



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

**POSGRADO EN CIENCIAS BIOLÓGICAS**

**INSTITUTO DE BIOLOGÍA**

**BIOLOGÍA EVOLUTIVA**

**UNA EVALUACIÓN EMPÍRICA DEL MODELO DE NACIMIENTO Y  
MUERTE FOSILIZADO**

**TESIS**

QUE PARA OPTAR POR EL GRADO DE:

**MAESTRA EN CIENCIAS BIOLÓGICAS**

PRESENTA:

**KARINA YAÑEZ AROCHE**

**TUTORA PRINCIPAL DE TESIS: DRA. SUSANA AURORA MAGALLÓN PUEBLA**  
INSTITUTO DE BIOLOGÍA, UNAM

**COMITÉ TUTOR: DR. ARNO SIRI-JEGOUSSE**  
INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN SISTEMAS, UNAM  
**COMITÉ TUTOR: DR. SANTIAGO ALEJANDRO RAMÍREZ BARAHONA**  
INSTITUTO DE BIOLOGÍA, UNAM

**CIUDAD UNIVERSITARIA, CD. MX., Enero, 202**



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.





**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

**POSGRADO EN CIENCIAS BIOLÓGICAS**

**INSTITUTO DE BIOLOGÍA**

**BIOLOGÍA EVOLUTIVA**

**UNA EVALUACIÓN EMPÍRICA DEL MODELO DE NACIMIENTO Y  
MUERTE FOSILIZADO**

**TESIS**

QUE PARA OPTAR POR EL GRADO DE:

**MAESTRA EN CIENCIAS BIOLÓGICAS**

PRESENTA:

**KARINA YAÑEZ AROCHE**

**TUTORA PRINCIPAL DE TESIS: DRA. SUSANA AURORA MAGALLÓN PUEBLA**  
INSTITUTO DE BIOLOGÍA, UNAM

**COMITÉ TUTOR: DR. ARNO SIRI-JEGOUSSE**  
INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN SISTEMAS, UNAM

**COMITÉ TUTOR: DR. SANTIAGO ALEJANDRO RAMÍREZ BARAHONA**  
INSTITUTO DE BIOLOGÍA, UNAM

**CIUDAD UNIVERSITARIA, CD. MX., Enero, 2024**

COORDINACIÓN GENERAL DE ESTUDIOS DE POSGRADO  
COORDINACIÓN DEL POSGRADO EN CIENCIAS BIOLÓGICAS  
INSTITUTO DE BIOLOGÍA  
OFICIO: CGEP/CPCB/ IB /0892/2023  
ASUNTO: Oficio de Jurado

M. en C. Ivonne Ramírez Wence  
Directora General de Administración Escolar, UNAM  
P r e s e n t e

Me permito informar a usted que en la reunión ordinaria del Comité Académico del Posgrado en Ciencias Biológicas, celebrada el día **19 de junio de 2023** se aprobó el siguiente jurado para el examen de grado de **MAESTRA EN CIENCIAS BIOLÓGICAS** en el campo de conocimiento de **BIOLOGÍA EVOLUTIVA** de la estudiante **YAÑEZ AROCHE KARINA** con número de cuenta **310564519** con la tesis titulada **“UNA EVALUACIÓN EMPÍRICA DEL MODELO DE NACIMIENTO Y MUERTE FOSILIZADO”**, realizada bajo la dirección de la **DRA. SUSANA AURORA MAGALLON PUEBLA**, quedando integrado de la siguiente manera:

Presidente: DR. ALEJANDRO EMMANUEL GONZÁLEZ VOYER  
Vocal: DR. ARTURO CARLOS II BECERRA BRACHO  
Vocal: DRA. CAROLINA GRANADOS MENDOZA  
Vocal: DR. EMILIO ESTRADA RUÍZ  
Secretario: DR. SANTIAGO ALEJANDRO RAMÍREZ BARAHONA

Sin otro particular, me es grato enviarle un cordial saludo.

**ATENTAMENTE**  
**“POR MI RAZA HABLARÁ EL ESPÍRITU”**  
Ciudad Universitaria, Cd. Mx., a 14 de noviembre de 2023

**COORDINADOR DEL PROGRAMA**



**DR. ADOLFO GERARDO NAVARRO SIGÜENZA**

c. c. p. Expediente del alumno

AGNS/RGA/EARR/rga



## **Agradecimientos Institucionales**

Al Posgrado en Ciencias Biológicas de la Universidad Nacional Autónoma de México.

Al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT) por la beca 1103836 otorgada para realizar mis estudios de maestría.

A mi tutora principal, la Dra. Susana Aurora Magallón Puebla por la guía que me ha ofrecido para formarme como investigadora, por su tiempo, enseñanzas y sugerencias durante mis estudios y la realización de este trabajo de investigación.

A los miembros del comité tutor, el Dr. Santiago A. Ramírez Barahona y el Dr. Arno Siri-Jegousse, por su guía y valiosos comentarios que enriquecieron este trabajo de investigación.

## **Agradecimientos personales**

A la Dra. Susana Magallón, por sus enseñanzas, por su motivación y dedicación durante estos años. Por creer en mí y aceptar guiar mi formación. Por ser una gran inspiración personal y profesional.

A los miembros del jurado, Dr. Alejandro González Voyer, Dra. Carolina Granados, Dr. Santiago Ramírez Barahona, Dr. Arturo Becerra y Dr. Emilio Estrada, por sus revisiones y comentarios que mejoraron mi trabajo. Al Dr. Ricardo García Sandoval, por su ayuda durante etapas importantes de este proyecto.

Al Instituto de Biología de la UNAM por las facilidades y espacio brindados durante mis estudios de maestría.

A mi familia, porque han estado presentes en cada paso que he dado, apoyando mis objetivos y aspiraciones. Gracias mamá porque nunca has soltado mi mano.

A Carlos, porque siempre has creído en mí, incluso cuando yo no lo hago, por apoyarme y motivarme siempre.

A Adriana, Andrea, César y Santiago por todas las pláticas y risas en el laboratorio, porque me han inspirado y he aprendido mucho de ustedes. Gracias por su apoyo y motivación, gracias por su amistad.

A mis roomates, María y Gloria, por las pláticas, fiestas y cafeterías que hemos conocido juntas. Gracias por aguantar mi buen y pésimo ánimo.

# Índice

|  |    |
|--|----|
| Resumen .....  | 1  |
| Abstract.....  | 3  |
| Introducción.....  | 5  |
| Métodos de fechamiento.....                                  | 6  |
| Marco filogenético bayesiano.....                            | 9  |
| Modelo Fossilizado de Nacimiento y Muerte.....               | 19 |
| Importancia de los muestreos aleatorio y diversificado ..... | 21 |
| Familia Lauraceae.....                                       | 22 |
| Objetivos.....   | 25 |
| Materiales y métodos.....                                    | 25 |
| Muestreo taxonómico.....                                     | 25 |
| Datos moleculares.....                                       | 26 |
| Muestreo de fósiles.....                                     | 27 |
| Reconstrucción filogenética.....                             | 28 |
| Calibración de nodos .....                                   | 28 |
| Modelo Fossilizado de Nacimiento y Muerte.....               | 30 |
| Resultados.....  | 33 |
| Reconstrucción filogenética.....                             | 33 |
| Estimación de tiempos de divergencia.....                    | 36 |
| Discusión.....   | 41 |
| Reconstrucción filogenética .....                            | 41 |
| Estimación de tiempos de divergencia .....                   | 43 |
| Muestreo aleatorio versus muestreo diversificado .....       | 47 |
| Perspectivas .....   | 51 |
| Referencias bibliográficas.....                              | 52 |
| Anexo I.....   | 60 |

## Índice de Figuras y Tablas

|  |    |
|--|----|
| Figura 1. Representación del modelo tripartito para la estimación del tiempo de divergencia bayesiana. Se presentan los datos utilizados para el análisis, los modelos de sustitución, reloj y de árbol. El teorema de Bayes en la parte inferior ilustra cómo se une todo para la estimación bayesiana de los tiempos de divergencia. Ilustración obtenida y modificada de Warnock & Wright 2020. ....  | 13 |
| Figura 2. Fotografías de flores de especies vivientes y fósiles pertenecientes a la familia Lauraceae. En la parte superior, de izquierda a derecha las especies vivientes: <i>Ocotea porosa</i> (Trofimov et al. 2019); <i>Damburneya coriacea</i> ; <i>Nectandra barbellata</i> (Trofimov et al. 2016); <i>Persea americana</i> (Rohwer et al 2009). En la parte inferior de izquierda a derecha las especies fósiles: <i>Pothomacanthus lobatus</i> (von Balthazar et al. 2007); <i>Perseanthus crossmanensis</i> (Herendeen et al. 1994); <i>Neuseniania tetrasporangiata</i> (Eklund 2000); <i>Mauldinia mirabilis</i> (Drinnan et al 1990). .... | 24 |
| Figura 3. Árbol filogenético de máxima verosimilitud estimado con tres marcadores moleculares, obtenido con RAxML. ....  | 35 |
| Figura 4. Tiempos de divergencia de Lauraceae con método de calibración de nodos incluyendo 13 fósiles (CN13). Las barras rojas representan el 95 % de HPD. ....   | 36 |
| Figura 5. Tiempos de divergencia de Lauraceae con el proceso Nacimiento-Muerte Fossilizado incluyendo 62 fósiles e implementando A. muestreo aleatorio y B. muestreo diversificado. Las barras rojas representan el 95 % de HPD. Los números romanos se refieren a los seis principales grupos: I. <i>Neolitsea-Laurus</i> ; II. <i>Cinnamomum-Ocotea</i> ; III. <i>Machilus-Persea</i> ; IV. <i>Mezilaurus-Chlorocardium</i> ; V. <i>Neocinnamomum-Caryodaphnopsis</i> ; VI. <i>Beilschmiedia-Cryptocarya</i> . ....  | 37 |
| Figura 6. Figura 5. Tiempos de divergencia de Lauraceae con el proceso Nacimiento-Muerte Fossilizado incluyendo 19 fósiles e implementando A. muestreo aleatorio y B. muestreo diversificado. Las barras rojas representan el 95 % de HPD. Los números romanos se refieren a los seis principales grupos: I. <i>Neolitsea-Laurus</i> ; II. <i>Cinnamomum-Ocotea</i> ; III. <i>Machilus-Persea</i> ; IV. <i>Mezilaurus-Chlorocardium</i> ; V. <i>Neocinnamomum-Caryodaphnopsis</i> ; VI. <i>Beilschmiedia-Cryptocarya</i> . ....  | 38 |
| Figura 7. Raincloud plots que permiten visualizar y comparar la distribución de las edades de los nodos en los cuatro diferentes análisis realizados con el proceso de Nacimiento y Muerte Fossilizado. ....   | 40 |
| <b>Tabla 1.</b> Edades de los nodos troncal, corona y de los seis principales grupos identificados de la familia Lauraceae con las diferentes estrategias de estimación de tiempos de divergencia y muestreo (edades en Ma). ....  | 39 |

## Resumen

Reconstruir el árbol de la vida es el objetivo más importante y ambicioso de la sistemática filogenética. Como resultado de todos los esfuerzos de reconstrucciones filogenéticas de los diferentes grupos de seres vivos encontramos que la distribución de la riqueza de especies en el árbol de la vida es desigual. Algunos grupos tienen una gran diversidad de especies, mientras que otros cuentan con muy pocas especies. Actualmente contamos con una gran cantidad de datos filogenéticos, moleculares y morfológicos, lo que ha hecho posible inferir una gran cantidad de filogenias utilizando modelos matemáticos que permiten explicar la evolución y relación entre las especies.

La inferencia de las filogenias de las especies para tratar de describir la dinámica del proceso de diversificación (especiación y extinción) es uno de los principales objetivos de la biología evolutiva. Uno de los avances metodológicos más recientes y significativos para la estimación de los tiempos de divergencia ha sido el modelo de nacimiento y muerte fosilizado (FBD, por sus siglas en inglés), propuesto en el 2014. El proceso FBD considera que las especies vivas y fósiles son parte del mismo proceso de diversificación evolutiva. La precisión de las estimaciones de edad obtenidas con este modelo se ha evaluado principalmente mediante simulaciones. También se han llevado a cabo estudios empíricos basados en muestreo aleatorio de especies vivas y fósiles, aunque es posible que la mayoría de los conjuntos de datos empíricos no estén completamente representados o que se ignore un sesgo debido al muestreo incompleto y no aleatorio de especies. Se ha propuesto una variación del FBD conocida como muestreo diversificado que tiene en cuenta los sesgos derivados por un inevitable muestreo incompleto de especies, asumiendo que hemos

muestreado un descendiente existente de cada grupo principal dentro del clado estudiado. Por lo tanto, seleccionar el esquema de muestreo más adecuado de especies vivas y fósiles puede contribuir a mejores estimaciones para conjuntos de datos empíricos. En este proyecto utilicé a la familia Lauraceae para investigar la influencia de los esquemas de muestreo aleatorio y diversificado en las estimaciones de tiempo de divergencia utilizando el modelo FBD. Lauraceae incluye ca. 3.000 especies vivas y está abundantemente representada en el registro fósil por flores, hojas y madera. Los resultados muestran que la familia Lauraceae se originó ca. 118 Ma durante el Aptiano, Cretácico Inferior. Las relaciones filogenéticas de los principales clados son consistentes en todos los árboles estimados. Además, los resultados utilizando muestreo aleatorio son consistentes con estudios previos (distribución más amplia de las edades de los nodos). En comparación con el esquema de muestreo aleatorio, el muestreo diversificado recuperó árboles con ramas terminales más cortas, es decir, los tiempos de divergencia son más recientes, por lo que he considerado la posibilidad de que exista una influencia de las edades fósiles utilizadas, pues la mayoría de los fósiles son del Cenozoico.

## **Abstract**

Reconstructing the tree of life is the most important and ambitious goal of phylogenetic systematics. As a result of all the efforts in phylogenetic reconstructions of different groups of living organisms, we find that the distribution of species richness on the tree of life is uneven. Some groups have a high diversity of species, while others have very few species. Currently, we have a wealth of phylogenetic, molecular, and morphological data, which has made it possible to infer numerous phylogenies using mathematical models that explain the evolution and relationships between species.

The inference of species phylogenies, in an attempt to describe the dynamics of the diversification process (speciation and extinction), is one of the main goals of evolutionary biology. One of the most recent and significant methodological advances for estimating divergence times has been the fossilized birth-death (FBD) model, proposed in 2014. The FBD model considers that living and fossil species are part of the same evolutionary diversification process. The accuracy of age estimates obtained with this model has been primarily evaluated through simulations. Empirical studies based on random sampling of living and fossil species have also been conducted, although it is possible that most empirical datasets are not fully representative or that a bias due to incomplete and non-random sampling of species is present. A variation of the FBD, known as diversified sampling, accounts for biases arising from the inevitable incomplete sampling of species, assuming that a descendant of each major group within the studied clade has been sampled. Therefore, selecting the most suitable sampling scheme for living and fossil species can contribute to better estimates for empirical datasets.

In this project, I used the Lauraceae family to investigate the influence of random and diversified sampling schemes on divergence time estimates using the FBD model. Lauraceae includes approximately 3,000 living species and is well represented in the fossil record by flowers, leaves, and wood. The results show that the Lauraceae family originated approximately 118 million years ago during the Aptian, Early Cretaceous. The phylogenetic relationships of the major clades are consistent across all estimated trees. Furthermore, the results using random sampling are consistent with previous studies (wider distribution of node ages). In comparison to random sampling, diversified sampling recovered trees with shorter terminal branches, indicating more recent divergence times. Therefore, I have considered the possibility of an influence from the fossil ages used, as most fossils are from the Cenozoic era.

## Introducción

Reconstruir el árbol de la vida es el objetivo más importante y ambicioso de la filogenética y sistemática. Algunos ejemplos son los esfuerzos para reconstruir la filogenia de las plantas verdes o Viridiplantae (Ruhfel et al., 2014) o de las plantas terrestres o Embryophyta (Doyle, 1996; Soltis et al., 2002), así como el árbol de vida de animales (Dunn et al., 2008). Como resultado de todos los esfuerzos de reconstrucciones filogenéticas de los diferentes grupos de seres vivos, encontramos que uno de los patrones más interesantes, es la distribución desigual de la riqueza de especies en el árbol de la vida. Algunos grupos tienen una gran diversidad de especies vivientes, mientras que otros cuentan con muy pocas especies vivientes - por ejemplo, entre las familias de plantas está Amborellaceae que es monotípica y Orchidaceae, con ca. 25 000 especies. Se han propuesto algunas explicaciones para esta disparidad; una de ellas es que la edad de los clados es diferente, por lo que los grupos más antiguos han tenido más tiempo para acumular especies (McPeck y Brown 2007). Otra propuesta es que la tasa de diversificación –el balance entre la especiación y la extinción-, varía entre los diferentes clados, aquellos con una alta tasa de diversificación acumulan más especies (Pyron y Burbrink 2012; Quental y Marshall 2013, Stadler et al 2014, Rabosky 2009). Sin embargo, ese no siempre es el caso, existen clados antiguos que tienen muy pocas especies, que tal vez en el pasado tuvieron una diversidad mayor que la que encontramos en la actualidad, entonces, a lo largo del tiempo la riqueza de especies puede aumentar o disminuir como resultado de diferentes factores.

Actualmente, contamos con una gran cantidad de datos filogenéticos, moleculares y morfológicos, lo que ha hecho posible inferir una gran cantidad de filogenias utilizando

modelos matemáticos que permiten explicar la evolución y relación entre las especies. Aun cuando hay cada vez mayor disponibilidad de datos moleculares completos y de alta calidad, no es posible obtener datos moleculares de una gran cantidad de organismos, representando un gran obstáculo para la reconstrucción del árbol de la vida o, simplemente, reconstruir las filogenias representativas de algunos grupos. También debe ser considerado que secuencias moleculares demasiado extensas y un gran número de especies pueden provocar ciertas complicaciones computacionales y de tiempo necesario para obtener resultados. Además de la cantidad de datos disponibles, la inferencia de los parámetros y filogenias también son susceptibles a los efectos del muestreo de especies, por lo que la elección de una estrategia de muestreo adecuada suele ser conveniente para incrementar la precisión de los resultados obtenidos (Bininda-Emonds & Stamatakis, 2006).

### **Métodos de fechamiento**

La inferencia de las relaciones filogenéticas de los organismos es de gran importancia en la biología evolutiva, pero contar con estimaciones precisas de los tiempos absolutos de divergencia de las especies, así como de las tasas de los diferentes procesos (especiación, extinción y diversificación), nos permite comprender los eventos evolutivos en el contexto de una escala temporal global e identificar las causas de la variación en las tasas de evolución molecular.

El desarrollo de modelos matemáticos ha permitido explicar la evolución de las secuencias y a partir de esto inferir filogenias. El método de fechamiento molecular asume que el tiempo de divergencia entre dos especies puede medirse a partir de la estimación de número

de diferencias entre dos secuencias moleculares (Duchene & Bromham, 2013). Sin embargo, las divergencias de secuencias moleculares solo pueden proporcionar una escala de tiempo relativa, pero no brindan información sobre la edad absoluta. Emile Zuckerkandl (1962) y Linus Pauling (1965) se basaron en la suposición de que las tasas de evolución molecular son constantes entre linajes y a lo largo del tiempo, desarrollando el concepto de reloj molecular, bajo el cual todas las ramas del árbol filogenético se pueden caracterizar por una tasa única. Este modelo no es realista cuando se consideran diferentes especies porque las tasas moleculares son variables entre ellos (Ho & Duchêne, 2014). Se han propuesto diferentes modelos de relojes moleculares que toman en cuenta cómo se podría distribuir la heterogeneidad de las tasas moleculares entre los linajes. El modelo de reloj molecular que actualmente es más utilizado es el modelo de reloj relajado (Drummond et al., 2006; Magallón, 2020), que asigna una tasa de sustitución única a cada rama del árbol filogenético. Las tasas se pueden modelar siguiendo el principio de autocorrelación temporal, que considera que los atributos que determinan las tasas moleculares se transmiten de los linajes ancestrales a los descendientes y, en consecuencia, que las tasas entre linajes estrechamente relacionados son similares (Drummond et al., 2006; Gillespie, 1991; Rannala & Yang, 2007). Otra propuesta es el modelo de reloj molecular no correlacionado, en él se consideran las diferencias de tasa molecular entre linajes que divergieron en el pasado lejano, donde se produjo una extinción sustancial o cuando el muestreo es incompleto (Drummond et al., 2006).

