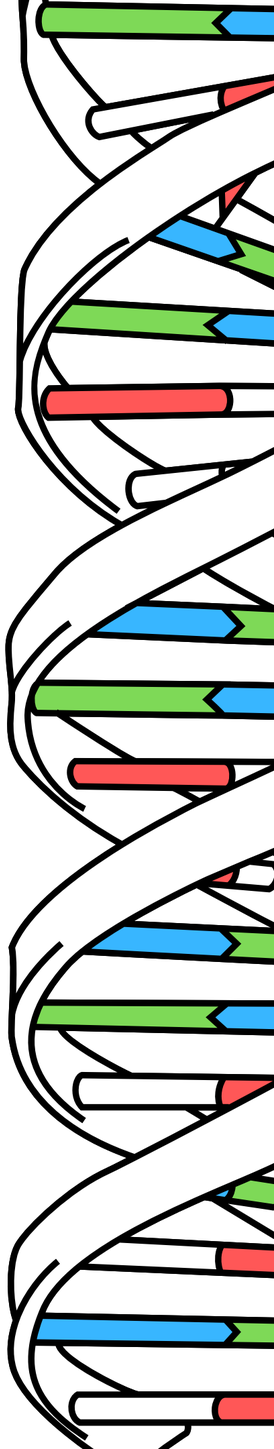
Introducción a la filogenómica



CENTRO DE INVESTIGACIONES BIOLÓGICAS DEL NOROESTE, S.C.

Introducción

Día	Temas
23 de septiembre	• **Introducción a la Filogenómica. Conceptos básicos** • Introducción al sistema operativo Linux
24 de septiembre	• Métodos de construcción de librerías genómicas • Control de calidad, ensamble y alineamiento de secuencias
25 de septiembre	• Método de Máxima Verosimilitud • Método de Inferencia Bayesiana
26 de septiembre	• Modelo coalescente multiespecies • Evaluación de árboles de especies
27 de septiembre	• Estimación de tiempos de divergencia • Evaluación de flujo genético

Introducción

Millones de años de evolución compartida con otras formas de vida

Introducción

Gran cadena del ser

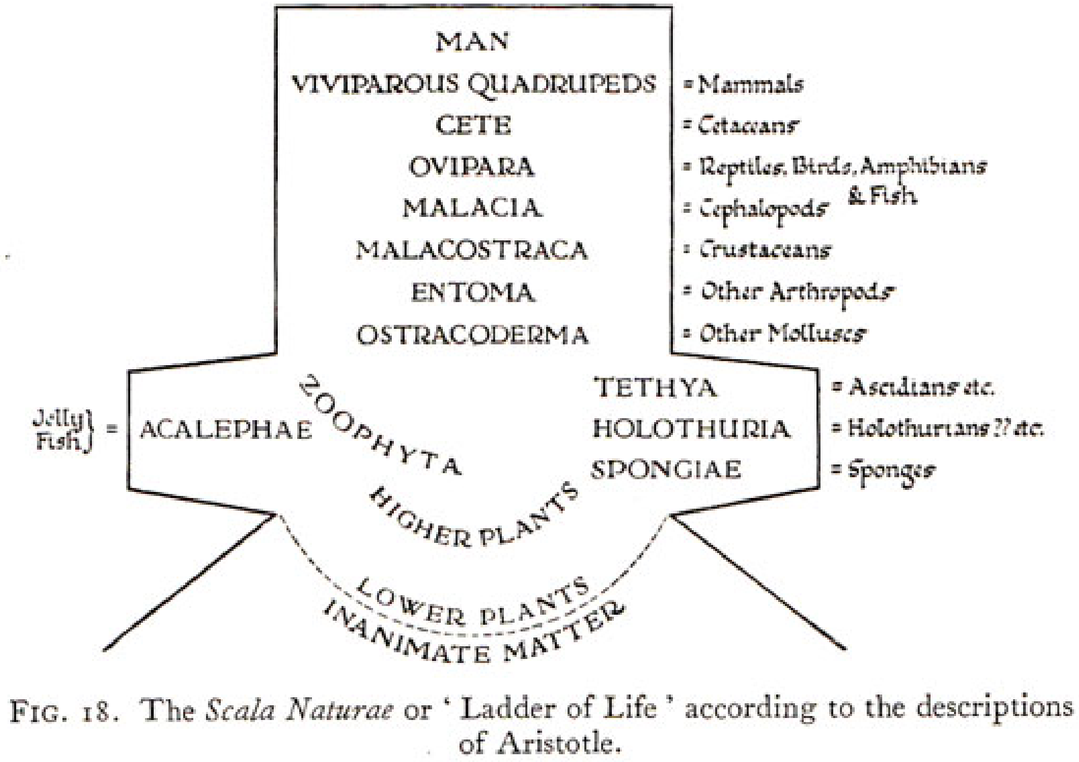


FIG. 18. The *Scala Naturae* or 'Ladder of Life' according to the descriptions of Aristotle.

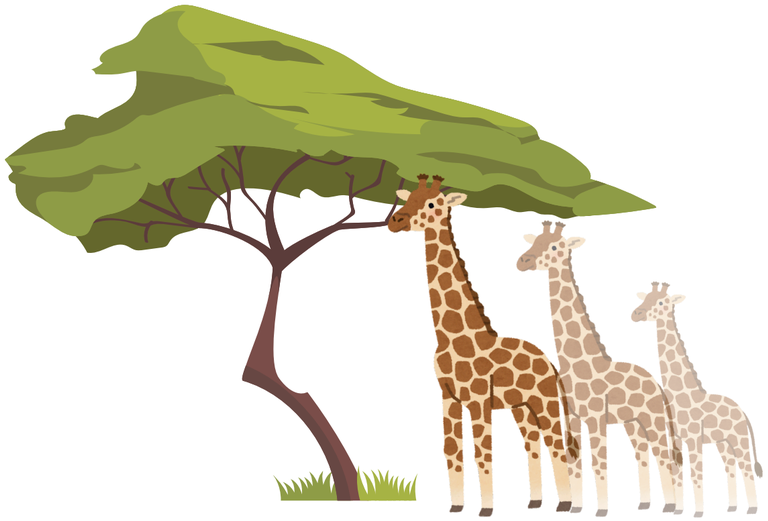
Jerarquía lineal de la vida

Mundo estático

Introducción

Jean-Baptiste Lamarck

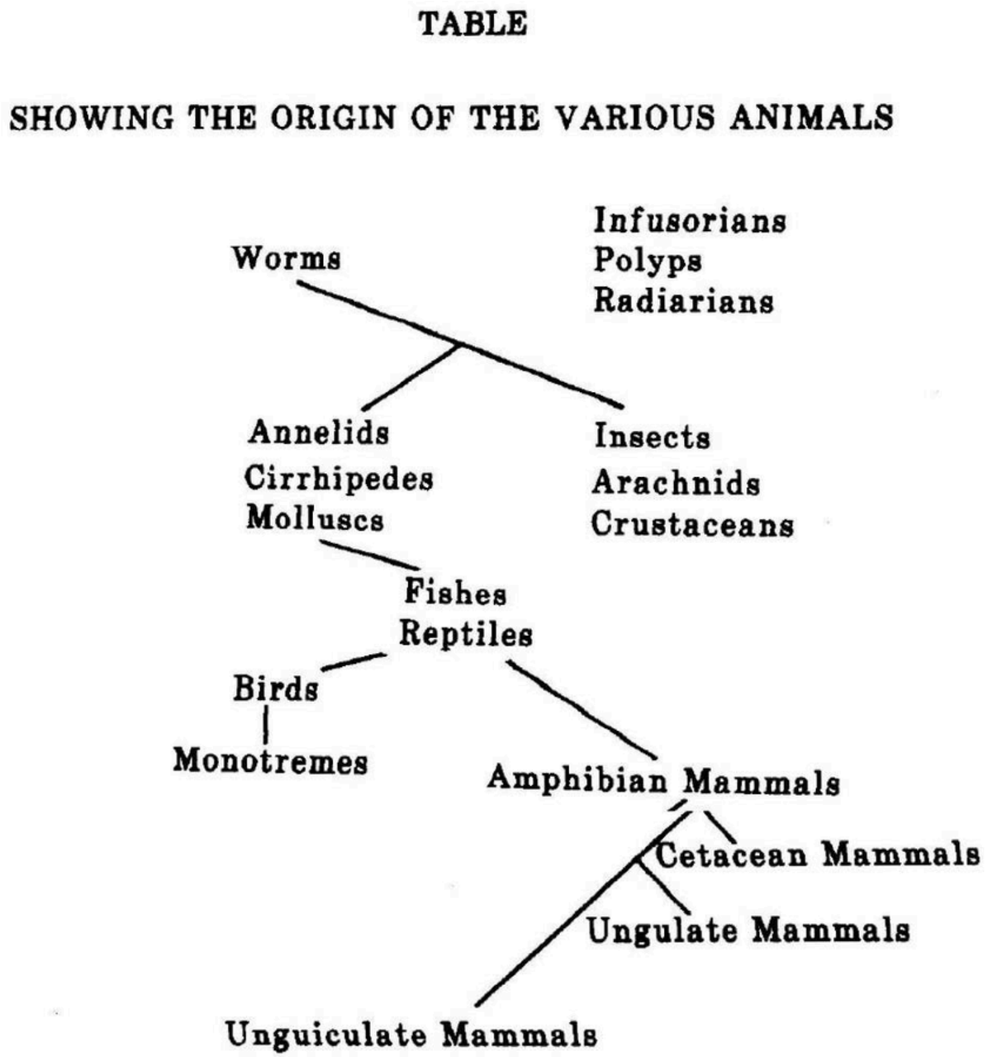
Philosophie Zoologique (1809)



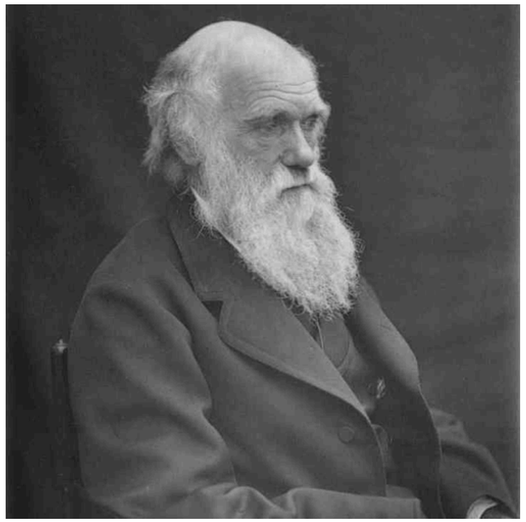
La vida evoluciona

Árbol de animales

Diferentes orígenes
No lineal
Progreso

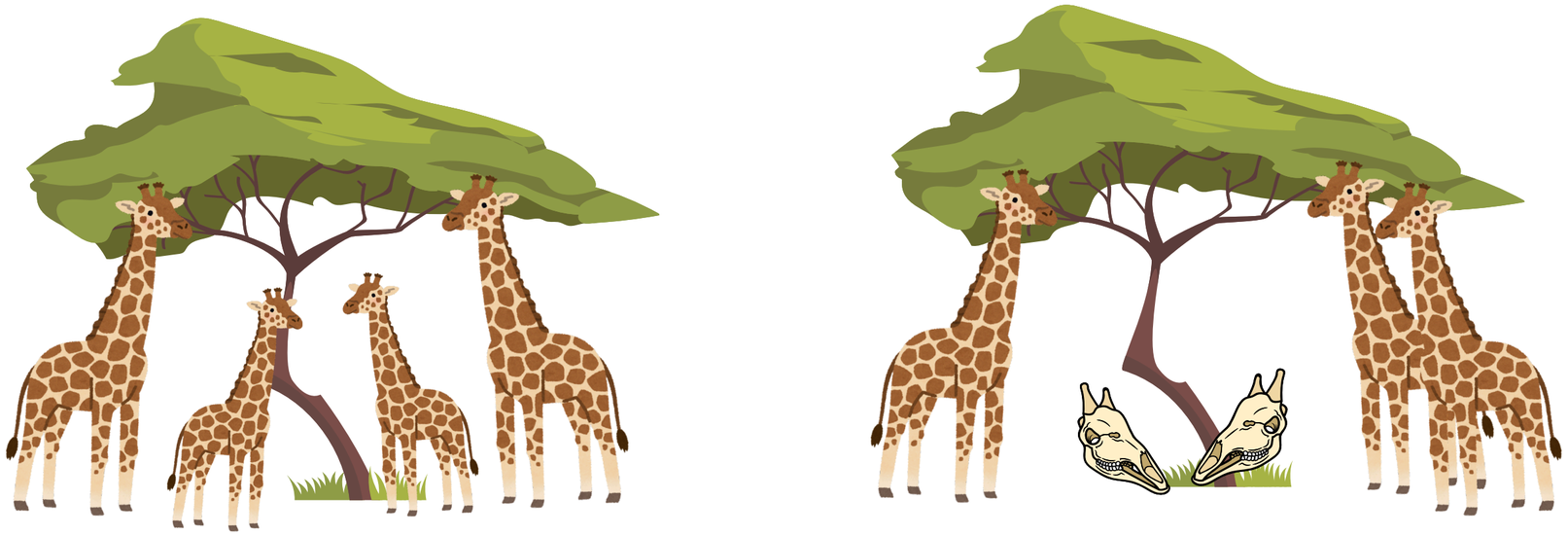


Introducción



Charles Darwin

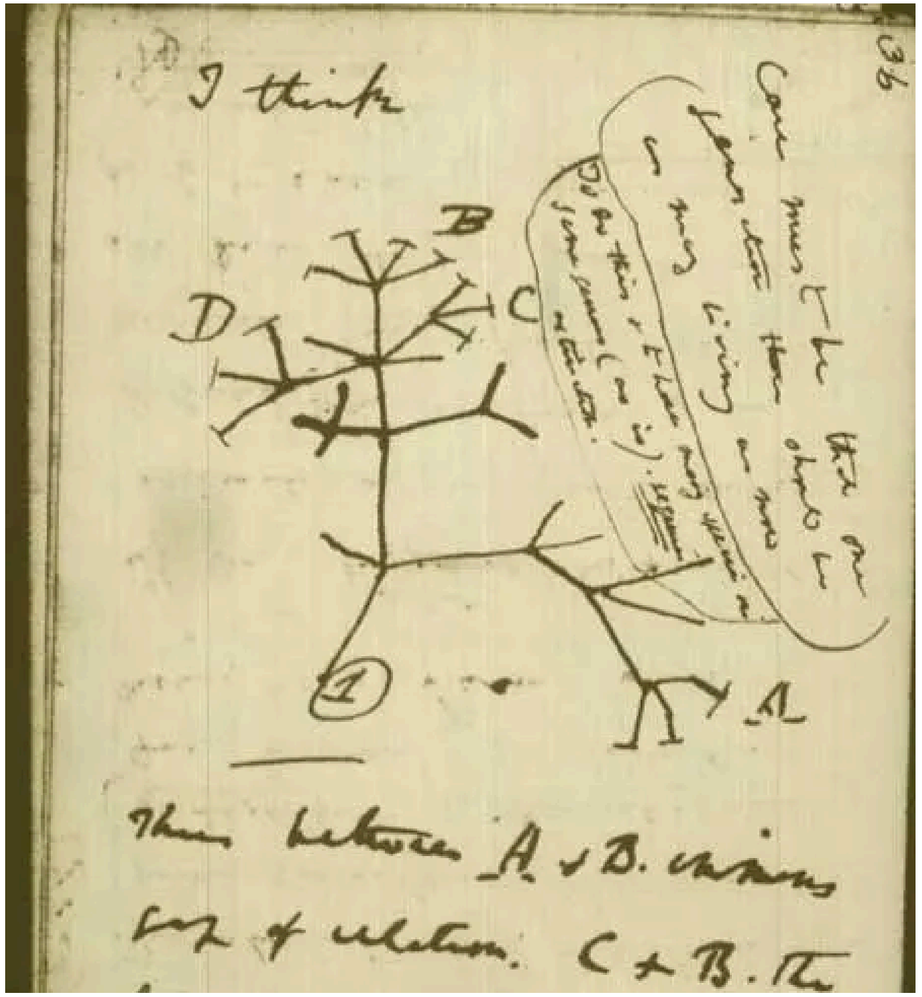
*Sobre **el origen de las especies** por medio de la selección natural* (1859).



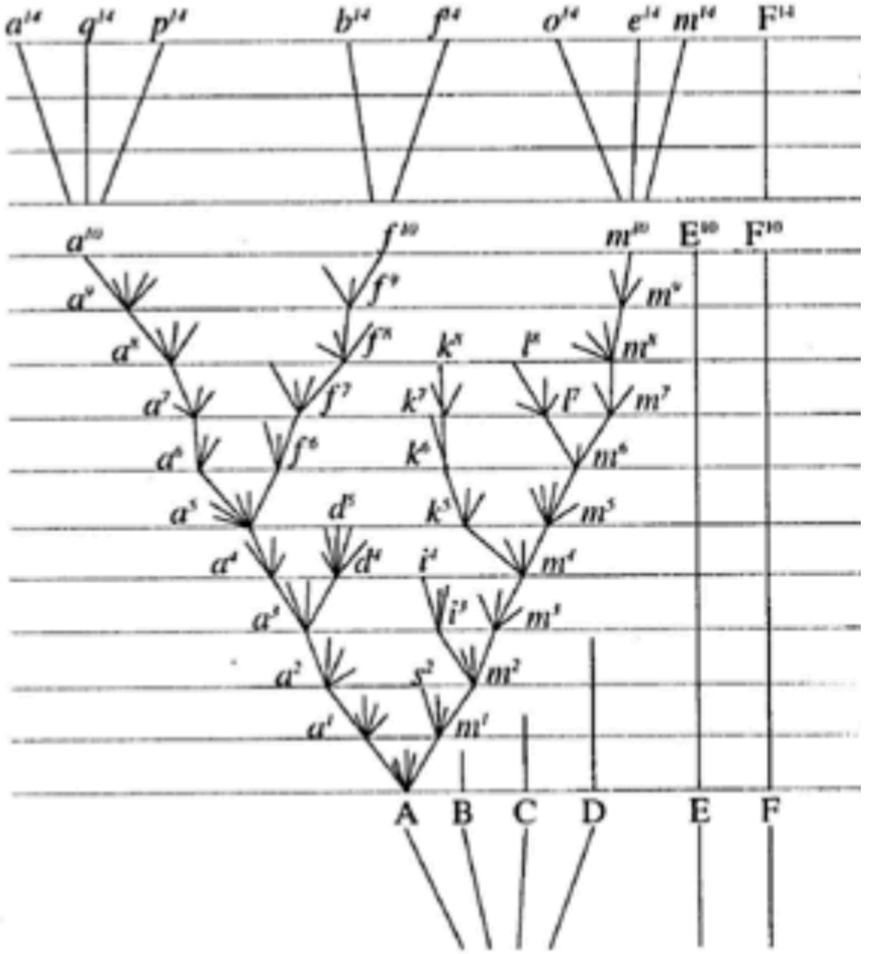
Evolución por selección natural

Concluyó que las diversas especies vivientes se remontan a los mismos antepasados comunes.

Introducción



Darwin, 1837



Darwin, 1859

...una generación sería tantas como las que viven ahora.

Introducción

Al considerar el origen de las especies se concibe perfectamente que un naturalista, reflexionando sobre las afinidades mutuas de los seres orgánicos, sobre sus relaciones embriológicas, su distribución geográfica, sucesión geológica y otros hechos semejantes, puede llegar a la conclusión de que las especies no han sido independientemente creadas, sino que han descendido, como las variedades, de otras especies. Sin embargo, esta conclusión, aunque estuviese bien fundada, no sería

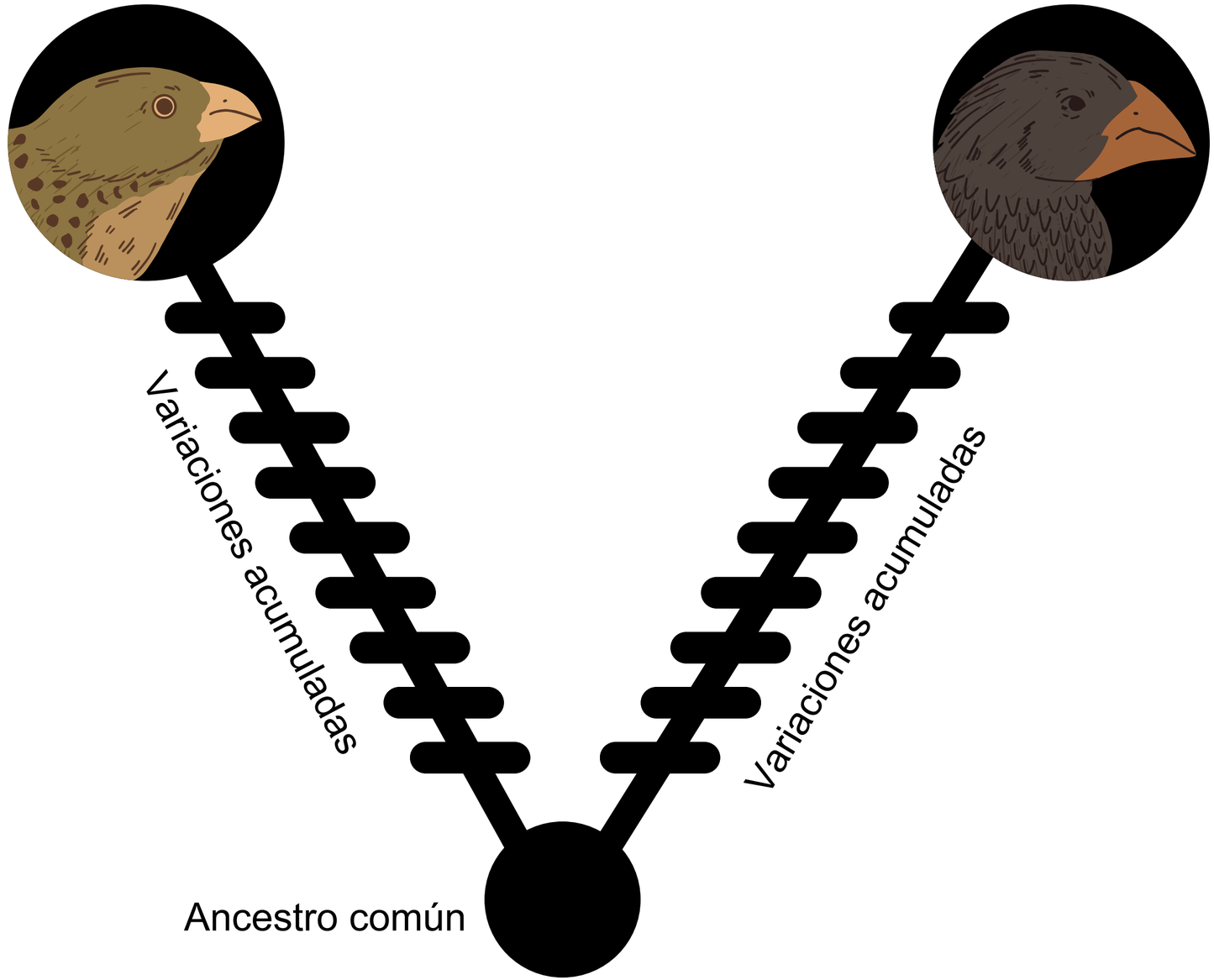
Introducción

De las muchas ramitas que florecieron cuando el árbol era un simple arbolillo, sólo dos o tres, convertidas ahora en ramas grandes, sobreviven todavía y llevan las otras ramas; de igual modo, de las especies que vivieron durante períodos geológicos muy antiguos, poquísimas han dejado descendientes vivos modificados. Desde el primer crecimiento del árbol, muchas ramas de todos tamaños se han secado

