



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN CIENCIAS BIOLÓGICAS

INSTITUTO DE BIOLOGÍA
BIOLOGÍA EVOLUTIVA

**ESTUDIO MACROEVOLUTIVO DEL ORIGEN, DIVERSIFICACIÓN Y EVOLUCIÓN FLORAL
EN MALVALES (EUDICOTYLEDONEAE), CON ÉNFASIS EN MALVACEAE *SENSU LATO***

TESIS

QUE PARA OPTAR POR EL GRADO DE:

DOCTORA EN CIENCIAS

PRESENTA:

REBECA HERNÁNDEZ GUTIÉRREZ

TUTORA PRINCIPAL DE TESIS: DRA. SUSANA AURORA MAGALLÓN PUEBLA

INSTITUTO DE BIOLOGÍA, UNAM

COMITÉ TUTOR: DRA. SONIA VÁZQUEZ SANTANA

FACULTAD DE CIENCIAS, UNAM

COMITÉ TUTOR: DR. ALFONSO OCTAVIO DELGADO SALINAS

INSTITUTO DE BIOLOGÍA, UNAM

MÉXICO, Cd. Mx.

Febrero, 2020.



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

**ESTUDIO MACROEVOLUTIVO DEL ORIGEN, DIVERSIFICACIÓN Y EVOLUCIÓN FLORAL
EN MALVALES (EUDICOTYLEDONEAE), CON ÉNFASIS EN MALVACEAE *SENSU LATO***

TESIS

QUE PARA OPTAR POR EL GRADO DE:

DOCTORA EN CIENCIAS

PRESENTA:

REBECA HERNÁNDEZ GUTIÉRREZ

TUTORA PRINCIPAL DE TESIS: DRA. SUSANA AURORA MAGALLÓN PUEBLA

INSTITUTO DE BIOLOGÍA, UNAM

COMITÉ TUTOR: DRA. SONIA VÁZQUEZ SANTANA

FACULTAD DE CIENCIAS, UNAM

COMITÉ TUTOR: DR. ALFONSO OCTAVIO DELGADO SALINAS

INSTITUTO DE BIOLOGÍA, UNAM

MÉXICO, Cd. Mx. Febrero, 2020.

COORDINACIÓN DEL POSGRADO EN CIENCIAS BIOLÓGICAS

INSTITUTO DE BIOLOGÍA

OFICIO CPCB/039/2020

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 del Subcomité por Campo de Conocimiento de Ecología, Manejo Integral de Ecosistemas, Biología Evolutiva y Sistemática del Posgrado en Ciencias Biológicas, celebrada el día 30 de septiembre de 2019, se aprobó el siguiente jurado para el examen de grado de **DOCTORA EN CIENCIAS** de la alumna **HERNÁNDEZ GUTIÉRREZ REBECA** con número de cuenta **513014648** con la tesis titulada: "**Estudio macroevolutivo del origen, diversificación y evolución floral en Malvales (Eudicotyledoneae), con énfasis en Malvaceae sensu lato**", realizada bajo la dirección de la **DRA. SUSANA AURORA MAGALLÓN PUEBLA**, quedando integrado de la siguiente manera:

Presidente:	DR. LUIS ENRIQUE EGUIARTE FRUNS
Vocal:	DRA. MARÍA DEL CARMEN MANDUJANO SÁNCHEZ
Secretario:	DR. ALFONSO OCTAVIO DELGADO SALINAS
Suplente:	DR. ALEJANDRO EMMANUEL GONZÁLEZ VOYER
Suplente	DRA. SILVANA MARTEN RODRÍGUEZ

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

ATENTAMENTE
"POR MI RAZA HABLARA EL ESPIRITU"
Cd. Universitaria, Cd. Mx., a 14 de enero de 2020.

COORDINADOR DEL PROGRAMA


DR. ADOLFO GERARDO NAVARRO SIGÜENZA

c.c.p. Expediente alumno(a).



COORDINACIÓN DEL POSGRADO EN CIENCIAS BIOLÓGICAS
UNIDAD DE POSGRADO
Edificio D, 1º Piso. Circuito de Posgrados, Ciudad Universitaria
Alcaldía Coyoacán. C. P. 04510 CDMX
Tel. (+5255)5623 7002 <http://pcbiol.posgrado.unam.mx/>

AGRADECIMIENTOS

Posgrado en Ciencias Biológicas, UNAM.

CONACyT Beca de Doctorado 507682/288658

PAPIIT UNAM IG200316

CONACyT Fronteras de la Ciencia 1602701-72

Elizabeth Bascom Fellowship, Missouri Botanical Garden, 2016

American Society of Plant Taxonomist (ASPT) Graduate Student Research Grant 2017

International Association for Plant Taxonomy (IAPT) Research Grant 2017

A mi tutora principal Dra. Susana A. Magallón Puebla

A los miembros del comité tutor, Dra. Sonia Vázquez Santana y el Dr. Alfonso Octavio Delgado Salinas

AGRADEZCO PERSONALMENTE A

Susana Magallón Puebla, por su enorme ayuda para desarrollar el proyecto y la tesis, por su extraordinario apoyo y enseñanza en todas las fases de mi doctorado, por la confianza que me tuvo para proponer y desarrollar este proyecto, por su comprensión, por su motivación constante para avanzar y crecer científicamente, y por su interés en mi formación profesional. Además, agradezco enormemente su interés y capacidad de crear un laboratorio donde se desarrollan y discuten ideas y preguntas sobre macroevolución de plantas.

Alfonso Delgado Salinas, por su extraordinaria contribución al desarrollo de mi proyecto de doctorado, por revisar y comentar la descripción de la morfología floral de Malvaceae s.l. y por las interesantes discusiones sobre flores, plantas y ciencia.

Sonia Vázquez Santana, por su interés y apoyo en este proyecto, en especial por su ayuda con la descripción de caracteres florales.

Luis Eguiarte Fruns por leer y hacer comentarios que mejoraron substancialmente el artículo de FBD de Malvales. Además, por enriquecer el contenido de la tesis con preguntas interesantes sobre evolución en plantas.

Los miembros de mi jurado de tesis, Alejandro González Voyer (I. Ecología UNAM), María del Carmen Mandujano Sánchez (I. Ecología UNAM), Silvana Martén Rodríguez (ENES Morelia UNAM), por sus comentarios y aportes para mejorar la tesis en forma y contenido, especialmente el contenido teórico.

Missouri Botanical Garden.

Royal Botanic Garden Edinburgh.

Royal Botanic Gardens, Kew.

Hong Kong Herbarium.

Herbario Nacional del Ecuador (QCNE).

Herbario Nacional de México (MEXU).

Servidor BIOLINUX Biomolecular, IB-UNAM.

Rocío González Acosta, IB-UNAM.

Arturo De Nova, Instituto de Investigación de Zonas Desérticas (UASLP).

Álvaro Campos, Estación Biológica Los Tuxtlas, IB-UNAM.

María del Rosario García Peña, MEXU.

Blanca Verónica Juárez Jaimes, MEXU.

Laura Calvillo, MEXU.

María Angélica de la Paz Ramírez Roa, MEXU.

Gilda Ortiz Calderón, MEXU.

Rafael Torres Colín, MEXU.

Alberto Reyes García, MEXU.

David Gernandt, MEXU.

Itzi Frago y Gerardo Salazar por las muestras colectadas en Ecuador y en República Dominicana.

Mario Ishiki, por muestras de *Sterculia apetala*.

Lidia Cabrera, Laboratorio de Biología Molecular de Plantas, IB-UNAM.

Laura Márquez Valdelamar, Laboratorio de Secuenciación Genómica de la Biodiversidad y la Salud (LSG), IB-UNAM.

Carolina Granados Mendoza, IB-UNAM, por su ayuda con el proyecto de genes minados y coordinación del proyecto de AHE.

Jim Solomon, Missouri Botanical Garden.

Alan Graham, Missouri Botanical Garden, por darme y comentar sus dos libros sobre paleofloras neotropicales.

Ricardo Perdiz, Instituto Nacional de Pesquisas da Amazônia (IMPA), por acceso a la base de datos de localidades en Brasil.

Cássio Van Den Berg, Universidade Estadual de Feira de Santana, Brasil, por su colaboración y apoyo extraordinario en campo y en el trabajo de laboratorio.

Alex Popovkin, por contribuir substancialmente a la colecta de *Hydrogaster trinervis*.

Toby Pennington, Royal Botanic Garden Edimburgh.

Royal Botanic Gardens, Kew: Alan Paton, Felix Forest, Olivier Murin, Bente Klitgaard, Sara Edwards.

Frank Van Caekenberghe, Botanic Garden Meise, Bélgica.

Marcia Peñafiel, Herbario Nacional del Ecuador (QCNE), INABIO.

Efraín Freire, Herbario Nacional del Ecuador (QCNE), INABIO.

David Neil, Herbario Amazónico del Ecuador (ECUAMZ).

Georgina Leite, Biblioteca IB-UNAM.

Alfredo Wong, IB-UNAM.

Maria von Balthazar, Universidad de Viena, Austria, por colaborar con entusiasmo en el proyecto de la morfología de Malvaceae s.l., por las largas discusiones para resolver la morfología de este grupo de “weirdos”, y por todo el trabajo que ha dedicado a este proyecto.

Jürg Schönenberger, Universidad de Viena, Austria, por revisar y contribuir en la descripción de los caracteres florales.

Hervé Sauquet, Royal Botanic Gardens and Domain Trust, Australia, por su ayuda para incorporar el proyecto de Malvales a Proteus.

Gerardo Contreras, por sus consejos académicos y largas pláticas motivadoras.

Itzi Fragoso y Patricia Rivera, por su amistad y apoyo en gran parte de mi posgrado.

Santiago Ramírez Barahona, IB-UNAM, por sus contribuciones en diversas fases de esta tesis.

Adriana Benítez Villaseñor, Andrea López Martínez, Luna Sánchez Reyes, por sus enormes contribuciones en diversas fases de esta tesis, y sobre todo por su linda amistad.

Brau, por su apoyo e interés en el desarrollo de mi posgrado durante los últimos seis años.

Amigas y amigos de León, Aguascalientes, CdMx, y los que están regados por todos lados, por su interés en mis proyectos y sobre todo por su amistad que siempre me mantiene a flote.

Jaqueline Armenta (Cuano), Isabel Loza, Celeste Morales, por todo su apoyo, extraordinaria amistad y hacer una familia armoniosa en el Palomar.

Mi familia, Julia, Yuz y Memo por todo su cariño, comprensión, apoyo y cuidado. Gracias por estar conmigo siempre y apoyarme incondicionalmente, gracias a eso esta tesis y todo el doctorado fue posible.

Dedico esta tesis a Julia, Yuz y Memo

ÍNDICE

RESUMEN	1
ABSTRACT	3
INTRODUCCIÓN.....	5
<i>Diversificación.....</i>	<i>5</i>
<i>El proceso de diversificación en angiospermas</i>	<i>7</i>
<i>El orden Malvales como modelo de estudio</i>	<i>9</i>
<i>Preguntas y organización de la investigación</i>	<i>12</i>
OBJETIVOS	18
CAPÍTULO I	19
RESUMEN	19
MATERIAL SUPLEMENTARIO	35
ARCHIVO SUPLEMENTARIO: LISTA DE FÓSILES.....	72
CAPÍTULO II.....	86
RESUMEN	86
MATERIALS AND METHODS.....	88
RESULTS.....	93
DISCUSSION	98
CAPÍTULO III	104
RESUMEN	104
INTRODUCTION	106
MATERIAL AND METHODS.....	109
RESULTS.....	115
DISCUSSION	127
MATERIAL SUPLEMENTARIO	140
DISCUSIÓN GENERAL.....	152
<i>Conclusiones</i>	<i>155</i>
<i>Perspectivas</i>	<i>156</i>

RESUMEN

Los estudios macroevolutivos (sobre el nivel de especie) son relevantes para contestar preguntas sobre procesos y patrones evolutivos a grandes escalas, para los cuales actualmente las filogenias moleculares son la base. A partir de una filogenia, se pueden estimar los tiempos de origen y diversificación de clados, y posteriormente investigar los procesos evolutivos que han dado lugar a la diversidad actual, cuya distribución entre clados, regiones y temporalmente es marcadamente dispar. En las angiospermas (plantas con flor) se han postulado diferentes factores que han generado la alta diversidad de especies, siendo la evolución floral uno relevante. El orden Malvales (Eudicotyledoneae) está conformado por 10 familias en su mayoría tropicales, todas con una notable diversidad floral y marcadas diferencias en su distribución geográfica y de riqueza de especies. Las relaciones filogenéticas de Malvales no están del todo resueltas, lo cual impide entender cómo ha sido su evolución. Dentro de Malvales, la familia Malvaceae *sensu lato* es la más rica en especies y en diversidad floral, particularmente del androceo (parte masculina de la flor). Malvaceae *s.l.* está conformada por nueve subfamilias que corresponden a las tradicionalmente familias Bombacaceae, Malvaceae *sensu stricto*, Sterculiaceae y Tiliaceae. Las relaciones filogenéticas entre las subfamilias no son consistentes cuando se utilizan diferentes marcadores moleculares, por lo que no se han abordado preguntas sobre la evolución de esta familia. Los objetivos de esta investigación fueron estimar el tiempo de origen y diversificación de Malvales y subclados, estimar si existen cambios en la tasa de diversificación y de evolución floral en Malvaceae *s.l.* Para la filogenia de Malvales se utilizaron ocho marcadores moleculares y se estimaron los tiempos de divergencia utilizando dos métodos alternativos de fechamiento molecular (*Node-dating* y *Fossilized Birth-Death process*), utilizando información de 80 fósiles. En particular para Malvaceae *s.l.* se realizaron dos muestreos genéticos para generar una filogenia de la familia. Por un lado, se llevó a cabo un minado de transcriptomas y genomas, a partir de cual se obtuvieron cuatro loci nucleares de una o bajo número de copias. Por otro lado, se realizó una técnica de secuenciación masiva, Enriquecimiento Híbrido Anclado (*Anchored Hybrid Enrichment*; AHE), a partir de la cual se obtuvo un muestreo de 150 loci, con los cuales se generó la filogenia base para los análisis macroevolutivos. Se estimaron las tasas de diversificación con el método de momentos y con un

método Bayesiano que calcula el número de cambios en la tasa de diversificación en una filogenia y a lo largo de las ramas (*Bayesian Analysis of Macroevolutionary Mixtures*; BAMM). Se construyó una matriz de caracteres morfológicos, en su mayoría caracteres florales y particularmente del androceo. Se hizo reconstrucción de estados ancestrales con Mapeo Estocástico y se midió la variación (heterogeneidad) en la tasa de evolución morfológica da cada carácter. Los resultados indican que el orden Malvales se originó en el Cretácico Superior, hace aproximadamente 100 millones de años, y las familias se originaron entre 90 y 56 millones de años. Se encontró que Malvaceae s.l. tienen una tasa de diversificación mayor al resto de las familias, junto con Sarcolaenaceae. Con los genes minados se obtuvo nueva evidencia de discordancia entre genes nucleares, que probablemente es la causa de la inconsistencia en las relaciones filogenéticas de la familia. La tasa de diversificación de Malvaceae s.l. es heterogénea, pues se estimaron cinco cambios a lo largo de la filogenia. Se detectaron de 5 a 27 cambios en esta tasa de evolución morfológica. En conjunto, estos resultados son un punto de partida para entender la historia evolutiva de Malvales y Malvaceae s.l. En general, la heterogeneidad en la evolución floral podría explicar la heterogeneidad en la diversificación de Malvaceae s.l.

