University of Pennsylvania

BIOL4536 Fall 2023

Professor: Gregory R. Grant

QUIZ#4

(BLAST/DNA-Seq)

Question 1. (2 pts.) BLAST is (circle all that apply)
(A) Global Aligner
(B) Local Aligner ⟵ **THIS ONE**
(C) Protein Aligner ⟵ **THIS ONE**
(D) DNA Aligner ⟵ **THIS ONE**
(E) Low homology aligner ⟵ **THIS ONE**

Question 2. (2 pts.) True or False: The BLAST E-value is the expected number of alignments, with the observed score or higher, between the query and a database of random sequence.

ANSWER: True.

Question 3. (2 pts.) Suppose the scoring scheme for a BLAST DNA search is $+1$ for match and -1 for mismatch. The expected score for an ungapped alignment of two *random* sequences (background probability $1/4$ per nucleotide) of length N is:

(A) 0
(B) $-\frac{1}{4}N$
(C) $-\frac{1}{2}N$
(D) $-\frac{3}{4}N$ ⟵ **THIS ONE**
(E) $-N$

Question 4. (2 pts.) A fastQ file has how many rows per read?
(A) One
(B) Two
(C) Three
(D) Four ⟵ **THIS ONE**

Question 5. (1 pts.) True or False. In paired-end data, we get the sequence of an entire fragment whenever the length of the fragment is less than three times the length of the reads.

ANSWER: False. It would be true if that said "two times the length" not "three times the length".

Question 6. (2 pts.) In your own words, in high throughput sequencing, what's the difference between single-end and paired-end data?

ANSWER: In sequencing, long DNA or RNA molecules are fragmented into small fragments on the order of 200-500 bases. If reads have length N then we sequence N bases from either one end (single-end) or both ends (paired-end) of the fragments.