

Filogenias: conceptos y generalidades

Phylogenies: concepts and generalities

Carlos Luis Leopardi-Verde^{1*} y Guadalupe Jeanett Escobedo-Sarti²

Fecha de recepción: 16 de noviembre de 2020

Fecha de aceptación: 5 de enero de 2021

Resumen - Cualquier investigación en biología es un ejercicio de comparación y eso incluye el estudio de la evolución. La indagación de patrones evolutivos en cualquiera de sus dos enfoques (micro o macroevolutivo) establece retos metodológicos para cualquier persona interesada en estos temas. Dichos enfoques tienen el interés común de comprender el origen de las relaciones de parentesco entre los organismos estudiados, aunque las escalas temporales y el nivel de organización en el que se concentran son diferentes. Actualmente, las filogenias son la herramienta más robusta para elucidar las hipótesis de relaciones ancestro-descendiente entre un conjunto de organismos. Estas representaciones consisten en proyecciones diagramáticas bidimensionales (cladogramas) o multidimensionales (redes) que pueden ser estimadas con diferentes aproximaciones (máxima parsimonia, máxima verosimilitud e inferencia bayesiana), según los datos disponibles y el propósito de la investigación. En esta revisión se presenta una introducción a los métodos disponibles para la construcción de filogenias, incluyendo la perspectiva tradicional que utiliza diagramas basados en dicotomías y las nuevas tendencias que tratan de visualizar patrones más complejos a través de redes evolutivas.



Palabras clave: Evolución, máxima parsimonia, máxima verosimilitud, inferencia bayesiana, redes filogenéticas.

Abstract - Any research in biology is an exercise of comparison that includes the study of evolution. The investigation of evolutionary patterns in either of its two approaches (micro or macroevolutionary) raises methodological challenges for any researcher interested in these topics. These approaches have a common interest in understanding the origin the relationships between the studied organisms, although the temporal scales and the level of organization in which they focus are different. Currently, phylogenies are the most robust tool to explain ancestor-descendant relationships between a set of organisms. These diagrams, which are two-dimensional (cladograms) or multidimensional (networks), can be estimated with different approximations (maximum parsimony, maximum likelihood, and bayesian inference) according to the data available and the purpose of the investigation. This review presents an introduction to the methods available for the construction of phylogenies, including the traditional perspective that uses diagrams based on dichotomies and the new trends that try to visualize more complex patterns through evolutionary networks.



Keywords: Evolution, maximum parsimony, maximum likelihood, bayesian inference, phylogenetic networks.

¹ Facultad de Ciencias Biológicas y Agropecuarias de la Universidad de Colima, Km. 40 Autopista Colima-Manzanillo, Crucero de Tecmán, Tecmán, Colima, México, C.P. 28930. *Autor de correspondencia: cleopardi@uclm.mx. ORCID: <https://orcid.org/0000-0001-5172-5114>

²ORCID: <https://orcid.org/0000-0002-4901-971X>

Introducción

Una parte importante del pensamiento humano se basa en comparaciones, por ello podemos saber si hace más o menos calor en un día soleado, si un alimento es dulce o salado, opinar sobre un sistema político, etcétera. De la misma manera, el pensamiento comparado es una parte esencial de cualquier persona que se dedica a la investigación de la biología evolutiva en sus distintos ámbitos, debido a que a través de las comparaciones se pueden descubrir y describir los patrones que caracterizan a los grandes fenómenos evolutivos (Harvey & Pagel, 1991).

La biología evolutiva comparte con otras ciencias -como la astronomía y la geología- la tarea de interpretar fenómenos que son imposibles de comprender si no se conoce su pasado (Harvey & Pagel, 1991). De hecho, la biología evolutiva puede considerarse parte de la sistemática, que para autores como Simpson (2019) incluye tanto a la taxonomía como al estudio de la evolución. La taxonomía se encarga de identificar a los seres vivos, describirlos, nombrarlos y clasificarlos, todo esto con reglas establecidas; en el caso de la botánica, en el Código Internacional de Nomenclatura de Algas, Hongos y Plantas (Turland *et al.*, 2018).

El estudio de la evolución se centra en comprender los patrones y procesos que dieron origen a los organismos tal como los vemos hoy, o en otras palabras, busca elucidar la historia de la vida. Esa historia puede analizarse a grandes rasgos en dos escalas: la microevolutiva, que está enfocada al estudio del cambio genético que acumulan las distintas poblaciones que conforman una metapoblación; normalmente, la escala temporal de esta aproximación está prácticamente circunscrita al presente y quizás a algunas decenas o cientos de años. La segunda escala es la macroevolutiva, en donde se estudian los cambios de linajes completos (conjuntos de especies) a lo largo de periodos de tiempo que corresponden más a la escala geológica, desde miles hasta millones de años (Herron & Freeman, 2014). La herramienta básica con la que se cuenta en biología evolutiva para expresar las hipótesis de relaciones ancestro-descendiente que conforman esa historia y para hacer inferencias sobre los patrones y procesos que pudieran explicarlas son las filogenias.

El enfoque microevolutivo utiliza marcadores moleculares altamente variables o secuencias y numerosas muestras de miembros de la misma especie. Según Hayward, Tollenaere, Dalton-Morgan & Batley (2015), un marcador molecular es un loci genético (una parte del genoma) que puede ser fácilmente rastreado y cuantificado en una población (o en un linaje) y que podría estar asociado con un gen o un carácter de interés. Esta información es procesada con herramientas estadísticas que permiten conocer el grado de variación o diferenciación entre las poblaciones e interpreta las relaciones entre éstas, básicamente utilizando un modelo de red en el que potencialmente todas las poblaciones pueden intercambiar información genética entre sí, lo que se expresa a través de alelos (las diferentes versiones que puede tener un gen) o haplotipos (conjunto de polimorfismos que se heredan juntos) compartidos o únicos. Este enfoque presupone que las especies están conformadas por poblaciones genéticamente vinculadas a través del intercambio ocasional de material genético, sea como gametos o en forma de individuos (Templeton, 2006; Herron & Freeman, 2014).

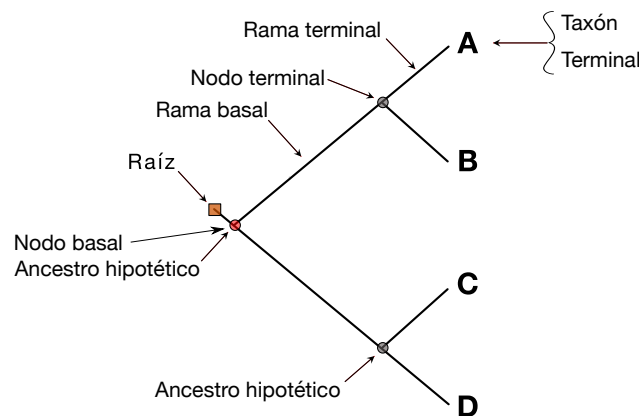
El enfoque macroevolutivo utiliza caracteres morfológicos y/o moleculares, estos últimos, aunque pueden ser de diversa índole, usualmente son secuencias de ADN de regiones específicas con grados de variación proporcionales a la profundidad con la que se quiera hacer el estudio (Wiley & Lieberman, 2011). En plantas, por ejemplo, para investigar grupos con divergencias profundas se han utilizado secuencias de regiones con un alto

grado de conservación, como *rbcL*, pero para estudiar grupos de reciente divergencia se ocupan secuencias de regiones altamente variables como ITS o ETS (Soltis & Soltis, 1999).

Los estudios macroevolutivos emplean herramientas estadísticas de diversa naturaleza para tratar de elucidar el grado de relación que guardan dos entidades (usualmente especies) entre sí. Comúnmente expresan la información generada en forma de filogenias, que son diagramas en forma de árboles en los que se presentan las relaciones hipotéticas entre un conjunto de ancestros y sus descendientes (Figura 1). Las filogenias idealmente son árboles dicotómicos en los que la raíz representa el origen o ancestro común a todas las entidades incluidas, cada punto de bifurcación (nodo) se interpreta como un ancestro e implica la aparición de nuevos linajes. La línea que conecta dos nodos se conoce como rama y la parte final en donde se encuentran las entidades incluidas se llama ramas terminales o simplemente terminales. El enfoque dicotómico tradicional presupone que, una vez formados los linajes, ya nunca más intercambian información genética entre sí (Knowles & Carstens, 2007). Aunque esta es la interpretación tradicional, en la literatura hay propuestas discordantes con tal visión y por ello se han planteado formas diferentes de análisis, de las cuales algunas son análogas a las redes con las que se interpretan las relaciones entre las poblaciones (Huson, Rupp & Scornavacca, 2010).

Figura 1.

