

Lezione 8: La filogenesi dei batteri – 2° parte

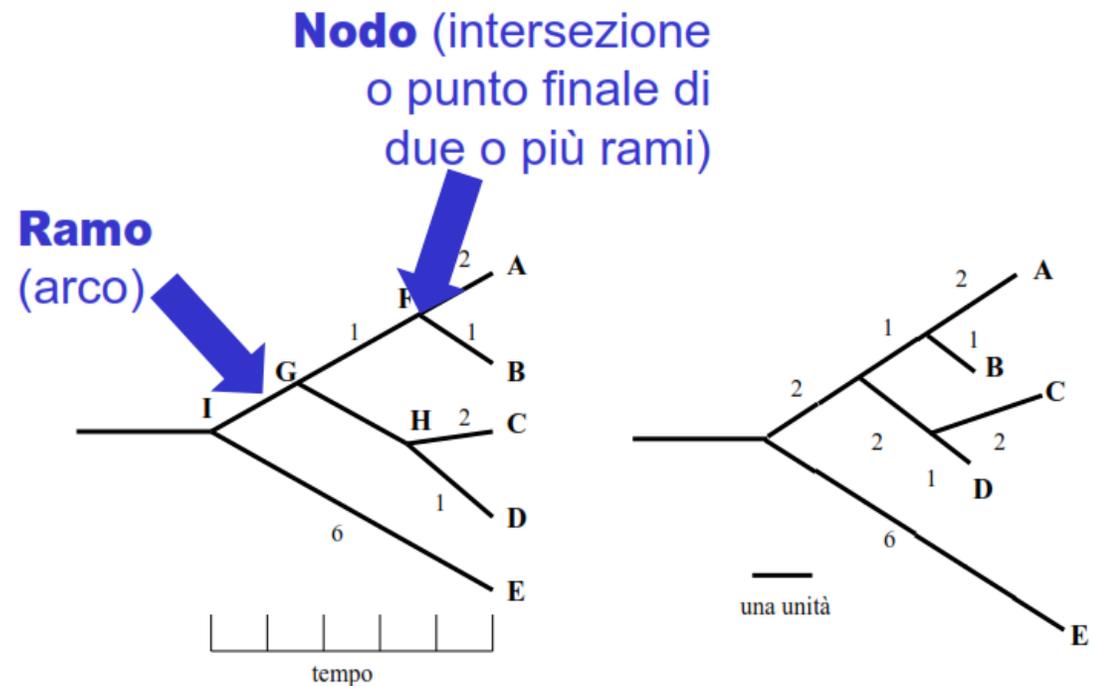
DOTT. ROSA ANNA NASTRO – STANZA 425° - 4°PIANO, LATO NORD
EMAIL: ROSA.NASTRO@UNIPARTHENOPE.IT

ALBERI FILOGENETICI

La filogenesi molecolare è usata per tracciare gli alberi evolutivi delle relazioni tra organismi. Questi alberi sono basati su dati di sequenza nucleotidici o proteici.

Un albero filogenetico è un diagramma che disegna la storia evolutiva di un organismo e di fatto somiglia ad un albero genealogico. Esso è definito da due informazioni: topologia e lunghezze dell'albero (somma della lunghezza dei rami).

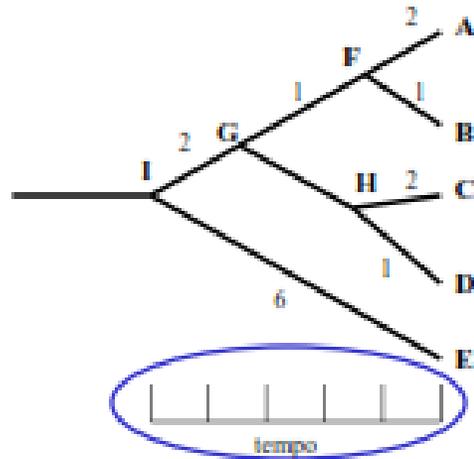
Un albero è un grafico formato da nodi e rami.



Gli alberi, se opportunamente disegnati, danno informazioni sulle distanze e sui tempi evolutivi.

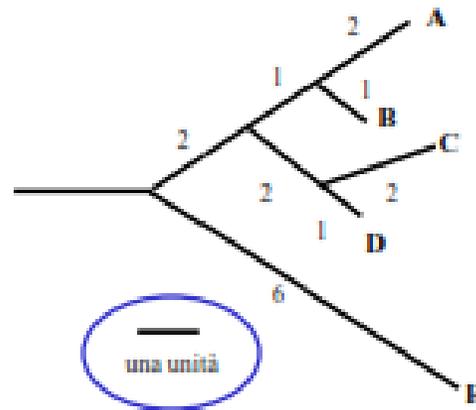
Nomenclatura degli alberi

I rami sono non in scala ...



... I nodi riflettono il tempo

I rami sono in scala ...



...La lunghezza dei rami è proporzionale al numero di aminoacidi che cambiano

Le estremità dei rami rappresentano le specie odierne.

I nodi rappresentano un passaggio dell'evoluzione in corrispondenza del quale da un antenato comune hanno avuto origine due linee separate

Sono due rappresentazioni alternative, dipende da cosa si vuole comunicare...

La rotazione intorno al nodo non cambia la topologia dell'albero.

Da uno stesso numero di sequenze è possibile ottenere più alberi filogenetici: per esempio da otto sequenze si possono ottenere 2×10182 alberi filogenetici!

L'analisi filogenetica utilizza i dati molecolari per provare ad identificare l'unico albero corretto che rappresenti in modo più accurato la storia evolutiva di un gruppo di sequenze.

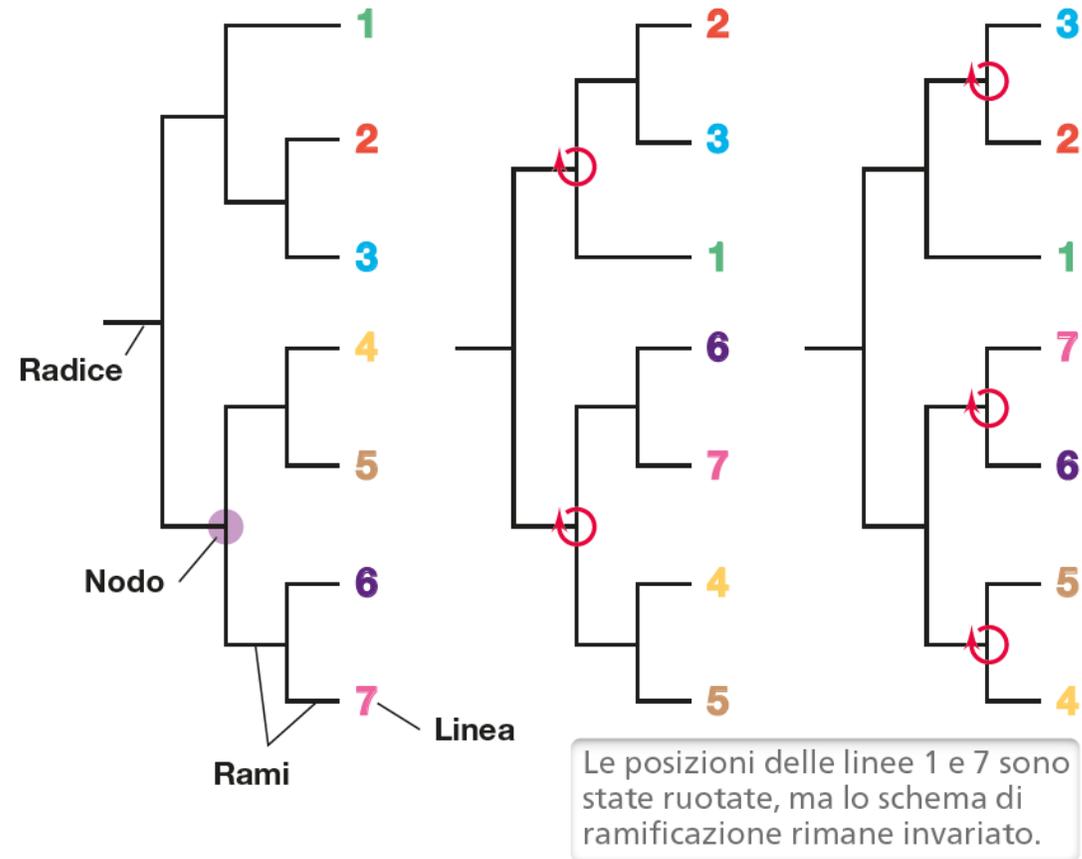


Figura 13.27 Alberi filogenetici e loro interpretazione.

