

Topics and references

References:

[D] Biological sequence analysis: probabilistic models of proteins and nucleic acids, Richard Durbin et al., Cambridge University Press.

[KR] Fundamental Concepts of Bioinformatics, D.E. Krane and M.L. Raymer, Pearson Education.

[G] Algorithms on strings, trees, and sequences, D. Gusfield, Cambridge University Press.

[E] Statistical Methods in Bioinformatics: An Introduction (Statistics for Biology and Health), W. J. Ewens and G. Grant, Springer.

- Evolution
 - Orthologs and paralogs [papers A.1, A.2, B.2]
 - Causes of genetic variations [papers A.1, A.2]
 - Reasons for differing rates of evolution [papers A.1, A.2, A.3, B.3]
 - Molecular clock hypothesis
 - Hardy-Weinberg equilibrium
 - Neutral evolution and genetic drift
 - Modeling with Markov Chains [D]
 - Substitution matrices [KR, E, B.1, B.4, B.5]
 - DNA (Kimura and Jukes-Cantor)
 - Protein (BLOSUM and PAM)
- Sequences
 - The idea of local alignment and global alignment [D, G]
 - Log odds score [G]
 - Gap penalty (linear and affine) [G]
 - Dynamic programming formulation [G]
 - Score
 - Edit distance
 - Suffix tree idea [G]
 - BLAST (basic idea) [B.8]
 - Statistics of alignment [B.9, E]
 - Extreme value distribution
 - p-value and e-value
- Motif finding
 - Structure of genes [KR]
 - Coding and non-coding regions
 - Promoters and other control elements
 - Differences between eukaryotes and prokaryotes

- EM formulation [D]
- Gibbs sampling formulation [D]
- Hidden Markov Model [D]
- Profile HMM [D]
- Multiple sequence alignment
 - SP score [G]
 - Steiner string [G]
 - Star alignment [G]
 - Dynamic programming for SP score [G]
 - Idea of Clustal-W [B.10]
- Phylogeny [KR,D]
 - Rooted and unrooted trees
 - Character-based methods
 - Parsimony
 - Maximum likelihood
 - Distance based methods
 - Distance metric
 - Additive trees and 4-point condition
 - UPGMA and ultrametric property
 - NJ
- Protein structures
 - General idea [KR]
 - Phi and Psi angles
 - Forces driving structure formation
 - Kinds of secondary structure elements
 - Prediction of SSE [KR]
 - Prediction of tertiary structure [KR]
 - Lattice model
 - Homology modeling
 - Protein threading
 - Structure alignment [papers at end of lecture notes]
 - Dynamic programming
 - Geometric hashing
 - SCOP classification [KR]
- Microarrays [E]
 - Idea of Affymetrics chips and cDNA arrays
 - Analysis of expression levels