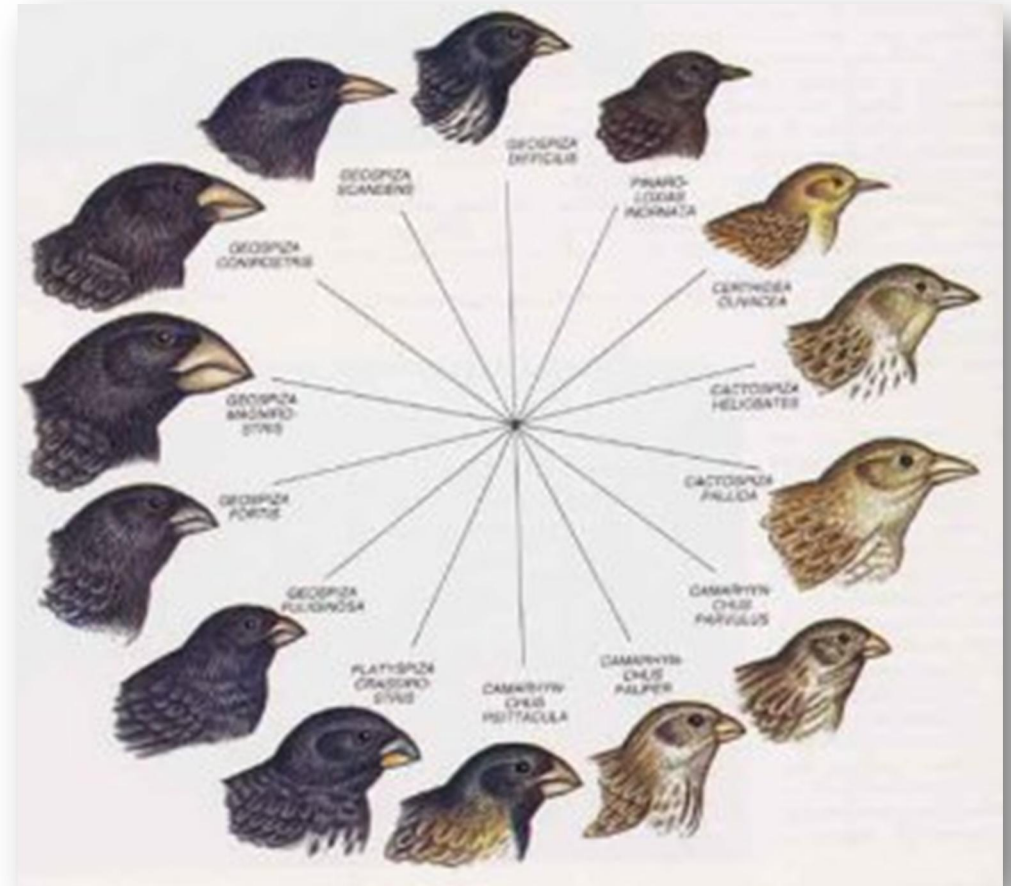


Alineamientos de Secuencias

Análisis comparativo

- El alineamiento de secuencias es similar a otros tipos de análisis comparativo. En ambos es necesario cuantificar las similitudes y diferencias (scoring) entre un grupo relacionado de entidades.
- Es la herramienta más usada en bioinformática y uso va desde la identificación de regiones o dominios en un par de secuencias hasta la comparación de un grupo de secuencias para análisis evolutivos y de filogenia.



Finches of the Galápagos Islands observed by Charles Darwin on the voyage of HMS Beagle

