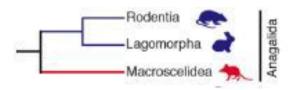


- 1. Introduzione
- Definizione di albero filogenetico
- Descrizione delle principali applicazioni degli alberi filogenetici in bioinformatica
- Struttura di un albero filogenetico
- Tipologie di albero filogenetico
- 2. Metodi per la costruzione degli alberi filogenetici
- 3. test statistici per valutazione dell'affidabilità
- 4. Applicazioni degli alberi filogenetici

- 1. Introduzione
- Definizione di albero filogenetico

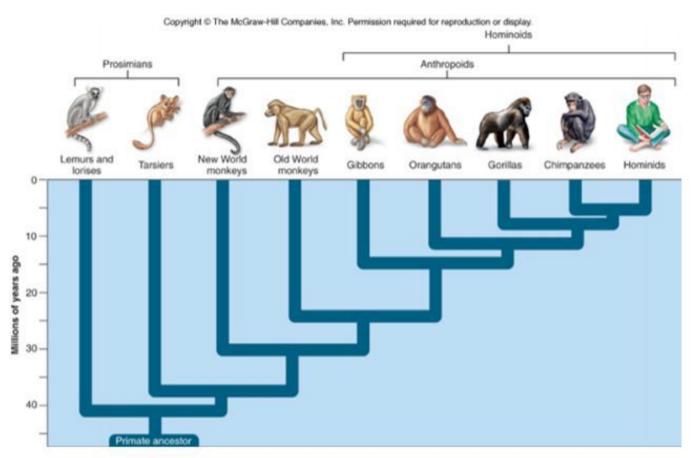


Un albero filogenetico, o albero evolutivo, è un diagramma ad albero che rappresenta la **relazione filogenetica** tra gli organismi o i loro geni sulla base di informazioni genetiche o morfologiche.

Questi alberi mostrano come gli organismi sono strettamente correlati evolutivamente, con rami che rappresentano gli eventi di speciazione e con un nodo comune che rappresenta l'ultimo antenato comune di questi organismi o geni.

Gli alberi filogenetici sono utilizzati per studiare **l'evoluzione** delle specie, l'origine della vita e la diversità biologica.

1. Introduzione



Le relazioni evolutive tra gli organismi, o più in generale tra geni o proteine omologhe, possono essere rappresentate attraverso alberi filogenetici o dendrogrammi.

Il termine "albero filogenetico" e "dendrogramma" sono spesso usati nel contesto delle analisi filogenetiche e della classificazione, ma possono avere sfumature leggermente diverse a seconda del contesto in cui vengono utilizzati.

Albero Filogenetico

Un albero filogenetico, è una rappresentazione grafica delle relazioni evolutive tra diverse specie, sottospecie, o gruppi genomici, basata sulle loro sequenze genomiche. Questi alberi sono costruiti sulla base di dati genomici come sequenze di DNA, RNA, o proteine e mirano a riflettere come le specie o i gruppi si sono divergenze nel tempo da un antenato comune.

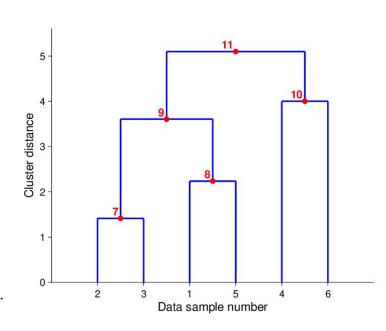
Gli alberi filogenetici sono spesso radicati, il che significa che hanno un punto designato come radice che rappresenta l'antenato comune più recente.

Dendrogramma

Un dendrogramma, d'altra parte, è un tipo di diagramma utilizzato per illustrare l'arrangiamento dei cluster formati da metodi di clustering gerarchico. (Il clustering gerarchico è un metodo di analisi dei cluster utilizzato per raggruppare insiemi di oggetti simili in cluster.)

Mentre può assomigliare a un albero filogenetico nella forma, il dendrogramma è tipicamente utilizzato in un contesto più ampio di analisi statistiche, non limitato solo alle relazioni evolutive. Può essere usato per mostrare come gli oggetti, come individui, specie, o campioni sperimentali, sono raggruppati basandosi su misure di somiglianza o distanza, come nelle analisi di espressione genica o in studi di mercato.

I dendrogrammi possono essere non radicati, il che significa che non implicano necessariamente una direzione evolutiva o un'origine



Gli alberi filogenetici sono strumenti fondamentali in bioinformatica e vengono utilizzati in diverse applicazioni.

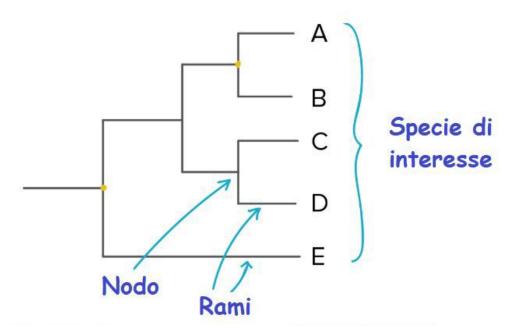
Ecco alcune delle principali applicazioni degli alberi filogenetici in bioinformatica:

- Classificazione tassonomica: gli alberi filogenetici vengono utilizzati per classificare e organizzare le specie in gruppi tassonomici. Ciò permette di identificare le relazioni evolutive tra le specie e di comprendere meglio la diversità biologica.
- Identificazione delle sequenze omologhe: gli alberi filogenetici possono essere utilizzati per identificare sequenze di DNA o proteine simili tra specie diverse. Ciò consente di identificare geni o proteine che svolgono funzioni simili in specie diverse e di capire meglio la loro evoluzione.