ABSTRACT

Macroevolutionary studies (above the species level) are relevant to answer questions regarding evolutionary processes and patterns at large scales, for which molecular phylogenies are currently the basis. With a phylogeny it is possible to estimate the time of origin and diversification of clades, and subsequently to address the evolutionary processes that underlie the current biological diversity, whose distribution among clades, regions and through time is markedly uneven. Numerous factors have been proposed to have generated the high species diversity in angiosperms (flowering plants), being floral evolution a relevant one. The order Malvales (Eudicotyledoneae) comprises 10 families mostly tropical, all of them with a notable floral diversity and marked differences in geographic distribution and species richness. Phylogenetic relationships of Malvales are not completely resolved, which hinders the understanding of its evolution in general. Within Malvales, Malvaceae *s.l.* has the highest species richness and floral diversity compared to the remaining families, particularly observed in the androecium (male element of the flower). Malvaceae *s.l.* comprises nine subfamilies that correspond to the traditionally families Bombacaceae, Malvaceae *sensu stricto*, Sterculiaceae, and Tiliaceae. Phylogenetic relationships among the subfamilies are not consistent when different molecular markers are applied, therefore further research on the evolution of the group has not been performed. The objectives of this research were to estimate the origin and diversification time of Malvales and subclades, to estimate whether there are shifts in diversification and morphological evolutionary rate in Malvaceae *s.l.* For Malvales, a phylogeny from eight molecular markers was generated, and two alternative methods of molecular dating were applied (Node-dating and the Fossilized Birth-Death process), using the information of 80 fossils. Particularly for Malvaceae *s.l.* two genetic samplings were applied to build a phylogeny of the family. On one hand, a transcriptome-genome mining was carried out, from which four single- or low-copy nuclear loci were obtained. On the other hand, a method of next generation sequencing was performed (Anchored Hybrid Enrichment; AHE), from which 150 loci were obtained to build a phylogenetic tree, which was the basis for the macroevolutionary analyses. Diversification rates were estimated with the method-of-moments and the number of rate shifts was calculated with a Bayesian method that estimates diversification rate along the

phylogeny and branches (Bayesian Analysis of Macroevolutionary Mixtures; BAMM). A morphological database was built, mostly including floral characters and with special focus on those from the androecium. Ancestral state reconstructions were performed with Stochastic Character Mapping for the morphological characters and the variation (heterogeneity) in the morphological evolutionary rate was measured. Results indicate that Malvales was originated in the Upper Cretaceous, around 100 million years ago, and the families originated between 90 and 56 million years ago. It was found that Malvaceae *s.l.* has a higher diversification rate than the remaining families, jointly with Sarcolaenaceae. The mining loci gave new evidence on the discordance among nuclear genes, which probably explains the inconsistency of the relationships among the subfamilies. Diversification rate of Malvaceae *s.l.* is heterogenous with five shifts in different places of the phylogeny. Morphological evolutionary rate shifted from 5 to 27 times. Jointly, these results are a starting point towards the understanding of the evolutionary history of Malvales and Malvaceae *s.l.* In general, floral evolutionary rate could explain the heterogeneity in diversification rate of Malvaceae *s.l.*

INTRODUCCIÓN

Diversificación

La inmensa diversidad biológica actual es el producto de procesos evolutivos que involucran aspectos bióticos (p. ej., plasticidad fenotípica, competencia) y abióticos (p. ej., adaptación a variables ambientales), a escalas poblacionales (microevolución) y sobre el nivel de especie (macroevolución) (Simpson, 1953; Erwin, 2000; Ricklefs, 2004; Benton, 2009). La diversidad biológica, la cantidad de especies, no está distribuida equitativamente en clados, regiones o en el tiempo. Una manera de explicar cómo se origina y mantiene la diversidad biológica, y las causas de la heterogeneidad en su distribución (entre linajes, geográfica o temporal), ha sido a través de la macroevolución, donde se considera que la diversidad biológica es el resultado de la interacción entre el proceso de especiación (mediante el cual se forman dos especies distintas a partir de un ancestro común) y el de extinción (desaparición de especies; Nee et al., 1994). En concreto, la diferencia entre la tasa de especiación y la de extinción, origen y extinción, respectivamente, de especies por unidad de tiempo, da como resultado la tasa de diversificación neta (Sepkoski, 1998).

El punto de partida para calcular la tasa de diversificación son los estudios macroevolutivos clásicos, desarrollados a partir del registro fósil (Simpson, 1944, 1953; Gould y Eldredge, 1977; Erwin, 2000). Con la evidencia fósil, Simpson (1944, 1953) encontró que podría calcularse en qué tiempo y de qué manera se originan y se extinguen las especies y de esta manera propuso que existe una tasa de especiación y de extinción. Sin embargo, dado que el tiempo a considerar es de millones de años atrás y que el registro fósil es poco abundante para muchos grupos de organismos, se consideraron modelos matemáticos para calcular la diversificación, uno de ellos es el modelo de nacimiento y muerte (*Birth-Death*; Kendall, 1948; Nee, 2006). Este modelo involucra la relación entre la tasa de especiación y la de extinción de especies, dichas tasas siendo constantes a través del tiempo (revisión en Nee, 2006). Uno de los estimados de diversificación utilizando el modelo de nacimiento-muerte es el método de momentos de Magallón y Sanderson (2001), el cual estima la tasa neta de diversificación a partir de la edad de origen (edad troncal) o la edad de diversificación (edad corona) de los clados, y calcula la relación de la edad con la riqueza de especies. Para conocer estas edades se utilizan comúnmente las

filogenias fechadas derivadas de un modelo de reloj molecular (revisión de métodos en Rutschmann, 2006).

Con el avance de la sistemática molecular, se ha podido integrar la evidencia fósil y las filogenias moleculares de especies vivientes (Morlon et al., 2011). Las filogenias fechadas han revolucionado las hipótesis sobre la evolución de grupos de especies, pues se ha estimado en tiempo absoluto el momento en el que surgieron determinadas innovaciones, transiciones, divergencias, etc. (revisión en dos Reis et al., 2016). Aunado a esto, el uso de filogenias moleculares fechadas permite investigar los procesos que dieron lugar al surgimiento de eventos de especiación (Ricklefs, 2007).

Recientemente, se han desarrollado numerosos modelos para estimar la tasa de diversificación neta, de especiación y de extinción utilizando como base filogenias moleculares. Estos modelos suelen involucrar complejos conceptos ecológicos y evolutivos que tienen relación con la manera en que las especies se originan, se mantienen en el tiempo y en regiones, y se extinguen. Morlon (2014) revisa estos modelos propuestos, encontrando que varían dependiendo de si la diversificación es dependiente del tiempo, de la diversidad, de variables ambientales, y de caracteres morfológicos, entre otros modelos complejos que incluyen la dinámica de metacomunidades.

Con respecto a la extinción, este proceso está fuertemente influenciado por factores abióticos (climáticos), y menos por factores bióticos o ecológicos (Ezard et al., 2011). La manera en la que los factores bióticos influyen en la tasa de extinción es favoreciendo que las especies no se extingan (baja tasa de extinción) (Rabosky, 2015). Rabosky (2015) propone que la especiación debe ser estudiada, no sólo en términos de aislamiento reproductivo, sino también en términos de mantenimiento de especies a largo plazo. La estimación de la tasa a la cual las especies se extinguen se dificulta si no se toma en cuenta la información del registro fósil (Morlon et al., 2011; Silvestro et al., 2014), por lo tanto, la tasa de diversificación neta tampoco podría estimarse con precisión.

La diversificación es clave para entender los procesos que generan y mantienen la diversidad en tiempos geológicos, en áreas geográficas y en clados específicos del árbol de la vida (Morlon, 2014 y referencias ahí citadas). La filogenia de especies representa el proceso

macroevolutivo que está dado por procesos microevolutivos (p. ej., selección natural). Procesos microevolutivos pueden causar un patrón observado en análisis macroevolutivos. Sin embargo, encontrar cambios en la tasa de diversificación no significa que se conozcan las causas de dicho cambio. La tasa de diversificación puede aumentar o disminuir a través del tiempo y de los linajes de un clado (Donoghue y Sanderson, 2015). Es por ello que se han planteado diferentes hipótesis de las causas de dicha heterogeneidad para diferentes linajes. Teniendo información sobre la biología, ecología y distribución geográfica de los linajes, y al conjuntar las hipótesis filogenéticas y tiempos de divergencia, se pueden investigar las posibles causas del surgimiento de eventos de especiación y de extinción (Barraclough y Nee, 2001).

Desde hace varias décadas los estudios de evolución han descrito las diferentes formas en las que se originan las especies, de manera que existen varios mecanismos mediante los cuales puede ocurrir un evento de especiación (Mayr, 1963). La hipótesis de la especiación ecológica describe que el aislamiento reproductivo evoluciona como una consecuencia de la selección natural disruptiva de los caracteres en diferentes ambientes (Schluter, 2001). Armbruster (2017) discute el papel de la especialización en los procesos macroevolutivos de las plantas, en donde el desplazamiento de caracteres genera: barreras prezygóticas, reducción de competidores (baja extinción) y ocurrencia de especies cercanas (riqueza de especies). Entonces, en especies simpátricas el desplazamiento de caracteres tiene un papel relevante sobre la tasa de especiación y también de extinción.

El proceso de diversificación en angiospermas

La alta riqueza de especies de las angiospermas ha estado sujeta a cambios evolutivos moldeados por posibilidades y limitantes genéticas, estructurales, funcionales y ecológicas (Endress, 2011). Se ha observado que las angiospermas no tienen una tasa homogénea de diversificación, sino que ésta difiera entre clados (Magallón et al., 2018). Dichas diferencias, podrían estar relacionados con una gran variedad de adaptaciones diferentes, por ejemplo, la polinización biótica (van der Niet y Johnson, 2012), la forma de vida herbácea (Eriksson y Bremer, 1992), la simetría floral (Sargent, 2004), presencia de nectarios extraflorales (Marazzi y Sanderson, 2010), fruto carnoso

(Lagomarsino et al., 2016), o la colonización de nuevos hábitats (Moore y Donoghue, 2007). Dentro de los caracteres florales evaluados destacan los espolones con nectario (Hodges y Arnold, 1995), polinización biótica, polinización especializada (Lagomarsino et al., 2016), simetría floral (Reyes et al., 2014), polinios (Givnish et al., 2015). Adicionalmente, las diferencias en tasa de diversificación podrían deberse a diferencias ecológicas, por ejemplo, tasa generacional, adaptación a variables climáticas (p. ej., temperatura), o factores no adaptativos (Morlon, 2014).

Se ha mencionado que los caracteres morfológicos florales están sujetos a una alta presión de selección, pues cada parte, y la flor como unidad, está directamente asociada a la polinización y finalmente a la reproducción (Fenster et al. 2004; Fenster et al., 2015; pero la florivoría puede estar involucrada también, Jogesh et al., 2016). Dicha presión selectiva se deriva de la influencia que genera la disposición de alguna o varias partes florales sobre la eficiencia de la polinización (La Rosa y Conner, 2017). La Rosa y Conner (2017) en particular enfatizan la importancia que tiene la disposición del androceo, en este caso de polinios, en la eficiencia reproductiva. El intercambio de polen para la polinización debe ser eficiente para asegurar la menor pérdida posible de polen (Muchhala y Thomson, 2012). Por lo tanto, la importancia de la evolución del androceo radica en que cualquier modificación en la posición, forma, estructura o número de estambres, anteras o tecas determina el éxito reproductivo.

En cuanto al desarrollo, recientemente se propuso que este es un factor a considerar en la diversificación, pues se observa en la naturaleza una relación entre la robustez en el desarrollo y el número de especies (*developmental robustness*: Melzer y Theißen, 2016; Endress, 2011). Esta hipótesis indica que la organización conservadora en el desarrollo, por ejemplo, el plan corporal de los insectos o la organización del perianto (cáliz y corola) en las flores de ciertas angiospermas, permite que haya una organización coordinada entre todas las partes y de esta manera generar nuevos caracteres. La clave de esta hipótesis es que una vez adquirido y fijado un carácter (o conjunto de caracteres), subsecuentemente se puede originar variación y mediante selección disruptiva resulte en divergencia de especies (Melzer y Theißen, 2016).