Homología vs. Similitud

Homología entre dos entes biológicos implica una herencia compartida

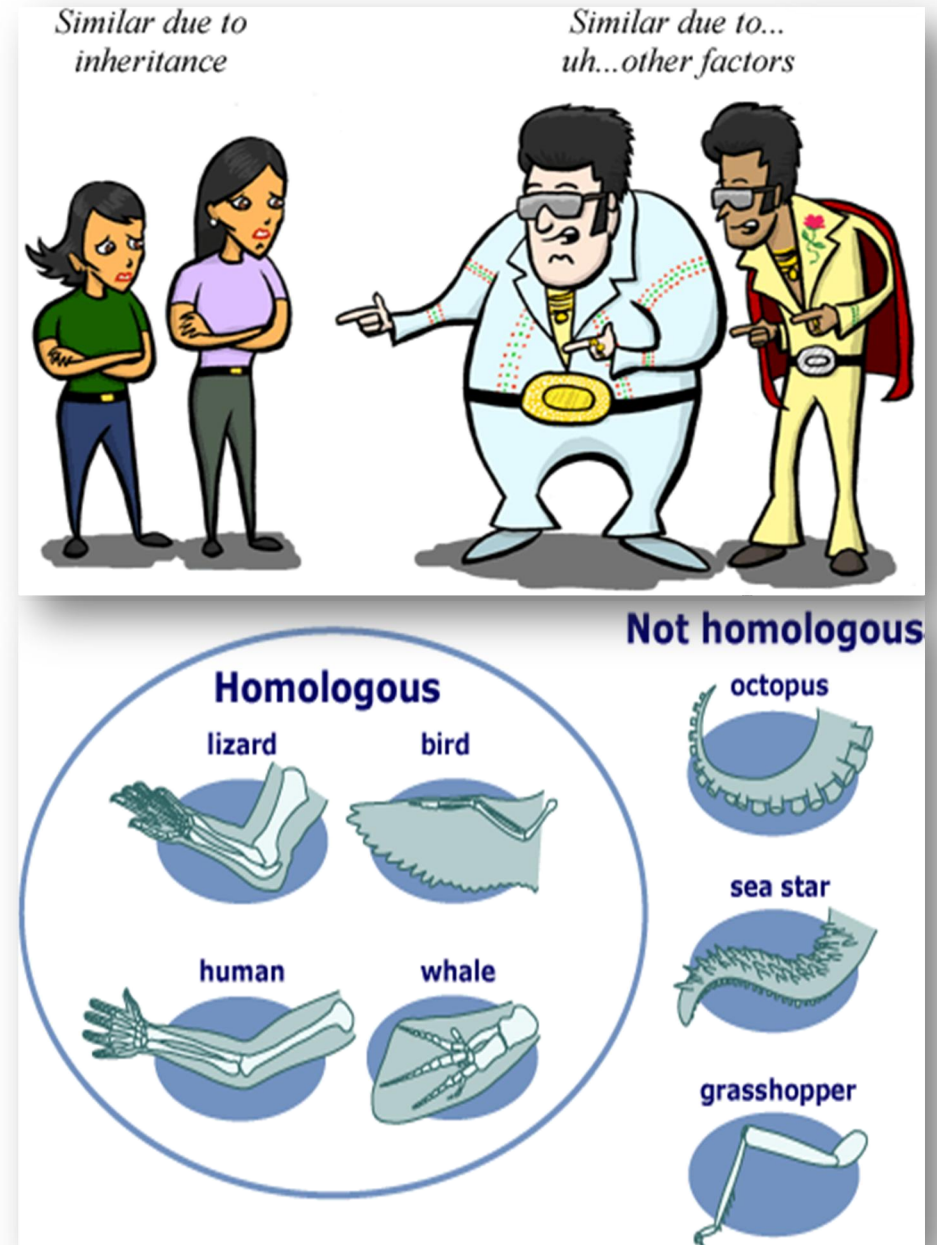
Homología es un término cualitativo:

