Assignment – IV
(MSA and Phylogeny)

Deadline: 14th April

1) What are the applications of multiple alignment? How does a multiple alignment carry more information than a pair-wise alignment? Explain briefly.

2) What is sum-of-pairs score? What are the drawbacks of sum-of-pairs score? Can you think of alternative scoring system?

3) Describe the steps involved in the progressive alignment approach. What are the drawbacks of this approach? How can the shortcomings of this approach overcome?

4) Assume we have several sequences that are L=50 residues long. The pairwise alignment of two sequences takes 1second on our computer. The alignment of 4 sequences takes $(2L)^{N-2} = 10^{2N-4} = 10^4$ seconds. If we had unlimited memory and were willing to wait for the answer until just before the sun burns out in 5 billion years, how many sequences could our computer align?

6) Obtain MSA of the following four sequences using the progressive approach and compute the final score of the alignment using sum-of-pairs. Using the scoring scheme: match = 1, mismatch/indel = -1

   S1: GATTCA
   S2: GTCTGA
   S3: GATATT
   S4: GTCAGC

7) Given below is the multiple sequence alignment of 6 species:

Species:	Site: 1 2 3 4
1	T C A A
2	G C A T
3	T T T T
4	G A T A
5	G A A C
6	A T A G

Find the parsimony score for the tree ((((1,2), (3,4)), 5), 6). Indicate the F sequence (base assignments at the nodes) at each vertex of the node.

8) Give reasons why the phylogeny of data can be different from the phylogeny of species. Describe in brief methods to detect gene duplication, homoplasy and horizontal gene transfer.

9) For *n* terminal taxa, the number of unrooted trees is given by

$$\frac{(2n-5)!}{2^{n-3}(n-3)!}$$

and the number of rooted trees is given as

$$\frac{(2n-3)!}{2^{n-2}(n-2)!}$$

Make a table of values and graph the two functions for $n \leq 10$.

10) Consider the trees in the figure below:
   a. Which of them are the same, as rooted metric trees?
   b. Which of them are the same, as unrooted metric trees?
   c. Which of them are the same, as rooted topological trees?
   d. Which of them are the same, as unrooted topological trees?
   e. For which trees does a molecular clock appear to be operating?