Bioinformatics Lecture 3

Presented by Douglas

Topics covered:

Applications and relevance of sequence alignment.

Alignment types: global, local and ends-free

Alignment scoring:

- matches, substitutions and indels
- gap scoring: constant, affine, convex, arbitrary

Substitution matrices, understand concepts behind PAM, BLOSUM

Needleman-Wunsch and Smith-Waterman algorithms, ends-free alignment, understand how they work and be able to perform alignments

Outcomes/Details:

You should be able to explain what sequencing alignment is used for giving one or two relevant examples either from the lectures, related content or your own experience/searches.

Know that sequence alignment is performed on both nucleotide and amino acid sequences.

Explain that sequence alignment provides two major things; the alignment itself and a score for the alignment that is proportional to the similarity, or, inversely the distance.

You should be able to name and explain in brief the four gap penalty models commonly used to score indels (insertions or deletions).

You should know that substitution (or identity) matrices are used to score the alignment of two nucleotides or amino acid residues (i.e. not for indels).

Substitution matrices may be based on basic chemical principles (e.g. similarity amino acid properties) but the most common ones are based on observed probability of closely aligned sequences.

Know that there are three main types of alignment (global/Needman-Wunsch, local/Smith-Waterman and ends-free), be able to explain where each is used and be able to use. You should be able to explain how each algorithm works using test or

pseudocode or a formal definition, whichever you prefer. Given a simple example including a gap rule and some kind of substitution matrix you should be able to demonstrate its use on two short sequences of nucleotides or amino acids.

You should be able to explain how manually aligned closely related sequences were used to build the first PAM matrix and how similar principles have been adapted since then for larger more comprehensive matrices especially BLOSSUM.

You should be able to comment on spatial and temporal complexity but we would not expect you to prove or go into details.