

Curso de Análisis Evolutivo de Genomas en el Linaje Verde, casos prácticos con Orthofinder y otras herramientas complementarias

Marcos Ramos González
Francisco J. Romero-Campero

Research Group on Microalgae Systems Biology and Biotechnology
Dpt Computer Science and Artificial Intelligence
Institute for Plant Biochemistry and Photosynthesis
(Universidad Sevilla – CSIC)

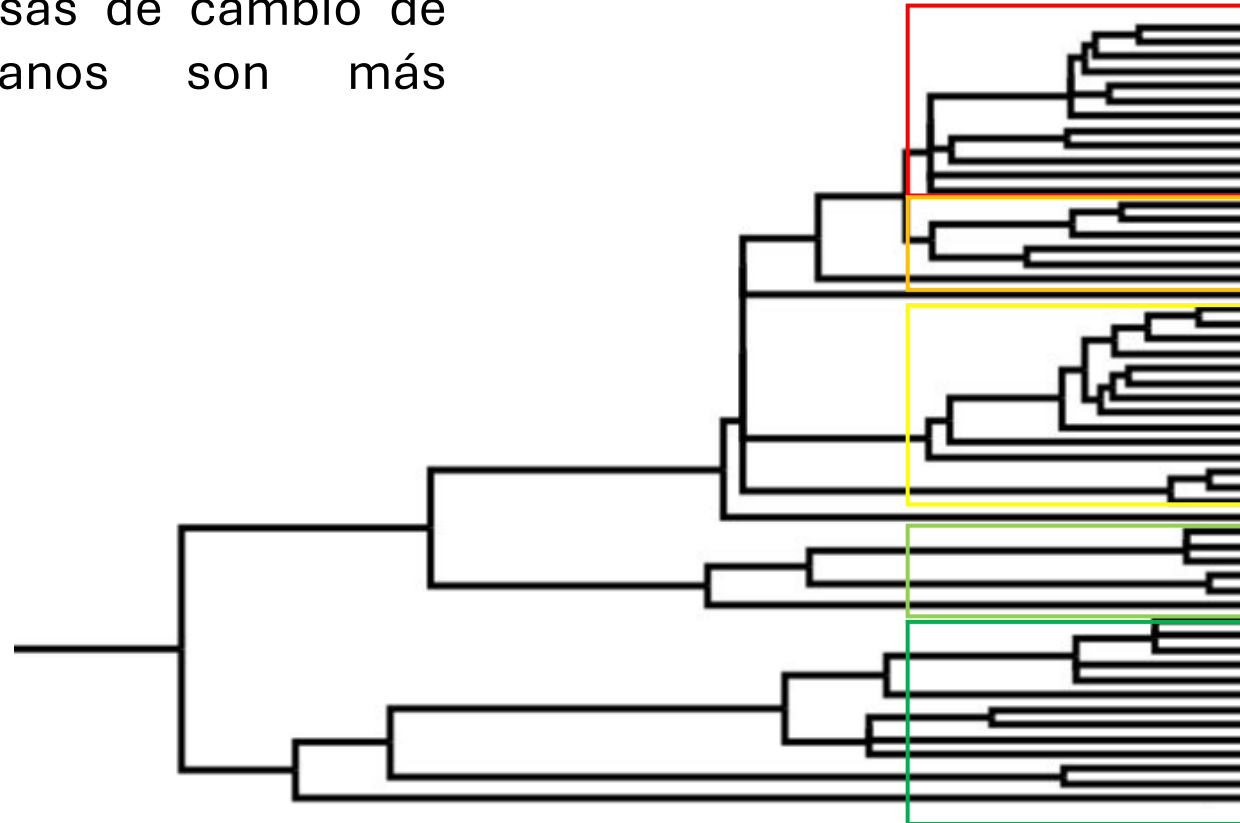
Calibración de árboles de especies

Un reloj molecular es una técnica que utiliza la tasa de mutación de las biomoléculas para deducir el momento en el que divergieron las distintas especies de un árbol.

Sin embargo, para utilizarlos con el objetivo de datar los árboles, es necesario calibrar estos relojes, ya que por sí solos no tienen información de tiempos absolutos, sólo de relación entre tasas de mutación. Esta calibración se puede realizar utilizando el registro fósil para imponer limitaciones temporales a los nodos internos y relacionar esos tiempos absolutos con las tasas de evolución.

