



Filogenética molecular y su efecto en la clasificación de los organismos

Nataly Ruiz Quiñones

nruizq@gmail.com



¿Dónde estamos hoy?

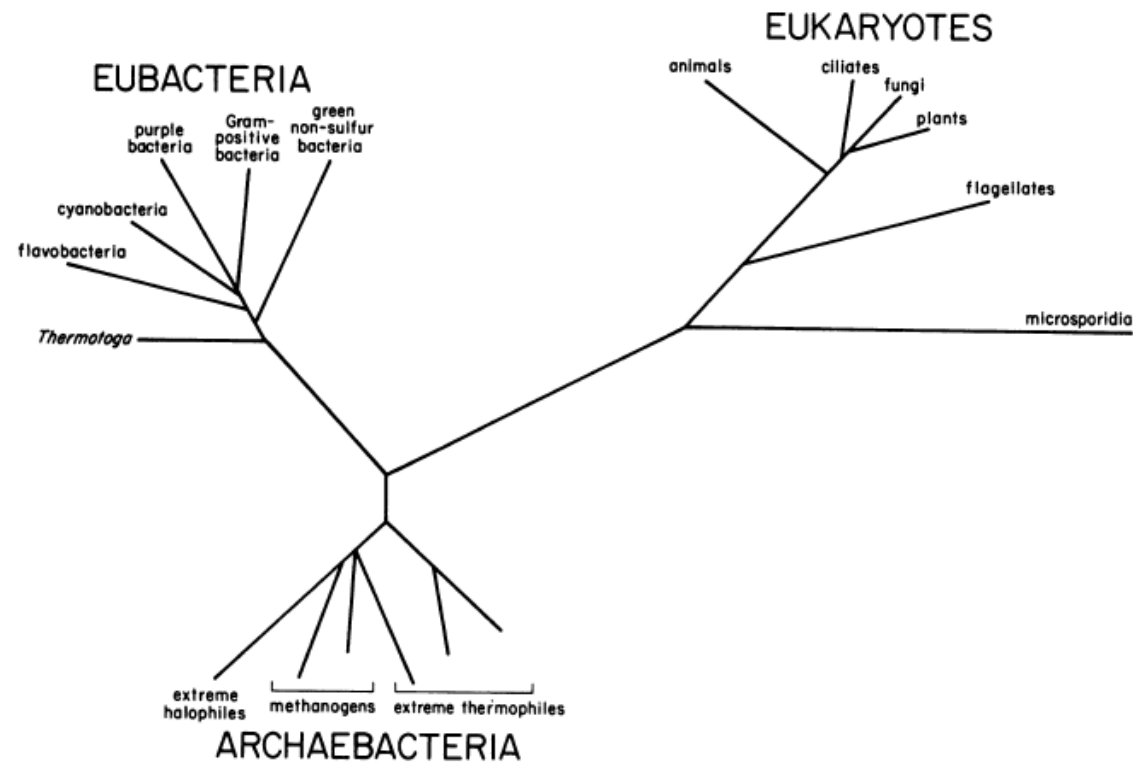
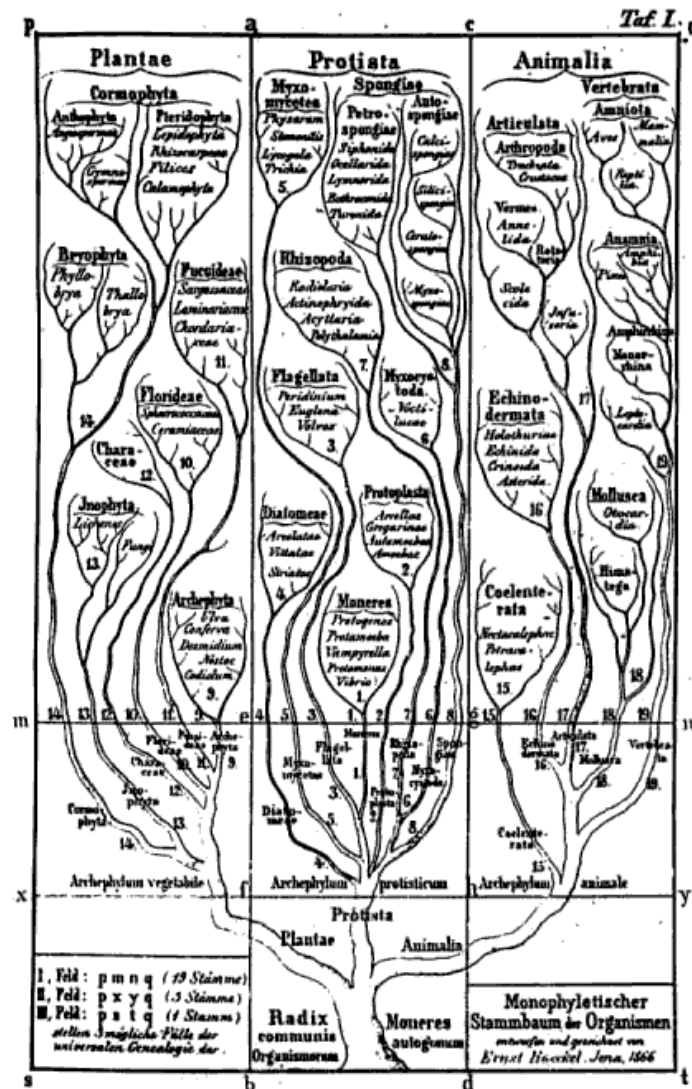
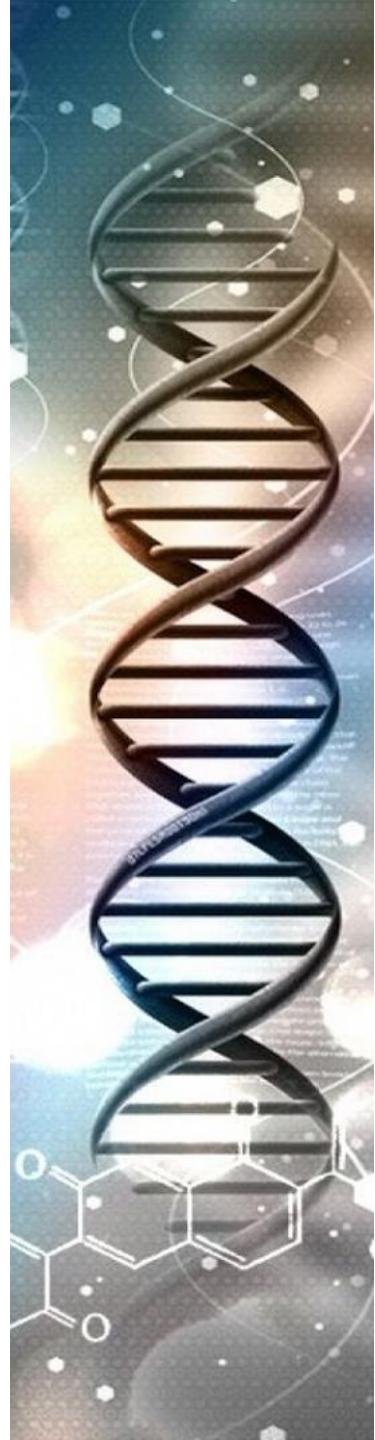


FIG. 4. Universal phylogenetic tree determined from rRNA sequence comparisons. A matrix of evolutionary distances (99) was calculated from an alignment (260) of representative 16S rRNA sequences from each of the three urkingdoms. This was used to construct a distance tree (36), based upon those positions represented in all sequences in the alignment in homologous secondary structural elements (75). Line lengths on the tree are proportional to calculated distances. The alignment includes the following eubacterial sequences: *Thermotoga maritima* (1); green non-sulfur bacteria, *Thermomicrobium roseum* (162); flavobacteria, *Flavobacterium heparinum* (234); cyanobacteria, *Anacystis nidulans* (224); gram-positive bacteria, *Bacillus subtilis* (68); and purple bacteria, *Escherichia coli* (19); the following archaeobacterial sequences: extreme halophiles, *Halobacterium volcanii* (72); methanogens, *Methanococcus vannielii* (96) and *Methanobacterium formicicum* (124); and extreme thermophiles, *Thermococcus celer* (Woese et al., unpublished data), *Desulfurococcus mobilis* (R. Garrett, personal communication), and *Thermoproteus tenax* (126); and the following eucaryotic sequences: microsporidia, *Vairimorpha necatrix* (226a); flagellates, *Euglena gracilis* (196); cellular slime molds, *Dictyostelium discoideum* (145); ciliates, *Paramecium tetraurelia* (195); fungi, *Saccharomyces cerevisiae* (179); plants, *Zea mays* (147); and animals, *Xenopus laevis* (181). Branching order within each kingdom is correct to a first approximation only. See the trees for the individual kingdoms for precise branching orders.



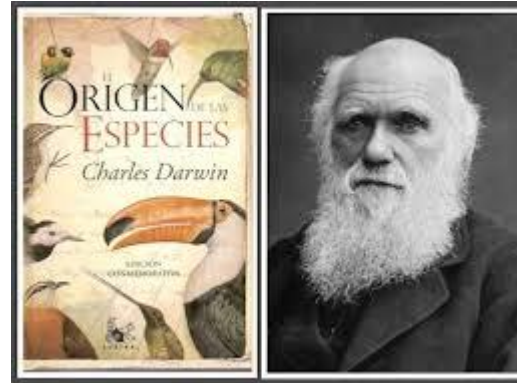
¿Cómo funciona la evolución?

1859- Charles Darwin

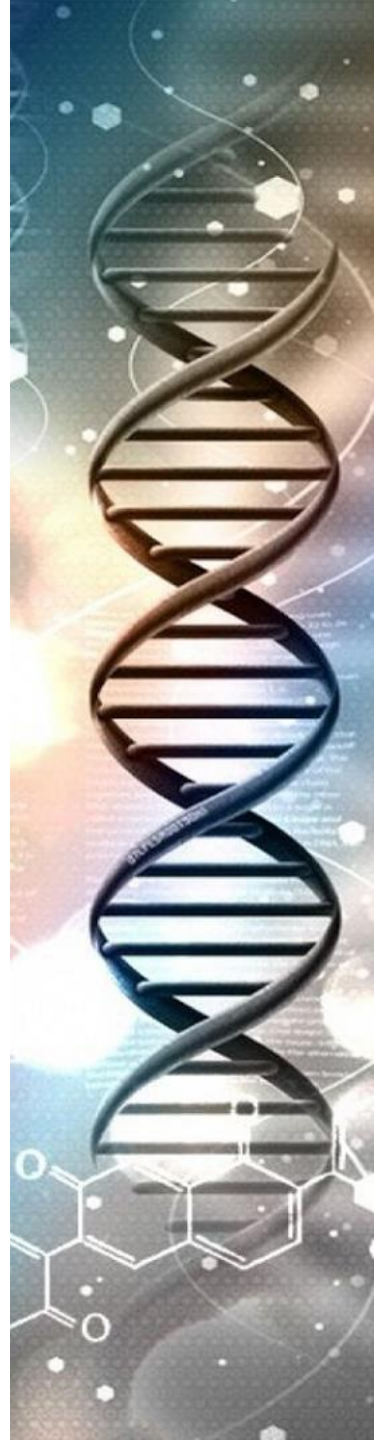
- Todos los organismos tienen un ancestro en común

