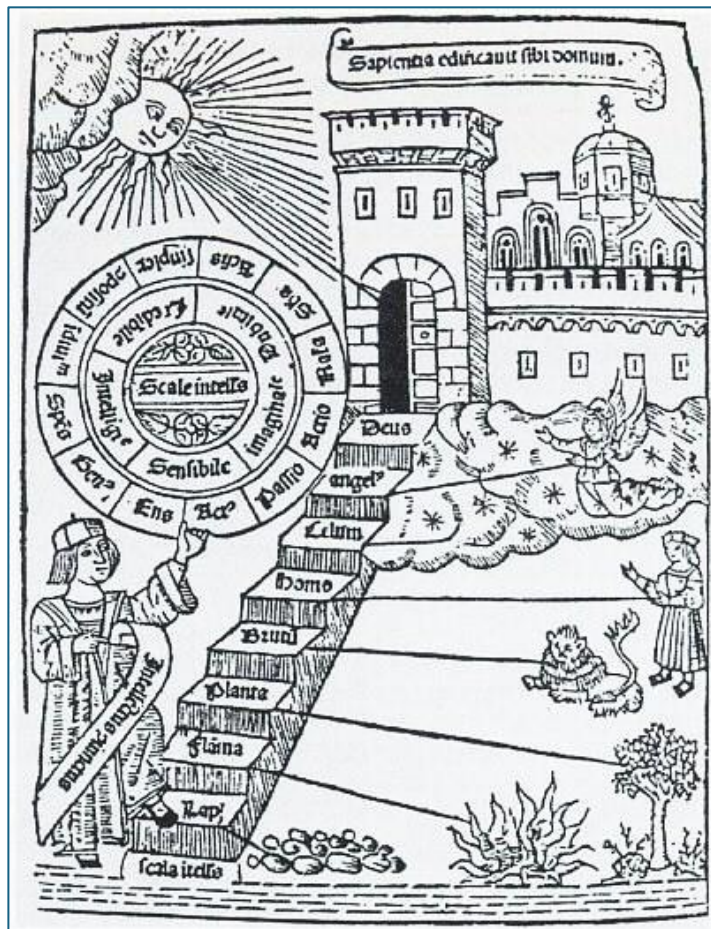


Split Decomposition

Properties and Algorithms of the Split
Decomposition Method in Phylogenetics

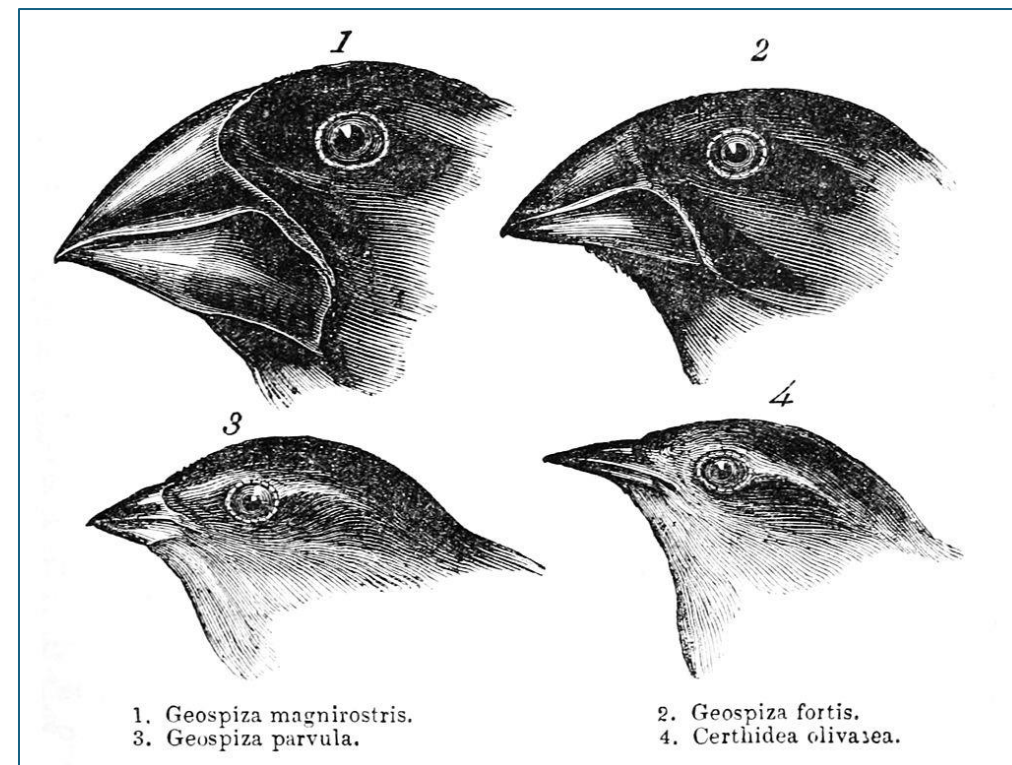


