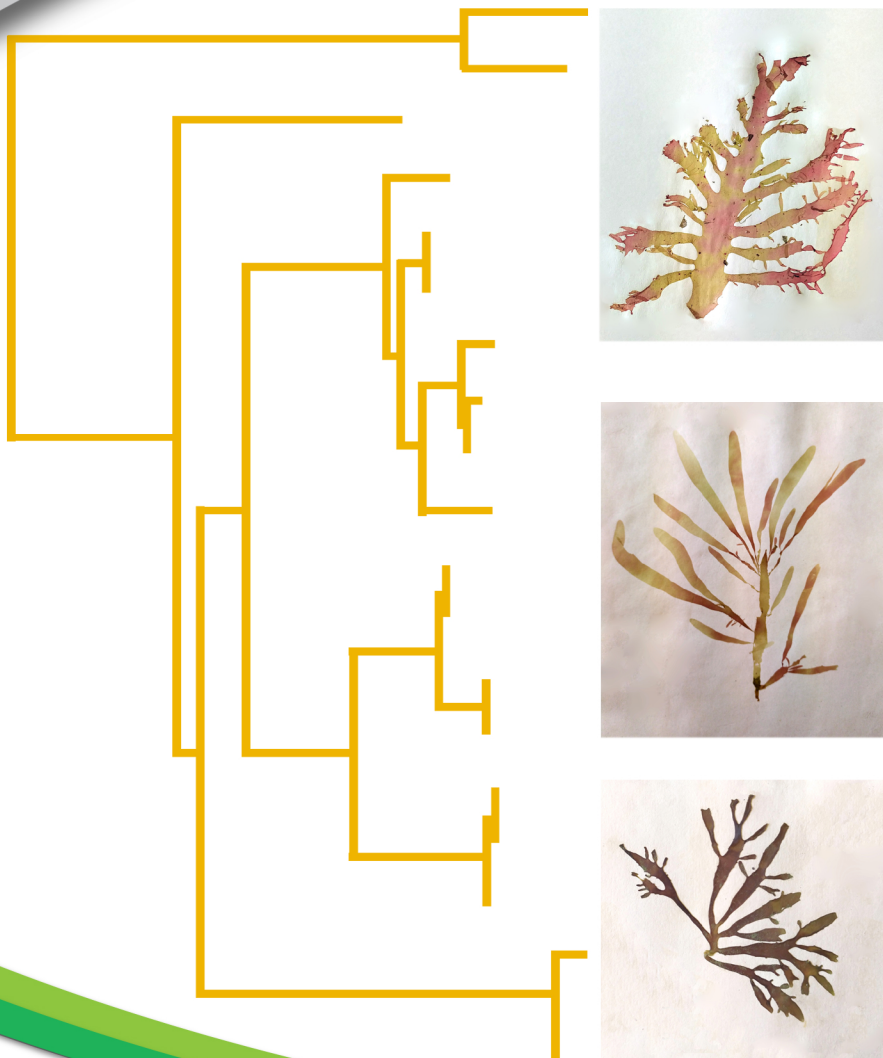




Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA
UNIDAD IZTAPALAPA

Sistemática Filogenética aplicada a la Ficología Marina



María Luisa **Núñez Resendiz**

Abel **Sentíes**

Kurt M. **Dreckmann**

María Esther
Meave del Castillo

María Eugenia
Zamudio Resendiz



UNIVERSIDAD AUTÓNOMA METROPOLITANA UNIDAD IZTAPALAPA

RECTOR GENERAL

Dr. Eduardo Abel Peñalosa Castro

SECRETARIO GENERAL

Dr. José Antonio De Los Reyes Heredia

UNIDAD IZTAPALAPA

RECTOR

Dr. Rodrigo Díaz Cruz

SECRETARIO

Dr. Andrés Francisco Estrada Alexanders

DIRECTORA DE LA DIVISIÓN DE C.B.S.

Dra. Sara Lucía Camargo Ricalde

COORDINADOR DE EXTENSIÓN UNIVERSITARIA

Mtro. Federico Bañuelos Bárcena

JEFE DE LA SECCIÓN DE PRODUCCIÓN EDITORIAL

Lic. Adrián Felipe Valencia Llamas

Primera edición 2019

ISBN: 978-607-28-1691-6

UNIVERSIDAD AUTÓNOMA METROPOLITANA UNIDAD IZTAPALAPA

Av. Michoacán y La Purísima

Iztapalapa, 09340. México, D. F.

Impreso y hecho en México/Printed in Mexico

Índice

Prólogo	5
<i>Software</i> requerido	6
Breves consideraciones para realizar estudios de sistemática filogenética con algas marinas: genomas y genes	7
Práctica 1. Edición y ensamble de secuencias nucleotídicas con BioEdit	9
Práctica 2. Construcción de matrices: bases de datos genómicos y alineamiento múltiple con BioEdit	19
Práctica 3. Estimación de distancias genéticas y modelos de evolución con MEGA-X	27
Práctica 4. Construcción de árboles filogenéticos I: UPGMA, Neighbor-Joining y Máxima Parsimonia con MEGA-X	41
Práctica 5. Construcción de árboles filogenéticos II: Máxima verosimilitud con RaxML	57
Práctica 6. Visualización y edición de árboles filogenéticos con FigTree	65
Práctica 7. Construcción de árboles filogenéticos III: Probabilidad Posterior Bayesiana con MrBayes	73
Práctica 8. Lectura e interpretación de un árbol filogenético	89
Glosario de términos	95
Bibliografía general	99

Prólogo

Esta obra surge como resultado de la necesidad de poner al alcance de los estudiantes y profesores de la división de CBS, el conocimiento y la práctica de los análisis de sistemática filogenética (SF), particularmente en grupos algales marinos. Se reconoce que este tipo de prácticas no son ex profeso para los grupos algales, porque pueden emplearse para los diversos grupos que conforman la biodiversidad del planeta.

La Sistemática Filogenética constituye una de las muchas disciplinas de la Biología Evolutiva (también conocida como Biología Sistemática o Comparada), y se encarga de analizar las relaciones de ancestría-descendencia entre los taxa incluidos en los sistemas de clasificación biológica (fundamentalmente las categorías genéricas -grupos de especies- y de familia -grupos de géneros-). Otras disciplinas son la Ecología Evolutiva, la Filogeografía (genética de poblaciones) y la Biogeografía (la dimensión espacial del proceso evolutivo). La SF es una disciplina incluyente, pues su objetivo cubre, con un mismo método y análisis, a Procariontes y Eucariontes, dado que opera en el marco de la teoría evolutiva. Los análisis filogenéticos, a través del entendimiento de los grupos monofiléticos (linajes que comparten un ancestro común inmediato) han dado lugar a avances sustanciales en el estudio de la biodiversidad, de la conservación de la misma, y de sus usos. Los análisis filogenéticos de un grupo económicamente importante, por ejemplo, han reducido significativamente, el tiempo del proceso ensayo-error en la selección del recurso (de la especie, realmente, importante económicamente). Esto porque las especies filogenéticamente relacionadas, sobre todo si son grupos hermanos (o sea, comparten un ancestro inmediato común), también comparten muchas otras características. De tal manera que, si una especie que se distribuye en Chile, y que es un recurso importante, y es filogenéticamente el taxon hermano de una especie presente en el litoral mexicano; entonces, la especie mexicana debe ser considerada como un recurso potencial.

Por otro lado, está el estudio de la biodiversidad en sí misma. La SF nos ha acercado meridianamente a un entendimiento más real, más actual de la diversidad biológica. Gracias a la SF, ahora manejamos conceptos evolutivos nuevos: especies crípticas (varios genotipos independientes usando un único fenotipo); y especies plásticas fenotípicamente (un mismo genotipo usando varios fenotipos). Esto ha dado lugar, por supuesto, a un incremento en la diversidad alfa.

La manera de abordar el tema de la SF en este manual, fue tanto metodológica como teórica. Las implicaciones del método de los análisis filogenéticos pueden estar sujetas al contexto teórico en el que la SF se origina y fortalece su filosofía, desde los conceptos hasta las aplicaciones en el universo de las clasificaciones biológicas. En este sentido, lo operativo se aterriza en los diferentes procesos que se aplican en los análisis filogenéticos. Cabe resaltar que la mayoría de dichos análisis dependen y son ejecutados por los sistemáticos (con o sin experiencia), con la ayuda computacional (sean programas en línea o de protocolos ya establecidos en el mercado). Hoy en día en la SF se utilizan métodos basados en parsimonia y probabilísticos, con una relevancia creciente de los caracteres moleculares.

Resultado de la experiencia de los autores en estudios de taxonomía alfa y sistemática molecular en los grupos de algas marinas, presentamos este manual que consiste de 8 prácticas, las cuales consideramos que reúnen los pasos a seguir para generar cladogramas o hipótesis filogenéticas, sin descuidar o estar conscientes de que la interpretación final de los análisis para cada grupo biológico estará en manos del(os) especialista(s) en el mismo.

Los protocolos moleculares de extracción, amplificación, purificación y secuenciación de ADN son procedimientos técnicos y rutinarios que, aunque fundamentales para la obtención de los datos (secuencias), no forman parte, propiamente, del análisis filogenético. Por esta razón, el orden de las prácticas aquí presentadas sigue estrictamente la secuencia del análisis. Así, después de la séptima práctica se cuenta con la más robusta hipótesis filogenética o cladograma, de tal modo que en la octava y última se aborda la interpretación del cladograma.

Es importante mencionar que todos los programas de cómputo que están involucrados en esta obra se encuentran de forma libre y gratuita en la WEB. Al no tratarse como laboratorio experimental, en esta obra no se presenta un reglamento estricto de laboratorio. En este sentido, el manual es operativo tanto en una sala de cómputo estudiantil como en un salón de clases con un video-proyector y la computadora del profesor.

Este manual está dirigido principalmente a estudiantes no graduados y graduados y puede ser utilizado para el trabajo de las siguientes unidades de enseñanza aprendizaje: Fundamentos de Sistemática, Evolución y Biogeografía de la licenciatura en Hidrobiología; así como Sistemática y Taxonomía y Sistemática Filogenética de la licenciatura en Biología y en la UEA Teoría Ecológica y Evolutiva I de la Maestría en Biología, todas ellas carreras de la División CBS en la UAM-I. Cabe resaltar que el trabajo de este tipo de prácticas se puede desarrollar parcialmente en temas netamente teóricos o con un enfoque de talleres, así pues, aunque las UEA no cuenten con un laboratorio experimental formal se puede hacer uso de las prácticas aquí planteadas. Por otro lado al ser un manual de prácticas, constituye también un libro de texto acerca de los análisis en Sistemática Filogenética.

Las secciones que aborda cada práctica son:

- INTRODUCCIÓN
- OBJETIVOS
- *SOFTWARE* REQUERIDO
- DESARROLLO
- RESULTADOS
- ACTIVIDADES COMPLEMENTARIAS
- BIBLIOGRAFÍA

Software requerido

1. BioEdit versión 7.0.5.3 disponible en <http://www.mbio.ncsu.edu/BioEdit/bioedit.html>
2. Base de datos del GenBank disponible en <https://www.ncbi.nlm.nih.gov/>
3. Mega-X Molecular Evolutionary Genetics Analysis disponible en <http://www.megasoftware.net>
4. RAxML v0.6.0 Randomized Axelerated Maximum Likelihood disponible en <https://raxml-ng.vital-it.ch/#/>
5. igTree v1.4.3 disponible en <http://tree.bio.ed.ac.uk/software/figtree/>
6. MrBayes 3.2.6 Bayesian Inference of Phylogeny disponible en <http://mrbayes.sourceforge.net/download.php>
7. DnaSP version 5.10.1 DNA Sequence Polymorphism disponible en http://www.ub.edu/dnasp/index_v5.html

Breves consideraciones para realizar estudios de Sistemática filogenética con algas marinas: genomas y genes

Las algas marinas son, en sentido amplio, un grupo artificial de eucariontes fotosintéticos que comparten niveles de organización y ambientes. Sin embargo, no forman un grupo monofilético, sino que los diferentes grupos provienen de diversos ancestros no algales. En tanto que las algas muestran una gran similitud morfológica y estructural, la sistemática en general de casi cualquier grupo de algas es compleja y controvertida. En la actualidad, la introducción de **marcadores moleculares** (genes) para el estudio de las relaciones evolutivas de los diferentes grupos de algas marinas se ha convertido en una condicionante, ya que estos marcadores no sólo proporcionan una fuente ilimitada de caracteres comparables no subjetivos (a diferencia de la morfología), sino que su grado de resolución de las filogenias ha dado buen resultado para esclarecer las relaciones evolutivas entre las algas a diferentes niveles jerárquicos.

Dada su condición de eucariontes fotosintéticos, las algas son organismos privilegiados que poseen tres fuentes de ADN o genomas: el del núcleo, el de la mitocondria y el del cloroplasto. Cada uno de estos genomas refleja una historia evolutiva particular, dado que para cada grupo de algas tienen un origen distinto. Por lo tanto, la elección de un marcador determinado debe realizarse cuidadosamente, considerando el nivel jerárquico en el cuál será comparado. De este modo, para el estudio de clados profundos (es decir, clados a niveles superiores como familia, orden, subclase, clase, etc.) se recomienda el uso de marcadores nucleares ya que el origen del núcleo es más antiguo en la historia evolutiva; sin embargo, si bien su ADN es conservado en ciertas regiones, en otras es altamente variable y casi imposible de comparar. En contraste, aunque los marcadores plastidiales son genes conservados, proporcionan variación suficiente y estable para comparar a nivel interespecífico; en el estudio de las macroalgas marinas, por ejemplo, estos marcadores han sido utilizados exitosamente para la discriminación de especies. Sin embargo, aunque los marcadores plastidiales son más conservados que los marcadores nucleares, no es posible utilizarlos para el estudio de clados profundos ya que su origen es polifilético, es decir, no tienen un origen común. Por otro lado, aunque la mitocondria fue adquirida en un evento evolutivo previo a la adquisición del cloroplasto, el surgimiento del organismo que dio origen a las mitocondrias actuales fue posterior al origen de los ancestros del cloroplasto (el metabolismo autótrofo permitió la evolución de la mitocondria); por lo tanto, el ADN mitocondrial es más joven que el ADN plastidial. Consecuentemente, para estudios de evolución temprana y estudios de variación genética en poblaciones, se recomienda el uso de genes mitocondriales.

Los principales marcadores que han sido utilizados en estudios filogenéticos de algas marinas se resumen en la Tabla 1.

Tabla 1. Principales marcadores utilizados en los estudios de filogenia de las algas marinas.

Grupo / genoma	Nucleares	Plastidiales	Mitocondriales
Microalgas	ITS ¹ , LSU ² , <i>psbO</i> ³ , SSU ⁴	<i>Ic1</i> ⁷ , <i>psbA</i> ⁸ , <i>psbC</i> ⁹ , <i>rbcL</i> ¹⁰	COB ¹² , COI-5P ¹³
Macroalgas	16S ⁵ , 18S ⁶ , LSU	<i>psbA</i> , <i>rbcL</i> , <i>tufA</i> ¹¹	<i>atp9</i> ¹⁴ , COI-5P, <i>Cox2</i> ¹⁵ , <i>Cox3</i> ¹⁶ , NADH ¹⁷

1. Espaciador del transcriptor interno del RNA (Internal Transcribed Spacer RNA); 2. Subunidad larga RNA ribosomal (Large Subunit ribosomal RNA); 3. Proteína potenciadora que evoluciona con el oxígeno (Oxygen-evolving enhancer protein 1); 4. Subunidad corta RNA ribosomal (Small Subunit ribosomal RNA); 5. 16S RNA ribosomal; 6. 18S RNA ribosomal; 7. Luciferasa; 8. Fotosistema II proteína D1; 9. Fotosistema II proteína 44 kDa; 10. Ribulosa-1,5-bisfosfato carboxilasa/oxigenasa subunidad larga; 11. Factor de elongación de Traducción Tu1 (Translation elongation factor Tu1); 12. Citocromo b; 13. Citocromo c oxidasa subunidad I; 14. ATPasa subunidad 9; 15. Citocromo oxidasa subunidad 2; 16. Citocromo oxidasa subunidad 3; 17. Dinucleótido de nicotinamida adenina

El incremento en las últimas décadas del número de estudios de filogenia para los diferentes grupos de algas, basados en diversos marcadores moleculares, ha conllevado a la creación de un acervo robusto de secuencias moleculares de especies de algas de todo el mundo disponibles en bases de datos públicas. A partir de estas secuencias, se han generado filogenias cada vez más robustas de los diferentes grupos. Por lo tanto, el empleo de marcadores moleculares ha jugado un papel prioritario en (a) el reconocimiento de taxa nuevos y de diversidad críptica, con el consecuente incremento de la diversidad alfa, (b) el reconocimiento de complejos morfológicos en especies morfológicamente monótonas, (c) el reconocimiento de especies auténticamente plásticas, (d) propuestas de nuevas combinaciones taxonómicas, (e) el restablecimiento de especies, (f) rearrreglos taxonómicos, (g) la identificación de los pro-

cesos evolutivos responsables de los principales eventos de especiación, (h) la detección de problemáticas taxonómicas por resolver y (i) la identificación de patrones biogeográficos que permitan hacer propuestas de conservación, determinar áreas de endemismo y correlacionar a las áreas con la evolución de las especies.

Práctica 1

Edición y ensamble de secuencias nucleotídicas con BioEdit

INTRODUCCIÓN

Una secuencia nucleotídica es una sucesión de al menos tres y hasta varios millones de nucleótidos (representados por bases nitrogenadas) covalentemente ligados. Las bases nitrogenadas se clasifican en (1) **bases púricas (o purinas)** que son Adenina (A) y Guanina (G), y en (2) **bases pirimídicas** (o pirimidinas) que son Timina (T) (o Uracilo (U) para el caso del ARN) y Citosina (C). En las moléculas de ADN, tanto purinas como pirimidinas se aparean. El **apareamiento de bases** se refiere a la interacción entre bases nitrogenadas que da origen al plegamiento de los ácidos nucleicos formando el ADN. Las interacciones entre las bases se dan a través de puentes de hidrógeno entre regiones específicas. Dicho apareamiento se da entre G-C (guanina-citosina) y A-T (adenina-timina), dando como consecuencia la estructura de doble hélice formada por dos cadenas antiparalelas. De este modo, la **cadena líder o molde** (también denominada **forward**) correrá en dirección 5' a 3', mientras que, la cadena complementaria (también denominada **reverse**) correrá en el sentido contrario 3' a 5'. Por lo tanto, cuando amplificamos secuencias de ADN para una identificación molecular, obtendremos dos secuencias en formato **AB1** (archivos de secuencia que incluyen, además de la secuencia de nucleótidos en formato fasta, un electroferograma de las bases que conforman a la secuencia), una forward y una reverse. Durante la edición de secuencias, ambas cadenas deben ser ensambladas a fin de obtener una **secuencia consenso** que sea comparable con otras secuencias y permita la identificación molecular. La secuencia consenso será entonces el resultado del ensamble entre forward y reverse.

Editar una secuencia nucleotídica es indispensable para:

- Detectar y eliminar las zonas de baja calidad en la secuencia.
- Detectar secuencias de **primer**.
- Corregir las ambigüedades mediante la comparación de los cromatogramas forward y reverse.
- Obtener una secuencia limpia.

En una secuencia de ADN, puede haber regiones codificantes (exones) o no codificantes (intrones). Las regiones codificantes están organizadas en tripletes de bases (codones que codifican para un aminoácido) que juntos codifican para sintetizar una proteína (conjunto de aminoácidos), por lo que la correcta edición de una secuencia es indispensable para no alterar el orden de los aminoácidos que componen a dichas proteínas.

Existen múltiples *softwares* sumamente amigables para la edición de secuencias, desafortunadamente no son libres. Sin embargo, hay algunos *softwares* libres que pueden ser descargados en línea y que permiten la edición de secuencias. Estos son Clustal X, MEGA, Mesquite o BioEdit.

OBJETIVO

- Obtener una secuencia consenso a partir de dos secuencias complementarias (forward y reverse) en formato AB1, utilizando como herramienta el programa BioEdit.

SOFTWARE REQUERIDO

1. BioEdit versión 7.0.5.3, disponible para su descarga en el sitio del programa, siguiendo la liga <http://www.mbio.ncsu.edu/BioEdit/bioedit.html>

Nota: asegúrate de descargar la versión compatible con el sistema operativo de tu computadora. Una vez descargado el archivo zip, da clic en el archivo ejecutable y sigue las instrucciones para la instalación del programa.

DESARROLLO

1. Abre el programa BioEdit, aparecerá una ventana en negro con una barra principal de herramientas, ubicada en la parte superior izquierda.
2. Para abrir las secuencias que vas a editar, da clic en el menú **File**; al hacerlo será desplegado un nuevo submenú, selecciona la opción **open** (Fig. 1). Te abrirá una ventana de diálogo que te permitirá buscar entre tus archivos el par de secuencias complementarias que deseas editar. Del directorio Prácticas de Filogenia, abre la subcarpeta Práctica 1, en ella encontrarás 10 secuencias (5 forward y 5 reverse) en formato **ABI**. Selecciona primero la secuencia **Reverse** nombrada **1_RrbcsStart** y presiona el botón **abrir**.

Nota: se recomienda que el proceso de edición de secuencias sea individual, es decir, un par de secuencias complementarias, forward y reverse, a la vez, no un conjunto grande de secuencias en una misma edición.

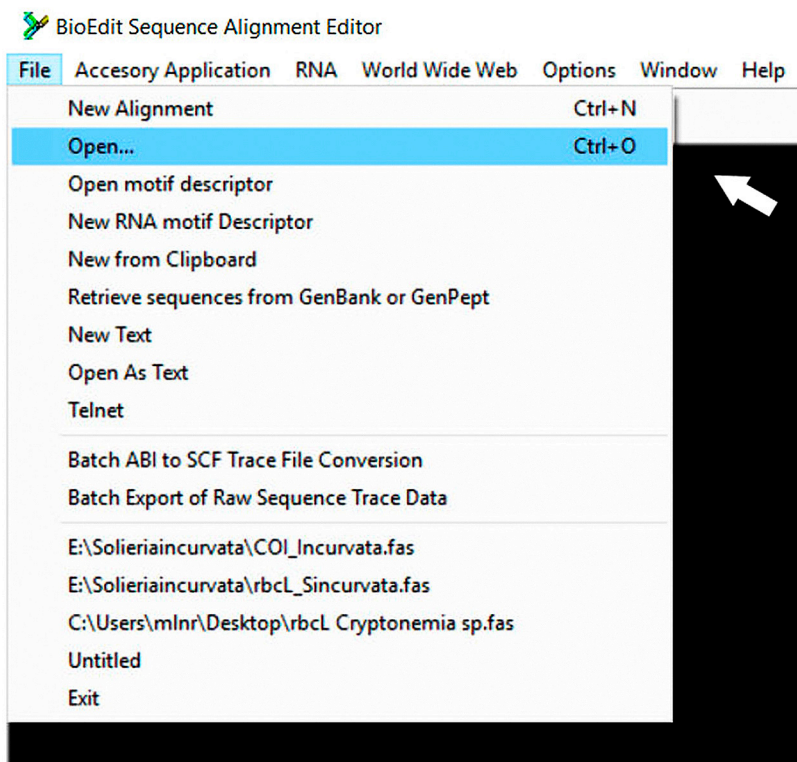


Fig. 1. Captura de pantalla paso 2: menú principal de herramientas del programa

3. En la pantalla negra del programa, verás aparecer un par de cuadros, uno dividido en dos cajas, la de la izquierda con el nombre de la secuencia y la caja de la derecha con una secuencia de bases (Fig. 2a, flecha), y otro cuadro con un cromatograma (Fig. 2b, flecha).

Nota: Este cromatograma corresponde a tu secuencia de nucleótidos y te mostrará tanto la frecuencia de tus bases como la calidad de tu secuencia. Cuando una secuencia es de baja calidad, los picos del cromatograma no tendrán una forma curva y estarán encimados. Una secuencia con calidad será aquella que sólo te muestre un cromatograma con picos muy curvos y sin otros picos encimados (Fig. 2b). A lo largo de tu cromatograma notarás zonas donde los picos están muy bien definidos y otras donde parecen muy sucios. Normalmente esto coincide con los extremos, contrarios para forward y reverse. De esta manera, cuando los picos en la cadena líder están sucios y, por tanto, presentan mayor ambigüedad, los picos en reverse comenzarán a estar limpios y permitirán reducir dicha ambigüedad y viceversa.



Fig. 2. Captura de pantalla paso 3: archivos de secuencias en formato AB1, (a) secuencia de nucleótidos y (b) cromatograma

4. Minimiza el cuadro con el cromatograma (no lo cierras) y sólo mantén abierto el cuadro con las dos cajas. Da clic sobre el nombre de tu secuencia (1_RrbcsStart) en la caja izquierda de tu cuadro. Al hacerlo, el nombre se pondrá en color negro (Fig. 3).

Nota: Dado que forward y reverse corren en diferentes direcciones ($5'-3'$ y $3'-5'$, respectivamente), las bases no son complementarias por lo tanto no pueden aparearse. Para poder ensamblarlas, necesitamos que ambas secuencias corran en la misma dirección ($5'-3'$). Adicionalmente, dado que las bases son complementarias, cuando en la cadena forward encontremos una T, en reverse encontraremos una A; por lo tanto, debemos obtener la cadena complementaria de reverse, así cuando en forward encontremos una T, en reverse, en lugar de encontrar la base complementaria (A), encontraremos el complemento de reverse para esa base, que será T, igual que en forward. De este modo ambas secuencias serán completamente comparables.

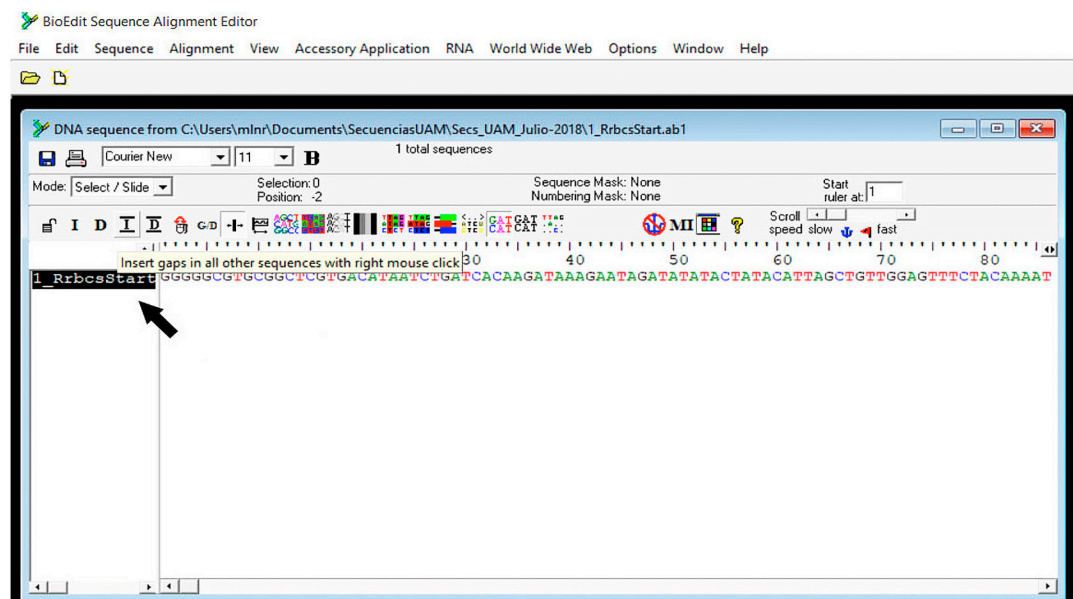


Fig. 3. Captura de pantalla paso 4: selección de secuencias

- Para obtener una cadena inversa complementaria de reverse, selecciona de la barra principal del programa el menú **Sequence**, al hacerlo será desplegado un nuevo submenú con varias opciones, selecciona la opción **Nucleic Acid**. De la nueva ventana de sub-submenú que te mostrará, selecciona la opción **Reverse Complement** (Fig. 4).

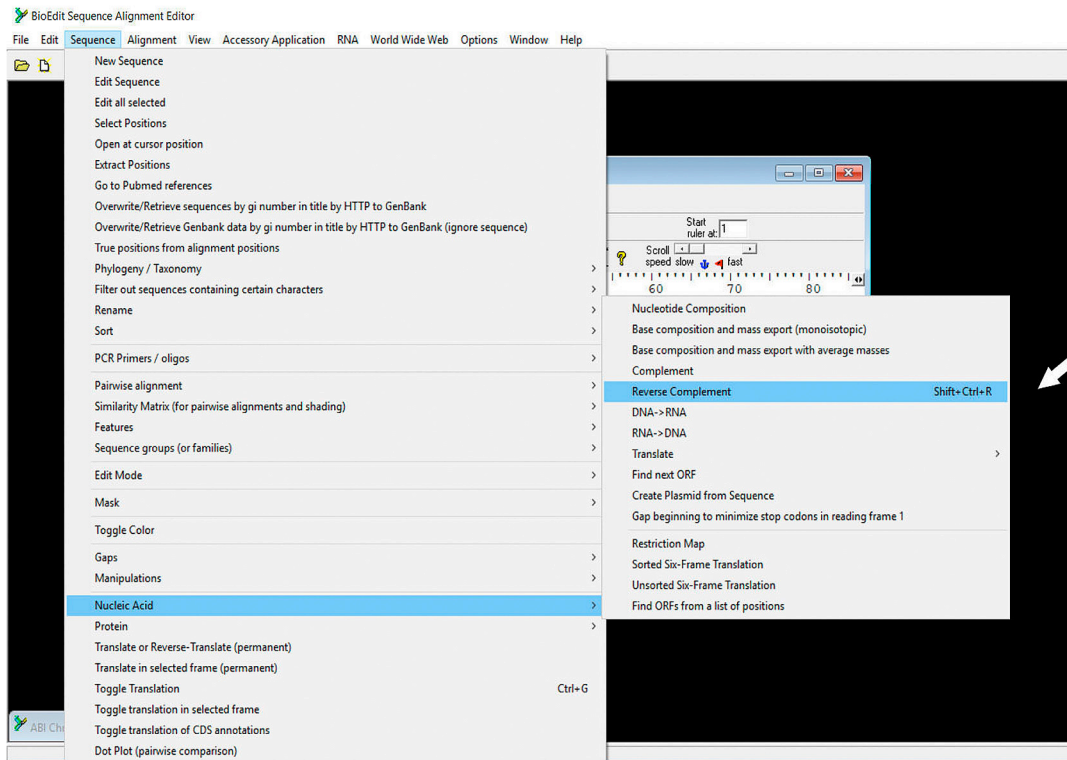


Fig. 4. Captura de pantalla paso 5: Cadena reversa complementaria en una cadena de nucleótidos

- Ahora minimiza el cuadro con las dos cajas en el que está la secuencia (no lo cierres) y maximiza únicamente el cuadro del cromatograma. En la barra principal, selecciona el menú **View**. Al hacerlo te desplegará un submenú, da clic en la opción **Reverse complement** (Fig. 5). Notarás que tanto tu secuencia de nucleótidos de la caja derecha del cuadro de secuencia como la secuencia en tu cromatograma habrán cambiado, ahora estarán mostrando la cadena complementaria inversa de reverse. Ahora minimiza también el cuadro del cromatograma.

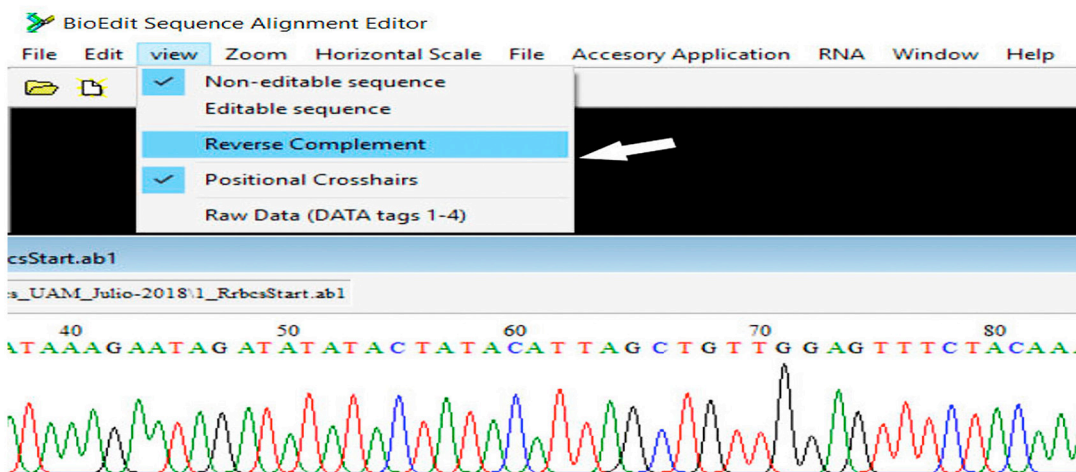


Fig. 5. Captura de pantalla paso 6: Cadena reversa complementaria de un cromatograma

7. Sin cerrar ninguno de los dos cuadros de la secuencia 1_RrbcsStart, abre ahora la secuencia forward nombrada **1_FrbcsStart**. Como con reverse, igualmente verás para forward el cuadro de la secuencia y el cuadro del cromatograma. Como forward es la cadena líder, no necesitas obtener el complemento inverso, sin embargo, para poder ensamblar las secuencias, necesitarás tener, en un mismo cuadro de nucleótidos, ambas secuencias.
8. Minimiza el cromatograma de 1_FrbcsStart. Del cuadro de la secuencia que quedó maximizado, vamos a copiar la secuencia de nucleótidos. Para ello, selecciona la primera base de tu secuencia, después, con ayuda de la barra espaciadora que se encuentra al inferior del cuadro de secuencia (Fig. 6a, flecha), recorre la secuencia hasta la última base. Presionando el botón shift de tu computadora, da clic sobre la última base. La secuencia aparecerá resaltada en negro (Fig. 6b, flecha).

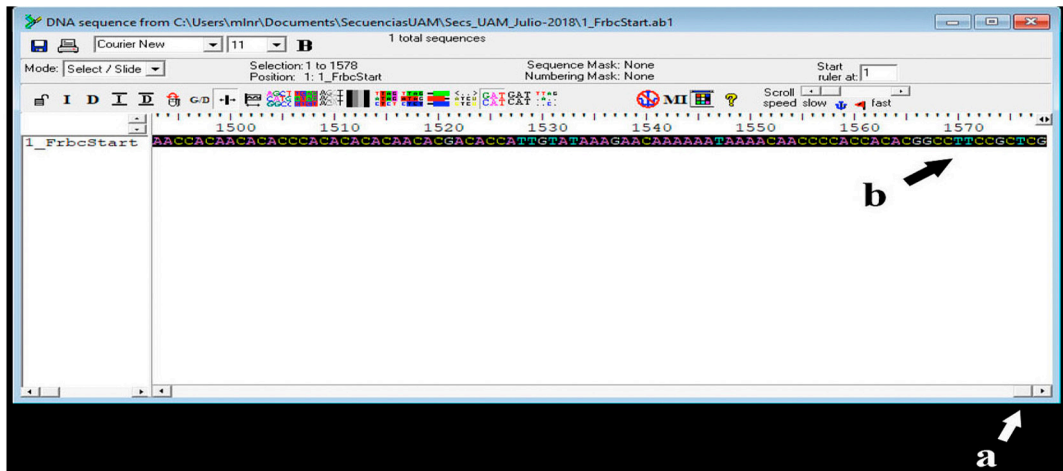


Fig. 6. Captura de pantalla paso 8: (a) barra espaciadora y (b) selección de secuencias

9. En la barra principal, selecciona el menú **Edit**, del submenú que desplegará, selecciona la opción **Copy**.
10. Minimiza el cuadro de la secuencia de forward y maximiza el cuadro de la secuencia de reverse.
11. En la barra principal, selecciona el menú **Sequence**, del submenú que te mostrará, selecciona la opción **New sequence**. Te abrirá un nuevo cuadro que contendrá tres recuadros en blanco, dos chicos y una central muy grande (Fig. 7).

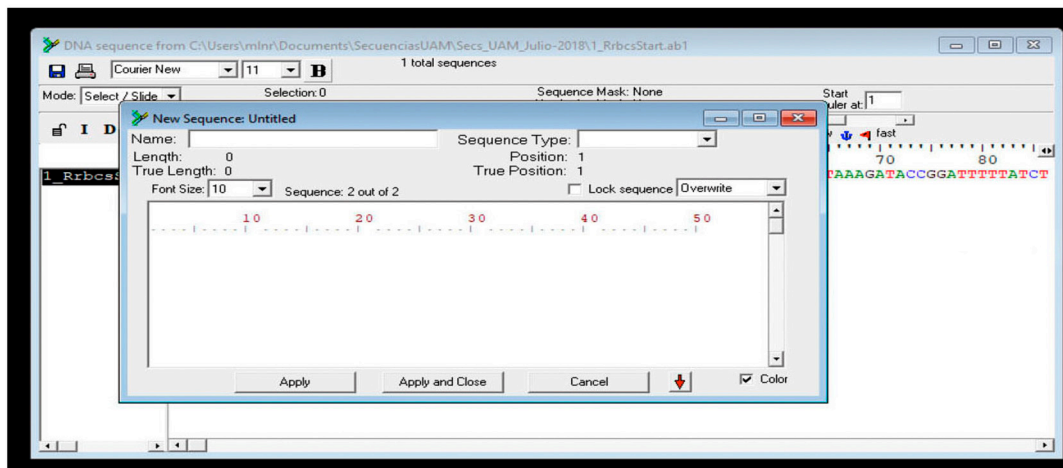


Fig. 7. Captura de pantalla paso 11: cuadro de diálogo para la adición de una nueva secuencia a un archivo preexistente

12. Posiciona el cursor sobre el recuadro más grande, haciendo clic sobre él con el mouse. De la barra principal, selecciona nuevamente el menú **Edit**, y del submenú la opción **Paste**. Notarás que la secuencia que copiaste de forward ahora está en el recuadro grande (Fig. 8a, flecha). De los dos recuadros superiores, el del lado izquierdo te pide poner el nombre de la secuencia, copia en él `1_FrbcStart` (Fig. 8b); el recuadro del lado derecho te pregunta sobre el tipo de secuencia que estás copiando, con la flecha despliega el menú de datos y selecciona la opción **DNA** (Fig. 8c, flecha). Una vez que hayas llenado los recuadros con la información requerida, presiona la opción **Apply** (Fig. 8d). Notarás que ahora ambas secuencias están en un mismo recuadro (Fig. 8e).

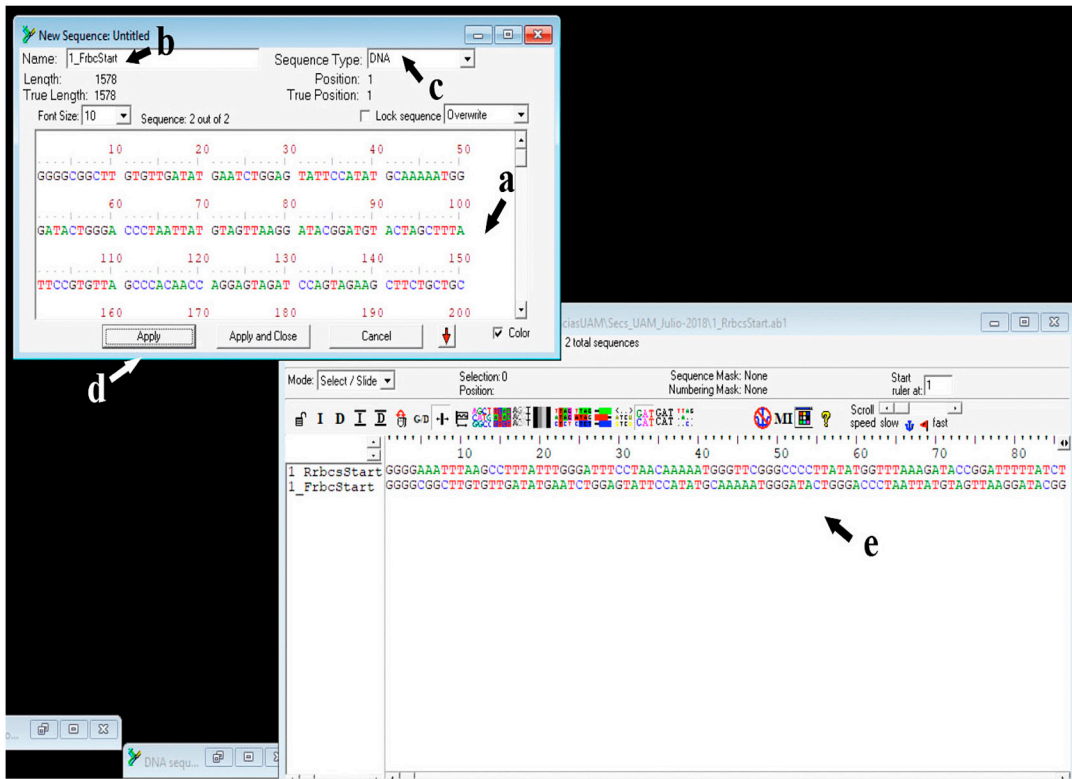


Fig. 8. Captura de pantalla paso 12: (a) cuadro para agregar una nueva secuencia, (b) cuadro para nombrar a la secuencia, (c) selección del tipo de datos, (d) aplicación a un archivo preexistente y (e) archivo de secuencias conjuntas

13. Cierra el cuadro en donde pegaste la secuencia de forward para generar una nueva en la ventana de reverse y el cuadro original de la secuencia forward, puesto que ésta ya se encuentra en el mismo archivo que reverse. Sólo conserva abiertos los dos cromatogramas y el cuadro con ambas secuencias.
14. Para ensamblarlas, selecciona ambas secuencias dando clic sobre sus nombres en la caja del lado izquierdo, como lo hiciste en el paso 4 (Fig. 3). De la barra principal selecciona el menú **Accessory Application**, del submenú que te mostrará selecciona la opción **Clustal W Multiple Alignment**.
15. Te mostrará un nuevo cuadro con diversas opciones. Considera los valores que el programa asigna a los recuadros por default y da clic en la opción **Run ClustalW** (Fig. 9). En el nuevo cuadro que te mostrará presiona **OK**.

