



UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Bioinformatica

2627-1-F1802Q114

Obiettivi

Conoscenze e comprensione

Il corso si propone di introdurre lo studente alla Bioinformatica. Questa disciplina nasce dalla necessità di sviluppare adeguati strumenti computazionali per la soluzione di molteplici problemi che sono ispirati dalla biologia. Tali strumenti trovano comunque applicazione anche in altri ambiti caratterizzati da elementi comuni alla Bioinformatica, quali il trattamento di dati massivi principalmente su sequenze.

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, il corso si occupa di genomica informatica e aspetti di pangenomica computazionale.

Capacità di applicare conoscenza e comprensione

In particolare verranno 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,
- progettare strutture dati adeguate a trattare dati massivi di sequenze,
- modellare la soluzione di problemi su sequenze genomiche mediante la formulazione di problemi combinatori,
- utilizzare le banche dati genomiche e il software disponibile per estrarre le informazioni di interesse dal genoma umano, quali le variazioni che caratterizzano una popolazione di individui.

Autonomia di giudizio

Lo studente sarà in grado di comprendere la rilevanza di un metodo computazionale dal punto di vista delle richieste di risorse spazio e tempo.

Dovrà essere in grado di comprendere se una data struttura dati è adeguata per la soluzione di problemi su sequenze genomiche, in funzione della dimensione del dato

Capacità comunicativa

Lo studente acquisirà il linguaggio e formalismo tipo della disciplina informatica

Capacità di apprendere

Lo studente deve essere in grado di leggere e capire autonomamente un articolo di bioinformatica che tratta le tematiche affrontate nel corso

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 evolutivi. Ricostruzione della storia evolutiva di specie con alberi evolutivi: 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 evolutivi. Ricostruzione della storia evolutiva di specie con alberi evolutivi: 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

Nessuno

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. Durante il corso verranno presentati degli articoli su temi affrontati nel corso e lo studente dovrà acquisire la capacità di comprenderne il contenuto

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 vengono assegnate delle prove intermedie da svolgere relativa alle diverse parti del corso. E' prevista poi una prova scritta finale (opzionale sulla base dell'esito delle prove in itinere). Il voto finale è determinato dalla valutazione della prova intermedia e della prova scritta finale (opzionale).

Orario di ricevimento

Per appuntamento

Sustainable Development Goals

SALUTE E BENESSERE
