

La Estadística detrás del Análisis Filogénico

Hugo Flores Arguedas

Departamento Ciencias y Matemáticas
Arkansas State University Campus Queretaro
hfloresarguedas@astate.edu

Escuela de Otono, Biología Matemática, Oct 7-11, 2024



Estadística y Filogenias 1

¿Qué hicimos ayer? Hablamos de...

Estadística y Filogenias 1

¿Qué hicimos ayer? Hablamos de...

- Similitud entre secuencias.

Estadística y Filogenias 1

¿Qué hicimos ayer? Hablamos de...

- Similitud entre secuencias.
- BLAST

Estadística y Filogenias 1

¿Qué hicimos ayer? Hablamos de...

- Similitud entre secuencias.
- BLAST

El BLAST es un algoritmo que funciona gracias a:

Estadística y Filogenias 1

¿Qué hicimos ayer? Hablamos de...

- Similitud entre secuencias.
- BLAST

El BLAST es un algoritmo que funciona gracias a:

- Referencia 1: Karlin and Altschul
- Referencia 2:

The Annals of Statistics
1990, Vol. 18, No. 2, 571–581

**STATISTICAL COMPOSITION OF HIGH-SCORING
SEGMENTS FROM MOLECULAR SEQUENCES**

By SAMUEL KARLIN¹, AMIR DEMBO AND TSUTOMU KAWABATA

*Stanford University, Stanford University and University of
Electro-Communications*

Tutorial | Published: 27 September 2001

Having a BLAST with bioinformatics (and avoiding BLASTphemy)

[Alexander Pertsemlidis](#)  & [John W Fondon III](#)

[Genome Biology](#) **2**, Article number: reviews2002.1 (2001) | [Cite this article](#)

59k Accesses | **59** Citations | **7** Altmetric | [Metrics](#)

Abstract

Searching for similarities between biological sequences is the principal means by which bioinformatics contributes to our understanding of biology. Of the various informatics tools developed to accomplish this task, the most widely used is BLAST, the basic local alignment search tool. This article discusses the principles, workings, applications and potential pitfalls of BLAST, focusing on the implementation developed at the National Center for Biotechnology Information.

BLASTphemy

Árboles Filogenéticos

¿Qué es un árbol filogenético?

¿Qué es un árbol filogenético?

Un Árbol Filogenético es una representación gráfica de la historia evolutiva de secuencias biológicas, que nos permite visualizar las relaciones evolutivas entre ellas.

¿Qué es un árbol filogenético?

Un Árbol Filogenético es una representación gráfica de la historia evolutiva de secuencias biológicas, que nos permite visualizar las relaciones evolutivas entre ellas.

[nature](#) > [nature ecology & evolution](#) > [articles](#) > [article](#)

Article | [Open access](#) | Published: 12 July 2024

The nature of the last universal common ancestor and its impact on the early Earth system

[Edmund R. R. Moody](#) ✉, [Sandra Álvarez-Carretero](#), [Tara A. Mahendrarajah](#), [James W. Clark](#), [Holly C. Betts](#), [Nina Dombrowski](#), [Lénárd L. Szánthó](#), [Richard A. Boyle](#), [Stuart Daines](#), [Xi Chen](#), [Nick Lane](#), [Ziheng Yang](#), [Graham A. Shields](#), [Gergely J. Szöllősi](#), [Anja Spang](#), [Davide Pisani](#) ✉, [Tom A. Williams](#) ✉, [Timothy M. Lenton](#) ✉ & [Philip C. J. Donoghue](#) ✉

[Nature Ecology & Evolution](#) **8**, 1654–1666 (2024) | [Cite this article](#)

102k Accesses | **7** Citations | **1381** Altmetric | [Metrics](#)

Luca Nature

¿Qué es un árbol filogenético?

¿Qué es un árbol filogenético?

Un Árbol Filogenético es una representación gráfica de la historia evolutiva de secuencias biológicas, que nos permite visualizar las relaciones evolutivas entre ellas. Tenemos diferentes formas para hacer árboles:

¿Qué es un árbol filogenético?

Un Árbol Filogenético es una representación gráfica de la historia evolutiva de secuencias biológicas, que nos permite visualizar las relaciones evolutivas entre ellas. Tenemos diferentes formas para hacer árboles:

- **Distance-Based**

- Unweighted Pair Group Method using Arithmetic average (UPGMA)
- Neighbor Joining (NJ)

¿Qué es un árbol filogenético?

