GlobalAlignment Documentation

Module name: GlobalAlignment
Description: Pairwise alignments of two DNA sequences by dynamic programming.
Author: Michael Kamal (Broad Institute) kamal@broad.mit.edu
Date: 12/5/03
Release: 1.0

Summary: Three dynamic programming algorithms are implemented:

- Global/Needleman-Wunsch
- Local/Smith-Waterman
- Iterative Local.

In the iterative local approach, a collection of non overlapping local alignments is found in a greedy, iterative fashion:

1. Find the best local alignment
2. If the best local alignment scores below the minimum threshold end the analysis for the region.
3. Otherwise, accept the alignment, exclude the aligning region from further analysis and the remaining region into subregions:

4.Repeat for the analysis for each of the subregions, larger than a minimum size cutoff

We use standard, affine scoring parameters: match (>0), mismatch(<0), gapOpen (<0) and gapExtend (>0). For sensible results, one should have : gapOpen<gapExtend and 2*gapOpen<mismatch. The first gap opened incurs gapOpen, subsequent gaps score gapExtend.

The input sequences should be supplied as separate, single record FASTA files. Valid sequence characters are :A,a,T,t,G,g,C,c,N,n. Lower and uppercase are treated the same. "N" characters (unknown bases) will not be aligned to real (non-N) bases but will be allowed to fall into gaps.

Only alignments in the forward ('+' relative orientation) are considered, no reverse complement searching is carried out.

References:

- Michael S. Waterman. Introduction to Computational Biology: Maps, Sequences and Genomes, 1995, CRC Press.

Usage/Example:

```
java DemoStandardLocal query_fasta_file subject_fasta_file match
mismatch gapOpen gapExtend output_file

java DemoStandardGlobal query_fasta_file subject_fasta_file match
mismatch gapOpen gapExtend output_file


java DemoStandardIterativeLocal query_fasta_file  subject_fasta_file
match mismatch gapOpen minSize gapExtend output_file


java DemoStandardLocal m.fa h.fa 1 -1 -5 -1 alignment.out
java DemoStandardGlobal m.fa h.fa 1 -1 -5 -1 alignment.out
java DemoIterativeLocal m.fa h.fa 1 -1 -5 -1 20 10 alignment.out
```

Parameters:

Name	Description
query_fasta_filename	query sequence
subject_fasta_filename	subject file name
match	integer score for a match
mismatch	integer score for a mismatch (<0)
gapOpen	cost for opening a gap (<0)
gapExtend	cost for extending a gap (<0)
minSize	Iteratlive Local only: smallest subregion to consider
minScore	Iterative Local only: smallest acceptable alignment score.
output_file	Text file

Return Value: Text file with alignments.

Platform dependencies:

Task type:	Sequence Analysis
CPU type:	any
OS:	any
Java JVM level:	any
Language:	Java
Support files:	none

Native command line: <java> <java_flags> -cp <libdir>DPJava.jar
DemoStandardGlobal <query_filename>
<subject_fasta_filename> <match> <mismatch> <gapopen>
<gapextend> <output_file>