

Introduzione all'analisi dei caratteri attraverso la parsimonia

Relazioni Genetiche

- **Fra individui all'interno delle popolazioni esistono relazioni genetiche.**
- **Queste comprendono relazioni fra antenato e discendente e relazioni più indirette basate sulla comune origine.**
- **All'interno di popolazioni che si riproducono sessualmente, esiste una rete di relazioni.**
- **Le relazioni genetiche all'interno delle popolazioni possono essere misurate con un coefficiente di affinità genetica.**

Relazioni Filogenetiche

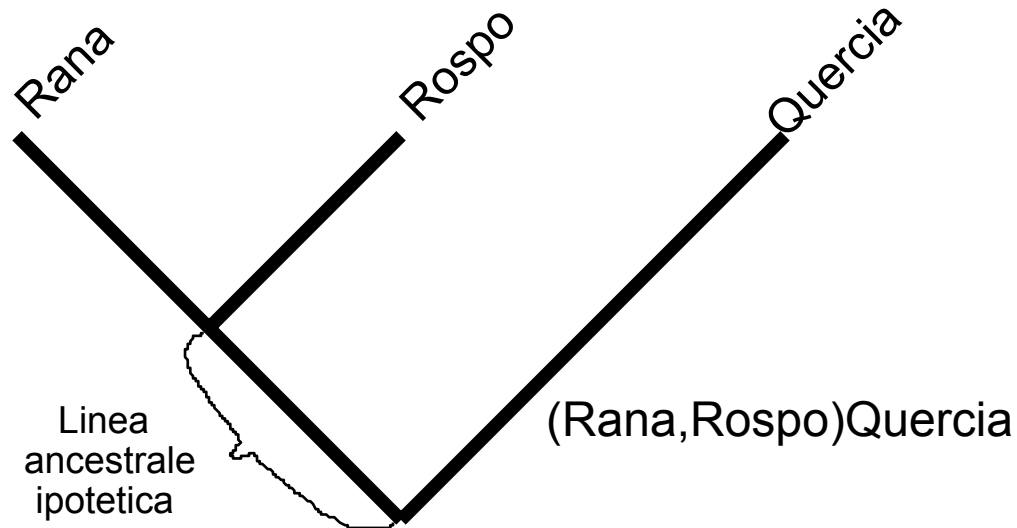
- **Fra linee evolutive (specie, geni) esistono relazioni filogenetiche.**
- **Queste comprendono relazioni fra antenato e discendente e relazioni più indirette basate sulla comune origine.**
- **Le relazioni filogenetiche fra specie o linee sono attese essere a forma di albero.**
- **Le relazioni filogenetiche non vengono misurate con un semplice coefficiente.**

Relazioni Filogenetiche

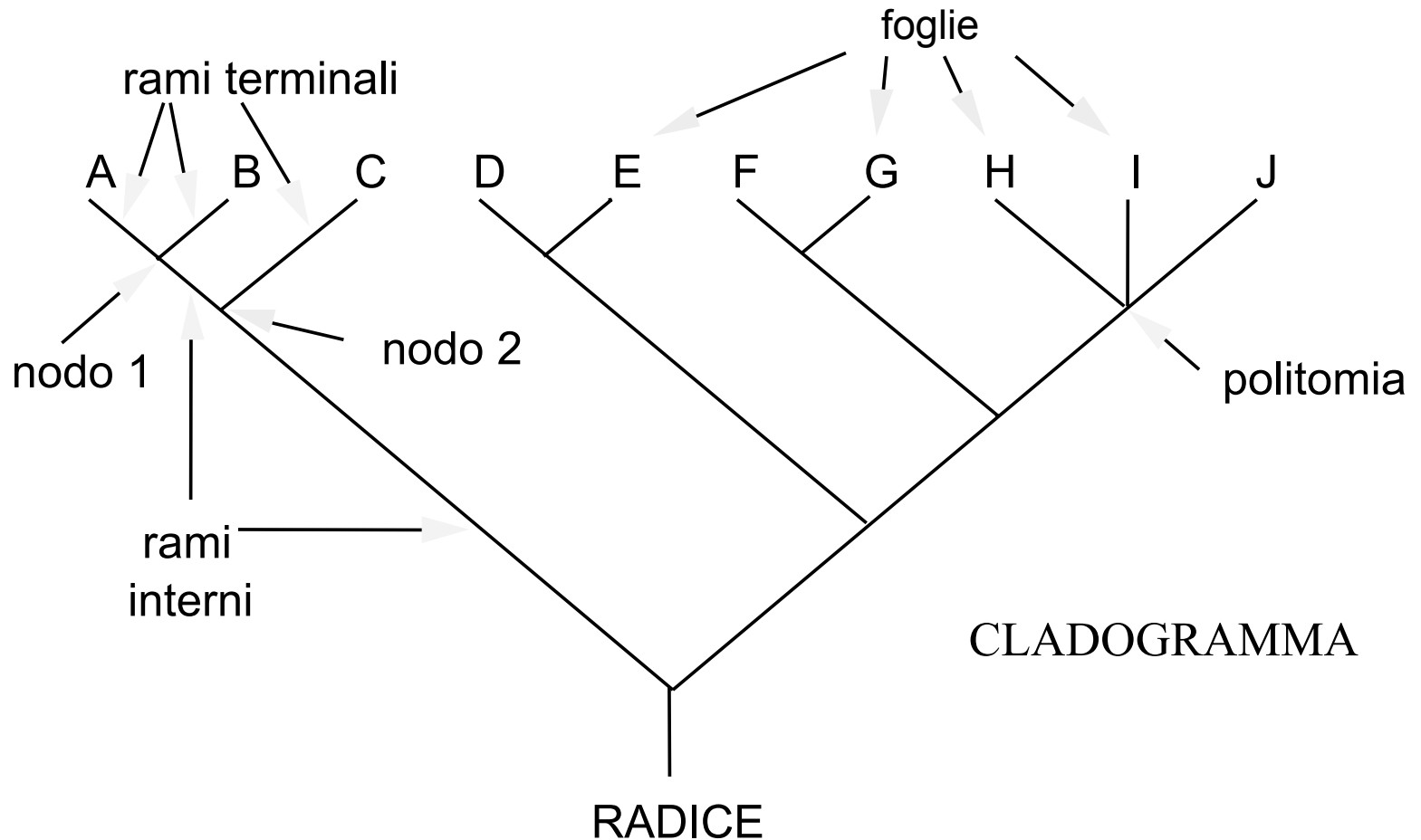
- **Tradizionalmente la ricostruzione filogenetica è stata dominata dalla ricerca degli antenati e dalla relazione antenato-discendente.**
- **Nella filogenesi moderna l'enfasi è spostata sulle relazioni indirette.**
- **Dal momento che tutte le linee sono correlate, l'affinità delle relazioni filogenetiche è un concetto relativo.**

Relazioni Filogenetiche

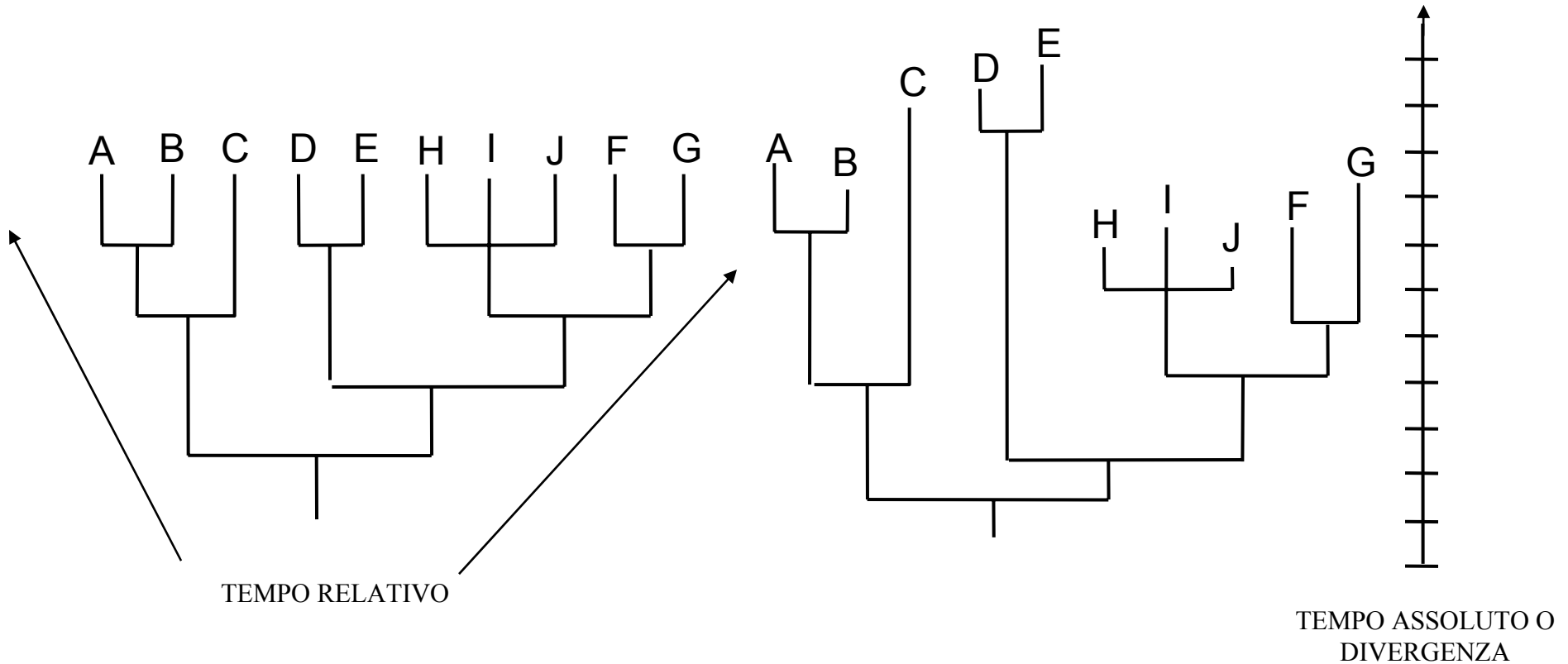
- Due linee sono più strettamente affini l'una all'altra che non a qualsiasi altra linea, se condividono un antenato più recente – Questo è un concetto cladistico.
- Le ipotesi filogenetiche sono ipotesi di origine comune.