Un Árbol Filogenético es una representación gráfica de la historia evolutiva de secuencias biológicas, que nos permite visualizar las relaciones evolutivas entre ellas. Tenemos diferentes formas para hacer árboles:

- **Distance-Based**

- Unweighted Pair Group Method using Arithmetic average (UPGMA)
- Neighbor Joining (NJ)

- **Character-Based**

- Parsimony
- Maximum Likelihood
- Bayesian Inference

Distance-Based:

Distance-Based:

Los métodos de construcción basados en distancia involucran calcular distancias evolutivas entre secuencias usando modelos de substitución, los cuáles son usados a su vez para construir una matriz de distancia.

Distance-Based:

Los métodos de construcción basados en distancia involucran calcular distancias evolutivas entre secuencias usando modelos de substitución, los cuáles son usados a su vez para construir una matriz de distancia.

Los dos métodos basados en distancia más populares son UPGMA y NJ. Estos métodos están inspirados en **técnicas de clustering**.

- Unweighted Pair Group Method using Arithmetic average (UPGMA)
- Neighbor Joining (NJ)

Métodos basados en distancia

UPGMA es el más simple de los métodos basados en distancia que construye un árbol filogenético con raíz usando un clustering secuencial.

Métodos basados en distancia

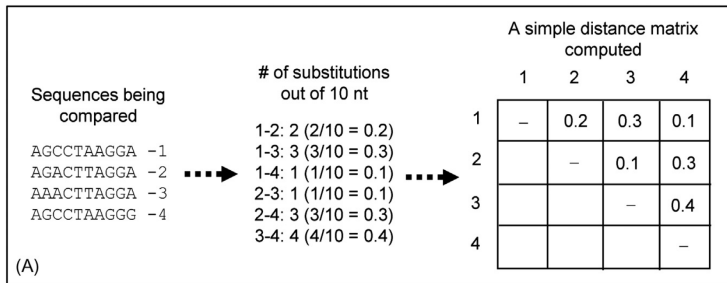
UPGMA es el más simple de los métodos basados en distancia que construye un árbol filogenético con raíz usando un clustering secuencial.

Primero, todas las secuencias son comparadas usando el resultado del alineamiento para calcular la matriz de distancia.

Métodos basados en distancia

UPGMA es el más simple de los métodos basados en distancia que construye un árbol filogenético con raíz usando un clustering secuencial.

Primero, todas las secuencias son comparadas usando el resultado del alineamiento para calcular la matriz de distancia.



Métodos basados en distancia

UPGMA es el más simple de los métodos basados en distancia que construye un árbol filogenético con raíz usando un clustering secuencial.

Métodos basados en distancia

UPGMA es el más simple de los métodos basados en distancia que construye un árbol filogenético con raíz usando un clustering secuencial.

Primero, todas las secuencias son comparadas usando el resultado del alineamiento para calcular la matriz de distancia.

Métodos basados en distancia

UPGMA es el más simple de los métodos basados en distancia que construye un árbol filogenético con raíz usando un clustering secuencial.

Primero, todas las secuencias son comparadas usando el resultado del alineamiento para calcular la matriz de distancia.

Usando esta matriz, **las dos secuencias con la menor distancia por pares se agrupan como un solo par**. Un nodo se coloca en el punto medio entre ellas.

Métodos basados en distancia

UPGMA es el más simple de los métodos basados en distancia que construye un árbol filogenético con raíz usando un clustering secuencial.

Primero, todas las secuencias son comparadas usando el resultado del alineamiento para calcular la matriz de distancia.

Usando esta matriz, **las dos secuencias con la menor distancia por pares se agrupan como un solo par**. Un nodo se coloca en el punto medio entre ellas.

Checa el siguiente video, empezando en el minuto 1:55,

https://www.youtube.com/watch?v=_b82GJhx8VM

Métodos basados en distancia

O bien, considera el siguiente ejemplo:

Métodos basados en distancia

O bien, considera el siguiente ejemplo:

	A	B	C	D
A	0			
B	3	0		
C	5	4	0	
D	7	1	2	0

Matrix 1

Métodos basados en distancia

O bien, considera el siguiente ejemplo:

	A	B	C	D
A	0			
B	3	0		
C	5	4	0	
D	7	1	2	0

Matrix 1

Usando esta matriz, **las dos secuencias con menor distancia son B y D**. Así, estas dos secuencias se agrupan en un par.