Nota: Clustal W alineará tu par de secuencias y hará empatar las bases apareándolas. Sin embargo, en algunos casos, las bases no van a coincidir.

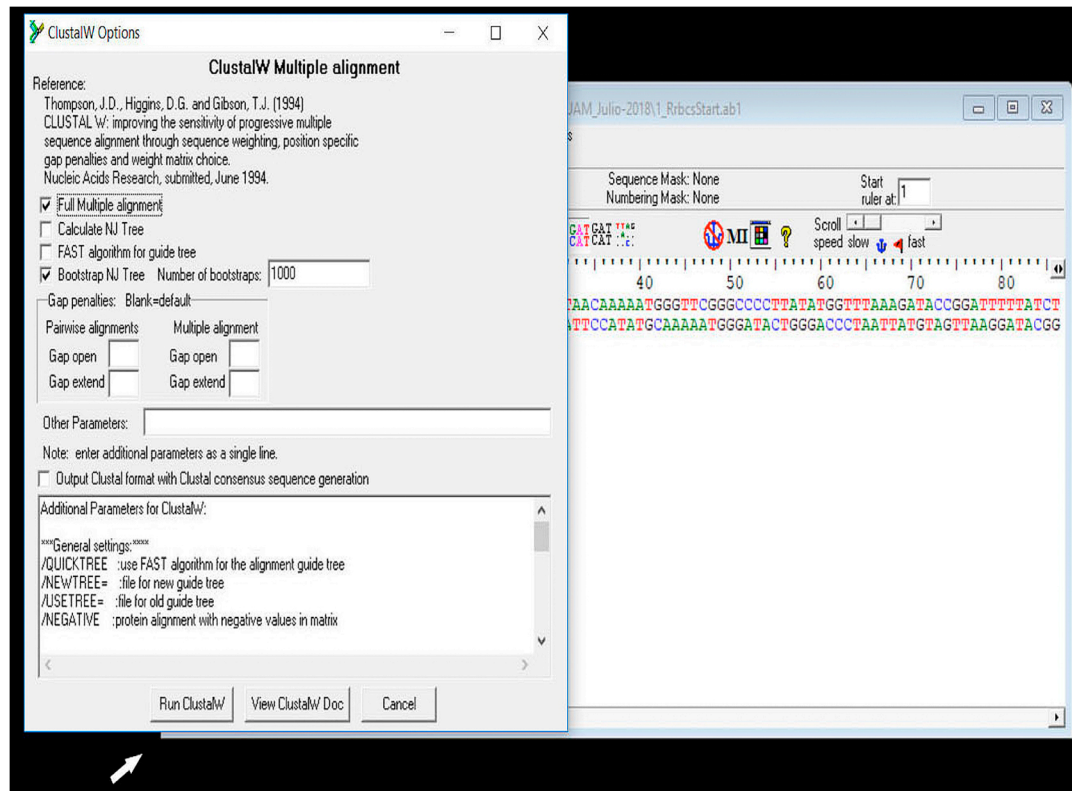


Fig. 9. Captura de pantalla paso 15: parámetros de ClustalW

16. En la parte superior del nuevo cuadro de secuencias, generado en clustal W, aparecerá una nueva subbarra de herramientas (Fig. 10a). Para detectar las bases que no coinciden entre sí, da clic en la opción **Shade identities and similarities alignmet window** de esta nueva subbarra (Fig. 10b). Notarás que los sitios en que las bases coinciden estarán resaltados en color, mientras que los sitios en que fueron diferentes no. Incluso notarás que en algunos sitios en lugar de una base aparecerá un **guión**. Este guion corresponde a un **GAP** (hueco, espacio en blanco) (Fig. 10c, flecha). Para editar tu secuencia debes colocar en los GAPS una base, si es que el cromatograma de mejor calidad la muestra, o borrar el GAP. En los sitios en que las bases no coinciden, debes decidir colocar sólo una base, esta deberá ser aquella que en los cromatogramas se vea más clara, debes guiarte en el cromatograma que tenga la mejor calidad. Notarás que tanto el cuadro de las secuencias como en el cuadro de los cromatogramas, muestran en la parte superior un intervalo de números de 10 en 10 (Fig. 10d). Estos números indican la posición de la base, por lo que, si notas que ambas secuencias no coinciden en una posición, debes fijarte el número de la posición que ocupan y ubicarte en la misma posición en el cromatograma, así podrás ver la calidad de la banda y tomar una decisión acerca de la base que debes conservar.

Nota: para editar la secuencia da clic en la opción **Mode** de la nueva barra de herramientas (Fig. 10e) y selecciona la opción **Edit**. Para que al momento de sustituir o borrar una base o un GAP no se desfasen las posiciones de los nucleótidos, selecciona la opción **Overwrite**, de este modo sobrescribirá en el espacio que ocupa la base que estás editando o no insertará una nueva base, por lo que el alineamiento no se desfasará.

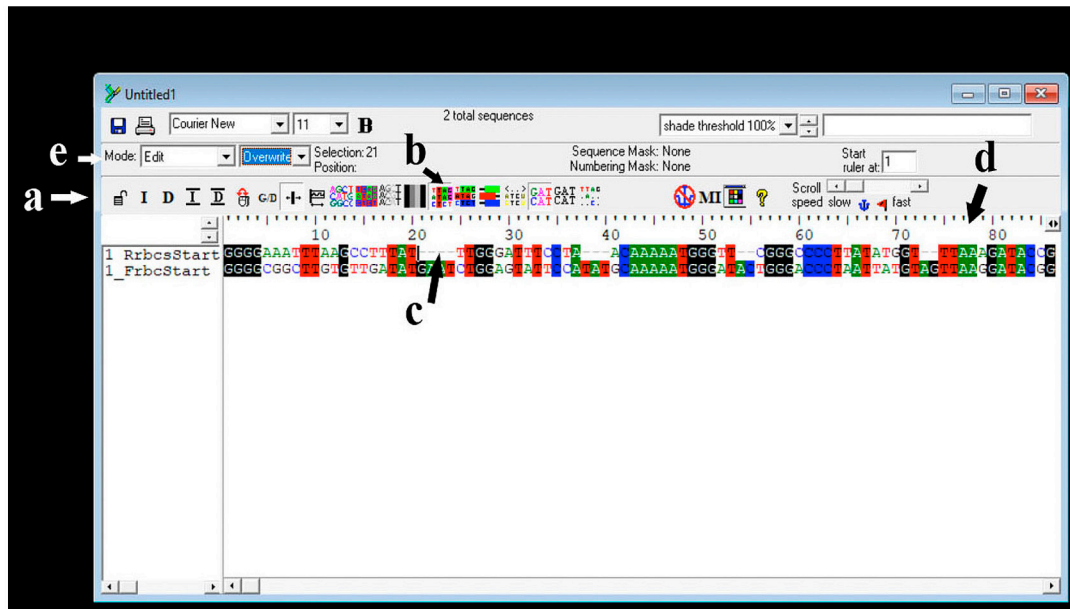


Fig. 10. Captura de pantalla paso 16: (a) sub barra de herramientas (b) edición de secuencias. (c) GAPS, (d) intervalo numérico de las posiciones en las bases y (e) comandos para la edición de secuencias

17. Una vez que hayas terminado con la edición, tus bases deberían coincidir totalmente, en consecuencia, todas estarán resaltadas en color, ya que se trata de la misma secuencia. Antes de guardarla elimina alguna de las dos (ya que ahora son iguales) y conserva sólo una para el siguiente paso.
18. Para guardar tu secuencia consenso selecciona en la barra principal del programa el menú **File**, del submenú que desplegará selecciona la opción **Save As**. Al dar clic en esta opción, se mostrará un nuevo cuadro de diálogo en el que podrás elegir el directorio, así como el formato en los que quieres guardarla (en donde desees guardarla en tu computadora fasta, phylip, txt, etc).
19. Guarda tu secuencia consenso en el subdirectorio **Práctica 1**, nombrándola **1Consenso**, con extensión **Fasta (.fas)**, de esta manera será compatible con todos los programas y puedes volver a abrirla y editarla con BioEdit.

Nota: también puedes guardar tu secuencia en el block de notas con extensión **txt**. Esto puedes hacerlo seleccionándola y copiándola directamente del cuadro de secuencia como en los pasos 8 y 9. Abre el block de notas y pega la secuencia a renglón corrido y sin espacios. No debes escribir nada más que la secuencia. Todos los datos adicionales que quieras colocar a la secuencia como sitio, fecha, gen o número de campo, etc., debes colocarlos en el nombre del archivo. Si colocas información adicional a tu secuencia, esta no podrá ser leída por otros programas.

RESULTADOS

Al concluir la edición de tu secuencia obtendrás dos archivos: **1Consenso.fas** que podrá ser leído con BioEdit o MEGA, etc., y **1Consenso.txt** en block de notas, este podrá ser importado y leído por otros programas de edición.

ACTIVIDADES COMPLEMENTARIAS

1. Obtén la secuencia consenso, en los formatos fasta y txt, de las siguientes secuencias en formato ab1, están disponibles en la subcarpeta **Práctica 1** del directorio **Prácticas de Filogenia**.

Archivos de secuencias:

2_FrbcStart

2_RrbcStart

3_FrbcStart

3_RrbcStart




4_FrbcStart

4_RrbcStart

5_FrbcStart

5_RrbcStart

BIBLIOGRAFÍA RECOMENDADA

-  Graur D. & Wen-Hsiung L. 2000. *Fundamentals of molecular evolution*, ed. 2. Sinauer Associates, Massachusetts.
-  Hall T.A. 1999. "BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT". *Nucleic Acids Symposium* 41: 95–98.
-  Simmons M.P. & Ochoterena H. 2000. "Gaps as characters in sequence based phylogenetic analyses". *Systematic Biology* 49: 369–381.

Práctica 2

Construcción de matrices: bases de datos genómicos y alineamiento múltiple con BioEdit

INTRODUCCIÓN

Las bases de datos genómicas son grandes reservorios de información de biología molecular, tal como secuencias nucleotídicas, artículos científicos, enfermedades genéticas u otros datos biotecnológicos de relevancia. El **Centro Nacional para la Información Biotecnológica** (NCBI por sus siglas en inglés) es la fuente de información genética más importante, ya que además de almacenar y mantener actualizada la base de datos de secuencias de ADN del GenBank, también se encarga de hacerla accesible en línea de manera gratuita y ordenada. Toda secuencia de ADN disponible en el GenBank tiene un único **número de acceso**, mediante el cual puede ser consultada, comparada y referida. GenBank se coordina con laboratorios individuales y otras bases de datos de secuencias como las del European Molecular Biology Laboratory (EMBL), la base de datos de ADN de Japón (DDBJ) o Barcode of Life Data System (BOLD), por lo que desde su portal podemos tener acceso a todas las secuencias existentes en el mundo para un organismo determinado.

Para que un grupo de secuencias de diferentes organismos que hipotéticamente están relacionados sean comparables entre sí y permitan identificar eventos que pudieran haber ocurrido entre linajes desde que estas secuencias divergieron a partir de un ancestro común, es necesario hacer empatar sus bases. Cuando empezamos a conformar una matriz de datos a partir de diferentes secuencias de un mismo gen, sus bases no siempre van a coincidir, esto debido a que las secuencias tienen diferentes longitudes que van generando el desplazamiento de las bases. Para ello es necesario realizar un **alineamiento**. Un **alineamiento múltiple** es el proceso en el que se comparan dos o más secuencias de ADN. Estos alineamientos son utilizados para la identificación de **regiones conservadas** (regiones que no varían) en un grupo de secuencias o eventos de mutación, tales como **sustitución de bases** (cambio de una base por otra), **inserciones** (inserción de una base que no se encuentra en otras secuencias) y **deleciones** (pérdida de una base). Las inserciones y **deleciones** en un alineamiento de secuencias están representadas por **Gaps**.

Existen múltiples *softwares* de descarga libre en línea que permiten realizar alineamientos múltiples de secuencias. Algunos de estos son Clustal X, MEGA, Mesquite o BioEdit.

OBJETIVOS

- Construir una matriz de secuencias de ADN a partir de secuencias propias y secuencias del GenBank.
- Aprender a importar a una matriz de datos, secuencias del GenBank.
- Aprender a alinear secuencias múltiples utilizando como herramienta BioEdit.

SOFTWARE REQUERIDO

1. BioEdit versión 7.0.5.3, disponible para su descarga en el sitio del programa, siguiendo la liga <http://www.mbio.ncsu.edu/BioEdit/bioedit.html>
2. Base de datos del GenBank disponible en línea, siguiendo la liga <https://www.ncbi.nlm.nih.gov/>

DESARROLLO

1. En el navegador de tu preferencia, copia la siguiente liga <https://www.ncbi.nlm.nih.gov/> para poder acceder a la base de datos del NCBI. Te mostrará la página de inicio en donde puedes explorar la información y herramientas que esta base de datos ofrece, desde acervos bibliográficos, proyectos biotecnológicos, secuencias moleculares o herramientas para subir o bajar archivos de secuencias (Fig. 11).

- En la barra de menú del lado derecho, encontrarás una barra de menú llamada **Popular Resources**. Selecciona la opción **BLAST** (Fig. 11, flecha).

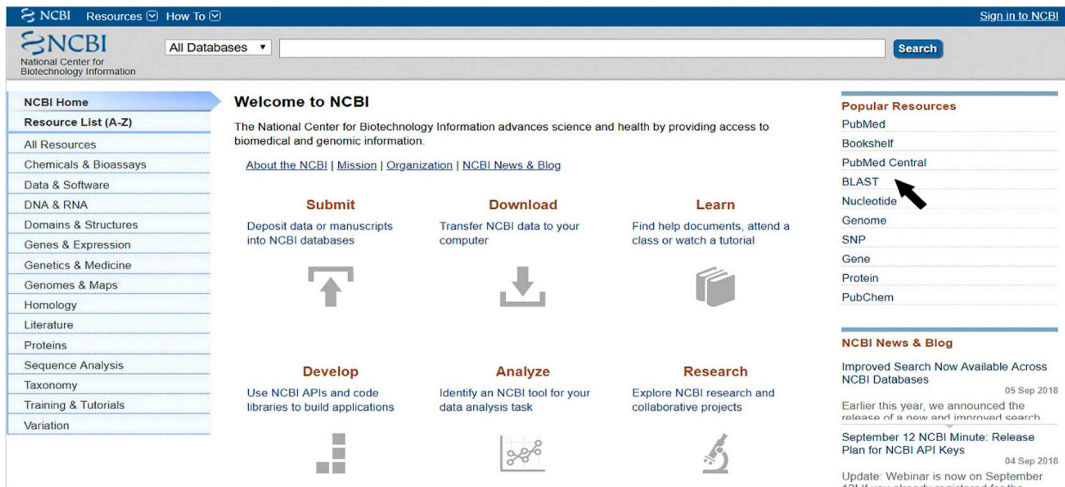


Fig. 11. Captura de pantalla pasos 1 y 2: página de inicio de la base de datos NCBI. Blast indicado con flecha

Nota: la herramienta de BLAST (Basic Local Alignment Search Tool) es un recurso de la base de datos del GenBank que te permite comparar una secuencia de ADN con todas las secuencias que almacena, permitiendo así una **identificación molecular**. Siempre que edites una secuencia molecular, es muy importante que realices un BLAST de tu secuencia consenso, a fin de verificar que tu secuencia corresponda efectivamente con un organismo relacionado a tu objeto de estudio, en este caso algas marinas. Si tu secuencia está contaminada, el resultado del BLAST corresponderá con un hongo o una bacteria o quizá con otra alga distinta a aquella que tú pensabas, que pudo haber estado epífita en tu organismo. De ser el caso, tu secuencia no servirá para tu análisis y deberás repetirla, para ello es importante siempre realizar un BLAST.

- Aparecerá en tu pantalla la página de BLAST (Fig. 12) que te mostrará cinco recuadros con diferentes opciones de BLAST de acuerdo con el tipo de datos que deseas comparar. Elige el recuadro con la opción **Nucleotid Blast**, ya que estamos comparando secuencias de nucleótidos (Fig. 12, flecha).

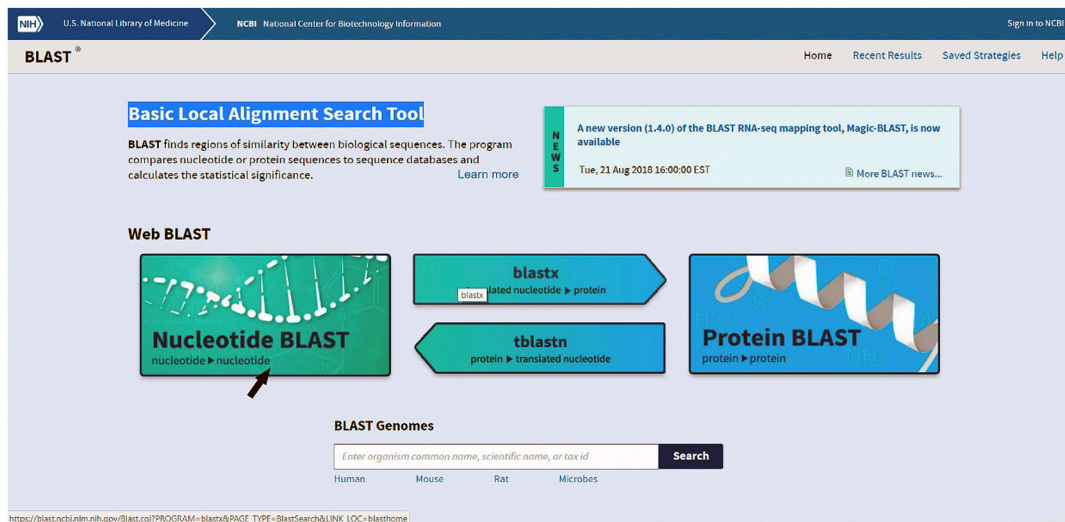


Fig. 12. Captura de pantalla paso 1: opciones de datos moleculares para BLAST

- Aparecerá una pantalla llamada **Standard Nucleotide BLAST** con tres secciones distintas: **Enter Query Sequence**, **Choose Search Set** y **Program Selection** (Fig. 13). En el primer recuadro en blanco de la sección Enter Query Sequence debes ingresar la secuencia que deseas comparar. Abre la secuencia **1Consenso.txt** (block de notas) que editaste en la práctica 1 y cópiala en ese recuadro (Fig. 13a, flecha).
- En la sección **Choose Search Set** deberás especificar el tipo de datos del genoma al que corresponde tu organismo a fin de restringir la búsqueda y hacerla más eficiente. Selecciona la opción **Others** (Fig. 13b, flecha). En la sección Program Selection deja el parámetro que el portal sugiere. Da clic en el botón **BLAST** (13c, flecha).

Fig. 13. Captura de pantalla pasos 4 y 5: captura de datos y selección de parámetros para BLAST:
(a) captura de la secuencia (b) tipo de genoma de la secuencia y (c) botón de análisis

- El resultado será, por un lado, un cuadro de barras de identidad con las 100 secuencias más parecidas a la tuya (Fig. 14a, flecha), y, por otro lado, un listado de esas especies por orden de identidad, es decir, las primeras secuencias de la lista son las más parecidas a tu secuencia (Fig. 14b, flecha). Del lado derecho del listado de identidad, te mostrará seis columnas adicionales con diferentes valores de identidad, como la puntuación de identidad mínima y máxima, el porcentaje de la secuencia que correspondió con tu secuencia, el porcentaje de similitud (identidad) y en la última columna el número de acceso de las secuencias que estuvieran más cercanas a la tuya (Fig. 14c, flecha).

Nota: notarás que tu secuencia tiene un 100 % de similitud con *Meristotheca cylindrica*, una especie de alga roja de la familia *Solieriaceae*. El 100 % de identidad indica que tu secuencia es sin lugar a dudas la especie *Meristotheca cylindrica*. Sabiendo esto ahora, trabajaremos en el alineamiento de secuencias de especies de *Solieriaceae*. Para realizar un alineamiento múltiple de las secuencias de *Solieriaceae* disponibles en el GenBank para un mismo gen, es indispensable concentrarlas en un solo archivo de secuencias. Para ello, las buscaremos en la base de datos del GenBank mediante sus números de acceso.

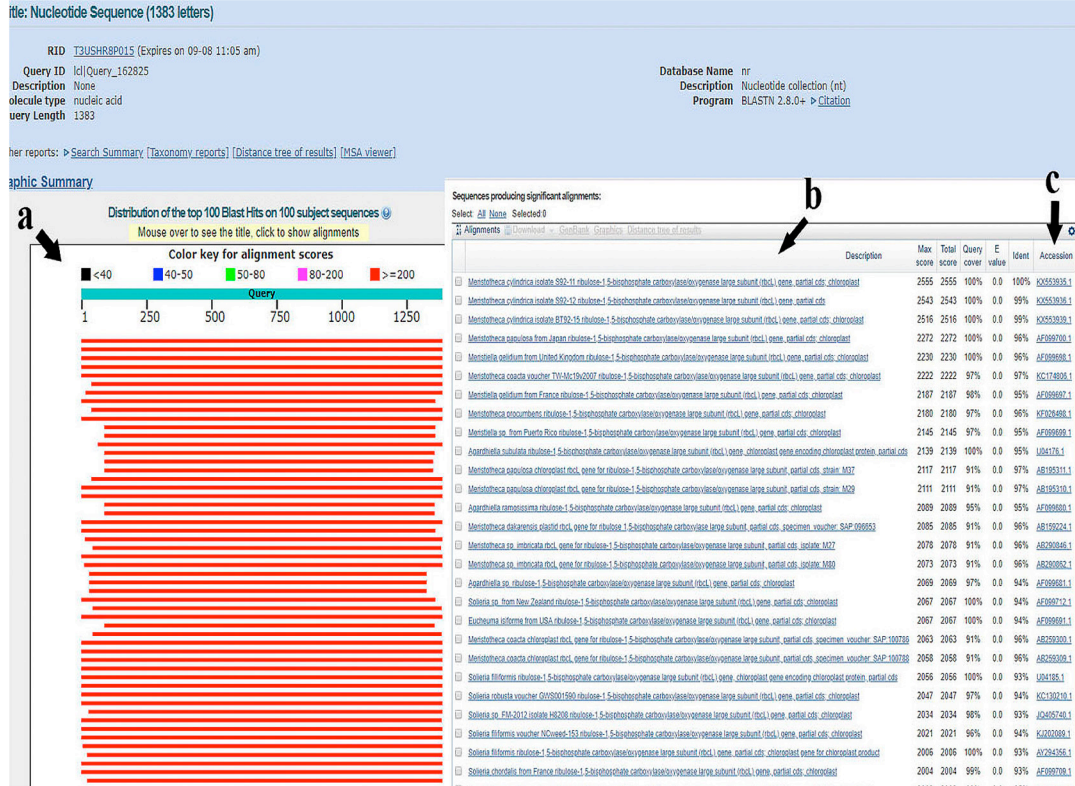


Fig. 14. Captura de pantalla paso 6: presentación de resultados del BLAST: (a) barras de identidad, (b) listado de especies por porcentaje de similitud y (c) valores de similitud y números de acceso

- Busca en la base de datos del GenBank, mediante sus números de acceso, las especies listadas en la Tabla 2. Para ello, regresa a la página de inicio de NCBI.

Tabla 2. Números de acceso al GenBank para las secuencias del gen *rbCL* de especies de Solieriaceae

Especies	Número de acceso
<i>Agardhiella subulata</i>	U04176
<i>Agardhiella ramosissima</i>	AF099680
<i>Anatheca mantagnei</i>	AB122015
<i>Betaphycus speciosus</i>	AF099685
<i>Betaphycus gelatinus</i>	JX069174
<i>Betaphycus philippinensis</i>	AF099684
<i>Euchematopsis isiformis</i>	AF099691
<i>Gracilaria flabelliformis</i>	AY049343
<i>Kappaphycus alvarezii</i>	JX069175
<i>Kappaphycus striatus</i>	JX623999
<i>Kappaphycus cottonii</i>	AF099695
<i>Merisothea coacta</i>	AB259300
<i>Merisothea darkarensis</i>	AB159224
<i>Merisothea gelidium</i>	AF099698
<i>Merisothea imbricata</i>	AB290846

<i>Merisotheca papulosa</i>	AF099700
<i>Merisotheca procumbens</i>	KF026498
<i>Merisotheca sp.</i>	AF099703
<i>Palisada perforata</i>	AF465807
<i>Sarcodiotheca gaudichaudii</i>	KM253822
<i>Sarcodiotheca furcata</i>	KM253835
<i>Sarcodiotheca filiforme</i>	AF099708
<i>Solieria filiforme</i>	AF099709
<i>Solieria filiformis</i>	KJ202089
<i>Solieria chordalis</i>	AF099709
<i>Solieria incurvata</i>	MG018944
<i>Solieria pacifica</i>	AF0997109
<i>Solieria robusta</i>	KC130210
<i>Tepoztequiella rhizoidea</i>	KY952886
<i>Tepoztequiella sp.</i>	AF099699
<i>Tacanoosca uncinata</i>	AF099693

8. En la parte superior verás dos barras, un menú desplegable con una flecha con opciones para restringir la búsqueda (Fig. 15a, flecha) y un recuadro en blanco para introducir el texto de la búsqueda (Fig. 15b, flecha). En la barra con menú desplegable busca y selecciona la opción **Nucleotide**. En el recuadro en blanco, escribe el primer número de acceso de la Tabla 2. A continuación da clic sobre el botón **Search** (Fig. 15c, flecha).

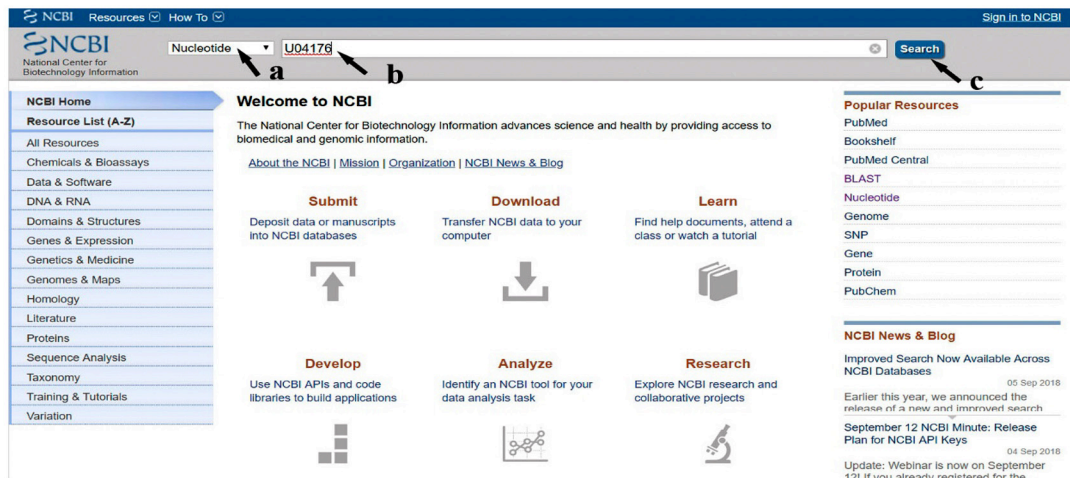


Fig. 15. Captura de pantalla paso 8: menú de búsqueda de la base de datos NCBI: (a) categorías de búsqueda, (b) introducción de datos de búsqueda y (c) botón de búsqueda

9. Aparecerá una pantalla con toda la información asociada a esa secuencia, como el gen al que corresponde, el número de acceso, el organismo al que pertenece, la referencia de la publicación en donde fue incluida y los datos propiamente de la secuencia (Fig. 16).
10. Da clic sobre la opción **FASTA** (Fig. 16, flecha).

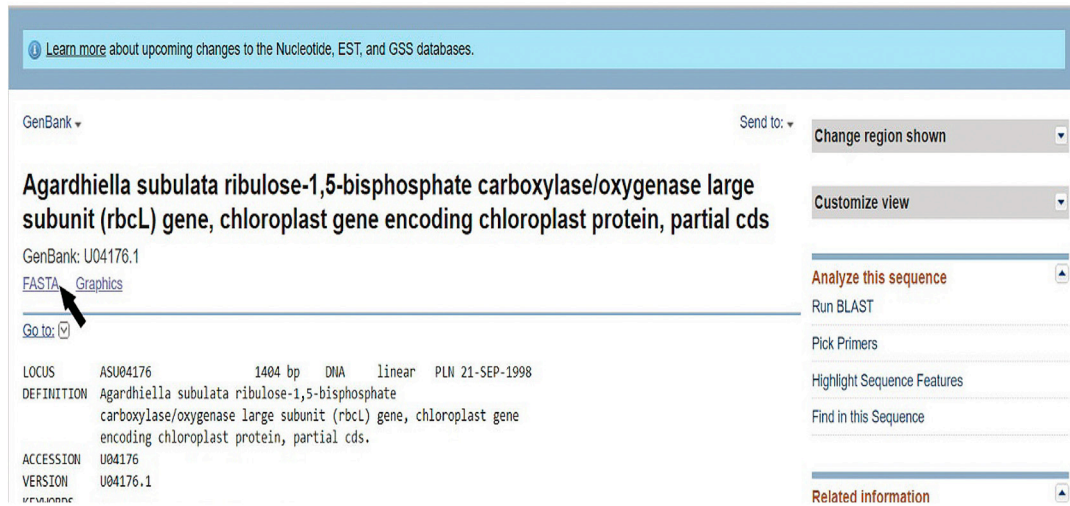


Fig. 16. Captura de pantalla pasos 9 y 10: presentación de los datos de secuencias en la base de datos del NCBI

11. Aparecerá únicamente la secuencia en formato FASTA (Fig. 17). Selecciónala y cópiala, sólo la secuencia de bases sin ninguna otra información adicional.

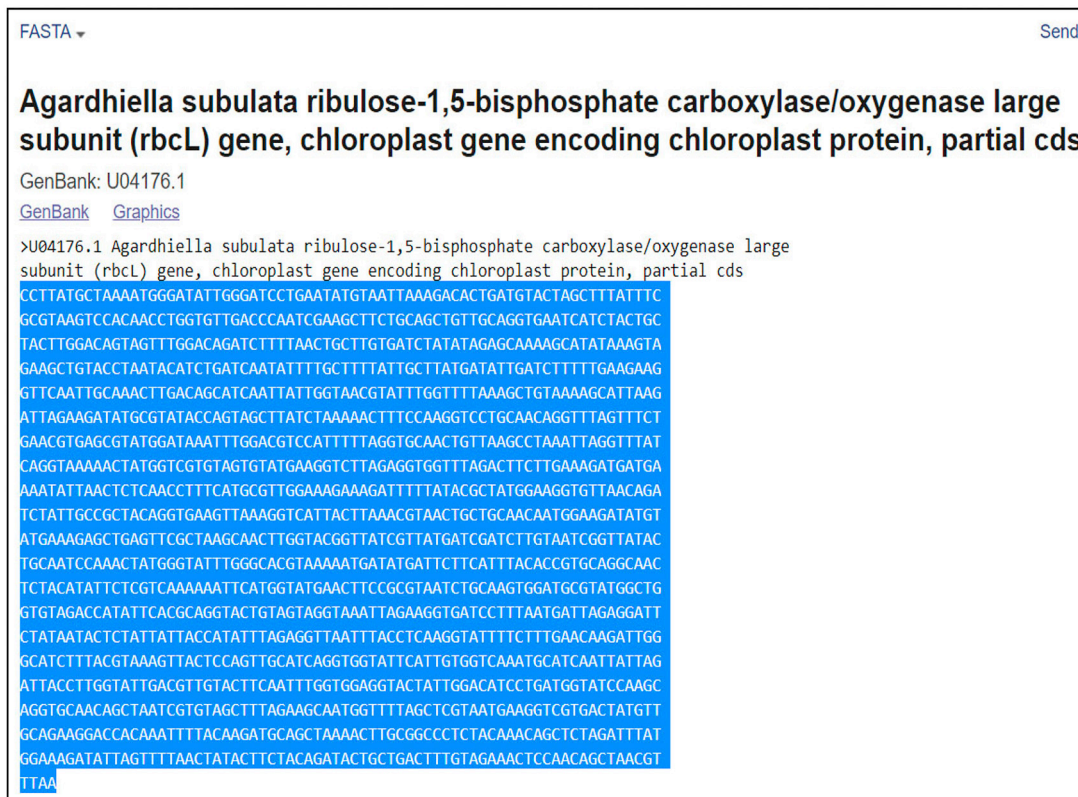


Fig. 17. Captura de pantalla paso 11: presentación de las secuencias fasta en la base de datos del NCBI

12. Del directorio Prácticas de Filogenia, subcarpeta Práctica 1, abre con BioEdit el archivo **1Consenso.fas** que editaste en la Práctica 1. Sobre este archivo concentraremos todas las secuencias de la Tabla 2.

RESULTADOS







Al final del alineamiento, obtendrás una matriz de secuencias **Solieriaceae.fas**, del gen *rbcl* para las especies de la familia Solieriaceae más dos secuencias de especies de otras familias que funcionarán en el futuro como **grupo externo**. Este archivo será utilizado en las prácticas futuras para inferencia filogenética.

Nota: para la elección adecuada del grupo externo, debes tener en cuenta el nivel taxonómico al que estás realizando el análisis. Si bien es cierto que el grupo externo debe ser tan lejano a tu grupo de interés (o grupo interno) como para que funcione realmente como externo, es decir, que no resulte agrupándose en tu grupo interno, tampoco puede ser tan lejano que genere atracción de ramas largas que resten resolución a tu análisis. Por lo tanto, si en tu análisis sólo estás comparando especies de único género, un buen grupo externo puede ser otro género de la familia con el que tu grupo de interés tenga la mayor distancia genética, y si en tu análisis incluyes a varios géneros de una misma familia, un buen grupo externo puede ser un género de una familia que pertenezca al mismo orden, pero considera siempre al género de mayor distancia genética dentro del nivel taxonómico superior al que pertenezca tu grupo de interés.

ACTIVIDADES COMPLEMENTARIAS

1. Agrega a tu archivo *rbcl_Solieriaceae.fas* las 20 secuencias restantes de la Tabla 2.
2. Repite el alineamiento con las nuevas secuencias añadidas y salva tu matriz en el mismo directorio que la anterior. Esta será utilizada en todas las prácticas futuras.

BIBLIOGRAFÍA RECOMENDADA

-  Hall T.A. 1999. "BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT." *Nucleic Acids Symposium* 41: 95–98.
-  Higgins D. 2003. *Multiple alignment. In: The phylogenetic handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing* (Ed. by P. Lemey, M. Salemi & A. Vandamme), pp. 45–71. Cambridge University Press, Cambridge, UK.
-  National Centre for Biotechnology Information, U.S. National Library of Medicine 8600 Rockville Pike, Bethesda MD, 20894 USA <https://www.ncbi.nlm.nih.gov/nuccore/U04176.1?report=fasta>
-  Simmons M.P. & Ochoterena H. 2000. "Gaps as characters in sequence based phylogenetic analyses". *Systematic Biology* 49: 369–381.
-  Thompson J.D., Higgins D.G. & Gibson T.J. 1994. "CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice". *Nucleic Acids Research* 22: 4673–4680
-  Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F. & Higgins D. 1997. "The CLUSTAL X windows interface: flexible strategies for multiple sequences aided by quality analysis tools". *Nucleic Acids Research* 25: 4876–4882.

Práctica 3

Estimación de distancias genéticas y modelos de evolución con Mega-X

INTRODUCCIÓN

La **distancia genética** se define como el porcentaje de diferencia (o no coincidencia) entre dos o más secuencias de nucleótidos. Estas secuencias pueden corresponder a individuos de la misma especie (**distancia intraespecífica**), o a individuos de especies distintas (**distancia interespecífica**). En términos generales, la distancia genética refleja, hipotéticamente, el tiempo transcurrido desde la divergencia de dos especies a partir de un ancestro común. Sin embargo, durante el proceso evolutivo, un nucleótido podría haber regresado varias veces a un estado ancestral después de haber experimentado otras mutaciones previas. La simple distancia genética entre dos secuencias (**distancias *p* o distancias no corregidas**), aunque proporciona información sobre que tanto difieren entre ellas, no considera estas imprecisiones, por lo que no refleja con precisión el tiempo de divergencia de dos especies. Por otro lado, cuando se emplean **modelos de evolución molecular** para establecer la distancia entre dos secuencias (**distancias corregidas**), mediante la corrección de imprecisiones y estimación de diversos parámetros (frecuencia de bases, sustituciones, tasas de heterogeneidad, etc.), las distancias genéticas si reflejarán un proceso evolutivo que podrá esquematizarse en una **filogenia**.

La mayoría de los **modelos de evolución molecular** (algoritmos que miden la tasa de cambios nucleotídicos (transiciones/transversiones) entre las secuencias de un alineamiento) conocidos, asumen que: (a) todos los sitios nucleotídicos cambian de forma independiente, lo que implica que el cambio en un sitio no afecta la probabilidad de cambio en otro sitio; (b) la tasa de sustitución es constante en el tiempo y entre los distintos linajes; (c) la frecuencia de las bases está en equilibrio, lo que implica que la composición de las bases es la misma a lo largo del tiempo y en distintos linajes y (d) las probabilidades condicionales de sustitución nucleotídicas son las mismas para todos los sitios y estos no cambian con el tiempo, lo que implica que todos los sitios cambian con la misma probabilidad, sin embargo, en toda secuencia genética hay una proporción de sitios variables y una proporción de **sitios invariables**.

Algunos de los principales modelos de evolución molecular son:

- **Modelo de Jukes-Cantor (JC)**: este modelo considera que todos los cambios entre nucleótidos son igualmente probables.
- **Kimura 2 parameter model (K2P)**: este modelo considera que las transiciones y transversiones tienen diferente probabilidad.
- **Hasegawa, Kishino & Yano (HKY)**: este modelo considera que las transiciones y transversiones tienen diferente probabilidad y frecuencia de bases desiguales.
- **General Time Reversible model (GTR)**: este modelo considera una probabilidad diferente para cada sustitución y frecuencias de bases desiguales.

El cálculo de distancias genéticas corregidas y no corregidas, puede llevarse a cabo con diferentes *softwares* libres tales como DnaSP, Mega o Mega-X, mientras que la estimación de modelos de evolución también puede llevarse a cabo con *softwares* libres como TNT, Mega, Mega-X o en portales en línea como TOPALI. El cálculo de los modelos en estos *softwares* o portales se realiza mediante el criterio de información de Akaike (Akaike Information criterion, AIC), que es una medida de la calidad relativa de un modelo estadístico para un conjunto dado de datos.

OBJETIVOS

- Calcular las distancias genéticas no corregidas entre las especies de *Solieriaceae*, utilizando como herramienta el programa Mega-X.
- Estimar el modelo de evolución nucleotídica que mejor explique las sustituciones en las secuencias de rbcL de las especies de Solieriaceae, utilizando como herramienta el programa Mega-X.
- Calcular las distancias genéticas corregidas de las especies de Solieriaceae, utilizando como herramienta el programa Mega-X.

SOFTWARE REQUERIDO

1. Mega-X Molecular Evolutionary Genetics Analysis (Kumar et al. 2018), disponible para su descarga en el sitio del programa, siguiendo la liga <http://www.megasoftware.net>

Nota: asegúrate de descargar la versión compatible con el sistema operativo de tu computadora. Una vez descargado el archivo zip, da clic en el archivo ejecutable y sigue las instrucciones para la instalación del programa.

DESARROLLO

Nota: aunque el programa MEGA-X trabaja únicamente con matrices en formato **MEG**, puede leer archivos en formato FASTA. Sin embargo, para que MEGA-X te permita realizar análisis, antes es necesario transformar el archivo Fasta en archivo MEG. Esto lo realiza el propio programa al abrir el archivo FASTA. La transformación se describe en los pasos 1-6 (Figs 19-24).

1. Abre el programa Mega-X. Aparecerá una ventana en blanco con dos barras de menú ubicadas en la parte superior izquierda, una general (Fig. 19a) y otra pictográfica de herramientas de análisis (Fig. 19b). En la parte inferior izquierda, aparecerá una tercera barra pictográfica con opciones exclusivas del *software*, no de análisis. (Fig. 19c).



Fig. 19. Captura de pantalla paso 1: opciones de menú en MEGA-X: (a) menú principal, (b) herramientas de análisis y (c) menú de *software*

- De la barra de herramientas de análisis, da clic sobre la opción DATA, al hacerlo será desplegado un nuevo sub-menú, selecciona la opción **Open a File/Session** (Fig. 20). Te abrirá una ventana de diálogo que te permitirá buscar entre tus archivos y seleccionar tu conjunto de datos. Del directorio Prácticas de Filogenia subdirectorío Práctica 2, selecciona la matriz de datos de secuencias **rbCL_Solieriaceae.fas** que generaste en la práctica 2.

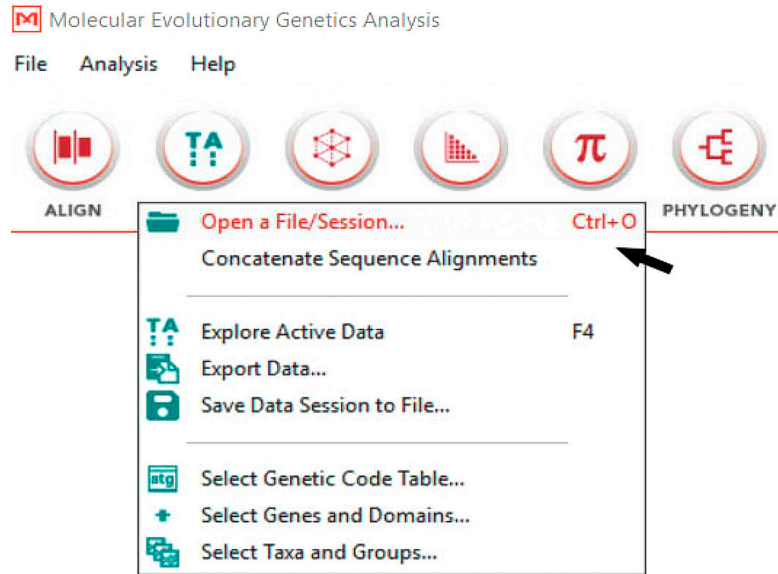


Fig. 20. Captura de pantalla paso 2: importación de una matriz de datos a MEGA-X

- Aparecerá un pequeño cuadro diálogo que te preguntará que deseas hacer con tus datos y te dará las opciones alinear y analizar. Selecciona la opción **Analyze** (Fig. 21).

