# Filogenia

## Evolución Molecular

Muchos Biólogos quieren hacer Filogenia

Pero la pregunta es por qué?

Porque me falta una figura en el manuscrito (Wrong answer)

Taxonomía

Genómica Comparativa

Relaciones Estructura-Función

Ortó- o parálogo

....

El "¿Por qué?" es importante porque afecta tus elecciones

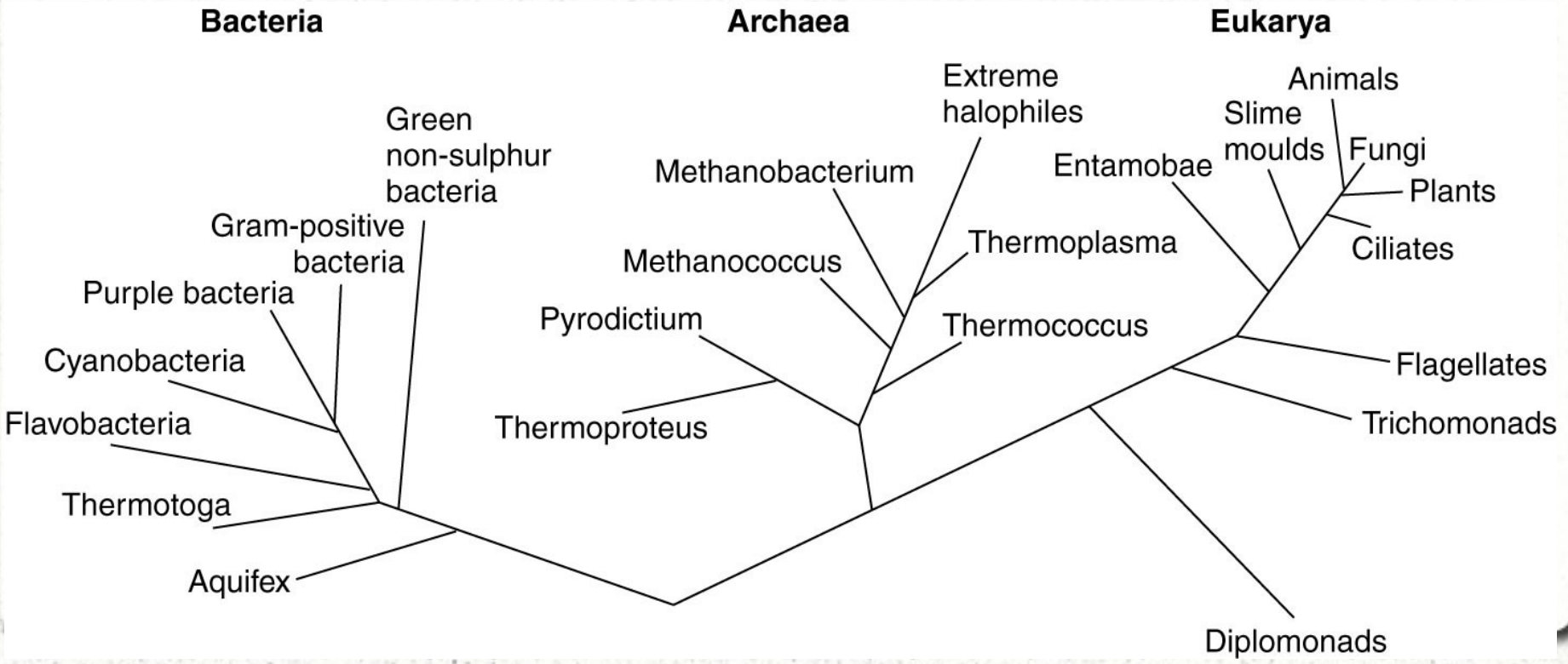
# Filogenia

## Introducción

El campo de la Filogenia tiene como blanco estudiar las relaciones entre especies, poblaciones, individuos o genes

Genealogía, oscurecida por un número de factores

Típicamente mostrado por árboles

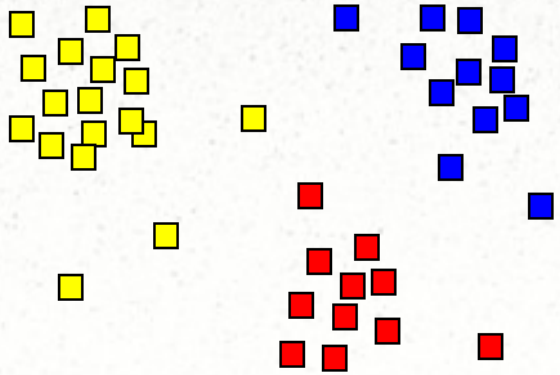


# Filogenia

## Clustering

### Clustering o agrupamiento

El análisis de agrupamiento o clustering es la tarea de asignar un conjunto de objetos en grupos (llamados clusters) de modo que los objetos en el mismo grupo son más similares (en uno u otro sentido) entre sí que a los de otros grupos.



# Filogenia

## Clustering

Filogenia es una forma de Hierarchical Clustering:

Clustering of clusters of ....

**Aprox fenética:** Por distancia vía método de "Hierarchical clustering"

**Aprox cladística:** Considerando “pathways” de evolución

La genealogía se basa en un ancestro común

Homología: Analogía por herencia

Dos cosas son homólogos o no son.....

Es un poco más complicado

# Homología no es tan fácil para entender...

Basado en porcentaje similitud

- También puede ser resultado de Evolución Convergente
- También puede ser aleatoria

Homología parcial

- Cuando sola una fracción de la secuencia viene del ancestro común
- Por ej. desde un evento de "gene fusion"

O.....Perdida de un intron que genera un "Loop" adicional

- Inserción de un sólo aminoácido???
- Mutación C por W????

Filogenia no es una ciencia exacta!

# Filogenia

## Paradoja

El MSA está basado en similitud y la filogenia en el MSA más variación

**Un MSA de buena calidad muestra altos niveles de conservación pero por otro lado la Filogenia requiere diferencias!**

Al final quieres tener un MSA que está alineado correctamente

-3D más conservado, usala información estructural

# Filogenia

## Homología Parcial

Ortólogos: Secuencias homólogas por especiación

- Generalmente tienen la misma función

Parálogos: Secuencias homólogas por duplicación

- Tendría que suponer que algunos en aspectos funcionales son diferentes
- El problema es decidir cuales son las subsecuencias homologas

Xenólogas: Vienen de HGT

- Tendría que suponer que no tienen la misma historia evolutiva

El datamining y la reconstrucción del MSA son importantes

# Filogenia

## Preparación

Datamining: Refseq o algo parecido, esplicing alternativo?

MSA: Si posible incluir información estructural
Promals3D
3D Coffee

MSA: Manual check!!

Trimming: No es importante si dos subsecuencias son homólogas o no, es importante si están alineadas correctamente, si no hay que sacar las columnas

# Block Mapping and Gathering with Entropy

Criscuolo and Gribaldo *BMC Evolutionary Biology* 2010, **10**:210
http://www.biomedcentral.com/1471-2148/10/210

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | L | C | F | K | L | R | R |
| | V | C | E | E | P | R | H |
| | I | F | T | D | Q | R | K |
| | F | F | G | R | A | R | R |
| H, Entropy | 100 | 50 | 33 | 100 | 100 | 0 | 33 |
| H Similarity Corrected | 10 | 50 | 50 | 50 | 100 | 0 | 10 |

BMGE is a good method for pre-phylogeny alignment trimming

# Filogenia

## Definiciones aburridas

**OTU:** Operational Taxonomic Unit

**Graph:** estructura abstracta hecha de "nodes" (nodos) y "edges" (conexiones)

**Path:** juego de "edges" consecutivos

**Connected graph:** "graph" con por lo menos un "path" entre "nodes"

**Tree (Árbol):** "Connected graph" con exactamente 1 "path" entre todo los "nodes" (mutuo)

**Edge length:** número asignado por "edge", significa distancia

**Path length:** Suma de "Edge length" de los "edges" del "path"

# Filogenia

## Distancia

La construcción de árboles esta hecho por disimilitud de secuencias

Las próximas reglas corresponden:

| | | |
|---|---|---|
| 1 | $d(A,B) \geq 0$ | No-negatividad |
| 2 | $d(A,B) = d(B,A)$ | Simetría |
| 3 | $d(A,C) < d(A,B) + d(B,C)$ | Desigualdad de triangulo |
| 4 | $d(A,B) = 0$ solo cuando $A = B$ | Distinción |

