



# Análisis Filogenéticos

**Evolución Molecular**



# Contexto

- Las mutaciones se considera proceso aleatorio.
- Genes distintos tienen tasas evolutivas distintas que dependen de la estructura y función de las proteínas que codifican.
- En los primeros análisis comparativos de secuencias de proteínas se extrapolaron los principios evolutivos que se conocían del estudio de la morfología → En morfología todo cambio es objeto de selección natural:
  - **Selección positiva (o evolución adaptativa)** se refiere al proceso por el que los cambios que ofrecen cierta ventaja son favorecidos y tienden a fijarse en la población.
  - **Selección negativa (o purificadora)** se refiere a la eliminación de aquellos cambios que perjudican la adaptación, y tienden a desaparecer de la población al reducir el éxito reproductivo.



# Contexto

**Zuckermandl y Pauling compararon globinas** de distintas especies, observaron que: el nº cambios aminoacídicos entre linajes era aproximadamente proporcional al tiempo de divergencia estimado a partir del registro fósil → **Reloj Molecular**

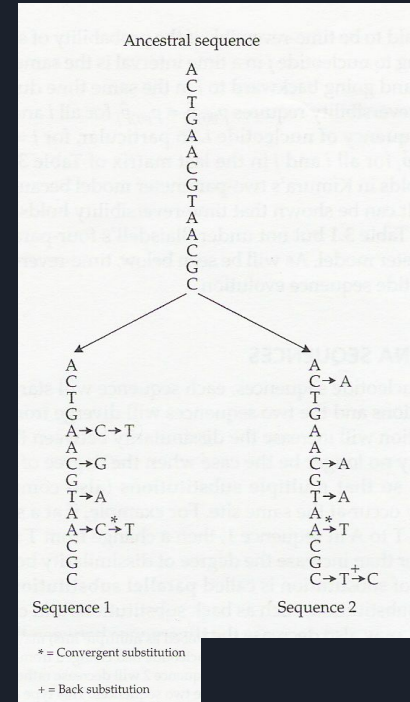
**Reloj Molecular → Asume velocidad constante del cambio evolutivo** (tasa de cambio constante, es decir la divergencia a partir de una secuencia ancestral debería ser la misma para todas las secuencias) → No se cumple para todos los genes.

**Kimura (1983) formuló la teoría neutral de la evolución molecular** → mayormente los cambios no afectan el éxito reproductor del individuo y no están sujetos a selección, estos cambios se acumulan a un ritmo aproximadamente constante a lo largo del tiempo.

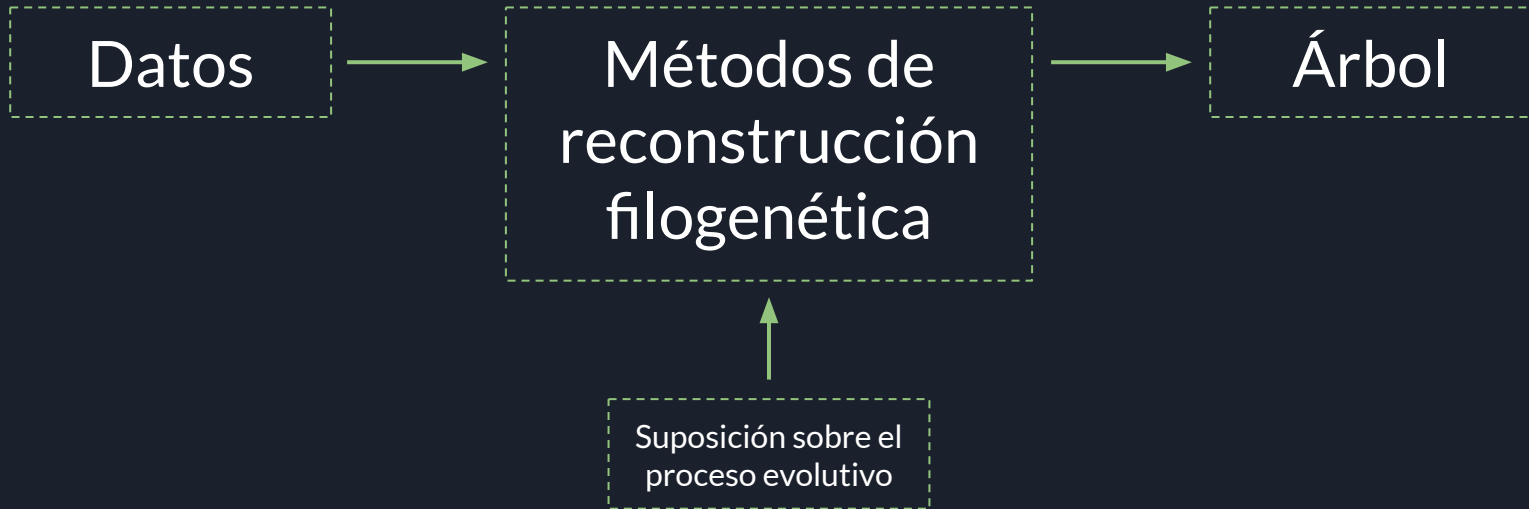
**En la actualidad se asume que la evolución molecular está gobernada tanto por procesos neutrales, como por la selección natural.**