Tre versioni equivalenti dello stesso albero filogenetico. L'unica differenza tra i tre alberi è rappresentata dalla rotazione intorno al punto indicato dalle frecce rosse. La posizione verticale delle specie è differente nei vari alberi, ma lo schema di discendenza (i nodi condivisi da ogni specie) rimane invariato.

1. Il primo passo per costruire un albero è l'allineamento delle sequenze.

2. Viene elaborata una matrice di distanza, sulla base del numero di differenze di sequenza.

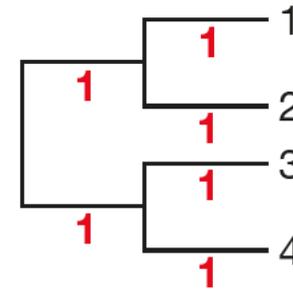
3. L'albero è costruito aggiungendo nodi per legare i rami che presentano le differenze minori.

1 A C T G A C
 2 A C T C A T
 3 A C A T G G
 4 A C A A G A

(a)

	1	2	3	4
1	-			
2	2	-		
3	4	4	-	
4	4	4	2	-

(b)



(c)

Figura 13.28 Costruzione degli alberi filogenetici. Il numero delle differenze nucleotidiche tra le sequenze geniche può essere utilizzato per costruire un albero filogenetico. Nell'allineamento di sequenza (a) possiamo contare il numero delle differenze tra ogni coppia di sequenze per costruire una matrice di distanza. (b) Questa matrice di distanza può essere utilizzata per costruire un albero (c) dove le lunghezze cumulative dei rami orizzontali (segnati con un "1" rosso) tra due specie qualsiasi sono proporzionali al numero delle differenze nucleotidiche tra le specie in esame.

Limiti degli alberi filogenetici

E' possibile valutare se un dato albero filogenetico è probabilmente il più rappresentativo dell'evoluzione di una data specie applicando un metodo statistico, il *bootstrapping*. Si immagina al di effettuare un gran numero (ordine del centinaio) di allineamenti simulati e di costruire per ciascuno di essi un albero filogenetico, in modo da avere alla fine un numero di alberi pari al numero di multiallineamenti simulati effettuati. A questo punto si cerca di ricostruire l'albero consenso: si calcola per ciascun nodo la percentuale di alberi simulati in cui viene riprodotto lo stesso nodo. Due alberi riproducono lo stesso nodo quando esso condivide le stesse unità tassonomiche discendenti. La percentuale così calcolata viene definita il valore di bootstrap del nodo considerato. Un alto valore di bootstrap corrisponde ad un'alta significatività statistica del nodo. Normalmente si applica la «majority rule consensus», vale a dire la regola per cui un nodo viene considerato significativo se presenta almeno il 50% di bootstrap.

Questo perché è possibile ottenere un numero considerevole di alberi filogenetici, soprattutto se l'analisi coinvolge più alleli: per esempio se per 4 sequenze arbitrarie si possono ottenere solo tre alberi, con 8 sequenze se ne possono ottenere 10395!

Limiti degli alberi filogenetici

Omoplasia o evoluzione convergente si verifica quando gli organismi condividono un carattere che non è stato ereditato da un antenato comune (per esempio le ali negli uccelli e negli insetti)

In genetica, l'omoplasia si ottiene quando due sequenze geniche risultano essere uguali a causa di mutazioni casuali occorse durante il processo evolutivo: in questo modo l'omoplasia può «mascherare» le distanze evolutive tra le specie e, quindi, portare alla formulazione di alberi filogenetici meno accurati.

Allo stesso modo, il trasferimento genico orizzontale (TGO), attraverso lo scambio genico tra individui appartenenti a specie differenti, può essere alla base della presenza di sequenze uguali in genomi tra organismi non correlati filogeneticamente.

E' importante distinguere la filogenesi molecolare che ricostruisce l'evoluzione di un singolo gene dalla filogenesi dell'organismo che illustra la storia evolutiva di un entità biologica.

Le mutazioni ricorrenti possono cancellare le informazioni evolutive, determinando differenze di sequenza in grado di far sottostimare le distanze reali.

1 A C T G
 ↓
 2 A G T G
 ↓ ↓
 3 A G A C
 ↓
 4 C G A C

Vengono osservate quattro differenze tra le sequenze 1 e 4.

1 A C T G
 ↓
 2 A G T G
 ↓ ↓ Mutazioni di sostituzione
 3 A A T C
 ↓
 4 A C T C

Viene osservata una sola differenza tra le sequenze 1 e 4.

(a)



(b)

Figura 13.29 Problema dell'omoplasia dovuta a mutazioni ricorrenti. È possibile che mutazioni ricorrenti arrivino a nascondere il reale numero di mutazioni intervenute a carico di due sequenze, a partire dal momento in cui queste condividevano un antenato comune. (a) Confronto tra due serie di mutazioni durante l'evoluzione di una sequenza genica. Sul lato sinistro, il numero di mutazioni è uguale al numero osservato tra le specie 1 e 4. Nel caso in cui intervenga una mutazione ricorrente (lato destro), il numero delle mutazioni osservate tra le specie 1 e 4 può risultare inferiore al numero delle mutazioni realmente avvenute. (b) La probabilità di mutazioni ricorrenti aumenta progressivamente con l'accumulo di mutazioni nel tempo.