Se es homólogo o no se es

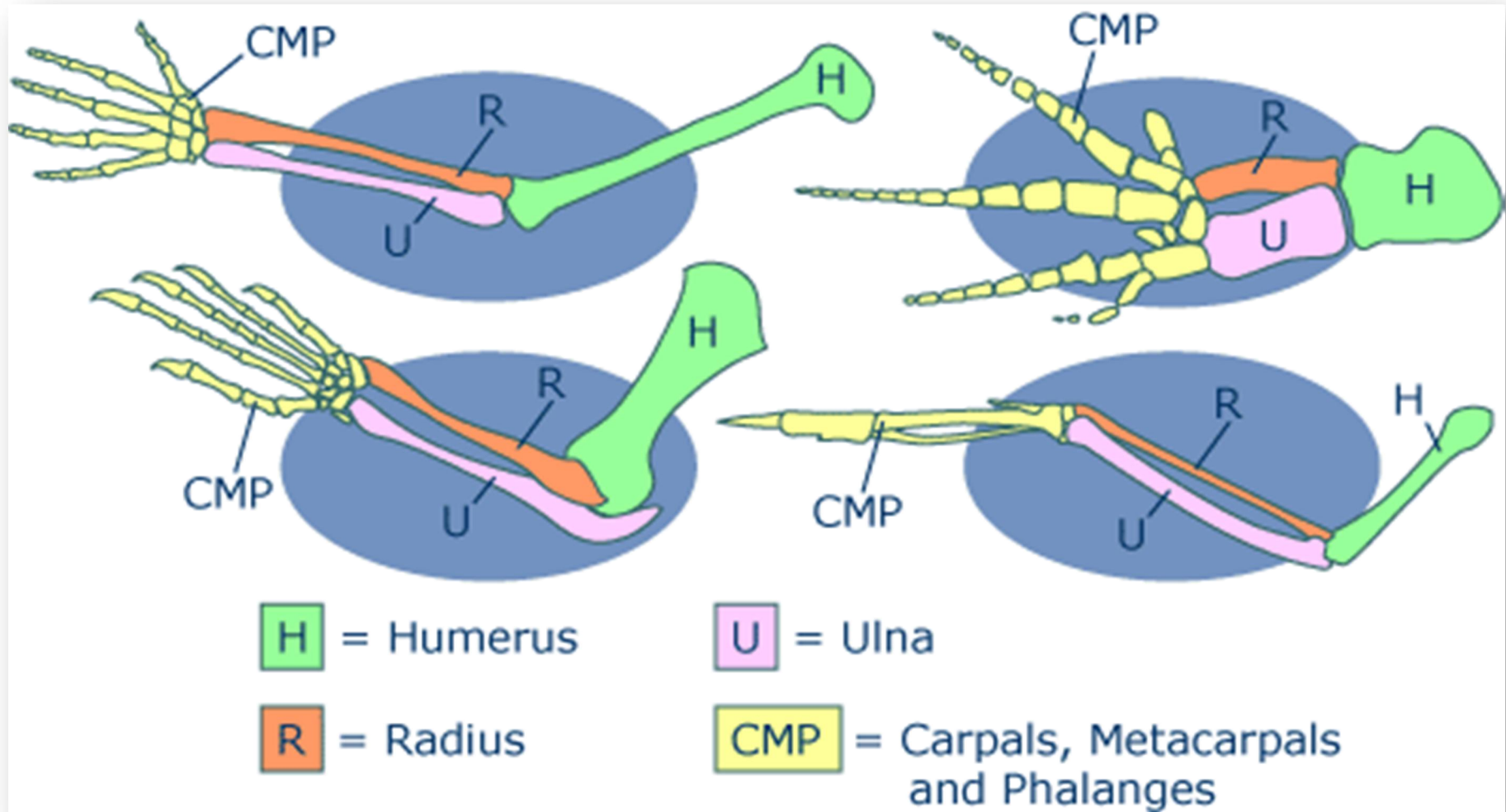
Similitud implica una apreciación cuantitativa o una cuantificación directa de algún carácter.

Podemos usar una medida de similitud para inferir homología

- **Homología**, similaridad debido a herencia de un ancestro común



- **Analogía**, similaridad debida a evolución convergente.