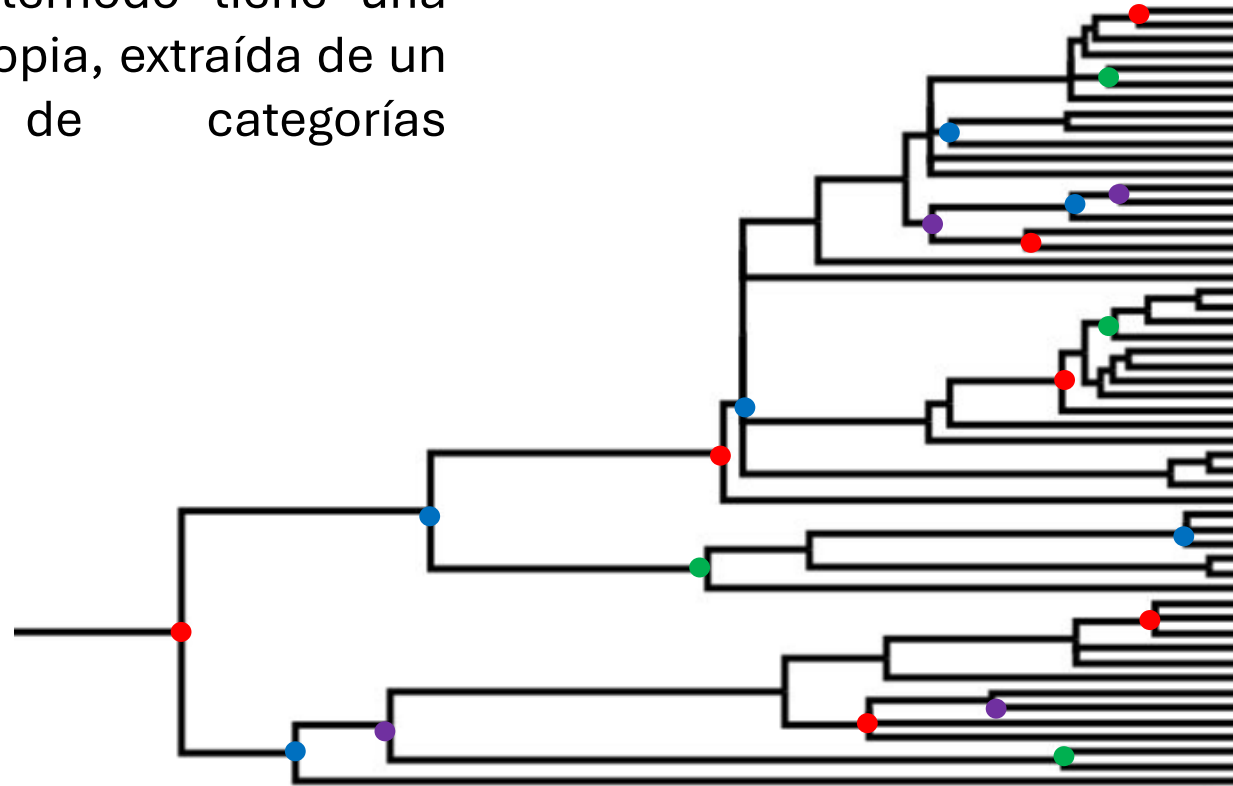
Modelos de calibración de árboles

Correlated: las tasas de cambio de internodos cercanos son más similares



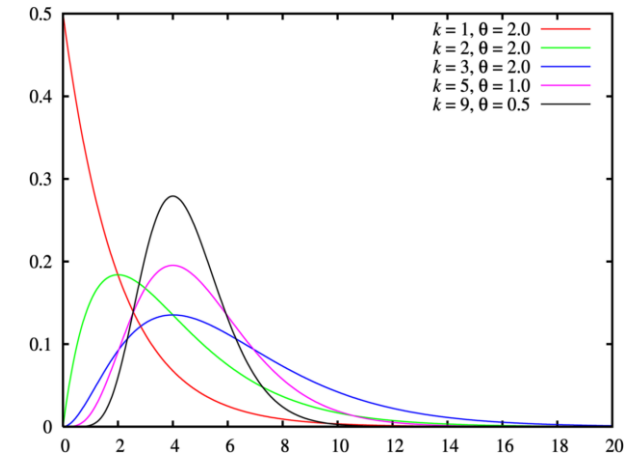
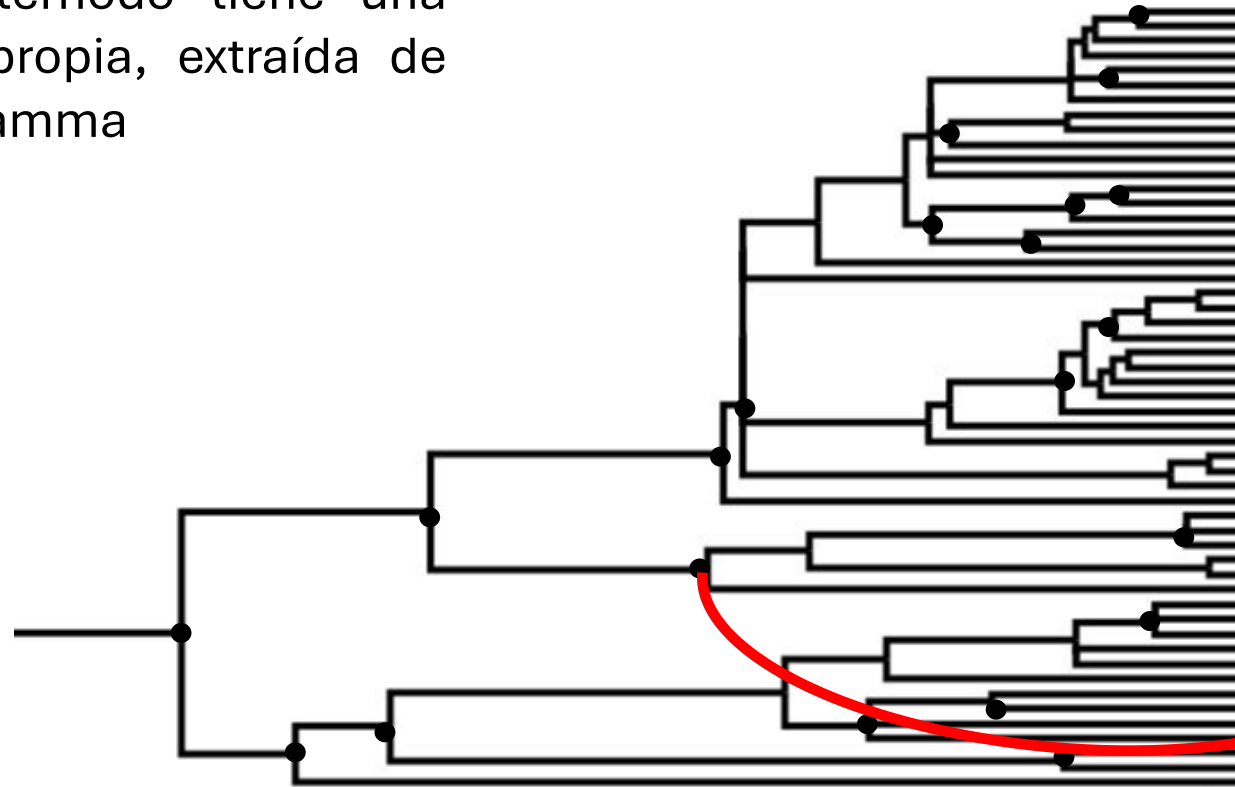
Modelos de calibración de árboles

Discrete: cada internodo tiene una tasa de cambio propia, extraída de un número n de categorías preseleccionadas



Modelos de calibración de árboles

Relaxed: cada internodo tiene una tasa de cambio propia, extraída de una distribución gamma



TimeTree

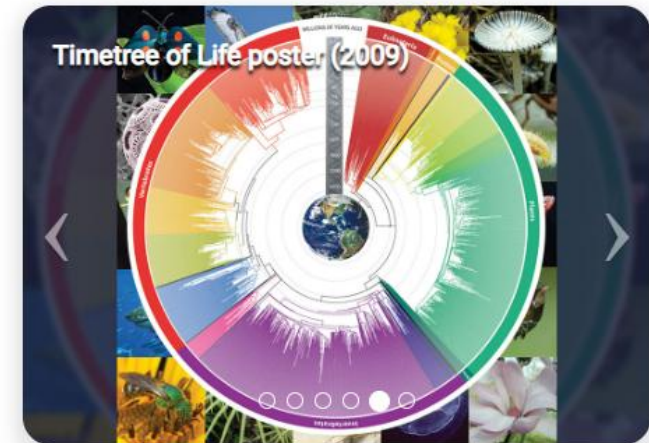
TimeTree es una base de datos con información sobre los tiempos de divergencia de diferentes especies. Se puede acceder desde: <https://timetree.org/>.