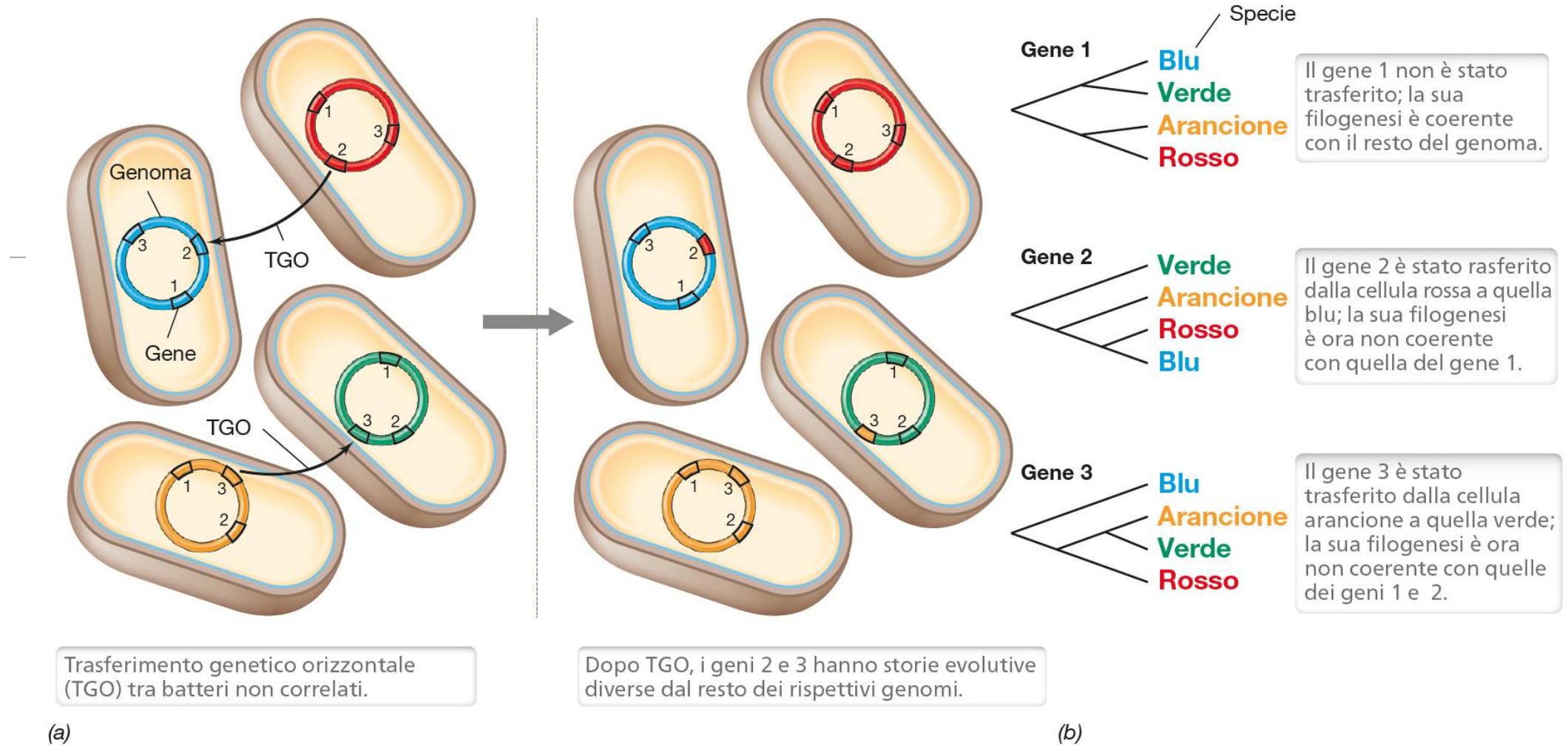
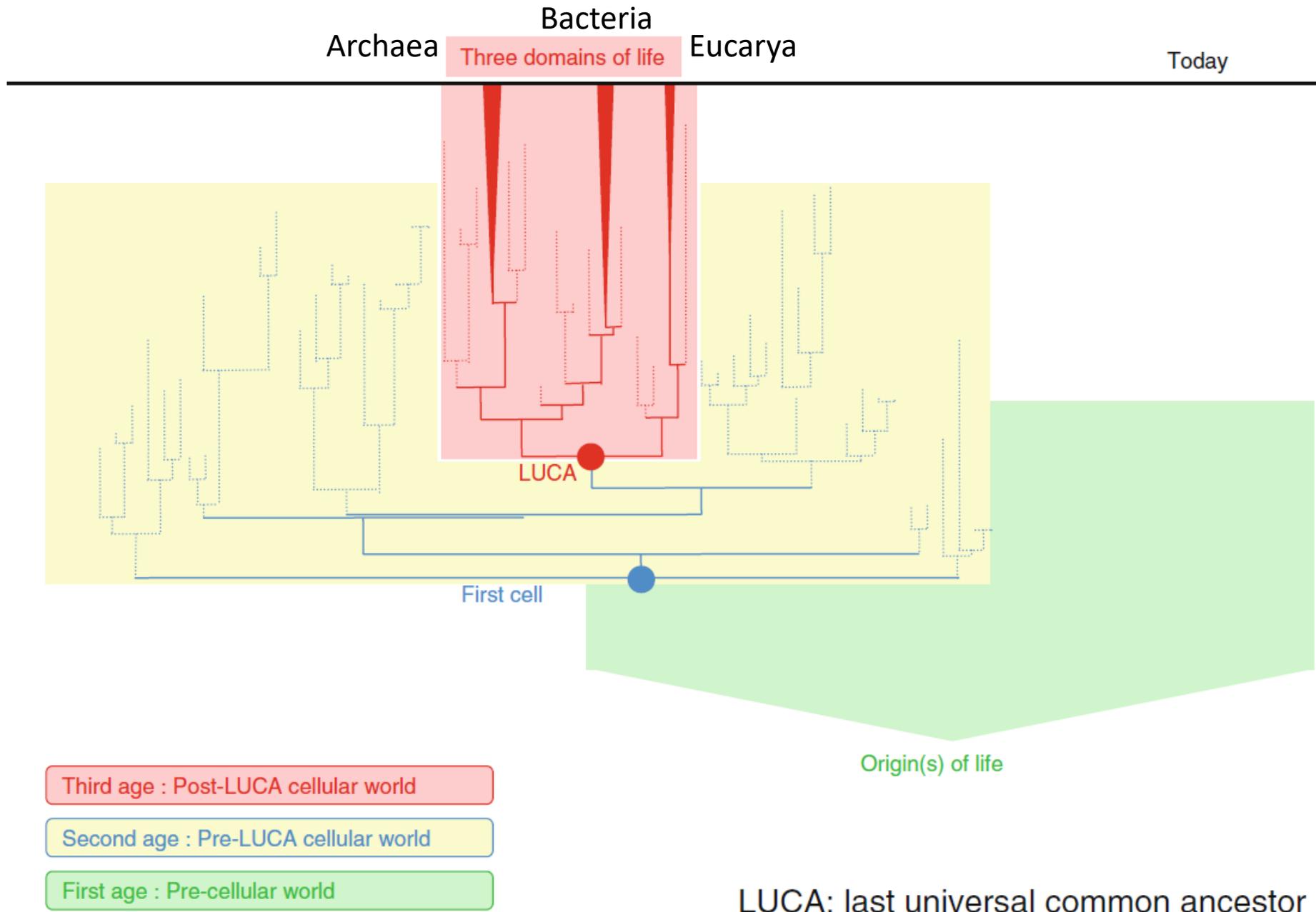


Figura 13.30 Trasferimento genico orizzontale. La storia evolutiva di un gene scambiato per trasferimento orizzontale può essere diversa da quella del resto del genoma. (a) I geni vengono trasferiti orizzontalmente tra microrganismi geneticamente distanti. I diversi colori sono utilizzati per identificare i microrganismi e i rispettivi genomi. (b) A causa del trasferimento orizzontale mostrato in (a) si ottengono differenti alberi filogenetici per i geni 1, 2 e 3. Solo l'albero del gene 1, che non è stato trasferito, rimane coerente con la filogenesi dell'organismo.

Ricostruzione della filogenesi di batteri, archea ed eukarya: meccanismi,
ipotesi e prove.

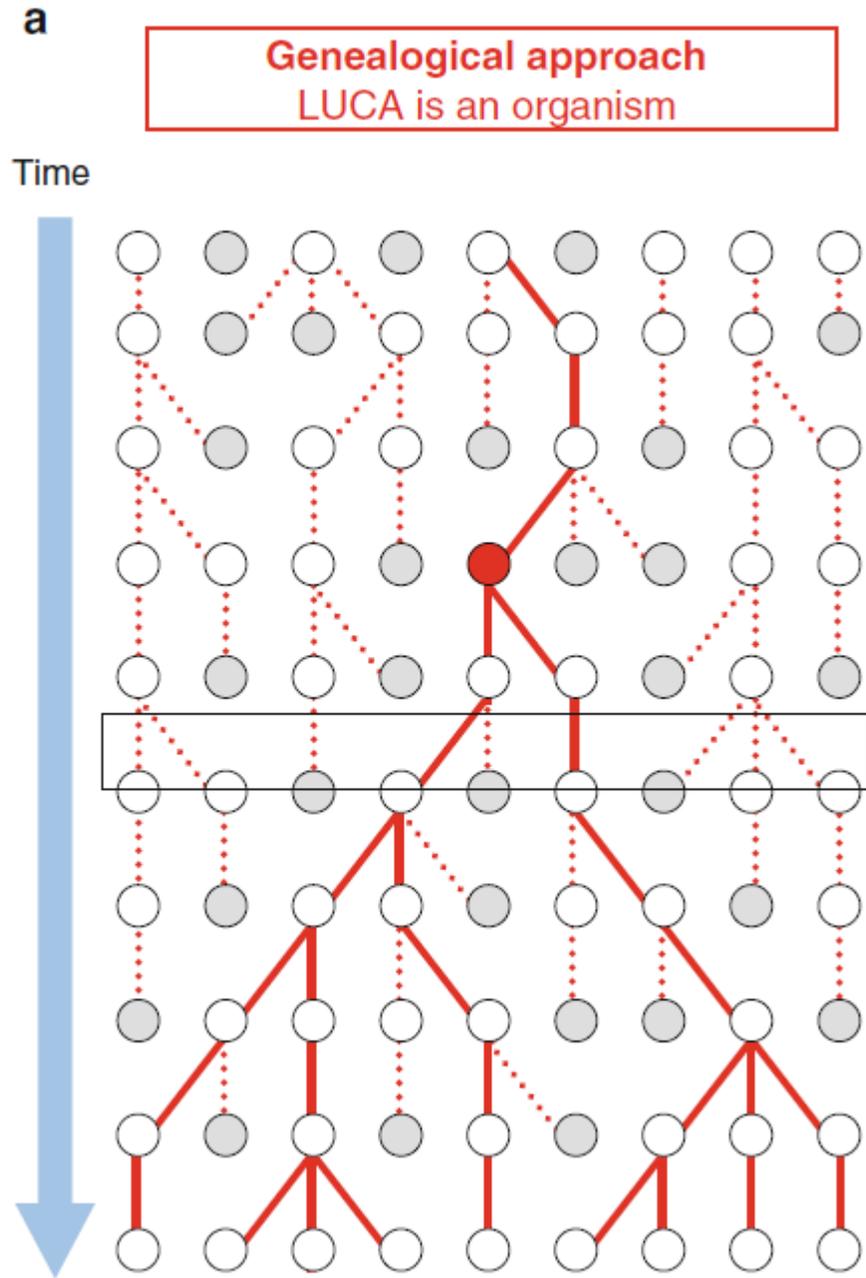
Il Last Universal Common Ancestor (LUCA)