3: Debe ser posible formar triángulo

$d(A,C) = d(A,B) + d(B,C)$: Es una línea

$d(A,C) > d(A,B) + d(B,C)$: Geométricamente imposible

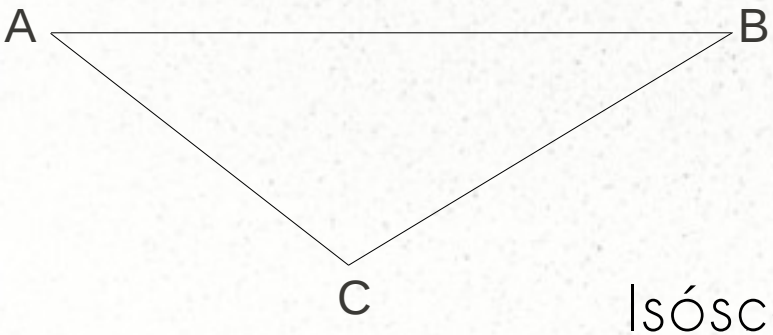
# Filogenia

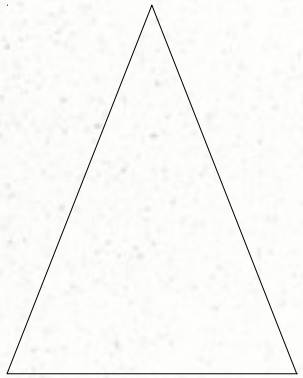
## Distancia

Para árboles ultamétricos:

5 d(A,B) ≤ Max [d(A,B), d(B,C)]

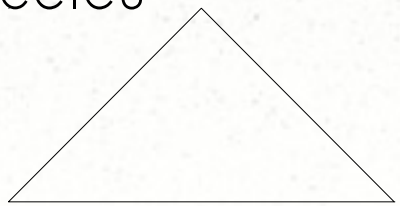
Triangulo Isósceles con las líneas de igual tamaño ≥ la línea pequeña



Isósceles y ultramétrico

Isósceles



# Phenetic, Hierarchical clustering

## Unweighthed Pair Group Method with Arithmic Mean (UPGMA)

Matriz de distancia:

| | A | B | C | D |
|---|---|---|---|---|
| A | 0 | 2 | 6 | 10 |
| B | | 0 | 6 | 10 |
| C | | | 0 | 10 |
| D | | | | 0 |

Matriz reducida:

| | AB | C | D |
|---|---|---|---|
| AB | 0 | 5 | 9 |
| C | | 0 | 10 |

Matriz rereducida:

| | ABC | D |
|---|---|---|
| ABC | 0 | 7 |

-Empezar con las dos más cercanas

A y B en una rama con ancestro virtual (AB)

-Calcular Matriz reducida con nodo AB

d(AB,C) = d(A,C)-d(AB-A) = 6-1 = 5

-Determinar las dos más cercanas (5 ± 1)/2

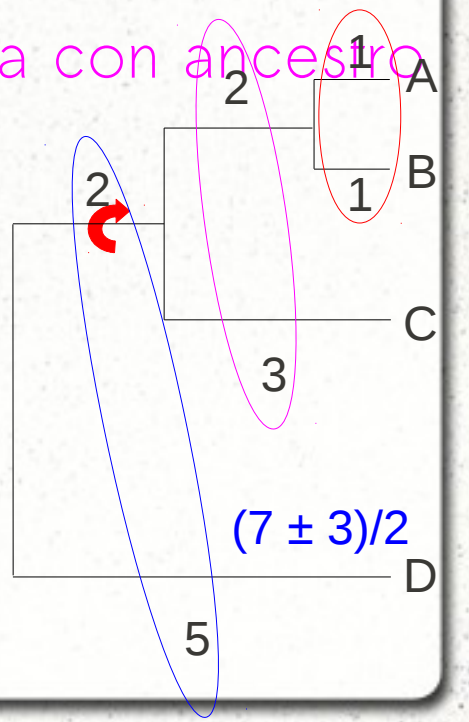
AB y C quedan en una rama con ancestro virtual (ABC)

Et cetera

d(ABC,D) = d(A,D)-d(ABC-A) =

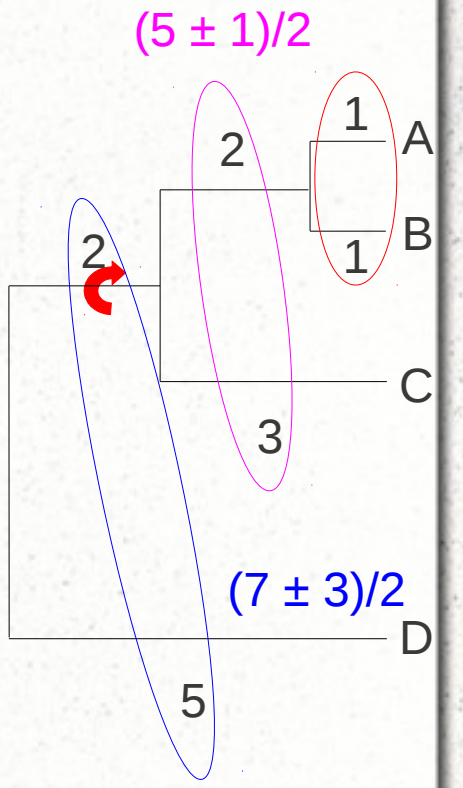
10-3 = 7

Ramas se pueden girar



# Phenetic, Hierarchical clustering

## Unweighthed Pair Group Method with Arithmic Mean (UPGMA)




Ultramétrico:

d(ABC,D) = d(A,D)-d(ABC-A) = 10-3 = 7

es igual a

d(ABC,D) = d(B,D)-d(ABC-B) = 10-3 = 7

Solo cuando la velocidad de evolución es igual

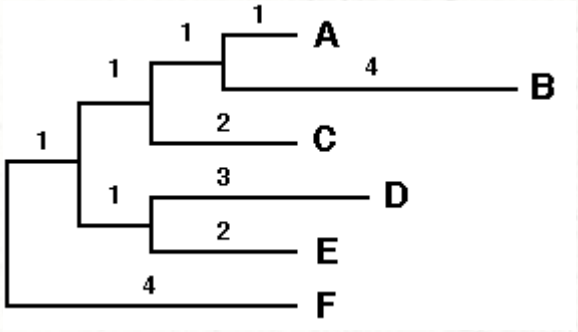
Ojo: El algoritmo no sabe!

# UPGMA

## Problemas

Muy sensible por velocidades de evolución desiguales

UPGMA depende de datos ultramétricos




distBD ≤ max (distBA,distAD):10 ≤ max (5,7): False!

| | A | B | C | D | E |
|---|---|---|---|---|---|
| B | 5 | | | | |
| C | **4** | 7 | | | |
| D | 7 | 10 | 7 | | |
| E | 6 | 9 | 6 | 5 | |
| F | 8 | 11 | 8 | 9 | 8 |

4 es la distancia mas pequeña entonces si usamos UPGMA A y C quedan en una rama etc

# Neighbor Joining: Considerando distancia por rama

| | A | B | C | D | E |
|---|---|---|---|---|---|
| B | 5 | | | | |
| C | **4** | 7 | | | |
| D | 7 | 10 | 7 | | |
| E | 6 | 9 | 6 | 5 | |
| F | 8 | 11 | 8 | 9 | 8 |

r(A) = 5+4+7+6+8=30
r(B) = 42
r(C) = 32
r(D) = 38
r(E) = 34
r(F) = 44

Paso 1: Calculamos la divergencia neta por cada OTU

Paso 2: Calculamos Matriz nueva: d'(ij)=d(ij) - [r(i) + r(j)]/(N-2)

Por ej: d'(AB)=d(AB) -[(r(A) + r(B)]/(N-2) = 5-[30 + 42]/4 = 3

| | A | B | C | D | E |
|---|---|---|---|---|---|
| B | **-13** | | | | |
| C | -11,5 | -11,5 | | | |
| D | -10 | -10 | -10,5 | | |
| E | -10 | -10 | -10,5 | **-13** | |
| F | -10,5 | -10,5 | -11 | -11,5 | -11,5 |

# Phenetic, Hierarchical clustering

## Neighbor Joining

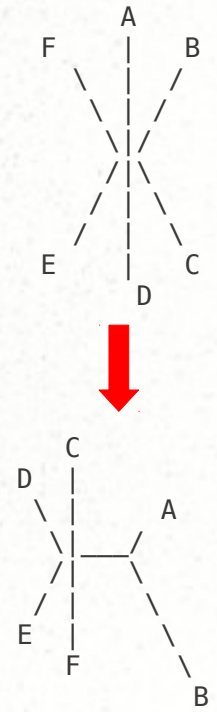
Paso 3: Elegir d'(ij) mas pequeña: -13: Node U con A y B (E y D)

Determinar matriz nueva con U

S(AU) = d(AB) / 2 + [r(A)-r(B)] / 2(N-2)

= 5/2 + [30 - 42]/8= 1

| | U | C | D | E |
|---|---|---|---|---|
| C | 3 | | | |
| D | 6 | 7 | | |
| E | 5 | 6 | 5 | |
| F | 7 | 8 | 9 | 8 |




Repite procedimiento con N=N-1 hasta N = 0

# Phenetic, Hierarchical clustering

## Neighbor Joining, Transformed distance method

### Transformed distance method

Introducir outgroup F: emergencia desde el encestador común antes tus secuencias

dist'AB = [(distAB - distAF - distBF)/2] + dist(av)F  Distancia transformadaVentajas

Rápidas:  Usar si tenes grandes cantidades de seqs

Calcula distancia

¿Por qué no es el método mejor?