Una filogenia hipotética y sus partes. En la figura se representa un cladograma perfectamente dicotómico con todas las ramas con la misma longitud. Note que tiene una raíz que marca el origen de la filogenia, el cual representa un ancestro hipotético al igual que cada nodo. Están las ramas que van conectando los nodos de la filogenia con los ancestros hipotéticos hasta que se llega a las ramas terminales que representan a los taxones incluidos en el muestreo.

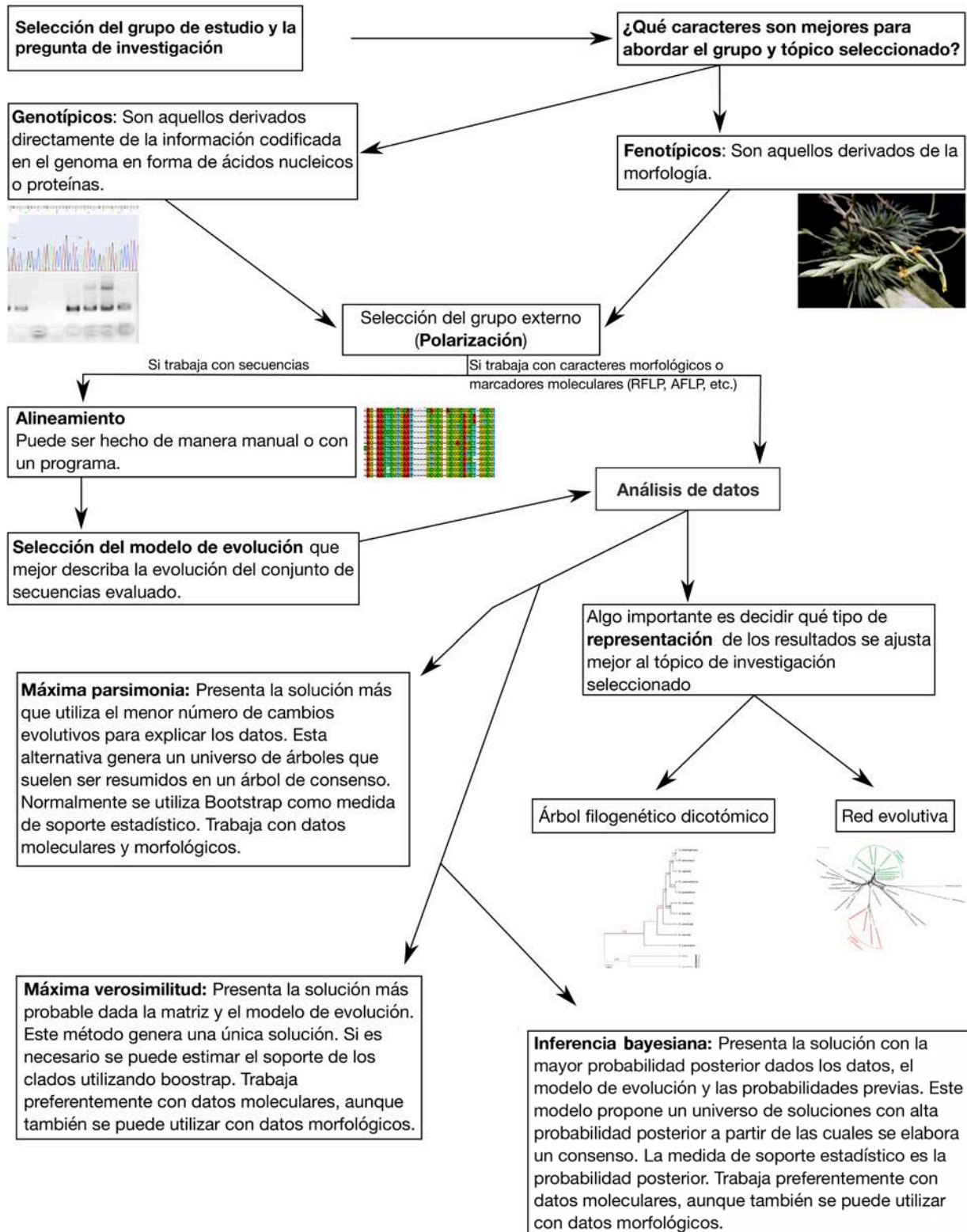


Es importante recordar que las filogenias expresan una hipótesis de relaciones evolutivas entre linajes y por lo mismo no deben ser confundidas con los dendrogramas o fenogramas. Estos últimos son diagramas dicotómicos que muestran similitud entre entidades en función de algo, por ejemplo, morfología (fenogramas) o en ecología se utilizan para comparar qué tanto se parecen dos comunidades (análisis de conglomerados).

Dado que esta contribución versa sobre el uso de métodos filogenéticos en un contexto tradicional, a continuación, se presentarán los elementos teóricos relacionados con una reconstrucción filogenética utilizando caracteres moleculares y/o morfológicos desde una perspectiva amplia, como se resume en la Figura 2. En una segunda contribución se aborda la utilidad de las herramientas filogenéticas para estudiar fenómenos evolutivos (evolución de caracteres, por mencionar alguno).

Figura 2.

Diagrama de flujo de una reconstrucción filogenética. Para detalles, consulte el documento.



La reconstrucción de una filogenia

La reconstrucción de una filogenia es un proceso cuyo primer paso requiere definir el grupo de estudio y su circunscripción. Tradicionalmente esto se hacía con criterios empíricos, sin un enfoque evolutivo y usualmente con el propósito de poner a prueba determinado esquema de clasificación. Actualmente, con base en la información que existe en la literatura, se hacen cada vez con más frecuencia propuestas menos conservadoras, con la intención de responder preguntas concretas que pueden ser prácticamente de cualquier índole imaginable (p. e. Givnish *et al.*, 2011; Smith & Hendricks, 2013).

Para elaborar una reconstrucción filogenética luego de la selección del grupo de estudio es necesario precisar el grupo externo. Este punto es esencial, pues se ocupa para definir la polaridad de la filogenia y por lo mismo es la base para establecer qué es lo “ancestral” y lo “derivado”. Lo que sigue es la elección del tipo de caracteres, que pueden ser de las dos fuentes más comunes, a las que por simplicidad se denomina aquí caracteres de origen fenotípico y los de origen genotípico. Lo anterior es relevante porque no es lo mismo utilizar caracteres fenotípicos o genotípicos codificados como una matriz binaria o multiestado, que utilizar secuencias de ADN, cuyo abanico de opciones de análisis es mucho mayor.

En el caso de las secuencias de ADN, el primer paso es establecer un alineamiento en el cual se maximice la comparabilidad (y compatibilidad) entre las secuencias que conforman la muestra, luego se debe especificar el modelo evolutivo que se utilizará para analizar la matriz de secuencias alineadas (ver sección de modelos más adelante). Posteriormente se hace el análisis, que debe estar guiado por la pregunta que se desea responder, ya que cada uno de los paradigmas para el análisis de datos (máxima parsimonia, máxima verosimilitud³ e inferencia bayesiana) presenta fortalezas y debilidades. Además, los productos de cada una de estas formas de análisis tienen propiedades diferentes, por lo que pueden ser usados para actividades tan distintas como establecer el grado de discrepancia de los datos utilizados, la tasa de evolución de un linaje, elaborar una escala temporal o entender cómo ha sido el proceso de cambio de un atributo.

Caracteres fenotípicos

Los caracteres fenotípicos son atributos que podemos observar o medir de la información contenida en el genoma de los organismos a analizar; pueden ser desde la presencia de determinado tipo de indumento, cambios en la forma, los tamaños, las proporciones, los ángulos, los colores, etcétera (Poe & Wiens, 2000). En los caracteres fenotípicos se pueden incluir también elementos como las fragancias, las características fisiológicas, los compuestos químicos, entre otros. Por supuesto, el uso de características con una homología discutible como las fragancias o los colores requiere de una justificación basada en investigación previa. Un carácter homólogo es aquel que tiene el mismo origen entre los taxa incluidos en el estudio (Herron & Freeman, 2014).

En este sentido, un carácter puede definirse como cualquier atributo que presenta variación entre dos o más grupos de organismos, rasgos que para ser empleados en los análisis filogenéticos deben ser constantes (Wagner, 2001). Este concepto es apropiado para describir cualquier actividad de la vida cotidiana y es una parte esencial de la biología comparada, que tiene por función examinar y capturar los patrones biológicos, y elaborar teorías sobre los procesos que podrían explicarlos (Eldredge & Cracraft, 1980).