Capítulo IV

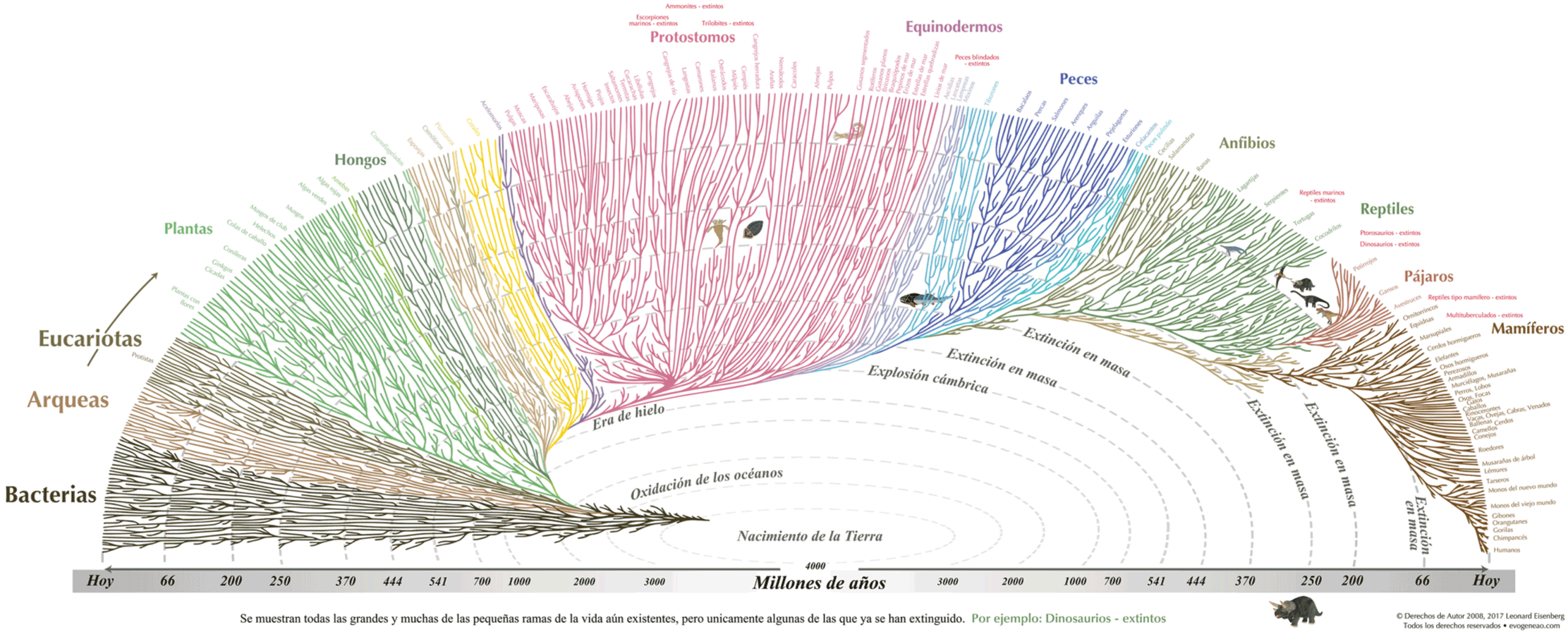
Introducción

- Similitudes estructurales
- Similitudes moleculares
- Distribuciones geográficas
- Fósiles
- La distribución de ciertas características



Introducción

Representación de las relaciones de ancestría-descendencia de los linajes



Se muestran todas las grandes y muchas de las pequeñas ramas de la vida aún existentes, pero unicamente algunas de las que ya se han extinguido. Por ejemplo: Dinosaurios - extintos

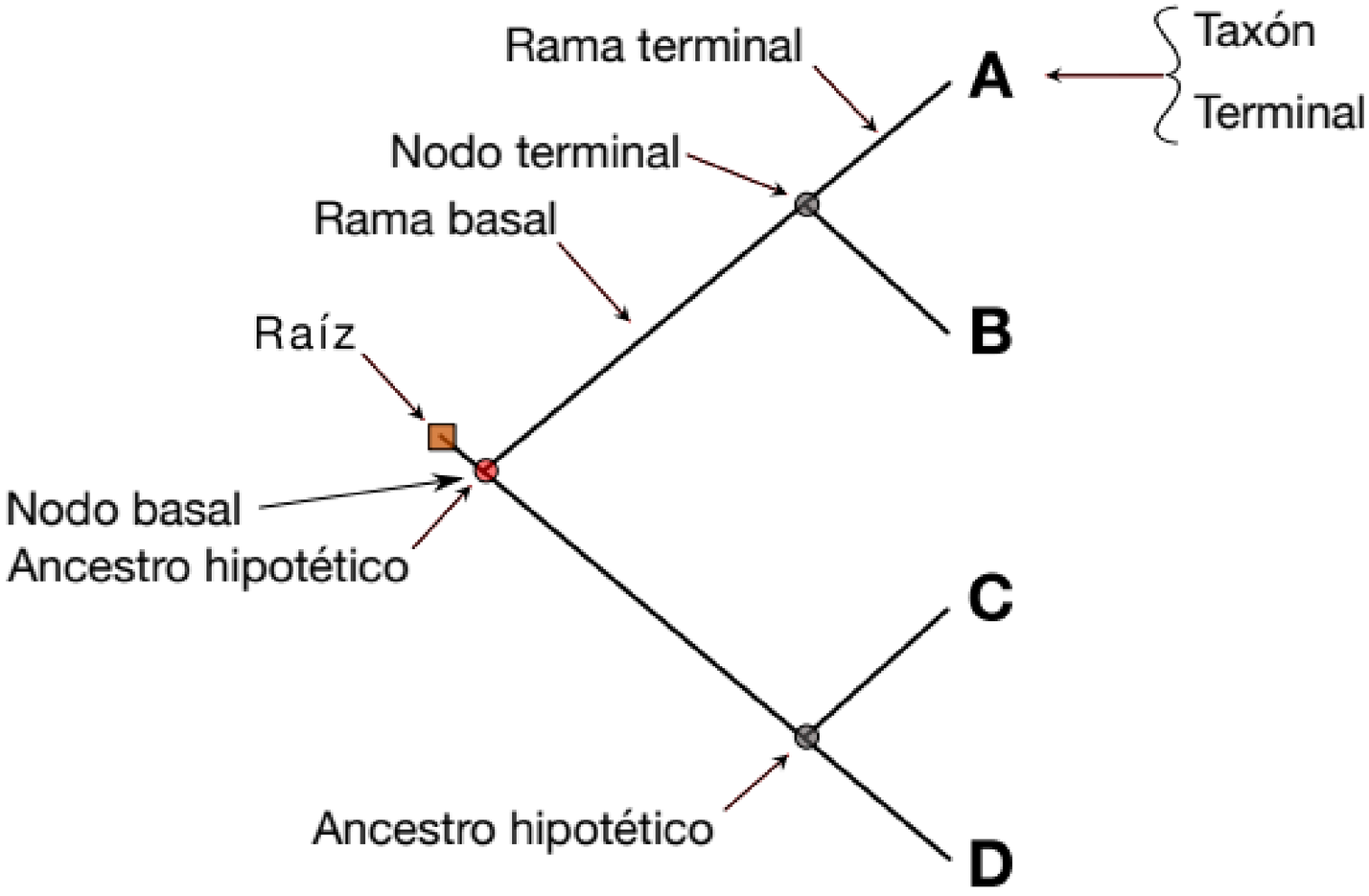
Introducción

Arbol filogenético
Filogenia
Árbol evolutivo

Diagrama ramificado que representa las relaciones entre organismos u otras entidades.

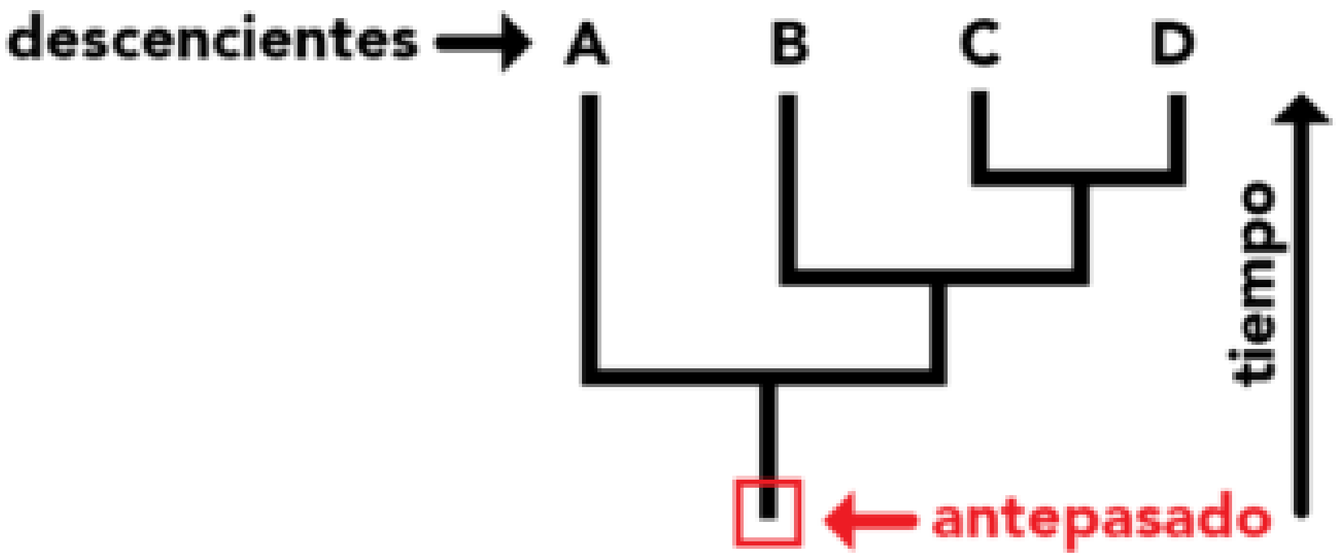
Es una hipótesis

Contexto

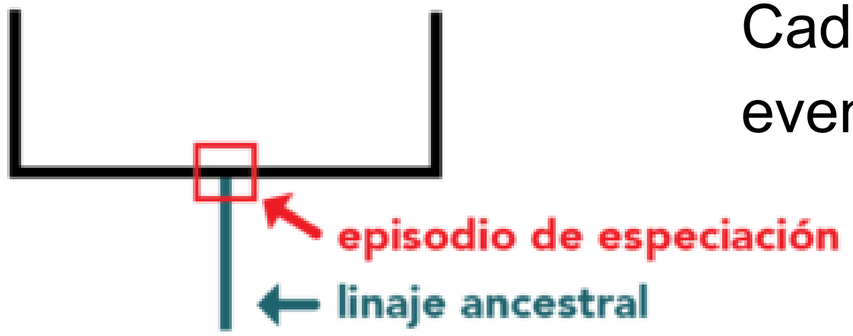


Leopardi-Verde & Escobedo-Sarti, 2021

Introducción



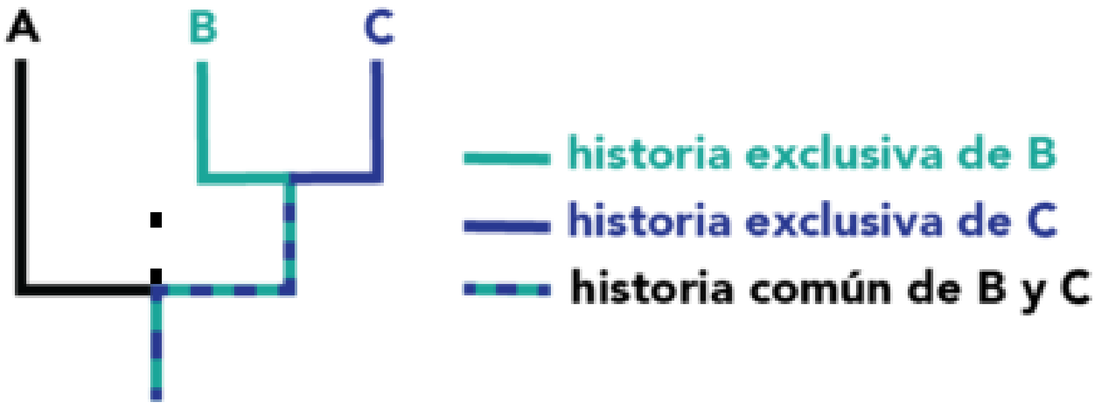
Descendientes que comparten un ancestro común.



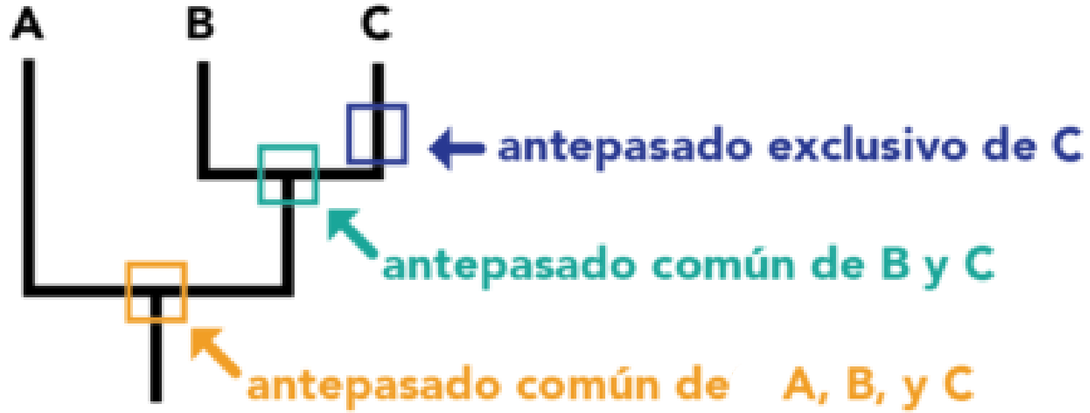
Cada ramificación indica un evento de divergencia.

evolution.berkeley.edu

Introducción



B y C comparten historia evolutiva común antes de divergir

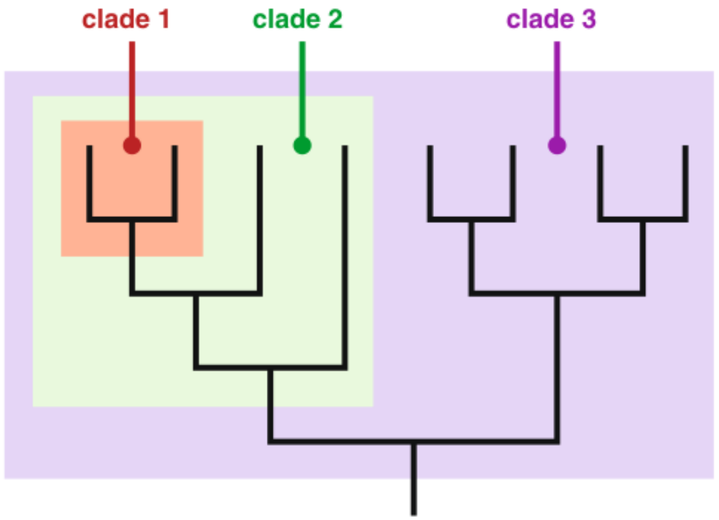


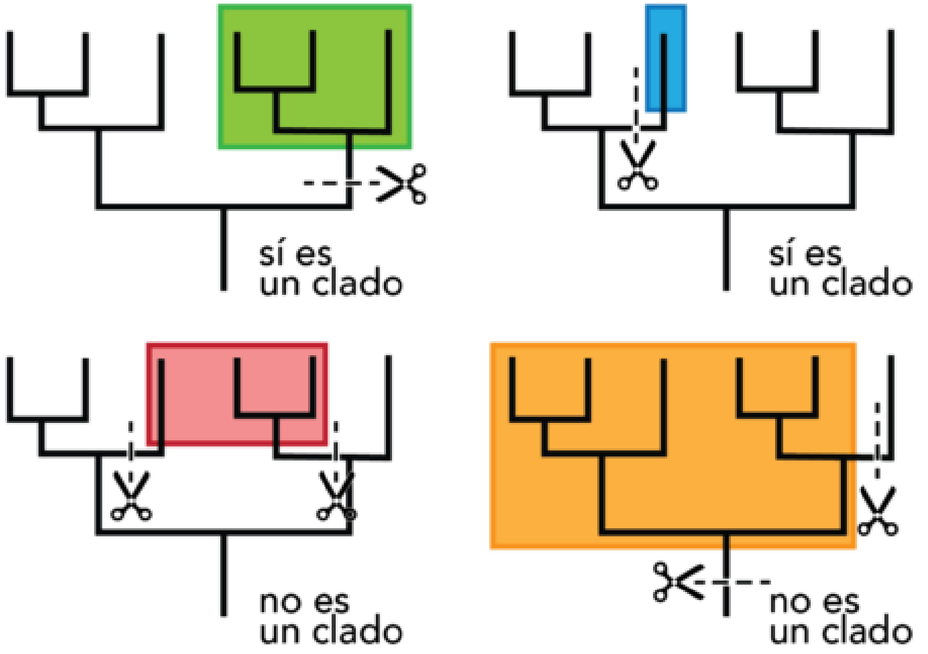
Divergencia continua
cada linaje tiene una historia propia y una historia compartida.

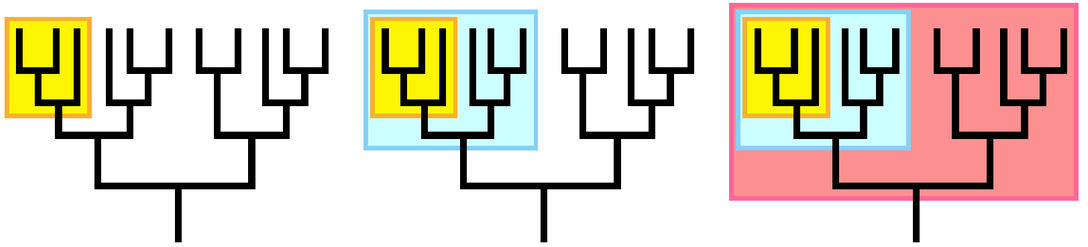
evolution.berkeley.edu

Introducción

Clado. Grupo de organismos que incluye un ancestro común y a todos sus descendientes.
Se puede identificar al realizar un único corte en una rama, que separa un grupo completo sin dejar fuera a ningún descendiente de ese ancestro.