Métodos basados en distancia

O bien, considera el siguiente ejemplo:

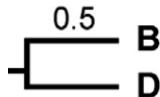
	A	B	C	D
A	0			
B	3	0		
C	5	4	0	
D	7	1	2	0

Matrix 1

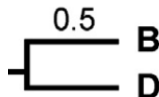
Usando esta matriz, **las dos secuencias con menor distancia son B y D**. Así, estas dos secuencias se agrupan en un par.

Posteriormente, la distancia entre este par y todas las otras secuencias se recalculan para formar una nueva matriz (ver siguiente diapositiva).

Métodos basados en distancia



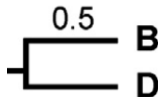
Métodos basados en distancia



$$d(A,BD) = \{d(A,B)+d(A,D)\}/2 = (3+7)/2 = 5$$

$$d(BD,C) = \{d(B,C)+d(C,D)\}/2 = (4+2)/2 = 3$$

Métodos basados en distancia



$$d(A, BD) = \{d(A, B) + d(A, D)\} / 2 = (3 + 7) / 2 = 5$$

$$d(BD, C) = \{d(B, C) + d(C, D)\} / 2 = (4 + 2) / 2 = 3$$

	A	BD	C
A	0		
BD	5	0	
C	5	3	0

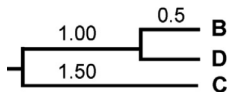
Matrix 2

Métodos basados en distancia

Con esta nueva matriz, se identifica y se agrupa la secuencia más cercana al primer par.

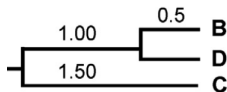
Métodos basados en distancia

Con esta nueva matriz, se identifica y se agrupa la secuencia más cercana al primer par.



Métodos basados en distancia

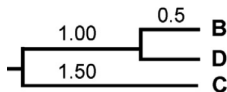
Con esta nueva matriz, se identifica y se agrupa la secuencia más cercana al primer par.



El proceso se repite hasta que todas las secuencias hayan sido posicionadas en el árbol.

Métodos basados en distancia

Con esta nueva matriz, se identifica y se agrupa la secuencia más cercana al primer par.



El proceso se repite hasta que todas las secuencias hayan sido posicionadas en el árbol.

$$\boxed{d(A, BDC) = \{d(A, B) + d(A, D) + d(A, C)\} / 3 = (3 + 7 + 5) / 3 = 5}$$

Métodos basados en distancia

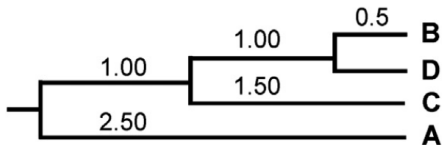
	A	BDC
A	0	
BDC	5	0

Matrix 3

Métodos basados en distancia

	A	BDC
A	0	
BDC	5	0

Matrix 3

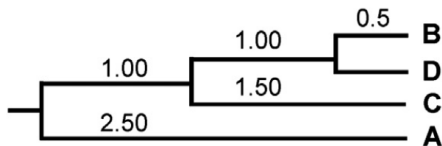


The Tree

Métodos basados en distancia

	A	BDC
A	0	
BDC	5	0

Matrix 3



The Tree

El método UPGMA **asume** que la **tasa evolutiva** de todos los taxons es **constante**, por lo que son equidistantes a la raíz.

Métodos basados en distancia

El método neighbor-joining (NJ) es **el más comúnmente utilizado entre los métodos basados en distancia.**

Métodos basados en distancia

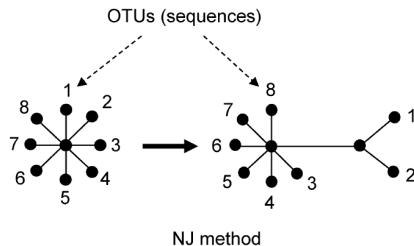
El método neighbor-joining (NJ) es **el más comúnmente utilizado entre los métodos basados en distancia.**

Es similar al método UPGMA , sin embargo, no asume la tasa constante, por lo que produce un árbol sin raíz.

Métodos basados en distancia

El método neighbor-joining (NJ) es **el más comúnmente utilizado entre los métodos basados en distancia.**

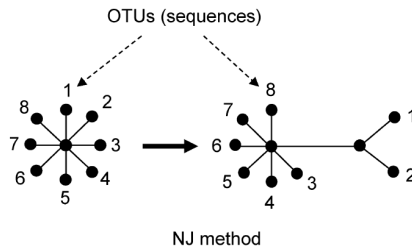
Es similar al método UPGMA , sin embargo, no asume la tasa constante, por lo que produce un árbol sin raíz.