La principal complicación con la datación molecular es que necesitamos separar las longitudes de las ramas en dos componentes: las tasas evolutivas y el tiempo (Bromham et al., 2018). Para hacer esto, incorporamos información temporal del registro fósil, eventos

biogeográficos, o estimaciones de tasas de estudios previos. Esto ha permitido la inferencia de árboles fechados, con ramas medidas en unidades de tiempo en lugar de unidades de tasa de sustitución relativa (Ho & Duchêne, 2014). Esto se conoce como calibrar el reloj molecular y la manera en la que ese hace tiene un gran impacto en las estimaciones de fecha que se obtienen (Guindon, 2020; Sauquet, 2013).

El método más común es la datación de nodos (Rannala & Yang, 1996), que consiste en usar la evidencia del registro fósil (u otras fuentes, por ejemplo, eventos biogeográficos) para definir los límites mínimos y/o máximos de la edad de un nodo específico y usar una distribución de probabilidad que refleje la incertidumbre de la calibración. La cantidad máxima de calibraciones que se pueden usar es igual a la cantidad de nodos internos en el árbol. Además, si contamos con varias fuentes de calibración para un mismo nodo, la más antigua informará la edad del nodo y el resto dejará de ser informativa. Por ejemplo, si contamos con muchos fósiles para un mismo nodo, utilizaremos el más antiguo para hacer la calibración. En este contexto, la calibración es un punto crítico y puede tener una influencia drástica en los resultados (Massoni et al., 2015; Sauquet et al., 2012), por lo que una revisión geológica exhaustiva es esencial para proporcionar una edad mínima precisa (Parham et al., 2012). La decisión sobre qué nodo será calibrado por un fósil debe tener en cuenta la incertidumbre de su posición filogenética y la incertidumbre de las relaciones entre las especies existentes (Massoni et al., 2015; Parham et al., 2012). Las calibraciones incorrectas, basadas en fósiles cuyas relaciones filogenéticas son incorrectas o a las que se les han asignado edades incorrectas, introducen errores en un análisis. Se ha demostrado que las estimaciones de tiempo de divergencia son sensibles a la elección de fósiles, su edad y la distribución elegida para la calibración (Heath, 2012; Ho & Phillips, 2009;

Warnock et al., 2012). Es necesaria una justificación exhaustiva y explícita de la edad y la relación filogenética para todos los fósiles utilizados para la calibración (Parham et al., 2012; Sauquet et al., 2012).

En sistemática existen varios métodos de distancia y de máxima verosimilitud para estimar tasas evolutivas y tiempos de divergencia de las especies (Sanderson, 2002). Sin embargo, el desarrollo del marco filogenético bayesiano ha sido un avance metodológico muy importante, en particular por la fácil interpretación de las probabilidades posteriores de los árboles y clados producidos y a la disponibilidad de algoritmos computacionales eficientes (Yang & Rannala, 2006).

### **Marco filogenético bayesiano**

Los métodos filogenéticos bayesianos se introdujeron en la década de 1990. y se han vuelto muy populares debido a los avances en la implementación de modelos evolutivos complejos y disponibilidad del software fácil de usar, lo que ha permitido analizar conjuntos de datos filogenéticos de gran tamaño (Nascimento et al., 2017; Yang & Rannala, 1997). Algunas de las utilidades de los métodos bayesianos en la filogenética son la reconstrucción de relaciones entre las especies, estimación de tiempos de divergencia, inferir la historia geográfica de las especies y la descripción del flujo de genes entre especies (CITA).

Los métodos bayesianos han sido preferidos por algunas razones. Las más importantes son que estos métodos estiman una muestra de valores de parámetros plausibles dado un modelo y un set de datos y proporcionan una incertidumbre asociada con cualquier

parámetro del modelo inferido, además permiten incorporar conocimiento existente sobre los valores de los parámetros (Warnock & Wright, 2020).

Una descripción más formal es que el método bayesiano es una metodología de inferencia estadística, que utiliza distribuciones de probabilidad para describir la incertidumbre de todos los parámetros del modelo. Se basa en el teorema de Bayes, que tiene tres componentes importantes, la verosimilitud del modelo, la probabilidad previa (o *prior*) y la probabilidad posterior. El teorema supone que la distribución de probabilidad posterior es el resultado del producto de la distribución previa y la verosimilitud, dividido por la probabilidad de los datos. En filogenética, el teorema toma la siguiente forma:

$$P(\theta|D) = P(\theta) P(D|\theta) / P(D)$$

donde  $P(\theta|D)$  es la probabilidad posterior, es decir, la distribución de probabilidad de los parámetros dados los datos,  $P(\theta)$  es la probabilidad previa, la distribución de probabilidad de los parámetros,  $P(D|\theta)$  es la verosimilitud, es decir, la probabilidad de los datos dados los parámetros del modelo y  $P(D)$  es la probabilidad de los datos (Figura 1).

De manera más explícita, la verosimilitud del modelo se refiere al cálculo de la probabilidad de los datos observados dado un modelo particular. En este contexto, es importante explicar qué consideramos como un modelo matemático, según Warnock y Wright (2020) un modelo es la descripción matemática de un fenómeno, este está compuesto de parámetros que van a representar puntos clave para describir el fenómeno y que van a ser tratados como variables aleatorias, es decir, un valor desconocido. En el caso de la filogenética molecular, la verosimilitud se calcula considerando el modelo de sustitución, de reloj molecular y el de árbol. Representa la probabilidad de que obtengamos

las secuencias seleccionadas, si el modelo que elegimos para describirlas fuera verdadero (Bromham 2017; Zhu 2020).

La probabilidad previa es una distribución de probabilidad a partir de la cual se puede extraer el valor de un parámetro en particular. Pueden representar nuestras creencias sobre los parámetros en el modelo de verosimilitud, sin considerar ninguna información de los datos. Es común utilizar distribuciones como Gamma o Exponencial como distribuciones previas, que pueden ser muy flexibles, debido a que la información confiable con la que los investigadores pueden fundamentar la elección de los previos podría no estar disponible o poderse asegurar con demasiada certeza, de manera que la flexibilidad otorgada por las distribuciones Gamma o Exponencial puede considerarse deseable (Warnock & Wright, 2020). Por otro lado, cuando no tenemos información sobre los valores que podría tomar un parámetro, podemos especificar una distribución previa "no informativa" que otorga la misma probabilidad a una amplia gama de valores posibles. Un ejemplo es la distribución uniforme, que otorga la misma probabilidad a los valores extremos de un rango elegido como a los valores intermedios (Bromham et al., 2018). Las probabilidades previas son parte del modelo, pues al restringir los valores que puede tomar un parámetro, es posible dirigir la estimación hacia o desde ciertos conjuntos de valores, sin cambiar el proceso de generación de árboles que se está implementando.

El posterior es el resultado de un análisis bayesiano, incluye una distribución de valores posibles para todos los parámetros especificados en nuestros modelos (modelos de sustitución, de reloj molecular y de árbol), incluida la topología del árbol y los tiempos de divergencia. Algo importante es que no podemos calcular fácilmente la probabilidad posterior debido a la cantidad de información y complejidad de los modelos filogenéticos.

Una muestra de la probabilidad posterior se genera a través del algoritmo conocido como *Markov Chain Monte Carlo* (MCMC). Bajo este algoritmo, se proponen valores para los parámetros y se califica la verosimilitud de los datos bajo este modelo. Luego, los parámetros del modelo se cambian y los datos se vuelven a calificar bajo este nuevo modelo. MCMC no realiza un seguimiento de los valores que ya se han puntuado, por lo tanto, un parámetro que se ajusta bien puede recuperarse varias veces. Los valores de cada parámetro aparecerán en la distribución posterior en proporción a su probabilidad, dado nuestro modelo y las distribuciones previas. El punto o puntos más altos de la distribución posterior representan las estimaciones de parámetros más probables. El algoritmo visita valores de parámetros con una distribución posterior alta con más frecuencia que aquellos con una distribución posterior baja. De hecho, el algoritmo visita los valores de los parámetros exactamente en proporción a su distribución posterior (Nascimento et al., 2017). La varianza de la distribución posterior para un parámetro dado refleja la incertidumbre en esa estimación. Si la varianza en nuestra distribución posterior es relativamente alta, esto refleja mucha incertidumbre en nuestra estimación de parámetros. Por el contrario, si la varianza es relativamente baja, la incertidumbre en nuestra estimación es baja, por lo que hay un alto respaldo en la estimación dados los datos, el modelo y las distribuciones previas.

A partir de la distribución posterior de los parámetros del modelo es posible obtener los intervalos de credibilidad (análogos a los intervalos de confianza). Por lo general, usamos el intervalo de densidad posterior más alta (HDP) del 95%, que es la dispersión de los valores posteriores que contiene el 95% del posterior y son una forma de dar la incertidumbre asociada con los parámetros o edades de los nodos. Por otra parte, resumir la

distribución posterior de árboles significa resumir los árboles estimados y obtener las relaciones filogenéticas que están mejor respaldadas por el posterior así como los valores de soporte para cada nodo (Bromham et al., 2018; O'Reilly & Donoghue, 2020; Warnock & Wright, 2020) (Bomham 2017; O'Reilly y Donoghue, 2018 Warnock & Wright 2020).

Una vez que tenemos una explicación del método bayesiano y sabemos que la inferencia posterior de un árbol filogenético bayesiano es una distribución de árboles y parámetros de un modelo específico, es importante comprender los modelos subyacentes utilizados para generar el árbol (Bromham et al., 2018; Nascimento et al., 2017).

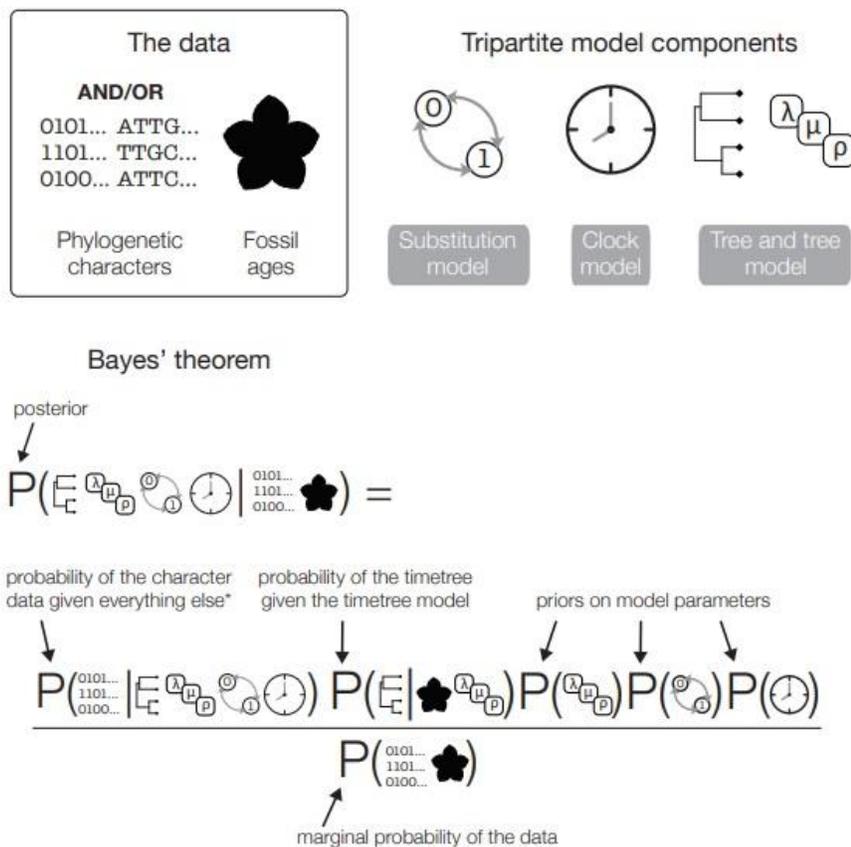


Figura 1. Representación del modelo tripartito para la estimación del tiempo de divergencia bayesiana. Se presentan los datos utilizados para el análisis, los modelos de sustitución, reloj y de árbol. El teorema de Bayes en la parte inferior ilustra cómo se une todo para la estimación bayesiana de los tiempos de divergencia. Ilustración obtenida y modificada de Warnock & Wright 2020.

Una manera en la que se ha desglosado el problema de estimar las relaciones filogenéticas, es decir, la topología de un árbol y las edades de los nodos, es asumir un modelo tripartito de evolución. El primer modelo describe la acumulación de diferencias entre los caracteres (modelo de sustitución), el segundo modelo describe la distribución de las tasas evolutivas en todo el árbol (modelo del reloj molecular) y el tercer modelo que describe la distribución de eventos de especiación, y por ende las edades de los nodos a lo largo del árbol (modelo de árbol; Warnock & Wright, 2020; Yang & Rannala, 1997).

El modelo de sustitución nos permite estimar la cantidad de cambio genético que ha ocurrido a lo largo del tiempo. La forma más sencilla es contar las diferencias observadas entre las secuencias de ADN alineadas. Sin embargo, es posible que estas diferencias no nos permitan considerar cambios pasados, o sustituciones múltiples en un mismo sitio, entonces el modelo de sustitución utiliza creencias sobre las probabilidades de diferentes tipos de eventos de sustitución para inferir cuántas sustituciones probablemente hayan ocurrido en el pasado, las cuales no se ven reflejadas en las diferencias entre las secuencias actuales. Los modelos de sustitución suelen incluir tres componentes principales: la matriz de sustitución que describe las tasas relativas de cambio entre diferentes nucleótidos; las frecuencias de los cuatro nucleótidos; y la forma en que las frecuencias varían entre los sitios (Bromham et al., 2018). Los modelos de sustitución más utilizados son los miembros de la familia general reversible en el tiempo (GTR, por sus siglas en inglés). El modelo GTR es el más complejo, ya que permite una tasa diferente para cada tipo de cambio de nucleótido. Otros modelos dentro de la familia GTR son el modelo más simple, el modelo Jukes-Cantor (JC), que supone que todos los tipos de sustituciones son igualmente probables, por lo que puede describirse mediante un único parámetro de tasa de sustitución

(Jukes & Cantor, 1969); El modelo HKY que incorpora patrones desiguales de tipos de sustitución, es decir, otorga diferentes tasas de transiciones y transversiones (Kimura, 1980) o el modelo F81 que considera diferentes frecuencias de nucleótidos (Felsenstein, 1981). En el caso de los caracteres morfológicos el más utilizado es el modelo Mk (Lewis, 2001), que es una generalización del modelo de sustitución JC que permite diferentes números de estados de caracteres.

En el caso del modelo de reloj, su función es describir la forma en que la tasa de cambio de caracteres varía, o no varía, a lo largo del árbol. Se han descrito diferentes modelos que hacen diferentes suposiciones acerca de cómo se distribuye la variación de esta tasa entre las ramas, estos van desde que cada rama tenga la misma tasa de evolución hasta que cada rama tenga su propia tasa.

El modelo de reloj molecular más simple asume tasas homogéneas entre las ramas de la filogenia. Esto generalmente se conoce como un reloj molecular estricto o global. El modelo de reloj estricto tiene un solo parámetro: la tasa de evolución,  $k$ , que se expresa en sustituciones por sitio por unidad de tiempo, según contexto temporal del análisis. Sin embargo, es más realista que esta tasa no sea igual en todas las ramas del árbol, por lo que se han propuesto una serie de modelos de reloj molecular que relajan la suposición de homogeneidad de tasas (Bromham et al., 2018; Ho & Duchêne, 2014). Para clasificar los modelos de reloj molecular se distingue entre el número de tasas distintas ( $k$ ) que se asignan a las ramas del árbol. Entonces,  $k = 1$  representa el modelo de reloj estricto. Cuando  $1 < k < n$ , algunas ramas comparten la misma tasa y el número de tasas distintas suele ser mucho menor que el número de ramas, estos son conocidos como modelos de reloj de frecuencia múltiple. Y un modelo de reloj relajado es aquel en el que  $k$  es igual al

número de ramas en el árbol ( $n$ ), de modo que cada rama puede tener una tasa distinta (Ho & Duchêne, 2014).

Otro punto importante de los relojes moleculares, es que, para que proporcionen edades de los nodos en una escala de tiempo absoluta en lugar de proporcionar tiempos relativos, tienen que ser calibrados. Las calibraciones pueden provenir de varias fuentes independientes, incluido el registro fósil, la evidencia geológica o edades obtenidas de una fuente secundaria. En el caso del método bayesiano filogenético esta información se agrega al análisis especificando una distribución previa informativa para la edad del nodo correspondiente. Es posible implementar múltiples calibraciones en todo el árbol. En el caso del uso de fósiles, estos dan una restricción mínima sobre la edad de un clado y generalmente se utiliza el representante fósil más antiguo y se aplican como restricciones de edad o una distribución previa para un nodo en el árbol, cuya elección suele ser subjetiva y se basa en una interpretación general del investigador (Parham et al., 2012; Sauquet, 2013).

Por último, los modelos de árbol incorporan suposiciones sobre los procesos de generación de árboles y nos permiten describir la probabilidad de observar un árbol calibrado a partir de la evidencia temporal que alimentamos a nuestros análisis. Esta información sobre la edad se utiliza para calibrar los modelos de sustitución y de reloj. Se trata de una familia de modelos utilizados en la inferencia de tiempo de divergencia bayesiana que combinados con la información de calibración tienen un impacto importante en las estimaciones de las edades de los nodos. Lo anterior se logra a partir de dos principales enfoques en los cuales la información temporal se incorpora al modelo de árbol por medio del fechamiento de nodos (*node-dating*) o el fechamiento de puntas (*tip-dating*). En el caso del fechamiento de

nodos, el resultado es un árbol fechado que incluye solo las relaciones entre especies vivas. Las calibraciones forman parte de la estimación de los tiempos de divergencia al informar sobre la edad de un nodo interno utilizando los registros más antiguos, pero no juegan un papel directo en la estimación de otros parámetros del modelo (Magallón, 2020; Warnock & Wright, 2020). Por el contrario, en el fechamiento de puntas las especies vivientes y fósiles pueden ser consideradas para la estimación de las relaciones filogenéticas, y tiempos de divergencia. Este enfoque se basa en análisis conocidos como evidencia total, en el que secuencias moleculares de especies vivientes y caracteres morfológicos tanto de especies vivientes como fósiles son utilizados, siendo los datos morfológicos utilizados para estimar las relaciones filogenéticas de las especies fósiles, evitando la necesidad de especificar la ubicación de los fósiles con respecto a las especies vivientes. Las calibraciones son proporcionadas por los fósiles, pero también éstos son representados como terminales en ramas filogenéticas extintas. Los fósiles utilizados no solo deben ser los más antiguos, sino que es deseable incluir tantos fósiles como sea posible, porque proporcionarán mayor información temporal para fechar el árbol. La utilidad de los fósiles dependerá de la disponibilidad de datos morfológicos filogenéticamente informativos (Arcila et al., 2015; Magallón, 2020).

