



SCHEMA DELL'INSEGNAMENTO (SI)

“LABORATORIO DI BIOINFORMATICA”

/

“PRACTICAL BIOINFORMATICS”

SSD BIO/10

DENOMINAZIONE DEL CORSO DI STUDIO: LAUREA MAGISTRALE IN SCIENZE BIOLOGICHE

ANNO ACCADEMICO 2021-2022

INFORMAZIONI GENERALI - DOCENTE

DOCENTE: DOTT. BRUNO HAY MELE

TELEFONO: +39 081 679 118

EMAIL: BRUNO.HAYMELE@UNINA.IT

INFORMAZIONI GENERALI - ATTIVITÀ

ANNO DI CORSO : I / II

SEMESTRE: I

CFU: 6

INSEGNAMENTI PROPEDEUTICI

Biochimica, Biologia Molecolare, Bioinformatica.

EVENTUALI PREREQUISITI

Nessuno

OBIETTIVI FORMATIVI

L'insegnamento si propone di fornire agli studenti nozioni approfondite che li rendano autonomi e consapevoli nella pianificazione, esecuzione ed interpretazione delle analisi bioinformatiche, con particolare attenzione verso la capacità' di comunicare il risultato all'esterno del proprio contesto scientifico.

The course aims to provide students with in-depth knowledge that makes them autonomous and aware of the planning, execution, and interpretation of bioinformatics analyses, focusing on communicating the result outside their scientific context.

RISULTATI DI APPRENDIMENTO ATTESI (DESCRITTORI DI DUBLINO)

Conoscenza e capacità di comprensione

Lo studente dovrà dimostrare di aver acquisito le conoscenze e gli strumenti metodologici necessari per estrarre informazioni e testare ipotesi dai/sui dati in suo possesso. Dovrà inoltre mostrare di aver compreso in linea generale potenzialità, limiti e ambiti di applicazione dei più comuni test statistici per la valutazione del risultato bioinformatico, e di aver acquisito le conoscenze di base necessarie per un uso consapevole degli stessi.

The student must demonstrate that he/she has obtained the knowledge and methodological tools necessary to extract information and test hypotheses from/on the data in her/his possession. They must also show that they have understood the potential, limits and applicability of the most common statistical tests in bioinformatics. Finally, they must demonstrate that they have acquired the basic knowledge necessary for informed use of the same.

Capacità di applicare conoscenza e comprensione

Lo studente dovrà dimostrare di sapersi orientare in un ambiente a linea di comando e di possedere le capacità di base per l'elaborazione degli output in tale contesto. Dovrà inoltre aver acquisito familiarità con il linguaggio R in termini generali e nello specifico per le applicazioni bioinformatiche, ed essere in grado di sviluppare flussi di lavoro che comprendano l'interrogazione di due o più banche dati.

Students must demonstrate that they know how to orient themselves in a command-line environment and that they have the basic skills for processing outputs in this context. They will also have to be familiar with the R language in general terms and for bioinformatics applications. Finally, they must be able to develop workflows that include querying two or more bioinformatics data sources.

PROGRAMMA-SYLLABUS

Richiami di statistica (1 CFU)

- variabili casuali, statistica descrittiva, distribuzioni di probabilità, correlazione e causalità, regressione lineare, test delle ipotesi. Differenza tra deterministico e stocastico, introduzione al concetto di probabilità condizionata e di verosimiglianza (likelihood).

Fonti di dati (1 CFU)

- organizzazione di dati molecolari, concetto di database e databank, interazione tra databases, curated

and automatic entries, fonti primarie e secondarie/specializzate. Struttura e utilizzo delle fonti. Bioinformatic mining. I database NCBI, UCSC genome browser.

Introduzione ai servizi bioinformatici fondamentali (2 CFU)

- file formats (fasta, genbank), allineamento locale/globale (blast*) e multiplo (clustalw), annotazione (Ensembl). Visualizzazione di strutture proteiche, rudimenti di protein modelling, cenni di bioinformatica delle proteine attraverso RCSB PDB, previsione degli effetti di mutazione con Sift/Polyphen

Rudimenti di command line (1 CFU)

- concetti di base; struttura gerarchica; permessi; muoversi tra le directory, principali comandi (curl, awk, sed), grep e le espressioni regolari, pipes, redirection

Introduzione all'uso di R (1 CFU)

- operazioni di base, funzioni statistiche, estensibilità, Bioconductor, pacchetti di interfaccia alle principali banche dati.

Basics of statistics (1 CFU)

- *random variables, descriptive statistics, probability distributions, correlation and causality, linear regression, hypothesis testing. Difference between deterministic and stochastic, introduction to the concept of conditional probability and likelihood.*

Sources of data (1 CFU)

- *organization of molecular data, the concept of database and databank, interaction between databases, curated and automatic entries, primary and secondary/specialized sources. Structure and use of sources. Bioinformatic mining. NCBI databases, UCSC genome browser.*

Introduction to fundamental bioinformatics services (2 CFU)

- *file formats (fasta, genbank), local / global alignment (blast *) and multiple alignment (clustalw), annotation (Ensembl). Visualization of protein structures, overview on protein modelling, fundamentals of protein bioinformatics through RCSB PDB, prediction of mutation effects with Sift / Polyphen*

Command line fundamentals (1 CFU)

- *basic concepts; hierarchical structure; permissions; moving between directories, main commands (curl, awk,sed), grep and regular expressions, pipes, redirection*

Introduction to the R environment for statistical computing (1 CFU)

- *basic operations, statistical functions, extensibility, Bioconductor, interfacing with reference databases.*

MATERIALE DIDATTICO

- Appunti del corso; (*Course notes*)
- Articoli scientifici e documentazione associata ai software utilizzati durante il corso; (*Scientific papers, software manuals*)
- Collezione di link di riferimento; (*Reference links*)
- Pseudocodici, templates e scripts di riferimento. (*Pseudocodes, templates, reference scripts*)

MODALITÀ DI SVOLGIMENTO DELL'INSEGNAMENTO

La parte teorica e quella applicativa del corso saranno equilibrate (50:50), svolte in concerto ed in maniera interattiva: gli studenti saranno sollecitati a sviluppare la narrativa durante la lezione.
Theory and practice will be balanced (50:50), interaction-based and concerted; students will be encouraged to contribute to develop the lesson.

VERIFICA DI APPRENDIMENTO E CRITERI DI VALUTAZIONE

- A)** L'esame di fine corso mira a verificare e valutare il raggiungimento degli obiettivi didattici elencati in grassetto nella sezione contenuti del programma.
- B)** Lo studente verrà interrogato sugli argomenti del corso e verranno valutati il grado di completezza della sua risposta, il livello di integrazione tra i vari contenuti del corso, il raggiungimento da parte dello studente di una visione organica dei temi affrontati, la padronanza espressive e la proprietà nel linguaggio scientifico.
La frequenza assidua ed il grado di partecipazione attiva in aula saranno elementi di valutazione positiva.

- A)** *The final examination is aimed to verify and evaluate the achievement of the educational learning targets listed in bold in the program contents section.*
- B)** *The student will be asked to answer questions in order to evaluate the degree of completeness of the answers, the level of integration between the different topics of the course and the appropriateness of the scientific language used.*
Regular attendance to the lessons and active participation during the classroom activities will be positively considered.

a) Modalità di esame:

L'esame si articola in prova	
scritta e orale	
solo scritta	
solo orale	X
discussione di elaborato progettuale	
altro	