



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

4. Align the following two sequences with the Needleman-Wunsch global alignment algorithm. Show the complete dynamic programming matrix, and indicate an optimal traceback.

Sequence 1: **GAATTCA**

Sequence 2: **GACTTAC**

Use the following scoring scheme in the score matrix:

Match: +5

Mismatch: -3

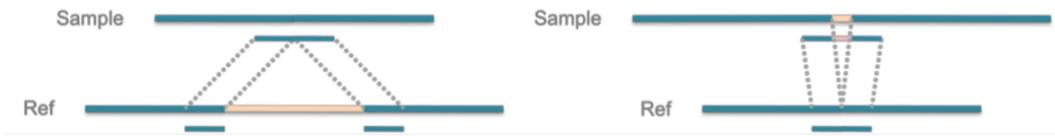
Gap: -4

	--	G	A	A	T	T	C	A
--								
G								
A								
C								
T								
T								
A								
C								

5. How does BLAST speed up the identification of high similarity sequences relative to Needleman Wunsch and Smith Waterman? (5 points)

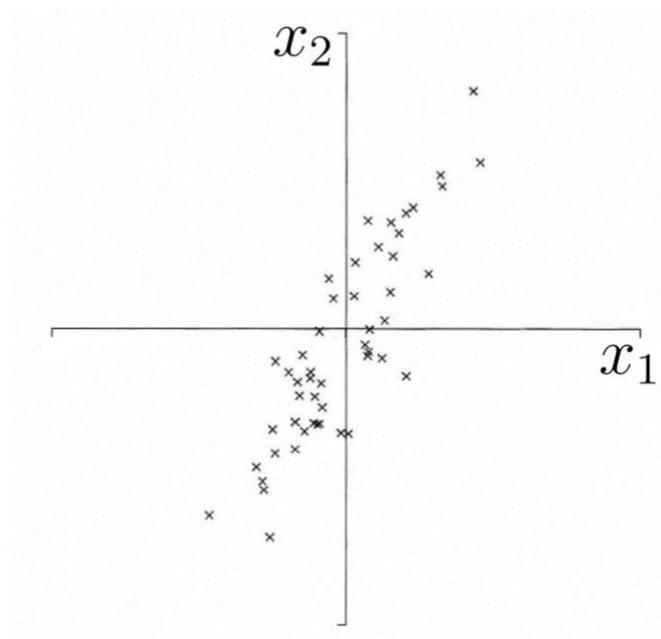
6. Which of the following diagrams illustrates an insertion and which shows a deletion? (5 points)

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



7. What are the differences between the reads produced by Illumina sequencing and those produced by Pacific Biosciences or Oxford Nanopore? How do the reads from PacBio and Nanopore aid in genome assembly? (5 points)

8. Draw the principal components for the data on the graph below: (5 points)



9. In SVD, the data matrix  $A$  is decomposed as  $A = USV^T$ . Suppose  $A$  is a  $4 \times 5$  matrix. What are the dimensions of  $U$ ,  $A$ ,  $V$  respectively? (5 points)

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

10. Which of the following statements is/are correct for general matrix  $A$ ? (5 points)
- Column vectors of  $U$  form a basis for the column space of  $A$
  - Column vectors of  $U$  form a basis for the row space of  $A$
  - Column vectors of  $V$  form a basis for the column space of  $A$
  - Column vectors of  $V$  form a basis for the row space of  $A$
11. Put the following steps in a proteomics experiment in the correct order: (5 points)
- MS
  - Enzyme digestion
  - Isolate and fragment peptide
  - MS/MS
  - Liquid chromatography
12. Name two aspects of the genome that need to be taken into account when calling CHIP-seq signal peaks and suggest a control. (5 points)
13. In relational databases, which of the following is required for second normal form but not first normal form? (5 points)
- No two rows in a table can be identical.
  - All of the non-key columns are dependent on all of the key
  - Every non-primary key column must be dependent on primary key

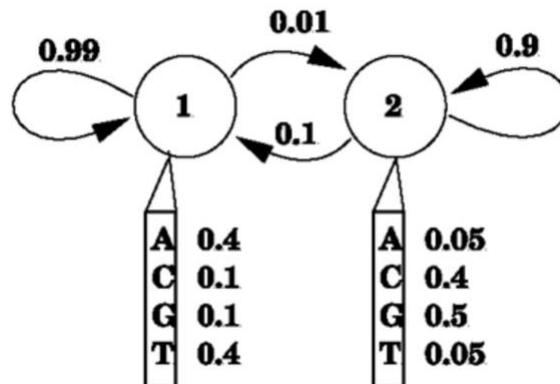
Name: \_\_\_\_\_

- d. All values for a given column must be of the same data type.

14. What do the the lines drawn in a biplot represent? (5 points)

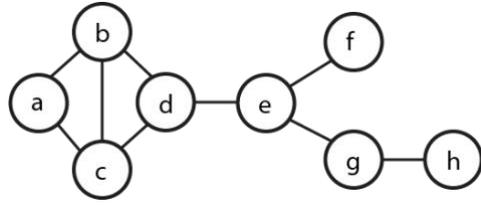
- a. The observation vectors
- b. The variables
- c. The principal components

15. In the hidden Markov model diagrammed below label the following: states, symbols, transition probabilities, and emission probabilities. (5 points)



16. Fill in the adjacency matrix for the following unweighted network: (7 points)

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



	a	b	c	d	e	f	g	h
a								
b								
c								
d								
e								
f								
g								
h								