Timetree of Life

TimeTree is a public knowledge-base for information on the evolutionary timescale of life. Data from thousands of published studies are assembled into a searchable tree of life scaled to time.

Timepanels showing events in geological time and astronomical history are provided for comparison with timelines and timetrees. Results can be exported in different formats for additional analyses and publication.

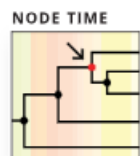


TimeTree

Para conocer las restricciones en el tiempo de divergencia de dos especies utilizamos Get Divergence Time.

Search

Get Divergence Time



Specify 2 Taxon Names [?]

Taxon 1:

Taxon 2:

Search



Resolve Ambiguity [?]

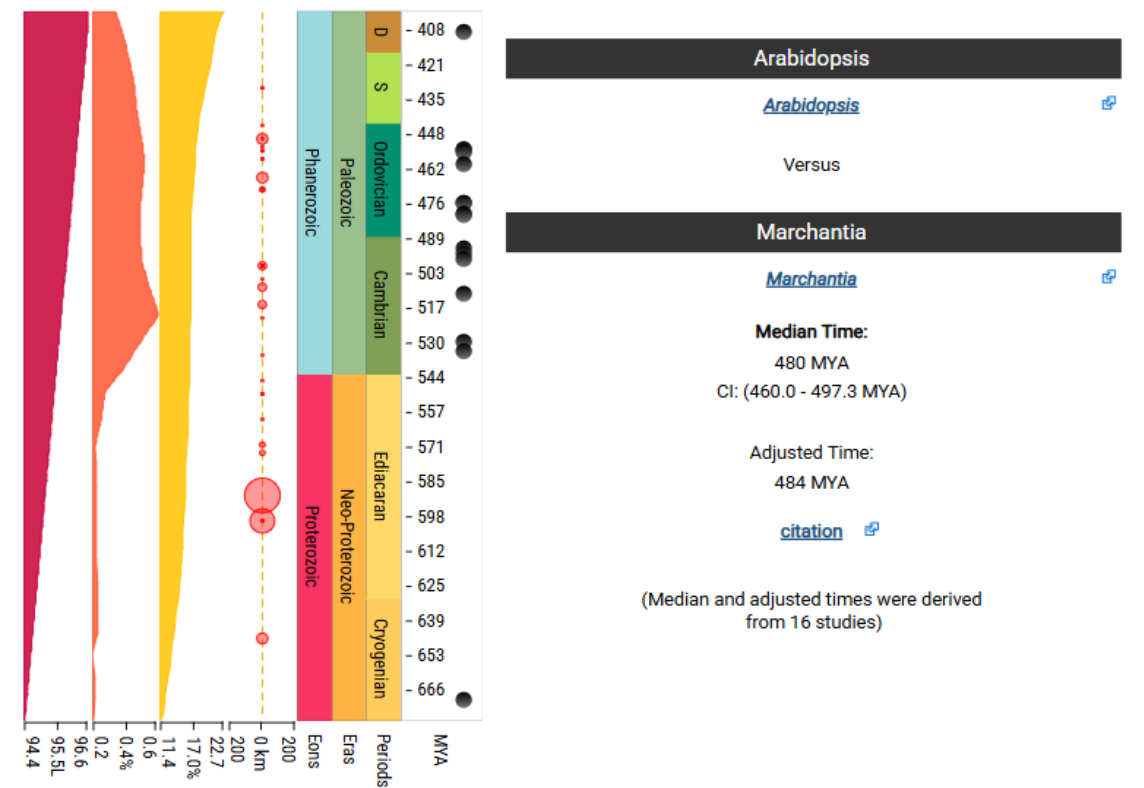
Taxon 1:

Taxon 2:

Show Time

TimeTree

TimeTree devuelve estimaciones sobre el intervalo de confianza de tiempos de divergencia de las dos especies y los estudios de donde sale la información.

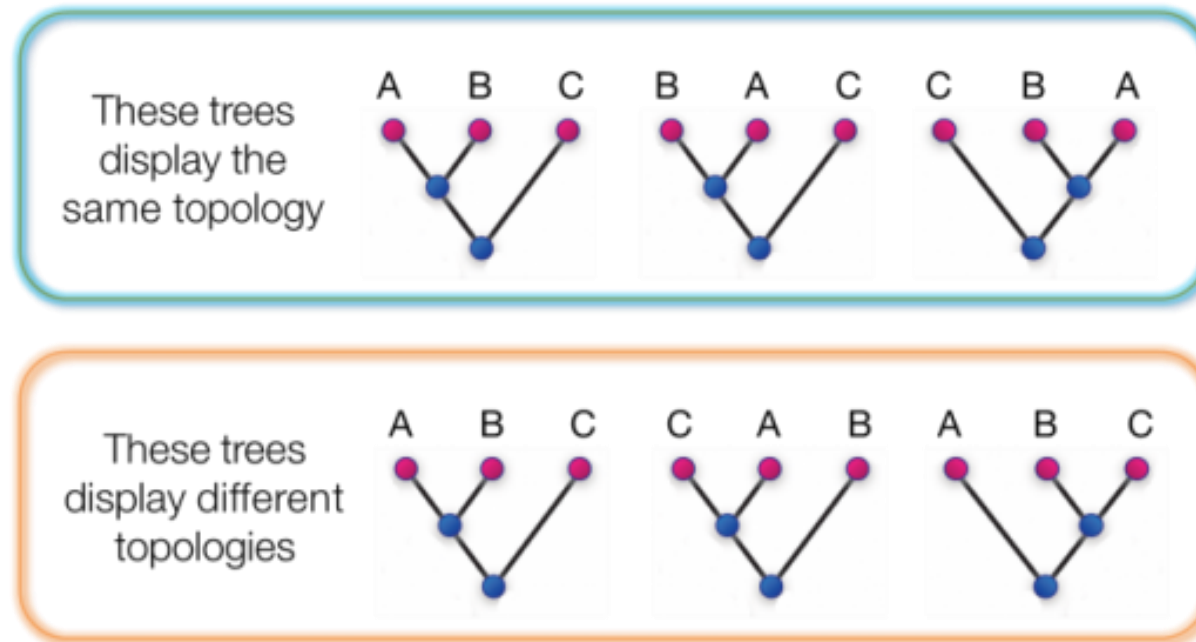


Molecular Time Estimates (16 Studies)

Copy	Excel	PDF	Print	Search:		Show	10	entries
Time	Title	Reference	Year	Data				
408.4	Divergence time estimates and the evolution of major lineages in the florideophyte red algae.	Yang et al.	2016	1 tree				
454.0	A nonparametric approach to estimating divergence times in the absence of rate constancy	Sanderson et al.	1997	1 tree				

