Homework # 7

Handed out: 10/26/06
Due: 11/7/06

The current homework requires you to be logged onto poppy.poplee.org, or on a machine that has the Seqaln package (http://hto-13.usc.edu/software/seqaln/index.html) installed.

1. Familiarize yourself with the Seqaln package (see “man seqaln-intro” to get started). Answer the following questions:
    a. What program would you use for local alignment with similarity scores?
    b. What program would you use for global alignment with similarity scores?
    c. How do you specify match/mismatch scores and gap open/extension penalties?
    d. How do you tell these programs to output the top k alignments?
2. Perform the local sequence alignment of the following two sequences:
   >A
   TCAGTTGCCCAAACCCGCT
   >B
   AGGGTTGACATCCGTTTT
    a. First set alpha (gap open) = beta (gap extension) = 1000 and perform the alignment with match score = 10, and with mismatch penalty score = 10, 7, 5, 3. Observe the effect of reducing the penalty score by examining the ten highest-scoring alignments for each case. What trend do you observe? Explain it.
    b. Now with match score = 10 and mismatch penalty score = 10, explore the effect of changing alpha from 15 to 10, then 5 (holding beta at the default 3). What trend do you observe? Explain it.
3. Download sequences contained in accession numbers P21189 and NP_143776.1 from the NCBI GenBank database (http://www.ncbi.nlm.nih.gov).
    a. Perform global alignment using the “blosum62” as the penalty matrix and the default indel settings (alpha = 15, beta = 3). Examine your result: can you discern a region that likely will produce high-scoring local alignments?
    b. Perform local alignment on the same sequences. Did the result from local alignment agree with your prediction based on the global alignment?
    c. Use the contiguous sequence of the alignment found at step b (remove gaps from NP_143776.1 portion of the alignment) as a query for an NCBI Blast search of the non-redundant databases. Are any putative conserved domains found? Check to see if there are any significant hits to *Danio rerio* (zebrafish) or *Arabidopsis* sequences.

For all these exercises don’t go too overboard on the answers, short answers are more than OK. Give me a quick printout of the alignments with the interesting regions highlighted in pen. Similarly, just print out the significant hits for 3.c.