

Midterm: CAP 6515 Algorithms for Computational Biology

10/17/2018 3:00pm - 10/22/2019 3:00pm

Note: Do not discuss the exams with anyone except the instructor. If you look for answers to the exam problems on the internet, please make sure acknowledge that. You should treat it as an in-class exam with only the notes/slides from the Webcourses.

1 Suffix tree and suffix array

- (1) Consider to build the suffix tree of the string *ACAGCTCACAGCTC* using Ukkonen's algorithm. Illustrate the **implicit** suffix trees after each phase with suffix links (20 pts).
- (2) Given a suffix tree of a string (note that it is not a implicit suffix tree), present a linear algorithm to retrieve the original string (20 pts).

2 Dynamic programming

- (1) Illustrate the alignment graph and the optimal path for computing the edit distance between two strings *CTGGCCATGAC* and *CGGCTCATCAC*. A replacement/insertion/deletion will cost 1. A match will cost 0 (15 pts).
- (2) Devise an algorithm to compute the number of distinct **optimal** alignments (the number of the optimal paths in alignment edit graph) between a pair of DNA sequences. A replacement/insertion/deletion will cost 1. A match will cost 0. Note that there may be multiple optimal alignments. Note that the trace-back is not needed. The output is a number. Please write down the recursive functions and describe the notations you use (20 pts).
- (3) For two RNA sequences $s_1s_2...s_m$ and $t_1t_2...t_n$, two inter-RNA base pairs (s_i, t_j) and (s_k, t_l) are called *crossing* if (1) $i < k$ and $j > l$; or (2) $i > k$ and $j < l$. Devise an algorithm to compute the maximal number of non-crossing base pairs between two RNA sequences (20 pts).
- (4) Who invented the technique of dynamic programming? Why did he name it? (5 pts)