Entonces, en el contexto del método bayesiano, el modelo de árbol se denomina previo de árbol (*tree prior*), pues se asigna una probabilidad previa para cada topología de árbol y a los parámetros que componen este modelo (Bromham et al., 2018). La función del previo de árbol es dar más peso a árboles de cierta apariencia y compatibles con un modelo particular del proceso de ramificación.

Uno de los modelos más utilizados para la datación molecular es el proceso de nacimiento-muerte (Kendall 1948). Este modelo tiene parámetros que describen la tasa de especiación (nacimiento,  $\lambda$ ) y la tasa de extinción (muerte,  $\mu$ ). El proceso de Yule es su forma más simple, también se conoce como modelo de nacimiento puro porque solo modela la especiación sin permitir la extinción de linajes (propuesta que no es realista), asumiendo que la tasa de especiación es constante (Yule, 1924). El modelo de nacimiento y muerte, además de suponer una tasa constante de especiación en cada linaje, permiten que cada linaje tenga la misma posibilidad de extinguirse en cualquier momento. Rara vez tenemos información directa sobre las tasas de especiación y extinción, por lo que estos parámetros del modelo de nacimiento y muerte, generalmente, se estiman en el análisis. La elección de la probabilidad previa de árbol tiene un gran impacto en la precisión de las estimaciones de las tasas de especiación y extinción. La probabilidad previa del árbol también contribuye a la distribución posterior de todos los tiempos de los nodos, influyendo en la precisión de las estimaciones de fecha para cada nodo en el árbol. En algunas versiones del modelo se ha añadido un parámetro que representa la proporción de especies vivientes incluidas en los datos muestreados ( $\rho$ ), este es un parámetro fijo en el análisis, ya que no se puede estimar si no se conocen los otros parámetros (Stadler, 2010). El modelo se ha desarrollado de tal manera que puede permitir que las tasas de nacimiento y muerte cambien con el tiempo (Stadler & Yang, 2013). También, permite incorporar información de los fósiles, de manera que informen el proceso de ramificación y estimación de tiempos de divergencia, esta versión es conocida como modelo fosilizado de nacimiento y muerte (Heath et al., 2014). En este caso la distribución de los fósiles se incorpora en el modelo a través de un parámetro de recuperación de fósiles ( $\psi$ ) que refleja la fracción esperada de linajes que dejan evidencia fósil antes de extinguirse. La edad de cada fósil se especifica como una

edad puntual o como un intervalo temporal (Barido-Sottani et al., 2019). El modelo de nacimiento y muerte fosilizado será utilizado en este trabajo y explicado con más detalle a continuación.

### **Modelo Fosilizado de Nacimiento y Muerte.**

El modelo de nacimiento y muerte fosilizado (FBD por sus siglas en inglés) es un enfoque de estimación de tiempos de divergencia alternativo a la calibración de nodos, que integra las muestras fósiles en el mismo proceso de diversificación que describe la dinámica de las especies vivientes, descrito en Stadler (2010) y Heath et al. (2014) bajo un marco de inferencia bayesiano. Esto implica que los fósiles pueden interpretarse como terminales de ramas extintas o como ancestros muestreados en ramas filogenéticas (Gavryushkina et al., 2017).

El modelo FBD describe la probabilidad de que el árbol de especies vivientes y los fósiles estén condicionados por los parámetros nacimiento-muerte:  $f [T | \lambda, \mu, \rho, \psi, x_c]$ , donde  $T$  denota la topología del árbol, los tiempos de divergencia, los tiempos de aparición de fósiles y la configuración de cómo los fósiles se adhieren al árbol, es decir, determinar si los fósiles se encuentran como ancestros muestreados en ramas con especies vivientes o como terminales de ramas extintas al unirse al árbol después de un evento de especiación (Heath et al., 2014). Los parámetros del modelo son: tasa de especiación ( $\lambda$ ), tasa de extinción ( $\mu$ ), la probabilidad de muestrear una especie existente ( $\rho$ , cuantas especies son muestreadas con respecto al total de especies descritas en el clado), tasa de recuperación de fósiles ( $\psi$ , tasa que toma en cuenta aspectos de los fósiles, tales como la fosilización, si son

encontrados, colectados y curados) y el momento en el que se originó el proceso ( $x_c$ ). El modelo FBD asume tasas constantes de especiación y extinción a lo largo del tiempo.

Este enfoque puede proporcionar mejores estimaciones de las tasas y tiempos de divergencia, aumenta en gran medida la cantidad de evidencia fósil que se puede usar durante la inferencia y producir edades más precisas para los nodos (Heath et al., 2014). Además de las secuencias moleculares y la información temporal que proveen los fósiles, es posible utilizar este modelo con caracteres morfológicos para las especies vivientes y muestras fósiles (enfoque de evidencia total), pero cuando estos caracteres morfológicos no están disponibles, el modelo permite imponer una estructura de tiempo en el árbol al permitir al usuario definir todos los puntos de unión posibles para los fósiles en el árbol de especies vivientes (Heath et al., 2014). Por lo tanto, uno de los desafíos es vincular los taxones fósiles con la filogenia de los taxones vivientes, aunque el método estima la ubicación de cada fósil y estos no están fijos dentro de un clado en particular, es importante tener conocimiento previo de las relaciones filogenéticas y tener en cuenta la incertidumbre de su posición filogenética (Sauquet, 2013).

El modelo FBD se ha ido desarrollando cada vez más al permitir la integración de información de rango estratigráfico como la edad de los fósiles, en lugar de seleccionar solo un punto medio o aleatorio de ese rango, así como diferentes modos de especiación (Silvestro et al., 2018; Stadler et al., 2018). Otra de las variaciones permite que tasas de muestreo, especiación y extinción sean variables en el tiempo (Gavryushkina et al., 2014), entre otras.

### **Importancia de los muestreos aleatorio y diversificado**

Es posible que la mayoría de los conjuntos de datos empíricos no estén completamente muestreados o que se ignore un sesgo por la elección de especies no aleatoria. Por ejemplo, pueden faltar especies vivientes debido a la disponibilidad limitada de secuencias moleculares o cuando trabajamos con clados grandes, donde la inclusión de todas las especies sería computacionalmente impráctico (Matschiner, 2019). El modelado incorrecto del proceso de muestreo puede causar problemas al inferir tasas de diversificación, así como en la estimación de edades en las filogenias (Zhang et al., 2016).

En los diferentes análisis de datación se suponga un muestreo completo o uniformemente aleatorio, tanto de los taxones vivientes como de los fósiles (en el caso del proceso FBD). Esto puede ser conveniente desde una perspectiva matemática, pero en la práctica generalmente la selección de especies no es aleatoria y el muestreo suele estar asociado con algún sesgo. La selección de especies para los análisis filogenéticos rara vez es uniformemente aleatoria, donde cada taxón tiene la misma probabilidad de ser incluido en el conjunto de datos, en cambio, una estrategia diferente denominada muestreo "diversificado" de especies vivientes es más común (Höhna, 2014; Höhna et al., 2011). Höhna et al. (2011) presentaron la importancia del muestreo de especies vivientes en los análisis de datación y exploran los efectos que tiene el muestreo sesgado utilizando simulaciones. Demostraron que el modelado incorrecto del proceso de muestreo puede causar problemas importantes al inferir las tasas de especiación y extinción en los modelos de nacimiento y muerte. Comúnmente, el objetivo de los biólogos al realizar un muestreo es maximizar la representación de los taxones considerados en el estudio. Por ejemplo, en un análisis de las relaciones dentro de una familia, trataremos de incluir ejemplares de todas

las tribus o todos los géneros. El muestreo diversificado el que más se asemeja a esta estrategia, haciéndolo, implícitamente, más común que un muestreo aleatorio (Zhang et al., 2016). En esta estrategia de muestreo se toma en consideración que el muestreo es incompleto y no aleatorio, busca adaptarse a la manera en la que los investigadores a menudo diseñan los muestreos de especies en los que buscan incluir representantes de cada uno de los principales grupos dentro del clado estudiado (por ejemplo, Barba-Montoya et al., 2018; Dos Reis et al., 2015; Jarvis et al., 2014).

Zhang et al. (2016) y Matschiner (2019) ampliaron el proceso FBD para tener en cuenta el muestreo diversificado de taxones existentes, evaluaron el desempeño de ambos tipos de muestreo, aleatorio y diversificado por medio de simulaciones y datos empíricos. Sus resultados coincidieron en que los árboles resultados de un muestreo diversificado tienen ramas terminales más largas, por lo tanto, edades más antiguas; mientras que el muestreo aleatorio produce árboles con un amplio rango de edades de los nodos, además de que en promedio se obtienen ramas terminales más cortas. Por lo tanto, estos enfoques brindan la oportunidad de examinar los impactos potenciales del muestreo en diferentes escalas filogenéticas en las inferencias de tiempos de divergencia y otros parámetros macroevolutivos (Luo et al., 2021).

### **Familia Lauraceae**

Las lauráceas forman una gran familia de plantas leñosas, es una de las familias de angiospermas tempranamente divergentes con más especies, con ca. 2500 especies clasificadas en más de 50 géneros. Están ampliamente distribuidas en latitudes tropicales y

subtropicales con unas pocas especies en áreas templadas (Rohwer, 1993). Se distribuyen principalmente en bosques húmedos en todas las elevaciones (van der Werff, 1991). Se consideran como una de las familias neotropicales económicamente más importantes (Rohwer, 1993). La madera de Lauraceae tiene una amplia gama de propiedades, desde especies blandas y de baja densidad hasta maderas duras y densas, y se utilizan en muebles, herramientas agrícolas o instrumentos musicales (Song et al., 2017). También, tienen importantes usos en la medicina herbal, perfumes y especias. Además, los frutos de las lauráceas, son ricos en lípidos y un elemento importante de la dieta de humanos y otros animales (Mazer & Wheelwright, 1993). Aunque la familia Lauraceae es monofilética (Renner, 1999; Renner & Chanderbali, 2000), la delimitación genérica sigue sin estar clara (Chanderbali et al., 2001; Rohwer, 2000; Rohwer & Rudolph, 2005). Uno de los motivos es que las secuencias utilizadas en los diferentes trabajos tienen una capacidad limitada para resolver problemas filogenéticos a nivel de especie o dentro de algunos géneros, por ejemplo, los géneros *Litsea* o *Persea* (Fijridiyanto & Murakami, 2009; J. Li et al., 2004; L. Li et al., 2011). Además, a pesar de los esfuerzos realizados por varios grupos de trabajo, los muestreos normalmente se limitan a unas pocas especies de los géneros más grandes, lo que puede causar un sesgo en los resultados. Por otro lado, esta familia tiene un amplio registro fósil de flores, hojas y madera cuyas edades van desde el Cretácico temprano, aunque son más abundantes durante el Cenozoico (Drinnan et al., 1990; Eklund & Kvacek, 1998).

Por lo anterior, las lauráceas proporcionan un excelente sistema modelo para investigar la influencia de los supuestos del modelo de nacimiento y muerte fosilizado en los tiempos de divergencia inferidos, que suponen un muestreo aleatorio y un muestreo diversificado.

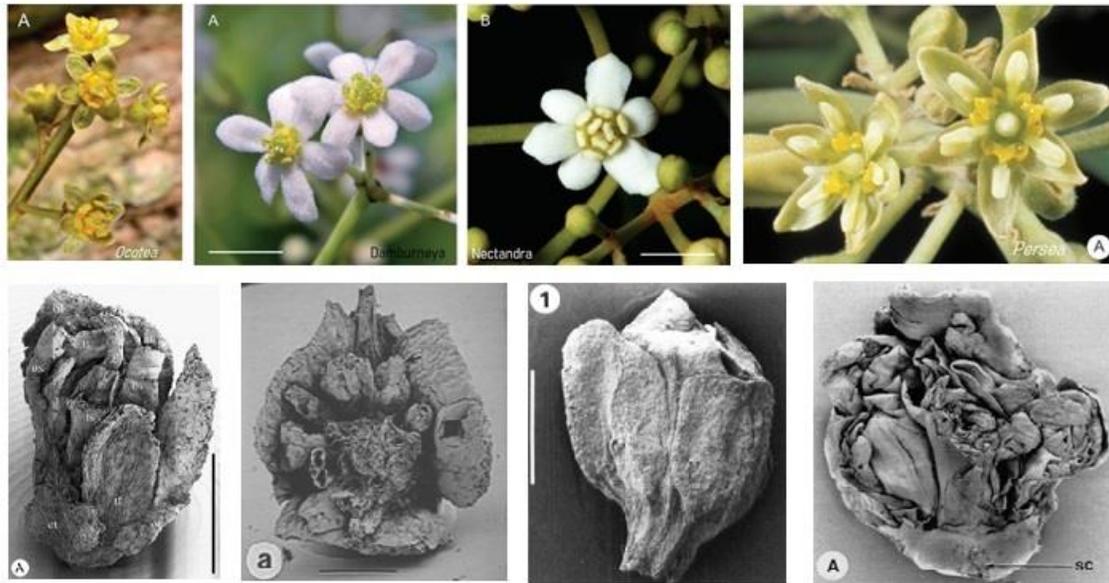


Figura 2. Fotografías de flores de especies vivientes y fósiles pertenecientes a la familia Lauraceae. En la parte superior, de izquierda a derecha las especies vivientes: *Ocotea porosa* (Trofimov et al. 2019); *Damburneya coriacea*; *Nectandra barbellata* (Trofimov et al. 2016); *Persea americana* (Rohwer et al. 2009). En la parte inferior, de izquierda a derecha las especies fósiles: *Pothomacanthus lobatus* (von Balthazar et al. 2007); *Perseanthus crossmanensis* (Herendeen et al. 1994); *Neusenias tetrasporangiata* (Eklund 2000); *Mauldinia mirabilis* (Drinnan et al. 1990).

El propósito de este proyecto es investigar la influencia de los supuestos de muestreo aleatorio y muestreo diversificado en el modelo de nacimiento y muerte fosilizado en los tiempos de divergencia inferidos, usando a las lauráceas como sistema modelo. De acuerdo con trabajos previos que han evaluado el desempeño de estos modelos de muestreo (Matschiner, 2019; Zhang et al., 2016) esperamos: a) que los árboles obtenidos de un muestreo diversificado presenten ramas terminales más largas (edades más antiguas) y b) que el muestreo aleatorio produzca árboles con un amplio rango de edades de los nodos y que en promedio se obtengan ramas terminales más cortas en comparación con el muestreo diversificado. Además, esperamos observar diferencias en los árboles resultantes de ambos modelos de muestreo según las edades de los fósiles utilizados en cada análisis, es decir, al incorporar una mayoría de fósiles de edades más recientes estos producirían edades de los

nodos más jóvenes, y al incluir fósiles más antiguos estos influirían en la estimación de los tiempos de divergencia produciendo edades más antiguas.

## **Objetivos.**

- Evaluar el efecto del muestreo aleatorio y diversificado en los tiempos de divergencia estimados con el modelo de nacimiento y muerte fosilizado de la familia Lauraceae.

## **Materiales y métodos.**

### **Muestreo taxonómico.**

Obtuve un muestreo taxonómico de 321 especies vivientes, de las cuales siete pertenecen al grupo externo (cuatro especies de la familia Calycanthaceae y tres especies de la familia Monimiaceae) y 102 representantes fósiles de la familia Lauraceae. El muestreo taxonómico se concentró en obtener la representación del mayor número de géneros de la familia Lauraceae de acuerdo a la disponibilidad de secuencias en GenBank (método explicado en la siguiente sección). Algunos géneros, tales como *Litsea* está representado por 36 spp., *Beilshmiedia* por 29 spp. y *Cinnamomum* por 26 spp., pero otros como *Sassafras* solo está representado por 2 spp. o *Umbellularia*, el cual es un género monotípico. En el caso del grupo externo, la familia Monimiaceae es una de las más cercanas a la familia Lauraceae (junto con Hernandiaceae) según un análisis de seis

regiones del genoma plastídico realizado por Renner (1999; Renner & Chanderbali, 2000), mientras que Calycanthaceae es hermana de todas las demás Laurales (Renner, 1999). Incluí las especies *Hedycarya angustifolia*, *Mollinedia schottiana* y *M. widgrenii* de la familia Monimiaceae y las especies *Calycanthus chinensis*, *C. floridus*, *C. occidentalis* e *Idiospermum australiense* de la familia Calycanthaceae porque en GenBank tienen disponibles los mismos marcadores moleculares utilizados para las especies de la familia Lauraceae utilizadas en este estudio. En el caso de los fósiles incluí registros cuyas descripciones presentaban asignaciones confiables a la familia Lauraceae. De manera general, el muestreo taxonómico estuvo basado en la disponibilidad de datos moleculares en GenBank y registros fósiles en la literatura, los cuales cumplieron con nuestros criterios para ser consideradas en este estudio, esta selección se describe en las siguientes secciones.

#### **Datos moleculares.**

Utilicé el software PyPHLAWD (Smith & Walker, 2019) para construir una base de datos molecular de la familia Lauraceae. Con el objetivo de construir una matriz molecular con el mayor número de géneros realicé un análisis de agrupamiento, en el que todos los genes disponibles en la base de datos de GenBank para diferentes especies se identifican y agrupan en archivos por separado. Obtuve secuencias de dos regiones de ADN de cloroplasto, *rbcL*, *trnK-matK*, y la región de los espaciadores internos transcritos (16S - ITS1 - 5,8S - ITS2 - 28S) del ADN ribosomal. Las matrices descargadas incluían en varios casos más de una especie por género. Logré una alineación a nivel de género seleccionando las especies para las que estaban disponibles las secuencias de todos o la mayoría de los marcadores moleculares utilizados. PyPHLAWD produce alineamientos de cada región genética con MAFFT (Katoh & Standley, 2013). Sin embargo, solo conservé el

alineamiento del marcador nuclear, pues para los dos marcadores plásticos realicé un alineamiento manual en Mesquite v.3.70 (Maddison & Maddison, 2021) tomando en cuenta el marco de lectura de ambos marcadores. Posteriormente, construí una matriz concatenada con las tres regiones genéticas, verifiqué la alineación visualmente y eliminé los extremos de las secuencias que excedían sustancialmente la longitud del resto de las secuencias.

Agregué manualmente los grupos externos, que consisten en cuatro especies de la familia Calycanthaceae (*Calycanthus chinensis*, *C. floridus*, *C. occidentalis* e *Idiospermum australiense*) y tres especies de la familia Monimiaceae (*Hedycarya angustifolia*, *Mollinedia schottiana* y *M. widgrenii*).

### **Muestreo de fósiles.**

Obtuve información de descripciones fósiles de la familia Lauraceae a partir de una búsqueda en la Paleobiology Database (<http://paleodb.org>) y la revisión de literatura. Se seleccionaron los registros que presentaban asignaciones confiables a la familia Lauraceae, es decir, aquellos registros cuya asignación fue explícitamente discutida en un trabajo publicado. La lista completa de fósiles contiene información sobre tipo de fósil y preservación, similitud con especies o géneros actuales, métodos y descripción utilizada para definir la pertenencia a la familia Lauraceae, edad del fósil y formación geológica donde fue encontrado, así como la referencia (Anexo I). Las edades se actualizaron de acuerdo con la carta cronoestratigráfica internacional de Cohen et al. (2013, actualizada el 2022).