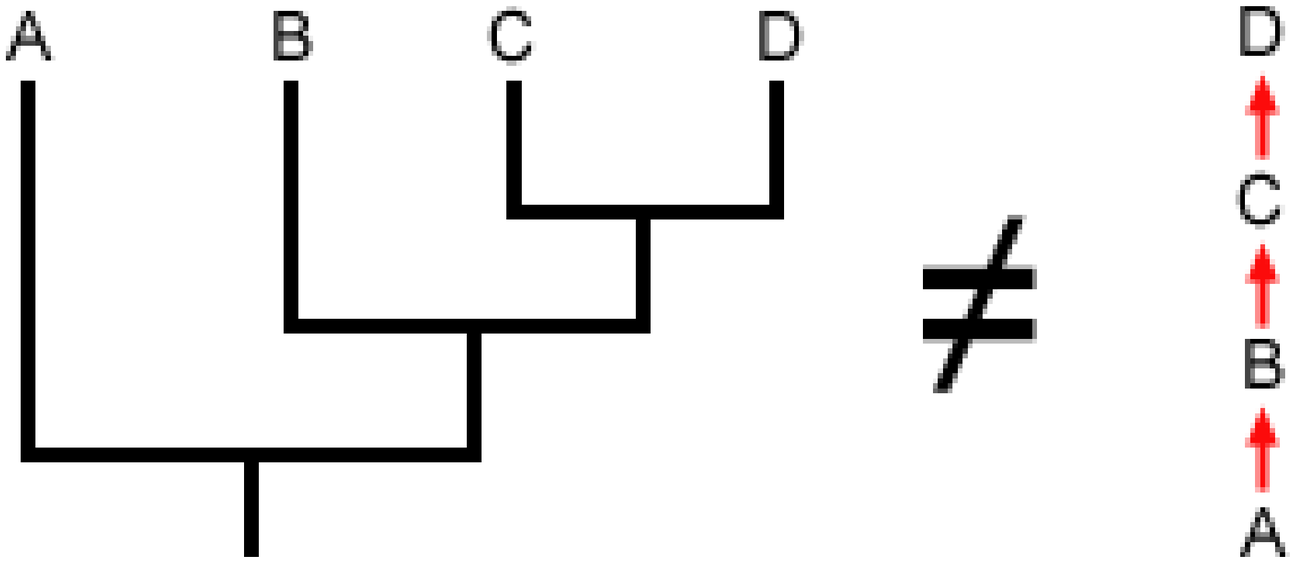




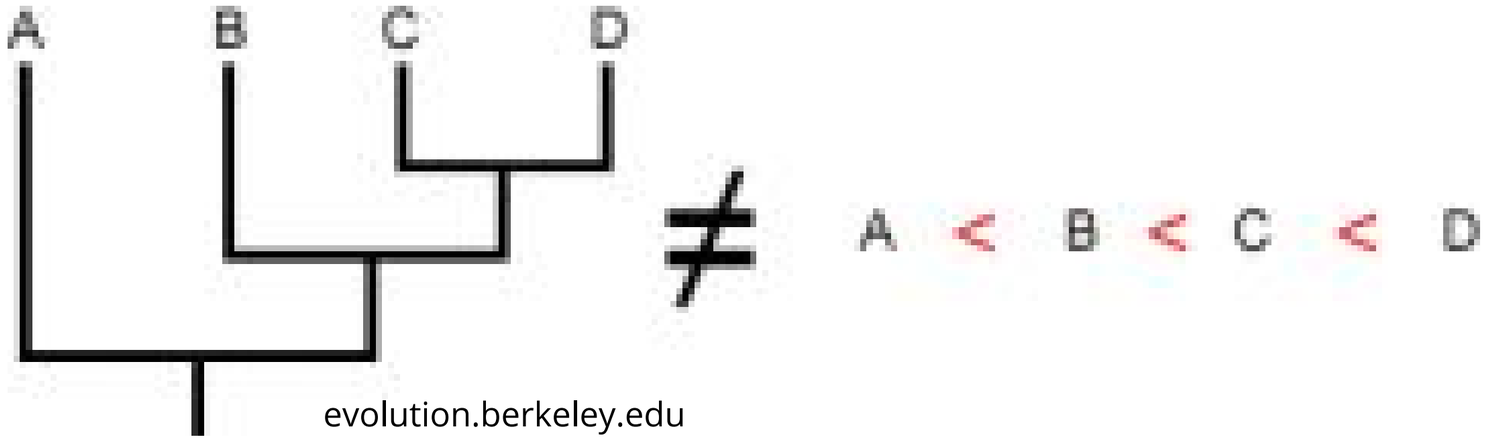


Thanukos, 2009

Introducción



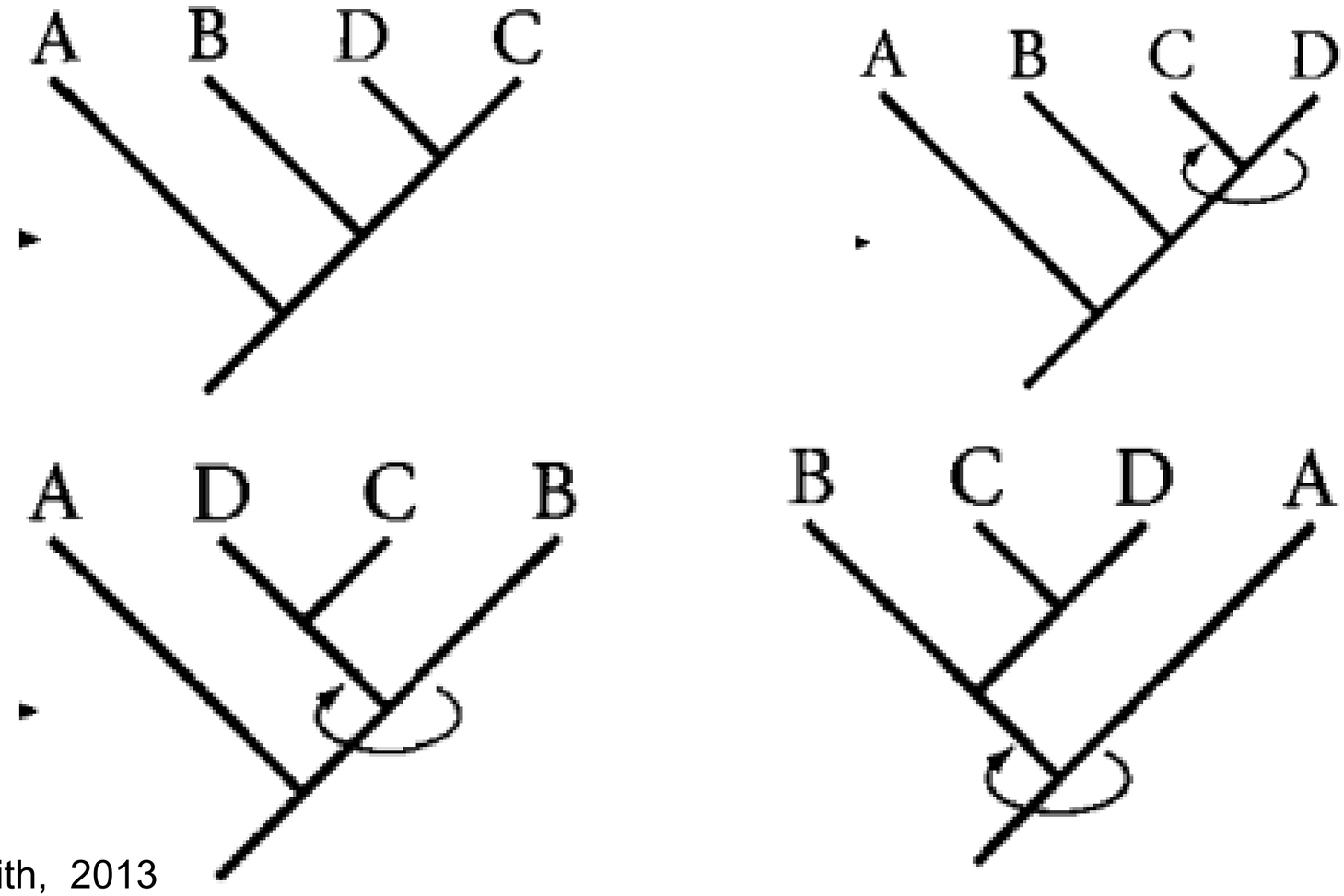
Orden no representa una secuencia lineal de evolución



Orden de las ramas no indica superioridad o progreso

evolution.berkeley.edu

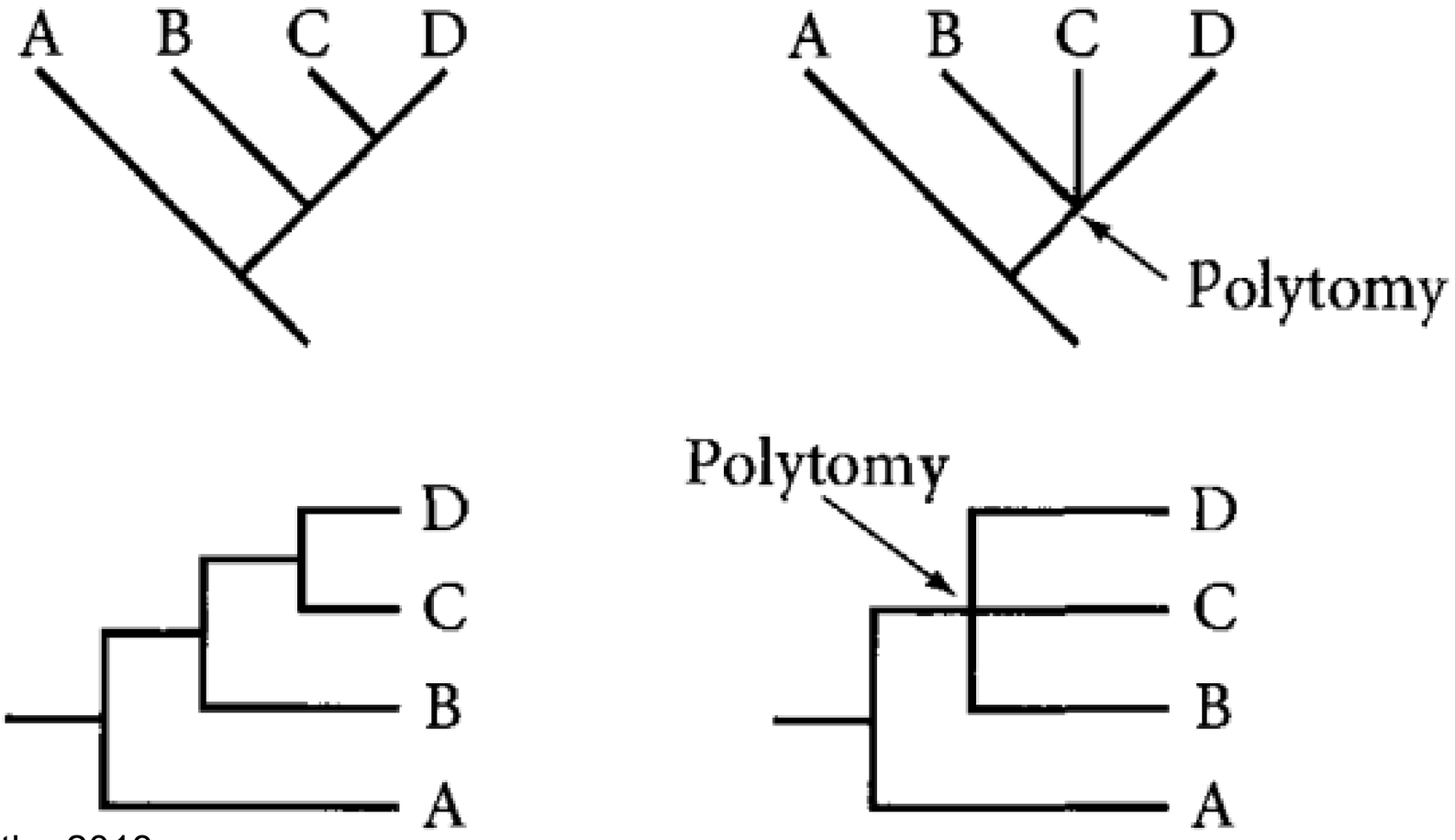
Introducción



Baum & Smith, 2013

Las ramas pueden rotarse alrededor de los nodos sin alterar las relaciones evolutivas. El orden visual no importa; sino cómo los linajes se conectan.

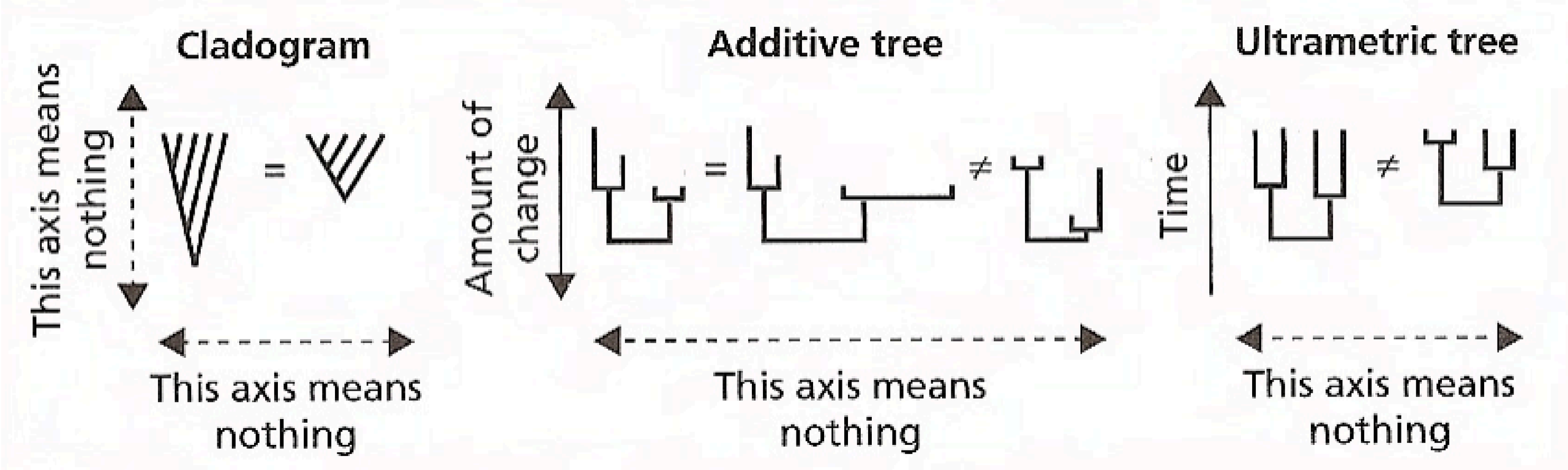
Introducción



Baum & Smith, 2013

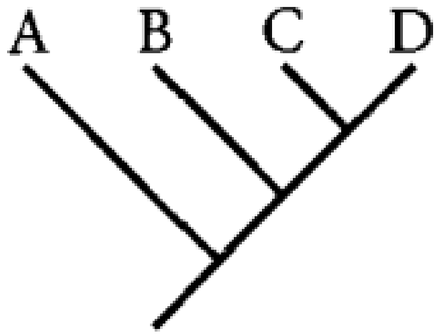
Politomías reflejan eventos evolutivos complejos o falta de información suficiente para resolver las relaciones exactas entre los linajes.

Introducción

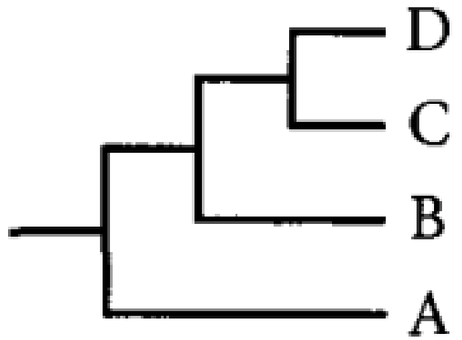


Interpretar los ejes y las longitudes de las ramas es clave

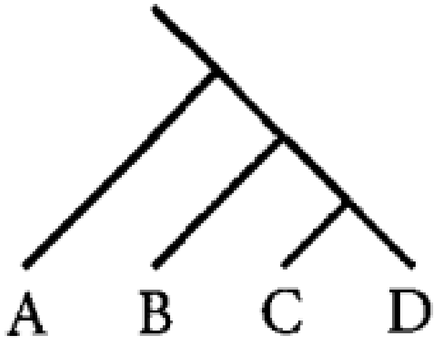
Introducción



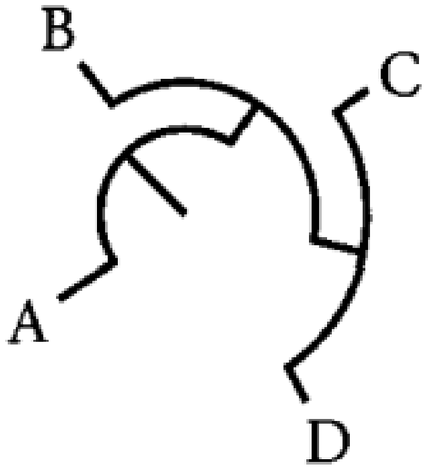
Diagonal-up



Rectangular-right



Diagonal-down

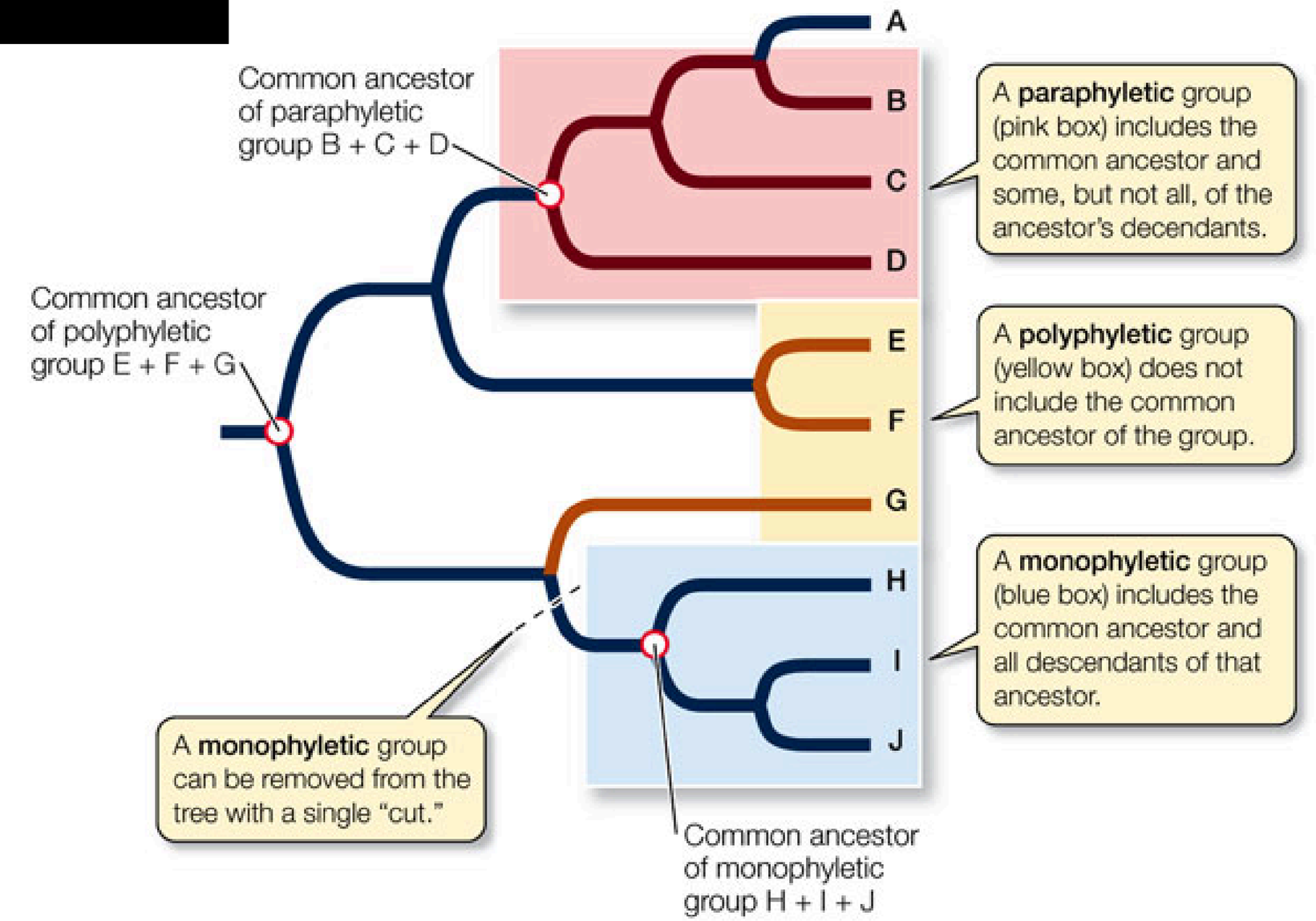


Circle

Baum & Smith, 2013

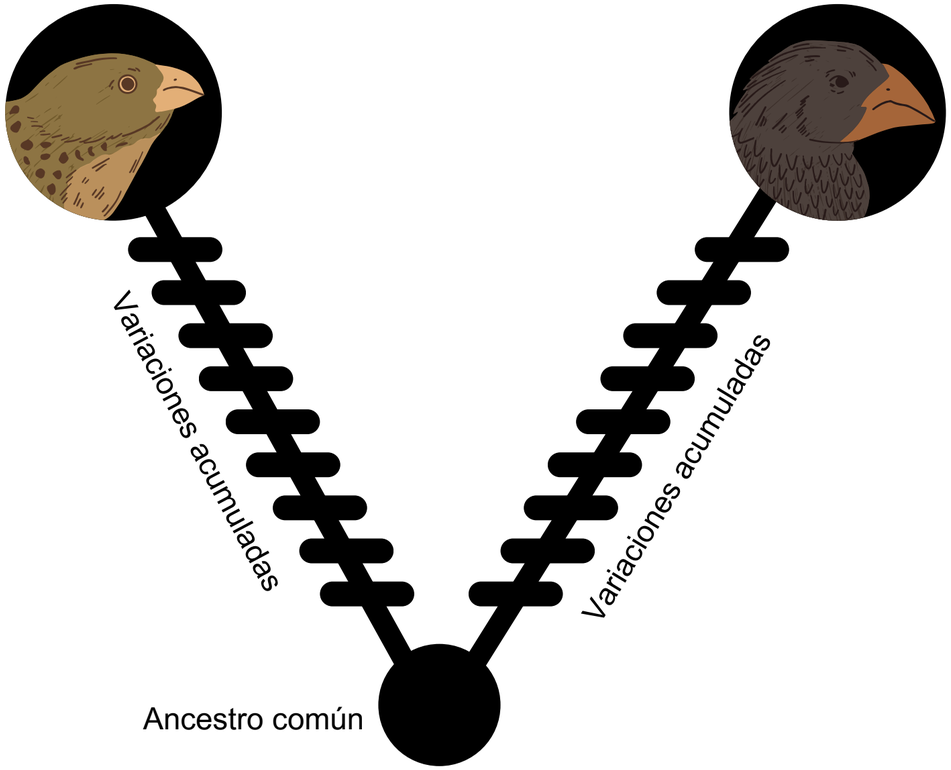
Representación visual cambia, relaciones evolutivas permanecen iguales.