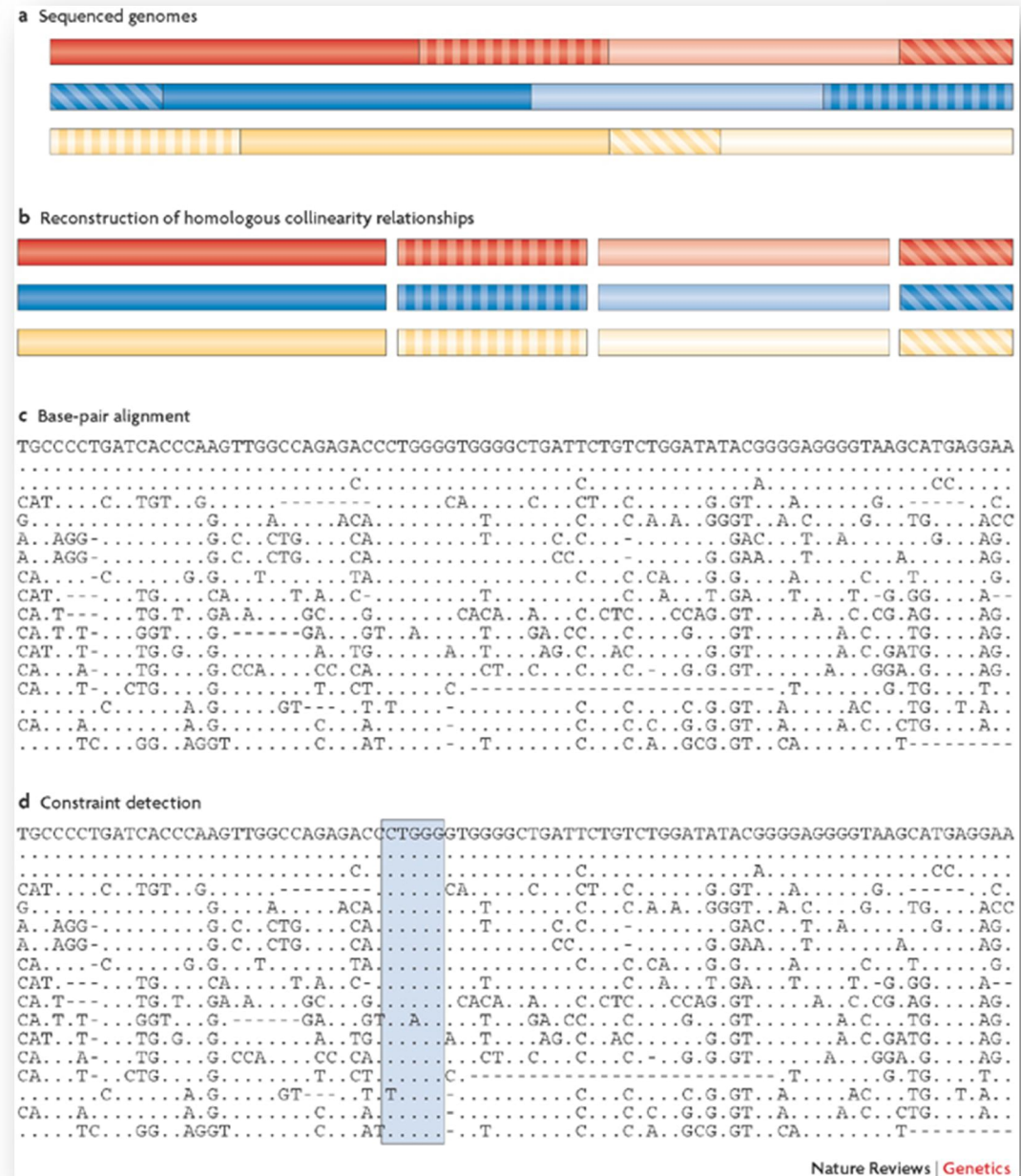


Análisis comparativo

Los algoritmos que alinean secuencias modelan procesos Evolutivos

Deriva de un ancestro común a través de cambios incrementales debido a errores en la replicación del ADN, mutaciones, daño o crossing-over desigual.

Sólo las secuencias actuales son conocidas, las secuencias ancestrales se postulan.