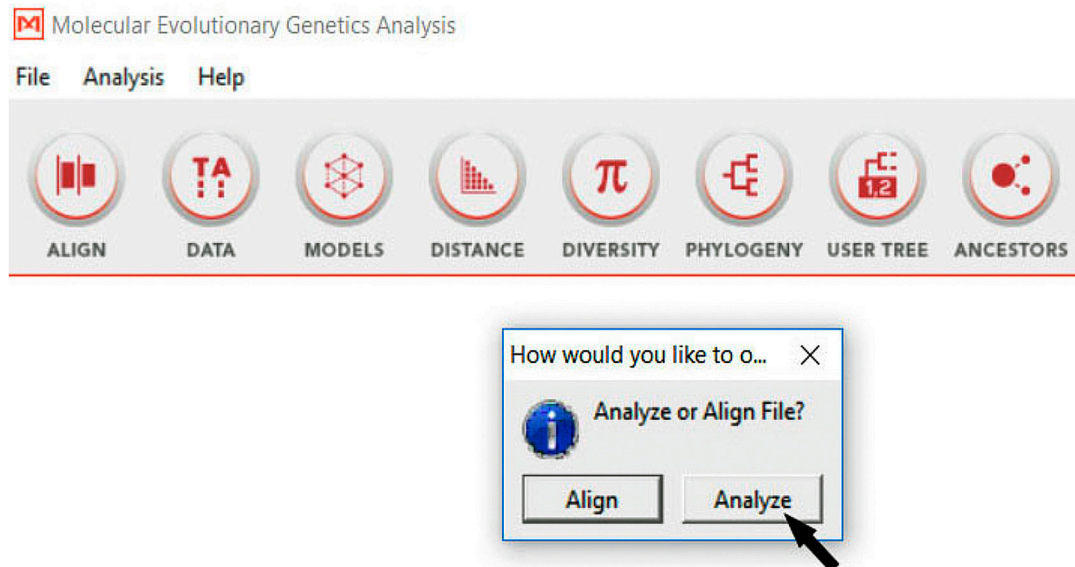


Fig. 21. Captura de pantalla paso 3: opciones de entrada de datos en MEGA-X

- Aparecerá otro pequeño cuadro de diálogo presentando tres diferentes opciones de tipos para tus datos (Fig. 22). Selecciona la opción **Nucleotide Sequences**, y presiona el botón **OK**.

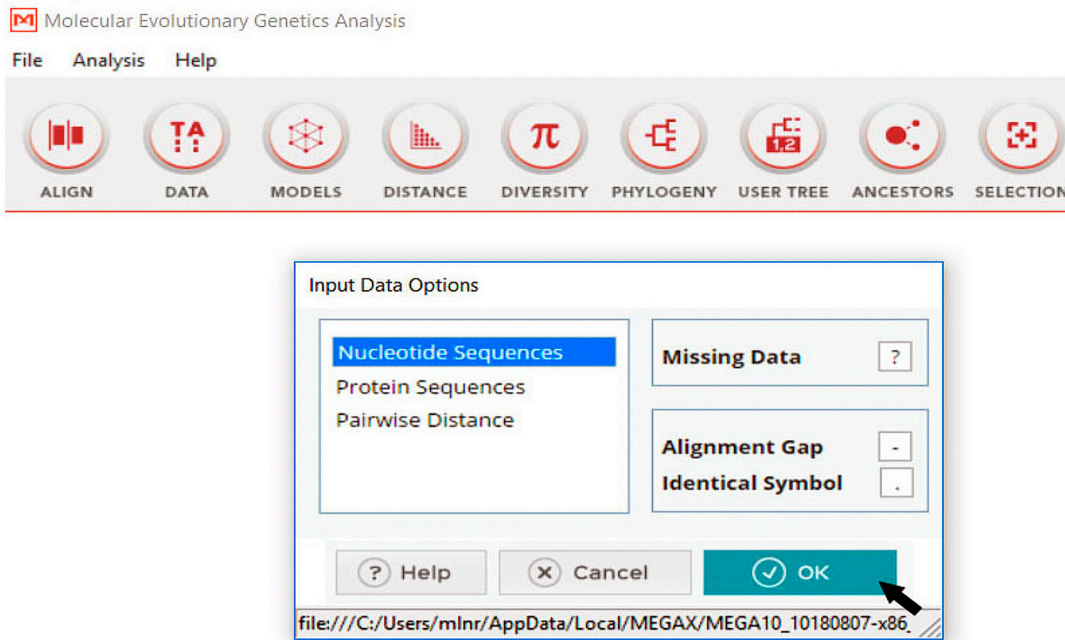


Fig. 22. Captura de pantalla paso 4: registro del tipo de datos de entrada en MEGA-X

- Un nuevo cuadro de diálogo te preguntará si tus secuencias codifican o no para alguna proteína. En este caso, como se trata de secuencias de rbcL, codifican para la subunidad larga de la Ribulosa 1,5-Bifosfato, Carboxilasa-Oxigenasa, por lo tanto, da clic en la opción **Yes** (Fig. 23).

Nota: no todas las secuencias de ADN son codificantes, en esos casos la opción deberá ser **No**.

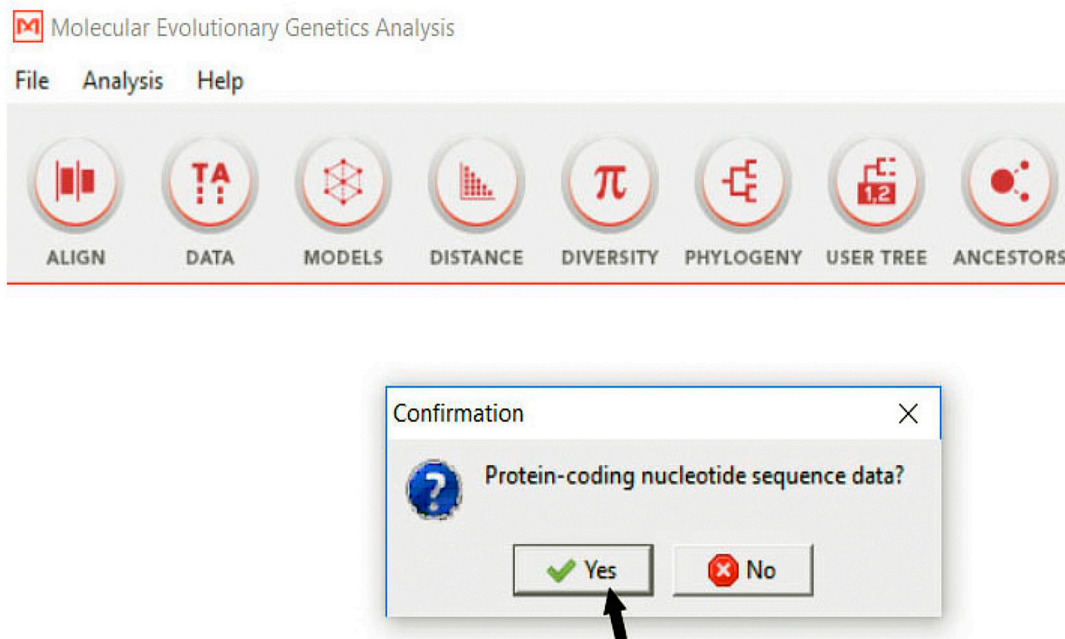


Fig. 23. Captura de pantalla paso 5: definición de datos codificantes en MEGA-X

6. En seguida, una nueva ventana de diálogo te preguntará sobre el código genético que es más apropiado (dependiendo de si tu secuencia es de una planta, animal, hongo, protozooario, etc.) para leer tu secuencia como una proteína. Selecciona la opción **Standard** y presiona el botón **OK**. (Fig. 24).

Nota: el código genético que el programa ofrece varía dependiendo del tipo de organismo, un tipo de código es más apropiado que otro para tu organismo. Por ejemplo, para las algas marinas en general el código genético será siempre estándar.

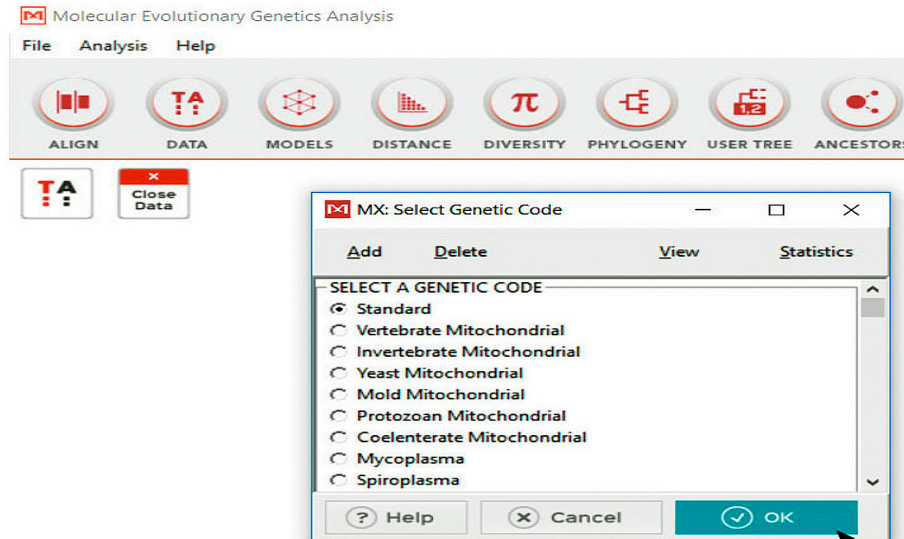


Fig. 24. Captura de pantalla paso 6: definición de código genético para la lectura de datos en MEGA-X

7. Finalmente, la matriz de datos está cargada en MEGA. Aparecerán en la pantalla en blanco del programa, dos archivos, la matriz de datos en formato **MEG** activa (Fig. 25a) y una opción para cerrar la matriz que está activa sin cerrar el programa (Fig. 25b).

Nota: si quisieras ver el alineamiento en MEGA, da clic sobre el ícono de alineamiento (Fig. 25a) y te desplegará la matriz de datos en formato Meg (Fig. 25c).

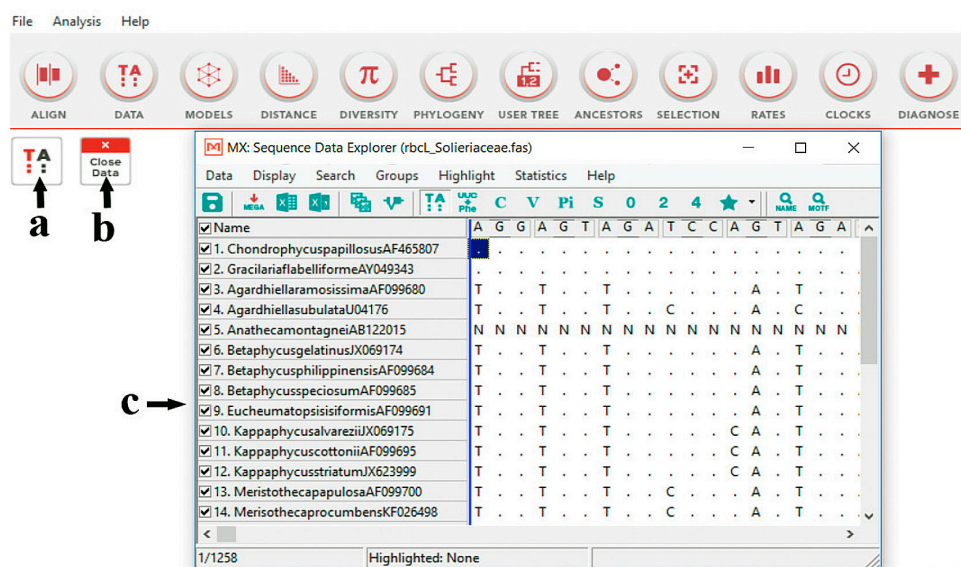


Fig. 25. Captura de pantalla paso 7: presentación de los datos de secuencia activos en la interfase de MEGA-X: (a) matriz de datos activa en mega, (b) opción para el cierre de datos activos sin cerrar el programa y (c) matriz de datos en formato MEGA

- En la barra de herramientas de análisis, da clic sobre la opción **DISTANCE**, al hacerlo será desplegado un nuevo submenú, selecciona la opción **Compute Pairwise Distances** (Fig. 26).

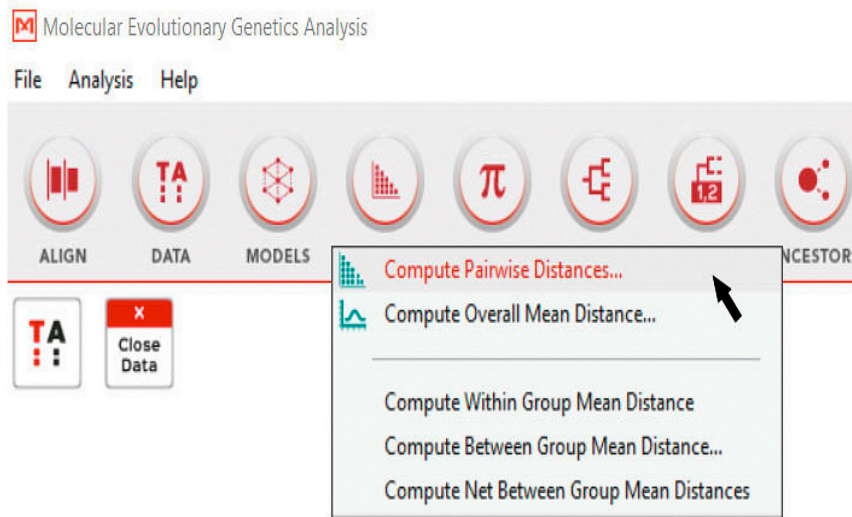


Fig. 26. Captura de pantalla paso 8: cálculo de distancias no corregidas

- Aparecerá un cuadro de diálogo preguntándote si quieres utilizar los datos actualmente activos y te mostrará el nombre de la matriz de datos que está activa (Fig. 27). Comprueba que sea el nombre de la matriz que deseas analizar y da clic en la opción **Yes**. Adicionalmente te pregunta si deseas que el programa recuerde reutilizar esos datos en futuros análisis, puedes activar la casilla o no, en este caso no la actives.

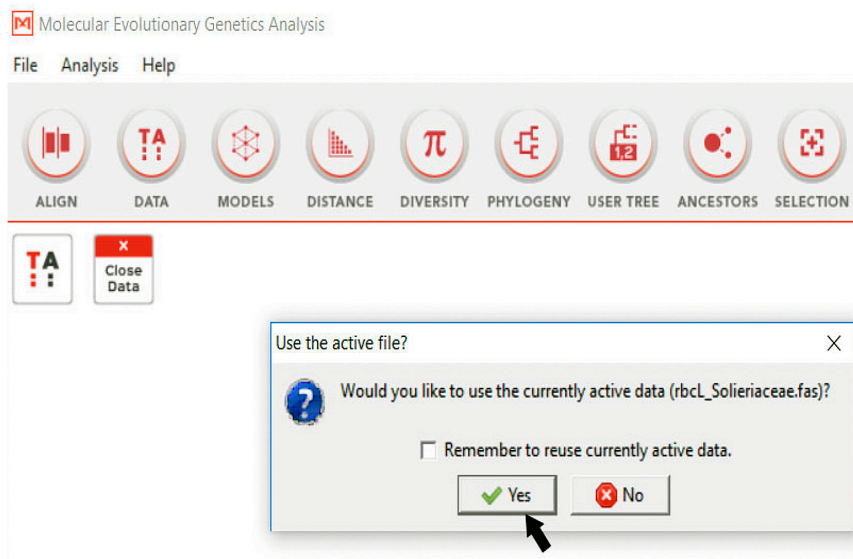


Fig. 27. Captura de pantalla paso 9: comando del programa para permitir el uso de los datos activos

- Aparecerá un cuadro de diálogo en donde debemos especificar las opciones de análisis para calcular la distancia genética (Fig. 28), comenzaremos calculando **distancias no corregidas**.

11. En la sección **Estimate Variance** debes especificar el método de estimación de la varianza, selecciona la opción **None** (Fig. 28a).
12. En la sección **Substitution Model** debes especificar primero el tipo de sustitución, elige la opción **Nucleotide** (Fig. 28b). Enseguida te preguntará por el modelo bajo el cual deseas calcular la distancia genética, como estamos estimando **distancias no corregidas**, selecciona la opción **p-distance** (Fig. 28c). El siguiente recuadro te preguntará por el tipo de mutaciones a considerar en el cálculo de la distancia, selecciona la opción **Transitions + Transversions** (Fig. 28d).
13. En la sección **Rates and Patterns**, te preguntará por la tasa de variación de los nucleótidos en tus secuencias, selecciona la opción **Uniform Rates** (Fig. 28e).
14. En la sección **Data Subset to Use**, te preguntará por el tratamiento que dará a los gaps en el alineamiento, selecciona la opción **Complete deletion** para que nos los considere en el cálculo (Fig. 28f).

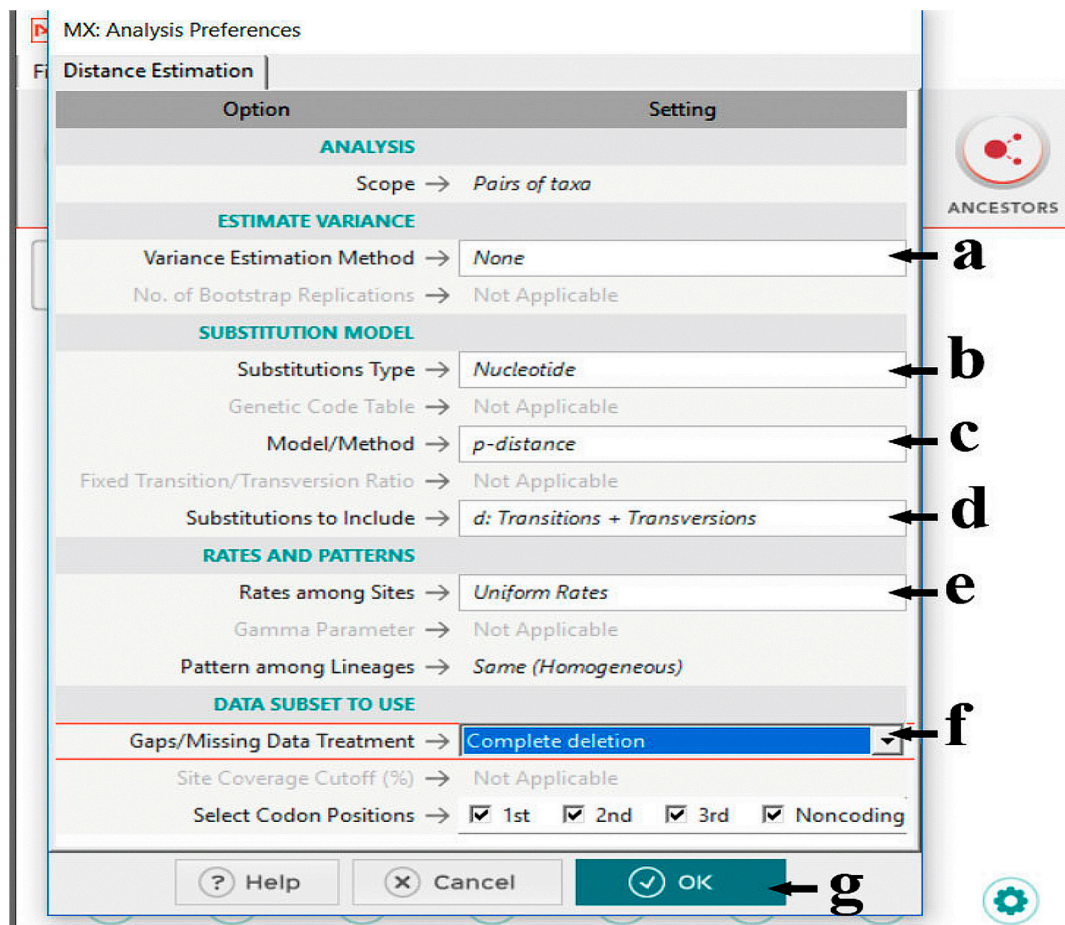


Fig. 28. Captura de pantalla pasos 10 a 14: parámetros para el cálculo de distancias no corregidas: (a) método de estimación, (b) tipo de sustitución de bases, (c) selección de modelo de estimación de distancia, (d) tipos de sustitución, (e) tasa de variación de nucleótidos, (f) tratamiento de los GAPS y (g) botón de ejecución del análisis

15. Presiona el botón **OK** (Fig. 28g). Te dará como resultado una **matriz triangular** con los porcentajes de distancia expresados en decimales (Fig. 29).

MX: Pairwise Distances (rbcl_Solieriaceae.fas)

File Display Average Caption Help

	1	2	3	4	5	6	7	8	9
1. ChondrophycuspapillosusAF465807	0								
2. GracilariafabelliformeAY049343	0.15087								
3. AgardhiellaramosissimaAF099680	0.14713	0.13716							
4. AgardhiellasubulataU04176	0.14838	0.13591	0.01122						
5. AnathecamontagneiAB122015	0.14214	0.13591	0.03865	0.04239					
6. BetaphycusgelatinusJX069174	0.13342	0.14214	0.07107	0.07232	0.06858				
7. BetaphycusphilippinensisAF099684	0.13342	0.14339	0.07232	0.07357	0.06733	0.01122			
8. BetaphycusspeciosumAF099685	0.13840	0.13965	0.06733	0.06858	0.06983	0.03616	0.03990		
9. EucheumatopsisiformisAF099691	0.13342	0.14090	0.04738	0.05112	0.04988	0.06608	0.06858	0.05611	
10. KappaphycusalvareziiJX069175	0.14214	0.14713	0.07107	0.07107	0.07357	0.07357	0.06983	0.07357	0.06983
11. KappaphycuscottoniiAF099695	0.14464	0.14838	0.07731	0.07731	0.07980	0.07980	0.07855	0.07980	0.07481
12. KappaphycusstriatumJX623999	0.14090	0.14838	0.07357	0.07357	0.07606	0.07606	0.07481	0.07606	0.07232
13. MeristothecapapulosaAF099700	0.14214	0.14214	0.06608	0.06733	0.05860	0.08354	0.08479	0.08728	0.06608
14. MeristothecaprocombensKF026498	0.14713	0.14464	0.06733	0.06858	0.05860	0.07980	0.08354	0.08603	0.06733
15. MeristothecaibricataAB290846	0.14464	0.13591	0.06234	0.06359	0.05237	0.08229	0.08603	0.08105	0.06608
16. MeristothecadakarensisAB159224	0.14339	0.13716	0.05985	0.06608	0.06110	0.07855	0.08728	0.08479	0.06359
17. MeristothecacoactaAB259300	0.14838	0.14339	0.06359	0.06484	0.05362	0.08479	0.08853	0.08354	0.06733
18. MeristothecaspAF099703	0.14713	0.14090	0.06484	0.06608	0.05860	0.08853	0.09227	0.08728	0.06858
19. MeristothecacylindricaKX553938	0.14339	0.13217	0.04863	0.05237	0.04863	0.07107	0.07481	0.06983	0.05362
20. MeristothecagelidiumAF099697	0.13965	0.13591	0.05611	0.05486	0.05985	0.07606	0.07980	0.07232	0.05860
21. SarcodiothecagaudichaudiiKM253822	0.14339	0.13466	0.05736	0.05736	0.06359	0.08105	0.08229	0.08603	0.07606
22. SarcodiothecacurcataKM253835	0.14464	0.13716	0.04988	0.05112	0.05860	0.07481	0.07980	0.07731	0.06733
23. SarconemafiliformeAF099708	0.13840	0.13965	0.06484	0.06858	0.06608	0.08354	0.08479	0.08229	0.06484
24. SolieriachordalisAF099709	0.14838	0.14090	0.05985	0.06110	0.06359	0.08479	0.08229	0.08229	0.06484
25. SolieriaformisKJ202089	0.14713	0.13342	0.05985	0.06359	0.04863	0.07357	0.07606	0.07481	0.06608
26. SolieriaincurvataMG018944	0.14339	0.12469	0.05860	0.06234	0.05112	0.07731	0.07481	0.07855	0.06608
27. SolieriapacificaAF099710	0.14963	0.15087	0.06110	0.06234	0.05985	0.07855	0.07980	0.07481	0.06858
28. SolieriarobustaKC130210	0.15461	0.14090	0.05611	0.05736	0.05486	0.08229	0.08479	0.07731	0.06234
29. TacanooscauncinataAF099693	0.15686	0.14464	0.05486	0.06110	0.06858	0.08479	0.08978	0.08853	0.07232
30. TepoztequellaspAF099699	0.14963	0.14090	0.04738	0.05112	0.04738	0.07731	0.08105	0.08229	0.06484
31. TepoztequellarhizoideaKY952886	0.15461	0.15337	0.05736	0.05985	0.05985	0.08728	0.09102	0.08479	0.06484

Fig. 29. Captura de pantalla paso 15: matriz triangular con los valores de distancia no corregida

16. En el menú superior de la matriz, te mostrará algunas herramientas de edición. Dando clic sobre el botón **Decrease the precision of Distances** (Fig. 30a, flecha) de esa barra de herramientas, reduce los valores a tres dígitos (Fig. 30).

Nota: los valores de porcentaje se leen como enteros después del primer 0. Ejemplo, la matriz triangular muestra un valor de distancia de 0.049 (Fig. 30b, flecha) entre *Agardhiella ramosissima* (taxa 3) y *Meristotheca cylindrica* (taxa 19). El porcentaje de **distancia no corregida** entre ambas especies es 4.9 %.

MX: Pairwise Distances (rbcl_Solieriaceae.fas)

File Display Average Caption Help

	1	2	3	4	5	6	7	8	9	10	11
1. ChondrophycuspapillosusAF465807											
2. GracilariafabelliformeAY049343	0.151										
3. AgardhiellaramosissimaAF099680	0.147	0.137									
4. AgardhiellasubulataU04176	0.148	0.136	0.011								
5. AnathecamontagneiAB122015	0.142	0.136	0.039	0.042							
6. BetaphycusgelatinusJX069174	0.133	0.142	0.071	0.072	0.069						
7. BetaphycusphilippinensisAF099684	0.133	0.143	0.072	0.074	0.067	0.011					
8. BetaphycusspeciosumAF099685	0.138	0.140	0.067	0.069	0.070	0.036	0.040				
9. EucheumatopsisiformisAF099691	0.133	0.141	0.047	0.051	0.050	0.066	0.069	0.056			
10. KappaphycusalvareziiJX069175	0.142	0.147	0.071	0.071	0.074	0.074	0.070	0.074	0.070		
11. KappaphycuscottoniiAF099695	0.145	0.148	0.077	0.077	0.080	0.080	0.079	0.080	0.075	0.009	
12. KappaphycusstriatumJX623999	0.141	0.148	0.074	0.074	0.076	0.076	0.075	0.076	0.072	0.005	0.004
13. MeristothecapapulosaAF099700	0.142	0.142	0.066	0.067	0.059	0.084	0.085	0.087	0.066	0.069	0.075
14. MeristothecaprocombensKF026498	0.147	0.145	0.067	0.069	0.059	0.080	0.084	0.086	0.067	0.074	0.080
15. MeristothecaibricataAB290846	0.145	0.136	0.062	0.064	0.052	0.082	0.086	0.081	0.066	0.067	0.074
16. MeristothecadakarensisAB159224	0.143	0.137	0.060	0.066	0.061	0.079	0.087	0.085	0.064	0.080	0.086
17. MeristothecacoactaAB259300	0.148	0.143	0.064	0.065	0.054	0.085	0.089	0.084	0.067	0.067	0.074
18. MeristothecaspAF099703	0.147	0.141	0.065	0.066	0.059	0.089	0.092	0.087	0.069	0.070	0.076
19. MeristothecacylindricaKX553938	0.143	0.132	0.049	0.052	0.049	0.071	0.075	0.070	0.054	0.066	0.072
20. MeristothecagelidiumAF099697	0.140	0.136	0.056	0.055	0.060	0.076	0.080	0.072	0.059	0.079	0.084
21. SarcodiothecagaudichaudiiKM253822	0.143	0.135	0.057	0.057	0.064	0.081	0.082	0.086	0.076	0.080	0.084
22. SarcodiothecacurcataKM253835	0.145	0.137	0.050	0.051	0.059	0.075	0.080	0.077	0.067	0.075	0.079
23. SarconemafiliformeAF099708	0.138	0.140	0.065	0.069	0.066	0.084	0.085	0.082	0.065	0.082	0.089
24. SolieriachordalisAF099709	0.148	0.141	0.060	0.061	0.064	0.085	0.082	0.082	0.065	0.085	0.091
25. SolieriaformisKJ202089	0.147	0.133	0.060	0.064	0.049	0.074	0.076	0.075	0.066	0.080	0.084
26. SolieriaincurvataMG018944	0.143	0.125	0.059	0.062	0.051	0.077	0.075	0.079	0.066	0.074	0.077
27. SolieriapacificaAF099710	0.150	0.151	0.061	0.062	0.060	0.079	0.080	0.075	0.069	0.079	0.085
28. SolieriarobustaKC130210	0.155	0.141	0.056	0.057	0.055	0.082	0.085	0.077	0.062	0.087	0.091
29. TacanooscauncinataAF099693	0.156	0.145	0.055	0.061	0.069	0.085	0.090	0.089	0.072	0.091	0.097
30. TepoztequellaspAF099699	0.150	0.141	0.047	0.051	0.047	0.077	0.081	0.082	0.065	0.081	0.087
31. TepoztequellarhizoideaKY952886	0.155	0.153	0.057	0.060	0.060	0.087	0.091	0.085	0.065	0.086	0.090

Fig. 30. Captura de pantalla paso 16: porcentajes de distancia en la matriz triangular: (a) opciones para el aumento o reducción de dígitos en los valores y (b) valores reducidos a tres dígitos

17. Para guardar la matriz de distancia selecciona el menú **File** y el submenú **Export/Print Distances**, aparecerá un recuadro que te preguntará a que formato deseas exportarla. Selecciona la opción **Excel Workbook**. Presiona el botón **OK**, de inmediato te abrirá una hoja de Excel con la matriz de distancia. Guárdala como **Distancias genéticas no corregidas** en el directorio Prácticas de Filogenia, subcarpeta Práctica 3.
18. Cierra la matriz triangular de **distancias no corregidas**, verás la pantalla como en el paso 7.
19. Para calcular las **distancias corregidas**, repite los pasos 8 y 9 y a continuación el 20.
20. En el cuadro de diálogo de las opciones de análisis para calcular la distancia genética (Fig. 28), mantén los mismos parámetros que para el cálculo de distancias no corregidas, **EXCEPTO**, la selección del modelo para el cálculo de la distancia (Fig. 28c) y la tasa de sustitución de nucleótidos (Fig. 28e).
21. En la sección **Substitution model**, opción **Model/Method**, selecciona el modelo **Jukes-Cantor** (Fig. 31a). en la sección Rates and Patterns, opción **Rates and patterns**, selecciona **Gamma Distributed** (Fig. 31b). Presiona **OK**.

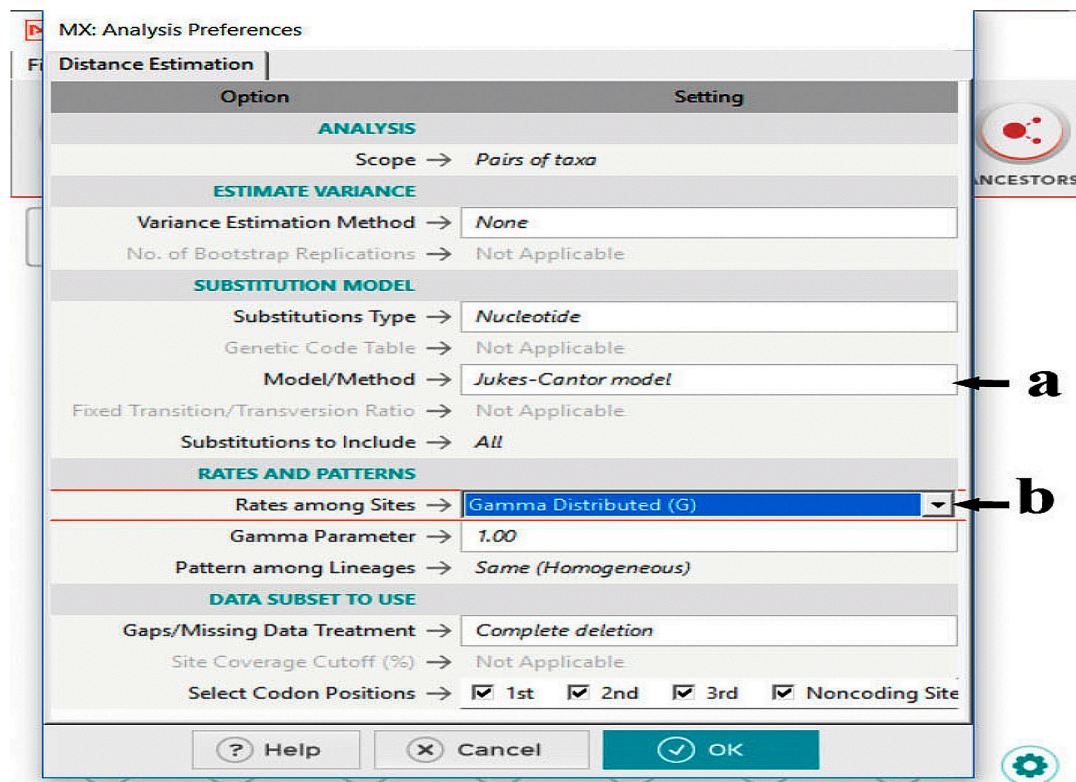


Fig. 31. Captura de pantalla paso 21: parámetros para el cálculo de distancias corregidas: (a) selección de un modelo de corrección y (b) selección de la tasa de variación entre sitios

22. Al igual que en el paso 15, te mostrará una matriz triangular de valores de distancias corregidas por el modelo Jukes-Cantor (Fig. 32).

Nota: notarás que habrá variaciones entre los valores de distancia corregida y distancia no corregida, pueden ser mínimos o significativos. La distancia genética corregida por el modelo de Jukes-Cantor entre *Agardhiella ramosissima* (taxa 3) y *Meristotheca cylindrica* (taxa 19), es de 5.2 % (Fig. 32, flecha).

MX: Pairwise Distances (rbcL_Solieriaceae.fas)

File Display Average Caption Help

	1	2	3	4	5	6	7	8	9
1. ChondrophycuspapillosusAF465807									
2. GracilariablabbiformeAY049343	0.189								
3. AgardhiellaramosissimaAF099680	0.183	0.168							
4. AgardhiellasubulataU04176	0.185	0.166	0.011						
5. AnathecamontagneiAB122015	0.175	0.166	0.041	0.045					
6. BetaphycusgelatinusJX069174	0.162	0.175	0.079	0.080	0.075				
7. BetaphycusphilippinensisAF099684	0.162	0.177	0.080	0.082	0.074	0.011			
8. BetaphycusspeciosumAF099685	0.170	0.172	0.074	0.075	0.077	0.038	0.042		
9. EucheumatopsisiformisAF099691	0.162	0.173	0.051	0.055	0.053	0.072	0.075	0.061	
10. KappaphycusalvareziiJX069175	0.175	0.183	0.079	0.079	0.082	0.082	0.077	0.082	0.077
11. KappaphycuscottoniiAF099695	0.179	0.185	0.086	0.086	0.089	0.089	0.088	0.089	0.083
12. KappaphycusstriatumJX623999	0.173	0.185	0.082	0.082	0.085	0.085	0.083	0.085	0.080
13. MeristothecapapulosaAF099700	0.175	0.175	0.072	0.074	0.064	0.094	0.096	0.099	0.072
14. MeristothecaprocumbensKF026498	0.183	0.179	0.074	0.075	0.064	0.089	0.094	0.097	0.074
15. MeristothecaimbricataAB290846	0.179	0.166	0.068	0.069	0.056	0.092	0.097	0.091	0.072
16. MeristothecadakarensisAB159224	0.177	0.168	0.065	0.072	0.067	0.088	0.099	0.096	0.069
17. MeristothecacoactaAB259300	0.185	0.177	0.069	0.071	0.058	0.096	0.100	0.094	0.074
18. MeristothecaspAF099703	0.183	0.173	0.071	0.072	0.064	0.100	0.105	0.099	0.075
19. MeristothecacylindricaKX553938	0.177	0.160	0.052	0.056	0.052	0.079	0.083	0.077	0.058
20. MeristothecagelidiumAF099697	0.172	0.166	0.061	0.059	0.065	0.085	0.089	0.080	0.064
21. SarcodithecaaudichaudiiKM253822	0.177	0.164	0.062	0.064	0.069	0.091	0.092	0.097	0.085
22. SarcodithecafurcataKM253835	0.179	0.168	0.053	0.055	0.064	0.083	0.089	0.086	0.074
23. SarconemafiliformeAF099708	0.170	0.172	0.071	0.075	0.072	0.094	0.096	0.092	0.071
24. SolieriachordalisAF099709	0.185	0.173	0.065	0.067	0.069	0.096	0.092	0.092	0.071
25. SolieriafiliformisKJ202089	0.183	0.162	0.065	0.069	0.052	0.082	0.085	0.083	0.072
26. Solieria incurvataMG018944	0.177	0.150	0.064	0.068	0.055	0.086	0.083	0.088	0.072
27. SolieriapacificaAF099710	0.187	0.189	0.067	0.068	0.065	0.088	0.089	0.083	0.075
28. SolieriarobustaKC130210	0.195	0.173	0.061	0.062	0.059	0.092	0.096	0.086	0.068
29. TacanooscauncinataAF099693	0.197	0.179	0.059	0.067	0.075	0.096	0.102	0.100	0.080
30. TepoztequiellaspAF099699	0.187	0.173	0.051	0.055	0.051	0.086	0.091	0.092	0.071
31. TepoztequiellarhizoideaKY952886	0.195	0.193	0.062	0.065	0.065	0.099	0.104	0.096	0.071

Fig. 32. Captura de pantalla paso 22: matriz triangular con los valores de distancia corregida por el modelo Jukes.Cantor

23. Guarda tu matriz de distancias corregidas igual que en el paso 17. No olvides cambiar el nombre del archivo a **Distancias genéticas corregidas**.
24. Cierra la matriz triangular de distancias corregidas, nuevamente regresarás al programa con los dos archivos activos.
25. Para calcular el modelo de evolución que utilizaremos para generar nuestros próximos análisis filogenéticos, selecciona del menú principal la opción **Models**, te desplegará un submenú en donde podrás seleccionar la opción **Find Best DNA/Protein Model** (Fig. 33).

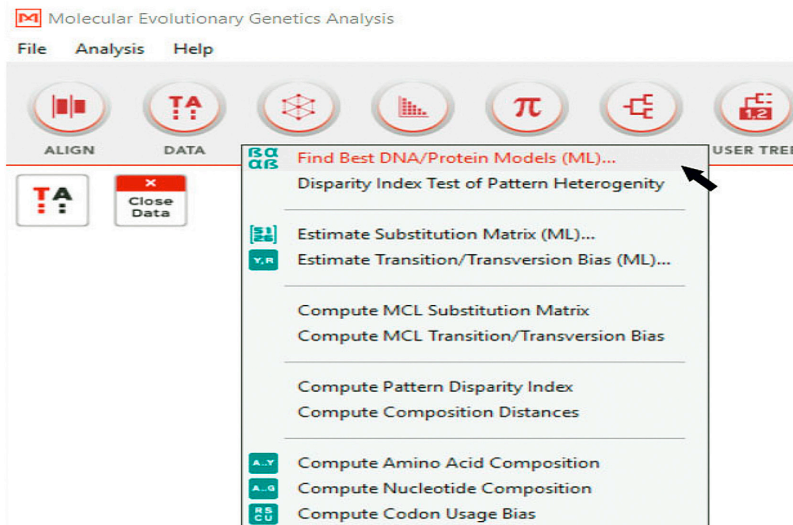


Fig. 33. Captura de pantalla paso 25: cálculo de modelos de evolución molecular

26. Nuevamente, como en el paso 9 te preguntará si quieres usar los datos activos, selecciona la opción **OK**.
27. Te mostrará un cuadro llamado **Analysis Preferences** en donde podrás modificar las opciones de análisis para la selección del modelo (Fig. 34).
28. En la sección **Analysis**, opción **tree to use**, selecciona **Neighbor-joining tree** (Fig. 34a). En la sección **Substitution Model**, opción **substitution type**, selecciona la opción **Nucleotide** (Fig. 34b). en la sección **Data subset to use**, opción **Gaps/Missing Data Treatment**, selecciona **Complete Deletion** (Fig. 34c). Para el resto de los datos considera los parámetros que asigna el programa de manera automática. Da clic en el botón **OK**.

