

# Introducción a la bioinformática

Laura Gómez-Romero

Instituto Nacional de Medicina Genómica

*lgomez@inmegen.gob.mx*

October 15, 2021

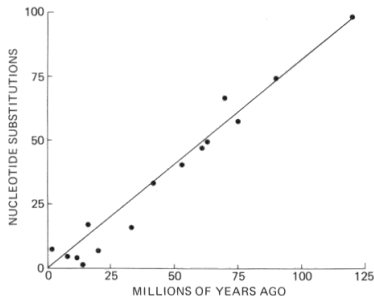
# Tabla de contenidos

- 1 Implicaciones de la similitud entre secuencias
- 2 Alineamientos
- 3 Matriz de puntos: Dot plot
- 4 Alineamientos globales
- 5 Alineamientos locales
- 6 Otros tipos de alineamientos

# Tabla de contenidos

- 1 Implicaciones de la similitud entre secuencias
- 2 Alineamientos
- 3 Matriz de puntos: Dot plot
- 4 Alineamientos globales
- 5 Alineamientos locales
- 6 Otros tipos de alineamientos

# El reloj molecular

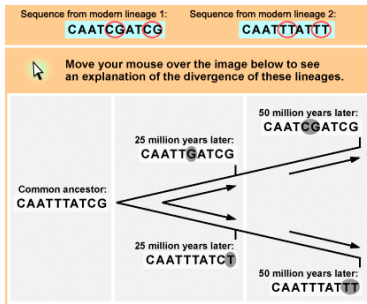


“The strong linear relationship suggests that molecular differences between pairs of species are proportional to the time of their separation, not the degree of organismal difference. Therefore, measures of genetic divergence can be used to date the time of divergence for species pairs for which no fossil data are available: genes function as Molecular Clocks

— A. C. Wilson 1976

[https://www.mun.ca/biology/scarr/Molecular\\_Clock.html](https://www.mun.ca/biology/scarr/Molecular_Clock.html)

# El reloj molecular



“Imagine that a length of DNA found in two species differs by four bases (as shown below) and we know that this entire length of DNA changes at a rate of approximately one base per 25 million years. That means that the two DNA versions differ by 100 million years of evolution and that their common ancestor lived 50 million years ago. Since each lineage experienced its own evolution, the two species must have descended from a common ancestor that lived at least 50 million years ago.

— [https://evolution.berkeley.edu/evolibrary/article/molecclocks\\_01](https://evolution.berkeley.edu/evolibrary/article/molecclocks_01)

# ¿Qué nos dice el grado de similitud entre dos secuencias?

- En general, mientras más similitud compartan dos secuencias más parecida será su función.
- Las secuencias para el mismo gene en un grupo de especies serán más diferentes mientras la distancia filogenética aumente.
- Las secuencias acumulan mutaciones con el paso del tiempo.
- Mientras más tiempo haya pasado desde que dos especies divergieron más mutaciones se habrán acumulado en sus genes.

¿Qué estamos asumiendo?

- Que existe un **ancestro común**

**DEFINICIÓN:** Dos secuencias que comparten un ancestro común son homólogos

# Tabla de contenidos

- 1 Implicaciones de la similitud entre secuencias
- 2 Alineamientos**
- 3 Matriz de puntos: Dot plot
- 4 Alineamientos globales
- 5 Alineamientos locales
- 6 Otros tipos de alineamientos



Un alineamiento entre dos secuencias puede ser utilizado para:

- Calcular el grado de similitud entre dos secuencias/calcular la distancia filogenética
- Buscar dominios funcionales
- Comparar un mRNA con su DNA de origen
- Identificar polimorfismos y mutaciones entre secuencias

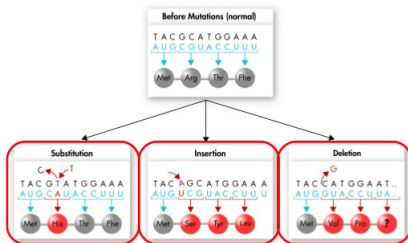
# Tipos de mutaciones

Mutaciones puntuales: cambios de un sólo nucleótido.

## Gene Mutations: Point Mutations

A point mutation is a change in a single nucleotide.

There are three types of point mutations:



Una mutación puntual a nivel de DNA puede tener efectos catastróficos a nivel de proteína.