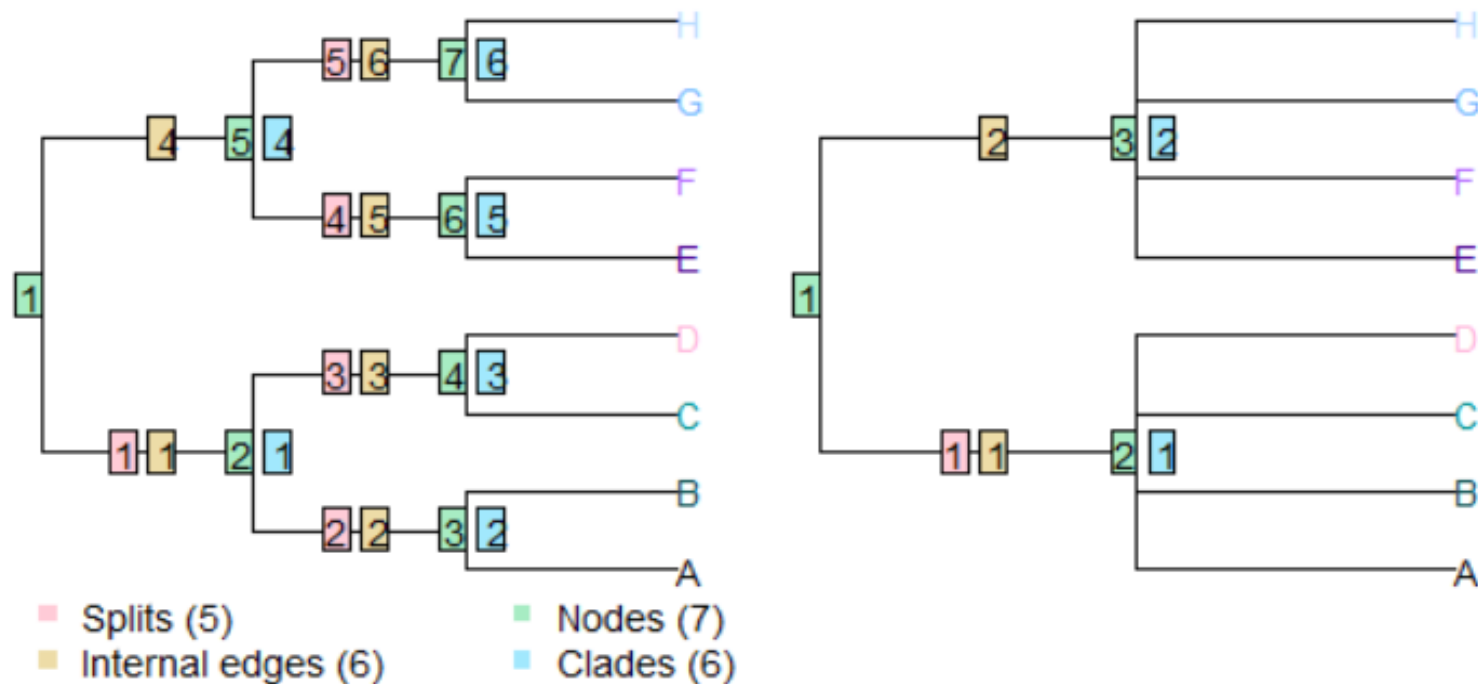
Comparación de topologías

La comparación de topologías de árboles nos permite calcular la diferencia entre dos árboles con las mismas hojas. Esto puede servir para relacionar, por ejemplo, un árbol de especies con un árbol que agrupe la información de un trait determinado en esas especies.



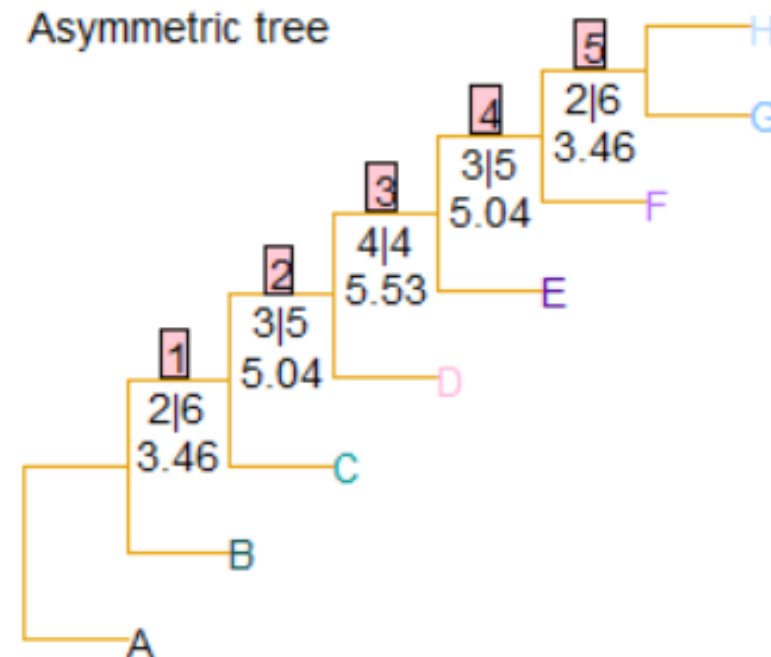
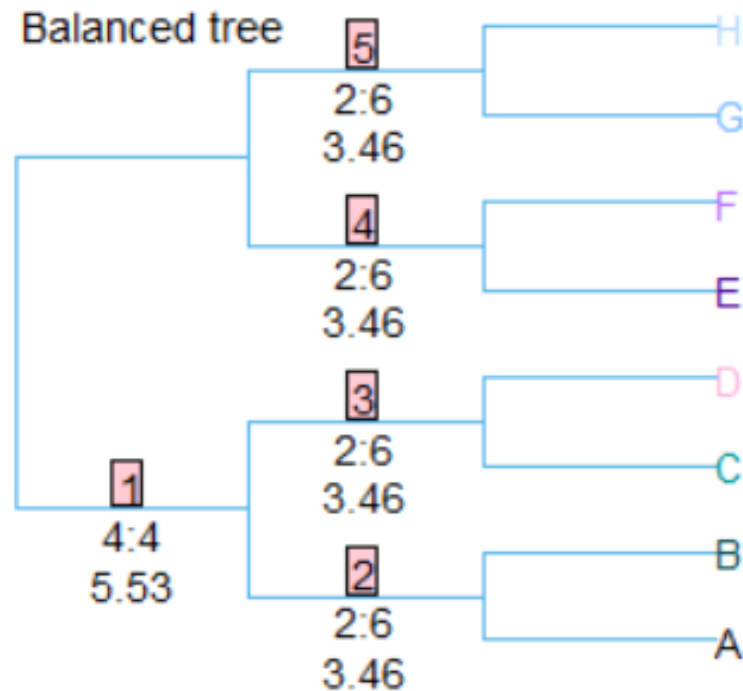
Comparación de topologías

Para compararlas, se usan los splits, que no son lo mismo que los nodos ni aristas internas. En el ejemplo izquierdo, como no hay otra forma de separar los nodos 2 y 5 (ya que 1 es sólo la raíz), esa rama corresponde a un único split.



