

# Alineamiento y modelos evolutivos

## Datos Moleculares II

Nelson R. Salinas

Septiembre 12, 2020

Nelson R. Salinas

[nrsalinas@gmail.com](mailto:nrsalinas@gmail.com)

<https://nrsalinas.github.io>

Instituto Alexander von Humboldt

Investigador Adjunto

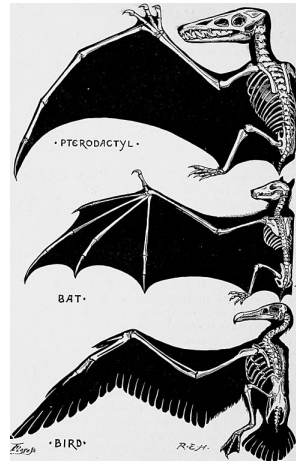
Repositorio del curso:

[https://github.com/nrsalinas/datos\\_moleculares\\_ii](https://github.com/nrsalinas/datos_moleculares_ii)

Allí encontrarán presentaciones, guías, conjuntos de datos, etc.

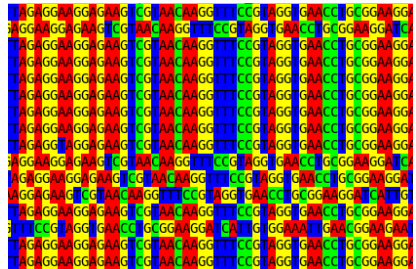
# Alineamiento

- Biología comparada requiere hipótesis de correspondencias entre los datos observados (=homologías).
- Necesario para cualquier clase de datos usados en sistemática filogenética: morfología, comportamiento, secuencias moleculares, señales acústicas, etc.



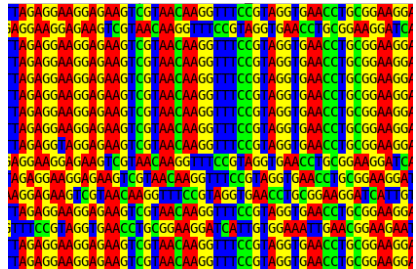
# Alineamiento

- En datos de secuencias (ADN, ARN, proteínas) estas hipótesis se establecen de una manera automatizada: alineamiento de secuencias múltiples.



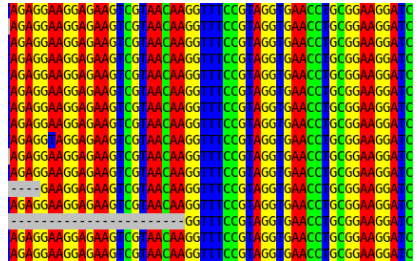
# Alineamiento

- En datos de secuencias (ADN, ARN, proteínas) estas hipótesis se establecen de una manera automatizada: alineamiento de secuencias múltiples.
- Alinear  $\rightarrow$  incluir espacios en blanco (*gaps*) en algunas posiciones de algunas secuencias.



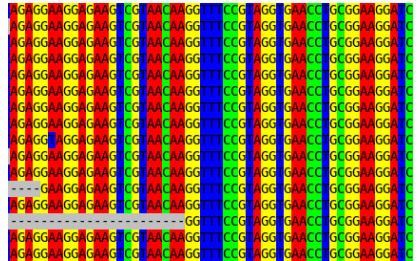
# Alineamiento

- En datos de secuencias (ADN, ARN, proteínas) estas hipótesis se establecen de una manera automatizada: alineamiento de secuencias múltiples.
- Alinear  $\rightarrow$  incluir espacios en blanco (*gaps*) en algunas posiciones de algunas secuencias.



# Alineamiento

- En datos de secuencias (ADN, ARN, proteínas) estas hipótesis se establecen de una manera automatizada: alineamiento de secuencias múltiples.
- Alinear  $\rightarrow$  incluir espacios en blanco (*gaps*) en algunas posiciones de algunas secuencias.
- Algunos sistemáticos realizan alineamientos de manera no automatizada ("al ojo"), pero entra en conflicto con método científico.



Computacionalmente, alinear secuencias es un un proceso intensivo.

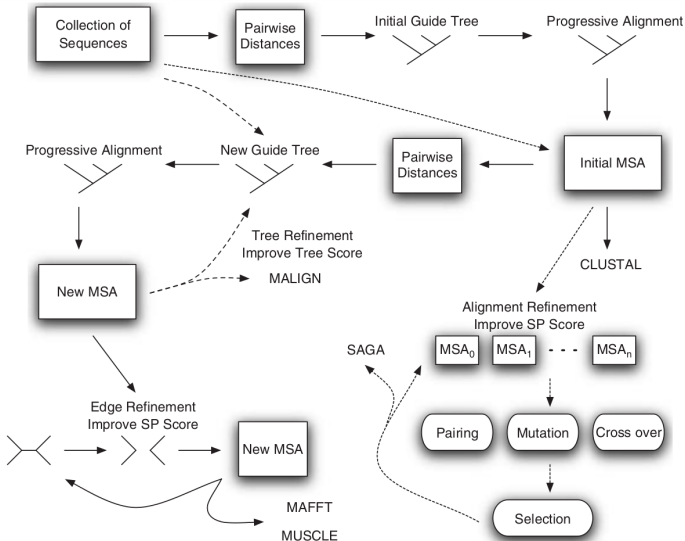
$n \backslash m$	2	3	4	5
1	3	13	75	541
2	13	409	23917	2244361
3	63	16081	10681263	14638756721
4	321	699121	5552351121	117629959485121
5	1683	32193253	3147728203035	$1.0 \times 10^{18}$
10	8097453	9850349744182729	$3.32 \times 10^{26}$	$1.35 \times 10^{38}$

Número posible de alineamientos para  $m$  número de secuencias cuya longitud es  $n$ .