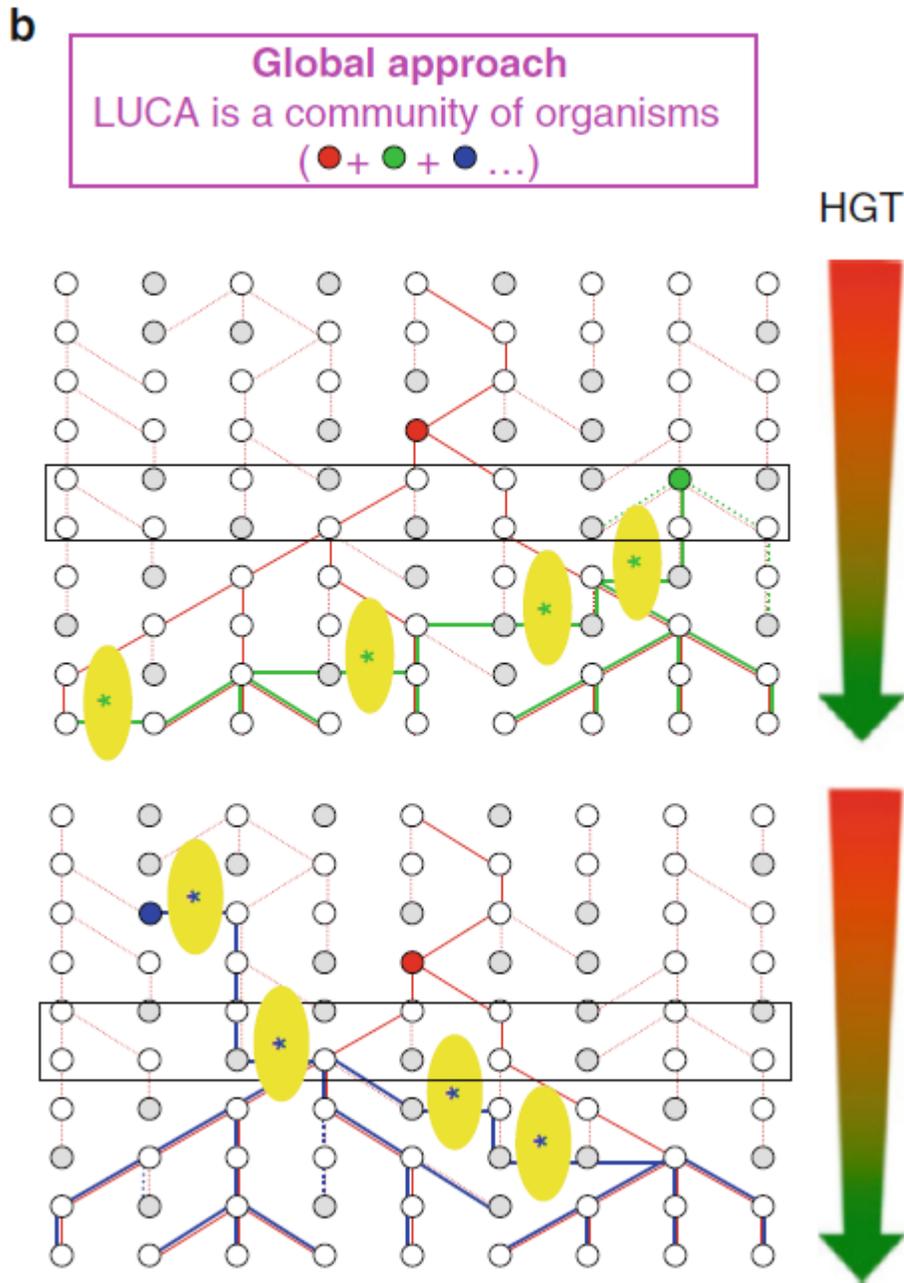
“LUCA is a theoretical construct—it might or might not have been something we today would call an organism. It helps to bridge the conceptual gap between rocks and water on the early Earth and ideas about the nature of the first cells.”

Weiss MC, Preiner M, Xavier JC, Zimorski V, Martin WF. The last universal common ancestor between ancient Earth chemistry and the onset of genetics. PLoS Genet. 2018 Aug 16;14(8):e1007518. doi: 10.1371/journal.pgen.1007518. PMID: 30114187; PMCID: PMC6095482.

Evoluzione della vita sulla Terra dalla sua origine. Si possono distinguere 3 stadi evolutivi: (1) il primo periodo: forme di vita non cellulari (mondo pre-cellulare); (2) il secondo periodo: discendenti della prima cellula (mondo cellulare pre-LUCA); (3) il terzo periodo: inizia con LUCA e continua ancora oggi (mondo cellulare post-LUCA). È importante notare che i rappresentanti delle diverse fasi hanno coesistito in determinati periodi. Pertanto, le prime cellule (rappresentanti del secondo periodo) molto probabilmente coesistevano con i loro cugini acellulari, discendenti del primo periodo. Allo stesso modo, i discendenti di LUCA probabilmente convissero per un periodo con i loro cugini del secondo periodo. La durata di queste tre fasi evolutive, e dei periodi possibili della convivenza, rimane del tutto sconosciuta. È probabile che la maggior parte dei lignaggi evolutivi che vennero alla luce si sono estinti senza lasciare discendenti.

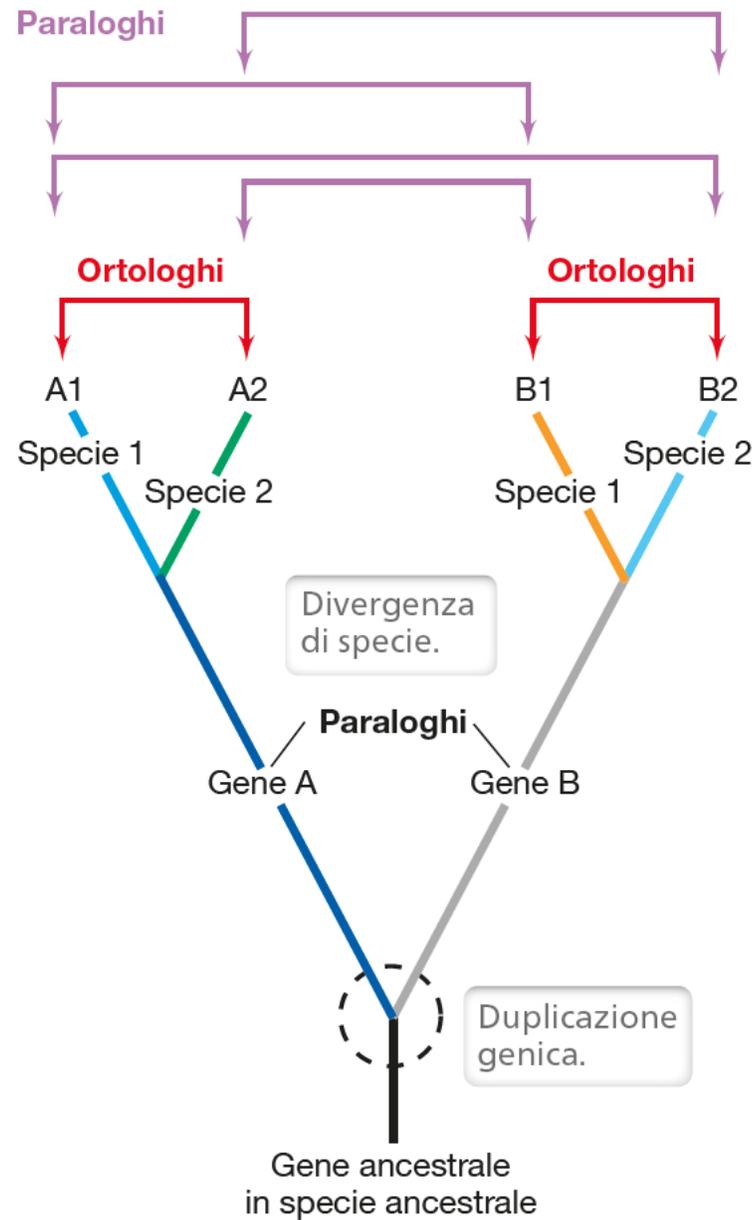


I cerchi rappresentano organismi che abitavano la Terra al tempo t . Ogni linea di cerchi rappresenta una generazione. Per ragioni di semplicità, si presume che (1) la riproduzione è asessuata e sincrona (non ci sono generazioni sovrapposte); (2) ogni organismo dà origine a zero, uno, due o tre discendenti; e (3) il numero di organismi che abitano la Terra ad ogni generazione è costante. LUCA è il comune antenato di tutti gli organismi viventi attualmente. A partire dagli organismi che abitano la Terra ora è possibile passare di generazione, in generazione, in generazione fino al cosiddetto punto di coalescenza (punto più recente da cui derivano tutti gli organismi esistenti). Questo punto di coalescenza rappresenta il LUCA. Ciò implica che nessuno degli organismi contemporanei del LUCA hanno lasciato discendenti che vivono oggi.