- **Predizione della funzione dei geni**: gli alberi filogenetici possono essere utilizzati per prevedere la funzione dei geni. Se un gene ha una relazione evolutiva stretta con un gene di cui si conosce la funzione, è probabile che abbia una funzione simile.
- Analisi della diversità genetica: gli alberi filogenetici possono essere utilizzati per analizzare la diversità genetica all'interno di una popolazione o tra popolazioni diverse. Ciò consente di studiare la distribuzione geografica delle specie e di comprendere meglio la loro evoluzione.
- Identificazione dei marcatori genetici: gli alberi filogenetici possono essere utilizzati per identificare marcatori genetici utili per lo studio di specifici fenomeni biologici, come la selezione naturale o l'adattamento alle condizioni ambientali

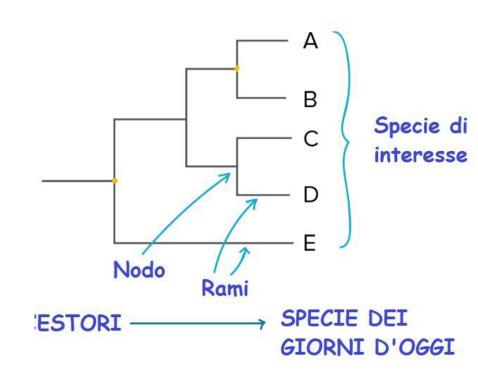
Struttura di un albero filogenetico
Un albero filogenetico è composto da nodi e rami.

I nodi rappresentano i punti di divergenza evolutiva, ovvero i punti in cui si sono separati due o più linee evolutive, mentre i rami rappresentano le linee evolutive stesse.



• Struttura di un albero filogenetico

La radice dell'albero è il **nodo più antico**, che rappresenta l'antenato comune più lontano delle specie o dei gruppi tassonomici analizzati. La lunghezza dei rami dell'albero può rappresentare la quantità di cambiamento evolutivo che è avvenuta lungo quel ramo, ad esempio, la quantità di mutazioni accumulate in una sequenza di DNA o la distanza evolutiva tra due sequenze di proteine.



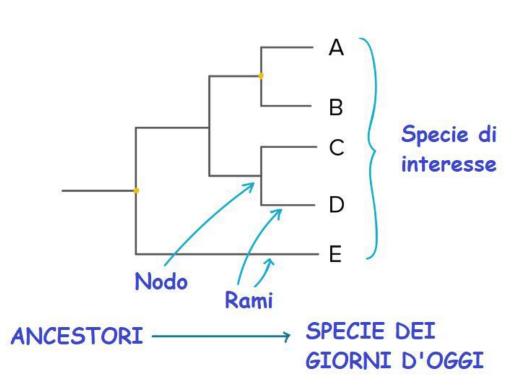
• Struttura di un albero filogenetico

Quando è presente un nodo con + di 2 rami si osserva la politomia.

Un albero senza politomie si definisce completamente risolto.

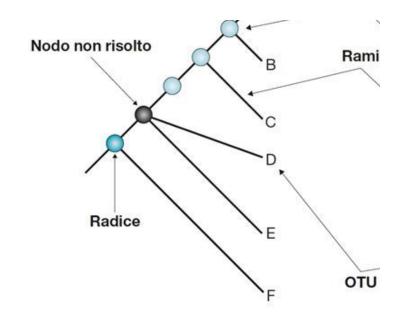
OTU: Unità Tassonomiche Operative (operational tassonomic unit)

Struttura dell'albero → **Topologia**



Struttura di un albero filogenetico

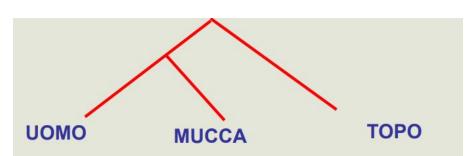
Un **nodo** di un albero filogenetico si dice **non risolto** quando non è possibile determinare con certezza la relazione evolutiva tra i cladi che si originano da esso. In altre parole, non è possibile stabilire se i cladi sono veramente correlati da un antenato comune o se la somiglianza tra le sequenze è dovuta al caso o ad altre ragioni.



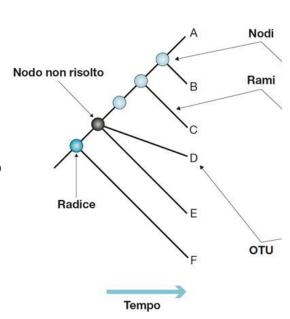
• Struttura di un albero filogenetico

In un albero filogenetico **con radice** (rooted tree), viene rappresentata una gerarchia di relazioni evolutive tra gli organismi, partendo da un antenato comune rappresentato dalla radice.

