

Taller 11 - Filogenética

Grupo 04 - sección 02

Integrantes:

- Nicolas Montoya Leon - 202310678
- John Anderson Acosta - 202212004
- Raquel Bautista Escobar – 202310296

1. Obtención y preparación de secuencias.

- Para este ejercicio debe obtener el siguiente conjunto de secuencias del portal de GenBank: números de acceso HQ122450.1 a HQ122478.1. Puede acceder al grupo de secuencias mediante la opción PopSet en el servidor del NCBI. Descargue estas secuencias en formato FASTA (en un mismo documento).

```
>AP010847.1 Myroconger compressus mitochondrial DNA, complete genome
GTTAATGTAGCTTAATCACAAAGCATGGCACTGAAAATGCCAAGATGAACCGTAAACGTTCCAATAACA
TAAAAGCTTGGTCCTGACTTTAGCGTCAGTTATTACCCATTACACATGCAAGTCCCCGCAACCCCGTG
AGAATGCCCTCTACCCCCGGTAGGGAAAGAGGGGCGGCATCAGGCACGCCACAAGGTAGCCAAAACGC
CTTGCTCAGCCACACCCCCAAGGGAAAATCGCAGTGATAAACATTAAGCCATAAGCGAAAGCTCGACTTA
GTTACGGTAAGTTTAAGGGCCCGTAAAACCTCGTGCAGCCACCGCGGTTATACGAGAGGCCAAGTTGA
CATTCTACGGCGTAAAGCGTGATTAGAGAACCTCAATTCCACTAAAGCCAAACACCCCTTAAGCTGTCAT
ACGCTGCCAGGGCACACGAAGCCCCACAACAAAGTAGCTTTATTAAACCTGAACTCACGACAACATAAGA
AACAAACTGGGATTAGATAACCCACTATGCTTAGTCTAAACTATGGCGAAATAATACAAATATGCCCG
CCAGGGAACTACGAGCATTAGCTTAAACCCAAAGGACTTGACGGTGCCTCAAACCCGCTAGAGGAGCC
TGTTCTATAACCGATACTCCCCGTTAACCTCACCAACCCCTGTTAATTCCGCTATATACCACCGTCGC
CAGTTTACCTTATGAAAGAATAATAGTAAGCTAAAGGGCATTGCCAAAACGTCAGGTGAGGTGTAGC
GAATGGGGTGGCAAGAAATGGGCTACATTCTGAAAACAGAAATATTACGGAAAGTGTAAATGAAAATCGC
ACGACTGAAGGTGGATTAGCAGTAAAAAGGAAAAAGAGTGTCTTTGAAGCCGGCTTGAGGGCGTA
CACACCGCCCCGTCACTCTCCTCAAATTACCCCTAATAATTAAACAATACTGGAATAAAACAAAGAGGAGG
CAAGTCGTAAACATGGTAAGTGTACCGGAAGGGTGCACCTGGAAATAATTAGGACGTAGCTAATAGTAAGCAT
CTCCCTTACACCGAGAGGATCTCGTGCATCGAGACGTTCTAGCAAAATAGCTAGCCTGAATACAAA
CAGTAAAAAAATAAACATAATAATCAGAAAAAGTTAACAAACTAAAACATTTCACCCCAAGTATGGC
CACACAAACCAATTATCACACCCATACACAAACCTACCCCAACCTCAACACAAATCAATAAC
```

- ¿Cuántas secuencias son?

En total se trabajó con 29 secuencias de un gen mitocondrial, más el outgroup contando con la totalidad del ADN mitocondrial dando un total de 30 secuencias.

- ¿Qué tipo de secuencias (nucleótidos/proteínas), a qué organismos pertenecen las secuencias y a qué función están asociadas? ¿En qué genomas está ubicado este gen?

Todas las secuencias corresponden a nucleótidos pertenecientes a genes mitocondriales. Los organismos a quienes pertenecen las secuencias, son anguilas de la familia Muraenidae; dicho gen se encuentra en la subunidad 1 y su función se relaciona con la constitución de la citocromo oxidasa.

- Puede encontrar toda esta información en la base de datos PopSet.
- Enraíce el árbol mediante la búsqueda de una secuencia outgroup. Responda ¿Qué y por qué escogió esa secuencia como outgroup?

En este caso se escogió como outgroup a *Myroconger compressus* ya que esta pertenece a una familia diferente a la del resto de las anguilas cuyas secuencias se analizan (*Myrocongridae* y *Muraenidae*) pero manteniendo el mismo orden (Anguilliformes) para mantener un grado de similitud y comparabilidad.