# Tipos de mutaciones

Existen otro tipo de mutaciones

## Types of Mutations

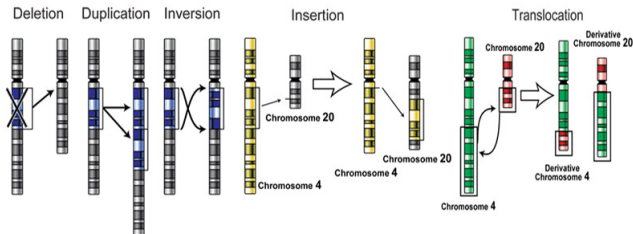


Imagen tomada de <https://microbenotes.com/types-of-mutations/>

# Comparación de dos secuencias

Una forma de comparar dos secuencias sería escribir una debajo de otra y contar las diferencias: 3 identidades en nuestro ejemplo

No alignment

CGATGCTAGCGTATCGTAGTCTATCGTAC

|        ||

ACGATGCTAGCGTTTCGTATCATCGTA

# Comparación de dos secuencias

Una forma de comparar dos secuencias sería escribir una debajo de otra y contar las diferencias: 3 identidades en nuestro ejemplo

No alignment

```
CGATGCTAGCGTATCGTAGTCTATCGTAC
      |      ||
ACGATGCTAGCGTTTCGTATCATCGTA
```

Sin embargo, este no es un alineamiento óptimo Si agregamos un gap al inicio, y algunos otros cambios obtenemos: 27 identidades

Aligned

```
-CGATGCTAGCGTATCGTAGTCTATCGTAC
||||| ||||| ||||| ||||| |||||
ACGATGCTAGCGTTTCGTA-TC-ATCGTA-
```

# No existe un alineamiento único entre secuencias

Dada la gran variedad de cambios que pueden ocurrir entre dos secuencias no existe un alineamiento único entre ellas

No gaps (10 matches)

a: ATATTGCTACGTATATCAT

| | | | | | | |

b: ATATATGCTACGTATCAT

With one gap (14 matches)

a: ATAT-TGCTACGTATATCAT

| | | | | | | | | |

b: ATATATGCTACGTATCAT

With two gaps (16 matches)

a: ATAT-TGCTACGTATATCAT

| | | | | | | | | | | |

b: ATATATGCTACG--TATCAT

**¿Cuál es el correcto?** No sabemos  
**¿Cuál es el mejor?** Depende del sistema de puntaje que elijamos

# Tipos de alineamientos

- **Globales:** Crean un alineamiento que cubra completamente ambas secuencias, agregando los gaps que sean necesarios
- **Locales:** Alinean únicamente los segmentos de alta identidad
- **Glocal:** Un alineamiento global que permite rearrreglos de la secuencia (desarrollado para alinear genomas completos)
- **Semi-globales:** Tratan de crear un alineamiento que cubra completamente ambas secuencias, sin penalizar por los gaps al inicio y al final del alineamiento

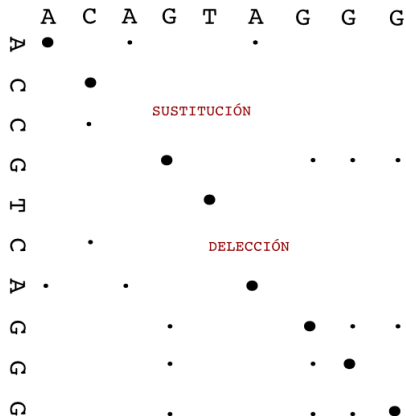
# Tabla de contenidos

- 1 Implicaciones de la similitud entre secuencias
- 2 Alineamientos
- 3 Matriz de puntos: Dot plot**
- 4 Alineamientos globales
- 5 Alineamientos locales
- 6 Otros tipos de alineamientos

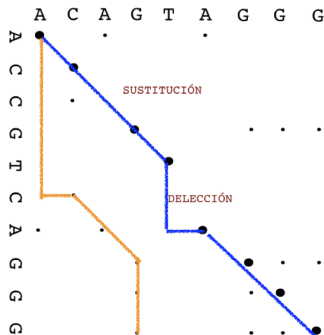


# Matriz de puntos

Una secuencia se representa en el eje horizontal y otra secuencia en el eje vertical. Los matches entre las dos secuencias se representan como puntos en la posición correspondiente. Un alineamiento se forma al conectar los puntos de la matriz.