Evolución Darwiniana

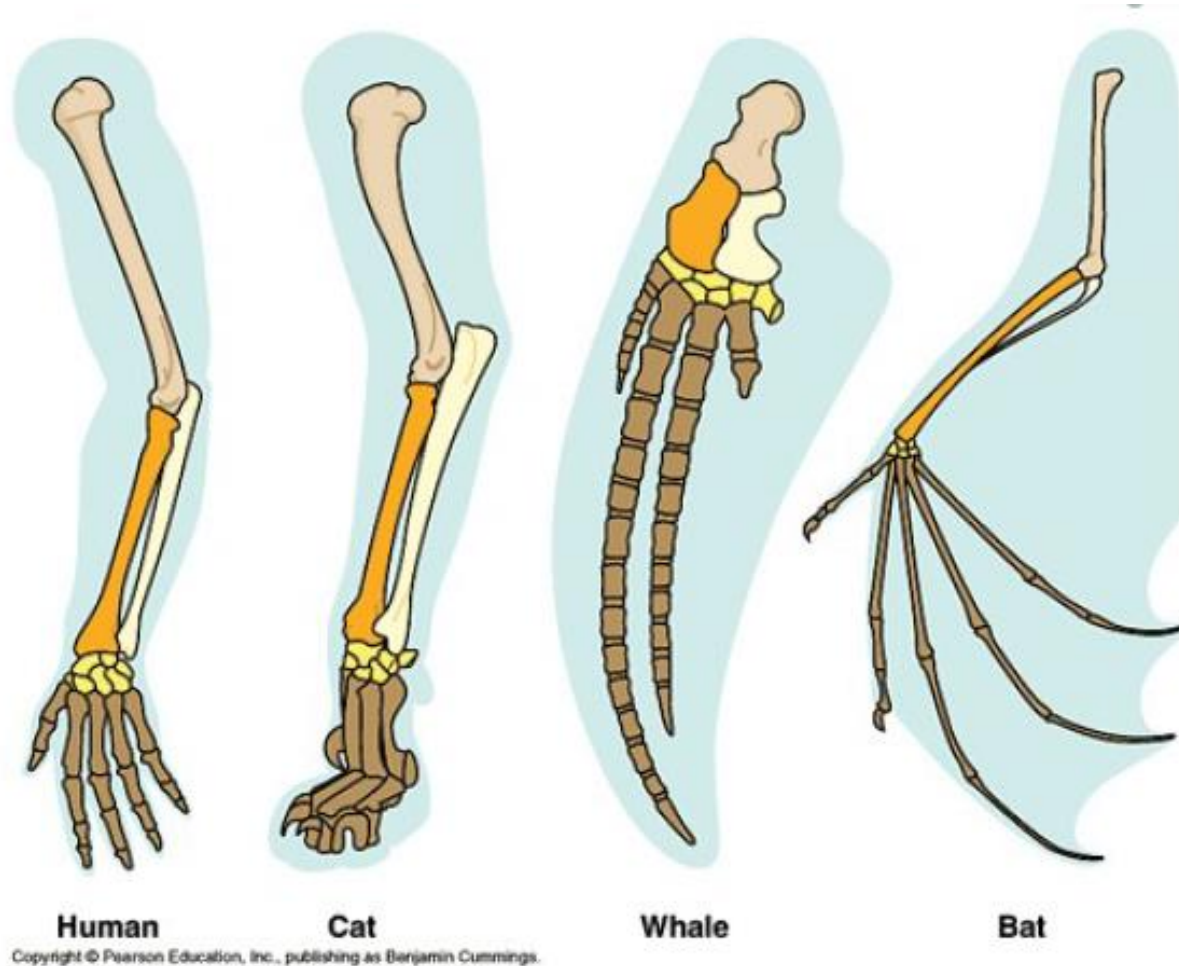
- Neutral
- Deriva genética
- Selección natural



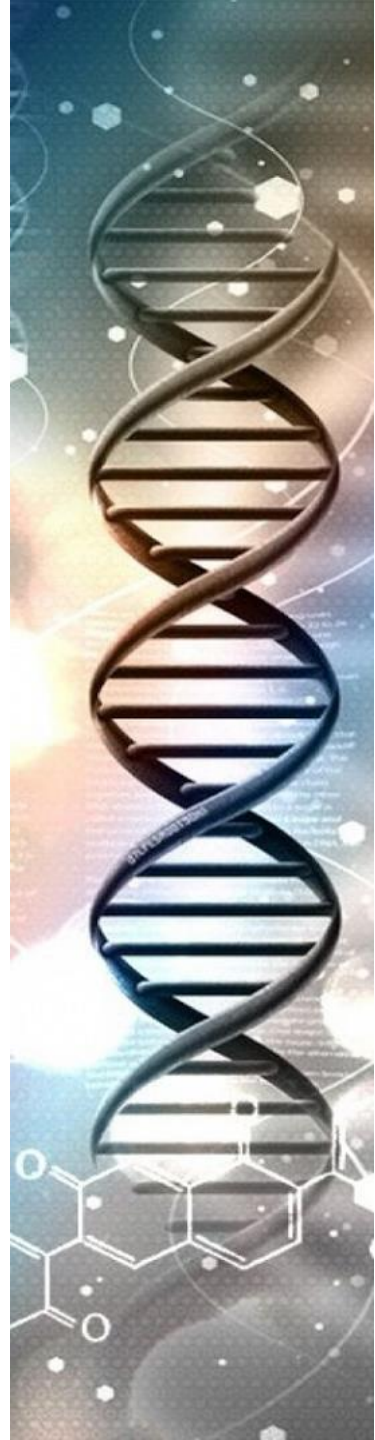
https://images.squarespace-cdn.com/content/v1/53a98e28e4b0356c592d592e/1452191912937-KJF342M917PL69TVX272/ke17ZwdGBToddI8pDm48kH1CVTYTI9Xgs88q_IK6NU1Zw-zPPgdn4jUwVcJE1ZvWQUxwkmyExglNqGp0l vTJZUJFbgE-7XRK3dMEBRBhUpzh1jRGfZrU0BugfrubTJnriCLtnl1FKfQknIOg4dNslZi8zA_-doRitpHFvGbG0cM/image-asset.jpeg?format=750w



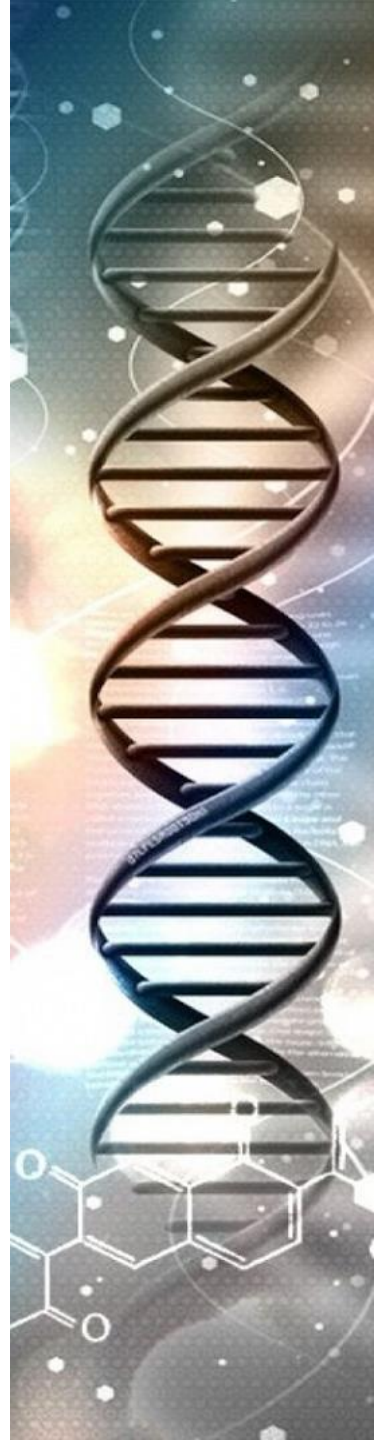
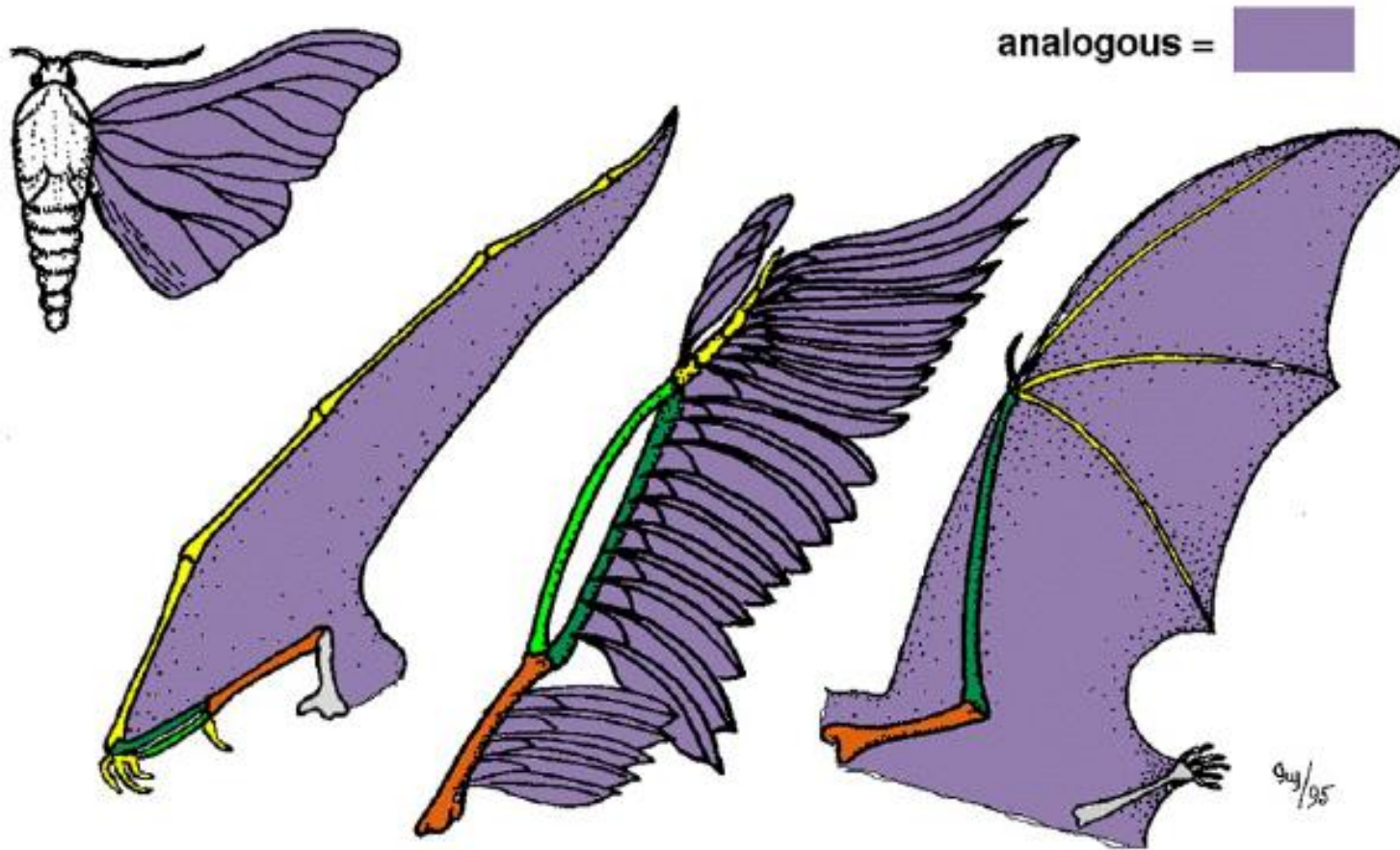
¿Cómo funciona la evolución? Homólogas



