN.html>) and FGENESH (<http://www.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind>) respectively and analyze the given gene sequence.
3. Select the parameters such as like organism, exon cutoff, print options etc in both Genscan and FGENESH.
4. Take appropriate screenshots of observations.
5. Analyze the results.

4 Observations

```
GENSCAN Output

View gene model output: PS | PDF

GENSCAN 1.0      Date run: 12-Feb-122    Time: 04:45:54

Sequence /tmp/02_12_22-04:45:54.fasta : 3031 bp : 43.68% C+G : Isochore 2 (43 - 51 C+G%)

Parameter matrix: HumanIso.smat

Predicted genes/exons:

Gn.Ex Type S .Begin ...End .Len Fr Ph I/Ac Do/T CodRg P.... Tscr..
-----
1.01 Sngl +   99   941  843  2  0 100  42  500 0.885 42.36
1.02 PlyA +  2732  2737   6                                -1.75

Suboptimal exons with probability > 1.000

Exnum Type S .Begin ...End .Len Fr Ph B/Ac Do/T CodRg P.... Tscr..
-----

NO EXONS FOUND AT GIVEN PROBABILITY CUTOFF

Predicted peptide sequence(s):
```

Figure 1: Genscan Output Screenshot 1/2

```

Predicted peptide sequence(s):

Predicted coding sequence(s):

>/tmp/02_12_22-04:45:54.fasta|GENSCAN_predicted_peptide_1|280_aa
MRFRAXIVDGACLNHFTRISN#IIAKLAKTCTLRISPDKLNFIILCDKLANGGVSMWCELEQ
ENFFNEYQMEGVSAENNEIYLELTSENL SRALKTAQ#ARALKIKLTHKHFPCLTVSVELL
SMSSSSRIVTHDIPKXIPRKLWKDLQEPVWPDPOVSIYLPVLKTHKSVVEKH#HISNHL
VIEANLDELNLKIELELCVTTTHFKDLGNPPLASESTHQDRNMEH#AEVHIDIRKLLQF
LAGQQVNPITKALCNIVNNKQVHFDLLHEDVSLQYFIPALS

>/tmp/02_12_22-04:45:54.fasta|GENSCAN_predicted_CDS_1|843_bp
atgagggttcgggccaagatcgtggacggggcctgtctgaaccacttcacacgaatcagt
aacatgatagccaagcttgccaaaacctgcacctccgcacgacctgataagcttaac
ttcatcctttgtgacaagctggcctaattgggggtgagcatgtgggtgagctgggaacag
gagaacttcttcaacgaatatcaaatggagggtgtctctgcagaaaacatgagatttat
ttagagctaacatcggaaaacttatctcgagccttgaaaactgccagaatgccagagcc
ttgaaaatcaactgactaataaacactttccctgcctcacagtctccgtggagctgtta
tctatgtcaagcagtagccgattgtgacacatgacatccataaagggtattcctagg
aaattgtggaaggacttgcaagaaccgggtggccagatcctgatgttagtatttatta
ccagtccttgaagactatgaaggtgttgggaaaaatgaaaaacatcagcaatcacctt
gttattgaagcaaacctagatggagaattgaattgaaaatagaactgaattagatgt
gttacaaactcatTTTaaagatcttggaatcctccttagcctctgaaagcaccatcaa
gacagaaacatggaacacatggcctgaagtgcacatagatattggaagctcctacagttt
cttgctggacaacaagtaaatcccaaaaggccttatgcaatattgtgaataacaagatg
gtgcattttgatctgcttcacgaagacgtctcccttcagtatttcacccctgcgtgtcc
tag

```

Figure 2: Genscan Output Screenshot 2/2

FGENESH 2.6 Prediction of potential genes in *Chlorocebus_sabaeus* genomic DNA
 Seq name: test sequence
 Length of sequence: 3031
 Number of predicted genes 1: in +chain 1, in -chain 0.
 Number of predicted exons 1: in +chain 1, in -chain 0.
 Positions of predicted genes and exons: Variant 1 from 1, Score:40.689874

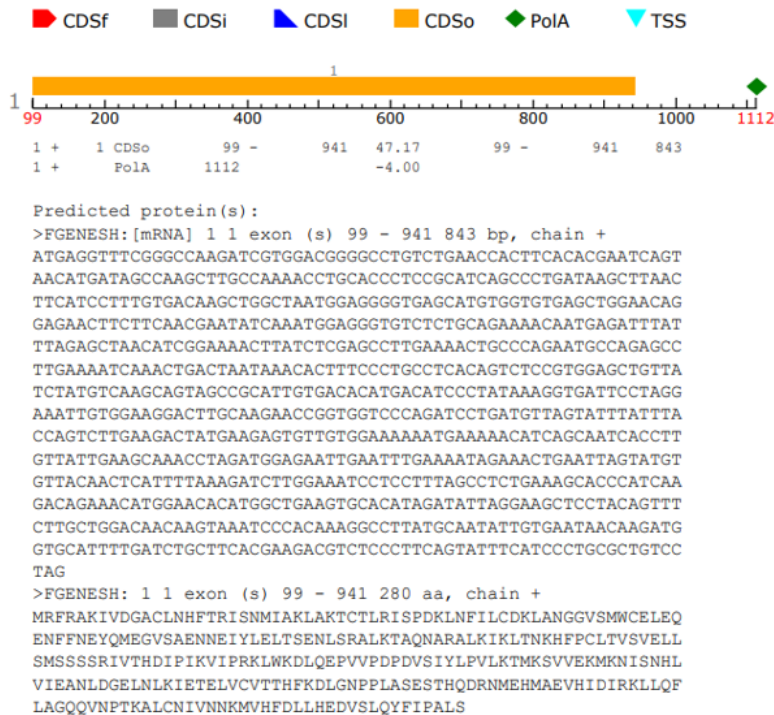


Figure 3: FGENESH Output Screenshot 1/1

5 Interpretation of Results

The given gene sequence of *Theropithecus gelada* HUS1 Checkpoint Clamp Component (HUS1) transcript variant X1 consists of:

3031 bp in length

43.61% C+G

2 number of Isochore

1 number of genes

0 number of Init = Initial exon (ATG to 5' splice site)

0 number of Intr = Internal exon (3' splice site to 5' splice site)

0 number of Term = Terminal exon (3' splice site to stop codon)

1 number of Sngl = Single-exon gene (ATG to stop)

0 Length of longest Prom = Promoter (TATA box / initiation site)

6 Length of longest PlyA = poly-A signal (consensus: AATAAA)

500 Maximum CodRg : coding region score (tenth bit units)

0.885 Maximum P : probability of exon (sum over all parses containing exon)

280 Length of Longest Amino Acid

Paste the Sequence of Longest Amino Acid:

**MRFRAKIVDGAACLNHFTRISNMIAKLAKTCTLRISPDKLNFILCDK
LANGGVSMWCELEQENFFNEYQMEGVSAENNEIYLELTSENLSRA
LKTAQNARALKIKLTNKHFPCLTVSVELLSMSSSSRIVTHDPIK
VIPRKLWKDLQEPVVPDPDVSIIYLPVLKTMKSVVEKMKNISNHLV
IEANLDGELNLKIETELVCVTTHFKDLGNPPLASESTHQDRNMEH
MAEVHIDIRKLLQFLAGQQVNPTKALCNIVNNKMVHFDLLHEDVS
LQYFIPALS**