UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Bioinformatic Methodologies

2122-1-F0802Q054

Obiettivi

L'obiettivo principale del corso è quello di introdurre gli studenti alle principali problematiche relative all'analisi computazionale di sequenze biologiche (DNA, RNA, proteine) e alla costruzione e simulazione di modelli matematici predittivi per la biologia. Gli studenti acquisiranno le conoscenze ed i concetti di base relativi alle metodologie e alle tecniche computazionali per la raccolta, la gestione e l'analisi di dati in biologia molecolare, e la padronanza dei principali strumenti computazionali necessari per estrarre informazioni di interesse per la ricerca biomedica dalle principali banche dati di sequenze.

Contenuti sintetici

1) Introduzione alla bioinformatica: motivazioni, problemi e metodologie

2) Allineamenti tra sequenze: allineamento globale, locale e multiplo

3) Organizzazione e gestione dei dati: le principali basi di dati; accesso, interrogazione, inserimento dati

4) Ricostruzione di alberi filogenetici

5) Studio del genoma: mappatura, sequenziamento, annotazione, ricostruzione

6) Reti di trascrizione: grafi e modelli computazionali; analisi del trascrittoma

Programma esteso

1) Elementi di Informatica: architettura di un computer, sistemi operativi, algoritmi, linguaggi di programmazione e basi di dati

2) Allineamento tra sequenze. Allineamento globale: similarità e omologia; matrice di punti; programmazione dinamica, traceback. Allineamento locale: matrici di sostituzione (PAM e BLOSUM); algoritmo di Smith & Waterman; ricerca euristica su database (BLAST, FASTA). Allineamenti multipli: allineamento progressivo (Clustal)

3) Basi di dati di sequenze molecolari. Basi di dati di sequenze genomiche (EMBL – GenBank) e proteiche (SwissProt, PDB). Sistemi di interrogazione delle basi di dati (NCBI, UCSC Genome Browser). Piattaforme di ricerca e analisi bioinformatiche (Galaxy).

4) Ricostruzione di alberi filogenetici. Metodi basati sulla distanza: clustering (UPGMUA e Neighbor Joining). Metodi “character-based” (massima parsimonia, massima verosimiglianza)

5) Studio del genoma. Mappatura. Sequenziamento: Sanger, Next Generation Sequencing (NGS), base calling. Annotazione: FASTA, FASTQ. Ricostruzione

6) Reti transcrizione. Analisi del trascrittoma: microarray, RNA-sequencing. Teoria dei grafi come strumento per la descrizione di una rete di trascrizione: distribuzione dei nodi in ingresso e in uscita. Motivi di rete. Modelli computazionali di regolazione dell’espressione genica: il caso dell’autoregolazione negativa

Prerequisiti

Nessuno: tutti i concetti necessari di informatica e matematica sono forniti nel corso

Modalità didattica

Lezioni frontali con slides.

Esercitazioni in aula di calcolo per l'utilizzo di banche dati biologiche e piattaforme di analisi bioinformatiche.

Materiale didattico

Slides delle lezioni e delle esercitazioni reperibili sulla pagina e-learning dell'insegnamento.

Periodo di erogazione dell'insegnamento

Primo semestre

Modalità di verifica del profitto e valutazione

Lo studente verrà valutato sulla presentazione di un articolo scientifico precedentemente assegnatogli + domande orale su tutto il programma svolto a lezione

Orario di ricevimento

Gli studenti sono invitati a contattare il docente per email per accordarsi su data e giorno (eventualmente via WebEx)