Lobo, 2014



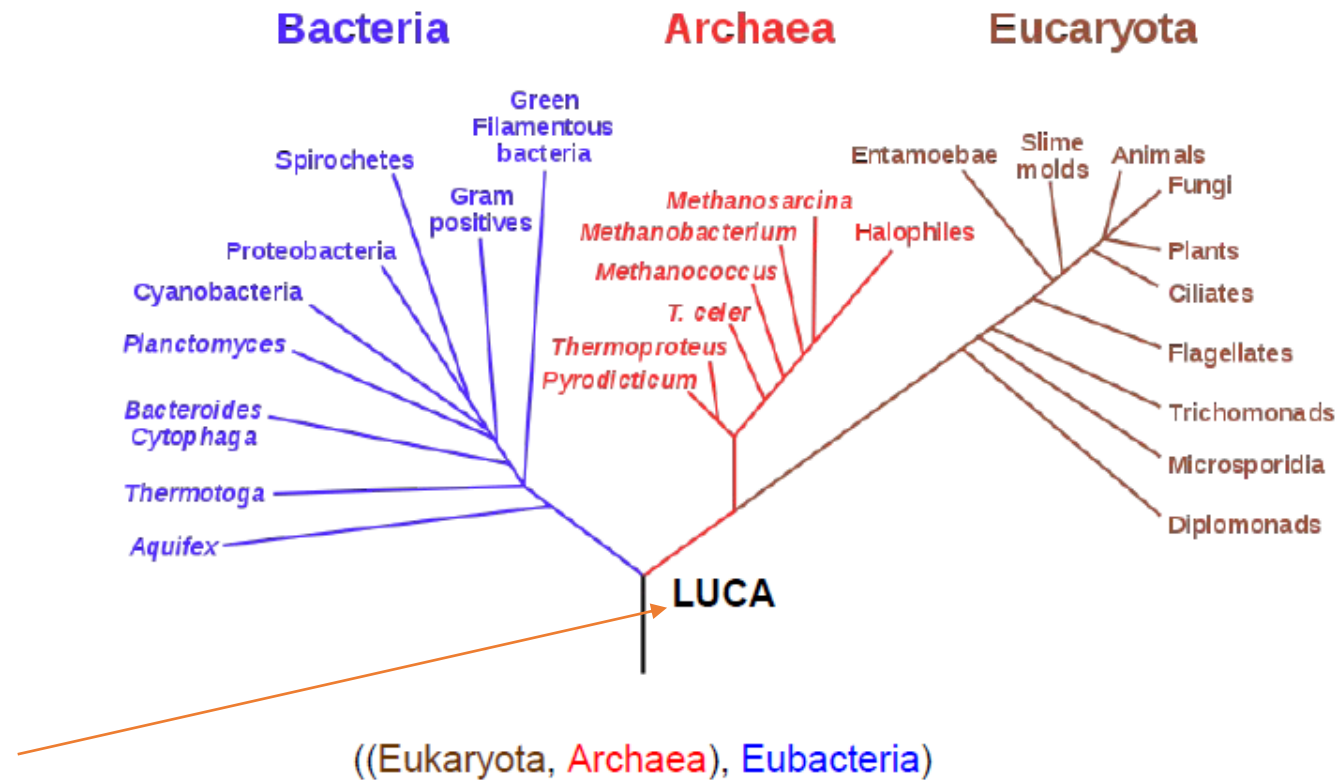
¿Cómo funciona la evolución? Analogías



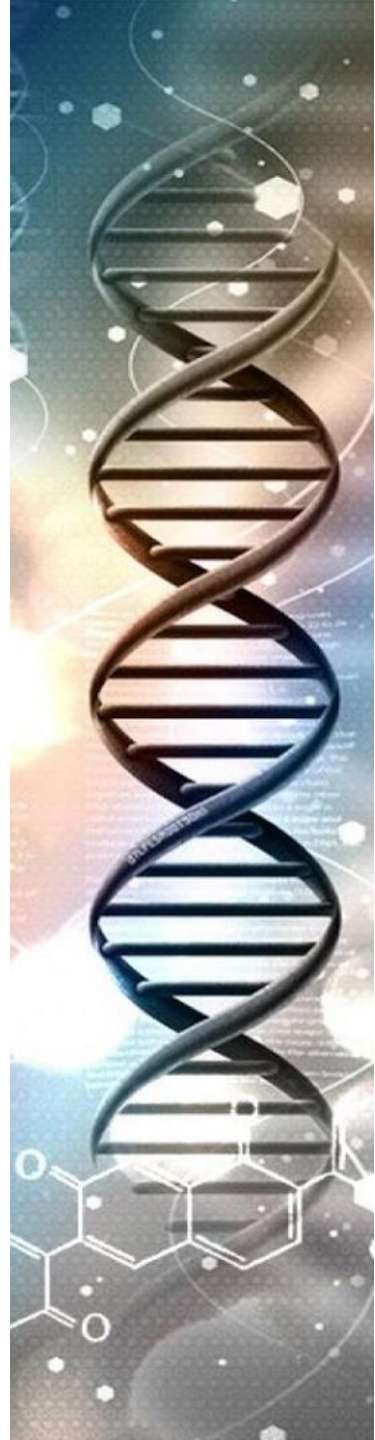
Características plesiomórficas

- DNA – RNA virus.
- Código genético conservado.
- Mismos mecanismos genéticos.
- Genes compartidos

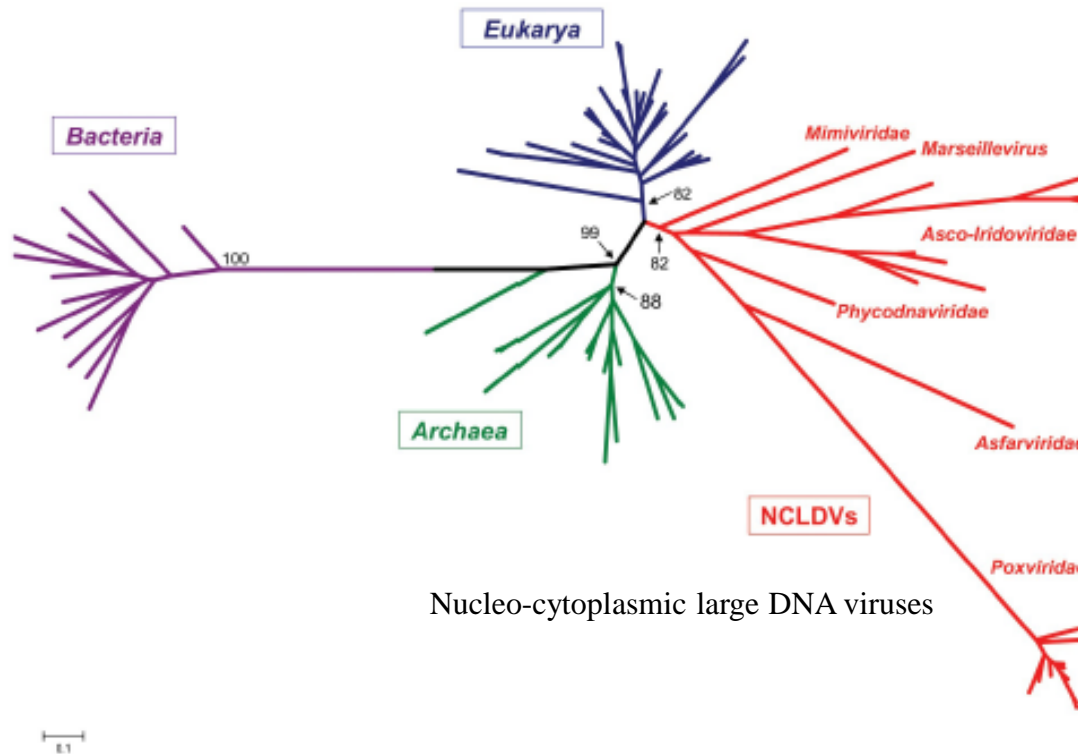
LUCA (Last Unique Common Ancestor)



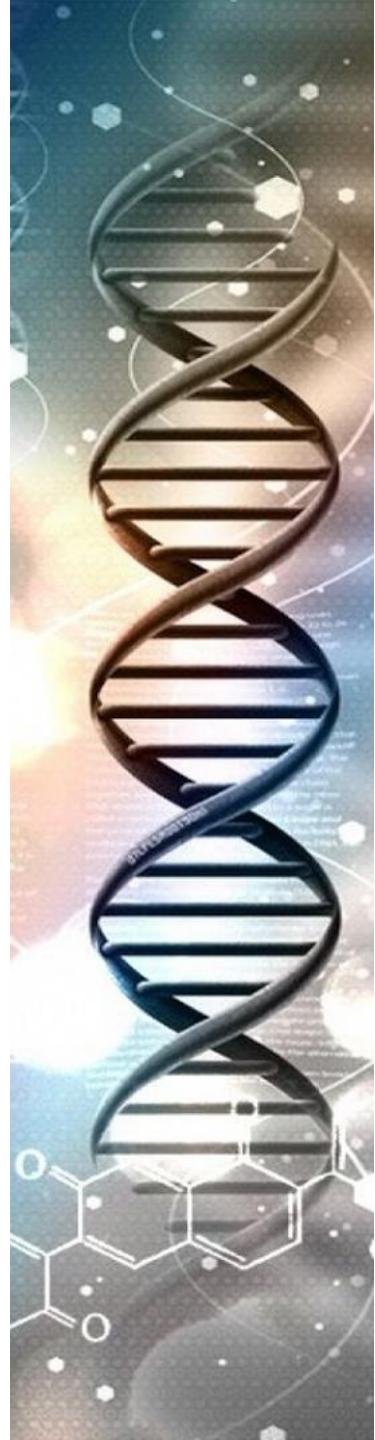
Lobo, 2014



¿El cuarto dominio?

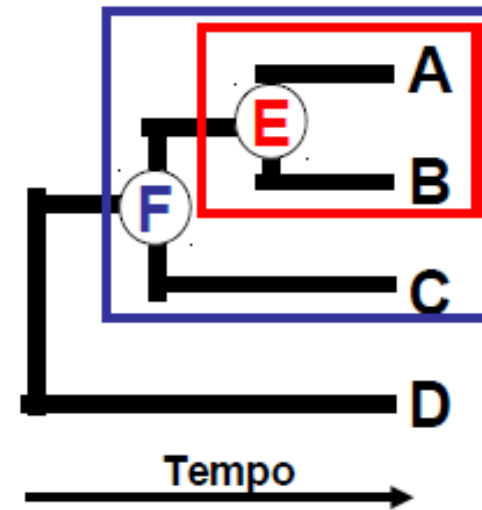
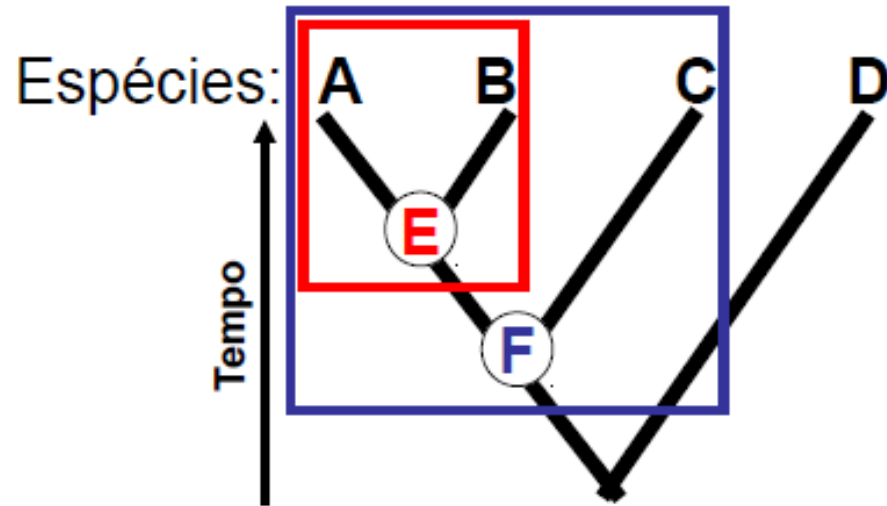


Stalking the Fourth Domain in Metagenomic Data: Searching for, Discovering, and Interpreting Novel, Deep Branches in Marker Gene Phylogenetic Trees. Wu D, Wu M, Halpern A, Rusch DB, Yooseph S, et al. (2011) Stalking the Fourth Domain in Metagenomic Data: Searching for, Discovering, and Interpreting Novel, Deep Branches in Marker Gene Phylogenetic Trees. PLoS ONE 6(3): e18011. doi:10.1371/journal.pone.0018011