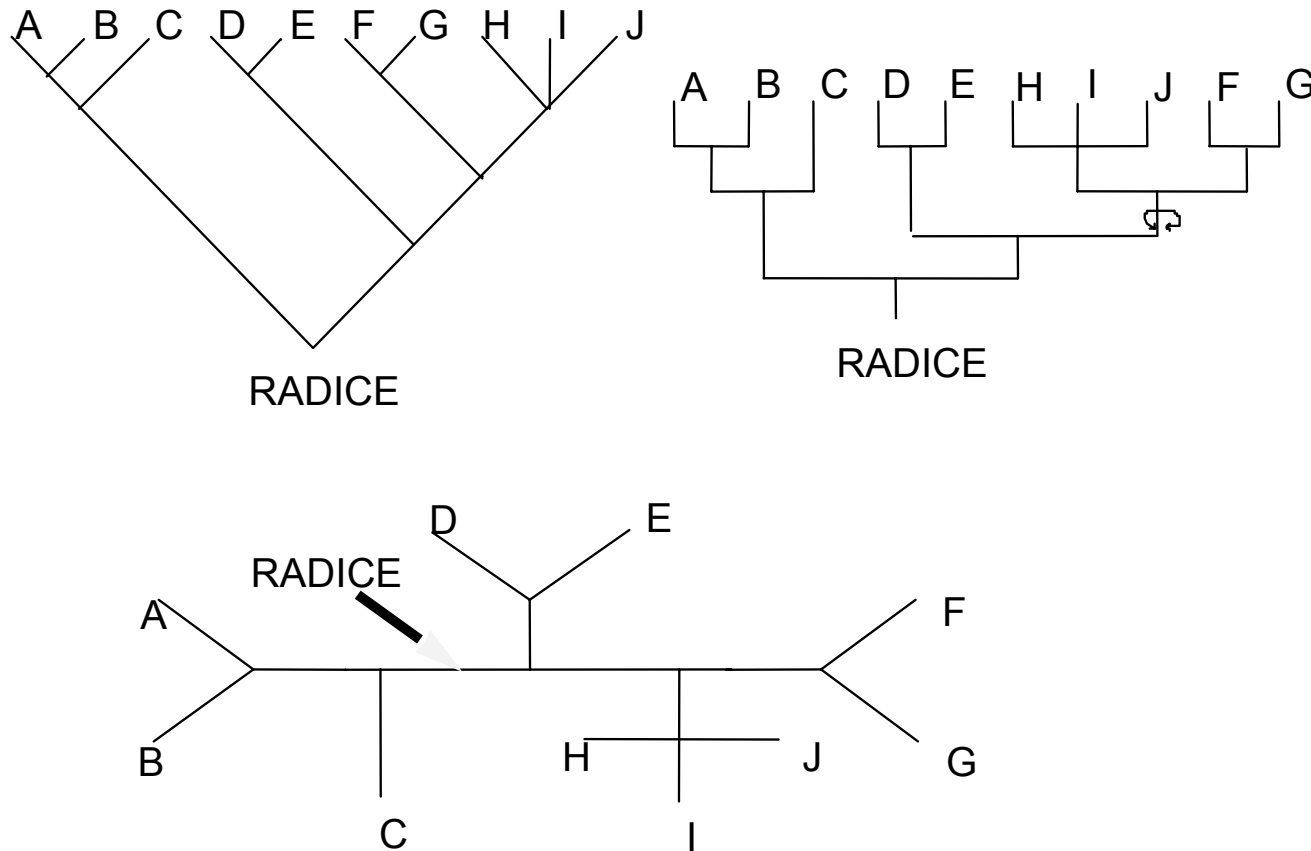
Alberi Filogenetici



CLADOGRAMMI E FILOGRAMMI



Alberi radicati e non



Caratteri e Stati del Carattere

- Gli organismi presentano gruppi di caratteristiche.
- Quando gli organismi/taxa differiscono per una caratteristica (per es., presenza o assenza o base nucleotidica diversa in un sito di una determinata sequenza) le condizioni diverse sono definite come *stati del carattere*.
- L'insieme degli stati di un carattere, rispetto ad una determinata caratteristica, costituisce un *carattere*.

Evoluzione del carattere

- **Cambiamenti ereditabili, nella morfologia, nelle sequenze geniche, etc., producono differenti stati del carattere.**
- **Le affinità e le differenze negli stati del carattere forniscono la base per comprendere la filogenesi (provvedono, cioè, le evidenze per le relazioni).**
- **L'utilità di queste evidenze dipende da quanto spesso i cambiamenti evolutivi che causano i diversi stati del carattere avvengano indipendentemente.**

Caratteri unici e immutabili

Dato un cambiamento evolutivo ereditabile che è unico e immutabile (per es., l'origine dei peli), in una specie ancestrale, la presenza di un nuovo stato del carattere in un taxon qualsiasi deve essere legato all'ereditarietà dall'antenato.

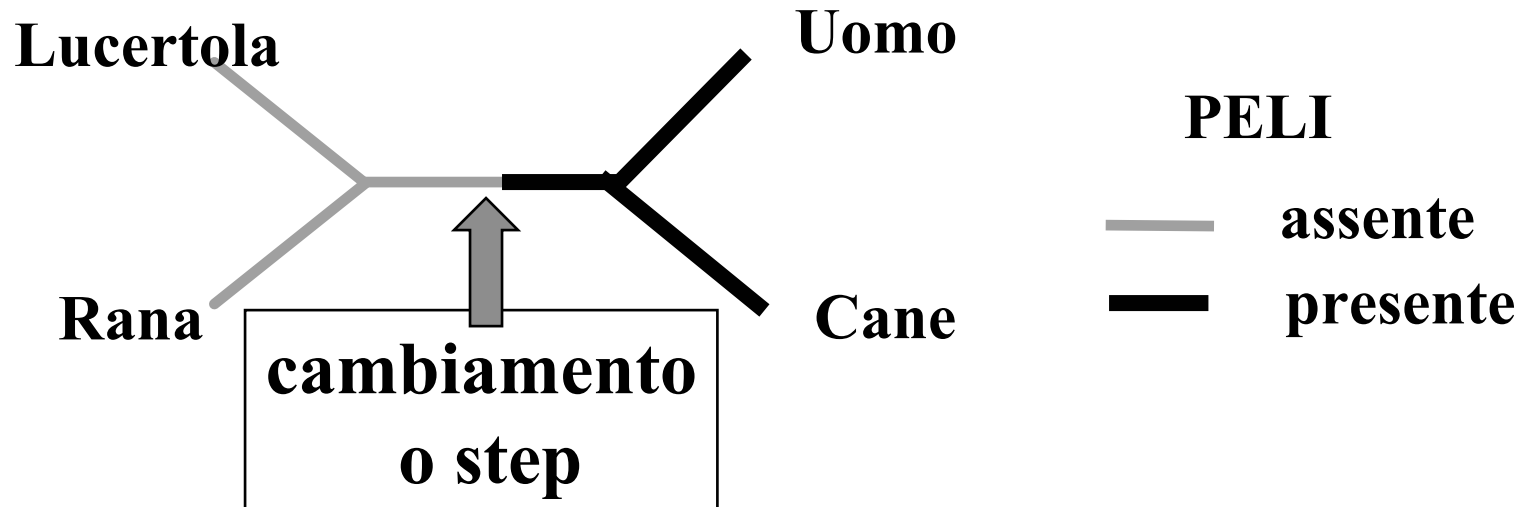
D'altra parte, la sua assenza indica un'origine diversa dei taxa sotto studio.

La novità è un'omologia che marca la discendenza.

I taxa con la novità rappresentano un clado (per es. I Mammiferi).

Caratteri unici e immutabili

- Poichè il carattere “peli” si è evoluto una volta sola ed è immutabile, ovvero tale novità non è retromutata e non è andata persa nel corso dell’evoluzione, rappresenta un carattere omologo e fornisce evidenze non ambigue sulle relazioni fra i taxa.

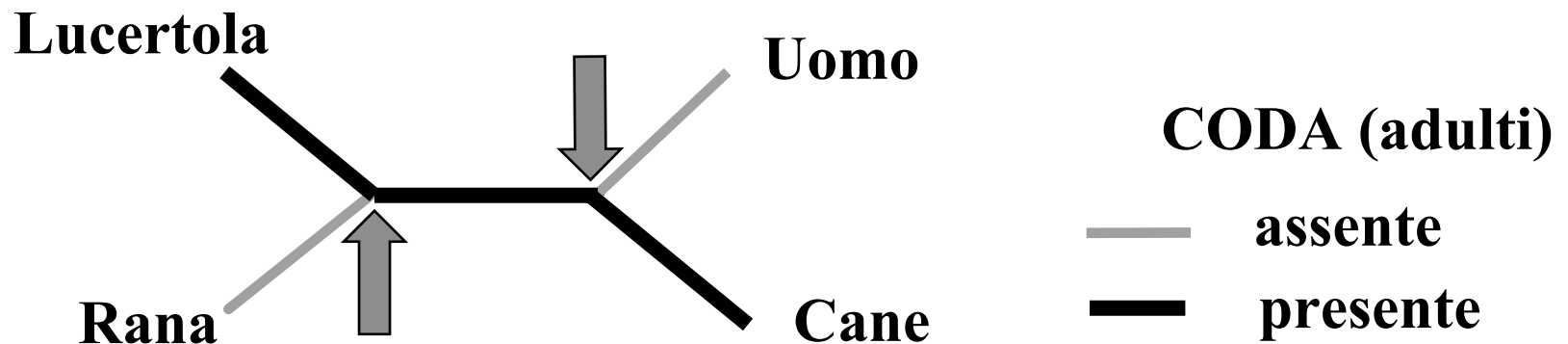


Omoplasia – Evoluzione Indipendente

- L'omoplasia è la similarità non omologa (dovuta a causa diverse dall'origine comune).**
- E' il risultato di evoluzione indipendente (convergenza, parallelismo, retromutazioni).**
- L'omoplasia può fornire evidenze errate circa le relazioni filogenetiche, se è interpretata erroneamente come omologia.**

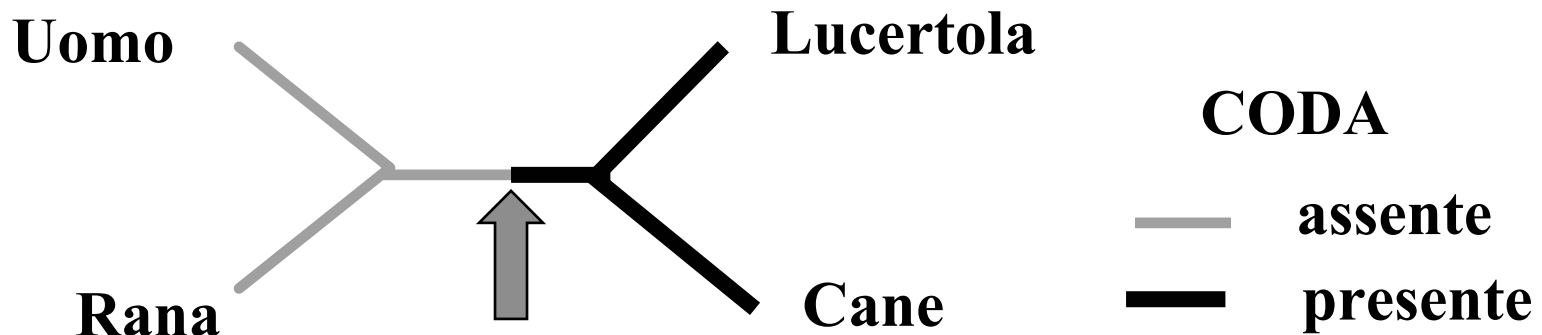
Omoplasia – Evoluzione Indipendente

- La perdita della coda si è evoluta indipendentemente nell'uomo e nella rana – ci sono due cambiamenti sul vero albero.