Llull, 1305

scala naturae

specie fisse e immutabili

Evoluzione

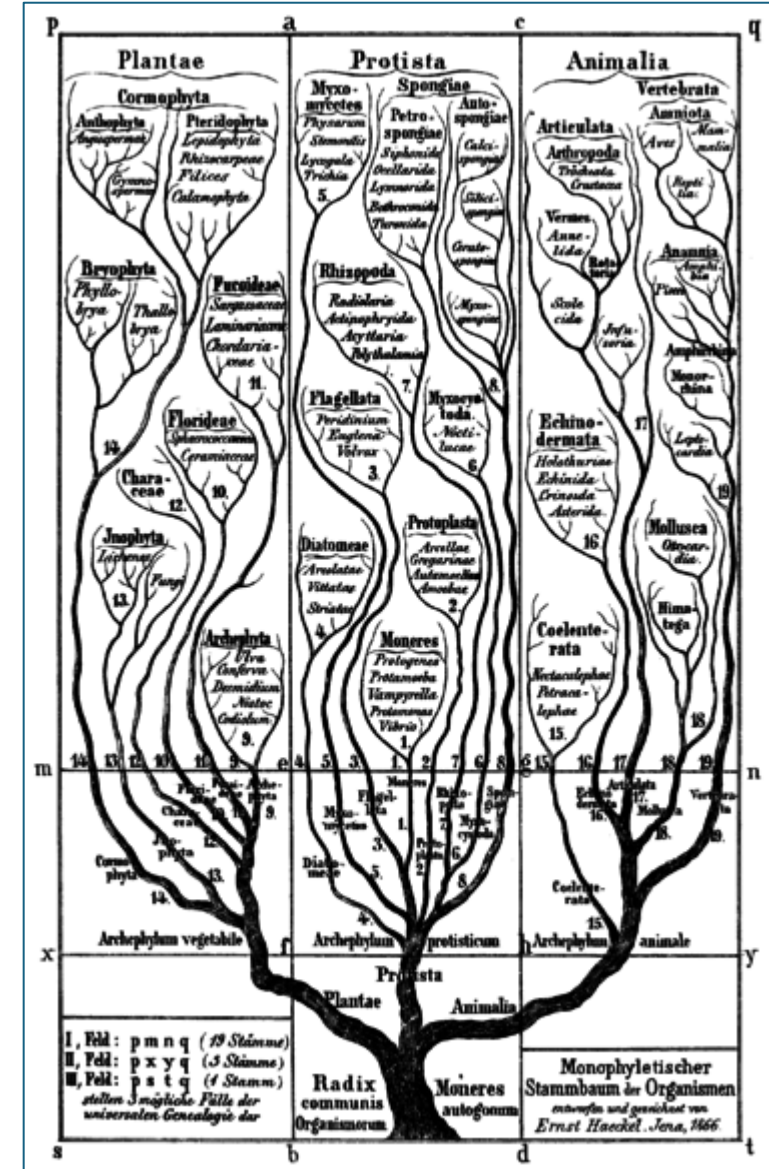
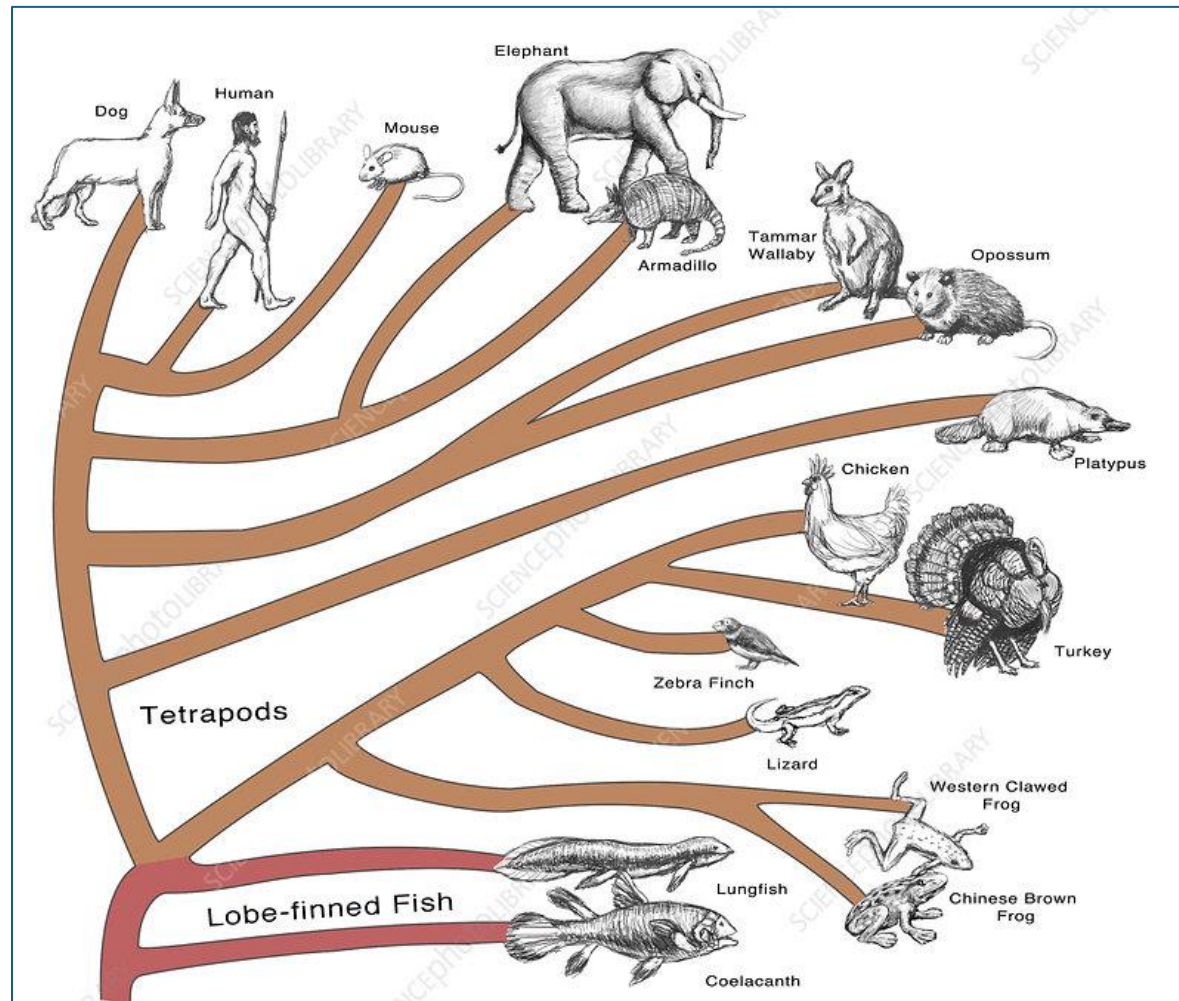


