BINF200 learning goals

Tom Michoel

November 5, 2025

1 Molecular biology

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

- The central dogma
- The genetic code
- Base pairing and complimentary sequences

2 Global pairwise sequence alignment

Understand and be able to apply the Needleman-Wunsch algorithm.

- Understand 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.

- Compiling a list of n-grams for a given input sequence.
- Creating a look-up table of n-grams for a given input sequence.
- Creating a score table for matching n-grams for a given score matrix.
- Creating a look-up table of matching n-grams.
- Searching a database for sequences with matching n-grams using a score threshold.
- Understand the concept of statistical significance and the E-value.

4 Multiple sequence alignment

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

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

5 Phylogenetics

Understand what is phylogenetics and how phylogenetic trees are constructed.

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