Métodos basados en distancia

El método neighbor-joining (NJ) es **el más comúnmente utilizado entre los métodos basados en distancia.**

Es similar al método UPGMA , sin embargo, no asume la tasa constante, por lo que produce un árbol sin raíz.



Checa el siguiente video, empezando en el minuto 7:51 para más detalles,

https://www.youtube.com/watch?v=_b82GJhx8VM

Métodos basados en distancia

Para un ejemplo concreto utilizando Python, se puede consultar el repositorio:

<https://github.com/hugofloresar/EOBM2024>

Carga el notebook PhylogeneticTrees.ipynb en Google Colab:

<https://colab.research.google.com>

Métodos basados en posición (Character-Based)

Los métodos character-based involucran analizar las secuencias propiamente. Estos métodos evalúan todas las secuencias a la vez analizando un sitio a la vez.

Métodos basados en posición (Character-Based)

Los métodos character-based involucran analizar las secuencias propiamente. Estos métodos evalúan todas las secuencias a la vez analizando un sitio a la vez.

Estos métodos son generalmente **considerados más exactos** que los métodos basados en distancia. Sin embargo, son **computacionalmente más intensivos** y requieren modelos estadísticos sofisticados.

Métodos basados en posición (Character-Based)

Los métodos character-based involucran analizar las secuencias propiamente. Estos métodos evalúan todas las secuencias a la vez analizando un sitio a la vez.

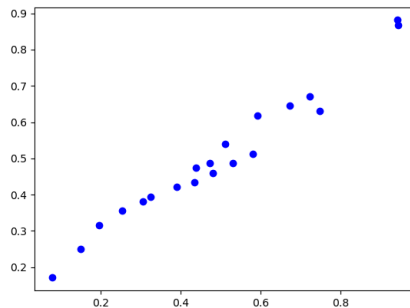
Estos métodos son generalmente **considerados más exactos** que los métodos basados en distancia. Sin embargo, son **computacionalmente más intensivos** y requieren modelos estadísticos sofisticados.

Máxima parsimonia (MP) y máxima verosimilitud (ML) son los métodos más conocidos.

Máxima Verosimilitud

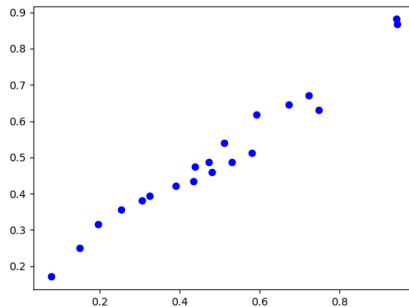
El caso de la regresión lineal

Considera el siguiente conjunto de datos



El caso de la regresión lineal

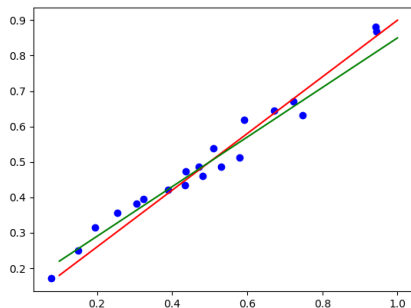
Considera el siguiente conjunto de datos



¿Pueden estos datos ser explicados por una función lineal?

El caso de la regresión lineal

¿Cuál recta describe mejor los datos?

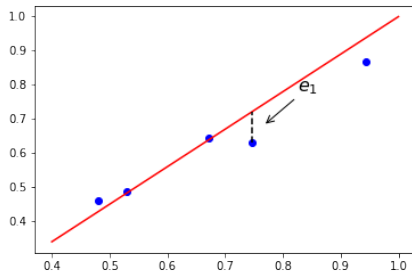


El caso de la regresión lineal

El método de los **mínimos cuadrados**:

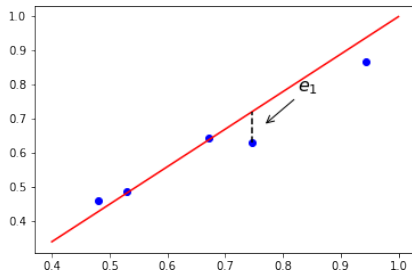
El caso de la regresión lineal

El método de los **mínimos cuadrados**:



El caso de la regresión lineal

El método de los **mínimos cuadrados**:



A cada recta,

$$y = mx + b$$

se le asocia la cantidad $E(m, b) = \sum e_i^2$

El caso de la regresión lineal

En el caso de la regresión lineal simple, se asume el modelo

$$Y = \beta_0 + \beta_1 X + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

El caso de la regresión lineal

En el caso de la regresión lineal simple, se asume el modelo

$$Y = \beta_0 + \beta_1 X + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

donde ϵ es una ruido aleatorio, independiente a X .