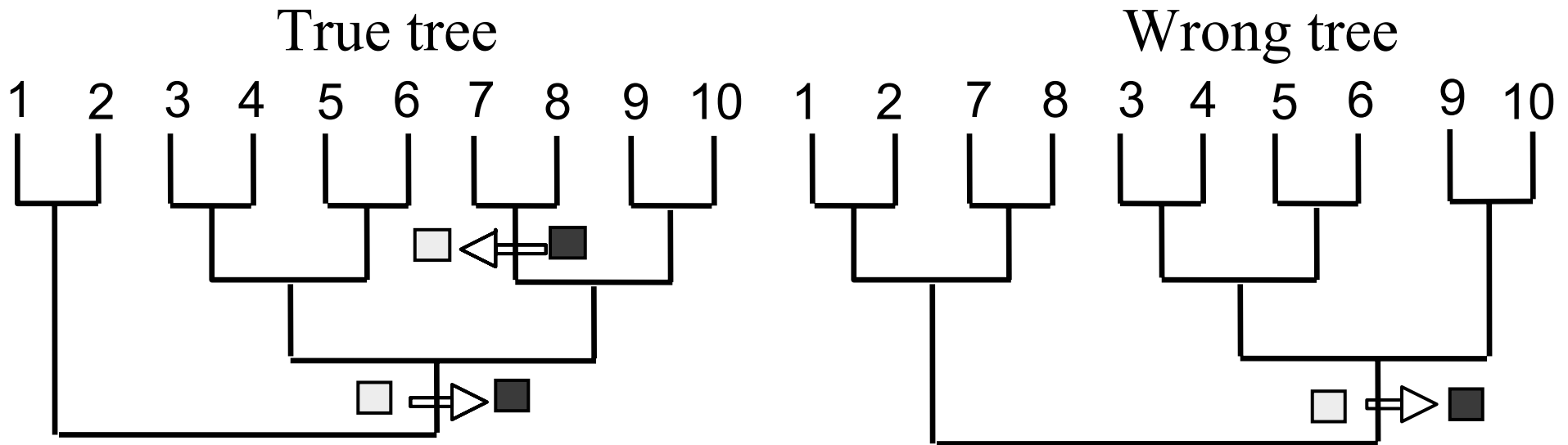
Omoplasia – deviante nelle evidenze di filogenesi

- Se interpretata erroneamente come omologia, l'assenza della coda sarebbe una prova per un albero sbagliato: raggrupperebbe l'uomo con la rana e la lucertola con il cane.



Omoplasia - retromutazioni

- Le retromutazioni sono cambiamenti evolutivi che tornano indietro alla condizione ancestrale.
- Come qualsiasi altro carattere omoplasico, le retromutazioni possono produrre evidenze errate di relazioni.



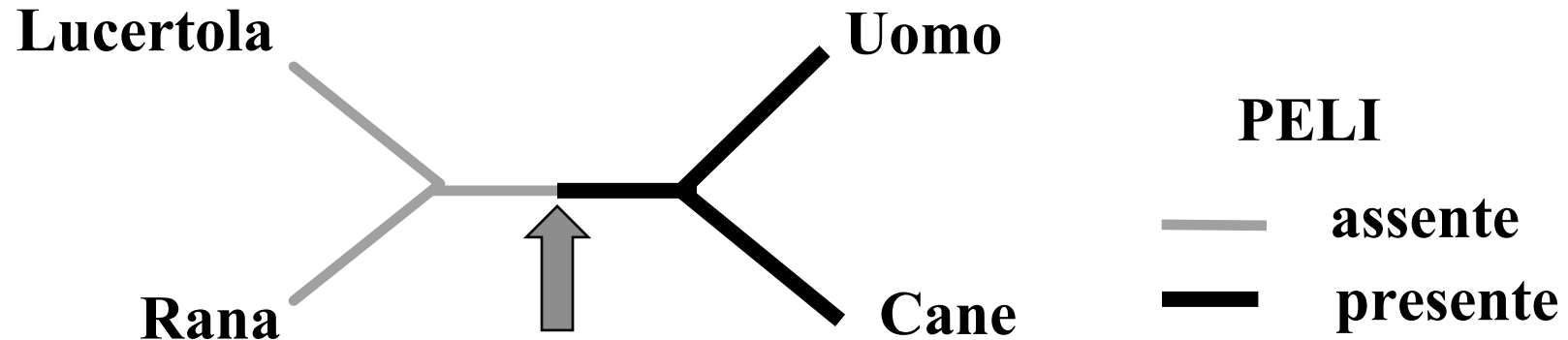
Omoplasia- un problema importante nello studio della filogenesi

- Se non ci fossero similarità dovute ad omoplasia, lo studio della filogenesi sarebbe facile – tutti i pezzi del mosaico coinciderebbero facilmente.**
- Distinguere l'omoplasia dall'omologia è quindi molto importante negli studi filogenetici.**

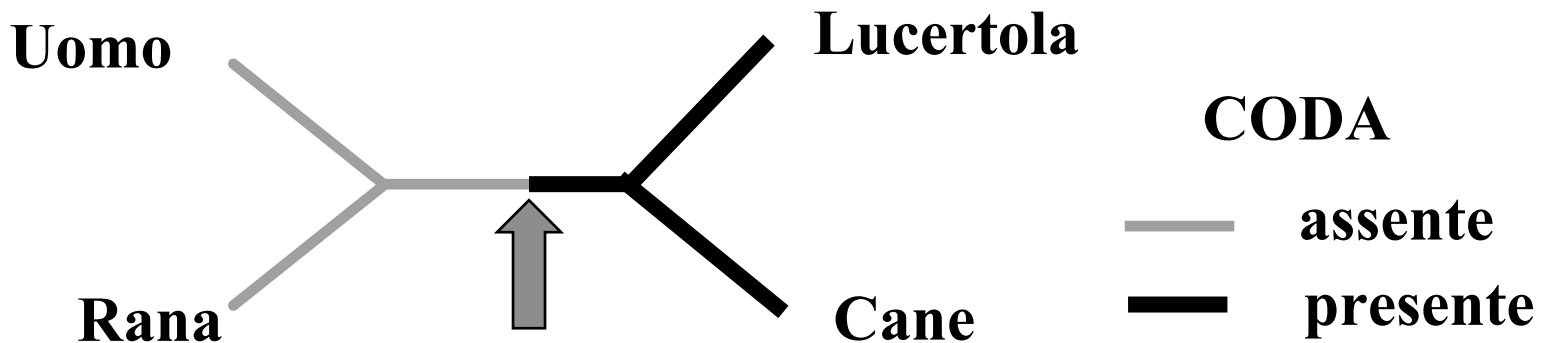
Omoplasia ed Incongruenza

- Se assumiamo che c'è un solo albero filogenetico corretto allora:**
- Quando i caratteri sostengono filogenesi in conflitto, sappiamo che ci deve essere una qualche evidenza di relazioni errata fra i caratteri incongruenti o incompatibili.**
- L'incongruenza fra due caratteri, significa che almeno uno dei caratteri è omoplasico e che almeno uno degli alberi sostenuto da quel carattere è sbagliato.**

Incongruenza o Incompatibilità



- Questi alberi e i caratteri sono incongruenti – non possono essere corretti entrambi gli alberi, almeno uno è sbagliato e almeno un carattere deve essere omoplasico.



Omologia e Omoplasia

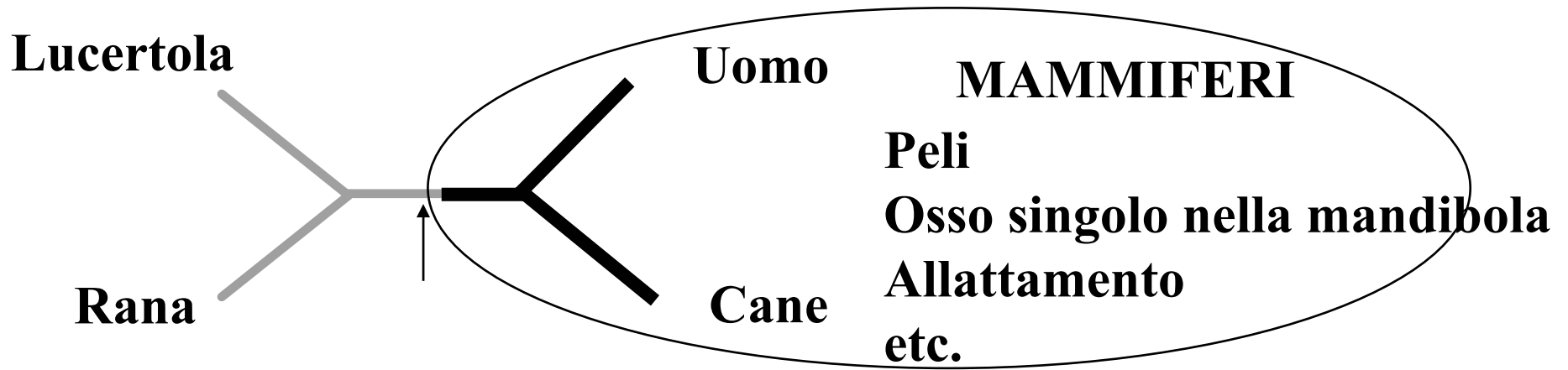
- **I morfologi usano molte tecniche per distinguere fra omoplasia e omologia.**
- **Caratteristiche omologhe sono attese mostrare similarità nei dettagli (nella posizione, nella struttura, nello sviluppo) mentre caratteristiche omoplasiche mostrano similarità più superficiali.**
- **Come detto da Charles Darwin la congruenza con altri caratteri fornisce la più completa evidenza di omologia.**

L'importanza della Congruenza

- **“The importance, for classification, of trifling characters, mainly depends on their being correlated with several other characters of more or less importance. The value indeed of an aggregate of characters is very evident a classification founded on any single character, however important that may be, has always failed.”**
- **Charles Darwin: Origin of Species, Ch. 13.**

Congruenza

- Il “vero” albero è sostenuto dalla congruenza di più caratteri.