Darwin, 1845

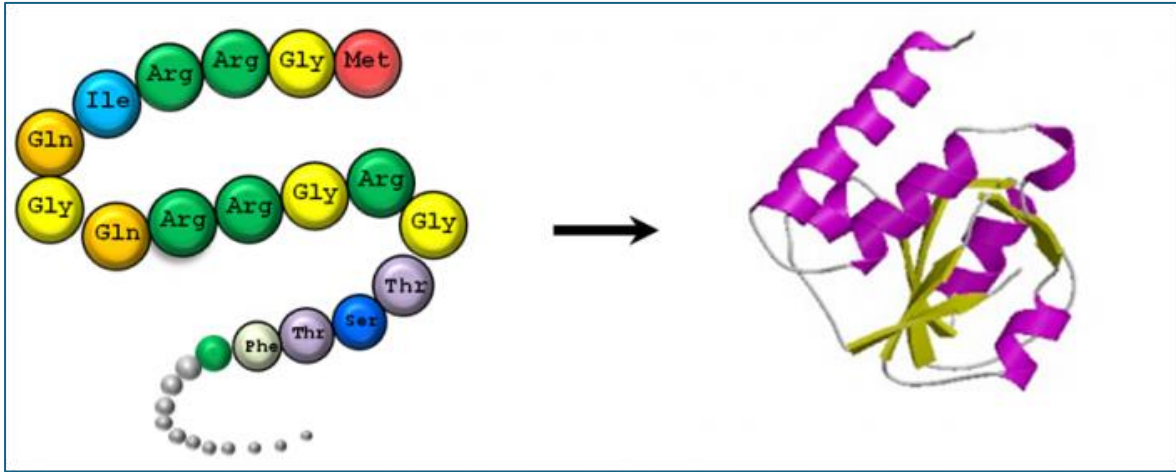
evoluzione

le specie possono cambiare

Filogenetica

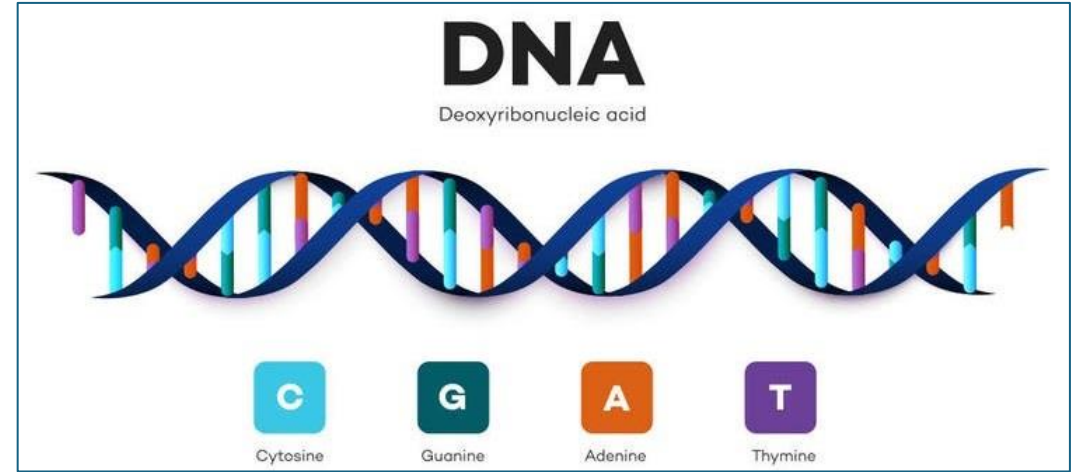


Haeckel, 1866



amminoacidi

proteina



nucleotidi

a: CAGGATTAG
b: CAGGTTTAG
c: CATTTTAG
d: ACGTTAA
e: ATGTTAA

sequenze allineate

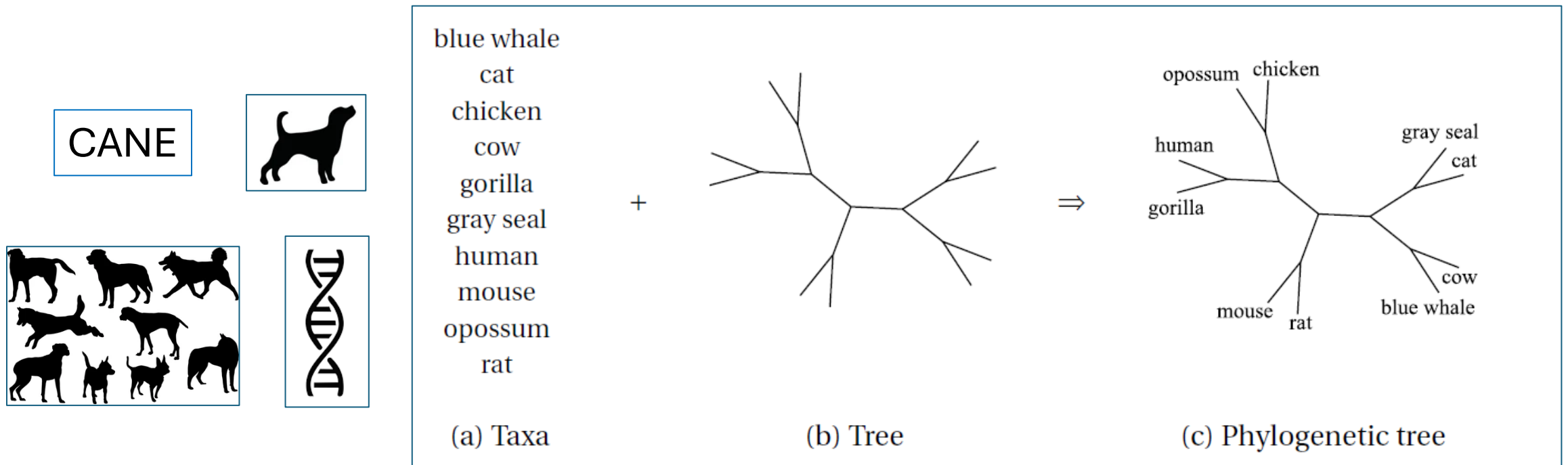


	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>
<i>a</i>	0	1	3	4	4
<i>b</i>	1	0	2	4	4
<i>c</i>	3	2	0	5	5
<i>d</i>	4	4	5	0	1
<i>e</i>	4	4	5	1	0

matrice delle distanze

Problema:

a partire da una matrice di distanze,
ricostruire l'albero filogenetico che le induce
(o una sua approssimazione)



