

Biological Data Analysis (CSE 182) : Assignment 2

Logistics

Submit a zip archive of your code, written answers, output and graphs by email.

1 Sequence Alignment and Gap penalties

The first question is simply to understand the space-saving alignment algorithm. Hopefully, the implementation will clarify some of the questions you may have had. Also, we have had quite a bit of discussion about the impact of different scoring functions on the alignment. In the assignment that follows this one, you will look at the impact of scoring functions on the length of the optimal alignment.

1.1 Problems

1. Implement a space-efficient program *locAL* to align two long DNA sequences with linear gap-penalty. The program should take two DNA sequences as input, along with user defined values for *match-score*, *mismatch-score*, *gapopen*, and *gapextend*. Its should be invoked as follows:

```
locAL <seq_file1> <seq_file2> -m <match> -s <mismatch> -go <gap-open> -ge  
<gap-extend>
```

and should output the following:

- Score of the best local-alignment.
- Length of the best local-alignment
- The alignment itself.

You can build *locAL* in stages as follows:

- (a) (20 pts.) Implement an efficient algorithm to compute the score of the best local-alignment using linear gap penalties (each gap is penalized independently).
- (b) (18 pts.) Extend the code to affine gap penalties by maintaining additional matrices D, I .
- (c) (20 pts.) Extend the code to generating the local-alignment itself.
- (d) (40 pts.) Apply the program to aligning the two pairs of sequences (available on the course web-site) with the following parameters: *match*:1, *mismatch*: -3, *gap-open*: -2, *gap-extend*: -2. Note that the sequences are long, and you can either develop a recursive space-efficient alignment approach, or try another trick, where you just obtain the boundaries of the local alignment (without the information to compute an alignment), and run your code a second time on reduced substrings to get the alignment.

Submit the code, and the output.

2. (2pts.) What language did you use? How much time did you take to do the assignment? Who did you discuss your homework with?