# Representación gráfica de un alineamiento



**ACAGT - AGGG**  
**ACCGTCAGGG**

**A - - - - CAG - - TAGGG**  
**ACCGTCAGGG - - - - -**

¿Cuál es mejor?

# Sistema de puntaje

- Asignemos a cada tipo de cambio un puntaje
- Calculemos el score de cada alineamiento

Tipo de cambio	Recompensa/ Penalización
Match	4
Mismatch	-2
Gap open	-5
Gap extension	-1

**ACAGT - AGGG**  
**ACCGT CAGGG**

**4 4 -2 4 4 -5 4 4 4 4**

**Score = 25**

**A - - - CAG - - TAGGG**  
**ACCGT CAGGG - - - -**

**4 -5 -1 -1 -1 4 4 4 -5 -1 -5 -1 -1 -1 -1**

**Score = -7**

# El score cambia si cambiamos el sistema de puntaje

- Asignemos a cada tipo de cambio un puntaje
- Calculemos el score de cada alineamiento

**¡Su turno!**

Tipo de cambio	Recompensa/ Penalización
Match	9
Mismatch	-1
Gap open	-2
Gap extension	-1

# El score cambia si cambiamos el sistema de puntaje

- Asignemos a cada tipo de cambio un puntaje
- Calculemos el score de cada alineamiento

Tipo de cambio	Recompensa/ Penalización
Match	9
Mismatch	-1
Gap open	-2
Gap extension	-1

**ACAGT - AGGG**  
**ACCGTCAGGG**

**9 9 -1 9 9 -2 9 9 9 9**

**Score = 69**

**A - - - CAG - - TAGGG**  
**ACCGTCAGGG - - - -**

**9 -2 -1 -1 -1 9 9 9 -2 -1 -2 -1 -1 -1 -1**

**Score = 22**

# Variación estructural: inserciones o deleciones

## Inserciones o deleciones

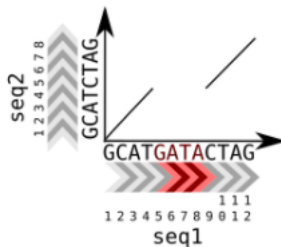


Imagen tomada de [https://bioinf.comav.upv.es/courses/biotech3/theory/sequence\\_alignment.html](https://bioinf.comav.upv.es/courses/biotech3/theory/sequence_alignment.html)

# Variación estructural: duplicaciones

## Duplicaciones

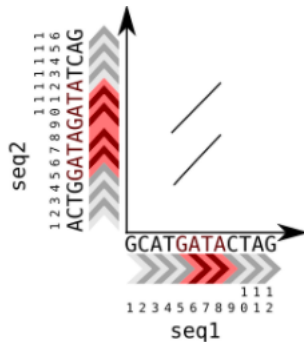


Imagen tomada de [https://bioinf.comav.upv.es/courses/biotech3/theory/sequence\\_alignment.html](https://bioinf.comav.upv.es/courses/biotech3/theory/sequence_alignment.html)

# Variación estructural: inversiones

## Inversiones

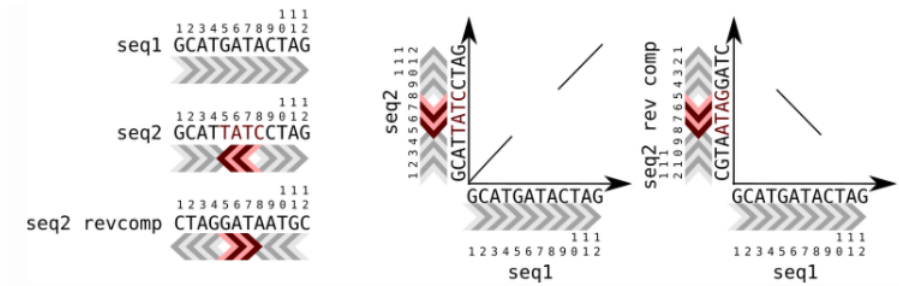


Imagen tomada de [https://bioinf.comav.upv.es/courses/biotech3/theory/sequence\\_alignment.html](https://bioinf.comav.upv.es/courses/biotech3/theory/sequence_alignment.html)