Utilizar caracteres fenotípicos normalmente ofrece dos restricciones que deben salvarse; la primera es determinar, más allá de la duda razonable, si el atributo seleccionado cumple con el criterio de homología. La

³ También se conoce como máxima probabilidad.

segunda tiene que ver con la definición precisa (sin ambigüedades) de los estados de carácter; esto es, de las diferentes formas en las que puede presentarse ese atributo (Wagner, 2001; Poe & Wiens, 2000). En muchos casos, los rasgos elegidos marcan discontinuidades muy claras, por ejemplo “verticilos trímeros” y “verticilos tetra- o pentámeros”, el primero es un carácter que define a las monocotiledóneas, mientras que el segundo distingue a las eudicotiledóneas (Simpson, 2019).

Los caracteres fenotípicos pueden ser utilizados de varias maneras: (i) codificando de forma discreta a aquellos rasgos que podrían resultar informativos (Wagner, 2001; Poe & Wiens, 2000); (ii) empleando caracteres continuos discretizados a través de alguno de los métodos propuestos a lo largo de los últimos 30 años (para una revisión, consultar García-Cruz & Sosa, 2006); (iii) haciendo uso de caracteres continuos de forma directa; a diferencia de los enfoques anteriores, éstos pueden ser analizados mediante la parsimonia cuadrática propuesta por Goloboff, Mattoni & Quinteros (2006) o con inferencia bayesiana modelados bajo difusión browniana (*brownian Diffusion*), o una de las variantes del modelo de procesos estocásticos conocido en inglés como *Random Walks* (Lemey, Rambaut, Welch & Suchard, 2010).

Caracteres genotípicos

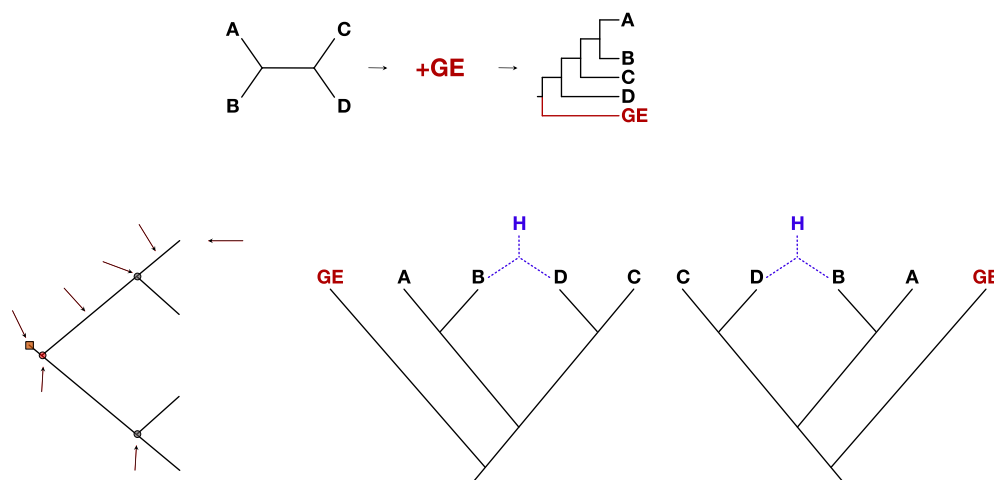
Los caracteres genotípicos se derivan del estudio directo de la información contenida en el genoma; esto puede ser a través de marcadores moleculares como los RAPD, AFLP, SSR, entre otros, así como del estudio directo de las secuencias de ADN. En la actualidad la fuente predilecta de información para construir filogenias es el uso de secuencias de ADN, que puede ser de regiones específicas del núcleo, cloroplasto o mitocondrias o una combinación de éstas, según la profundidad de las relaciones o grupo de organismos a estudiar (p. e. Jiang, Zhu, Song, Li, Yang & Yu, 2014; Leopardi-Verde, Carnevali & Romero-González, 2017). La elección de las regiones a usar normalmente está relacionada con la tasa de evolución; por ejemplo, en plantas la tasa de evolución del genoma del cloroplasto es mucho mayor que la de las mitocondrias y por lo mismo es más informativo (Soltis & Soltis 1999). En contraste, en los animales hay genes mitocondriales que son lo suficientemente variables como para que puedan ser utilizados como códigos de barras (p. e. Prado, Pozo, Valdez-Moreno & Hebert, 2011). Cabe mencionar que en los estudios más recientes ya se empiezan a ocupar cantidades masivas de datos, al punto que en algunos casos ya se emplean conjuntos genómicos completos (Kim *et al.*, 2020; Koenen *et al.*, 2020).

Polarización

Se dice que dos o más estados de un carácter están polarizados cuando se establece, bajo algún criterio, cuál fue el primero en evolucionar (Wiley & Lieberman, 2011); la misma idea aplica al escoger el grupo externo en una reconstrucción filogenética. La polarización es fundamental para el análisis filogenético, pues a través de ella es posible dar dirección a un árbol filogenético (Figura 3). Una consecuencia de esto es que se identifican las sinapomorfías que diagnostican a los grupos monofiléticos que emergen de un estudio. Antes de continuar, es importante aclarar que una apomorfía es un carácter único de un grupo que se considera una sinapomorfía cuando es compartido por todos los miembros de un grupo evolutivamente derivado. Mientras que si es compartido por todos los miembros del grupo analizado se le denomina simplesiomorfía, por ejemplo, el pelo es un carácter sinapomórfico de los mamíferos si se ve en el contexto de la filogenia de los vertebrados; sin embargo, si nos enfocamos exclusivamente en la filogenia de los mamíferos este carácter es considerado simplesiomórfico, porque todos lo comparten.

Figura 3.

El efecto de polarizar. De izquierda a derecha se muestra un árbol no enraizado, al que se añade un grupo externo (GE) para dar dirección a la filogenia.



Se han descrito muchos métodos para la polarización, no obstante, el más ampliamente aceptado es la comparación con un grupo externo, por dos razones: la primera, porque sus supuestos son confiables y sencillos; la segunda es porque la información que se requiere para su aplicación en la mayor parte de los casos está disponible y se recolecta al mismo tiempo que para el grupo de organismos de interés (Bryan, 2001). La comparación con un grupo externo establece la ancestría identificando el estado compartido con el taxón o taxa designados para este fin.

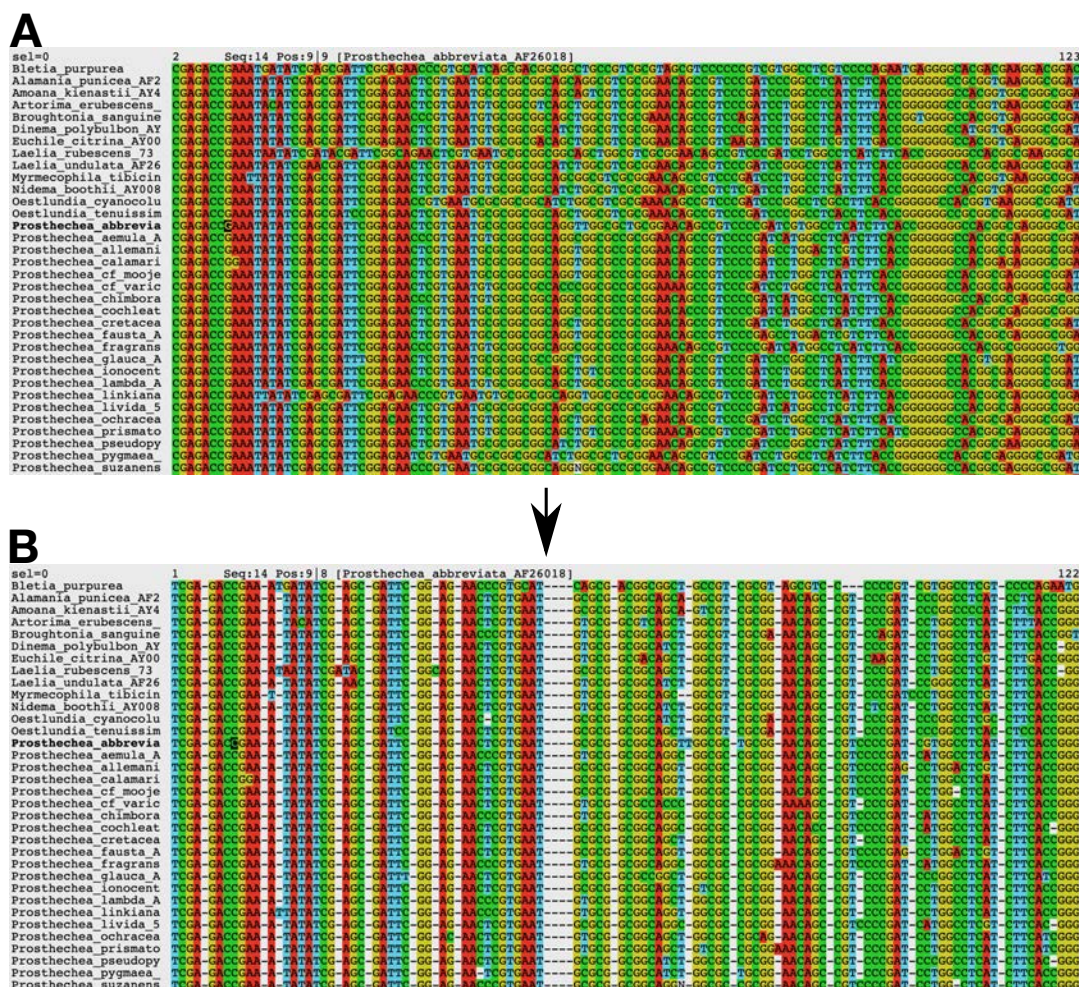
El método ontogénico se basa en la comparación de los patrones de desarrollo de las especies y establece la ancestría de un estado de carácter utilizando como criterio el orden en que los distintos estados de ese carácter aparecen en la ontogenia de los organismos. A pesar de que este método se basa en evidencia primaria, debido a lo difícil que es hacer estudios de patrones de desarrollo para muchos organismos y a que existen diversas corrientes de interpretación para esos resultados, se considera que es un criterio poco práctico (Harvey & Pagel, 1991; Bryan, 2001). Un tercer método es el del uso del registro fósil, que establece que los estados ancestrales son aquellos que aparecen primero en los fósiles. Debido a que la evidencia fósil es fragmentaria, este criterio ha caído en desuso (Bryan, 2001). Cabe notar que el único criterio válido para la polarización de una filogenia, cuando se trabaja usando como información base secuencias de ADN, es la selección del grupo externo; los otros dos son útiles cuando se intenta polarizar sólo caracteres morfológicos.