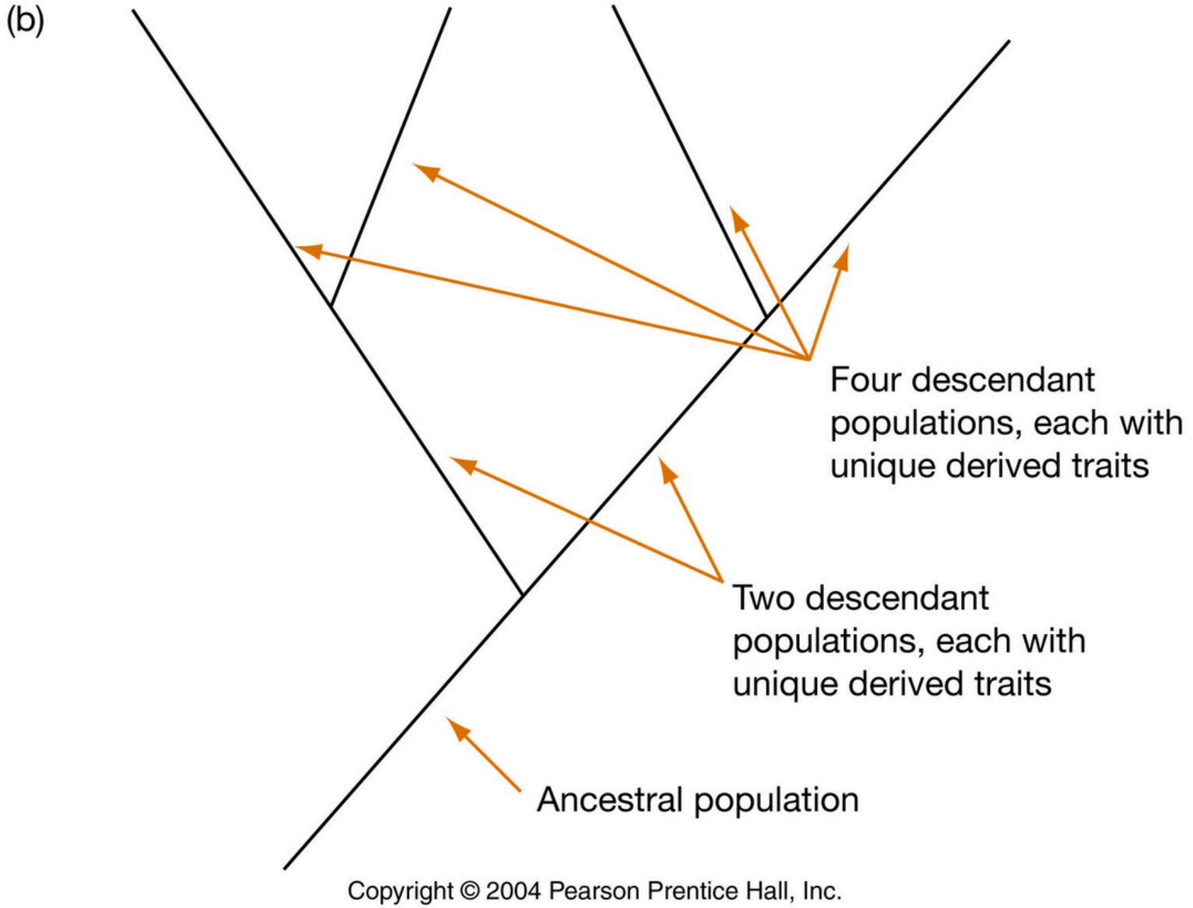
Introducción



***LIFE 8e*, Figure 25.12**

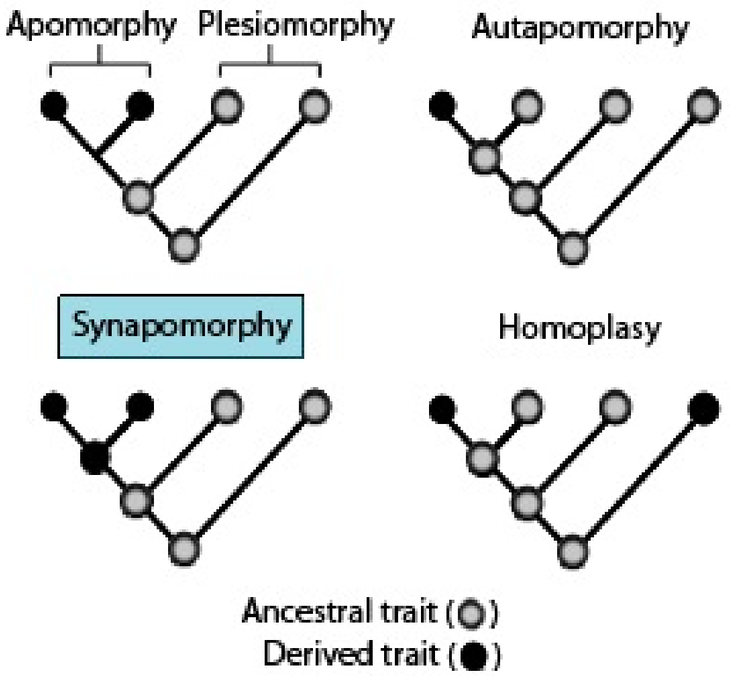
Introducción

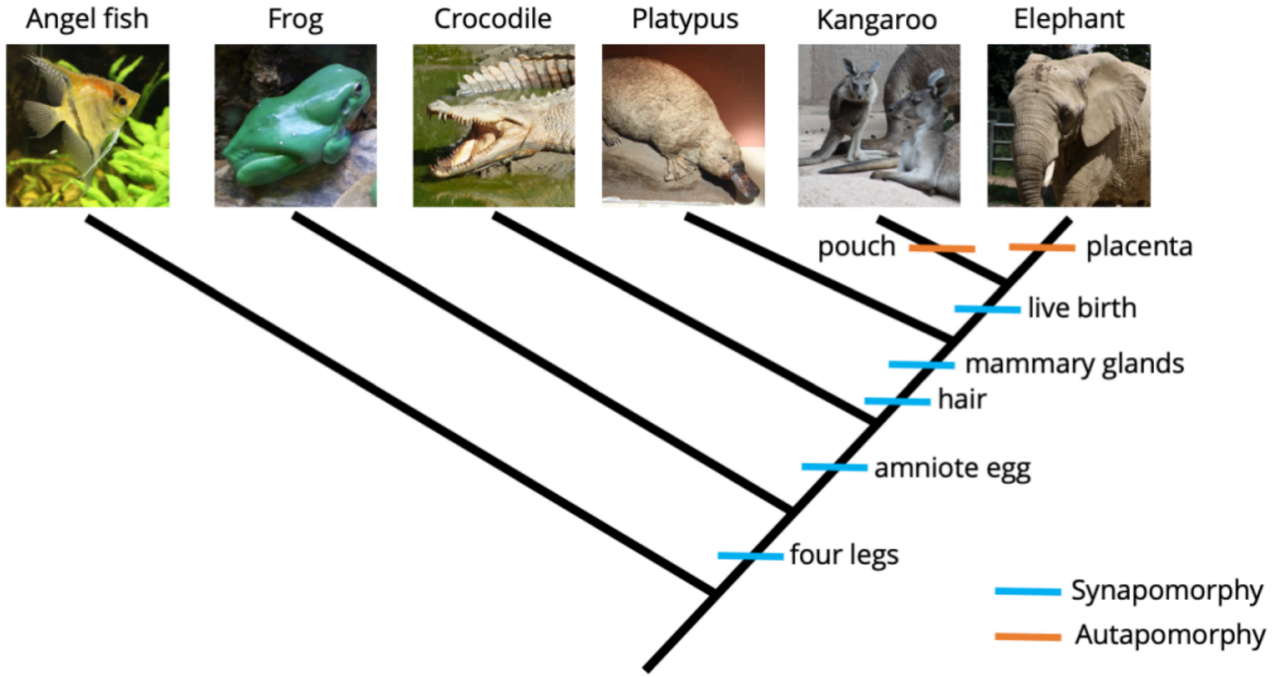




Los linajes resultantes del proceso de diversificación acumulan rasgos derivados únicos.

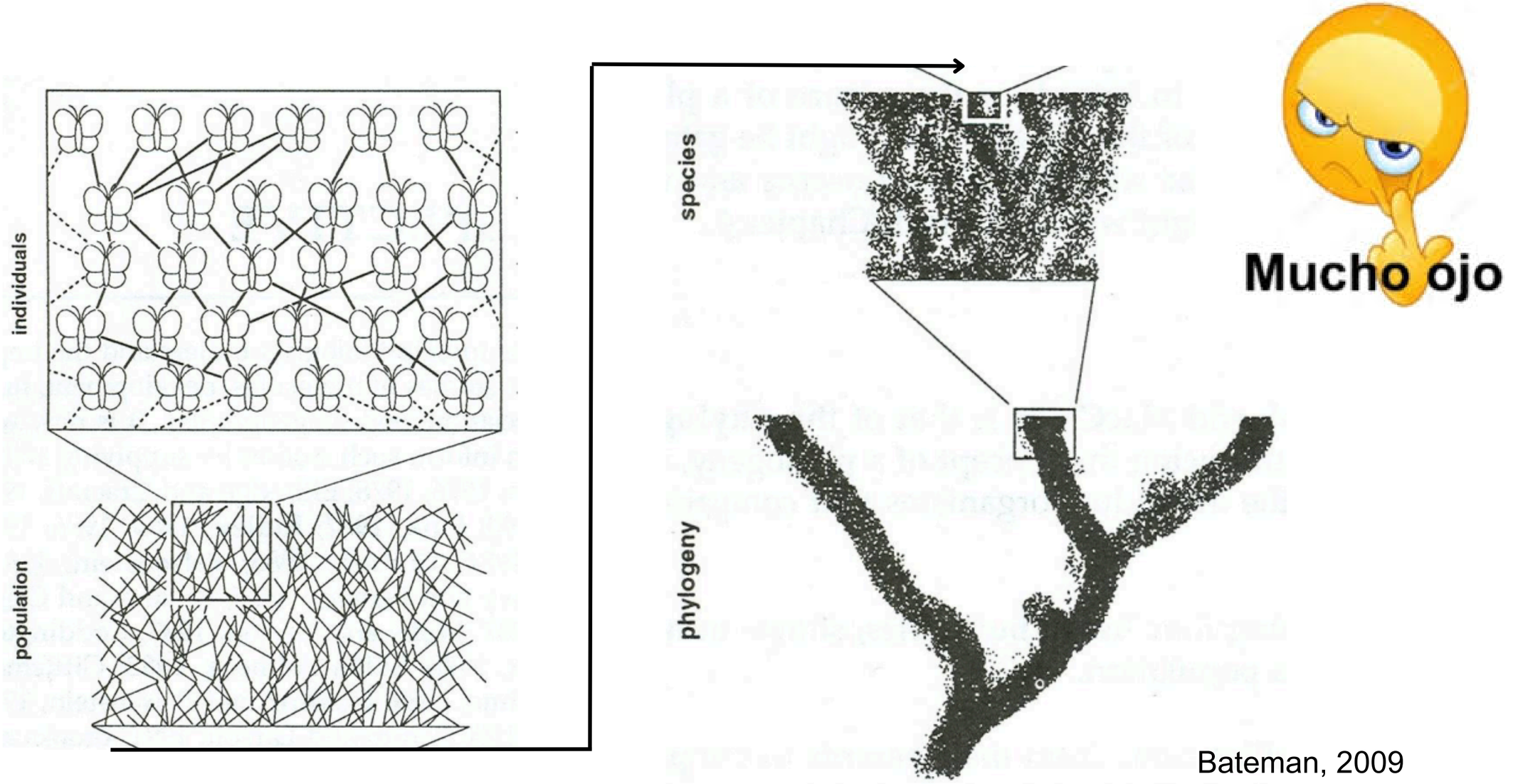
Introducción





Apomorfía: Caracter derivado (nuevo) en un linaje.
Plesiomorfía: Caracter ancestral presente en el ancestro común.
Sinapomorfía: Apomorfía compartida por varios linajes, indica un ancestro común.
Autapomorfía: Caracter derivado único de un solo linaje.
Homoplasia: Aparición independiente de un caracter en diferentes linajes (Convergencia).

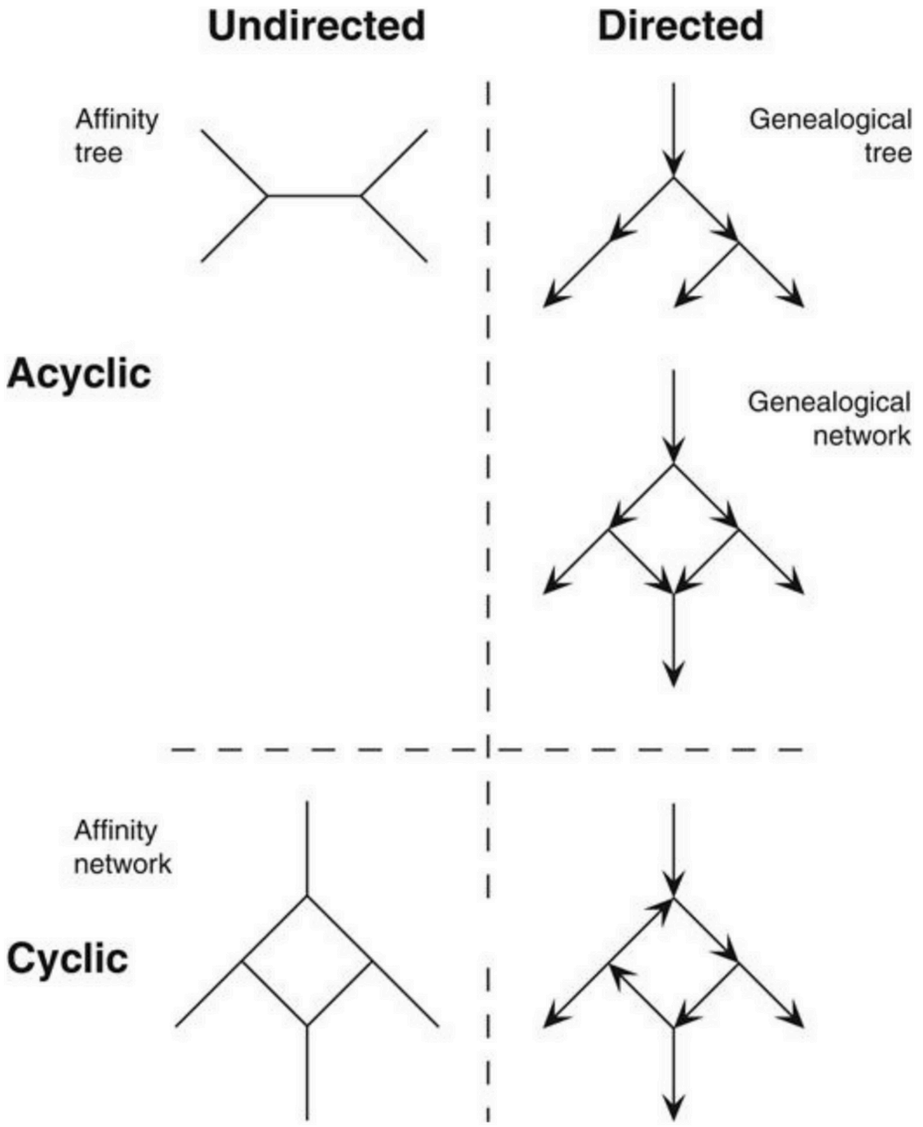
Introducción



Bateman, 2009

Introducción

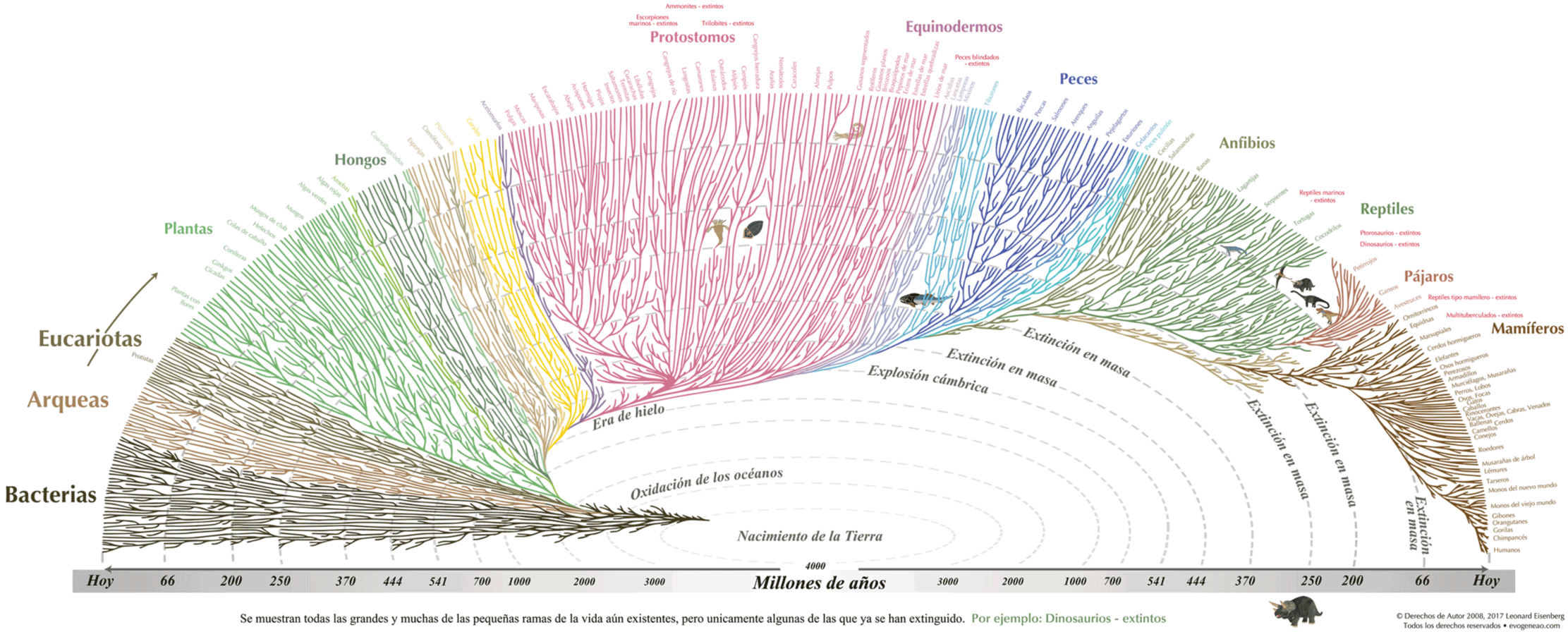
Procesos complejos, como el flujo genético o la hibridación, que no pueden ser representados solo con árboles bifurcados.



Morrison, 2013

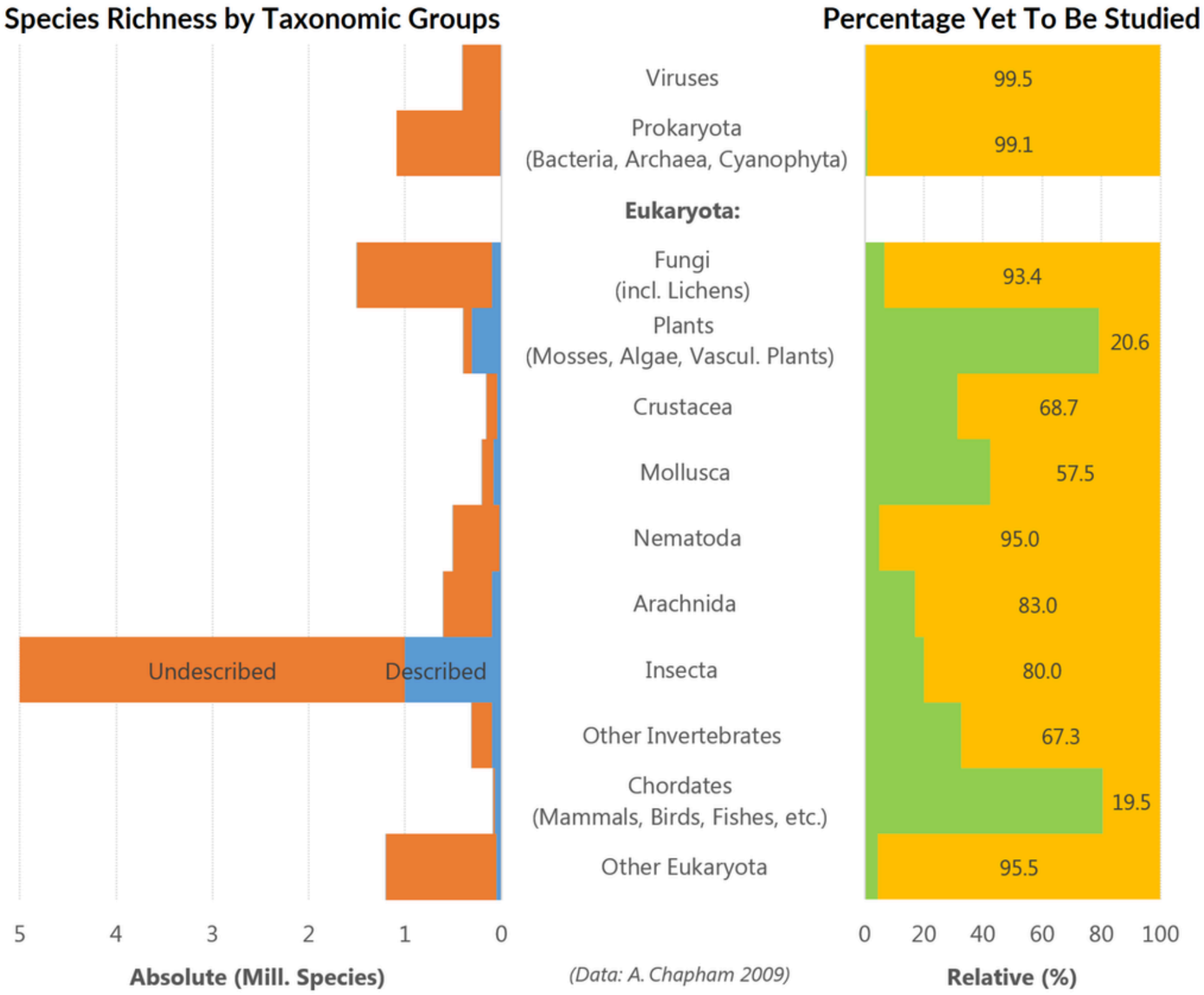
Introducción

Representación de las relaciones de ancestría-descendencia de los linajes



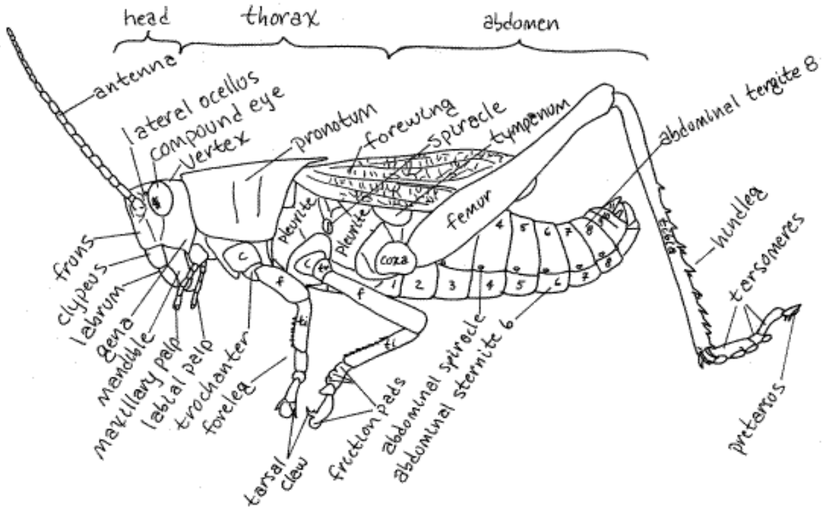
Se muestran todas las grandes y muchas de las pequeñas ramas de la vida aún existentes, pero unicamente algunas de las que ya se han extinguido. Por ejemplo: Dinosaurios - extintos