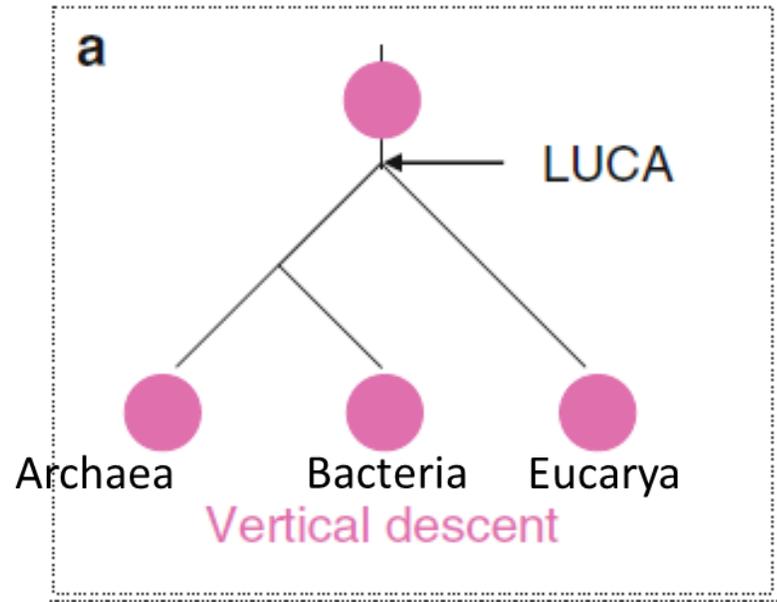
I cerchi rappresentano organismi che abitavano la Terra al tempo t . Ogni linea di cerchi rappresenta una generazione. Per ragioni di semplicità, si presume che (1) la riproduzione da una generazione all'altra è asessuata e sincrona (non ci sono generazioni sovrapposte); (2) ogni organismo dà origine a zero, uno, due o tre discendenti; e (3) il numero di organismi che abitano la Terra ad ogni generazione è costante.

La storia degli organismi è delineata in rosso. Altri colori (verde e blu) rappresentano la storia di due geni. Lo scambio di materiale genetico tra organismi contemporanei (asterischi) determina l'origine e la storia evolutiva di ciascun gene, che può essere distinta e differire da quelle degli organismi. Esistono virtualmente tanti LUCA quanti sono i geni (cerchi colorati). La freccia situata a destra mostra la frequenza dei trasferimenti genici orizzontali (HGT): inizialmente alta e gradualmente più bassa.



Geni omologhi: discendono da un unico gene ancestrale
Geni ortologhi: geni omologhi che svolgono la stessa funzione
Geni paraloghi: geni omologhi che svolgono funzioni differenti

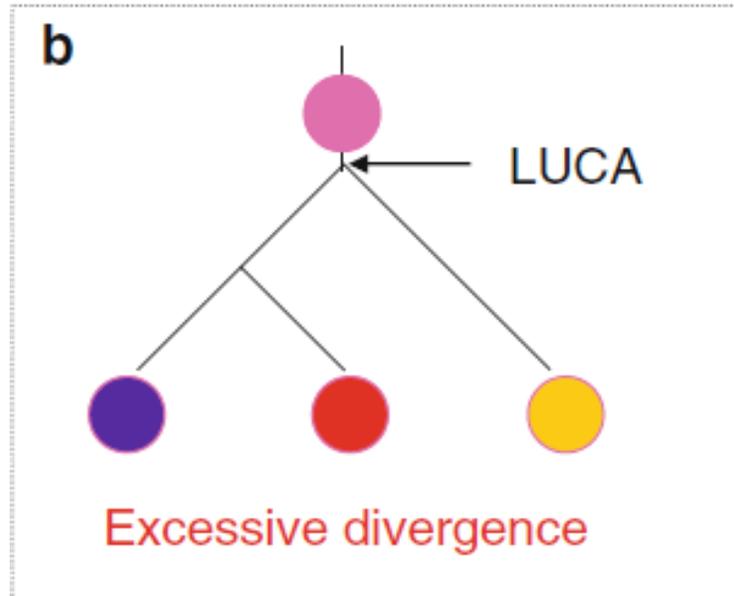
Figura 13.16 Omologhi. Questo albero genealogico illustra l'evoluzione di una famiglia genica costituita da geni omologhi (ortologhi e paraloghi), cioè discendenti da un unico gene ancestrale presente in un progenitore. Nella specie ancestrale si verifica la duplicazione di un gene che nel corso dell'evoluzione permette alla copia ridondante, il gene B, di sviluppare una nuova funzione all'interno del genoma della stessa specie ancestrale (Figura 13.17 a). Con il passare del tempo, la specie ancestrale si evolve in due nuove specie le cui sequenze geniche si sono separate a causa dell'evoluzione. I geni A1 e A2 svolgono le stesse funzioni e derivano dallo stesso gene ancestrale, perciò sono ortologhi, come anche B1 e B2. I geni derivanti dallo stesso gene ancestrale che svolgono però funzioni differenti sono paraloghi.



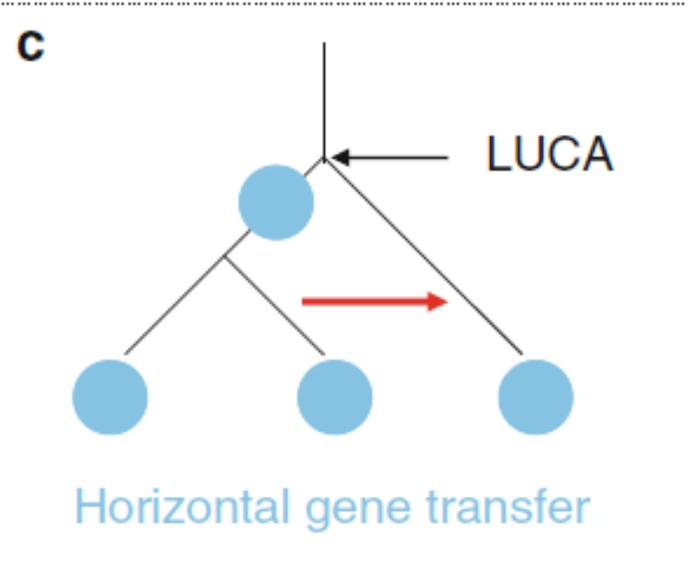
Possibili scenari evolutivi che spiegano la presenza o assenza di un carattere fenotipico nei tre domini della vita.

Uno specifico gene appare in un antenato di LUCA (cerchio rosa) ed è presente in LUCA e conservato nei tre domini della vita durante l'evoluzione .

Possibili scenari evolutivi che spiegano la presenza o assenza di un carattere fenotipico nei tre domini della vita.