L'albero può essere orientato, con la radice in basso e i rami che si estendono verso l'alto, oppure invertito, con la radice in alto e i rami che si estendono verso il basso.



Albero con radice



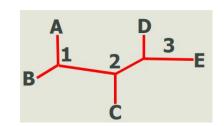
Corso Bioinformatica 1 -2024 Prof.ssa Tiziana Castrignanò email: tiziana.castrignano@unitus.it

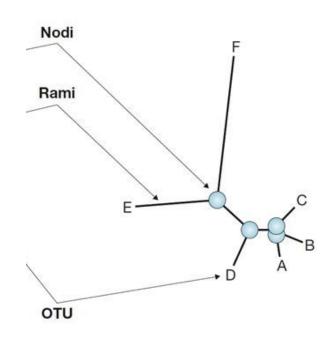
Struttura di un albero filogenetico

Albero senza radice

In un albero filogenetico **senza radice** (unrooted tree), la gerarchia di relazioni evolutive tra gli organismi viene rappresentata senza la presenza di un antenato comune definito come radice.

Gli organismi sono quindi collegati da relazioni di parentela basate sulla somiglianza dei loro caratteri, senza che sia possibile distinguere tra loro una relazione di discendenza diretta.

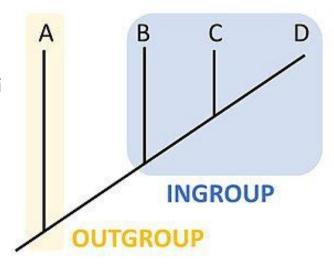




Struttura di un albero filogenetico

Un altro aspetto importante da conoscere per l'interpretazione di un albero sono gli **outgroup**: un gruppo di organismi più distante che funge da gruppo di riferimento quando si determinano le relazioni evolutive dell'ingroup, l'insieme di organismi in studio.

L'outgroup viene utilizzato come punto di confronto per **l'ingroup** e consente specificamente di radicare la filogenesi (creare un albero con radice). Poiché la direzione del cambio di carattere può essere determinata solo su una filogenesi con radice, la scelta del gruppo esterno è essenziale per comprendere l'evoluzione dei tratti lungo una filogenesi.



Un semplice cladogramma che mostra le relazioni evolutive tra quattro specie: A, B, C e D. Qui, la specie A è il gruppo esterno (outgroup) e le specie B, C e D formano il gruppo interno (ingroup).

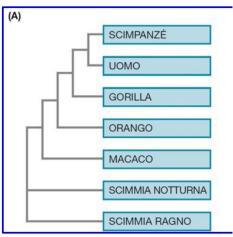
• Tipologie di albero filogenetico

Se un albero filogenetico descrive solo la topologia, ovvero le relazioni filogenetiche tra i vari nodi, allora la lunghezza dei rami non ha alcun significato e l'albero si definisce **cladogramma**.

Un cladogramma è un tipo di albero filogenetico che rappresenta le relazioni evolutive tra gruppi di organismi sulla base delle loro caratteristiche condivise, o caratteri derivati.

Un cladogramma indica solo l'ordine relativo di divergenza tra i rami dell'albero, ma non fornisce informazioni sulla durata degli eventi evolutivi o sulla quantità di cambiamento evolutivo tra i

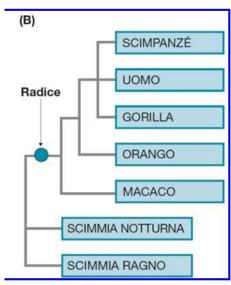
rami.



Tipologie di albero filogenetico

Un **filogramma**, d'altra parte, è un tipo di albero filogenetico che rappresenta la storia evolutiva dei gruppi di organismi sulla base di informazioni sul tempo e sulla quantità di cambiamento evolutivo.

Un filogramma può essere costruito utilizzando tecniche di datazione molecolare o altri metodi per stimare il tempo di divergenza tra i rami dell'albero e per quantificare la quantità di cambiamento evolutivo tra i gruppi di organismi.

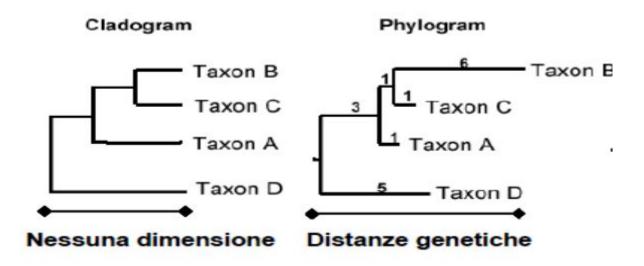


Tipologie di albero filogenetico

Quindi...

Cladogramma: albero filogenetico che descrive esclusivamente la topologia

Filogramma: albero filogenetico in cui la lunghezza dei rami è proporzionale alla distanza evolutiva tra i nodi.



Tipologie di albero filogenetico

Per la costruzione di un albero filogramma è importante introdurre l'ipotesi dell'orologio molecolare.

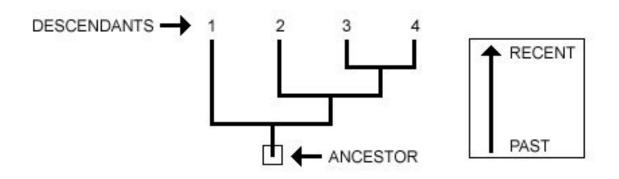
L'orologio molecolare è un concetto importante nell'evoluzione biologica che si riferisce alla **velocità** con cui **mutazioni** genetiche casuali si accumulano in sequenze di DNA o RNA nel corso del tempo.

