

BIIN 200: Bioinformatics I

Dr. Craig Struble

Fall 2004, Second Midterm Exam

closed book, closed notes, one US Letter cheat sheet, calculators OK

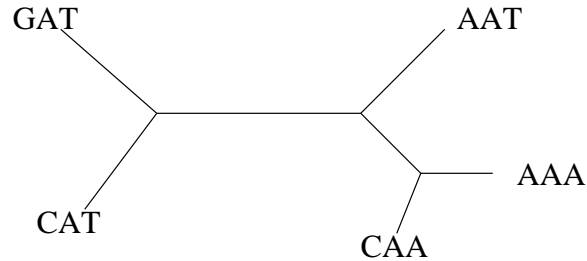
100 points

7 pages

Name: _____

Answer each question in the space provided. Do not use more space than provided. Write neatly. You may use the back of your test pages for scratch paper.

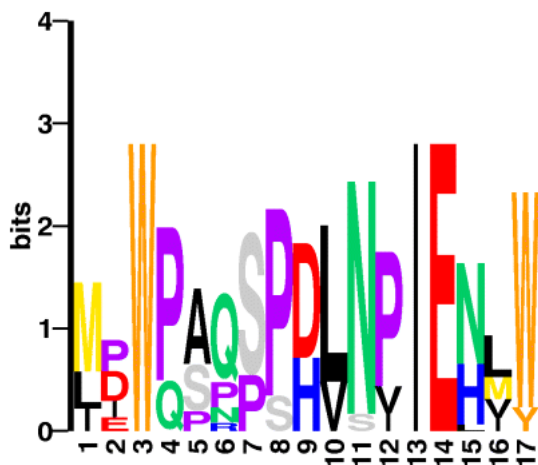
1. [8 pts.] Given the phylogenetic tree below, with sequences at the leaves, assign sequences to the interior nodes that minimize the number of changes between sequences.



2. [5 pts.] Is the tree in the previous question rooted or unrooted?
3. [5 pts.] What phylogenetic method tries to minimize the changes between sequences in a phylogenetic tree?
4. [7 pts.] Does the method in the previous question work well when the rate of sequence change differs along branches? Why or why not?
5. [5 pts.] Using the quick determination of significance of sequence alignment scores, would a local alignment score of 12 bits between two sequences of length 140 be considered significant at the 0.05 level?
6. [5 pts.] What does an expect value $E = 10^{-10}$ mean? Why is this a good value to use for partial alignments of truly related sequences?
7. [5 pts.] What is the complexity of the DNA sequence below?

ACAAACAATA

8. [5 pts.] Given the sequence logo below, which amino acid enriches the region described by the logo?



Logo of Gibbs Block D (Tc1) 9 sequences

9. [10 pts.] Given the position specific scoring matrix PSSM containing log odds scores below, find the position and the ODDS score of the best match in the sequence ACAGGAT.

Position	A	C	G	T
1	-1.7	-1.7	-1.7	1.3
2	-1.7	-1.7	2.1	-1.7
3	-1.7	-1.7	-1.7	1.3
4	-1.7	-1.7	2.1	-1.7

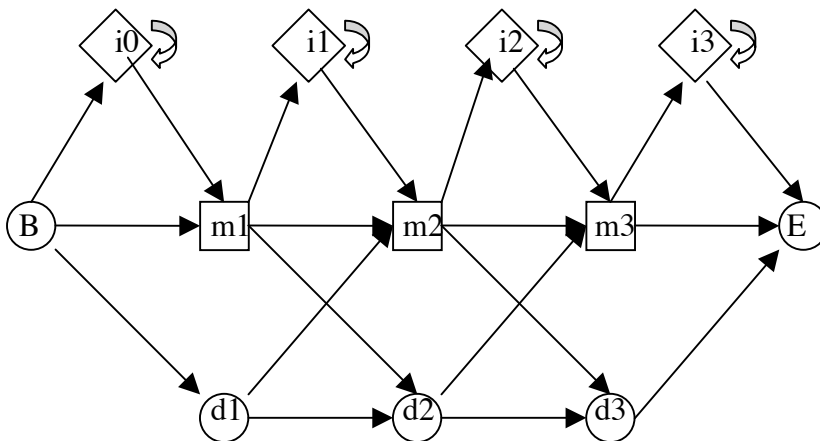
10. [10 pts.] Below is a simplified profile hidden Markov model (HMM) for DNA sequences without transition and emission probabilities assigned. Suppose you have two sequences: ACAT and ACCG. For the first sequence, the path through the HMM is