El caso de la regresión lineal

En el caso de la regresión lineal simple, se asume el modelo

$$Y = \beta_0 + \beta_1 X + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

donde ϵ es una ruido aleatorio, independiente a X . La pdf condicional de Y para cada x está dada por

$$\prod_i^n p(y_i | x_i; \beta_0, \beta_1, \sigma^2) = \prod_i^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}}$$

El caso de la regresión lineal

En el caso de la regresión lineal simple, se asume el modelo

$$Y = \beta_0 + \beta_1 X + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

donde ϵ es una ruido aleatorio, independiente a X . La pdf condicional de Y para cada x está dada por

$$\prod_i^n p(y_i | x_i; \beta_0, \beta_1, \sigma^2) = \prod_i^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}}$$

Los parámetros del modelo son β_0, β_1 y a veces, σ .

Máxima Verosimilitud

El método de Máxima Verosimilitud (ML) usa cada posición en un alineamiento y evalúa todos los posible árboles.

Máxima Verosimilitud

El método de Máxima Verosimilitud (ML) usa cada posición en un alineamiento y evalúa todos los posible árboles.

Calcula la verosimilitud para cada árbol y busca el que tenga la mayor de todas.

Máxima Verosimilitud

El método de Máxima Verosimilitud (ML) usa cada posición en un alineamiento y evalúa todos los posible árboles.

Calcula la verosimilitud para cada árbol y busca el que tenga la mayor de todas.

¿Cuáles son los parámetros de un árbol filogenético?

Máxima Verosimilitud

El método de Máxima Verosimilitud (ML) usa cada posición en un alineamiento y evalúa todos los posible árboles.

Calcula la verosimilitud para cada árbol y busca el que tenga la mayor de todas.

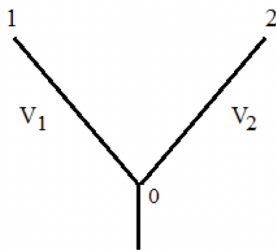
¿Cuáles son los parámetros de un árbol filogenético?

La idea principal está en determinar:

- La topología del árbol
- La longitud de las ramas
- Los parámetros del modelo evolutivo

que maximicen la probabilidad de observar las secuencias con las que se cuentan.

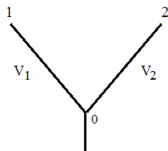
Máxima Verosimilitud: Un ejemplo



Puntas o nodos terminales 1 y 2, la raíz 0, y la longitud de sus ramas v_1 y v_2 . Supongamos que se conoce el estado del nodo 0, S_0 . Entonces:

$$L = \pi_{S_0} P_{S_0 S_1}(v_1) P_{S_0 S_2}(v_2)$$

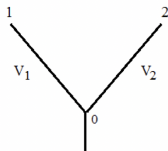
Máxima Verosimilitud: Un ejemplo



$$L = \pi_{S_0} P_{S_0 S_1}(v_1) P_{S_0 S_2}(v_2)$$

donde

Máxima Verosimilitud: Un ejemplo

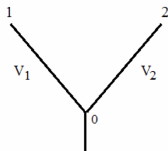


$$L = \pi_{S_0} P_{S_0 S_1}(v_1) P_{S_0 S_2}(v_2)$$

donde

- π_{S_0} es la probabilidad de la que la raíz se encuentre en el estado S_0 .

Máxima Verosimilitud: Un ejemplo

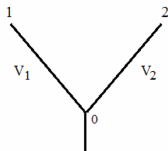


$$L = \pi_{S_0} P_{S_0 S_1}(v_1) P_{S_0 S_2}(v_2)$$

donde

- π_{S_0} es la probabilidad de la que la raíz se encuentre en el estado S_0 .
- $P_{S_0 S_1}(v_1)$ es la probabilidad de que la raíz transicione de S_0 a la punta 1, en el estado S_1 , con una rama de longitud v_1

