

Grundlagen der Bioinformatik

SoSe 2018

Assignment 02

Submit electronically in Ilias by 30.4.2016, 10h

To solve the following problems, please download the file `download-02.zip` from Ilias or here:
<http://ab.inf.uni-tuebingen.de/teaching/sose2016/gbi>.

1 Calculation of edit distance using DP (3 points)

Write a program that computes the edit distance between two strings using Dynamic Programming as discussed in the lecture. In more detail, your program should expect a FastA file containing two sequences as input and report their edit distance as output. Please implement your solution by editing the file called `EditDistance.java`.

Apply your program to the provided input file called `sequences2.fasta` and print out both your code and the result (to a pdf).

2 Traceback (3 points)

Extend your implementation of `EditDistance.java` so that it uses trace-back to compute an alignment that achieves the optimal score computed in the previous task. Apply the program to the file `sequences2.fasta` and print out both your code and the result (to a pdf).

3 Application to DNA and amino acid sequences (4 points)

Go to the NCBI website and identify a gene for which you can download the DNA sequence and the translated protein sequence, for two different organisms. Let's call these two pairs of sequences A_{dna} , B_{dna} , and A_{aa} , B_{aa} , respectively.

Report the edit distance for each of the two pairs, A_{dna} vs B_{dna} , and A_{aa} vs B_{aa} , respectively. How do the two distances compare with each other? Discuss what features of DNA might cause the edit distance between DNA sequences to be larger than for amino acid sequences? Are there features of amino acid sequences that might lead to elevated distances between amino acid sequences? Is the edit distance equally useful for both DNA and amino acid sequences? (If you can't get your own implementation to work, then please use an online tool.)