Desventaja: Reducción de información de secuencias

Usas solamente la distancia

Distancia anda bien a pares pero no en MSAs!!!

# Cladistic clustering

## Maximum Parsimony

Parsimonia: El principio es que uno no debe multiplicar entidades innecesariamente, o hacer más asunciones que las necesarias, y en general que uno debe perseguir la hipótesis más simple

Sobre-simplificación

Árbol de parsimonia máxima es UN árbol basado en la cantidad de mutaciones menores

Distancia total mínima: puede existir más que uno

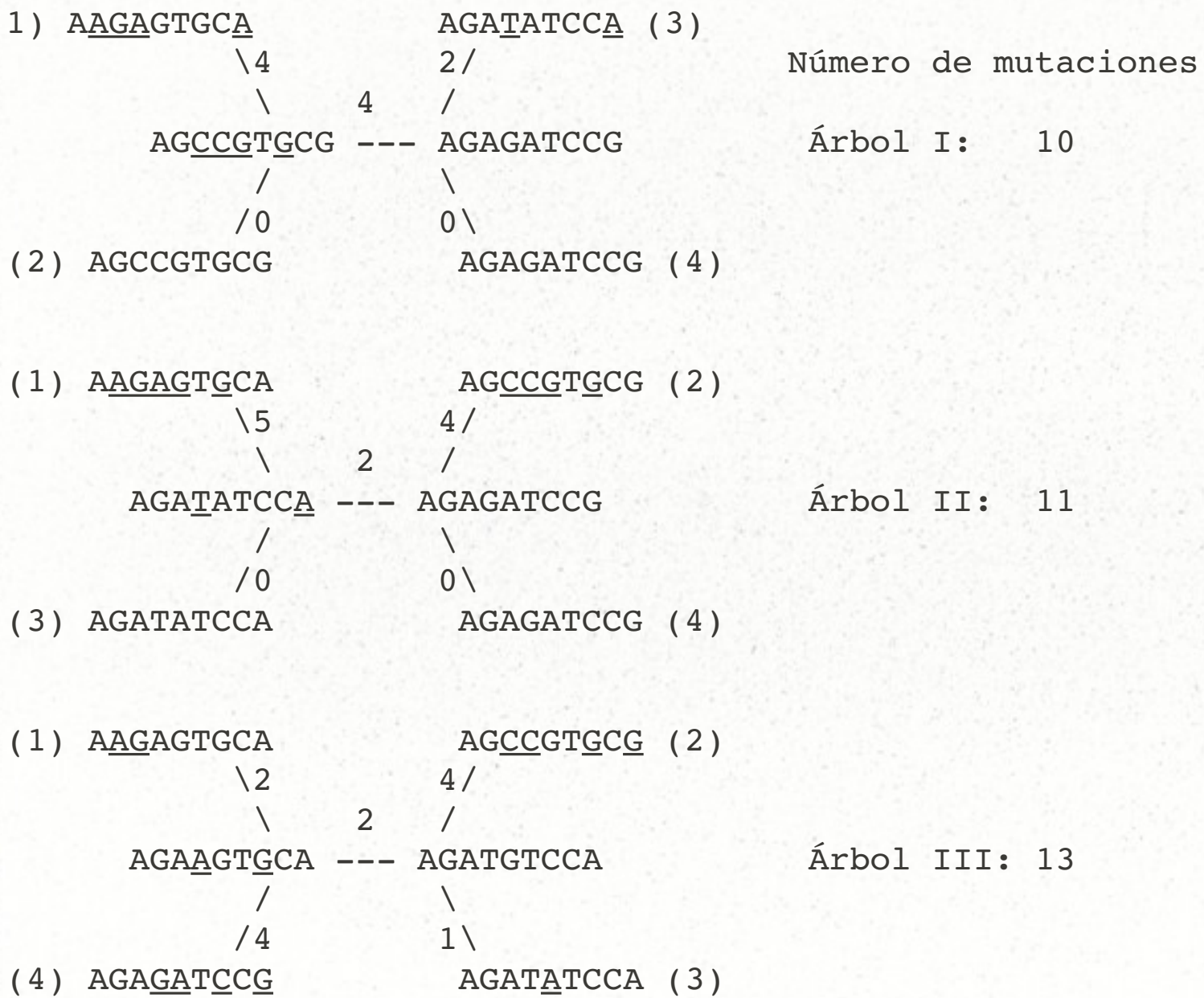
Desventaja: Busca todos los árboles posibles

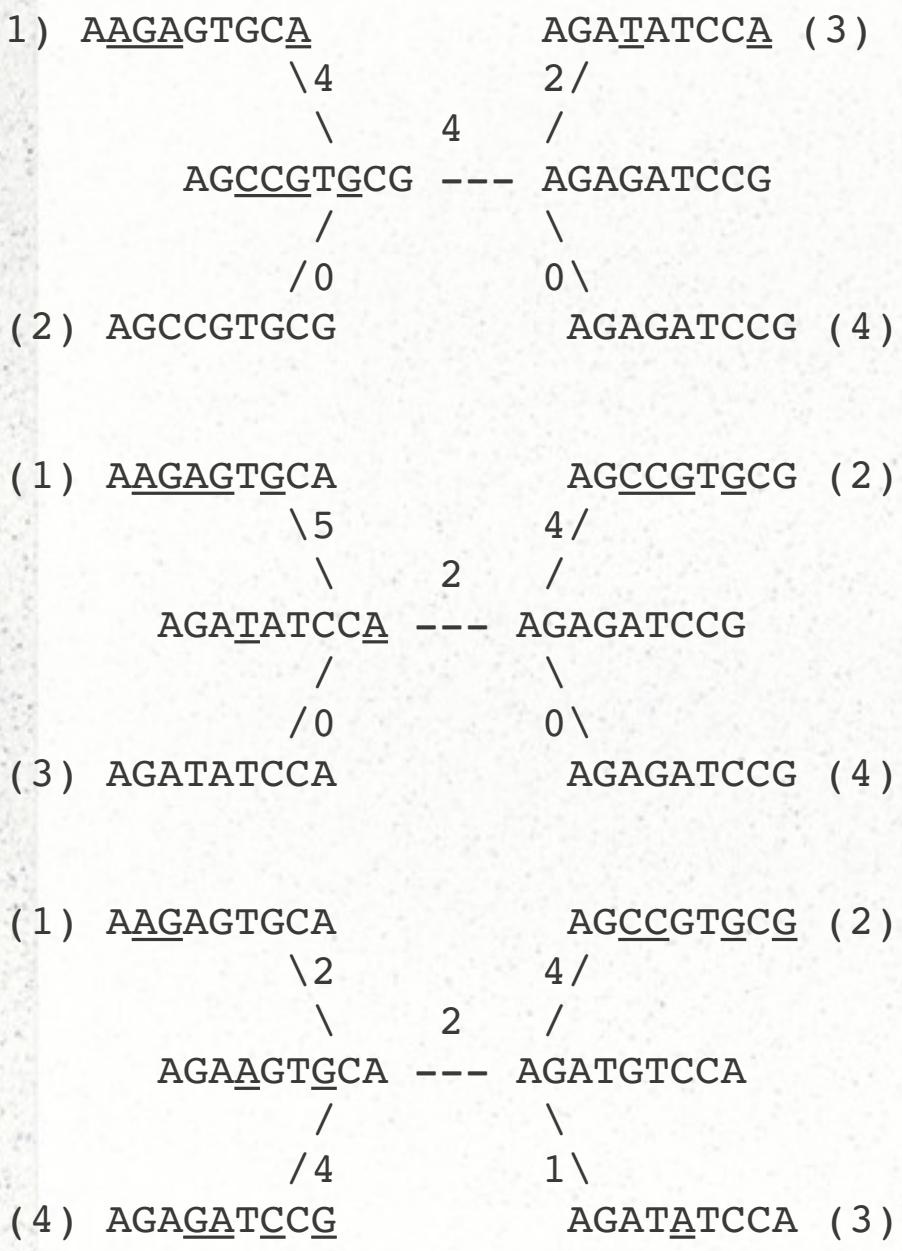
# Maximum Parsimony

| | Sitio | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Seq | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 1 | A | A | G | A | G | T | G | C | A |
| 2 | A | G | C | C | G | T | G | C | G |
| 3 | A | G | A | T | A | T | C | C | A |
| 4 | A | G | A | G | A | T | C | C | G |

| Número de OTUs | Número de árboles sin raiz | Número de árboles con raiz |
|---|---|---|
| 2 | 1 | 1 |
| 3 | 1 | 3 |
| 4 | 3 | 15 |
| 5 | 15 | 105 |
| 6 | 105 | 945 |
| 7 | 954 | 10,395 |
| 8 | 10,395 | 135,135 |
| 9 | 135,135 | 34,459,425 |
| 10 | 34,459,425 | 2.13E15 |
| 15 | 2.13E15 | 8.E21 |

