Part 3. Local sequence alignment with affine gap scores

Due: Thursday November 8th, 2012

Overall weight: 30% of total project score

Specification: **Extend** the program you wrote in Part 2 to achieve the following:

- Perform local alignment (find best matching substring)
- Accept alignment scores formatted as a HOXD matrix (detail below)
- Gap penalties are "affine" - the combination of a gap opening penalty (paid once per group of gaps) plus a gap extension penalty (proportional to number of gaps within a group).

Output format: The output of your program must follow the format shown below - note the coordinates within the two strings of the aligned region. Just as before, if the strings are too long to be printed on one line (assume one line is 80 characters) you will need to wrap the alignment around on multiple lines.

```
Score = 2, Identities = 14/18 (78%), Gaps = 3/18 (17%)
Seq1  ATTC-TCAT--TAGGACCGGC
      |||  ||||  ||| ||| ||
Seq2  ATTGATCATGGTAG-ACC-GC
```

Interface: follow the same rules as for Part 2, except that you must now accept two new values - gap opening and gap extension penalties - and a new file - the HOXD matrix specifying the substitution/match scores.

HOXD matrices: A sample matrix is provided below.

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).

HOXD matrix
47.5% G+C

A	C	G	T
91	–114	–31	–123
–114	100	–125	–31
–31	–125	100	–114
–123	–31	–114	91