L'idea alla base dell'orologio molecolare è che le mutazioni genetiche avvengono ad un **ritmo costante** all'interno di una specie, in modo che il numero di mutazioni che si accumulano in una determinata regione del genoma di due organismi correlati possa essere utilizzato per calcolare l'età dell'antenato comune più

recente.

• Tipologie di albero filogenetico

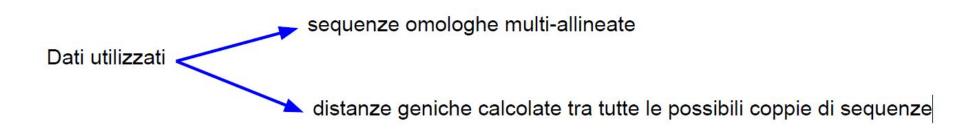
Ad esempio, se due specie hanno sequenze di DNA molto diverse, ci si aspetta che siano evolutivamente distanti, mentre se le loro sequenze sono molto simili, ci si aspetta che siano evolutivamente vicine. L'orologio molecolare viene utilizzato per quantificare questa distanza evolutiva in termini di tempo, utilizzando una scala temporale come anni o milioni di anni.



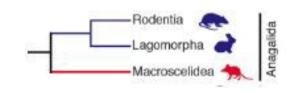
- Rodentia - Rodentia

1. Metodi per la costruzione degli alberi filogenetici

Ci sono vari metodi per la costruzione degli alberi filogenetici, tra cui quelli basati sul clustering e quelli basati sulle massime verosimiglianze.



1. Metodi per la costruzione degli alberi filogenetici

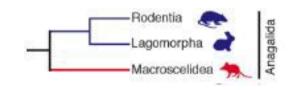


Tra quelli basati su algoritmi di clustering il più semplice è quello noto come **UPGMA** (Unweighted Pair Group Method with Arithmetic Mean).

L'algoritmo UPGMA è un metodo basato sul clustering per costruire alberi filogenetici. L'idea principale dell'algoritmo è quella di raggruppare **sequenze simili** in gruppi successivi, a partire dalle sequenze più simili e procedendo in ordine di **dissimilarità crescente.**

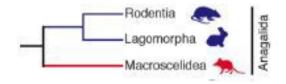
Possiamo quindi affermare che agisce sulla base delle distanze genetiche.

1. Metodi per la costruzione degli alberi filogenetici



Come funziona UPGMA:

- Calcola una matrice di distanza tra tutte le sequenze delle diverse specie.
- Raggruppa le due sequenze più simili in un cluster.
- Calcola la distanza tra questo nuovo cluster e tutti gli altri cluster o sequenze.
- Ripete i passaggi 2 e 3 finché tutte le sequenze sono raggruppate in un unico cluster.



UPGMA piu' nel dettaglio:

 Calcolo delle Distanze: Inizia calcolando una matrice di distanza tra tutte le coppie di specie o sequenze, usando una misura di dissimilarità genetica.

2. Formazione dei Cluster:

- Trova le due specie più vicine nella matrice di distanza e uniscile in un nuovo cluster.
- L'altezza del punto in cui queste specie si uniscono nel dendrogramma è pari alla metà della loro distanza. Questo rappresenta il momento della loro ultima divergenza comune.

3. Aggiornamento della Matrice:

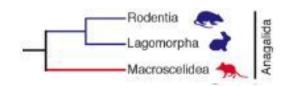
- Calcola la distanza tra il nuovo cluster e gli altri cluster/specie come la media aritmetica delle distanze preesistenti.
- Aggiorna la matrice di distanza per riflettere l'inclusione del nuovo cluster e la rimozione delle specie originali.

4. Iterazione:

 Ripeti il processo di formazione dei cluster e aggiornamento della matrice fino a che non resta un solo cluster che include tutte le specie.

5. Costruzione dell'Albero:

 Completa l'albero filogenetico, che mostra le relazioni evolutive tra le specie con i nodi che indicano gli antenati comuni e i rami che rappresentano la divergenza genetica.



Esempio di applicazione di UPGMA:

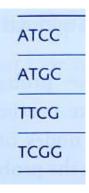
Supponiamo di avere le seguenti quattro sequenze di DNA:

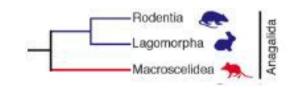
Sequenza 1: ATCC

Sequenza 2: ATGC

Sequenza 3: TTCG

Sequenza 4: TCGG



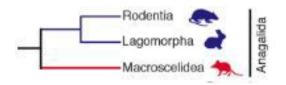


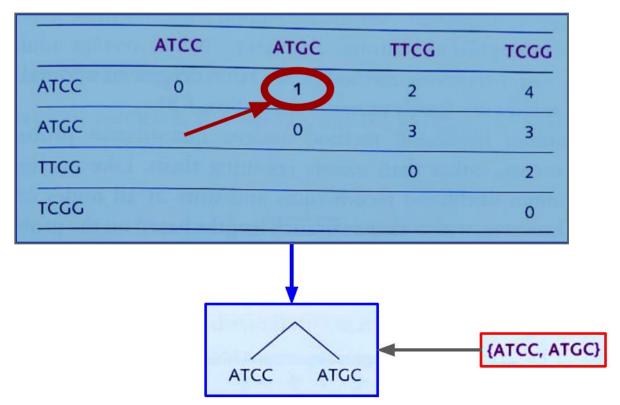
Matrice di distanza:

