A Course on Selected Topics in Computational Molecular Biology

Giuseppe Lancia
University of Padova and Celera Genomics

Mon 6/5 (2h) **Preliminaries and Alignment Problems**

DNA, amino acids and (Shotgun) sequencing. Assembly and the Shortest Superstring Problem. Basic Alignment. Sum-of-Pairs Multiple Sequence Alignment: Carillo and Lipman's dynamic programming. Kececioglu et al.'s MSA program.

Tue 7/5 (2h) **Still Alignments. The Tree Alignment Problem**

Gusfield's 2-approx for SP and Bafna, Lawler and Pevzner's improvements. Heuristics for SP-Alignment. Feng and Doolittle's "once a gap, always a gap" algorithm. Aligning along the min routing-cost tree. Tree Alignment and philogenies. Sankoff's dynamic programming algorithm. Hints to Gusfield's and Wang Lifted Alignments for Approx guarantee and to Ravi and Kececioglu approx algorithm.

Wed 8/5 (3h) **End Alignments. Single Nucleotide Polymorphisms (SNPs)**

Steiner Alignments: A heuristic approach to Tree alignment and sequence alignmment. Integer Programming approaches to alignment. Non crossing matchings. Reinert et al. I.P. algorithm for RNA alignment.

Single Nucleotide Polimorphisms. Haplotypes and genotypes. The population haplotyping problem. Clark's rule for haplotype inference. Gusfield's graph-theoretic model and Integer Programming approach.

Thu 9/5 (3h) **SNPs problems for population and individuals**

Inferring Ancestral history from haplotypes: the Haplotype Coloring Problem and the Perfect Phylogeny SNP Haplotyping Problem for populations.

The single individual haplotyping problem. Complexity, polynomial algorithms and heuristic/branch and bound algorithms.

Mon 13/5 (3h) **Genome rearrangements**

Rearrangements and evolutionary distances. Inversions, Transpositions, Translocations, Duplications and Deletions. Single chromosome rearrangements. Breakpoints and breakpoint graph. Sorting by reversals. Bafna and Pevzner's bound for unsigned SBR and its tightness. Caprara, Lancia and Ng's algorithm for SBR.

Wed 15/5 (3h) **End rearrangements. Protein structure problems**

The diameter of SBR.Cayley graphs and expected reversal distance. Sorting by reversals-and-transpositions. Simple approx algorithm. Signed Sorting by Reversals. Polynomial results. A linear-time algorithm for the signed reversal distance.

Protein structures. The PDB. The folding determination problem. The HP model and its complexity. CASP and CAFASP competitions and folding strategies.

Mon 16/5 (3h) **Protein fold comparisons**

Comparing folds, and structural data bases (SCOP, DALI, ...). Programs for structure alignments. Contact maps and contact map alignments. An I.P. approach to contact map overlap. Caprara and Lancia's Lagrangian algorithm for contact map overlap. Genetic Algorithms and local search heuristics.