

Data Science Lab in Biosciences

UniShare

Davide Cozzi
@dlcgold

Indice

1	Introduzione	2
2	Big Data in Biotechnology & Biosciences	3
3	Making Sense of Biological Data	4

Capitolo 1

Introduzione

Questi appunti sono presi a lezione. Per quanto sia stata fatta una revisione è altamente probabile (praticamente certo) che possano contenere errori, sia di stampa che di vero e proprio contenuto. Per eventuali proposte di correzione effettuare una pull request. Link: <https://github.com/dlccgold/Appunti>.

Capitolo 2

Big Data in Biotechnology & Biosciences

Capitolo 3

Making Sense of Biological Data

La **biologia** è una *scienza della vita* che include:

- zoologia
- citologia
- ecologia
- botanica
- microbiologia
- fisiologia
- genetica
- biochimica
- ...

Tale scienza si occupa di capire le strutture, le funzioni, le origini, le interazioni e la tassonomia (ovvero della classificazione) delle creature viventi, non solo dell'uomo.

Si hanno vari livelli d'informazione con tanti approcci per lo studio delle informazioni stesse. È come se avessi un enorme *grafo multilayer* di informazioni, molto *complesso*. In questa scienza non si hanno regole universali, si ha sempre l'eccezione. La validità dei metodi analitici dipende fortemente da caratteristiche, conosciute o meno, dei dati stessi. Alcuni algoritmi hanno sunti diversi a seconda della tipologia del dato biologico. I dati sono tendenzialmente poco conosciuti. La *biologia* prevede quindi studi davvero molto

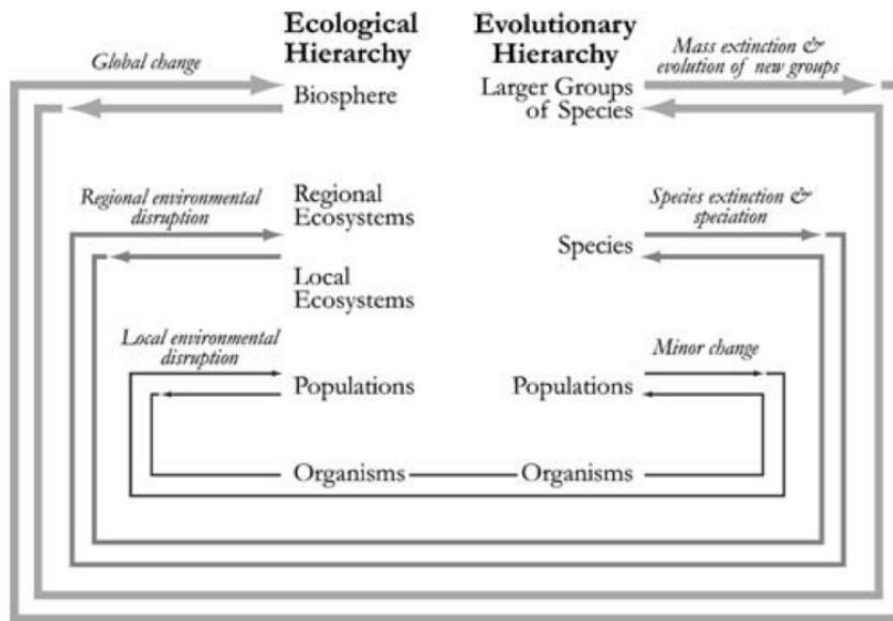


Figura 3.1: Vecchio schema grafico delle gerarchie in biologia e delle loro interazioni.

complicati, più di quanto sembri. Inoltre alcuni fenomeni sono difficilissimi da misurare e hanno una grande incertezza.

Si ha una *struttura gerarchica evolutiva della biologia* ma solo perché l'uomo ha strutturato le conoscenze sulla tassonomia. In realtà la vita è meno schematica e più “fluida” di quel che si vuole credere. Si hanno quindi:

- raggruppamento di specie simili
- specie
- popolazione
- organismi

Un'altra gerarchia si basa sui fenomeni che interconnettono i vari livelli appena elencati. Si hanno quindi i cosiddetti **ecosistemi**, avendo una *gerarchia ecologica*. Per l'uomo si ha una *reale*, ovvero dove vive un organismo, praticamente globale, vivendo praticamente ovunque. Si hanno quindi interazioni tra queste varie gerarchie nella biologia, che descrive sia le caratteristiche evolutive che ambientali, come visibile in figura 3.1.

La biologia inoltre evolve molto in fretta anche nel definire i suoi formalismi e i suoi limiti. Le pubblicazioni “invecchiano” molto in fretta.

In biologia si hanno delle *regole*, influenzate comunque dalla percezione umana che crea dei *bias descrittivi* che si ripercuotono sui dati stessi. Anche in questo le nuove tecnologie di sequenziamento permettono di avere dati migliori, in quanto permettono di osservare il fenomeno in modo corretto.