Calcoliamo la matrice di distanza tra le sequenze

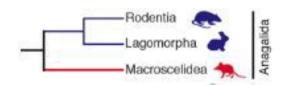
bli tajii t	ATCC	ATGC	TTCG	TCGG
ATCC	0	1	2	4
ATGC		0	3	3
ттсс	il med po	Accies ylano.	0	2
TCGG	III TI SHOKE	Att expression		0

Raggruppiamo le sequenze 1 e 2, con una distanza di 1:





1. Metodi per la costruzione degli alberi filogenetici

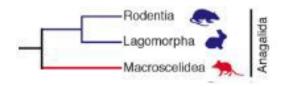


Ricalcolo della distanza:

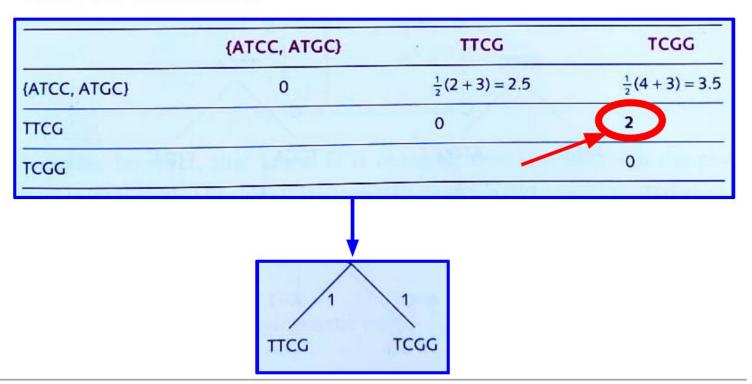
Calcoliamo la distanza tra la sequenza 1,2 e la sequenza 3 e 4:

	{ATCC, ATGC}	TTCG	TCGG
(ATCC, ATGC)	0	$\frac{1}{2}(2+3)=2.5$	$\frac{1}{2}(4+3)=3.5$
TTCG		0	2
TCGG	21, ship persed for a s	A LANGUAGE	0

1. Metodi per la costruzione degli alberi filogenetici



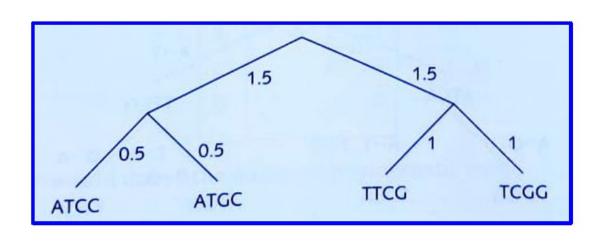
Matrice delle distanze ridotta:



1. Metodi per la costruzione degli alberi filogenetici

Rodentia Lagomorpha Macroscelidea

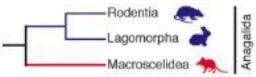
Raggruppiamo



Siamo arrivati all'albero finale avendo seguito la seguente regola:

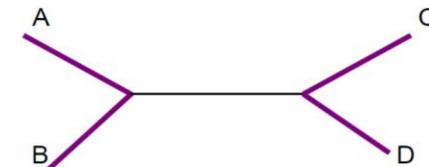
la lunghezza dei rami tra i nodi X e Y è uguale alla metà della distanza tra i nodi X e Y.

1. Metodi per la costruzione degli alberi filogenetici

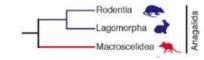


L'algoritmo **Neighbor-joining** è un metodo basato sul clustering per costruire alberi filogenetici.

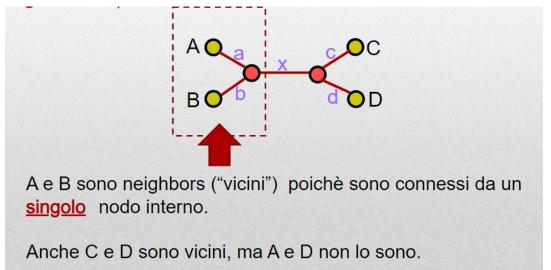
A differenza dell'algoritmo UPGMA, Neighbor-joining tiene conto della diversità all'interno dei gruppi e cerca di minimizzare la somma delle distanze tra i nodi dell'albero.



Come funziona Neighbor-joining:



Il metodo del Neighbor-joining utilizza un algoritmo di clustering che parte da una topologia a stella (albero completamente non risolto) e, attraverso iterazioni successive, unisce coppie di OTUs vicine (neighbor), separandole dal resto dell'albero ancora non risolto, in modo da minimizzare la somma delle lunghezze di tutti i rami dell'albero.