Recopilé información de 102 fósiles, varios de los cuales pueden asignarse con confianza a géneros de la familia Lauraceae. Además, asigné un puntaje a los fósiles que me permitió

seleccionarlos en las siguientes fases de mi investigación. Los fósiles con puntuación de 3 son aquellos que cuentan con una descripción y comparación morfológica detallada con las especies vivientes de la familia; se asignó una puntuación de 2 a los fósiles que se asignaron a la familia basándose en información disponible en la literatura; y la puntuación de 1 se le dio a los fósiles obtenidos en estudios de paleofloras.

### **Reconstrucción filogenética.**

Para evaluar qué modelo de sustitución se ajusta mejor a los datos moleculares realicé una prueba basada en el Criterio de Información de Akaike (AIC) implementada en jModelTest (Darriba et al., 2012). El modelo de sustitución de nucleótidos General Time Reversible con una distribución Gamma para considerar la heterogeneidad de tasas entre sitios (GTR +  $\Gamma$ ) fue el de mejor ajuste. Los datos moleculares fueron subdivididos en tres particiones con base en los resultados de jModelTest, cada partición corresponde a un marcador molecular. Hice una reconstrucción filogenética con máxima verosimilitud usando RAxML v.8 (Stamatakis, 2014). Implementé la aproximación GTRCAT y un bootstrap no paramétrico derivado de 1000 réplicas. La aproximación GTRCAT se refiere a una solución computacional para el modelo GTR+  $\Gamma$ , esta aproximación cumple el mismo propósito, permitiendo la integración de la heterogeneidad de las tasas en los análisis filogenéticos con un menor costo de tiempo computacional y consumo de memoria (Stamatakis, 2016).

### **Calibración de nodos**

En este método, el modelo de generación de árboles es el proceso de nacimiento-muerte, en el que la estimación de las relaciones filogenéticas y de tiempos de divergencia es independiente del tipo y la fuente de las calibraciones. Como resultado, la información

proporcionada por las calibraciones solo tiene influencia en las edades de los nodos del árbol, pero el resultado es un árbol filogenético con terminales existentes.

Para obtener el árbol fechado, implementé el método de calibración de nodos en BEAST v. 2.6.7 (Bouckaert et al., 2014), utilizando un reloj molecular relajado no correlacionado, con las diferentes tasas de sustitución de las ramas obtenidas de una distribución lognormal (UCLN). Se consideró una partición por cada marcador utilizado, la misma que utilicé para los análisis filogenéticos. Se vinculó el modelo de reloj que se aplicó a la alineación molecular, suponiendo que las diferentes particiones de datos se rigen por el mismo tipo de variación entre linajes. También vinculé el modelo del árbol, suponiendo que un solo modelo de generación de árboles rige los tiempos de especiación. Sin embargo, permití que hubiera variación de la tasa entre sitios, mediante la implementación de diferentes modelos de sustitución molecular en diferentes particiones de datos, es decir, los modelos de sustitución de sitio se desvincularon entre las particiones del árbol. En cada caso, se asignó un modelo GTR +  $\Gamma$  con frecuencias de base empíricas y tasas de sustitución estimadas. La distribución Gamma, que permite tomar en cuenta la heterogeneidad de la tasa entre sitios, se subdividió en cuatro categorías discretas y se estimó su forma. Las probabilidades previas para cada tipo de transición se obtuvieron a partir de una distribución gamma con  $\alpha = 2.0$  y  $\beta = 0.5$ . La probabilidad previa para la media UCLN se obtuvo de una distribución exponencial con media = 10, y la desviación estándar de una distribución gamma con  $\alpha = 0,5396$  y  $\beta = 0,3819$  (por *default*). Se definieron 13 nodos internos no redundantes para ser calibrados, los cuales fueron ubicados directamente por la posición filogenética de los fósiles en el análisis filogenético de máxima verosimilitud. A cada nodo se le asignó la edad mínima derivada del límite superior del intervalo estratigráfico en el que el fósil fue

asignado. Las calibraciones fueron implementadas como valores previos (*priors*) obtenidos de una distribución lognormal limitada por la edad mínima del fósil (*offset*), una media ( $\mu$ ) igual a la edad mínima + 15%, y una desviación estándar ( $s$ ) de 1.0. La raíz del árbol, es decir, el nodo troncal, fue calibrada secundariamente con una distribución Uniforme en el intervalo de la edad máxima y mínima estimado para el orden Laurales por Ramírez-Barahona et al. (2020).

### **Modelo Fosilizado de Nacimiento y Muerte.**

Realicé los análisis con el FBD que supone un muestreo aleatorio para la estimación del tiempo de divergencia y reconstrucción de las relaciones filogenéticas en BEAST2 v.2.6.7 (Bouckaert et al., 2014), utilizando el paquete Sampled Ancestors (Gavryushkina et al., 2014), que permite modelar las especies fósiles como taxones terminales o como ancestros muestreados en ramas filogenéticas.

En los análisis se incluyeron 321 taxones vivientes, de los cuales siete corresponden al grupo externo. Llevé a cabo dos análisis con diferente número de fósiles: FBD62, que contiene las especies fósiles asignadas con suficiente grado de confianza a la familia Lauraceae, es decir, los fósiles en las categorías 2 y 3; y FBD19, que contiene las especies fósiles asignadas a la familia Lauraceae con un alto grado de confianza, de la categoría 3. Las edades de las especies vivientes y fósiles se especificaron como millones de años antes del presente y se incorporaron en BEAUti a través de un archivo de texto que incluía la edad de todos los taxones vivientes (0 millones de años [Ma]) y fósiles (Anexo 1). Las relaciones filogenéticas entre especies vivientes y fósiles se especificaron con el comando TAXSETS en el archivo NEXUS de entrada. Realicé análisis con cada set de fósiles que me permitiera considerar la incertidumbre de la edad de los fósiles. Por un lado se

consideraron las edades de acuerdo al intervalo de tiempo definido por el rango estratigráfico en el que se encontró al fósil y por otro utilizamos la media de ese intervalo (Material suplementario 2).

El conjunto de datos moleculares es el mismo que utilicé en los análisis de reconstrucción filogenética y *node-dating* y se dividió en (1) el *ITS* nuclear; los genes plastídicos (2) *matK* y (3) *rbcL*. Desvinculé los modelos de sustitución de cada sitio para asumir diferentes modelos de evolución de secuencias para cada gen, los tres marcadores fueron analizados con un modelo GTR +  $\Gamma$ . Los previos para cada tipo de transición de nucleótidos se obtuvieron de una distribución gamma, con  $\alpha = 2.0$  y  $\beta = 0.5$  para la transición AG, y el resto de transiciones con  $\alpha = 2.0$  y  $\beta = 0.25$ . Por otra parte, especificué el modelo de heterogeneidad de tasa molecular entre linajes, en este caso vinculé los modelos de reloj para que los tres genes tuvieran las mismas tasas relativas de sustitución entre las ramas. Implementé un reloj molecular relajado no correlacionado, con las tasas de las ramas extraídas de una distribución lognormal (UCLN). El número de tasas discretas se codificó como -1, lo que indica que el número de tasas discretas en el modelo de reloj relajado es igual al número de ramas en el árbol ( $k = n$ ). Además, a los previos para los dos hiperparámetros: la media y la desviación estándar (*uclMean* y *uclStdev*, respectivamente) se obtuvieron de una distribución exponencial; *uclMean* con la media igual a 10.0; y *uclStdev* con media 0.3337, ambos con un offset igual a 0.

Finalmente, vinculé los árboles para asegurar que las tres particiones compartan la misma topología de árbol y tiempos de ramificación, que fueron especificados con el Proceso de Nacimiento y Muerte Fosilizado (Heath et al., 2014). El modelo es condicionado al origen (tiempo de inicio) del proceso, correspondiente a la raíz del árbol porque todos los fósiles

fueron asignados al grupo corona de Lauraceae. A la probabilidad previa de la edad del nodo raíz le asigné una distribución uniforme entre 122.3 y 172.0 Ma, correspondiente al 95% de la densidad posterior (HPD) más alta de la edad estimada para el nodo de Laurales por Ramírez-Barahona et al. (2020). La tasa de diversificación ( $r$ ) previa se obtuvo a partir de una distribución exponencial con media = 1 (según Heath, 2022). La tasa de recambio ( $\epsilon$ ) previa fue obtenida de una distribución uniforme entre 0.0 y 1.0. La proporción ( $s$ ) de muestreo fósil, la probabilidad de que un linaje entre en el registro fósil antes de extinguirse, se obtuvo a partir de una distribución beta con  $\alpha = 2.0$  y  $\beta = 2.0$  (según Heath, 2022). La proporción de especies vivientes muestreadas ( $\rho$ ), correspondiente a la proporción de especies muestreadas en relación con el número total de especies de Lauraceae, se fijó en 0.129.

Para los análisis se implementó el modelo FBD que supone un muestreo diversificado, que reconoce que el muestreo es incompleto y sesgado, utilicé el paquete BDSKY v.1.3.3 (*Birth-Death Skyline*; Stadler et al., 2013) que permite seleccionar el modelo generador del árbol como *SAFBDSkyline* (*Sampled Ancestors Fossilized Birth Death Skyline*), especificando el muestreo diversificado como *BrithDeathSkylineDiversifiedSampling*. Los conjuntos de datos de especies vivientes y fósiles (FBD62 y FBD19), así como las configuraciones de las edades para considerar la incertidumbre de las edades de los fósiles (edades en intervalo y edades medias) y valores de parámetros fueron las mismas que las que utilizamos con el muestreo aleatorio.

Los análisis de calibración de nodo y FBD se realizaron en CIPRES Science Gateway (Miller et al., 2010) y en el laboratorio B-123 del Instituto de Biología, Universidad Nacional Autónoma de México. Los análisis de calibración de nodo consistieron en dos

ejecuciones independientes de Cadena de Markov Monte Carlo (MCMC) de 150 millones de generaciones y 200 millones de generaciones muestreando cada 1000 generaciones, y para los diferentes FBD62 y FBD19 una ejecución de 600 millones de generaciones muestreando una cada 10000 generaciones. Examiné los resultados de las diferentes ejecuciones para determinar la convergencia de la cadena y el tamaño de muestra efectivo (ESS) de los diferentes parámetros estimados con Tracer v.1.6 (Rambaut et al., 2014). Los árboles de diferentes ejecuciones de los análisis de calibración de nodo se combinaron con LogCombiner v.2.6.6 (Bouckaert et al., 2014). Para cada uno de los nueve análisis obtuve el Árbol de Máxima Credibilidad (Maximum Clade Credibility; MCC) con TreeAnnotator v.2.6.6 (Bouckaert et al., 2014), sobre el cual se resumieron las estimaciones de edad, y fueron visualizados con FigTree (Rambaut, 2014) y editados en rstudio .

## **Resultados.**

### **Reconstrucción filogenética.**

El conjunto de datos moleculares incluye tres marcadores, que comprende a 2621 pares de bases para 321 especies vivientes, de las cuales 314 pertenecen a la familia Lauraceae, tres a la familia Monimiaceae y cuatro a la familia Calycanthaceae. Los 314 taxones corresponden a al menos el 95% de los géneros en Lauraceae. Las relaciones filogenéticas entre la familia Calycanthaceae y las familias Monimiaceae + Lauraceae obtenidas en el análisis de máxima verosimilitud en RAxML tienen un apoyo de bootstrap (BS) del 100%,

y el apoyo bootstrap obtenido entre la familia Monimiaceae y Lauraceae también es del 100% (Fig. 3). Este análisis confirma la monofilia de la familia Lauraceae, que coincide con estudios filogenéticos publicados previamente (Chanderbali et al., 2001, Rohwer y Rudolph, 2005). La topología obtenida muestra relaciones sólidas y suficientemente respaldadas entre linajes profundos de Lauraceae, aunque es necesario considerar que en la mayoría de los géneros con más de una especie no encontramos monofilia. Se identificaron seis clados entre los linajes profundos de Lauraceae, estos son consistentes con trabajos previos, y es importante notar que la mayoría tienen valores de apoyo bajos, con la excepción de *Machilus-Persea* y *Mezilaurus-Chlorocardium* (Chanderbali et al., 2001, Rohwer y Rudolph, 2005, Song et al., 2019): *Neolitsea-Laurus* (BS=32%), *Cinnamomum-Ocotea* (BS=53%), *Machilus-Persea* (BS=84%), *Mezilaurus-Chlorocardium* (BS=86%), *Neocinnamomum-Caryodaphnopsis* (BS=48%) y *Beilschmiedia-Cryptocarya* (BS=10%).

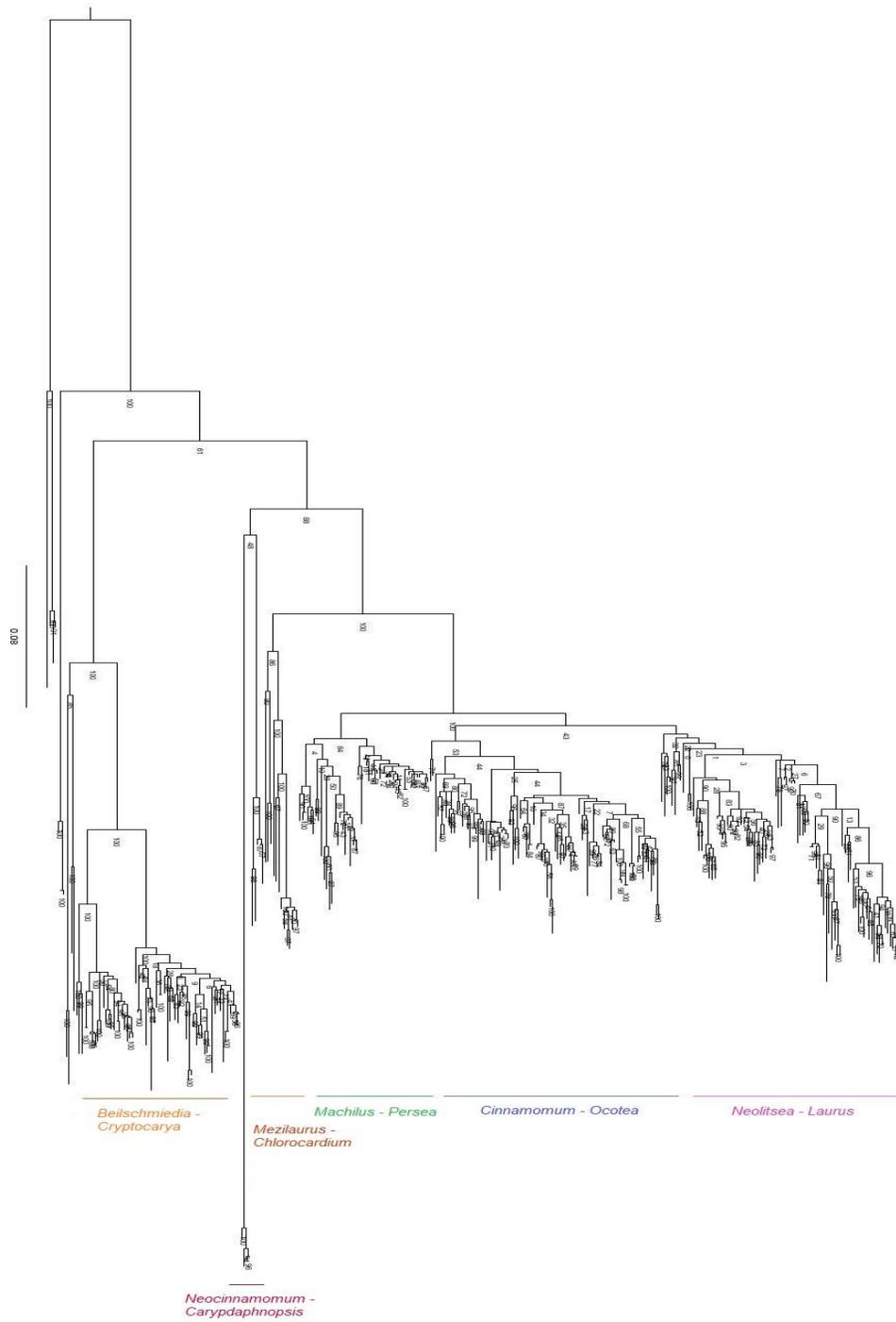


Figura 3. Árbol filogenético de máxima verosimilitud estimado con tres marcadores moleculares, obtenido con RAxML.

## Estimación de tiempos de divergencia.

En general, las edades estimadas para la familia Lauraceae con los métodos de calibración de nodos y el proceso de nacimiento y muerte fosilizado fueron similares (Fig. 4-6). Es importante notar que las relaciones filogenéticas estimadas con ambos métodos se mantienen similares a la reconstrucción filogenética obtenida con máxima verosimilitud.

El árbol fechado obtenido con el método de calibración de nodos (CN13) se presenta en la Fig. 4. La edad estimada para el inicio de diversificación de Lauraceae con este método es hace 130.26 millones de años (95% HPD 120.15-140.59), durante el Hauteriviano (Cretácico temprano).

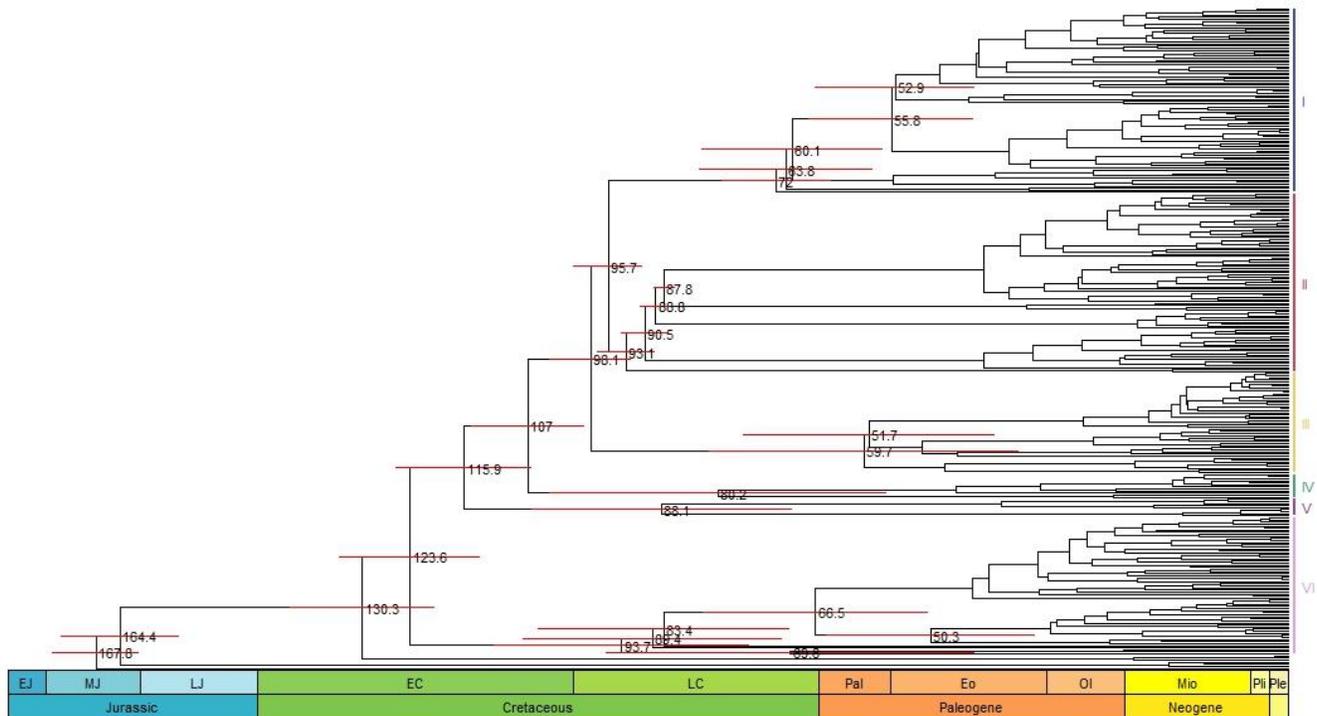


Figura 4. Tiempos de divergencia de Lauraceae con método de calibración de nodos incluyendo 13 fósiles (CN13). Las barras rojas representan el 95 % de HPD.

