

# Alineamientos de secuencias (Pairwise alignments).2

BCH441 - Bioinformatics  
Boris Steipe

[http://steipe.biochemistry.utoronto.ca/abc/index.php/Bioinformatics\\_Main\\_Page](http://steipe.biochemistry.utoronto.ca/abc/index.php/Bioinformatics_Main_Page)

<http://circos.ca/>



# Alineamiento de Secuencias

---

## Alineamiento global = Alineamiento óptimo

Los alineamientos globales basados en el algoritmo Needleman-Wunsch se utilizan cuando necesitamos que toda la secuencia sea alineada.

¿Qué hacemos cuando necesitamos encontrar fragmentos de la secuencia con alta similitud, comparar secuencias mucho más largas (e.g. genomas) o buscar en bases de datos?

# Alineamiento de Secuencias

---

## Alineamiento local

Smith-Waterman algorithm (1981):

Usado para buscar similitud entre dos secuencias altamente divergentes. Permite realizar alineamientos entre algunas secciones de la matriz y no ser obligado desde un principio al fin.

# Alineamiento de Secuencias

## Smith-Waterman algorithm (1981):

Solo hay dos diferencias con el alineamiento global:

1. Cualquier valor que sea menor que 0 es rellenado con 0.
2. Un valor igual a 0 significa que puede empezar otro alineamiento.

Initialize the scoring matrix

	T	G	T	T	A	C	G	G
G	0	0	0	0	0	0	0	0
G	0							
T	0							
T	0							
G	0							
A	0							
C	0							
T	0							
A	0							

$$\text{Substitution matrix: } S(a_i, b_j) = \begin{cases} +3, & a_i = b_j \\ -3, & a_i \neq b_j \end{cases}$$

$$\text{Gap penalty: } \begin{aligned} W_k &= kW_1 \\ W_1 &= 2 \end{aligned}$$

# Buenas prácticas en **Bioinformatics**

¿Cómo llevan a diario el seguimiento de sus experimentos de wet-lab?

¿Qué prácticas tienen regularmente en el laboratorio?

¿Cómo hacer esto mediante análisis bioinformáticos?

## 1. Anoten todo lo que hacen!!!

Organizar ideas; publicaciones con metodologías interesantes; programas, scripts y códigos usados; números de acceso y líneas de comando utilizadas!!



Remember everything important.



# Buenas prácticas en **Bioinformatics**

---

¿Cómo llevan a diario el seguimiento de sus experimentos de wet-lab?

¿Qué prácticas tienen regularmente en el laboratorio?

¿Cómo hacer esto mediante análisis bioinformáticos?

## 2. Reproducibilidad de resultados:

No lo hagan solo una vez! Repitan el experimento!, en general los programas poseen una variedad de parámetros o están basados en modelos probabilísticos que no entregan el exactamente el mismo resultado cada vez que es ejecutado.

Verificar errores

Mejorar sus códigos o los datos utilizados

Probar en diferentes condiciones (usar otros parámetros)

# Buenas prácticas en **Bioinformatics**

---

¿Cómo llevan a diario el seguimiento de sus experimentos de wet-lab?

¿Qué prácticas tienen regularmente en el laboratorio?

¿Cómo hacer esto mediante análisis bioinformáticos?

## 3. Diversidad de metodologías:

¿Existe más de una forma para corroborar lo que estoy estudiando? ¿Algunas de ellas tienen mayor validez que otras o son todas igual de válidas? La técnica que conocen no siempre va a ser la mejor. Actualizar, buscar y utilizar más de una metodología es necesario al estudiar un proceso biológico. Esto les dará más robustez a sus conclusiones.

(no utilicen solo el método que les dio el resultado que quieren).

# Buenas prácticas en **Bioinformatics**

---

¿Cómo llevan a diario el seguimiento de sus experimentos de wet-lab?

¿Qué prácticas tienen regularmente en el laboratorio?

¿Cómo hacer esto mediante análisis bioinformáticos?

## 4. Entiendan lo que hacen y utilizan:

No es necesario saber todos los modelos probabilísticos de fondo o los algoritmos utilizados, pero tengan una noción de por qué y cuáles son los motivos por los que utilizan ese programa.

(No usaré BLAST/MEGA porque es lo que todos hacen).



# Buenas prácticas en **Bioinformatics**

---

¿Cómo llevan a diario el seguimiento de sus experimentos de wet-lab?

¿Qué prácticas tienen regularmente en el laboratorio?

¿Cómo hacer esto mediante análisis bioinformáticos?

## 5. Conocer tus recursos (hardware/computador).

¿Puedo realizar un árbol filogenético de inferencia bayesiana de un alineamiento de 90 secuencias con un largo de 18.000 caracteres en mi PC? ¿Irá a usar toda la RAM? Deben entender las limitaciones de cada uno de sus PCs para entender que cosas pueden o no pueden hacer.

(y no pitiarse su PC).

# Buenas prácticas en **Bioinformatics**

---

¿Cómo llevan a diario el seguimiento de sus experimentos de wet-lab?

¿Qué prácticas tienen regularmente en el laboratorio?

