UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Bioinformatics

2122-1-F1801Q108

Aims

The main goal of the course is to introduce the student to the new discipline that is Bioinformatics. This new research field is strongly motivated by the need to understand the mechanisms involved in biological processes in order to find the solution to computational problems arising from them. The course will provide the main algorithmic techniques and data structures that the student can use to solve computational problems related to sequence analysis or to the reconstruction of the evolutionary history. The student will achieve the ability of solving simple problems of sequence analysis, phylogenetic reconstruction and will learn how to get information from the genome databases.

Contents

Introduction to computational biology: motivations and methodologies. Sequence camparison and analysis and its relevance. Multiple alignment techniques (local and global alignment). The prediction of the gene structure and the alternative splicing prediction problem. Assembly from NGS data, de Bruijn graphs, overlap graphs and their applications.

The search of motifs and patterns in biological sequences. The pattern matching problem and data structures, such as suffix arrays, suffux trees and BWT transform and their application to pattern search in biological sequences. Applications to cancer genomics. The study of genomic variations in the population. The reconstruction of the evolutionary history. Phylogenetic trees and different methods for reconstructing trees. Haplotyping: methods based on maximum parsimony and the coalescent model.

The genome databases and the use of software tools for genome analysis.

Detailed program

1. Introduction to computational biology: motivations and methodologies. Sequence comaprison and analysis and its relevance- Multiple alignment techniques (locan and global alignment)- The prediction of the gene structure and the alternative splicing prediction problem.

2. Algorithms and data structures in bioinformatics: De Brujin graphs, overlap graphs and indexing of NGS data. Applications to the assembly from NGS data.

3. The prediction of the gene structure and the alternative splicing prediction problem. The search of motifs and patters in biological sequences. The pattern matching problem and data structures, such as suffix arrays, suffux trees and BWT tranform and their application to pattern search in biological sequences.

4. The study of genomic variations in the population.The reconstruction of the evolutionary history. Phylogenetic trees and different methods for reconstructing trees. Haplotyping: methods based on maximum parsimony and the coalescent model.

5. Metodi combinatori in Cancer genomics.

6. The genome databases and the use of software tools for genome analysis. Computational pangenomics.

Prerequisites

None

Teaching form

Lectures and practice exercises held in Italian language.

In case of restrictions related to Covid-19 emergency, lectures and practice exercises could be done by means of recorded or online video, with specific non-recorded online meetings for discussing and answering possible questions from students.

Textbook and teaching resource

(1) Slides and notes

(2) An Introduction to Bioinformatics Algorithms N.C Jones, P.A. Pevzner.

(3) Introduction to Computational molecular biology - Carlos Setubal, Joao Meidanis.

(4) Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology -Dan Gusfield.

Semester

Second semester

Assessment method

Written assignments are given during the course, concerning the different topics of the course. There is a final oral discussion of the written assignments. The final grade is determined by the evaluation of the witten assignments.

Office hours

By appointment