Option	Setting
ANALYSIS	
Tree to Use →	Automatic (Neighbor-joining tree) ← a
User Tree File →	Not Applicable
Statistical Method →	Maximum Likelihood
SUBSTITUTION MODEL	
Substitutions Type →	Nucleotide ← b
Genetic Code Table →	Not Applicable
DATA SUBSET TO USE	
Gaps/Missing Data Treatment →	Complete deletion ← c
Site Coverage Cutoff (%) →	Not Applicable
Select Codon Positions →	<input checked="" type="checkbox"/> 1st <input checked="" type="checkbox"/> 2nd <input checked="" type="checkbox"/> 3rd <input checked="" type="checkbox"/> Noncoding Sites
Branch Swap Filter →	None
SYSTEM RESOURCE USAGE	
Number of Threads →	7

Fig. 34. Captura de pantalla paso 28: preferencias para el cálculo de modelos de evolución nucleotídica: (a) tipo de análisis, (b) tipo de sustitución y (c) tratamiento de los GAPS

29. Te mostrará un cuadro de análisis con el porcentaje de avance (Fig. 35).

MX: Progress

Progress

Details ✖ Stop

Status/Options

Run Status

Start time	07/09/2018 18:25:23
Status	Making initial tree
Thread-3	T3 + I
Thread-6	TN93 + G + I
Thread-1	HKY + G + I
Thread-2	HKY + G

Analysis Options

```

Tree to Use           : Automatic (Neighbor-joining tree)
Statistical Method    : Maximum Likelihood
Substitution Model
Substitutions Type    : Nucleotide
Data Subset to Use
Gaps/Missing Data Treatment : Complete deletion
Select Codon Positions : 1st,2nd,3rd,Non-Coding
Branch Swap Filter    : None
System Resource Usage
Number of Threads     : 7
    
```

Fig. 35. Captura de pantalla paso 29: monitoreo del porcentaje de avance en el análisis

30. Como resultado del análisis, te mostrará una nueva ventana con un listado de modelos y los valores de los diferentes parámetros estimados para cada uno. Los modelos estarán enlistados en orden de prioridad, es decir, el modelo que mejor explica el conjunto de datos proporcionado está en primer lugar y tendrá el valor de **BIC** más bajo (Fig. 36, flecha). En la leyenda inferior de la tabla, explica los parámetros medidos, los valores considerados y la nomenclatura para los modelos. De acuerdo con el cálculo del modelo, el mejor modelo para explicar la variación de nuestros datos es **TN93+G+I** (Tajima Nei 93 + Gamma+Sítios Invariables).

MEGA Caption Expert: Find Best-Fit Substitution Model (ML)

File Edit View Help

TAT

Results

Table. Maximum Likelihood fit of 24 different nucleotide substitution models

Model	Parameters	BIC	AICc	lnL	(+I)	(+G)	R	f(A)	f(T)	f(C)	f(G)	r(AT)	r(AC)	r(AG)	r(TA)	r(TC)	r(TG)	r(CA)	r(CT)	r(CG)	r(GA)	r(GT)	r(GC)
TN93+G+I	66	9958.587	9422.952	-4645.298	0.47	0.67	3.88	0.293	0.330	0.149	0.227	0.031	0.014	0.103	0.027	0.180	0.021	0.027	0.398	0.021	0.133	0.031	0.014
TN93+G	65	9962.184	9434.659	-4652.157	n/a	0.21	3.83	0.293	0.330	0.149	0.227	0.031	0.014	0.104	0.027	0.178	0.021	0.027	0.396	0.021	0.134	0.031	0.014
T92+G+I	63	9969.241	9457.937	-4665.806	0.50	0.71	3.88	0.312	0.312	0.188	0.188	0.030	0.018	0.152	0.030	0.152	0.018	0.030	0.251	0.018	0.251	0.030	0.018
GTR+G+I	69	9971.946	9411.980	-4636.795	0.45	0.64	3.98	0.293	0.330	0.149	0.227	0.044	0.012	0.105	0.039	0.182	0.012	0.025	0.403	0.014	0.136	0.018	0.009
T92+G	62	9975.798	9472.605	-4674.145	n/a	0.20	3.86	0.312	0.312	0.188	0.188	0.031	0.018	0.151	0.031	0.151	0.018	0.031	0.251	0.018	0.251	0.031	0.018
GTR+G	68	9976.319	9424.463	-4644.042	n/a	0.22	3.93	0.293	0.330	0.149	0.227	0.044	0.013	0.105	0.039	0.181	0.013	0.025	0.401	0.015	0.136	0.019	0.010
HKY+G+I	65	9995.163	9467.638	-4668.646	0.51	0.71	4.12	0.293	0.330	0.149	0.227	0.030	0.014	0.186	0.027	0.122	0.021	0.027	0.270	0.021	0.239	0.030	0.014
HKY+G	64	10000.456	9481.041	-4676.353	n/a	0.20	4.10	0.293	0.330	0.149	0.227	0.031	0.014	0.185	0.027	0.122	0.021	0.027	0.269	0.021	0.239	0.031	0.014
T92+I	62	10066.941	9563.748	-4719.717	0.64	n/a	3.56	0.312	0.312	0.188	0.188	0.033	0.020	0.149	0.033	0.149	0.020	0.033	0.247	0.020	0.247	0.033	0.020
GTR+I	68	10083.551	9531.695	-4697.658	0.63	n/a	3.55	0.293	0.330	0.149	0.227	0.048	0.014	0.123	0.043	0.163	0.014	0.027	0.361	0.017	0.159	0.020	0.011
TN93+I	65	10092.102	9564.576	-4717.115	0.62	n/a	3.34	0.293	0.330	0.149	0.227	0.034	0.015	0.085	0.030	0.186	0.024	0.030	0.412	0.024	0.110	0.034	0.015
HKY+I	64	10098.769	9579.354	-4725.509	0.64	n/a	3.70	0.293	0.330	0.149	0.227	0.033	0.015	0.182	0.029	0.119	0.023	0.029	0.264	0.023	0.234	0.033	0.015
K2+G+I	62	10110.619	9607.426	-4741.555	0.45	0.63	3.23	0.250	0.250	0.250	0.250	0.030	0.030	0.191	0.030	0.191	0.030	0.030	0.191	0.030	0.191	0.030	0.030
K2+G	61	10120.977	9625.895	-4751.795	n/a	0.23	3.18	0.250	0.250	0.250	0.250	0.030	0.030	0.190	0.030	0.190	0.030	0.030	0.190	0.030	0.190	0.030	0.030
K2+I	61	10213.627	9718.545	-4798.120	0.63	n/a	2.99	0.250	0.250	0.250	0.250	0.031	0.031	0.187	0.031	0.187	0.031	0.031	0.187	0.031	0.187	0.031	0.031
JC+G+I	61	10608.918	10113.836	-4995.766	0.45	0.72	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC+G	60	10615.720	10128.749	-5004.227	n/a	0.25	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC+I	60	10698.391	10211.421	-5045.563	0.63	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
GTR	67	10800.766	10257.020	-5061.326	n/a	n/a	2.13	0.293	0.330	0.149	0.227	0.071	0.012	0.095	0.063	0.156	0.027	0.024	0.346	0.027	0.122	0.039	0.018
TN93	64	10807.020	10287.606	-5079.635	n/a	n/a	2.76	0.293	0.330	0.149	0.227	0.040	0.018	0.092	0.035	0.171	0.027	0.035	0.378	0.027	0.118	0.040	0.018
T92	61	10877.652	10382.570	-5130.132	n/a	n/a	2.72	0.312	0.312	0.188	0.188	0.040	0.024	0.140	0.040	0.140	0.024	0.040	0.232	0.024	0.232	0.040	0.024
K2	60	10894.312	10407.341	-5143.523	n/a	n/a	2.69	0.250	0.250	0.250	0.250	0.034	0.034	0.182	0.034	0.182	0.034	0.034	0.182	0.034	0.182	0.034	0.034
HKY	63	10930.736	10419.432	-5146.553	n/a	n/a	2.72	0.293	0.330	0.149	0.227	0.042	0.019	0.170	0.037	0.111	0.029	0.037	0.246	0.029	0.219	0.042	0.019
JC	59	11352.601	10873.742	-5377.728	n/a	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083

Fig. 36. Captura de pantalla paso 30: listado de parámetros y cálculos para la selección del mejor modelo

31. Guarda el listado de modelos de evolución y sus valores en una hoja de Excel, igual que en el paso 17. Nombra a tu archivo como **ModelosEvolucion**.

RESULTADOS











Al final de la práctica, obtendrás tres archivos de Excel: dos matrices triangulares de distancia, una de **distancias no corregidas.exe** y una de **distancias corregidas.exe**; el tercer archivo corresponderá a los modelos de evolución calculados llamado **ModelosEvolucion.exe**. Estos archivos serán utilizados en las prácticas futuras para la inferencia filogenética.

ACTIVIDADES COMPLEMENTARIAS

- Indica el porcentaje de divergencia o la distancia genética **no corregida**, entre *Meristotheca cylindrica* y cada una de las especies restantes de *Meristotheca*.
- Genera una matriz de distancia genética corregida utilizando el modelo **Tamura-Nei**.

3. Indica el porcentaje de distancia genética **corregida** con los modelos Jukes-Cantor y Tamura Nei, entre *Meristothecha cylindrica* y cada una de las especies restantes de *Meristothecha*.
4. Discute las diferencias entre distancias corregidas y no corregidas y entre distancias corregidas con dos modelos diferentes.
5. Calcula el mejor modelo de evolución nucleotídica considerando los GAPS. Discute si hay diferencias en el mejor modelo considerando GAPS o No.

BIBLIOGRAFÍA RECOMENDADA

-  Jukes T. & Cantor C.R. 1969. "Evolution of protein molecules". In: *Mamalian protein metabolism* (Ed. by H.N. Munro), pp. 21–132. Academic press, New York, USA.
-  Kimura M. 1980. "A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences". *Journal of Molecular Evolution* 16: 111–120.
-  Kimura M. 1981. "Estimation of evolutionary distances between homologous nucleotide sequences". *Proceedings of the National Academy of Sciences USA* 78: 454–458.
-  Kumar S., Stecher G., Li M., Knyaz C. & Tamura K. 2018. "MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms". *Molecular Biology and Evolution* 35: 1547–1549.
-  Nei M. & Li W.H. "1979. Mathematical model for studying genetic variation in terms of restriction endonucleases". *Proceedings of the National Academy of Sciences USA* 76: 5269–5273.
-  Posada D. 2001. "The effect of branch length variation on the selection of models of molecular evolution". *Journal of Molecular Evolution* 52: 434–444.
-  Posada D. & Crandall K.A. 1998. "MODELTEST: testing the model of DNA substitution". *Bioinformatics* 14: 817–818.
-  Tajima F.M.N. 1984. "Estimation of evolutionary distance between nucleotide sequences". *Molecular Biology and Evolution* 1: 269–285.
-  Tamura K. & Nei M. 1993. "Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees". *Molecular Biology and Evolution* 10: 512–526.
-  Upholt W.B. 1977. "Estimation of DNA sequence divergence from comparison of restriction endonuclease digests". *Nucleic Acids Research* 4: 1257–1265.

Práctica 4

Construcción de árboles filogenéticos I: UPGMA, Neighbor-Joining y Máxima Parsimonia con MEGA-X

INTRODUCCIÓN

Un **árbol filogenético** es un **dendograma** que muestra las relaciones evolutivas entre diferentes taxa que tienen un **ancestro común**. Actualmente existen diversos métodos para construir árboles filogenéticos, dichos métodos asumen que los **grupos monofiléticos** son indicadores de las relaciones evolutivas, sin embargo, difieren en el manejo de un mismo conjunto de datos y en los procesos de construcción de una hipótesis filogenética. Estos métodos de análisis se dividen en dos tipos principales: los métodos de distancia y los métodos discretos.

Los **métodos de distancia** consisten en relacionar la distancia genética entre los taxa con su relación filogenética, asumiendo que dicha distancia es reflejo de la relación entre taxa; sin embargo, dicha relación es válida únicamente en casos con tasas de cambio constantes y ausencia de **homoplasias**. Al emplear estos métodos usando datos moleculares, es posible emplear modelos evolutivos para corregir las distancias calculadas considerando la homoplasia, disminuyendo así estas imprecisiones. Los métodos de distancia más comunes son: el método de **UPGMA** (Unweighted Pair Group Method Using Arithmetic Average o Pares No Ponderados Utilizando Media Aritmética) y el método de **Neighbor-Joining o vecino más cercano** (NJ).

Los **métodos discretos** o métodos de búsqueda de árboles basados en criterios de optimización están basados directamente en el análisis de los caracteres. Estos métodos cuentan los eventos de mutación acumulados en las secuencias y pueden eliminar la pérdida de información que se da cuando los caracteres son transformados a distancias. Estos métodos a su vez utilizan fundamentalmente dos tipos de métodos en la construcción de árboles filogenéticos: búsquedas exactas y búsquedas heurísticas o estocásticas. Las **búsquedas exactas**, utilizan algoritmos (exhaustivos o **Branch and bound (corte y empalme)**) que garantizan la obtención de todos los árboles óptimos a partir de un conjunto de datos; sin embargo, estas búsquedas están limitadas a un conjunto de datos no mayor a 20 taxa. Las **búsquedas heurísticas** utilizan búsquedas limitadas conforme a un criterio seleccionado (**stepwise addition (adición de pasos)**, **Branch swapping (intercambio de ramas)**, **tree bisection and reconnection (bisección y reconexión de ramas)**, etc.), se utilizan cuando, por el exceso de taxa, no es posible aplicar los algoritmos exactos. Sin embargo, no localizan necesariamente los árboles óptimos o al menos no todos. Los métodos discretos más comunes son: **Máxima Parsimonia** (MP), **Máxima Verosimilitud** (ML) y la **Inferencia Bayesiana** (BI).

UPGMA

El método de UPGMA es un algoritmo de **agrupamiento jerárquico** que, a partir de una matriz de distancias genera un árbol filogenético. Este método inicia con la agrupación de los dos taxa con la menor distancia, las distancias entre este nuevo taxon compuesto y el resto de los taxa se recalculan para añadir al siguiente par de grupos más cercanos. Esta iteración continúa hasta que todos los taxa son colocados en el árbol, de este modo, el último grupo añadido se considera como el grupo más lejano o **grupo externo**.

Neighbor-Joining (Vecino más cercano)

El método de Neighbor-Joining es un algoritmo de agrupamiento para reconstruir árboles filogenéticos a partir de datos evolutivos de distancia. Se basa en el algoritmo de **evolución mínima** y el principio de este método es encontrar pares (=vecinos) de unidades taxonómicas operacionales (OTU's) que minimicen la longitud total de la rama en cada etapa de agrupación de OTU's, comenzando con un árbol en forma de estrella. Este método parte de una matriz de distancias entre cada par de taxa, el primer paso será modificar dicha matriz de forma que la distancia, entre cada par de terminales, dependa tanto de la propia distancia como de la distancia de ambos con respecto al resto de terminales. Es importante destacar que este método genera árboles sin **raíz**, por lo que es importante incluir **grupos externos** que puedan servir de punto de partida para situar (enraizar) al **ancestro común**.

Máxima Parsimonia

El método de Máxima Parsimonia tiene como objetivo la búsqueda de un posible árbol (o conjunto de árboles) filogenético que explique la totalidad de un conjunto de datos con el menor número de pasos (menor número de cambios evolutivos). Se basa en el principio conocido como Navaja de Occam, el cual establece que, ante distintas soluciones posibles, la más sencilla es probablemente la correcta. Siguiendo este principio, un árbol con el menor número de sustituciones es probablemente la mejor opción para explicar las diferencias entre los taxa. Esta perspectiva se justifica por el hecho de que los cambios evolutivos que suceden dentro de lapsos de tiempo cortos son relativamente raros. La construcción del árbol filogenético de MP inicia contabilizando el número de **transformaciones** entre caracteres en el árbol y buscando todas las posibles topologías que requieren el mínimo número de cambios evolutivos, el árbol con el menor número de cambios es el elegido como el mejor. Sin embargo, en este análisis, sólo un pequeño número de **sitios** (sitios informativos) que tienen información filogenética importante son usados en la determinación del árbol. Los **sitios informativos** son sitios en una secuencia que tienen al menos dos tipos diferentes de caracteres, cada uno ocurriendo al menos dos veces. Por otro lado, los **sitios conservados** (sitios constantes o con cambios que ocurren sólo una vez) o los sitios muy variables, son los **sitios no informativos**, estos últimos no son considerados en los análisis de MP.

Los análisis de distancia (UPGMA y NJ) y el análisis discreto de MP pueden ser llevados a cabo con el *software* libre MEGA-X.

OBJETIVO

- Construir árboles filogenéticos mediante los métodos de UPGMA, Neighbor-Joining y Máxima Parsimonia, utilizando como herramienta el programa MEGA-X.

SOFTWARE REQUERIDO

1. MEGA-X Molecular Evolutionary Genetics Analysis, disponible para su descarga en el sitio del programa, siguiendo la liga <http://www.megasoftware.net>

DESARROLLO

1. Abre el programa MEGA-X. A continuación, abre tu matriz de datos **Solieriaceae.fas** siguiendo los **pasos 1-6** (Figs 19-24) descritos en la Práctica 3, hasta que aparezca tu archivo de datos, en formato MEG, cargado en la pantalla principal del programa (como aparece en el paso 7 de la práctica 3).
2. De la barra de herramientas de análisis, da clic sobre la opción **PHYLOGENY**, al hacerlo será desplegado un nuevo submenú con las diferentes opciones de análisis filogenéticos que puedes realizar con MEGA-X. Vamos a comenzar con los análisis de distancia, selecciona la opción **Construct Test/UPGMA Tree** (Fig. 37).

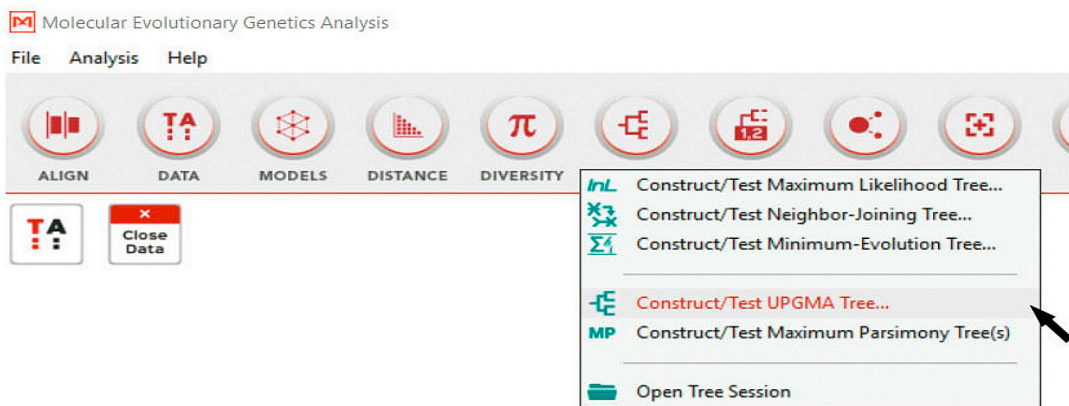


Fig. 37. Captura de pantalla paso 2: construcción de árboles filogenéticos con el método de UPGMA

3. Aparecerá un cuadro de diálogo preguntándote si quieres utilizar los datos actualmente activos, da clic en la opción **Yes** (Fig. 38).

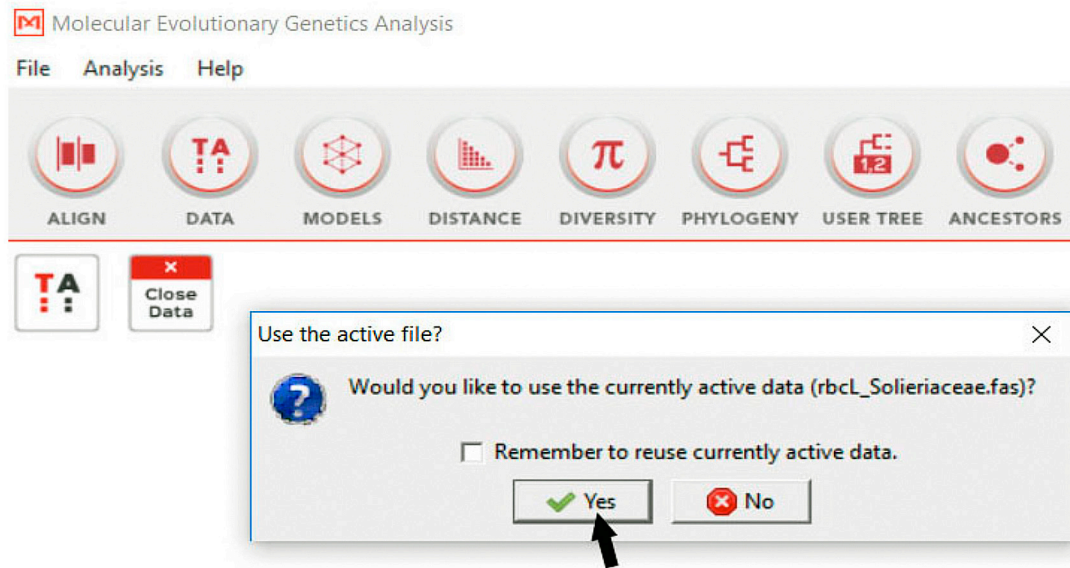


Fig. 38. Captura de pantalla paso 3: comando del programa para permitir el uso de los datos activos

4. Aparecerá un nuevo cuadro de diálogo preguntándote sobre las preferencias de análisis, subdividido a su vez en seis subsecciones (Fig. 39). En la sección **Analysis** te mostrará los taxa que incluirá en el análisis, en este caso todos los taxa, y el método de búsqueda **UPGMA** (Fig. 39a).
5. En la sección Phylogeny Test, selecciona la opción Bootstrap Method (Fig. 39b). A su vez, se activará una nueva opción No. of Bootstrap Replications, escribe el número 1000 (Fig. 39c).

Nota: para los árboles filogenéticos, se calcula una medida de **soporte de los clad**os que te indica la robustez y confiabilidad de la hipótesis que estos reflejan. **Bootstrap** es uno de los métodos más utilizados en filogenia, principalmente en MP y ML. Consiste en crear una matriz de datos cuyas columnas se van cambiando de forma aleatoria, e incluso se repiten en una misma matriz, y con cada una de las matrices resultantes crear un árbol filogenético, contando al final en cuántos de ellos aparece el clado de interés. Este proceso se repite varias veces, siendo lo más común repetirlo al menos 1000 veces, entre mayor número de réplicas sean corridas, incrementará el valor de soporte de los clados. Al final se indica el porcentaje de veces en que cada nodo del árbol se presentó en las matrices re-muestreadas. Al final todos estos árboles se representan en un árbol de consenso en el cual aparecen los porcentajes de la proporción de cada réplica. Así, entre más cercano sea el porcentaje a 100%, más robusta será la hipótesis representada.

6. En la sección **Substitution Model**, selecciona primero el tipo de datos a analizar, en este caso selecciona la opción **Nucleotide** (Fig. 39d). A continuación, selecciona el modelo de evolución, en tanto que el mejor modelo para explicar nuestros datos, calculado en la **Práctica 3**, fue **TN93+G+I**, selecciona el modelo **Tajima-Nei** (Fig. 39e); en tipo de sustituciones selecciona la opción **Transitions + Transversions** (Fig. 39f).
7. En la sección **Rates and Patterns** vamos a seleccionar la opción **Gamma Distributed**, ya que nuestro modelo calculado nos pide también una distribución Gamma (Fig. 39g), de inmediato te preguntará sobre los parámetros de gamma, selecciona la opción **1.00** (Fig. 39h).
8. En la sección **Data Subset to Use** selecciona el tipo de tratamiento de GAPS, selecciona la opción **Partial Deletion** (Fig. 39i). Para el resto de los parámetros, vamos a dejar activos los parámetros automáticos del programa. Presiona el botón OK.

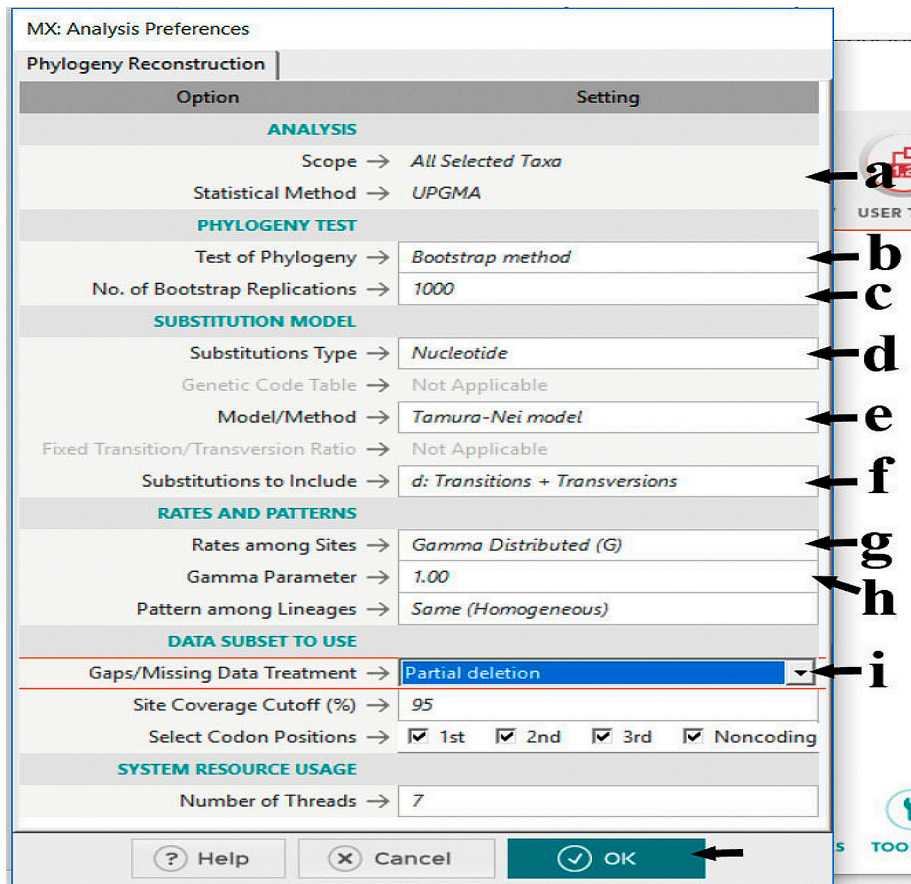


Fig. 39. Captura de pantalla pasos 4 a 8: preferencias de análisis para construir un árbol filogenético con UPGMA: (a) método de estimación, (b) test de filogenia, (c) asignación del número de réplicas, (d) tipos de sustitución, (e) modelo de estimación de distancia, (f) tipo de sustitución de nucleótidos, (g) distribución de las tasas de variación, (h) parámetros de gamma, (i) tratamiento de los GAPS y botón de ejecución del análisis

9. Aparecerá un cuadro de avance de la ejecución de tu análisis (Fig. 40).

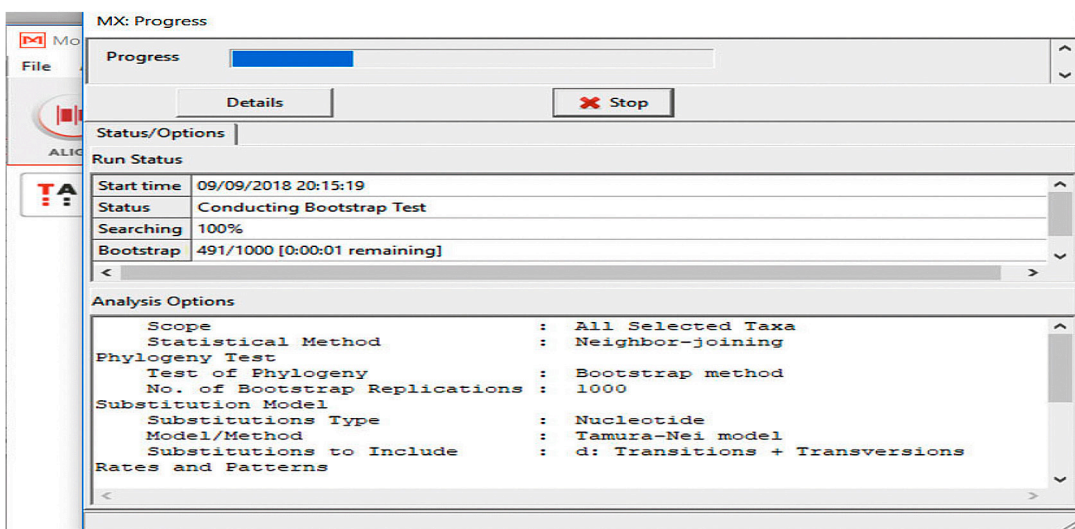


Fig. 40. Captura de pantalla paso 9: monitoreo del porcentaje de avance en el análisis

10. Al terminar el análisis, la pantalla te mostrará los resultados divididos en pestañas, en una primera pestaña el árbol óptimo (Fig. 41) y en una segunda pestaña el árbol de consenso de Bootstrap.

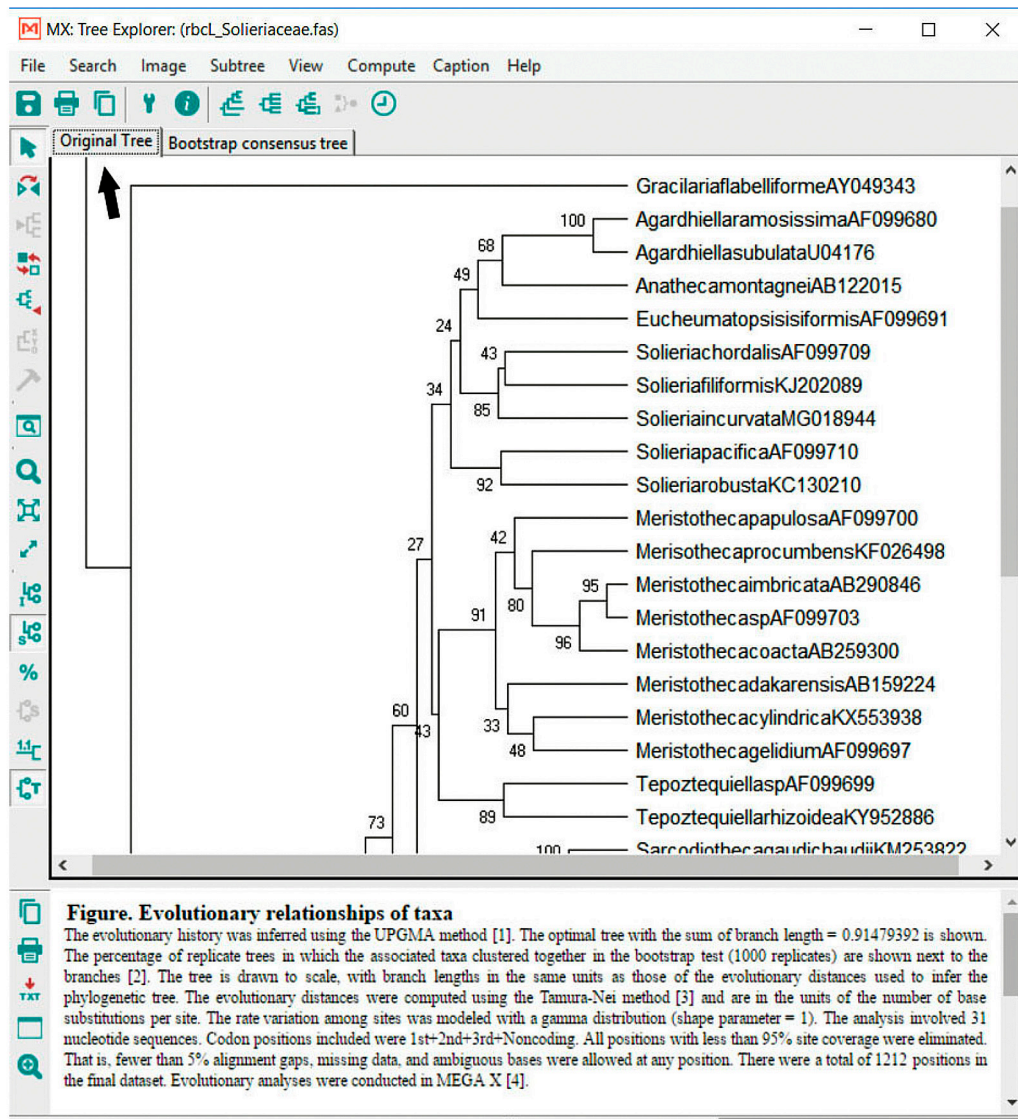


Fig. 41. Captura de pantalla paso 10: presentación de resultados del análisis con UPGMA mostrando el árbol óptimo

11. En una ventana doble, abierta en la parte inferior de la pantalla, te mostrará una leyenda de la figura que obtuvo en la que explicará detalladamente lo que la figura te está mostrando (Fig. 42b). Copia la información de esta leyenda en un archivo de block de notas en el directorio Prácticas de Filogenia carpeta Práctica 4, para que puedas consultarlo más adelante.

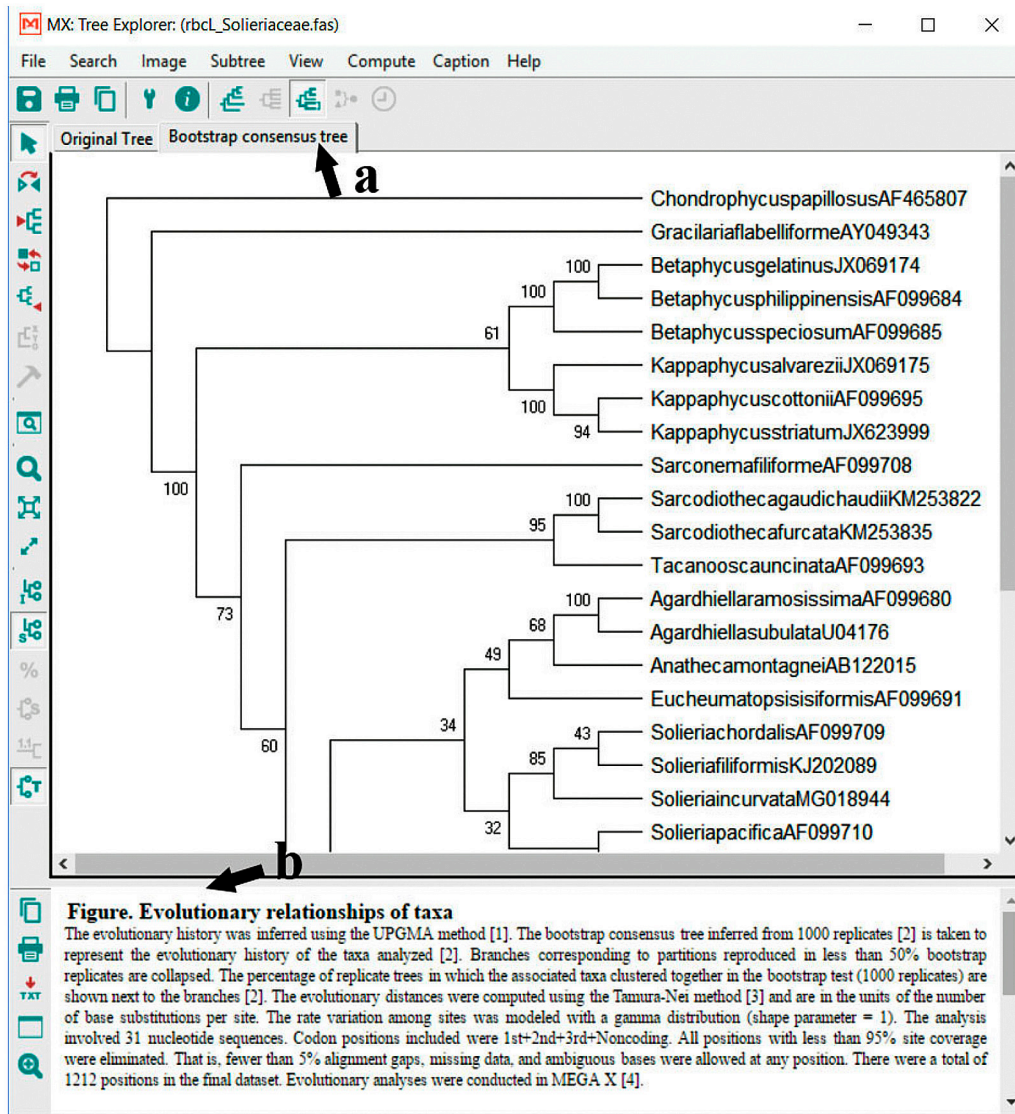


Fig. 42. Captura de pantalla pasos 10 y 11: (a) resultado para el árbol de consenso de Bootstrap y (b) pies de figura para la explicación del análisis

- Para guardar los árboles generados, en el menú principal en la parte superior de la pantalla, selecciona **Image**, da clic en la opción **Save as Image File** (Fig. 43). Te abrirá un pequeño cuadro de diálogo en el que deberás seleccionar el formato en que deseas guardar la imagen (Fig. 44) selecciona el formato **TIF**. A continuación, da clic en el ícono de archivo (Fig. 44a), para seleccionar el directorio donde quieres guardar la imagen. Guárdala en la carpeta **Práctica 4**, con el nombre **UPG-MA_Solieriaceae**, luego da clic en el botón **OK** (Fig. 44b).

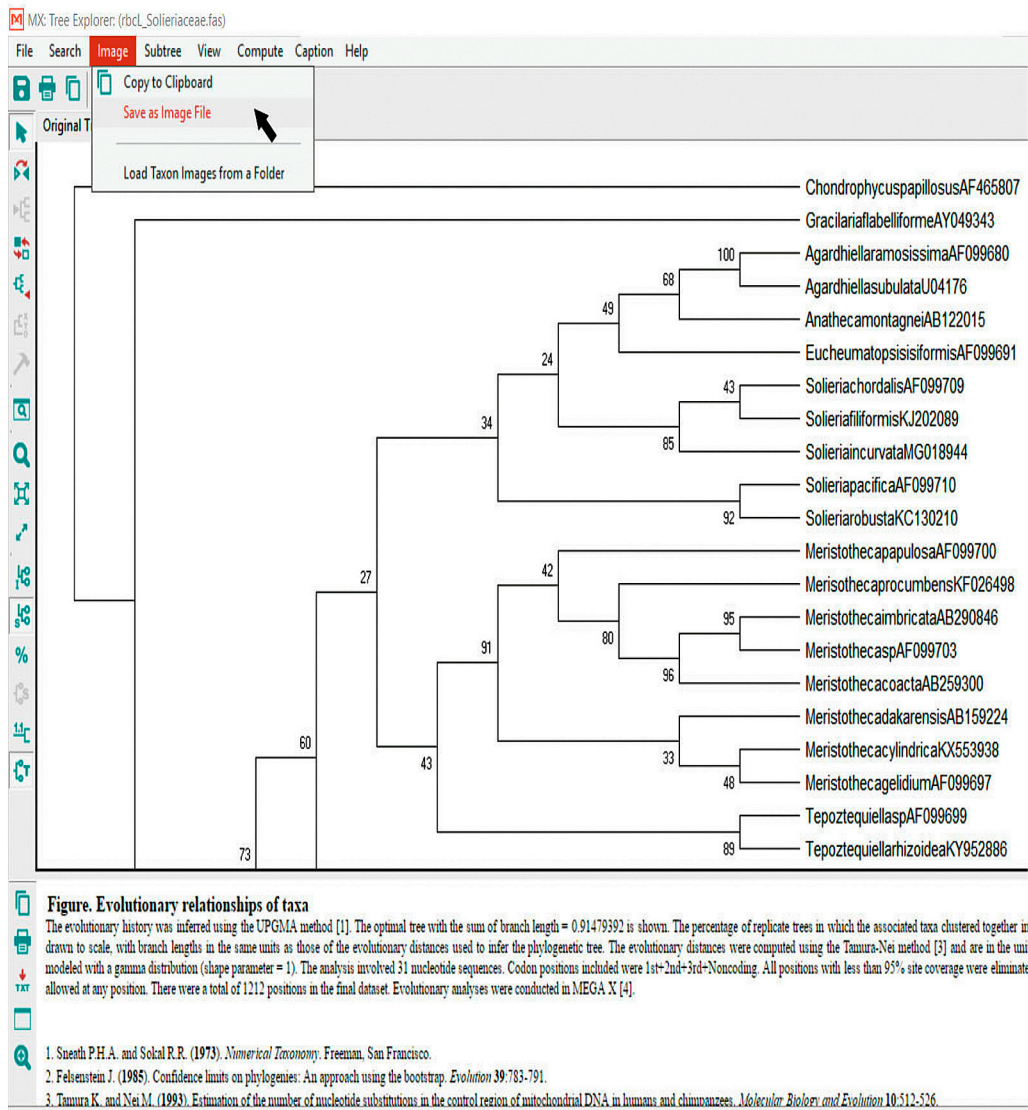


Fig. 43. Captura de pantalla paso 12: opciones para guardar árboles filogenéticos construidos en MEGA-X

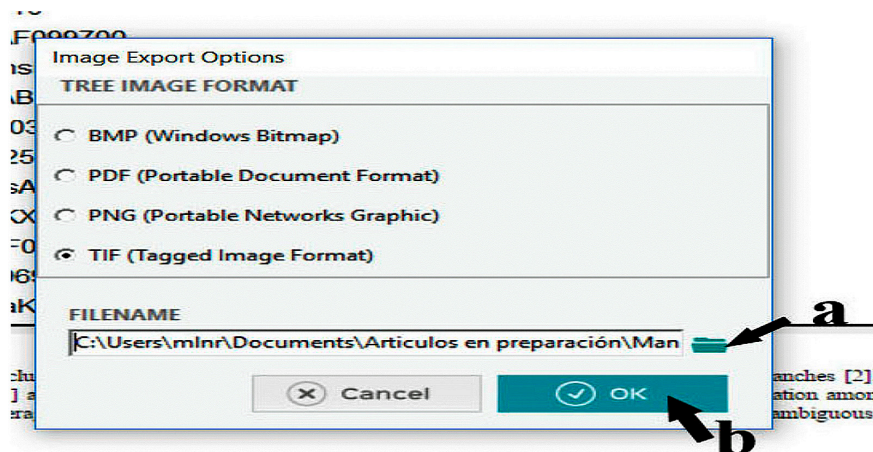


Fig. 44. Captura de pantalla paso 12: selección del formato de árboles filogenéticos construidos en MEGA-X

13. Guarda también el árbol de Bootstrap con el nombre **UPGMA-Bootstrap_Solieriaceae**.
14. Una vez que hayas guardado los dos árboles resultantes del análisis UPGMA, cierra la ventana de resultados. Regresarás a la ventana de inicio con la matriz de datos. En la barra de herramientas de análisis, vuelve a dar clic en la opción **PHYLOGENY**, ahora selecciona **Construct Test/Neighbor-Joining Tree** (Fig. 45).