# Reconstrucción Filogenética

- Una forma simple de **medir la divergencia entre 2 secuencias es contar el número de sustituciones en un alineamiento (distancia entre secuencias)**
- El número de sustituciones observadas pueden no reflejar los verdaderos eventos evolutivos que ocurrieron (más de un cambio intermedio al estado actual o un nucleótido idéntico observado en el alineamiento debido a mutaciones en paralelo en ambas secuencias) → **homoplasia**
- **Para aproximar mejor la historia evolutiva y las distancias entre las secuencias existen modelos estadísticos (modelos de evolución)** para inferir las verdaderas distancias evolutivas entre secuencias



# Reconstrucción Filogenética



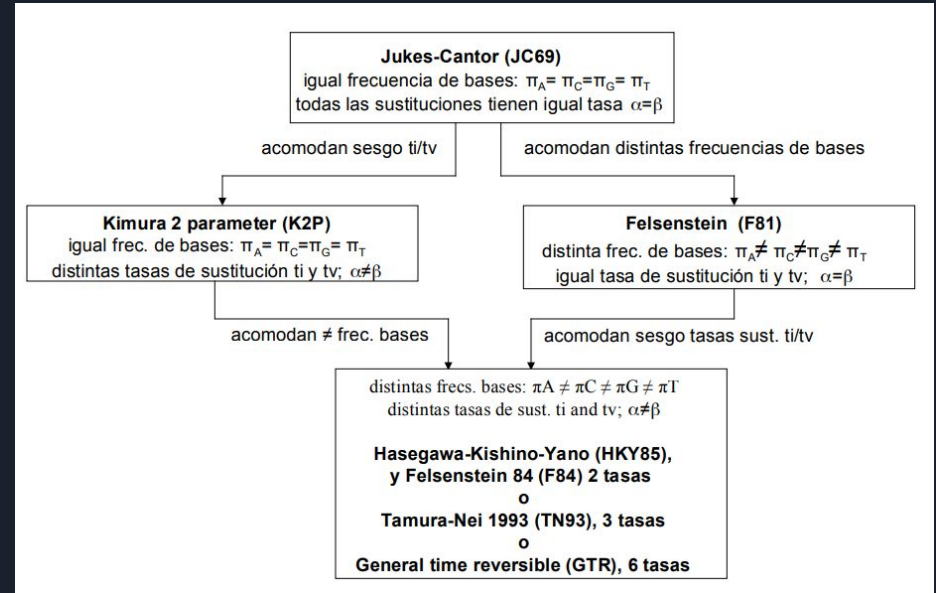
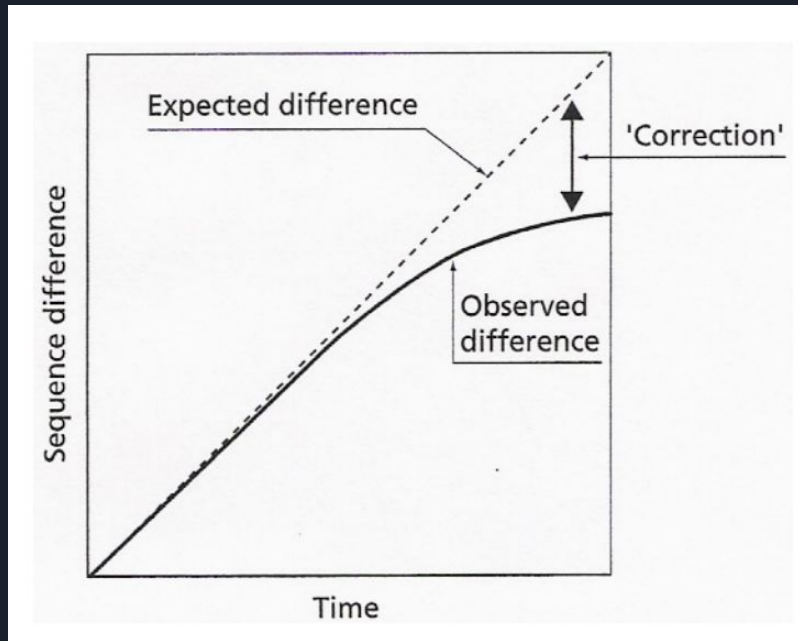
Caracteres que representen variaciones heredables



# Reconstrucción Filogenética

- Los **modelos evolutivos** son **descripciones matemáticas de la evolución** de las secuencias:
  - **Transiciones (ti)**: cambios entre **bases del mismo tipo** (mismas características químicas) → ocurren más fácilmente por razones químicas y estéticas (de espacio).
  - **Transversiones (tv)**: cambios entre **bases de distinto tipo**.
- Los **distintos modelos evolutivos pesan de manera distinta los distintos tipos de mutaciones puntuales** → Jukes-Cantor el más sencillo considera con igual probabilidad las transversiones que las transiciones.