Métodos para la reconstrucción de una filogenia

Alineamiento. Consiste en organizar las secuencias (ADN, ARN o proteínas) considerando cada posición comparable de la matriz (p.e. la posición 355 del alineamiento) como una hipótesis de homología posicional (Figura 4). Usualmente estas hipótesis deben ser lo más parsimoniosas posible (se prefieren las opciones más simples sobre las más complejas) y persiguen inferir las relaciones evolutivas sin necesidad de un conocimiento previo de los eventos evolutivos que dieron origen a esos cambios (Morrison, 2006)

Figura 4.

Ejemplo de un alineamiento. A) Segmento de una matriz de ITS sin alinear. B) Segmento de la matriz de ITS luego del alineamiento.



Los alineamientos pueden hacerse manualmente o con ayuda de programas de cómputo. Los manuales presentan el problema de ser difíciles de repetir; mientras que los elaborados con programas, aunque pueden ser reproducidos con facilidad, usualmente requieren ser refinados a mano. Los programas pueden seguir tres métodos básicos para alinear las secuencias: global, por bloques o combinados; también pueden ser progresivos, iterativos o ser una mezcla de ambos (Mount, 2001; Pivorano & Heringa, 2008).

Los métodos globales utilizan las secuencias completas durante el alineamiento (p. e. ClustalW y MUSCLE). Por su parte, los que funcionan por bloques identifican patrones en las secuencias y los usan para elaborar el alineamiento (p. e. MAFFT). Sólo los métodos más recientes, como T-Coffee, son capaces de incorporar estas dos metodologías (Mount, 2001; Pivorano & Heringa, 2008).

Los métodos progresivos intentan construir el alineamiento generando un “boceto” de árbol filogenético, para lo que usan una distancia genética y algoritmos como el método de grupo de pares no ponderados con

media aritmética (UPGMA); este boceto es empleado para ir añadiendo una a una las secuencias a ser alineadas. Por conveniencia, los algoritmos alinean las secuencias por pares. Por ello, cuando se utiliza un alineamiento progresivo, es de extrema importancia que las dos primeras secuencias se alineen correctamente, ya que su resultado afecta al resto del proceso (Mount, 2001). En los alineamientos progresivos, una vez que se coloca un gap, en esa posición siempre habrá un gap (Mount, 2001); ClustalW es un método que funciona de esta manera. Los gaps también se conocen como indels y son llamados así porque representan una inserción o una delección en el genoma, reciben este nombre porque no es posible *a priori* saber de qué se trata.

Los métodos iterativos se diferencian de los progresivos en que repiten tantas veces como sea necesario el alineamiento de las secuencias, presentando al final una secuencia de consenso. En la actualidad, es común el uso de algoritmos que combinan ambas aproximaciones (son globales e iterativos), como MUSCLE y T-Coffee (Pivorano & Heringa, 2008).

Recurrir a uno u otro tipo de algoritmos es una decisión multifactorial. Edgar & Batzoglou (2006) hacen una serie de recomendaciones al respecto, por ejemplo, si se es muy tradicional, la opción lógica es ClustalW que da resultados bastante buenos, aunque es un algoritmo lento y menos preciso que los métodos más modernos. Si se manejan 500-1000 secuencias en la matriz que será alineada, MUSCLE es más eficiente que ClustalW en el uso del tiempo y su carácter iterativo disminuye sustancialmente los errores de identidad que pueden ocurrir en los alineamientos de ClustalW. Otra opción muy buena cuando se tienen más de 500 secuencias o secuencias con patrones muy complejos es MAFFT. T-Coffee produce alineamientos muy buenos, pero es lento, demanda amplios recursos del sistema y no puede ser utilizado con más de 50 secuencias a la vez.

Para hacer los alineamientos hay muchas alternativas, desde utilizar los ejecutables de los algoritmos directamente, hasta emplear programas que los integran, como MEGA, Seaview, Geneious, entre otros. Otra opción es recurrir a servidores dedicados a alineamientos. En el sitio web del libro de Lemey, Salemi & Vandamme (2009)⁴ existe una lista de algoritmos disponibles e indica dónde pueden ser utilizados o descargados. Adicional a esto, el servidor del Laboratorio Europeo de Biología Molecular (EMBL-EBI) pone a disposición del público, libre y gratuitamente, algunos de los algoritmos más populares.⁵

Selección del modelo de evolución. Para reconstruir una hipótesis de relaciones filogenéticas es indispensable contar con un modelo que describa cómo las secuencias de ADN o proteínas evolucionan a través de mutaciones, deriva génica, selección, recombinación, etcétera (Huson *et al.*, 2010). En palabras más formales, un modelo de evolución es un conjunto de parámetros que se utilizan para describir el patrón de sustituciones que caracterizan a una matriz de secuencias de ADN o proteínas (Strimmer & von Haeseler, 2009). En este sentido, cuando se trabaja con secuencias, la selección del modelo es un paso crítico, pues tanto la subparametrización como la sobreparametrización pueden alterar los resultados del análisis (Lemmon & Moriarty, 2004).

Actualmente existen más de 88 modelos que reflejan distintos grados de complejidad en la evolución del ADN (Posada, 2008); entre ellos, los más importantes en orden de complejidad son Jukes y Cantor (JC), Felsenstein 81 (F81), Kimura dos parámetros (K2P), Hasegawa-Kishino-Yano (HKY), Kimura tres parámetros (K3P), Tamura-Nei (TrN), Simétrico (SYM) y el modelo general de tiempo reversible (GTR, por "*general time reversible*"). La diferencia entre cada uno de estos modelos es el número y tipo de parámetros usados; por ejemplo, JC sólo considera la tasa de sustitución, mientras que F81 añade a esto la proporción en la que están las bases y así se van incrementando los parámetros hasta llegar al más complejo, el GTR, que toma en cuenta todos los parámetros posibles como

⁴ <https://www.kuleuven.be/aidslab/phylogenybook/Table3.1.html>

⁵ <https://www.ebi.ac.uk/Tools/msa/>

sustituciones, frecuencias de las bases, transiciones, entre otros, y que puede hacerse más complejo al incluir la probabilidad de sitios invariables (I) y Gamma (γ); este último es una medida de la heterogeneidad entre los sitios que componen el alineamiento (Huson *et al.*, 2010).

Así, para un principiante quizás lo más práctico es dejar que un programa de computadora analice la matriz de datos y le sugiera tanto el modelo más apropiado, como los parámetros que mejor se ajustan a los datos; mientras que para un experto esto es una ayuda importante en el proceso, aunque pueden ir más allá si así lo desean. Para la selección de modelos hay varios programas, aunque uno de los más populares es Modeltest o su actualización Modeltest-NG (Posada, 2008; Strimmer y von Haeseler, 2009; Darriba *et al.*, 2019). Si se desea aprender más sobre los modelos de evolución, un par de lecturas excelentes son el capítulo 3 de Huson *et al.* (2010) y Strimmer & von Haeseler (2009).