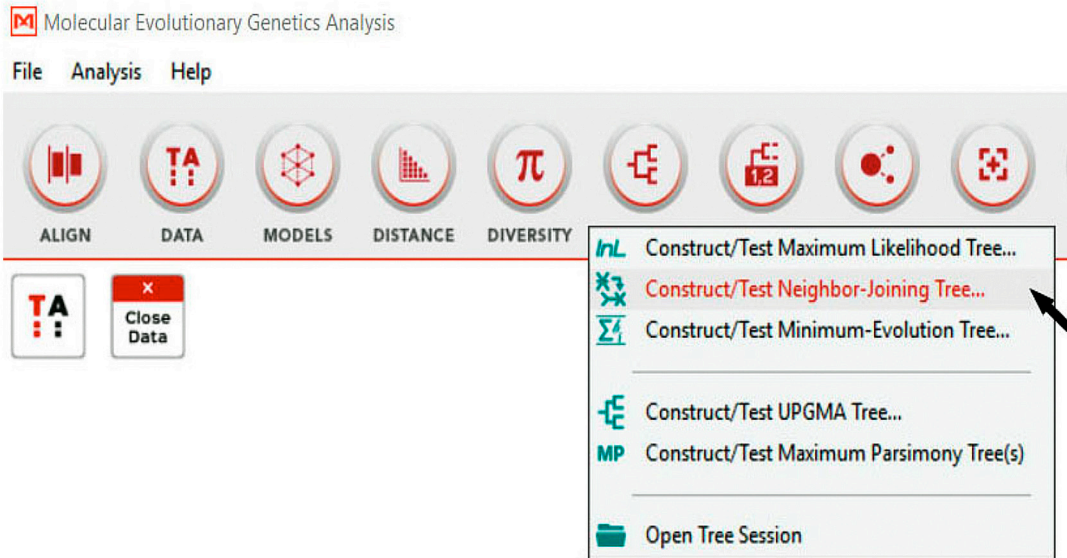


Fig. 45. Captura de pantalla paso 14: construcción de árboles filogenéticos por el método de Neighbor-Joining

15. Nuevamente da clic en el botón **Yes**, para utilizar la matriz de datos activa.
16. Aparecerá nuevamente el cuadro de diálogo preguntándote sobre las preferencias de análisis, como ya modificamos los parámetros para el análisis de **UPGMA** (pasos 4-8), el programa los recordará automáticamente y ya estarán asignados. Sin embargo, notarás que el método de búsqueda habrá cambiado a **Neighbor-joining**. También notarás que el primer parámetro **Test of Phylogeny**, no estará asignado; vuelve a seleccionar la opción **Bootstrap method** (Fig. 46, flecha). Da clic en el botón **OK**.

MX: Analysis Preferences

Phylogeny Reconstruction

Option	Setting
ANALYSIS	
Scope →	All Selected Taxa
Statistical Method →	Neighbor-joining
PHYLOGENY TEST	
Test of Phylogeny →	Bootstrap method
No. of Bootstrap Replications →	1000
SUBSTITUTION MODEL	
Substitutions Type →	Nucleotide
Genetic Code Table →	Not Applicable
Model/Method →	Tamura-Nei model
Fixed Transition/Transversion Ratio →	Not Applicable
Substitutions to Include →	d: Transitions + Transversions
RATES AND PATTERNS	
Rates among Sites →	Gamma Distributed (G)
Gamma Parameter →	1.00
Pattern among Lineages →	Same (Homogeneous)
DATA SUBSET TO USE	
Gaps/Missing Data Treatment →	Partial deletion
Site Coverage Cutoff (%) →	95
Select Codon Positions →	<input checked="" type="checkbox"/> 1st <input checked="" type="checkbox"/> 2nd <input checked="" type="checkbox"/> 3rd <input checked="" type="checkbox"/> Noncoding Site
SYSTEM RESOURCE USAGE	
Number of Threads →	7

? Help X Cancel ✓ OK

Fig. 46. Captura de pantalla paso 16: preferencias de análisis para la construcción de un árbol filogenético con Neighbor-Joining

17. Nuevamente te mostrará una pantalla con el resultado, en una pestaña el árbol óptimo y en otra pestaña el árbol de consenso de Bootstrap. También te mostrará la misma ventana inferior con la leyenda de figura con la explicación detallada. Del mismo modo que en los pasos 11-13, guarda la leyenda de figura y los dos árboles resultantes. Esta vez nómbralos **Neighbor-joining_Solieriaceae** y **Neighbor-joining-Bootstrap_Solieriaceae**.
18. Una vez que hayas guardado los dos árboles resultantes del análisis Neighbor-joining, cierra la ventana de resultados. Regresarás a la ventana de inicio con la matriz de datos. En la barra de herramientas de análisis, vuelve a dar clic sobre la opción **PHYLOGENY**, ahora selecciona la opción **Construct Test/Maximum Parsimony Tree(s)** (Fig. 47).

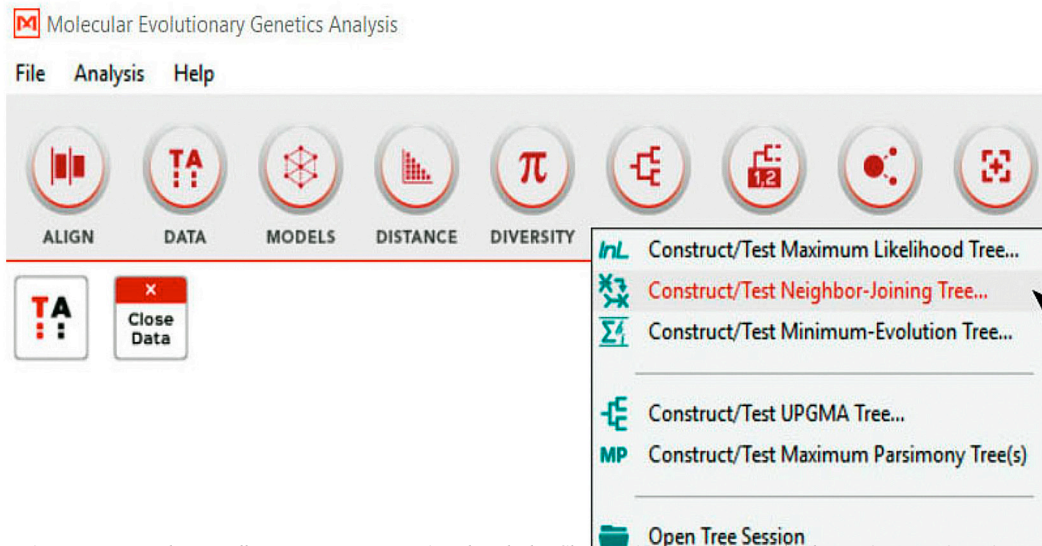


Fig. 47. Captura de pantalla paso 18: construcción de árboles filogenéticos por el método de Máxima Parsimonia

19. Nuevamente aparecerá el cuadro de diálogo preguntándote sobre las preferencias de análisis, esta vez algunas secciones estarán modificadas (Fig. 48). Notarás que el método de búsqueda habrá cambiado a **Maximum Parsimony**, tampoco el primer parámetro **Test of Phylogeny** estará asignado; vuelve a seleccionar la opción **Bootstrap method** (Fig. 48a). Una nueva sección habrá aparecido, **Tree Inference Options**. En esta sección deberás seleccionar las opciones de búsqueda que deseas para tu análisis, recuerda que esta opción no trabaja con distancias.
20. En el método de búsqueda de parsimonia selecciona la opción **Tree-Bisection.Reconnection (TBR)** (Fig. 48b). Para el resto de los parámetros, conserva los que el programa asignó automáticamente. Da clic en el botón **OK**.

MX: Analysis Preferences

Phylogeny Reconstruction

Option	Setting
ANALYSIS	
Statistical Method →	Maximum Parsimony
PHYLOGENY TEST	
Test of Phylogeny →	Bootstrap method
No. of Bootstrap Replications →	1000
SUBSTITUTION MODEL	
Substitutions Type →	Nucleotide
Genetic Code Table →	Not Applicable
DATA SUBSET TO USE	
Gaps/Missing Data Treatment →	Partial deletion
Site Coverage Cutoff (%) →	95
Select Codon Positions →	<input checked="" type="checkbox"/> 1st <input checked="" type="checkbox"/> 2nd <input checked="" type="checkbox"/> 3rd <input checked="" type="checkbox"/> Noncoding Sites
TREE INFERENCE OPTIONS	
MP Search Method →	Tree-Bisection-Reconnection (TBR)
No. of Initial Trees (random addition) →	10
MP Search level →	1
Max No. of Trees to Retain →	100
SYSTEM RESOURCE USAGE	
Number of Threads →	7

árbolos.r

REE

a

b

Help Cancel OK

Fig. 48. Captura de pantalla pasos 19 y 20: preferencias de análisis para la construcción de un árbol filogenético con Máxima Parsimonia: (a) test de parsimonia y (b) método de búsqueda

21. Te mostrará una ventana con el porcentaje de avance de tu análisis (Fig. 49). Notarás que este análisis tomará mayor tiempo que los dos anteriores, esto es porque el espacio de muestreo es mayor y la búsqueda más exhaustiva.

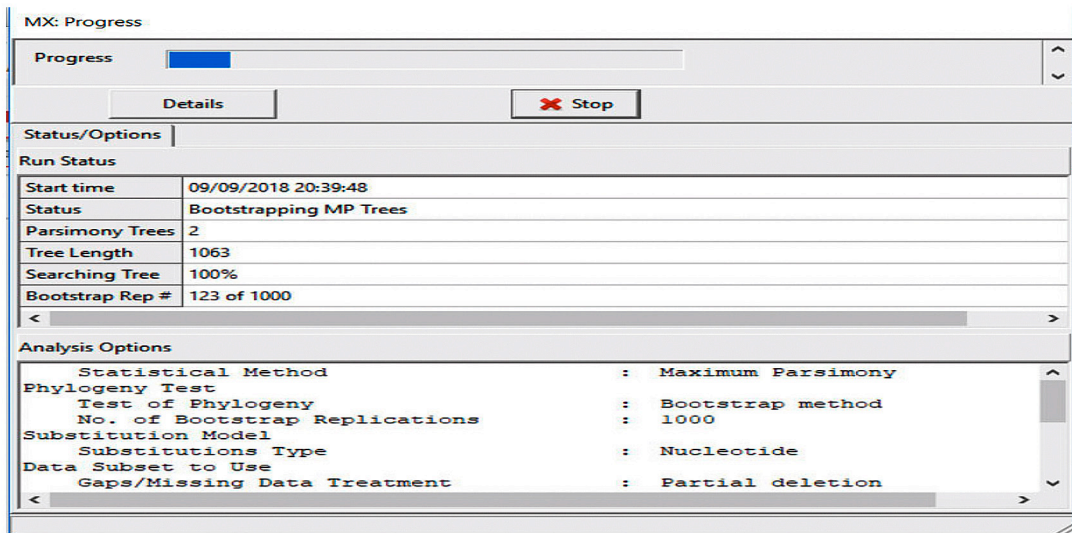


Fig. 49. Captura de pantalla paso 21: monitoreo del porcentaje de avance en el análisis

22. Al terminar el análisis se mostrará nuevamente el árbol óptimo y el árbol de consenso de Bootstrap. Si hubiera más de un árbol óptimo, que en MP puede haberlos, te mostraría todos los árboles igualmente parsimoniosos, entonces te mostraría más de un archivo.
23. Del mismo modo que en los pasos 11-13, guarda la leyenda de figura y los dos árboles resultantes. Esta vez nómbralos **MP_Solieriaceae** y **MP-Bootstrap_Solieriaceae**.
24. Podrás editar tus archivos de árbol en un editor gráfico de imagen como Paint, Power-Point, Adobe Photoshop o Corel Draw.

Nota: los árboles de Bootstrap no reflejan relaciones filogenéticas, sólo el soporte de las ramas, por lo que estos árboles no te servirán para inferencia filogenética, sólo te servirán para la asignación de valores a los nodos de las ramas en tus árboles. Por lo tanto, no los edites sólo úsalos para la edición.

RESULTADOS

Al final de la práctica, tendrás seis archivos de imagen TIF: dos árboles obtenidos con UPGMA (Figs. 50 y 51), dos árboles generados con Neighbor-joining (Figs 52 y 53) y dos árboles con Máxima Parsimonia (Figs. 54 y 55). Estos archivos serán utilizados en las prácticas futuras para la inferencia filogenética.

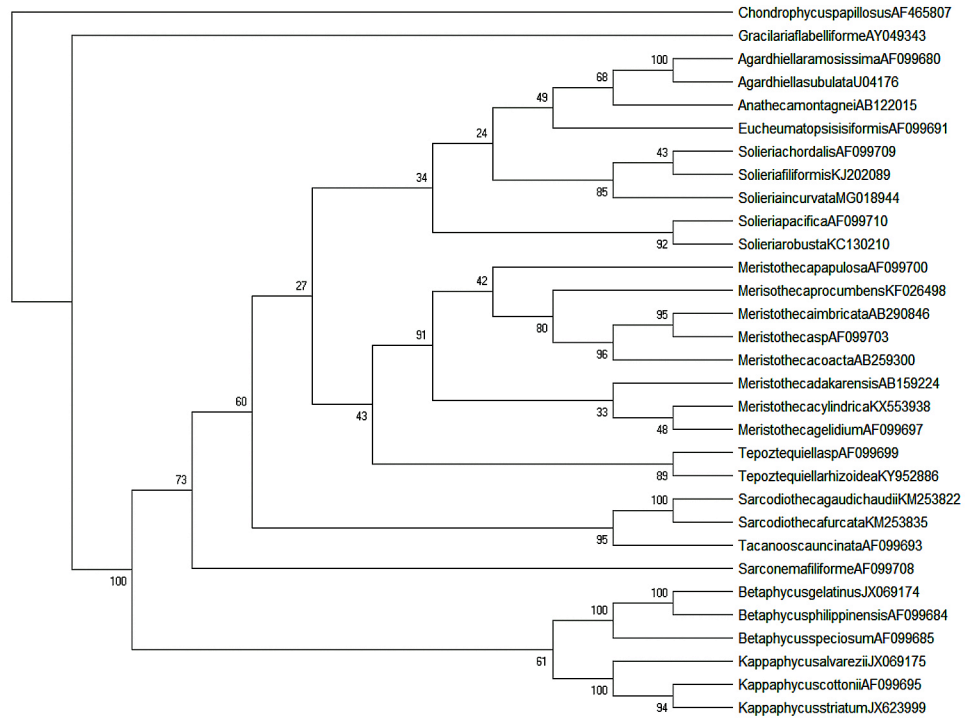


Fig. 50. Topología del árbol óptimo construida con el método de UPGMA

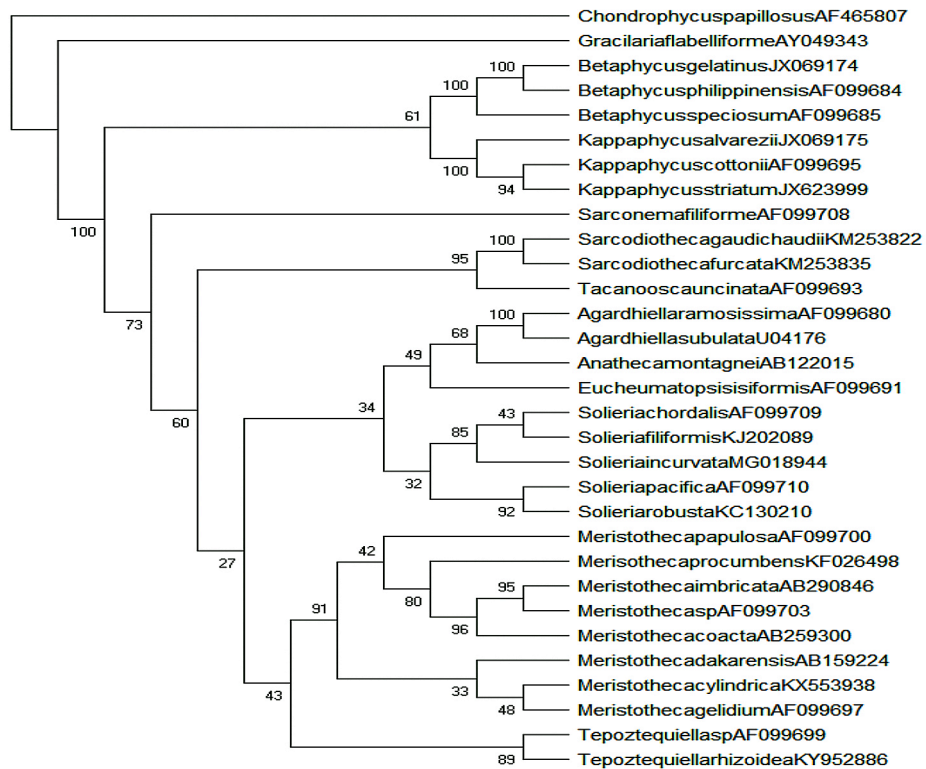


Fig. 51. Topología del árbol consenso de Bootstrap construida con el método de UPGMA

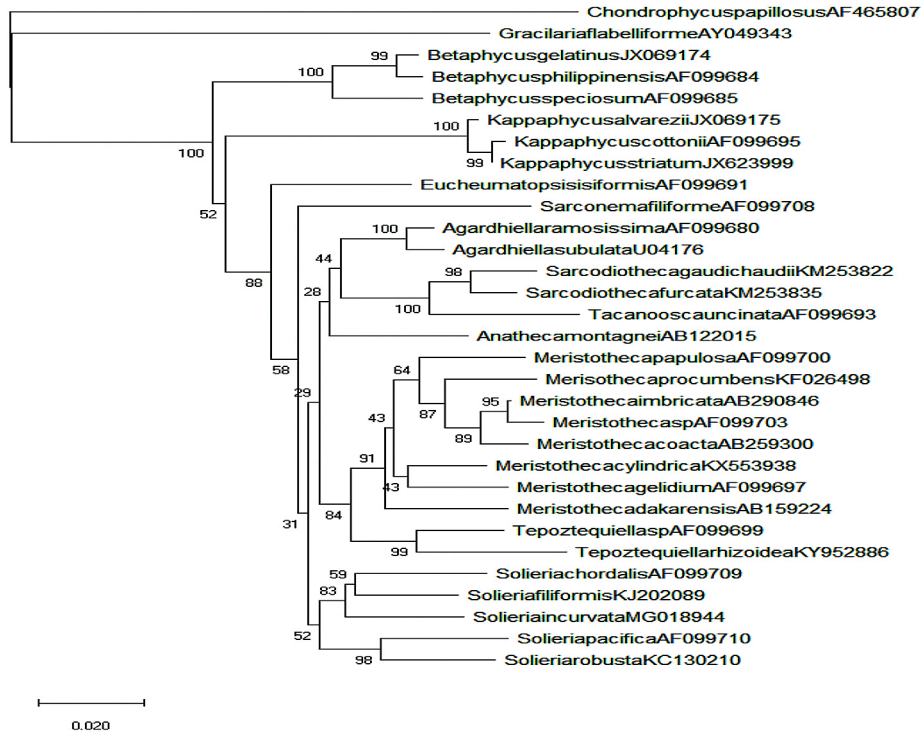


Fig. 52 Topología del árbol óptimo construida con el método de Neighbor-Joining

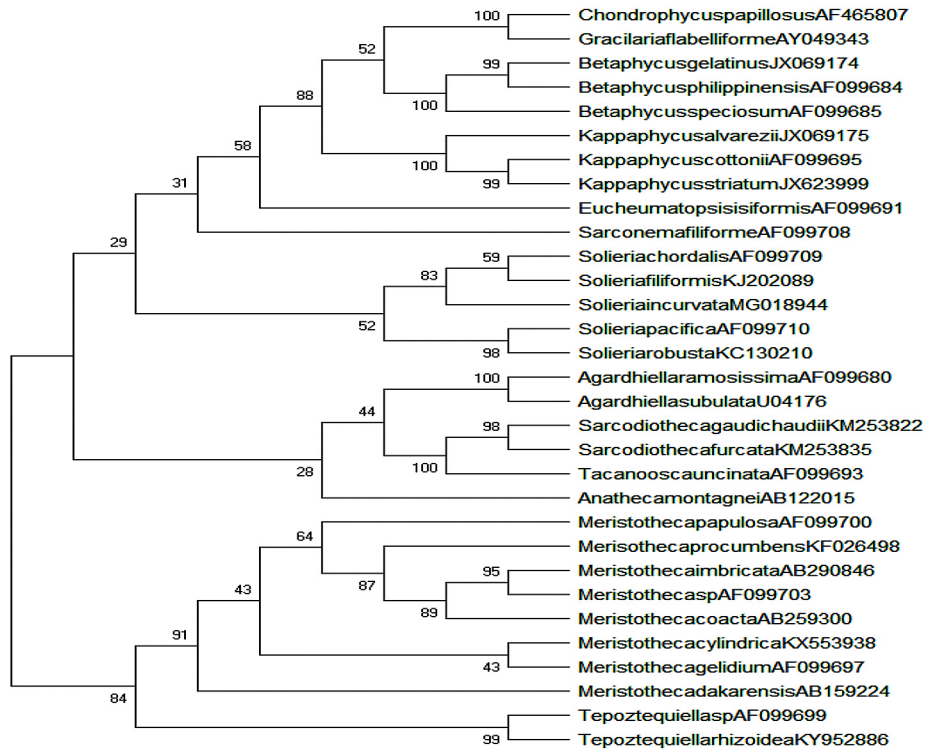


Fig. 53. Topología del árbol consenso de Bootstrap construida con el método de Neighbor-Joining

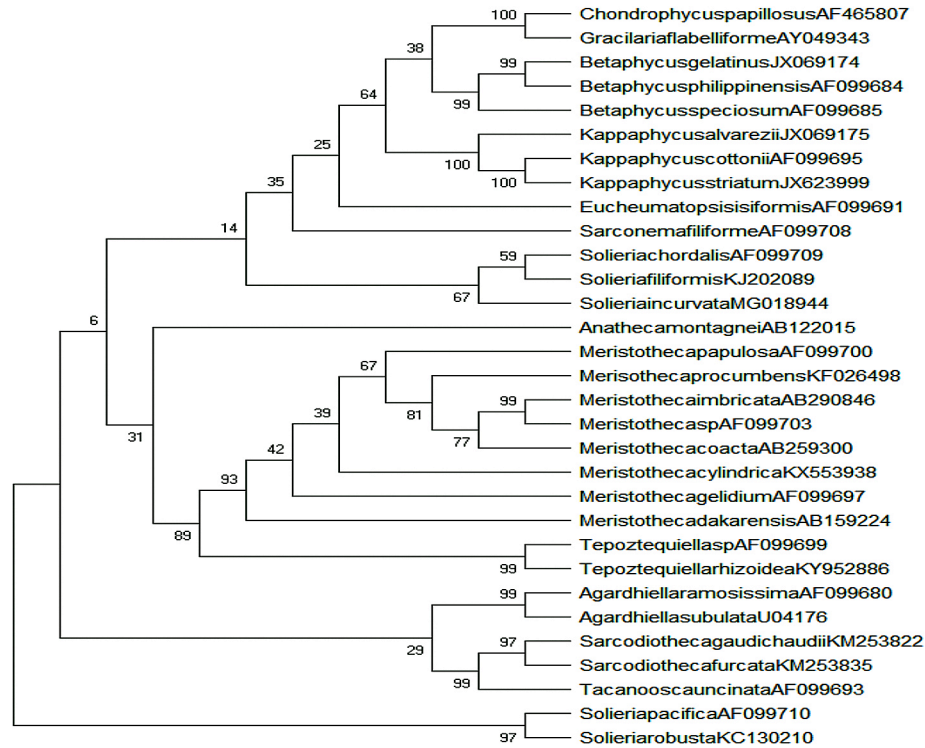


Fig. 54. Topología del árbol óptimo construida con el método de Máxima Parsimonia

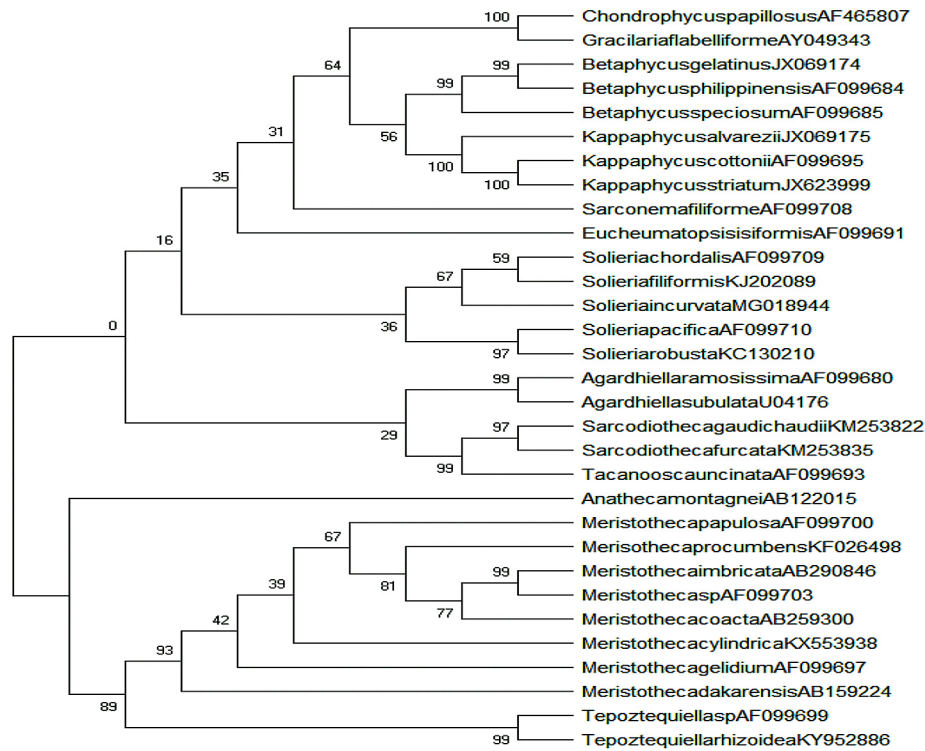












Fig. 55. Topología del árbol consenso de Bootstrap construida con el método de Máxima Parsimonia

ACTIVIDADES COMPLEMENTARIAS

1. Describe el árbol óptimo obtenido con los tres métodos estudiados en la práctica (en total tres árboles), considerando agrupamientos de especies y valores de Bootstrap.
2. Discute las diferencias que observas entre los tres diferentes métodos, ¿alguno fue mejor?
3. Discute los valores de Bootstrap obtenidos en los tres árboles, ¿qué análisis fue el más robusto?

BIBLIOGRAFÍA RECOMENDADA

-  Adams E.N. 1972. "Consensus techniques and the comparison of taxonomic trees". *Systematics Zoology* 37: 27–29.
-  Arnedo M.A. 1999. "Cladismo: la reconstrucción filogenética, basada en Parsimonia". *Boletín de la Sociedad Entomológica aragonesa* 26: 57–84.
-  Bandelt H.J. & A.W.M. Dress. 1992. "Split decomposition: A new and useful approach to phylogenetic analysis of distance data". *Molecular phylogenetics and evolution* 1: 242–252.
-  Bremer K. 1994. "Branch support and tree stability". *Cladistics* 10: 295–304.
-  Cummings M.P., Handley S.A., Myers D.S., Reed D.L., Rokas A. & Winka K. 2003. "Comparing bootstrap and posterior probability values in the four taxon case". *Systematic Biology* 52: 477–487.
-  Doyle J.J. 1992. "Gene trees and species trees: Molecular systematics as one character taxonomy". *Systematic Botany* 17: 144–163.
-  Hendy M.D. & Penny D. 1982. "Branch and bound algorithms to determine minimal evolutionary trees". *Mathematical Biosciences* 59: 277–290.
-  Nixon K.C. & Carpenter J.M. 1993. "On outgroups". *Cladistics* 9: 413–426.
-  Swofford D.L. & Olsen G.J. 1990. "Phylogeny reconstruction". In: *Molecular Systematic* (Ed. by D.M. Hillis & C. Moritz), pp. 411–501. Sinauer Associates, Massachusetts, USA.
-  Williams D.M. 1992. "DNA analysis: Theory". In: *Cladistics: A practical course in Systematics* (Ed. by P.L. Forey, C.J. Humphries, I.J. Kitching, R.W. Scotland, D.J. Siebert & D.M. Williams), pp. 89–101. Oxford University Press, Oxford, UK.

Práctica 5

Construcción de árboles filogenéticos II: Máxima Verosimilitud con RaxML

INTRODUCCIÓN

El análisis de **Máxima Verosimilitud** (ML por sus siglas en inglés, Maximum Likelihood) es un método exhaustivo que busca todas las posibles topologías y considera cada posición en un alineamiento, para encontrar la hipótesis filogenética que mejor explique la tasa de cambios nucleotídicos (transiciones/transversiones) en un alineamiento dado. Para ello, emplea **modelos probabilísticos** de selección del **mejor árbol** (aquel que tenga la más alta probabilidad (verosimilitud) de reflejar el proceso evolutivo real).

Debido a que la inferencia filogenética con el método de Máxima verosimilitud asume un modelo evolutivo de cambio (distancia entre secuencias), la selección del modelo evolutivo tendrá un gran impacto sobre la filogenia resultante. De este modo, si se elige un modelo simple, este puede subestimar la **longitud de las ramas**. El modelo propuesto por Stakatakis, el autor del *software*, para un análisis de ML es GTR.

En todo análisis de ML, para cada árbol se calcula la probabilidad de que refleje cada una de las posiciones en el alineamiento, el cálculo de esta probabilidad se realiza para cada sitio. Finalmente, se suman las probabilidades de todas las posibles reconstrucciones dando la verosimilitud (que es proporcional a la probabilidad de los datos dado un modelo de evolución) de cada sitio particular. El árbol resultante será el producto de las verosimilitudes de cada posición, que será la explicación más probable y parsimoniosa.

Las principales ventajas de este método son que intenta estimar la tasa real de cambio especificando un modelo de evolución, considera como posibles todas las hipótesis filogenéticas para la posición de cada aminoácido, toma en cuenta la longitud de las ramas y ofrece hipótesis robustas si las longitudes de rama son cortas (amplio muestreo). Por otro lado, dado que las tasas de cambio son mucho mayores en las ramas terminales que en el resto del árbol, este método es altamente afectado por la atracción de ramas largas, pudiendo generar filogenias erróneas. Adicionalmente, descarta caracteres que pueden ser relevantes tales como las **autoapomorfias**, subestima cambios múltiples y con ello puede subestimar también el número de sustituciones, además su ejecución es considerablemente más lenta que otros tipos de análisis.

Los análisis de ML pueden ser llevados a cabo con diferentes *softwares* libres tales como TNT, RaxML, Garlic, MEGA, o en portales en línea como Cypress o TOPALI.

OBJETIVO

- Construir un árbol filogenético con el criterio de Máxima Verosimilitud, utilizando como herramienta el portal en línea de RaxML.

SOFTWARE REQUERIDO

1. RAxML v0.6.0 Randomized Axelerated Maximum Likelihood, disponible para su descarga en el sitio del programa o para trabajar en línea, siguiendo la liga <https://raxml-ng.vital-it.ch/#/>

DESARROLLO

1. En el navegador de tu preferencia, copia la siguiente liga <https://raxml-ng.vital-it.ch/#/> para poder acceder al sitio en línea de RaxML. El programa estará dividido en cuatro secciones: Data, Evolutionary Model, Analysis y Notification.

- En la sección **Data**, deberás cargar la matriz de datos en el primer recuadro en blanco dando clic sobre el botón **seleccionar archivo**, te abrirá una ventana de diálogo que te permitirá buscar entre tus archivos y seleccionar tu conjunto de datos. En el directorio **Prácticas de Filogenia** subdirectorio **Práctica 2**, selecciona la matriz de datos de secuencias **rbcl_Solieriaceae.fas** que generaste en la práctica 2. Si el archivo fue cargado correctamente, junto al botón **seleccionar archivo**, aparecerá el nombre de tu archivo y una palomita en color verde (Fig. 56). Si cargaste por error el archivo equivocado, sólo da clic en el botón **Remove file**, y vuelve a cargar el archivo correcto.

Nota: los formatos reconocidos por el programa para la matriz de datos, son sólo FASTA O PHYLIP, por lo que debes asegurarte de que tu archivo tenga alguna de estas dos extensiones, de lo contrario no podrá ser leído por el programa. Recuerda que para transformar tu conjunto de datos a cualquiera de estos programas puedes abrirlo en el programa Bioedit y salvarlo en el formato correspondiente.

RAxML BlackBox

Data

Paste your **sequence alignment**

or upload a file

rbcl_Soli...ceae.fas

✔

Must be in [PHYLIP format](#), or FASTA format, or convertible by readseq

Paste your **constraint tree**

or upload a file

No se eligió archivo

This option allows you to specify an incomplete or comprehensive multifurcating constraint tree in NEWICK format. [More help here.](#)

Fig. 56. Captura de pantalla del paso 2: matriz de datos

- En la sección **Evolutionary model**, deberás seleccionar los parámetros con los cuales deseas explicar la sustitución nucleotídica de tu conjunto de datos (Fig. 57). Primero, selecciona si el modelo de evolución será particionado o no, en este último caso, da clic en la opción **unpartitioned model**. Enseguida, selecciona el tipo de datos que estás analizando, ya que estamos trabajando con una matriz construida a partir de secuencias de ADN, selecciona la opción **DNA**. En la opción **Substitution matrix**, debes seleccionar el modelo de evolución que mejor se ajustó a tu conjunto de datos, previamente estimado con el programa MEGA, buscándolo entre las opciones de modelos que despliega la flecha doble del recuadro (Fig. 57a). En este caso el modelo que usaremos será **GTR** aunque no haya sido el modelo que explicara mejor nuestros datos. En la opción **Stationary base frequencies** selecciona la opción **ML estimate**; a continuación, muestra una casilla en blanco para considerar o no la **proportion of invariable sites**; dado que en todo conjunto de datos puede haber una proporción de sitios que cambian libremente y una proporción de sitios invariables, selecciona la casilla para que ningún arreglo quede fuera del análisis (Fig. 57b). En la opción **Among-site rate heterogeneity**, selecciona una distribución **GAMMA** para especificar la tasa de variación entre sitios (Fig. 57c). Para el resto de los parámetros de esta sección, conserva los que te sugiere el portal.

Evolutionary model

Substitution matrix:

Stationary base frequencies:

Proportion of invariant sites:

Among-site rate heterogeneity:

Number of rate categories:

GAMMA category rates:

Ascertainment bias correction:

Model string is composed as described [here](#). Individual model modifiers are joined with "+" (plus sign).

Fig. 57. Captura de pantalla del paso 3: (a) selección de modelo, (b) proporción de sitios invariables y (c) distribución de la variación entre sitios

4. En la sección **Analysis** deberás especificar los criterios para el análisis. En la subopción **ML tree search**, el portal ofrece una serie de criterios, en la opción **optimize**, para optimizar la topología resultante, la longitud de las ramas y el modelo, marca con una paloma las tres opciones (Fig. 58a). en la opción **Starting trees**, mantén activadas las opciones **Parsimony** y **Random** (Fig. 58b). En los recuadros que se encuentran a un lado de cada una de estas opciones, puedes seleccionar el **número de árboles** subiendo o bajando con la flecha doble, el portal sugiere 10 árboles, puedes incrementarlo a 100. Finalmente, en la subopción **Bootstrapping**, selecciona la casilla en blanco (Fig. 58c). Al seleccionar la opción, aparecerán dos criterios adicionales, **Number of replicates**, que te permitirá fijarlo tú mismo o permitirle al programa seleccionarlo y **Bootstopping cutoff**, donde podrás seleccionar el número de repeticiones con un máximo de 100, o permitir que el programa lo determine (Fig. 58c).

Analysis

ML tree search

Optimize: Topology Branch lengths Model ← **a**

Starting trees

Parsimony: Number of trees: ← **b**

Random: Number of trees:

Paste your tree in newick format:

or upload a file: No se eligió archivo

Bootstrapping ← **c**

Number of replicates:

Number of replicates:

Fig. 58. Captura de pantalla del paso 4: (a) parámetros del modelo de evolución, (b) criterios de optimización de árboles y (c) Bootstrap

- Finalmente, en la sección **Notification** deberás proporcionar una dirección de correo electrónico para que, una vez terminado el análisis, recibas una notificación de que está listo y el enlace donde puedes encontrarlo (Fig. 59a). Una vez ingresado tu correo, da clic en la opción **Submit**. Aparecerá una leyenda que te informa que tu análisis está corriendo y un enlace en que estará disponible una vez esté finalizado (Fig. 59b).

Notification

Email

a ↗

↘

Submit

Enter your email if you want to be notified upon job completion

Your job is running and will be available at this url:

<https://raxml-ng.vital-it.ch/#/result/1403/code/3teTcsHwZrfn>

b ↗

Fig. 59. Captura de pantalla paso 5: (a) Notificación para recuperar el análisis y (b) enlace de recuperación

RESULTADOS

Una vez que el análisis terminó, el portal te enviará un correo de notificación para informarte que está listo. Al dar clic en el enlace adjunto al correo, te desplegará la topología resultante y una carpeta descargable llamada **tree** en formato zip. Descarga y descomprime los archivos de la carpeta copiándolos y pegándolos en una nueva dentro de tu directorio Prácticas de Filogenia subdirectorío Práctica 5. Al interior de esta carpeta encontrarás dos archivos con extensión txt que contendrán las probabilidades calculadas durante el análisis, tu alineamiento en formato FASTA y las cuatro topologías generadas con extensión TREE. El primer archivo corresponde al mejor árbol para el conjunto de datos (**result.raxml.bestTree**) (Fig. 60), el segundo con la topología de Bootstrap (**result.raxml.bootstraps**) (Fig. 61), el tercero con la topología de ML, es decir, la hipótesis más probable (**result.raxml.mlTrees**) (Fig. 62), y el último con la topología que contiene los valores de soporte de los nodos (**result.raxml.support**) (Fig. 63). Para establecer las relaciones filogenéticas de tu grupo, normalmente se selecciona el mejor árbol (**bestTree**); sin embargo, éste no siempre refleja mejor las relaciones filogenéticas. Cualquiera que elijas siempre debes especificar los criterios de tu selección en toda comunicación idónea de resultados. Estos archivos serán utilizados en la Práctica 6, para visualizarlos y editarlos con el programa FigTree (Práctica 6).