# Maximum Parsimony






Sitio 1 no muestra mutaciones

2: 1, 1 y 1 mutación: Sitio NO informativo

3: 2, 2, y 2: NO informativo

4: 3, 3 y 3: NO informativo

5: 1, 2 y 2: Informativo

6: no muestra mutaciones

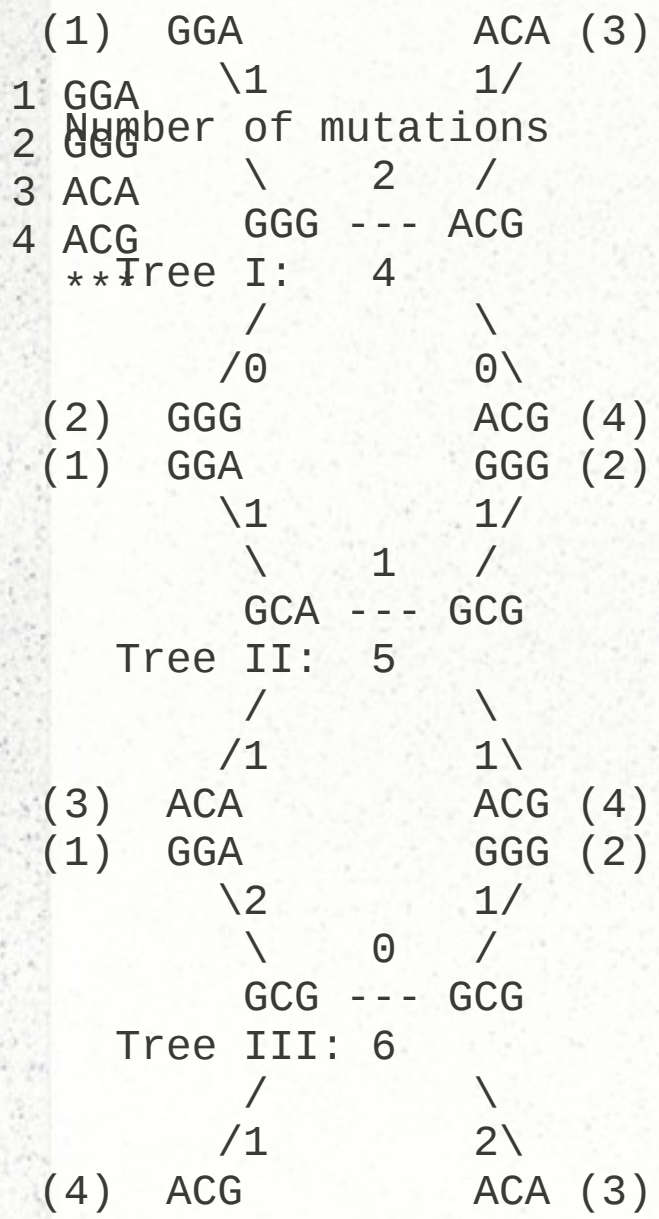
7: 1, 2, y 3: Informativo

8: no muestra mutaciones

9: 2, 1, y 2: Informativo

# Maximum Parsimony

## Sítios informativos




1 GGA
2 GGG
3 ACA
4 ACG
**

Mejoramos n

y N?

Igual tenemos que calcular todos los árboles

Todos?

Branch and Bound

# Metodos para optimizar la velocidad

## Branch and Bound

Usa la idea de “Optimistic Score”

Por ej un motivo de ADN

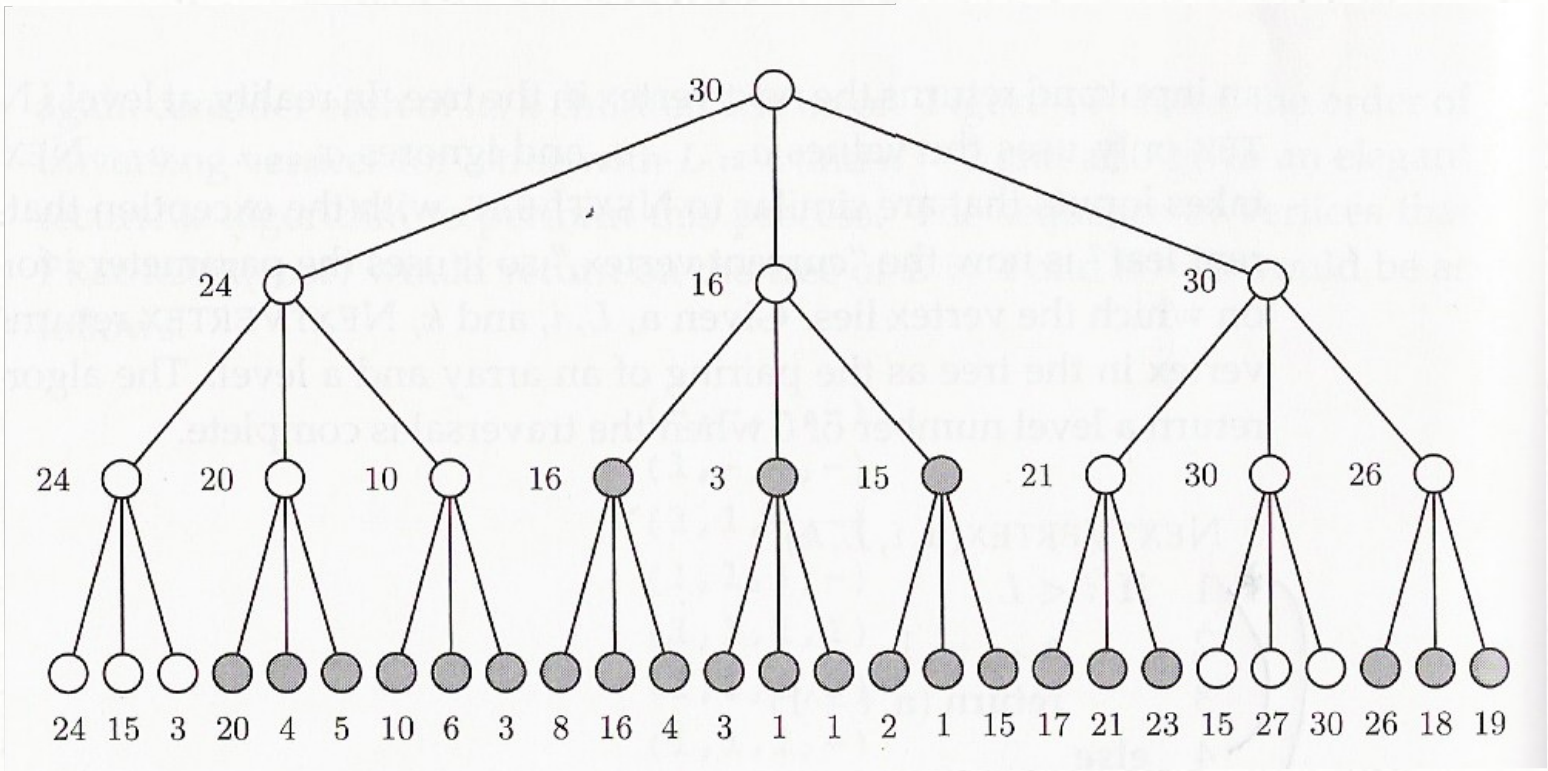
Sabes que la última posición no puede contribuir más que un número máximo

Desde la penúltima también

Se llama Branch and Bound porque estas dibujando un árbol, calculas el “Maximum bound”

# Metodos para optimizar la velocidad

## Branch and Bound




Cada número indica el "Optimistic Score" desde la posición correspondiente

# Maximum Parsimony

## Heuristic search

La aplicación de sitios informativos y Branch and Bound no es suficiente para tener un algoritmo con velocidad aceptable

Heuristic search: Enfoque a velocidad, menos a calidad/exactitud

Empezamos con dos o tres secuencias, calculamos

Agregar secuencia por secuencia y recalcular todo el árbol:

**Significa que el orden de las secuencias tiene un impacto**

# Maximum Parsimony

## Heuristic search

1 JUMBLE: Mezclar el orden DE LAS SECUENCIAS (POR REGLON)

Multiples jumbles: Sí requiere mas CPU pero igual combinado con el algoritmo heuristico es mucho mas rápido

2 Global Rearrangements

Subárboles: reorganizar los sub-arboles hasta que se encuentra el mejor árbol

Se incluye información de substitución en el MP!

# Maximum Parsimony

## Ojo!

Ojo: Los sítios no-informativos no tienen información por la forma de búsqueda!

Maximum parsimony no usa todo la información de las secuencias!!

El Branch and Bound NO es un truco heuriístico

Es posible tener por ej. cinco árboles MP: Hay que calcular un consenso

# Información y el paradigma

No será mejor usar un MSA de ADN?

Puede ser que es mejor usar un “codon-alignment”
La calidad de un MSA de secuencias proteícas >>> MSA and
La cantidad de información MSA and >>> MSA proteica

MSA de proteinas → Codon Alignment

T-Coffee tiene protogene

Mis experiencias son más o menos

Ojo!! Y los constraints al ambito de ADN?