Estrategias de análisis. La máxima parsimonia (MP) es un método que busca el árbol más corto que mejor explica los datos observados (Whelan, 2008; Swofford & Sullivan, 2009). La idea de esta búsqueda se basa en el principio conocido como la navaja de Ockham (*Occam's razor*), que sugiere que cuando hay dos o más hipótesis que proveen una respuesta igualmente probable, la más simple suele ser la mejor. En este sentido, en la MP el árbol más simple es aquel que minimiza la cantidad de cambios evolutivos que se requieren para explicar los datos (Swofford & Sullivan, 2009). Minimizar la cantidad de cambios evolutivos implica seleccionar el árbol con menor número de homoplasias, que son aquellos caracteres que se adquieren de forma independiente por grupos no relacionados desde el punto de vista evolutivo, este término también se conoce como evolución paralela (Sanderson & Hufford, 1996, Swofford & Sullivan, 2009). Ejemplo de caracteres homoplásicos son la partenogénesis en reptiles o la forma esférica de algunas *Cactaceae* y *Euphorbia* obesa (*Euphorbiaceae*).

En general, la MP tiene la ventaja de ser bastante intuitiva y de requerir pocos recursos computacionales en comparación con métodos estadísticos más demandantes; sin embargo, su problema es que si las ramas son muy largas, puede sugerir relaciones incorrectas con un soporte elevado.

Para buscar el árbol más corto, la parsimonia puede intentar generar todos los árboles, lo cual se conoce como búsqueda exhaustiva (es viable sólo hasta 12 terminales) o puede muestrear la población de árboles a través de una búsqueda heurística tradicional o utilizando el algoritmo Ratchet (Nixon, 1999). Independiente de cómo se establezca la búsqueda heurística, el programa utilizado muestreará la población de árboles a través de algoritmos como el "intercambio con el vecino más cercano" (NNI), el "cortado y redibujado de árboles" (SPR) o la "bisección y reconexión de árboles" (TBR), que respectivamente representan de la búsqueda más simple a la más completa (Swofford & Sullivan, 2009).

En MP el Bootstrap es empleado con frecuencia para evaluar el soporte estadístico de los árboles. Este análisis es una técnica estadística de remuestreo de datos con réplica, en la que la mitad del conjunto de caracteres es eliminado, los datos que permanecen son duplicados de tal manera que el grupo de caracteres es del mismo tamaño que al inicio; a partir de éste se realiza el análisis filogenético y se obtiene la filogenia más parsimoniosa. El proceso descrito se repite tantas veces como sea necesario, lo usual es al menos mil ocasiones y luego se estima a partir de todos los árboles generados con el conjunto de datos parcialmente simulados la proporción en la que se presenta cada clado, incluido en la filogenia basada en datos reales (Schmidt, 2009). Algunos programas populares para construir árboles utilizando máxima parsimonia son PAUP* (Swofford, 2002), Phylip (Felsenstein, 1989), TNT (Goloboff, Farris & Nixon, 2008), Nona (Goloboff, 1999) y WinClada (Nixon, 2002).

El método de máxima verosimilitud (ML, por sus siglas en inglés) toma un modelo de evolución de secuencias de ADN o proteínas y busca la combinación de valores de los parámetros que componen el modelo que genere el

árbol filogenético capaz de representar con la mayor probabilidad las secuencias que integran la matriz utilizada. En otras palabras, es la probabilidad de los datos dado un árbol (Lewis, 1998; Bromham & Penny, 2003).

Algo interesante de la ML es que entre los factores que se toman en cuenta al estimar un árbol filogenético están la longitud de las ramas y la topología. Por ello es posible que un árbol tenga una topología "correcta", pero si su longitud de ramas está mal calculada, no necesariamente tendrá una buena probabilidad. La precisión en la estimación de la longitud de ramas convierte a este método en una herramienta indispensable cuando se quiere evaluar hipótesis de evolución de caracteres o estimar tasas de diversificación. Algunos programas populares para construir árboles utilizando ML son RAxML-NG (Kozlov, Darriba, Flouri, Morel & Stamatakis, 2019), PhyML (Guindon *et al.*, 2010), entre otros. Para leer más al respecto, una introducción excelente es Lewis (1998).

La inferencia bayesiana (IB) es un método que, con base en un modelo específico de sustitución, dado los datos y las probabilidades previas, selecciona el árbol que tiene la mayor probabilidad posterior (Bromham & Penny, 2003). Este método utiliza modelos evolutivos acoplados a Cadenas de Markov Monte Carlo y a otros algoritmos como Metrópolis para poder hacer sus estimaciones (Ronquist, van den Mark & Huelsenbeck, 2009).

La IB funciona haciendo un muestreo a través de todas las poblaciones de árboles que se "cruzan" en el camino de la cadena de Markov. Generalmente el análisis se deja correr por un número de generaciones que sea suficientemente grande, como para que las cadenas alcancen la estabilidad y converjan. Se recomienda hacer varias corridas, a fin de asegurarse de que las cadenas realmente convergieron y que se muestrearon todas las poblaciones de árboles (Ronquist, van den Mark & Huelsenbeck, 2009). En este tipo de análisis es conveniente tener en cuenta que los árboles son escogidos conforme a una función de probabilidad y que todos están conectados (Whelan, 2008).

La probabilidad posterior (PP) puede considerarse como la probabilidad, valga la redundancia, de que un nodo de la filogenia obtenida bajo un determinado modelo de evolución sea cierto. De hecho, estudios han demostrado que estos valores pueden interpretarse directamente (Alfaro, Zoller & Lutzoni, 2003), por lo que una probabilidad posterior de 0.95 equivale a 95% de probabilidad de que sea cierto. El punto débil de la IB y sus probabilidades posteriores es la sensibilidad a parámetros como α (que indica la forma de la distribución) y a otros componentes intrínsecos del modelo, entre los que se encuentran la probabilidad de transversión/inversión (κ) y sitios invariables (i). De hecho, una sobreparametrización del modelo puede "inflar" los valores de las PP, aunque esto no es tan grave como la subparametrización, que sí puede alterar la topología y ocasionar un alto soporte a clados falsos (Alfaro *et al.*, 2003; Ronquist *et al.*, 2009). Algunos programas populares para construir árboles utilizando IB son MrBayes (Ronquist *et al.*, 2012), BEAST (Bouckaert *et al.*, 2019), entre otros.

Redes filogenéticas (*Phylogenetic networks*)

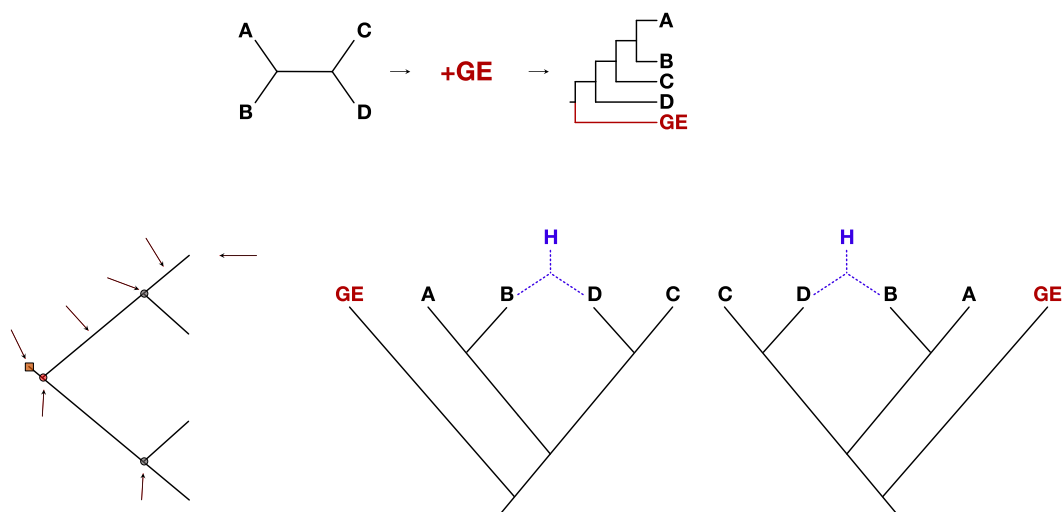
En la mayor parte de los trabajos de sistemática que involucran el uso de filogenias, la representación de las relaciones ancestro-descendiente se hace a través de dicotomías (Huson *et al.*, 2010). Este tipo de representación tiene una suposición subyacente: cada vez que ocurre un proceso de cladogénesis en un linaje, éste se divide sólo en dos nuevos linajes que nunca más interaccionan (Figura 1); no obstante, existe evidencia de que linajes diferentes pueden interactuar (Levin, 2004; Mallet, 2005; Via & West, 2008). Un ejemplo claro de este tipo patrón evolutivo son los híbridos (Soltis & Soltis, 2009; Russell *et al.*, 2010). Se han documentado procesos de recombinación o transferencia horizontal de genes entre especies (Richardson & Palmer, 2006; Czisłowski *et al.*, 2018) y también hay casos en que las relaciones son mucho más complicadas que una dicotomía, como las radiaciones evolutivas que caracterizan a virus, como al de Inmunodeficiencia Humana (VIH) o al de la influenza

(Chan, Carlsson & Rabadan, 2013). En los anteriores y otros similares las relaciones reconstruidas que usan el enfoque dicotómico aportan poca o ninguna información sobre los procesos evolutivos subyacentes al grupo de estudio.