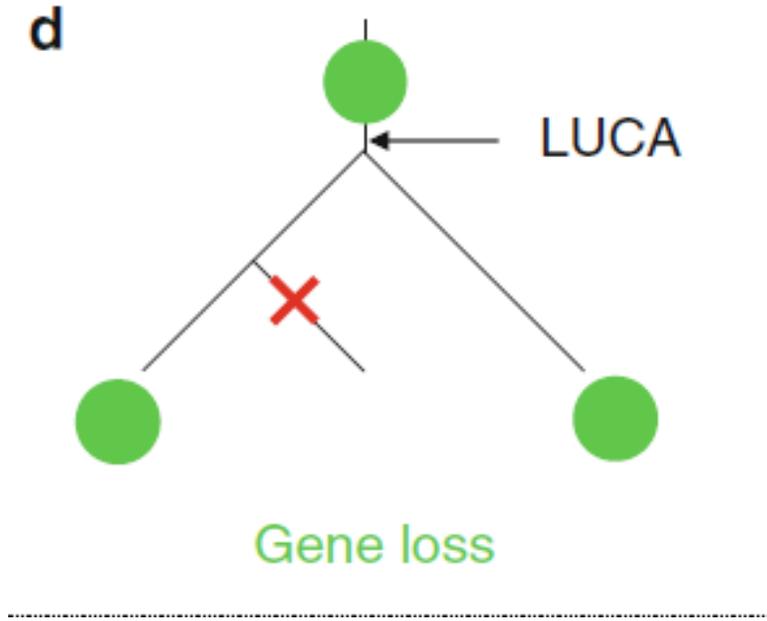


Un gene appare in un antenato di LUCA (cerchio rosa) ed è presente in LUCA e conservato nei tre domini della vita durante l'evoluzione, ma non può essere identificato come carattere universale a causa dell'eccessiva divergenza dovuta a mutazioni non conservative



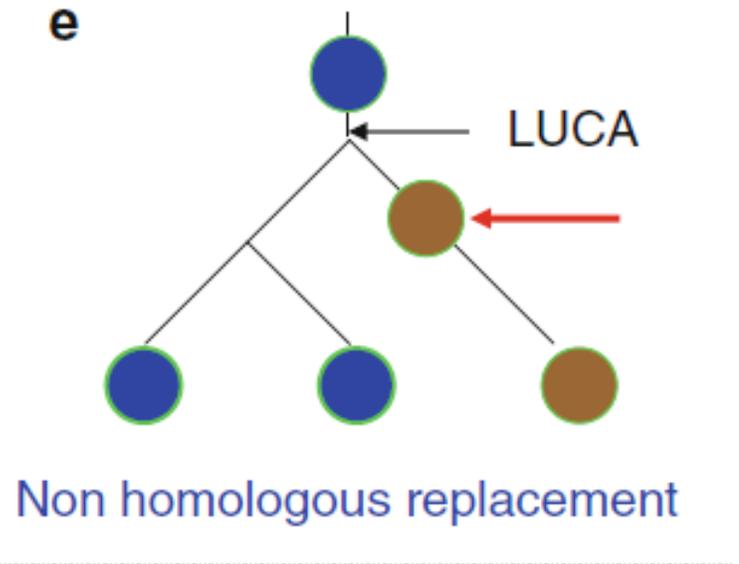
Possibili scenari evolutivi che spiegano la presenza o assenza di un carattere fenotipico nei tre domini della vita.

Un gene appare in un discendente di LUCA ed è trasmesso mediante trasferimento orizzontale (freccia rossa).



Possibili scenari evolutivi che spiegano la presenza o assenza di un carattere fenotipico nei tre domini della vita.

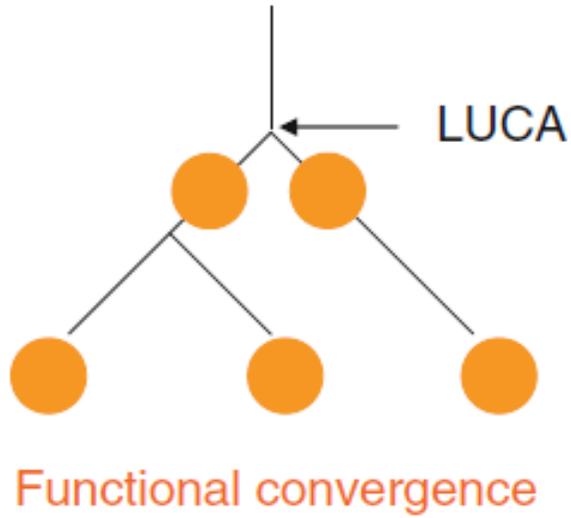
Perdita di un gene presente in un antenato di LUCA (cerchio verde)



Possibili scenari evolutivi che spiegano la presenza o assenza di un carattere fenotipico nei tre domini della vita.

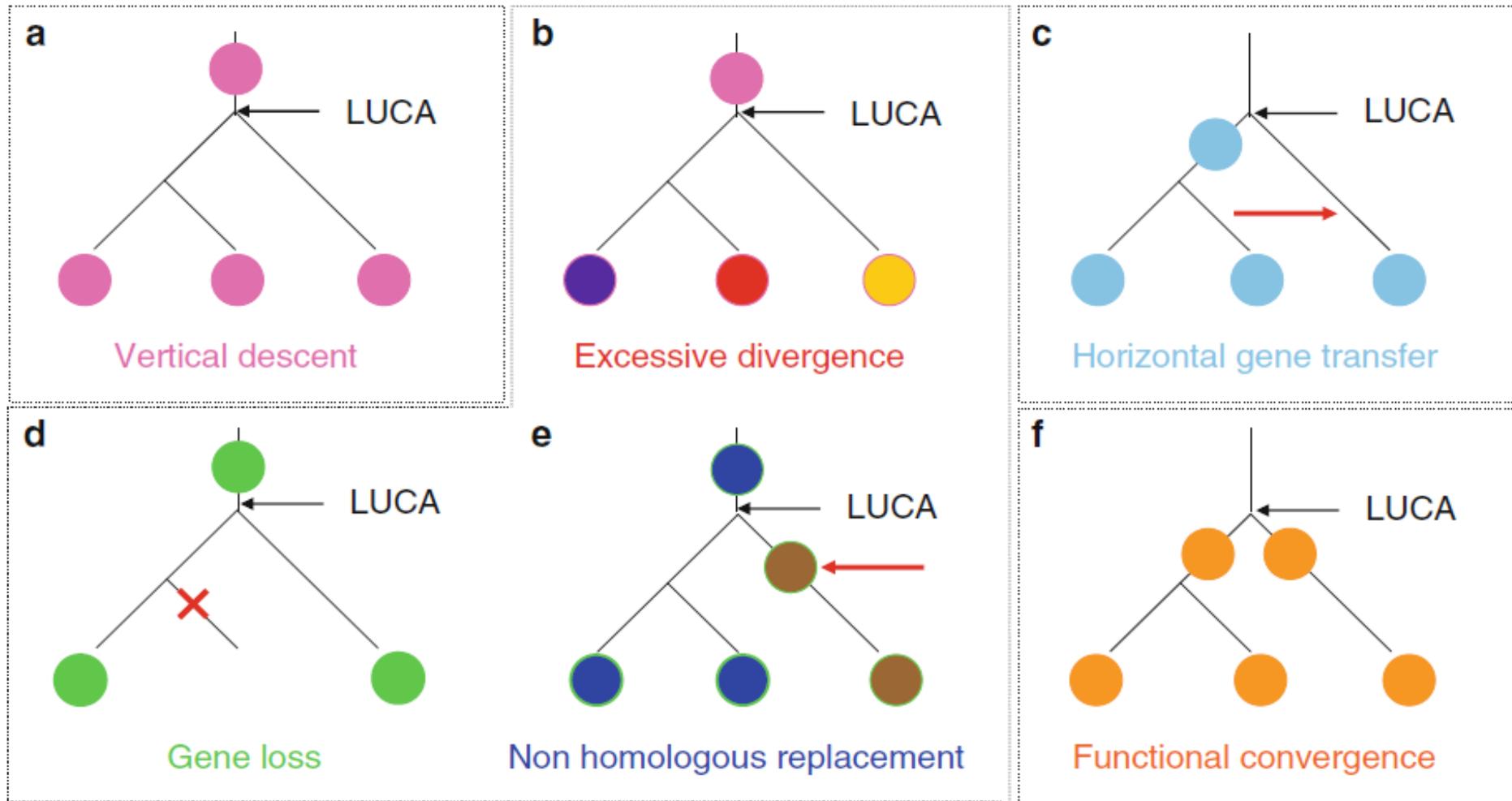
Sostituzione di un gene presente in un antenato di LUCA (cerchio marrone)

f



Possibili scenari evolutivi che spiegano la presenza o assenza di un carattere fenotipico nei tre domini della vita.