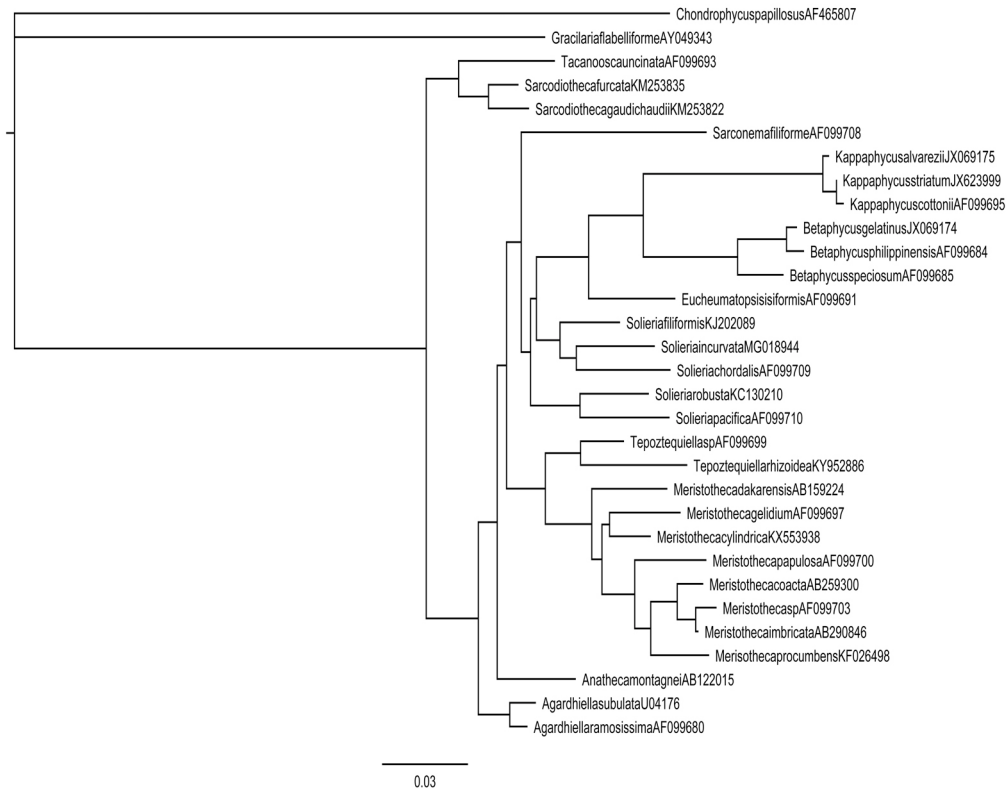


Fig. 60. Topología del mejor árbol construida con el método de Máxima Verosimilitud

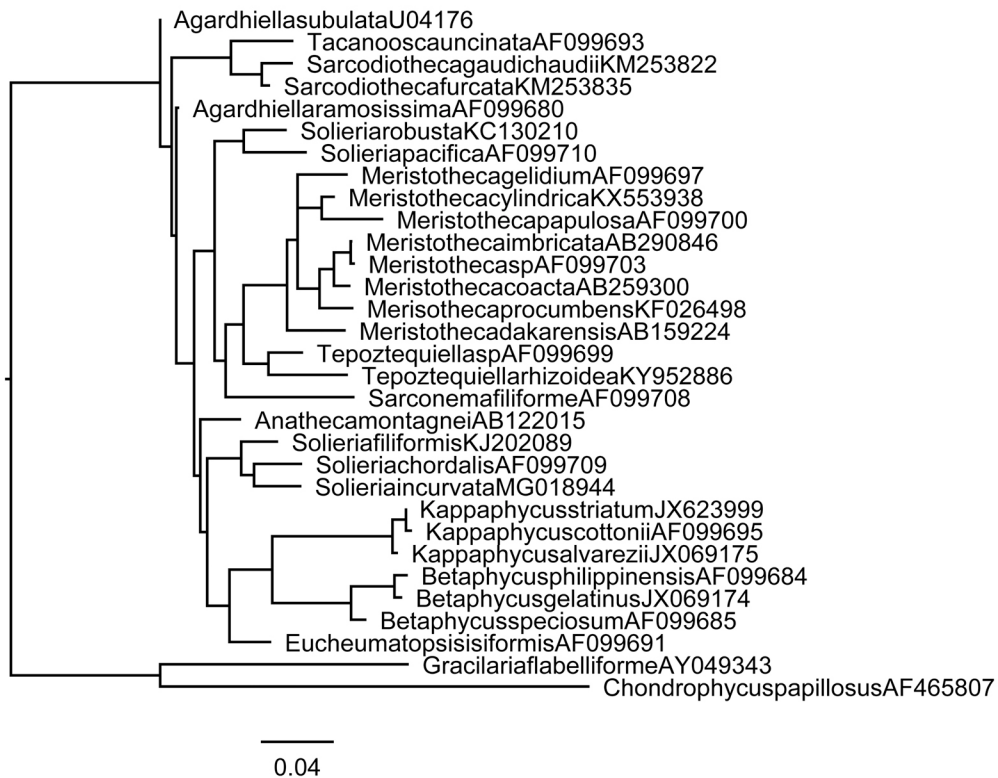


Fig. 61. Topología de Bootstrap construida con el método de Máxima Verosimilitud

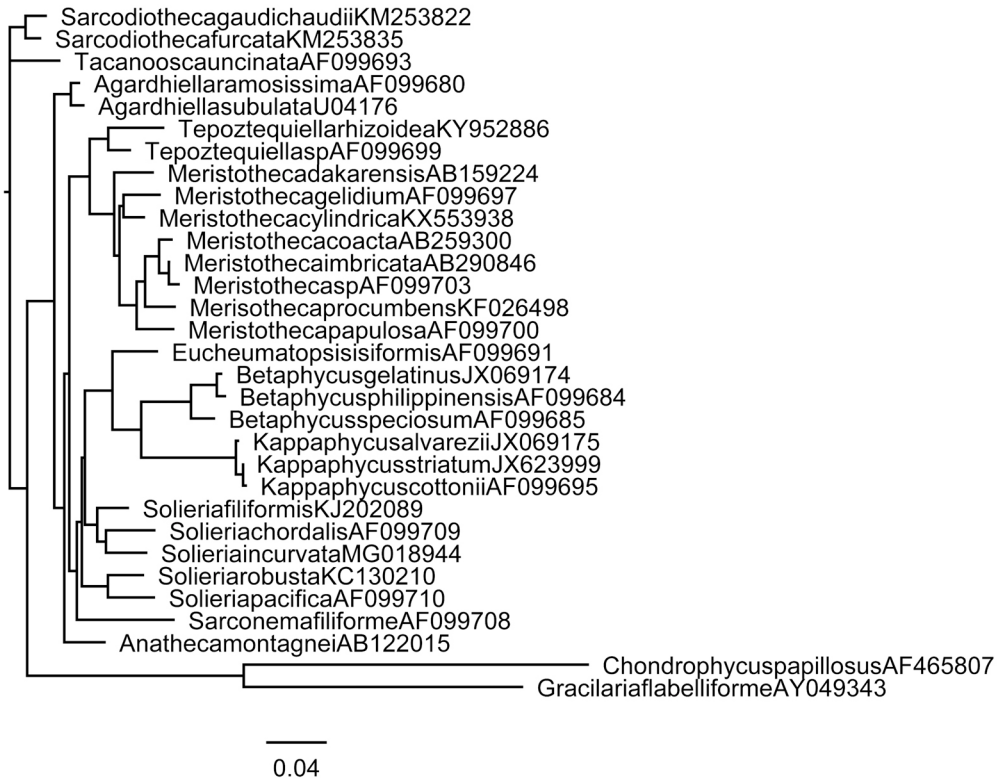


Fig. 62. Topología de Máxima Verosimilitud

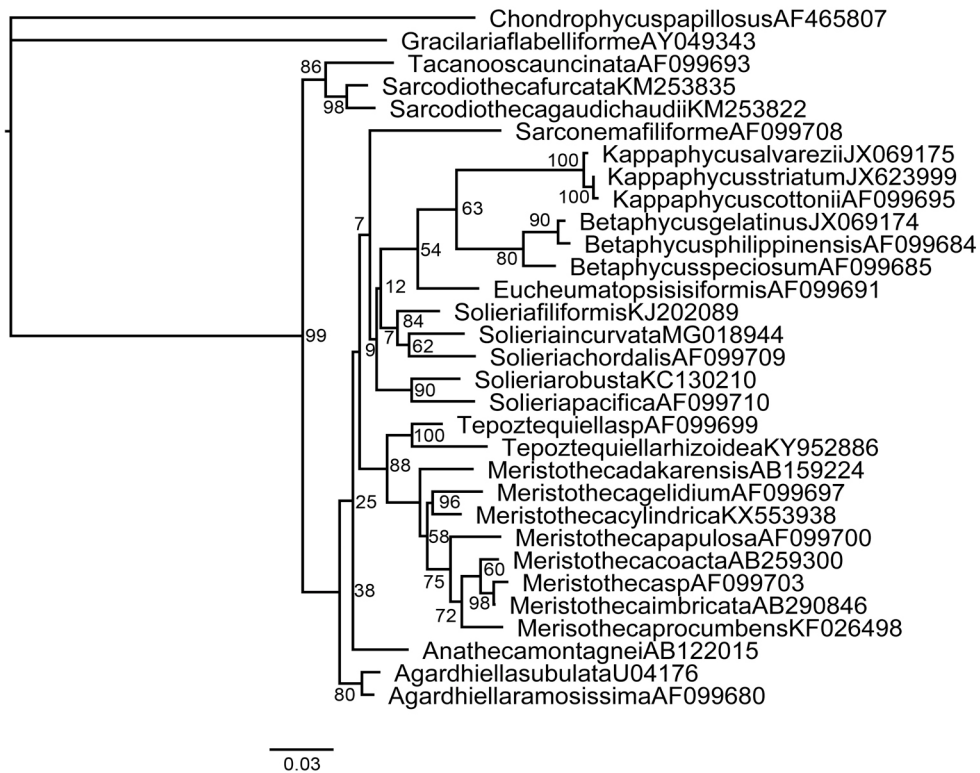








Fig. 63. Topología con los valores de soporte para el mejor árbol de Máxima Verosimilitud

ACTIVIDADES COMPLEMENTARIAS

1. Utilizando la misma matriz de datos, vuelve a ejecutar un análisis de ML intercambiando el modelo GTR por el modelo HKY.
2. Describe las diferencias entre ambas topologías considerando relaciones de parentesco de las especies representadas y soporte de los nodos.

BIBLIOGRAFÍA RECOMENDADA

-  Felsenstein J. 1981. "Evolutionary trees from DNA sequences: A maximum likelihood approach". *Journal of Molecular Evolution* 17: 368–376.
-  Felsenstein J. 1986. "Confidence limits on phylogenies: An approach using bootstrap". *Evolution* 39: 783–791.
-  Jukes T. & Cantor C.R. 1969. "Evolution of protein molecules". In: *Mamalian protein metabolism* (Ed. by H.N. Munro), pp. 21–132. Academic press, New York, USA.
-  Posada D. & Crandall K.A. 1998. "MODELTEST: testing the model of DNA substitution". *Bioinformatics* 14: 817–818.
-  Stamatakis A. 2006. "RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models". *Bioinformatics* 22: 2688–2690.
-  Stamatakis A., Hoover P. & Rougemont J. 2008. "A Rapid Bootstrap Algorithm for the RAxML Web-Servers". *Systematic Biology* 75: 758–771.

Práctica 6

Visualización y edición de árboles filogenéticos con FigTree

INTRODUCCIÓN

FigTree es un programa de libre descarga en la red diseñado como visor gráfico de árboles filogenéticos y como programa para producir figuras de calidad listas para publicación. Está diseñado para mostrar archivos resumidos y anotados, principalmente con extensión **tree**, generados a partir de una variedad de programas. El programa tiene una interfase gráfica que permite a los usuarios modificar varios componentes del árbol. Los archivos de figuras de árbol se pueden exportar como archivos PDF para ser publicados o para su edición posterior en otro programa de gráficos, tal como CorelDraw o Adobe Photoshop.

El uso de FigTree en la edición de árboles filogenéticos permite:

- Visualizar árboles gráficos multiplataforma.
- Tres diferentes estilos de árbol: rectangular, polar y radial.
- Visualizar las alturas de nodo, las longitudes de rama, los valores de soporte y otras anotaciones.
- Barras de rango de altura del nodo si están disponibles.
- Colapso de clados en triángulos.
- Coloración de ramas y etiquetas terminales.
- Colorear mediante anotación (por ejemplo, valores de soporte).
- Búsqueda rápida de etiquetas de consejos o etiquetas de consejos parciales.
- Impresión y exportación como gráficos PDF.

OBJETIVOS

- Visualizar gráficamente los árboles filogenéticos con extensión tree, utilizando como herramienta el programa FigTree.
- Explorar las herramientas de edición del programa FigTree y editar los árboles filogenéticos de ML obtenidos en la práctica 5.

SOFTWARE REQUERIDO

1. FigTree v1.4.3 Andrew Rambaut. Institute of Evolutionary Biology, University of Edinburgh (2006-2016), disponible para su descarga en el sitio del programa, siguiendo la liga <http://tree.bio.ed.ac.uk/software/figtree/>

DESARROLLO

1. Abre el programa FigTree.
2. Verás que el programa es un editor de imagen típico con diferentes opciones de menú (Fig. 64). Un menú principal en la parte superior izquierda de la pantalla (Fig. 64a), un submenú pictográfico debajo del menú principal (Fig. 64b) y una barra de herramientas de **Layout** con opciones diversas de edición en la columna izquierda (Fig. 64c).

Nota: FigTree sólo puede leer archivos con extensión **tree**. Archivos de árbol con formato **jpg** o **tif**, pueden ser abiertos y editados en otros visores gráficos que no reconocen el formato **tree**. También FigTree puede transformar archivos **tree** en archivos que puedan ser leídos en otros editores de imagen.

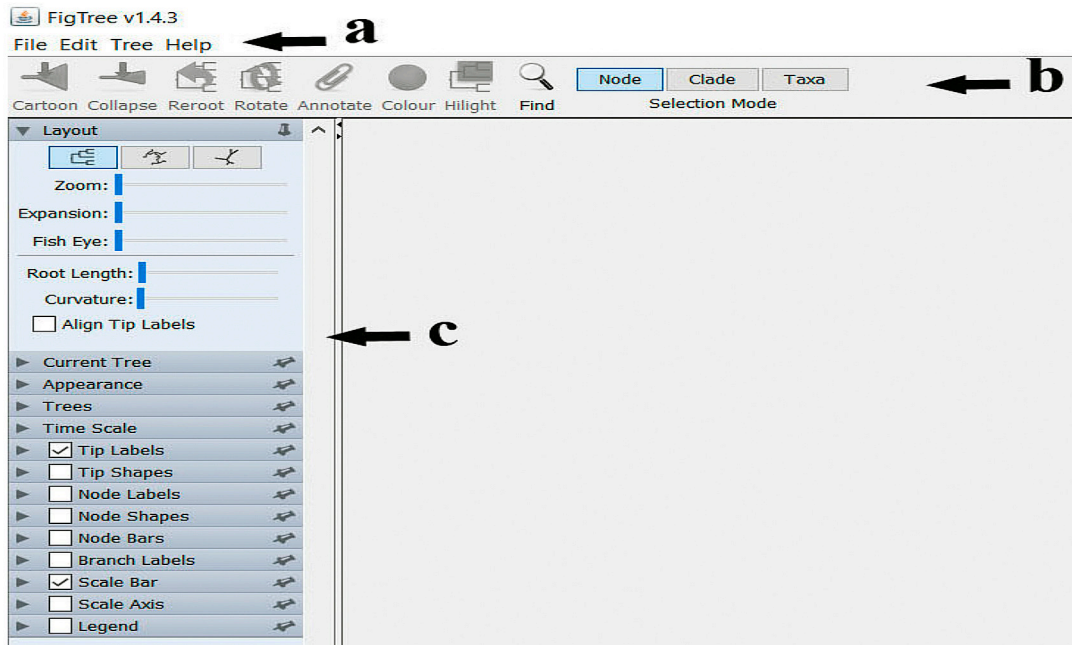


Fig. 64. Captura de pantalla paso 2: opciones de menú del programa FigTree (a) menú principal, (b) submenú pictográfico y (c) barra de herramientas de Layout

3. En el menú principal selecciona la opción **File**, del submenú que desplegará selecciona la opción **open** (Fig. 65). Te abrirá una ventana de diálogo que te permitirá buscar entre tus archivos y seleccionar uno. Vamos a editar el árbol de consenso que obtuvimos con Máxima Verosimilitud.

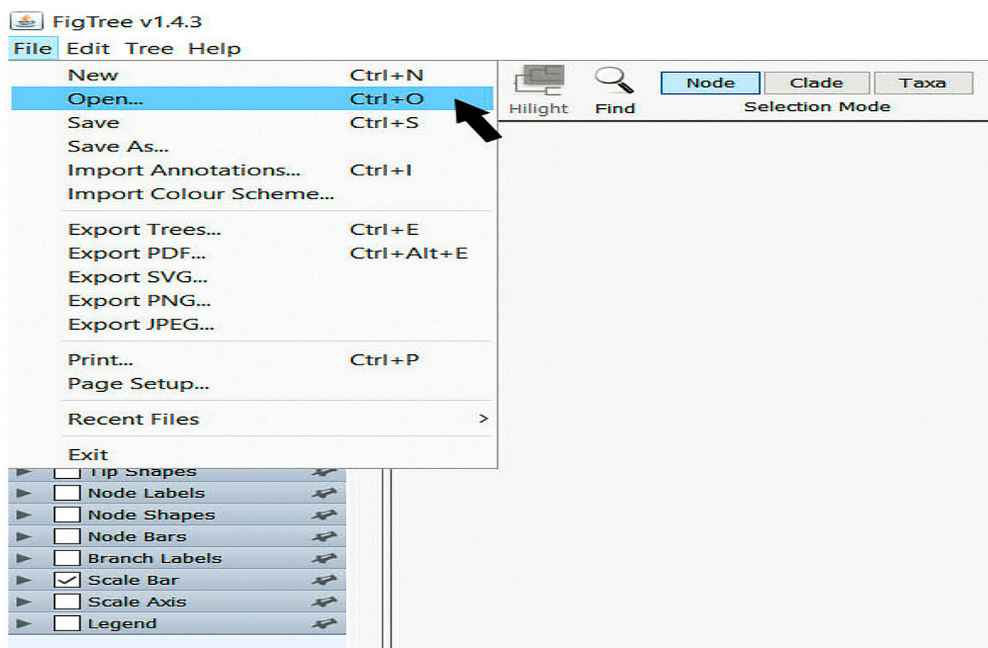


Fig. 65. Captura de pantalla paso 3: abrir un archivo tree en FigTree

4. En el directorio Prácticas de Filogenia, subcarpeta Práctica 5, selecciona el archivo **result.raxml.bestTree**. Aparecerá en la caja del lado derecho del programa la representación gráfica del mejor árbol construido con Máxima Verosimilitud (Fig. 66).

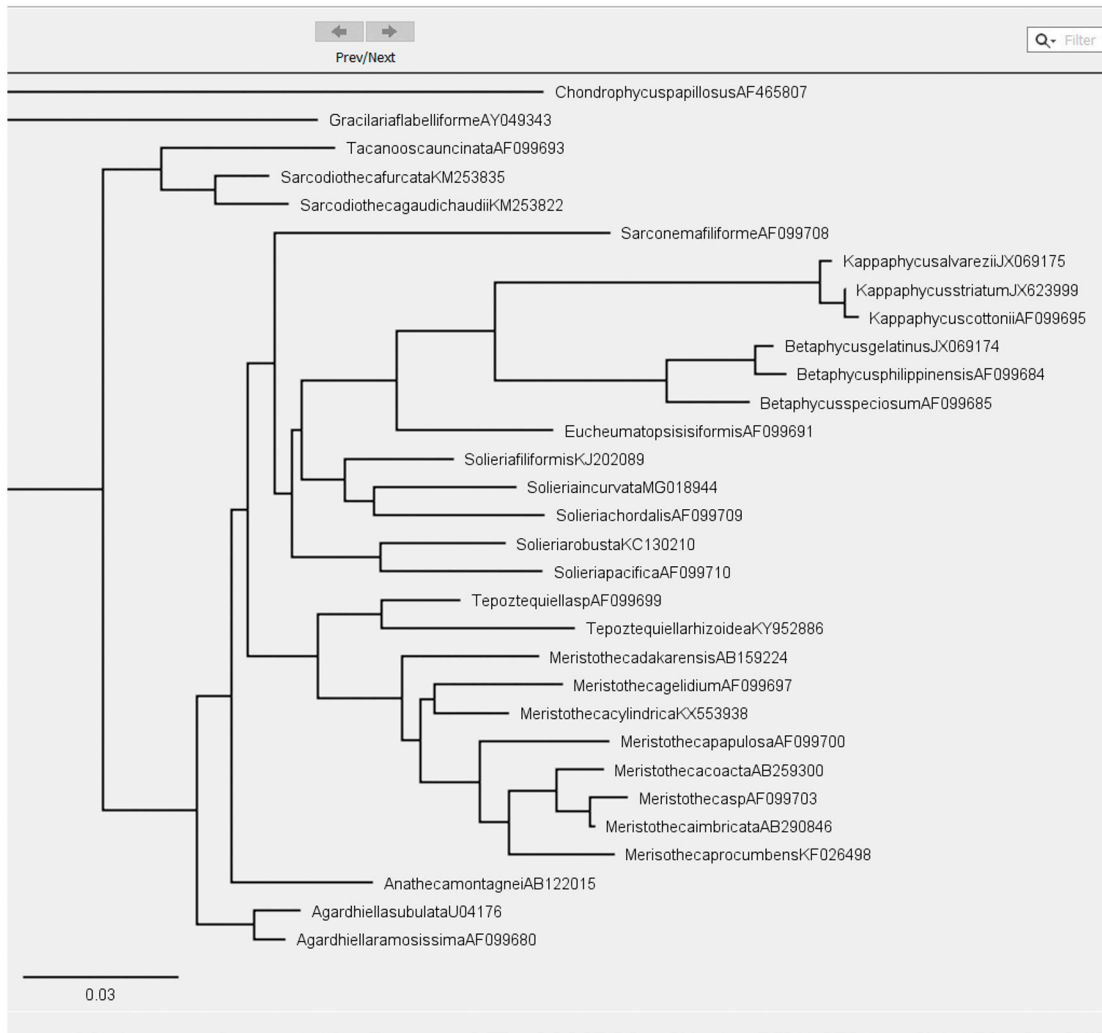


Fig. 66. Captura de pantalla paso 4: representación gráfica del mejor árbol de Máxima Verosimilitud abierto con FigTree

5. Notarás que la forma del árbol será rectangular, al igual que un típico cladograma. Sin embargo, puedes editar la forma del árbol con los primeros íconos de la barra de **Layout** (Fig. 67a). estas formas pueden ser **Polar** (Fig. 67b) o **Radial** (Fig. 67c).

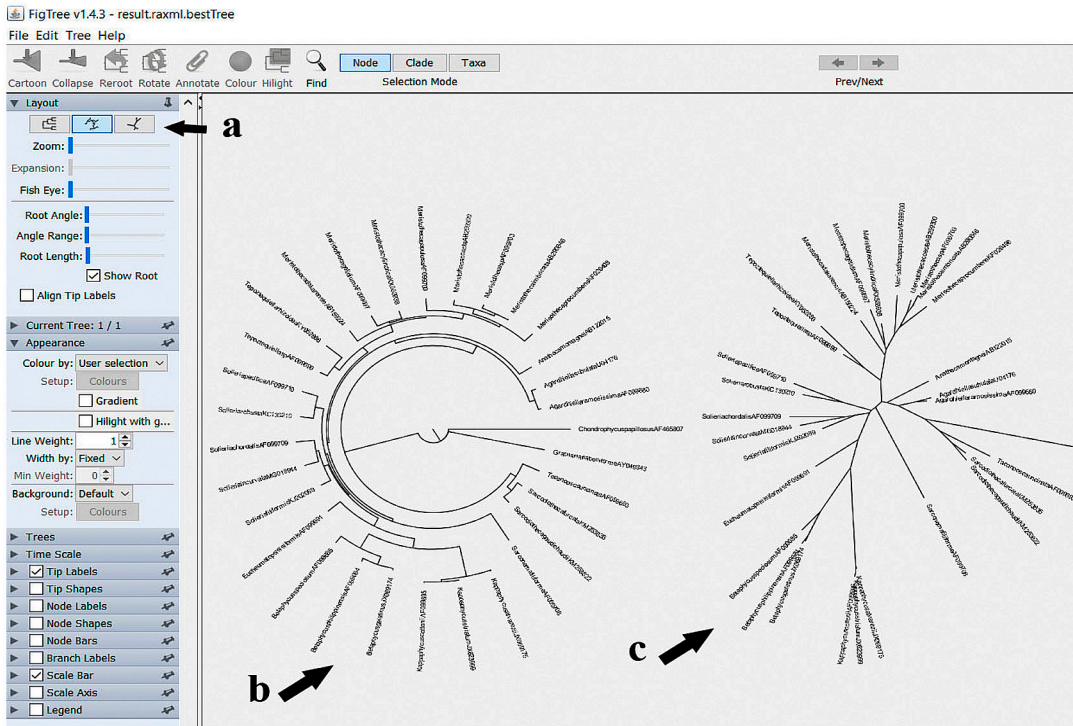


Fig. 67. Captura de pantalla paso 5: edición de la forma del árbol con Layout (a) íconos de edición, (b) forma polar y (c) forma radial

6. Selecciona la forma rectangular. Notarás que los nombres de los taxa son muy pequeños y no pueden ser leídos fácilmente. Puedes aumentar el tamaño de la fuente seleccionando con un clic el menú **Tip Labels** (Fig. 68a), a continuación, se desplegará otro submenú, en la opción **Font size** notarás que dice **8**, con la flecha ascendente incrementa el tamaño de la fuente a **14** (Fig. 68b). Notarás que ahora los nombres de los taxa son legibles (Fig. 68c).



Fig. 68. Captura de pantalla paso 6: (a) herramientas de edición de etiquetas terminales, (b) tamaño de fuente y, (c) etiquetas en la topología

7. En el submenú superior notarás las opciones **Node**, **Clade** y **Taxa** (Fig. 69, flecha). Las opciones de edición del submenú pictográfico estarán modificando el sitio en donde te encuentres. Es decir, si deseas cambiar el color de la tipografía para los nombres de taxa, deberás seleccionar primero la opción **Taxa** y luego hacer la edición (Fig. 70, flecha). Si realizas la edición sobre la opción **Node**, y luego aplicas el color azul, será el nodo el que cambiará al color azul (Fig. 71, flecha).

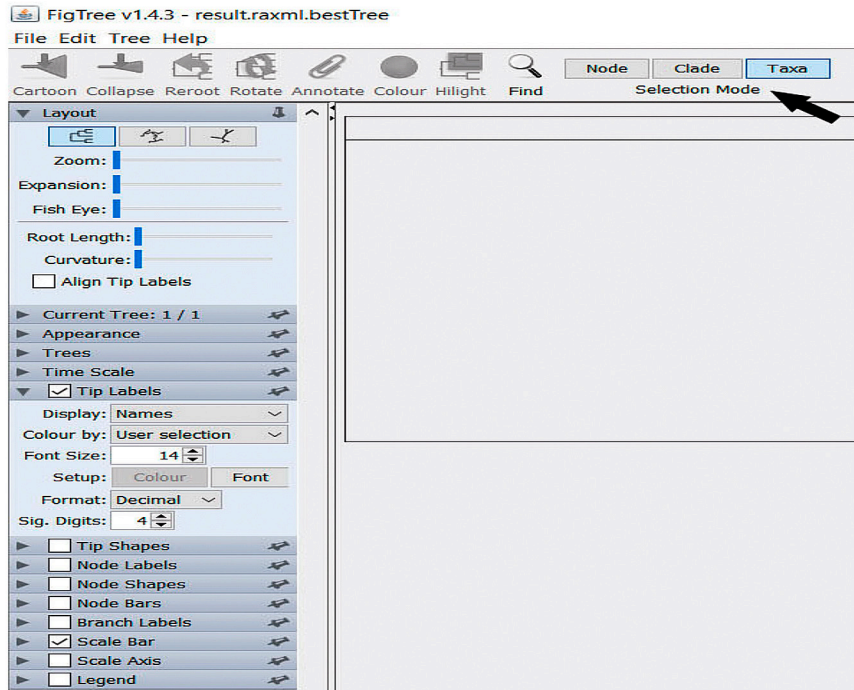


Fig. 69. Captura de pantalla paso 7: opciones de menú para la edición en sitios particulares de la topología como nodos, clados o taxa

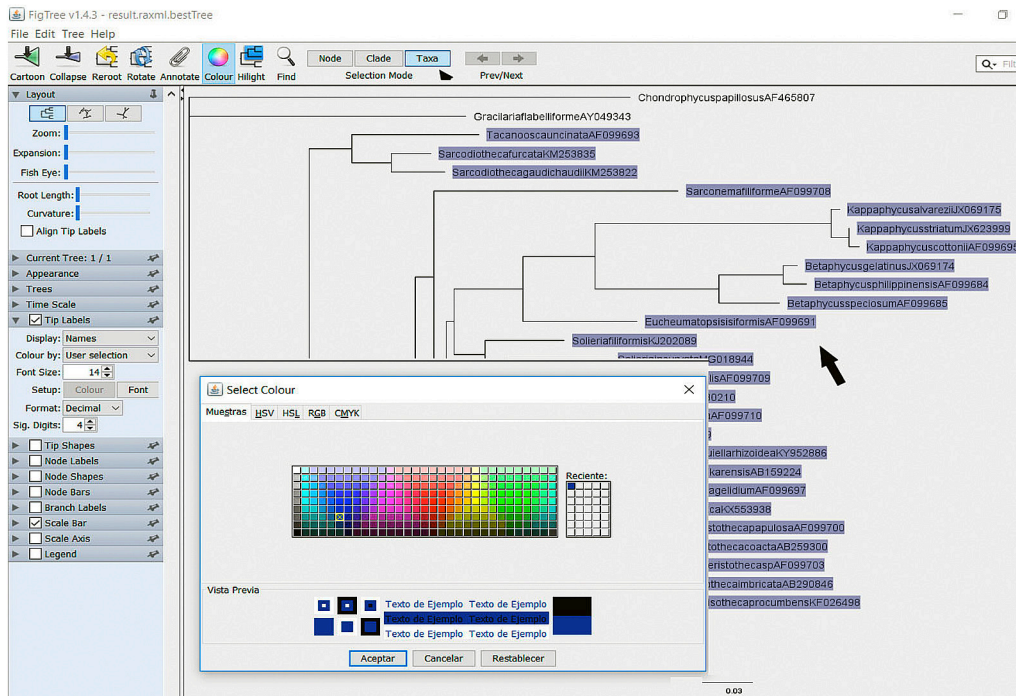


Fig. 70. Captura de pantalla paso 7: ejemplo de edición sobre los taxa

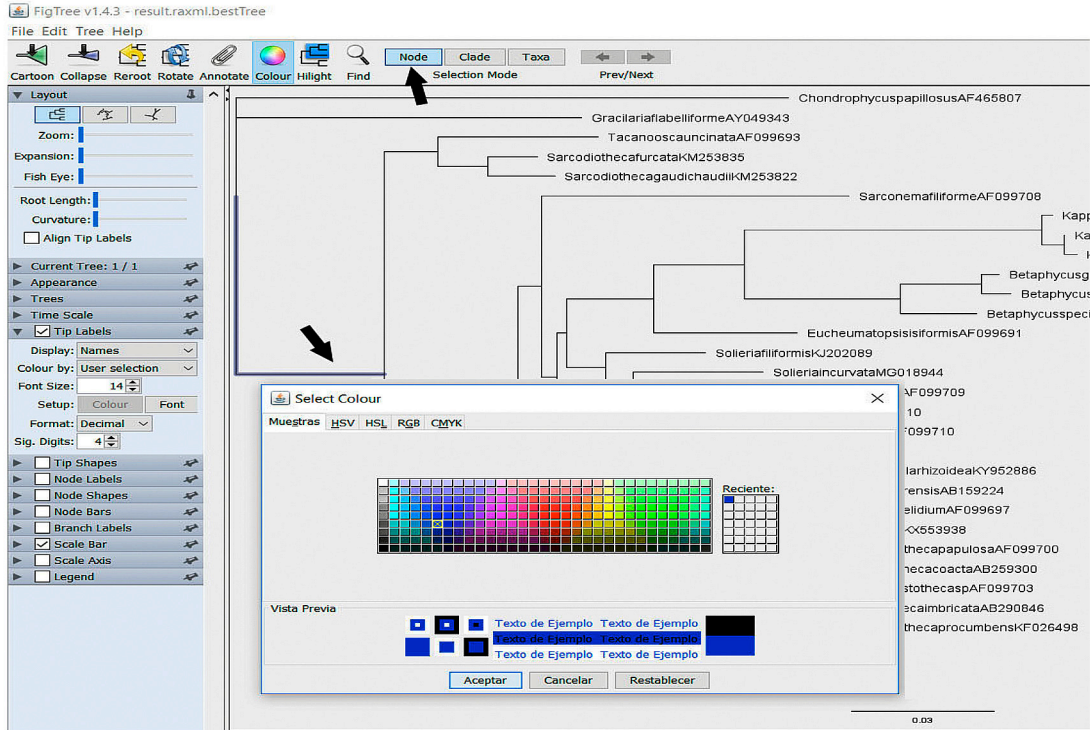


Fig. 71. Captura de pantalla paso 7: ejemplo de edición sobre los nodos

8. Continúa explorando las múltiples opciones de edición. Cuando hayas editado el árbol a tu gusto y necesidades, sávalo en un formato de imagen compatible con los lectores gráficos tradicionales. Para ello, ve a la barra de menú principal y selecciona el menú **File**, del submenú que te mostrará, selecciona la opción **Export JPEG** (Fig. 72).

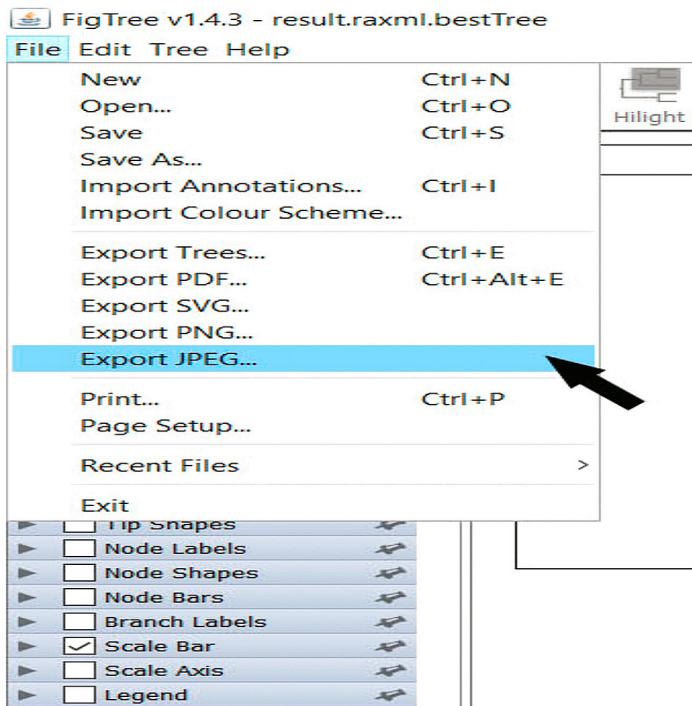
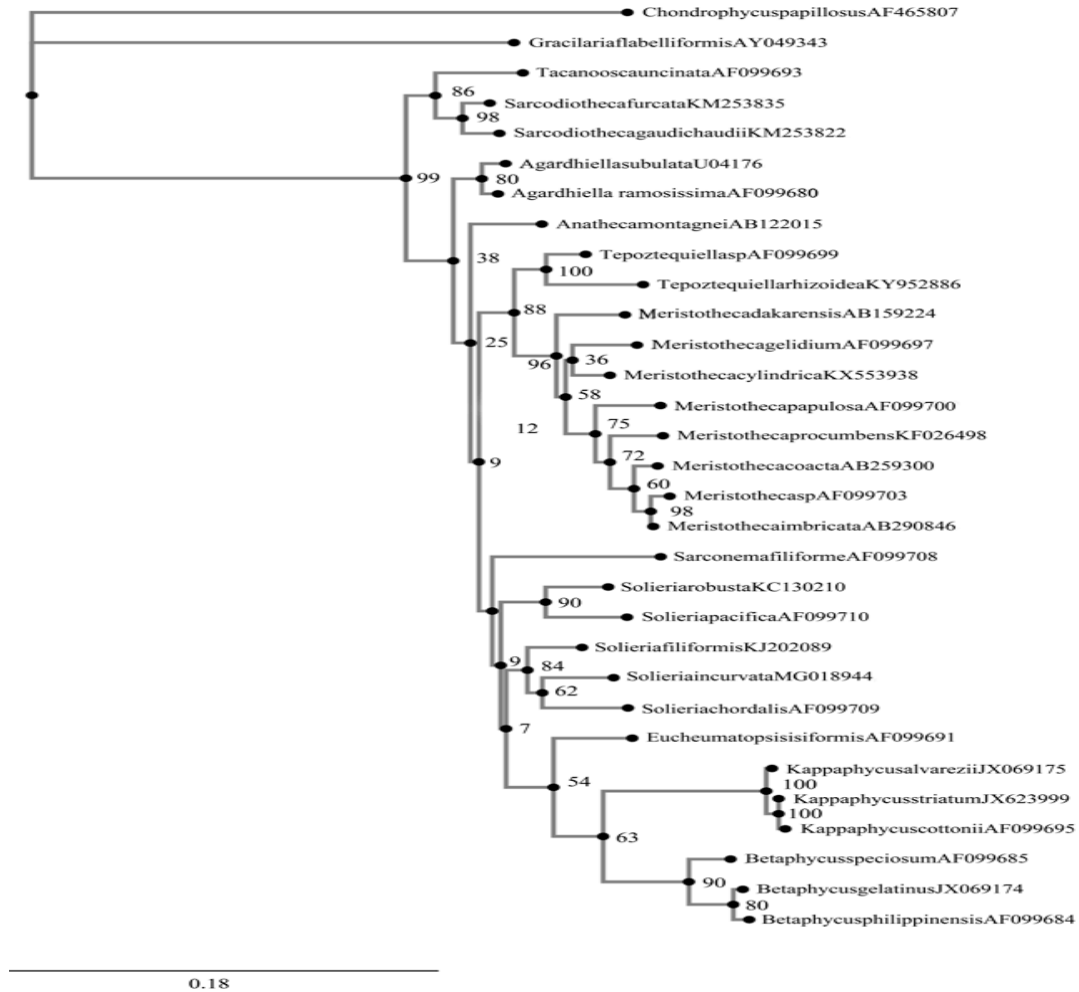


Fig. 72. Captura de pantalla paso 8: exportar un árbol editado en FigTree como archivo JPG

9. Te abrirá una ventana de diálogo para que selecciones el directorio en que deseas guardar tu árbol. Selecciona el directorio Prácticas de Filogenia, subcarpeta **Práctica 6**, y nómbralo **Solieriaceae-ML_editado**.

RESULTADOS


Al final de la práctica, obtendrás un archivo de imagen **Solieriaceae-ML_editado.jpg** (Fig. 73). No olvides incluir en tu edición del mejor árbol los valores de Bootstrap, estos podrás verlos en el archivo **result.raxml.support** y de ahí copiarlos en los nodos del mejor árbol. Recuerda que Bootstrap no es una hipótesis filogenética, no uses ese árbol para la edición. Este archivo será utilizado en la práctica de inferencia filogenética.



ACTIVIDADES COMPLEMENTARIAS

1. Edita en FigTree el archivo **result.raxml.mlTrees** y obtén un árbol en formato JPG con los valores de Bootstrap a partir de este archivo

BIBLIOGRAFÍA RECOMENDADA

-  Rambaut A. 2016. FigTree v1.4.3. Institute of Evolutionary Biology, University of Edinburgh. <http://tree.bio.ed.ac.uk/software/figtree/>

Práctica 7

Construcción de árboles filogenéticos III: Probabilidad Posterior Bayesiana con MrBayes

INTRODUCCIÓN

La inferencia bayesiana es un tipo de inferencia estadística basada en el cálculo de la **probabilidad posterior**, es decir, la probabilidad de que una hipótesis pueda ser cierta es calculada después de formulada la hipótesis. La mejor estimación filogenética es aquella que maximiza la probabilidad posterior de que un árbol sea correcto. El cálculo de la probabilidad posterior se efectúa mediante el teorema de Bayes, el cual combina la probabilidad previa de una filogenia con su verosimilitud. El cálculo de la probabilidad posterior implica evaluar todas las posibles combinaciones de longitud de rama y parámetros de un modelo.

En tanto que el teorema de Bayes no tiene solución matemática, la probabilidad posterior de los árboles filogenéticos (y otros parámetros del modelo de sustitución) no puede determinarse analíticamente. En cambio, **MCMC** (Markov chain Monte Carlo) se usa para aproximar las probabilidades posteriores de los árboles dibujando muestras (dependientes) de la distribución posterior. Este programa también puede implementar una variante de MCMC llamada "cadena de Markov acoplada a la metrópoli Monte Carlo", o MCMCMC para abreviar. El método de MCMC comienza seleccionando, normalmente al azar, parámetros tales como la topología, la longitud de las ramas, modelo evolutivo, etc. En segundo lugar, se hace un pequeño movimiento aleatorio (modificación de los parámetros) y se evalúa la función de probabilidad. Si los valores mejoran, se acepta el movimiento y se repite el proceso; si los valores no mejoran, el nuevo árbol se acepta con una probabilidad proporcional al empeoramiento producido. El fundamento es que, si se repite muchas veces este proceso se obtendrá una estimación de la probabilidad asociada a cada árbol.

El análisis de Inferencia Bayesiana puede realizarse con el programa MrBayes que es un *software* de libre descarga. Sin embargo, actualmente existen otros programas libres con interfases más amigables como el programa Beast, también pueden realizarse en portales en línea como Cypress o TOPALI.

OBJETIVO

- Construir un árbol filogenético con el criterio de Probabilidad Posterior Bayesiana, utilizando como herramienta el programa Mr. Bayes.

SOFTWARES REQUERIDOS

1. MrBayes 3.2.6 Bayesian Inference of Phylogeny, disponible para su descarga en el sitio del programa, siguiendo la liga <http://mrbayes.sourceforge.net/download.php>
2. DnaSP version 5.10.1 DNA Sequence Polymorphism, disponible para su descarga en el sitio del programa, siguiendo la liga http://www.ub.edu/dnasp/index_v5.html

DESARROLLO

Nota: para ejecutar un análisis en MrBayes es necesario transformar primero la matriz **fasta** a formato **nexus**. MrBayes es muy sensible a variaciones en la longitud de secuencias o caracteres extraños, por lo que se recomienda que no haya espacios entre los nombres de taxa ni otros caracteres. Adicionalmente, es indispensable que en las secuencias no haya etras diferente a A, T, G, C, o N para los indels o gaps. Para transformar la matriz a formato nexus, utilizaremos el programa **DnaSP**, este programa igualmente es sensible a todas las variaciones antes mencionadas y si hay algún carácter extraño en la matriz fasta, DnaSP lo detecta y te indica en donde está el problema, MrBayes no. Además, el formato de ambos programas es compatible.

1. Abre el programa DnaSP. Te mostrará dos barras de menú, el menú principal y debajo de este un submenú pictográfico (Fig. 74, flechas).



Fig. 74. Captura de pantalla paso 1: opciones de menú de DnaSP

2. En el menú principal selecciona **File**, de las opciones que desplegará selecciona **Open Data File** (Fig. 75). Abrirá una ventana de diálogo que te permitirá buscar entre tus archivos y seleccionar tu conjunto de datos. Del directorio **Prácticas de Filogenia** subdirectorio **Práctica 2**, selecciona la matriz de datos de secuencias **rbcl_Solieriaceae.fas** que generaste en la práctica 2.

