

BIIN 200: Bioinformatics I

Dr. Craig Struble

Fall 2004, Midterm Exam

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

100 points

5 pages

Name: \_\_\_\_\_

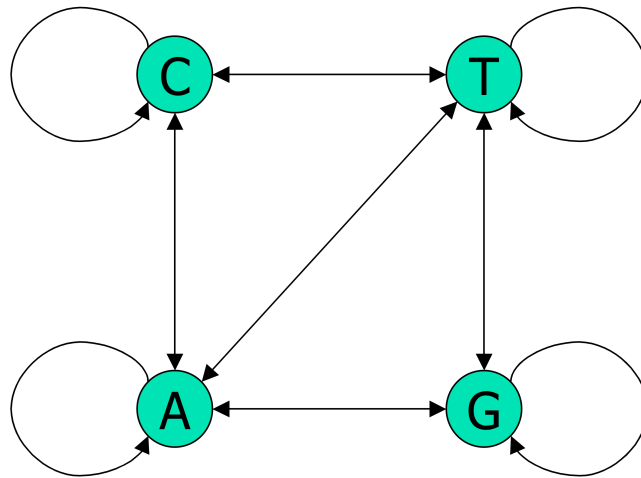
Answer each question in the space provided. Do not use more space than provided. Write neatly. You may use the back of your test pages for scratch paper.

1. [5 pts.] What does the 30 mean for the PAM30 scoring matrix?
2. [5 pts.] *Fill in the blank.* In the central dogma, RNA is \_\_\_\_\_ from DNA.
3. [5 pts.] What kind of mutations are represented by gaps in a sequence alignment?
4. [5 pts.] Provide the name of a database focusing on protein data.
5. [8 pts.] If you were performing pairwise comparisons of DNA sequences of the same gene from two closely related yeast species, which scoring matrix and pairwise alignment algorithm would you use and why?
6. [10 pts.] Given the sequences TAGAA and GTCAT, draw a dot matrix plot, assuming a window and stringency of 1.

7. [5 pts.] Given a DNA sequence, how many possible open reading frames (possible translations into an amino acid sequence) exist?

8. [8 pts.] Give an example of FASTA formatted data.

9. [5 pts.] A Markov model for DNA substitutions is shown below. Circle the edges that represent *transitional* changes.



10. [5 pts.] What kind of sequences contained in GenBank have the highest quality?

11. [5 pts.] Given the DNA sequence below, label the 3' and 5' ends in the typical orientation of sequences, and provide the reverse complement of the sequence.

TTACCGAACTAACTG

12. [8 pts.] Given the two DNA sequences **ACGT** and **ACT**, and using +2 for a match, -2 for a mismatch, and a gap penalty of -1, give an optimal global alignment and its score. *Note: You do not have to use a dynamic programming algorithm.*

13. [8 pts.] If a BLOSUM80 entry has a value of -7, what is the associated odds score?

14. [8 pts.] Convergence is one source of similar sequences. First, draw a diagram representing convergence. Second, would two genes that result from convergent evolution be considered homologous?

15. [10 pts.] Given the matrix below for the Smith-Waterman local alignment algorithm applied to two DNA sequences, provide the resulting alignment and alignment score.

	-	G	C	C	A	A	G	T	A	G	G
-	0	0	0	0	0	0	0	0	0	0	0
A	0	0	0	0	3	3	0	0	3	0	0
C	0	0	3	3	0	2	2	0	0	2	0
G	0	3	0	2	2	0	5	1	0	3	5
A	0	0	2	0	5	5	1	4	4	0	1
G	0	3	0	1	1	4	8	4	3	7	3
C	0	0	6	3	0	0	4	7	3	3	6
G	0	3	2	5	2	0	3	3	6	6	6
T	0	0	2	1	4	0	2	6	2	5	5
A	0	0	0	1	4	7	3	2	9	5	4
T	0	0	0	0	0	3	6	6	5	8	4
G	0	3	0	0	0	2	6	5	5	8	11
A	0	0	2	0	3	3	2	5	8	4	7