Week 4 Lab: BLAST worksheet

Name:

Instructions: A copy of this worksheet will be randomly selected from your group at the end of Week 4's lab. Answer questions completely and clearly, fully indicating your reasoning/ thoughts (often writing in complete sentences is helpful). I expect the responses to have no typos and minimal grammatical error. You are allowed/encouraged to discuss questions with your labmates.

A1. Which of the hits under 'Description' is the best hit for your unknown sequence (provide its name and accession ID)? Use the metrics provided in the search results to support your reasoning. ***Note: We will ignore all computationally predicted genes and focus only on experimentally verified genes. This is because the experimentally verified genes are the only ones that we can be sure actually exist.***

The following question is out of order to keep questions associated with tables on the same page. Skip it until you have reached the appropriate step in the protocol.

B1. What is the difference between "XM_##" versus "NM_##" versus "NP_##"?

A2. Fill in the following table (Table 1) with information for each alignment block from the *blastn* search. Make sure to add the total number of mRNA base pairs in the table title. If you do not have enough ranges to fill all rows, leave the remaining rows empty. Don't forget to sort your ranges by subject start position!

Table 1. *Blastn* results for the unknown sequence and the *D. melanogaster* legless mRNA (_____bp)

Range #	Subject Position Range	Query Position Range	E-value	% Identity

A3. Use data from Table 1 to assert whether the entire *D. melanogaster legless* mRNA aligns to our unknown sequence. That is, provide evidence to support or reject whether the sequences matching *D. melanogaster legless* mRNA in your unknown sequence show collinearity. Cite table(s) as appropriate.

A4. Write a hypothesis on what gene, if any, your *D. yakuba* unknown sequence contains.

B1. *Moved to the first page of the worksheet.*

B2. Fill in the following table (Table 2) with information for each alignment block from the *blastx* search. Make sure to add the total number of protein amino acids in the table title. If you do not have enough ranges to fill all rows, leave the remaining rows empty. Don't forget to sort your ranges by subject start position!

Table 2. *Blastx* results for the unknown sequence and the *D. melanogaster* legless protein (_____aa)

Range #	Subject Position Range	Query Position Range	E-value	% Identity	% Positives

B3. Use data from Table 2 to assert whether the entire *D. melanogaster* legless protein aligns to our unknown sequence. First add what support you find, using similar reasoning as in question A4. Then note any discrepancies, clearly distinguishing between supporting and conflicting evidence. Remember to be specific, assuming the reader has no image to rely on. Cite table(s) as appropriate.

C1. Fill in the following table (Table 3) with information for each CDS (query) you compared to the unknown sequence (subject) using the 2-sequence comparison in the *tblastn* search.

Table 3. *tblastn* for the coding sequences (CDS) for *legless* in *D. melanogaster* and the unknown sequence

CDS #	FlyBase ID	Query Position Range	Subject Position Range	E-value	% Identity	% Positives

C2. Use data from Table 3 to explain whether all CDS of the *D. melanogaster legless* gene are accounted for in our unknown sequence. Make sure to explain what this table resolves from the discrepancies discussed in question B3. Cite table(s) as appropriate.