Las estimaciones con el modelo FBD y un muestreo aleatorio (FBD62R y FBD19R; Fig. 5 y 6, respectivamente) indican un origen en el Aptiano (Cretácico temprano), hace aproximadamente 118.39 y 118.2 millones de años (115.95-120.46 y 115.69-120.56), respectivamente, para el inicio de la diversificación de la familia Lauraceae. El caso de las estimaciones con el modelo FBD y un muestreo diversificado es similar. Al incluir 62 fósiles (FBD62D; Fig.5) la edad estimada para el inicio de la divergencia de Lauraceae es hace aproximadamente 117.27 (113.8-120.34) millones de años. Con 19 fósiles (FBD19D; Fig.6) la edad es similar a las obtenidas para el muestreo aleatorio, hace 118.06 millones de años (115.2-120.36). Es importante mencionar que en ambos esquemas de muestreo se infiere que todos los fósiles son puntas en lugar de nodos ancestrales en los árboles.

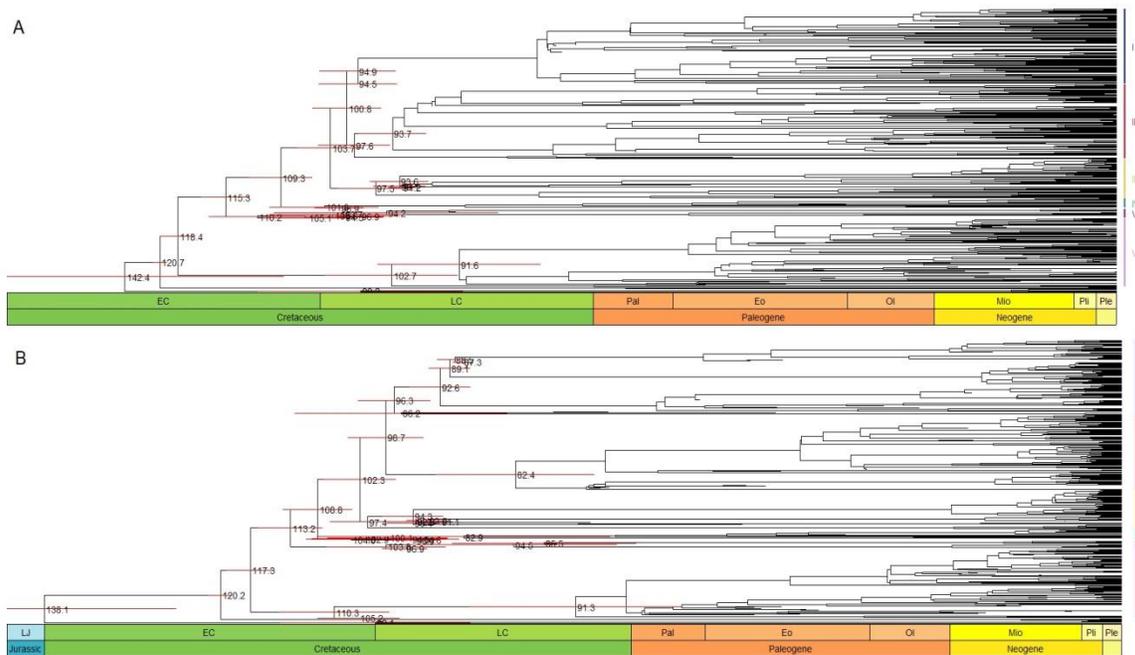


Figura 5. Tiempos de divergencia de Lauraceae con el proceso Nacimiento-Muerte Fossilizado incluyendo 62 fósiles e implementando A. muestreo aleatorio y B. muestreo diversificado. Las barras rojas representan el 95 % de HPD. Los números romanos se refieren a los seis principales grupos: I. *Neolitsea-Laurus*; II. *Cinnamomum-Ocotea*; III. *Machilus-Persea*; IV. *Mezilaurus-Chlorocardium*; V. *Neocinnamomum-Caryodaphnopsis*; VI. *Beilschmiedia-Cryptocarya*.

En la Tabla 1 se presentan las estimaciones de edad de los nodos troncal, corona de la familia Lauraceae y de los nodos corona de los seis principales grupos identificados de la familia Lauraceae con las diferentes estrategias de estimación de tiempos de divergencia y muestreo.

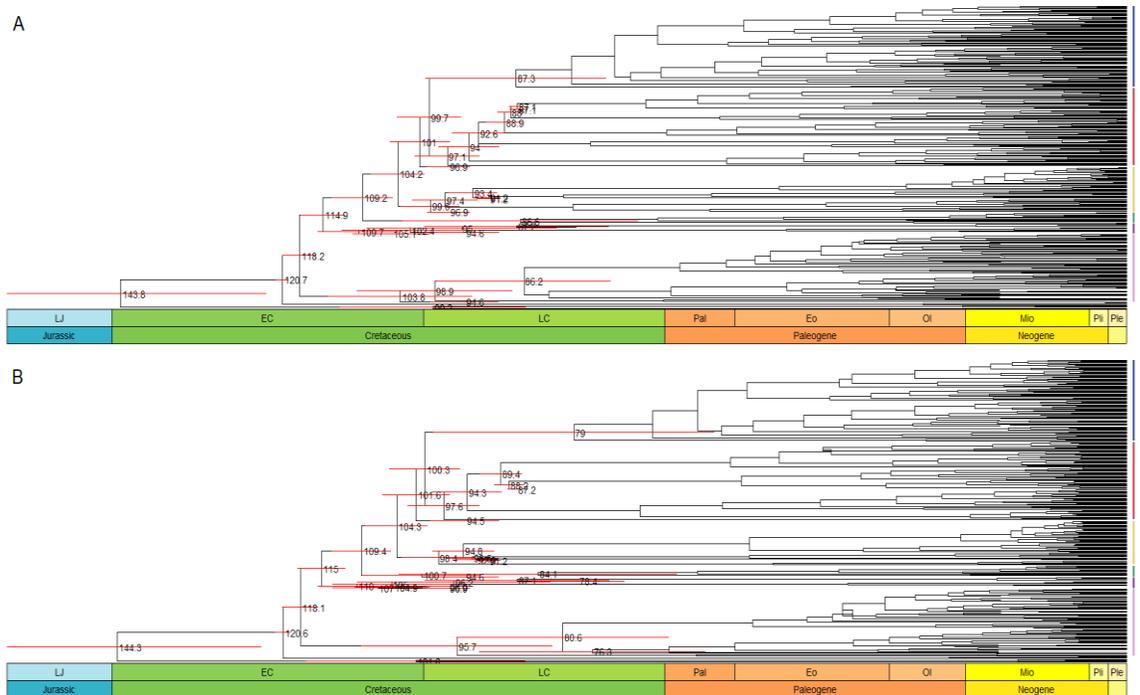


Figura 6. Tiempos de divergencia de Lauraceae con el proceso Nacimiento-Muerte Fossilizado incluyendo 19 fósiles e implementando A. muestreo aleatorio y B. muestreo diversificado. Las barras rojas representan el 95 % de HPD. Los números romanos se refieren a los seis principales grupos: I. *Neolitsea-Laurus*; II. *Cinnamomum-Ocotea*; III. *Machilus-Persea*; IV. *Mezilaurus-Chlorocardium*; V. *Neocinnamomum-Caryodaphnopsis*; VI. *Beilschmiedia-Cryptocarya*.

A pesar de que las edades del nodo troncal de Laurales y nodo corona de Lauraceae no difieren mucho, el modelado del proceso de muestreo tiene un impacto en los tiempos de divergencia inferidos de los nodos internos. Cuando se supone que los taxones existentes se muestrearon aleatoriamente, las edades inferidas de las lauráceas tienen una distribución homogénea a lo largo del tiempo, aunque vemos ramas terminales relativamente más largas (Fig. 5 y 6) y con intervalos HPD más amplios. Al cambiar del muestreo aleatorio al diversificado, la mayoría de las ramas terminales fueron más cortas, sin superar el Mioceno

tardío. Es posible observar que las ramas previas al periodo Neógeno son mucho más largas, tanto con 62 como con 19 fósiles (Fig. 5 y 6).

Los resultados utilizando un muestreo aleatorio son consistentes con los trabajos de Zhang et al. (2016) y Matschiner (2019) pues hay una distribución más amplia de las edades de los nodos. Esos trabajos también suponen que el muestreo diversificado tiene el efecto de estirar las ramas más recientes y encoger las más ancestrales, lo que da como resultado estimaciones de edad más jóvenes cerca de la raíz del árbol; no obstante, nuestros resultados recuperan árboles con ramas terminales más cortas, es decir, lo contrario a lo esperado.

Tabla 1. Edades de los nodos troncal, corona y de los seis principales grupos identificados de la familia Lauraceae con las diferentes estrategias de estimación de tiempos de divergencia y muestreo (edades en Ma).

|   | Calibración<br>de nodos | FBD62-<br>Muestreo<br>aleatorio | FBD62-<br>Muestreo<br>diversificado | FBD19-<br>Muestreo<br>aleatorio | FBD19-<br>Muestreo<br>diversificado |
|---|-------------------------|---------------------------------|-------------------------------------|---------------------------------|-------------------------------------|
| Nodo troncal-<br>Laurales                 | 167.75                  | 142.45                          | 138.09                              | 143.83                          | 141.31                              |
| Nodo corona<br>Lauraceae                  | 130.26                  | 118.39                          | 117.27                              | 118.2                           | 118.06                              |
| <i>Beilschmiedia-<br/>Cryptocarya</i>     | 93.74                   | 102.65                          | 91.28                               | 103.85                          | 95.7                                |
| <i>Neocinnamomum-<br/>Caryodaphnopsis</i> | 88.1                    | 94.21                           | 94.5                                | 109.72                          | 109.97                              |
| <i>Mezilaurus-<br/>Chlorocardium</i>      | 80.22                   | 89.54                           | 82.87                               | 86.62                           | 100.72                              |
| <i>Machilus-Persea</i>                    | 59.68                   | 97.52                           | 97.41                               | 99.49                           | 98.35                               |
| <i>Cinnamomum-Ocotea</i>                  | 93.13                   | 97.57                           | 82.38                               | 97.1                            | 97.58                               |
| <i>Neolitsea-Laurus</i>                   | 71.95                   | 88.58                           | 96.28                               | 87.29                           | 78.95                               |

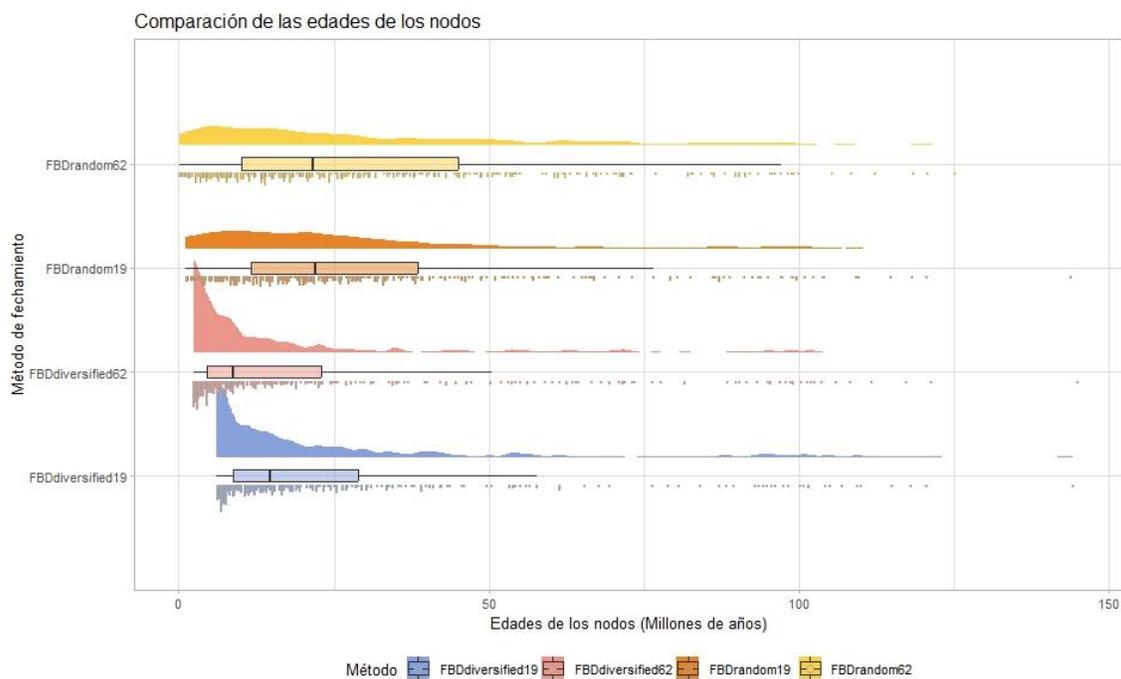


Figura 7. Raincloud plots que permiten visualizar y comparar la distribución de las edades de los nodos en los cuatro diferentes análisis realizados con el proceso de Nacimiento y Muerte Fossilizado.

Las Figuras 5 y 6, en los árboles B, correspondientes a los análisis que consideran un muestreo diversificado, es obvia la concentración de edades de las ramas terminales en el periodo Neógeno. Las Figuras 5 y 6, en los árboles A, nos muestra que en el caso en el que se considera un muestreo aleatorio las edades están muy uniformemente distribuidas a lo largo de los 100 millones de años, aunque la mayoría de los nodos no son más antiguos que 50 millones de años, mientras que en los casos en los que se considera un muestreo diversificado están muy concentradas entre el presente (0 millones de años) y los 50 millones de años.

## Discusión.

### Reconstrucción filogenética

El árbol de las especies vivientes estimada con máxima verosimilitud basada únicamente con datos moleculares es congruente con estudios filogenéticos previos sobre la familia Lauraceae. Estos trabajos han demostrado que no es fácil reconstruir las relaciones filogenéticas en la familia (Chanderbali et al., 2001; Fijridiyanto & Murakami, 2009; L. Li et al., 2011; Rohwer, 2000; Rohwer et al., 2009; Rohwer & Rudolph, 2005; Song et al., 2020), por lo que las afinidades de las especies y las relaciones genéricas dentro de muchos de los principales linajes de la familia aún están poco resueltas. La mayoría de los estudios filogenéticos sobre la familia están enfocados en la reconstrucción de relaciones dentro de ciertos géneros, tales como *Neolitsea*, *Persea*, *Sassafras*, *Lindera*, entre otros (Fijridiyanto & Murakami, 2009; L. Li et al., 2007, 2011; Nie et al., 2007).

Algo que se debe considerar es la capacidad de los marcadores moleculares utilizados para resolver las relaciones filogenéticas entre las especies vivientes de la familia Lauraceae que se ha visto reflejado en estudios previos (Chanderbali et al., 2001; Renner, 1999; Rohwer, 2000; Song et al., 2020). Generalmente, la información de las secuencias de nucleótidos del cpADN es útil para resolver las relaciones filogenéticas en plantas, sobre todo a niveles medios o profundos de la filogenia (Barraclough et al., 1996; Sang, 2002). Sin embargo, los marcadores moleculares de cpADN, como *rbcL*, tienen tasas de evolución molecular relativamente lentas (Clegg, 1993), por lo que su utilidad filogenética a niveles superficiales a veces es limitada. Algunas regiones de codificación de cpADN que evolucionan rápidamente, como *matK* y *ndhF*, se han utilizado para análisis filogenéticos

en niveles taxonómicos bajos y la resolución a menudo fue insatisfactoria (Sang, 2002), en el caso de la familia Lauraceae esto se puede corroborar en los trabajos de Rohwer (2000) y Chanderbali et al. (2001). Por otro lado, ha aumentado el uso del *nrITS* (*ITS1-5.8S-ITS2*) para reconstruir la filogenia a niveles genéricos o inferiores en plantas (Álvarez & Wendel, 2003). Esta región se ha vuelto popular en la sistemática de plantas porque es relativamente fácil de amplificar utilizando cebadores universales, la evolución concertada a menudo homogeneiza las secuencias de copias y los datos de *nrITS* han demostrado ser lo suficientemente variables como para resolver las relaciones en niveles más bajos en muchos grupos de plantas (Sang, 2002).

Los análisis de máxima verosimilitud proporcionaron apoyo estadístico alto para las relaciones de la familia Lauraceae, tanto a nivel genérico como intragenérico (Fig. 1), especialmente en linajes de los nodos más profundos, como el grupo *Cryptocarya* más *Beilschmiedia*, el grupo *Caryodaphnopsis-Neocinnamomum* y el grupo *Mezilaurus* más *Chlorocardium*. Dentro del grupo *Cryptocarya*, que es el más antiguo dentro de Lauraceae, *Cryptocarya* es hermana del clado *Beilschmiedia*, es un grupo en el que se anidan otros géneros relacionados (*Endiandra* y *Syndiclis*). *Caryodaphnopsis* y *Neocinnamomum* están asociados en el presente estudio y se ha encontrado previamente que tienen una relación relativamente cercana (L. Li et al., 2016; Wang et al., 2010).

De los clados restantes, el grupo *Persea-Machilus* con *Machilus* y *Phoebe* como subconjuntos del grupo, recibió un apoyo moderado (> 50), y esto coincide con los estudios de Chanderbali et al. (2001), Li et al. (2011) y Rohwer et al. (2009). Sin embargo, al igual que con estos estudios anteriores, obtuvimos bajo apoyo estadístico en las relaciones entre especies dentro del género *Machilus*. El género *Cinnamomum* aparece muy cercanamente

relacionada con el género *Ocotea*, y en ninguno de los casos se recupera la monofilia de los géneros. De los géneros restantes se obtuvo bajo apoyo estadístico, aunque se ha reconocido una estrecha relación entre los géneros *Actinodaphne*, *Lindera*, *Litsea*, *Neolitsea* y *Laurus* en casi todas las clasificaciones de Lauraceae (L. Li et al., 2007; Z. M. Li et al., 2006; Xiao et al., 2020). Esto sugiere que los marcadores moleculares multilocus aún no brindan filogenias bien resueltas para todas las Lauraceae, y, aunque podría ser necesario más trabajo de secuenciación enfocado en resolver las relaciones filogenéticas dentro de la familia, la taxonomía también debe ser revisada.

### **Estimación de tiempos de divergencia**

En un sentido estricto, por la manera en la que aplicamos este modelo en este proyecto podemos decir que el FBD es un método de calibración de nodos en el que la ubicación de los fósiles en la filogenia no fue estimada. Esto hubiera requerido datos morfológicos para las especies fósiles y vivientes, como lo hace un análisis de evidencia total. Sin embargo, el FBD puede incorporar todos los fósiles disponibles para un clado determinado, no solo el fósil más antiguo, lo que otorga más información para estimar las tasas de especiación y extinción. Además, los enfoques de calibración convencionales requieren que las probabilidades previas para cada nodo calibrado estén parametrizadas, lo que normalmente es complicado para los biólogos. El FBD supera esta limitación al permitir la ambigüedad en la ubicación filogenética precisa de los fósiles, sin dejar de considerarlos parte del proceso macroevolutivo (Arcila et al., 2015; Heath et al., 2014).