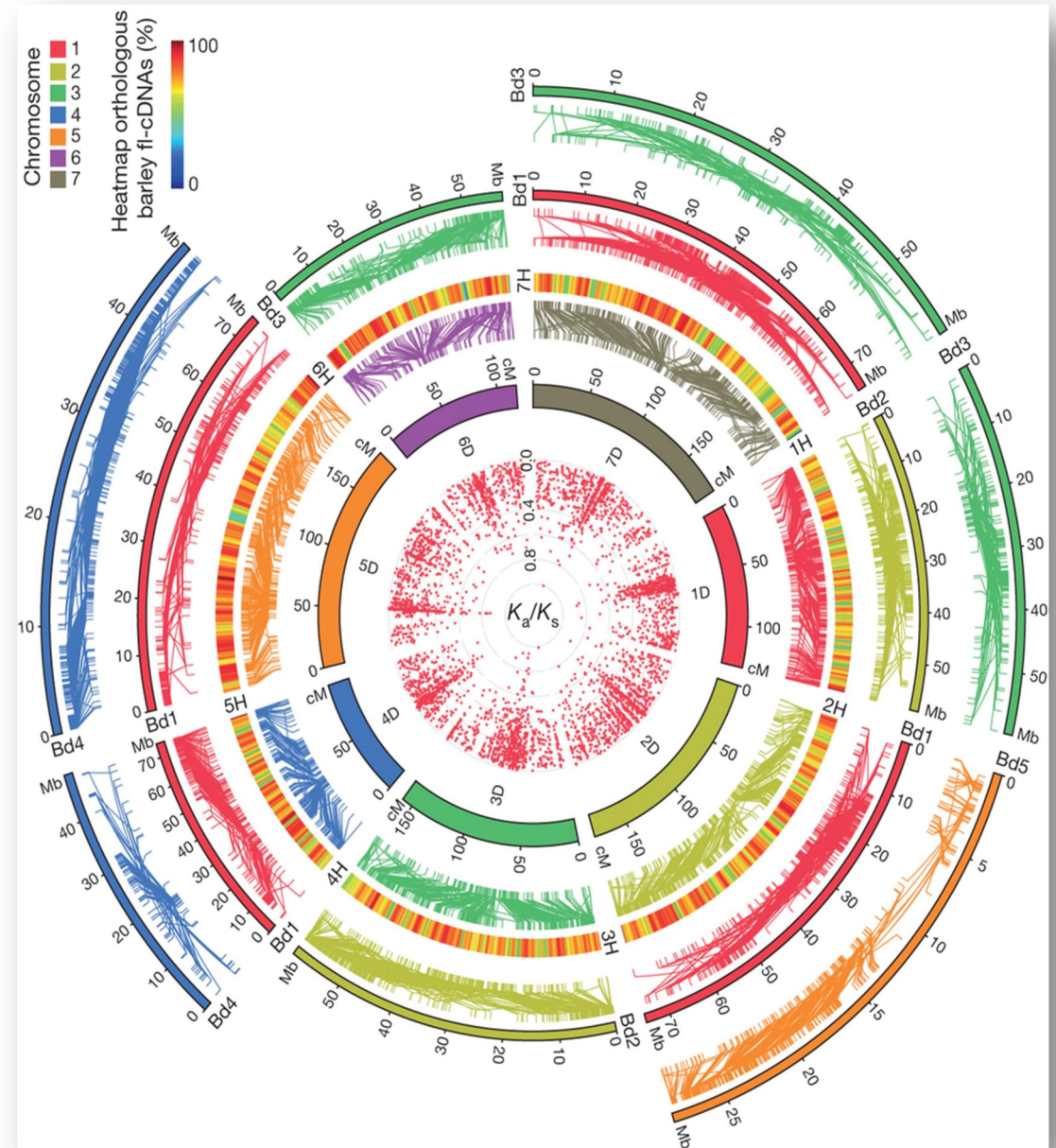
Alineamientos

Qué es un alineamiento?

El procedimiento de comparación de dos (o más) secuencias que busca una serie de caracteres individuales o patrones de caracteres que se encuentren en el mismo orden en ambas secuencias

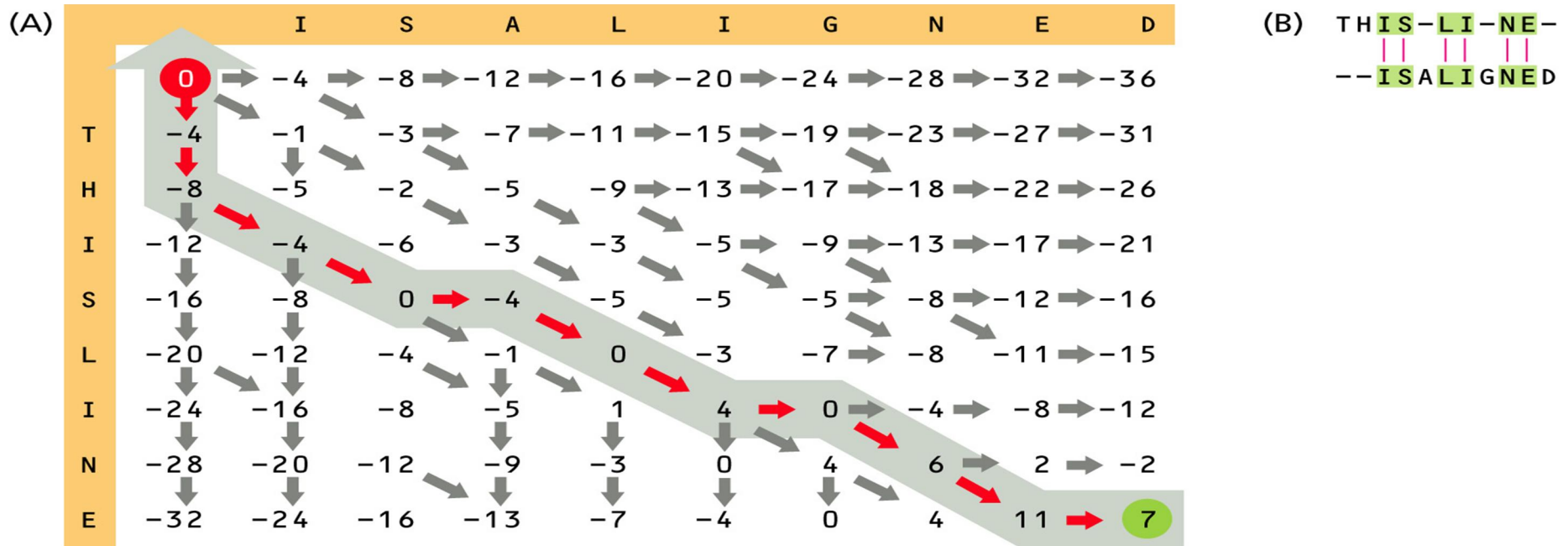
Cómo alineamos dos secuencias?