# Comparaciones genómicas: humano vs chimpancé

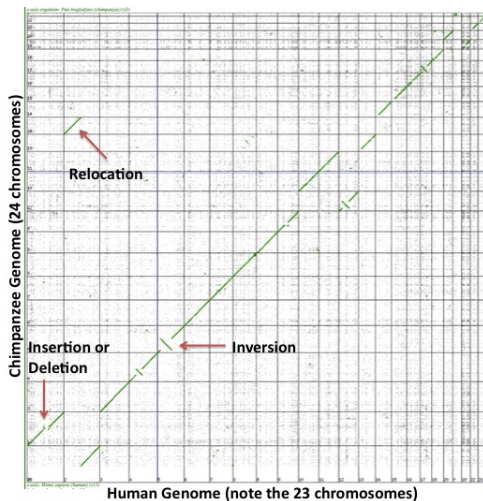


Imagen tomada de [https://genomeevolution.org/wiki/index.php/Synten:\\_Getting\\_the\\_Big\\_Picture](https://genomeevolution.org/wiki/index.php/Synten:_Getting_the_Big_Picture)

# Tabla de contenidos

- 1 Implicaciones de la similitud entre secuencias
- 2 Alineamientos
- 3 Matriz de puntos: Dot plot
- 4 Alineamientos globales**
- 5 Alineamientos locales
- 6 Otros tipos de alineamientos

# Alineamiento global

**¿Qué es un alineamiento global?** Es un alineamiento que cubre completamente ambas secuencias, agregando los gaps que sean necesarios

**¿Cuándo usarlo?** Sólo se recomienda su uso cuando se sospecha que dos secuencias son similares en toda su extensión

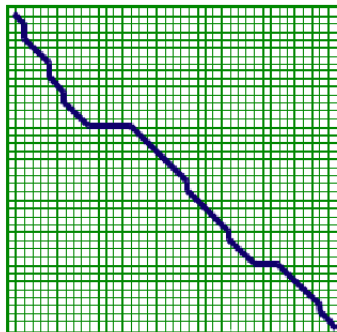


Imagen tomada de [https://bio.libretexts.org/Bookshelves/Computational\\_Biology/](https://bio.libretexts.org/Bookshelves/Computational_Biology/)

# Algoritmo Needleman–Wunsch

- Desarrollado por Saul B. Needleman and Christian D. Wunsch en 1970.
- Una de las primeras aplicaciones de la programación dinámica.
- El algoritmo divide un problema grande en problemas más pequeños y usa las soluciones de los problemas pequeños para encontrar la solución óptima del problema grande.
- El propósito del algoritmo es **encontrar el alineamiento global con el mejor score** de entre todos los alineamientos posibles.

# Algoritmo Needleman–Wunsch

Estas serán nuestras secuencias problema de ejemplo:

- GCATGCU
- GATTACA

Y este nuestro sistema de puntaje:

- Match: +1
- Mismatch o indel: -1

Inicialización de la matriz:

		G	C	A	T	G	C	U
G								
A								
T								
T								
A								
C								
A								

# Algoritmo Needleman–Wunsch

## Reglas del juego

- El camino desde la celda superior o de la celda de la izquierda representan un indel. Por lo tanto, toma los scores de las celdas superior o izquierda y suma el score del indel
- El camino en diagonal representa un match o mismatch, por lo tanto toma el score de la celda arriba-izquierda (diagonal) y agrega el score que corresponda (match o mismatch)
- La celda extrema superior izquierda se inicializa en 0

Inicialización de la matriz:

		G	C	A	T	G	C	U
G								
A								
T								
T								
A								
C								
A								

# Algoritmo Needleman–Wunsch

## Reglas del juego

- El camino desde la celda superior o de la celda de la izquierda representan un indel. Por lo tanto, toma los scores de las celdas superior o izquierda y suma el score del indel
- El camino en diagonal representa un match o mismatch, por lo tanto toma el score de la celda arriba-izquierda (diagonal) y agrega el score que corresponda

Rellenando la primer fila y la primer columna

		G	C	A	T	G	C	U
	0	-1	-2	-3	-4	-5	-6	-7
G	-1							
A	-2							
T	-3							
T	-4							
A	-5							
C	-6							
A	-7							

# Algoritmo Needleman–Wunsch

## Reglas del juego

- El camino desde la celda superior o de la celda de la izquierda representan un indel. Por lo tanto, toma los scores de las celdas superior o izquierda y suma el score del indel
- El camino en diagonal representa un match o mismatch, por lo tanto toma el score de la celda arriba-izquierda (diagonal) y agrega el score que corresponda