2. Obtención y preparación de secuencias.

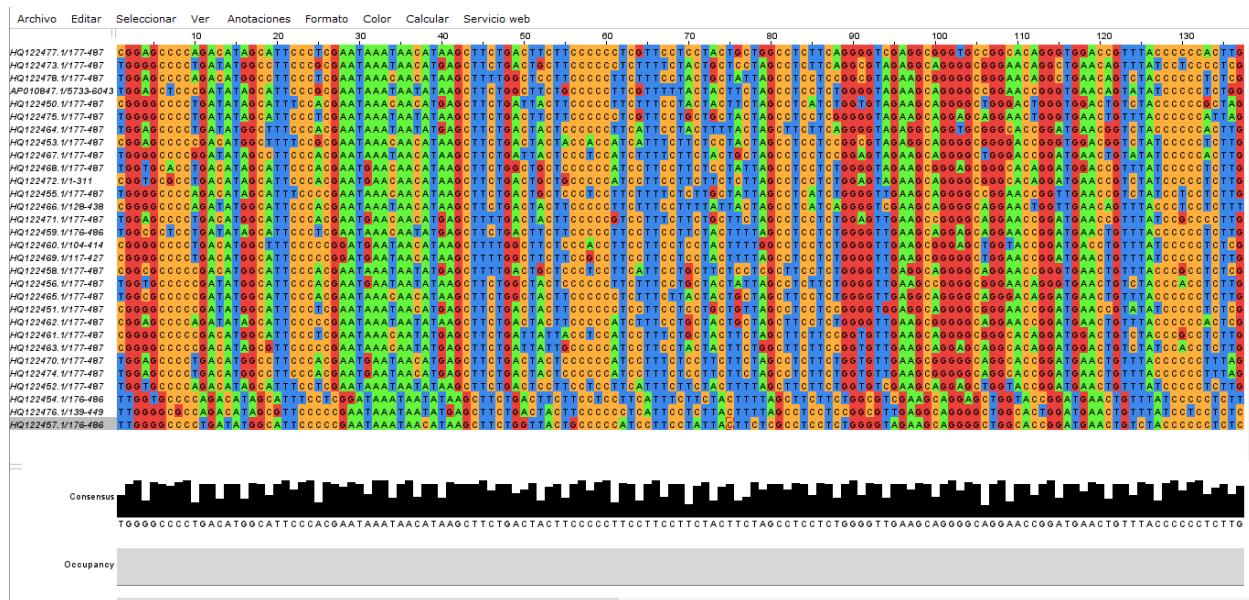
- En este punto antes de estimar una filogenia generamos un alineamiento que pueda reconocer el programa de reconstrucción. En nuestro caso, usaremos MUSCLE. Cargue el módulo de necesario en el cluster. Utilice el comando muscle para realizar el alineamiento múltiple con las secuencias del archivo fasta.
- ¿Cuál fue el comando completo que usó para generar el alineamiento múltiple?

```
muscle -in secuenciataller11.fasta -out alineado11.fasta
```

- Explique el principio de MUSCLE y en qué circunstancias suele aplicarse.

MUSCLE comienza alineando las secuencias en pares, lo cual le permite identificar las regiones mas homologas y así establecer relaciones de similitud y filogenéticas. El alineamiento se realiza de manera progresiva alineando primero las secuencias más similares y luego las menos similares. MUSCLE puede utilizarse para realizar alineamientos de secuencias múltiples en una variedad de contextos, desde la investigación filogenética hasta el diseño de herramientas moleculares para la biología molecular

- Realice la edición pertinentes del archivo de salida de MUSCLE en Aliview o Jalview. ¿Encontró posiciones con gaps para todas las secuencias? ¿Encontró alguna secuencia particularmente más larga o corta que las otras? ¿Dónde se encontraban? Reporte los cambios que usted decidió hacer al alineamiento.



En efecto se encontró una secuencia particularmente larga, que correspondió al outgroup; esto se debió a que cuando se descargó la secuencia de este, se descargó como la totalidad del genoma mitocondrial y por ende tenía una mayor extensión.

En la edición se recortaron todas las regiones que se extendían más allá del alineamiento, es decir, todas las regiones no correspondientes al gen citocromo oxidasa subunidad 1 y que no se alinearán con el resto de secuencias.

3. Reconstrucción de árboles NJ y Parsimonía.

- Calcule la matriz de distancias pareadas entre las secuencias utilizando la opción Distance. Según esta matriz, ¿cuáles son las especies más cercanas?

Las más cercanas serían la 10 y la 32 ya que tienen la misma distancia.