Robinson-Foulds

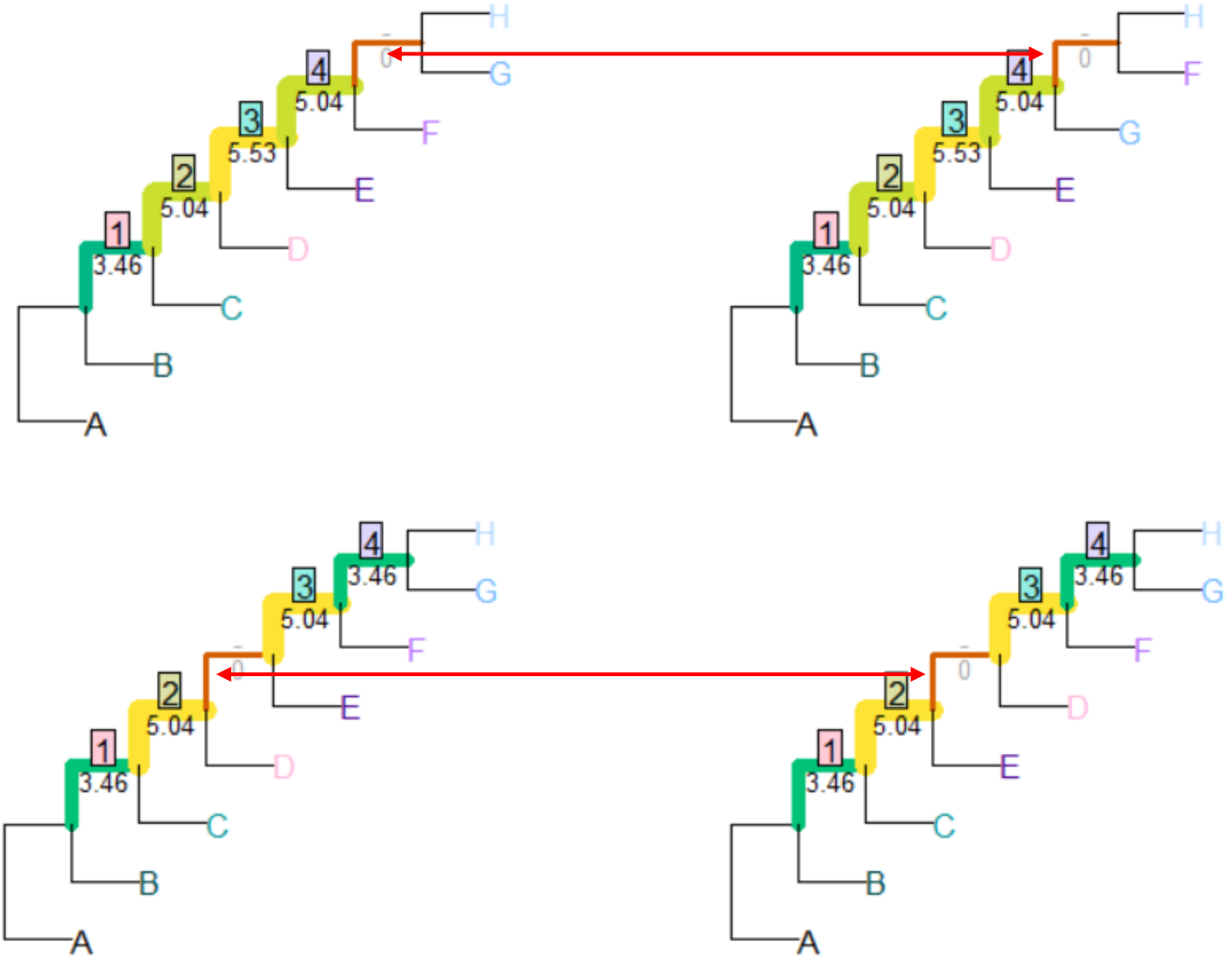
La forma más intuitiva de medirlo es la distancia de RB, que se basa en contar el número de splits únicos en cada árbol (que no aparecen en el otro). Sin embargo, esto asume una probabilidad idéntica de ocurrencia de cada split, y origina que árboles asimétricos exhiban mayores distancias que los equilibrados.



Robinson-Foulds corregida por información

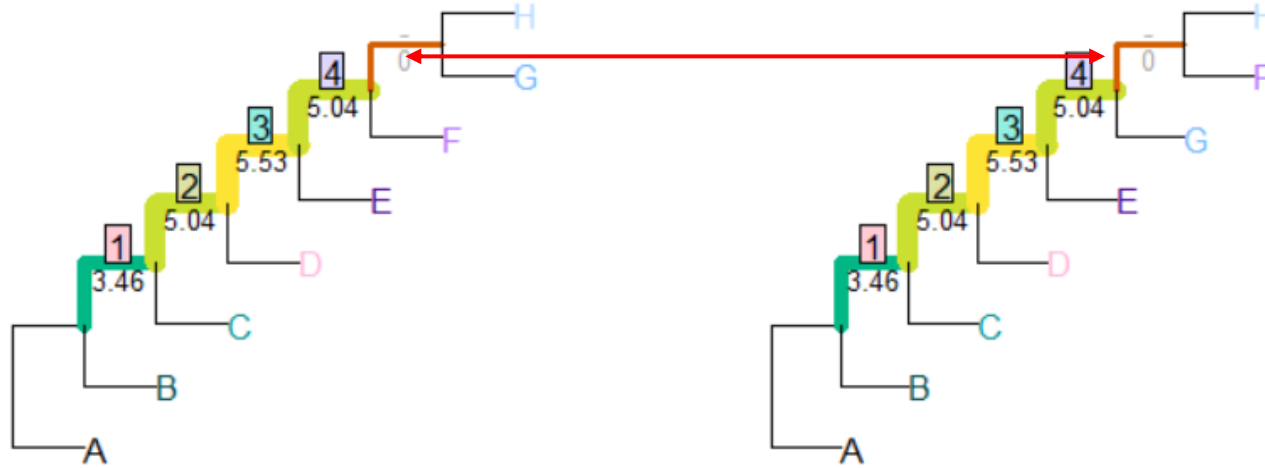
Para superar esa limitación, se introdujeron conceptos de teoría de la información, midiendo la cantidad de información que porta cada rama (en bits).