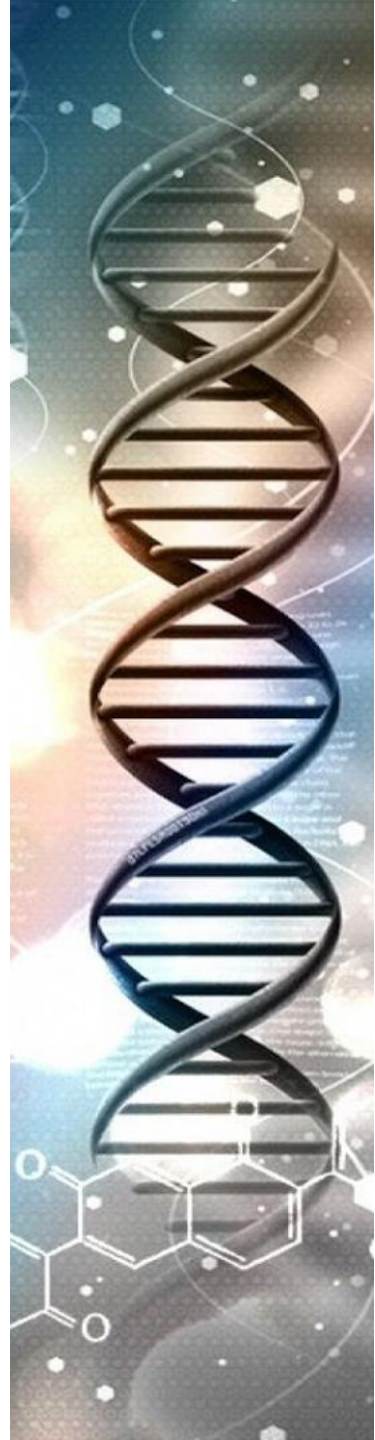


Construir una filogenia

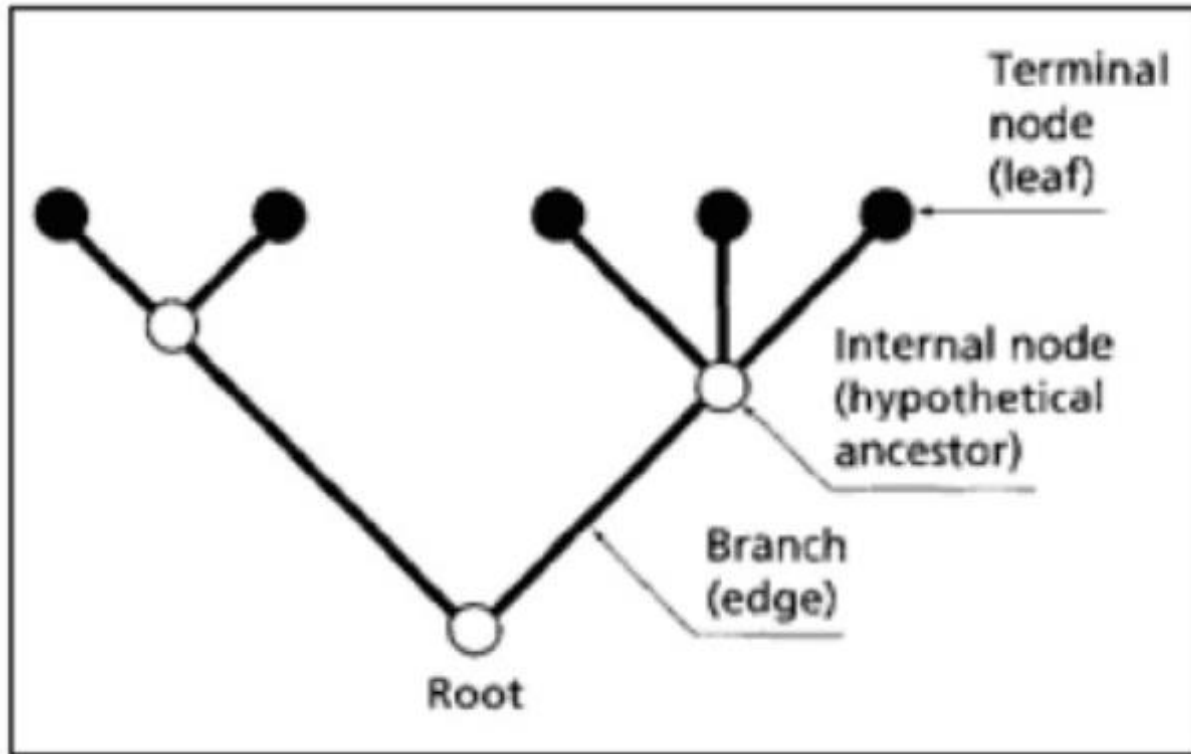
Las relaciones entre organismos a lo largo del tiempo



Lobo, 2014



Estructuras de un árbol



Lobo, 2014

362 Algebraic and Combinatorial Computational Biology

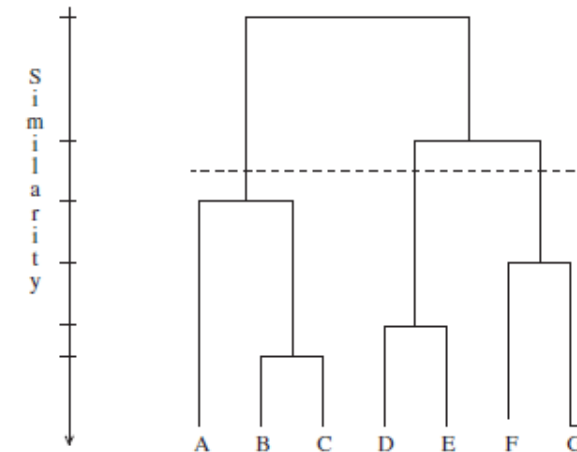


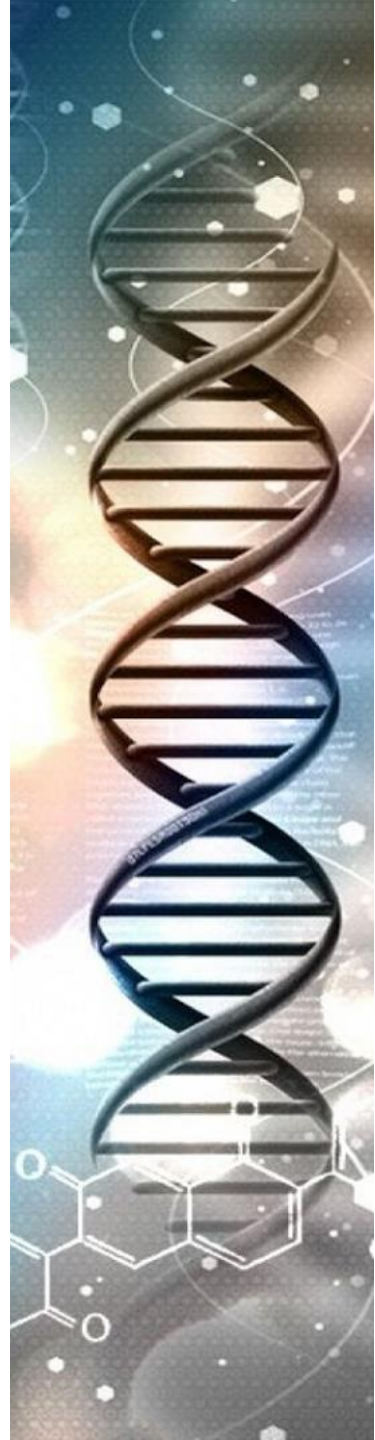
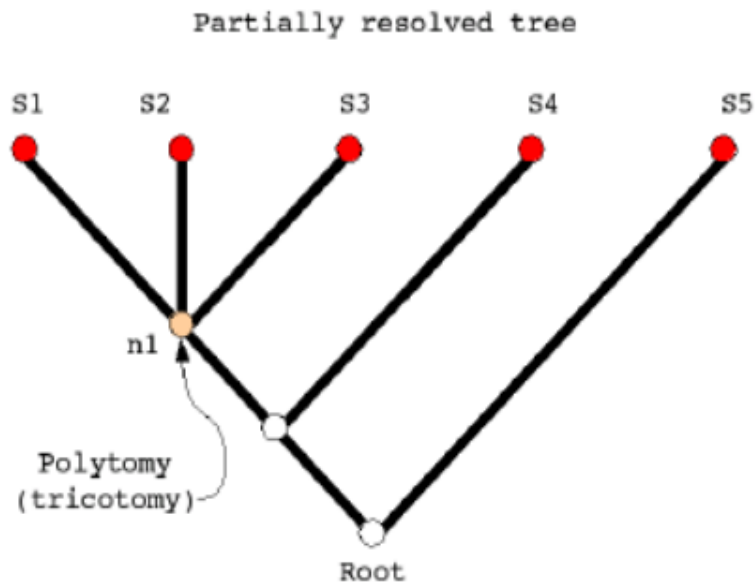
FIG. 11.5 The dendrogram representation of the results of a hierarchical agglomerative clustering algorithm. The *dashed line* serves as the stopping point of this particular algorithm and produces three clusters. One clusters contains A, B, and C. Another contains D and E. The final cluster contains F and G. An algorithm with a stricter similarity criterion would result in fewer clusters at the stopping point. Likewise, a looser criterion would result in more clusters. (Credit: Henriquerocha/PD-self.)



Estructuras de un árbol

Taxón = operational taxonomic unit (OTU) S1-S5

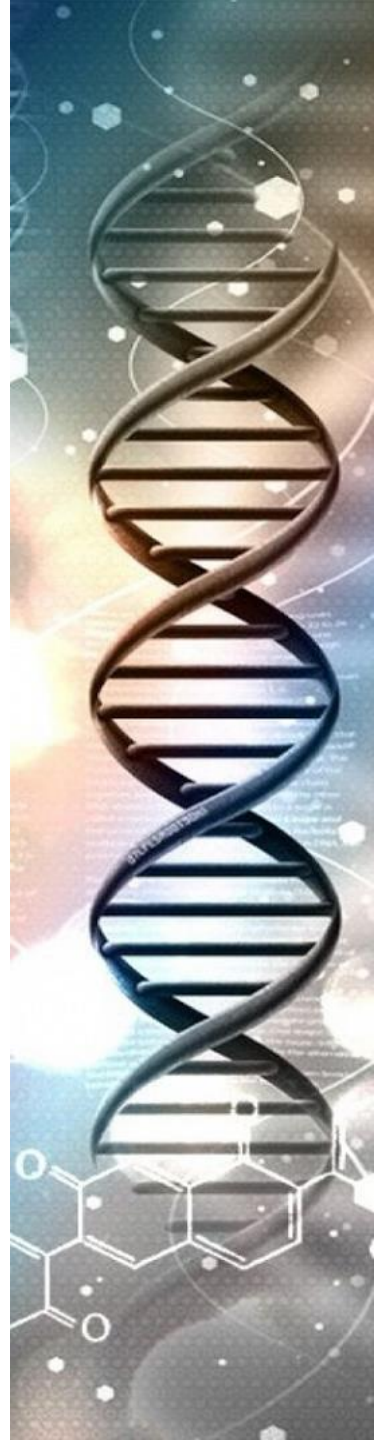
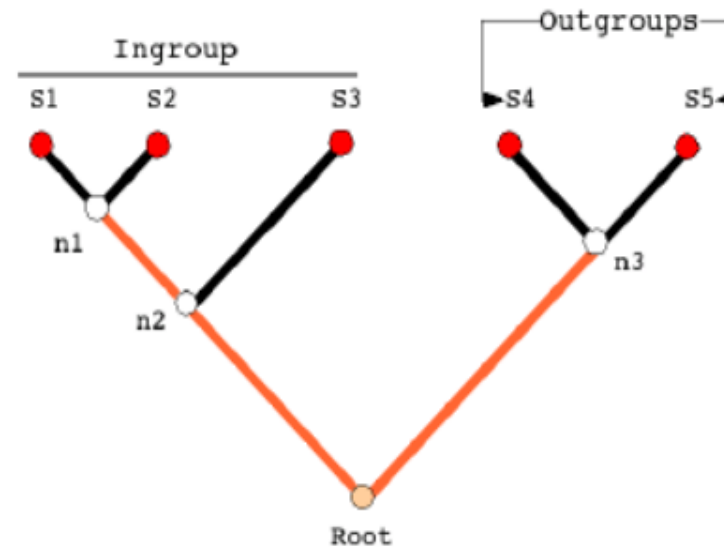
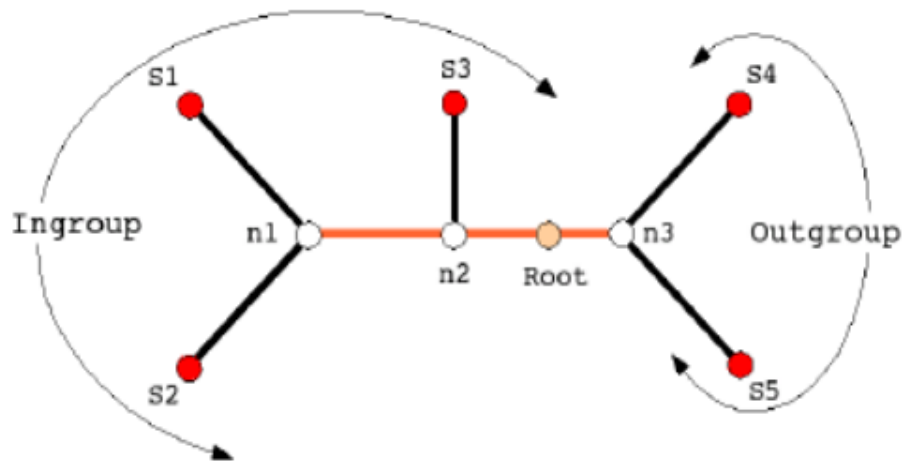
Politomía: árbol no resuelto (se puede intentar añadir más secuencias para resolverlo)



