UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Bioinformatics

2122-1-F0901D043-F0901D087M

Aims

The main objective of the module is to introduce students to the main problems related to the computational analysis of biological sequences (DNA, RNA, proteins). Students will acquire the basic knowledge and concepts related to computational methods and techniques for collecting, managing and analyzing data in molecular biology, such as sequence data generated by Next Generation Sequencing (NGS) technologies, and will master the main computational tools necessary to extract information of interest for biomedical research from the main sequencing databases.

Contents

Introduction to bioinformatics: motivations, problems and methodologies.
Data generation: NGS technologies
Organization and management of data: the main databases; access, query, data entry
Data analysis: genome reconstruction and annotation; sequence comparison: global, local and multiple alignment algorithms; reconstruction of phylogenies; transcriptome analysis.

Detailed program

1. Data management in life sciences

2. Basics of informatics

2.1. Algorithms and programs

2.2. Alphabets, word, graphs

2.3. Databases

3. The NGS technology

3.1. Second generation NGS platforms

3.2. Third generation NGS platforms

3.3. Genomic data formats

3.4. Genome reconstruction and annotation

4. Basi di dati di sequenze molecolari

4.1. Genomic databases (EMBL – GenBank)

4.2. Protein databases (SwissProt, PDB)

4.3. Database query systems

5. Sequence Analysis in molecular biology

5.1. Exact String matching algorithms

5.2. Sequence alignments

5.2.1. Motivations

5.2.2. Dot matrices

5.2.3. Substitution matrices (PAM, BLOSUM)

5.2.4. Global alignment: Needleman-Wunsch Algorithm

5.2.5. Local alignment: Smith-Waterman Algorithm

5.2.6. Euristic Algorithms: BLAST, Fasta, BWA

5.2.7. Multiple alignment algorithms; CLUSTALW

6. Functional motifs finding in sequences

6.1. Suffix trees

6.2. Pattern discovery algorithms

7. Transcriptome Analysis

7.1. Gene Annotation and d alternative transcripts

7.2. RNA-seq data analysis

8. Molecular evolution: philogenetic trees reconstruction

8.1. Clustering algorithms

8.1.1 k-means

8.1.2 Neighbor joining

8.2. UPGMA

8.3. Maximum parsimony methods

8.4. Maximum likelihood methods

Prerequisites

Basic knowledge of computer science and molecular biology

Teaching form

Standard classes (if possible)

Textbook and teaching resource

M. Helmer Citterich, F. Ferrè, G. Pavesi, C. Romualdi, G. Pesole, Fondamenti di bioinformatica (Zanichelli editore)

Teachers notes

Semester

First semester

Assessment method

Oral exam at the end of the course, on all the topics covered in the course. The evaluation will take into account the specific knowledge of the topics, the clarity of exposition and the ability to connect the computational tools to the biological problems to be solved.

Office hours

To be defined with the student by email contact