En una revisión sobre los factores que influyen en la diversificación, Wiens (2017) menciona que los caracteres “dinámicos” tienen mayor influencia en modificar las tasas de diversificación. Los caracteres “dinámicos”, explica Wiens, en realidad no son un carácter sino la tasa a la que

evoluciona. Es decir, la presencia o ausencia de un estado de carácter no está correlacionada con la diversificación a gran escala, en lugar de eso, es la tasa a la que evoluciona dicho carácter lo que está correlacionado con la tasa de diversificación. Adicionalmente, Onstein (2019) propone que la “evolvabilidad” de caracteres morfológicos a escala macroevolutiva puede ser la causa de la alta diversificación en las angiospermas y de clados internos. De acuerdo con esta propuesta, las modificaciones discretas de caracteres estables (con desarrollo robusto; Melzer y Theißen, 2016) podrían influir en la diversificación como caracteres “dinámicos”, es decir, es probable que la tasa a la que estos pequeños cambios surgen sea la causa de los cambios en las tasas de diversificación. La familia Proteaceae tiene simetría floral variable y se ha demostrado que está asociada con aumento en la diversificación en este grupo de plantas (Reyes et al., 2014; O’Meara et al., 2016), a pesar de tener un plan organizacional conservado (Reyes, 2014). Este es un indicador de que los caracteres que están conservados en su organización o conformación pueden tener una correlación con los eventos de especiación mediante variaciones que no afectan estructuralmente el atributo, y la tasa a la que evolucionan puede ser la causa del cambio en la tasa de diversificación observada a gran escala (macroevolución).

El orden Malvales como modelo de estudio

Un grupo de plantas relevante para estudiar procesos macroevolutivos es el orden Malvales. Este grupo es interesante porque está conformado por una alta diversidad de especies que son componentes importantes de diversos biomas, particularmente de los bosques tropicales lluviosos y los bosques tropicales estacionalmente secos, comprende una alta diversidad de morfologías, es cosmopolita, y muchas especies de Malvales son de interés económico y cultural a nivel mundial. El orden está formado por 10 familias, ca. 333 géneros y alrededor de 6,200 especies (The Plant List, 2017). El orden incluye en su mayoría árboles, sin embargo, también hay especies de arbustos, hierbas y enredaderas. Los caracteres distintivos de los miembros del orden son la presencia de células de mucílago, tricomas estrellados, floema estratificado en forma de cuña, margen de las hojas con dientes vascularizados (i.e., dientes malvoides) y aceites ciclopropenoides en las semillas (Le Péchon y Gigord, 2014). En general, en la mayoría de las

familias existe un patrón de polistemonía en los órganos del androceo (órganos masculinos; Fig. 1), es decir, la multiplicación en el número de los estambres con respecto del patrón general de las eudicotiledóneas del clado Pentapetales (i.e., androceo conformado por cinco estambres o múltiplos de cinco; von Balthazar et al., 2004). Sin embargo, la diversidad del androceo en el orden es considerablemente alta, pues cada familia tiene una conformación de estambres y anteras particular (Bayer et al., 1999) y se ha dicho que el androceo en Malvales es un punto caliente (*hot spot*) de eventos evolutivos (Endress, 2010).

A pesar de su importancia, no existe resolución filogenética consistente entre las familias que conforman el grupo, lo cual impide la investigación sobre la evolución de este orden de plantas. Específicamente, es interesante resolver el momento en el que se originaron las familias de Malvales, pues la distribución geográfica de algunas de ellas es relevante para la conformación de biomas y tipos de vegetación tropicales. Por ejemplo, los bosques de dipterocarpos son un bioma restringido de Asia y conocer su origen es relevante para entender la evolución en general de esta vegetación. Asimismo, las familias Sphaerosepalaceae y Sarcolaenaceae son endémicas a Madagascar, de las cuales se desconoce si se originaron en la isla o si migraron de África continental.

Malvaceae *sensu lato* es la familia más rica en especies del orden Malvales. Es un grupo conformado por nueve subfamilias (Malvoideae, Tilioideae, Bombacoideae, Sterculioideae, Brownlowioideae, Grewioideae, Dombeyoideae, Byttnerioideae y Helicteroideae) y en total incluye alrededor de 4,550 especies (The Plant List, 2017). Los caracteres que apoyan la incorporación en una sola familia a las tradicionalmente familias Bombacaceae, Malvaceae *sensu stricto*, Sterculiaceae y Tiliaceae, son los nectarios constituidos por tricomas glandulares localizados en la base del cáliz y el epicáliz (Fig. 1G), una estructura formada por brácteas rodeando el cáliz (Judd y Manchester, 1997; Bayer et al., 1999; Alverson et al., 1999). Una característica particular de la familia es la alta variedad de modificaciones del androceo (Fig. 1). Por ejemplo, las anteras monotecas, que se encuentran en dos subfamilias (Malvoideae y Bombacoideae), son el resultado de múltiples divisiones de los primordios de los estambres, pero esta característica que se ha desarrollado de manera independiente en los dos grupos (von Balthazar et al., 2004; von Balthazar et al., 2006; Janka et al., 2008). Estos estudios de desarrollo

floral han descrito parte de la evolución del androceo en esta familia. Sin embargo, permanece incompleto el conocimiento que se tiene sobre lo que ha ocurrido en las otras subfamilias, en las que también se observan modificaciones interesantes.

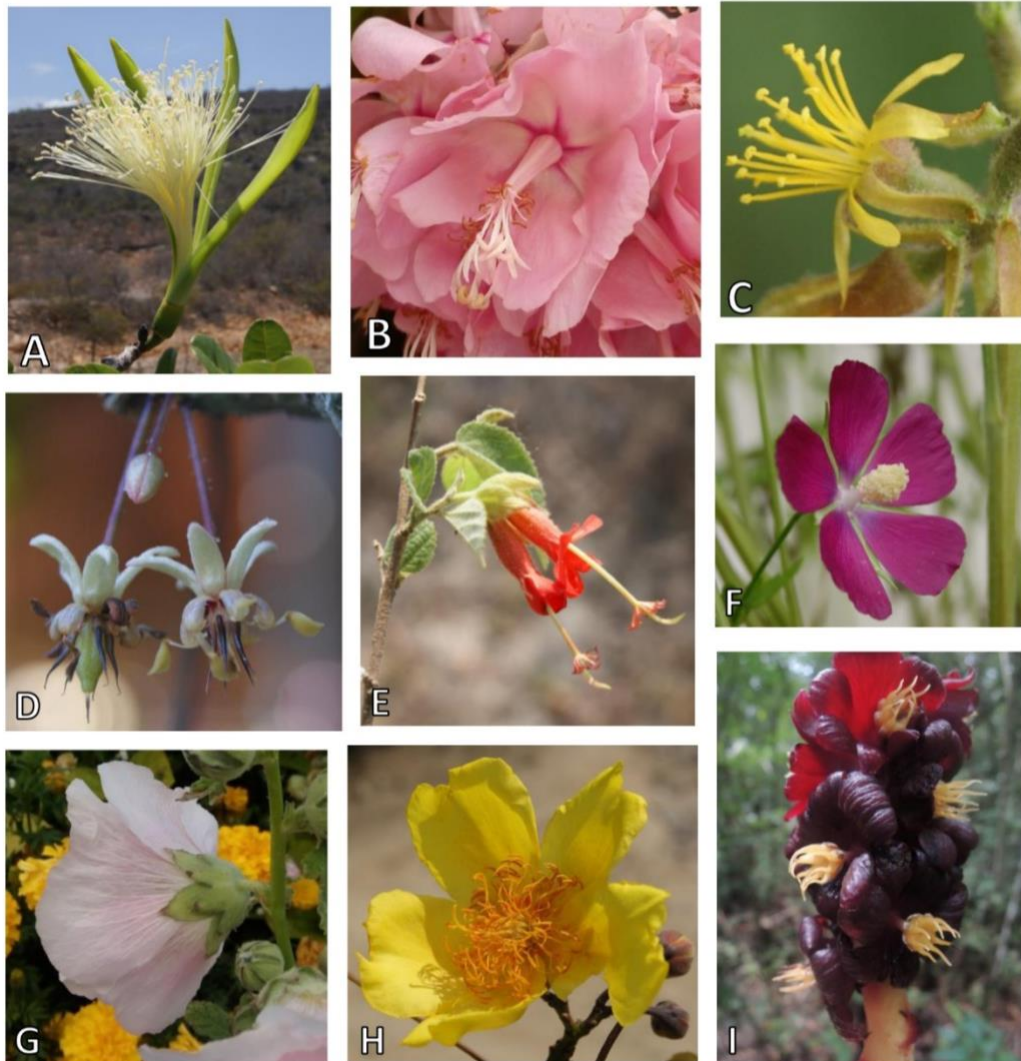


Figura 1. Diversidad floral de Malvales. A: *Pachira* sp. (Bombacoideae, Malvaceae); B: *Dombeya wallichii* (Dombeyoideae, Malvaceae); C: *Grewia* sp. (Grewioideae, Malvaceae); D: *Theobroma cacao* (Byttnerioideae, Malvaceae); E: *Helicteres guazumifolia* (Helicterioideae, Malvaceae); F: *Callirhoe bushii* (Malvoideae, Malvaceae); G: Epicáliz (*Alcea rosea*; Malvoideae); H: *Cochlospermum vitifolium* (Bixaceae); I: *Bdallophytum americanum* (Cytinaceae). Fotografías: RHG.

En un estudio sobre la diversificación en las angiospermas (Magallón et al., 2018), se detectó un aumento en la tasa de diversificación en Malvaceae s.l., lo cual podría estar relacionado con la capacidad que tienen estas especies de modificar la morfología floral. Dado que el androceo es la parte floral que más varía en Malvaceae s.l., una hipótesis es que el androceo tiene una alta tasa de evolución, potencialmente sería un carácter dinámico (*sensu* Wiens, 2017), y que esta “evolvabilidad” (Onstein, 2019) haya facilitado la diversificación de especies en esta familia. Las causas microevolutivas de dicha tasa de evolución pueden ser diversas, pero se ha sugerido que la modificación en la conformación de las anteras en Malvaceae s.l. está relacionada con la polinización (Janka et al., 2008). Siendo así que probablemente exista una relación entre la modificación del androceo, derivada de una adaptación a la polinización a nivel microevolutivo, y los cambios potenciales en las tasas de diversificación estimados a escala macroevolutiva.

Preguntas y organización de la investigación

La motivación de esta tesis fue contestar las siguientes preguntas generales ¿cuándo se originó y diversificó el orden Malvales y cómo ha sido su dinámica de diversificación? Dado que la familia Malvaceae s.l. es la más diversa en especies, distribución y morfología, las siguientes preguntas están relacionadas a esta familia: ¿ha habido cambios en la tasa de diversificación de Malvaceae s.l.?, ¿cómo ha evolucionado la flor en la familia Malvaceae s.l.? ¿Ha habido cambios en la tasa de evolución morfológica floral? Si es así, ¿los cambios en diversificación están relacionados con la evolución floral, en especial con el androceo? Para abordar estas preguntas se llevaron a cabo tres investigaciones individuales que corresponden a los tres capítulos que conforman esta tesis y cuyas preguntas particulares se describen en cada uno de ellos.

El primer capítulo trata del origen y diversificación de las familias de Malvales, incluyendo una evaluación de dos métodos de fechamiento molecular (*Node-dating* y *Fossilized Birth-Death process*; Hernández-Gutiérrez y Magallón, 2019). El segundo capítulo tiene como objetivo conocer las relaciones internas de Malvaceae s.l. a partir de genes de una sola copia obtenidos del minado de varios transcriptomas y tres genomas (manuscrito en preparación para *Systematic Botany*). Finalmente, el tercer capítulo aborda la pregunta de cuál es la dinámica de diversificación

de las malváceas y cómo ha sido la evolución floral a lo largo del tiempo, y si estas dinámicas están relacionadas entre sí (manuscrito en preparación para *American Journal of Botany*).

Literatura citada

- Alverson W, Whitlock B, Nyffeler R, Bayer C, Baum D. 1999. Phylogeny of the core Malvales: evidence from *ndhF* sequence data. *American Journal of Botany* 86(10): 147-1486.
- Armbruster W.S. 2017. The specialization continuum in pollination systems: diversity of concepts and implications for ecology, evolution and conservation. *Functional Ecology* 31: 88-100.
- Barracough TG y Nee S. 2001. Phylogenetics and speciation. *Trends in Ecology & Evolution* 16(7): 391-399.
- Bayer C, Fay MF, De Bruijn AY, Savolainen V, Morton CM, Kubitzki K, Alverson WS, Chase MW. 1999. Support for an expanded family concept of Malvaceae within a recircumscribed order Malvales: a combined analysis of plastid *atpB* and *rbcl* DNA sequences. *Botanical Journal of the Linnean Society* 129:267-303.
- Benton MJ. 2009. The Red Queen and the Court Jester: Species diversity and the role of biotic and abiotic factors through time. *Science* 323: 728-732.
- Donoghue MJ y Sanderson MJ. 2015. Confluence, synnovation, and depauperons in plant diversification. *New Phytologist* 207: 206-274.
- dos Reis M, Donoghue PCJ, Yang Z. 2016. Bayesian molecular clock dating of species divergences in the genomics era. *Nature Reviews Genetics* 17: 71-80.
- Endress PK. 2010. Flower structure and trends of evolution in eudicots and their major subclades. *Annals of the Missouri Botanical Garden* 97: 541-583.
- Endress PK. 2011. Evolutionary diversification of the flowers in angiosperms. *American Journal of Botany* 98(2): 370-396.
- Eriksson O y Bremer B. 1992. Pollination systems, dispersal modes, life forms, and diversification rates in angiosperm families. *Evolution* 46(1): 258-266.
- Erwin DH. 2000. Macroevolution is more than repeated rounds of microevolution. *Evolution & Development* 2(2): 78-84.

