Identification of functional sites in gene/genome

Aim:- Identification of functional sites in given genome using Genscan and ORF finder.

Introduction:-

Genscan is an online program to identify complete gene direction gene finder for human sequence. An important feature for human of genscan is that because on probablistic method of genomic model sequence composition or gene structure it is able to assign meaningful probability to particular events under which probability P(E) is defined as the sum of probability to particular events under the model of all possible passes which contain the exact exon E in the correct reading frame. Internal exons are predicted more accuratly than initial (or) terminal exons predicted.

Query:- Haemoglobin and Fibrinogen.

URL:- hollywood.mit.edu/GENSCAN.html
https://ncbi.nlm.nih.govt/

Procedure :-

a> ORF-finder :-

→ tool used for open reading

→ Log onto NCBI and retrieve a nucleotide sequence in FASTA format.

→ Log on to NCBI in the welcoming page of NCBI click on all. resources option present towards left.

→ Select tools option & click onto ORF, paste the sequence in sequence box, click on the submit result will be displayed.

b> GENSCAN :-

→ Log on to NCBI & retrieve 2 nucleotide sequence in FASTA format.

→ Log on to GENSCAN using the url. hollywood edu/genscan. html.

→ paste the sequence & submit.

Name of Experiment... Date............4/1/2024....................... Page No. 114

Experiment No... Experiment Result..............................

Result and Discussion :-

a) ORF - Open Reading Frame.

The number of ORF's for the given nucleotide by using ORF finder tool available in NCBI. Among all the ORF's, select one ORF such that, it length should be maximum & involves both start & stop codons.

On the left hand side of the result page, a sequence of particular ORF is given, this tool also allows us to BLAST for the given ORF. The result of BLAST is obtained for given ORF protein sequence as a query.

b) Genscan :-

A predicted protein sequence is obtained for the input nucleotide sequence with genscan software tool.

P-10
C-10
N-10
R-20
55

Phasianus colchicus voucher SYSb004985 fibrinogen (FGB) gene, intron 5 and partial cds

GenBank: MT842193.1

>MT842193.1 Phasianus colchicus voucher SYSb004985 fibrinogen (FGB) gene,
intron 5 and partial cds

```
GTTAGTGTGAAATAACAATGTCATTAAAATATTACTGTGAATACTAAAACAGAACAAATATACTAATAAT
TTTCAATTTATACAAACTAATATGGAGTTACAATTTTTGTAGTTATTTGAGAATTAAGTCTGACTCATTC
AAAAGAGTCATGTTTGTCAGATACCCAGCAGGATATTGCTATAATGAAACTGTAAATGAATAAGTGTGAA
NGGATGGGTAAATAAAAACATATCTTACTATAGCTAAAATTGTCAATGTAGTGGTGGACAGAAGTGGAAG
ATCTGGTTCTTATCTTACCCTGTTACTACATGCTATATAAAAAGACCTGAGGTGCTTCTTCACCTCCTGC
TTCTTCTACATTTTCTCATGAAGCAAGTCACTGCTTCAGTGTGATTAAAAATGGCAATGCTGTACTCATA
AGCAGTCTTGCTACAAAGAAGTCACTTGCTCCCTTCATCTTGAGATAATGGTAGAAAGGCTGTATCTTAA
GACTTAAAGCTCCTTTCATAATTTATCCAATTAAGATTAGATTATTTAATCTAAGCCCTGTTGTTCCTGT
AGGCTGGACTTTGATTCAGAATCGCCAGG
```
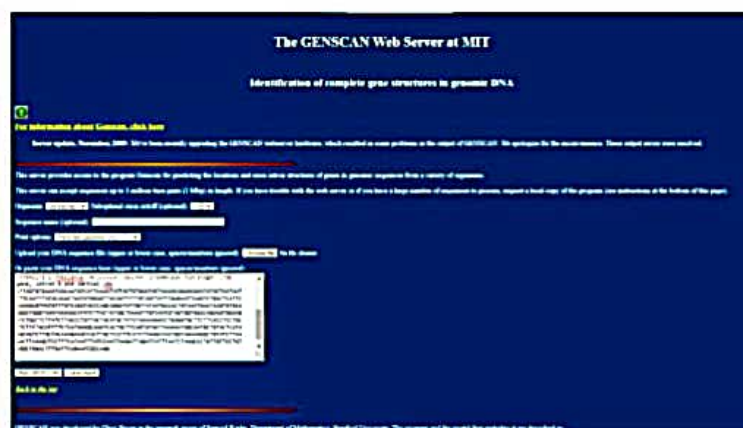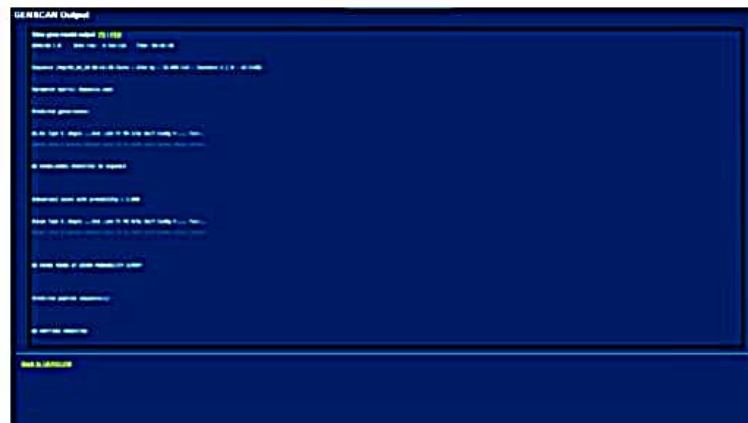
Phasianus versicolor voucher SYSb001826 fibrinogen (FGB) gene, intron 5 and partial cds

