



UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Elementi di Bioinformatica

2021-3-E3101Q116

Aims

The student will know some fundamental problems and algorithms in bioinformatics. The student will be able to write small size programs to solve some problems in bioinformatics, using also data originating from publicly available databases.

Knowledge and understanding

This course provides basic knowledge and understanding on:

- Algorithms on biological sequences
- Data structures to index biological sequencing
- Algorithms for phylogeny reconstruction
- C programming in bioinformatics
- Unix shell
- Version control
- Python programming in bioinformatics
- File Formats used in bioinformatics

Ability to apply knowledge and understanding

At the end of the course the students will be able to:

- decide which algorithms and data structures can be used to solve some problems in bioinformatics
- write efficient C programs for bioinformatics problems
- write Python programs for bioinformatics problems

Contents

Fundamental problems and algorithms in bioinformatics. Pattern matching. Sequence Alignment. DNA sequencing. Evolutionary histories. Managing biological data and databases.

Detailed program

1. Pattern matching: Algorithms and Data Structures. Karp-Rabin algorithm, Dömölki algorithm.
2. Suffix trees and suffix arrays: management, pattern matching and applications to the longest substring problem.
3. Sequence Alignment of two strings. Global alignment, local alignment, band alignment. Linear and generic gap cost. Multiple sequence alignment.
4. DNA Sequencing. Overlap graphs and de Bruijn graphs.
5. Evolutionary trees. Character-based models. Gusfield's algorithm for the perfect phylogeny. Distance-based models: ultrametrics and additive distance. UPGMA and Neighbor Joining algorithms. Max likelihood.
6. Genotypes and haplotypes. Single individual and pedigree cases.
7. Biological Data: file formats
8. Biological Databases
9. Bioinformatics open source software development methodologies
10. Linux shell
11. Git

Prerequisites

Algorithms and data structures; Programming Languages

Teaching form

Lectures and Laboratory. The individual study can use the e-learning platform to enrich the standard activities and to self assess the level of competence acquired during the course. During the emergency, lectures will be with

synchronous videoconferences.

This course is taught in Italian.

Textbook and teaching resource

The adopted textbook is "Algorithms on Strings, Trees and Sequences", by Daniel Gusfield, Cambridge Univ. Press. The library has some copies, also as [ebook](#).

The book "An Introduction to Bioinformatics Algorithms" by N. Jones, P. Pevzner is used only for some parts on phylogeny reconstruction and on genome sequencing.

The books "[Theoretical Evolutionary Genetics](#)" by J. Felsenstein and "[Population and Quantitative Genetics](#)" by Graham Coop are used for some topics on phylogeny reconstruction and on haplotypes.

Semester

First semester

Assessment method

The assessment has a written exam and a project work.

The written exam is taken individually, on the algorithms topics presented during the lectures. This part consists of open-ended questions. The written exam is 1 hour long and contains 4 questions. Of those questions, you have to answer to 3 of them.

The project work can be taken in two alternative ways.

1. a project work, mainly consisting of programming, that can be done by a single individual or a small group (max 3 students). Each student decides if they want to write Python or C code. Students decide the group composition. The specification of the project is decided by the lecturers. Each student (or each group) will receive a personalized text after the last lecture. The project will be discussed after the student has passed the written exam.
2. Solving three different exercises during the course. Such exercises have to be solved individually. Two exercises will consist of writing Python programs, while 1 exercise will be a C program. All exercises will have a strict deadline. To successfully solve an exercise, the program will have to pass some tests, possibly within a specific time limit or within a memory constraint. Who successfully completes all exercises will receive a 30/30 evaluation for the project part. No credit will be given for partial exercises (included if you

pass 2 exercises out of 3).

The final grade is obtained by weighting 50% of the degree of the written exam and 50% the project work, but you have to pass both parts.

There are no in-progress written exams.

Office hours

Please reserve a meeting via email. During the emergency, office hours will be via videoconference.

?