Máxima Verosimilitud: Un ejemplo

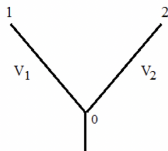


$$L = \pi_{S_0} P_{S_0 S_1}(v_1) P_{S_0 S_2}(v_2)$$

donde

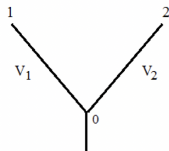
- π_{S_0} es la probabilidad de la que la raíz se encuentre en el estado S_0 .
- $P_{S_0 S_1}(v_1)$ es la probabilidad de que la raíz transicione de S_0 a la punta 1, en el estado S_1 , con una rama de longitud v_1
- $P_{S_0 S_2}(v_2)$ es la probabilidad de que la raíz transicione de S_0 a la punta 2, en el estado S_2 , con una rama de longitud v_2

Máxima Verosimilitud: Un ejemplo



Para este ejemplo,

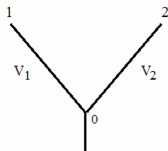
Máxima Verosimilitud: Un ejemplo



Para este ejemplo,

- Sabemos los estados S_1 y S_2 .

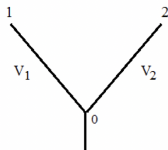
Máxima Verosimilitud: Un ejemplo



Para este ejemplo,

- Sabemos los estados S_1 y S_2 .
- Pero no necesariamente sabemos el estado S_0 .

Máxima Verosimilitud: Un ejemplo

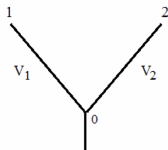


Para este ejemplo,

- Sabemos los estados S_1 y S_2 .
- Pero no necesariamente sabemos el estado S_0 .
- Supongamos que estamos considerando una secuencia de nucleótidos, entonces

$$L = \sum_{S_0} \pi_{S_0} P_{S_0 S_1}(v_1) P_{S_0 S_2}(v_2)$$

Máxima Verosimilitud: Un ejemplo



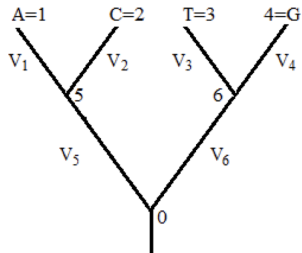
Para este ejemplo,

- Sabemos los estados S_1 y S_2 .
- Pero no necesariamente sabemos el estado S_0 .
- Supongamos que estamos considerando una secuencia de nucleótidos, entonces

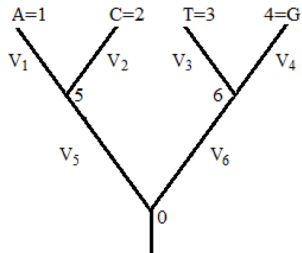
$$L = \sum_{S_0} \pi_{S_0} P_{S_0 S_1}(v_1) P_{S_0 S_2}(v_2)$$

con $S_0 \in \{A, C, T, G\}$.

Máxima Verosimilitud: Ejemplo 2

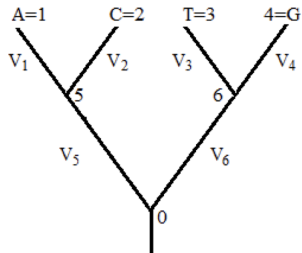


Máxima Verosimilitud: Ejemplo 2



LES TOCA A USTEDES!

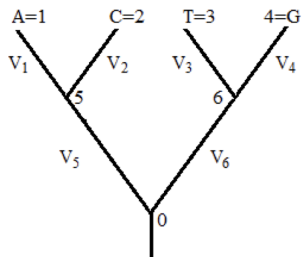
Máxima Verosimilitud: Ejemplo 2



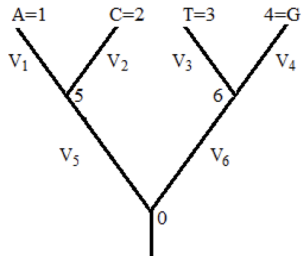
LES TOCA A USTEDES!

$L = ???$

Máxima Verosimilitud: Ejemplo 2

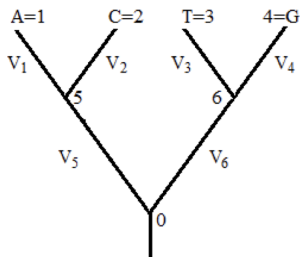


Máxima Verosimilitud: Ejemplo 2



Suponiendo que conocemos S_0 , S_5 y S_6

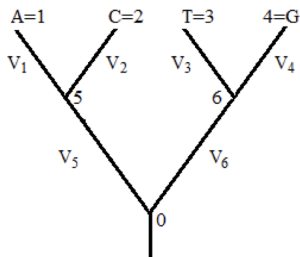
Máxima Verosimilitud: Ejemplo 2



Suponiendo que conocemos S_0 , S_5 y S_6

$$L = \pi_{S_0} P_{S_0 S_5}(v_5) P_{S_5 A}(v_1) P_{S_5 C}(v_2) P_{S_0 S_6}(v_6) P_{S_6 T}(v_3) P_{S_6 G}(v_4)$$

