

Bioinformatik für Biologen SS 2019

Complementary Homework 7

This homework will not be collected or graded, but its contents and the papers to read may be relevant in an upcoming lecture, test or assignment.

1. Define how specificity and sensitivity is computed.
Give reasons why one might optimize a search method for specificity and sensitivity?
In a BLAST or FASTA search, how do you change the cut-off E-value to make a search more specific or more sensitive?

2. You have computed the following Position-specific-Scoring-Matrix (PSSM) from an alignment of DNA sequences:

	1	2	3
A	0.9	0.3	-1.3
C	-3.4	-0.6	0.4
G	-0.2	0.1	1.1
T	-0.1	0.5	0.5

Which of the following sequences gets the highest PSSM score?

AAA, CCC, GGG, TTT

Which determine from the PSSM, what sequence would get the highest possible score?

3. Describe with your own words what DNA-sequences are matched by the following pattern: C-x(3,4)-[TG]-C (1 sentence)

Which BLAST algorithm is used to search for such pattern in databases?

4. What is the probability for a match in two equally long random sequences if they are aligned without gaps. We assume that the 4 nucleotide occur with the following frequencies for A=1/10, C=4/10, G=4/10, T=1/10
5. You have a sequence AAGT which you want to randomize.
(Assume you have a bag where you put these 4 nucleotides into and draw from it.)
What are the probabilities to sample GTAA if you re-place the letters back into the bag, and if you do not?
6. Use the marble example from the lecture with a bag containing 2 blue and 3 red marbles. You draw twice from the bag without putting the marbles back.
What is the overall probability to draw 2 marbles of equal color?