- Ezard TH, Aze T, Pearson PN, Purvis A. 2011. Interplay between changing climate and species' exology drives macroevolutionary dynamics. *Science* 332: 349-351.
- Fenster CB, Armbruster WS, Wilson P, Dudash MR, Thompson JD. 2004. Pollination syndromes and floral specialization. *Annual Review of Ecology and Systematics*. 35:375-403.
- Fenster CB, Reynolds RJ, Williams CW, Makowsky R, Dudash MR. 2015. Quantifying hummingbird preference for floral trait combinations: The role of selection on trait interactions in the evolution of pollination syndromes. *Evolution* 69(5): 1113-1127.
- Givnish TJ, Spalink D, Ames M, Lyon SP, Hunter SJ, Zuluaga A, Iles WJD., Clements MA, Arroyo MTK., Leebens-Mack J, Endara L, Kriebel R, Neubig KM, Whitten M, Williams NH, Cameron KM. 2015. Orchid phylogenomics and multiple drivers of their extraordinary diversification. *Proceedings of the Royal Society B* 282: 20151553.
- Gould SJ y Eldredge N. 1977. Punctuated Equilibria: The tempo and mode of evolution reconsidered. *Paleobiology* 3(2): 115-151.
- Hernández-Gutiérrez R y Magallón S. 2019. The timing of Malvales evolution: incorporating its extensive fossil record to inform about lineage diversification. *Molecular Phylogenetics and Evolution* 140: 106606.
- Hodges SA y Arnold ML. 1995. Spurring plant diversification: Are floral nectar spurs a key innovation? *Proceedings of the Royal Society B* 262: 343-348.
- Janka H, von Balthazar M, Alverson W, Baum D, Semir J, Bayer C. 2008. Structure, development and evolution of the androecium in *Adansonieae* (core *Bombacoideae*, *Malvaceae* s. l.). *Plant Systematics and Evolution* 275: 69-91.
- Jogesh T, Overson RP, Raguso RA, Skogen KA. 2016. Herbivory as an important selective force in the evolution of floral traits and pollinator shifts. *AoB PLANTS* plw088. DOI: <https://doi.org/10.1093/aobpla/plw088>.
- Judd W y Manchester S. 1997. Circumscription of *Malvaceae* (Malvales) as determined by a preliminary cladistic analysis of morphological, anatomical, palynological, and chemical characters. *Brittonia* 49:384405.
- Kendall DG. 1948. On the generalized "birth-and-death" process. *Annals of Mathematical Statistics* 19: 1-15.

- Lagomarsino LP, Condamine FL, Antonelli A, Mulch A, Davis CC. 2016. The abiotic and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae). *New Phytologist* 210: 1430-1442.
- La Rosa RJ y Conner JK. 2017. Floral function: effects of traits on pollinators, male and female pollination success, and female fitness across three species of milkweeds (*Asclepias*). *American Journal of Botany* 104(1): 150-160.
- Le Péchon P. y Gigord LDB. 2014. On the Relevance of Molecular Tools for Taxonomic Revision in Malvales, Malvaceae s.l., and Dombeyoideae. Chapt. 7. En *Methods in molecular Biology*. Ed. Besse, P. Springer, New York.
- Magallón S y Sanderson MJ. 2001. Absolute diversification rates in angiosperm clades. *Evolution* 55(9): 1762-1780.
- Magallón S, Sánchez-Reyes LL, Gómez-Acevedo S. 2018. Thirty clues to the exceptional diversification of flowering plants. *Annals of Botany*, mcy182
<http://doi.org/10.1093/aob/mcy182>.
- Marazzi B y Sanderson MJ. 2010. Large-scale patterns of diversification in the widespread legume genus *Senna* and the evolutionary role of extrafloral nectaries. *Evolution* 64: 3570-3592.
- Mayr E. 1963. *Populations, species, and evolution*. The Belknap Press of Harvard University Press. EUA.
- Melzer R y Theißen G. 2016. The significance of developmental robustness for species diversity. *Annals of Botany* 117(5): 725-732.
- Moore BR y Donoghue MJ. 2007. Correlates of diversification in the plant clade Dipsacales: geographic movement and evolutionary innovations. *The American Naturalist* 170(S2): 28-55.
- Morlon H, Parsons TL, Plotkin JB. 2011. Reconciling molecular phylogenies with the fossil record. *Proceedings of the National Academy of Sciences* 108(39): 16327-16332.
- Morlon H. 2014. Phylogenetic approaches for studying diversification. *Ecology Letters* 17: 508-525.
- Muchhala N y Thomson JD. 2012. Interspecific competition in pollination systems: costs to male fitness via pollen misplacement. *Functional Ecology* 26 (2): 476-482.

- Nee S, May RM, Harvey PH. 1994. The reconstructed evolutionary process. *Philosophical transactions of the Royal Society B* 344: 305-311.
- Nee S. 2006. Birth-Death models in Macroevolution. *Annual Review of Ecology, Evolution, and Systematics* 37: 1-17.
- O'Meara B, Smith SD, Armbruster WS, Lawrence DH, Hardy CR, Hileman LC, Hufford L, Litt A, Magallón S, Smith SA, Stevens PF, Fenster CB, Diggle PK. 2016. Non-equilibrium dynamics and floral trait interactions shape extant angiosperm diversity. *Proceedings of the Royal Society B* 283: 20152304.
- Onstein RE. 2019. Darwin's second 'abominable mystery': trait flexibility as the innovation leading to angiosperm diversity. *New Phytologist* doi:10.1111/nph.16294.
- Rabosky DL. 2015. Reproductive isolation and the causes of speciation rate variation in nature. *Biological Journal of the Linnean Society* 118: 13-25.
- Reyes E, Morlon H, Sauquet H. 2014. Presence in Mediterranean hotspots and floral symmetry affect speciation and extinction rates in Proteaceae. *New Phytologist* 207: 401-410.
- Ricklefs RE. 2004. A comprehensive framework for global patterns in biodiversity. *Ecology Letters* 7: 1-15.
- Ricklefs RE. 2007. Estimating diversification rates from phylogenetic information. *Trends in Ecology and Evolution* 22(1): 601-610.
- Rutschmann F. 2006. Molecular dating of phylogenetic trees: A brief review of current methods that estimate divergence times. *Diversity and Distributions* 12: 35-48.
- Sargent RD. 2004. Floral symmetry affects speciation rates in angiosperms. *Proceedings of the Royal Society of London B* 271: 603-608.
- Schluter D. 2001. Ecology and the origin of species. *Trends in Ecology & Evolution* 16 (7): 372-380
- Sepkoski JJ Jr. 1998. Rates of speciation in the fossil record. *Philosophical transactions of the Royal Society B* 353: 315-326.
- Silvestro D, Schnitzler J, Liow LH, Antonelli A, Salamin N. 2014. Bayesian estimation of speciation and extinction from incomplete fossil occurrence data. *Systematic Biology* 63: 349-367.
- Simpson GG. 1944. *Tempo and mode in evolution*. Columbia University Press, New York.
- Simpson GG. 1953. *The major features of evolution*. Columbia University Press, New York.

- van der Niet T y Johnson SD. 2012. Phylogenetic evidence for pollinator-driven diversification of angiosperms. *Trends in Ecology and Evolution* 27(6): 353-361.
- von Balthazar M, Alverson WS, Schönenberger J, Baum DA. 2004. Comparative floral development and androecium structure in Malvoideae (Malvaceae *s.l.*). *International Journal of Plant Sciences* 165(4):445-473.
- von Balthazar M, Schönenberger J, Alverson WS, Janka H, Bayer C, Baum DA. 2006. Structure and evolution of the androecium in the Malvatheca clade (Malvaceae *s.l.*) and implications for Malvaceae and Malvales. *Plant Systematics and Evolution* 260:171-197.
- Wiens JJ. 2017. What explains patterns of biodiversity across the Tree of Life? *Bioessays* 38: 1600128.

OBJETIVOS

General

Examinar los procesos macroevolutivos que han originado y mantenido la diversidad de especies y morfológica actual del orden Malvales.

Particulares

1. Estimar el tiempo de origen y de diversificación del orden Malvales y examinar la dinámica de diversificación de cada familia.
2. Estimar la dinámica de diversificación de la familia Malvaceae *s.l.*, medir la heterogeneidad en la tasa de evolución de caracteres morfológicos florales, con énfasis en la estructura del androceo.
 - a. Generar una hipótesis filogenética de Malvaceae *s.l.* con genes nucleares de bajo número de copias.
 - b. Estimar los tiempos de origen y diversificación, y evaluar si hay cambios en la tasa de diversificación de Malvaceae *s.l.*
 - c. Trazar la evolución floral, especialmente del androceo, de Malvaceae *s.l.* y sus subfamilias.

CAPÍTULO I

THE TIMING OF MALVALES EVOLUTION: INCORPORATING ITS EXTENSIVE FOSSIL RECORD TO INFORM ABOUT LINEAGE DIVERSIFICATION

Artículo publicado: 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: 106606.

Resumen

El registro fósil es una importante fuente de información la evolución de linajes, especialmente cuando se combina con hipótesis filogenéticas. En lugar de sólo ser calibraciones en nodos internos, como en métodos de *node-dating* (ND), el incluir fósiles como parte del mismo proceso macroevolutivo que ha generado a las especies vivientes, implementado en el *Fossilized Birth-Death process* (FBD), representa un avance substancial para la estimar tiempos de divergencia en grupos con un rico registro fósil. El orden Malvales (angiospermas eudicotiledóneas) comprende 10 familias, incluyendo especies icónicas como el algodón, el cacao, y los dipterocarpos. El orden completo abarca aproximadamente seis mil especies y un número substancial de fósiles han sido descritos para la mayoría de las familias. El conocimiento que se tiene sobre la evolución del orden ha estado limitado por la falta de resolución en las relaciones filogenéticas entre las familias. El objetivo de este estudio fue generar una hipótesis filogenética robusta con todas las familias del orden para estimar el tiempo de origen y de diversificación de sus linajes mayores y evaluar su dinámica de diversificación. Se ensambló una base de datos de fósiles de Malvales, la cual se utilizó para evaluar el efecto del número de fósiles incluido en los análisis de FBD. Se construyó con “tubería” bioinformática una base de datos molecular de ocho loci y un muestreo taxonómico de 278 taxa. Se estimaron los tiempos de divergencia de Malvales mediante ND y FBD; con este último método se evaluaron cinco conjuntos de datos que difieren en el número de fósiles incluidos (14, 50 y 80) y la certidumbre en la asignación taxonómica de cada fósil. Las edades obtenidas fueron utilizadas para estimar la tasa neta de diversificación con el método de momentos. Los resultados muestran altos

valores de apoyo en la mayoría de las relaciones internas del orden. Los tiempos de divergencia derivados de los diferentes conjuntos de datos son similares en su mayoría. La distancia entre la edad troncal y la corona tendió a ser mayor en los grupos pobremente muestreados y para estos grupos los intervalos de confianza fueron más amplios. Este estudio muestra que todos los linajes mayores de Malvales se originaron durante el Cretácico Superior (hace 100-66 millones de años) pero diversificaron en la riqueza de especies vivientes en diferentes momentos durante el Terciario. Para el Mioceno Temprano (hace 20.4 millones de años) todos los linajes mayores se habían diversificado. Los estimados de tasa neta de diversificación muestran que la mayor tasa fue para Malvaceae s.l. y Sarcolaenaceae, mientras que la menor tasa fue en Cytinaceae.



The timing of Malvales evolution: Incorporating its extensive fossil record to inform about lineage diversification

Rebeca Hernández-Gutiérrez^{a,*}, Susana Magallón^b

^a Posgrado en Ciencias Biológicas, Instituto de Biología, Universidad Nacional Autónoma de México, 3er Circuito de Ciudad Universitaria, Del. Coyoacán, Ciudad de México 04510, México

^b Departamento de Botánica, Instituto de Biología, Universidad Nacional Autónoma de México, 3er Circuito de Ciudad Universitaria, Del. Coyoacán, Ciudad de México 04510, México

1. Introduction

Species of Malvales are significant components of tropical biomes around the world (Oliveira-Filho, 2017; Sosef et al., 2017; Webb et al., 2008). For example, species of Dipterocarpaceae are primary structural elements in southeast Asian wet tropical forests, and other families include distinctive elements of wet and dry tropical forests, in particular, seasonally-dry tropical forests (e.g. Malvaceae s.l., Sarcolaenaceae; Pennington et al., 2009). Malvales also includes numerous species of economic, cultural and ecological importance such as cotton, *Gossypium hirsutum*; cocoa, *Theobroma cacao*; achiote or annatto, *Bixa orellana*; durian, *Durio zibethinus*. The evolution of lineages within Malvales has been studied for particular families and subclades (e.g., Thymelaeaceae: Van der Bank et al., 2002; Cistaceae: Guzmán and Vargas, 2009; Grewoioideae [Malvaceae s.l.]: Brunk and Muellner, 2012; Dipterocarpaceae: Heckenhauser et al., 2017), but the evolutionary history of the order is still poorly understood and the phylogenetic relationships within it are not well resolved (Le Pechón and Gigord, 2014).

Seminal systematic work conducted two decades ago marked the current circumscription of Malvales (e.g., Alverson et al., 1998; Bayer et al., 1999; Baum et al., 2004). The order is composed of 10 families: Bixaceae (21 spp.), Cistaceae (207 spp.), Cyrtaceae (10 spp.), Dipterocarpaceae (680 spp.), Malvaceae s.l. (4225 spp.), Muntingiaceae (3 spp.), Neuradaceae (10 spp.), Sarcolaenaceae (79 spp.), Sphaerosepalaceae (18 spp.), and Thymelaeaceae (891 spp.). Shared morphological characters among these families include the presence of mucilage cells and ducts, stellate hairs, valvate calyx aestivation, and cyclopropanoid seed oils (Kubitzki and Chase, 2003). Primarily, the family includes tropical, arboreal species, however, a great range of growth forms are represented (e.g., herbs, shrubs, vines), and some members are distributed in temperate regions (e.g., Cistaceae and subfamily Tilioideae [Malvaceae s.l.]). The 10 families are well supported as distinct monophyletic units, as well as the relationships among some of them, for example, Cistaceae as sister to

Dipterocarpaceae and Sarcolaenaceae. In contrast, other relationships are not well understood, for example the position of Malvaceae s.l. and Thymelaeaceae relative to the rest of the families (Le Pechón and Gigord, 2014).