Omoplasia nei dati molecolari

L'incongruenza e perciò l'omoplasia può essere comune nei dati di sequenze molecolari.

- C'è un numero limitato di stati del carattere alternativi (cioè, soltanto A, G, C e T nel DNA).**
- I tassi di evoluzione sono a volte alti.**

Gli stati del carattere sono chimicamente identici.

- Omologia e omoplasia sono egualmente simili.**
- Non possono essere distinte da studi dettagliati di similarità e differenza.**

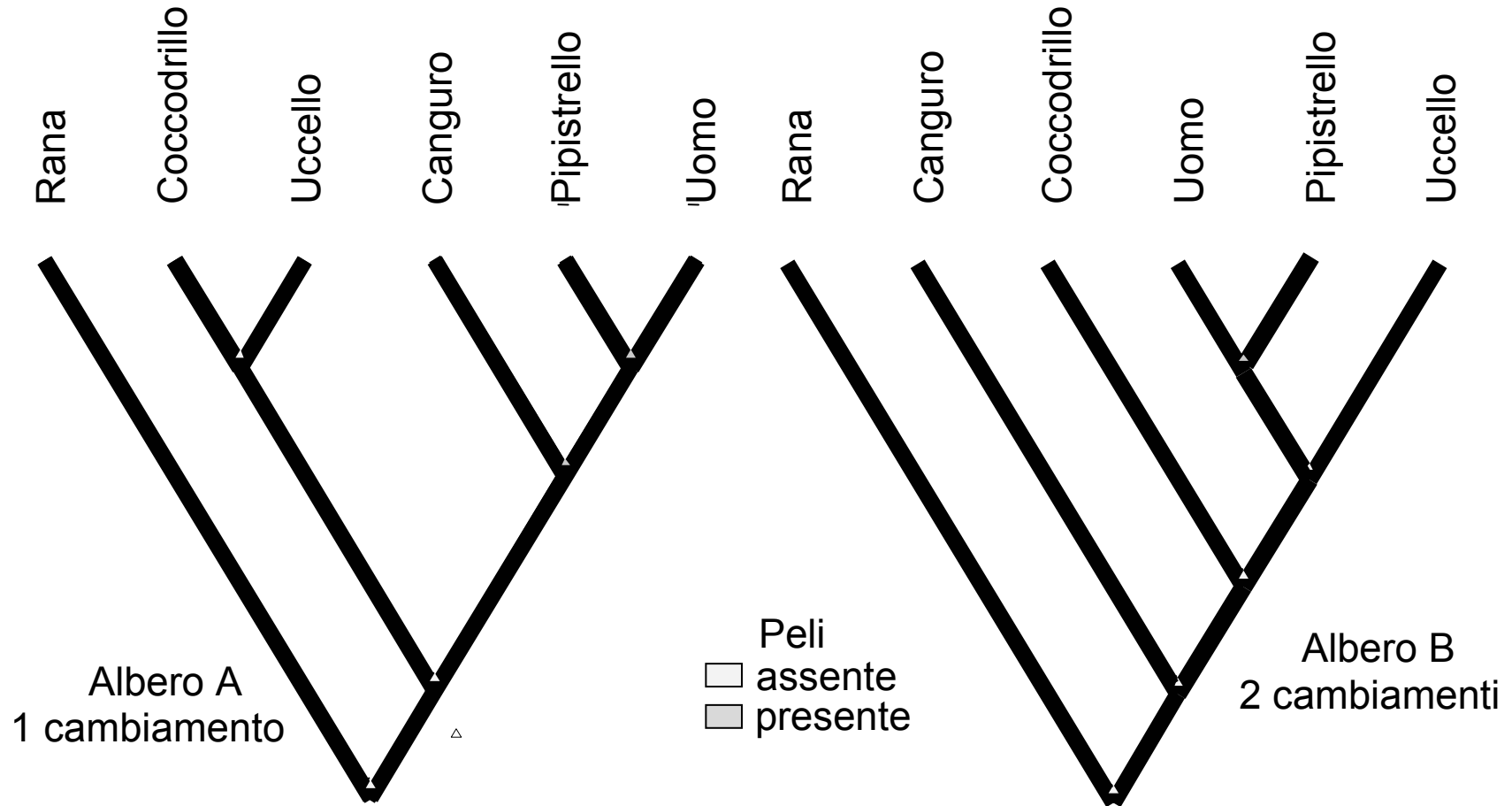
Analisi della Parsimonia

- **I metodi utilizzati dall'analisi della parsimonia forniscono un modo per scegliere fra ipotesi filogenetiche alternative.**
- **Il criterio su cui si basa la parsimonia favorisce ipotesi che massimizzano la congruenza e minimizzano l'omoplasia.**
- **Tale criterio dipende dalla corrispondenza fra un carattere ed un albero.**

Corrispondenza del Carattere

- Inizialmente, possiamo definire la corrispondenza fra un carattere ed un albero come il numero minimo di cambiamenti richiesti per spiegare la distribuzione osservata degli stati dei caratteri fra i taxa.
- Ciò è determinato dall'ottimizzazione del carattere parsimonioso.
- I caratteri differiscono per la corrispondenza con alberi diversi.

Corrispondenza del carattere

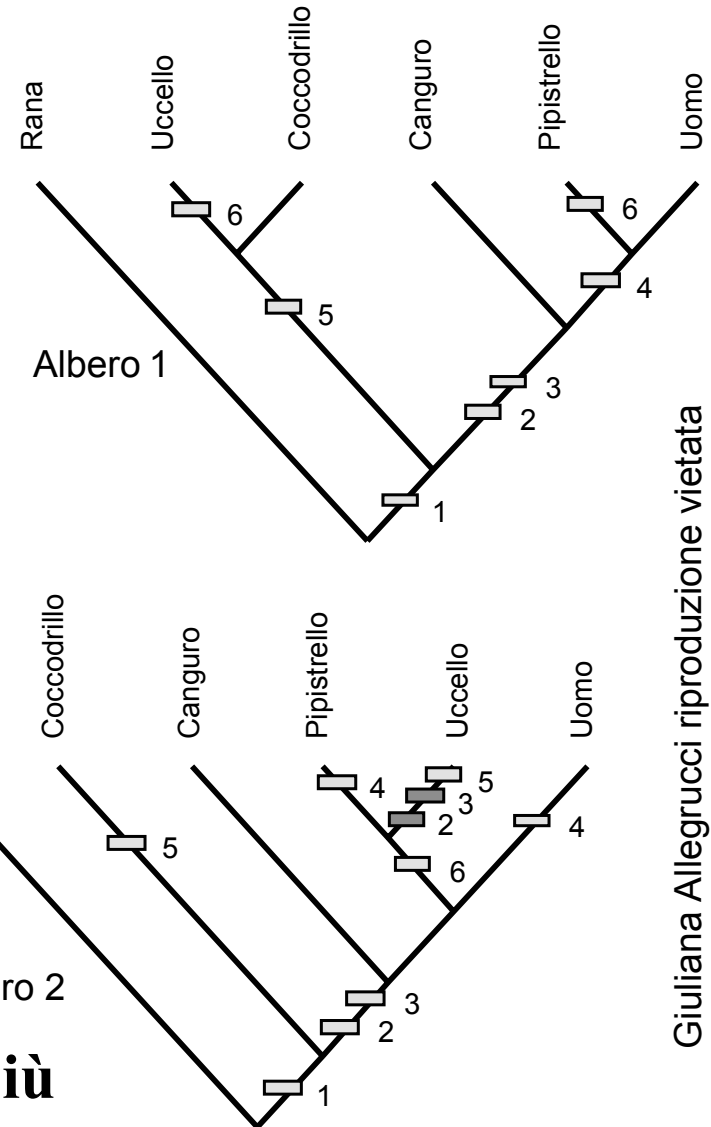


Analisi della Parsimonia

- Dato un gruppo di caratteri, come per esempio delle sequenze allineate, l'analisi della parsimonia lavora per determinare la corrispondenza (numero di cambiamenti) di ogni carattere su un determinato albero.
- La somma di tutti i caratteri è definita *Lunghezza dell'albero*.
- L'albero più parsimonioso ha la lunghezza minima necessaria a spiegare le distribuzioni osservate di tutti i caratteri.

Parsimonia in pratica

		CARATTERI						
		1	2	3	4	5	6	
		amnion	peli	allattamen	placenta	finestra antiorbitale	ali	
TAXA	Rana	-	-	-	-	-	-	
	Uccello	+	-	-	-	+	+	
	Coccodrillo	+	-	-	-	+	-	
	Canguro	+	+	+	-	-	-	
	Pipistrello	+	+	+	+	-	+	
	Uomo	+	+	+	+	-	-	
Corrispondenza	Tree 1	1	1	1	1	1	2	7
	Tree 2	1	2	2	2	2	1	10



Di questi due alberi Albero 1 ha la lunghezza più corta ed è il più parsimonioso.

Entrambi gli alberi presentano alcune omoplasie.

Risultati dell'analisi di parsimonia

- **Uno o più alberi egualmente parsimoniosi.**
- **Ipotesi dell'evoluzione del carattere associata con ogni albero (dove e come i cambiamenti sono avvenuti).**
- **Lunghezza dei rami (quantità di cambiamento associata ai rami).**
- **Diverse statistiche dell'albero e del carattere descrivono la corrispondenza fra albero e dati.**
- **Alberi sub-ottimali: opzionale.**

Tipi di carattere

- I caratteri possono differire nei costi (ovvero contributo alla lunghezza dell'albero) sulla base dei diversi tipi di cambiamenti.