# Alineamiento

## Principales programas de alineamiento progresivo



## **¿Existe una manera inequívoca de seleccionar el mejor alineamiento para un análisis?**

No.

Existen varios métodos para comparar alineamientos producidos por algoritmos distintos, se resumen en dos estrategias generales:

- 1 Similitud respecto a un alineamiento “verdadero”.
- 2 Comparación respecto a una función objetiva.

Probablemente la mejor alternativa es realizar una decisión informada en el efecto del alineamiento en el análisis posterior (v.g., reconstrucción filogenética).

¿Qué hacer con un alineamiento?

- Los alineamientos son medios para llegar a un fin.
- El fin es otro análisis.
- Son necesarios criterios para compartir alineamientos...
- ...Formatos de matrices.

- Un programa  $\rightarrow$  un formato.
- La mayoría emplean archivos legibles y pueden ser editados en cualquier editor de texto.
- Formatos binarios emplados en genómica (optimización de espacio).

## Fasta

```
1 >Mayaca_sellowiana_X0038
2 CTCCTGAGTATGAACTAAAGATACTGATATCTTGGCAGCAT
3 GCGGGAGCTGCAGTAGCTGCCGAATCTTCTACTGGTACATGG
4 TTACAAAGGGCGATGCTACCACATTGATCCTGTTCTGGGGA
5 ATCTTTTTGAAGAAGGTTTCAGTTACTAACATGTTTACTTCT
6 CTACGTCTGGAGGATTTGCGAATTCCTACTTCTTATTCAAAA
7 AGATAAATTGAACAAGTATGGTCGTCCTCTATTGGGATGTAC
8 GTAGGGCAGTTTATGAATGTTTACGTGGTGGACTTGATTTTA
9 CGTTGGAGAGACCGTTTCTTATTTTGTGCGGAAGCTATTTT
0 CTTGAATGCTACTGCGGGTACATGTGAAGAAATGCTTAAAG
1 TGCATGACTATTTAACTGGGGGATTCACTGCAAATACAAGTT
2 ATCCATCGCGCAATGCATGCAGTTATTGATAGACAAAAGAAT
```

Estandarizado por el programa de alineamiento Fasta.

### Nexus

```
1 #nexus
2 [Data processed on 2020 September 3 using 2matrix.pl -t Henri_alg.fasta -o n -n Henriettea
3 Indel characters coded using 2xread using the "simple gap coding" method of SIMMONS AND O
4 begin taxa;
5   dimensions ntax=16;
6   taxLabels
7     Henriettea_ekmanii
8     Henriettea_flavescens
9     Henriettea_gibberosa
10    Henriettea_macfadyenii
11    Henriettea_martiusii
12    Henriettea_multiflora
13    Henriettea_ramiflora
14    Henriettea_spruceana
15    Henriettea_squamata
16    Henriettea_squamulosa
17    Henriettea_uniflora
18    Henriettella_aggregata
19    Henriettella_fascicularis
20    Henriettella_odorata
21    Henriettella_rimosa
22    Henriettella_sessilifolia
23    ;
24 end;
25 begin characters;
26   dimenstons nchar=928;
27   format datatype=DNA missing=? gap=-;
28   charLabels
29     [1] sequence_0
30     [2] sequence_1
31     [3] sequence_2
```

Estandarizado por los programas de filogenética MacClade y  
PAUP\*.

## Nexus

Estandarizado por los programas de filogenética MacClade y PAUP\*.

El formato está resumido en un manual de 32 páginas que no es exhaustivo (Maddison et al. 1997. NEXUS: an extensible file format for systematic information. Systematic Biology 46: 590–621).

Otros programas lo han incorporado en su funcionamiento, pero usualmente con modificaciones no bien documentadas.

## Phylip

```
1 16 945
2 Henriettea_ekmanii TTTAGAGGAAGGAGAAGTCGTAAC
3 Henriettea_flavescens ---AGAGGAAGGAGAAGTCGTAAC
4 Henriettea_gibberosa TTTAGAGGAAGGAGAAGTCGTAAC
5 Henriettea_macfadyenii TTTAGAGGAAGGAGAAGTCGTAAC
6 Henriettea_martiusii TTTAGAGGAAGGAGAAGTCGTAAC
7 Henriettea_multiflora TTTAGAGGAAGGAGAAGTCGTAAC
8 Henriettea_raniflora TTTAGAGGAAGGAGAAGTCGTAAC
9 Henriettea_spruceana TTTAGAGGAAGGAGAAGTCGTAAC
10 Henriettea_squamata TTTAGAGGTAGGAGAAGTCGTAAC
11 Henriettea_squamulosa ---AGAGGAAGGAGAAGTCGTAAC
12 Henriettea_uniflora -TTAGAGGAAGGAGAAGTCGTAAC
13 Henriettella_aggregata -----GAAGGAGAAGTCGTAAC
14 Henriettella_fascicularis TTTAGAGGAAGGAGAAGTCGTAAC
15 Henriettella_odorata -----
16 Henriettella_rimosa TTTAGAGGAAGGAGAAGTCGTAAC
17 Henriettella_sessilifolia TTTAGAGGAAGGAGAAGTCGTAAC
```

Estandarizado por el paquete de programas de filogenética  
PHYLogeny Inference Package.