

Riassunto dell'elaborato finale

Nome e cognome del candidato: Matteo Tortoli – Matricola: 5604520 – Email: matteo.tortoli@stud.unifi.it

Titolo dell'elaborato: *Applicazioni dell'algoritmica alla biologia: alberi evolutivi*.

Nome e cognome della relatrice: Maria Cecilia Verri – Email: mariacecilia.verri@unifi.it

Con la presente tesi si intende fornire una panoramica sulla bioinformatica, scienza emergente negli ultimi anni. L'obiettivo è quello di mostrare le applicazioni degli algoritmi alla biologia, in particolar modo si illustrerà lo studio e la costruzione degli *alberi evolutivi*, argomento principale di questa tesi.

La bioinformatica è quella scienza multidisciplinare che coinvolge la biologia, l'informatica, la genetica e la statistica. Essa è volta a risolvere i problemi di natura biologica attraverso algoritmi e strumenti computazionali.

La struttura dei capitoli è stata pensata per fornire le nozioni base della biologia, per poi passare al nucleo della tesi, ovvero gli algoritmi per la costruzione degli alberi evolutivi (o filogenetici).

La struttura del lavoro complessivo è mostrata di seguito.

1. **Capitolo Uno:** informazioni biologiche di base. Classificazione delle forme di vita e descrizione del DNA, RNA e proteine. Tutte queste informazioni sono necessarie per poter comprendere i concetti spiegati nei capitoli successivi.
2. **Capitolo Due:** introduzione alla bioinformatica. Storia della bioinformatica e principali aree di ricerca, in particolar modo il focus è sulla filogenetica, che è quella branca della scienza che si occupa degli alberi evolutivi.
3. **Capitolo Tre:** albero evolutivo. Suddivisione in albero radicato e non radicato: il primo mostra le evoluzioni tra le entità biologiche, mentre il secondo illustra come queste entità siano legate tra loro.
4. **Capitolo Quattro:** Nucleo principale della tesi. Si parla del "*problema della costruzione degli alberi evolutivi*". Vengono illustrati i principali algoritmi che risolvono tale problema. In particolar modo i primi sono algoritmi "di base" ma presentano delle criticità, mentre gli ultimi sono i più efficaci, poiché risolvono tali criticità. Il risultato di questa ricerca sono gli algoritmi *Neighbor-Joining* ed il *Unweighted Pair Group Method with Arithmetic Mean (UPGMA)*, dove il primo permette la costruzione di alberi senza radice, mentre il secondo di quelli con radice.

Tutti gli algoritmi prendono in input una matrice D definita *matrice delle distanze*, chiamata così poiché ottenuta calcolando la distanza tra tutte le coppie di elementi di cui è composta.

Riportiamo, a titolo di esempio, un caso in cui l'applicazione di questi studi ha permesso di arginare la diffusione di una seria malattia: nei primi anni duemila, un virus chiamato SARS (Severe Acute Respiratory Syndrome), "Sindrome Acuta Respiratoria Grave" è stata la causa di diverse morti, soprattutto in Cina. Grazie alla costruzione degli alberi evolutivi è stato possibile studiare il virus e risalire "all'origine" della trasmissione della SARS, impendendo così la sua diffusione.