- Wagner (ordinati, additivi)

0 — 1 — 2 (morfologia, costi diversi)



- Fitch (non ordinati, non-additivi)

A — G (morfologia, molecole)



T — C

(costi eguali per tutti i cambiamenti)

— un cambiamento
— due cambiamenti

Tipi di Caratteri

- **Sankoff (generalizzato).**

A—G (morfologia, molecole).

T—C (pesi specificati dall'utilizzatore).

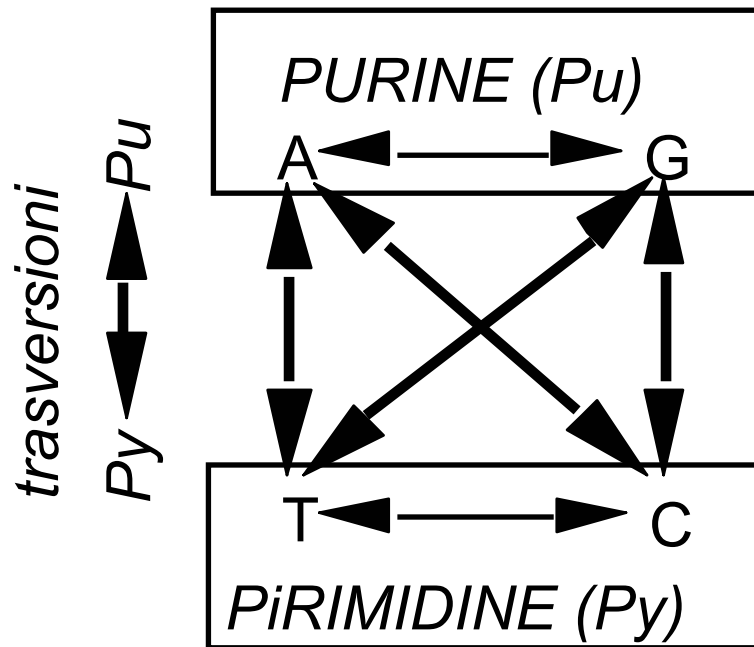
— un cambiamento

— cinque cambiamenti

- **Esempio, peso diverso per le transizioni e le trasversioni.**
- **I pesi sono specificati in una matrice di cambiamenti (*stepmatrix*).**
- **I pesi sono di solito simmetrici ma possono essere anche asimmetrici (per es., costa più ottenere un sito di restrizione che perderlo).**

Stepmatrices

- Una stepmatrix specifica i costi dei cambiamenti all'interno di un carattere



transizioni
 $Py \rightarrow Py$
 $Pu \rightarrow Pu$

Da

	A			
	A	C	G	T
A	0	5	1	5
C	5	0	5	1
G	1	5	0	5
T	5	1	5	0

Caratteri differenti (cioè, 1^a, 2^a e 3^a posizione nel codone) possono avere anche pesi diversi

Parsimonia pesata

- **Se tutti i tipi di cambiamenti di tutti i caratteri hanno egual peso, allora la parsimonia:**
 - **Minimizza l'omoplasia**
 - **Massimizza la quantità di similarità dovuta ad origine comune**
 - **Minimizza la lunghezza dell'albero**
- **Se i cambiamenti sono pesati in modo diverso, la parsimonia minimizza la lunghezza dell'albero che risulta essere una somma pesata del costo di ogni carattere.**

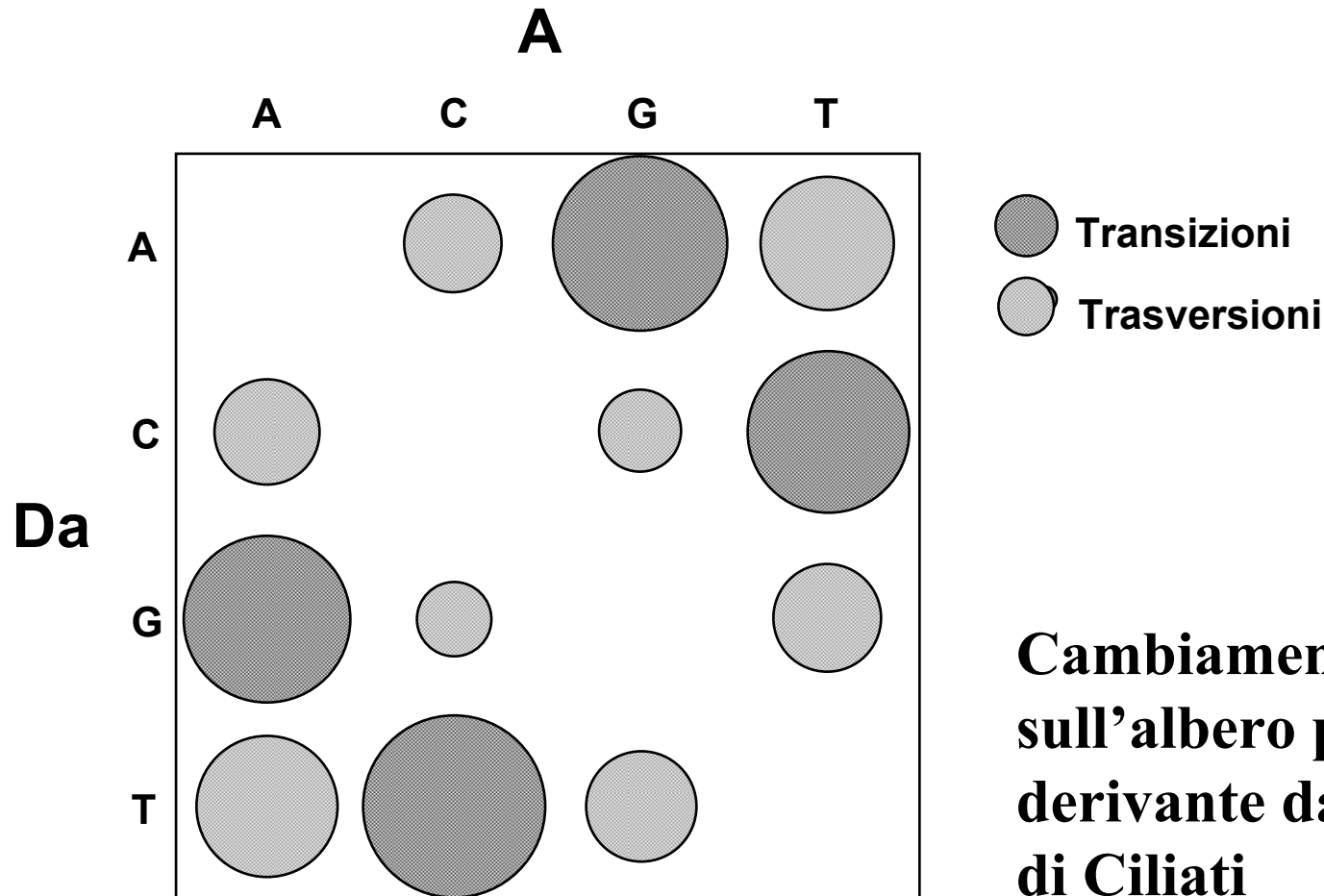
Perchè pesare i caratteri?

Molti sistematici considerano inaccettabile il peso, ma il peso spesso è inevitabile:

- **Le transizioni possono essere più comuni delle trasversioni**
- **Tipi differenti di transizioni e trasversioni possono essere più o meno comuni**
- **I tassi di cambiamento possono variare con la posizione sul codone**
- **La corrispondenza dei caratteri su alberi differenti può indicare differenze nella loro attendibilità**

Comunque, il considerare pesi eguali è la procedura più comune e più semplice, anche se probabilmente non è la migliore

I diversi tipi di cambiamento differiscono in frequenza



**Cambiamenti non ambigui
sull'albero più parsimonioso
derivante da dati SSUrDNA
di Ciliati**

Vantaggi della Parsimonia

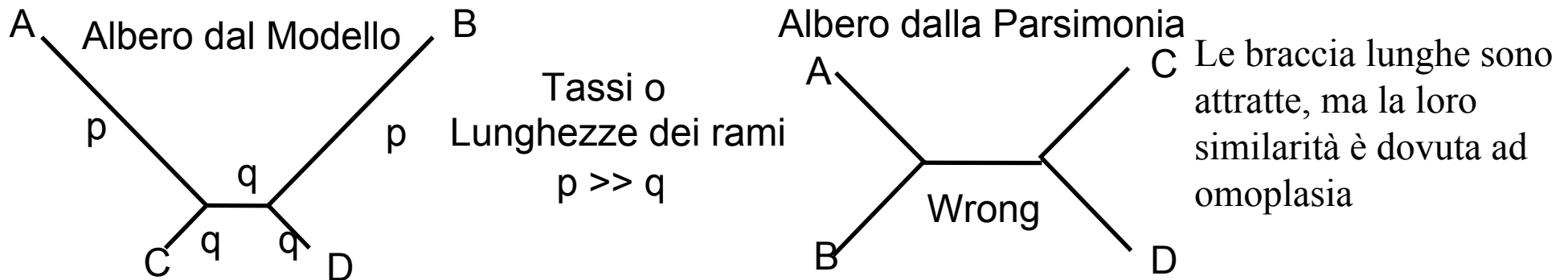
- **E' un metodo semplice – facilmente comprensibile**
- **Non sembra dipendere da un modello di evoluzione esplicito**
- **Restituisce sia gli alberi che le ipotesi associate all'evoluzione del carattere**
- **Dovrebbe dare risultati attendibili se i dati sono ben strutturati e l'omoplasia è o rara o distribuita a caso sull'albero**

Svantaggi della Parsimonia

- **Può dare risultati sbagliati se l'omoplasia è comune o concentrata in parti particolari dell'albero, per es.:**
 - **convergenza termofilica**
 - **errore nella composizione in basi**
 - **attrazione dei rami lunghi**
- **Sottostima le lunghezze dei rami**
- **Il modello di evoluzione è implicito**
- **La parsimonia è spesso giustificata su base puramente filosofica- noi dobbiamo preferire le ipotesi più semplici – soprattutto dai morfologi**
- **Per la maggior parte dei sistematici molecolari questo non è convincente**

