

University of Pennsylvania
BIOL4536 Fall 2023
Professor: Gregory R. Grant
Exam#2 (27 More Review Problems)

Question 1. Suppose you have two sequences S_1 and S_2 where $|S_1| < 2|S_2|$. Is it possible to have two optimal local alignments between S_1 and S_2 that do not overlap? In other words the two alignments involve completely different bases from both sequences.

Question 2. Assume a pairwise alignment scoring scheme that scores +1 for a match and -1 for a mismatch or indel. In the following multiple sequence alignment, what's the "sum of pairwise alignments" score? *Note: an indel/indel scores 0*

```

A T _ G C G
A _ C G T _
A T C A C _

```

Question 3. True or False. A greedy algorithm never finds an optimal solutions.

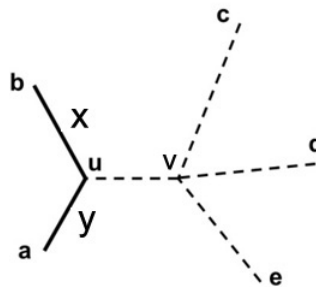
Question 4. A position weight is square. What are its dimensions?

Question 5. True or False. A distance matrix is symmetric.

Question 6. True or False. A distance matrix has ones on the diagonal.

Question 7. Consider the following distance matrix and unrooted tree. What is $d(a, c) + d(b, c) - 2d(u, v) - 2d(v, c)$?

	a	b	c	d	e
a	0				
b	5	0			
c	9	10	0		
d	9	10	8	0	
e	8	9	7	3	0



Question 8. True or False. In a SAM file, the CIGAR string determines the exact location of the alignment.

Question 9. Suppose we are clustering rows of the following block to construct a BLOSUM N matrix. Is there a value of N such that $0 < N < 100$ that will result in only one cluster?

A	A	C	G
C	C	A	A
A	G	C	G
G	G	C	G
G	G	G	G

Question 10. Suppose we are aligning to a DNA database. Is it possible that the background frequencies are different from 25% per nucleotide?

Question 11. True or False. Extension is the most time-consuming part of the BLAST algorithm.

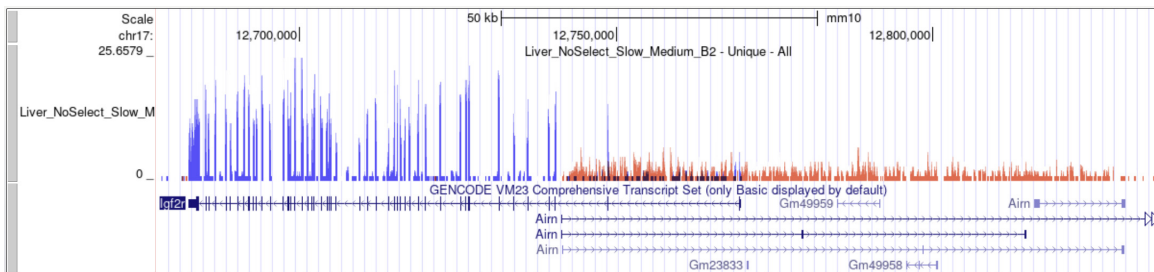
Question 12. True or False. In the BLAST random walk, it's possible that the max excursion is 0.

Question 13. True or False. An RNA-Seq read can be a multimapper to the genome but unambiguously increment the count of one gene.

Question 14. True or False. It's possible for an RNA-Seq read to align uniquely to the genome but be ambiguous about whether it aligns to an intron or an exon.

Question 15. True or False. It's possible for an RNA-Seq read to align uniquely to the genome but be ambiguous about whether it aligns to one of two different genes.

Question 16. Anti-sense transcription is when a gene is transcribed in the wrong direction. Does the following graphic show any anti-sense transcription?



Question 17. True or False. We use different aligners for RNA-Seq as for ChIP-Seq.

Question 18. Which of the following are relevant for ChIP-Seq

- (A) Epigenetics
- (B) DNA
- (C) RNA
- (D) SNP Calling
- (E) Peak Calling
- (F) Gene Regulation

Question 19. Suppose there are 30,000 genes. Suppose we do an RNA-Seq DE analysis and that the largest q -value over all genes is 0.4. How many genes do we expect are DE in total?

- (A) 12,000
- (B) 18,000
- (C) It cannot be determined

Question 20. You are performing 100 tests. You want to use the Bonferroni correction to control the FWER at the level 0.01. Suppose it turns out all genes are significant. What can you say about the maximum of the 100 p -values?

Question 21. True or False. If all p -values and all q -values are one, then all null hypotheses are true.

Question 22. True or False. The "F" in FWER and FDR stand for the same thing.

Question 23. Consider five consecutive SNPs. Suppose a population has exactly two haplotypes with no exceptions. True or False. If an individual is heterozygous at one of the SNPs then they are heterozygous at all five.

Question 24. Consider the following Markov transition matrix. True or False. This Markov Chain perfectly models repeatedly flipping a fair coin.

$$\begin{array}{c} A \\ B \end{array} \begin{array}{cc} A & B \\ \left[\begin{array}{cc} 1/2 & 1/2 \\ 1/2 & 1/2 \end{array} \right] \end{array}$$

Question 25. True or False. The BLAST algorithm one-hit method first extends hits as ungapped alignments.

Question 26. We perform two independent tests $T1$ and $T2$ and consider them significant if their p -value is $\leq C$. True or False. The probability that at both of them are false-positives is exactly double the probability that one (and only one) of them is a false-positive.

Question 27. Consider the following CIGAR string: 3S97M. True or False. The entire read aligned from end-to-end with no indels.