Overall, the evolutionary history of Malvales is still poorly understood, in part because molecular dating analyses have included small taxonomic samples, with an information gap between broad-scale and fine-scale (e.g., genus-level) dated phylogenies, and also the use of few fossils as calibrations. Nevertheless, the age of Malvales and of some of its clades has been discussed, for example the origin of Dipterocarpaceae in the Coniacian (88 Million years ago [Ma], Upper Cretaceous; Ducousso et al., 2004), or estimated, for example the onset of Cistaceae diversification (mean crown age) in the late Eocene (22.6 Ma; Aparicio et al., 2017). In addition, the divergence time estimation for Malvales as a whole has been conducted in angiosperm-wide studies (e.g. Wikström et al., 2001; Tank et al., 2015; Magallón et al., 2015).

The order Malvales has an abundant fossil record, and in particular, of the ten extant families, seven are securely represented. This dense fossil representation makes Malvales a suitable system to explore different dating methods that utilize the fossil record in fundamentally different ways. Most of the fossil record of Malvales has been collected in North America and Asia, followed by Europe and southern South America. Consideration of the evolutionary history of the group, documented by the fossil record, has an enormous potential to substantially improve our understanding of many aspects of evolutionary history, including the timing of origin and diversification of clades (e.g., Magallón et al., 2015), its morphological evolution (Slater et al., 2012), and its diversification dynamics (Silvestro et al., 2015). The absence of a solid framework of the timing of Malvalean diversification precludes the explicit investigation of the evolutionary history of the group. Hence, a first step towards a better understanding of the evolution of Malvales is to obtain a phylogenetic dated tree of the order.

Current phylogenetic studies allow the recognition of major lineages

Abbreviations: Ma, million years ago; Myr, million years

* Corresponding author.

E-mail addresses: rebecahdezgtz@gmail.com (R. Hernández-Gutiérrez), s.magallon@ib.unam.mx (S. Magallón).

<https://doi.org/10.1016/j.ympev.2019.106606>

Received 18 February 2019; Received in revised form 27 August 2019; Accepted 27 August 2019

Available online 28 August 2019

1055-7903/© 2019 Elsevier Inc. All rights reserved.

of Malvales and provide critical guidelines for more detailed analyses with an adequate taxonomic representation. The wealth of molecular sequence data can be analyzed in conjunction with the abundant and well-documented fossil record within the framework of methods such as the Fossilized Birth-Death process (FBD; Heath et al., 2014). This approach represents a novel and promising avenue to investigate macro-evolutionary processes in the order. The goals of this study are to investigate the evolutionary history of Malvales based on a phylogenetic analysis of a representative sample of the order, using divergence time-estimation methods that incorporate its rich fossil history. We use this information to obtain insights about the rate of diversification of major lineages within Malvales. In particular, we aimed to (1) generate a phylogeny with a taxonomic sample at the genus level; (2) implement the FBD model to estimate divergence times and diversification parameters, evaluating the effect of using different numbers of fossils; and (3) compare the dating results between FBD and node-calibrating method, as well as estimated diversification parameters based on time estimates of each method.

2. Materials and methods

2.1. DNA sequences, alignment and phylogenetic analyses

We used the PHLAWD pipeline (Smith et al., 2009) to build a molecular database of Malvales by assembling and aligning published gene sequences for the group. Our main goal was to build a molecular matrix with the highest number of genera. After an exploration of all the GenBank database by obtaining the highest coverage and identity of the genetic regions for Malvales, we downloaded eight molecular markers, six from the plastid genome: *rbcl*, *atpB*, *trnK-matK*, *trnH-psbA*, *trnL-F*, *ndhF*; one from the mitochondrial genome, *matR*; and the nuclear ITS. The PHLAWD pipeline produces an alignment with MAFFT (Katoh and Standley, 2013) for each genetic region. The matrix downloaded by PHLAWD included in several cases more than one species per genus. We achieved a genus-level alignment by selecting the species for which the sequences of all or most of the used molecular markers were available. After this, we re-aligned the sequences with MUSCLE 3.7 (Edgar, 2004), using default settings, to refine the alignment. We subsequently checked the alignment visually and deleted the ends of sequences that substantially exceeded the length of the rest of the sequences, but we retained ambiguously aligned regions. We built a concatenate matrix with the eight genetic regions (Supporting File 3).

We included taxa from all ten families of Malvales. Malvaceae s.l. is the largest clade in the order, so we also included representatives of its nine subfamilies (Table 1). For each genus, we aimed at including the sequences of different loci belonging to the same species, but this could not be achieved in all cases, and for 62 genera, it became necessary to include sequences of different species (Table S1). We manually added the outgroups, consisting of two species of Brassicales (Tropaeolaceae: *Tropaeolum* spp.; Caricaceae: *Carica papaya*), two species of Huerteales (Tapisciaceae: *Tapiscia* spp.; Gerrardinaceae: *Gerrardina foliosa*), and one species of Sapindales (Nitrariaceae: *Nitraria retusa*). In total, we included 258 genera and 12,592 aligned base pairs (bp) corresponding to eight molecular markers (Supplementary Material Table S1). There is a high proportion of missing data (Table S2), in general, for *matR* and *atpB*, and specifically for clades such as Sarcocaulaceae and Sphaerosepalaceae.

We evaluated model fitness for each genetic region with jModelTest (Durrant et al., 2012), where the best fit model for each locus was the General Time Reversible (GTR) model of nucleotide substitution, allowing rate variation according to a gamma distribution combined with a proportion of invariable sites (GTR + Γ + I). Phylogenetic relationships were estimated with Maximum Likelihood using RAXML (Stamatakis, 2014), implementing the GTRCAT model and 1000 Bootstrap replicates. We performed Bayesian inference using MrBayes v.3.2.6 (Ronquist et al., 2012). We partitioned the data according to the

Table 1

Taxon sampling for each of malvacean families and subfamilies of the Malvaceae s.l.

#	Family	Sampled genera	Total genera	% genera included
1	Malvaceae s.l.	187	244	76.6
2	Dipterocarpaceae	12	16	75
3	Sarcocaulaceae	2	9	22.2
4	Cistaceae	8	8	100
5	Bixaceae	2	4	50
6	Sphaerosepalaceae	2	2	100
7	Thymelaceae	34	51	66.7
8	Neuradaceae	3	3	100
9	Cytinaceae	2	2	100
10	Muntingiaceae	1	3	33.3
	TOTAL	253	342	74
#	Malvaceae s.l. subfamily	Sampled genera	Total genera	% genera included
1	Malvoideae*	92	119	77.3
2	Bombacoideae	13	17	76.5
3	Sterculioideae	13	13	100
4	Tiliioideae	3	3	100
5	Dombeyoideae	14	19	73.7
6	Brownlowioideae	7	7	100
7	Helicterioideae	10	13	76.9
8	Grewioideae	13	25	52
9	Byttnerioideae	22	28	78.6

* Includes genera with unclear subfamilial placement, such as *Chiranthodendron*, *Fremontodendron*, *Ochroma*, *Patinoa*, *Quararibea*, *Matisia* and *Septotheca*.

eight molecular markers and linked parameters for each partition. The analysis was set to run three heated chains and one cold chain, and to perform two independent parallel runs. The analysis ran for 2×10^6 generations, sampling every 1000 generations, discarding 20% as burn-in. We used Tracer v.1.6 (Rambaut et al., 2014) to corroborate chain convergence. Topologies were summarized in a 50% majority rule tree. All analyses were performed in the CIPRES Science Gateway (Miller et al., 2010). Mitochondrial genes of some parasites, and particularly Cytinaceae, have a higher substitution rate compared with their non-parasitic relatives (Zervas et al., 2019). In the present study we used the mitochondrial molecular marker, *matR*, for Cytinaceae species. Thus, we performed additional analyses excluding Cytinaceae to determine the impact of this parasitic family in the estimation of phylogenetic relationships.

2.2. Fossil information

We reviewed the literature in search of fossils of the ten malvacean families, and for Malvales in general. We obtained information from paleofloras, fossil descriptions and fossil databases (Supporting File 1). We included micro- and macro-fossils (e.g., pollen and macroscopic structures such as leaves and flowers). To select fossils that represent reliable assignments to Malvales, two main criteria had to be fulfilled: (1) a fossil was included if its taxonomic affinity with Malvales had been explicitly analyzed or discussed in published work; and (2) a fossil was included if the geological formation where it was collected was properly identified and the stratum was well dated. When several specimens of the same species are known, only the oldest one was included as a representative of the species, to avoid overestimation of speciation and extinction rates, which could be caused by erroneous interpretation of specimens of the same species (derived from a single branching event) as members of different species (corresponding to different branching events, but see Renner et al., 2017).

We compiled information for 80 fossils of different groups within Malvales, several of which can be confidently assigned to families, subfamilies and genera (Supporting File 1). We acknowledge that there are fossils that we did not include, but usually these are very young and

usually belong to already sampled species or genus. The complete list of fossils, their stratigraphic position, phylogenetic assignment and related information is summarized in [Supporting File 1](#). The ages were updated according to the international chronostratigraphic chart of [Cohen et al. \(2015\)](#). We additionally scored the 80 fossils based on the type of study from which each fossil was obtained. Fossils with score 3 are those that were explicitly, morphologically compared with extant species; score 2 was assigned to fossils that were taxonomically classified based on information available in the literature; and a score of 1 to fossils obtained from palynofloras.

Fossil ages were introduced in the FBD analyses (see below) as the midpoint of the time interval that covers the epoch where the fossils were found. In a recent study, [Renner et al. \(2017\)](#) found that assigning an age corresponding to the oldest or youngest bound of the stratigraphic interval of fossils causes a bias in the estimation of the probability of observing a lineage prior to its extinction (s), but in general the assignment of either age (oldest, youngest or random) resulted in similar outcomes. Thus, for the FBD analyses we assigned a fixed value for the age. To obtain a specific value we considered the narrowest stratigraphic interval to which the fossil can be securely assigned, and calculated the midpoint value between its upper and lower limits. For the node dating analysis, we used prior distributions to incorporate the age reference for each calibration point, as described below.

From the 80-fossil data set, we generated three subsets to include in different analyses. Sixty one of the included fossils belong to crown groups of taxonomically recognized groups, meaning that they are considered to have originated after the deepest split within the group, while 19 fossils are placed in the stem groups of the lineages, i.e., interpreted as having originated before the deepest split leading to extant species (Table S3).

2.3. Divergence time and diversification dynamics estimated with the Fossilized Birth Death process

It has recently become possible to estimate divergence times by integrating extant and extinct species in the same tree-generating diversification process, under the Fossilized Birth-Death (FBD; [Heath et al., 2014](#)) model. This model assumes that the fossils are part of the same macroevolutionary process that gave rise to the living species ([Heath et al., 2014](#)). This implies that fossils can be interpreted as terminals of extinct branches, or as sampled ancestors in phylogenetic branches ([Gavryushkina et al., 2017](#)). The sampling through time model was introduced by [Stadler \(2010\)](#) and later developed by [Heath et al. \(2014\)](#) as an explicit divergence time estimation method under a Bayesian inference framework. The model includes four major parameters: speciation rate (λ), extinction rate (μ), fossil recovery rate (ψ) and the probability of sampling an extant species (ρ). The FBD model assumes constant speciation and extinction rates through time.

An important property of the currently implemented FBD model, related with the way in which it integrates fossils in the phylogenetic history, is that it models the presence of sampled ancestors ([Gavryushkina et al., 2014](#)). Specifically, the model takes all the possible attachment points of the fossils on the extant tree ([Heath et al., 2014](#)), meaning that, given the fossil recovery rate, speciation and extinction rates, some of the fossils will be placed as ancestors of living lineages and others will be interpreted as terminal points of an extinct lineage ([Gavryushkina et al., 2014](#)). This is a relevant aspect to consider in cases where no morphological data set for fossils and extant species is used. In such cases, *a priori* knowledge of the phylogenetic placement of fossils is crucial. Nevertheless, the method estimates a distribution of the placement of each fossil, thus fossils are not necessarily fixed in a particular clade. In our study we used the taxonomic placement of the fossils to inform their assignment to clades ([Supporting File 1](#)).

In contrast to FBD, the most commonly applied dating method, in which internal nodes are calibrated (node dating, ND), consists of an *a priori* assignment of a probability distribution of the age of a fossil (or

other type of calibration source) assigned to a specified node. In ND, the tree-generating model (i.e., Birth-Death process) only considers extant species and is solely determined by the speciation and extinction rates. Fossils are only used to inform ages of pre-determined internal nodes. Moreover, the placement of fossils needs to be independently informed, whereas in the FBD model it may be estimated by including morphological data of extant and fossils species and applying a model of morphological evolution ([Gavryushkina et al., 2017](#)). When morphological data is not available fossils can be assigned to a given clade, and their position within the clade is estimated as a model parameter with an associated posterior distribution.

Divergence time estimation and diversification parameters were estimated with the FBD in BEAST2 v.2.4.6 ([Bouckaert et al., 2014](#)), including the Sampled Ancestors package ([Gavryushkina et al., 2014](#)), which allows fossil species to be modeled as extinct terminal taxa or as sampled ancestors in phylogenetic branches. The used molecular dataset is the same as the one used in phylogenetic estimation, and was partitioned into (1) the nuclear ITS; (2) the mitochondrial *matK*; and (3) the plastid *atpB*, *rbcL*, *trnK-matK*, *ndhF*, *trnH-psbA* and *trnL-F*. FBD analyses included only ingroup species (i.e., the five outgroup species were trimmed), therefore the dataset included 253 extant taxa.