La Parsimonia può essere inconsistente

- Felsenstein (1978) ha sviluppato un semplice modello di filogenesi comprendente quattro taxa e un miscuglio di rami corti e lunghi. I risultati della parsimonia su questa base sono sbagliati

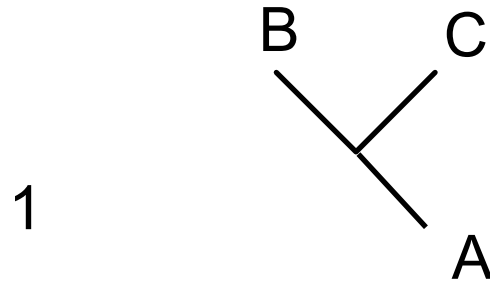


- Più aumentano i dati e più aumenta la certezza di risultati sbagliati con la parsimonia, quindi la parsimonia è statisticamente inconsistente
- I sostenitori della parsimonia risposero dichiarando che i risultati di Felsenstein dimostravano che il suo modello era irrealistico
- E' ora riconosciuto che l'attrazione dei rami lunghi è uno dei più seri problemi negli studi filogenetici

Ricerca degli alberi ottimali – soluzioni esatte

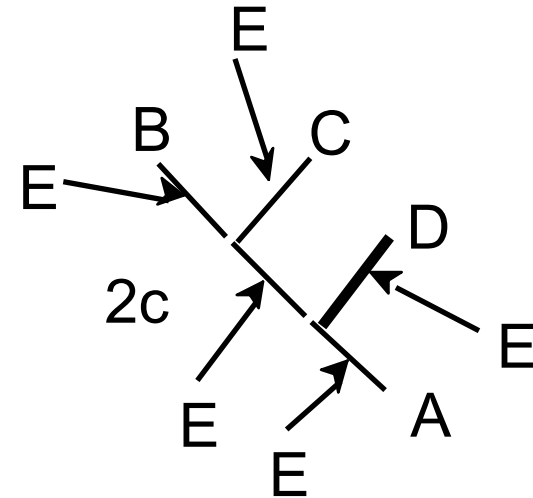
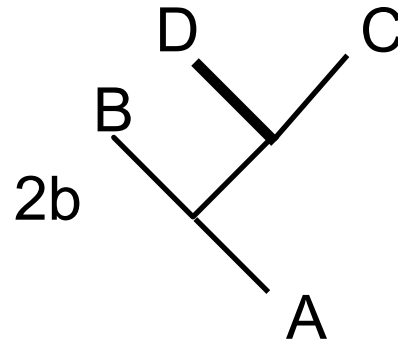
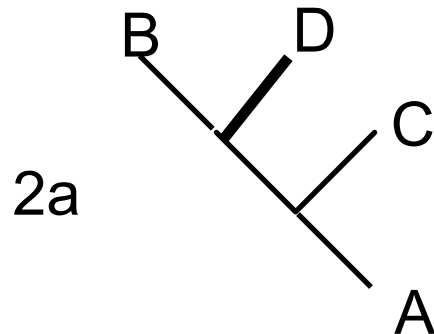
- **Le soluzioni esatte possono essere usate soltanto per un piccolo numero di taxa**
- **Exhaustive search esamina tutti i possibili alberi**
- **Normalmente è usata per problemi relativi a meno di 10 taxa**

Ricerca degli alberi ottimali - exhaustive search



Albero iniziale, 3 taxa qualsiasi

Aggiunge il quarto taxon (D) in ognuna delle tre possibili posizioni -> tre alberi

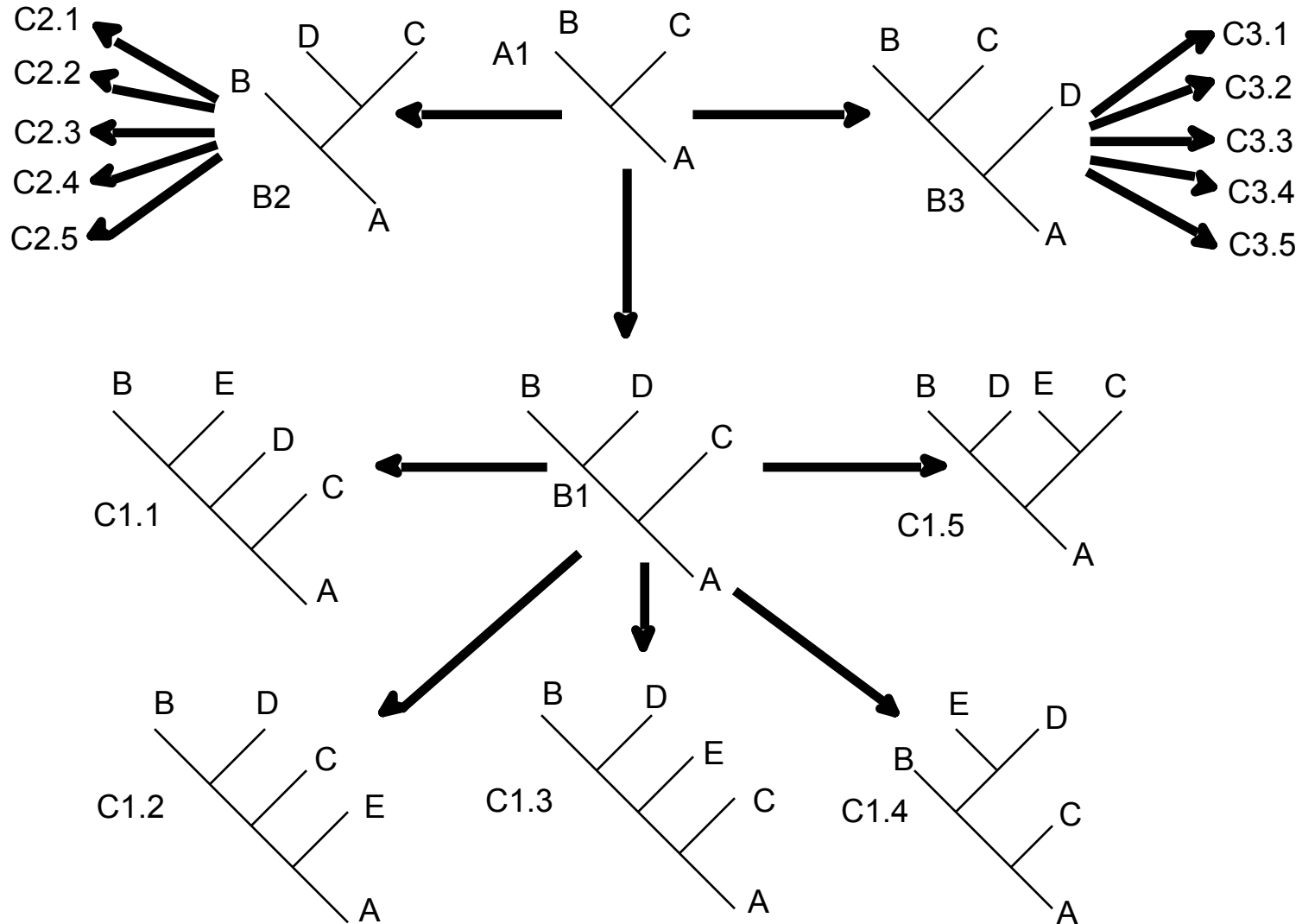


Aggiunge il quinto taxon (E) in ognuna delle cinque possibili posizioni su ognuno dei tre alberi -> 15 alberi, e così via

Ricerca degli alberi ottimali – soluzioni esatte

- La ricerca *branch and bound* risparmia tempo scartando, durante la costruzione dell'albero, le famiglie di alberi che non possono essere più corte dell'albero più corto trovato.
- Può essere migliorata specificando un limite iniziale superiore per la lunghezza dell'albero.
- Solitamente si usa per problemi con meno di 18 taxa.

Ricerca degli alberi ottimali -branch and bound



Ricerca degli alberi ottimali – heuristic search

- Il numero dei possibili alberi aumenta esponenzialmente con il numero di taxa. Questo fa sì che la *exhaustive search* sia di fatto impraticabile per set di dati grandi.
- I metodi euristici sono usati per cercare lo *spazio* albero per gli alberi più corti, costruendo o selezionando un albero iniziale e scambiando i rami per cercare l'albero migliore.
- Gli alberi trovati non sono garantiti essere i più parsimoniosi, sono i più *probabilmente parsimoniosi*.

Ricerca degli alberi ottimali - heuristic search

- **Stepwise addition**

Asis – l'ordine nella matrice di dati

Closest – inizia con l'albero più corto costituito da tre taxa, aggiunge taxa in un ordine che produce il minore aumento nella lunghezza dell'albero (*greedy heuristic*)

Simple – Il primo taxon della matrice è preso come riferimento – i taxa sono aggiunti in ordine di similarità al taxon di riferimento