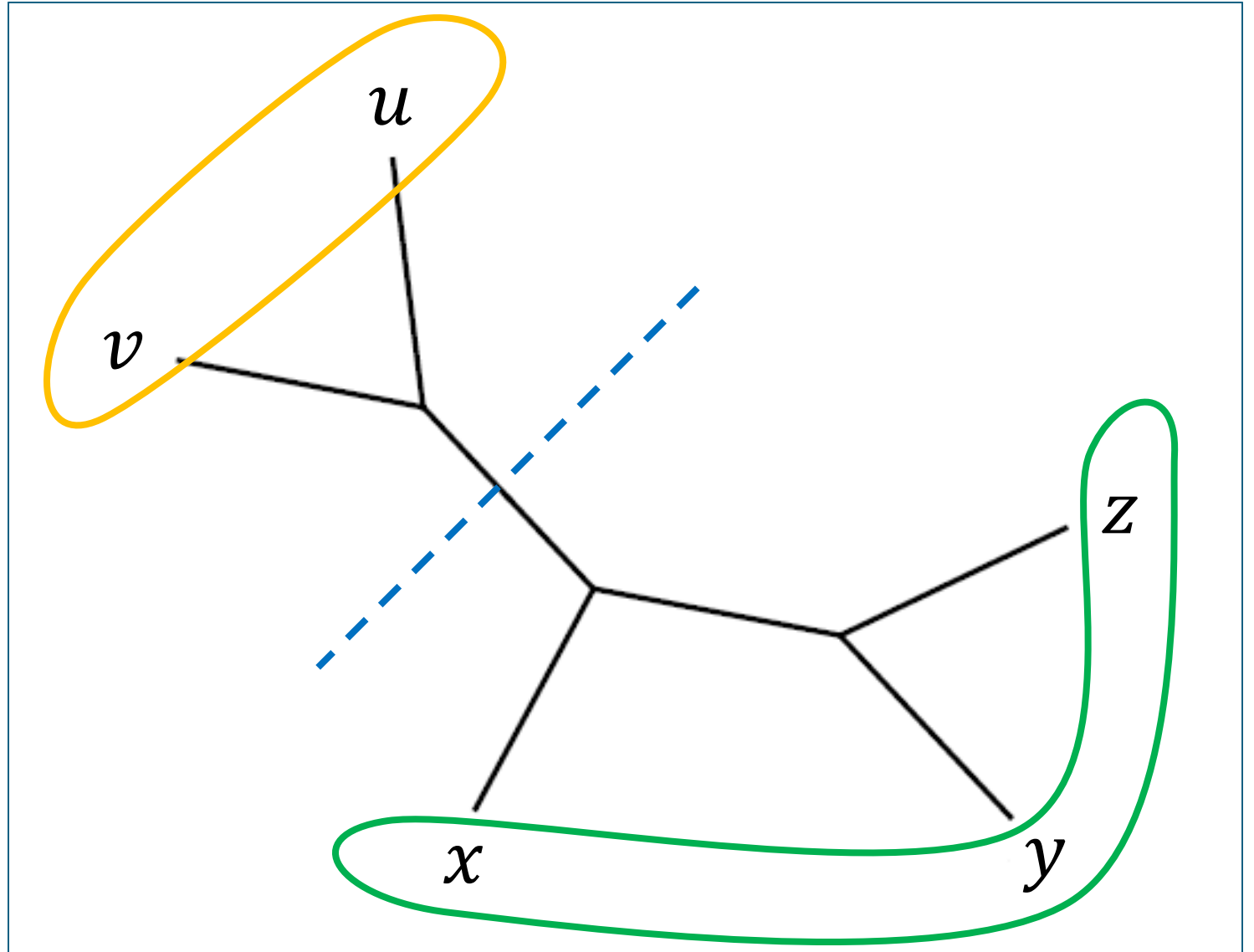
Split

bipartizione
dell'insieme
dei taxa X

es.

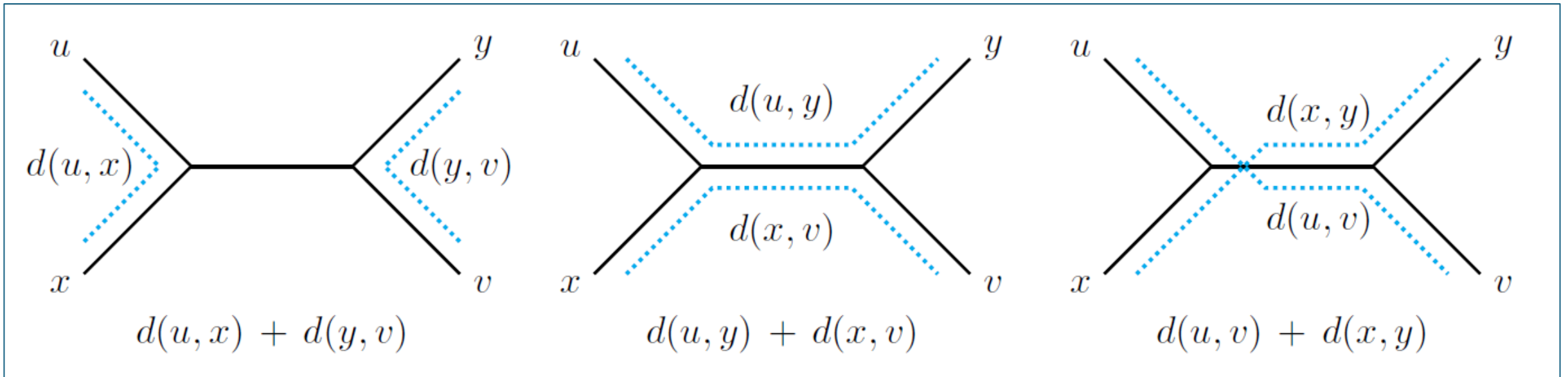
$$uv \mid xyz$$

$$X = \{u, v, x, y, z\}$$



Fatto Sia d una distanza indotta da un albero.
 Allora uno split $A|B$ fa parte dell'albero
 se e solo se per ogni $a, a' \in A$ e $b, b' \in B$

$$aa' + bb' < ab + a'b' = ab' + a'b$$

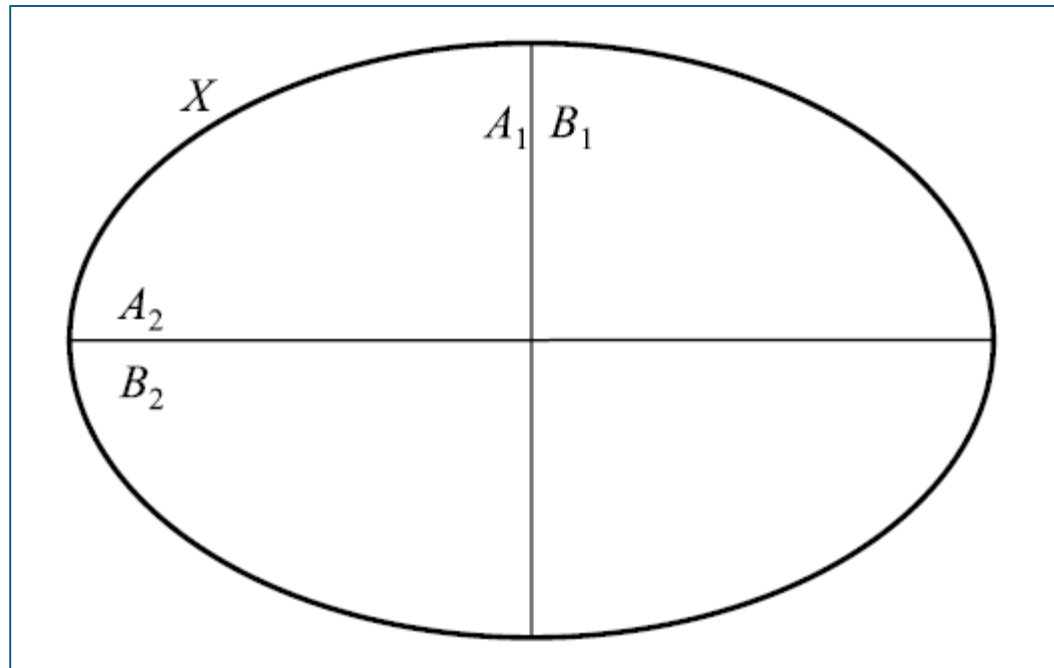


Fatto Un insieme di split è realizzato da un albero filogenetico se e solo se è **compatibile**,

cioè dati due split $A_1|B_1$ e $A_2|B_2$, almeno una tra

$$A_1 \cap A_2, \quad A_1 \cap B_2, \quad B_1 \cap A_2, \quad B_1 \cap B_2$$

è vuota.