# Contexto



Los modelos evolutivos corrigen la diferencia entre lo esperado y lo observado



# Interpretación de un árbol

- Un **árbol consta de ramas y nodos** que conectan estas ramas entre sí.
- **El largo de rama da la cantidad de cambios no sinónimos por sitio, cada cierto tiempo** (menos en UPGMA → árboles ultra métricos)
- Las **ramas terminales (hojas) corresponden a las secuencias actuales** (observadas), y **las ramas internas** representan sus **ancestros hipotéticos**. Su longitud refleja la cantidad de cambio acumulado.
- Los **nodos relacionan** estas **ramas** entre sí **según su relación ancestro-descendiente**.
- El **nodo más interno** es el **ancestro común**. La forma de **enraizar un árbol** es incluyendo **una secuencia externa al grupo de interés** que dote al árbol filogenético de sentido histórico (polaridad).





# Reconstrucción Filogenética

**Primer paso** para hacer un árbol, probablemente **el paso más crítico** del procedimiento **debido a que éste establece las correspondencias posicionales en la evolución, es construir el alineamiento** del que partimos. Sólo el alineamiento correcto produce inferencias filogenéticas correctas

Clasificación de los métodos de **reconstrucción filogenética**:

- **Los métodos de distancia** primero **convierten los alineamientos de secuencias en una matriz de distancias**, la cual es usada por el método algorítmico de reconstrucción para calcular el árbol.
- **Los métodos discretos** (MP, ML, Bayesianos) **consideran cada sitio del alineamiento** (o una función probabilística para cada sitio) directamente.

# Máxima Parsimonia: basados en caracteres

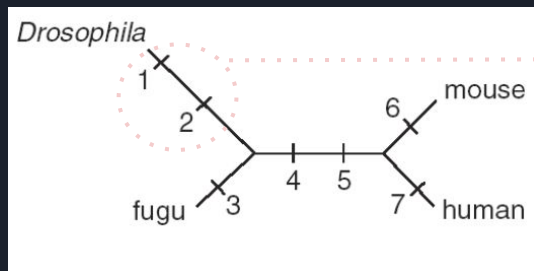
Busca la topología del que pueda ser explicada con el menor número de cambios posibles. Los datos a evaluar corresponden a cada posición o sitio en la secuencia de nucleótidos o aminoácidos.

El método cuenta el número mínimo de sustituciones que se habrían necesitado para lograr que la secuencia 2 derive de la secuencia 1 (No tiene en cuenta las mutaciones silenciosas).

Teniendo en cuenta el siguiente ejemplo veamos cómo construir un árbol:

	1	2	3	4	5	6	7
<i>Drosophila</i>	t	t	a	t	t	a	a
fugu	a	a	t	t	t	a	a
mouse	a	a	a	a	a	t	a
human	a	a	a	a	a	a	t

fugu	3
mouse	5 4
human	5 4 2
<i>Drosophila</i>	fugu mouse



Es más  
parsimonioso



# Optimización de Árboles

Una vez obtenidas las diferentes topologías, se le asigna a cada una de ellas un puntuación, y es el árbol con la topología de menor puntuación el que se elige como árbol más parsimonioso

- Algoritmos exactos: Se garantiza que se encontrará el árbol ó árboles óptimos, pero puede requerirse una gran cantidad de tiempo de cómputo
- Búsqueda exhaustiva: Se intenta encontrar soluciones óptimas, pero no hay garantía de hallarlas (en este caso el árbol óptimo). No obstante, se reduce el tiempo de análisis cuando se evalúan datos con un gran número de taxa (Búsqueda “branch and bound”, Métodos heurísticos (aproximados), Adición secuencial (Stepwise addition) , Permutación de ramas (Branch swapping))



# Máxima Verosimilitud

- El método de Máxima Verosimilitud emplea modelos probabilísticos para seleccionar el mejor árbol, es decir, aquel que tenga la más alta probabilidad (verosimilitud) de reflejar el proceso evolutivo real
- Es un método exhaustivo que busca todas las posibles topologías y considera cada posición en un alineamiento (no sólo sitios informativos)

# Bootstrap