Por ello, cuando se sospecha que los procesos evolutivos no siguen un patrón dicotómico, es conveniente explorar los juegos de evidencia con herramientas que permitan el análisis de las reticulaciones y su representación a través de diagramas multifurcados, conocidos como redes filogenéticas (Figura 5). Para diferenciar una filogenia estándar de una red evolutiva, lo primero que debe quedar claro es que un árbol filogenético es un diagrama en el que se representan las relaciones ancestro-descendiente entre los taxa, a través de una serie de líneas (ramas) y conexiones (nodos) en los que se puede sugerir (árbol enraizado) o no (árbol no enraizado) una dirección en los procesos evolutivos (Figuras 1 y 3). Una característica importante de este diagrama es que no se forman ciclos y es estrictamente bifurcado cuando está completamente resuelto. Una red filogenética, en cambio, es cualquier gráfico utilizado para representar las relaciones evolutivas entre un grupo de organismos (sea abstracta o explícitamente), en el que se da nombre a algunos nodos del diagrama mientras que otros funcionan como conectores (*edges*) (Huson *et al.*, 2010). Dependiendo de si la hipótesis es explícita o abstracta, los nodos pueden representar entidades ancestrales o no.

Figura 5.

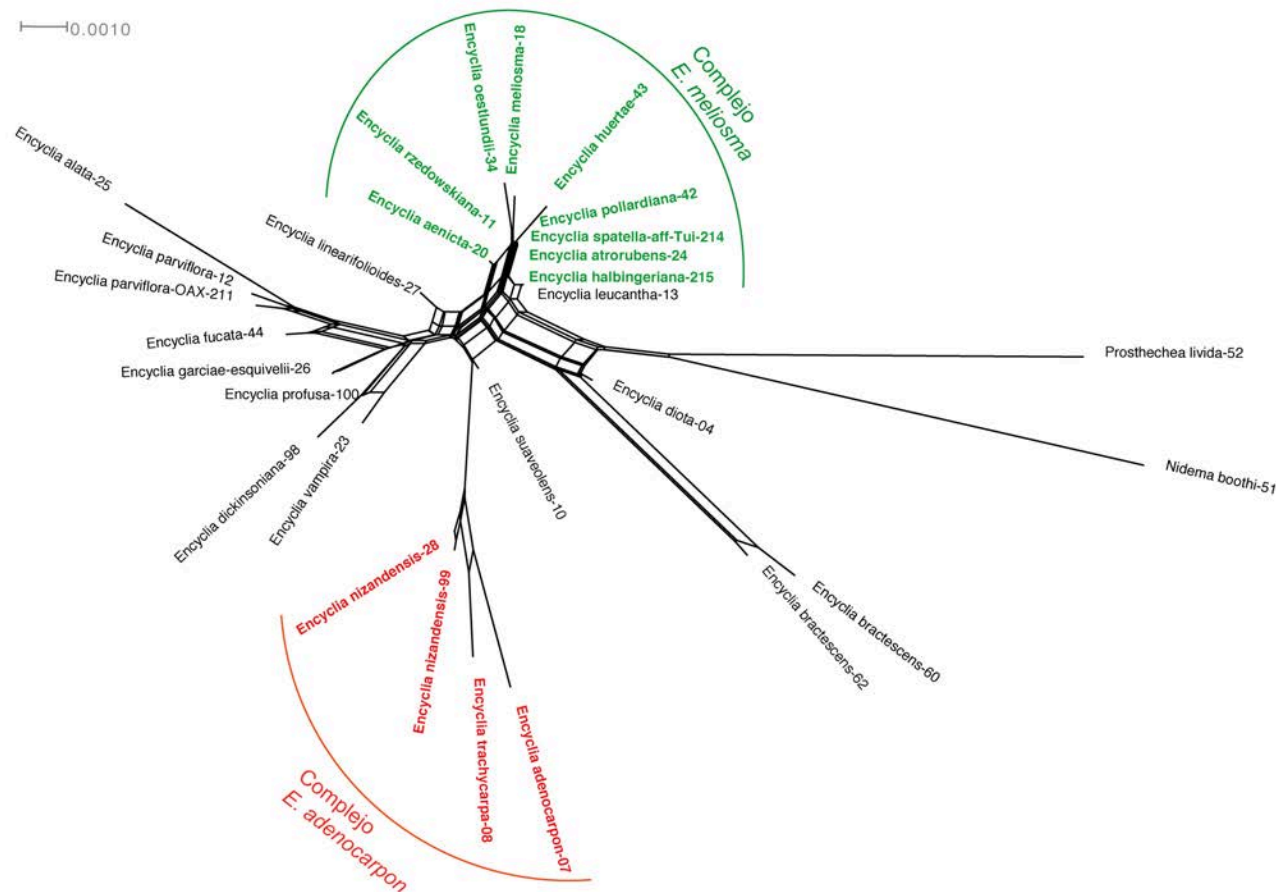
Filogenia hipotética en la que se muestra la posición de un taxón híbrido (H).



Las redes evolutivas pueden construirse utilizando como bloques de datos secuencias, distancias o árboles; asimismo, es posible procesar los bloques de datos empleando cualquier modelo evolutivo o paradigma de análisis (MP, ML, IB) y pueden ser enraizadas o no. Uno de los métodos más usados para elaborar redes es el conocido como redes divisivas (*splits networks*) (Figura 6). Para el caso de los esquemas enraizados, éstos pueden ser contruidos utilizando varias herramientas, una de las más interesantes son las filogenias reticuladas, llamadas en inglés *galled networks*, aunque hay redes enraizadas para casos específicos como las redes de hibridación (*hybridization networks*), de recombinación (*recombination networks*), entre otras (Huson *et al.*, 2010).

Figura 6.

Ejemplo de una red filogenética tipo *split network*. En la figura se aprecian las relaciones entre especies del género *Encyclia* (Orchidaceae: Laeliinae) basadas en el gen *PhyC*. Figura elaborada utilizando datos propios.



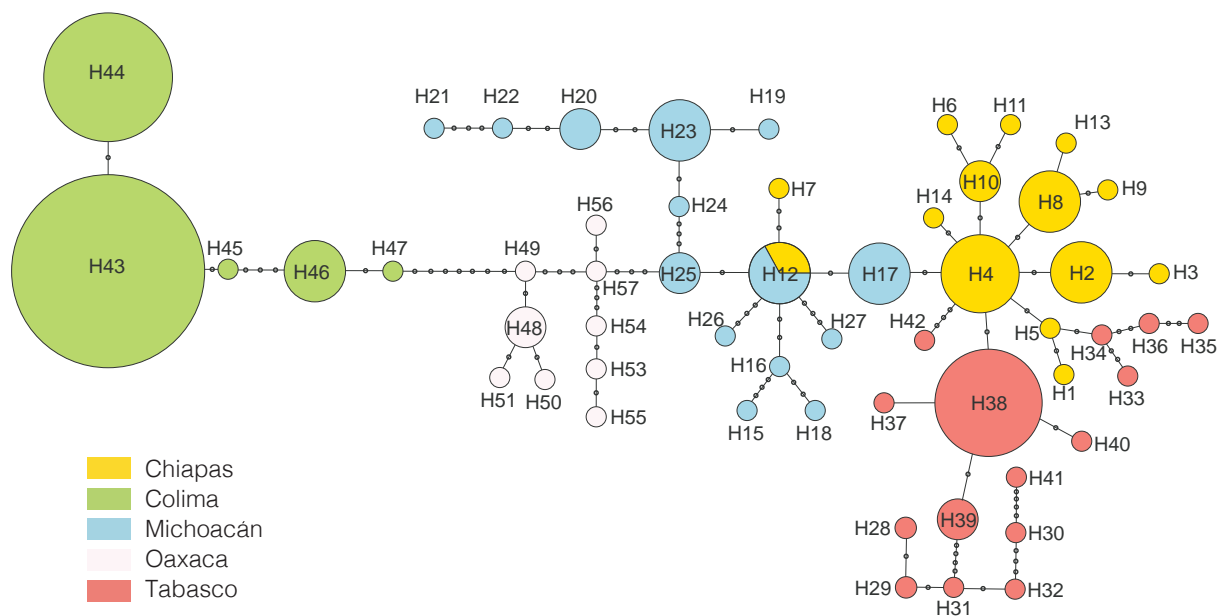
Las redes filogenéticas son especialmente útiles cuando se trata de entender historias evolutivas complejas en las que hay un alto intercambio de información entre linajes, por ejemplo, Stavrínides & Guttman (2004) estudiaron la evolución del coronavirus causante del Síndrome Respiratorio Agudo (SARS) en varios linajes de aves y mamíferos, con el fin de dilucidar su origen. La información base que utilizaron fueron secuencias de los genes que codifican para la replicasa, las espinas superficiales de la cápside, la matriz y las proteínas de la nucleocápside. Los análisis muestran que los linajes de estos virus relacionados con aves y mamíferos interactúan generando reticulaciones que podrían deberse a recombinación entre éstos. En un trabajo relacionado, Forster, Forster, Renfrew & Forster (2020) utilizaron redes filogenéticas para analizar el genoma del SARS-CoV-2 y encontraron que la forma inicial de la red es consistente con el primer patrón de contagio, pero se hace difuso a medida que se acumulan efectos fundadores y el virus va mutando.