Introducción

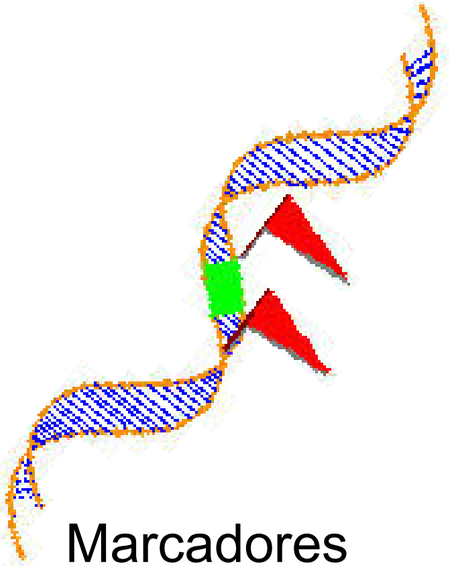


Introducción

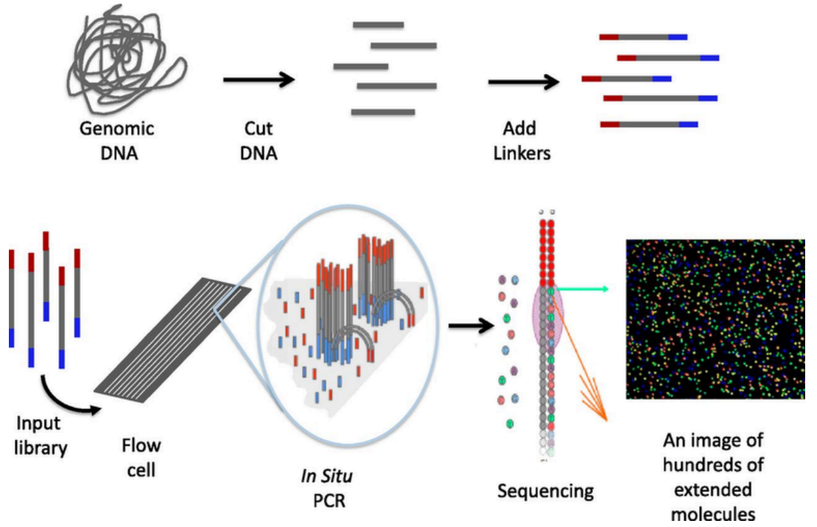
Carácter - Atributo que puede variar



Morfología



Marcadores puntuales



Datos genómicos

Líneas de evidencia

Introducción

Marcadores puntuales

- Primers universales
- Tasas de evolución conocidas
- Métodos estandarizados

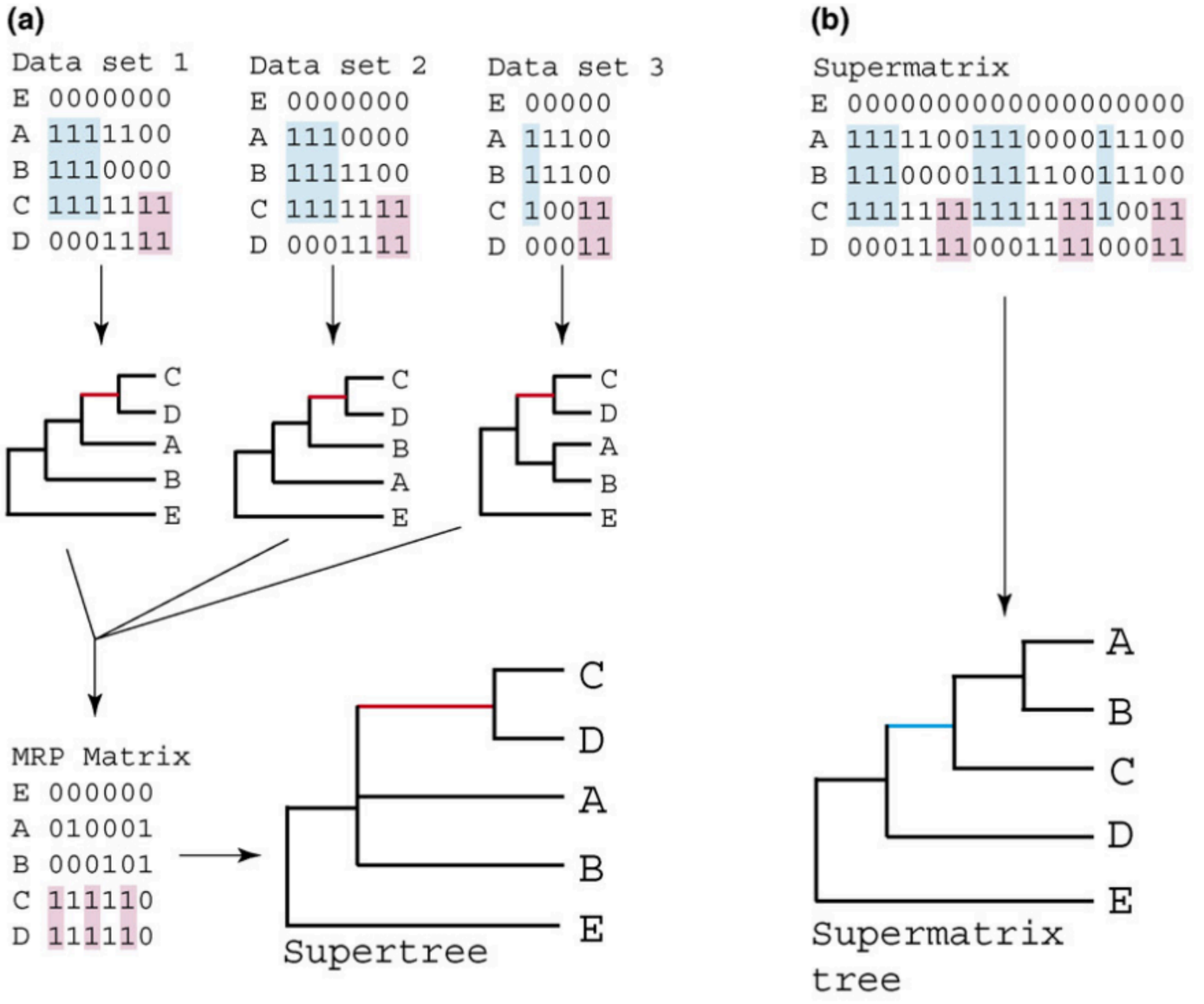
- Resolución
- Soporte
- Historia

Gene	Description	Reference
EF-1α	Elongation factor-1α, Role in protein synthesis.	[52]
rpoA gene	Encoding the alpha subunit of RNA polymerase	[53]
atpB	Encode the beta subunit of ATP synthase	[54]
dnaA	involved in DNA synthesis initiation	[55]
ftsZ	Role in cell division	[56]
gapA	Codes for glyceraldehyde phosphate dehydrogenase	[57]
groEL	Encodes bacterial heat shock protein.	[58]
gltA	Encoding citrate synthase	[59]
ITS	Piece of non-functional RNA situated between structural	[60]
	ribosomal RNAs precursor transcript.	
lux Gene	encode proteins involved in luminescence	[61]
PEPCK	Codes for phosphoenolpyruvate carboxykinase	[62]
pyrH genes	Codes for uridine monophosphate (UMP) kinases	[63]
recA	Role in recombination	[64]
U2 snRNA	Component of the spliceosome	[65]
Wsp gene	Encodes a major cell surface coat protein	[66]
Nuclear H3	Codes for protein which is associated with DNA	[67]
trnH-psbA	Non-coding intergenic spacer region located in plastid genome	[68]
rpoB, rpoC1	Coding region located in plastid genome	[69]

Table 1: List of some other molecular markers used in phylogeny research.

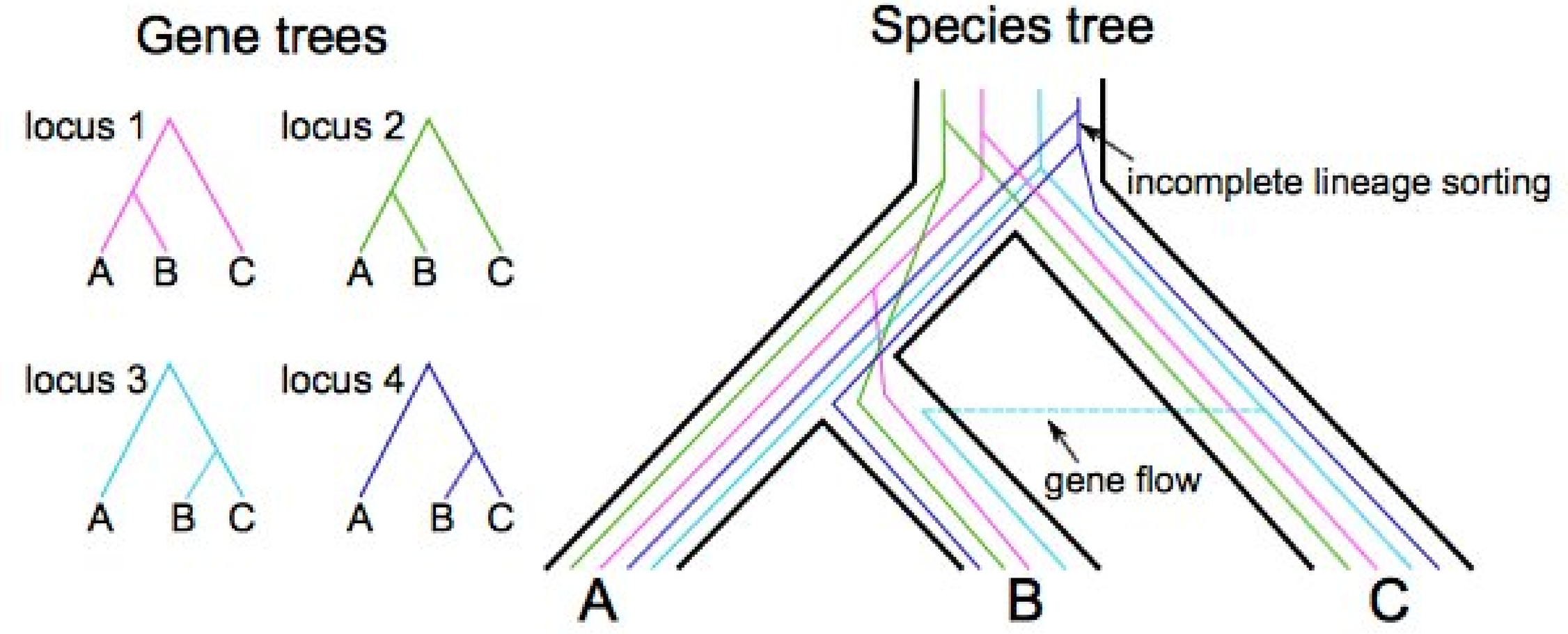
Patwardhan et al. 2014

Introducción



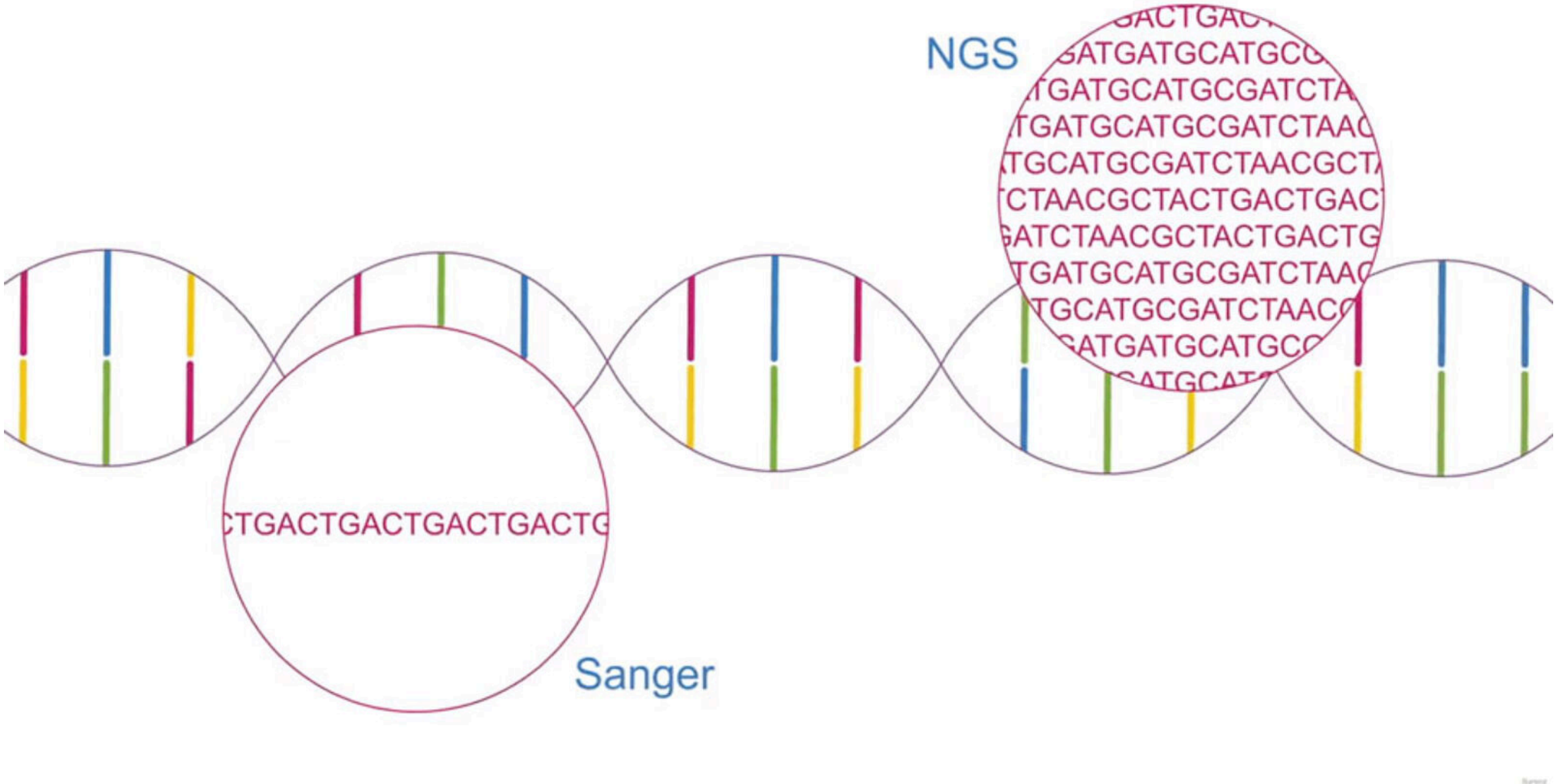
de Queiroz & Gatesy, 2006

Introducción

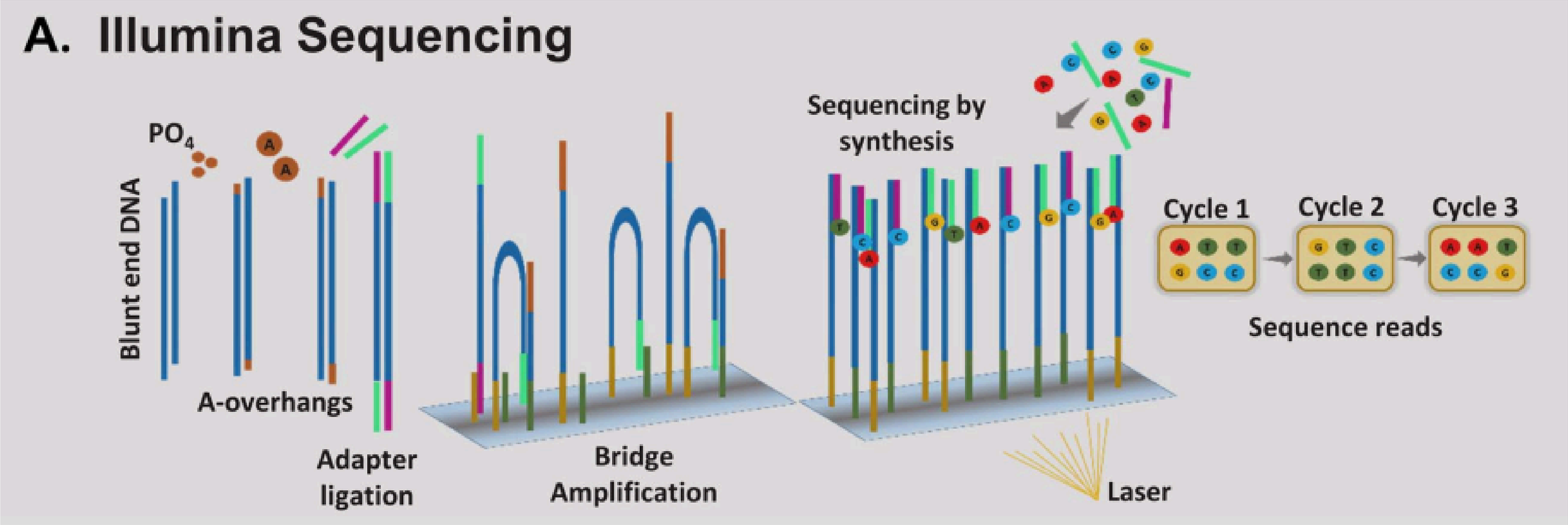


Marin et al., 2020

Introducción



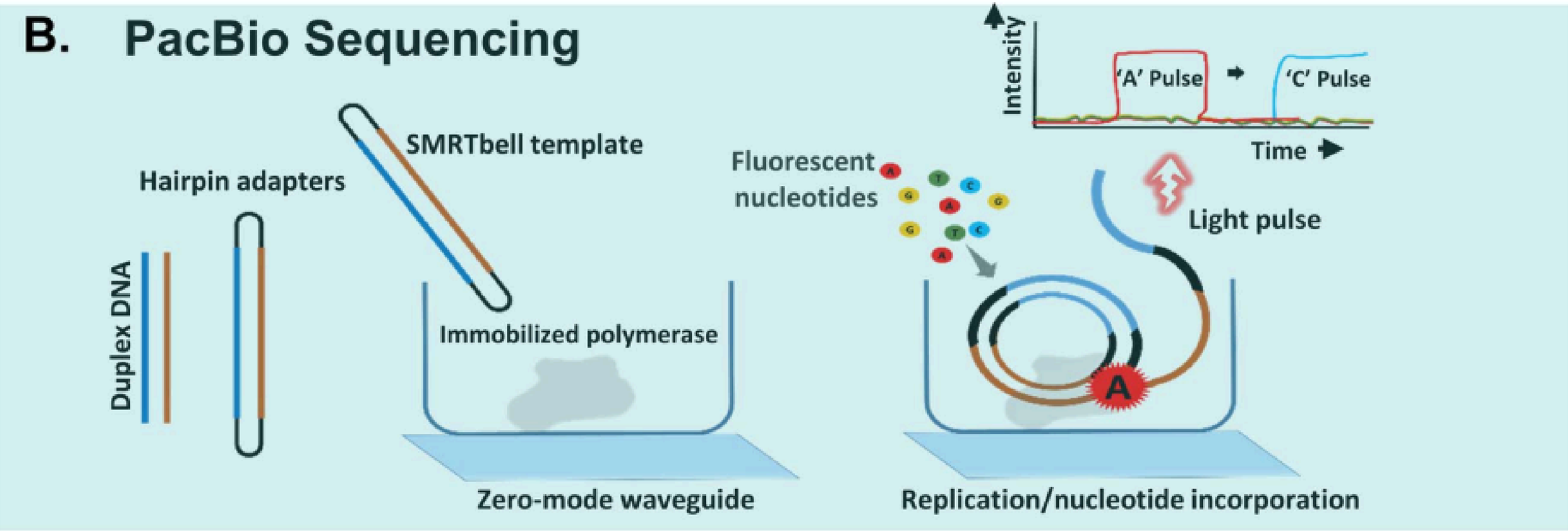
Introducción



- Lecturas Cortas: 50-300 pb
- Alta Precisión
- Millones de lecturas por corrida
- Bajo costo por base y tiempos cortos

Bharti & Grimm, 2019

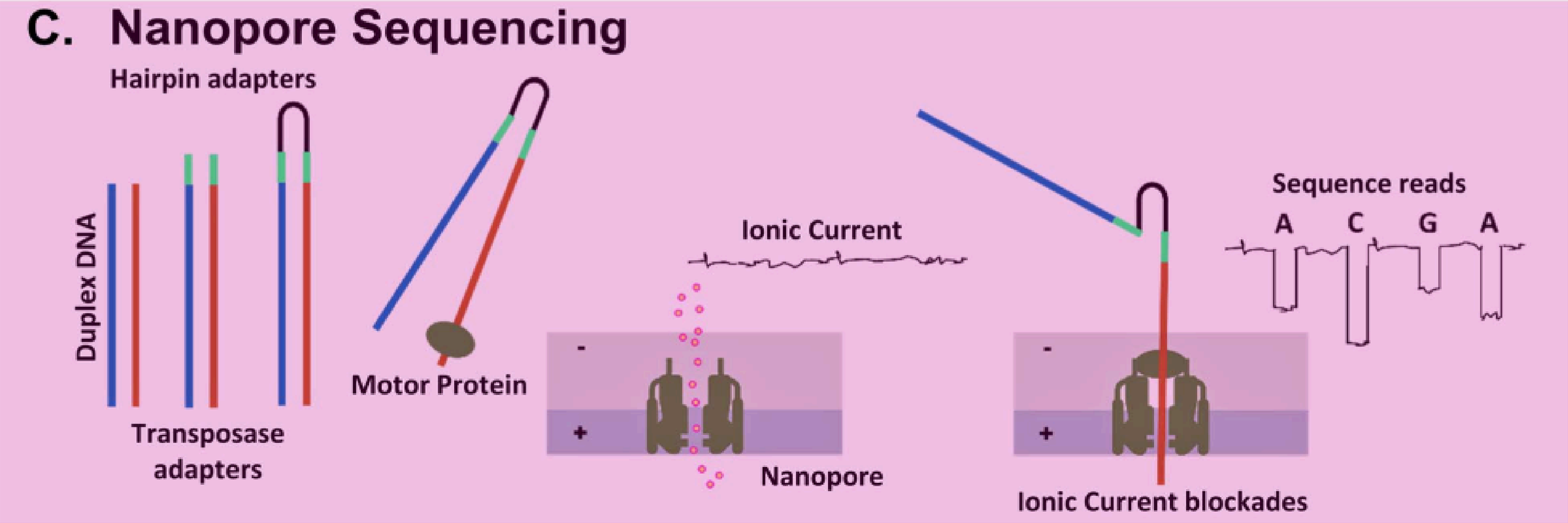
Introducción



- Lecturas largas: 10,000-100,000 pb.
- Secuenciación en tiempo real y detección epigenética.
- Errores inicial (10-15%) a >99.9% con consenso.
- Necesita más ADN y es más costosa

Bharti & Grimm, 2019

Introducción



- Lecturas ultra largas: Hasta >1 Mb.
- Detecta cambios en la corriente eléctrica mientras las bases de ADN/ARN pasan por un poro.
- Portátil y escalable
- Error alto(~5-15%) que otras tecnologías

Bharti & Grimm, 2019

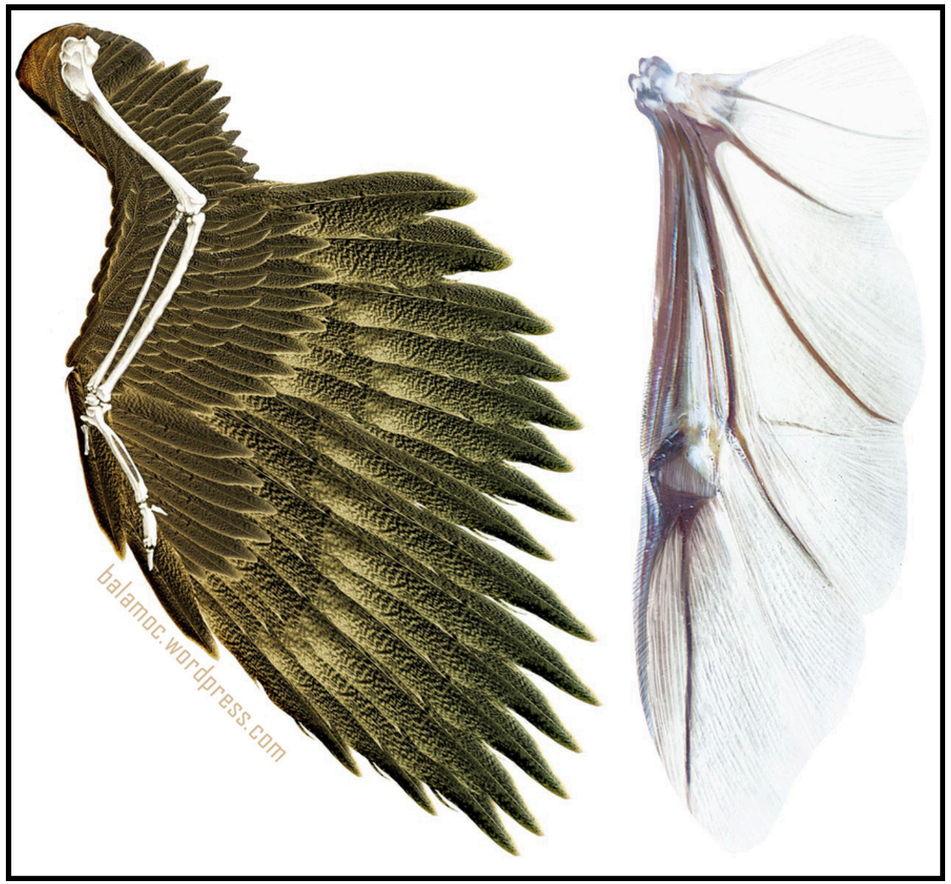
Introducción

Table 1. Summary of the main advantages of long-reads sequencing over short-read sequencing.

