



# UNIVERSITÀ DI PISA

---

## BIOINFORMATICA

**ROBERTO MARANGONI**

Academic year 2021/22  
Course BIOLOGIA MOLECOLARE E CELLULARE  
Code 178EE  
Credits 3

| Modules        | Area   | Type    | Hours | Teacher(s)        |
|----------------|--------|---------|-------|-------------------|
| BIOINFORMATICA | INF/01 | LEZIONI | 32    | ROBERTO MARANGONI |

### Obiettivi di apprendimento

#### *Conoscenze*

Il corso è finalizzato a fornire i principali concetti di Bioinformatica, con particolare riferimento alla bioinformatica per l'analisi di sequenze biologiche. Il corso è di tipo elementare e non richiede specifici prerequisiti, in particolare non di tipo informatico.

#### *Modalità di verifica delle conoscenze*

- esercizi durante il corso
- simulazione di esami finali
- prove a risposta multipla

#### *Capacità*

Uso dei principali applicativi per ricerche in banche dati, allineamenti di sequenze e problemi affini. Capacità di capire il correlato informatico dei problemi dell'elaborazione di dati biologici.

#### *Modalità di verifica delle capacità*

- esercizi su problemi-tipo

#### *Comportamenti*

- familiarità con l'uso di strumenti informatici
- capacità di capire i limiti e le corrette modalità d'impiego degli strumenti informatici

#### *Modalità di verifica dei comportamenti*

- capacità di risolvere autonomamente i problemi-tipo proposti

#### Prerequisiti (conoscenze iniziali)

Nessun prerequisito specifico, a parte una generale formazione in biologia e biologia molecolare e nozioni veramente basilari di statistica

#### Indicazioni metodologiche

Lezioni frontali con esercitazioni al computer. Per il corrente a.a., dati i problemi che la pandemia causa per l'accesso alle aule informatizzate, viene richiesto agli studenti di portare a lezione il proprio computer portatile o un tablet.

#### Programma (contenuti dell'insegnamento)

##### **Introduzione**

Cenni di teoria dell'informazione e di teoria algoritmica dell'informazione. Informazione, entropia e probabilità. Entropia condizionata. Complessità. Codici e canali. Problemi sulle stringhe. Algoritmi per il confronto e l'allineamento di stringhe.

##### **Parte I: la bioinformatica orientata alle sequenze**



## UNIVERSITÀ DI PISA

---

### Dati biologici e loro manipolazione

Basi di dati di sequenze di macromolecole biologiche. Problemi relativi all'archiviazione e alla ricerca di sequenze di macromolecole. Le banche dati esistenti: struttura dei record e strategie di interrogazione. Concetto di "database annotato". Algoritmi per sequenze biologiche. Algoritmi per la ricerca di somiglianze tra sequenze. Algoritmi per l'allineamento tra sequenze. BLAST, FASTA.

### Analisi comparativa ed evolutiva di sequenze biologiche

Multiallineamento. Costruzione di alberi filogenetici *rooted* e *unrooted*. PHYLIP. Problemi relativi alle stime temporali su alberi filogenetici.

### Analisi di genomi completi

Ortologie e paralogie. Autosomiglianze. Disegno di primers ottimali per PCR. Risultati dell'analisi statistica di alcuni dei genomi completi noti. Ipotesi delle duplicazioni ancestrali. Database di genomi completi e strumenti di pubblico dominio per l'analisi di genomi completi. Predizione di sequenze trascriventi e non trascriventi. Predizione di ORI, ORF, introni/esoni. Algoritmi per la ricerca e la predizione di TFBS. Analisi della variazione della complessità lungo sequenze genomiche. Cenni a processi markoviani e matrici di transizione. Gli Hidden Markov Models. I database genomici: Genome Browser@UCSC e Ensembl. Tipologia dei dati contenuti. Tracce e tabelle. Custom tracks e file BED. BLAT.

### Parte II: bioinformatica orientata alle strutture

Cenni alle tecniche sperimentali per la misurazione di strutture: diffrazione raggi X, NMR, crio-microscopia. Database strutturali: PDB e relativi tipi di dati. Profili di idropatia di Kyte e Doolittle. Metodi statistici di Chou e Fasman. Cenni alle reti neurali artificiali. Meccanismi di predizione basati su reti neurali. Il caso AlphaFold.

### Parte III: la bioinformatica orientata alle funzioni

Reti di controllo genico e *metabolic pathways*. Descrizione del flusso informazionale all'interno di una cellula. Ruolo dei sistemi formali. Database di reti metaboliche: KEGG/Pathway, struttura dei dati. I proteomi. Database di proteomi.

### Parte IV: esercitazioni al computer

Struttura di un computer generico. La struttura di Internet: ambienti e applicativi su rete. Interrogazioni ai database. Uso guidato dei vari software illustrati nel corso. Applicazione pratica di ciascuna nozione teorica illustrata.

### Bibliografia e materiale didattico

- Lesk, "Bioinformatica", McGraw-Hill.
- Pascarella e Paiardini, "Bioinformatica", Zanichelli
- Helmer-Citterich et al, "Fondamenti di Bioinformatica", Zanichelli

Materiale fornito dal docente

### Modalità d'esame

Test a risposta multipla, eventualmente integrato di colloquio orale. (se in remoto: colloquio orale)

Ultimo aggiornamento 23/08/2021 11:44