Algorithms for Computational Biology
Quiz 2
Time 1 Hr. 10 Min.,
Each question carries 5 points

1. Write any algorithm (or code snippet) with $\log_2 n$ time complexity.
2. How is base quality score computed (in the context of DNA sequencing)?

3. Use Smith-Waterman approach to find out the best local alignment of the following sequences: "ACCTAAGG" and "GGCTCAATCA". Consider +2 for match, -1 mismatch and -2 for gap. Also show the alignment.

4. Describe the steps involved in genome assembly.

5. I have N 1000 nucleotides long DNA sequences, wherein I am searching for a planted motif of length 10 using bruteforce. My scoring scheme is rather simple - I need to calculate the average Hamming Distance of the individual motifs (across DNA sequences) with the consensus one. The only challenge is that it's unknown how many of these N sequences have the motif in it. All that I know is that at least one of these sequences definitely has the motif. I will have to try X collections (configurations) of motifs. What is X (Explain clearly)?
6. Write the pseudo code of a greedy algorithm for motif discovery.