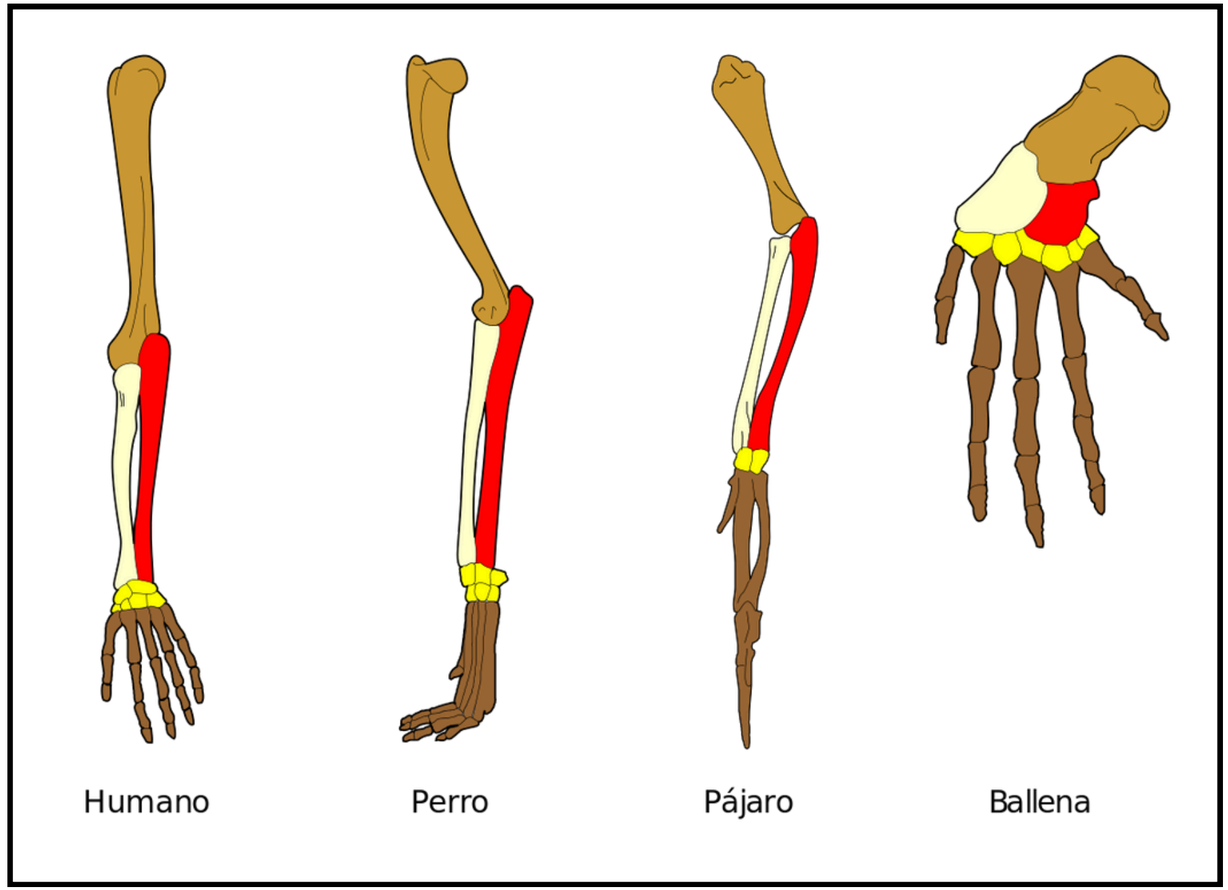
Short-Read Technologies	Long-Read Technologies
Fixed run time: - Increased time to results and inability to identify workflow errors before completed sequencing - Additional practical complexities associated with handling and storing large volumes of sequence data	Real-time data acquisition: - Achieve rapid turnaround with immediate access to results - Enrich single targets during sequencing, with no additional sample prep using adaptive sampling - Identify microbiome composition and resistance in real-time
Limited flexibility: - Sample batching often required for optimal efficiency - Potentially leads to long turnaround times - Benchtop devices confine sequencing to centralized locations	Scalable and flexible: - Scale to suit the throughput needs - Decentralize sequencing - No sample batching needed
Read length typically 50–300 bp	Unrestricted read length (>4 Mb achieved)
Limited genomic characterization: - Short reads do not span entire structural variants or important classes of genomic aberrations (repeat expansions and repeat-rich regions) - fragmented genome assemblies and ambiguous isoforms identification - Short sequencing reads may not span complex genomic regions such as genes duplications, transposons and prophage sequences - Potentially missing important genomic information	Comprehensive genomic characterization: - Identify mutations in complex and repetitive genomic regions - Accurately phase single nucleotide variants, structural variants, and base modifications - Can fully assemble genomes more easily - Simplify de novo assembly and correct microbial reference genomes - Possibility to completely assemble genomes and plasmids from metagenomic samples - Resolving complex genomic regions and similar species
Amplification required: - Amplification can introduce bias reducing uniformity of coverage and removes base modifications - Necessitating additional sample prep and sequencing runs	Amplification-free protocols: - Detect and phase base modifications as standard - No additional preparation required
Constrained to the lab: - Traditional sequencing technologies are typically expensive and require substantial site infrastructure - Usually limited its usage to well-resourced environments - Delay in transmitting the results	Sequence anywhere: - Sequence in your lab or in the field - Sequence at sample source and eliminate sample shipping delays - Scale-up with high-throughput

Khendher et al., 2022

Introducción

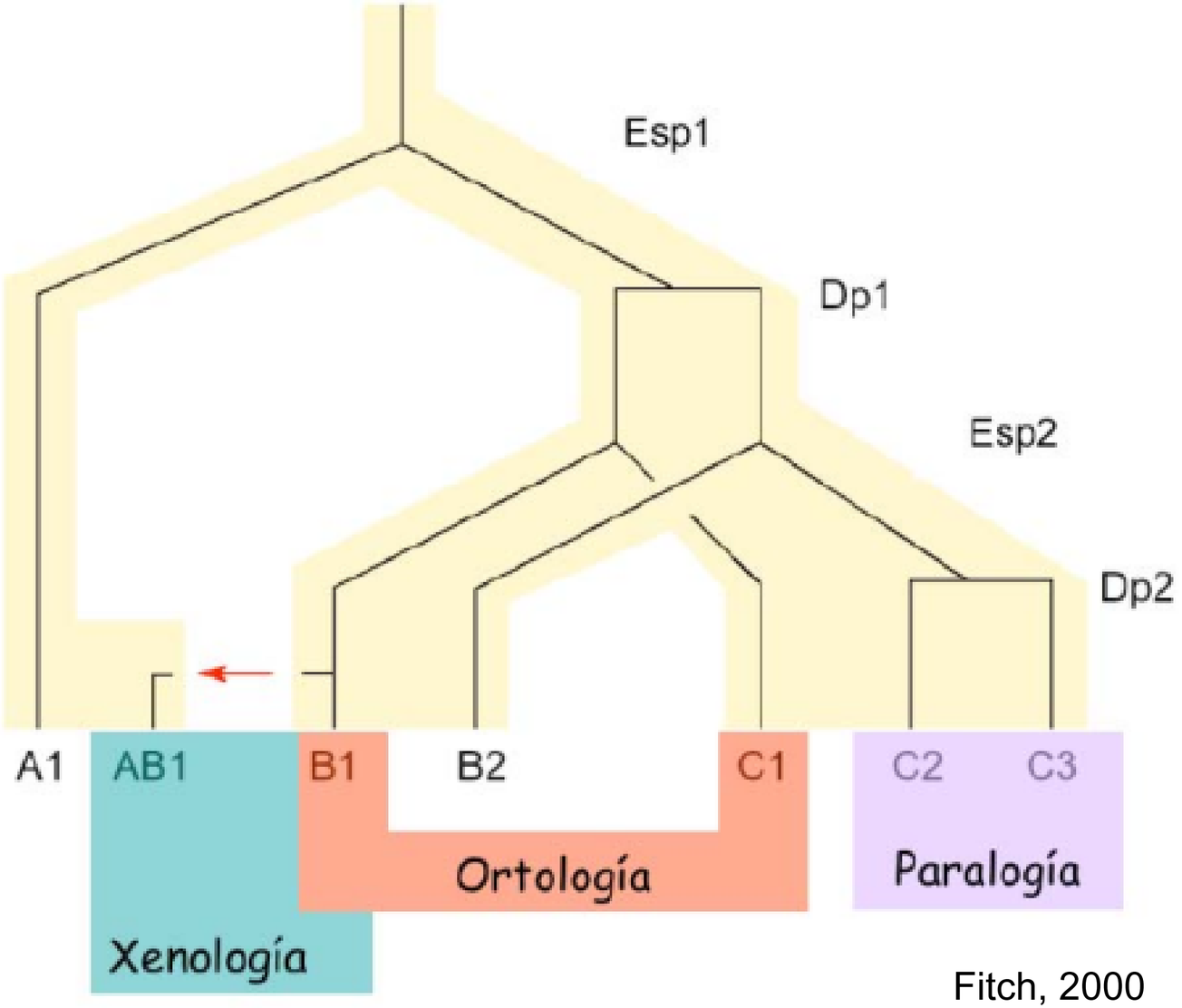


Analogía



Homología

Introducción



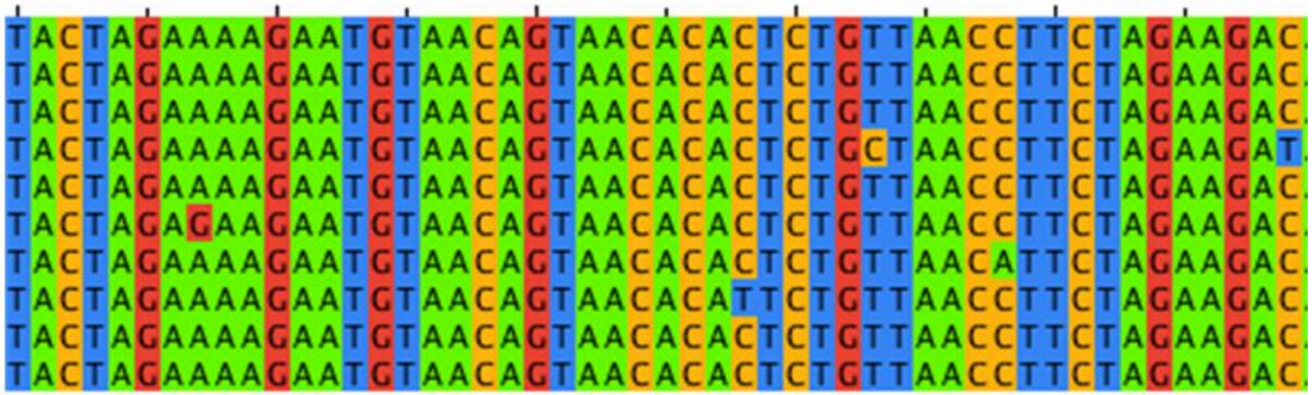
Fitch, 2000

Introducción

What is an Alignment?

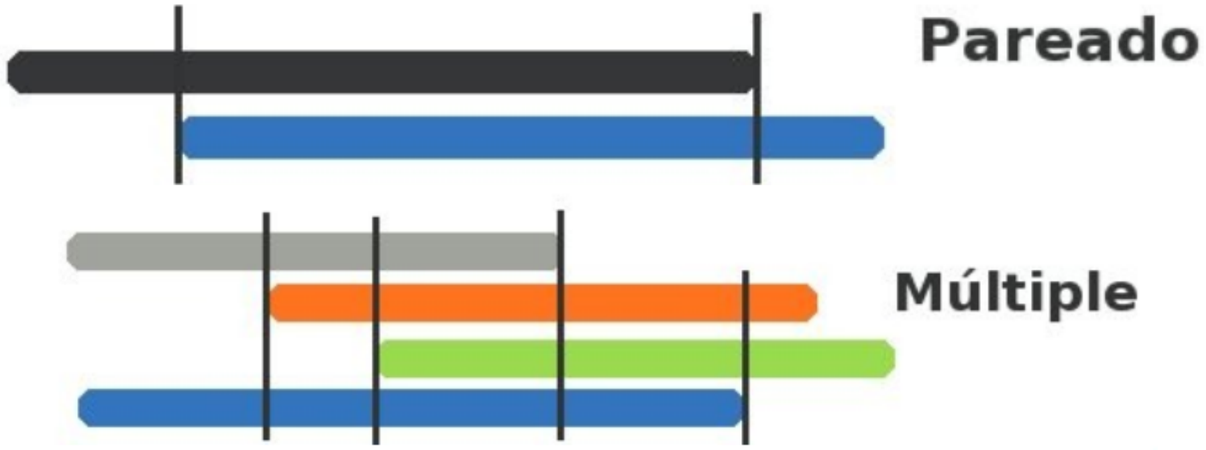
Unaligned

Aligned

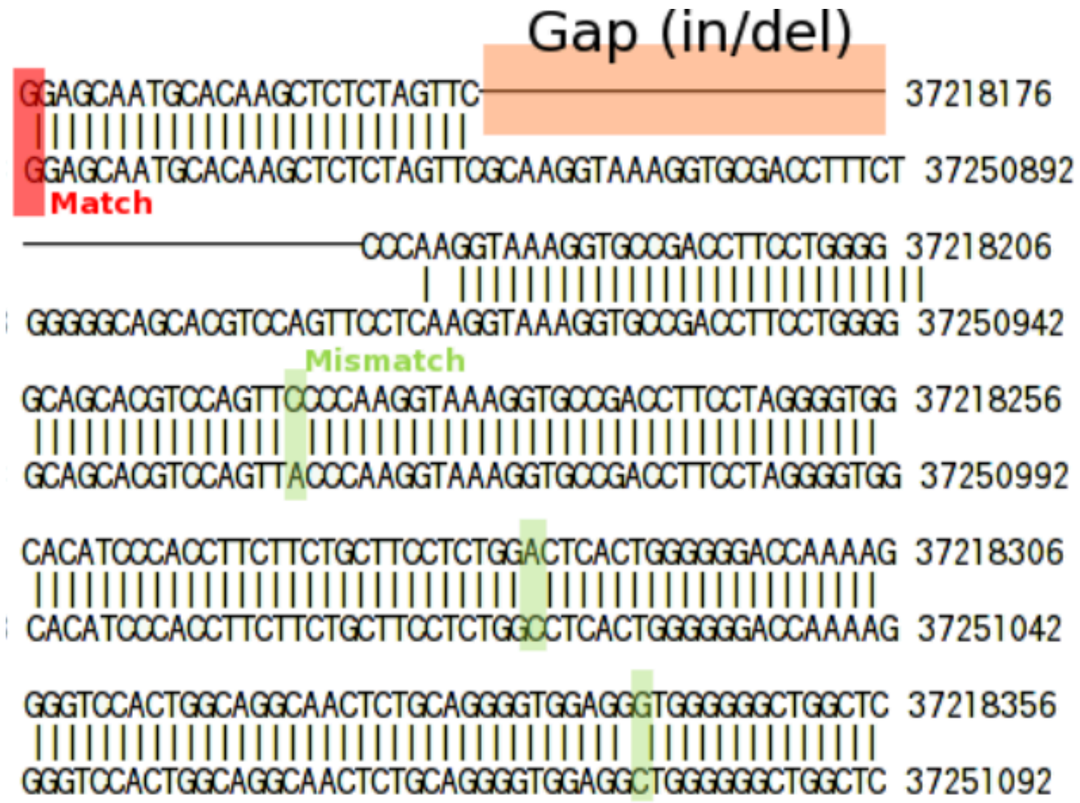


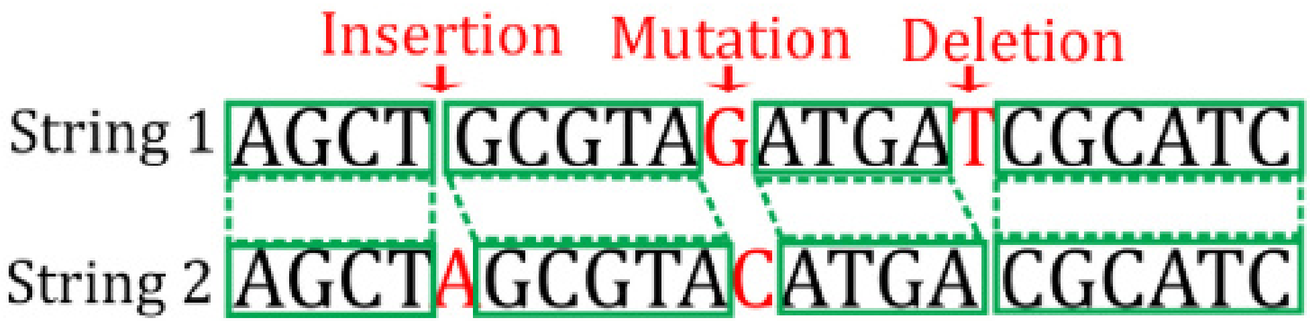
- Lining up related (homologous) positions
  - Allows comparison

Introducción



Introducción





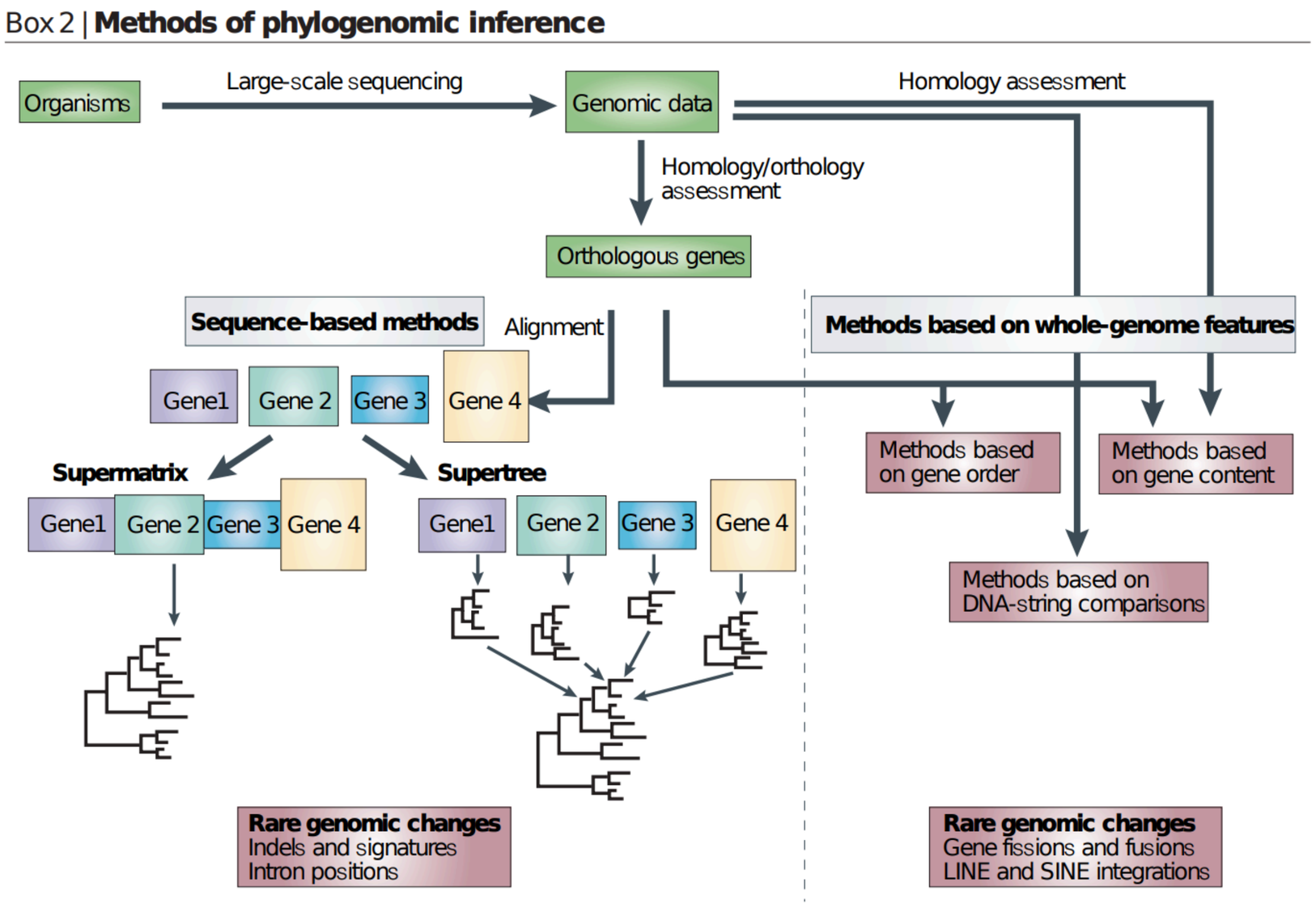
Introducción

FILOGENÓMICA

Intersección de la filogenética y la genómica: El uso de grandes conjuntos de datos genómicos para resolver preguntas filogenéticas.

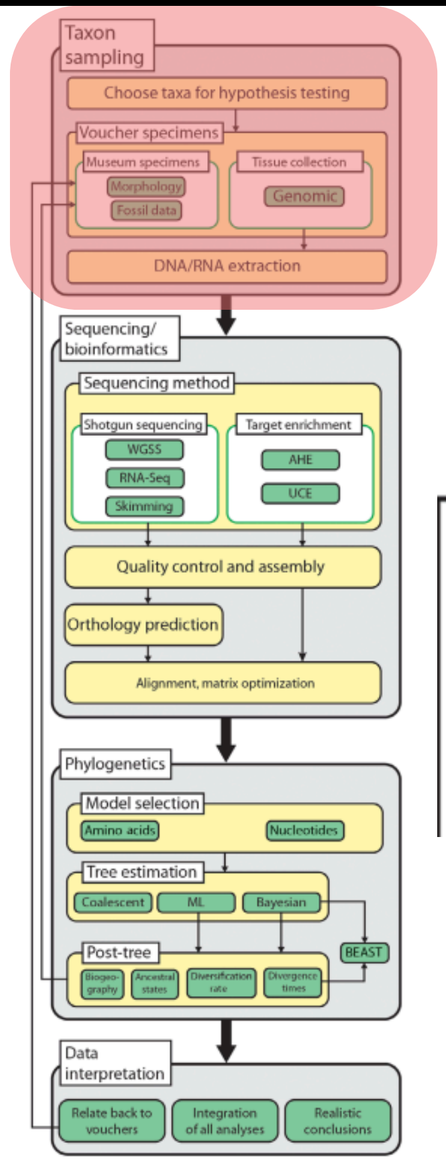
Las relaciones entre los taxones se infieren en función de la homología (herencia de un ancestro común, comúnmente observada como patrones de similitud de secuencias) a lo largo de genomas completos, ya sea mediante un enfoque comparativo gen a gen, concatenado multigen o de genoma completo. (Xin-Chan & Ragan, 2013; Burki, 2014).