12	13	14	15	16	17	18	19	20	21	22	23
0.1829648505											
0.2103638247	0.2223400157										
0.2183319943	0.2265641660	0.1791112812									
0.2223400157	0.267522385	0.223400157	0.1868333273								
0.1907168273	0.2304045756	0.1868333273	0.1606030862	0.0528175558							
0.2103638247	0.2103638247	0.2103638247	0.1907168273	0.2143399730	0.2143399730						
0.1829648505	0.2064034235	0.2263641660	0.1946154677	0.2143399730	0.1829648505	0.1752725049					
0.1985293670	0.1946154677	0.2103638247	0.1636438098	0.2263641660	0.2103638247	0.1413709532	0.1525424558				
0.2024568452	0.2143399730	0.1791112812	0.1946154677	0.2143399730	0.1714484084	0.1823319943	0.1907168273	0.2064034235			
0.1668333273	0.1905693670	0.1791112812	0.14000959317	0.1946154677	0.1714464084	0.1668333273	0.1791112812	0.1408059317	0.1562966034		
0.2304045706	0.2385347016	0.2183319943	0.1668333273	0.2508551660	0.2304045756	0.1907168273	0.1907168273	0.214339730	0.1907168273	0.2024568452	
0.2304045756	0.2385347016	0.2143399730	0.1985293670	0.2549959787	0.2467314688	0.2385347016	0.2064034235	0.2263641660	0.1829648505	0.1829648505	0.108581130
0.1907168273	0.2344613762	0.1800630862	0.1676368800	0.2024568452	0.1907168273	0.2024568452	0.1907168273	0.1907168273	0.183319943	0.1791112812	0.194615467
0.1907168273	0.2344613762	0.1800630862	0.1676368800	0.2024568452	0.1907168273	0.2024568452	0.1907168273	0.1907168273	0.183319943	0.1791112812	0.194615467
0.1791112812	0.2143399730	0.2064034235	0.1791112812	0.1946154677	0.1829648505	0.2064034235	0.1905693670	0.2024568452	0.1985293670	0.1600630862	0.182964850
1.1650819337	1.2738847935	1.1754447207	1.02071934190	1.1754447207	1.1446730620	1.2511565424	1.2624560937	1.2071341909	1.2289334057	1.2971416557	1.2071341909
1.1650819337	1.2854456159	1.1148020989	1.1146730620	1.1148200989	1.1050639329	1.1964981299	1.2289334057	1.1754447207	1.1549254335	1.2180043351	1.1650819337
1.0954020130	1.2071934190	1.0576616851	1.1050639329	1.0576616851	1.0484450300	1.1346227262	1.1446730620	1.1346227262	1.0954020130	1.1754447207	1.085832562

- Construya un árbol de similitud por Neighbor Joining utilizando la opción Phylogeny.
(También puede crear un árbol de este tipo utilizando la herramienta TreeMaker).



- Construya un nuevo árbol ahora utilizando el método de Máxima Parsimonia utilizando la opción Phylogeny.



4. Reconstrucción de árbol de máxima verosimilitud.

- Utilice la opción Models para encontrar el mejor modelo de sustitución para construir el árbol filogenético a partir de su alineamiento. (También puede crear un árbol de este tipo utilizando la herramienta FindModel). ¿Cuál es el mejor modelo? ¿Por qué?

Rta: el mejor modelo sería Jukes-Cantor, es decir, el primer modelo. Podemos deducirlo porque tiene mayor AIC y menor LnL.

MODELS CONSIDERED:

Model name	AIC	LnL
JC: Jukes-Cantor (model 1)	17271.332806	-8635.666403
JC: Jukes-Cantor plus Gamma (model 3)	15902.79109	-7950.395545
F81: Felsenstein 1981 (model 5)	17203.440694	-8598.720347
F81: Felsenstein 1981 plus Gamma (model 7)	15808.103736	-7900.051868
K80: Kimura 2-parameter (model 9)	16790.134692	-8394.067346
K80: Kimura 2-parameter plus Gamma (model 11)	15381.002018	-7688.501009
HKY: Hasegawa-Kishino-Yano (model 13)	16729.096998	-8360.548499
HKY: Hasegawa-Kishino-Yano plus Gamma (model 15)	15313.760776	-7651.880388
TrN: Tamura-Nei (model 21)	16729.897082	-8359.948541
TrN: Tamura-Nei plus Gamma (model 23)	15314.963972	-7651.481986
GTR: General Time Reversible (model 53)	16671.012486	-8327.506243
GTR: General Time Reversible plus Gamma (model 55)	15297.942578	-7639.971289

Click one of the model names in the table to display the parameter matrix for that model.