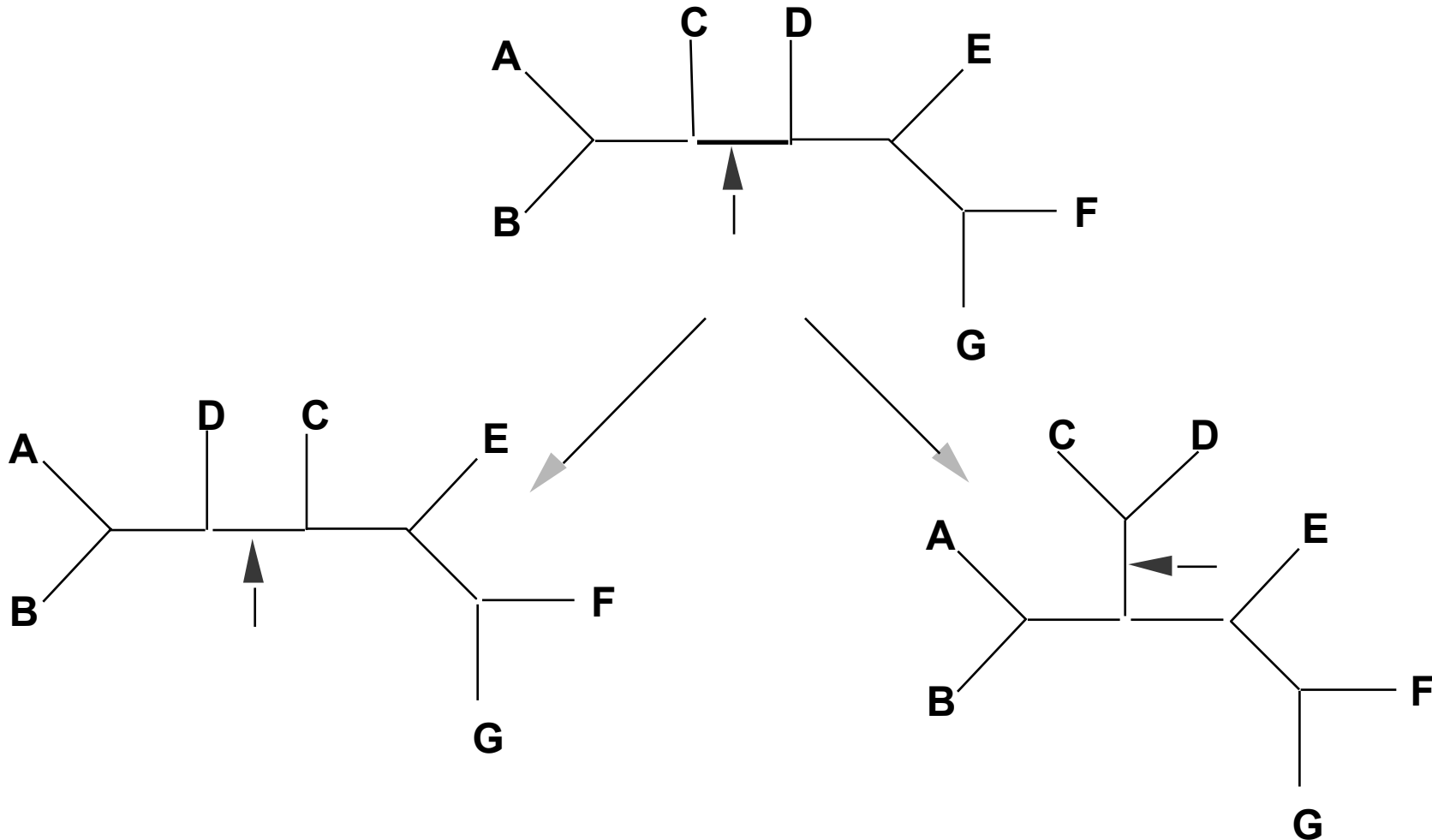
Random – I taxa vengono aggiunti a una sequenza presa a caso, possono essere usate molte sequenze differenti

Ricerca degli alberi più parsimoniosi – heuristic search

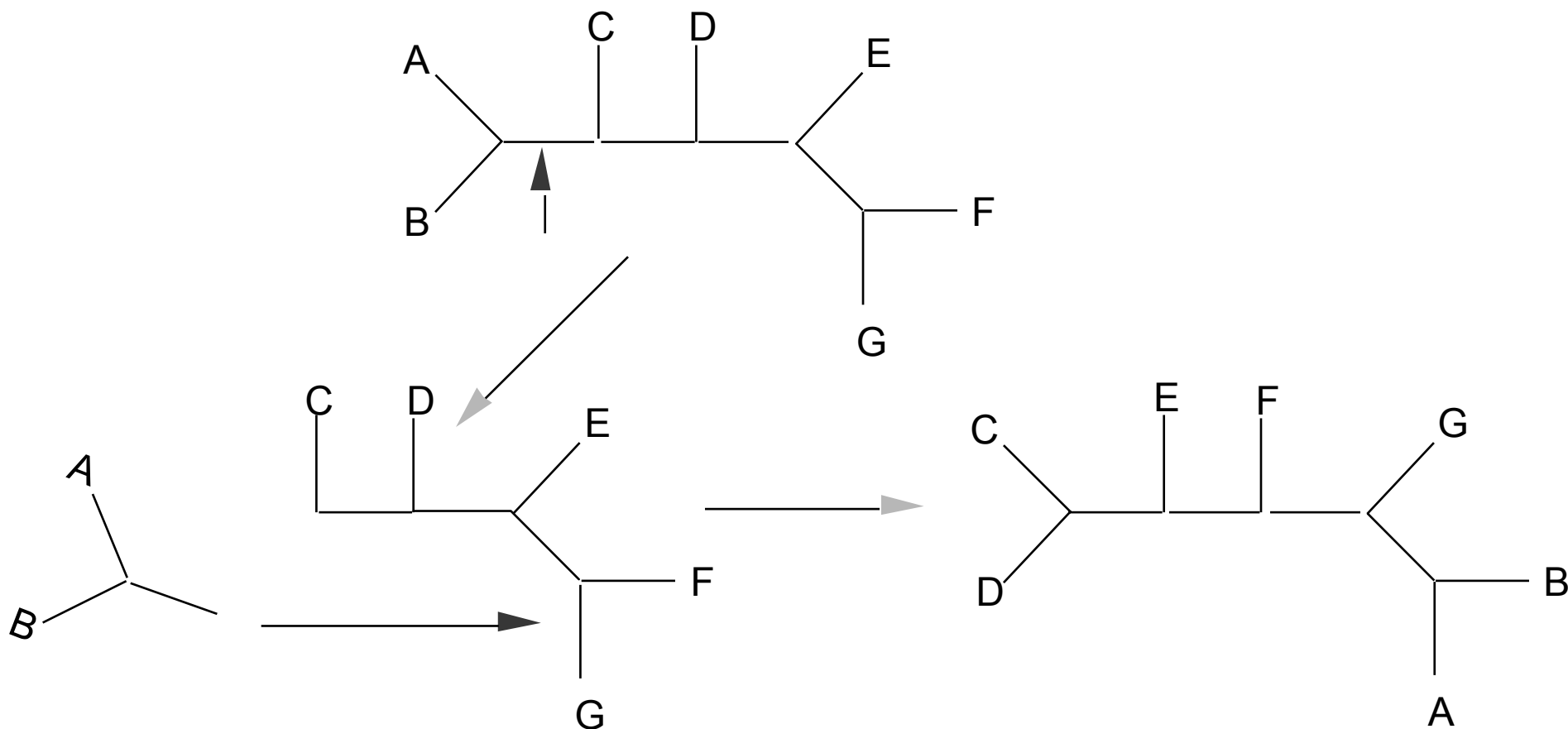
- **Scambio dei rami (*Branch Swapping*):**
 - Scambio fra i vicini più vicini (NNI)**
 - Taglio degli alberi secondari e ricostruzione (SPR)**
 - Bisezione dell'albero e ricongiungimento (TBR)**
- Altri metodi**

Ricerca degli alberi più parsimoniosi – heuristic search

- Scambio fra i vicini più vicini (NNI)

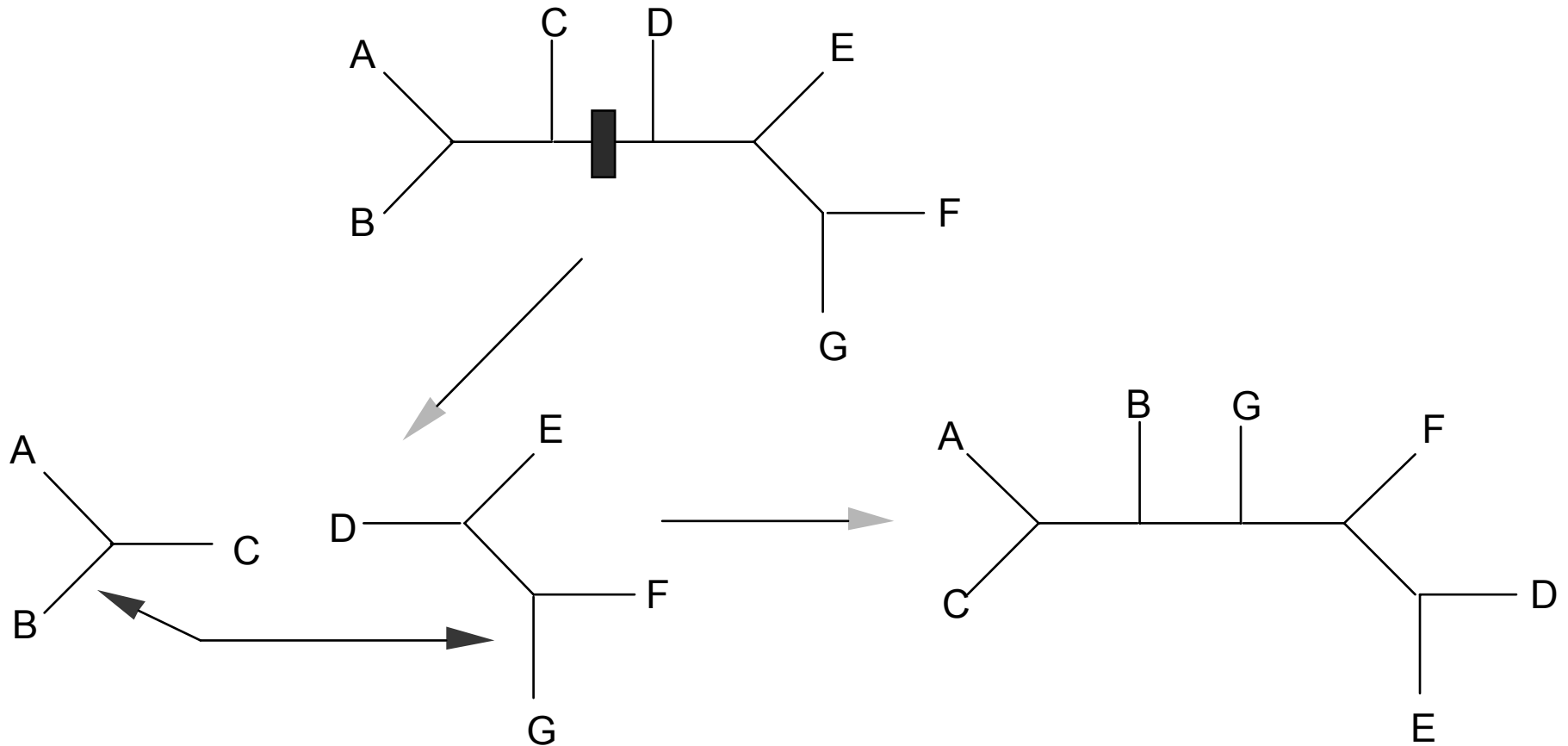


- **Taglio degli alberi secondari e ricostruzione (SPR)**



Ricerca degli alberi più parsimoniosi – heuristic search

- Bisezione dell'albero e ricongiungimento (TBR)