$$d(u, x) + d(y, v)$$

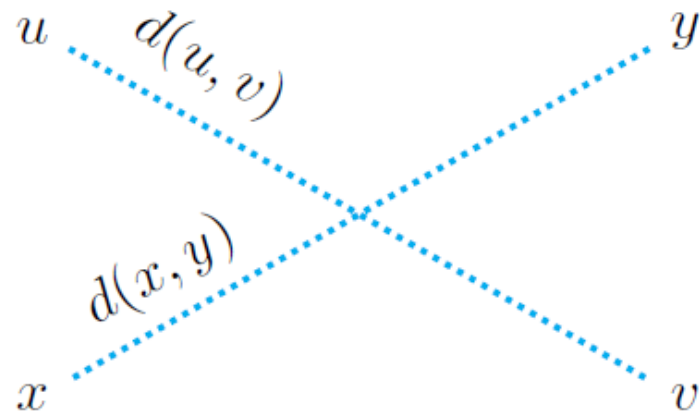


$$d(u, y)$$

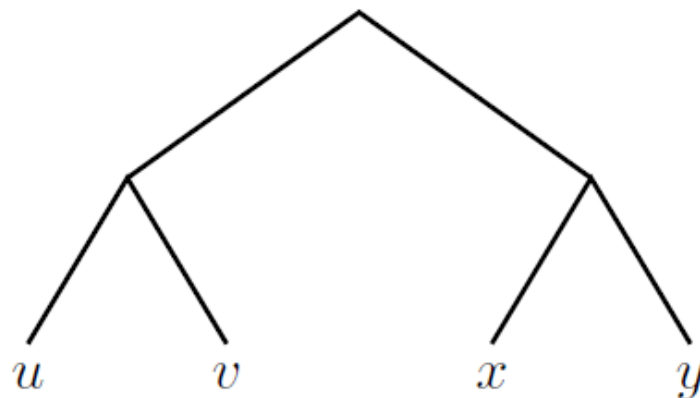
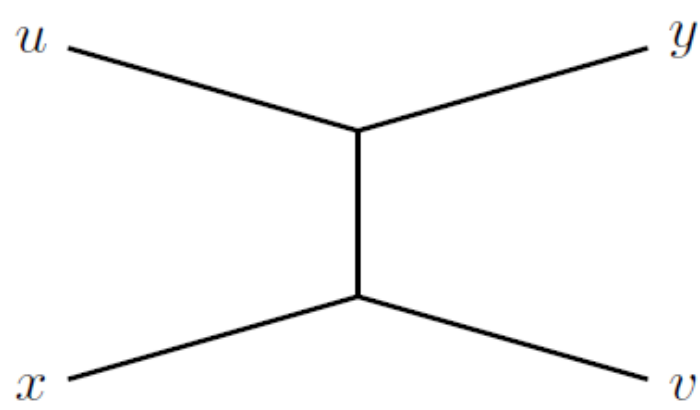
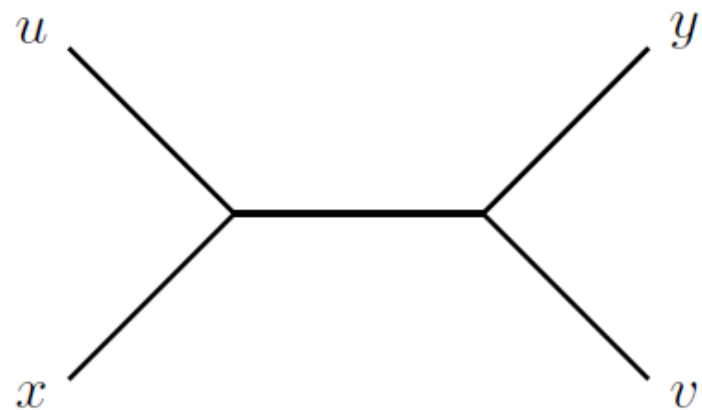


$$d(x, v)$$

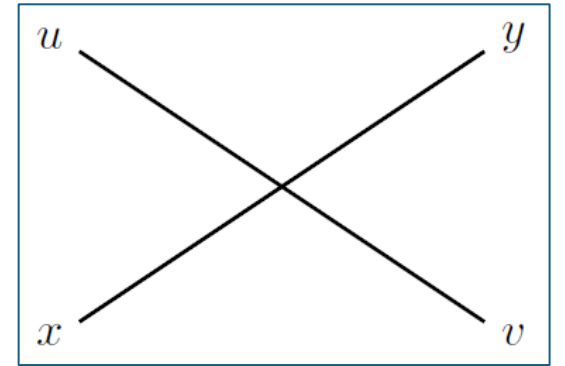
$$d(u, y) + d(x, v)$$



$$d(u, v) + d(x, y)$$

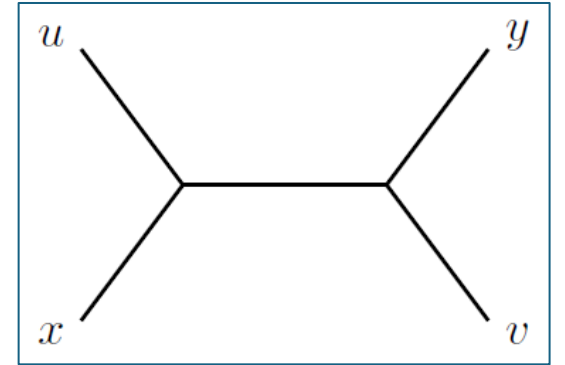


$$aa' + bb' < ab + a'b' = ab' + a'b$$

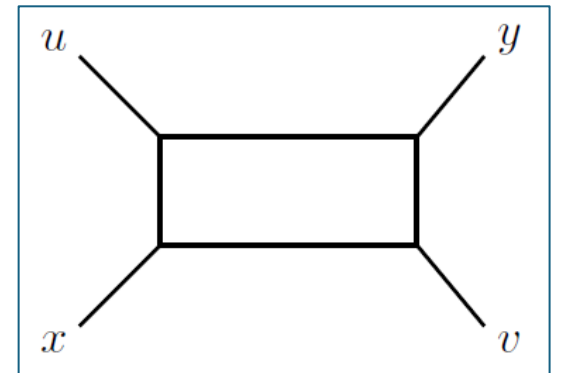


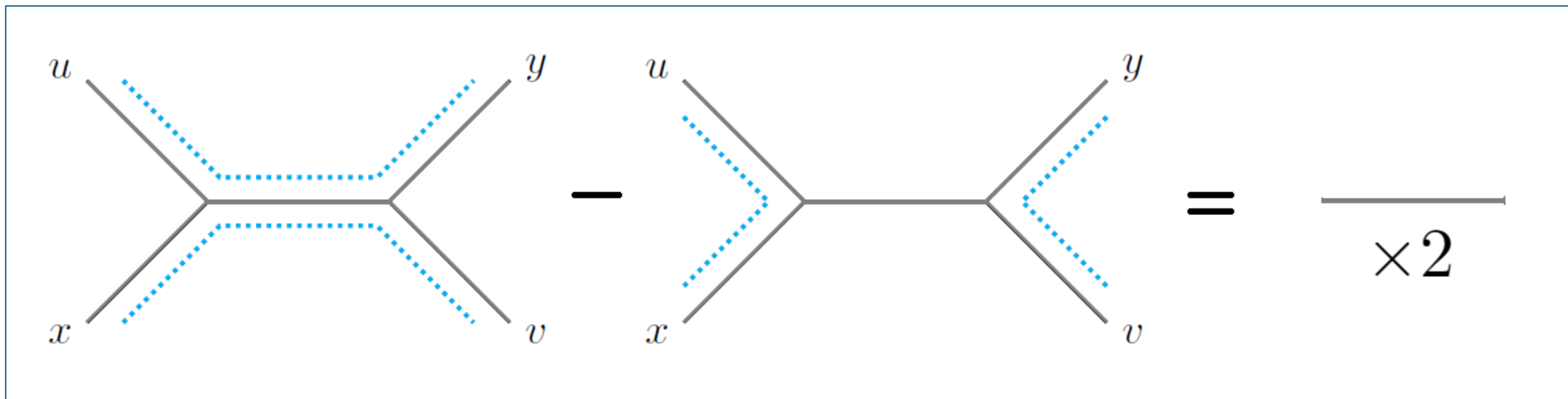
$$aa' + bb' < ab + a'b', \quad ab' + a'b$$