## Primer subproblema

		G
	0	-1
G	-1	X

- Camino desde la celda diagonal:  
match:  $0 + 1 = 1$
- Camino desde la celda superior: indel:  
 $-1 + (-1) = -2$
- Camino desde la celda izquierda:  
indel:  $-1 + (-1) = -2$



# Algoritmo Needleman–Wunsch

## Reglas del juego

- El camino desde la celda superior o de la celda de la izquierda representan un indel. Por lo tanto, toma los scores de las celdas superior o izquierda y suma el score del indel
- El camino en diagonal representa un match o mismatch, por lo tanto toma el score de la celda arriba-izquierda (diagonal) y agrega el score que corresponda

## Primer subproblema

		<b>G</b>
	0	-1
<b>G</b>	-1	1



- Se almacena el score más alto
- Se almacena el camino de donde vino

# Algoritmo Needleman–Wunsch

**X**

		<b>G</b>	<b>C</b>
	0	-1	-2
<b>G</b>	-1	1	<b>X</b>
<b>A</b>	-2	<b>Y</b>	

- Camino desde la celda diagonal:  
mismatch:  $-1 + (-1) = -1$
- Camino desde la celda superior: indel:  
 $-2 + (-1) = -3$
- Camino desde la celda izquierda: indel:  $1 + (-1) = 0$

**Y**

		<b>G</b>	<b>C</b>
	0	-1	-2
<b>G</b>	-1	1	<b>X</b>
<b>A</b>	-2	<b>Y</b>	

- Camino desde la celda diagonal:  
mismatch:  $-1 + (-1) = -2$
- Camino desde la celda superior: indel:  
 $1 + (-1) = 0$
- Camino desde la celda izquierda: indel:  $-2 + (-1) = -3$

# Algoritmo Needleman–Wunsch

X

		G	C
	0	-1	-2
G	-1	1	0
A	-2	0	

- Camino desde la celda diagonal:  
mismatch:  $-1 + (-1) = -1$
- Camino desde la celda superior: indel:  
 $-2 + (-1) = -3$
- Camino desde la celda izquierda:**  
**indel:  $1 + (-1) = 0$**

Y

		G	C
	0	-1	-2
G	-1	1	0
A	-2	0	

- Camino desde la celda diagonal:  
mismatch:  $-1 + (-1) = -2$
- Camino desde la celda superior:**  
**indel:  $1 + (-1) = 0$**
- Camino desde la celda izquierda:  
indel:  $-2 + (-1) = -3$

# Algoritmo Needleman–Wunsch

Cada camino de la celda extrema derecha inferior hacia la celda extrema izquierda superior representan alineamientos igualmente válidos.

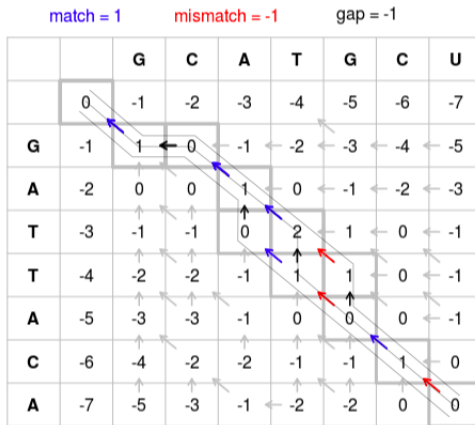
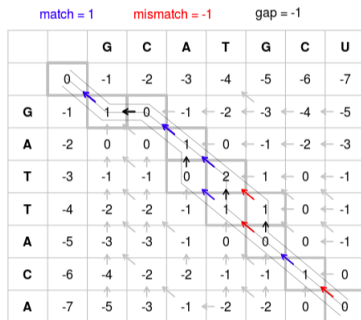


Imagen tomada de <https://en.wikipedia.org/>

# Algoritmo Needleman–Wunsch



- Una flecha diagonal representa un match o mismatch
- Flechas horizontales generan un gap en la secuencia lateral
- Flechas verticales generan un gap en la secuencia encabezado
- Múltiples flechas representan opciones de alineamientos

# Algoritmo Needleman–Wunsch

match = 1      mismatch = -1      gap = -1

		G	C	A	T	G	C	U
	0	-1	-2	-3	-4	-5	-6	-7
G	-1	1	0	-1	-2	-3	-4	-5
A	-2	0	0	1	0	-1	-2	-3
T	-3	-1	-1	0	2	1	0	-1
T	-4	-2	-2	-1	1	1	0	-1
A	-5	-3	-3	-1	0	0	0	-1
C	-6	-4	-2	-2	-1	-1	1	0
A	-7	-5	-3	-1	-2	-2	0	0

