UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Bioinformatics

2122-1-F0901D043-F0901D087M

Obiettivi

L'obiettivo principale del modulo è quello di introdurre gli studenti alle principali problematiche relative all'analisi computazionale di sequenze biologiche (DNA, RNA, proteine). 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, come i dati di sequenze generati dalle tecnologie Next Generation Sequencing (NGS), 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

Introduzione alla bioinformatica: motivazioni, problemi e metodologie.

Generazione dei dati: le tecnologie NGS

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

Analisi dei dati: ricostruzione e annotazione di genomi; confronto di sequenze: algoritmi di allineamento globale, locale e multiplo; ricostruzione di filogenie; analisi del trascrittoma.

Programma esteso

1. La gestione dei dati nelle scienze della vita

2. L’informatica essenziale

2.1. Algoritmi e programmi

2.2. Alfabeti, parole, grafi

2.3. Basi di dati

3. La tecnologia NGS

3.1. Piattaforme NGS di seconda generazione

3.2. Piattaforme NGS di terza generazione

3.3. Il formato dei dati genomici

3.4. Ricostruzione e annotazione di genomi

4. Basi di dati di sequenze molecolari

4.1. Basi di dati Genomiche (EMBL – GenBank)

4.2. Basi di dati di sequenze proteiche (SwissProt, PDB)

4.3. I sistemi di interrogazione delle Basi di Dati

5. Analisi di sequenze in biologia molecolare

5.1. Algoritmi di String matching esatto

5.2. Allineamento di sequenze

5.2.1. Motivazioni

5.2.2. Matrici a punti

5.2.3. Matrici di sostituzione PAM, BLOSUM

5.2.4. Allineamento globale: Algoritmo di Needleman-Wunsch

5.2.5. Allineamento locale: Algoritmo di Smith-Waterman

5.2.6. Algoritmi euristici: BLAST, Fasta, BWA

5.2.7. Allineamento multiplo; CLUSTALW

6. Ricerca di motivi funzionali in sequenze

6.1. Alberi di suffissi

6.2. Algoritmi di pattern discovery

7. Analisi del trascrittoma

7.1. Annotazione di geni e trascritti alternativi

7.2. Analisi di dati RNA-seq

8. Evoluzione molecolare: ricostruzione di alberi filogenetici

8.1. Algoritmi di Clustering

8.1.1. k-means

8.1.2. Neighbor joining

8.1.3. UPGMA

8.1.4. Metodi di massima parsimonia

8.1.5. Metodi di massima verosimiglianza

Prerequisiti

Conoscenze di base di informatica e di biologia molecolare

Modalità didattica

Lezioni in presenza (se possibile)

Materiale didattico

M. Helmer Citterich, F. Ferrè, G. Pavesi, C. Romualdi, G. Pesole, Fondamenti di bioinformatica (Zanichelli editore)

Dispense fornite dal docente

Periodo di erogazione dell'insegnamento

Primo semestre

Modalità di verifica del profitto e valutazione

Non sono previste verifiche intermedie. La verifica finale consiste in una prova orale, che verrà valutata tenendo conto della chiarezza espositiva e della completezza delle risposte a tre/quattro domande relative ai contenuti dell'insegnamento.

Orario di ricevimento

Da definire con lo studente via email