Homework 4 of Genomics 2019

Dehe Wang

December 4, 2019

Homework

Question 1: Sequence alignment

Dynamic programming (DP) algorithm is a very common model in bioinformatcs. Please compute the alignment score constrained by the penalty rules in the following table, and use the DP algorithm to find the highest scored alignment.

Table 1: Penalty rules

Contents	Score
Match	1' / base
Mismatch	-2' / base
N	-1' / base
Creat a gap	-3' / gap
Expand a gap	-2' / base

Example

```
Align1  : ACGTGGATCGTA----CGTACTGATCG-------
Align2  : ----TGTANCGTAGTTCGTAG-GATCGTACG
Label   :     MMXMNMMMMGEEMMMMMGMMMMM
Summary: M: 17, X: 1, N: 1, G: 2, E: 2
Score   : 1x17 + (-2)x1 + (-1)x1 + (-3)x2 + (-2)x1 = 6
Where,
    M stands for match, X stands for mismatch,
    N stands for N, G stands for create a gap,
    and E stands for expand a gap.
```

Please calculate the alignment score for the below sequences (10'), find the best alignment (20') and compute the highest alignment score (10').

sequence 1: ACGCNCTTTTCTCGCGTACCTTTACTAATAGAATGCAAAGACGTCCCCCG

sequence 2: CTCGCGTACCTTTACTAAGAGAATGCGNAGACGTCCCCCGGNGGACCGAT

Question 2: Data visualization

Please use the desktop version of JBrowse to visualize the omics data in the "omics" folder, and take a screenshot as pdf format. (Genome 5', BAM 5', BED 5', GTF 5', GFF 5', VCF 5', BigWig 5')

Question 2: Genomic ethics

Please read the second section of the fifth part of the textbook, and discuss the significance of genomic ethics. (500 words, 25')

Please answer question one in the the given "StudentID_Answer.tsv" file, save the screenshot as "StudentID_Screenshot.pdf" and write the answers for question three in "StudentID_Answer.txt". And please package the 3 files as "StudentID_HW4.zip".

You should answer in the format as in the given example. And you should submit your homework to Course website before Dec. 11, 2019.

Grading

The score S of this homework is calculated by the following formula:

$$S = (S_0 - P) * 0.97^d$$

where S_0 is the total score from all questions, P is a penalty and is obtained from the following table, and d is the late time (in day) calculated by the e-mail event stamp.

Table 2: Penalty list

Contents	Score
File naming error	-5' / file
Packaging error	-5'

Reference

1. **Stanford University: Dynamic programming**: `https://web.stanford.edu/class/cs97si/04-dynamic-programming.pdf`
2. **Carnegie Mellon University: Dynamic programming**: `https://www.cs.cmu.edu/~avrim/451f09/lectures/lect1001.pdf`
3. **Wikipedia: Dynamic programming**: `https://en.wikipedia.org/wiki/Dynamic_programming`
4. **JBrowse**: `https://jbrowse.org/`