UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Big Data in Biotechnology & Biosciences

2627-2-FDS02Q025-FDS02Q02501

Obiettivi

Fino a pochi anni fa, il collo di bottiglia per le analisi quantitative in Biologia era la generazione del dato, oggigiorno è l'estrazione della conoscienza biologica da grandi dataset. In particolare, l'avvento delle tecnologie Next Generation Sequencing (NGS) ha permesso ai ricercatori di osservare le variazioni a diverse scale biologiche, come le variazioni sull'intero genome, l'espressione dei singoli geni e l'abbondanze di piccoli frammenti di RNA, le modifiche epigenetiche, i motivi per il protein binding o le variazioni di conformazione nei cromosomi. Il tutto in maniera efficente e economicamente accessibile.

Fenomeni biologici, come malattie che affliggono l'uomo o altre caratteristiche di animali o piante descrivibili in maniera quantitativa, per essere compresi richiedono un enorme e variegato quantitativo di dati provenienti da studi complessi. Questi dati vengono organizzati e stoccati in banche dati pubbliche come la International Nucleotide Sequence Database Collaboration (INSDC), il database primario al mondo per le informazioni sulle sequenze biologiche.

Obiettivo principale di questo corso è quello di portare lo studente a comprendere la strutture dei dati biologici presenti in queste banche dati. Allo studente verrà richiesto di integrare queste informazioni molecolari con altre sorgenti di dati. Si esploreranno anche i limiti legati alla disponibilità ed all'integrità dei dati molecolari.

I due moduli sono fortemente integrati in tutta la fruizione del corso. Sia le lezioni che le esercitazioni verranno svolte in compresenza dei docenti. L'esame sarà contestuale per i due moduli e verrà svolto tramite un progetto basato sull'integrazione del contenuto dei due moduli.

Contenuti sintetici

- Origine del dato biologico.
- I livelli della variabilità biologica.
- Genotipo, fenotipo, ambiente e le loro relazioni.

- Tipi di dato: focus sul dato molecolare.
- Le basi di dati biologiche.
- Lab hands-on.

Programma esteso

- **Origine del dato biologico.** I dati biologici derivano da processi evolutivi. In questa sezione si richiameranno i principali processi evolutivi alla base della generazione dei dati biologici.
- **I livelli della variabilità biologica:** geni, individui, popolazioni e specie. In questa sezione si metterà in luce la variabilità nei dati biologici presente a tutti i livelli gerarchici.
- **Genotipo, fenotipo, ambiente e le loro relazioni.** La complessità biologica è riassunta dalle relazioni tra genotipo, fenotipo ed ambiente da cui gli individui hanno origine.
- **Tipi di dato**: focus sul dato molecolare. I principali dati biologici si basano sul DNA o le proteine. In questa sezione verranno analizzati i diversi tipi.
- **Le basi di dati biologiche:** le basi di dati biologiche sono collezioni delle scienze della vita, raccolte tramite esperimenti scientifici, letteratura scientifica, esperimenti con tecnologie high-throughput ed analisi computazioniali. Obiettivo di questo corso è la loro esplorazione.
- **Lab hands-on:** i dati biologici, la loro gestione ed analisi. Questa è la parte essenziale del modulo. Dopo la parte introduttiva gli studenti verranno guidati in esperienze di laboratorio in cui verranno maneggiati principalmente dati biologici e la loro integrazinoe con altre risorse.

Prerequisiti

Conoscenza di base di Python

Modalità didattica

Lezioni e sessioni in laboratorio informatico

Materiale didattico

Lucidi mostrati a lezione e articoli scientifici. Il materiale è a disposizione sulla pagina e-learning.

Periodo di erogazione dell'insegnamento

Secondo anno, secondo semestre

Modalità di verifica del profitto e valutazione

Gli studenti svilupperanno progetti sugli argomenti visti durante le lezioni. Gli argomenti del progetto devono essere analizzati integrando i concetti di entrambi i moduli.
I progetti saranno esposti oralmente in una data concordata con gli insegnanti.
L'esame sarà comprensivo per entrambi i moduli.
A causa della forte interdisciplinarietà e della natura applicata del corso, sarà incoraggiata la formazione di piccoli gruppo

La presentazione a conclusione dei progetti avverà tramite slides e dovrà avere la seguente struttura:

- Obiettivi principali del progetto
- Discussione delle strategie utilizzate
- Risultati

Orario di ricevimento

Per appuntamento tramite e-mail

Sustainable Development Goals

ISTRUZIONE DI QUALITÁ