Máxima Verosimilitud: Ejemplo 2

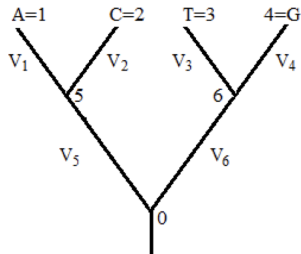


Suponiendo que conocemos S_0 , S_5 y S_6

$$L = \pi_{S_0} P_{S_0 S_5}(v_5) P_{S_5 A}(v_1) P_{S_5 C}(v_2) P_{S_0 S_6}(v_6) P_{S_6 T}(v_3) P_{S_6 G}(v_4)$$

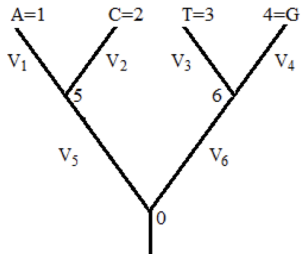
pero en general, no conocemos S_0 , S_5 y S_6 , ...

Máxima Verosimilitud: Ejemplo 2



$$L = \sum_{S_0} \sum_{S_5} \sum_{S_6} \pi_{S_0} P_{S_0 S_5}(v_5) P_{S_5 A}(v_1) P_{S_5 C}(v_2) P_{S_0 S_6}(v_6) P_{S_6 T}(v_3) P_{S_6 G}(v_4)$$

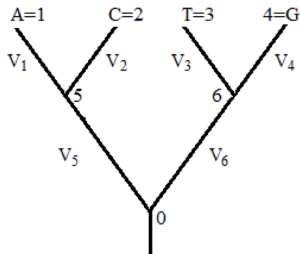
Máxima Verosimilitud: Ejemplo 2



$$L = \sum_{S_0} \sum_{S_5} \sum_{S_6} \pi_{S_0} P_{S_0 S_5}(v_5) P_{S_5 A}(v_1) P_{S_5 C}(v_2) P_{S_0 S_6}(v_6) P_{S_6 T}(v_3) P_{S_6 G}(v_4)$$

con $S_0, S_5, S_6 \in \{A, C, T, G\}$.

Máxima Verosimilitud: Ejemplo 2



$$L = \sum_{S_0} \sum_{S_5} \sum_{S_6} \pi_{S_0} P_{S_0 S_5}(v_5) P_{S_5 A}(v_1) P_{S_5 C}(v_2) P_{S_0 S_6}(v_6) P_{S_6 T}(v_3) P_{S_6 G}(v_4)$$

con $S_0, S_5, S_6 \in \{A, C, T, G\}$.

Así se calcula la verosimilitud en **un sitio determinado**.

Máxima Verosimilitud

La verosimilitud es determinada al evaluar la probabilidad de que cierto modelo evolutivo haya generado los datos observados.

Máxima Verosimilitud

La verosimilitud es determinada al evaluar la probabilidad de que cierto modelo evolutivo haya generado los datos observados.

Las verosimilitudes para cada sitio se multiplican para obtener la verosimilitud de cada árbol.

Máxima Verosimilitud

La verosimilitud es determinada al evaluar la probabilidad de que cierto modelo evolutivo haya generado los datos observados.

Las verosimilitudes para cada sitio se multiplican para obtener la verosimilitud de cada árbol.

El método ML es el más lento y computacionalmente intensivo de los métodos mencionados.

¿Qué es eso del Bootstrap?

¿Qué es eso del Bootstrap?

Sample
(Original set of sequences)

	1	2	3	4	5	6	7	8	9	10
1-	A	T	G	G	A	C	C	A	A	A
2-	A	T	G	C	C	A	A	C	T	T
3-	A	T	G	C	C	A	C	A	T	T
4-	A	T	G	G	A	C	A	C	A	A

¿Qué es eso del Bootstrap?

Sample
(Original set of sequences)

	1	2	3	4	5	6	7	8	9	10
1-	A	T	G	G	A	C	C	A	A	A
2-	A	T	G	C	C	A	A	C	T	T
3-	A	T	G	C	C	A	C	A	T	T
4-	A	T	G	G	A	C	A	C	A	A

Vamos a hacer un remuestreo (con reemplazo) de la muestra original para crear nuevos conjuntos de pseudomuestras:

¿Qué es eso del Bootstrap?

