

University of Pennsylvania
BIOL4536 Fall 2023
Professor: Gregory R. Grant
QUIZ#3 (MAKEUP SOLUTIONS)
(Alignment)

Question 1. (1 pt.) True or False. Substitution matrices are used for both protein alignment, DNA/RNA alignment.

ANSWER: True.

Question 2. (1 pt.) True or False. In a substitution matrix, off diagonal entries can be positive.

ANSWER: True.

Question 3. (2 pt.) Explain why PAM and BLOSUM substitution matrices are symmetric.

ANSWER: Since the data we infer from are all extant species, we do not know the directionality of changes. So we must assume they are the same.

Question 4. (2 pts.) Suppose the rows of a block of data have been clustered to generate a BLOSUM60 matrix. Suppose two rows of the block are in the same cluster. Which of the following are possible? Circle *all* that apply:

- (A) The two rows are more than 70% identical.
- (A) The two rows are more than 60% identical.
- (B) The two rows are less than 50% identical.
- (C) The two rows are less than 40% identical.

ANSWER: All of them are possible, since clusters are transitive closures, two sequences in the same cluster can be completely different.

Question 5. (2 pt.) Suppose there were only three amino acids *A*, *B* and *C* (as opposed to 20). The matrix below could be which of the following? Circle the ones that apply.

- (A) A Matrix of counts.
- (B) A Markov Chain Probability Transition matrix.
- (C) A Substitution matrix.
- (D) None of the above. ← **THIS ONE** *Each row of a Markov transition matrix must add up to one*

$$\begin{array}{c} A \quad B \quad C \\ A \quad \left[\begin{array}{ccc} 0.3 & 0.1 & 0.2 \end{array} \right] \\ B \quad \left[\begin{array}{ccc} 0.4 & 0.1 & 0.3 \end{array} \right] \\ C \quad \left[\begin{array}{ccc} 0.3 & 0.8 & 0.5 \end{array} \right] \end{array}$$

FLIP OVER FOR LAST QUESTION.

Question 6. (2 pts.) Suppose the following block of data has been clustered into four clusters (in order to construct a BLOSUM matrix), so that the first four sequences comprise one cluster and the bottom three are each in their own cluster. So one cluster of size 4 and three clusters of size 1.

Count the number of times a *B* aligns with a *C* for the goal of constructing a BLOSUM matrix. So you have to count properly taking into account the clusters, as is done in the BLOSUM method.

<i>A</i>	<i>B</i>	<i>A</i>	<i>A</i>
<i>A</i>	<i>B</i>	<i>A</i>	<i>A</i>
<i>A</i>	<i>A</i>	<i>A</i>	<i>A</i>
<i>A</i>	<i>B</i>	<i>A</i>	<i>A</i>
<i>A</i>	<i>A</i>	<i>B</i>	<i>D</i>
<i>A</i>	<i>C</i>	<i>B</i>	<i>A</i>
<i>D</i>	<i>A</i>	<i>B</i>	<i>A</i>

ANSWER:

$$\frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$$