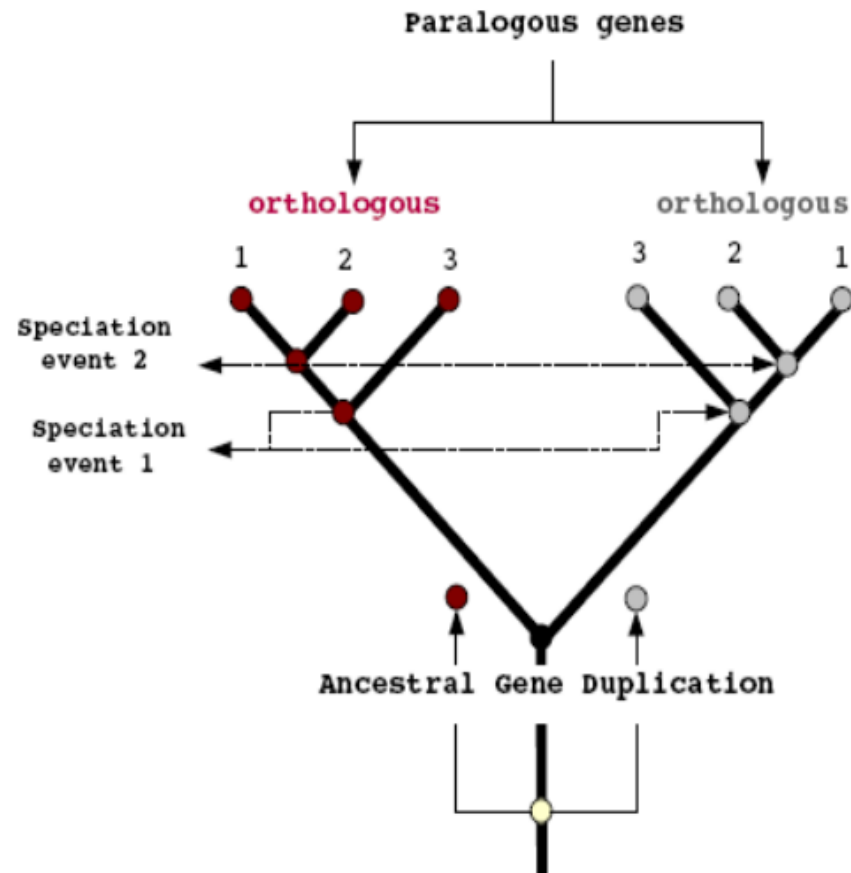
Estructuras de un árbol

Árboles sin raíz (unrooted) o enraizados (rooted)

Outgroup: secuencia/s relacionadas con las secuencias del árbol (ingroup)



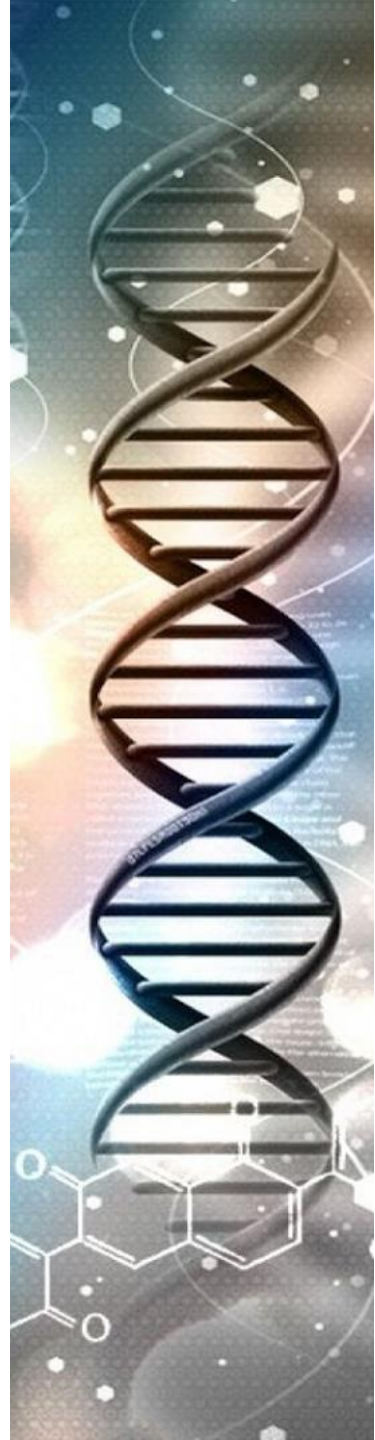
Estructuras de un árbol



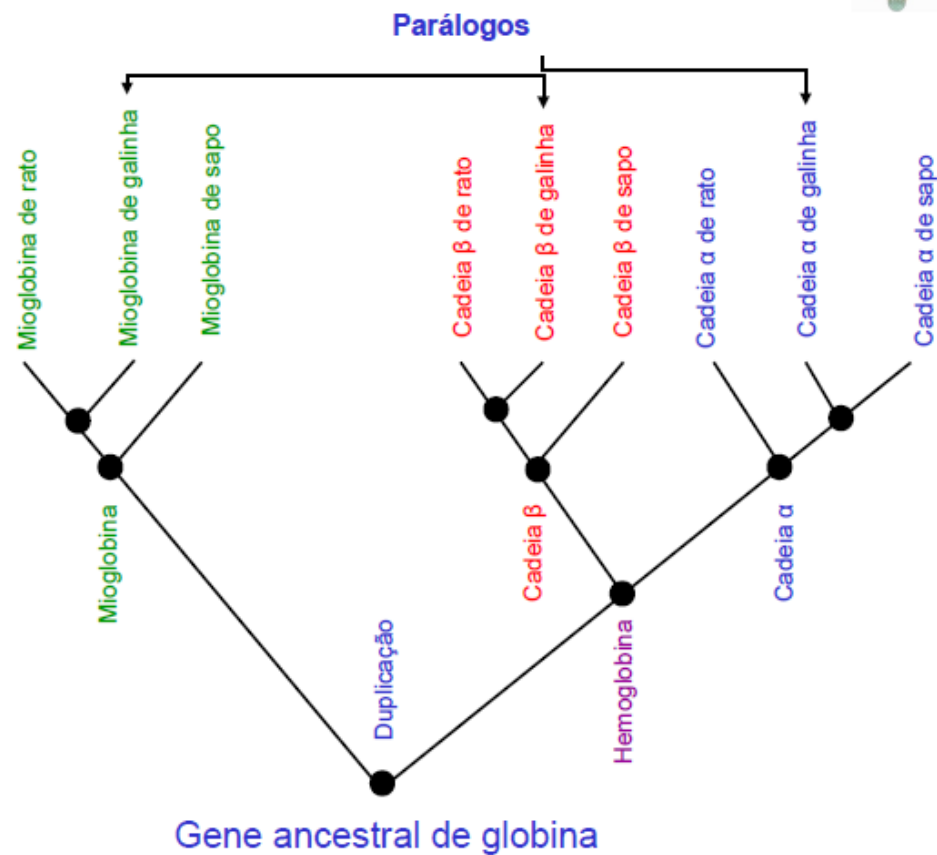
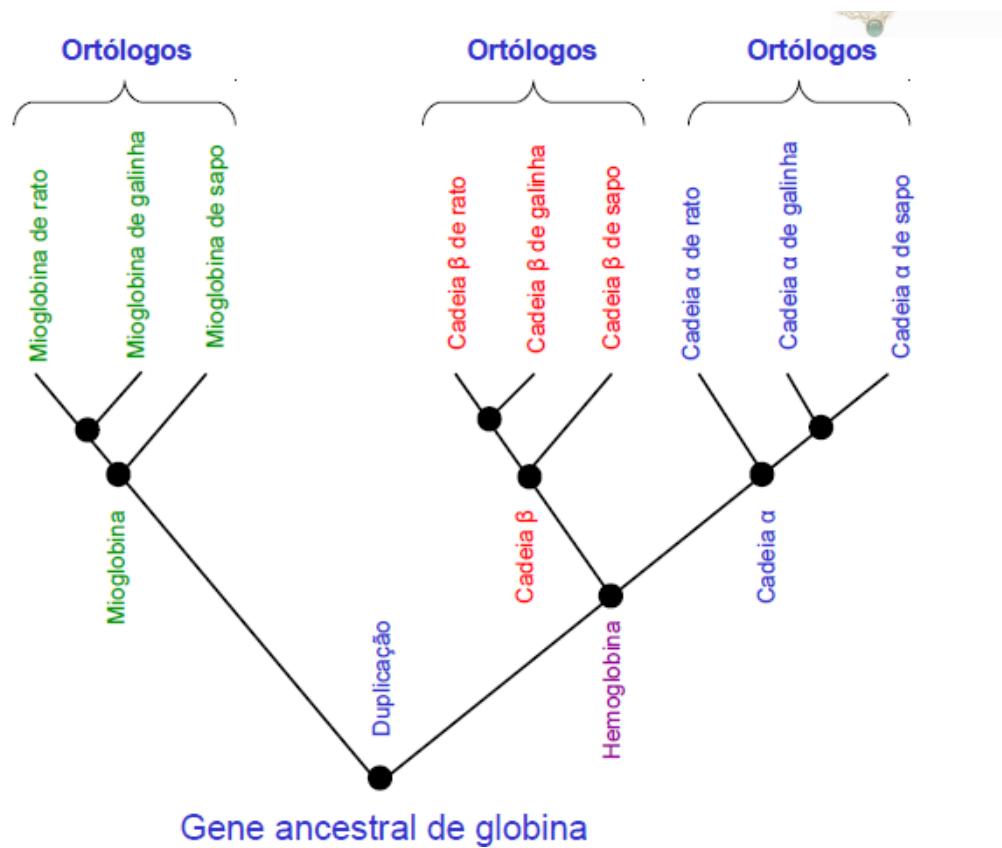
Parálogo: genes que sufrieron divergencia y ya no cumplen una misma función.

Ortólogo: genes que comparten un ancestro común y conservan su funcionalidad.

