

Bioinformatics: Life Sciences on Your Computer

Practice Problems - 1

Note: Softberry.com (FGENESH) has been down all day, Monday, July 7. It is usually reliable, and I expect it to be running soon.

1. Retrieve the sequence called “fly_gene_b.txt” at https://d396qusza40orc.cloudfront.net/bioinform/assignments/fly_gene_b.txt. It is genomic DNA covering one coding region from a *Drosophila* species that is not *D. melanogaster*. Be careful with Genscan--it reads the descriptive line of the FASTA-formatted sequence as DNA characters. Better to open the txt file and paste in the sequence in RAW format.
 - a. Go to GENSCAN (link below) and obtain results. Report the exon locations and state which exon is predicted with the highest probability.
 - b. Go to FGENESH (link below) and obtain those results. What are the differences between the FGENESH result and the GENSCAN result?
 - c. Where is the transcription start site in FGENESH? Is it known? Does the transcription start correspond with the beginning of exon 1 (as predicted by FGENESH)?

2. Retrieve the file called “chicken_gene.txt ” at https://d396qusza40orc.cloudfront.net/bioinform/assignments/chicken_gene.txt. Use these three algorithms to obtain results (see note above on Genscan and RAW format):

GenScan: <http://genes.mit.edu/GENSCAN.html>

FGENESH: <http://www.softberry.com/berry.phtml?topic=fgenes&group=programs&subgroup=gfind>

Augustus: <http://bioinf.uni-greifswald.de/augustus/submission>

Compare the results to the actual exon locations:

CDS#1 (interleukin-13)

284-424
932-982
1301-1414
1847-1957

CDS#2 (interleukin-4)

6339-6473
6615-6662
7613-7744
7997-8092

Report how close the predictions were to the actual locations and what those differences were.

3. Take the cow sequence (<https://d396qusza40orc.cloudfront.net/bioinform/assignments/cow.txt>) and use FGENESH. Save the predicted peptide sequence.
 - a. Use the predicted peptide (NOT the mRNA) from FGENESH and do a blastp search. Which (if any) could be the actual protein coded for by that cow DNA?

4. There are a series of problems on the web at http://www.biology.arizona.edu/human_bio/problem_sets/dna_forensics_2/dna_forensics.html. They are very good and should help you learn more about the process of DNA fingerprinting.

5. Which could be the father?

CODIS Locus	Mother	Child	Male #1	Male #2	Male #3	Male #4
CSF1PO	8, 10	10, 11	9, 11	7, 11	8, 11	7, 12
FGA	22, 25	21, 22	21, 23	19, 20	21, 23	25
VWA	18, 19	17, 19	17	16	16, 17	17, 18
DS13S317	12	11, 12	11, 12	11, 13	12	12, 13

Answers

1. a. GenScan

Initial: 178..450

Internal: 622..1429

Internal: 4272..4522

Terminal: 4585..4944

The third exon has the highest probability (0.999)

b. FGENESH

CDSf 601..1429

CDSi 4272..4522

CDSI: 4585..4944

FGENESH only predicts three exons, with a different start codon.

c. TSS should differ from first CDS, which is the start codon.

2. GenScan was 100% correct on both. (if you were off consistently by about 10, you likely included the top descriptive line)

FGENESH - correct

Augustus was correct on IL-13 but missed the final exon of IL-4--predicts alternate start codons and transcripts.

3. a. 12538-12603

b. 10692-10711

c. 13153

d. likely incorrect

4. Answers on tutorial site.

5. Male #4 is eliminated from the CSF1PO locus--he father must have provided the 11 allele.

For FGA, the mother provided 22 so 21 must have come from the father. Only #1 and #3 have a 21 allele so #2 is eliminated.

The father must have a 17 allele for VWA. ONLY #2 (already eliminated) does not.

Since the mother is homozygous for 12 at DS12S317, the 11 allele must have come from the father.

Only #1 and #2 have that.

Only #1 is still a paternity candidate after 4 loci.