El enfoque de redes evolutivas también es útil por debajo del nivel de especie, en donde los linajes suelen intercambiar información genética. Por ejemplo, el uso de redes en forma de árboles de expansión mínima permite

comprender el flujo de genes entre poblaciones de una misma especie. Estas herramientas no distinguen entre grupos, por lo que pueden servir lo mismo para elucidar los patrones de domesticación de cerdos (Wu *et al.*, 2007) que para comprender la diversidad genética de un hongo fitopatógeno (Figura 7) (Manzo-Sánchez *et al.*, 2019).

Figura 7.

Una red de expansión mínima de los 57 haplotipos encontrados en cinco poblaciones de *Pseudocercospora fijiensis* (M. Morelet) Deighton. Los círculos negros pequeños representan los haplotipos faltantes, los círculos coloreados de diferente tamaño son haplotipos; los colores indican las poblaciones a las que pertenecen y el tamaño representa la frecuencia del haplotipo. Figura reproducida de Manzo-Sánchez *et al.*, 2019.



Conclusión

El estudio de patrones evolutivos es un área fascinante de la biología comparada. Las herramientas disponibles son variadas y permiten utilizar casi cualquier fuente de datos disponibles, desde la morfología hasta datos genómicos. Con esta variedad de datos también han ido cambiando las herramientas, iniciando por la máxima parsimonia, máxima verosimilitud, hasta los métodos que usan inferencia bayesiana y que son los más recientes. Cada una de estas formas de análisis tiene sus fortalezas y debilidades. De igual manera, se han flexibilizado las opciones de representar las relaciones y ya no sólo se piensa en dicotomías, sino que hay maneras de ver patrones más complejos de evolución como las reticulaciones. Sin embargo, los usos de las filogenias son muy diversos, como las estimaciones de edad que permiten los relojes moleculares o las comparaciones ecológicas en un contexto evolutivo.

Referencias

- Alfaro, M. E., Zoller, S. & Lutzoni, F. (2003). Bayes or bootstrap? a simulation study comparing the performance of Bayesian Markov Chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Molecular Biology and Evolution*, 20(2), 255-266. doi: 10.1093/molbev/msg028
- Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., ... Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology*, 15(4), e1006650. doi: 10.1371/journal.pcbi.1006650
- Bromham, L. & Penny, D. (2003). The modern molecular clock. *Nature Reviews Genetics*, 4(3), 216-224. doi: 10.1038/nrg1020
- Bryan, H. N. (2001). Character polarity and the rooting of cladograms. En G. P. Wagner (Ed.), *The Character Concept in Evolutionary Biology* (319-341). San Diego, Estados Unidos: Academic Press.
- Chan, J. M., Carlsson, G. & Rabadan, R. (2013). Topology of viral evolution. *Proceedings of the National Academy of Sciences*, 110(46), 18566-18571. doi: 10.1073/pnas.1313480110
- Czislowski, E., Fraser-Smith, S., Zander, M., O'Neill, W. T., Meldrum, R. A., Tran-Nguyen, L. T. T., Batley, J. & Aitken, E. A. B. (2018). Investigation of the diversity of effector genes in the banana pathogen, *Fusarium oxysporum* f. sp. *cubense*, reveals evidence of horizontal gene transfer. *Molecular Plant Pathology*, 19(5), 1155-1171. doi: 10.1111/mpp.12594
- Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B. & Flouri, T. (2019). ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. *Molecular Biology and Evolution*, 37(1), 291-294. doi: 10.1101/612903
- Edgar, R. C. & Batzoglou, S. (2006). Multiple sequence alignment. *Current Opinion in Structural Biology* 16 (3), 368-373. DOI: 10.1016/j.sbi.2006.04.004
- Eldredge, N. & Cracraft, J. (1980). *Phylogenetic patterns and the evolutionary process, methods and theory in comparative biology*. New York: Columbia University Press.
- Felsenstein, J. (1989). PHYLIP-Phylogeny Inference Package (Version 3.2). *Cladistics* 5(2), 164-166. doi: 10.1111/j.1096-0031.1989.tb00562.x
- Forster, P., Forster, L., Renfrew, C. & Forster, M. (2020). Phylogenetic network analysis of SARS-CoV-2 genomes. *Proceedings of the National Academy of Sciences*, 117(17), 9241-9243. doi: 10.1073/pnas.2004999117
- García-Cruz, J. & Sosa, V. (2006). Coding quantitative character data for phylogenetic analysis: A comparison of five methods. *Systematic Botany*, 31(2), 302-309. doi: 10.1600/036364406777585739
- Givnish, T. J., Barfuss, M. H. J., Ee, B. V., Riina, R., Schulte, K., Horres, R., Gonsiska, P. A., Jabaily, R. S., Crayn, D. M., Smith, A. C., Winter, K., Brown, G. K., Evans, T. M., Holst, B. K., Luther, H., Till, W., Zizka, G., Berry, P. E. & Sytsma, K. J. (2011). Phylogeny, adaptive radiation, and historical biogeography in *Bromeliaceae*: Insights from an eight-locus plastid phylogeny. *American Journal of Botany*, 98(5), 872-895. doi: 10.3732/ajb.1000059

- Goloboff, P. (1999). *NONA (NO NAME)* (Version 2). Tucumán, Argentina: autor.
- Goloboff, P. A., Farris, J. S. & Nixon, K. C. (2008). TNT, a free program for phylogenetic analysis. *Cladistics*, 24(5), 774-786. doi: 10.1111/j.1096-0031.2008.00217.x
- Goloboff, P. A., Mattoni, C. I. & Quinteros, A. S. (2006). Continuous characters analyzed as such. *Cladistics*, 22(6), 589-601. doi: 10.1111/j.1096-0031.2006.00122.x
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, 59(3), 307-321. doi: 10.1093/sysbio/syq010
- Harvey, P. H. & Pagel, M. D. (1991). *The Comparative Method in Evolutionary Biology*. Oxford, Reino Unido: Oxford University.
- Hayward, A. C., Tollenaere, R., Dalton-Morgan, J. & Batley, J. (2015). Molecular marker applications in plants. En Batley (Ed.), *Plant genotyping: Methods and protocols* (13-27). New York: Springer. doi: 10.1007/978-1-4939-1966-6_2
- Herron, J. C. & Freeman, S. (2014). *Evolutionary analysis*. Glenview, Estados Unidos: Pearson.
- Huson, D., Rupp, R. & Scornavacca, C. (2010). *Phylogenetic Networks. Concepts, Algorithms and Applications*. New York: Cambridge University Press.
- Jiang, W., Zhu, J., Song, C., Li, X., Yang, Y. & Yu, W. (2014). Molecular phylogeny of the butterfly genus *Polytremis* (Hesperiidae, Hesperinae, Baorini) in China. *PLOS ONE*, 8(12), 1-15. doi: 10.1371/journal.pone.0084098
- Kim, Y.-K., Jo, S., Cheon, S.-H., Joo, M.-J., Hong, J.-R., Kwak, M. & Kim, K. J. (2020). Plastome evolution and phylogeny of Orchidaceae, with 24 new sequences. *Frontiers in Plant Science*, 11, 22. doi: 10.3389/fpls.2020.00022
- Knowles, L. L. & Carstens, B. C. (2007). Delimiting species without monophyletic gene trees. *Systematic Biology*, 56(6), 887-895. doi: 10.1080/10635150701701091
- Koenen, E. J. M., Ojeda, D. I., Steeves, R., Migliore, J., Bakker, F. T., Wieringa, J. J., Kidner, C., Hardy, O. J., Pennington, R. T., Bruneau, A. & Hughes, C. E. (2020). Large-scale genomic sequence data resolve the deepest divergences in the legume phylogeny and support a near-simultaneous evolutionary origin of all six subfamilies. *New Phytologist*, 225(3), 1355-1369. doi: 10.1111/nph.16290
- Kozlov, A. M., Darriba D., Flouri, T., Morel, B. & Stamatakis A. (2019). RAXML-NG: A fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, 35(21), 4453-4455. doi: 10.1093/bioinformatics/btz305
- Lemey, P., Rambaut, A., Welch, J. J. & Suchard, M. A. (2010). Phylogeography takes a relaxed random walk in continuous space and time. *Molecular Biology and Evolution*, 27(8), 1877-1885. doi: 10.1093/molbev/msq067
- Lemey, P., Salemi M. & Vandamme A. (Eds.) (2009). *The phylogenetic handbook: A practical approach to phylogenetic analysis and hypothesis testing*. New York: Cambridge University Press.