Jukes-Cantor

	T	C	A	G
T	f_N	a	a	a
C	a	f_N	a	a
A	a	a	f_N	a
G	a	a	a	f_N

Jukes-Cantor + &gamma;

	T	C	A	G
T	f_N	a	a	a
C	a	f_N	a	a
A	a	a	f_N	a
G	a	a	a	f_N

Felsenstein 1981

	T	C	A	G
T	f_T	a	a	a
C	a	f_C	a	a
A	a	a	f_A	a
G	a	a	a	f_G

Felsenstein 1981 + &gamma;

	T	C	A	G
T	f_T	a	a	a
C	a	f_C	a	a
A	a	a	f_A	a
G	a	a	a	f_G

Kimura 2-parameter

	T	C	A	G
T	f_N	a	b	b
C	a	f_N	b	b
A	b	b	f_N	a
G	b	b	a	f_N

Kimura 2-parameter + &gamma;

	T	C	A	G
T	f_N	a	b	b
C	a	f_N	b	b
A	b	b	f_N	a
G	b	b	a	f_N

Hasegawa-Kishino-Yano

	T	C	A	G
T	f _T	a	b	b
C	a	f _C	b	b
A	b	b	f _A	a
G	b	b	a	f _G

Hasegawa-Kishino-Yano + &gamma

	T	C	A	G
T	f _T	a	b	b
C	a	f _C	b	b
A	b	b	f _A	a
G	b	b	a	f _G

Tamura-Nei

	T	C	A	G
T	f _T	a	b	b
C	a	f _C	b	b
A	b	b	f _A	c
G	b	b	c	f _G

General Time Reversible

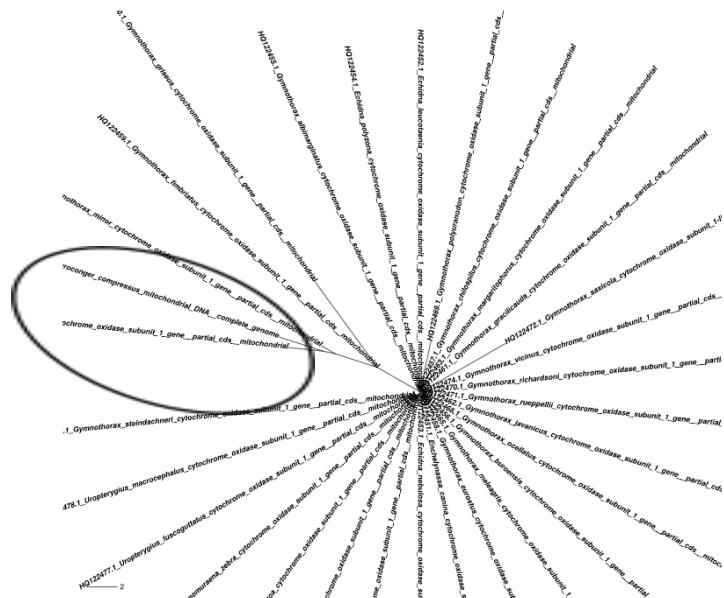
	T	C	A	G
T	f_T	a	b	c
C	a	f_C	d	e
A	b	d	f_A	f
G	c	e	f	f_G