Sample (Original set of sequences)

	1	2	3	4	5	6	7	8	9	10
1-	A	T	G	G	A	C	C	A	A	A
2-	A	T	G	C	C	A	A	C	T	T
3-	A	T	G	C	C	A	C	A	T	T
4-	A	T	G	G	A	C	A	C	A	A

Vamos a hacer un remuestreo (con reemplazo) de la muestra original para crear nuevos conjuntos de pseudomuestras:

Resampling-1

	1	1	1	2	2	3	3	8	8	10
1-	A	A	A	T	T	G	G	A	A	A
2-	A	A	A	T	T	G	G	C	C	T
3-	A	A	A	T	T	G	G	A	A	T
4-	A	A	A	T	T	G	G	C	C	A

¿Qué es eso del Bootstrap?

Se repite este proceso varias veces

¿Qué es eso del Bootstrap?

Se repite este proceso varias veces

Resampling-2

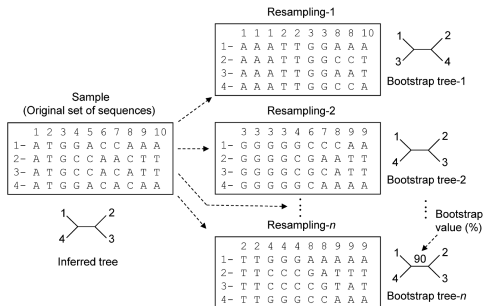
	3	3	3	3	4	6	7	8	9	9
1-	G	G	G	G	G	C	C	C	A	A
2-	G	G	G	G	C	G	A	A	T	T
3-	G	G	G	G	C	G	C	A	T	T
4-	G	G	G	G	G	C	A	A	A	A

¿Qué es eso del Bootstrap?

Se repite este proceso varias veces

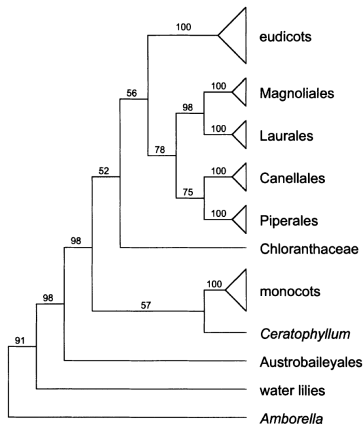
Resampling-2

	3	3	3	3	4	6	7	8	9	9
1-	G	G	G	G	G	C	C	C	A	A
2-	G	G	G	G	C	G	A	A	T	T
3-	G	G	G	G	C	G	C	A	T	T
4-	G	G	G	G	G	C	A	A	A	A



¿Qué es eso del Bootstrap?

Con cada pseudomuestra, se repite el análisis, hasta llegar a algo como esto




Los números representan los **bootstrap valores**. Indican el porcentaje de los árboles bootstrap que *apoyan* o sustentan la topología obtenida en el árbol original.

Reconstruction de Ancestros

https://www.youtube.com/watch?v=Siau2o_egGI

Reconstruction de Ancestros

https://www.youtube.com/watch?v=Siau2o_egGI




latest

Installation

TreeTime command line usage

Detailed command line documentation

API documentation



Private docs hosting for any Docs as Code tool. Start a trial with **Read the Docs for Business**.

Ad by EthicalAds

Read the Docs

v: latest

/ TreeTime: time-tree and ancestral sequence inference

[Edit on GitHub](#)

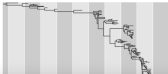
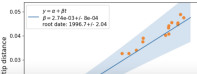
TreeTime: time-tree and ancestral sequence inference

CI passing

TreeTime provides routines for ancestral sequence reconstruction and inference of molecular-clock phylogenies, i.e., a tree where all branches are scaled such that the positions of terminal nodes correspond to their sampling times and internal nodes are placed at the most likely time of divergence.

To optimize the likelihood of time-scaled phylogenies, TreeTime uses an iterative approach that first optimizes branch lengths of the tree given the sequence data and date constraints, and then optimizes coalescent tree priors, relaxed clock parameters, or resolves polytomies. This cycle is repeated a few times. The only topology optimization are (optional) resolution of polytomies in a way that is most (approximately) consistent with the sampling time constraints on the tree.

The code is hosted on github.com/neherlab/treetime.

Muchas gracias