¿Cómo hacer esto mediante análisis bioinformáticos?

## 6. Colaborar con la comunidad (open science).

Si encuentran una forma de hacer más eficiente un proceso bioinformático, crean un nuevo software o análisis de datos, publiquen sus resultados abiertamente con licencias de código abierto (open-source licenses). La comunidad bioinformática, al igual que el mundo de las ciencias en la computación, se basan en open-source softwares para facilitar la investigación a otros grupos relacionados (como Linux).

[\(colaboren con el mundo para que el mundo colabore con ustedes\).](#)

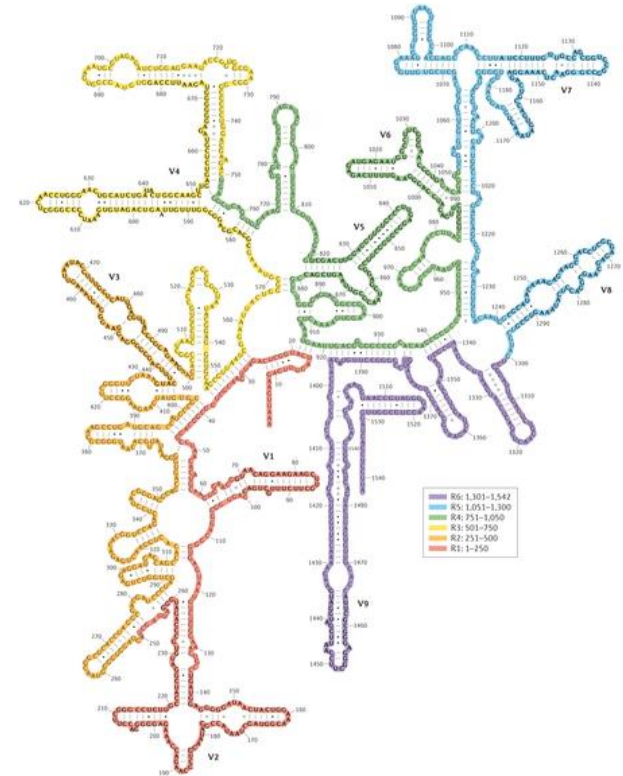
# Taller: Alineamientos de secuencias (PSA)

## 16S rRNA

1400-1600 bp de largo

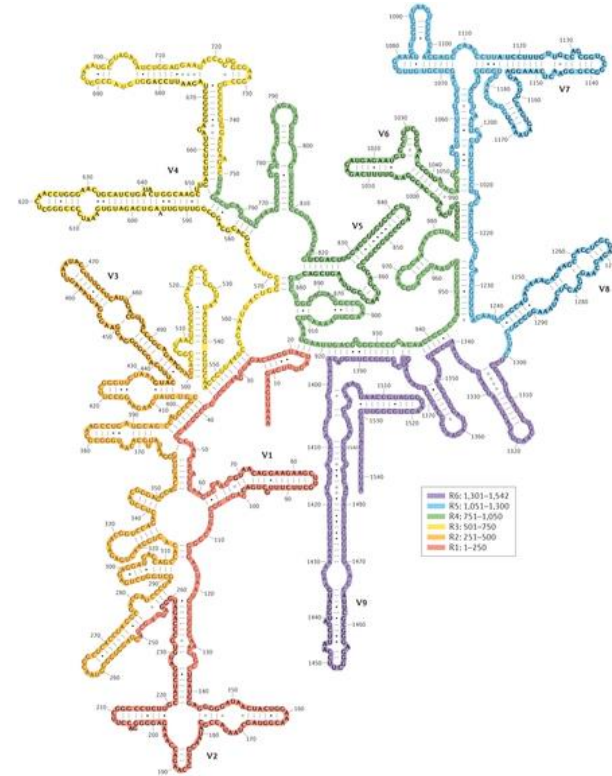
**Identificación de taxas  
bacterianas/archeas**

Ejercicio, fome pero informativo



# Taller: Alineamientos de secuencias (PSA)

1. Elegir dos bacterias o archeas.
2. Utilizar su secuencia en fasta como búsqueda en BLAST (blastn, usando como base de datos 16S rRNA)
3. Guardar 3 mejores resultados y sus porcentajes de identidad.
4. Extraer las 3 secuencias en fasta de esos resultados
5. Realizar un alineamiento global ([https://www.ebi.ac.uk/Tools/psa/emboss\\_needle/](https://www.ebi.ac.uk/Tools/psa/emboss_needle/)) entre la bacteria original y los tres resultados obtenidos.
6. Guardas las identidades obtenidas.
7. ¿Qué conclusiones pueden sacar de esto?



# Taller: Tareas para próximo taller

---

## Sistema operativo Windows/MacOS:

1. **Descargar e instalar máquina virtual (e.g. VirtualBox)**
2. **Descargar e instalar CONDA**
3. **Descargar e instalar BLAST local**
4. **Descargar e instalar MAFFT**

## Sistema operativo Linux:

1. **Descargar e instalar CONDA**
2. **Descargar e instalar BLAST local**
3. **Descargar e instalar MAFFT**