## Part 2. Global sequence alignment

**Assigned: September 25th, 2012**
**Due: October 9th, 2012**

**Overall weight: 10% of total project score**

**Specification:** You must implement the dynamic programming algorithm described in class to construct the global (end-to-end) alignment of two DNA sequences. Your program must accept two DNA sequences in FASTA format and output a global alignment of these sequences in the following format:

```
Edit distance = 7

Seq1  ATTC-TCAT--TAGGACCGGC
       ||  ||||  ||| ||| ||
Seq2  -TTGATCATGGTAG-ACC-GC
```

Note: the vertical bars indicate characters that match between the two sequences.

If the sequences are too long to be displayed on one line (assume a line has 80 characters), the alignment should wrap around as shown below:

```
Edit distance = 12

Seq1  ATTC-TCAT--TAGGACCGGC
       ||  ||||  ||| ||| ||

Seq2  -TTGATCATGGTAG-ACC-GC

Seq1  GCACATCA-G-TAGGACC
      | | |||| | ||| |||

Seq2  GTAGATCATGGTAG-ACC
```

**Interface:** Your program should accept 5 parameters on the command line: the names of the files containing the two sequences, and the scores for a match, mismatch, or gap in the alignment. Below are two examples:

Simple: `myProg file1.fa file2.fa 3 -1 -2` (parameters are simply listed in order)

With options:
`myProg -s1 file1.fa -s2 file2.fa -match 3 -mismatch -1 -gap -2` (use command-line options)

You can pick any option you wish, and even allow certain parameters to be missing (in which case they would be assigned default values), however you must indicate in a README file how to run your program (and what the default parameters are if they are not specified).

**Additional details:** Any questions about this assignment should be sent to both myself and the TA.
You can assume that the two FASTA files contain exactly one sequence (or if they contain more than one just use the first sequence in each file).

You can now use any of the Bio* libraries to read the FASTA files (but not to perform the alignments).