We conducted a set of analyses to assess how the number of included fossils influenced the time estimates. For this, we performed three FBD analyses with different numbers of included fossils: FBD14, with 14 fossil species corresponding to the oldest fossils in each clade, and which are the same as those implemented in the ND analysis (see section 2.4); FBD80, containing all fossil species that could be reliably assigned to different clades within Malvales (see Section 2.2; Table S3); and FBD50, with 50 fossil species. This analysis was performed to observe the variation in the age estimates with an intermediate number of fossils between 14 and 80. The 50 fossils included the first 14 and the rest were selected with the criterion of keeping the distribution of fossils in the 80 analysis.

We additionally performed a set of analyses aimed to determine the potential influence on time estimates of fossils which phylogenetic position is less secure. We conducted two FBD analyses, one with fossils score 2 and 3 (FBDs2-3; 50 fossils) and one with fossils score 3 (FBDs3; 22 fossils) ([Supporting File 1](#)). In both analyses we excluded Cytinaceae to observe how time estimates vary when this parasitic group is absent.

Ages of extant and fossil species were specified as (million) years before the present and incorporated in BEAUTi through a text file that included the age of all the extant (0 million years [Myr]) and fossil taxa ([Supporting File 2](#)). Phylogenetic relationships among extant and fossil species were specified with the TAXSETS command in the input NEXUS file ([Supporting File 3](#)). We constrained the families to be monophyletic as observed in the results from the Bayesian phylogram (Section 2.1). Although *Muntingia calabura* was found to be contained within Cytinaceae, we did not include this arrangement as a constraint for the dating analyses, given the low support of this relationship. Instead, in those analyses in which Cytinaceae was included, it was constrained as a clade. Relationships of fossils with extant species were based on the published analyses of their taxonomic affinities ([Supporting File 1](#)).

Site substitution models were unlinked among the tree partitions, and in each case, were assigned a GTR + Γ model with empirical base frequencies, and estimated substitution rates. The gamma distribution to account for among-site rate heterogeneity was subdivided into four discrete categories, and its shape was estimated. The priors for each type of transition were obtained from a gamma distribution with $\alpha = 2.0$ and $\beta = 0.5$ for the transition AG, and the rest of the nucleotide transitions had $\alpha = 2.0$ and $\beta = 0.25$, following [Heath \(2017\)](#). To specify the model of lineage-specific substitution rate heterogeneity we implemented an uncorrelated relaxed molecular clock linked across partitions, with rates obtained from a lognormal distribution (UCLN). The prior for the UCLN mean was obtained from an exponential distribution with mean = 10, and standard deviation from a gamma distribution with $\alpha = 0.5396$ and $\beta = 0.3819$ (as default). The Fossilized

Birth-Death process (Heath et al., 2014) was the generating model for the topology and the internal branching times.

The model was conditioned on the origin (starting time) of the process, corresponding to the root of the tree (i.e., crown node) because all fossils were assigned to the crown group and not to the stem group of Malvales (following Heath, 2017). Its prior age was assigned a uniform distribution between 89.44 and 103.2 Ma, corresponding to the 95% highest posterior density (HPD) of the age estimated for Malvales stem node by Magallón et al. (2015). We did not use a crown age to calibrate the root because most of the studies including Malvales estimate a younger age than or overlapping with the age of the oldest fossils in our sampling. The diversification rate (r) prior was obtained from a log-normal distribution with mean = 0.09, derived from the rate estimated for Eudicotyledoneae in Magallón et al. (2019), and standard deviation of 1.25. The turnover rate (ϵ) prior obtained from a uniform distribution between 0.0 and 1.0. The fossil sampling proportion (s), the probability that a lineage enters the fossil record before it becomes extinct, was obtained from a beta distribution with $\alpha = 2.0$ and $\beta = 2.0$ (following Heath, 2017). The proportion of sampled extant species (p), corresponding to the proportion of sampled species relative to the total number of malvacean species, was fixed at 0.04.

The FBD analyses were conducted in the CIPRES Science Gateway (Miller et al., 2010) and the Biolinux server (Instituto de Biología, Universidad Nacional Autónoma de México). The FBD14 and FBD50 analyses consisted of two independent Markov chain Monte Carlo (MCMC) runs of 200 million generations sampling one every 2000 generations, and for the FBD80 one run of 450 million generations sampling one every 5000 generations. The outputs of the two independent runs were examined for chain convergence and effective sample size (ESS) of different estimated parameters with Tracer v.1.6 (Rambaut et al., 2014). Trees from different runs of each analysis were combined with LogCombiner v.2.4.6 (Bouckaert et al., 2014). For each one of the four analyses we obtained the Maximum Clade Credibility (MCC) tree with TreeAnnotator v.2.4.6 (Bouckaert et al., 2014), onto which estimates of age and of other parameters were summarized.

The FBD model estimates the diversification rate (d) and turnover rate (r ; i.e., relative extinction [ϵ]) for the entire phylogeny. With these two rates we calculated the speciation (λ) and extinction (μ) rates according to the birth-death model of diversification, where $\lambda = d/1 - \epsilon$ and $\mu = d(\epsilon)/1 - \epsilon$ (Nee, 2006). The FBD model also estimates the probability that a fossil taxon is the ancestor of a lineage rather than the tip of an extinct branch. BEAST2 includes an application that extracts the frequencies of all the sampled ancestors in a sample of dated trees (SampledAncestorTreeAnalyser). We entered a sample of 800–1000 trees from each of the three FBD analyses into the SampledAncestorTreeAnalyser application.

2.4. Divergence time estimation with node-calibrated relaxed molecular clock

A node-calibrated relaxed molecular clock was implemented to establish a comparison with FBD-estimated divergence times. This analysis was conducted in BEAST2 v.2.4.6 (Bouckaert et al., 2014) following the general procedures described above for FBD analyses, but with the following differences: (1) a Birth-Death process was selected as the topology and branching times generating model; and (2) 14 clades, corresponding to five families of Malvales and eight subfamilies of Malvaceae s.l., were constrained with fossil-derived ages (Table S1; Supplementary File 1). As described above, each clade was calibrated with the oldest fossil that can be confidently assigned to that clade. Calibrations were implemented following the conservative criteria of Parham et al. (2012). Constraints of clades were specified in the NEXUS input file. Each calibration was implemented as a uniform prior distribution (Foster et al., 2016), with the lower (youngest) bound equal to the minimum age of the stratigraphic range of the fossil, and the upper (oldest) bound for all the calibrations equal to 103.2 Ma, which is the

maximum limit in the 95% HPD estimated for the stem Malvales in Magallón et al. (2015). The node-calibrating analyses were run for 200 million of MCMC generations, sampling one every 1000 generations. The output log file was analyzed with Tracer, summarized with LogCombiner and TreeAnnotator, and visualized with FigTree v.1.4.3 (Rambaut, 2014).

2.5. Diversification parameters derived from node-calibrated chronogram

To establish a comparison of diversification parameters derived from the node-calibrated (ND14) chronogram and those estimated with the FBD, we used estimated ages of the families of Malvales and subfamilies of Malvaceae s.l. to calculate net diversification rates with the method-of-moments (m-o-m) estimator of Magallón and Sanderson (2011) using the *bd.ms* function of the R package *geiger* v.2.0.6 (Pennell et al., 2014; R Foundation for Statistical Computing). We used stem and crown ages of clades estimated with the ND14 and FBD80 analyses, combined with the total species richness derived from The Plant List (www.theplantlist.org). We considered all accepted names and counted the species for genera not included in our study. Additionally, we calculated the diversification rate using the estimates from FBDs2-3. We preferred this analysis over FBDs3 because it comprises a wider representation of fossils across the families than the analysis that includes only fossils with score 3 (i.e., only Malvaceae s.l. is represented in FBDs3). Diversification rates were estimated for the ND14 with two extreme values for the relative extinction rate ($\epsilon = 0.01$, and $\epsilon = 0.9$); for FBD80 and FBDs2-3 analyses diversification rates were obtained with the estimated relative extinction rate from the FBD model ($\epsilon = 0.919$ and $\epsilon = 0.906$, respectively).

3. Results

3.1. Phylogenetic relationships of Malvales

The molecular data set includes eight molecular markers representing the three genomic compartments, comprising a total 12,592 (bp). The 258 taxa correspond to 74% of the genera in Malvales (Table S2). The proportion of genera sampled in each family of Malvales and subfamily of Malvaceae s.l. is shown in Table 1 and the proportion of taxon and locus sampling is shown in Table S2. Phylogenetic relationships among Malvacean families obtained in the analysis with MrBayes are shown in Fig. 1, including Bootstrap support from the Maximum Likelihood analysis (internal PP/BS values are shown in Fig. S1-4). All the families were found to be monophyletic, except in the Bayesian analysis when including Cytinaceae because it contains *Muntingia calabura* (Fig. S3-4), but this relationship is supported by 0.56 PP.

Fig. 1 shows that Neuradaceae is the sister to the rest of the families (1 PP/100 BS) and successively Malvaceae s.l. is sister to the remaining families (0.87 PP/81 BS). Within Malvaceae s.l., all subfamilies were found to be monophyletic. However, we considered *Quararibea* and *Pentaplaris* within Malvoideae, and *Chiranthodendron*, *Ochroma* and *Septotheca* to be genera without subfamilial circumscription. We considered Bombacoideae a clade with primarily palmately compound leaves, as “core Bombacoideae” defined by Baum et al. (2004).

3.2. Fossil data set

Mining the published paleobotanical literature resulted in a dataset of 80 fossils for Malvales (Supporting File 1). Seven malvacean families were represented by at least one record. The most abundantly represented family is Malvaceae s.l. with 70% of the included fossils, followed by Dipterocarpaceae with 15%, and Thymelaeaceae with 6.25%. Neuradaceae, Cytinaceae and Sphaerosepalaceae are not represented in the fossil record. Within Malvaceae s.l., eight out of nine subfamilies are represented by at least one record, with Bombacoideae being the most abundantly represented (29.8% of the crown fossils),

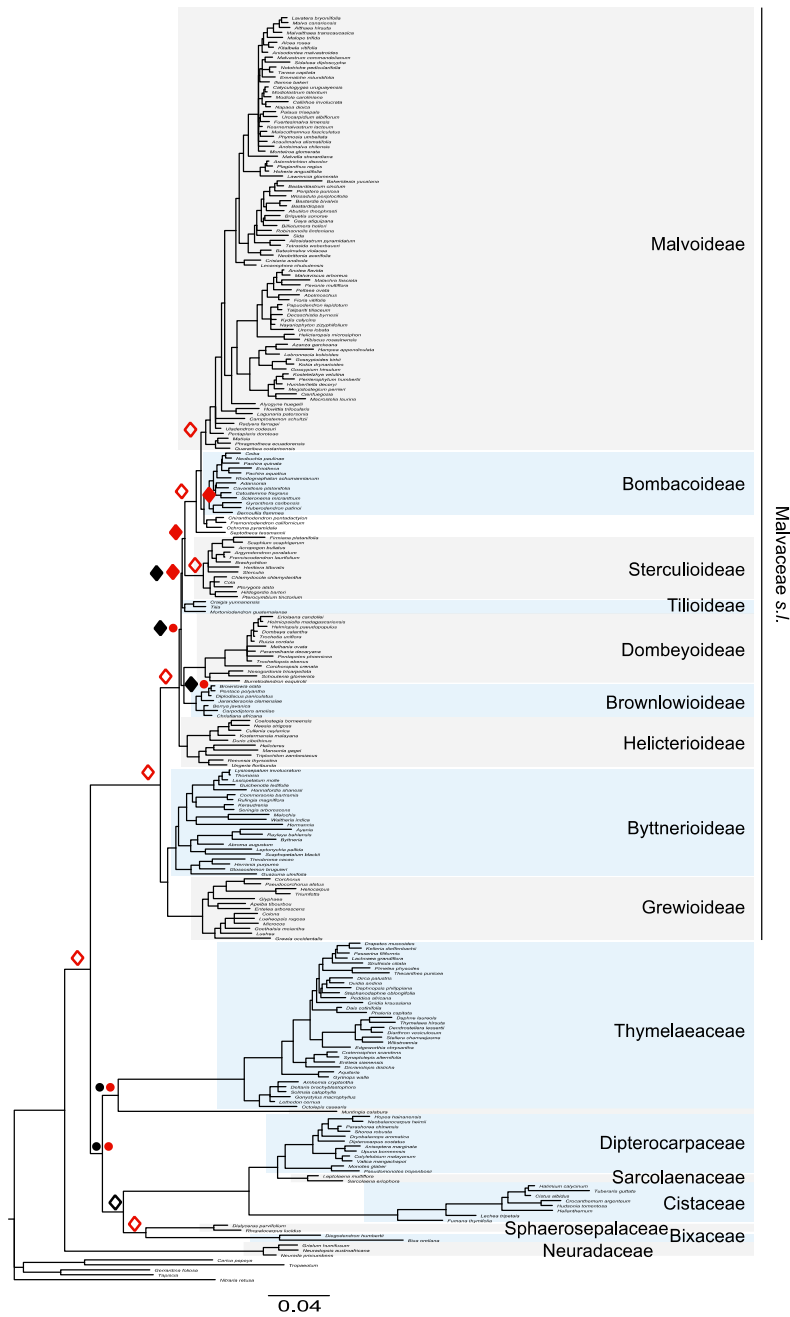


Fig. 1. Bayesian phylogenetic tree of order Malvales estimated with eight molecular markers. Black symbols indicate Bayesian posterior probabilities (PP) and red symbols indicate maximum likelihood bootstrap support (BS). Open diamonds correspond to nodes with 0.8–0.99 PP/80–99 BS; solid diamonds to nodes with 0.7–0.8 PP/70–80 BS; circles to nodes with < 0.7 PP/ < 70 BS. The rest of the major lineages have 1 PP/100 BS. For support of intrafamilial relationships see the Supplementary Material. The red branch represents *Muntingia calabura* within paraphyletic Cytinaceae.