# Models of nt Sequence evolution

**Jukes and Cantor**

| | A | C | G | T |
|---|---|---|---|---|
| A | - | a | a | a |
| C | a | - | a | a |
| G | a | a | - | a |
| T | a | a | a | - |

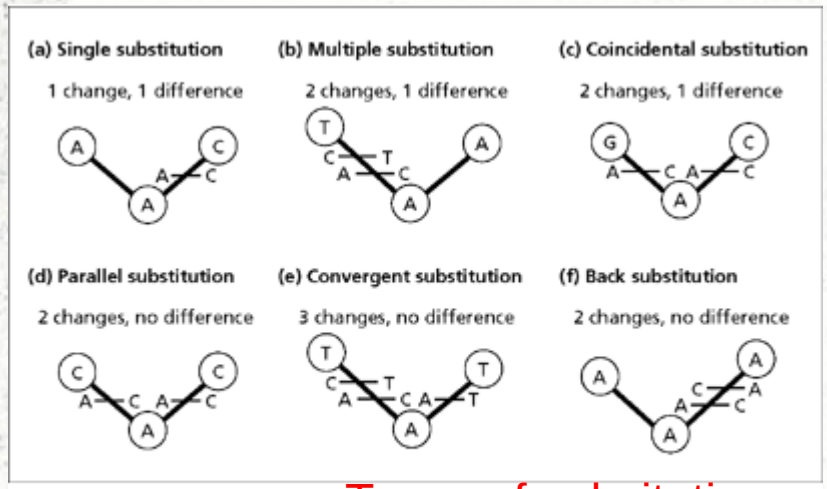
$$P_t = \begin{bmatrix} . & \alpha & \alpha & \alpha \\ \alpha & . & \alpha & \alpha \\ \alpha & \alpha & . & \alpha \\ \alpha & \alpha & \alpha & . \end{bmatrix} \qquad f = [¼, ¼, ¼, ¼]$$

Matriz de substitución

Vector del estado inicial

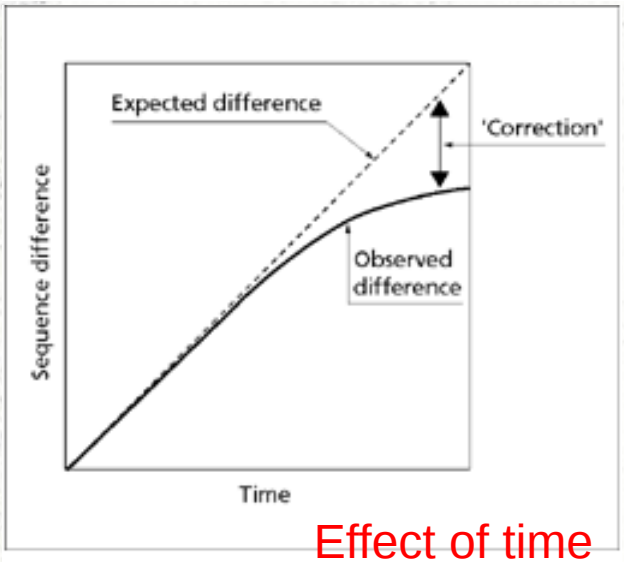
# Models of nt Sequence evolution

## Transition bias




Types of subsitution

Page, Holmes
**Molecular Evolution**



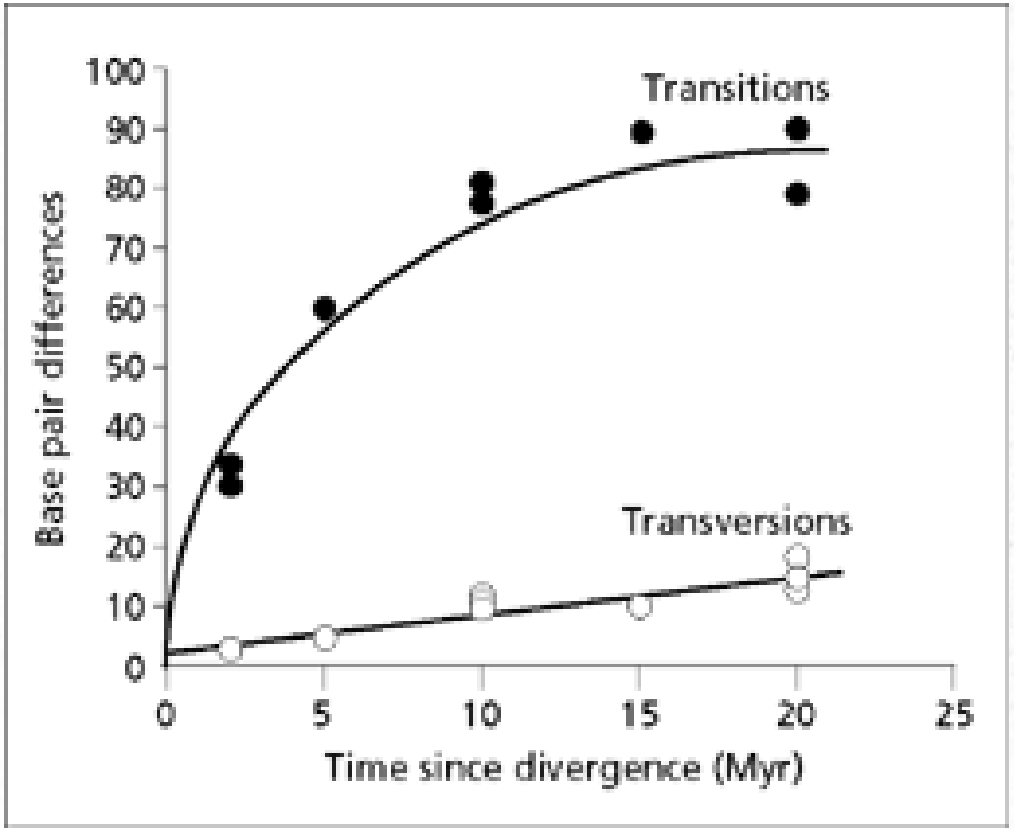

Effect of time

Page, Holmes
**Molecular Evolution**

**Kimura's 2 parameter**

| | A | C | G | T |
|---|---|---|---|---|
| A | - | b | a | b |
| C | b | - | b | a |
| G | a | b | - | b |
| T | b | a | b | - |

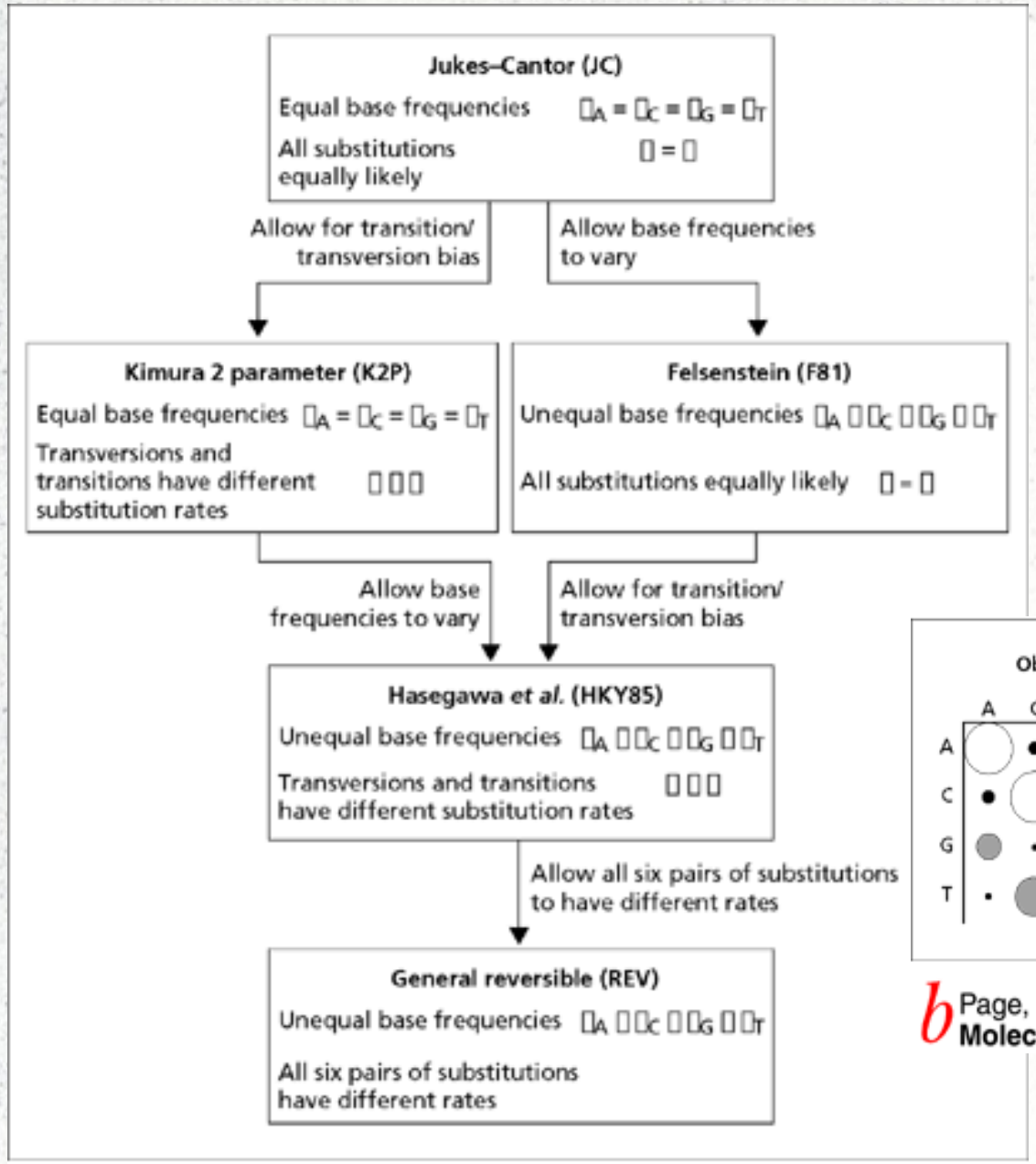
$$P_t = \begin{bmatrix} . & \beta & \alpha & \beta \\ \beta & . & \beta & \alpha \\ \alpha & \beta & . & \beta \\ \boldsymbol{\alpha} & \alpha & \beta & . \end{bmatrix} \quad f = [¼, ¼, ¼, ¼]$$