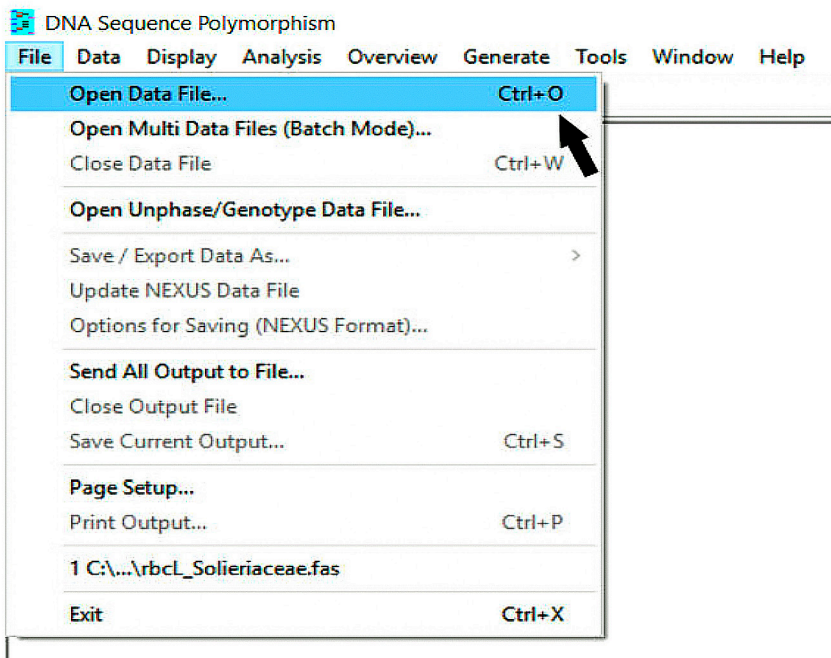


Fig. 75. Captura de pantalla paso 2: importación de una matriz de datos a DnaSP

3. Te mostrará una ventana de diálogo con el resumen de tu conjunto de datos (Fig. 76). Presiona el botón **Close**.

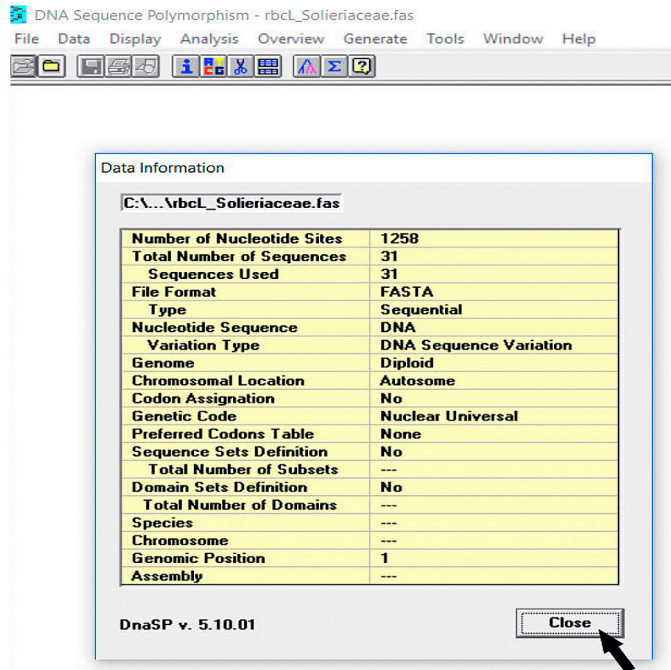


Fig. 76. Captura de pantalla paso 3: resumen de un conjunto de datos en DnaSP

4. Vuelve a seleccionar el menú **File**, ahora selecciona la opción **Save/Export Data As** y selecciona **NEXUS File Format** (Fig. 77). Abrirá una ventana de diálogo que te permitirá seleccionar el directorio en donde deseas guardar la matriz en formato nexus. Selecciona el directorio Prácticas de Filogenia subdirectorío Práctica 7, carpeta **MrBayes** para guardar la matriz, nómbrala **Solieriaceae**.

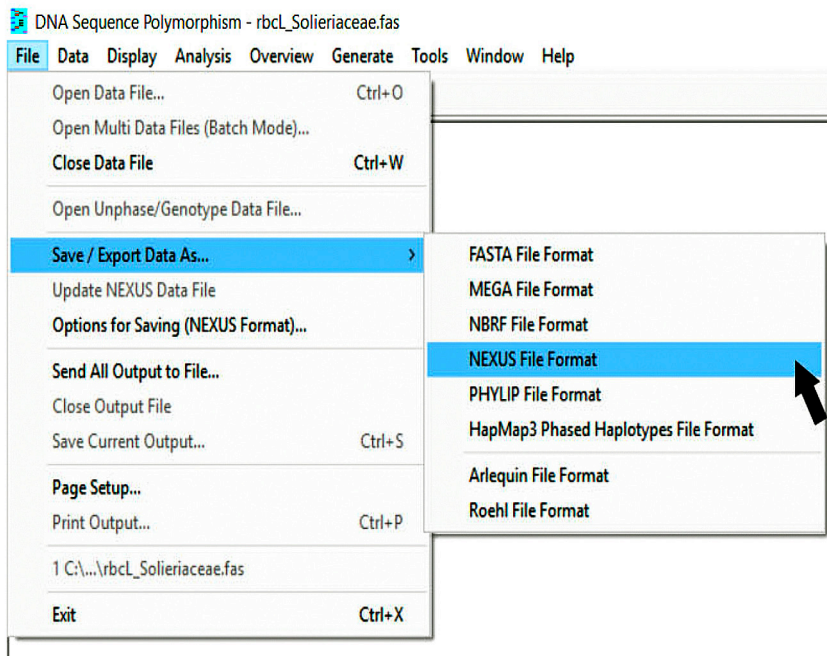



Fig. 77. Captura de pantalla paso 4: exportación de una matriz de datos a formato nexus

5. Cierra DnaSP.
6. Ve a la carpeta Práctica 7. En ella encontrarás una subcarpeta llamada **MrBayes**, en esta carpeta estarán los archivos necesarios para ejecutar el programa.

Nota: el archivo de datos que deseas ejecutar en MrBayes, SIEMPRE debe estar guardado en la misma carpeta del programa, en la raíz. Si no se encuentran en la misma carpeta, MrBayes no reconocerá el archivo y no podrá abrirlo.

7. Abre el archivo ejecutable de MrBayes. Notarás que, a diferencia de todos los programas anteriores, MrBayes no tiene una interfase amigable, sino que deberás teclear manualmente cada una de las órdenes para ejecutar el programa (Fig. 78, flecha).

 C:\Users\mlnr\Documents\Articulos en preparación\Manual\Práctica7\mrbayes32\mrbayes.exe

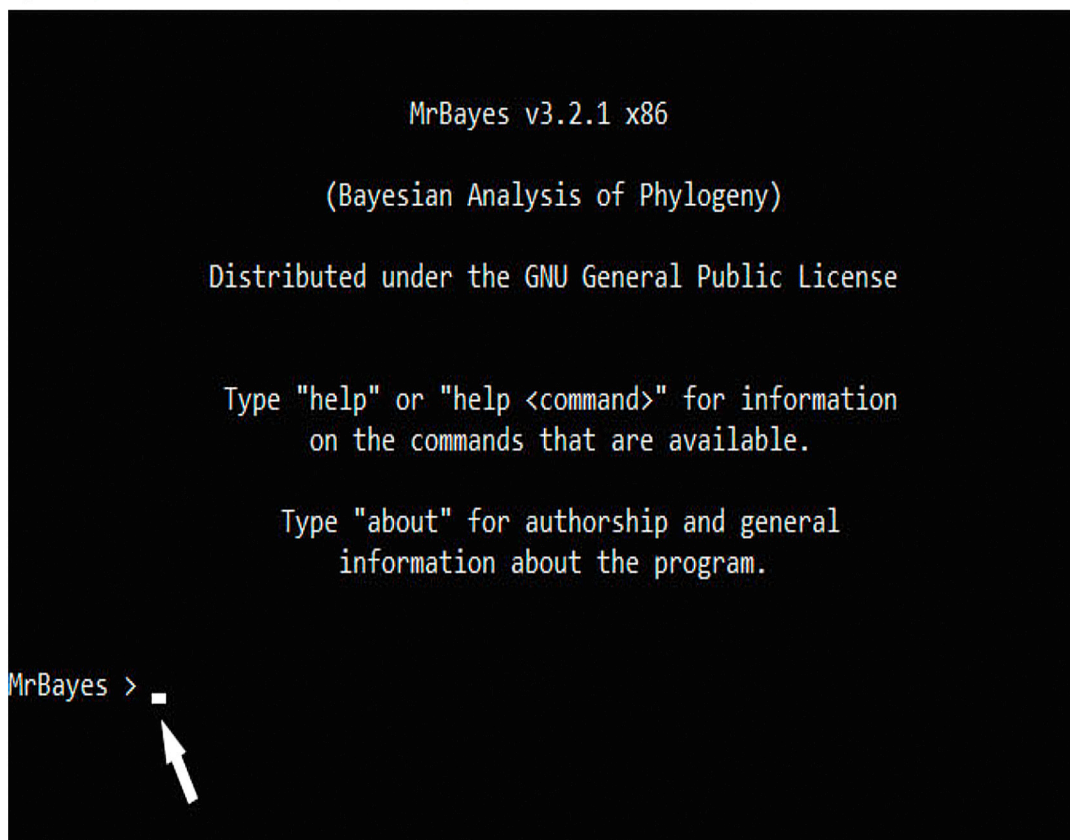


Fig. 78. Captura de pantalla paso 7: interfase del programa Mr. Bayes

8. Para abrir la matriz de datos en el programa deberás teclear el comando **exe** seguido del nombre de tu archivo y la extensión **.nex**. En este caso escribe **exe Solieriaceae.nex** y presiona la tecla enter.
9. Aparecerá en la interfase el resumen de tus datos y al final de la lista de taxa aparecerá la leyenda **successfully read matrix** (Fig. 79, cabeza de flecha). Esta leyenda indicará que el programa leyó correctamente la matriz y que podrá ejecutar el análisis, de haber un error en la matriz, en lugar de esta leyenda encontrarás una leyenda de error.

```

C:\Users\mlnr\Documents\Articulos en preparación\Manual\Práctica7\mrbayes32\mrbayes.exe
MrBayes > exe Solieriaceae.nex

Executing file "Solieriaceae.nex"
DOS line termination
Longest line length = 94
Parsing file
Expecting NEXUS formatted file
Reading taxa block
  Allocated taxon set
  Defining new set of 31 taxa
Exiting taxa block
Reading characters block
  Allocated matrix
  Defining new character matrix with 1258 characters
  Data is Dna
  Missing data coded as ?
  Gaps coded as -
  Matching characters coded as .
  Taxon 1 -> Chondrophycuspapillo
  Taxon 2 -> Gracilariaflabellifo
  Taxon 3 -> Agardhiellaramosissi
  Taxon 4 -> AgardhiellasubulataU
  Taxon 5 -> AnathecamontagneiAB1
  Taxon 6 -> BetaphycusgelatinusJ
  Taxon 7 -> Betaphycusphilippine
  Taxon 8 -> BetaphycusspeciosumA
  Taxon 9 -> Eucheumatopsisisisifor
  Taxon 10 -> Kappaphycusalvarezii
  Taxon 11 -> KappaphycuscottoniiA
  Taxon 12 -> KappaphycusstriatumJ
  Taxon 13 -> Meristothecapapulosa
  Taxon 14 -> Meristothecaprocumben
  Taxon 15 -> Meristothecaimbricat
  Taxon 16 -> Meristothecadakarens
  Taxon 17 -> MeristothecacoactaAB
  Taxon 18 -> MeristothecaspAF0997
  Taxon 19 -> Meristothecacylindri
  Taxon 20 -> Meristothecagelidium
  Taxon 21 -> Sarcodiothecagaudich
  Taxon 22 -> Sarcodiothecafurcata
  Taxon 23 -> SarconemafiliformeAF
  Taxon 24 -> SolieriachordalisAF0
  Taxon 25 -> SolieriafiliformisKJ
  Taxon 26 -> Solieria incurvataMG0
  Taxon 27 -> SolieriapacificaAF09
  Taxon 28 -> SolieriarobustaKC130
  Taxon 29 -> TacanooscauncinataAF
  Taxon 30 -> TepoztequiellaspAF09
  Taxon 31 -> Tepoztequiellarh01
Successfully read matrix

```

Fig. 79. Captura de pantalla pasos 8 y 9: importación exitosa de una matriz nexus al programa MrBayes

10. Para determinar el modelo general de sustitución nucleotídica, teclea el comando `lset Nucmodel=4by4` y la tecla enter. Si introduces un parámetro que el programa no reconozca, la leyenda que mostrará será **Could not find command** (Fig. 80a), pero no tienes que volver a iniciar el proceso, sólo escribe el parámetro correcto nuevamente y te mostrará la leyenda **Successfully set likelihood model parameters** (Fig. 80b).

Nota: **4by4** es el modelo estándar de sustitución de ADN en el que hay sólo cuatro estados (A, T, C, G); las otras opciones posibles para el comando `lset Nucmodel=` son **doublet** (para un modelo apropiado para modelar las regiones madre de los genes ribosomales en donde el espacio de estado son los 16 dobles de nucleótidos) o **codon** (el modelo de sustitución se expande alrededor de los tripletes de nucleótidos - un codón).

```

C:\Users\mlnr\Documents\Articulos en preparación\Manual\Práctica7\mrbayes32\mrbayes.exe
Taxon 7 -> Betaphycusphilippine
Taxon 8 -> BetaphycusspeciosumA
Taxon 9 -> Eucheumatopsisisisifor
Taxon 10 -> Kappaphycusalvarezii
Taxon 11 -> KappaphycuscottoniiA
Taxon 12 -> KappaphycusstriatumJ
Taxon 13 -> Meristotheccapapulosa
Taxon 14 -> Meristotheccaprocumben
Taxon 15 -> Meristotheccaimbricat
Taxon 16 -> Meristotheccadakarens
Taxon 17 -> MeristotheccacoactaAB
Taxon 18 -> MeristotheccaspAF0997
Taxon 19 -> Meristotheccacylindri
Taxon 20 -> Meristotheccagelidium
Taxon 21 -> Sarcodiotheccagaudich
Taxon 22 -> Sarcodiotheccafurcata
Taxon 23 -> SarconemafiliformeAF
Taxon 24 -> SolieriachordalisAF0
Taxon 25 -> SolieriafiliformiskJ
Taxon 26 -> Solieria incurvataMG0
Taxon 27 -> SolieriapacificafAF09
Taxon 28 -> SolieriarobustakKC130
Taxon 29 -> TacanooscauncinataAF
Taxon 30 -> TepoztequiellaspAF09
Taxon 31 -> Tepoztequiellarhizoi
Successfully read matrix
Setting default partition (does not divide up characters)
Setting model defaults
Seed (for generating default start values) = 1536608520
Setting output file names to "Solieriaceae.nex.run<i>.<p|t>"
Exiting characters block
Skipping "CODONS" block
Skipping "CODONUSAGE" block
Skipping "DnaSP" block
Reached end of file

MrBayes > tn93
      Could not find command "tn93"

MrBayes > lset Nucmodel=4by4
      Setting Nucmodel to 4by4
      Set state frequency prior to default
      Successfully set likelihood model parameters

MrBayes >
  
```

Fig. 80. Captura de pantalla paso 10: ejemplo de leyendas al introducir un parámetro en MrBayes: (a) leyenda de parámetro mal introducido y (b) leyenda de parámetro exitosamente introducido

11. Para determinar la tasa de sustitución de bases teclea **lset Nst=1** y la tecla enter (Fig. 81a).

Nota: **1** limita todas las tasas para ser la misma; las otras opciones posibles para el comando **lset Nst=** son **2** (permite que las transiciones y transversiones tengan tasas potencialmente diferentes) y **6** (permite que todas las tasas sean diferentes, sujeto a la restricción de reversibilidad temporal).

12. Para determinar el tipo de código genético teclea el comando **lset Code=Universal** y la tecla enter (Fig. 81b).

13. Para determinar la tasa de variación entre sitios teclea el comando **lset Rates=Gamma** y la tecla enter (Fig. 81c).

Nota: **gamma** es la opción para un único parámetro de distribución de las tasas; las otras opciones posibles para este comando son **equal** (cuando no hay variación de tasas entre los sitios), **adgamma** (cuando la distribución de la tasa marginal es gamma, pero los sitios adyacentes tienen tasas correlacionadas) y **propinv** (cuando una proporción de los sitios son invariables).

14. Para determinar el número de generaciones que deseas muestrear, teclea el comando **mcmc ngen=1000000** y la tecla enter (Fig. 81d).

Nota: para un análisis robusto se recomienda mínimo 1000000 de generaciones.

15. Para determinar el número de cadenas que se ejecutarán para la variante MCMCMC, teclea el comando **mcmc nchains=4** y la tecla enter (Fig. 81e).

Nota: el valor predeterminado es 4: 1 cadena fría y 3 cadenas calientes, es posible elegir más de 4 cadenas.

16. Para determinar la frecuencia de muestreo de las cadenas de Markov, teclea el comando **mcmcp samplefreq=500** y la tecla enter (Fig. 81f).

Nota: puedes muestrear la cadena en cada ciclo, pero esto daría como resultado archivos de salida muy grandes. Acortar la frecuencia de muestreo de las cadenas es una forma de reducir estos archivos y hacer que las muestras sean independientes.

17. Para determinar la frecuencia con que se imprime la información sobre las cadenas en la pantalla, teclea el comando **mcmcp printfreq=500** y la tecla enter (Fig. 81g).

Notas: se recomienda que entre mayor sea el número de generaciones, mayor sea la frecuencia de impresión. Por ejemplo, si se corren 50 000 000 de generaciones, se recomienda una frecuencia de muestreo de 1 000.

18. Para determinar el grupo externo con que enraizaremos el árbol, teclea el comando **outgroup Chondrophycuspapillo** y la tecla enter (Fig. 81h).

Nota: normalmente el programa reconoce de manera automática al primer taxon de la matriz como el grupo externo. En caso de que olvides definirlo o quieras obviarlo, siempre coloca al inicio de la matriz la o las secuencias que vayan a funcionar como tu grupo externo.

19. Una vez definidos los parámetros ejecutaremos el análisis tecleando el comando **mcmc** y la tecla enter (Fig. 81i).

```

C:\Users\mlnr\Documents\Articulos en preparación\Manual\Práctica7\mrbayes32\mrbayes.exe
MrBayes > lset Nucmodel=4by4
Setting Nucmodel to 4by4
Set state frequency prior to default
Successfully set likelihood model parameters
MrBayes > lset Nst=1 ← a
Setting Nst to 1
Successfully set likelihood model parameters
MrBayes > lset Code=Universal ← b
Setting Code to Universal
Successfully set likelihood model parameters
MrBayes > lset Rates=Gamma ← c
Setting Rates to Gamma
Successfully set likelihood model parameters
MrBayes > mcmcp ngen=1000000 ← d
Setting number of generations to 1000000
Successfully set chain parameters
MrBayes > mcmcp nchains=4 ← e
Setting number of chains to 4
Successfully set chain parameters
MrBayes > mcmcp samplefreq=500 ← f
Setting sample frequency to 500
Successfully set chain parameters
MrBayes > mcmcp printfreq=500 ← g
Setting print frequency to 500
Successfully set chain parameters
MrBayes > outgroup Chondrophycuspapillo ← h
Setting outgroup to taxon "Chondrophycuspapillo"
MrBayes > mcmc_ ← i
  
```

Fig. 81. Captura de pantalla pasos 11 a 19: definición de parámetros de análisis: (a) sustitución de bases, (b) código genético, (c) tasa de variación entre sitios, (d) número de generaciones, (e) cadenas de MCMC, (f) frecuencia de muestreo, (g) frecuencia de impresión del muestreo, (h) grupo externo y (i) ejecutar análisis

20. Notarás que el análisis se está ejecutando. En la pantalla te mostrará la información de las cadenas cada 500 como definimos el parámetro (Fig. 82).

Nota: este paso tomará un buen rato, la rapidez dependerá del número de generaciones definidas y la velocidad de tu computadora. Se paciente puede tomar incluso horas.

```
C:\Users\mlnr\Documents\Articulos en preparación\Manual\Practica7\mrbayes32\mrbayes.exe
982000 -- (-7640.522) (-7638.635) (-7623.092) [-7615.215] * (-7621.330) (-7625.054) [-7622.590] (-7633.246) -- 0:00:52
982500 -- (-7651.561) (-7631.032) (-7622.850) [-7626.826] * [-7619.730] (-7616.822) (-7623.558) (-7631.248) -- 0:00:51
983000 -- (-7640.400) (-7626.621) [-7622.380] (-7630.265) * [-7618.802] (-7616.841) (-7632.430) (-7621.998) -- 0:00:49
983500 -- (-7633.369) (-7631.154) (-7624.429) [-7628.456] * [-7625.760] (-7632.087) (-7626.268) (-7630.906) -- 0:00:48
984000 -- (-7628.292) (-7618.608) [-7626.598] (-7639.441) * [-7616.079] (-7622.129) (-7621.626) (-7624.203) -- 0:00:46
984500 -- (-7622.537) (-7633.973) (-7644.062) [-7620.245] * (-7617.757) (-7632.486) [-7626.488] (-7633.580) -- 0:00:45
985000 -- (-7624.867) (-7632.227) (-7629.492) [-7622.068] * [-7620.969] (-7620.462) (-7629.702) (-7626.950) -- 0:00:43

Average standard deviation of split frequencies: 0.006920

985500 -- [-7618.557] (-7633.244) (-7644.593) (-7618.360) * (-7621.009) (-7630.267) (-7627.134) [-7623.375] -- 0:00:42
986000 -- (-7629.472) (-7625.814) (-7638.368) [-7621.816] * (-7626.746) (-7640.666) (-7648.634) [-7617.334] -- 0:00:40
986500 -- (-7617.928) [-7624.995] (-7635.262) (-7634.799) * [-7623.890] (-7640.087) (-7635.981) (-7625.462) -- 0:00:39
987000 -- (-7614.716) (-7628.492) [-7623.161] (-7614.842) * (-7637.645) (-7632.184) (-7634.405) [-7624.611] -- 0:00:38
987500 -- (-7619.645) [-7625.658] (-7624.631) (-7632.470) * (-7625.483) (-7621.360) (-7642.091) [-7620.410] -- 0:00:36
988000 -- (-7630.279) (-7626.768) (-7643.793) [-7623.234] * (-7629.895) (-7627.561) [-7625.116] (-7632.153) -- 0:00:35
988500 -- (-7640.249) [-7627.761] (-7640.855) (-7641.068) * (-7631.952) (-7623.918) (-7625.882) [-7627.007] -- 0:00:33
989000 -- (-7623.712) (-7621.946) [-7630.758] (-7629.529) * (-7637.949) (-7635.900) [-7620.780] (-7618.107) -- 0:00:32
989500 -- [-7628.977] (-7628.042) (-7633.403) (-7640.968) * (-7634.182) (-7623.808) (-7624.537) [-7613.426] -- 0:00:30
990000 -- (-7634.839) (-7624.358) (-7627.237) [-7624.452] * [-7626.982] (-7624.169) (-7639.766) (-7626.517) -- 0:00:29

Average standard deviation of split frequencies: 0.007138

990500 -- (-7644.548) (-7617.732) [-7624.444] (-7633.440) * [-7632.042] (-7631.179) (-7655.120) (-7633.088) -- 0:00:27
991000 -- (-7645.192) [-7619.592] (-7620.449) (-7624.990) * (-7625.939) [-7624.761] (-7643.499) (-7619.558) -- 0:00:26
991500 -- (-7635.039) (-7622.009) (-7627.121) [-7625.831] * (-7627.305) [-7619.218] (-7636.062) (-7644.871) -- 0:00:24
992000 -- (-7627.539) (-7620.037) (-7628.553) [-7620.535] * [-7622.648] (-7637.444) (-7617.337) (-7631.672) -- 0:00:23
992500 -- (-7625.397) (-7618.757) (-7629.604) [-7617.537] * [-7618.866] (-7629.275) (-7618.773) (-7641.462) -- 0:00:21
993000 -- (-7631.614) (-7634.128) (-7628.256) [-7622.383] * (-7610.246) (-7635.973) [-7636.730] (-7644.676) -- 0:00:20
993500 -- (-7635.015) [-7613.223] (-7639.680) (-7622.038) * (-7636.454) (-7625.563) [-7634.108] (-7648.920) -- 0:00:19
994000 -- (-7633.485) (-7633.496) [-7631.407] (-7618.864) * (-7638.641) [-7623.062] (-7641.984) (-7637.764) -- 0:00:17
994500 -- (-7631.593) (-7622.159) (-7643.486) [-7615.680] * (-7644.127) [-7610.651] (-7618.682) (-7634.180) -- 0:00:16
995000 -- (-7637.857) (-7626.702) (-7624.696) [-7620.704] * (-7631.772) (-7623.327) [-7617.829] (-7630.987) -- 0:00:14

Average standard deviation of split frequencies: 0.007012

995500 -- (-7615.068) [-7621.269] (-7639.121) (-7621.799) * [-7613.971] (-7637.569) (-7627.330) (-7621.960) -- 0:00:13
996000 -- (-7626.809) (-7635.490) (-7628.811) [-7616.756] * [-7614.645] (-7622.976) (-7623.252) (-7624.235) -- 0:00:11
996500 -- (-7622.121) (-7617.571) (-7637.950) [-7623.750] * [-7613.051] (-7634.963) (-7625.770) (-7644.386) -- 0:00:10
997000 -- [-7615.727] (-7626.212) (-7636.550) (-7628.462) * (-7622.575) [-7621.742] (-7630.926) (-7643.897) -- 0:00:08
997500 -- [-7610.123] (-7620.333) (-7624.061) (-7632.481) * (-7630.368) [-7621.987] (-7628.294) (-7634.107) -- 0:00:07
998000 -- (-7631.976) [-7617.585] (-7641.895) (-7628.579) * (-7637.289) [-7621.853] (-7616.468) (-7638.071) -- 0:00:05
998500 -- (-7641.891) (-7614.168) [-7622.132] (-7642.986) * (-7635.054) [-7629.042] (-7618.195) (-7627.357) -- 0:00:04
999000 -- (-7626.175) (-7624.310) [-7614.726] (-7635.172) * (-7639.912) (-7626.184) (-7628.138) [-7624.769] -- 0:00:02
999500 -- (-7636.359) [-7618.885] (-7624.593) (-7621.161) * [-7625.167] (-7628.995) (-7635.061) (-7623.134) -- 0:00:01
1000000 -- (-7629.339) [-7614.294] (-7640.356) (-7622.301) * (-7625.911) (-7629.211) [-7626.821] (-7614.567) -- 0:00:00

Average standard deviation of split frequencies: 0.007252

Continue with analysis? (yes/no):  ←
```

Fig. 82. Captura de pantalla paso 20: análisis en ejecución

21. Cuando el programa termina de muestrear las generaciones que definiste te preguntará si deseas que continúe corriendo más generaciones, tecléa **no** y la tecla enter (Fig. 83).


```

C:\Users\mlnr\Documents\Articulos en preparación\Manual\Práctica7\mrbayes32\mrbayes.exe
982500 -- (-7651.561) (-7631.032) (-7622.850) [-7626.826] * [-7619.730] (-7616.822) (-7623.558) (-7631.248) -- 0:00:51
983000 -- (-7640.400) (-7626.621) [-7622.380] (-7630.265) * [-7618.802] (-7616.841) (-7632.430) (-7621.998) -- 0:00:49
983500 -- (-7633.369) (-7631.154) (-7624.429) [-7628.456] * [-7625.760] (-7632.087) (-7626.268) (-7630.906) -- 0:00:48
984000 -- (-7628.292) (-7618.608) [-7626.598] (-7639.441) * [-7616.079] (-7622.129) (-7621.626) (-7624.203) -- 0:00:46
984500 -- (-7622.537) (-7633.973) (-7644.062) [-7620.245] * [-7617.757] (-7632.486) [-7626.488] (-7633.580) -- 0:00:45
985000 -- (-7624.867) (-7632.227) (-7629.492) [-7622.068] * [-7620.969] (-7620.462) (-7629.702) (-7626.950) -- 0:00:43

Average standard deviation of split frequencies: 0.006920

985500 -- [-7618.557] (-7633.244) (-7644.593) (-7618.360) * (-7621.009) (-7630.267) (-7627.134) [-7623.375] -- 0:00:42
986000 -- (-7629.472) (-7625.814) (-7638.368) [-7621.816] * (-7626.746) (-7640.666) (-7648.634) [-7617.334] -- 0:00:40
986500 -- (-7617.928) [-7624.995] (-7635.262) (-7634.799) * [-7623.890] (-7640.087) (-7635.981) (-7625.462) -- 0:00:39
987000 -- (-7614.716) (-7628.492) [-7623.161] (-7614.842) * (-7637.645) (-7632.184) (-7634.405) [-7624.611] -- 0:00:38
987500 -- (-7619.645) [-7625.658] (-7624.631) (-7632.470) * (-7625.483) (-7621.360) (-7642.091) [-7620.410] -- 0:00:36
988000 -- (-7630.279) (-7626.768) (-7643.793) [-7623.234] * (-7629.895) (-7627.561) [-7625.116] [-7632.153] -- 0:00:35
988500 -- (-7640.249) [-7627.761] (-7640.855) (-7641.068) * (-7631.952) (-7623.918) (-7625.882) [-7627.007] -- 0:00:33
989000 -- (-7623.712) (-7621.946) [-7630.758] (-7629.529) * (-7637.949) (-7635.900) [-7620.780] (-7618.107) -- 0:00:32
989500 -- [-7628.977] (-7628.042) (-7633.403) (-7640.968) * (-7634.182) (-7623.808) (-7624.537) [-7613.426] -- 0:00:30
990000 -- (-7634.839) (-7624.358) (-7627.237) [-7624.452] * [-7626.982] (-7624.169) (-7639.766) (-7626.517) -- 0:00:29

Average standard deviation of split frequencies: 0.007138

990500 -- (-7644.548) (-7617.732) [-7624.444] (-7633.440) * [-7632.042] (-7631.179) (-7655.120) (-7633.088) -- 0:00:27
991000 -- (-7645.192) [-7619.592] (-7620.449) (-7624.990) * (-7625.939) [-7624.761] (-7643.499) (-7619.558) -- 0:00:26
991500 -- (-7635.039) (-7622.009) (-7627.121) [-7625.831] * (-7627.305) [-7619.218] (-7636.062) (-7644.871) -- 0:00:24
992000 -- (-7627.539) (-7620.037) (-7628.553) [-7620.535] * [-7622.648] (-7637.444) (-7617.337) (-7631.672) -- 0:00:23
992500 -- (-7625.397) (-7618.757) (-7629.604) [-7617.537] * [-7618.866] (-7629.275) (-7618.773) (-7641.462) -- 0:00:21
993000 -- (-7631.614) (-7634.128) (-7628.256) [-7622.383] * (-7610.246) (-7635.973) [-7636.730] (-7644.676) -- 0:00:20
993500 -- (-7635.015) [-7613.223] (-7639.680) (-7622.038) * (-7636.454) (-7625.563) [-7634.108] (-7648.920) -- 0:00:19
994000 -- (-7633.485) (-7633.496) (-7631.407) [-7618.864] * (-7638.641) [-7623.062] (-7641.984) (-7637.764) -- 0:00:17
994500 -- (-7631.593) (-7622.159) (-7643.486) [-7615.680] * (-7644.127) [-7610.651] (-7618.682) (-7634.180) -- 0:00:16
995000 -- (-7637.857) (-7626.702) (-7624.696) [-7620.704] * (-7631.772) (-7623.327) [-7617.829] (-7630.987) -- 0:00:14

Average standard deviation of split frequencies: 0.007012

995500 -- (-7615.068) [-7621.269] (-7639.121) (-7621.799) * [-7613.971] (-7637.569) (-7627.330) (-7621.960) -- 0:00:13
996000 -- (-7626.809) (-7635.490) (-7628.811) [-7616.756] * [-7614.645] (-7622.976) (-7623.252) (-7624.235) -- 0:00:11
996500 -- (-7622.121) (-7617.571) (-7637.950) [-7623.750] * [-7613.051] (-7634.963) (-7625.770) (-7644.386) -- 0:00:10
997000 -- [-7615.727] (-7626.212) (-7636.550) (-7628.462) * (-7622.575) [-7621.742] (-7630.926) (-7643.897) -- 0:00:08
997500 -- [-7610.123] (-7620.333) (-7624.061) (-7632.481) * (-7630.368) [-7621.987] (-7628.294) (-7634.107) -- 0:00:07
998000 -- (-7631.976) [-7617.585] (-7641.895) (-7628.579) * (-7637.289) [-7621.853] (-7616.468) (-7638.071) -- 0:00:05
998500 -- (-7641.891) (-7614.168) [-7622.132] (-7642.986) * (-7635.054) [-7629.042] (-7618.195) (-7627.357) -- 0:00:04
999000 -- (-7626.175) (-7624.310) [-7614.726] (-7635.172) * (-7639.912) (-7626.184) (-7628.138) [-7624.769] -- 0:00:02
999500 -- (-7636.359) [-7618.885] (-7624.593) (-7621.161) * [-7625.167] (-7628.995) (-7635.061) (-7623.134) -- 0:00:01
1000000 -- (-7629.339) [-7614.294] (-7640.356) (-7622.301) * (-7625.911) (-7629.211) [-7626.821] (-7614.567) -- 0:00:00

Average standard deviation of split frequencies: 0.007252

Continue with analysis? (yes/no): no

```

Fig. 83. Captura de pantalla paso 21: fin de muestreo de las generaciones definidas

22. Notarás que desplegará un resumen del análisis (Fig. 84).

C:\Users\minr\Documents\Articulos en preparación\Manual\Práctica7\mrbayes32\mrbayes.exe

```

Analysis completed in 48 mins 47 seconds
Analysis used 2926.42 seconds of CPU time
Likelihood of best state for "cold" chain of run 1 was -7604.02
Likelihood of best state for "cold" chain of run 2 was -7604.64

Acceptance rates for the moves in the "cold" chain of run 1:
  With prob. (last 100) chain accepted proposals by move
    19.6 % ( 24 %) Dirichlet(Pi)
    25.0 % ( 29 %) Slider(Pi)
    25.5 % ( 26 %) Multiplier(Alpha)
     4.4 % (  3 %) ExtSPR(Tau,V)
     1.6 % (  0 %) ExtTBR(Tau,V)
     6.9 % (  3 %) NNI(Tau,V)
     7.9 % (  4 %) ParsSPR(Tau,V)
    26.0 % ( 24 %) Multiplier(V)
    31.9 % ( 27 %) Nodeslider(V)
    23.9 % ( 28 %) TLMultiplier(V)

Acceptance rates for the moves in the "cold" chain of run 2:
  With prob. (last 100) chain accepted proposals by move
    19.3 % ( 30 %) Dirichlet(Pi)
    25.4 % ( 23 %) Slider(Pi)
    26.3 % ( 24 %) Multiplier(Alpha)
     4.3 % (  3 %) ExtSPR(Tau,V)
     1.6 % (  1 %) ExtTBR(Tau,V)
     6.8 % (  3 %) NNI(Tau,V)
     7.8 % ( 13 %) ParsSPR(Tau,V)
    26.1 % ( 24 %) Multiplier(V)
    31.8 % ( 35 %) Nodeslider(V)
    23.6 % ( 22 %) TLMultiplier(V)

Chain swap information for run 1:
      1      2      3      4
-----
1 |          0.66   0.39   0.21
2 | 167189          0.68   0.43
3 | 166433 166122          0.70
4 | 167241 166827 166188

Chain swap information for run 2:
      1      2      3      4
-----
1 |          0.65   0.39   0.22
2 | 166479          0.68   0.42
3 | 167026 166947          0.70
4 | 166632 166574 166342

Upper diagonal: Proportion of successful state exchanges between chains
  
```

Fig. 84. Captura de pantalla paso 22: resumen del análisis

23. Para descartar los árboles de los extremos teclea el comando `sump burnin=250 000` (Fig. 85).

Nota: de todo el universo de árboles muestreados en el análisis, normalmente se descartan aquellos del inicio y el final, para garantizar que los árboles de entre los que seleccionará al mejor este en la media. El comando `burnin` determina el número de muestras (árboles) que serán descartadas del archivo de entrada antes de calcular las estadísticas de resumen. Si hay varios archivos de entrada, se descartará el mismo número de muestras de cada uno. Normalmente se descarta el 25 % de los árboles muestreados, en este caso el 25 % de 1 000 000.

```

C:\Users\mlnr\Documents\Articulos en preparación\Manual\Práctica7\mrbayes32\mrbayes.exe

Acceptance rates for the moves in the "cold" chain of run 2:
  With prob.   (last 100)   chain accepted proposals by move
  19.3 %      ( 30 %)      Dirichlet(Pi)
  25.4 %      ( 23 %)      Slider(Pi)
  26.3 %      ( 24 %)      Multiplier(Alpha)
  4.3 %       ( 3 %)       ExtSPR(Tau,V)
  1.6 %       ( 1 %)       ExtTBR(Tau,V)
  6.8 %       ( 3 %)       NNI(Tau,V)
  7.8 %       ( 13 %)      ParsSPR(Tau,V)
  26.1 %      ( 24 %)      Multiplier(V)
  31.8 %      ( 35 %)      Nodeslider(V)
  23.6 %      ( 22 %)      TLMultiplier(V)

Chain swap information for run 1:

      1      2      3      4
-----
1 |           0.66   0.39   0.21
2 | 167189          0.68   0.43
3 | 166433 166122          0.70
4 | 167241 166827 166188

Chain swap information for run 2:

      1      2      3      4
-----
1 |           0.65   0.39   0.22
2 | 166479          0.68   0.42
3 | 167026 166947          0.70
4 | 166632 166574 166342

Upper diagonal: Proportion of successful state exchanges between chains
Lower diagonal: Number of attempted state exchanges between chains

Chain information:

ID -- Heat
-----
1 -- 1.00 (cold chain)
2 -- 0.91
3 -- 0.83
4 -- 0.77

Heat = 1 / (1 + T * (ID - 1))
      (where T = 0.10 is the temperature and ID is the chain number)

MrBayes > sump burnin=250000 ←

```

Fig. 85. Captura de pantalla paso 23: comando burnin

24. Te desplegará una pantalla mostrándote la tendencia de muestreo y sus parámetros (Fig. 86). Notarás que la distribución es homogénea (Fig. 86, flecha) lo que quiere decir que el muestreo es bueno.

Nota: si la distribución no fuera homogénea, puedes teclear nuevamente el comando `sump burnin=` y reducir o aumentar el porcentaje.


```

C:\Users\mlnr\Documents\Articulos en preparación\Manual\Práctica7\mrbayes32\mrbayes.exe
File "Solieriaceae.nex.lstat" already exists
Overwrite information in this file? (yes/no): no
Appending to file "Solieriaceae.nex.lstat"

Estimated marginal likelihoods for runs sampled in files
  "Solieriaceae.nex.run1.p" and "Solieriaceae.nex.run2.p":
(Use the harmonic mean for Bayes factor comparisons of models)

(Values are saved to the file Solieriaceae.nex.lstat)

Run   Arithmetic mean   Harmonic mean
-----
  1    -7612.06         -7637.30
  2    -7610.78         -7639.23
-----
TOTAL  -7611.22         -7638.67
-----

Model parameter summaries over the runs sampled in files
  "Solieriaceae.nex.run1.p" and "Solieriaceae.nex.run2.p":
Summaries are based on a total of 3002 samples from 2 runs.
Each run produced 2001 samples of which 1501 samples were included.
Parameter summaries saved to file "Solieriaceae.nex.pstat".

File "Solieriaceae.nex.pstat" already exists
Overwrite information in this file? (yes/no): no
Appending to file "Solieriaceae.nex.pstat"

                                95% HPD Interval
                                -----
Parameter      Mean      Variance  Lower      Upper      Median  min ESS*  avg ESS  PSRF+
-----
TL              1.300844  0.004576  1.170442  1.431652  1.299634  1273.65  1387.32  1.000
pi(A)           0.254850  0.000087  0.237375  0.273124  0.254861  1255.13  1279.27  1.000
pi(C)           0.203022  0.000068  0.186585  0.218906  0.203006  1232.58  1252.08  1.000
pi(G)           0.183758  0.000059  0.169612  0.199391  0.183599  912.34   1095.39  1.000
pi(T)           0.358371  0.000121  0.335867  0.378624  0.358296  1054.86  1111.54  1.000
alpha           0.248389  0.000281  0.217015  0.281645  0.248037  1029.78  1189.56  1.000
-----
* Convergence diagnostic (ESS = Estimated Sample Size); min and avg values
  correspond to minimal and average ESS among runs.
  ESS value below 100 may indicate that the parameter is undersampled.
+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
  and Rubin, 1992) should approach 1.0 as runs converge.

MrBayes > sumt ←

```

Fig. 87. Captura de pantalla paso 25: comando sumt

26. Al teclear el comando sumt calculará las probabilidades con los datos que no fueron descartados, y generará dos gráficos (árboles), uno con los valores de credibilidad de la hipótesis o probabilidad posterior (Fig. 88) y otro de la hipótesis filogenética para los datos de la matriz (Fig. 89).