$$aa' + bb' < \min \{ab + a'b', ab' + a'b\}$$

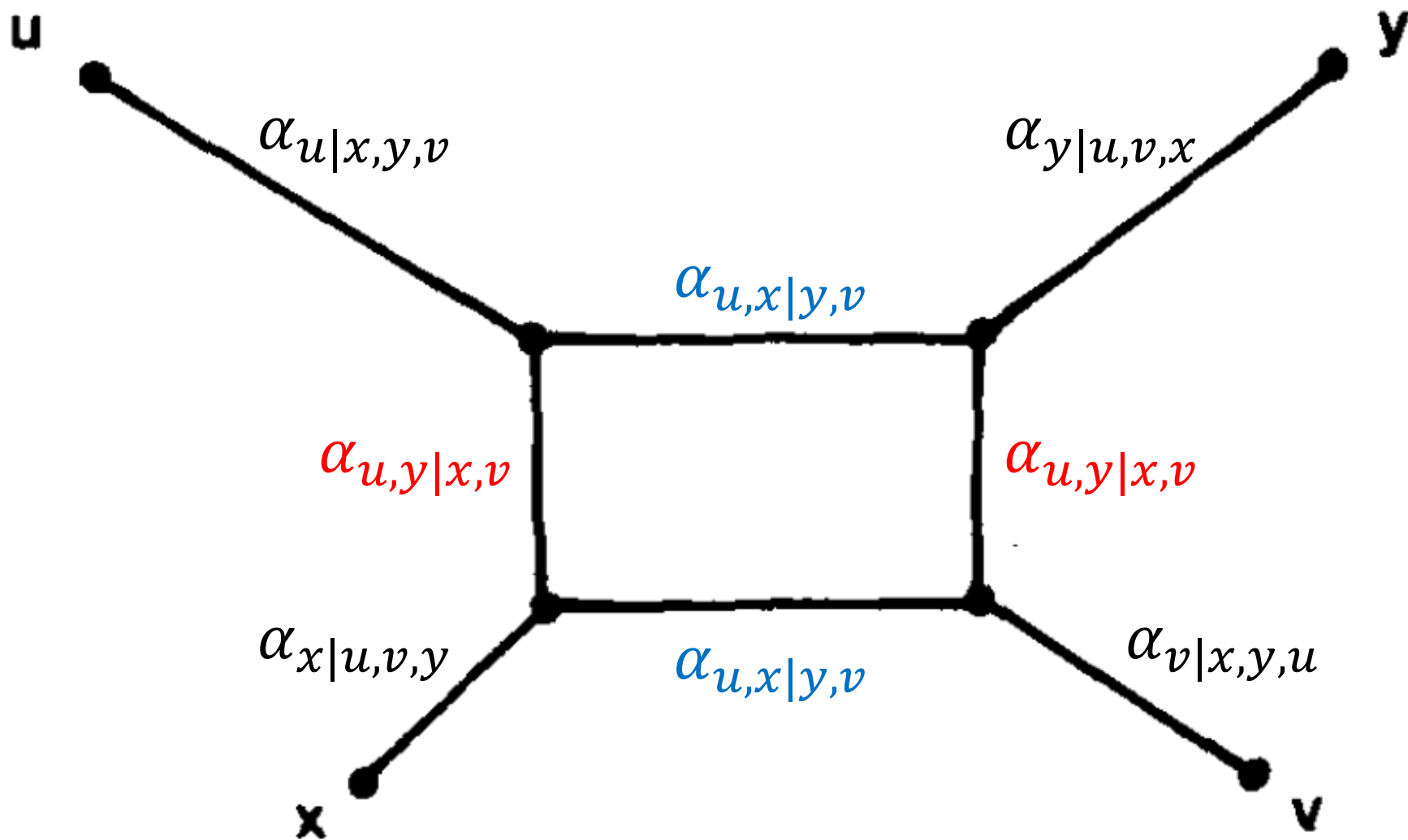


$$aa' + bb' < \max \{ab + a'b', ab' + a'b\}$$





$$\alpha_{A|B} := \min_{\substack{a, a' \in A \\ b, b' \in B}} \frac{1}{2} \left(\max \{ab + a'b', a'b + ab', aa' + bb'\} - aa' - bb' \right)$$



$$\alpha_{u,v|x,y} = 0$$

Split Decomposition

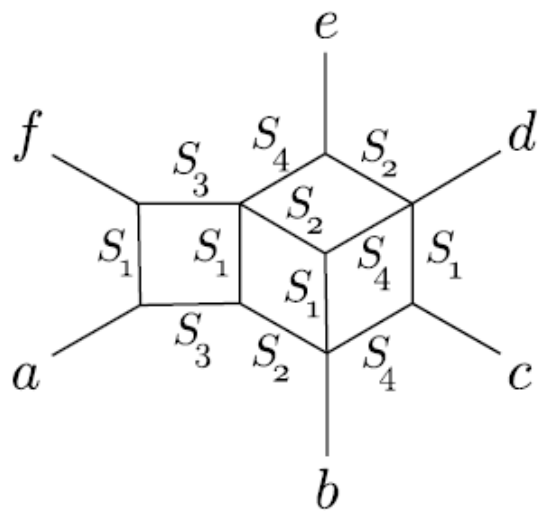
Una qualsiasi funzione simmetrica $d : X \times X \rightarrow \mathbb{R}$
si può scrivere come

$$d = d_0 + \sum_{A|B} \alpha_{A|B} \cdot \delta_{A|B}$$

$$\delta_{A|B}(x, y) = \begin{cases} 0, & \text{se } x, y \in A \text{ o } x, y \in B \\ 1, & \text{altrimenti} \end{cases}$$