- Una flecha diagonal representa un match o mismatch
- Flechas horizontales generan un gap en la secuencia lateral
- Flechas verticales generan un gap en la secuencia encabezado
- Múltiples flechas representan opciones de alineamientos

```

U → CU → GCU → -GCU → T-GCU → AT-GCU → CAT-GCU → GCAT-GCU
A → CA → ACA → TACA → TTACA → ATTACA → -ATTACA → G-ATTACA
      ↓
(branch) → TGCU → ...
           → TACA → ...
    
```

# Existen distintos sistemas de puntaje

- Match = 1
- Mismatch = -1
- Indel = -1

- Match = 0
- Mismatch = 1
- Indel = 1

Mientras más grande el score, más parecidas las secuencias

**¿Y en este caso?**

# Existen distintos sistemas de puntaje

- Match = 1
- Mismatch = -1
- Indel = -1

Mientras más grande el score, más parecidas las secuencias

- Match = 0
- Mismatch = 1
- Indel = 1

**Mientras más grande el score más diferentes las secuencias**



# Matrices de sustitución

	A	G	C	T
A	1	-1	-1	-1
G	-1	1	-1	-1
C	-1	-1	1	-1
T	-1	-1	-1	1

- Match = 1
- Mismatch = -1

	A	G	C	T
A	1	-1	-1	-1
G	-1	1	-1	-1
C	-1	-1	1	-1
T	-1	-1	-1	4

- Match A-A, G-G o C-C = 1
- Match T-T = 4
- Mismatch = -1

# Matrices de sustitución: PAM y BLOSUM

- Existen aminoácidos químicamente más similares entre ellos
- Una forma de pesar una sustitución de un aminoácido por otro es contar cuántas veces ocurre cada sustitución en posiciones homólogas en alineamientos de proteínas homólogas
- Sustituciones neutrales o con poco impacto funcional ocurrirán más frecuentemente que sustituciones con un alto impacto funcional

# Matrices de sustitución: PAM

- Existen varias matrices PAM
- Su nombre se deriva de la cantidad de mutaciones cada 100 aminoácidos que diferencian las secuencias comparadas para crear las matrices
- PAM60 significa que existen 60 mutaciones por cada 100 aminoácidos de secuencia
- Es importante notar que cada posición de la secuencia puede estar involucrada en más de una mutación (sustituciones hacia distintos aminoácidos), por lo que existen matrices PAM160 o incluso PAM200.

# Matrices de sustitución: BLOSUM

- Las matrices BLOSUM se construyen a partir de segmentos altamente conservados cubriendo rangos evolutivos más distantes que las matrices PAM
- El nombre de estas matrices deriva del mínimo porcentaje de identidad permitido entre las secuencias utilizadas
- BLOSUM70 se construyó a partir de secuencias más similares que BLOSUM50

# Herramienta online: EMBOSS Needle

← → ↻ ebi.ac.uk/Tools/psa/emboss\_needle/ [Icons] En pausa

EMBL-EBI Services Research Training Industry About us 🔍 EMBL-EBI Hinxton ▾

## EMBOSS Needle

Input form Web services Help & Documentation Bioinformatics Tools FAQ Feedback Share

Tools > Pairwise Sequence Alignment > EMBOSS Needle

## Pairwise Sequence Alignment

EMBOSS Needle reads two input sequences and writes their optimal global sequence alignment to file.

### STEP 1 - Enter your protein sequences

Enter a pair of

PROTEIN ▾

sequences. Enter or paste your first **protein** sequence in any supported format:

Or, upload a file: Seleccionar archivo No se eligió archivo

Use an [example sequence](#) | [Clear sequence](#) | [See more example inputs](#)

# Tabla de contenidos

- 1 Implicaciones de la similitud entre secuencias
- 2 Alineamientos
- 3 Matriz de puntos: Dot plot
- 4 Alineamientos globales
- 5 Alineamientos locales**
- 6 Otros tipos de alineamientos



# Algoritmo Smith-Waterman

- Propuesto por Temple Smith y Michael Waterman en 1981.
- También se basa en principios de programación dinámica.
- Garantiza que el alineamiento que se encuentre es óptimo (el mejor) con respecto al sistema de puntaje establecido.
- Es una técnica exhaustiva y lenta, lo que la hace imposible de aplicar para secuencias muy largas.
- No recomendable si se quieren hacer un gran número de comparaciones.