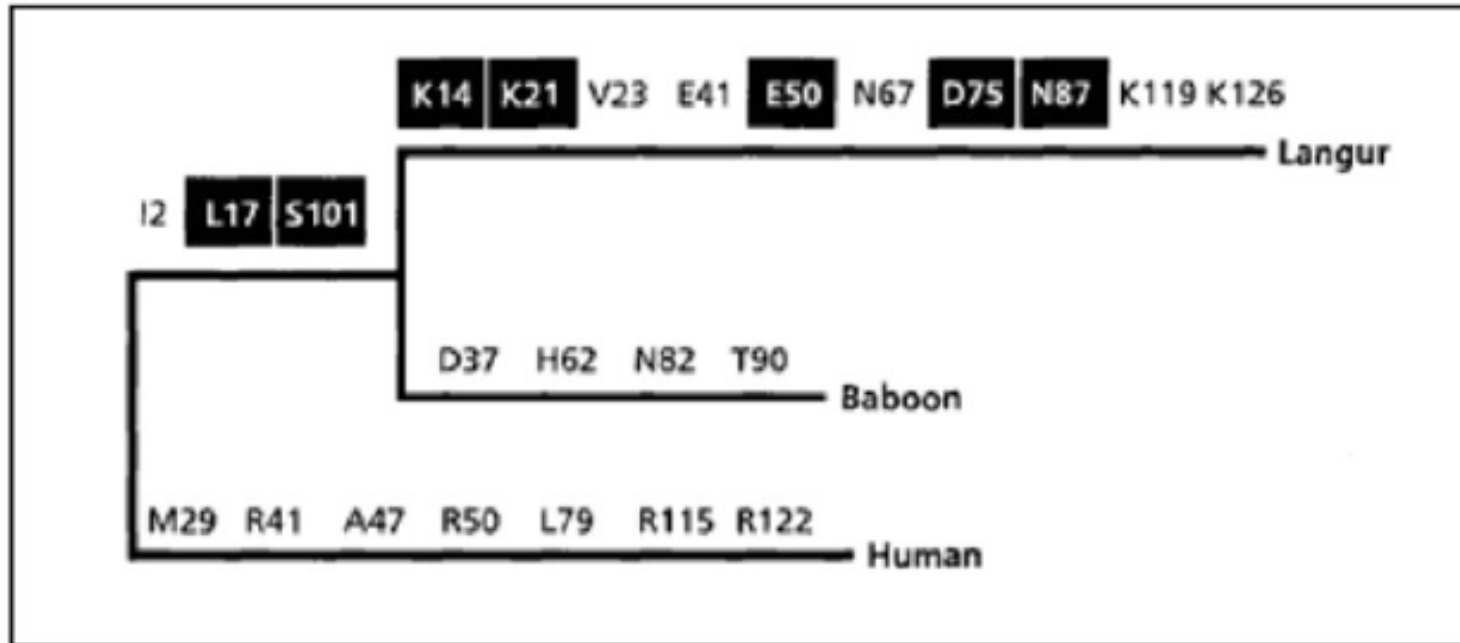
Ejemplo: los rojos son ortólogos entre ellos, pero parálogos respecto a los grises.



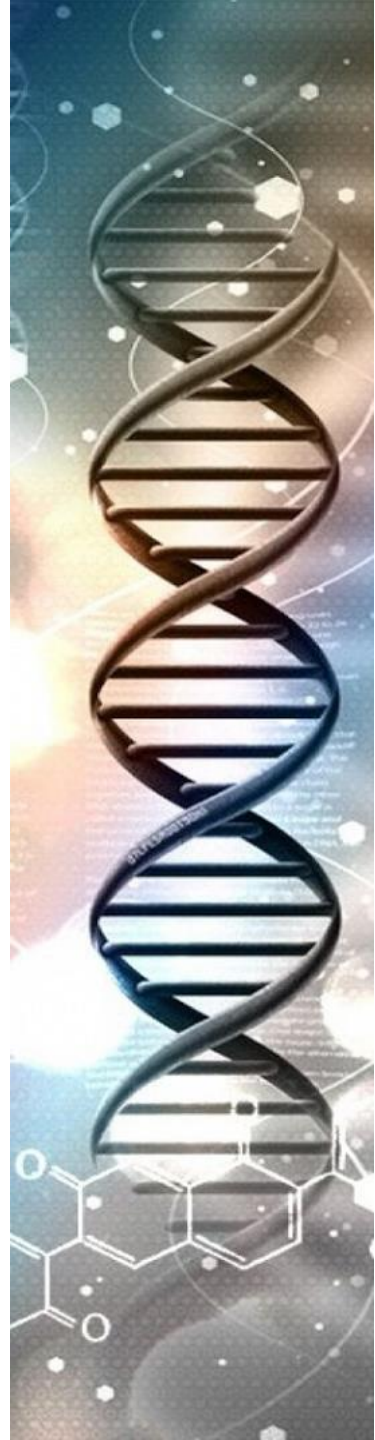
Estruturas de um árbol



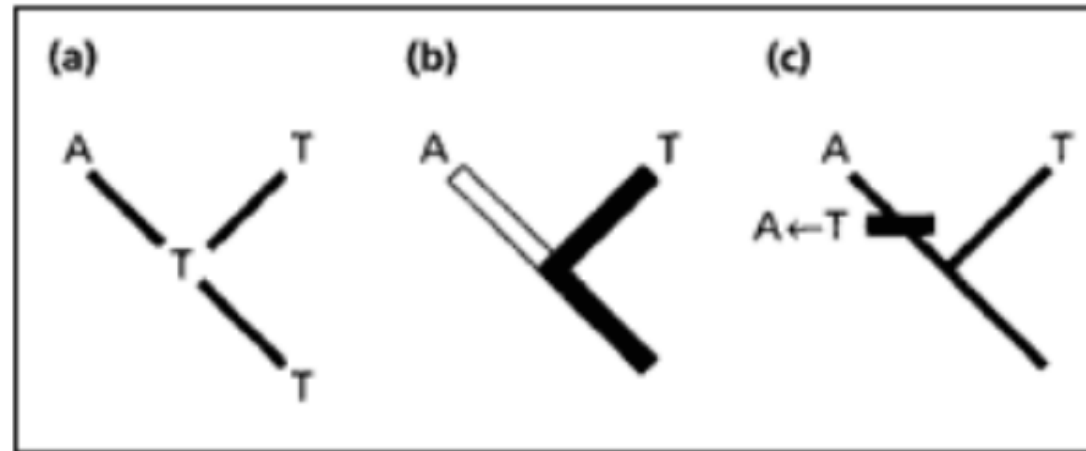
Representaciones en el árbol



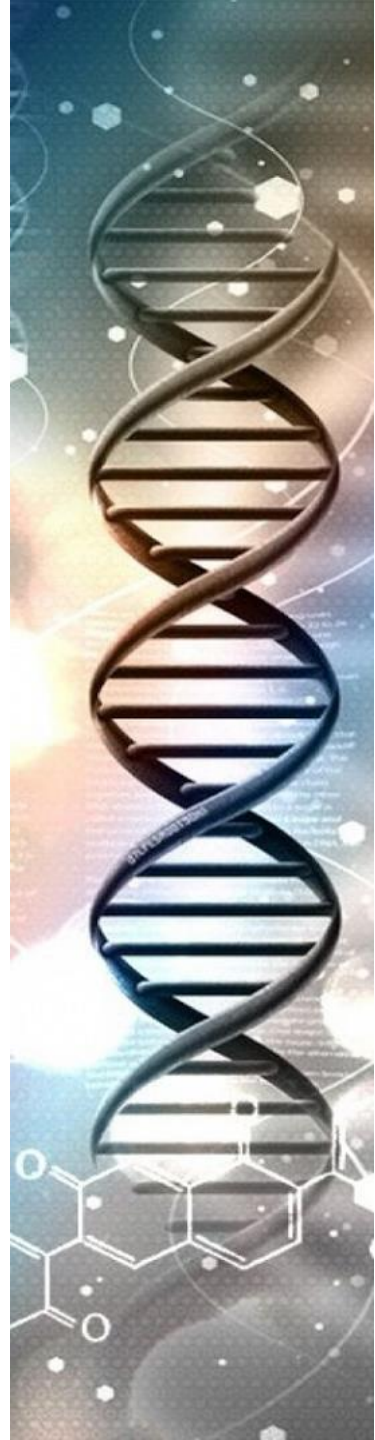
Las distancias en el árbol sólo se estima entre genes homólogos