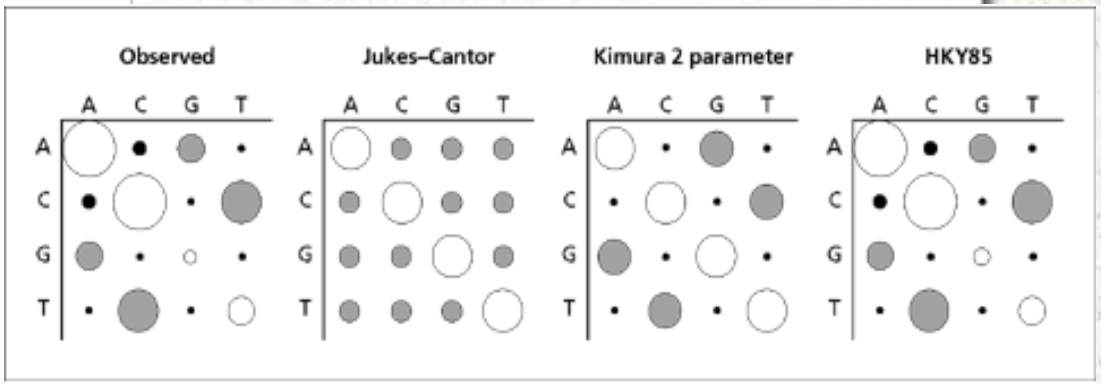

b Page, Holmes
**Molecular Evolution**

¿Que significa para estudios de Evolución Molecular?

Palabra Clave: Gene family




b Page, Holmes
Molecular Evolution

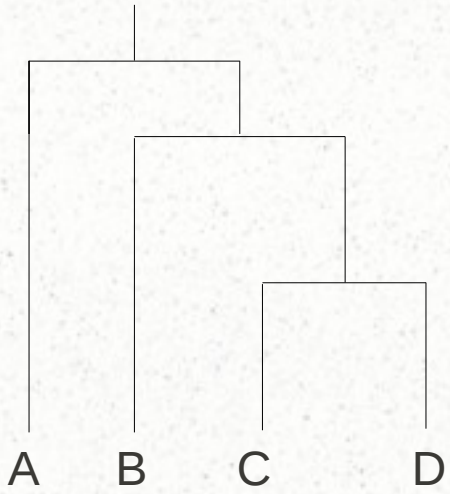



b Page, Holmes
Molecular Evolution

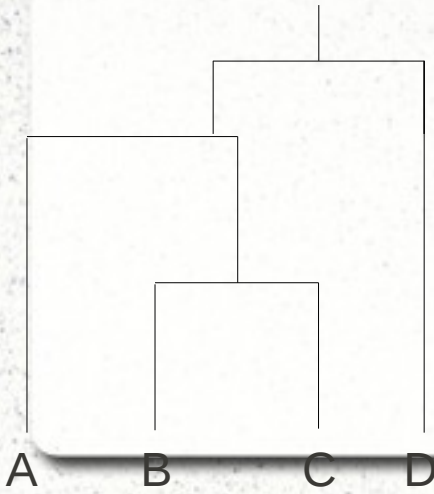
# Outgroup

## El raíz de un árbol F es estructural, no nutricional

No es siempre necesario tener una raíz, La raíz de un árbol filogenético es comparable con un raíz estructural (sin el rol nutricional)




| | A | B | C | D |
|---|---|---|---|---|
| A | 0 | 3 | 3 | 3 |
| B | | 0 | 2 | 2 |
| C | | | 0 | 1 |
| D | | | | 0 |




Después la especiación C y D, D muta mas rápido

| | A | B | C | D |
|---|---|---|---|---|
| A | 0 | 3 | 3 | 20 |
| B | | 0 | 2 | 20 |
| C | | | 0 | 20 |
| D | | | | 0 |

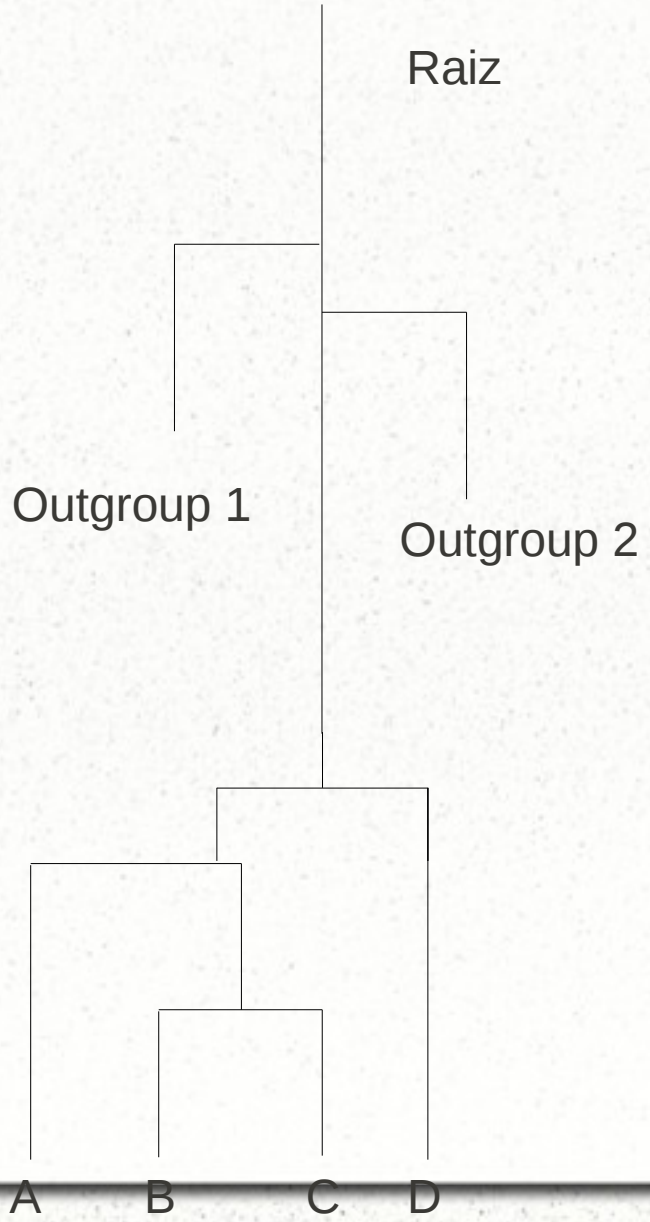


Si una filogenia (árbol) esta "fuera" de la balanza por

# Outgroup

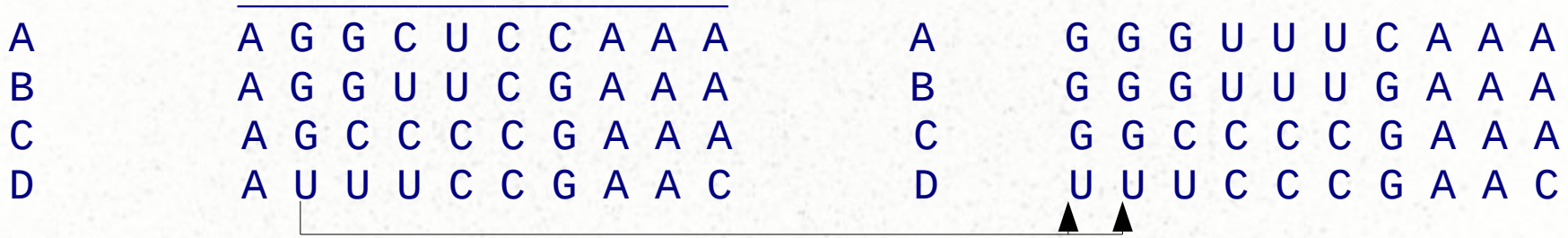
El raíz de un árbol F es estructural, no nutricional

## Outgroup y Raiz



# Pruebas

## Bootstrap y Jackknife




El procedimiento de un bootstrap construye más alineamiento

Algunas columnas o sitios se duplican y reemplazan a otros

La cantidad de sitios sera identico

Cuantos? 1 400 aa: 400 bootstraps

2 Más pragmatico: Empirico

El Jackknife es similar pero con secuencias contiguas: dirigida a proteínas multi-dominio

## Caracter Heurística

## Pruebas

1. Otra filogenia (Consensos!)
2. Filogenia por sub-juego: Debe ser igual
3. Pruebas estadísticas
   - Jackknifing
   - Bootstrapping
4. Edge largas: Incluir outgroup
   - Por qué no probar 2 outgroups diferentes?