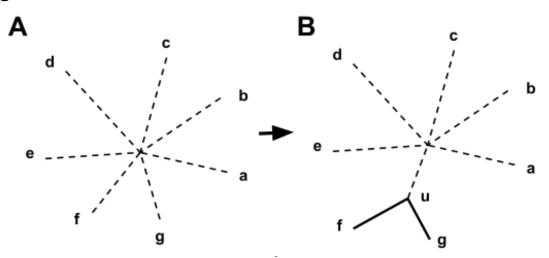


Corso Bioinformatica 1 -2024 Prof.ssa Tiziana Castrignanò email: tiziana.castrignano@unitus.it

1. Metodi per la costruzione degli alberi filogenetici

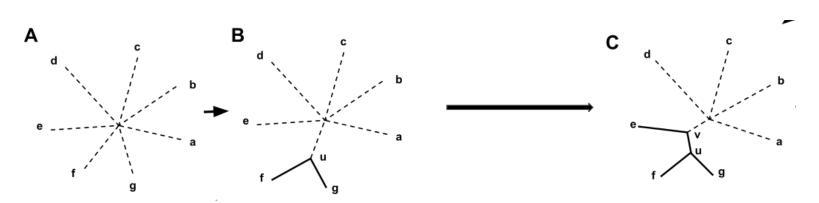
Esempio di costruzione: Partendo da un albero a stella (A), la matrice Q viene calcolata e utilizzata per scegliere una coppia di nodi da unire, in questo caso f e g. Questi sono uniti a un nodo appena creato, u, come mostrato in (B).

La parte dell'albero mostrata come linee continue è ora fissa e non verrà modificata nei successivi passaggi di unione.



1. Metodi per la costruzione degli alberi filogenetici

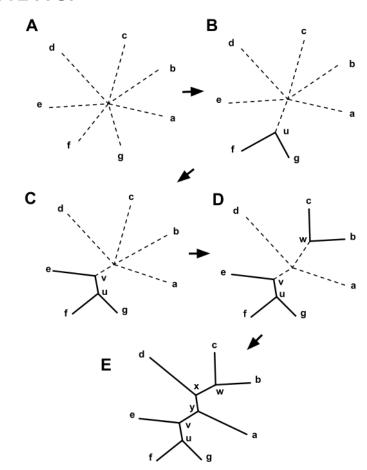
Le distanze dal nodo u ai nodi ae sono calcolate dall'equazione (3). Questo processo viene quindi ripetuto, utilizzando una matrice delle sole distanze tra i nodi, a, b, c, d, e e u, e una matrice Q derivata da essa. In questo caso u ed e sono uniti alla v appena creata, come mostrato in (C).



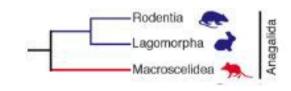
Corso Bioinformatica 1 -2024 Prof.ssa Tiziana Castrignanò email: tiziana.castrignano@unitus.it

1. Metodi per la costruzione degli alberi filogenetici

Altre due iterazioni portano prima a (D) e poi a (E), a quel punto l'algoritmo è terminato, poiché l'albero è completamente risolto.



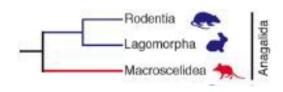
1. Metodi per la costruzione degli alberi filogenetici



Il metodo di neighbor Johnning è molto rapido, tiene conto della possibilità di differenti velocità di evoluzione lungo diversi rami dell'albero ed è generalmente affidabile.

L'applicazione di questo metodo consente di stimare la lunghezza di tutti i rami dell'albero filogenetico. In questo modo è possibile ricalcolare una matrice delle distanze tra tutte le coppie di OTU considerate nell'analisi.

1. Metodi per la costruzione degli alberi filogenetici



L'obiettivo della costruzione degli alberi filogenetici è quello di rappresentare le relazioni evolutive tra gli organismi. Il metodo della **massima parsimonia** è uno dei metodi usati per costruire alberi filogenetici e si basa sull'idea di trovare l'albero che richiede il minor numero di cambiamenti evolutivi (o mutazioni) per spiegare le relazioni tra le sequenze.

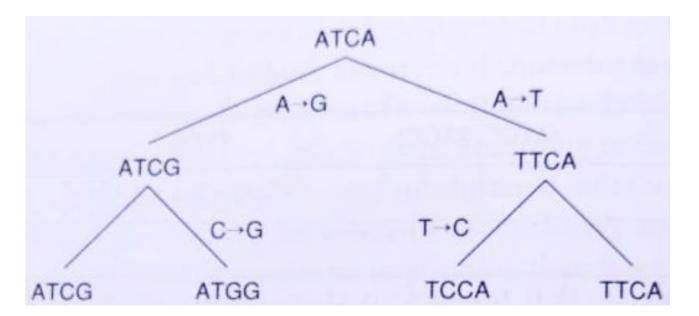
Metodo qualitativo → consente di determinare la topologia dell'albero che descrive le relazioni filogenetiche tra le sequenze in esame.

Per calcolare l'albero di massima parsimonia bisogna prima selezionare i siti informativi del multi-allineamento.

Un sito è considerato informativo solo se contiene almeno due differenti caratteri (nucleotidi o aminoacidi), ciascuno dei quali sia presente almeno in due delle sequenze in esame.