Aún así, satura muy rápido, moviendo dos hojas se puede alcanzar una similitud de 0.

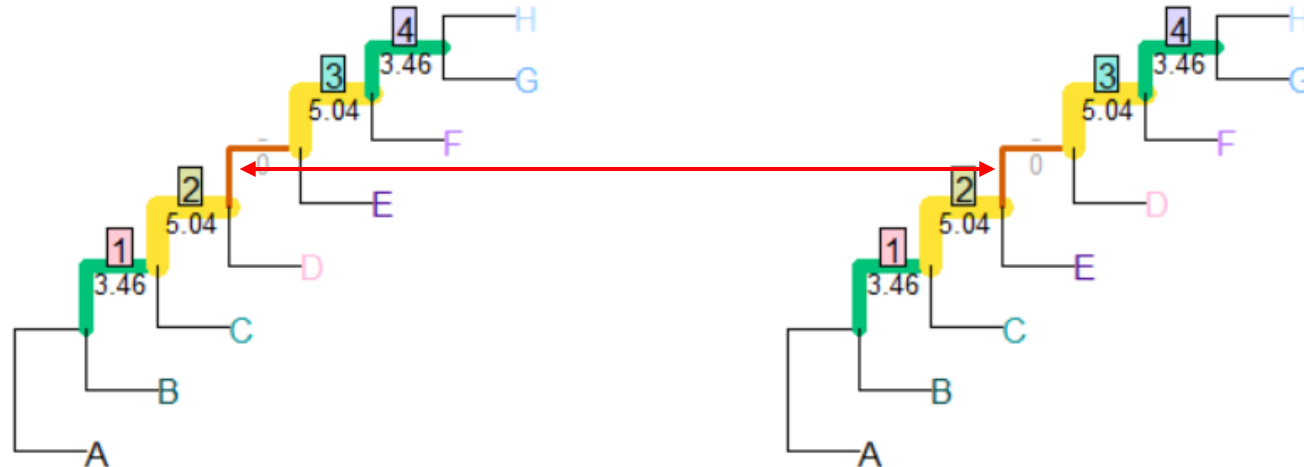


Robinson-Foulds corregida por información

RB = 2

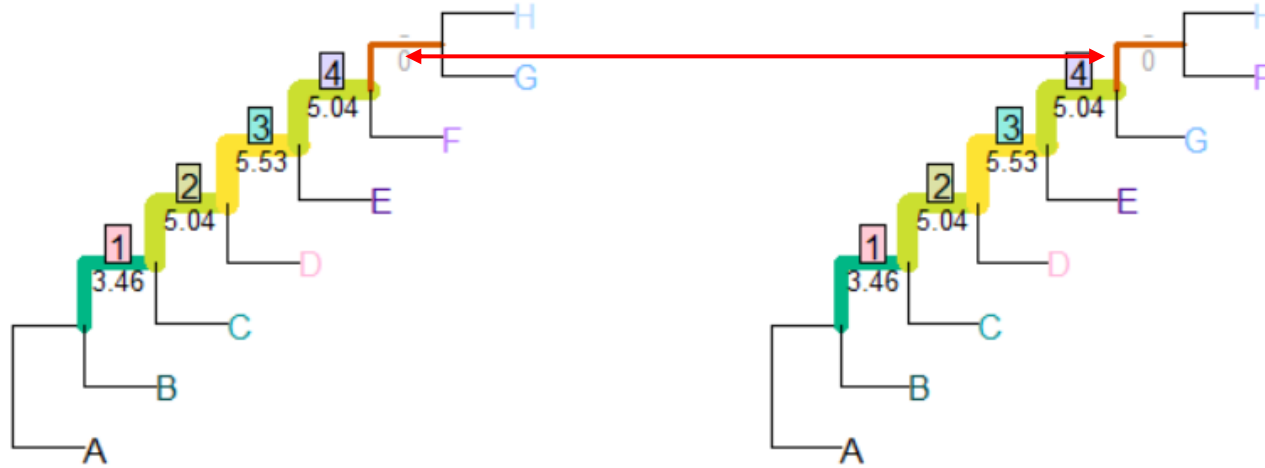


RB = 2

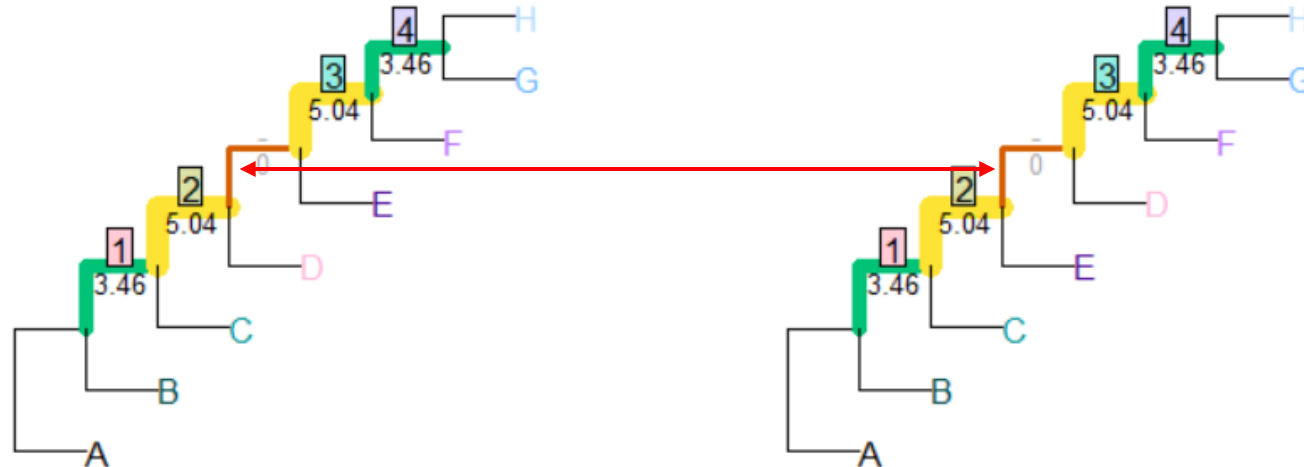


Robinson-Foulds corregida por información

RBC = 3.46



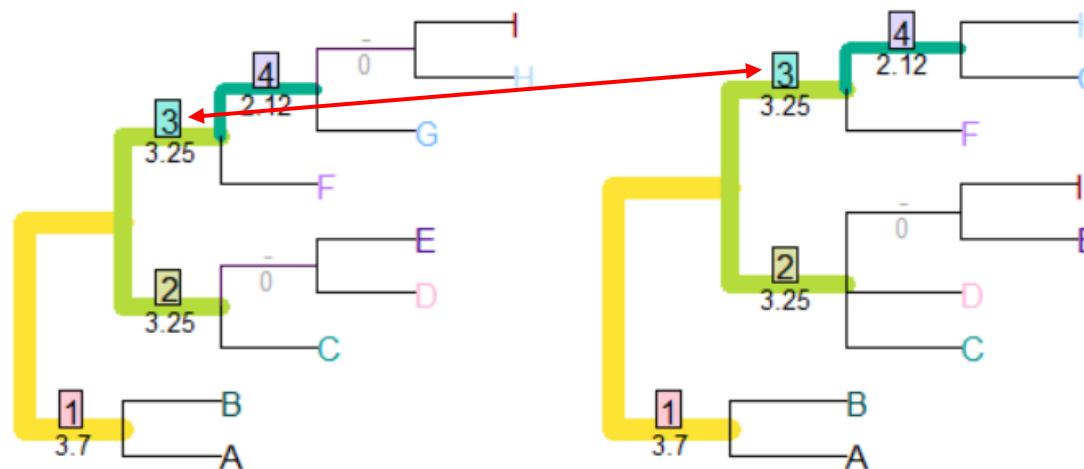
RBC = 5.53



Robinson-Foulds generalizadas. Shared Phylogenetic Information

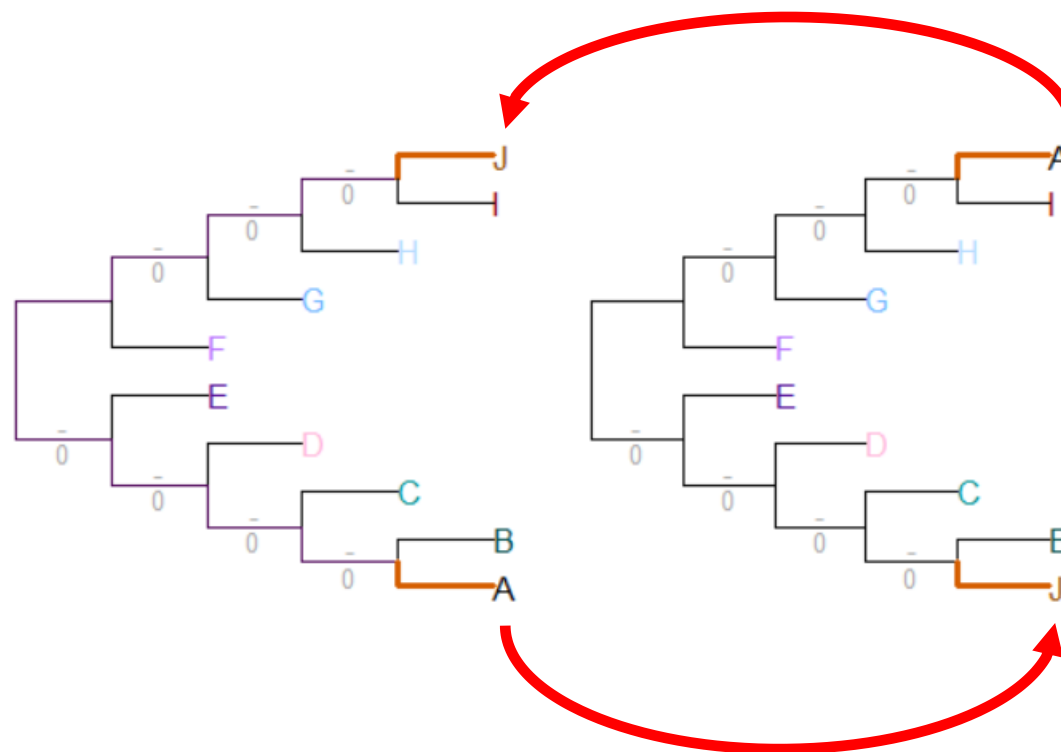
Las distancias RF generalizadas se basan en asignar similitudes distintas de 0 a pares de splits que no sean idénticos, pero que presenten cierto parecido. Para esto, hay que realizar un mapeo entre los splits de un árbol y el del otro.