GenBank: MT842178.1

>MT842178.1 Phasianus versicolor voucher SYSb001826 fibrinogen (FGB) gene,
intron 5 and partial cds

```
GTTAGTGTGAAATAACAATGTCATTAAAATATTACTGTGAATACTAAAACAGAACGAATATACTAATAAT
TTTCAATTTATACAAACTAATATGGAGTTACAATTTTTGTAGTTATTTGAGAATTAAGTCTGACTCATTC
AAAAGAGTMATGTTTGTCAGATACCCAGCAGGATATTGCTATAATGAAACTGTAAATGAATAAGTGTGAA
GGGATGGGTAAATAAAAACATATCTTACTATAGCTAAAATTGTCAATGTAGTGGTGGACAGAAGTGGAAG
ATCTGGTTCTTATCTTACCCTGTTACTACATGCTATATAAAAAGACCTGAGGTGCTTCTTCACCTCCTGC
TTCTTCTACATTTTCTCATGAAGCAAGTCACTGCTTCAGTGTGATTAAAAATGGCAATGCTGTACTCATA
AGCAGTCTTGCTACAAAGAAGTCACTTGCTCCCTTCATCTTGAGATAATGGTAGAAAGGCTGTATCTTAA
GACTTAAAGCTCCTTTCATAATTTATCCAATTAAGATTAGATTATTTAATCTAAGCCCTGTTGTTCCTGT
AGGCTGGACTTTGATTCAGAATCGCCAGG
```

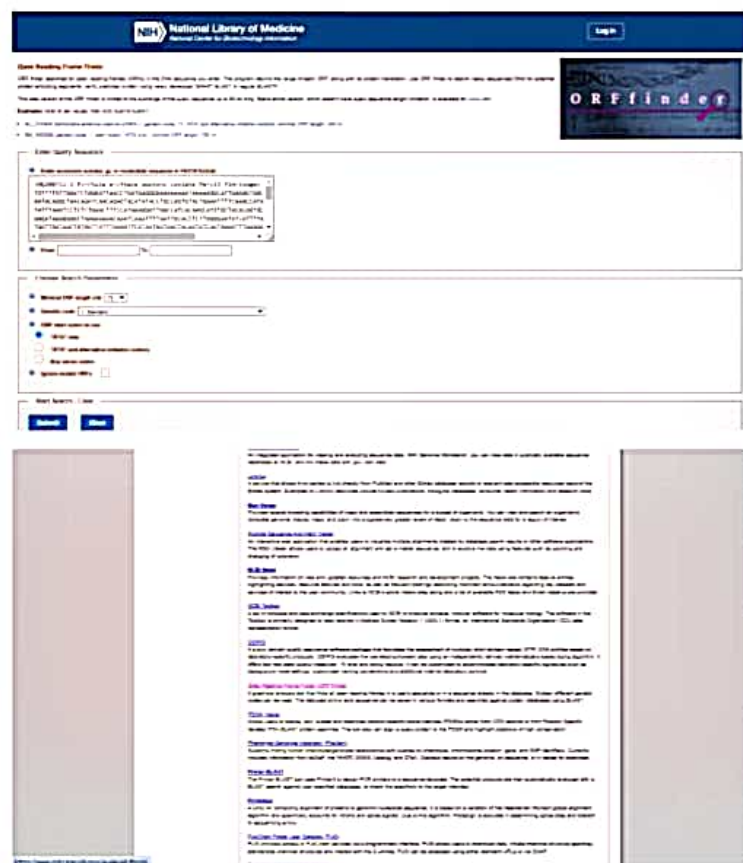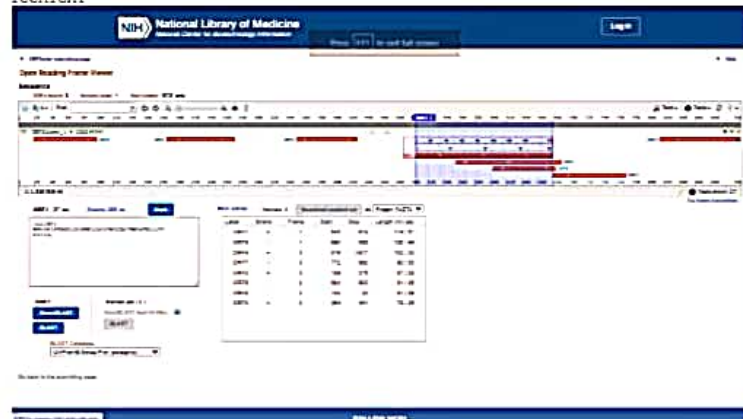# Pyrrhula erythaca owstoni isolate Pery13 fibrinogen (fib-7) gene, partial cds

