UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Big Data in Biotechnology & Biosciences

2122-2-F9101Q026-F9101Q026M

Aims

Few years ago, data generation was the bottleneck, now it is data mining, or extracting useful biological insights from large datasets. In particular, next generation sequencing (NGS) technologies have allowed researchers to screen changes at varying biological scales, such as genome-wide genetic variation, gene expression and small RNA abundance, epigenetic modifications, protein binding motifs, and chromosome conformation in a high-throughput and cost-efficient manner.

Biological phenomena, such as human diseases or quantitative traits in animals/plants need massive amounts and multiple types of ‘big’ data from complex studies to be described.

These data are mostly organized in public resources such as the International Nucleotide Sequence Database Collaboration (INSDC), the world's primary database of biological sequence (and related) information.

The main purpose of this course is to lead the student to understand the biological data structure in the public databases. The student will be required to integrate the biological data acquired with other data sources. The limits linked to the availability and integrity of biological data will be explored.

The two modules are highly integrated in all the contents delivered. Both lectures and workshops are delivered in co-presence by both teachers. The exam will be carried out in co-presence on a project based on the integrated contents of both modules.

Contents

- Origin of biological data
- The levels of biological variability: genes, individuals, populations, species
- Genotype, phenotype and environment and their relationship
- Type of data: focus on molecular data
- Biological database
- Practical experiences (laboratories): biological data, their management and analysis

Detailed program

- **Origin of biological data.** Biological data derives from evolutionary processes. In this section the principal processes in evolution that are generating biological data will be summarised.
- **The levels of biological variability:** genes, individuals, populations, species. Variability in biological data is shown at all the hierarchic levels that will be uncovered in this section.
- **Genotype, phenotype and environment and their relationship.** Biological complexity is summarised by the relationships among genotype, phenotype and environment from which the individuals are originated.
- **Type of data:** focus on molecular data. The main types of biological data are based on DNA or proteins. In this section we will show the different kinds.
- **Biological database:** Biological databases are libraries of biological sciences, collected from scientific experiments, published literature, high-throughput experiment technology, and computational analysis. The aim of this course is to explore these resources.
- **Practical experiences (laboratories):** biological data, their management and analysis. This is the core of this module. Following the introduction, students will be led by the teachers in lab experiences to manage biological data and integrate with other data resources.

Prerequisites

Python basic knowledges

Teaching form

Frontal lessons and lab practices.

During the Covid-19 emergency course will be held in a remote manner with video recorded lectures and with web video conferences events.

Textbook and teaching resource

Slides and scientific papers. Materials are available on the e-learning page.

Semester

Second year, second semester

Assessment method

The students, will develop on-going projects on the topics seen during the lectures. The topics of the project must be analyzed by integrating the concepts of both modules.
The projects will be exposed orally on a date agreed with the teachers.
The exam will be comprehensive for both modules.
Due to the strong interdisciplinary and the applied nature of the course, the formation of small groups will be encouraged.

The expected report will be a slides presentation with the following structure:

- The main goals of the student’s project
- Explanation of the strategies used to report the data
- The results

Office hours

Please reserve a meeting via email