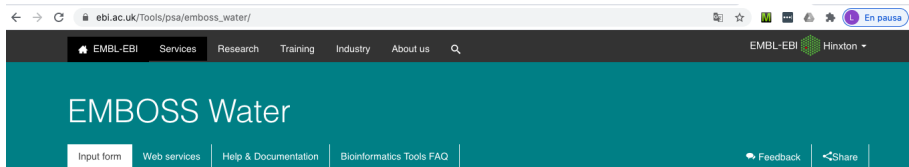


## Procedimiento

- Se genera una matriz de puntaje siguiendo el mismo procedimiento que en el algoritmo Needleman–Wunsch, con una diferencia: las celdas negativas se convierten a ceros.
- El procedimiento de traceback inicia en la celda con el puntaje más alto
- El procedimiento de traceback termina cuando una celda con puntaje cero es alcanzada

Ver animación:

[https://en.wikipedia.org/wiki/Smith%E2%80%93Waterman\\_algorithm](https://en.wikipedia.org/wiki/Smith%E2%80%93Waterman_algorithm)



Tools &gt; Pairwise Sequence Alignment &gt; EMBOSS Water

## Pairwise Sequence Alignment

EMBOSS Water uses the Smith-Waterman algorithm (modified for speed enhancements) to calculate the local alignment of two sequences.

STEP 1 - Enter your protein sequences

Enter a pair of

PROTEIN

sequences. Enter or paste your first **protein** sequence in any supported format:

Or, upload a file: 

Seleccionar archivo

 No se eligió archivo

Use a example sequence | Clear sequence | See more example inputs

## Basic Local Alignment Search Tool

Stephen F. Altschul<sup>1</sup>, Warren Gish<sup>1</sup>, Webb Miller<sup>2</sup>  
Eugene W. Myers<sup>3</sup> and David J. Lipman<sup>1</sup>

<sup>1</sup>*National Center for Biotechnology Information  
National Library of Medicine, National Institutes of Health  
Bethesda, MD 20894, U.S.A.*

<sup>2</sup>*Department of Computer Science  
The Pennsylvania State University, University Park, PA 16802, U.S.A.*

<sup>3</sup>*Department of Computer Science  
University of Arizona, Tucson, AZ 85721, U.S.A.*

*(Received 26 February 1990; accepted 15 May 1990)*

A new approach to rapid sequence comparison, basic local alignment search tool (BLAST), directly approximates alignments that optimize a measure of local similarity, the maximal segment pair (MSP) score. Recent mathematical results on the stochastic properties of MSP scores allow an analysis of the performance of this method as well as the statistical significance of alignments it generates. The basic algorithm is simple and robust; it can be implemented in a number of ways and applied in a variety of contexts including straight-forward DNA and protein sequence database searches, motif searches, gene identification searches, and in the analysis of multiple regions of similarity in long DNA sequences. In addition to its flexibility and tractability to mathematical analysis, BLAST is an order of magnitude faster than existing sequence comparison tools of comparable sensitivity.

## Algoritmo

- Utiliza un algoritmo conocido como "seed and extend"
- Identifica en la secuencia problema (query) una serie de palabras cortas (semillas) no superpuestas
- Busca coincidencias exactas de estas semillas en la base de datos (subject)
- A partir de estas semillas extiende los alineamientos permitiendo mismatches e indels
- La fase de extensión termina cuando el score del alineamiento es menor a un umbral previamente elegido

## Aplicaciones

- Secuencias "cortas" contra secuencias muy largas:
  - Genes contra genomas
  - Exones contra genes
- Una secuencia o conjunto de secuencias contra una base de datos con un gran número de secuencias (todos los genes o proteínas anotados)
  - Un gene o genes contra todos los genomas conocidos
  - Un transcrito o transcritos contra todos los transcriptomas conocidos
  - Una proteína o proteínas contra todos los proteomas conocidos