GenBank: HQ284711.1

GenBank Graphics PopSet

>HQ284711.1 Pyrrhula erythaca owstoni isolate Pery13 fibrinogen (fib-7) gene,
partial cds

TGTTTTGTTGGATCTAGGGTTAACCTGATGAGGGGAAAAAAAATAAAAAGGCATTGAGAGTGACATCTGT
AATACAGGCTAACAGATCAACAGAGTACATATACCTGCCAGTGTACTGAAATTTTCAAACCATAAGAAAC
TATTTAAGTCCTCTCTGAACTTTCCATAAAGGATTGACCATCACAAGCATGTGCTACACAGTGCACTCAG
AAGATAAAGGAGCTAAGAAAAACAAATCAAATTTTAATTGCACCTCTTGGGGAATATCATTTTAGTTTAC
TGATTGATAACTGTGGTTATTTGAAGTTCATAATGATAACTACAATATCAATGAAATTTAAAGATAGAAA
GCAATGAGAAATGTATATCTTCCCCAGCCTAAACAATTCCTTTATTCATTAATATGTTATCAGGTATTAA
GCACAGGTTAGAGATACAGTGCTCTATTATGTACTTTTACAACTGAGCTCCTGTCTTCTGTGTAAGCAGA
ACATTGCTTTTTTAATGGATATGTTACCTGCTTTTTCCTTCAGGATCCATTATATTTTGCTGGAGTTAGT
ACCAATGTTCCTGCTTAGTGTCCTCAGCACTCACTGCTGCTGTCAATCTCTGACTATGAATATACTGCAT
ACAGACTTGCTGCTTTGCTTTTTTTTTTTTTTTTTTTTCTTGCTGGTAAGTCATTCAGCATGCTTGCTTCT
TACTTGGTCGTGGACTGTCACTAACCTAACATATAGTTTTACCTGTCCTTGGTTATGCCAAGAGAACTCC
ATAGATATTTAGCTGGTAACATGCAGGGACATTACATATTTTGACATGCTACAAAATGCAGTGAAAACTA
ATGGCAAATTAAANAAATATTATATATTCAGTTCTATTTTTATTTTAAATTCTAGTGCAAGTACATACCA
TCCATCAT

NIH) National Library of Medicine
National Center for Biotechnology Information

Log in

| Nucleotide ⌄ | fibrinogen | | Search |

All Resources

| All | Databases | Downloads | Submissions | Tools | How To |

Databases

Assembly
A database providing information on the structure of assembled genomes, assembly names and other meta-data, statistical reports, and links to genomic sequence data.

BioCollections
A curated set of metadata for culture collections, museums, herbaria and other natural history collections. The records display collection codes, information about the collections' home institutions, and links to relevant data at NCBI.

BioProject (formerly Genome Project)
A collection of genomics, functional genomics, and genetics studies and links to their resulting datasets. This resource describes project scope, material, and objectives and provides a mechanism to retrieve datasets that are often difficult to find due to inconsistent annotation, multiple independent submissions, and the varied nature of diverse data types which are often stored in different databases.

BioSample
The BioSample database contains descriptions of biological source materials used in experimental assays.

Bookshelf
A collection of biomedical books that can be searched directly or from linked data in other NCBI databases. The collection includes biomedical textbooks, other scientific titles, genetic resources such as GeneReviews, and NCBI help manuals.

ClinVar
A resource to provide a public, tracked record of reported relationships between human variation and observed health status with supporting evidence. Related information in the NIH Genetic Testing Registry (GTR), MedGen, Gene, OMIM, PubMed and other sources is accessible through hyperlinks on the records.

ClinicalTrials.gov
A registry and results database of publicly- and privately-supported clinical studies of human participants conducted around the world.

Computational Resources from NCBI's Structure Group
A centralized page providing access and links to resources developed by the Structure Group of the NCBI Computational Biology Branch (CBB). These resources cover databases and tools to help in the study of macromolecular structure, conserved domains and protein classification, small