UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Biologia Computazionale

2627-1-F0602Q124

Obiettivi

L'insegnamento si propone di fornire agli studenti una solida preparazione teorica e pratica nell'ambito della biologia computazionale. Particolare attenzione sarà rivolta alla predizione della struttura e della funzione delle proteine a partire dalla loro sequenza amminoacidica, attraverso l'applicazione di metodi basati sia su approcci classici che su tecniche di intelligenza artificiale. Verranno introdotti e analizzati i principali database di sequenze e strutture biologiche, con un'enfasi specifica sull'integrazione e sull'interpretazione critica dei risultati generati dai diversi metodi computazionali.
Gli obiettivi generali includono:
Conoscenze e capacità di comprensione
Al termine dell'insegnamento, lo studente avrà acquisito conoscenze sui principali metodi computazionali per l'analisi di sequenze biologiche. L'apprendimento comprenderà sia la comprensione teorica degli algoritmi sottostanti, sia la capacità di applicarli concretamente. Lo studente sarà inoltre guidato a interpretare i dati computazionali all'interno di un contesto biologico, sviluppando la capacità di formulare ipotesi coerenti e sperimentabili.

Capacità di applicare conoscenze e comprensione
Lo studente sarà in grado di progettare e implementare in autonomia una pipeline di analisi bioinformatica per la predizione strutturale e funzionale di una proteina di funzione ignota.

Autonomia di giudizio
Lo studente sarà stimolato a valutare in modo critico l'affidabilità delle predizioni computazionali, sviluppando la capacità di riconoscere i principali limiti dei metodi utilizzati. Sarà inoltre incoraggiato a integrare diverse fonti di informazione, come dati strutturali e funzionali, al fine di ottenere un'interpretazione biologica coerente, completa e scientificamente fondata.

Abilità comunicative
Al termine dell'insegnamento, lo studente padroneggerà il lessico appropriato per descrivere con precisione un'intera analisi computazionale, illustrandone in modo chiaro obiettivi, metodi, risultati e limiti.

Capacità di apprendimento
Lo studente sarà in grado di affrontare problemi biologici nuovi, non trattati durante le esercitazioni, applicando in modo autonomo le competenze acquisite e utilizzando con flessibilità i metodi appresi durante l'insegnamento.

Contenuti sintetici

1. Database biologici: sequenze, strutture e strategie di interrogazione
2. Allineamento di sequenze
3. Alberi filogenetici
4. Predizione funzionale: dalla sequenza amminoacidica alla funzione
5. Predizione strutturale: dalla sequenza amminoacidica alla struttura
6. Metodi computazionali avanzati per lo studio delle interazioni proteina-ligando
7. Applicazione di metodi computazionali per predire la struttura e la funzione di una proteina ignota.

Programma esteso

1. Database biologici: sequenze, strutture e strategie di interrogazione
Strategie di interrogazione e recupero dati dalle principali risorse bioinformatiche.
Esplorazione dei database di sequenza (GenBank, UniProt) e di metagenomica (MGnify).
Esplorazione dei database strutturali (PDB, AlphaFold DB) e comprensione dei formati standard (PDB, mmCIF).
Esplorazione dei database basati su famiglie, domini e classificazione strutturale (Pfam, PROSITE, CATH).

2. Allineamento di sequenze
Principi e applicazione delle matrici di punteggio per l'allineamento (PAM, BLOSUM, Position Specific Scoring Matrix).
Tecniche di allineamento a coppie (locale) e multiplo per l'identificazione di regioni conservate.
Ricerca per similarità in larga scala tramite BLAST: interpretazione critica dei risultati (E-value, bit-score).

3. Ricostruire la Storia Evolutiva: Filogenesi Molecolare
Metodi per la costruzione di alberi filogenetici (Neighbor-Joining, Maximum Likelihood).
Interpretazione di un albero: lettura della topologia, valutazione del supporto statistico (bootstrap) e identificazione di eventi evolutivi.
Principi di ricostruzione di sequenze ancestrali per inferire le proprietà delle proteine del passato.

4. Principi di Machine Learning
Introduzione al Machine Learning: preparazione dei dati, training, validazione, testing e principali paradigmi di apprendimento (supervisionato, non supervisionato e cenni di reinforcement learning).
Limiti dei modelli classici e introduzione al Deep Learning.
Architettura del neurone artificiale e delle reti neurali profonde (Deep Neural Networks, DNN).
Il meccanismo di attenzione (self-attention) e i modelli transformer nelle applicazioni di biologia computazionale.