$$B \rightarrow m1 \rightarrow m2 \rightarrow m3 \rightarrow i3 \rightarrow E$$

and for the second sequence the path is

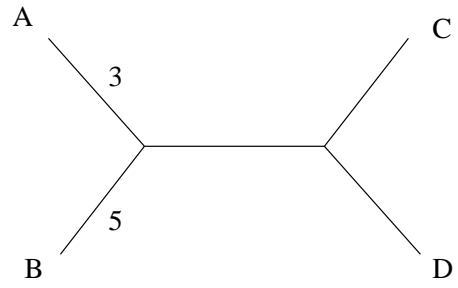
$$B \rightarrow m1 \rightarrow i1 \rightarrow m2 \rightarrow m3 \rightarrow E.$$

What is the resulting alignment of the two sequences?



11. [10 pts.] Given the distance matrix and the partially determined phylogenetic tree below, use Fitch-Margoliash to determine the branch lengths on the remainder of the tree.

	B	C	D
A	8	19	18
B		21	20
C			9



12. [10 pts.] In BLAST, the sequence database is scanned for query words derived from the query sequence. Briefly describe either what BLAST or BLAST2 does after a query word is found in a sequence in the database.

13. [5 pts.] Assuming uniform distribution of nucleotides, what is the probability of matching a DNA sequence of length 11 at random?

14. [10 pts.] Using the BLOSUM62 scoring matrix at the end of this test and a gap penalty of -5, what is the multiple sequence alignment score for the alignment

N-YLS
NKYLS
NK-LS
N-F-S
N-FLS

BLOSUM62 Scoring Matrix

	a	r	n	d	c	q	e	g	h	i	l	k	m	f	p	s	t	w	y	v
a	4	-1	-2	-2	0	-1	-1	0	-2	-1	-1	-1	-2	-1	1	0	-3	-2	0	
r	-1	5	0	-2	-3	1	0	-2	0	-3	-2	2	-1	-3	-2	-1	-1	-3	-2	-3
n	-2	0	6	1	-3	0	0	0	1	-3	-3	0	-2	-3	-2	1	0	-4	-2	-3
d	-2	-2	1	6	-3	0	2	-1	-1	-3	-4	-1	-3	-3	-1	0	-1	-4	-3	-3
c	0	-3	-3	-3	9	-3	-4	-3	-3	-1	-1	-3	-1	-2	-3	-1	-1	-2	-2	-1
q	-1	1	0	0	-3	5	2	-2	0	-3	-2	1	0	-3	-1	0	-1	-2	-1	-2
e	-1	0	0	2	-4	2	5	-2	0	-3	-3	1	-2	-3	-1	0	-1	-3	-2	-2
g	0	-2	0	-1	-3	-2	-2	6	-2	-4	-4	-2	-3	-3	-2	0	-2	-2	-3	-3
h	-2	0	1	-1	-3	0	0	-2	8	-3	-3	-1	-2	-1	-2	-1	-2	-2	2	-3
i	-1	-3	-3	-3	-1	-3	-3	-4	-3	4	2	-3	1	0	-3	-2	-1	-3	-1	3
l	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4	-2	2	0	-3	-2	-1	-2	-1	1
k	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5	-1	-3	-1	0	-1	-3	-2	-2
m	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5	0	-2	-1	-1	-1	-1	1
f	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6	-4	-2	-2	1	3	-1
p	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7	-1	-1	-4	-3	-2
s	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4	1	-3	-2	-2
t	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5	-2	-2	0
w	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11	2	-3
y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	-1
v	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4