A mano o haciendo uso de métodos o algoritmos estadísticos



Métodos de Alineamientos

Globales (Needleman-Wunsch 1970): Alineamiento de secuencias completas, Maximiza el numero de correspondencias entre las secuencias para todo el largo de las mismas.



Locales (Smith-Waterman 1981): Alineamiento de subsecuencias, Es una modificación del algoritmo de programación dinámica, dándole mayor peso a correspondencias locales entre dos secuencias.

Resultados de un alineamientos

A: T C A G A C G A G T G

B: T C G G A G C T G

I T C A G A C G A G T G

T C G G A - - G C T G

II T C A G A C G A G T G

T C G G A - G C - T G

III T C A G A C G A G T G

T C G G A - G - C T G

Gap penalties

gap opening penalty = -5 / gap extension penalty = -1

1- Abrir un gap es costoso

GCTACTAG-T-T--CGC-T-TAGC
GCTACTAGCTCTAGCGCGTATAGC

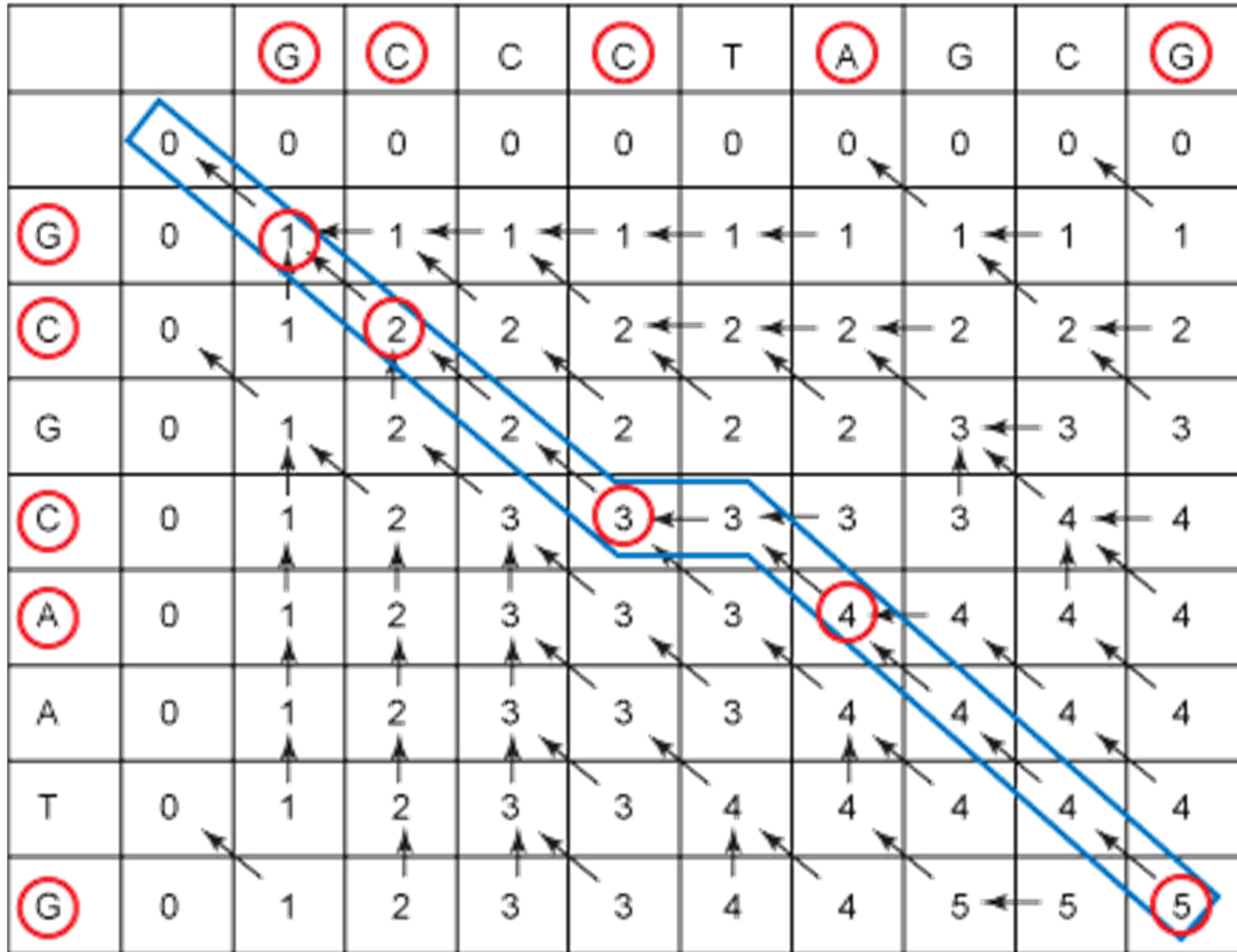
Penalty = 5 * (-5) + 6 * (-1) = -31

2 - Extender un gap es menos costoso

GCTACTAGTT-----CGCTTAGC
GCTACTAGCTCTAGCGCGTATAGC

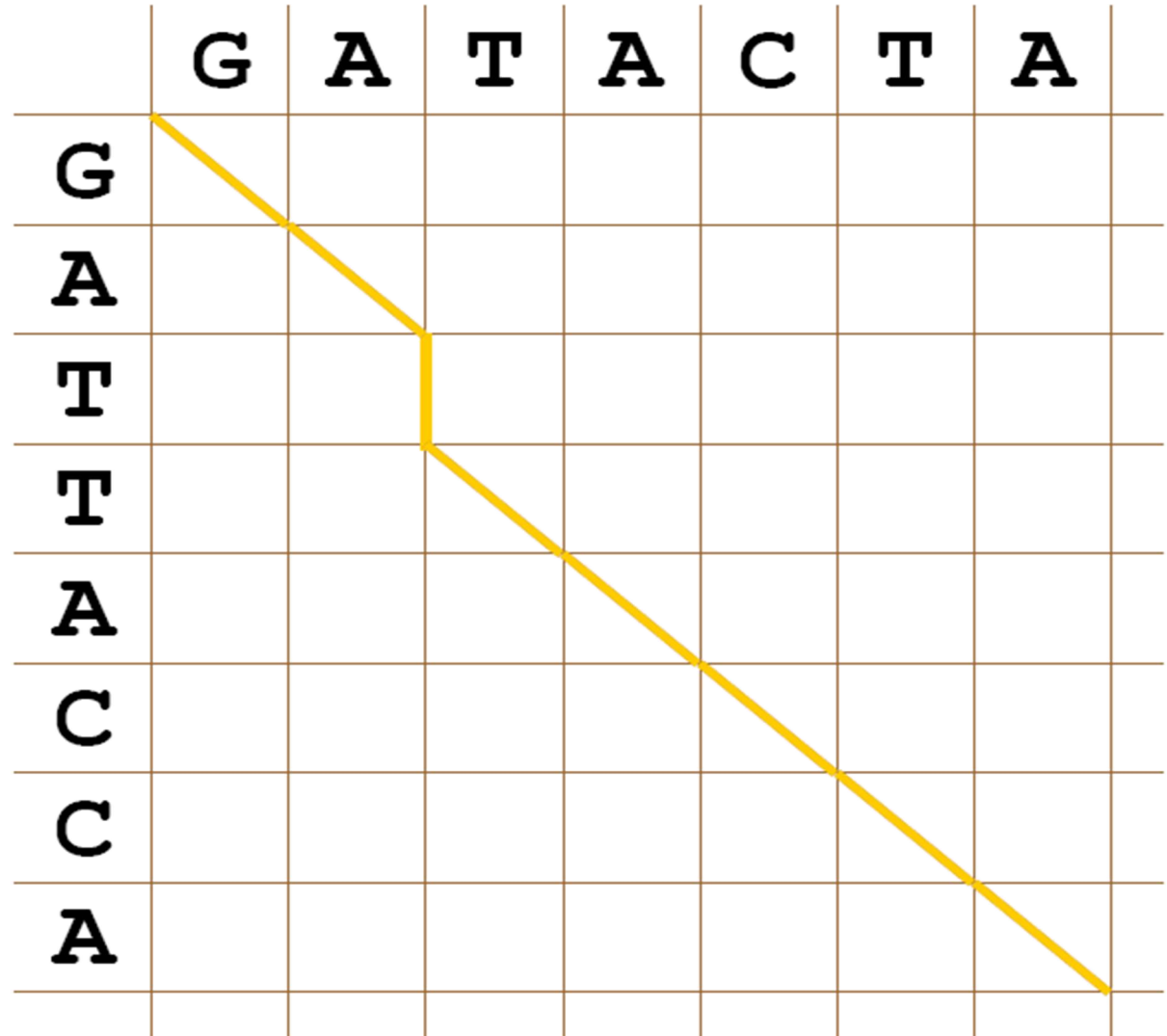
Penalty = 1 * (-5) + 6 * (-1) = -11

Algoritmos de programación dinámica



Imprimir el alineamiento

GA - TACTA
GATTACCA

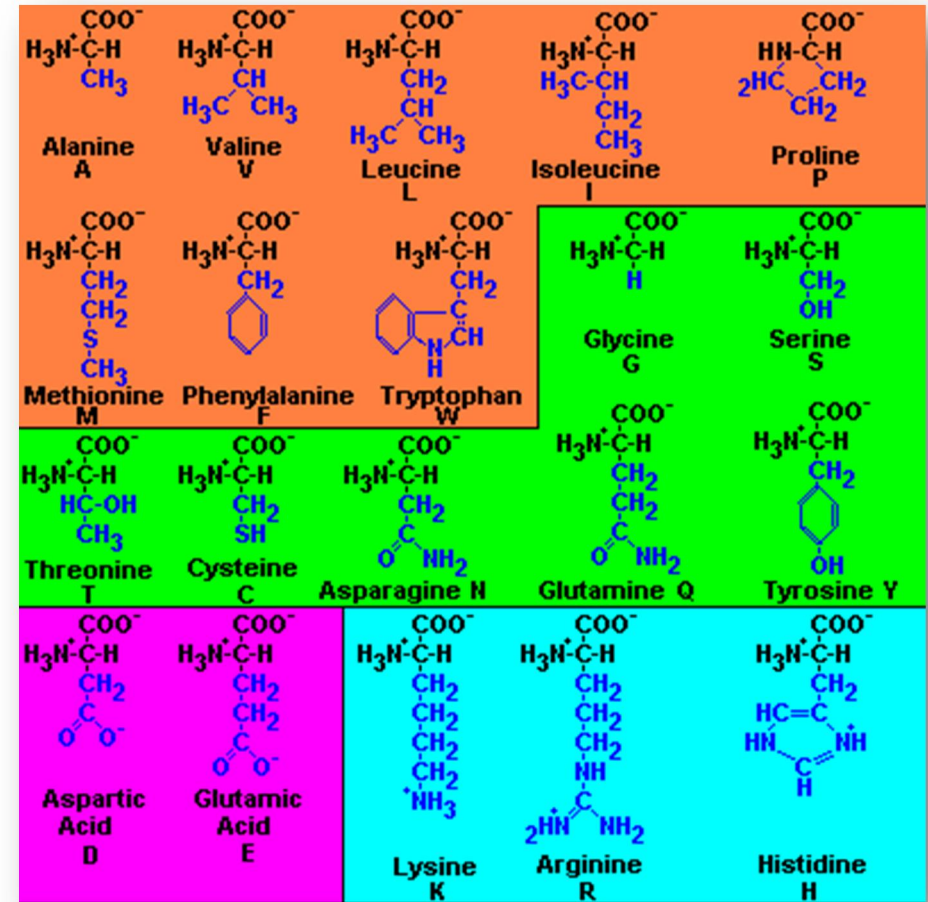


Matrices

Un ejemplo de matriz de scoring podría ser el clásico ejemplo de penalizar más los cambios que alteran las propiedades químicas de un residuo (aa)

- **Hidrofóbicos:** Ile, Val, Leu, Ala
- **Polares (+):** Lys, Arg, His
- **Polares (-):** Glu, Asp
- **Aromáticos:** Phe, Tyr, Trp .etc.

$$\begin{aligned} \text{Ile} \quad \times \quad \text{Val} &= -1 \\ \text{Ile} \quad \times \quad \text{Asp} &= -5 \\ \text{Phe} \quad \times \quad \text{Tyr} &= -1 \end{aligned}$$



Phe x Gly = -8

Matrices para Proteínas

- **PAM** - point accepted mutation based on *global* alignment [evolutionary model]

PAM 250 (80) - PAM 120 (66) PAM 90 (50)

% change



-
- **BLOSUM** - block substitutions based on *local* alignments [similarity among conserved sequences]

BLOSUM 30 - BLOSUM 62 - BLOSUM 80

% identity



Métodos de alineamientos

- Alineamiento progresivo
 - ClustalW
 - T-Coffee
 - PSAlign (semiprogresivo, mejor calidad)
- Métodos iterativos
 - PRRN/PRRP
 - CHAOS/DIALIGN
 - MUSCLE (*multiple sequence alignment by log-expectation*)
- *Hidden Márkov Models* (**Modelos de Márkov ocultos**)
 - POA (*Partial-Order Alignment*)
 - SAM (*Sequence Alignment and Modeling System*)
 - CASP (*Critical Assessment Tech for Prot Structure Prediction*)
 - HMMe