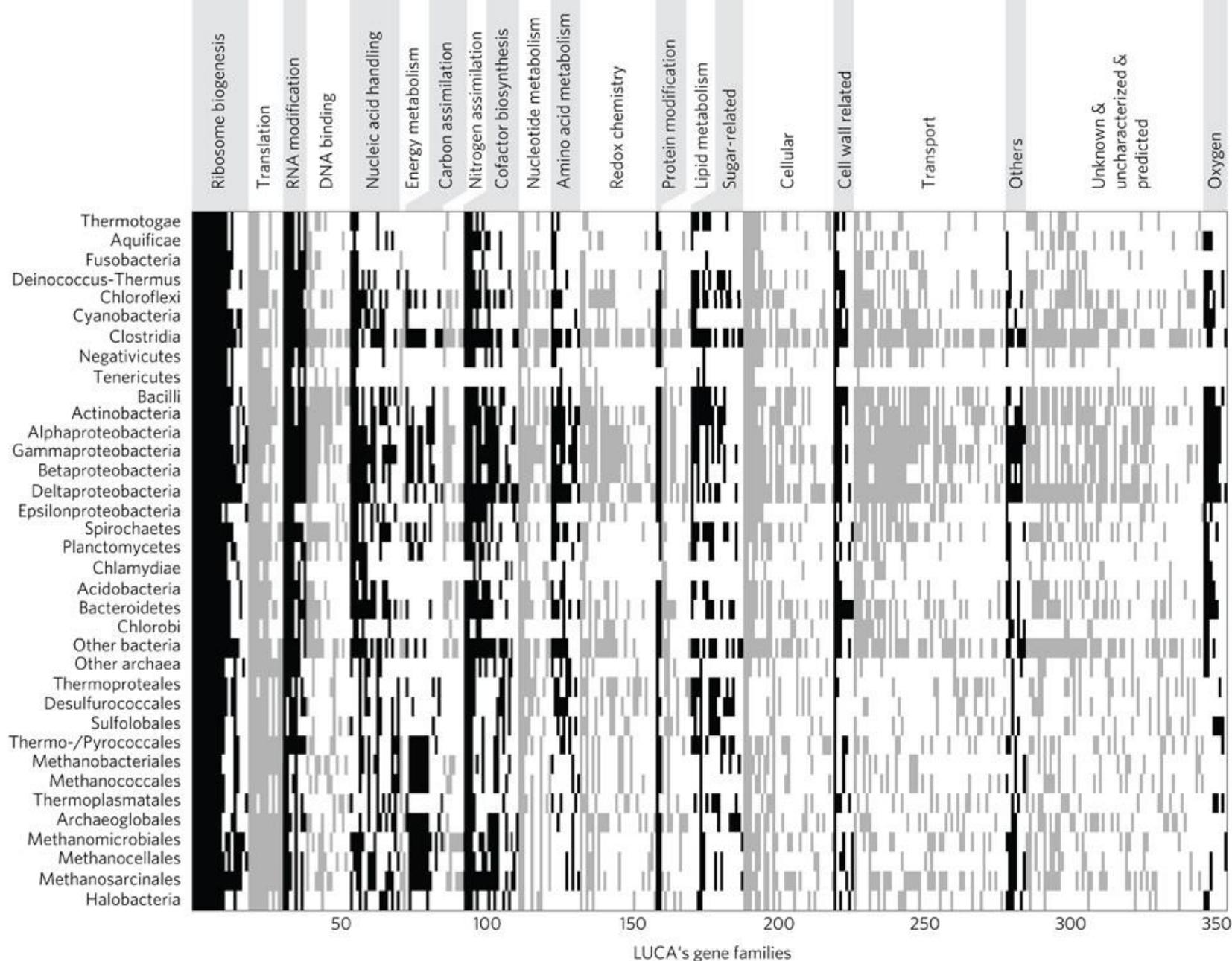
Apparizione indipendente (convergenza) della stessa funzione a carico di geni non omologhi.



(b, d, e) un gene presente in LUCA che non è rilevato simultaneamente in tutti e tre i domini della vita (c) gene universale che non esiste in LUCA, (f) una funzione universale ma codificata da geni diversi (convergenza funzionale)

Il punto di partenza di questi studi si basa sul fatto che gli attuali membri dei tre domini della vita hanno a numero di tratti comuni, come un genoma a DNA, che definisce uno stesso codice genetico, un ribosoma complesso e una membrana citoplasmatica. **La spiegazione più semplice per la presenza delle caratteristiche comuni a tutti gli organismi è quella di considerare che siano state tutte ereditate dal loro ultimo antenato comune, cioè LUCA.**

In tutto 355 geni possono essere ricondotti ad un unico antenato comune.



For each protein family (columns), ticks indicate the presence (black or grey) or absence (white) of the gene in the corresponding group shown on the left (rows). Within each functional category, clusters were sorted by the total sum of genes present (x axis). T

La cellula ancestrale da cui sono derivati tutti gli organismi viventi, il LUCA, aveva una membrana diversa da quelle delle cellule di oggi: "perdeva". Questa particolarità le permetteva di rifornirsi di energia sfruttando i gradienti di concentrazione di particelle cariche presenti nel suo ambiente, quello delle bocche idrotermali marine. Le pompe protoniche usate da tutte le cellule odierne si sono evolute nella colonizzazione di ambienti più "difficili«

Secondo la ricostruzione di Victor Sojo, Nick Lane e colleghi – descritta in [un articolo su “PLoS Biology”](#), l'organismo ancestrale sarebbe vissuto in prossimità delle bocche idrotermali degli antichi mari (note anche come fumarole nere, o *black smoker*), un ambiente in cui le calde acque alcaline, povere di protoni, che fuoriescono si mescolano alle acque oceaniche più ricche di protoni.

LUCA sfruttava la differenza di concentrazione dei protoni in questi due ambienti per procurarsi l'energia necessaria alla produzione di adenosintrifosfato (ATP), la molecola che ancora oggi alimenta la crescita delle cellule. Questo meccanismo funziona, però, solo se la cellula è dotata di una membrana “porosa”, permettendo un flusso naturale di protoni sia in entrata, sia in uscita.