- Lemmon, A. R. & Moriarty, E. C. (2004). The importance of proper model assumption in bayesian phylogenetics. *Systematic Biology*, 53(2), 265-277. doi: 10.1080/10635150490423520
- Leopardi-Verde, C. L., Carnevali, G. & Romero-González, G. A. (2017). A phylogeny of the genus *Encyclia* (Orchidaceae: Laeliinae), with emphasis on the species of the Northern Hemisphere. *Journal of Systematics and Evolution*, 55(2), 110-123. doi: 10.1111/jse.12225
- Levin, D. (2004). Ecological speciation: Crossing the divide. *Systematic Botany*, 29(4), 807-816. doi: 10.1600/0363644042451134
- Lewis, P. O. (1998). Maximum likelihood as an alternative to parsimony for inferring phylogeny using nucleotide sequence data. En D. E. Soltis, P. S. Soltis & J. J. Doyle (Eds.), *Molecular Systematics of Plants II-DNA Sequencing* (132-163). New York: Springer.
- Mallet, J. (2005). Hybridization as an invasion of the genome. *TRENDS in Ecology and Evolution*, 20(5), 229-237. doi: 10.1016/j.tree.2005.02.010
- Manzo-Sánchez, G., Orozco-Santos, M., Islas-Flores, I., Martínez-Bolaños, L., Guzmán- González, S., Leopardi-Verde, C. L. & Canto-Canché, B. (2019). Genetic variability of *Pseudocercospora fijiensis*, the black Sigatoka pathogen of banana (*Musa* spp.) in Mexico. *Plant Pathology*, 68(3), 513-522. doi: 10.1111/ppa.12965
- Morrison, L. (2006). Multiple sequence alignment for phylogenetic purposes. *Australian Systematic Botany*, 19(6), 479-539. doi: 10.1071/sb06020
- Mount, D. (2001). *Bioinformatics sequence and genome analysis*. New York: Cold Spring Laboratory Press.
- Nixon, K. C. (1999). The parsimony Ratchet, a new method for rapid parsimony analysis. *Cladistics*, 15(4), 407-414. doi: 10.1111/j.1096-0031.1999.tb00277.x
- Nixon, K. C. (2002). *WinClada* (Versión 1.0). Ithaca, Estados Unidos: autor.
- Pivorano, W. & Heringa, J. (2008). Multiple sequence alignment. En J. Keith. (Ed.), *Bioinformatics, vol. I: data, sequence analysis, and evolution* (143-161). Berlín: Springer.
- Poe, S. & Wiens, J. J. (2000). Character selection and the methodology of morphological phylogenetics. En J. J. Wiens (Ed.), *Phylogenetic Analysis of Morphological Data* (20-36). Washington: Smithsonian Institution Press.
- Posada, D. (2008). JModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution*, 25(7), 1253-1256. doi: 10.1093/molbev/msn083
- Prado, B. R., Pozo, C., Valdez-Moreno, M. & Hebert, P. D. N. (2011). Beyond the colours: Discovering hidden diversity in the Nymphalidae of the Yucatan Peninsula in Mexico through DNA barcoding. *PLoS ONE*, 6(11), e27776. doi: 10.1371/journal.pone.0027776
- Richardson, A. O. & Palmer, J. D. (2006). Horizontal gene transfer in plants. *Journal of Experimental Botany*, 58(1), 1-9. doi: 10.1093/jxb/erl148
- 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(3), 539-542. doi: 10.1093/sysbio/sys029

- Ronquist, F., van den Mark, P. & Huelsenbeck, J. P. (2009). Bayesian phylogenetics analysis using mrbayes. En P. Lemey, M. Salem & A. Vandamme (Ed.), *Phylogenetic Handbook: A Practical Approach to Phylogenetic and Hypothesis Testing* (210-236). Oxford, Reino Unido: Cambridge University Press.
- Russell, A., Samuel, R., Klejna, V., Barfuss, M. H. J., Rupp, B. & Chase, M. W. (2010). Reticulate evolution in diploid and tetraploid species of *Polystachya* (Orchidaceae) as shown by plastid DNA sequences and low-copy nuclear genes. *Annals of Botany*, 106(1), 37-56. doi: 10.1093/aob/mcq092
- Sanderson, M. J. & Hufford, L. (Eds.). (1996). *Homoplasy, the recurrence of similarity in evolution*. San Diego: Academic Press.
- Schmidt, H. A. (2009). Testing tree topologies. En P. Lemey, M. Salemi & A. Vandamme (Eds.), *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*, (381-396). Cambridge, Reino Unido: Cambridge University Press.
- Simpson, M. G. (2019). *Plant systematics*. San Diego: Academic Press-Elsevier.
- Smith, U. & Hendricks, J. (2013). Geometric morphometric character suites as phylogenetic data: Extracting phylogenetic signal from gastropod shells. *Systematic Biology*, 62(3), 366-385. doi: 10.1093/sysbio/syt002
- Soltis, D. E. & Soltis, P. S. (1999). Choosing an approach and an appropriate gene for phylogenetic analysis. En D. E. Soltis, P. S. Soltis & J. J. Doyle. (Ed.), *Molecular Systematics of Plants II-DNA Sequencing* (1-42). New York: Springer.
- Soltis, P. S. & Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annual Review of Plant Biology*, 60(1), 561-588. doi: 10.1146/annurev.arplant.043008.092039
- Stavrinides, J. & Guttman, D. S. (2004). Mosaic evolution of the severe acute respiratory syndrome Coronavirus. *Journal of Virology*, 78(1), 76-82. doi: 10.1128/jvi.78.1.76-82.2004
- Strimmer, K. & von Haeseler, A. (2009). Genetic distances and nucleotide substitution models. En P. Lemey, M. Salem & A. Vandamme. (Ed.), *The Phylogenetic Handbook: A Practical Approach to Phylogenetic and Hypothesis Testing* (111-141). Cambridge, Reino Unido: Cambridge University Press.
- Swofford, D. L. (2002). PAUP*. *Phylogenetic Analysis Using Parsimony (*and Other Methods)* (Version 4). Sunderland, Estados Unidos de América: Sinauer Associates.
- Swofford, D. & Sullivan, J. (2009). Phylogeny inference based on parsimony and other methods using PAUP*. En P. Lemey, M. Salemi & A. Vandamme (Eds.), *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing* (267-312). Cambridge, Reino Unido: Cambridge University Press.
- Templeton, A. R. (2006). *Population genetics and microevolutionary theory*. Danvers, Estados Unidos de América: Wiley.
- Turland, N. J., Wiersema, J. H., Barrie, F. R., Greuter, W., Hawksworth, D. L., Herendeen, P. S., Knapp, S., Kusber, W.-H., Li, D.-Z., Marhold, K., May, T. W., McNeill, J., Monroe, A. M., Prado, J., Price, M. J. & Smith, G. F. (Eds.). (2018). *International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by*

the Nineteenth International Botanical Congress Shenzhen, China, July 2017. Regnum Vegetabile 159. Glashütten: Koeltz Botanical Books. doi: 10.12705/Code.2018

Via, S. & West, J. (2008). The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, 17(19), 4334-4345. doi: 10.1111/j.1365-294x.2008.03921.x

Wagner, G. P. (2001). *The Character Concept in Evolutionary Biology*. San Diego: Academic Press.

Whelan, S. (2008). Inferring trees. En Keith, J. M. (Ed.), *Bioinformatics, vol I: Data, sequence analysis, and evolution* (287-309). New York: Humana Press.

Wiley, E. O. & Lieberman, B. S. (2011). *Phylogenetics: Theory and Practice of Phylogenetic Systematics*. Hoboken, Estados Unidos de América: Willey-Blackwell.

Wu, G., Yao, Y., Qu, K., Ding Z., Li, H., Palanichamy M., Duan, Z., Li, N., Chen, Y. S. & Zhang, Y.-P. (2007). Population phylogenomic analysis of mitochondrial DNA in wild boars and domestic pigs revealed multiple domestication events in east Asia. *Genome Biology*, 8(11), R245. doi: 10.1186/gb-2007-8-11-r245



"Mandala
(Capullus)"
Raku
16.5 x 25 x 25 cm
2014