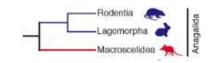
Come funziona la massima parsimonia:

L'algoritmo della massima parsimonia calcola l'albero filogenetico che minimizza il numero totale di cambiamenti evolutivi. In altre parole, cerca di trovare l'albero più parsimonioso (ovvero, che richiede il minor numero di mutazioni) per spiegare le relazioni tra le sequenze

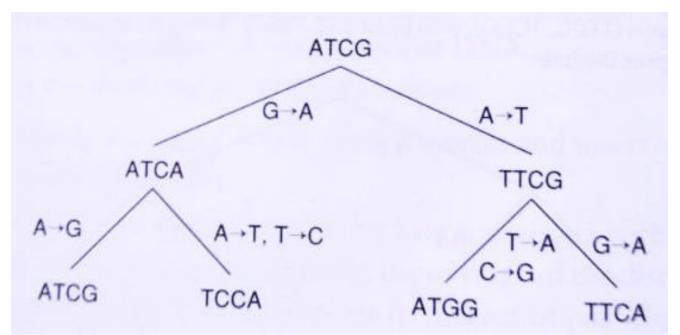


Albero ottimale ottenuto con la massima parsimonia cercando il piu' basso numero di sostituzioni pari a 4 (in questo caso).

1. Metodi per la costruzione degli alberi filogenetici

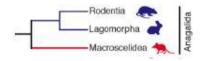


Albero alternativo prodotto con la massima parsimonia che produce un numero di sostituzioni pari a 7 (che è quindi da escludersi rispetto a quello ottimale).



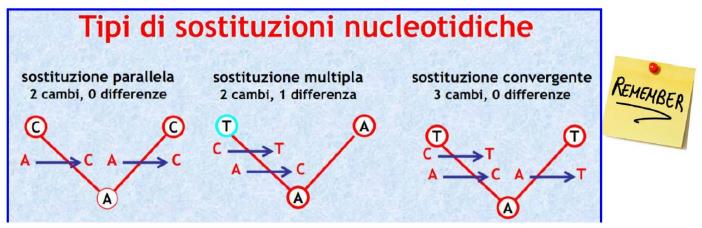
Albero alternativo prodotto con la massima parsimonia che produce un numero di sostituzioni pari a 7 (che è quindi da escludersi rispetto a quello ottimale).

I. Metodi per la costruzione degli alberi filogenetici

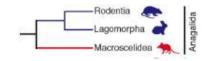


Problemi dei metodi basati sulla massima parsimonia

- 1. Non considera alcun modello esplicito di evoluzione molecolare
- 2. Non tiene conto di sostituzioni parallele, convergenti o multiple che sono tanto più frequenti quanto maggiore è la distanza tra le sequenze(*).

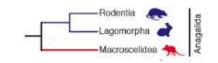


1. Metodi per la costruzione degli alberi filogenetici



- (*) La limitazione 2. ha maggiore impatto sulle sequenze nucleotidiche poiché usano un alfabeto di soli 4 caratteri.
- 3. Tutte le sostituzioni vengono considerate equivalenti e questo non corrisponde al reale processo evolutivo subito dalle sequenze.
- 4. Non sempre viene trovata una soluzione univoca.

1. Metodi per la costruzione degli alberi filogenetici



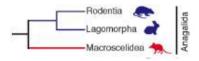
Il metodo della **massima verosimiglianza** è uno dei metodi usati per costruire alberi filogenetici e si basa sull'idea di trovare l'albero più verosimile (ovvero, più probabile) per spiegare le relazioni tra le sequenze.

Il metodo della massima verosimiglianza è il più consistente con i dati del multi-allineamento.

$$L = \Pr(D|T)$$

L'albero che ottiene la massima probabilità rappresenta la stima di massima verosimiglianza della filogenesi tra le OTUs considerate.

1. Metodi per la costruzione degli alberi filogenetici

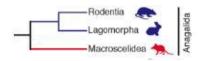


Come funziona la massima verosimiglianza:

L'algoritmo della massima verosimiglianza calcola l'albero filogenetico che **massimizza** la **probabilità** di osservare le sequenze date, considerando un modello di evoluzione appropriato.

In altre parole, cerca di trovare l'albero più verosimile per spiegare le relazioni tra le sequenze.

1. Metodi per la costruzione degli alberi filogenetici

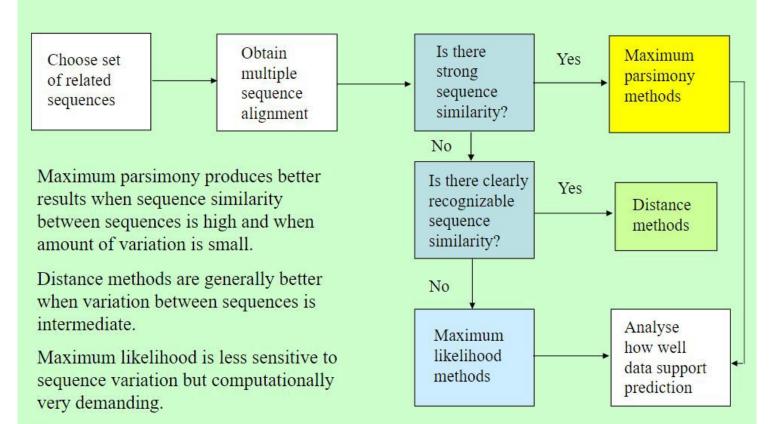


Limiti del metodo della massima verosimiglianza

Il principale limite del metodo della massima verosimiglianza è di natura computazionale. Con un numero di sequenze > 20 il calcolo delle probabilità diventa non affrontabile in tempi ragionevoli.

