

# CMPT 441/711

## Bioinformatics Algorithms: Introduction to Computational Molecular Biology

Fall 2014

**Course Coordinates:** WMC2202, MWF 11:30-12:20

**Instructor:** Cenk Sahinalp, [cenk@cs.sfu.ca](mailto:cenk@cs.sfu.ca)

**Instructor's office hours:** TASC I 9003, MW 12:30-1:30

**Teaching assistant:** Salem Malikic, TASC I 9003, [smalikic@sfu.ca](mailto:smalikic@sfu.ca)

### Course outline

- *1:* A brief introduction to algorithmic problems in computational biology. Elementary algorithm analysis; big-oh notation and its limitations.
- *2:* Knuth-Morris-Pratt algorithm for exact string matching.
- *3:* Suffix trees and suffix arrays. Linear time suffix sorting and Burrows-Wheeler transform.
- *4:* Karp-Rabin fingerprinting.
- *5:* The notion of global string/sequence alignment with mismatches and indels. Applications of alignments to genome analysis. Fundamentals of dynamic programming (DP). Edit distance computation; approximate string matching.
- *6:* Longest common subsequence problem.
- *7:* Multiple sequence alignment (MSA); NP-hardness and pseudo-polynomial algorithms for MSA.
- *8:* A constant factor approximation method for MSA. A crash course in Molecular Biology.
- *9:* Practical issues related to sequence alignment. Weight functions and their effect on DP; gap penalties and alignment under linear, convex and concave penalty functions; speeding up alignment via four russians trick. DNA archive search: BLAST local alignment search heuristic, short-cuts and short-comings. Flash and PatternHunter. Protein alignment and search; discussion on probabilistic weight matrices; PAM and BLOSUM.
- *10:* Introduction to probabilistic methods. Hidden markov models. CpG island determination via HMMs.

- *11*: Sequence alignment via HMM. Vitterbi method. Backward-forward probabilities. Expectation maximization. Baum-Welch procedure. Gibbs sampling.
- *12*: Phylogenetic tree (re)construction; maximum likelihood; parsimony, neighbor-join method.
- *13*: High throughput sequencing; assembly through shortest common superstring problem; reduction to TSP and hardness of approximation; heuristic solutions through Euler tours on deBruijn graph.
- *14*: Genome rearrangements and structural alterations; mapping and implied tools.
- *15*: RNA structure prediction through binding energy minimization.
- *16*: RNA-RNA interaction prediction. miRNA target prediction.
- *17*: Protein-protein interactions. Properties of interaction networks.