

Korábbi zh-k feladati

Define the Jaccard coefficient that expresses the similarity of property sets A and B! Calculate the Jaccard coefficient, if $A=[2,4,6,8,10,12]$ and $B=[3,6,9,12]$

Define the Levenstein edit-distance! What is the Levenstein edit-distance of the following sequences: $x=ATTACCA$, $y=ACATTACC$? (gap=-1, mismatch=0, match=1) Write down each step of your computation!

Calculate the entropy of the following multiple alignment!

ATTC
TTCG
TTTG
ATCG

Given the reads and the corresponding de Bruijn graph:

$\{ATG, TGG, TGC, GTG, GGC, GCA, GCG, CGT\}$ (+ábra)

What is the original sequence? Use the de Bruijn graph based algorithm for the computation! (4)

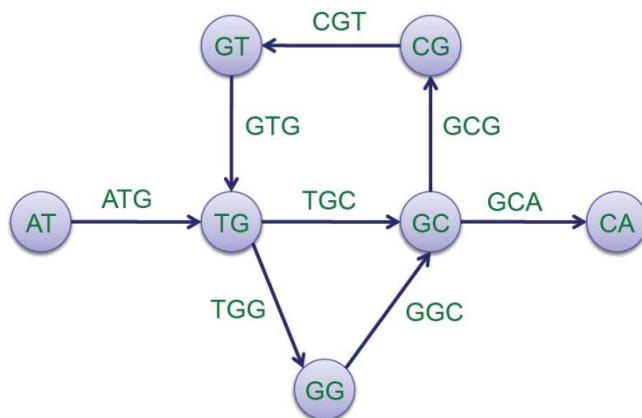
The following distance matrix is given:

	A	B	C	D
B	2			
C	6	6		
D	6	6	2	
E	6	6	4	4

Draw the dendrogram from it based on UPGMA method!

Given the reads and the corresponding de Bruijn graph:

{ATG, TGG, TGC, GTG, GGC, GCA, GCG, CGT}



What is the original sequence? Use the de Bruijn graph based algorithm for the computation!

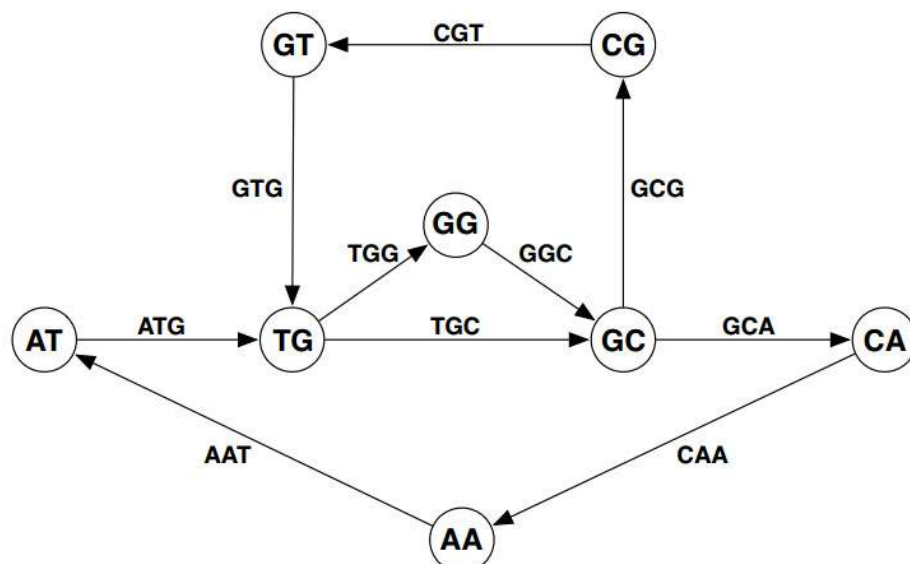
(4)

1. Align the following two DNA sequences (use Needleman-Wunsch algorithm)! The substitution matrix is the following: (X-X: 4, X-Y:2/-4 gap:-4). seq1: ATGC, seq2: AGAGC

2. What is the profile representation of this multiple alignment? (3)
 ATTC-G
 ACTC-G
 CTTGAT
 ACTG-G

3. Given the following Newick formula: ((D,(A,(B,C)))(E,F)). Draw a tree from it! (2)

1. Given the reads and the corresponding de Bruijn graph:
 {ATG,AAT,CAA,GTG,CGT,GCG,GCA,TGC,TGG,GGC}



How would you calculate the two distance values for the following sequences? (4)

v = ATTAATAT

w = TTTATATA

Prepare the global alignment of the words 'APPLE' and 'HAPPY' using the following matrix and a penalty -1 for indels! (7)

	A	E	H	L	P	Y
A	1	-1	-2	-1	-1	-3
E		1	-1	-2	-1	-1
H			1	-3	-1	-1
L				1	-1	-1
P					1	-1
Y						1

We have a dataset and 2 trees about some animals. Which tree represents better the evolution of the given attribute of the animals? (6)

A	B	C	D	E	F	G	H
1	0	1	1	0	1	1	0