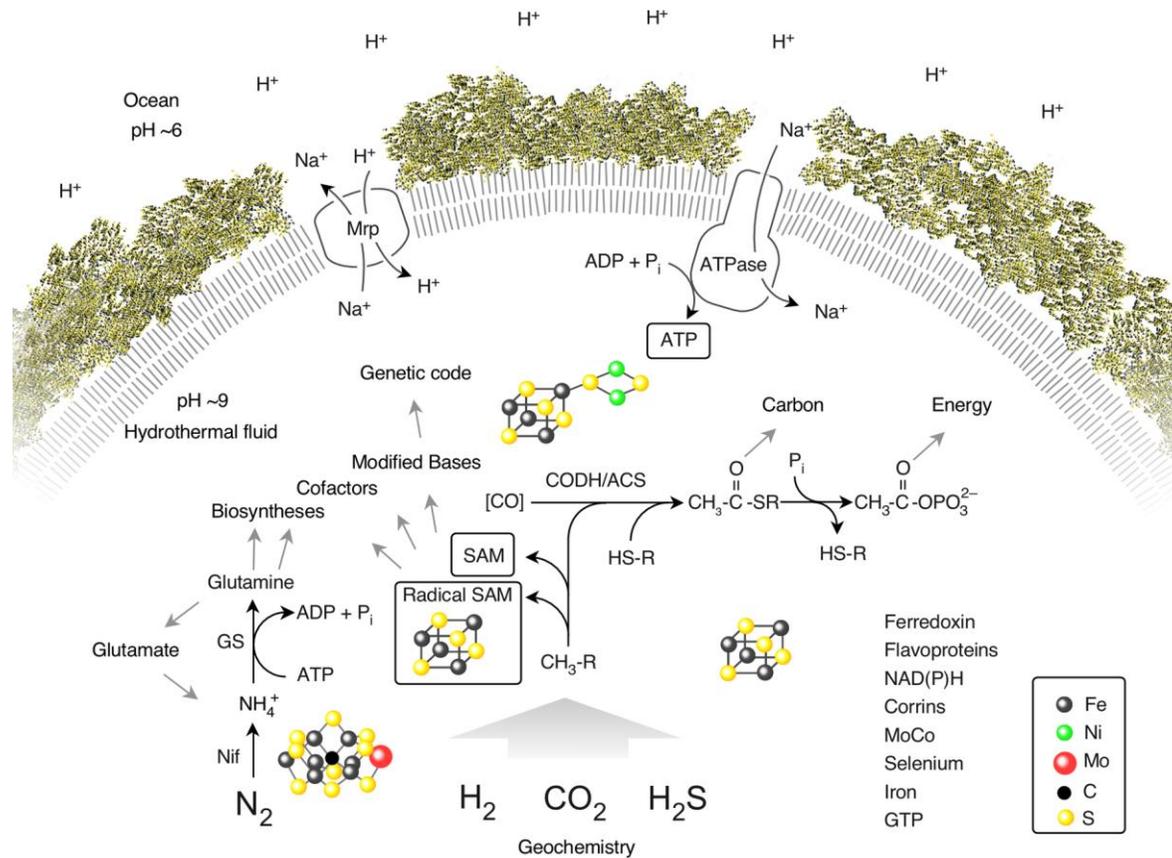


Fig 2. Three ways to infer genes present in LUCA. The gene presence is indicated with a plus sign, absence with a minus sign. a) Genes found universally in both domains, regardless of their tree, trace to LUCA. About 30 fulfil this criterion. b) Another way to trace genes to LUCA is to say that any gene found in both archaea and bacteria was present in LUCA. However, thousands of these genes will have been transferred between bacteria and archaea by LGT so were not necessarily present in LUCA. c) Genes present in only one bacterial or archaeal phylum could easily be the result of LGT and are removed. But presence in two phyla per domain while preserving domain monophyly yields good candidates to have been present in LUCA. Such phylogenies would only result from LGT under very specific and restrictive conditions. They require exactly one transdomain transfer followed by either i) one additional transdomain LGT from the same donor lineage to a different recipient phylum or ii) retention during phylum divergence in the recipient domain, plus—in addition to either criteria i) or ii)—an additional, more subtle but highly restrictive criterion: No further transdomain LGTs occurred during all of evolution. Subsequent transdomain LGT would violate domain monophyly for the gene. Indeed, transdomain LGT is common, and 97% of the trees examined by Weiss and colleagues [78] did not exclude transdomain LGT (remaining 3%, 355 trees, provided in S1 Appendix). LGT, lateral gene transfer; LUCA, last universal common ancestor.

Il LUCA doveva essere un organismo anaerobico che viveva in un ambiente ad alta temperatura, ricco di ferro e zolfo, e utilizzava H_2 come fonte di energia e CO_2/CO come fonte di carbonio. Con tutta probabilità cresceva in sorgenti idrotermali, forse utilizzando la Wood-Ljungdahl pathway per la fissazione della CO_2 come i moderni clostridi

Per colonizzare ambienti differenti dalle fumarole, LUCA ha però dovuto adattare la sua membrana in modo da pompare attivamente i protoni fuori dalla cella e consentire il flusso di protoni e la produzione di ATP. Batteri e archeobatteri abbiano sviluppato strutture della membrana cellulare e pompe protoniche completamente differenti, pur mantenendo gli stessi macchinari per alimentare la crescita.

Secondo una nuova ricerca, la più antica "struttura" comune a tutti gli esseri viventi - eucarioti, batteri e molto probabilmente anche archaea - aveva già un'organizzazione simile a una cellula. L'elemento chiave è rappresentato da una struttura, presente nei tre domini della vita: batteri, archaea ed eucarioti (piante, animali, funghi, alghe e tutto il resto), che andrebbe a tutti gli effetti considerata un organello, l'acidocalcisoma.

“Acidocalcisomes are acidic organelles containing calcium and a high concentration of phosphorus in the form of pyrophosphate (PP_i) and polyphosphate (poly P). Acidocalcisomes function as storage sites for cations and phosphorus, and participate in PP_i and poly P metabolism, calcium homeostasis, maintenance of intracellular pH, and osmoregulation. Acidocalcisomes interact with other organelles in a number of organisms suggesting their association with the endosomal/lysosomal pathway, and are considered part of the lysosome-related group of organelles.”

https://www.lescienze.it/news/2011/10/06/news/luca_l_ultimo_antenato_comune_di_tutti_gli_organismi-559369/

Identification of Organelles in Bacteria Similar to Acidocalcisomes of Unicellular Eukaryotes*

Received for publication, April 30, 2003, and in revised form, May 29, 2003
Published, JBC Papers in Press, June 3, 2003, DOI 10.1074/jbc.M304548200

Manfredo Seufferheld, Mauricio C. F. Vieira, Felix A. Ruiz, Claudia O. Rodrigues,
Silvia N. J. Moreno, and Roberto Docampo‡

From the Laboratory of Molecular Parasitology, Department of Pathobiology and Center for Zoonoses Research, University of Illinois at Urbana-Champaign, Urbana, Illinois 61802

Storia evolutiva della H⁺-pirofosfatasi vacuolare, comune negli acidocalcisomi delle cellule eucariotiche e batteriche, ma presente anche negli Archaea.



Acidocalcisomi e H⁺ pirofosfatasi presenti in LUCA (?)

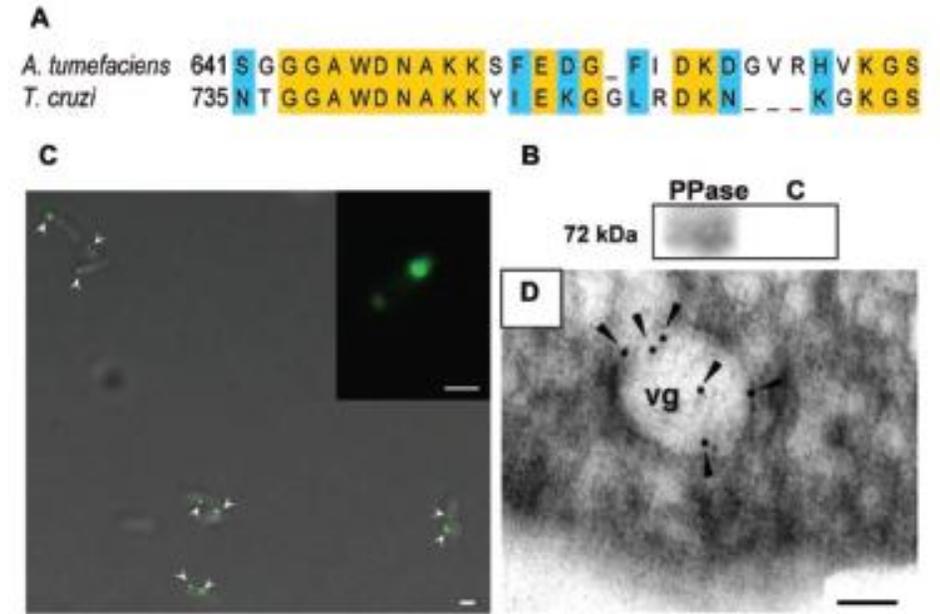


FIG. 5. **Western blot analysis, immunofluorescence analysis, and immunoelectron microscopy of H-PPase in *A. tumefaciens*.** A, CLUSTAL W alignment of the C-terminal region of putative H-PPases from *A. tumefaciens* (GenBank accession number AAL42186) and *T. cruzi* (AF159881). Homologous residues are in yellow, and conserved residues are in blue. B, detection of the H-PPase by immunoblot, using a polyclonal antibody against the *T. cruzi* H-PPase. *A. tumefaciens* proteins (14 g) were separated by SDS-PAGE and transferred to nitrocellulose. *Left lane*, immunoblot probed with antibody against the H-PPase recognized a polypeptide of apparent molecular mass of 72 kDa. The *right lane* shows immunoblot probed with pre-immune serum. C, confocal immunofluorescence analysis of H-PPase in *A. tumefaciens*. Arrowheads show labeling at one pole of the cells. Confocal fluorescence image was overlaid on bright-field image of the same cells. The *inset* shows greater magnification of one of the cells by fluorescence microscopy. Bars, 1 μ m. D, immune electron microscopy of the cells. Labeling of a cytoplasmic vacuole (vg) is evident (arrowheads). Bar, 40 nm.

Mettendo a confronto le sequenze genetiche relative alla H⁺-pirofosfatasi vacuolare in centinaia di organismi che rappresentano i tre domini della vita, i ricercatori ne hanno ricostruito un "albero genealogico" che mostra le relazioni delle sue diverse versioni in organismi diversi. Molto significativamente, osservano i ricercatori, questo albero è molto simile nei dettagli all'albero della vita universale creato sulla base dell'analisi di numerosi geni. L'enzima e l'acidocalcisoma sono quindi molto antichi e risalgono a LUCA, ossia a prima della separazione dei tre rami principali dell'albero della vita

La radice dell'albero universale della vita

Chi tra Archaea, Batteri ed Eucarioti è emerso per primo?

Questo problema è ancora oggetto di accesi dibattiti nella comunità scientifica.

Per sapere in che ordine i tre domini di vita divergente, sarebbe necessario:

1. Ricostruire l'albero della vita universale (cioè basato su marcatori molecolari presenti in tutti gli organismi viventi)
2. Risalire alla radice questo albero (vale a dire definire la direzione di tempo evolutivo lungo i rami dell'albero)