- Cobertura (query coverage)
  - Porcentaje de la secuencia query cubierta por el alineamiento
- Identidad (percent identity)
  - Porcentaje del alineamiento que coincide con la secuencia en la base de datos (match)
  - Número de bases idénticas del alineamiento / Longitud del alineamiento
- E valor (E value)
  - Probabilidad de tener el alineamiento observado al azar dada la longitud de la secuencia y el tamaño de la base de datos
  - Es proporcional al número de alineamientos que esperas por azar con ese score
  - Entre más cercano sea a 1 es más probable que el resultado sea al azar

# Cobertura



Secuencia query  
Secuencia subject  
Alineamiento local reportado

Cobertura

Longitud secuencia query = 11

Número de bases de la secuencia query incluidas en el alineamiento = 7

Cobertura =  $7/11 = 0.63$

# Identidad



Secuencia query  
Secuencia subject  
Alineamiento local reportado

## Identidad

Número de bases idénticas del alineamiento = 6  
Longitud de alineamiento = 8

Identidad =  $6/8 = 0.75$



← → ↺ blast.ncbi.nlm.nih.gov/Blast.cgi

U.S. National Library of Medicine NCBI National Center for Biotechnology Information Sign in to NCBI

## BLAST®

Home Recent Results Saved Strategies Help


COVID-19 is an emerging, rapidly evolving situation.  
Get the latest public health information from CDC: <https://www.coronavirus.gov>.  
Get the latest research from NIH: <https://www.nih.gov/coronavirus>.  
Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>.

### Basic Local Alignment Search Tool

**BLAST** finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)

**A new version BLAST+ (2.11.0) is here.**  
This version supports a new usage reporting service and a new multi-threading feature.  
Tue, 05 Nov 2020 12:00:00 EST [More BLAST news...](#)

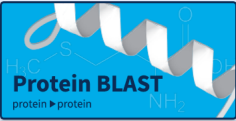
### Web BLAST



**Nucleotide BLAST**  
nucleotide ► nucleotide

**blastx**  
translated nucleotide ► protein

**tblastn**  
protein ► translated nucleotide



**Protein BLAST**  
protein ► protein

# Tipos de BLAST

- **blastn** Búsqueda de nucleótidos en base de datos de nucleótidos
- **blastp** Búsqueda de una secuencia de aminoácidos en base de datos de proteínas
- **blastx** Búsqueda de los 6 marcos de lectura de una secuencia de nucleótidos en base de datos de proteínas
- **tblastn** Búsqueda de una secuencia de aminoácidos en los 6 marcos de lectura de base de datos de nucleótidos

Para cada una de las siguientes secuencias:

- TP53a.fasta
- TP53b.fasta
- TP53c.fasta

Realiza un BLAST contra las siguientes bases de datos:

- Reference RNA sequences
- RefSeq Genome Database
- Reference proteins

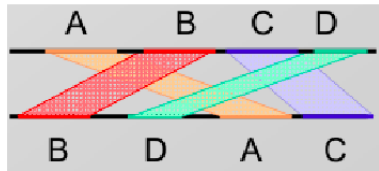
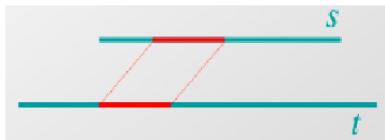
¿Qué puedes concluir de las distintas secuencias FASTA? ¿Cómo interpretas los valores de identidad, cobertura y e.valor para cada BLAST?

# Tabla de contenidos

- 1 Implicaciones de la similitud entre secuencias
- 2 Alineamientos
- 3 Matriz de puntos: Dot plot
- 4 Alineamientos globales
- 5 Alineamientos locales
- 6 Otros tipos de alineamientos**

# Alineamientos glocal

Alinean ambas secuencias de inicio a fin permitiendo rearrreglos genómicos



# Alineamientos semi-globales

Se deriva de un alineamiento global, con la diferencia de que los gaps no se penalizan al inicio o final de las secuencias. Si tenemos las secuencias  $s$  y  $t$  existen los siguientes ocho casos [Sequence Alingment, Dannie Durand, 2015]:

- Gaps no tienen penalización al inicio de  $s$
- Gaps no tienen penalización al inicio de  $t$
- Gaps no tienen penalización al final de  $s$
- Gaps no tienen penalización al final de  $t$
- Gaps no tienen penalización al inicio y final de  $s$
- Gaps no tienen penalización al inicio y final de  $t$
- Gaps no tienen penalización al inicio de  $s$  y final de  $t$
- Gaps no tienen penalización al inicio de  $t$  y final de  $s$

# References I