Metodo della *Split Decomposition*

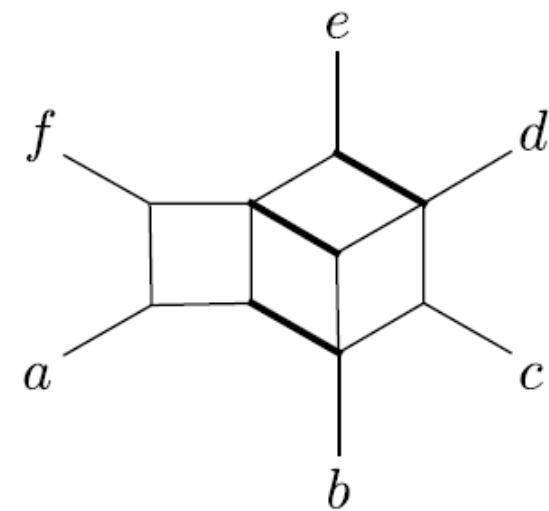
- calcolare la decomposizione di d
- tenere gli split con indice non nullo
- costruire lo split network



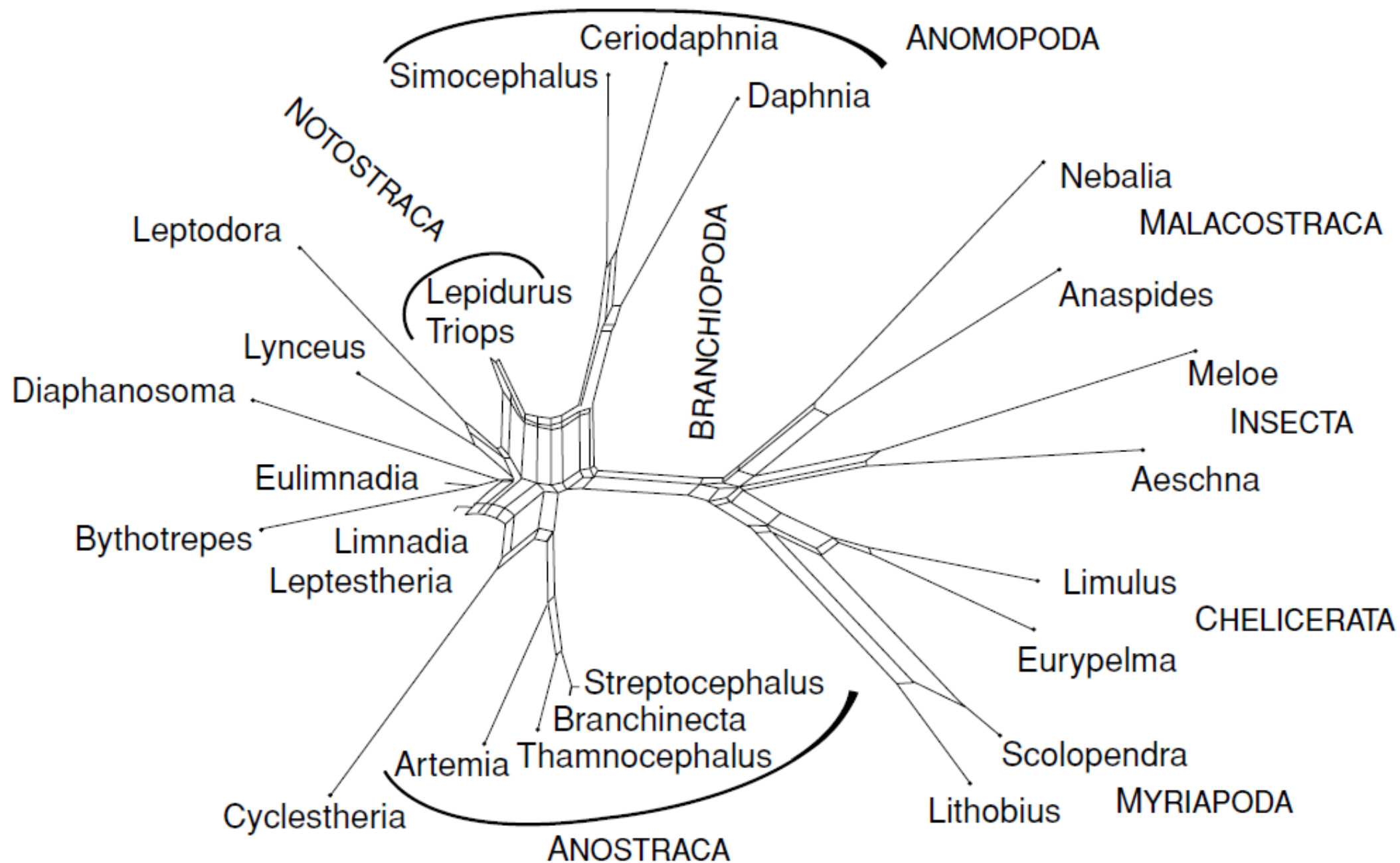
(a) Split network N

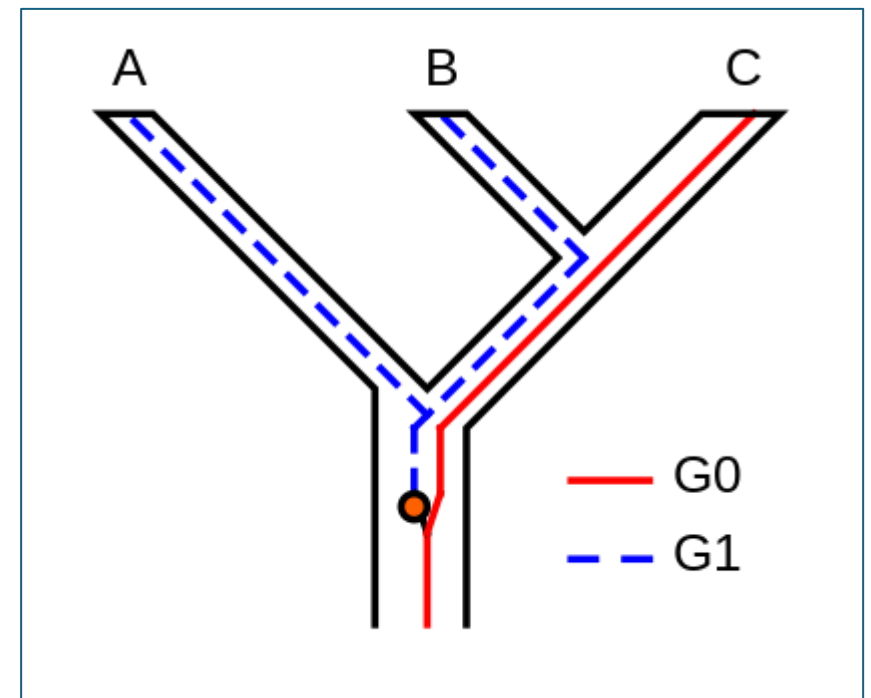
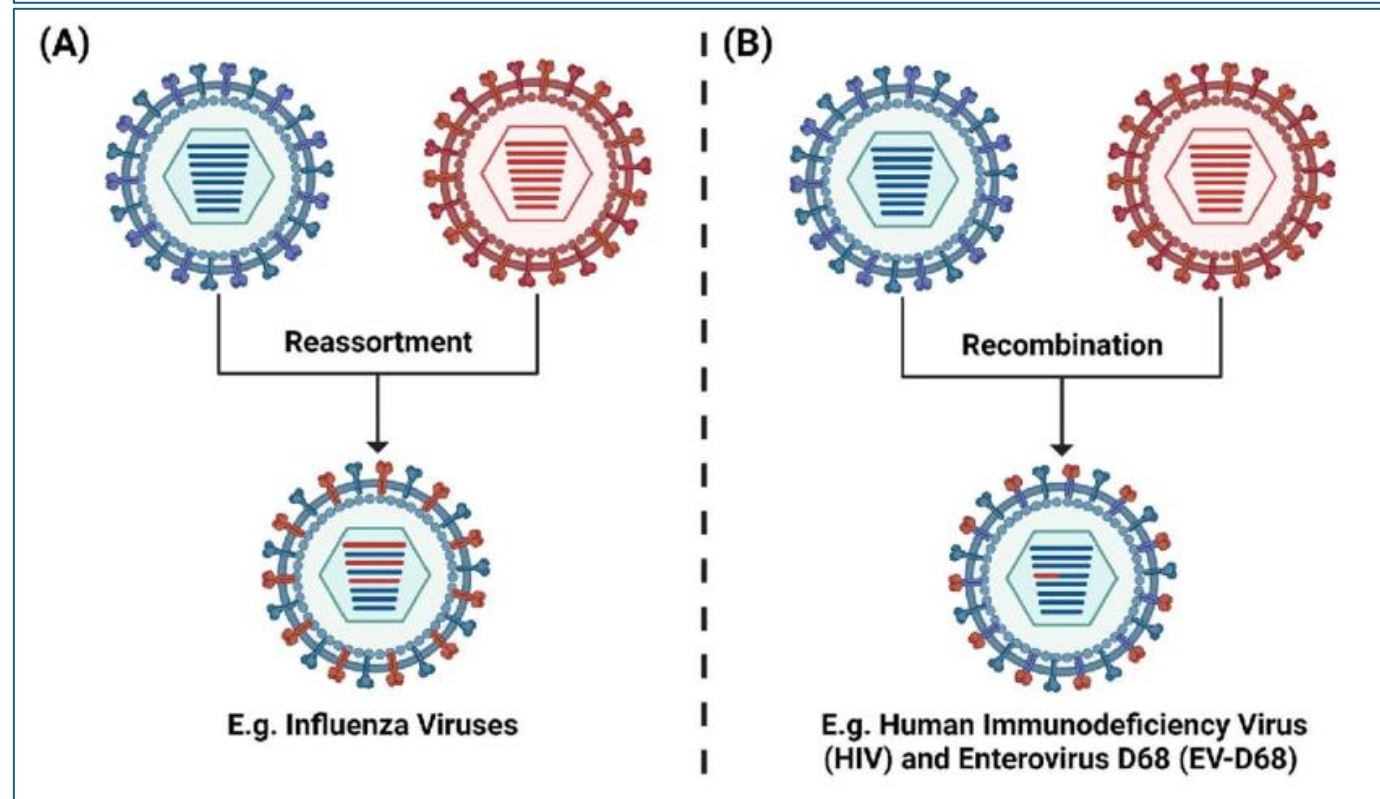
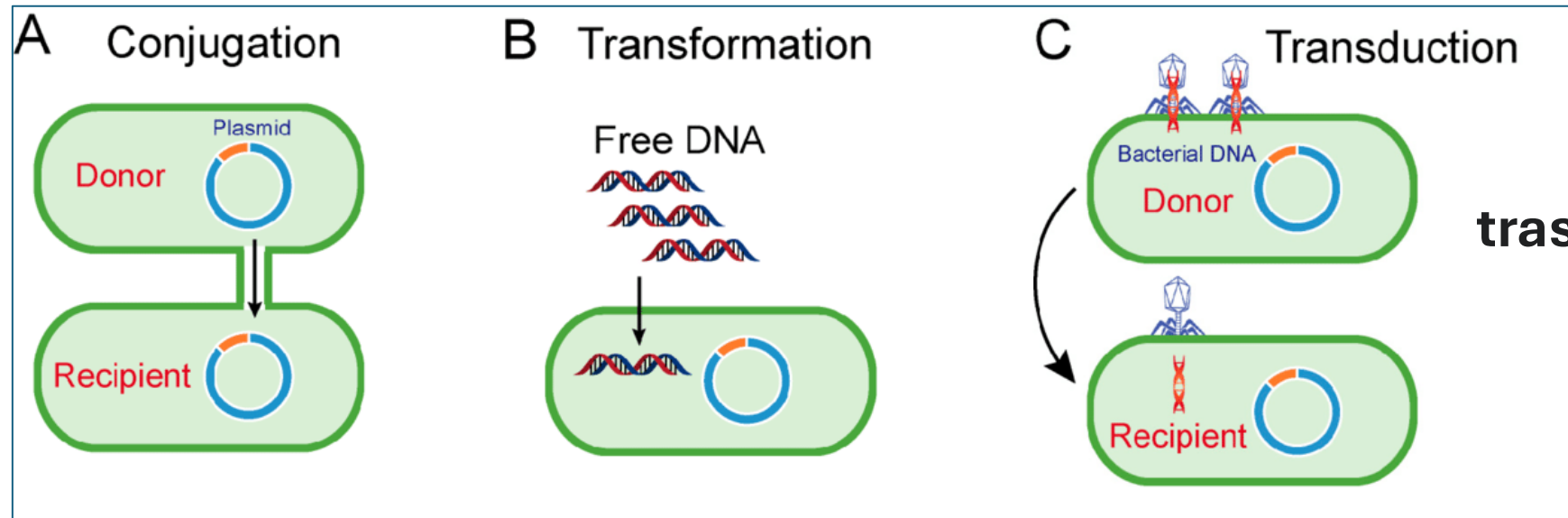
$$\begin{aligned} S_1 &= \{a, b, c\} | \{d, e, f\} \\ S_2 &= \{a, e, f\} | \{b, c, d\} \\ S_3 &= \{a, f\} | \{b, c, d, e\} \\ S_4 &= \{a, b, f\} | \{c, d, e\} \end{aligned}$$

(b) Non-trivial splits



(c) Split $S_2 = \frac{\{a, e, f\}}{\{b, c, d\}}$





incomplete lineage sorting

Essay

Mathematics Is Biology's Next Microscope, Only Better; Biology Is Mathematics' Next Physics, Only Better

Joel E. Cohen

Can Biology Lead to New Theorems?

CAN BIOLOGY LEAD TO NEW THEOREMS?

BERND STURMFELS

ABSTRACT. This article argues for an affirmative answer to the question in the title. In future interactions between mathematics and biology, both fields will contribute to each other, and, in particular, research in the life sciences will inspire new theorems in “pure” mathematics. This point is illustrated by a snapshot of four recent contributions from biology to geometry, combinatorics and algebra.

Feature article

Grazie