Ricerca degli alberi più parsimoniosi –heuristic search

- **Scambio dei rami**

Nearest neighbor interchange (NNI)

Subtree pruning and redrafting (SPR)

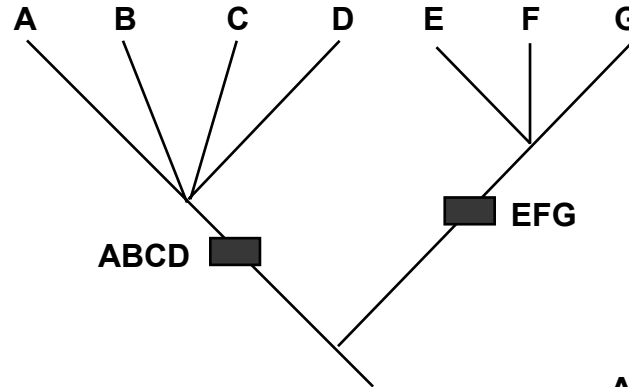
Tree bisection and reconnection (TBR)

- **Insito nella natura della ricerca euristica è il fatto che non possiamo sapere quale è il metodo che produrrà l'albero più parsimonioso**
- **TBR è il metodo che utilizza la routine più ampia di sostituzione e il suo utilizzo con l'addizione multipla di sequenze a caso dovrebbe funzionare bene.**

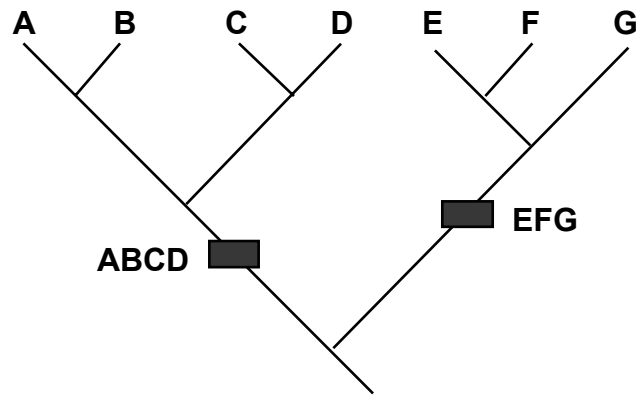
Ricerca con vincoli topologici

- **I vincoli topologici sono ipotesi filogenetiche definite dall'utilizzatore**
- **Utilizzo:**
 - 1. Comprende un raggruppamento specifico o una serie di relazioni**
 - 2. Esclude un raggruppamento specifico o una serie di relazioni**

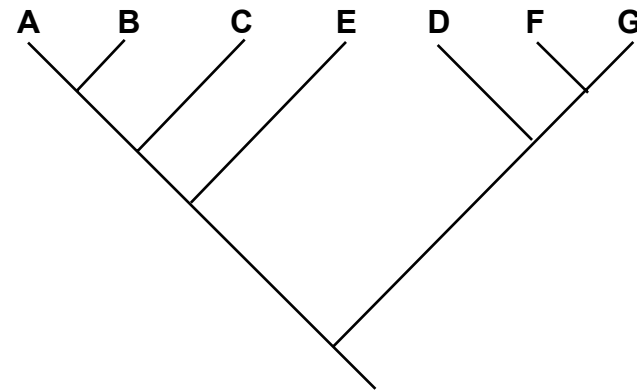
Ricerca con vincoli topologici



ALBERO VINCOLATO
((A,B,C,D)(E,F,G))



Compatibile



Incompatibile

Alberi ottimali multipli

- **Molti metodi possono produrre alberi multipli egualmente ottimali**
- **Fra questi alberi può essere operata una selezione usando ulteriori criteri, ma**
- **Tipicamente, le relazioni comuni a tutti gli alberi ottimali, vengono riassunte in un *albero di consenso***

Metodi di Consenso

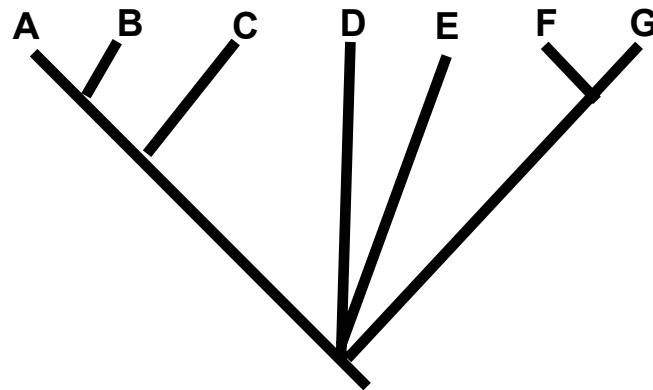
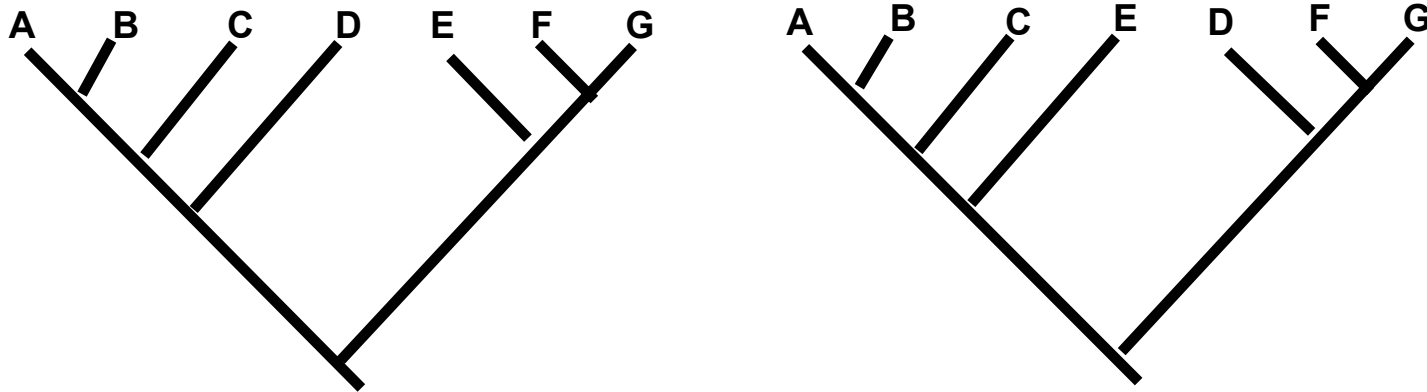
- **Un albero di consenso riassume la concordanza fra un gruppo di alberi primari**
- **Esistono diversi metodi di consenso che differiscono per:**
 1. **Il *tipo* di concordanza**
 2. **Il *livello* di concordanza**
- **I metodi di consenso possono essere usati con alberi multipli derivanti da una singola analisi o da più analisi**

Metodi di consenso esatti (*Strict*)

- **Tali metodi richiedono concordanza fra tutti gli alberi primari**
- **Mostrano solo le relazioni che sono sostenute in modo non ambiguo dall'interpretazione parsimoniosa dei dati**
- **Il metodo più comune (*strict component consensus*) focalizza l'attenzione su gruppi/componenti/completa divisione**
- **Produce un albero di consenso che comprende le divisioni trovate in tutti gli alberi primari**
- **Le altre relazioni (quelle per cui gli alberi primari non concordano) vengono mostrate come politomie non risolte**

Metodi di consenso esatti (*Strict*)

DUE ALBERI PRIMARI



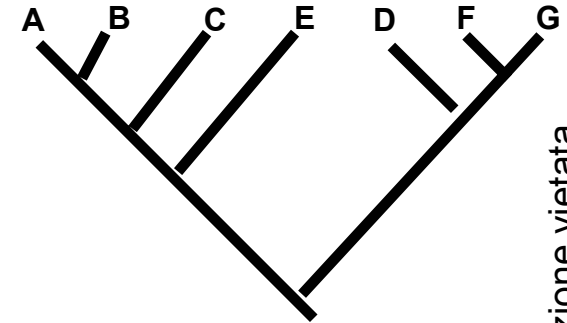
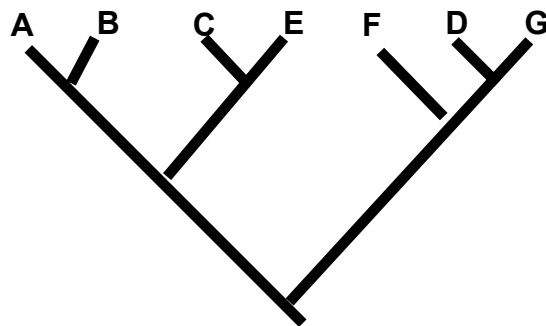
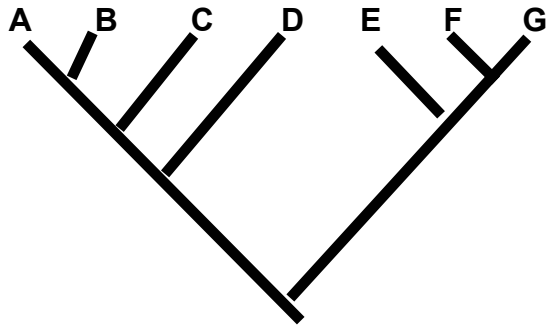
STRICT COMPONENT CONSENSUS TREE

Metodi di consenso – regola della maggioranza

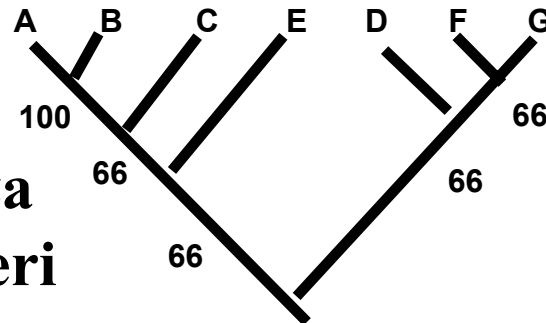
- Tali metodi richiedono concordanza fra la maggioranza degli alberi primari
- Possono comprendere relazioni non sostenute dall'interpretazione più parsimoniosa dei dati
- Il metodo più comune si concentra su gruppi/componenti/divisioni complete
- Produce un albero di consenso che comprende tutte le divisioni trovate nella maggior parte (>50%) degli alberi primari
- Le altre relazioni sono mostrate come politomie non risolte

Majority rule consensus

TRE ALBERI PRIMARI



I numeri indicano la frequenza
dei raggruppamenti negli alberi
primari



MAJORITY-RULE COMPONENT CONSENSUS TREE