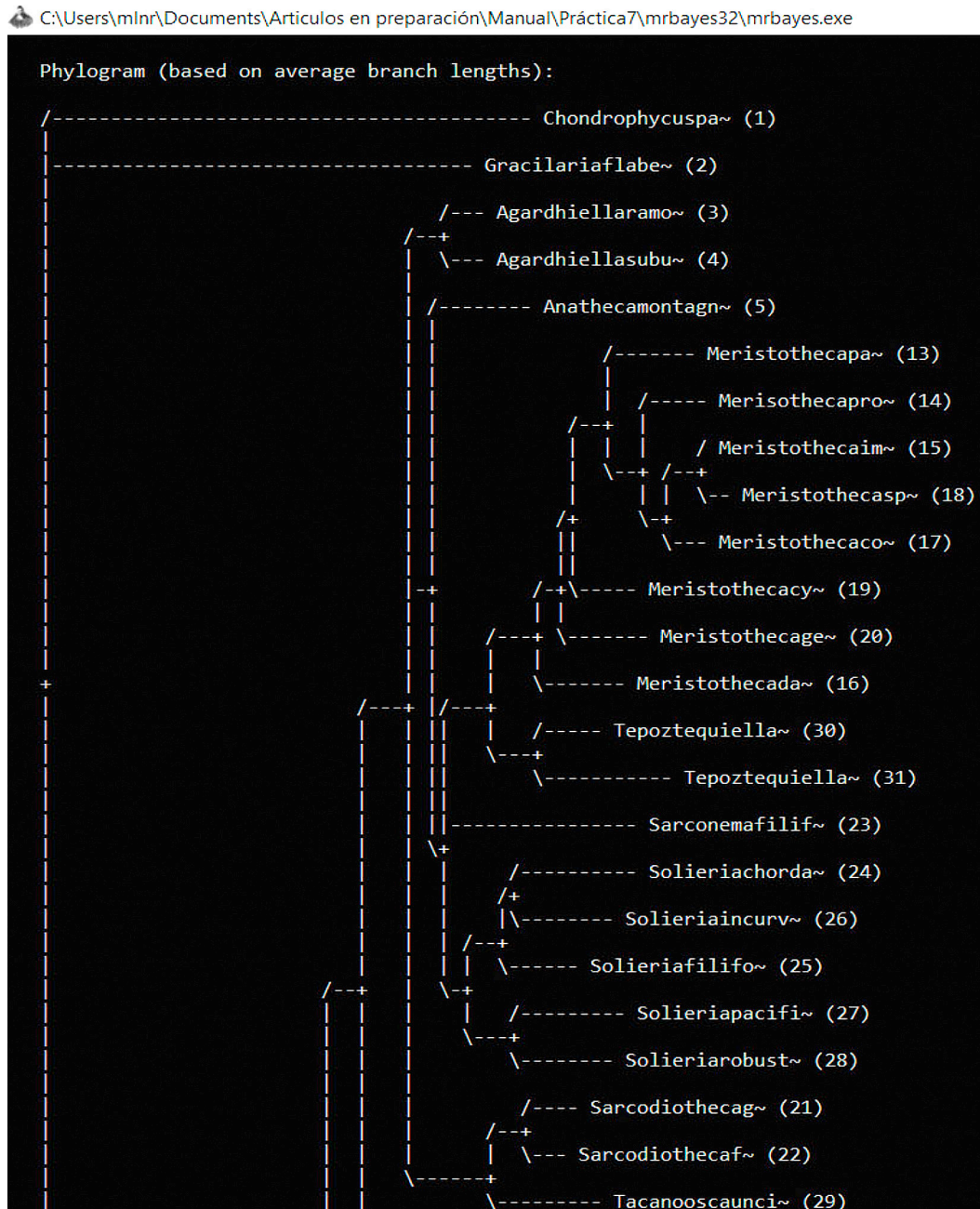


Fig. 89. Captura de pantalla paso 26: gráfico de árbol con la hipótesis filogenética de MrBayes

27. Cierra el programa MrBayes y abre el directorio Prácticas de Filogenia, subdirectorío Práctica 7. En la carpeta del programa MrBayes, notarás que se generaron 14 archivos nuevos, todos con el nombre de la matriz de datos y diferentes extensiones. Esos archivos fueron generados por el programa conforme fuimos designando parámetros. En estos archivos podrás encontrar la información de todos los datos que el programa generó y podrás abrirlos con block de notas, adicionalmente encontrarás el árbol filogenético con extensión **tree**.







RESULTADOS

Al final de la práctica, obtendrás un archivo de árbol **Solieriaceae.nex.con**. Este archivo será utilizado, junto con los obtenidos en prácticas anteriores, para la práctica de inferencia filogenética.

ACTIVIDADES COMPLEMENTARIAS

1. Edita en FigTree el archivo **Solieriaceae.nex.con** añadiendo los valores de probabilidad posterior a los nodos.
2. Guarda el árbol editado en formato JPG.

BIBLIOGRAFÍA RECOMENDADA

-  Huelsenbeck J.P. & Ronquist F. 2001. MRBAYES: "Bayesian inference of phylogenetic trees. Department of Biology", University of Rochester, Rochester, NY. *Bioinformatics* 17: 754–75.
-  Huelsenbeck J.P., Ronquist F., Nielsen R. & Bollback J.P. 2001. "Bayesian Inference of Phylogeny and Its Impact on Evolutionary Biology". *Science* 294: 2310–2314.
-  Librado P. & Rozas J. 2009. "DnaSP v5: a software for comprehensive analysis of DNA polymorphism data". *Bioinformatics* 25: 1451–1452.
-  Rannala B. & Yang Z.H. 1996. "Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. Journal of Molecular". *Evolution* 43: 304–311.
-  Ronquist F., Teslenko M., van der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A. & Huelsenbeck J.P. 2012. "MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space". *Systematic Biology* 61: 539–542.
-  Yang Z. & Rannala B. 2010. "Bayesian species delimitation using multilocus sequence data. Proceedings of the National". *Academy of Sciences USA* 107: 9264–9269.

Práctica 8

Lectura e interpretación de un árbol filogenético

INTRODUCCIÓN

La filogenia es el establecimiento de las relaciones de parentesco entre especies a través del estudio de su historia evolutiva. La representación gráfica de dichas relaciones son los árboles filogenéticos. Existe una terminología específica para la lectura e interpretación de los árboles filogenéticos. A través del reconocimiento de esas partes en nuestro grupo biológico de interés podremos establecer su historia evolutiva.

En términos técnicos, un árbol filogenético (Fig. 90) está conformado por un diagrama ramificado que indica las relaciones filogenéticas de un grupo biológico. Este grupo biológico está conformado por un número determinado de **taxa terminales** (especies) representados por **ramas** de diferentes longitudes (**dicotómicas** o **politómicas**), las cuales a su vez estarán conectadas por **nodos** (ancestros hipotéticos) o puntos de divergencia; estos taxa terminales constituirán el **grupo interno** de nuestro análisis. Adicionalmente, hay un taxon o un grupo de taxa que servirán como raíz para el grupo interno (conectado por la **rama basal**), este es el grupo externo. A su vez, las especies del grupo interno estarán organizadas en grupos de dos o más especies llamados clados. Los clados se definen del interior al exterior del árbol tomando como punto de partida el nodo basal; de este modo, los clados más cercanos a la raíz son los **clados basales** (clados antiguos/especies antiguas) y los clados más lejanos a la raíz son los **clados derivados** (clados recientes/especies de reciente origen). Los **clados menores** al interior de los clados mayores también se denominan **subclados**. Un clado es siempre un grupo monofilético, sin embargo, cuando diferentes taxa de una misma **categoría taxonómica** se encuentran distribuidas en diferentes clados monofiléticos, el grupo de esa categoría será un **grupo parafilético**. El (Los) **grupo(s) hermano(s)** (especie(s) hermana(s) o clado(s) hermano(s)) es(son) el(los) taxon(es) que está(n) más cercanamente emparentado(s) con otro.

ARCHIVOS REQUERIDOS

- Matriz de **distancias no corregidas.exe** resultado de la Práctica 3.
- Matriz de **distancias corregidas.exe** resultado de la Práctica 3.
- Tabla de **ModelosEvolucion.exe** resultado de la Práctica 3.
- Árbol filogenético **UPGMA_Solieriaceae** resultado de la Práctica 4.
- Árbol de bootstrap **UPGMA-Bootstrap_Solieriaceae** resultado de la Práctica 4.
- Árbol filogenético **Neighbor-joining_Solieriaceae** resultado de la Práctica 4.
- Árbol de bootstrap **Neighbor-joining-Bootstrap_Solieriaceae** resultado de la Práctica 4.
- Árbol filogenético **MP_Solieriaceae** resultado de la Práctica 4.
- Árbol de bootstrap **MP-Bootstrap_Solieriaceae** resultado de la Práctica 4.
- Árbol filogenético **result.raxml.bestTree** resultado de la Práctica 5.
- Árbol de valores de soporte **result.raxml.support** resultado de la Práctica 5.
- Árbol filogenético **Solieriaceae.nex.con** resultado de la Práctica 7.

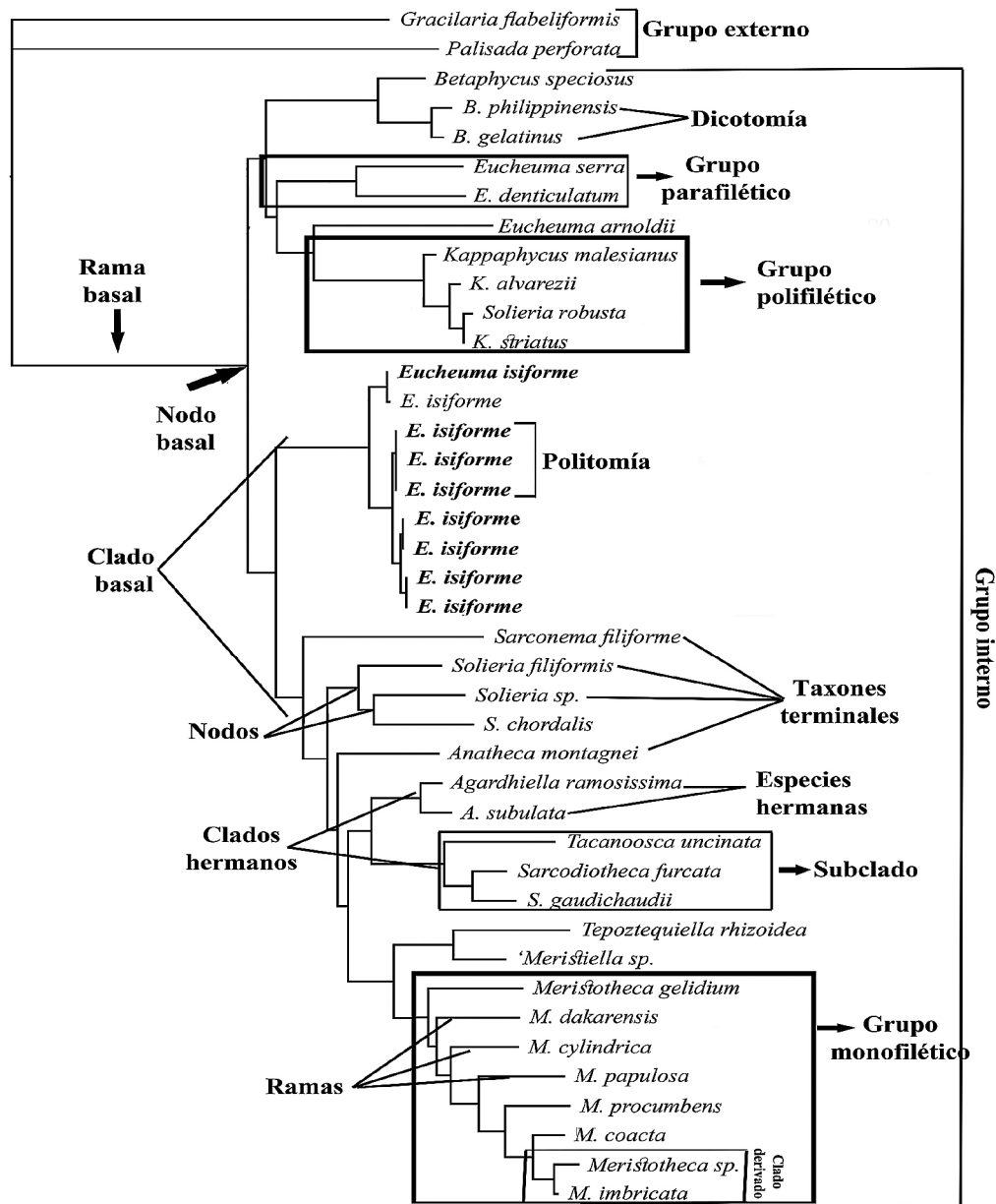


Fig. 90. Componentes de un árbol filogenético

OBJETIVOS

- Conocer los componentes de un árbol filogenético y aplicar a dichos componentes la terminología usada en sistemática.
- Identificar en los árboles filogenéticos de UPGMA, NJ, MP, ML y BI sus componentes.
- Relacionar e interpretar los componentes de un árbol filogenético (inferencia filogenética).
- Integrar la información de los distintos análisis para establecer las relaciones e historia evolutiva de un grupo biológico.

DESARROLLO

1. En los árboles resultantes de cada uno de los diferentes análisis (únicamente uno por análisis, mejor árbol o árbol óptimo), señala:
 - a. Raíz
 - b. Grupo interno
 - c. Grupo externo
 - d. Clados basales
 - e. Clados derivados
 - f. Clados monofiléticos
 - g. Clados parafiléticos
 - h. Subclados
 - i. Especies hermanas
 - j. Clados hermanos
 - k. Nodo basal
 - l. Clados basales
 - m. Clados derivados
 - n. Rama basal
2. Para cada una de las topologías descritas en el paso 1, indica el valor de soporte de los nodos.
3. Con base en los cinco árboles filogenéticos responde lo siguiente:
 - a. ¿Cuál es la especie hermana de *Meristotheca cylindrica*?
 - b. ¿Cuál es la especie hermana de *Kappaphycus striatum*?
 - c. ¿Cuál es la especie hermana de *Meristotheca coacta*?
 - d. ¿Cuál es la especie (o especies) más antigua (basal) de la familia *Solieriaceae*?
 - e. ¿Cuál es el clado hermano o grupo hermano de las especies del género *Meristotheca*?
 - f. ¿Cuál es un género parafilético?
 - g. ¿Cuál es el clado de las especies del género *Betaphycus* y el clado de las especies de *Kappaphycus*?
 - h. ¿Qué especies conforman el clado más antigua de la familia *Solieriaceae*?

- i. ¿Qué tipo de clado conforman las especies del género *Meristotheca*?
 - j. ¿Cuál fue el clado que tuvo los menores valores de soporte?
 - k. ¿En cuántos clados se encuentran repartidas las especies del género parafilético?
 - l. ¿Qué relación de parentesco tiene los géneros *Tacanoosca* y *Sarcofiotheca*?
 - m. ¿Cuáles son los géneros más distantes en la filogenia?
 - n. ¿Cuál es el género más robusto (con valores de soporte más altos) en la filogenia?
 - o. ¿Qué relación tiene *Eucheumatopsis isiformis* con los géneros *Kappaphycus* y *Betaphycus*?
 - p. ¿Con qué especies y con qué tipo de relación se relaciona *Anatheca montagnei*?
 - q. ¿Cuántos clados tuvieron un valor de soporte por debajo del 50 %?
 - r. ¿Qué especies conforman al clado más joven?
 - s. ¿Cuál es el ancestro inmediato de *Meristotheca procumbens*?
 - t. ¿Cuáles el ancestro inmediato de las especies del género *Solieria*?
4. Con base en la matriz de distancias no corregidas responde, ¿cuál es el porcentaje de distancia genética entre:
- a. *Meristotheca cylindrica* y *Meristotheca coacta*?
 - b. *Tepoztequiella rhizoidea* y *Meristotheca dakarensis*?
 - c. *Solieria incurvata* y *S. pacifica*?
 - d. *Agardhiella ramosissima* y *Betaphycus gelatinus*?
 - e. *Solieria chordalis* y *S. filiformis*?
 - f. *Sarcoiotheca gaudichaudii* y *Sa. furcata*?
 - g. *Eucheumatopsis isiformis* y *Meristotheca cylindrica*?
 - h. *Eucheumatopsis isiformis* y *Solieria incurvata*?
 - i. *Tepoztequiella sp.* y *Kappaphycus cottonii*?
 - j. *Tacanoosca uncinata* y *Betaphycus philippinensis*?
5. Con base en la matriz de distancias corregidas con el modelo de Jukes-Cantor responde, ¿cuál es el porcentaje de distancia genética entre:
- a. *Tepoztequiella rhizoidea* y *Solieria incurvata*.
 - b. *Eucheumatopsis isiformis* y *Tepoztequiella rhizoidea*?
 - c. *Tacanoosca uncinata* y *Meristotheca sp.*?

- d. *Kappaphycus cottonii* y *K. alvarezii*?
- e. *Solieria pacifica* y *S. robusta*?
- f. *Agardhiella subulata* y *A. ramosissima*?
- g. *Anatheca montagnei* y *Sarconema filiforme*?
- h. *Betaphycus speciosum* y *B. gelatinus*?
- i. *Meristotheca gelidium* y *M. apullosa*?
- j. *Meristotheca dakarensis* y *Meristotheca* sp.?

RESULTADOS

Al final de la práctica obtendrás las relaciones filogenéticas de las especies de la familia Solieriaceae a partir del marcador rbcL del cloroplasto, inferida a partir de análisis de distancia: UPGMA y Neighbor-Joining, y análisis discretos: Máxima Parsimonia, Máxima Verosimilitud e Inferencia Bayesiana. Adicionalmente, obtendrás porcentajes de distancia genética para sustentar tu inferencia filogenética.

ACTIVIDADES COMPLEMENTARIAS

Presenta un informe con los resultados de todas las prácticas siguiendo el formato de un artículo científico. Para ello debes separa los componentes de las prácticas en **Material y Método**, **Resultados y Discusión**. Guíate respondiendo las siguientes preguntas en cada sección:






Material y métodos: ¿qué gen utilizaste para la obtención de tus secuencias?, ¿con qué *software* fueron editadas y alineadas las secuencias?, ¿cuántas y cuáles especies del GenBank fueron usadas (puedes presentarlas en una tabla con sus números de acceso)?, ¿qué especies fueron usadas como grupo externo?, ¿qué criterio fue usado para seleccionarlas como grupo externo (revisa el concepto de grupo externo)?, ¿qué métodos fueron usados para la inferencia filogenética?, ¿qué modelo de evolución se usó para los análisis y bajo qué criterio fue calculado?, ¿con qué *softwares* se realizaron los análisis y el cálculo del modelo?, ¿cuál fue el test para calcular el soporte de las ramas y cuántas réplicas fueron corridas? En caso de un análisis Bayesiano, ¿cuántas cadenas de Monte Carlo fueron usadas?, ¿cuál fue la frecuencia de muestreo de las cadenas?, ¿cuántas generaciones fueron corridas?, ¿cuántos árboles fueron descartados por burnin?, ¿qué tipo de distancias genéticas fueron calculadas? Para el caso de distancias corregidas, ¿qué modelo fue usado?, ¿en qué programa fueron calculadas?

Resultados: ¿cuántas especies conformaron el alineamiento?, ¿qué longitud (número de nucleótidos) tuvo el alineamiento? Después, describe detalladamente cada una de las topologías obtenidas especificando el número de grandes clados en cada una, el número de subclados que constituyen cada gran clado, ¿qué especies constituyen cada subclado?, ¿qué clados fueron hermanos?, ¿qué especies fueron hermanas dentro de cada subclado?, ¿qué clados, y en consecuencia qué especies, fueron los más basales?, ¿qué valores soportan a cada clado?, ¿cuál es el porcentaje de distancia genética entre las especies de cada clado?

Discusión: discute las diferencias en las topologías resultantes de cada análisis, ¿hubo diferencias en el número de clados?, ¿hubo diferencias entre las especies que conformaron cada clado?, ¿hubo diferencias en las relaciones de parentesco (especies hermanas/grupos hermanos) de cada clado?, ¿hubo diferencias en los valores de soporte?, ¿qué análisis tuvo los mayores valores de soporte?, ¿a qué nivel del cladograma se encontraron los valores más bajos?, ¿el valor de soporte del clado tiene alguna relación con el número de especies que agrupó?, ¿los porcentajes de distancia fueron congruentes con los clados?, ¿entre más basales los clados, la distancia fue mayor?, ¿entre más derivados los clados, el porcentaje de distancia fue menor?, ¿entre especies hermanas y especies del mismo clado el porcentaje de distancia fue mayor o menor?, ¿los porcentajes de distancia entre especies de un mismo subclado fueron similares a los de especies de diferentes subclados?

Nota: cuando se presentan los resultados filogenéticos en una comunicación idónea de resultados (artículo científico), las preguntas de la sección de resultados y discusión se reducen sólo al grupo de interés; es decir, si se trabaja con las especies del género *Meristotheca*, la descripción del clado y de las relaciones filogenéticas así como los porcentajes de distancia genética, serán descritos y discutidos sólo para el clado que involucre a las especies de este género y no a todos los clados del análisis, a menos que se esté estudiando a la familia completa. Dichos resultados a su vez deben ser discutidos mediante la comparación con otros estudios publicados relacionados al grupo de interés.

BIBLIOGRAFÍA RECOMENDADA

-  Graur D. & Wen-Hsiung L. 2000. *Fundamentals of molecular evolution*, ed. 2. Sinauer Associates. Massachusetts.
-  Morrone J.J. 2005. *Sistemática, Biogeografía, Evolución los patrones de la biodiversidad en tiempo-espacio*. Las prensas de ciencias. Facultad de Ciencias UNAM. Ciudad de México.
-  Swofford D.L. & Olsen G.J. 1990. "Phylogeny reconstruction". In : *Molecular Systematics* (Ed. by D.M. Hillis & C. Mortiz), pp. 411–501. Sinauer Associates, Massachusetts, USA.
-  Wiley E.O. 1981. *Phylogenetics. Theory and practice of phylogenetic systematic*. John Wiley and sons, New York.
-  Wiley E.O., Siegel-Causey D., Brooks D.R. & Funk V.A. 1991. *The compleat cladist: a primer of Phylogenetic procedures*, vol. 19. University of Kansas Museum of Natural History, Kansas, USA.

Glosario de términos

A

Alineamiento: disposición en línea de los pares de bases de dos secuencias de ADN.

Alineamiento múltiple: es una extensión del alineamiento de pares que incorpora más de dos secuencias al mismo tiempo, con el fin de alinear todas las secuencias de un conjunto dado. Se utiliza para la identificación de regiones conservadas en un grupo de secuencias que hipotéticamente están relacionadas evolutivamente.

Aminoácido: molécula orgánica, péptido o compuesto químico que forma las proteínas.

Ancestro: especie de la que desciende otra. El antepasado directo por parentesco.

Ancestro común: especie de la cual descienden dos o más especies. Estas especies compartirán al ancestro.

Ancestro inmediato: el ancestro más reciente en una bifurcación.

Apareamiento de bases: interacción entre bases nitrogenadas, mediante la formación de puentes de hidrógeno, que da origen a las formas híbridas o plegadas de los ácidos nucleicos.

Apomorfía: carácter derivado de un estado ancestral.

Árbol filogenético: gráfico que representa las relaciones filogenéticas entre diferentes taxa.

Autoapomorfía: apomorfía exclusiva de un único taxon.

B

Bases nitrogenadas: compuestos orgánicos cíclicos que incluyen dos o más átomos de nitrógeno.

C

Cadena molde: hebra que sirve de plantilla en el proceso de transcripción para hacer una nueva cadena complementaria.

Carácter: rasgo o atributo observable de un organismo.

Categoría taxonómica: rangos o grupos jerárquicos en que se clasifica a los seres vivos.

Cladograma: esquema dicotómico que muestra una hipótesis de relación entre varios taxa.

Clado: grupo monofilético en un cladograma que contiene a un antepasado común y a todos sus descendientes.

Clado basal: conjunto de dos especies o más que se separaron más tempranamente que otros miembros del mismo grupo.

Clado derivado: es el conjunto de especies de deriva (origen) más reciente a partir de un clado ancestral.

Clado hermano: los dos subclados de una bifurcación.

Convergencia: parecido adquirido independientemente en diferentes taxa, por consecuencia no heredados de la especie ancestral.

D

Delección: pérdida de genes o de secciones cromosómicas como resultado de uno o dos rompimientos cromosómicos. Pérdida de bases en una secuencia.

Dendograma: representación gráfica o diagrama de datos en forma de árbol.

Dicotomía: nodo en un árbol que conecta tres ramas: un linaje ancestral y dos ramas que representan a los linajes descendientes actuales.

Distancia genética: porcentaje de divergencia entre dos taxa. Distancia no corregida: porcentaje de divergencia entre dos taxa sin corrección por modelo evolutivo que considere estados plesiomórficos en el proceso evolutivo. Distancia corregida: porcentaje de divergencia en el que se considera un modelo evolutivo para la corrección de porcentajes considerando eventos evolutivos que pudieron originar el regreso a un estado plesiomórfico.

Distancia intraespecífica: porcentaje de divergencia entre individuos de una misma población.

Distancia interespecífica: porcentaje de divergencia entre dos especies.

E

Especies hermanas: los dos taxa de cada bifurcación en un cladograma.

Evolución mínima: algoritmo para encontrar el árbol más corto, usado en el método de neighbor-joining.

G

Gap: espacio en una secuencia en que una base se perdió o se ganó para sólo una secuencia o un grupo de secuencias en el alineamiento.

Gen: unidad estructural y funcional fundamental. Secuencia en el ADN compuesta por una región reguladora y una región que se transcribe como unidad.

Genoma: constitución genética de un organismo.

Grupo externo: cualquier grupo usado en el análisis que no es incluido en el grupo interno (grupo bajo estudio). Sirve para enraizar el árbol.

Grupo hermano: taxon filogenéticamente más próximo al grupo interno.

Grupo interno: grupo formado por todas las especies bajo estudio.

Grupo monofilético: grupo formado por un ancestro y todos sus descendientes.

Grupo parafilético: grupo que incluye al ancestro pero no a todos sus descendientes. Se basa en simplesiomorfias.

Grupo polifilético: grupo formado por descendientes de varios ancestros diferentes. Se basa en convergencias.

H

Hipótesis filogenética: inferencia de las relaciones evolutivas de un grupo de taxa a partir de un árbol filogenético.

Homología: caracteres diferentes que tienen una relación ancestral descrita como preexistente.

Homología de posición: sitios en los que dos secuencias son similares entre sí debido a que presentan un mismo origen evolutivo.

Homoplasia: caracteres similares que no tienen un origen común. Son el resultado de la convergencia evolutiva o paralelismo.

I

Indel: contracción de las palabras inserción y delección, representadas en una secuencia por GAPS.

Inserción: ganancia de bases en una secuencia, producto de una mutación.

L

Longitud de rama: sustituciones de nucleótidos por sitio. Número de cambios evolutivos entre especies.

M

Marcador molecular: segmento de ADN con una ubicación física conocida en un cromosoma. Segmento de un gen o un gen completo.

Mutación: cambio que se produce en la secuencia de un gen.

N

Nodo: punto en donde cada rama se bifurca o termina.

Nucleótido: base nitrogenada, púrica o pirimidica, unida covalentemente a una azúcar y a un grupo fosfato.

P

Paralelismo: parecido adquirido independientemente en taxa relacionados.

Parsimonia: regla científica que postula que, si existen dos respuestas a un problema, la más simple es la correcta.

Pirimidina: compuesto nitrogenado que se encuentra en los ácidos nucleicos.

Plesiomorfía: estado ancestral de un carácter.

Politomía: un nodo sin bifurcación en un cladograma.

Porcentaje de distancia: valor de no correspondencia genética entre dos taxa.

Porcentaje de identidad: valor de correspondencia genética entre dos taxa.

Primer: cadena de ácido nucleico, conformada por 16 a 24 pares de bases altamente conservadas en un organismo dado, que sirve como punto de partida para la replicación del ADN.

Purina: compuesto nitrogenado que se encuentra en los ácidos nucleicos.

R

Raíz: nodo que da origen a todos los terminales.

Rama: línea que une a un nodo con otro.

Rama basal: la rama más antigua en una filogenia. Es la de mayor longitud.

Rama terminal: representación de las especies más derivadas o de especiación reciente en una filogenia. Son las de menor longitud.

S

Secuencia: porción de una molécula de ácidos nucleicos.

Secuencia consenso: secuencia ideal que representan los nucleótidos o aminoácidos que se encuentran con mayor frecuencia en cada posición de un fragmento de DNA.

Simplesiomorfía: carácter plesiomórfico presente en dos o más taxones.

Sitios: posición fija en una secuencia que determina la posición de una base.

Sitios conservados: regiones en un genoma (secuencias) que se mantienen estables, no cambian.

Sitios informativos: regiones en un genoma (secuencia) que tienen al menos dos estados de carácter, repetidos en al menos dos taxa.

Subclado: clado pequeño al interior de un clado mayor.

Sustitución: mutación en la que cambia un nucleótido por otro. Mutación que se fija en una población (especie).

T

Tasa de sustitución: velocidad de cambio de una base por otra en una secuencia.



















Taxon: grupo de organismos pertenecientes a una categoría taxonómica.





















Taxon terminal: corresponde a los taxa observados.





















Transición: mutación puntual en la que una purina cambia por otra purina o una pirimidina cambia por otra pirimidina.

Transversión: mutación puntual en la que una purina cambia por una pirimidina o viceversa.







Bibliografía general

-  Adams E.N. 1972. "Consensus techniques and the comparison of taxonomic trees". *Systematics Zoology* 37: 27–29.
-  Arnedo M.A. 1999. "Cladismo: la reconstrucción filogenética, basada en Parsimonia". *Boletín de la Sociedad Entomológica aragonesa* 26: 57–84.
-  Bandelt H.J. & Dress A.W.M. 1992. "Split decomposition: A new and useful approach to phylogenetic analysis of distance data". *Molecular phylogenetics and evolution* 1: 242–252.
-  Bandelt H.J. & Dress A.W.M. 1994. "An Order Theoretic Framework for Overlapping Clustering". *Discrete Mathematics* 136: 21–37.
-  Barraclough T.G. & Vogler A. . 2000. "Detecting the geographical pattern of speciation from species-level phylogenies". *American Naturalist* 155: 419–434.
-  Bollback J.P. 2006. "SIMMAP: Stochastic character mapping of discrete traits on phylogenies". *Bioinformatics* 7: 88.
-  Bremer K. 1990. "Combinable component consensus". *Cladistics* 6: 369–372.
-  Arnedo M.A. 1999. "Cladismo: la reconstrucción filogenética, basada en Parsimonia". *Boletín de la Sociedad Entomológica aragonesa* 26: 57–84.
-  Bremer K. 1994. "Branch support and tree stability". *Cladistics* 10: 295–304.
-  Bryant D. & Moulton V. 2004. "Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks". *Molecular Biology and Evolution* 21: 255–265.
-  Castillo-Cerón J.M. & Goyenechea I. 2007. "Conceptos básicos en sistemática filogenética: los Deuterostomados como ejemplo". In: *La sistemática: base del conocimiento de la biodiversidad* (Ed. by A. Contreras-Ramos, M.C. Cuevas-Cardona, I. Goyenechea & U. Iturbe), pp. 145–157. Universidad Autónoma del estado de Hidalgo. Hidalgo. México.
-  Camin J.H. & Sokal R.R. 1965. "A method for deducing branching sequences in phylogeny". *Evolution* 19: 311–326.
-  Doyle J.J. 1992. "Gene trees and species trees: Molecular systematics as one character taxonomy". *Systematic Botany* 17: 144–163.
-  Farris J.S. 1969. "A successive approximations approach to character weighting". *Systematics Zoology* 18: 374–385.
-  Farris J.S. 1970. "Methods of computing Wagner trees". *Systematics Zoology* 23: 548–554.
-  Farris J.S. 1989. "The retention index and homoplasy excess". *Systematics Zoology* 38: 406–407
-  Farris, J. Abert V., Kallersjo M., Lipscomb D. & Kluge A. 1996. "Parsimony jackknifing outperforms neighbor Joining". *Cladistics* 12: 99–124.
-  Felsenstein J. 1978. "Cases in which parsimony or compatibility methods will be positively misleading". *Systematic Zoology* 27: 401–410.
-  Felsenstein J. 1981. "Evolutionary trees from DNA sequences: A maximum likelihood approach". *Journal of Molecular Evolution* 17: 368–376.

-  Felsenstein J. 1986. "Confidence limits on phylogenies: An approach using bootstrap". *Evolution* 39: 783–91.
-  Fitch W.M. 1971. "Distinguishing homologous from analogous proteins". *Systematics Zoology* 20: 406–416.
-  Folder O., Black M., Hoeh W., Lutz R. & Vrijenhoek R. 1994. "DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates". *Molecular Marine Biology and Biotechnology* 3: 294–299.
-  Graur D. & Wen-Hsiung L. 2000. *Fundamentals of molecular evolution*, ed. 2. Sinauer Associates, Massachusetts.
-  Hall T.A. 1999. "BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT". *Nucleic Acids Symposium* 41: 95–98.
-  Hendy M.D. & Penny D. 1982. "Branch and bound algorithms to determine minimal evolutionary trees". *Mathematical Biosciences* 59: 277–290.
-  Higgins D. 2003. "Multiple alignment". In: *The phylogenetic handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing* (Ed. by P. Lemey, M. Salemi & A. Vandamme), pp. 45–71. Cambridge University Press, Cambridge, UK.
-  Hillis D.M. 1997. "Primer: phylogenetic analysis". *Current Biology* 7: R129–R131.
-  Hillis D.M. & Bull J.J. 1993. "An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis". *Systematic Biology* 42: 182–192.
-  Huelsenbeck J.P. & Ronquist F. 2001. "MRBAYES: Bayesian inference of phylogenetic trees. NY". *Bioinformatics* 17: 754–75. Department of Biology, University of Rochester, Rochester,
-  Huelsenbeck J.P., Ronquist F., Nielsen R. & Bollback J.P. 2001. "Bayesian Inference of Phylogeny and Its Impact on Evolutionary Biology". *Science* 294: 2310–2314.
-  Huson D.H. & Bryant D. 2006. "Application of phylogenetic networks in evolutionary studies". *Molecular Biology and Evolution* 23: 254–267.
-  Jukes T. & Cantor C.R. 1969. *Evolution of protein molecules*. In: *Mamalian protein metabolism* (Ed. by H.N. Munro), pp. 21–132. Academic press, New York, USA.
-  Kimura M. 1980. "A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences." *Journal of Molecular Evolution* 16: 111–120.
-  Kimura M. 1981. "Estimation of evolutionary distances between homologous nucleotide sequences". *Proceedings of the National Academy of Sciences USA* 78: 454–458.
-  Kluge A.G. 1998. "Total evidence or taxonomic congruence: cladistics or consensus classification". *Cladistics* 14: 151–158.
-  Kumar S., Stecher G., Li M., Knyaz C. & Tamura K. 2018. "MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms". *Molecular Biology and Evolution* 35: 1547–1549.
-  Lanyon S.M. 1985. "Detecting internal inconsistencies in distance data". *Systematics Zoology* 34: 397–403.
-  Lee M.S.Y. 2001. "Unalignable sequences and molecular evolution". *Trends in Ecology and Evolution* 16: 681–685.
-  Leliaert F., Verbruggen H., Vanormelingen P., Steen F., López-Bautista J.M., Zuccarello G.C. & De Clerck O. 2014. "DNA-based species delimitation in algae". *European Journal Phycology* 49: 179–196.

-  Librado P. & Rozas J. 2009. "DnaSP v5: a software for comprehensive analysis of DNA polymorphism data". *Bioinformatics* 25: 1451–1452.
-  Maddison W.P. 1997. "Gene trees in species trees". *Systematic Biology* 46: 523–536.
-  Maddison W.P. & Maddison D.R. 1992. "MacClade: Analysis of Phylogeny and Character Evolution". *Sinauer Associates Massachusetts*.
-  Morgenstern B., Frech K., Dress A. & Werner T. 1998. "DIALIGN: Finding local similarities by multiple sequence alignment". *Bioinformatics* 14: 290–294.
-  Morrone J.J. 2005. *Sistemática, Biogeografía, Evolución los patrones de la biodiversidad en tiempo-espacio*. Las prensas de ciencias, Facultad de Ciencias UNAM, Ciudad de México.
-  Morrone J.J. 2013. *Sistemática. Fundamentos, métodos, aplicaciones*. Las prensas de ciencias, Facultad de Ciencias UNAM, Ciudad de México.
-  National Centre for Biotechnology Information, U.S. National Library of Medicine 8600 Rockville Pike, Bethesda MD, 20894 USA <https://www.ncbi.nlm.nih.gov/nuccore/U04176.1?report=fasta>
-  Nei M. & Li W.H. 1979. "Mathematical model for studying genetic variation in terms of restriction endonucleases." *Proceedings of the National Academy of Sciences USA* 76: 5269–5273.
-  Nelson G. 1979. "Cladistic analysis and synthesis: Principles and definitions with a historical note on Adamson's Familles des Plantes" (1763–1764). *Systematics Zoology* 28: 1–29.
-  Nixon K.C. 1999. "The parsimony ratchet, a new method for rapid parsimony analysis". *Cladistics* 15: 407–415.
-  Nixon K.C. & Carpenter J.M. 1993. "On outgroups". *Cladistics* 9: 413–426.
-  Page R.D.M. & Holmes E.C. 1998. *Molecular evolution: A Phylogenetic Approach*. Blackwell Science, Cambridge.
-  Piganeau G., Eyre-Walker A., Grimsley N. & Moreau H. 2011. "How and why DNA barcodes underestimate the diversity of microbial eukaryotes". *PLoS One* 6: e16342.
-  Posada D. 2001. "The effect of branch length variation on the selection of models of molecular evolution". *Journal of Molecular Evolution* 52: 434–444.
-  Posada D. & Crandall K.A. 1998. "MODELTEST: testing the model of DNA substitution". *Bioinformatics* 14: 817–818.
-  Rambaut A. 2016. FigTree v1.4.3. Institute of Evolutionary Biology, University of Edinburgh. <http://tree.bio.ed.ac.uk/software/figtree/>
-  Rannala B. & Yang Z.H. 1996. "Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference". *Journal of Molecular Evolution* 43: 304–311.
-  Reid N. & Carstens B. 2012. "Phylogenetic estimation error can decrease the accuracy of species delimitation: a Bayesian implementation of the general mixed Yule-coalescent model". *BMC Evolutionary Biology* 12: 196.
-  Ronquist F. 2004. "Bayesian inference of character evolution". *Trends in Ecology and Evolution* 19: 475–481.
-  Ronquist F. & Huelsenbeck J.P. 2003. "MrBayes 3: Bayesian phylogenetic inference under mixed models". *Bioinformatics* 19: 1572–1574.

- 📖 Ronquist F, Teslenko M., van der Mark P., Ayres D.L., Darling A., Hohna S., Larget B., Liu L., Suchard M.A. & Huelsenbeck J.P. 2012. "MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space". *Systematic Biology* 61: 539–542.
- 📖 Simmons M. . & Freudenstein J.V. 2003. "The effects of increasing genetic distance on alignment of, and tree construction from, rDNA internal transcribed spacer sequences". *Molecular Phylogenetics and Evolution* 26: 444–451.
- 📖 Smith A.B. 1994. "Rooting molecular trees Problems and strategies". *Biological Journal of the Linnean Society* 51: 279–292.
- 📖 Stamatakis A. 2006. "RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models". *Bioinformatics* 22: 2688–2690.
- 📖 Stamatakis A., Hoover P. & Rougemont J. 2008. "A Rapid Bootstrap Algorithm for the RAxML Web-Servers". *Systematic Biology* 75: 758–771.
- 📖 Swofford D.L. 2003. PAUP*. Phylogenetic analysis using parsimony (*and other methods), Versión 4.0b10. Sinauer Associates, Massachusetts.
- 📖 Swofford D.L. & Olsen G.J. 1990. "Phylogeny reconstruction". In: *Molecular Systematic* (Ed. by D.M. Hillis & C. Moritz), pp. 411–501. Sinauer Associates, Massachusetts, USA.
- 📖 Swofford D.L., Olsen G.J., Waddle P.J. & Hillis D.M. 1996. "Phylogenetic inference". In: *Molecular Systematics* (Ed. by D.M. Hillis & C. Moritz), pp. 407–514. Sinauer Associates, Massachusetts, USA.
- 📖 Tajima F.M.N. 1984. "Estimation of evolutionary distance between nucleotide sequences". *Molecular Biology and Evolution* 1: 269–285.
- 📖 Tamura K. & Nei M. 1993. "Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees". *Molecular Biology and Evolution* 10: 512–526.
- 📖 Tamura K., Peterson D., Peterson N., Stecher G., Nei M. & Kumar S. 2011. "MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance and Maximum Parsimony Methods". *Molecular Biology and Evolution* 28: 2731–2739.
- 📖 Templeton A.R. 2001. "Using phylogeographic analyses of gene trees to test species status and processes". *Molecular Ecology* 10: 779–791.
- 📖 Upholt W.B. 1977. "Estimation of DNA sequence divergence from comparison of restriction endonuclease digests". *Nucleic Acids Research* 4: 1257–1265.
- 📖 Verbruggen H. 2014. "Morphological complexity, plasticity, and species diagnosability in the application of old species names in DNA-based taxonomies". *Journal of Phycology* 50: 26–31.
- 📖 Wheeler W. 1990. "Nucleic acid sequence phylogeny and random outgroups". *Cladistics* 6: 363–367.
- 📖 Wiens J.J. 2007. "Species delimitation: new approaches for discovering diversity". *Systematic Biology* 56: 875–878.
- 📖 Wiley E.O. 1981. *Phylogenetics. Theory and practice of phylogenetic systematic*. John Wiley and sons. New York.
- 📖 Wiley E.O., Siegel-Causey D., Brooks D.R. & Funk V.A. 1991. *The compleat cladist: a primer of Phylogenetic procedures*, vol. 19. University of Kansas Museum of Natural History, Kansas, USA.

-  Wilkinson M. 1995. "A comparison of two methods of character construction". *Cladistics* 11: 297–308.
-  Williams D.M. 1992. "DNA analysis: Theory". In: *Cladistics: A practical course in Systematics* (Ed. by P.L. Forey, C.J. Humphries, I.J. Kitching, R.W. Scotland, D.J. Siebert & D.M. Williams), pp. 89–101. Oxford University Press, Oxford, UK.
-  Jukes T. & Cantor C.R. 1969. *Evolution of protein molecules*. In: *Mamalian protein metabolism* (Ed. by H.N. Munro), pp. 21–132. Academic press, New York, USA.
-  Yang Z. 1994. "Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods". *Journal of Molecular Evolution* 39: 306–314
-  Yang Z. & Rannala B. 2010. "Bayesian species delimitation using multilocus sequence data". *Proceedings of the National Academy of Sciences USA* 107: 9264–9269.
-  Zimmermann J., Jahn R. & Gemeinholzer B. 2011. "Barcoding diatoms: evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols". *Organisms Diversity and Evolution* 11: 173–192.



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA

UNIDAD IZTAPALAPA

División de Ciencias Biológicas y de la Salud

Av. San Rafael Atlixco No.186, Col. Vicentina
C.P. 09340, Del. Iztapalapa, México D.F.
Tel.: (01) 58044600