Introduction to bioinformatics, Autumn 2006, Exercise 3

6.10.2006

1. (a) Download sequences contained in accession numbers NM_054111 and NM_173027 from the NCBI GenBank database. How do the sequences relate to each other according to the database?

   (b) Perform global alignment for the above sequences using the 'align' tool in the sequence analysis toolbox of European Bioinformatics Institute at http://www.ebi.ac.uk with default parameters (choose parameter Molecule = DNA). Describe the results.

   (c) Perform global alignment again with parameters Gap Open = 1.0 and Gap Extend = 1.0. Compare the results with the previous alignment.

2. (a) Perform local alignment for the sequences NM_054111 and NM_173027 using the align tool (see problem 1) with default parameters.

   (b) Perform local alignment again with parameters Gap Open = 1.0 and Gap Extend = 1.0. Compare the results with the previous alignment.

3. Using the CLUSTALW tool at `http://www.ebi.ac.uk`, perform multiple alignment on the following sequences (also available at `http://www.cs.helsinki.fi/mbi/courses/06-07/itb/data/sequences.txt`) using default parameters.

   ```
   cggtagctgcctaggcatac
   tcgtagtgcctagcatat
   ggttagctgctagcatta
   cagtgcggtgagtgaata
   aagctatcgatgaagattat
   gagtagcgcaggatgattac
   tagtagcgaagagatttac
   aggcagctggctagcataa
   ```

   How good was the alignment? What can you say about the possible evolutionary history of the sequences by inspecting the cladogram?

4. (Chapter 12, Exercise 5) Find out the parsimony score for the tree
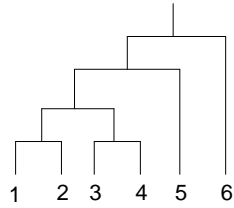
given the following set of sequences

```
1 TCAA
2 GCAT
3 TTTT
4 GATA
5 GAAC
6 ATAG
```

Indicate the F set at each vertex of the tree.

5. (Chapter 12, Exercise 4) Find the unique tree with distances given by

	a	b	c	d
a	0	3	6	5
b		0	7	6
c			0	3
d				0

Hint: consider additivity.