Los splits que no coinciden pueden puntuarse según la cantidad de **información filogenética** que tienen en **común**, que es una función de la proporción de árboles que son coherentes con ambas divisiones.



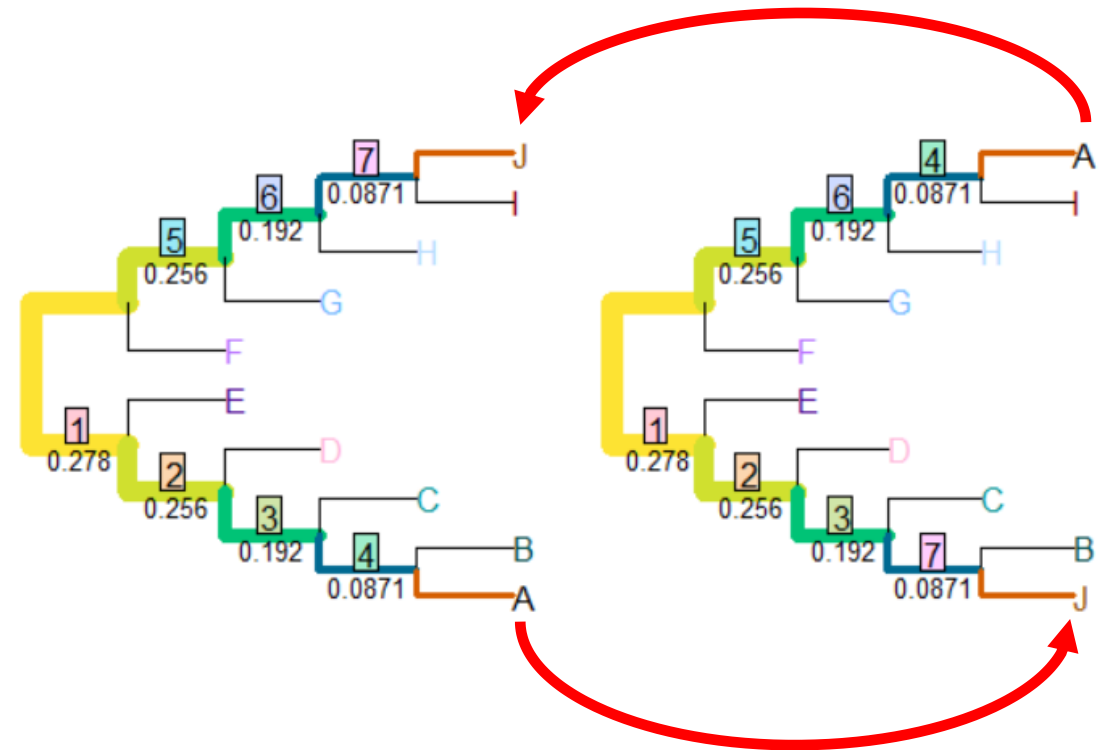
Robinson-Foulds generalizadas. Shared Phylogenetic Information

Esta métrica asigna probabilidades de 0 a los splits incompatibles, es decir, aquellos que no pueden darse en un mismo árbol. Esto provoca que la inversión de dos hojas lejanas sea suficiente para reducir a 0 el valor de esta métrica en una comparación, a pesar de la similitud del resto de los árboles.