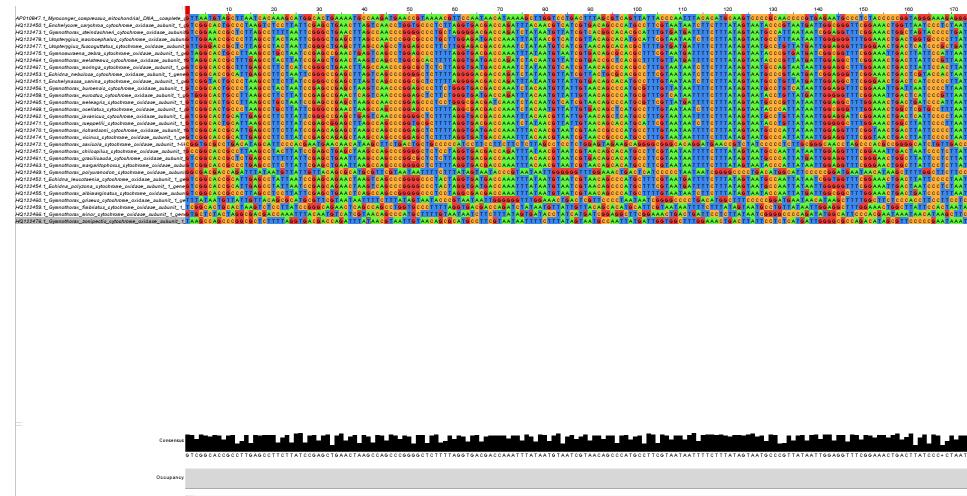
General Time Reversible + γ

	T	C	A	G
T	f_T	a	b	c
C	a	f_C	d	e
A	b	d	f_A	f
G	c	e	f	f_G

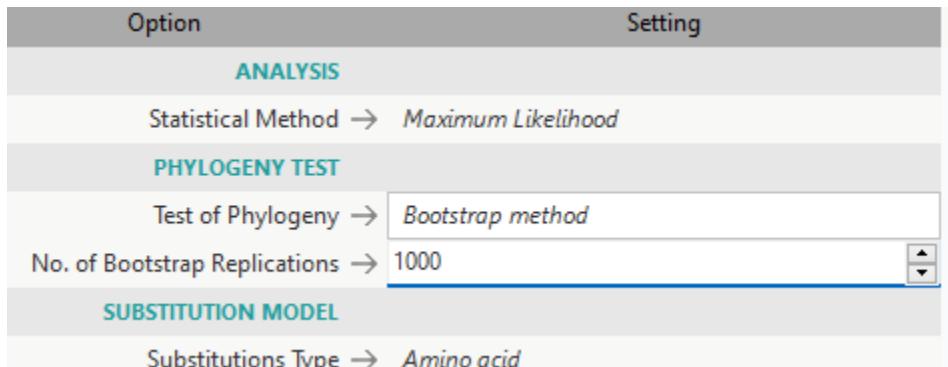
- Utilice la opción Phylogeny para construir un árbol mediante el método de Máxima Verosimilitud (deje la opción Test of Phylogeny en None) utilizando el mejor modelo de sustitución que encontró en el punto anterior (no olvide también modificar los parámetros del modelo Rates and Patterns). (También puede crear un árbol de este tipo utilizando la herramienta IQ-TREE). Modifique la visualización del árbol para mostrar el largo de las ramas. ¿El outgroup en el árbol resultó ser el mismo que usted escogió?

Rta: Teniendo en cuenta el árbol generado por IQ-TREE podemos afirmar que si es el mismo outgroup con el que se estaba trabajando.





- Construya nuevamente un árbol mediante el método de Máxima Verosimilitud (utilizando el mejor modelo de sustitución que encontró en el punto anterior), pero esta vez incluya soporte de bootstrap de 1000 corridas. Modifique la visualización del árbol para incluir la longitud de las ramas y el soporte de bootstrap.



- Ahora vamos a utilizar la línea de comandos para realizar todos los pasos que realizamos anteriormente. Cargue el alineamiento al clúster (en caso de que ya no lo haya hecho). Cargue el módulo iqtree/2.2.0_b (no se le olvide iniciar sesión interactiva). Utilicé el comando iqtree2 para construir un árbol utilizando el método de Máxima Verosimilitud con un soporte de bootstrap de 10000 repeticiones. Además también especifique el outgroup para enraizar el árbol. Note que por defecto, iqtree va a evaluar cuál es el mejor modelo de sustitución y lo va a utilizar. ¿Cuál fue el modelo de sustitución que se utilizó?
- Descargue el árbol obtenido en el punto anterior y carguelo en iTol (o en phylo.io, que es una herramienta más básica). En iTol se encuentran muchas herramientas tipo para

mejorar la representación gráfica de tu árbol, por ejemplo, cambiar la raíz, ordenar las ramas, cambiar la fuente de los tipos, incluir una escala, cambiar la forma de las ramas y la distancia vertical entre ramas, incluir medidas de soporte para cada rama, entre muchas otras opciones. Explore las diferentes opciones del programa y genere un archivo de imagen para presentar la filogenia estimada mostrando la longitud de las ramas y el soporte de bootstrap.

- Gymnomuraena y Echidna son géneros de peces cuyos miembros son muy parecidos en su morfología craneana y en su dentición. Teniendo en cuenta este criterio se ha propuesto agruparlos como un mismo género. Según los tres métodos utilizados para construcción de árboles que aprendimos ¿puede afirmar usted que estos datos soportan o rechazan el agrupamiento en un mismo género? ¿Por qué?

Rta: Para poder afirmar si los datos soportan o rechazan el agrupamiento en el mismo género se deben tener en cuenta 3 métodos de construcción de árboles filogenéticos.