En ambos enfoques, calibración de nodos y FBD, las estimaciones de las edades de los nodos en el modelo se basan en la distribución, el muestreo y la ubicación filogenética de

los fósiles. Sin embargo, a diferencia del enfoque de calibración de nodos, la inferencia mejora a medida que aumenta el número de fósiles en los nodos ya calibrados. Esto depende en gran medida de que los fósiles estén correctamente ubicados filogenéticamente y cubran una amplia gama de edades de nodos (Heath et al., 2014).

A pesar de las diferencias entre los métodos de fechamiento de nodos (CN y FBD), todos los análisis probados arrojaron resultados similares. Sin embargo, los tiempos de divergencia estimados para los nodos raíz y el nodo Monimiaceae + Lauraceae son significativamente más antiguos en el caso del análisis de calibración de nodos que con el FBD (aleatorio; (Fig. 5 y 6, árboles A). Esta diferencia entre los análisis podría estar influenciada por el número de fósiles incluidos. No obstante no se puede hacer una comparación estricta entre los dos métodos (CN y FBD), porque tanto las distribuciones previas y el número de los parámetros son diferentes, también las edades de los fósiles son aplicadas de manera diferente, es decir, como distribuciones previas lognormal y uniforme para la calibración secundaria del nodo raíz en el caso de CN y como un intervalo en el caso del FBD.

Además, es importante mencionar que otro aspecto importante es la manera en la que se tratan las edades de los fósiles en un análisis de FBD. Barido-Sottani et al. (2019) argumentan que los fósiles rara vez están fechados con precisión, sino que están asociados con un intervalo de edades posibles, generalmente vinculado al rango estratigráfico en el que se encontró el fósil. Este es un procedimiento que genera incertidumbre pues el vínculo entre lito-, bio- y cronoestratigrafía puede ser difícil de establecer (Benton & Donoghue, 2007; Hopkins et al., 2018; Parham et al., 2012). Hasta hace unos años, y aun actualmente, en las aplicaciones del modelo FBD se asignan edades a los fósiles utilizando el punto

medio del intervalo conocido de incertidumbre de edad o una edad aleatoria extraída de ese intervalo (rango estratigráfico). Sin embargo, Barido-Sottani et al. (2019) logran modelar explícitamente la incertidumbre de la edad de los fósiles colocando una probabilidad previa en las edades de los fósiles y al tratar las fechas de muestreo de fósiles como parámetros adicionales y co-estimándolas junto con otros parámetros del modelo.

Las edades de los fósiles obtenidas en nuestros análisis con el modelo FBD fueron tratadas de esta manera y aunque no existe un trabajo que haga una comparación directa en la manera en que se tratan las edades de los fósiles en los análisis de CN y FBD, es posible que fijar las edades de los fósiles puede dar lugar a estimaciones erróneas o poco precisas de los tiempos de divergencia. Incorporar la incertidumbre de la edad estratigráfica explícitamente aumenta sustancialmente las posibilidades de recuperar edades más precisas de los nodos.

Por otro lado, Heath et al. (2014) demostraron a través de simulaciones la precisión de las estimaciones de tiempo de divergencia al representar el ancho promedio del intervalo de credibilidad (IC) del 95 %. Sus resultados muestran que las estimaciones con el modelo FBD son menos precisas, pues presentan IC del 95% más grandes, en relación con el método convencional de calibración de nodos. Y esto justifica que la inferencia de los tiempos de divergencia con el modelo produce estimaciones conservadoras de las edades de los nodos. Sin embargo, es importante destacar que a medida que se agregan más fósiles, aumenta la precisión bajo el modelo FBD, pues los anchos promedio de los IC del 95% son más pequeños.

Este efecto es claro en nuestros resultados, los IC del 95% son, generalmente, más amplios en los nodos más jóvenes de los análisis FBD62R y FBD19R que en el caso de CN13, pero observamos que estos intervalos son ligeramente más amplios en FBD19R. Otra coincidencia con el trabajo de Heath et al. (2014) es que a medida que aumenta la edad, los intervalos de credibilidad del 95% se hacen más amplios en el análisis CN13, pero tanto en FBD62R y FBD19R el nodo raíz es mucho más amplio que en CN13. Esta tendencia probablemente esté relacionada con la interacción de la calibración secundaria de la raíz y la relativa antigüedad de los fósiles de estos grupos, lo que resulta en un intervalo de credibilidad del 95% muy amplio.

Hernández-Gutiérrez y Magallón (2019), en un estudio para estimar el tiempo de origen del orden Malvales, obtienen otro resultado interesante, y es que a medida que se muestrean más taxones existentes, los intervalos de credibilidad del 95 % se vuelven más estrechos, y esto parece ser independiente del número de fósiles incluidos, por lo tanto, la precisión de las estimaciones derivadas de FBD y CN está más influenciada por la densidad del muestreo de taxones existentes que por el muestreo de fósiles. En nuestros resultados también es posible observar esto, por ejemplo, el clado *Cryptocarya-Beilschmiedia* es el que cuenta con más especies vivientes, aunque también es el que cuenta con el mayor número de fósiles, sus intervalos de credibilidad del 95% son en promedio más estrechos en comparación con intervalos de credibilidad del 95% observados en los clados *Neocinnamomum-Caryodaphnopsis* y *Mezilaurus-Chlorocardium*, que son los clados con menor número de especies vivientes. Sin embargo, este efecto no se observa en el caso de FBD19R. Al respecto, Warnock et al. (2015) discuten que aunque hay una modesta ganancia de precisión al incrementar el número de secuencias moleculares, es decir,

taxones existentes, esta puede ser incrementada al aumentar la precisión de las calibraciones fósiles (lo que se ha discutido a lo largo de este proyecto), y una manera de lograrlo es la evaluación *a priori* de la evidencia fósil, lo que significa que es necesario asignar una restricción (conservadora) de edad mínima a la edad de un clado y minimizar la incertidumbre filogenética. Entonces, es también importante la relación que existe entre la distribución previa o *a priori* que otorgamos a las diferentes calibraciones de nuestro estudio, que, dependiendo de la información que tengamos de los diferentes fósiles, se puede asignar una densidad de probabilidad uniforme o no uniforme (por ejemplo, lognormal o exponencial). De manera general, cuando se asigna una distribución uniforme, que fue el tratamiento que le dimos a los fósiles en este proyecto, existe una probabilidad previa igual por unidad de tiempo para el evento de divergencia que abarca los límites mínimo y máximo de la información estratigráfica del fósil. De acuerdo con los resultados de Warnock et al. (2015) el uso de una distribución no uniforme provoca un aumento de la incertidumbre y conduce a un aumento en las edades estimadas, y provoca que los intervalos de credibilidad del 95% sean más difusos; mientras que el uso de una distribución uniforme tiene el efecto contrario.

### **Muestreo aleatorio versus muestreo diversificado**

Höhna et al. (2011) incorporaron el muestreo diversificado en el proceso de nacimiento-muerte al suponer que el investigador logró maximizar la diversidad, es decir, que se ha muestreado exactamente un descendiente existente para cada rama que estuvo presente en un momento específico en el pasado ( $x_{cut}$ ) dando lugar a las especies vivientes.

En este estudio seleccionamos a la familia Lauraceae por su amplio registro fósil, pero nos enfocamos en obtener la representación de más del 90% de los géneros existentes de la familia. Definimos un análisis con muestreo aleatorio cuando se incluyeron a las 314 especies vivientes que coincidieron en los mismos marcadores moleculares (*ITS*, *trnK-matK* y *rbcL*). Zhang et al. (2016) ampliaron el proceso FBD para tener en cuenta el muestreo diversificado propuesto por Höhna et al. (2011), el cual implementamos con el paquete BDSKY en BEAST2.

Con base en estudios previos sabemos que bajo el esquema de muestreo diversificado el árbol tiene ramas terminales largas, la mayoría de los tiempos de especiación se agrupan cerca de la raíz del árbol y con esto el árbol equilibra el número de descendientes vivientes a ambos lados de cada uno de los eventos de los nodos más profundos y esto es dependiente del punto en el que el investigador decide colocar el  $x_{cut}$  (Höhna et al., 2011; Luo et al., 2021). Por otro lado, si los taxones vivientes son elegidos aleatoriamente los eventos de especiación deberían distribuirse de manera más uniforme en el tiempo y la longitud de las ramas, no hay una restricción de tiempo y se están considerando todas las especies vivientes con secuencias moleculares disponibles (Höhna et al., 2011; Luo et al., 2021; Matschiner, 2019; Zhang et al., 2016).

Ramírez-Barahona et al. (2020) realizaron un análisis de datación utilizando un modelo de reloj bayesiano log-normal no correlacionado y el método de calibración de nodos para datar la filogenia de las angiospermas utilizando un conjunto de 238 fósiles. Este análisis de datación estimó que los taxones vivientes de Laurales comenzaron a diversificarse alrededor de 148.97 Ma (122.3-1772.02 Ma), entre el Jurásico medio y el Cretácico inferior y en el caso de Lauraceae fue alrededor de 83.03 Ma (56.39-105.50 Ma), entre el Cretácico

superior y el Paleoceno (Ramírez-Barahona et al., 2020). En contraste, las reconstrucciones filogenéticas obtenidas en este estudio con un muestreo diversificado y aleatorio utilizando 62 fósiles con el proceso FBD indicaron una edad de diversificación de Laurales de 138.09 y 142.45 Ma, respectivamente, y para el nodo corona de Lauraceae fue de 117.27 y 118.39 Ma, respectivamente. Debido a que utilizamos la edad de Laurales como calibración secundaria en nuestro estudio, hace sentido que las edades estimadas se encuentren dentro del intervalo de credibilidad del 95%. Sin embargo, en la edad del nodo corona Lauraceae hay un retraso de casi 40 Ma con respecto a los resultados de Ramírez-Barahona et al.(2020). Esto puede explicarse por el muestreo taxonómico, pues aumentamos el número de especies vivientes y fósiles de la familia. Como se muestra en la Tabla 1, no hay una diferencia substancial en las edades estimadas para los nodos de los principales grupos identificados en la familia entre los muestreos diversificado y aleatorio, estos son los nodos más profundos de los árboles filogenéticos estimados. Sin embargo, en las Figuras 3-6 cuando implementamos un muestreo diversificado las edades de los nodos más jóvenes se acumulan entre el Paleógeno y Neógeno. Los resultados bajo el modelo FBD de muestreo diversificado (Fig. 5 y 6) son diferentes a los esperados y a lo obtenido en trabajos previos (Matschiner, 2019; Zhang et al., 2016). Una explicación a este resultado podría ser la influencia de la edad de los fósiles utilizados, siendo la mayoría pertenecientes al Cenozoico. Sin embargo, utilizando 19 fósiles, entre ellos todos los que pertenecen al Cretácico y otros del Cenozoico, los resultados fueron similares.

Podría existir alguna razón para sospechar que las estimaciones de edad obtenidas con el modelo FBD con un muestreo diversificado pueden estar sesgadas hacia el presente debido a las imperfecciones del modelo y su implementación en el programa BEAST2. Aunque no

es el objetivo del presente proyecto, sería conveniente repetir nuestro análisis utilizando otro software, por ejemplo, MrBayes (Ronquist et al., 2012), y comparar los resultados obtenidos. López-Martínez et al. (en preparación) utilizaron el método de evidencia total implementado con el FBD (TE-FBD) que considera un muestreo diversificado para investigar la evolución de atributos morfológicos de la estructura floral de eudicotiledóneas tempranamente divergentes en un contexto temporal. Los resultados obtenidos por Loper-Martínez et al. (en preparación) coinciden con los obtenidos en el presente estudio, pues indican que las edades obtenidas con TE-FBD suponiendo un muestreo diversificado fueron ligeramente más jóvenes que las que obtuvieron con un método convencional de calibración de nodos que supone que el muestreo es aleatorio.

Por conveniencia matemática, el modelo del muestreo diversificado asume que el investigador puede encontrar la muestra que representa la cantidad máxima de diversidad filogenética dado el número elegido de puntas (Zhang et al., 2016). Sin embargo, en la práctica, es probable que nuestro muestreo esté inevitablemente sesgado, ya que no es posible pasar por alto el efecto que tienen la disponibilidad de secuencias moleculares, bases de datos morfológicos, así como la preservación y registro de los fósiles. También, es importante considerar el punto en el que se selecciona el corte ( $x_{cut}$ ) para efectuar el muestreo, podemos ignorar algunas de las ramificaciones más antiguas del árbol e incluir las más jóvenes, modificando por completo las estimaciones de edad.

Por lo tanto, aunque la estrategia de muestreo diversificado puede corregir en gran medida un muestreo taxonómico incompleto (Luo et al., 2021; Zhang et al., 2016), no puede modelar completamente los esquemas de muestreo taxonómicamente diversificados que se utilizan en la realidad.

## Perspectivas

Es importante considerar que las estimaciones de tiempo de divergencias obtenidas en este trabajo no son necesariamente precisas, por lo que estos resultados no deben ser tomados como una afirmación absoluta y final. Una de las sugerencias para mejorar aún más la datación de la familia Lauraceae sería considerar abordar el análisis en un marco de datación de evidencia total con el proceso FBD que considere un muestreo diversificado. Más datos de secuencias moleculares podrían ayudar a expandir la cantidad de terminales existentes, aumentar la precisión de las estimaciones de la longitud de las ramas del árbol filogenético y mejorar el modelado de especiación y extinción en el árbol. Además, una mejor comprensión de la evolución morfológica de la familia ayudaría a ubicar a los fósiles con mayor certidumbre en la filogenia entre los taxones vivientes y, por ende, a aumentar la precisión de los resultados. Por otro lado, realizar estos análisis en el software RevBayes y comparar los resultados con los obtenidos con BEAST2 podría ayudarnos a descartar errores en la implementación del modelo o considerar otros factores que podrían haber comprometido los resultados del presente trabajo. Otra posibilidad para evaluar la precisión y veracidad de nuestros resultados sería realizar análisis con datos simulados, considerando diferentes cantidades de especies vivientes y fósiles.

## Referencias bibliográficas.

- Álvarez, I., & Wendel, J. F. (2003). Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution*, 29(3), 417–434.  
[https://doi.org/10.1016/S1055-7903\(03\)00208-2](https://doi.org/10.1016/S1055-7903(03)00208-2)
- Arcila, D., Alexander Pyron, R., Tyler, J. C., Ortí, G., & Betancur, R. (2015). An evaluation of fossil tip-dating versus node-age calibrations in tetraodontiform fishes (Teleostei: Percomorphaceae). *Molecular Phylogenetics and Evolution*, 82(PA), 131–145. <https://doi.org/10.1016/j.ympev.2014.10.011>
- Barba-Montoya, J., dos Reis, M., Schneider, H., Donoghue, P. C. J., & Yang, Z. (2018). Constraining uncertainty in the timescale of angiosperm evolution and the veracity of a Cretaceous Terrestrial Revolution. *New Phytologist*, 218(2), 819–834.  
<https://doi.org/10.1111/nph.15011>
- Barido-Sottani, J., Aguirre-Fernández, G., Hopkins, M. J., Stadler, T., & Warnock, R. (2019). Ignoring stratigraphic age uncertainty leads to erroneous estimates of species divergence times under the fossilized birth-death process. *Proceedings of the Royal Society B: Biological Sciences*, 286(1902). <https://doi.org/10.1098/rspb.2019.0685>
- Barracough, T. G., Harvey, P. H., & Nee, S. (1996). Rate of rbcL gene sequence evolution and species diversification in flowering plants (angiosperms). *Proceedings of the Royal Society B: Biological Sciences*, 263(1370), 589–591.  
<https://doi.org/10.1098/rspb.1996.0088>
- Benton, M. J., & Donoghue, P. C. J. (2007). Paleontological evidence to date the tree of life. *Molecular Biology and Evolution*, 24(1), 26–53.  
<https://doi.org/10.1093/molbev/msl150>
- Bininda-Emonds, O. R. P., & Stamatakis, A. (2006). Taxon sampling versus computational complexity and their impact on obtaining the tree of life. *Reconstructing the Tree of Life: Taxonomy and Systematics of Species Rich Taxa*, 77–95.  
<https://doi.org/10.1201/9781420009538>
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., Suchard, M. A., Rambaut, A., & Drummond, A. J. (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Computational Biology*, 10(4), 1–6.  
<https://doi.org/10.1371/journal.pcbi.1003537>
- Bromham, L., Duchêne, S., Hua, X., Ritchie, A. M., Duchêne, D. A., & Ho, S. Y. W. (2018). Bayesian molecular dating: opening up the black box. *Biological Reviews*, 93(2), 1165–1191. <https://doi.org/10.1111/brv.12390>
- Chanderbali, A. S., van der Werff, H., & Renner, S. S. (2001). Phylogeny and Historical Biogeography of Lauraceae : Evidence from the Chloroplast and Nuclear Genomes. *Annals of the Missouri Botanical Garden*, 88(1), 104–134.

- Clegg, M. T. (1993). Chloroplast gene sequences and the study of plant evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 90(2), 363–367. <https://doi.org/10.1073/pnas.90.2.363>
- Cohen Finney, S C, Gibbard, P L & Fan, J -X, K. M. (2016). International Chronostratigraphic Chart. *The ICS International Chronostratigraphic Chart*, 36(3), 199–204. <http://www.stratigraphy.org/ICSchart/ChronostratChart2016-04.pdf>
- Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). JModelTest 2: More models, new heuristics and parallel computing. *Nature Methods*, 9(8), 772. <https://doi.org/10.1038/nmeth.2109>
- Dos Reis, M., Thawornwattana, Y., Angelis, K., Telford, M. J., Donoghue, P. C. J., & Yang, Z. (2015). Uncertainty in the Timing of Origin of Animals and the Limits of Precision in Molecular Timescales. *Current Biology*, 25(22), 2939–2950. <https://doi.org/10.1016/j.cub.2015.09.066>
- Doyle, J. (1996). Seed Plant Phylogeny and the Relationships of Gnetales. *International Journal of Plant Sciences*, 157(6).
- Drinnan, A. N., Crane, P. R., Friis, E. M., & Pedersen, K. R. (1990). Lauraceous flowers from the Potomac Group (mid-Cretaceous) of eastern North America. *Botanical Gazette*, 151(3), 370–384. <https://doi.org/10.1086/337838>
- Drummond, A. J., Ho, S. Y. W., Phillips, M. J., & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology*, 4(5), 699–710. <https://doi.org/10.1371/journal.pbio.0040088>
- Duchene, D., & Bromham, L. (2013). Rates of molecular evolution and diversification in plants: Chloroplast substitution rates correlate with species-richness in the Proteaceae. *BMC Evolutionary Biology*, 13(1). <https://doi.org/10.1186/1471-2148-13-65>
- Dunn, C. W., Hejnol, A., Matus, D. Q., Pang, K., Browne, W. E., Smith, S. A., Seaver, E., Rouse, G. W., Obst, M., Edgecombe, G. D., Sørensen, M. V., Haddock, S. H. D., Schmidt-Rhaesa, A., Okusu, A., Kristensen, R. M., Wheeler, W. C., Martindale, M. Q., & Giribet, G. (2008). Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature*, 452(7188), 745–749. <https://doi.org/10.1038/nature06614>
- Eklund, H., & Kvacek, J. (1998). LAURACEOUS AND FROM THE ( CZECH REPUBLIC , CENTRAL. 159(February), 668–686.
- Fijridiyanto, I. A., & Murakami, N. (2009). Phylogeny of Litsea and related genera (Laureae-Lauraceae) based on analysis of rpb2 gene sequences. *Journal of Plant Research*, 122(3), 283–298. <https://doi.org/10.1007/s10265-009-0218-8>
- Gavryushkina, A., Heath, T. A., Ksepka, D. T., Stadler, T., Welch, D., & Drummond, A. J. (2017). Bayesian total-evidence dating reveals the recent crown radiation of penguins. *Systematic Biology*, 66(1), 57–73. <https://doi.org/10.1093/sysbio/syw060>
- Gavryushkina, A., Welch, D., Stadler, T., & Drummond, A. J. (2014). Bayesian Inference

- of Sampled Ancestor Trees for Epidemiology and Fossil Calibration. *PLoS Computational Biology*, 10(12). <https://doi.org/10.1371/journal.pcbi.1003919>
- Gillespie, J. H. (1991). *The causes of molecular evolution (Vol. 2)*. Oxford University Press.
- Guindon, S. (2020). Rates and Rocks: Strengths and Weaknesses of Molecular Dating Methods. *Frontiers in Genetics*, 11(May). <https://doi.org/10.3389/fgene.2020.00526>
- Heath, T. A. (2012). A hierarchical Bayesian model for calibrating estimates of species divergence times. *Systematic Biology*, 61(5), 793–809. <https://doi.org/10.1093/sysbio/sys032>
- Heath, T. A. (2022). Divergence Time Estimation using BEAST v2. Dating Species Divergences with the Fossilized Birth-Death Process. *Beast Tutorial*, 1–49.
- Heath, T. A., Huelsenbeck, J. P., & Stadler, T. (2014). The fossilized birth-death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences of the United States of America*, 111(29). <https://doi.org/10.1073/pnas.1319091111>
- Hernández-Gutiérrez, R., & Magallón, S. (2019). The timing of Malvales evolution: Incorporating its extensive fossil record to inform about lineage diversification. *Molecular Phylogenetics and Evolution*, 140(August), 106606. <https://doi.org/10.1016/j.ympev.2019.106606>
- Ho, S. Y. W., & Duchêne, S. (2014). Molecular-clock methods for estimating evolutionary rates and timescales. *Molecular Ecology*, 23(24), 5947–5965. <https://doi.org/10.1111/mec.12953>
- Ho, S. Y. W., & Phillips, M. J. (2009). Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology*, 58(3), 367–380. <https://doi.org/10.1093/sysbio/syp035>
- Höhna, S. (2014). Likelihood Inference of Non-Constant Diversification Rates with Incomplete Taxon Sampling. *PloS One*, 9(1). <https://doi.org/10.1371/journal.pone.0084184>
- Höhna, S., Stadler, T., Ronquist, F., & Britton, T. (2011). Inferring speciation and extinction rates under different sampling schemes. *Molecular Biology and Evolution*, 28(9), 2577–2589. <https://doi.org/10.1093/molbev/msr095>
- Hopkins, M. J., Bapst, D. W., Simpson, C., & Warnock, R. C. M. (2018). The inseparability of sampling and time and its influence on attempts to unify the molecular and fossil records. *Paleobiology*, 44(4), 561–574. <https://doi.org/10.1017/pab.2018.27>
- Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., Ho, S. Y. W., ..., & Zhang, G. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, 346(6215), 469–478.