Jumbles y bootstrap: Cuántos?

Empírico

## Caracter Heurística

## Pruebas

Empírico

1°: Árbol sin incluir Jumbles

2°: Árbol con N/10 Jumbles

3°: Árbol con N Jumbles

4°: Comparar y trabajar con la cantidad de Jumbles menor (X)

5°: Árbol con X Jumbles pero con otra semilla para iniciar el proceso de los jumbles

La semilla es un número para empezar el "Random generator"

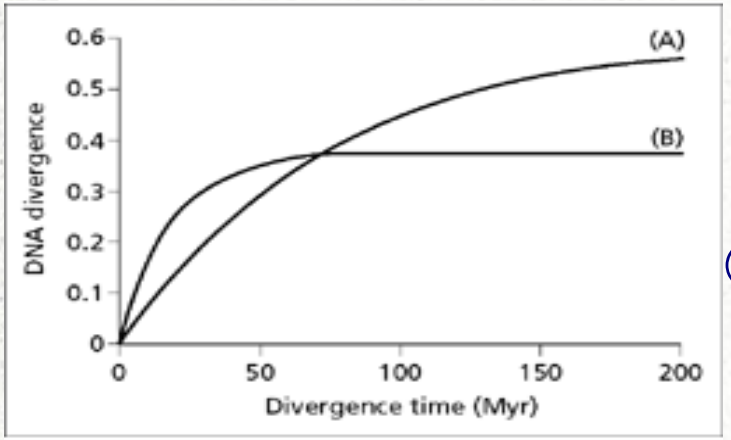
6°: Comparar para confirmar la cantidad de Jumbles menor (X)

$7^0$: Repetir con booststraps

# Variación en "substitution rate"

## Distribución Gamma

Asumimos que todos los sitios están evolucionando a la misma velocidad.




Page, Holmes
**Molecular Evolution**

0.5%/MYr, 80% sitios

2%/MYr, 50% sitios

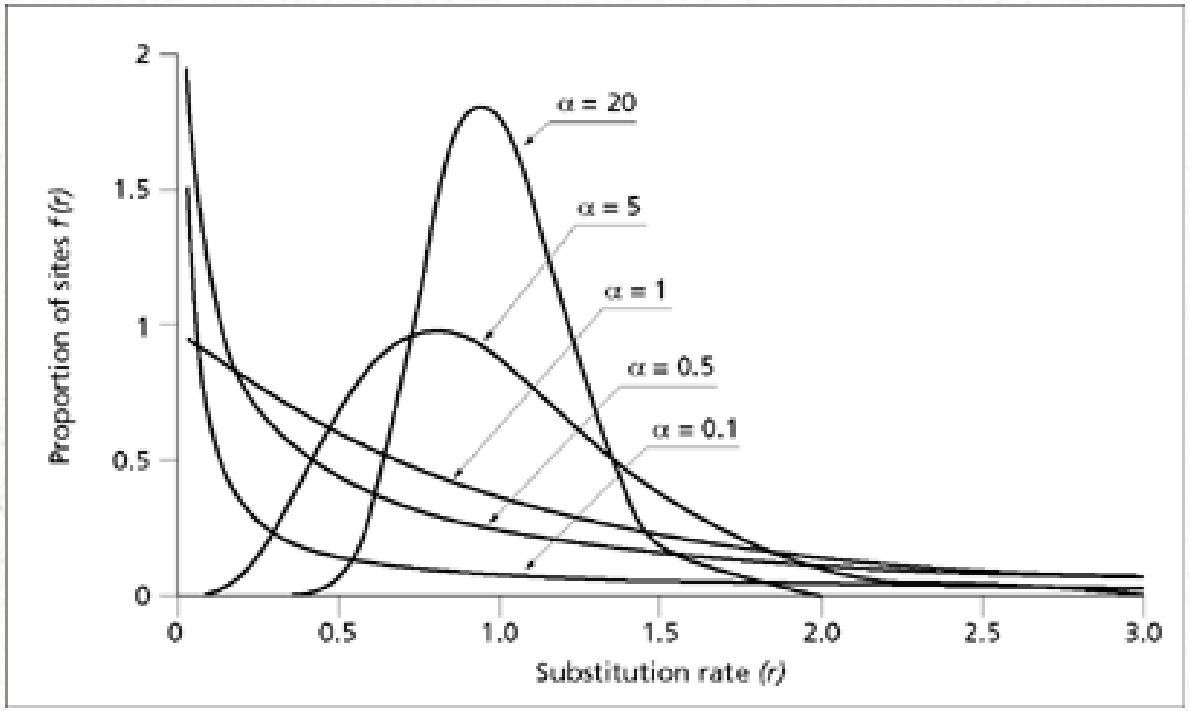
Claramente NO una hipótesis realista.

Variación en la velocidad de subsititución afecta la divergencia y de aquí la filogenia

La heterogeneidad puede ser dibujada mediante la distribución gamma.

# Variación en "substitution rate"

## Distribución Gamma




Page, Holmes
**Molecular Evolution**

La forma de la distribución gamma se define como un parámetro llamado alfa (α).

Si conocemos α podemos coregir el modelo de evolución

## Software

## Varios....

PHYLIP, Joe Felsenstein

PAUP, 100 us$

MEGA:  No tengo la idea de los algoritmos

PHYML es otro software para calcular Maximum Likelihood

- -Empieza con NJ tree
- -Computa verosimilitud el árbol
- -Computa verosimilitud de sub-árbolel (ramas)
- -Intercambiar ramas, calcular verosimilitud, buscar óptimo

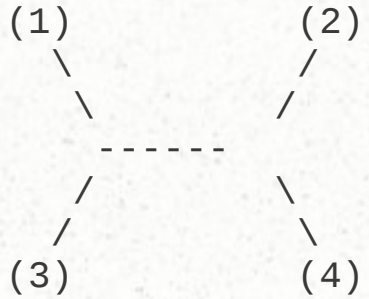
Truco Heurístico: Calculas ML de sub-juegos (aunque varias veces)

Paper

# Maximum Likelihood

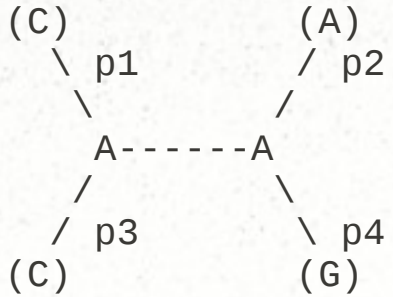
## Árbol más probable basado en modelo evolucionario

```
          1          j          ....N
(1)       A G G C U  C C A A ....A
(2)       A G G U U  C G A A ....A
(3)       A G C C C  A G A A.... A
(4)       A U U U C  G G A A.... C
```




Sitio J: Cual es la probabilidad que el árbol corresponde a la evolución del sitio?

Independiente del raíz: calculamos todos los posibles sitios ancestrales

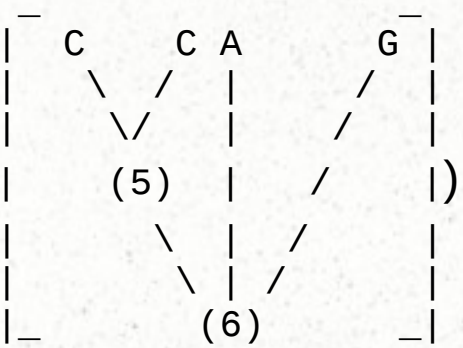



Por ej. A como sitio ancestral:

$P_{J(A\_A)} = p1 * p2 * p3 * p4$

# Maximum Likelihood

## Árbol más probable basado en modelo evolucionario

$P_{j(X-Y)} = \sum P$ 

$)$

con a X (5) y Y (6) los cuatro nucleótidos posibles

$$P_{Tree} = P_j(j=1 \rightarrow N) = P_1 * P_2 * P_j * \ldots\ldots P_N *$$

# Maximum Likelihood

## Verosimilitud máxima

DNAML (PHYLIP): 4*4

PROTML (PHYLIP) : 20 * 20!!

Con DNAML hay que calcular 16 posibilidades por posicion (columna!).

Con PROTML hay que calcular 400 posibilidades pero por 3 veces sitios menos

¿Cual es mejor MP o ML?

¿Cual requiere mas CPU, MP o ML?