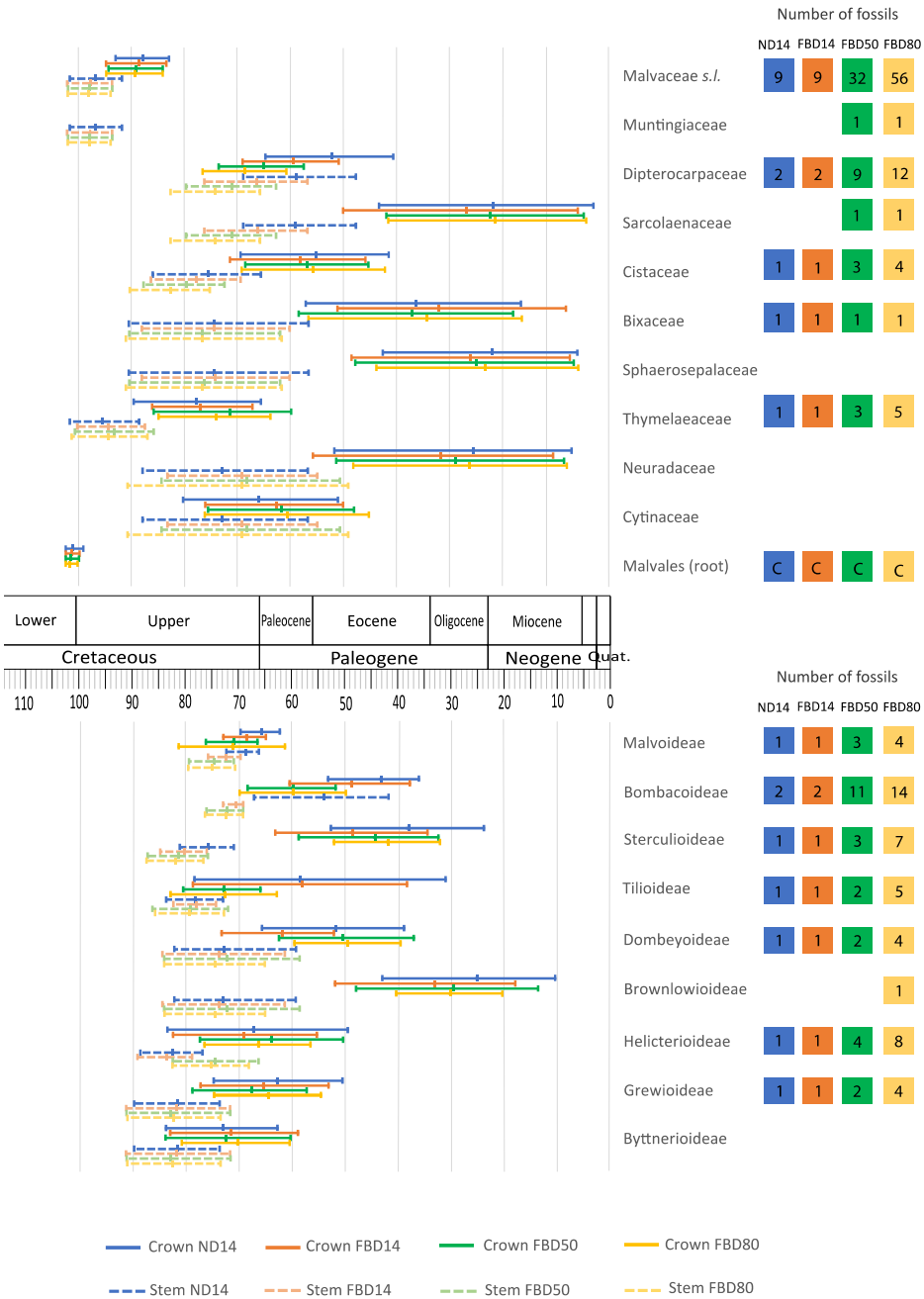


Fig. 2. Divergence time estimation compared among four analyses with different number of fossils. Lines show the mean and the 95% HPD intervals. The number of fossils for each group are shown in the right panel. C: secondary calibration. ND: node dating; FBD: Fossilized Birth-Death.

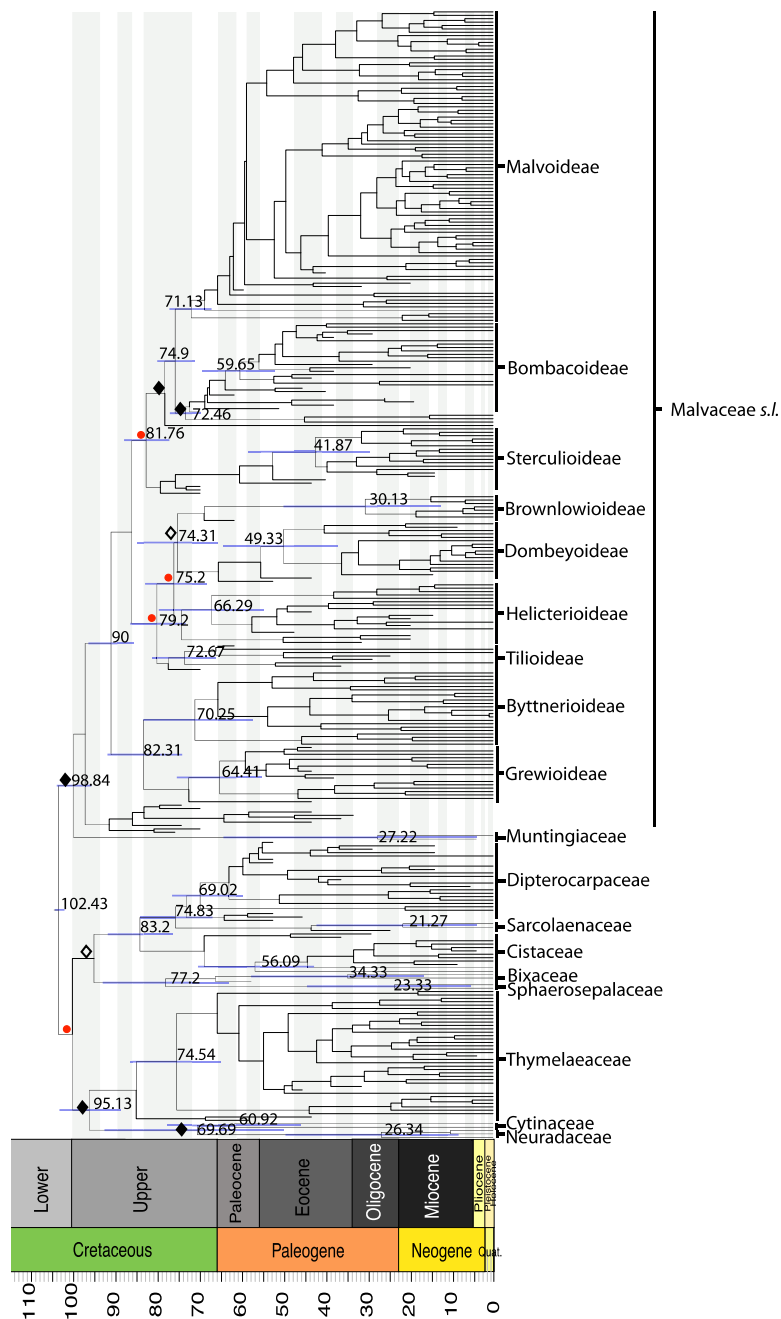


Fig. 3. Divergence times of Malvales with the Fossilized Birth-Death process including 80 fossils (FBD80). Branches that do not reach the present correspond to fossil taxa. Blue bars represent the 95% HPD. White diamonds correspond to nodes with 0.8–0.99 PP; black diamonds to nodes with 0.7–0.8 PP; red circles to nodes with < 0.5 PP. For support of intrafamilial relationships and divergence times see the Supplementary Material.

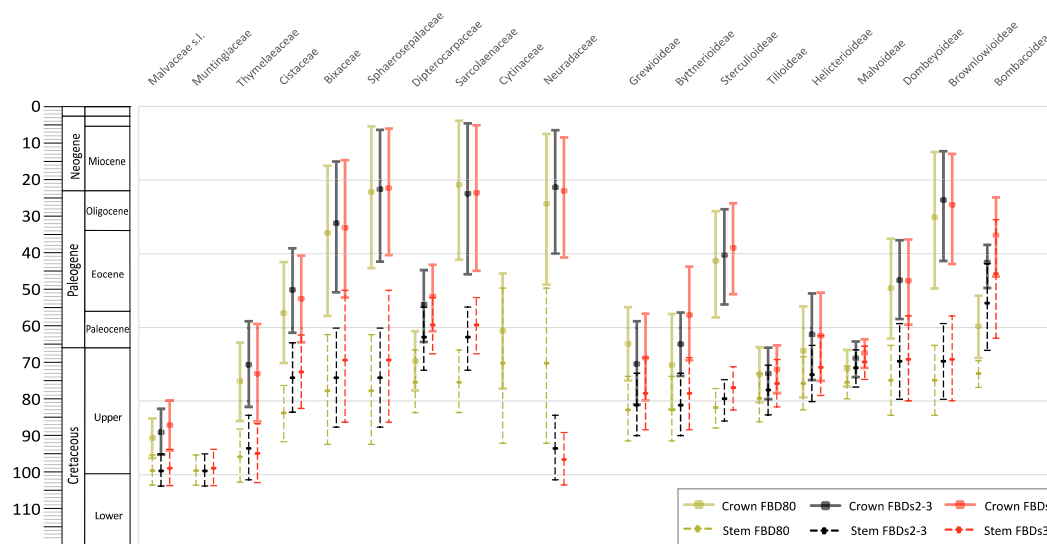


Fig. 4. Crown and stem age estimated with FBD80 and the highly-scored fossils, FBDs2-3 and FBDs3. Bars represent the 95% HPD.

followed by Grewioideae (17%) and Sterculioideae (14.9%). Subfamily Byttnerioideae is represented by very recent fossils, which we did not include in the present study.

3.3. Divergence time and diversification dynamics estimated with the Fossilized Birth Death process

Parameter estimation in the analyses of ND and FBD was evaluated by corroborating the convergence of MCMC chains, where most ESS values were equal or higher than 200, with a few parameters having > 100. Phylogenetic relationships estimated with the FBD were mostly similar to those estimated with MrBayes (see Section 4.1 for differences; Figs. 1, 3; Figs. S1–10). Fig. 2 and Table S4 show the mean age and 95% HPD for the stem and crown nodes of the families of Malvales and subfamilies of Malvaceae s.l. estimated in the FBD14, FBD50 and FBD80 analyses, except for Muntingiaceae, which was represented only by one species, hence only its stem age was inferred. Given the narrow calibration interval assigned to the root of the tree, corresponding to the crown node of Malvales, age estimates for this node exhibited very small variation in the different analyses (Fig. 2; Table S4).

The pattern of estimates among the four analyses is different for the stem and the crown ages. Among the three FBD analyses that include Cytinaceae and differ in number of fossils, FBD80 estimated the oldest stem ages (76.9% of the 13 different stem ages, because the members of each of three sister pairs share their stem age) and FBD14 the youngest (69.2%), whereas for crown ages FBD80 estimated the youngest (47.36%) and FBD14 the oldest (52.6%). Estimates of FBD80 and FBD50 are mostly similar. Fig. 2 shows that age estimates obtained from three FBD and ND analyses are highly overlapped. When considering the two FBD analyses that exclude Cytinaceae and incorporate only highly-scored fossils (FBDs2-3 and FBDs3), the ages tended to be younger than those estimated in analyses that also included fossils with low score (Fig. 4). Specifically, for crown age estimates, analyses including high-scoring fossils (FBDs3-2) yielded younger (35%) and FBD14 yielded the oldest (47%), whereas for stem ages, FBDs3 yielded the youngest (41%) and FBD80 de oldest ages (47%) (Table S4). Comparing the six dating analyses, most of the HPD intervals overlap,

except in particular clades and analyses.

Estimated 95% HPD intervals are narrower for stem ages than for crown ages, except in Cytinaceae and Neuradaceae, where the magnitude of intervals is similar (Table S4). The results suggest that, as more living taxa are sampled for a group, the narrower its age HPD interval (Figs. 2–4). This pattern seems independent from the number of fossils as, for example, Malvaceae s.l. and Malvoideae exhibit narrow HPD intervals, but the number of fossils in each clade is very different (56 and 4, respectively).

The FBD model estimates phylogenetic relationships together with divergence times and net diversification rate. However, the latter value can only be accurately estimated with a species-level sampling. Table S5 and Figure S11 show the estimated net diversification rate for Malvales in the five FBD analyses. We found that diversification rate estimates tend to decrease as more fossils are included, whereas turnover rates tend to increase as observed from the FBD80 compared with the other four analyses (Table S5). Speciation and extinction rates increase with the number of included fossils, resulting in a net diversification rate smaller than those estimated with the other four analyses (Table S5, Fig. S11).

Fossils that were estimated as ancestors (Gavryushkina et al., 2014) varied among the FBD analyses. In Table S7 we show only those fossils identified as ancestors with the highest sampled frequency (10–66% of the sampled chronograms). These results are tentative, as it is risky to consider a fossil as ancestor of a lineage, especially without analyzing morphological characters.

Our results show that all malvacean families originated in the Upper Cretaceous (ca. 83–66 Ma) (Dipterocarpaceae-Sarcobalanaceae stem age at 62.6 [71.6–54.5] Ma in FBDs2-3). A short time period elapsed between the crown node of Malvales (102.4 [103.2–100.9] Ma) and the origin of Malvaceae s.l. and Thymelaeaceae (98.9 [102.8–94.6] and 95.1 [102–87.6] Ma, respectively), during the Cenomanian (Figs. 2–4). The stem age of the rest of the families is ca. 20 Myr younger than the diversification of the order as a whole (Figs. 2–4). The diversification of Malvaceae s.l. took place shortly after its origin (ca. 8 Myr, Fig. 4; Table S4) but Thymelaeaceae took longer to diversify (ca. 18 Myr, Fig. 4; Table S4). The rest of the families originated between 84 and 69 Ma, from the Santonian to the Maastrichtian-Early Paleocene. The

diversification of families, on the other hand, took place at very different evolutionary times (Figs. 2–4).