Si usano metodi approssimati (euristici) per ottenere le probabilità in tempi ragionevoli. Uno di questi metodi è il metodo dei quartetti (Quartet Puzzling, QP) che cerca l'albero di massima verosimiglianza tra tutti i possibili alberi determinati da gruppi di 4 sequenze.

A strategy of phylogenetic reconstruction

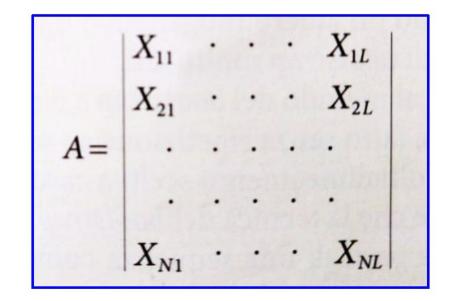


Test statistici per valutare il grado di affidabilità delle ricostruzioni filogenetiche

Problema → Come misurare la robustezza di un'ipotesi filogenetica?

Tecnica più usata → ricampionamento casuale dei siti del multiallineamento utilizzato nell'analisi (*bootstrap*).

Consideriamo *N* sequenze nucleotidiche multi-allineate, *Si* (i=1,N) di lunghezza *L*. Il loro allineamento può essere rappresentato dalla matrice A. Un sito del multiallineamento corrisponde ad una colonna di questa matrice.



Test statistici per valutare il grado di affidabilità delle ricostruzioni filogenetiche

Scegliamo a caso N siti, anche piu' volte (ricampionamento) → multiallineamento simulato

Generiamo migliaia di allineamenti multipli con questa procedura → migliaia di alberi filogenetici ottimali.

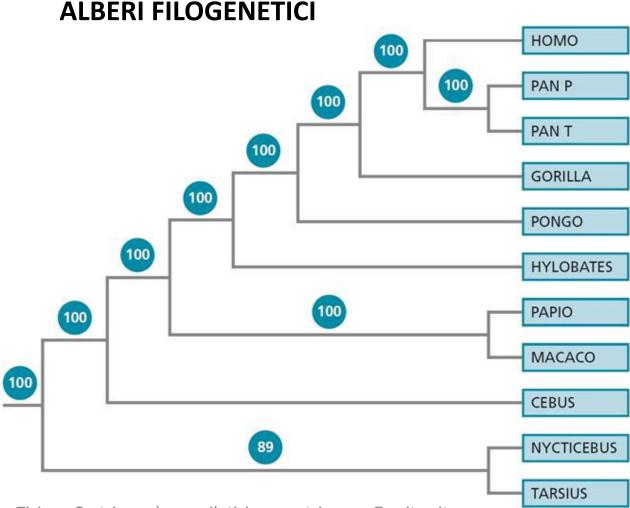
Costruiamo un albero consenso con uno score per nodo pari alla percentuale di alberi simulati in cui viene riprodotto il nodo (percentuale di bootstrap).

Due alberi riproducono uno stesso nodo quando questo, in ambedue gli alberi, il nodo condivide esattamente lo stesso insieme di OTUs discendenti.

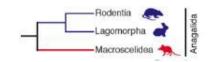
Più' è alto il nodo di bootstrap più il nodo considerato risulterà statisticamente significativo.

Figura 6.6

Albero filogenetico di alcune specie di primati determinato dal programma MrBayes sulla base dell'analisi evolutiva dei geni mitocondriali con l'indicazione dei valori di bootstrap a supporto dei diversi nodi.



1. Applicazioni degli alberi filogenetici

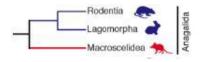


Riassumendo...

Gli alberi filogenetici trovano numerose applicazioni in diversi campi della biologia e della bioinformatica, tra cui:

- 1. **Evoluzione** e tassonomia: gli alberi filogenetici forniscono una rappresentazione visiva dell'evoluzione delle specie e della loro relazione tassonomica.
- Classificazione e identificazione: gli alberi filogenetici possono essere utilizzati per classificare nuove specie e per identificare relazioni tra organismi di diverse specie.
- 3. Biologia molecolare: gli alberi filogenetici possono essere utilizzati per studiare l'evoluzione delle sequenze molecolari (ad esempio, geni, proteine, RNA) e per identificare relazioni tra sequenze molecolari.

1. Applicazioni degli alberi filogenetici



- 4. Biologia **evolutiva**: gli alberi filogenetici possono essere utilizzati per studiare i meccanismi dell'evoluzione, ad esempio la selezione naturale e la deriva genetica.
- Biologia delle popolazioni: gli alberi filogenetici possono essere utilizzati per studiare la struttura e la dinamica delle popolazioni e per identificare sottopopolazioni e flussi genici.
- 6. **Epidemiologia**: gli alberi filogenetici possono essere utilizzati per studiare la diffusione di malattie infettive e per identificare le origini e le rotte di trasmissione.