Introducción

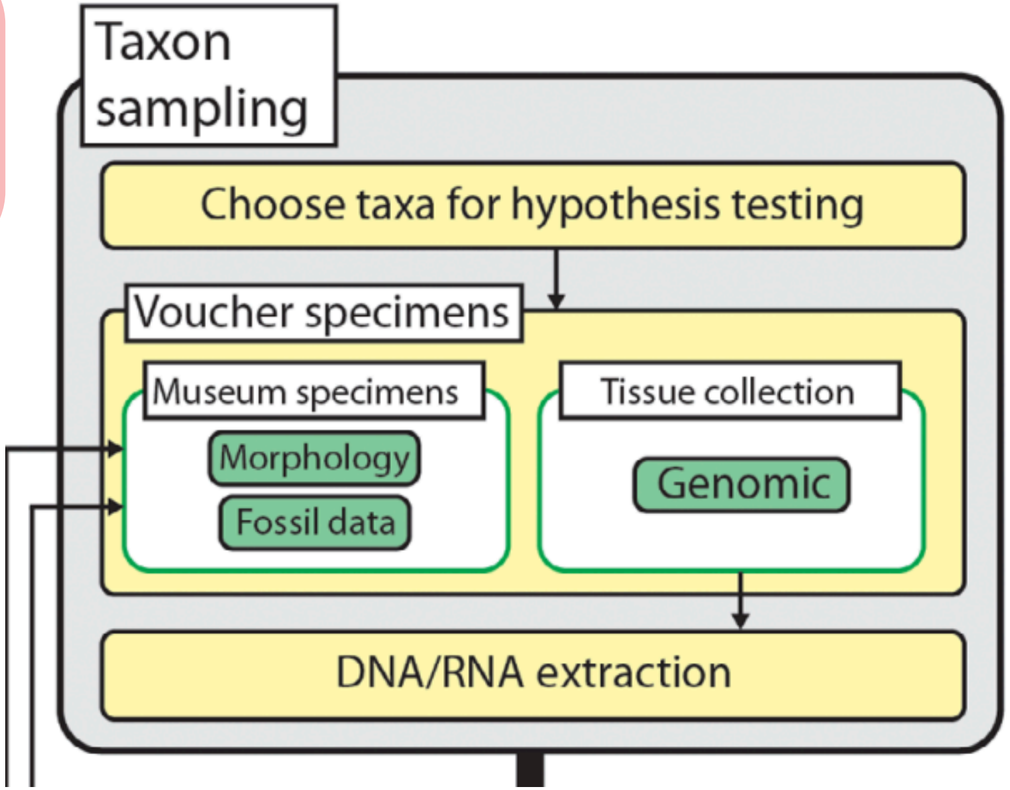


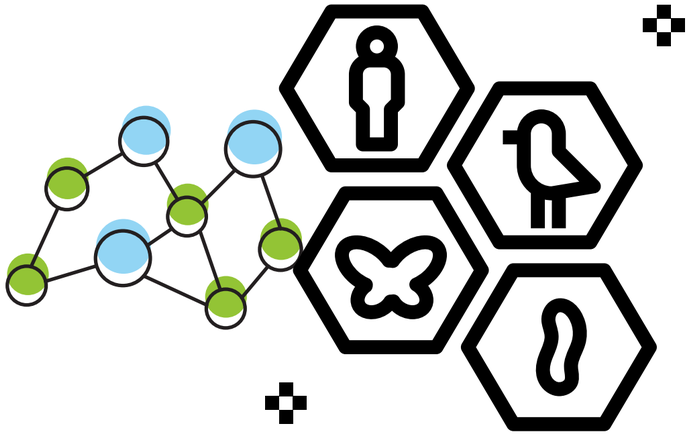
Delsuc et al., 2005

Introducción

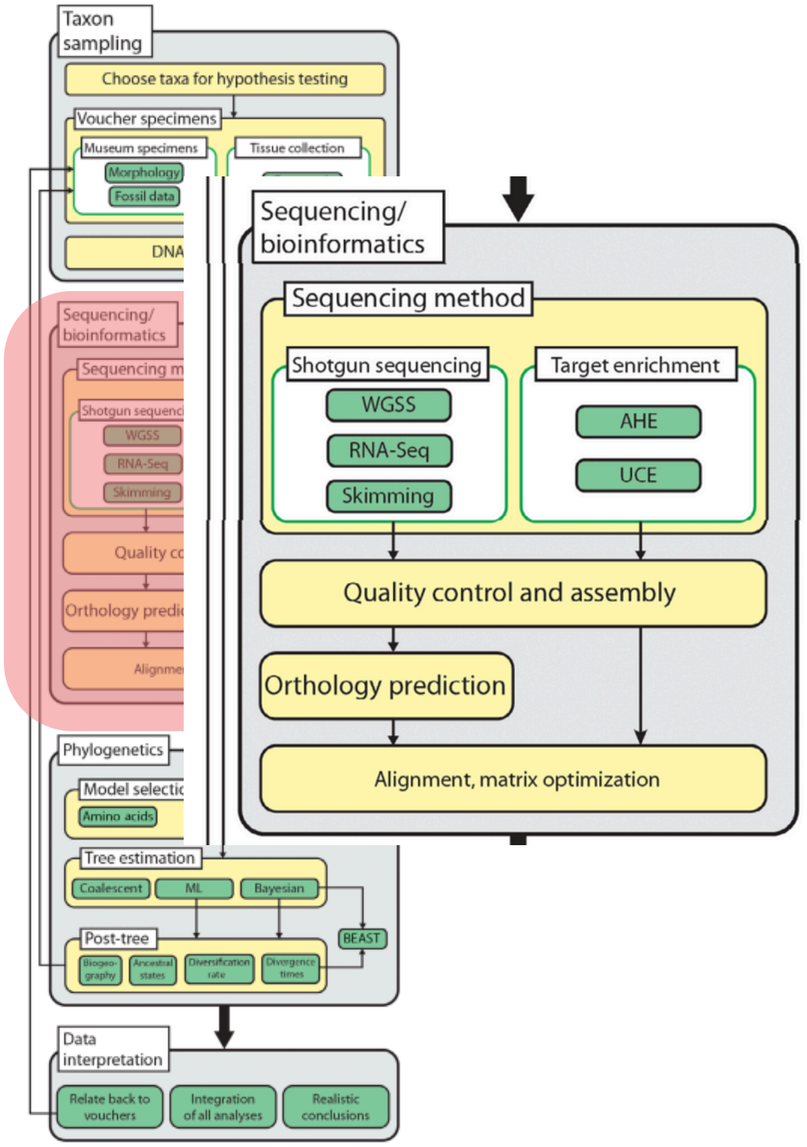


Young & Gillung, 2019

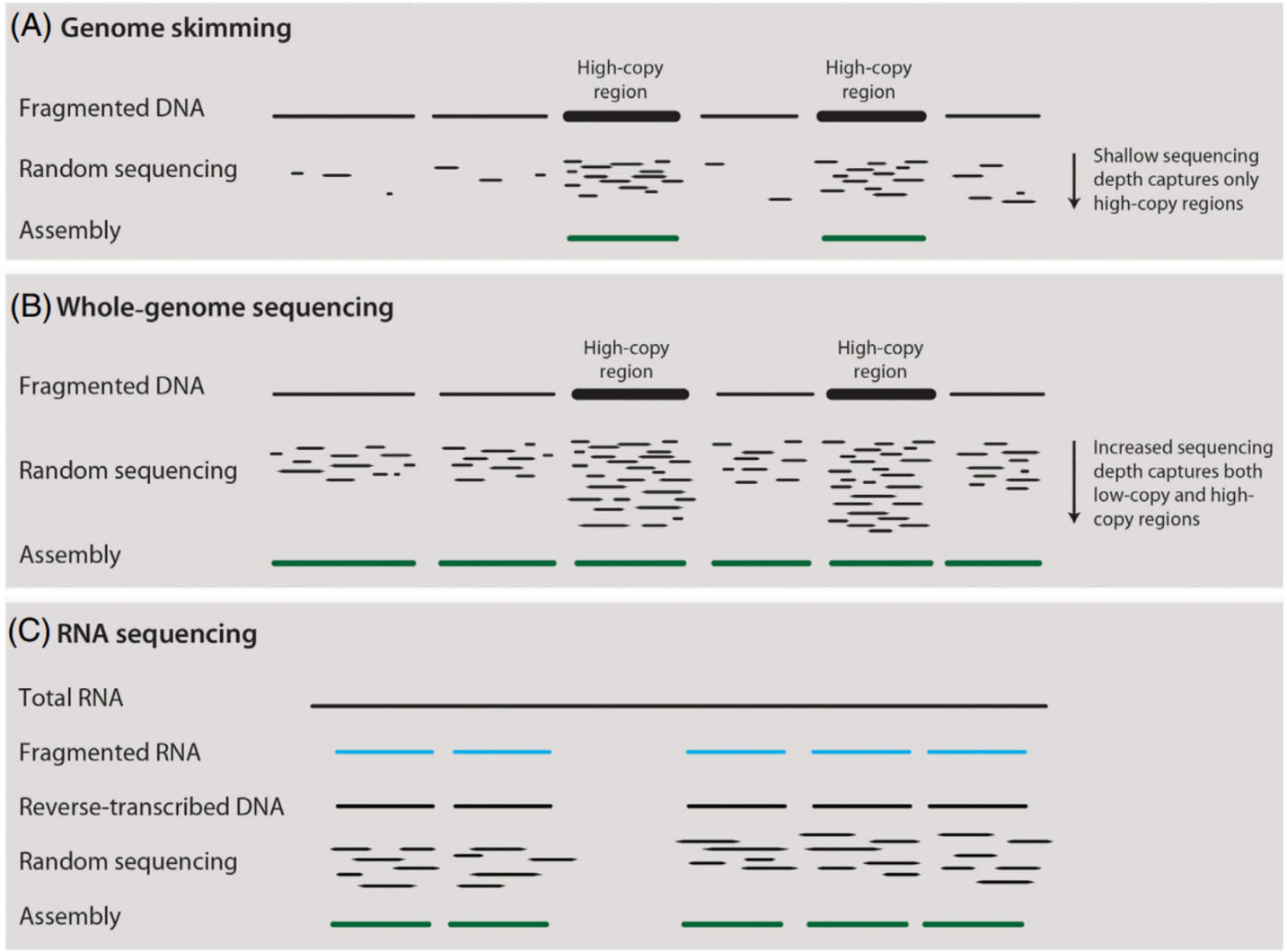




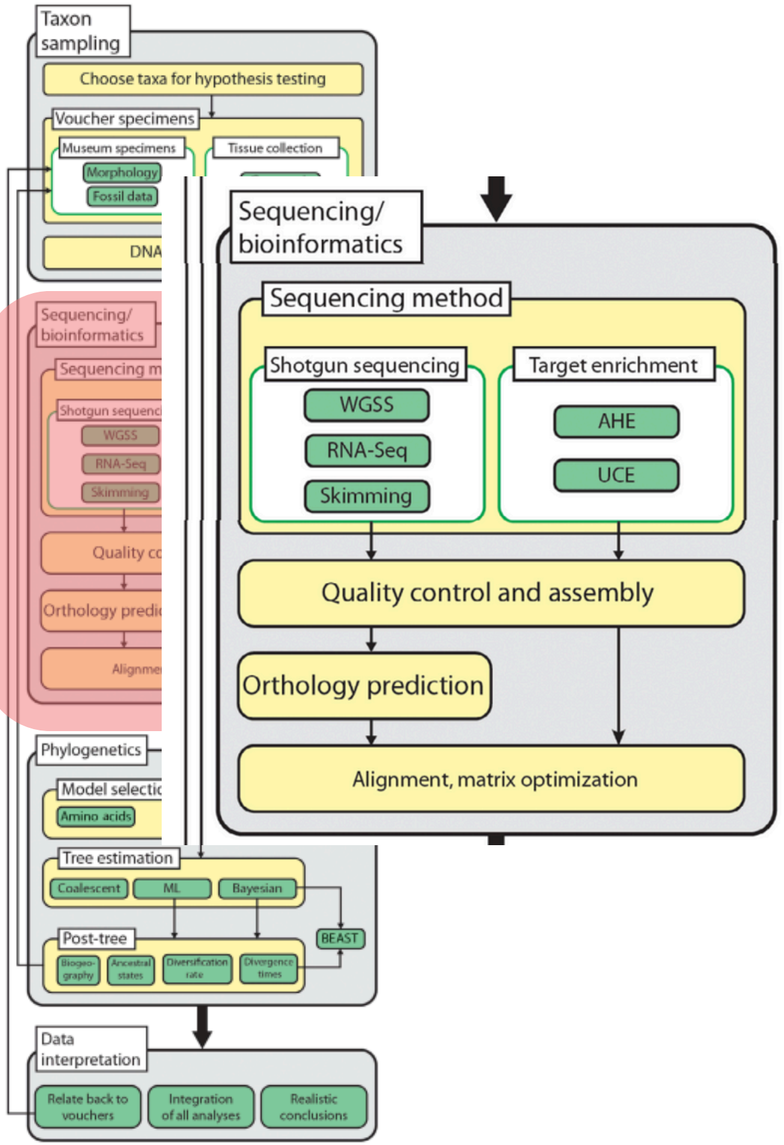
Introducción



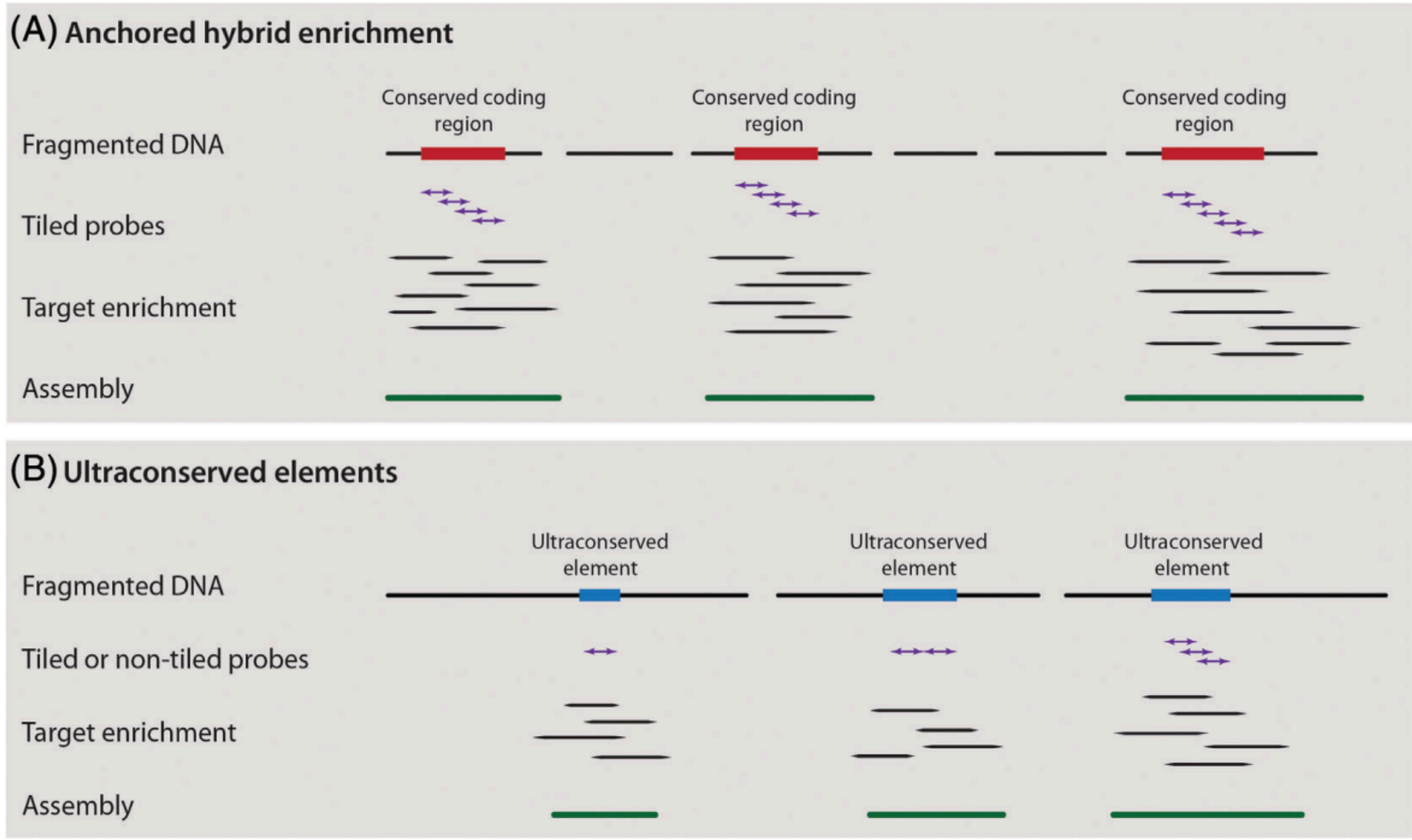
Young & Gillung, 2019



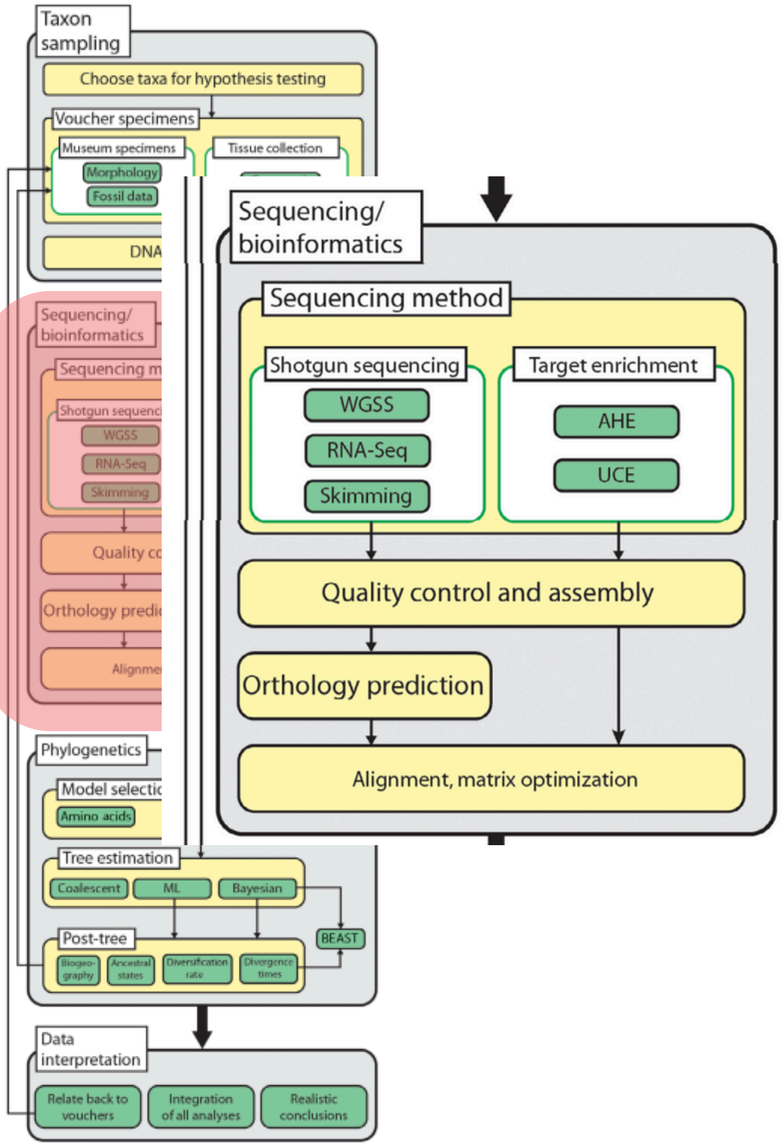
Introducción



Young & Gillung, 2019



Introducción



Young & Gillung, 2019

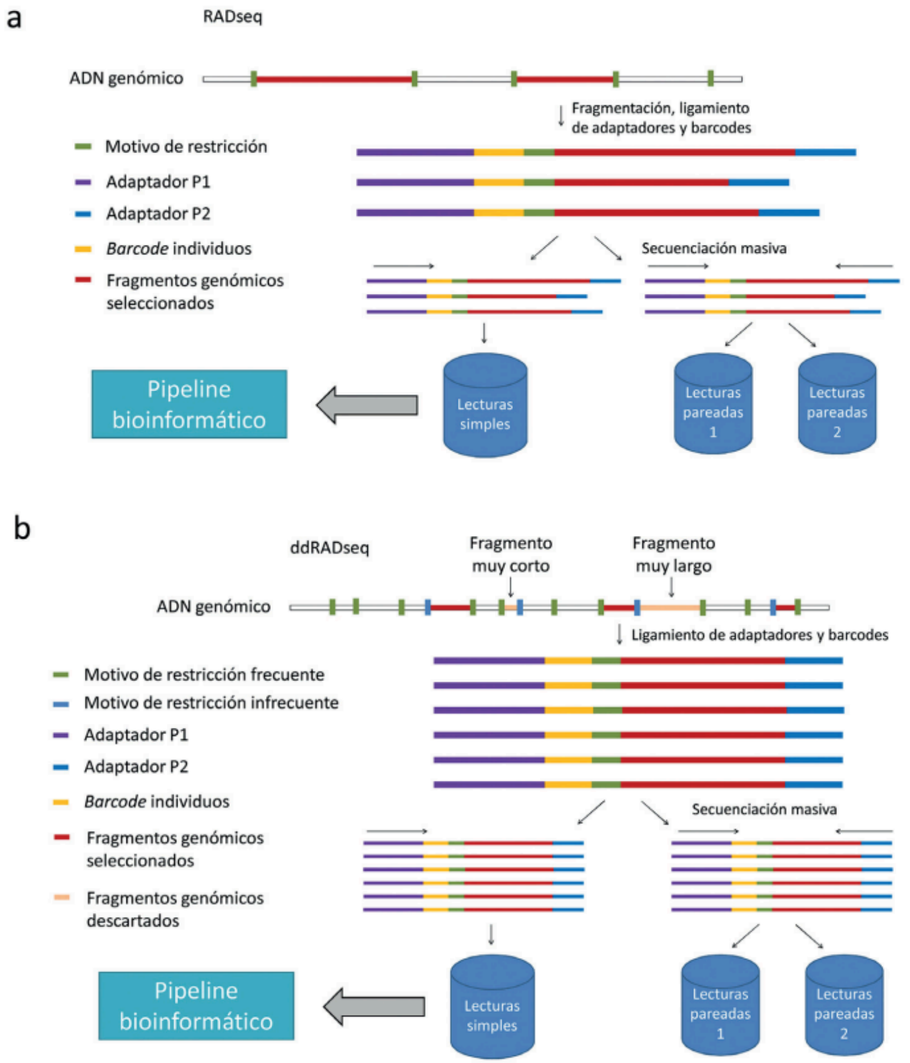
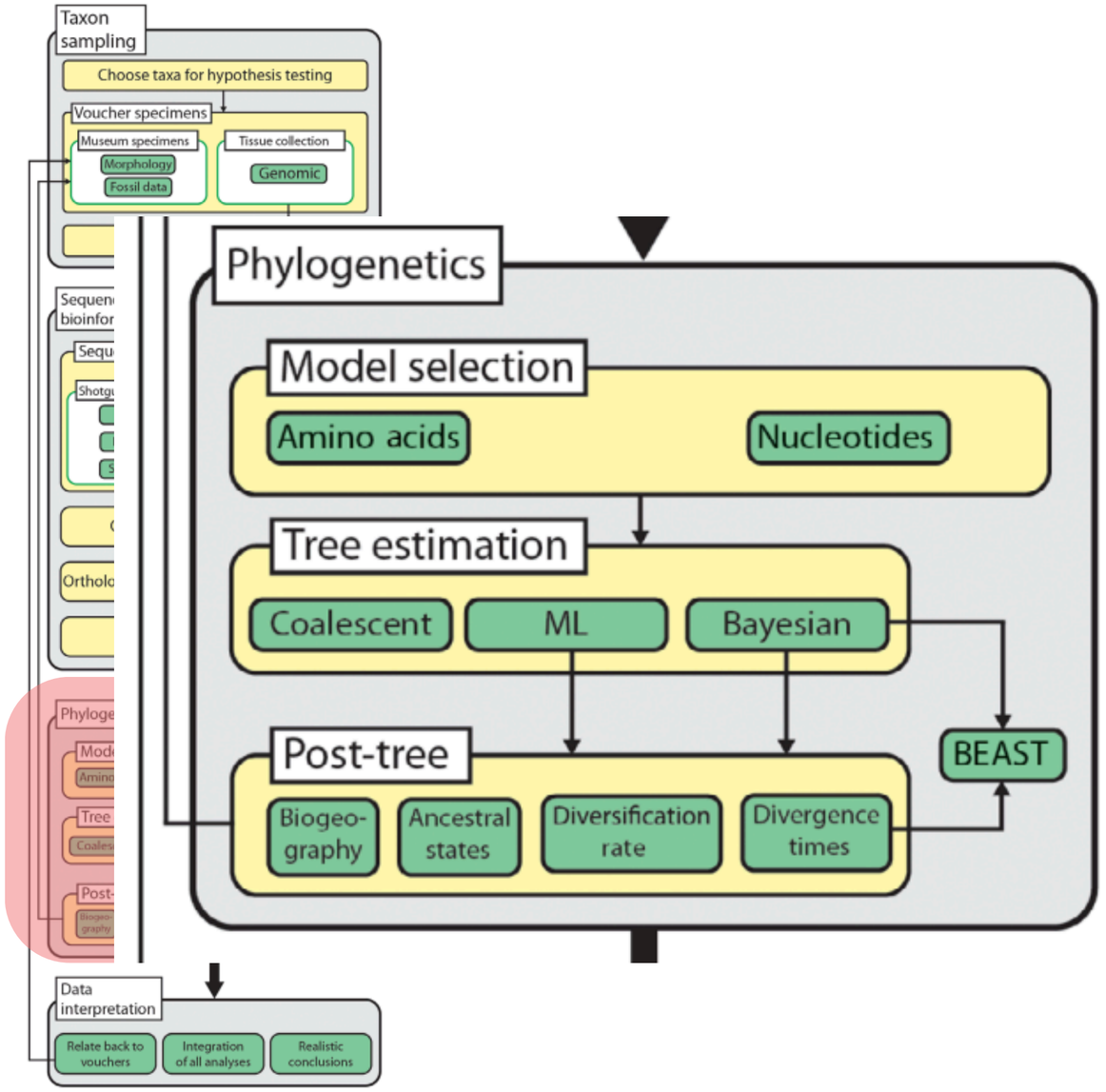


Fig. 2.- Esquema de las metodologías de genotipado RAD-seq (a) y ddRAD-seq (b).

Fig. 2.- Outline of RAD-seq (a) and ddRAD-seq (b) genotyping methodologies.

Loópez de Heredia, 2016

Introducción



Young & Gillung, 2019

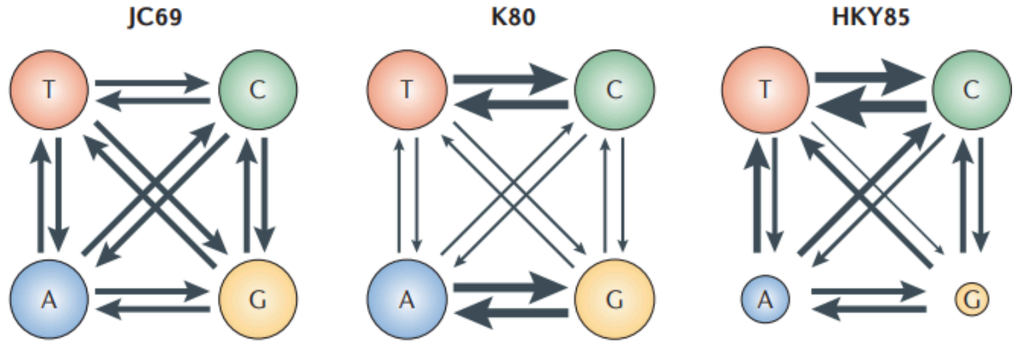