Representaciones en el árbol



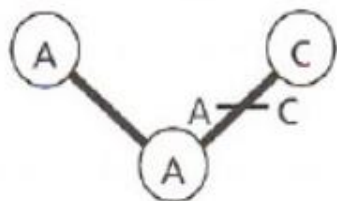
Lobo, 2014



Representaciones en el árbol

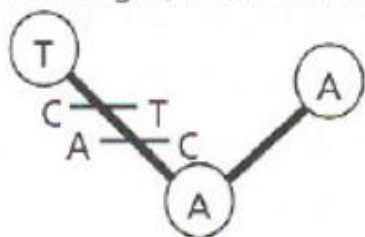
(a) Single substitution

1 change, 1 difference



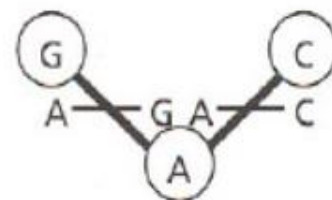
(b) Multiple substitution

2 changes, 1 difference



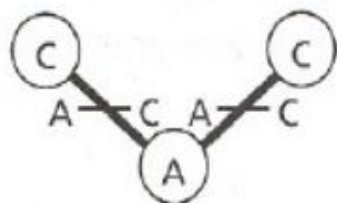
(c) Coincidental substitution

2 changes, 1 difference



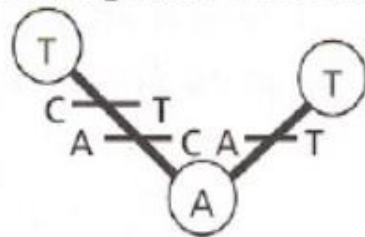
(d) Parallel substitution

2 changes, no difference



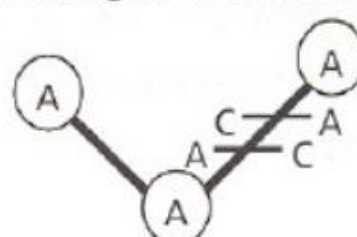
(e) Convergent substitution

3 changes, no difference

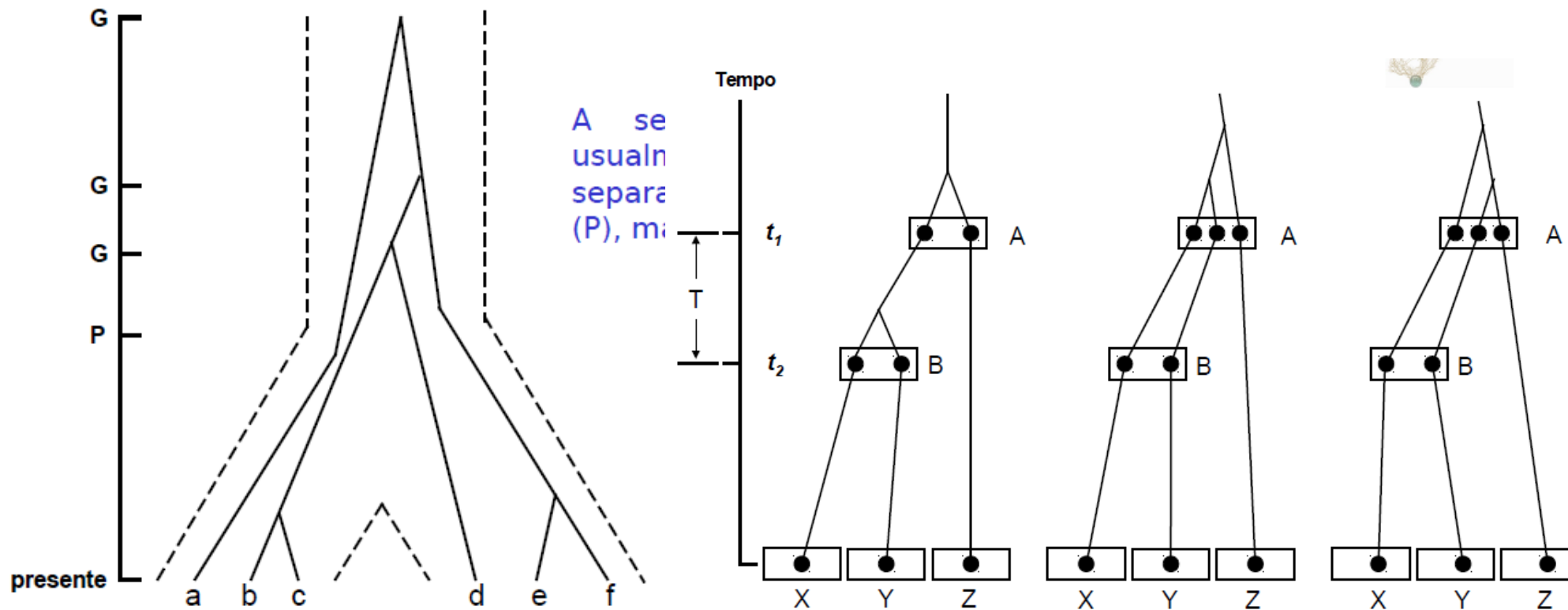


(f) Back substitution

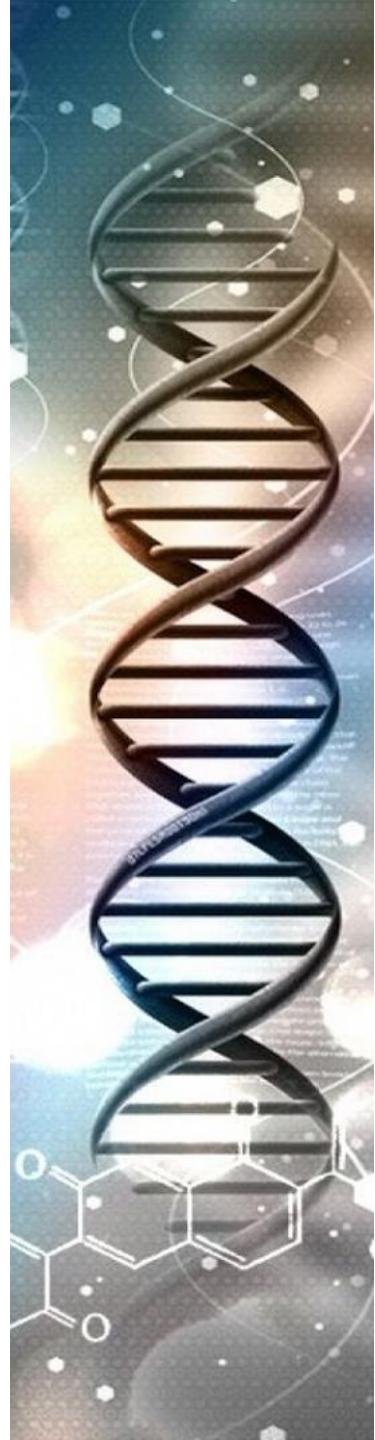
2 changes, no difference



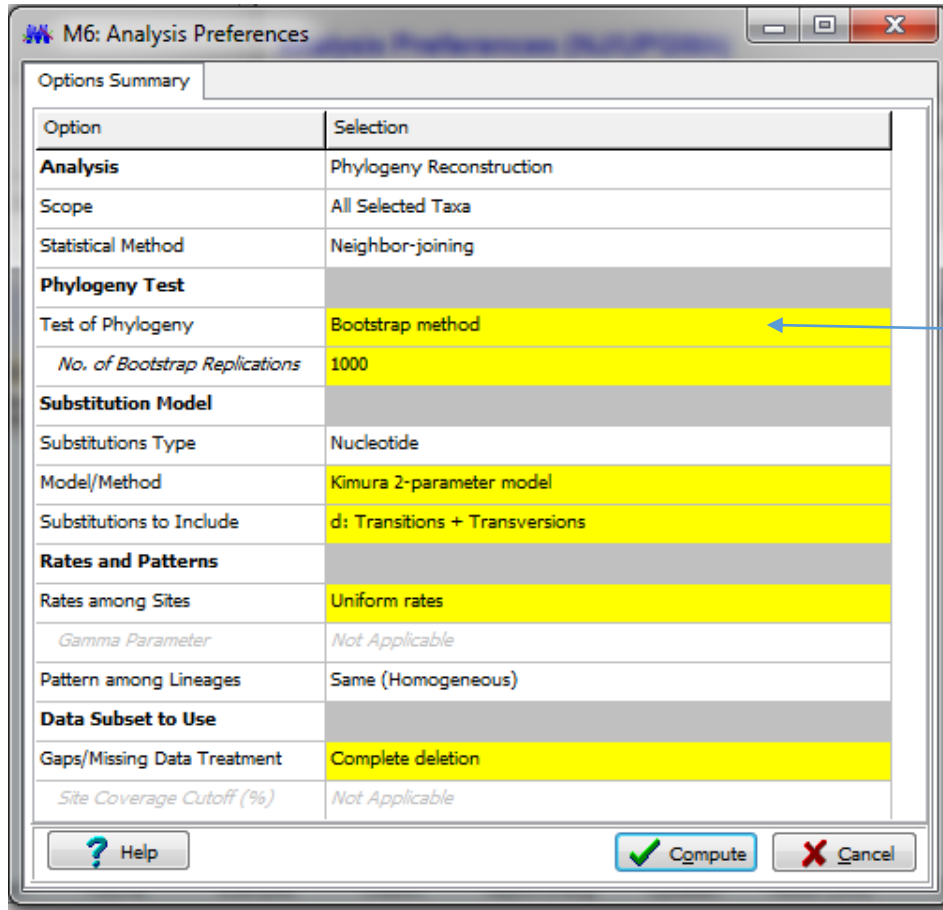
Representaciones en el árbol



Árboles coalescentes.



Construyendo árboles filogenéticos



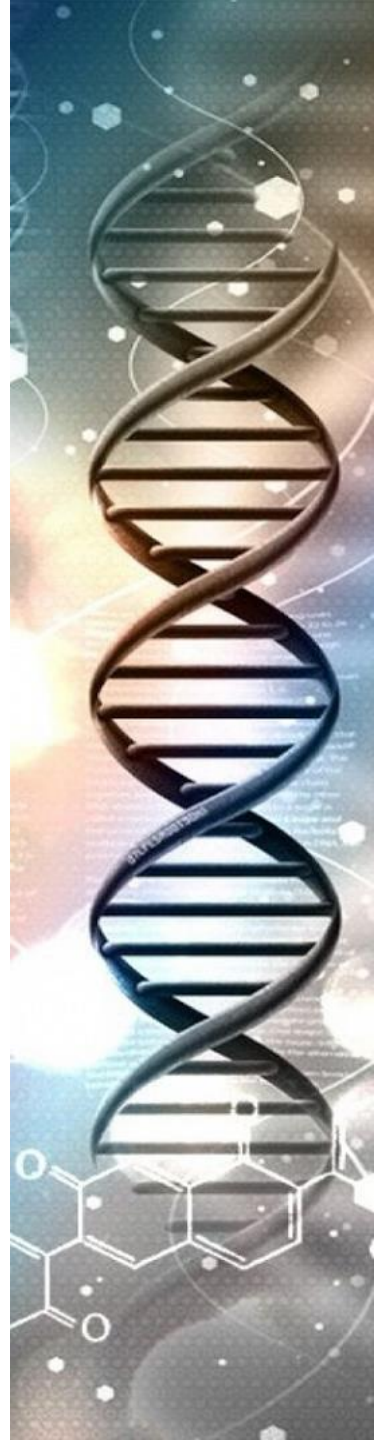
M6: Analysis Preferences

Options Summary

Option	Selection
Analysis	Phylogeny Reconstruction
Scope	All Selected Taxa
Statistical Method	Neighbor-joining
Phylogeny Test	
Test of Phylogeny	Bootstrap method
No. of Bootstrap Replications	1000
Substitution Model	
Substitutions Type	Nucleotide
Model/Method	Kimura 2-parameter model
Substitutions to Include	d: Transitions + Transversions
Rates and Patterns	
Rates among Sites	Uniform rates
Gamma Parameter	Not Applicable
Pattern among Lineages	Same (Homogeneous)
Data Subset to Use	
Gaps/Missing Data Treatment	Complete deletion
Site Coverage Cutoff (%)	Not Applicable

? Help ✓ Compute ✗ Cancel

Modelo estadístico:
Confianza de los ramos
realizando un remuestreo.

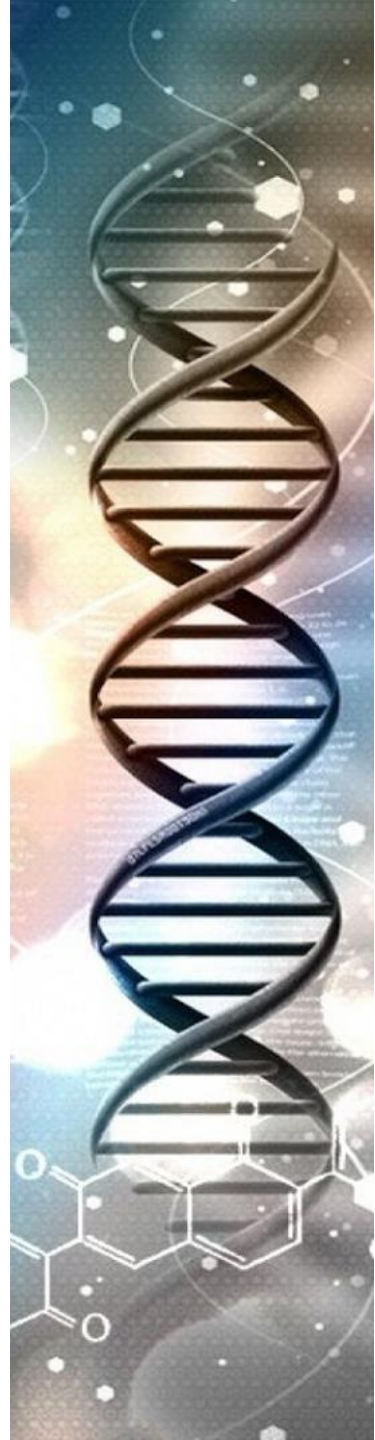


Construyendo árboles filogenéticos

Métodos basados en distancias:

UPGMA (unweighted pair group method with arithmetic mean):
Agrupa los pares que tienen la menor diferencia entre ellos y recalcula la matriz a partir de cada grupo formado.

Neighbor-joining: Crea una matriz de distancias para resolver un árbol que presenta topología de estrella.



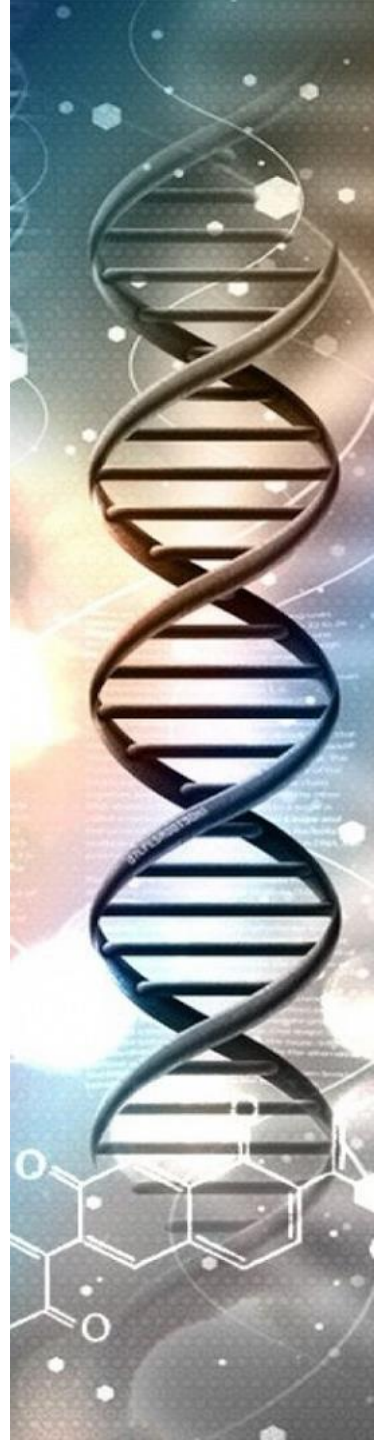
Construyendo árboles filogenéticos

Métodos basados en caracteres:

Buscan el mejor árbol

Parsimonia: incluya menos ramos, no tiene modelo evolutivo.

