

Biology 191 - Week 8 - BLAST Activity Answer Key

CDS # (Number of complete codons)	Protein Alignment (Start-End)	Our Sequence Alignment (Start-End)
6 (60)	1-60	9344-9168
5 (96)	1-96	9104-8817
4 (69)	1-69	8311-8105
3 (667)	1-667	5371-3365
2 (63)	1-63	2800-2606
1 (510)	1-510	2550-1018

Questions

1. Return to p. 110 and recall the problem with the first exon (should say 'alignment block'), which spanned from 9344-8814, in the initial blastx search (which compared the translated *D. yakuba* unknown DNA sequence to the RefSeq protein database). What do the results in the table on p. 114 indicate about a possible solution to this problem?

The alignment block identified on p. 110 that spans from 9344-8814 actually consists of two separate exons, one that extends from 9344-9168 and the other from 9104-8817. Remember that this alignment block also contained a large gap of unaligned amino acids, including a stop codon. If the first exon ends at nt 9168, and the next one begins at 9104, the bases in between are intron sequences. The large gap of unaligned bases is because non-coding intron sequences are much less conserved than exon sequences.

2. What other problems did you encounter with the blastx search results on p. 110, which were resolved by the results in the table on p. 114?

There was significant overlap for two pairs of adjacent alignment blocks: 1-158 (9344-8814) overlaps 148-229 (8332-8099) and 228-897 (5374-3359) overlaps 890-959 (2821-2606). If you look back at the blastx alignments, the beginning of the 148-229 alignment is much worse than the end of the 1-158 alignment. Thus, the 148-229 alignment block has likely been extended into the preceding intron. The same is likely true for the beginning of the 890-959 alignment block. The data from the above table confirm this, as the 3rd exon begins at 8311 and not 8332, and the 5th exon begins at 2800 and not 2821.

3. Examine the table on p. 114 (and the bl2seq alignments). Were all of the alignments full-length, e.g. were all of the amino acids at the beginning and end of each *D. melanogaster* (Query) CDS accounted for in the tblastn alignments?

No, two of the six CDS alignments were not full length.

4. Which of the CDS were missing amino acids (aa) at the end(s) of the tblastn alignments and which amino acids were missing in each (by number order, such as aa 1-5, 665-667, etc.)?

The alignment for CDS 5 was missing the last (96th) amino acid and the alignment for CDS 4 was missing the first 22 amino acids.