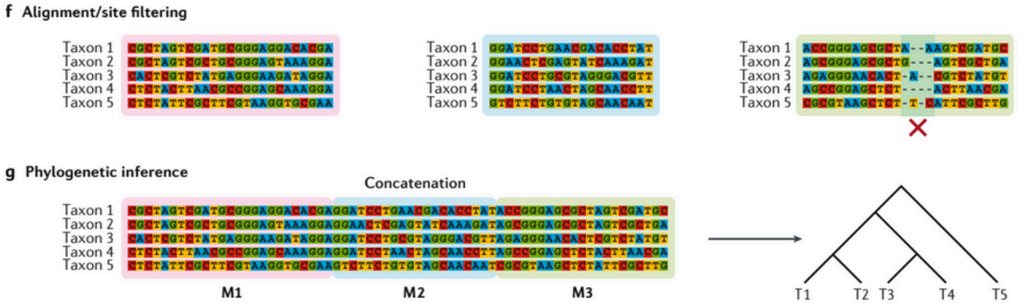


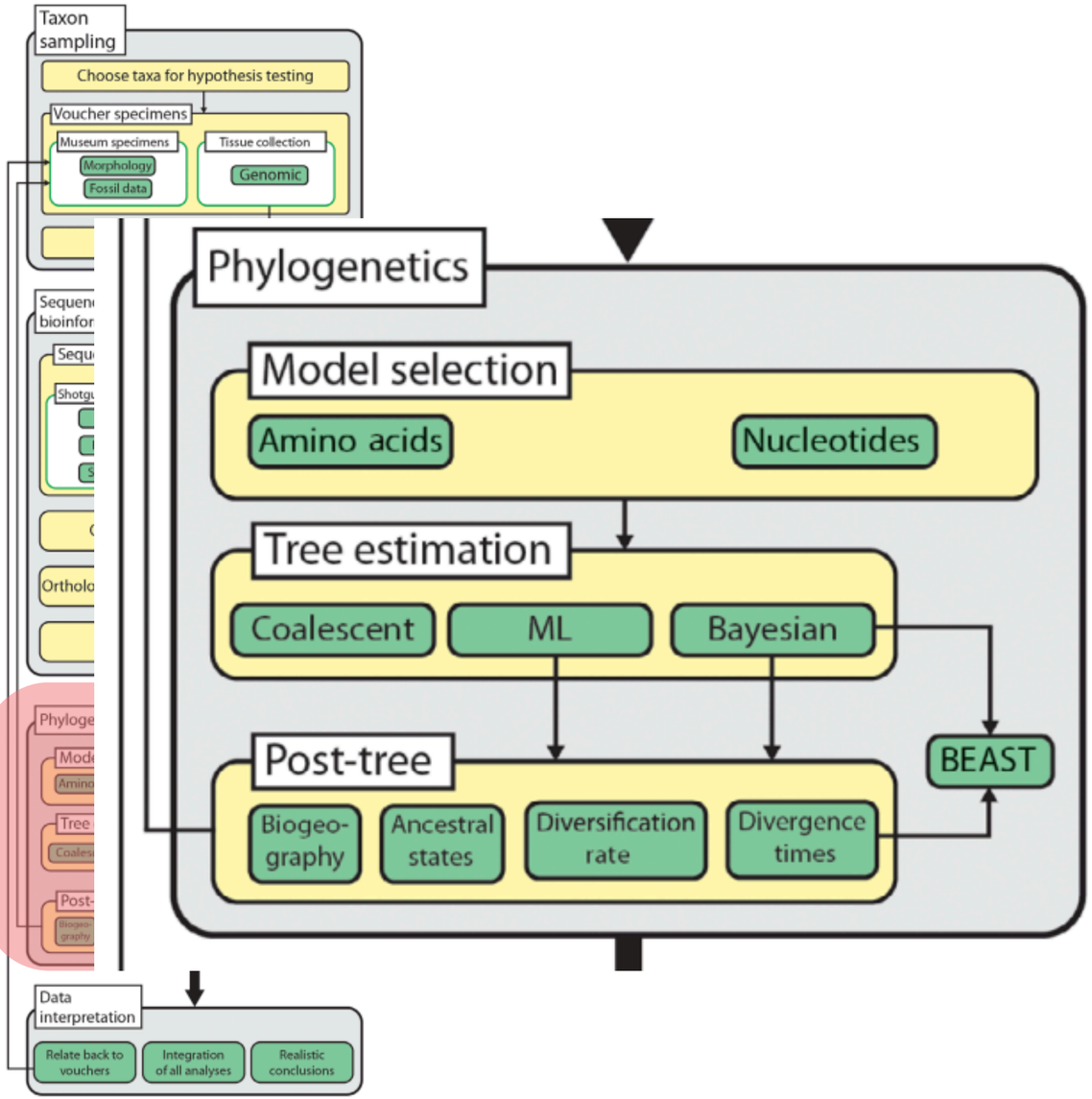
Figure 1 | **Markov models of nucleotide substitution.** The thickness of the arrows indicates the substitution rates of the four nucleotides (T, C, A and G), and the sizes of the circles represent the nucleotide frequencies when the substitution process is in equilibrium. Note that both JC69 and K80 predict equal proportions of the four nucleotides.

Yang & Rannala, 2012

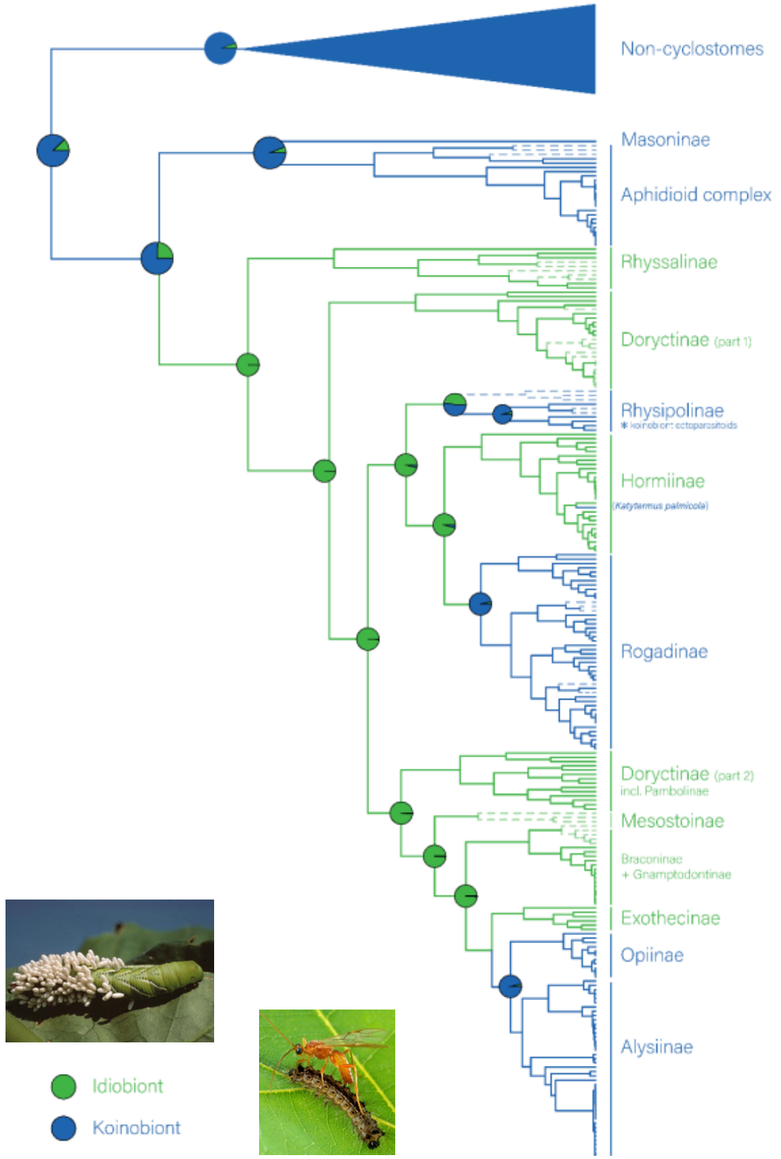


Kapli et al., 2020

Introducción

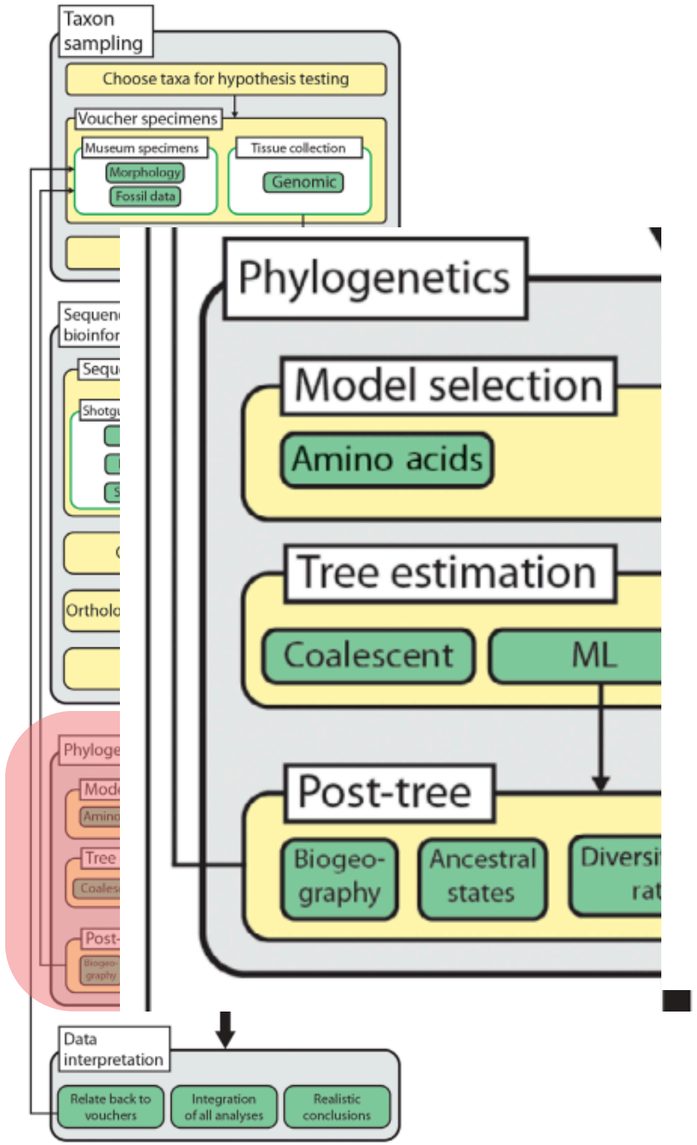


Young & Gillung, 2019

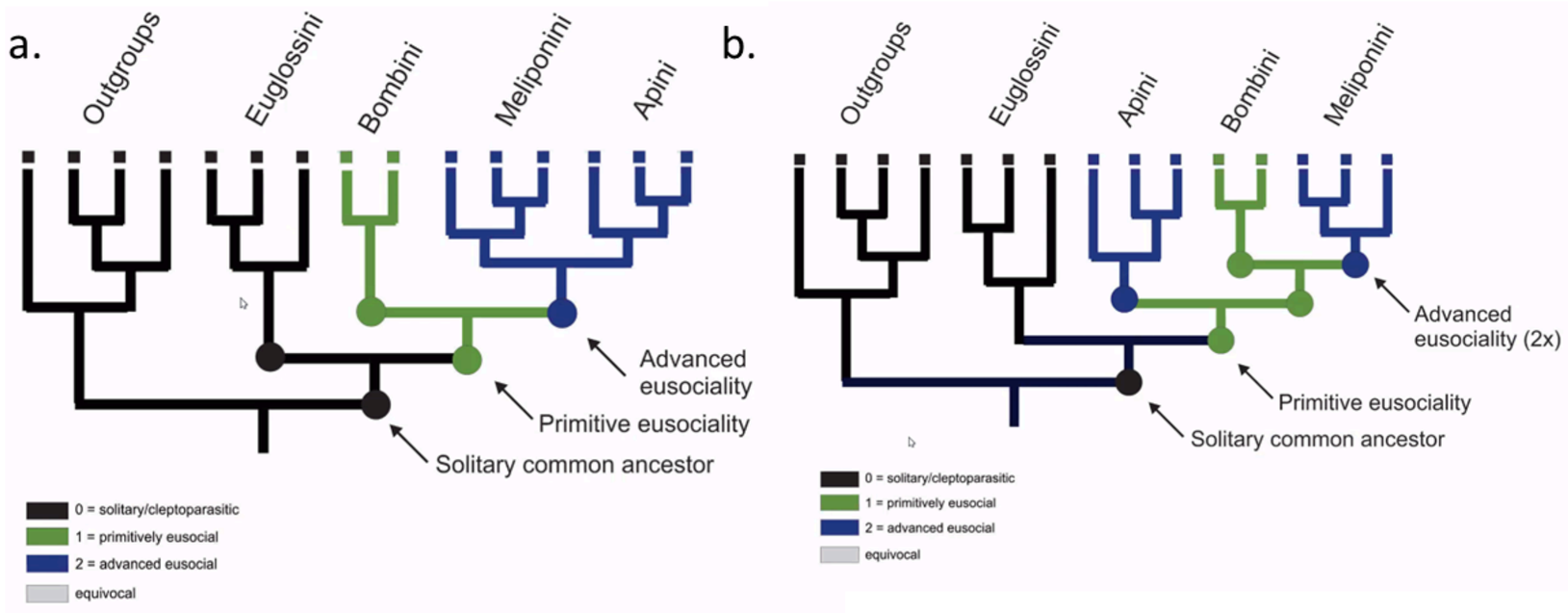


Jasso-Martínez et al., 2022

Introducción

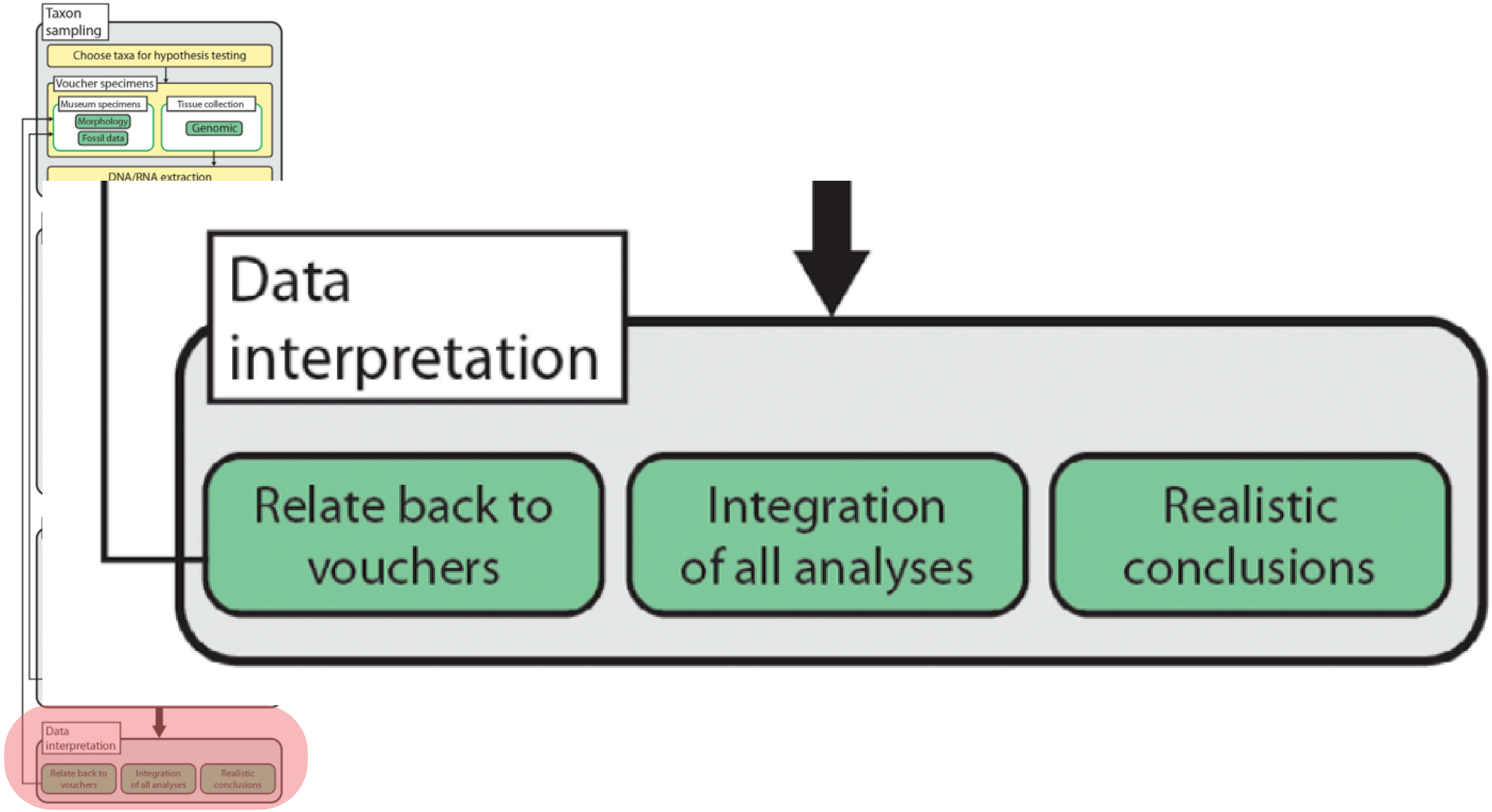


Young & Gillung, 2019



Bossert et al., 2017

Introducción



Young & Gillung, 2019