UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Bioinformatics

2425-1-F1801Q108

Obiettivi

Il corso si propone di introdurre lo studente ad una recente disciplina, la Bioinformatica. Questa disciplina nasce dalla necessità di comprendere, dal punto di vista dell'Informatica, i meccanismi che regolano le computazioni nei processi biologici per poi sviluppare adeguati strumenti computazionali per la soluzione di molteplici problemi ad essi legati.

L'obiettivo principale del corso è quello di fornire allo studente la padronanza delle tecniche algoritmiche e delle strutture dati per poter affrontare lo studio e la soluzione di problemi computazionali di analisi e confronto di genomi multipli anche considerando gli aspetti di genomica comparativa mediante la costruzione di alberi filogenetici.

In particolare verrano forniti allo studente gli strumenti per:

- progettare la soluzione algoritmica di problemi di analisi di genomi multipli o di confronto e ricostruzione di alberi filogenetici.

- modellare la soluzione di problemi biologici su sequenze genomiche mediante la formulazione di problemi combinatori.

- utilizzare le banche dati genomiche per estrarre le informazioni di interesse sul genoma umano.

Contenuti sintetici

Introduzione alla biologia computazionale: motivazioni e metodologie.

L'importanza del confronto e dell'analisi di sequenze biologiche. Tecniche di allineamento di sequenze (allineamento globale e locale e allineamento multiplo). Algoritmi per l'allineamento di sequenze nella predizione della struttura di un gene (splicing alternativo). Assemblaggio di dati di sequenziamento NGS con grafi di de Bruijn

compressi. Grafi di de Bruijn e grafi di overlap e loro applicazioni.

La ricerca di motivi in sequenze biologiche. Il problema generale del matching esatto. Gli alberi suffisso, i suffix array e la trasformata di Burrows-Wheeler e la loro applicazione nella ricerca di pattern nelle sequenze biologiche.

Lo studio delle variazioni (mutazioni) geniche nella popolazione. Alberi evoluzionari. Ricostruzione della storia evolutiva di specie con alberi evoluzionari: metodi principali. Applicazione alla genomica tumorale. L'aplotipizzazione di individui: metodi combinatori basati sul modello coalescente e il criterio di massima parsimonia.

Internet e la post-genomica. Le banche dati e il software per l'analisi del genoma.

Programma esteso

1. L'importanza del confronto e dell'analisi di sequenze biologiche. Tecniche di allineamento di sequenze (allineamento globale e locale e multiplo). Algoritmi per l'allineamento di sequenze nella predizione della struttura di un gene (splicing alternativo).

2. Strutture dati e algoritmi su grafi in bioinformatica: grafi di de Bruijn, grafi di overlap e indicizzazione. Applicazioni al problema del sequenziamento di dati NGS e assemblaggio dati.

3. La ricerca di motivi in sequenze biologiche. Il problema generale del matching esatto. Gli alberi suffisso, i suffix array e la trasformata di Burrows-Wheeler e la loro applicazione nella ricerca di pattern nelle sequenze biologiche.

4. Lo studio delle variazioni (mutazioni) geniche nella popolazione. Alberi evoluzionari. Ricostruzione della storia evolutiva di specie con alberi evoluzionari: metodi principali. Assemblaggio di aplotipi: metodi combinatori.

5. Metodi combinatori in Cancer genomics.

6. Internet e la post-genomica. Le banche dati e il software per l'analisi del genoma. La pangenomica computazionale.

Prerequisiti

Nesssuno

Modalità didattica

Tutte le ore di lezione si compongono di una prima parte svolta in modalità erogativa (erogazione dei contenuti) e una seconda parte invece svolta in modalità interattiva.

Materiale didattico

(1) Slides e dispense

(2) An Introduction to Bioinformatics Algorithms N.C Jones, P.A. Pevzner.

(3) Introduction to Computational molecular biology - Carlos Setubal, Joao Meidanis.

(4) Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology -Dan Gusfield.

Periodo di erogazione dell'insegnamento

Secondo semestre

Modalità di verifica del profitto e valutazione

In itinere viene assegnata una prova intermedia da svolgere relativa alle diverse parti del corso. E' prevista poi una prova scritta finale. Il voto finale è determinato dalla valutazione della prova intermedia e della prova scritta,

Orario di ricevimento

Per appuntamento

Sustainable Development Goals

SALUTE E BENESSERE