Máxima verosimilitud: Busca el árbol que ha generado las diferencias encontrada en los datos pero depende del modelo evolutivo.



Calculando distancias evolutivas

Distancia-p: Sólo compara presencia ausencia



Certhidea olivacea



Tiaris obscura

<i>C. olivacea</i>	CATGCACTACACAGCAGACACCAACCTAGCCTTCTCCTCCGTGCCCCATATATGCCGAGA
<i>T. obscura</i>	CATACACTACACAGCAGACACTAACCTAGCCTTCTCTTCCGTGCCCCACATATGCCGAGA
<i>C. olivacea</i>	CGTCCAATTTCGGATGACTCATCCGCAACCTCCACGCAAACGGAGCCTCCTTCTTCTTCAT
<i>T. obscura</i>	CGTCCAATTTCGGCTGACTCATCCGCAACCTCCACGCAAACGGAGCCTCCTTCTTCTTCAT
<i>C. olivacea</i>	CTGCATCTACCTACACATCGGACGAGGAATCTACTACGGCTCATACTAAATAAAGAAAC
<i>T. obscura</i>	CTGCATCTACCTACACATCGGACGAGGAATCTACTACGGCTCATACTTAAACAAAGAAAC
<i>C. olivacea</i>	CTGAAACATTGGAGTCATCCTCCTAACCCTCATAGCAACAGCCTTTGTAGGATACGT
<i>T. obscura</i>	CTGAAACATTGGAGTCATATCCTCCTAACCCTCATAGCAACAGCCTTTGTAGGATACGT
<i>C. olivacea</i>	CCTACCATGAGGCCTAAATATCCTTCTGAGGGGCTACAGTAATCACAAACCTGTTCTCAGC
<i>T. obscura</i>	CCTACCATGAGGCTCAAATGTCCTTCTGAGGGGCTACCGTAATCACAAACCTATTCTCAGC

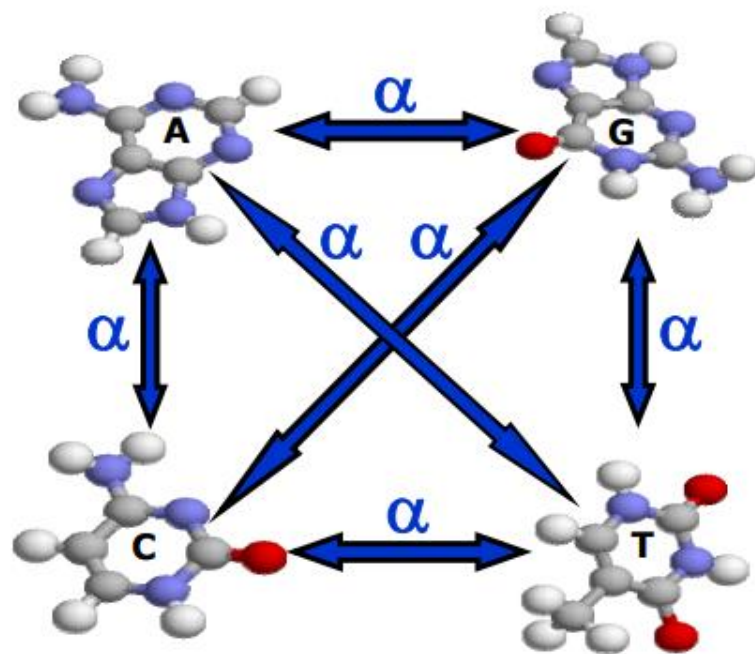
$$\text{Distânciagenética observada}(\hat{p}) = \frac{\text{número de sítios ocupados por nucleotídeos diferentes nas 2 sequências}}{\text{número total de sítios nas 2 sequências}}$$

$$\text{Distância genética observada}(\hat{p}) = \frac{15}{300} = 0,05$$



Calculando distancias evolutivas

Jukes-Cantor: frecuencia y tasa de substitución para todas las 4 bases es la misma.



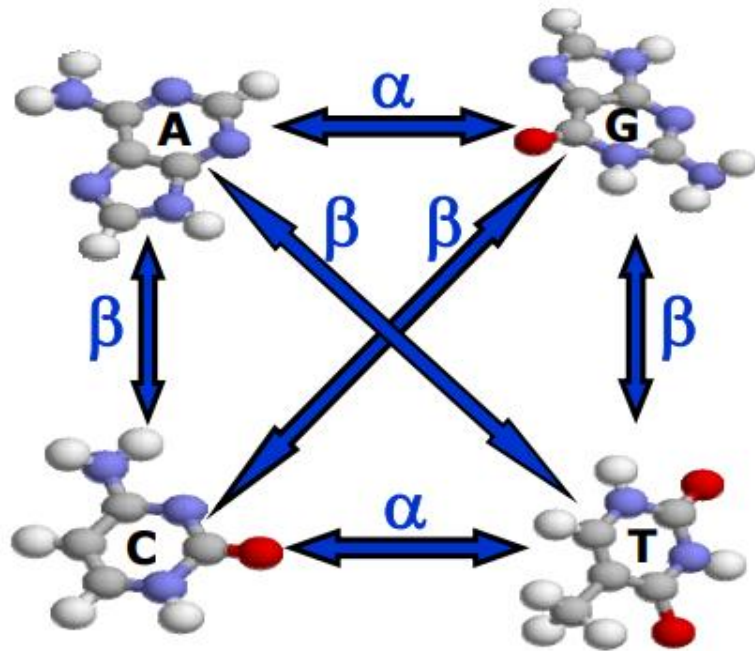
$$\text{Distância observada } \hat{p} = \frac{15}{300} = 0,05$$

$$\hat{d}_{JC} = -\frac{3}{4} \ln \left(1 - \frac{4}{3} \hat{p} \right) = -\frac{3}{4} \ln \left(1 - \frac{4}{3} \times 0,05 \right) = 0,05$$



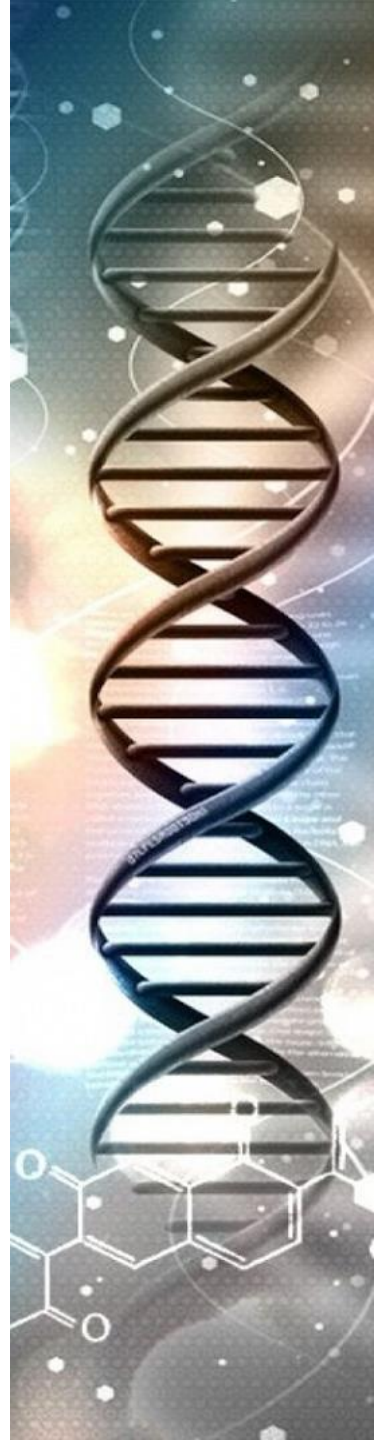
Calculando distancias evolutivas

Kimura 2 parámetros: la frecuencia de las bases son iguales.
La tasa de sustitución es diferente.

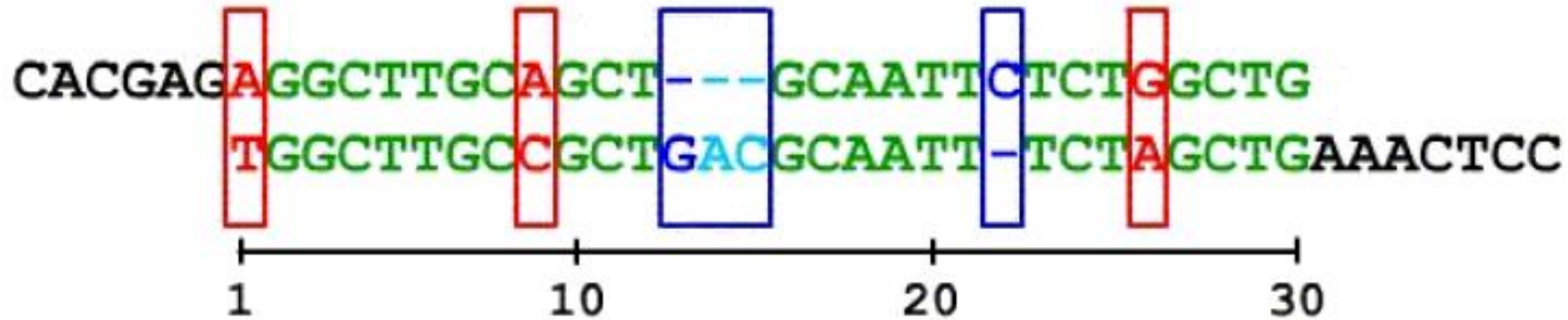


$$\hat{d}_{\text{Kimura}} = \frac{1}{2} \ln \left(\frac{1}{1 - 2\hat{P} - \hat{Q}} \right) + \frac{1}{4} \ln \left(\frac{1}{1 - 2\hat{Q}} \right)$$

P= Transiciones acumuladas
Q= Transversiones acumuladas



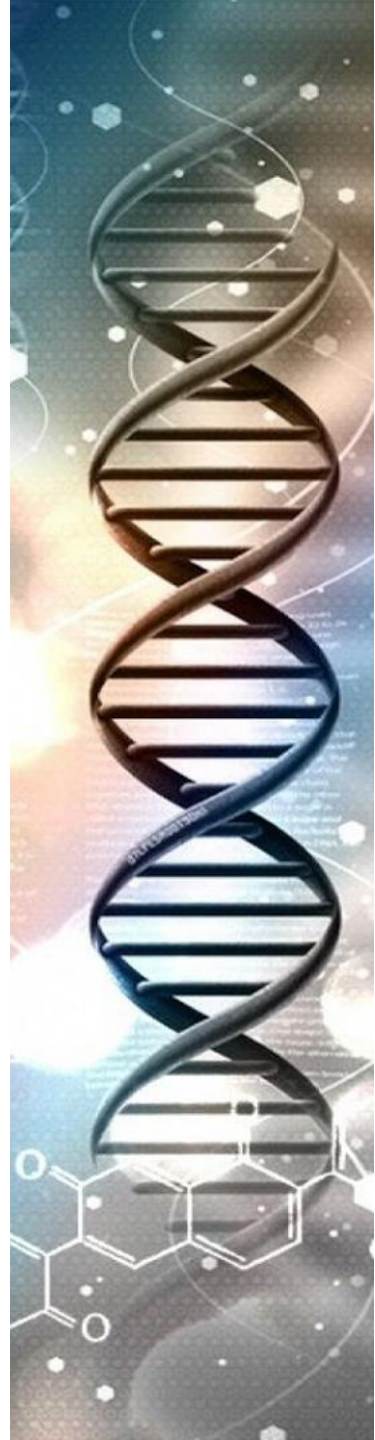
Alineamientos



$$\begin{array}{rcl} 3 \times (-2) & = & -6 \\ 2 \times (-4) & = & -8 \\ 2 \times (-3) & = & -6 \\ 23 \times (+1) & = & 23 \\ \hline & & 3 \end{array}$$

3

Gap opening	-4 pontos
Gap extension	-3 pontos
Mismatch	-2 pontos
Match	+1 ponto



Alineamientos