- Jukes, T. H., & Cantor, C. R. (1969). Evolution of protein molecules. *Mammalian Protein Metabolism* 3, 21–132.
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780. <https://doi.org/10.1093/molbev/mst010>
- Li, J., Christophel, D. C., Conran, J. G., & Li, H. W. (2004). Phylogenetic relationships within the “core” Laureae (Litsea complex, Lauraceae) inferred from sequences of the chloroplast gene matK and nuclear ribosomal DNA ITS regions. *Plant Systematics and Evolution*, 246(1–2), 19–34. <https://doi.org/10.1007/s00606-003-0113-z>
- Li, L., Li, J., Conran, J. G., & Li, X. W. (2007). Phylogeny of Neolitsea (Lauraceae) inferred from Bayesian analysis of nrDNA ITS and ETS sequences. *Plant Systematics and Evolution*, 269(3–4), 203–221. <https://doi.org/10.1007/s00606-007-0580-8>
- Li, L., Li, J., Rohwer, J. G., van der Werff, H., Wang, Z. H., & Li, H. W. (2011). Molecular phylogenetic analysis of the Persea group (Lauraceae) and its biogeographic implications on the evolution of tropical and subtropical Amphi-Pacific disjunctions. *American Journal of Botany*, 98(9), 1520–1536. <https://doi.org/10.3732/ajb.1100006>
- Li, L., Madriñán, S., & Li, J. (2016). Phylogeny and biogeography of caryodaphnopsis (Lauraceae) inferred from low-copy nuclear gene and ITS sequences. *Taxon*, 65(3), 433–443. <https://doi.org/10.12705/653.1>
- Li, Z. M., Li, J., & Li, X. W. (2006). Polyphyly of the genus Actinodaphne (Lauraceae) inferred from the analyses of nrDNA ITS and ETS sequences. *Acta Phytotaxonomica Sinica*, 44(3), 272–285. <https://doi.org/10.1360/aps040150>
- Luo, A., Zhang, C., Zhou, Q.-S., Ho, S. Y. W., & Zhu, C.-D. (2021). Impacts of Taxon-Sampling Schemes on Bayesian Molecular Dating under the Unresolved Fossilized Birth-Death Process. *BioRxiv*, 2021.11.16.468757. <https://www.biorxiv.org/content/10.1101/2021.11.16.468757v1%0Ahttps://www.biorxiv.org/content/10.1101/2021.11.16.468757v1.abstract>
- Maddison, W. P., & Maddison, D. R. (2021). *Mesquite: a modular system for evolutionary analysis*. (Version 3.70). <http://www.mesquiteproject.org>
- Magallón, S. (2020). Principles of Molecular Dating. In S. Y. W. Ho (Ed.), *The Molecular Evolutionary Clock* (pp. 67–81). Springer.
- Massoni, J., Doyle, J., & Sauquet, H. (2015). Fossil calibration of magnoliidae, an ancient lineage of angiosperms. *Palaeontologia Electronica*, 18(1), 1–25. <https://doi.org/10.26879/435>
- Matschiner, M. (2019). Selective sampling of species and fossils influences age estimates under the fossilized birth–death model. *Frontiers in Genetics*, 10(OCT), 1–10. <https://doi.org/10.3389/fgene.2019.01064>
- Mazer, S. J., & Wheelwright, N. T. (1993). Fruit size and shape: Allometry at different

- taxonomic levels in bird-dispersed plants. *Evolutionary Ecology*, 7(6), 556–575.  
<https://doi.org/10.1007/BF01237821>
- Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *Proceedings of the Gateway Computing Environments Workshop, GCE* (pp. 1–8). <https://doi.org/10.1109/GCE.2010.5676129>
- Nascimento, F. F., Reis, M. Dos, & Yang, Z. (2017). A biologist’s guide to Bayesian phylogenetic analysis. *Nature Ecology and Evolution*, 1(10), 1446–1454.  
<https://doi.org/10.1038/s41559-017-0280-x>
- Nie, Z. L., Wen, J., & Sun, H. (2007). Phylogeny and biogeography of *Sassafras* (Lauraceae) disjunct between eastern Asia and eastern North America. *Plant Systematics and Evolution*, 267(1–4), 191–203. <https://doi.org/10.1007/s00606-007-0550-1>
- O’Reilly, J. E., & Donoghue, P. C. J. (2020). The effect of fossil sampling on the estimation of divergence times with the fossilized birth–death process. *Systematic Biology*, 69(1), 124–138. <https://doi.org/10.1093/sysbio/syz037>
- Parham, J. F., Donoghue, P. C. J., Bell, C. J., Calway, T. D., Head, J. J., Holroyd, P. A., Inoue, J. G., Irmis, R. B., Joyce, W. G., Ksepka, D. T., Patané, J. S. L., Smith, N. D., Tarver, J. E., Van Tuinen, M., Yang, Z., Angielczyk, K. D., Greenwood, J. M., Hipsley, C. A., Jacobs, L., ... Benton, M. J. (2012). Best practices for justifying fossil calibrations. *Systematic Biology*, 61(2), 346–359.  
<https://doi.org/10.1093/sysbio/syr107>
- Rambaut, A. (2014). *Figtree, a graphical viewer of phylogenetic trees*.  
<http://tree.bio.ed.ac.uk/software/figtree/>
- Rambaut, A., Suchard, M. A., Xie, D., & Drummond, A. J. (2014). *Tracer v1.6*.  
<http://tree.bio.ed.ac.uk/software/tracer/>
- Ramírez-Barahona, S., Sauquet, H., & Magallón, S. (2020). The delayed and geographically heterogeneous diversification of flowering plant families. *Nature Ecology and Evolution*, 4(9), 1232–1238. <https://doi.org/10.1038/s41559-020-1241-3>
- Rannala, B., & Yang, Z. (1996). Probability Distribution of Molecular Evolutionary Trees: A New Method of Phylogenetic Inference. *Journal of Molecular Evolution*, 43(3), 304–311. <https://doi.org/10.1007/pl00006090>
- Rannala, B., & Yang, Z. (2007). Inferring speciation times under an episodic molecular clock. *Systematic Biology*, 56(3), 453–466.  
<https://doi.org/10.1080/10635150701420643>
- Renner, S. S. (1999). Circumscription and phylogeny of the laurales: Evidence from molecular and morphological data. *American Journal of Botany*, 86(9), 1301–1315.  
<https://doi.org/10.2307/2656778>
- Renner, S. S., & Chanderbali, A. S. (2000). What is the relationship among Hernandiaceae,

- Lauraceae, and Monimiaceae, and why is this question so difficult to answer? *International Journal of Plant Sciences*, 161(6 SUPPL.).  
<https://doi.org/10.1086/317574>
- Rohwer, J. G. (1993). Lauraceae. In *Flowering Plants· Dicotyledons* (pp. 366–391). Springer, Berlin, Heidelberg.
- Rohwer, J. G. (2000). Toward a Phylogenetic Classification of the Lauraceae : Evidence from matK Sequences. *American Society of Plant Taxonomists*, 25(1), 60–71.
- Rohwer, J. G., Li, J., Rudolph, B., Schmidt, S. A., van der Werff, H., & Li, H. W. (2009). Is Persea (Lauraceae) monophyletic? evidence from nuclear ribosomal ITS sequences. *Taxon*, 58(4), 1153–1167. <https://doi.org/10.1002/tax.584009>
- Rohwer, J. G., & Rudolph, B. (2005). Jumping Genera: The Phylogenetic Positions of Cassytha , Hypodaphnis , and Neocinnamomum (Lauraceae) Based on Different Analyses of trnK Intron Sequences. *Annals of the Missouri Botanical Garden*, 92(2), 153–178.
- 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. <https://doi.org/10.1093/sysbio/sys029>
- Ruhfel, B. R., Gitzendanner, M. A., Soltis, P. S., Soltis, D. E., & Burleigh, J. G. (2014). From algae to angiosperms-inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evolutionary Biology*, 14(1).  
<https://doi.org/10.1186/1471-2148-14-23>
- Sanderson, M. J. (2002). Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. *Molecular Biology and Evolution*, 19(1), 101–109. <https://doi.org/10.1093/oxfordjournals.molbev.a003974>
- Sang, T. (2002). Utility of low-copy nuclear gene sequences in plant phylogenetics. *Critical Reviews in Biochemistry and Molecular Biology*, 37(3), 121–147.  
<https://doi.org/10.1080/10409230290771474>
- Sauquet, H. (2013). A practical guide to molecular dating. *Comptes Rendus - Palevol*, 12(6), 355–367. <https://doi.org/10.1016/j.crpv.2013.07.003>
- Sauquet, H., Ho, S. Y. W., Gandolfo, M. A., Jordan, G. J., Wilf, P., Cantrill, D. J., Bayly, M. J., Bromham, L., Brown, G. K., Carpenter, R. J., Lee, D. M., Murphy, D. J., Sniderman, J. M. K., & Udovicic, F. (2012). Testing the impact of calibration on molecular divergence times using a fossil-rich group: The case of nothofagus (Fagales). *Systematic Biology*, 61(2), 289–313. <https://doi.org/10.1093/sysbio/syr116>
- Silvestro, D., Warnock, R. C. M., Gavryushkina, A., & Stadler, T. (2018). Closing the gap between palaeontological and neontological speciation and extinction rate estimates. *Nature Communications*, 9(1). <https://doi.org/10.1038/s41467-018-07622-y>

- Smith, S. A., & Walker, J. F. (2019). PyPHLAWD: A python tool for phylogenetic dataset construction. *Methods in Ecology and Evolution*, *10*(1), 104–108. <https://doi.org/10.1111/2041-210X.13096>
- Soltis, D. E., Soltis, P. S., & Zanis, M. J. (2002). Phylogeny of seed plants based on evidence from eight genes. *American Journal of Botany*, *89*(10), 1670–1681. <https://doi.org/10.3732/ajb.89.10.1670>
- Song, Y., Yu, W. Bin, Tan, Y. H., Jin, J. J., Wang, B., Yang, J. B., Liu, B., & Corlett, R. T. (2020). Plastid phylogenomics improve phylogenetic resolution in the Lauraceae. In *Journal of Systematics and Evolution* (Vol. 58, Issue 4). <https://doi.org/10.1111/jse.12536>
- Song, Y., Yu, W. Bin, Tan, Y., Liu, B., Yao, X., Jin, J., Padmanaba, M., Yang, J. B., & Corlett, R. T. (2017). Evolutionary comparisons of the chloroplast genome in Lauraceae and insights into loss events in the magnoliids. *Genome Biology and Evolution*, *9*(9), 2354–2364. <https://doi.org/10.1093/gbe/evx180>
- Stadler, T. (2010). Sampling-through-time in birth-death trees. *Journal of Theoretical Biology*, *267*(3), 396–404. <https://doi.org/10.1016/j.jtbi.2010.09.010>
- Stadler, T., Gavryushkina, A., Warnock, R. C. M., Drummond, A. J., & Heath, T. A. (2018). The fossilized birth-death model for the analysis of stratigraphic range data under different speciation modes. *Journal of Theoretical Biology*, *447*, 41–55. <https://doi.org/10.1016/j.jtbi.2018.03.005>
- Stadler, T., Kühnert, D., Bonhoeffer, S., & Drummond, A. J. (2013). Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proceedings of the National Academy of Sciences of the United States of America*, *110*(1), 228–233. <https://doi.org/10.1073/pnas.1207965110>
- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Wang, Z. hua, Li, J., Conran, J. G., & Li, H. wen. (2010). Phylogeny of the Southeast Asian endemic genus *Neocinnamomum* H. Liu (Lauraceae). *Plant Systematics and Evolution*, *290*(1), 173–184. <https://doi.org/10.1007/s00606-010-0359-1>
- Warnock, R. C. M., Parham, J. F., Joyce, W. G., Lyson, T. R., & Donoghue, P. C. J. (2015). Calibration uncertainty in molecular dating analyses: There is no substitute for the prior evaluation of time priors. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1798). <https://doi.org/10.1098/rspb.2014.1013>
- Warnock, R. C. M., & Wright, A. M. (2020). Understanding the Tripartite Approach to Bayesian Divergence Time Estimation. *Cambridge University Press*, 1–44. <https://doi.org/10.1017/9781108954365>
- Warnock, R. C. M., Yang, Z., & Donoghue, P. C. J. (2012). Exploring uncertainty in the calibration of the molecular clock. *Biology Letters*, *8*(1), 156–159.

<https://doi.org/10.1098/rsbl.2011.0710>

- Xiao, T. W., Xu, Y., Jin, L., Liu, T. J., Yan, H. F., & Ge, X. J. (2020). Conflicting phylogenetic signals in plastomes of the tribe Laureae (Lauraceae). *PeerJ*, 8, 1–23. <https://doi.org/10.7717/peerj.10155>
- Yang, Z., & Rannala, B. (1997). Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo method. *Molecular Biology and Evolution*, 14(7), 717–724. <http://mbe.oxfordjournals.org/cgi/content/abstract/14/7/717>
- Yang, Z., & Rannala, B. (2006). Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution*, 23(1), 212–226. <https://doi.org/10.1093/molbev/msj024>
- Zhang, C., Stadler, T., Klopstein, S., Heath, T. A., & Ronquist, F. (2016). Total-Evidence Dating under the Fossilized Birth-Death Process. *Systematic Biology*, 65(2), 228–249. <https://doi.org/10.1093/sysbio/syv080>

## Anexo I.

### Registro fósil

1. Especie: *Persites argutus*  
Estructura: Hoja  
Localidad: Fort Union, North Dakota, EE.UU.  
Estratigrafía: Paleoceno  
Tiempo absoluto: 55.8-50.3 Ma.  
Referencia: Hickey 1977.  
Análisis:  
Asociación: *Persea*  
Grado: 2
2. Especie: *Sassafrasoxylon gottwaldii*  
Estructura: Madera  
Localidad: López de Bertodano, Seymour Island, Antártica.  
Estratigrafía: Maastrichtiano  
Tiempo absoluto: 72.1-66 Ma.  
Referencia: Poole et al. 2000  
Análisis:  
Asociación: *Sassafras*  
Grado: 2
3. Especie: *Beilschmiedia eocenica*  
Estructura: Hoja  
Localidad: Formación Green River, Colorado, EE.UU.  
Estratigrafía: Eoceno  
Tiempo absoluto: 56-33.9 Ma.  
Referencia: MacGinitie 1969.  
Análisis:  
Asociación: *Beilschmiedia*  
Grado: 2
4. Especie: *Lindera varifolia*  
Estructura: Hoja  
Localidad: Formación Green River, Colorado, EE.UU.  
Estratigrafía: Eoceno  
Tiempo absoluto: 56-33.9 Ma.  
Referencia: MacGinitie 1969.  
Análisis:  
Asociación: *Lindera*  
Grado: 2
5. Especie: *Ocotea coloradensis*  
Estructura: Hoja  
Localidad: Formación Green River, Colorado, EE.UU.  
Estratigrafía: Eoceno  
Tiempo absoluto: 56-33.9 Ma.  
Referencia: MacGinitie 1969.  
Análisis:  
Asociación: *Ocotea*  
Grado: 2
6. Especie: *Beilschmiedioxylon africanum*  
Estructura: Madera  
Localidad: Bassin du Haut-Djerem, Tignère, Cameroun.  
Estratigrafía: Mioceno  
Tiempo absoluto: 23-5.33 Ma.  
Referencia: Dupéron-Laudoueneix & Dupéron, 2005.  
Análisis:  
Asociación: *Beilschmiedia*  
Grado: 2

7. Especie: *Paraperseoxylon septatum*  
Estructura: Madera  
Localidad: Formación Puerto Yerúa, Entre Ríos, Argentina.  
Estratigrafía: Campaniano-Maastrichtiano  
Tiempo absoluto: 83.6-66 Ma.  
Referencia: Franco et al. 2015.  
Análisis:  
Asociación: *Persea*  
Grado: 2

8. Especie: *Laurinoxylon artabeae*  
Estructura: Madera  
Localidad: Formación Ituzaingó, Entre Ríos, Argentina.  
Estratigrafía: Plioceno-Pleistoceno  
Tiempo absoluto: 5.33- 0.0117 Ma.  
Referencia: Franco 2012.  
Análisis:  
Asociación: *Ocotea*  
Grado: 2

9. Especie: *Curtiembrexylon poledrii*  
Estructura: Madera  
Localidad: Formación Puerto Yerúa, Entre Ríos, Argentina.  
Estratigrafía: Campaniano-Maastrichtiano  
Tiempo absoluto: 83.6-66 Ma.  
Referencia: Franco et al. 2015.  
Análisis:  
Asociación: *Beilschmiedia*  
Grado: 2

10. Especie: *Cryptocaryoxylon oleiferum*  
Estructura: Madera  
Localidad: Formación El Palmar, Entre Ríos, Argentina.  
Estratigrafía: Pleistoceno tardío  
Tiempo absoluto: 2.58-0.129 Ma.  
Referencia: Ramos et al. 2015.  
Análisis:  
Asociación: *Cryptocarya*  
Grado: 2