Robinson-Foulds generalizadas. Mutual Clustering Information

Para solucionar lo anterior, se puede asignar el valor del mutual clustering information a cada par de splits para reconocer la similitud entre las estructuras de los dos árboles incluso cuando todos los posibles emparejamientos entre splits son conflictivos.

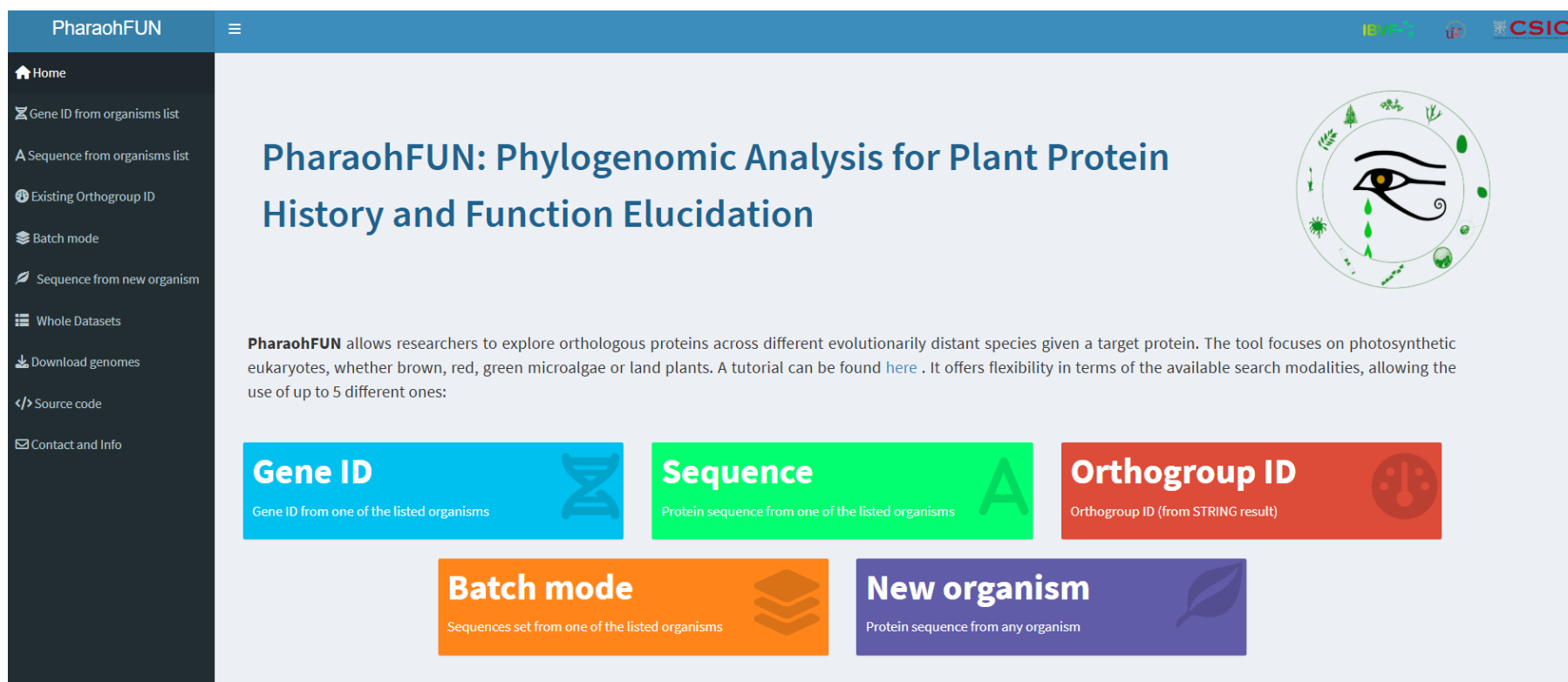


Comparación de topologías

Para más información, consultar <https://cran.r-project.org/web/packages/TreeDist/vignettes/Robinson-Foulds.html>

PharaohFUN

PharaohFUN integra información de ortología, función, interacciones físicas y reconstrucción de estados ancestrales para el análisis de las historias evolutivas de los ortogrupos, con 5 formas de búsqueda distintas. Está disponible en: <https://greennetwork.us.es/PharaohFUN/>



The screenshot displays the PharaohFUN web application interface. The top navigation bar is blue and includes the PharaohFUN logo, a hamburger menu icon, and logos for IBVF, UZ, and CSIC. A dark sidebar on the left contains a list of navigation options: Home, Gene ID from organisms list, Sequence from organisms list, Existing Orthogroup ID, Batch mode, Sequence from new organism, Whole Datasets, Download genomes, Source code, and Contact and Info. The main content area has a light blue background and features the title "PharaohFUN: Phylogenomic Analysis for Plant Protein History and Function Elucidation" in large blue text. To the right of the title is a circular logo depicting an eye surrounded by green plant icons. Below the title, a paragraph describes the tool's purpose: "PharaohFUN allows researchers to explore orthologous proteins across different evolutionarily distant species given a target protein. The tool focuses on photosynthetic eukaryotes, whether brown, red, green microalgae or land plants. A tutorial can be found [here](#). It offers flexibility in terms of the available search modalities, allowing the use of up to 5 different ones:". Below this text are five colored buttons, each with an icon and a description: "Gene ID" (blue, DNA helix icon, "Gene ID from one of the listed organisms"), "Sequence" (green, letter 'A' icon, "Protein sequence from one of the listed organisms"), "Orthogroup ID" (red, circular icon with three dots, "Orthogroup ID (from STRING result)"), "Batch mode" (orange, stacked layers icon, "Sequences set from one of the listed organisms"), and "New organism" (purple, leaf icon, "Protein sequence from any organism").