Bioinformatics Notes

Artem Vasenin

Last updated December 25, 2016

Contents

Chapter 1

Biology background

1.1 DNA

Structure of DNA DNA is structured as a double helix. Each helix is a sequence of nucleotides, there are four types of nucleotides: A,C,T and G. The sequences are aligned with each other such that at each position the nucleotide on one helix is paired with nucleotide on the other helix in one of the four possibilities: A-T, T-A, C-G or G-C.

Structure of proteins DNA uses a triplet code, triplet of nucleotides is called a "codon". Each codon is a code for a specific amino acid. When multiple amino acids join together they form a protein. There are 20 types of amino acids which can be used to form a protein. Some codons are used for punctuation. A gene is a sequence of codons between start and stop codons. Each gene is responsible for one protein.

DNA - Protein translation To copy DNA, it is first untwisted and unzipped inside the nucleus. Then free RNA nucleotides form

complementary base pairs with one of the DNA strands. RNA uses a U nucleotide instead of DNA's T, so they can be paired as follows: A-U, T-A, C-G and G-C; DNA on the left, RNA on the right. When multiple RNA nucleotides bond next to each other, weak hydrogen bons form between those nucleotides. mRNA transcribes amino acid code using codons. mRNA strand is then synthesised from the sequence of RNA nucleotides. mRNA strand peels off the DNA and moves out of the nucleus into the cytoplasm.

Translation takes place on the ribosome in the cytoplasm. The ribosomes are sites of protein synthesis. mRNA strand attaches to a ribosome and tRNA molecules transport specific amino acids to the ribosome. The anti-codons and codons match up to form complementary base pairs. Peptide bonds form between the adjacent amino acids to form the polypeptide, which is a protein.

Chapter 2

Sequence Alignment

We have a problem of aligning two sequences. We can do two types of alignment: "global" and "local". During alignment there are two actions we can take:

- Remove the first symbol from each sequence, thereby creating a "match" or "mismatch".

- Remove the first symbol from one of the sequences, thereby creating "deletion" or "insertion".

2.1   Local Alignment

We can represent the alignment search space as a graph. We start with a