UNIVERSITÀ DI PISA

# BIOINFORMATICS

## NADIA PISANTI

| | |
|---|---|
| Academic year | 2021/22 |
| Course | INFORMATICA |
| Code | 644AA |
| Credits | 6 |

| Modules | Area | Type | Hours | Teacher(s) |
|---|---|---|---|---|
| BIOINFORMATICS | INF/01 | LEZIONI | 48 | NADIA PISANTI |

### Obiettivi di apprendimento

*Conoscenze*
This course has the goal to give the student an overview of algorithmic methods that have been conceived for the analysis of genomic sequences. We will focus both on theoretical and combinatorial aspects as well as on practical issues such as whole genomes sequencing, sequences alignments, the search of patterns in biological sequences, the inference of repeated patterns and of long approximated repetitions, the computation of genomic distances, and several biologically relevant problems for the management and investigation of genomic data.

*Modalità di verifica delle conoscenze*
All students are requested to make a final oral report on an assigned topic (see below for full description of assessment methods).
Students understanding is evaluated during the classes as well as with such final oral report.
Students who did not attend the classes will be additionally asked to undergo an oral exam.
The exam (see below for its description) has the main goal of evaluating the students understanding of the problems and the methods described in the course.

*Capacità*
The students will additionally learn how a scientific paper is like and how to prepare slides for an oral presentation.

*Modalità di verifica delle capacità*
Besides evaluating student understanding, the exam is additionally meant as a chance to learn how a scientific paper is like, and how to make an oral presentation on scientific/technical topics, that is designed for a specific audience.

*Comportamenti*
The student will learn how to present a scientific result.

*Modalità di verifica dei comportamenti*
The lecturer will evaluate the student oral report.

### Prerequisiti (conoscenze iniziali)
A Basic course on algorithms

### Indicazioni metodologiche
Learning activities include:

- attending lectures and participation in discussions
- possible participation in seminars
- preparation of an oral report
- individual study
- bibliography search for the preparation of the oral report

Attendance: Advised
Teaching Methods are mainly face to face lectures. Seminars may be offered as well.

UNIVERSITÀ DI PISA

## Programma (contenuti dell'insegnamento)

A brief introduction to molecular biology: DNA, proteins, the cell, the synthesis of a protein.

Sequences Alignments: Dynamic Programming methods for local, global, and semi-local alignments. Computing the Longest Common Subsequences. Multiple Alignments.

Pattern Matching: Exact Pattern Matching: algorithms (Knuth-)Morris-Pratt, Boyer-Moore, Karp-Rabin with preprocessing of the pattern. Algorithm with preprocessing of the text: use of indexes.

Motifs Extraction: KMR Algorithm for the extracion of exact motifs and its modifications for the inference of approximate motifs.

Finding Repetitions: Algorithms for the inference of long approximate repetitions. Filters for preprocessing.

Fragment Assembly: Genomes sequencing: some history, scientific opportunities, and practical problems. Some possible approaches for the problem of assembling sequenced fragments. Link with the “Shortest common superstring” problem, the Greedy solution. Data structures for representing and searching sequencing data.

New Generation Sequencing: Applications of High Throughput Sequencing and its algorithmic problems and challenges. Investigating data types resulting from the existing biotechnologies, and the possible data structures and algorithms for their storage and analysis.

## Bibliografia e materiale didattico

Course material is available in the course web site.
The following book covers most of the topics of the syllabus:
Veli Mäkinen, Djamal Belazzougui, Fabio Cunial, Alexandru I. Tomescu:
Genome-Scale Algorithm Design: Biological Sequence Analysis in the Era of High-Throughput Sequencing. Cambridge University Press 2015, ISBN 9781107078536

## Indicazioni per non frequentanti

Students not attending the classes are additionally requested to undergo an oral exam on the topics of the Syllabus, in addition to the regular (that is, for attending students) assessment method.

## Modalità d'esame

Each student is assigned a paper that is a very recent scientific work on topics related to those of the course (tipically it is a paper accepted for publication in the proceedings of an international conference that is going to be held in a few weeks/months). The paper is part of a pool of possible papers selected by the lecturer. The paper assignment follows a brief description of all papers in the pool made by the lecturer, and a bidding phase of the students over such papers. Once the student has his/her paper assigned, the task is to prepare and make a presentation of that work that: (1) describes the results presented in that paper, (2) is suited for the actual audience (that will be the course class) as for comprehension opportunity, (3) sticks to the allowed time slot.
Students presentations usually take place all together somewhen at the end of the course. Exceptions are possible upon request for specific needs. Once the course is over, students can undergo the examination anytime during the academic year by agreeing an appointment: please, send an email to the teacher.

## Pagina web del corso

https://classroom.google.com/u/0/c/MTU5MjI1OTAxNDA5

## Altri riferimenti web

http://didawiki.di.unipi.it/doku.php/bio/start

*Ultimo aggiornamento 16/07/2021 11:53*