11. Especie: *Beilschmiedioxylon parataubertiana*  
Estructura: Madera  
Localidad: Formación El Palmar, Entre Ríos, Argentina.  
Estratigrafía: Pleistoceno tardío  
Tiempo absoluto: 2.58-0.129 Ma.  
Referencia: Ramos et al. 2012.  
Análisis:  
Asociación: *Beilschmiedia*  
Grado: 2

12. Especie: *Laurinoxylon atlanticum*  
Estructura: Madera  
Localidad: Formación Santa Cruz, Santa Cruz, Argentina.  
Estratigrafía: Mioceno temprano  
Tiempo absoluto: 23-5.33 Ma.  
Referencia: Brea et al. 2012.  
Análisis:  
Asociación: *Persea*  
Grado: 2

13. Especie:  
*Cryptocaryoxylon hancockii*  
Estructura: Madera  
Localidad: Formación Clarno, Oregon, EE.UU.  
Estratigrafía: Eoceno  
Tiempo absoluto: 56-33.9 Ma.  
Referencia: Wheeler & Manchester, 2002  
Análisis:  
Asociación: *Cryptocarya*  
Grado: 2

14. Especie: *Cryptocaryoxylon meeksii*  
Estructura: Madera  
Localidad: Formación Clarno, Oregon, EE.UU.  
Estratigrafía: Eoceno  
Tiempo absoluto: 56-33.9 Ma.  
Referencia: Wheeler & Manchester, 2002  
Análisis:  
Asociación: *Cryptocarya*  
Grado: 2

15. Name: *Cryptocaryoxylon radiporosum*  
Structure: Wood  
Locality: Clarno Formation, Oregon, USA.  
Stratigraphy: Eocene  
Absolute time: 56-33.9 Ma.  
Reference: Wheeler & Manchester, 2002  
Analysis:  
Assignment: *Cryptocarya*  
Score: 2

16. Especie: *Mezilaurinoxylon eiporosum*  
Estructura: Madera  
Localidad: Formación Clarno, Oregon, EE.UU.  
Estratigrafía: Eoceno  
Tiempo absoluto: 56-33.9 Ma.  
Referencia: Wheeler & Manchester, 2002  
Análisis:  
Asociación: *Mezilaurus*  
Grado: 2

17. Especie: *Ulminium magnioleiferum*  
Estructura: Madera  
Localidad: Formación Clarno, Oregon, EE.UU.  
Estratigrafía: Eoceno  
Tiempo absoluto: 56-33.9 Ma.  
Referencia: Wheeler & Manchester, 2002  
Análisis:  
Asociación: *Cryptocarya*  
Grado: 2

18. Especie: *Laurinoxylon chalatanangensis*  
Estructura: Madera  
Localidad: Formación Chalatenango, Chalatenango, El Salvador  
Estratigrafía: Oligoceno tardío – Mioceno temprano  
Tiempo absoluto: 27.82-15.97 Ma.  
Referencia: Cevallos-Ferriz et al. 2016  
Análisis:  
Asociación: *Ocotea*  
Grado: 2

19. Especie: *Laurinoxylon czechense*  
Estructura: Madera  
Localidad: Bohemian Massif, Czech Republic  
Estratigrafía: Oligoceno – Mioceno temprano  
Tiempo absoluto: 33.9-15.97 Ma.  
Referencia: Sakala et al. 2010  
Análisis:  
Asociación: *Cinnamomum*  
Grado: 2

20. Especie: *Cinnamomoxylon seemannianum*  
Estructura: Madera  
Localidad: Bohemian Massif, Czech Republic.  
Estratigrafía: Eoceno tardío – Mioceno temprano  
Tiempo absoluto: 56-15.97 Ma.  
Referencia: Koutecký & Sakala 2015  
Análisis:  
Asociación: *Cinnamomum*  
Grado: 2

21. Especie: *Ulminium mucilaginosum*  
Estructura: Madera  
Localidad: Formación El Palmar, Entre Ríos, Argentina  
Estratigrafía: Pleistoceno tardío  
Tiempo absoluto: 2.58-0.129 Ma.  
Referencia: Brea 1998  
Análisis:  
Asociación: *Ocotea*  
Grado: 2

22. Especie: *Ulminium artabaeae*  
Estructura: Madera  
Localidad: Formación El Palmar, Entre Ríos, Argentina  
Estratigrafía: Pleistoceno tardío  
Tiempo absoluto: 2.58-0.129 Ma.  
Referencia: Brea 1998  
Análisis:  
Asociación: *Nectandra*  
Grado: 2

23. Especie: *Sassafras hesperia*  
 Estructura: Hoja  
 Localidad: Klondike Mountain, Washington, USA.  
 Estratigrafía: Eoceno  
 Tiempo absoluto: 56-33.9 Ma.  
 Referencia: Wolfe & Manchester 1997  
 Análisis:  
 Asociación: *Sassafras*  
 Grado: 2
24. Especie: *Cinnamomophyllum bendirei*  
 Estructura: Hoja  
 Localidad: Formación John Day, Oregon, EE.UU.  
 Estratigrafía: Oligoceno  
 Tiempo absoluto: 33.9-23.03 Ma.  
 Referencia: Meyer & Manchester 1997  
 Análisis:  
 Asociación: *Cinnamomum*  
 Grado: 1
25. Especie: *Cinnamomophyllum knowltonii*  
 Estructura: Hoja  
 Localidad: Formación John Day, Oregon, EE.UU.  
 Estratigrafía: Oligoceno  
 Tiempo absoluto: 33.9-23.03 Ma.  
 Referencia: Meyer & Manchester 1997  
 Análisis:  
 Asociación: *Cinnamomum*  
 Grado: 1
26. Especie: *Mauldinia mirabilis*  
 Estructura: Flor  
 Localidad: Formación Patapsco, Mauldin Mountain, Maryland, EE.UU.  
 Estratigrafía: Cenomaniano  
 Tiempo absoluto: 100.5-93.9 Ma.  
 Referencia: Meyer & Manchester 1997  
 Análisis:  
 Asociación: *Lauraceae*  
 Grado: 3
27. Especie: *Neusenienia tetrasporangiata*  
 Estructura: Flor  
 Localidad: Formación Black Creek, North Carolina, EE.UU.  
 Estratigrafía: Santoniano-Campaniano  
 Tiempo absoluto: 86.3-72.1 Ma.  
 Referencia: Eklund 2000  
 Análisis:  
 Asociación: *Neocinnamomum*  
 Grado: 3
28. Especie: *Mauldinia bohémica*  
 Estructura: Flor  
 Localidad: Bohemian Massif, Czech Republic  
 Estratigrafía: Cenomaniano  
 Tiempo absoluto: 100.5-93.9 Ma.  
 Referencia: Eklund & Kvaček 1998  
 Análisis:  
 Asociación: *Lauraceae*  
 Grado: 3
29. Especie: *Mauldinia hirsuta*  
 Estructura: Flor  
 Localidad: Sarbay quarry, Rudnyy, Kazakhstan  
 Estratigrafía: Cenomaniano - Turoniano  
 Tiempo absoluto: 100.5-89.8 Ma.  
 Referencia: Frumin et al. 2004  
 Análisis:  
 Asociación: *Lauraceae*  
 Grado: 3
30. Especie: *Perseanthus crossmanensis*  
 Estructura: Flor  
 Localidad: Formación Raritan, New Jersey, EE.UU.  
 Estratigrafía: Turoniano  
 Tiempo absoluto: 93.9-89.8 Ma.  
 Referencia: Herendeen et al. 2004  
 Análisis:  
 Asociación: *Persea*  
 Grado: 3
31. Especie: *Ocotea obtusifolia*  
 Estructura: Hoja  
 Localidad: Formación Puryear, Tennessee, EE.UU.

Estratigrafía: Eoceno  
Tiempo absoluto: 56-33.9 Ma.  
Referencia: Dilcher 1963  
Análisis:  
Asociación: *Ocotea*  
Grado: 1

32. Especie: *Lauranthus futabensis*  
Estructura: Flor  
Localidad: Formación Ashizawa,  
Fukushima Prefecture, Japón  
Estratigrafía: Coniaciano temprano  
Tiempo absoluto: 89.8-86.3 Ma.  
Referencia: Takahashi et al. 2001  
Análisis:  
Asociación: *Ocotea*  
Grado: 3

33. Especie: *Laurophyllum maarensis*  
Estructura: Hoja  
Localidad: Formación Foulden Maar,  
Otago, Nueva Zelanda  
Estratigrafía: Mioceno temprano  
Tiempo absoluto: 23.03-13.82 Ma.  
Referencia: Banister et al. 2012  
Análisis:  
Asociación: *Cryptocarya*  
Grado: 3

34. Especie: *Laurophyllum  
microphyllum*  
Estructura: Hoja  
Localidad: Formación Foulden Maar,  
Otago, Nueva Zelanda  
Estratigrafía: Mioceno temprano  
Tiempo absoluto: 23.03-13.82 Ma.  
Referencia: Banister et al. 2012  
Análisis:  
Asociación: *Cryptocarya*  
Grado: 3

35. Especie: *Laurophyllum waipiata*  
Estructura: Hoja  
Localidad: Formación Foulden Maar,  
Otago, Nueva Zelanda  
Estratigrafía: Mioceno temprano  
Tiempo absoluto: 23.03-13.82 Ma.

Referencia: Banister et al. 2012  
Análisis:  
Asociación: *Cryptocarya*  
Grado: 3

36. Especie: *Laurophyllum taieriensis*  
Estructura: Hoja  
Localidad: Formación Foulden Maar,  
Otago, Nueva Zelanda  
Estratigrafía: Mioceno temprano  
Tiempo absoluto: 23.03-13.82 Ma.  
Referencia: Banister et al. 2012  
Análisis:  
Asociación: *Cryptocarya*  
Grado: 3

37. Especie: *Laurophyllum  
longfordiensis*  
Estructura: Hoja  
Localidad: Formación Foulden Maar,  
Otago, Nueva Zelanda  
Estratigrafía: Mioceno temprano  
Tiempo absoluto: 23.03-13.82 Ma.  
Referencia: Banister et al. 2012  
Análisis:  
Asociación: *Cryptocarya*  
Grado: 3

38. Especie: *Laurophyllum  
calicarioides*  
Estructura: Hoja  
Localidad: Formación Foulden Maar,  
Otago, Nueva Zelanda  
Estratigrafía: Mioceno temprano  
Tiempo absoluto: 23.03-13.82 Ma.  
Referencia: Banister et al. 2012  
Análisis:  
Asociación: *Litsea*  
Grado: 3

39. Especie: *Laurophyllum sylvestris*  
Estructura: Hoja  
Localidad: Formación Foulden Maar,  
Otago, Nueva Zelanda  
Estratigrafía: Mioceno temprano  
Tiempo absoluto: 23.03-13.82 Ma.

Referencia: Banister et al. 2012

Análisis:

Asociación: *Beilschmiedia*

Grado: 3

40. Especie: *Laurophyllum lacustris*

Estructura: Hoja

Localidad: Formación Foulden Maar,  
Otago, Nueva Zelanda

Estratigrafía: Mioceno temprano

Tiempo absoluto: 23.03-13.82 Ma.

Referencia: Banister et al. 2012

Análisis:

Asociación: *Beilschmiedia*

Grado: 3

41. Especie: *Laurophyllum vulcanicola*

Estructura: Hoja

Localidad: Formación Foulden Maar,  
Otago, Nueva Zelanda

Estratigrafía: Mioceno temprano

Tiempo absoluto: 23.03-13.82 Ma.

Referencia: Banister et al. 2012

Análisis:

Asociación: *Beilschmiedia*

Grado: 3

42. Especie: *Laurophyllum otagoensis*

Estructura: Hoja

Localidad: Formación Foulden Maar,  
Otago, Nueva Zelanda

Estratigrafía: Mioceno temprano

Tiempo absoluto: 23.03-13.82 Ma.

Referencia: Banister et al. 2012

Análisis:

Asociación: *Beilschmiedia*

Grado: 3

43. Especie: *Litseopsis nova-zelandiae*

Estructura: Hoja

Localidad: Formación Foulden Maar,  
Otago, Nueva Zelanda

Estratigrafía: Mioceno temprano

Tiempo absoluto: 23.03-13.82 Ma.

Referencia: Banister et al. 2012

Análisis:

Asociación: *Litsea*

Grado: 3

44. Especie: *Beilschmiedia tarairoides*

Estructura: Hoja

Localidad: Longford Formación,  
Murchinson, Nueva Zelanda

Estratigrafía: Mioceno tardío

Tiempo absoluto: 13.82-5.33 Ma.

Referencia: Holden 1982

Análisis:

Asociación: *Beilschmiedia*

Grado: 1

45. Especie: *Cryptocarya tutakiae*

Estructura: Hoja

Localidad: Longford Formación,  
Murchinson, Nueva Zelanda

Estratigrafía: Mioceno tardío

Tiempo absoluto: 13.82-5.33 Ma.

Referencia: Holden 1982

Análisis:

Asociación: *Cryptocarya*

Grado: 1

46. Especie: *Cryptocarya*

*murchisoniensis*

Estructura: Hoja

Localidad: Longford Formación,  
Murchinson, Nueva Zelanda

Estratigrafía: Mioceno tardío

Tiempo absoluto: 13.82-5.33 Ma.

Referencia: Holden 1982

Análisis:

Asociación: *Cryptocarya*

Grado: 1

47. Especie: *Cryptocarya longfordiensis*

Estructura: Hoja

Localidad: Longford Formación,  
Murchinson, Nueva Zelanda

Estratigrafía: Mioceno tardío

Tiempo absoluto: 13.82-5.33 Ma.

Referencia: Holden 1982

Análisis:

Asociación: *Cryptocarya*

Grado: 1

48. Especie: *Cryptocarya bulleriana*  
 Estructura: Hoja  
 Localidad: Longford Formación,  
 Murchinson, Nueva Zelanda  
 Estratigrafía: Mioceno tardío  
 Tiempo absoluto: 13.82-5.33 Ma.  
 Referencia: Holden 1982  
 Análisis:  
 Asociación: *Cryptocarya*  
 Grado: 1
49. Especie: *Cinnamomum miocenicum*  
 Estructura: Hoja  
 Localidad: Longford Formación,  
 Murchinson, Nueva Zelanda  
 Estratigrafía: Mioceno tardío  
 Tiempo absoluto: 13.82-5.33 Ma.  
 Referencia: Holden 1982  
 Análisis:  
 Asociación: *Cinnamomum*  
 Grado: 1
50. Especie: *Litsea dawsoniana*  
 Estructura: Hoja  
 Localidad: Longford Formación,  
 Murchinson, Nueva Zelanda  
 Estratigrafía: Mioceno tardío  
 Tiempo absoluto: 13.82-5.33 Ma.  
 Referencia: Holden 1982  
 Análisis:  
 Asociación: *Litsea*  
 Grado: 1
51. Especie: *Lindera clarnensis*  
 Estructura: Fruta  
 Localidad: Formación Clarno, Oregon,  
 EE.UU.  
 Estratigrafía: Eoceno  
 Tiempo absoluto: 56-33.9 Ma.  
 Referencia: Manchester 1994  
 Análisis:  
 Asociación: *Lindera*  
 Grado: 2
52. Especie: *Laurinoxylon stichkai*  
 Estructura: Madera
- Localidad: Formación Bridger, Wyoming,  
 EE.UU.  
 Estratigrafía: Eoceno  
 Tiempo absoluto: 56-33.9 Ma.  
 Referencia: Boonchai & Manchester 2012  
 Análisis:  
 Asociación: *Actinodaphne*  
 Grado: 2
53. Especie: *Bandulskaiia aestuaria*  
 Estructura: Hoja  
 Localidad: Macquarie Harbor, Tasmania,  
 Australia  
 Estratigrafía: Eoceno temprano  
 Tiempo absoluto: 56-41.2 Ma.  
 Referencia: Carpenter et al. 2007  
 Análisis:  
 Asociación: *Cryptocarya*  
 Grado: 2
54. Especie: *Laurophyllum acuminatum*  
 Estructura: Hoja  
 Localidad: Formación Nerriga, New South  
 Wales  
 Estratigrafía: Eoceno  
 Tiempo absoluto: 56-33.9 Ma.  
 Referencia: Hill 1986  
 Análisis:  
 Asociación: *Cryptocarya*  
 Grado: 1
55. Especie: *Laurophyllum sinuatum*  
 Estructura: Hoja  
 Localidad: Formación Nerriga, New South  
 Wales  
 Estratigrafía: Eoceno  
 Tiempo absoluto: 56-33.9 Ma.  
 Referencia: Hill 1986  
 Análisis:  
 Asociación: *Laurus*  
 Grado: 1
56. Especie: *Laurophyllum  
aerocryptocaryoides*  
 Estructura: Hoja  
 Localidad: Formación Nerriga, New South  
 Wales

Estratigrafía: Eoceno  
Tiempo absoluto: 56-33.9 Ma.  
Referencia: Conran & Christophel 1998  
Análisis:  
Asociación: *Cryptocarya*  
Grado: 2

57. Especie: *Cinnamomum sezannense*  
Estructura: Hoja  
Localidad: Formación Fort Union,  
Montana, EE.UU.  
Estratigrafía: Paleoceno  
Tiempo absoluto: 66-56 Ma.  
Referencia: Brown 1962  
Análisis:  
Asociación: *Cinnamomum*  
Grado: 2

58. Especie: *Marmarthia johnsonii*  
Estructura: Hoja  
Localidad: North Dakota, EE.UU.  
Estratigrafía: Maastrichtiano  
Tiempo absoluto: 72.1-66 Ma.  
Referencia: Peppe et al. 2007  
Análisis:  
Asociación: *Lindera*  
Grado: 2

59. Nombre: *Potomacanthus lobatus*  
Estructura: Flor  
Localidad: Potomac Group, Virginia,  
EE.UU.  
Estratigrafía: Albiano temprano - medio  
Tiempo absoluto: 113-100.5 Ma.  
Referencia: von Balthazar et al. 2007  
Análisis:  
Asociación: *Lauraceae*  
Grado: 3

60. Especie: *Marmarthia pearsonii*  
Estructura: Hoja

Localidad: Hell Creek, South Dakota,  
EE.UU.  
Estratigrafía: Maastrichtiano  
Tiempo absoluto: 72.1-66 Ma.  
Referencia: Johnson 2002  
Análisis:  
Asociación: *Lindera*  
Grado: 2

61. Especie: *Marmarthia trivialis*  
Estructura: Hoja  
Localidad: Hell Creek, South Dakota,  
EE.UU.  
Estratigrafía: Maastrichtiano  
Tiempo absoluto: 72.1-66 Ma.  
Referencia: Johnson 2002  
Análisis:  
Asociación: *Lindera*  
Grado: 2

62. Especie: *Mauldinia angustiloba*  
Estructura: Flor  
Localidad: Rhenish Massif, Alemania  
Estratigrafía: Cenomaniano - Turoniano  
Tiempo absoluto: 100.5-89.8 Ma.  
Referencia: Viehofen et al. 2008  
Análisis:  
Asociación: *Lauraceae*  
Grado: 3

63. Especie: *Cryptocaryoxylon*  
*radiporosum*  
Estructura: Madera  
Localidad: Formación Clarno, Oregon,  
EE.UU.  
Estratigrafía: Eoceno  
Tiempo absoluto: 56-33.9 Ma.  
Referencia: Wheeler & Manchester 2002  
Análisis:  
Asociación: *Cryptocarya*  
Grado: 2