5. Predizione funzionale: dalla sequenza alla funzione
Metodi di annotazione funzionale basati su omologia e analisi filogenetica.
Strategie e limiti dell'annotazione funzionale automatica.
Applicazione dei Modelli Nascosti di Markov (HMM) per il riconoscimento di motivi funzionali e l'assegnazione a famiglie proteiche.
Predizioni di modifiche post-traduzionali

6. Predizione strutturale: dalla sequenza alla struttura tridimensionale

Predizione delle strutture secondarie e delle regioni intrinsecamente disordinate (IDRs).
Metodi tradizionali di predizione: approcci basati su homology modelling, fold-recognition/threading e metodi ab initio.
Valutazione critica della qualità dei modelli 3D: analisi del pLDDT, punteggi QMEAN e utilizzo del grafico di Ramachandran.
Principi di allineamento strutturale per il confronto tra proteine.

7. Metodi computazionali per lo studio della dinamica e delle interazioni proteina-ligando
Predizione di complessi biomolecolari mediante approcci di ultima generazione (AlphaFold 3 e Boltz2).
Introduzione al docking molecolare
Introduzione alla dinamica molecolare

8. Laboratorio: predizione della struttura e della funzione di una proteina ignota
Analisi di sequenza. Esplorazione dei principali database di sequenze e strutture proteiche, ricerca di omologhi mediante BLAST, allineamento multiplo e identificazione di domini conservati tramite InterPro.
Predizione strutturale. Generazione di modelli tridimensionali della proteina target e dei suoi complessi biomolecolari mediante AlphaFold 3.
Validazione dei modelli. Valutazione della qualità strutturale mediante analisi del pLDDT e grafici di Ramachandran.
Analisi delle interazioni. Identificazione dei residui chiave coinvolti nelle interazioni proteina-DNA o proteina-ligando.
Visualizzazione molecolare. Utilizzo di software dedicati (ChimeraX) per l'analisi e la rappresentazione tridimensionale di biomolecole.

Prerequisiti

Basi di biochimica, biochimica delle proteine e biologia molecolare

Modalità didattica

L'insegnamento è strutturato in due moduli integrati per un totale di 45 ore (6 CFU), progettati per combinare basi teoriche con attività pratica.

1. Modulo teorico (18 lezioni in aula, 35 ore - 5 CFU). La didattica integra sessioni frontali, supportate da slide e video, con sessioni interattive, in cui gli studenti saranno guidati nell'analisi di dati computazionali, applicando da subito i concetti appresi.

2. Modulo Pratico (5 lezioni in laboratorio informatico, a frequenza obbligatoria, 10 ore - 1 CFU). Il laboratorio prevederà una fase introduttiva con slide e tutorial sull'uso degli strumenti software. Successivamente, gli studenti applicheranno queste competenze in modo pratico per estrarre dati e si confronteranno in discussioni interattive per interpretare i risultati ottenuti.

Materiale didattico

Slides e videolezioni: reperibili sulla pagina Moodle dell’insegnamento (http://elearning.unimib.it/).

Testi consigliati:
M.H. Citterich et al. “Fondamenti di Bioinformatica” Zanichelli
Pascarella e Paiardini “Bioinformatica, dalla sequenza alla struttura delle proteine” Zanichelli

Periodo di erogazione dell'insegnamento

Secondo semestre

Modalità di verifica del profitto e valutazione

Prova orale. La prova orale verterà su:
- Presentazione con diapositive PowerPoint dei risultati ottenuti durante il modulo di laboratorio. La valutazione considera il grado di comprensione degli strumenti computazionali utilizzati, l’interpretazione critica dei risultati ottenuti e l’uso del linguaggio appropriato.
- Colloquio sui contenuti del modulo teorico. La valutazione si baserà sulla conoscenza degli argomenti trattati, sulla capacità di stabilire connessioni tra i contenuti teorici e gli strumenti computazionali applicati nel laboratorio, sull’uso di un linguaggio adeguato e sulla chiarezza espositiva.

Non sono previste prove in itinere

Orario di ricevimento

Su appuntamento previa e-mail al docente

Sustainable Development Goals

SALUTE E BENESSERE | ISTRUZIONE DI QUALITÁ | CONSUMO E PRODUZIONE RESPONSABILI