# Maximum Likelihood o Parismony

En MP suponemos la minima de mutación

Es como decir: No hay “backmutations” INCORRECTO

En ML suponemos que el ancestro tenia como las cuatro (o veinte) posible residuos

Es como decir: “Aunque tengo en 49 de reglones una W y en 1 Y, igual el ancestro podría tener Q” IMPROBABLE

Igual, porque usamos modelos de P empíricos la IMPROBABILIDAD es pequeña ($\neq 0$)

# Maximum Likelihood o Parismony

Situación 1: Tenemos 50 secuencias de una proteina órtologa, desde 50 cepas de un hongo: MP o ML?

MP porque la probabilidad de tener backmutations es pequeña, la probabilidad que un 49 W y un Y deriven de una Q es bastante alto

# Maximum Likelihood

## Sí, pero de qué

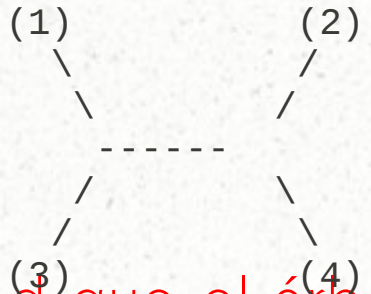
¿Que es el árbol de verosimiltud máxima?

Tiene “dirección”, Hay un modelo y data

El MSA = Data, Árbol = Modelo

Entonces el árbol de ML es el árbol con mayor probabilidad?

```
          1          j          ....N
(1)       A G G C U  C  C A A  ....A
(2)       A G G U U  C  G A A  ....A
(3)       A G C C C  A  G A A.... A
(4)       A U U U C  G  G A A.... C
```




Sitio J: Cual es la probabilidad que el árbol corresponde a la evolución del sitio?

En otras palabras: El árbol explique los datos!

# Bayesian Phylogeny

## Prior and Posterior Probability

Maximum likelihood, looks for the model with the highest probability of producing the data.

Bayesian statistics look for the model that is most likely to have produced the data.

Y NO es lo mismo!

Un médico le dice a un paciente que tiene una probabilidad de 999 en 1000 de tener SIDA

El ensayo dio positivo

Tenemos 1 falso positivo por mil ensayos

Generalmente lo que dice el médico es incorrecto porque NO tiene en cuenta el prior

search for tree that maximizes the chance of seeing the data (P (Data | Tree))

Maximum likelihood

search for tree that maximizes the chance of seeing the tree given the data (P (Tree | Data))

Bayesian inference

# Bayesian Phylogeny

| | Sí | No |
|---|---|---|
| Paciente con cancer | 100% | 0% |
| Paciente sin cancer | 10% | 90% |

1% tiene cancer

Pr(+|Test+) = 100/(100+10)= 0.91 Likelihood

Supongamos una población de 1000 y ponemos la tabla

| | Sí | No |
|---|---|---|
| 10 pacientes con cancer | 10 | 0 |
| 990 pacientes sin cancer | 99 | 891 |

Pr(+|Test+|1%)= 10/99+10= 0.092 Posterior Probability

# Bayesian Phylogeny

Bayes Theorem: $P(A|B) = P(B|A) * P(A) / P(B)$

Que en nuestro caso significa:

Prob (usted tiene cáncer si se obtiene un resultado positivo) = Prob (se obtiene un resultado positivo si usted tiene cánce) * Prob (usted tiene cáncer, independientemente del resultado del ensayo) / Prob (se obtiene un resultado positivo, independientemente de si usted tiene cáncer)

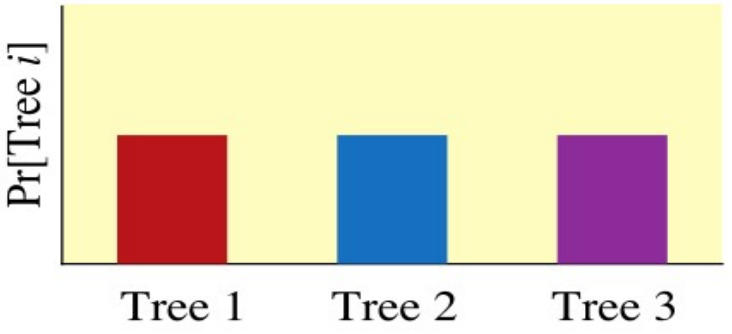
1.0 * 0.01 / 0.109 = 0.092

El conocimiento del porcentaje positivos reales es el prior!

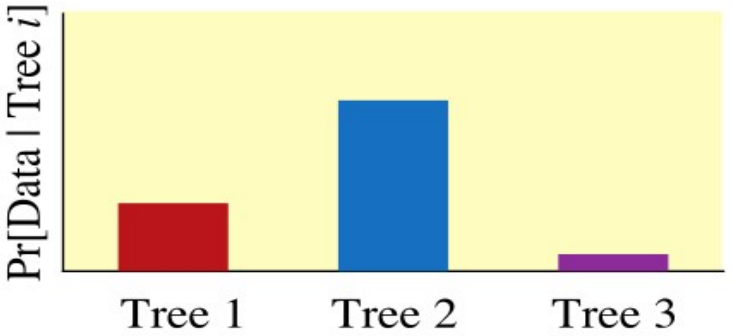
# Filogenia

## Mr Bayes

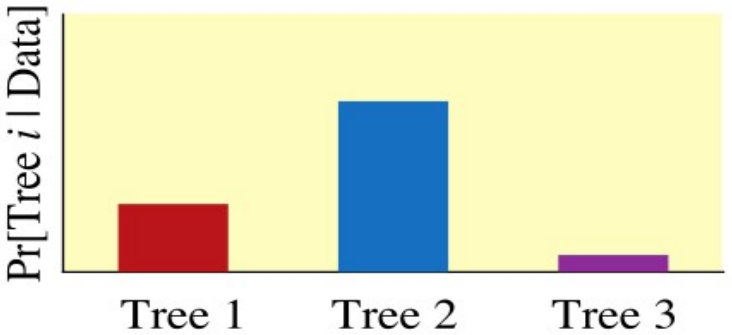
The **prior probability** of a tree represents the probability of the tree before the observations have been made. Typically, all trees are considered equally probable, a priori. However, other information can be used to give some trees more prior probability (e.g., the taxonomy of the group).




The **likelihood** is proportional to the probability of the observations (often an alignment of DNA sequences) conditional on the tree. This probability requires making specific assumptions about the processes generating the observations.




The **posterior probability** of a tree is the probability of the tree conditional on the observations. It is obtained by combining the prior and likelihood for each tree using Bayes' formula.




ig. 1. The main components of a Bayesian analysis.

# Filogenia

## Mr Bayes

Bayesian inference is statistical inference in which evidence or observations are used to update or to newly infer the probability that a hypothesis may be true.

$$Pr[Tree \mid Data] = \frac{Pr[Data \mid Tree] * Pr[Tree]}{Pr[Data]}$$

Prior probability of a phylogeny

Probability of a phylogeny
= Likelihood

Posterior probability of a phylogeny
=Probability that the most probable tree is correct

**Table 1.** The Bayesian approach to problems in phylogeny.

| Problem | Bayesian approach | Ref. |
|---|---|---|
| Inferring phylogeny | Find tree with maximum posterior probability; evaluate features in common among the sampled trees | (*1–3*) |
| Evaluating uncertainty in phylogenies | Evaluate clade probabilities; form credible set containing trees whose cumulative probability sums to 0.95 | (*3*, *40*) |
| Detecting selection | Model substitution process on the codon and calculate probability of being in purifying or positively selected class; sample substitutions and count number of synonymous and nonsynonymous changes | (*29*, *32*) |
| Comparative analyses | Perform analysis on many trees, and weight results by the probability that each tree is correct | (*41–43*) |
| Divergence times | Use fossils as a calibration. Infer divergence times by using a strict or relaxed molecular clock | (*44*) |
| Testing molecular clock | Calculate Bayes factor for the clock versus no branch length restrictions | (*24*) |

## Como determinar la “prior probability?”

## Mr Bayes

Bayes Theorem predice el futuro en base del pasado

Flip a coin: podemos determinar las dos P por empírica

Flip a coin in the future and gamble based on empirical/historical data

Sín empirica decis es 50-50, es la Probabilidad sin prior

Pero si la empírica te muestra algo diferente? 75-25!

Seguis con 50-50?

Como aplicar en filogenía?

Es dificil

# Mr Bayes

Cual es el prior?

Generalmente una “Flat Distribution

Supongamos que el árbol ML tiene verosimilitud 72, No 2 tiene 70. Nro, 2 nos puede dar información?

Distribución A: 0-1-2-4-70-72-70-68-2-0

Distribución B: 0-(.....)12-22-44-70-72-70-68-22-12-1-0

Distribución A me da mas confianza que el árbol ML es, al lado el árbol más probable, el correcto

Puede ser que es tan grave que cambia la ML.

# Mr Bayes

## Implicaciones

Es una onda bastante nueva, no es nada necesario dibujar un árbol Bayes

1 Bayes con su matemática PUEDE ser más rápido que ML con bootstrap
2 Bayes no require bootstrap
Phyml tiene “approximate likelihood-ratio test (aLRT) based on the log ratio between the likelihood value of the current tree and that of the best alternative
3 Por ahora solo con flat distribution