

BINF200 H2025 learning goals

Tom Michoel

November 6, 2025

1 Molecular biology

Know essential concepts of molecular biology necessary for understanding biological sequences and structures.

- Be able to explain the central dogma
- Be able to explain what base pairing, genes, introns, exons, codons, etc. are.
- Be able to explain the genetic code.
- Be able to construct the complementary sequence of a given sequence.
- Be able to translate a given RNA sequence into a protein sequence.

2 Global pairwise sequence alignment

Understand the Needleman-Wunsch algorithm.

- Be able to explain the principle of dynamic programming and why it can be used to maximize a pairwise alignment score.
- Be able to fill a dynamic programming table given a scoring scheme.
- Be able to backtrack in a dynamic programming table to find one or more optimal global alignments.

3 BLAST

Understand the basic steps of the BLAST algorithm.

- Be able to compile a list of n-grams for a given input sequence.
- Be able to create a look-up table of n-grams for a given input sequence.

- Be able to create a score table for matching n-grams for a given score matrix.
- Be able to create a look-up table of matching n-grams.
- Be able to search a database for sequences with matching n-grams using a score threshold.
- Be able to explain the concept of statistical significance and the E-value.

4 Multiple sequence alignment

Understand the principle and main approach of multiple sequence alignment (MSA).

- Be able to explain different MSA scoring methods (sum of pairs, entropy-based).
- Be able to explain why a dynamic programming solution exists in theory but is not practical.
- Be able to explain and apply the progressive multiple sequence alignment algorithm.

5 Phylogenetics

Understand what is phylogenetics and how phylogenetic trees are constructed.

- Be able to explain what evolutionary relationships are.
- Be able to explain the difference between physiological trait-based and molecular sequence-based phylogenetics.
- Be able to explain the major assumptions of molecular phylogenetics.
- Be able to define the root of a tree using an outgroup or midpoint rooting.
- Be able to explain the different types of trees.
- Be able to explain the different steps in tree construction.
- Be able to explain and apply the different steps of the UPGMA algorithm for constructing a tree from a matrix of pairwise distances.
- Be able to explain and apply the principle of maximum parsimony.

6 Sequence motifs

Understand what sequence motifs are, how known motifs can be detected in a sequence, and how new motifs can be discovered.

- Be able to give examples of types of regulatory sites in a genome and their biological function.
- Be able to explain what position-specific count, probability, and log-odds score matrices are.

- Be able to detect known motifs in a sequence by scoring position using a position-specific probability or log-odds score matrix.
- Be able to explain the concept of statistical significance for deciding a score cutoff.
-