The subfamilies of *Malvaceae s.l.* originated within a short period of time, between 82 and 72 Ma, during the Campanian (Bombacoideae stem age at 53.4 [66.3–42.6] Ma according to FBDs2-3 analyses), but they diversified at different times (Fig. 4), all before the early Miocene (20 Ma) all the subfamilies had diversified. Within *Malvaceae s.l.*, five subfamilies (Byttnerioideae, Grewioideae, Helicterioideae, Tilioideae and Sterculioideae) originated during the Campanian (ca. 83.6–72.1 Ma), slightly after the diversification of the family (90.91 Ma; Figs. 2–4), followed by the remaining subfamilies at the end of the Campanian and the Maastrichtian (ca. 75–66 Ma; Figs. 2–4). The times of origin and diversification of Malvoideae and Tilioideae are very close, ca. 3 and ca. 9 Myr, respectively; whereas for the rest of the subfamilies the origin and diversification are separated by at least ca. 12 Myr (Fig. 4; Table S3), with the largest differences found in Brownlowioideae and Sterculioideae (ca. 45 and 40 Myr, respectively).

3.4. Divergence time estimation with node-calibrated relaxed molecular clock

Fig. S5 shows the Malvales time-tree obtained with the node-calibrated UCLN relaxed clock (ND14), and Table S4 shows the estimated ages and 95% HPD. Stem age confidence intervals overlap among the six analyses, where 84% of mean ages of major lineages differ by 10 Myr (33% considering FBDs2-3 and FBDs3) and 53% differ by less than 5 Myr (42% considering FBDs2-3 and FBDs3; Figs. 2 and 4; Table S4). Crown age confidence intervals also overlap among the six analyses, where 72% of the estimates differ by less than 10 Myr (50% considering FBDs2-3 and FBDs3) and 33% differ by less than 5 Myr (17% considering FBDs2-3 and FBDs3). For stem estimates, the greatest differences are found between FBD80 and ND14, and between FBD80 and FBDs3, where the former generated the oldest estimates, and the latter the youngest, suggesting that stem ages tend to be older as the number of fossils increases. This pattern is not apparent for crown ages, where the oldest estimates were generated by FBD14 and the youngest by ND14 and FBDs2-3 (Table S4). The greatest differences among the six analyses and between stem and crown ages are observed for the clade formed by Dipterocarpaceae plus Sarcocaulaceae, and in subfamily Bombacoideae. These numbers show that estimates of crown node age tend to differ more than those for stem ages.

3.5. Diversification parameters derived from ND and FBD

Rates of diversification estimated with the m-o-m and with the FBD model can only be compared for Malvales as a whole, as the FBD model estimates a single, constant rate for the entire phylogeny. Table S6 shows the net diversification (r) rates for Malvales families and *Malvaceae s.l.* subfamilies estimated with the m-o-m for the stem and crown node age estimates obtained with ND14 and FBD80. Fig. 5 shows the diversification rate of the stem and crown node of each lineage obtained with the m-o-m estimator on ND14 under two extreme rates of relative extinction (ϵ), as well as estimates obtained in the FBD80 analysis under the mean value estimated for the relative extinction rate obtained with the FBD model (Table S5).

The analyses with low and high relative extinction ($\epsilon = 0.01$ and 0.9 , respectively) produced similar diversification rates, and these results are observed using both stem and crown ages. Differences between the results obtained with the two relative extinction fractions are more pronounced in Neuradaceae and Cytinaceae when stem age is used, whereas for the crown ages the greatest differences are observed in Sarcocaulaceae and Sphaerosepalaceae (Fig. 5; Table S6). Similar results were obtained from m-o-m results applied to ND14, FBD80, and FBDs2-3, therefore we only show in Fig. 5 the results from ND14 and FBD80 (Table S6 shows results from the three analyses). As expected, the greatest difference occurs in clades with a long temporal difference

between the stem and the crown ages (e.g. Dipterocarpaceae).

The highest crown diversification rate was obtained for Sarcocaulaceae (Table S6), but the estimated rates vary from ND14 ($\epsilon = 0.01$: $r = 0.16$; $\epsilon = 0.9$: $r = 0.08$) to FBD80 ($\epsilon = 0.919$: $r = 0.08$) while the family with highest stem diversification rate is *Malvaceae s.l.* (ND14 $\epsilon = 0.01$: $r = 0.08$; $\epsilon = 0.9$: $r = 0.06$; FBD80 $\epsilon = 0.919$: $r = 0.06$). The families with lowest crown and stem diversification rate are Cytinaceae (ND14 $\epsilon = 0.01$: $r = 0.02$; $\epsilon = 0.9$: $r = 0.01$; FBD80 $\epsilon = 0.919$: $r = 0.009$) and Muntingiaceae (ND14 $\epsilon = 0.01$: $r = 0.004$; $\epsilon = 0.9$: $r = 0.001$; FBD80 $\epsilon = 0.919$: $r = 0.0009$) (Fig. 5; Table S6). Within this family the highest diversification rate is estimated for Brownlowioideae ($\epsilon = 0.01$: $r = 0.15$; $\epsilon = 0.9$: $r = 0.09$) when using the ND14 crown age, or Sterculioideae ($\epsilon = 0.919$: $r = 0.08$), when using the FBD80 crown age. In contrast, when considering the stem age, subfamily Malvoideae has the highest rates (ND14 $\epsilon = 0.01$: $r = 0.1$; $\epsilon = 0.9$: $r = 0.08$; FBD80 $\epsilon = 0.919$: $r = 0.07$). The lowest diversification rate is for Tilioideae with crown and stem estimates and also between ND14 and FBD80 (Fig. 5; Table S6).

4. Discussion

4.1. Phylogenetic relationships

Our study is based on the largest molecular data set for Malvales, and provides an improved understanding of phylogenetic relationships at the family level. We confirm some previously known relationships, for example, that Dipterocarpaceae and Sarcocaulaceae form the sister clade to Cistaceae, and document strongly supported relationships that in previous studies were poorly supported or unresolved, in particular that Bixaceae and Sphaerosepalaceae as sister clades.

All the families were found to be monophyletic and agree with previous studies regarding their relationships. An important result in this study is the resolution of previously uncertain relationships, for instance, the placement of Bixaceae and Sphaerosepalaceae, and *Malvaceae s.l.* as the sister of a clade comprising the rest of the families (excluding Neuradaceae). Within *Malvaceae s.l.*, we detected the two main clades of Byttneriina and Malvadendrina, as well as the Malvatheca clade (Malvoideae, Bombacoideae, *Ochroma*, *Septotheca*, *Chiranthodendron* and *Fremontodendron*; Alverson et al. 1999; Baum et al. 2004; Baum et al. 1998). Previous studies reported a polytomy among the lineages in Malvadendrina. Here, we show that Dombeyoideae is sister to Brownlowioideae (Fig. 1), and that Sterculioideae may be (MrBayes: 1 PP; BEAST: ca 0.5 PP; RAxML: 75 BS) sister to Malvatheca (Figs. 1 and 3), contrasting with the results of Conover et al. (2018) where Sterculioideae and Tilioideae are sister clades.

The phylogenetic relationships obtained from MrBayes and BEAST analyses are mostly similar, but some differences are discussed. In this study, the following relationships are strongly supported (> 0.9 PP): (1) Neuradaceae as sister to the rest of the families (1 PP) (2) Cistaceae, Sarcocaulaceae plus Dipterocarpaceae (Dipterocarpacean clade); (3) Bixaceae plus Sphaerosepalaceae (Bixalean clade); and (4) a clade formed by the Dipterocarpacean and Bixalean clades.

The persistent inconsistency in the placement of Muntingiaceae and Neuradaceae, and their separate association with Cytinaceae, is probably caused by the low amount of molecular data (only *matR*) available for the latter family, and, most importantly, by the high molecular evolutionary rates caused by its parasitic habit (Bromham et al., 2013). Although there is evidence for *Muntingia* to be the sister group of Cytinaceae, as shown by Nickrent (2007), that analysis only includes *Muntingia* as the single representative of the family, and the relationship was moderately supported (72% bootstrap). If future analyses provide increased support for this relationship, it would be very interesting to investigate the early floral evolution in the order by incorporating its distinct morphology (Bayer et al., 1998).

Neuradaceae has been previously resolved as sister to the rest of Malvales (Alverson et al., 1998; Nickrent, 2007). However, its

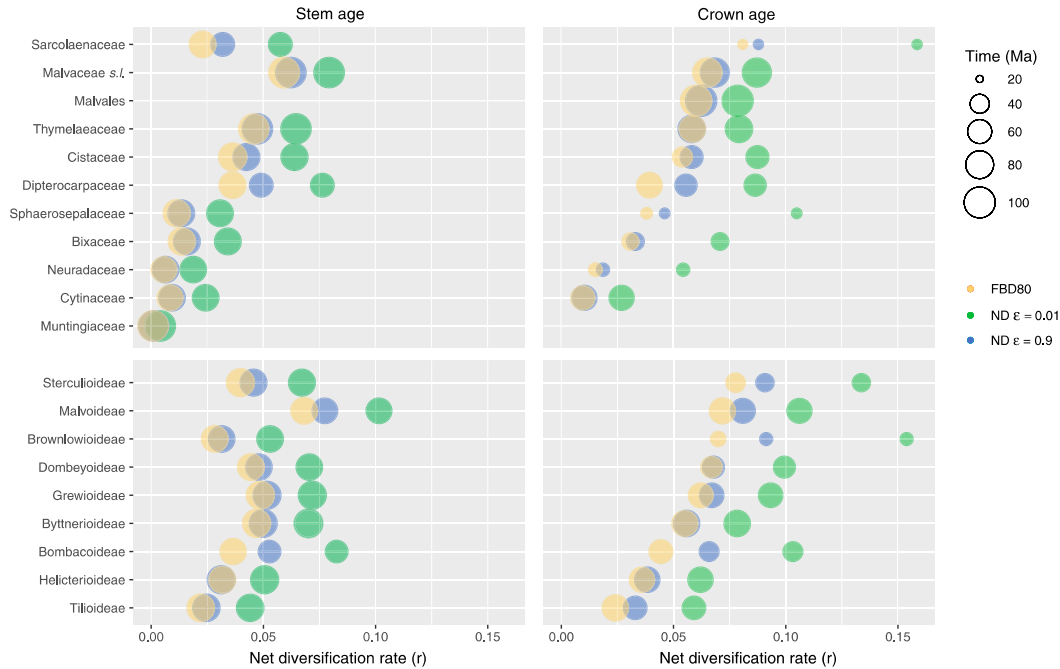


Fig. 5. Net diversification rates obtained with the method-of-moments estimator using stem ages (left panels) and crown ages (right panels) for the Malvacean families and subfamilies of Malvaceae s.l. For ND dates, the relative extinction fraction was set to two extreme values ($\epsilon = 0.01$; $\epsilon = 0.9$). For FBD80, the value was set according to the estimated by the FBD model ($\epsilon = 0.919$). Clade ages are represented by circle size.

placement varies depending on the molecular markers and the taxon sampling (Alverson et al., 1998; Bayer et al., 1999; Sun et al., 2016). The close relationship among Dipterocarpaceae, Sarcolaenaceae and Cistaceae has been well documented in previous studies (Alverson et al. 1998; Ducousso et al. 2004; Heckenhauer et al. 2017).

Within Malvaceae s.l., the placement of Tilioideae and Helicterioideae is not yet resolved. The resulting relationships show subfamily Tilioideae as the sister of a clade consisting of the Malvaceae clade (Bombacoideae plus Malvoideae) plus Sterculioideae, with low support (0.72 PP/26 BS), however Tilioideae is the sister of a group comprising subfamilies Helicterioideae, Brownlowioideae and Dombeyoideae in FBD80 and FBDs2-3 analyses, but with low support (0.39 and 0.83 PP, respectively).

4.2. Timing of malvacean diversification

Divergence times of the major lineages obtained with the FBD estimates are similar among each other, and the differences among them are mostly influenced by the number of fossils included (Fig. 2). Moreover, as shown in Fig. 2, age estimates among the four different analyses (ND and FBD) are remarkably congruent, even when considering analyses including the highly-scored fossils, a result already obtained in different studies for different organisms (Heath et al., 2014; Arcila et al., 2015; Saladin et al., 2017). However, Saladin et al. (2017) showed that the mean values were significantly different among the different techniques and constraints, even when most of their estimates differ in less than 10 Myr and the HPD intervals overlap. Here, among the four analyses that differ only in number of fossils, the greatest variations were mostly observed in the 95% HPD intervals, which are slightly wider in ND14, except for Bombacoideae (Fig. 2; Table S4). Here we were unable to make a strict comparison between the two

dating methods (i.e., ND and FBD) because fossil ages were applied differently, i.e., as uniform priors on the age of internal node in ND14, and as fixed age (middle-point approach) in FBD analyses. Considering the two analyses including only highly-scored fossils (FBDs2-3 and FBDs3) ages are overall younger than the four analyses in which fossils with high and low scores were included (Fig. 4; Table S4). A likely reason is that the latter analyses included “early but risky”, low-score fossils (Sauquet et al., 2012), which may push clades towards older ages (Supplementary File 1).

Heath et al. (2014) showed that the FBD model is less precise than the ND methods as observed in the larger 95% HPD intervals from FBD results, but this effect is not apparent in our results. Here, the HPD intervals among the six different analyses are generally wider in the ND14 crown ages, and in the stem ages the width of the intervals varies among groups, although wider intervals were slightly more frequent in FBD50. Although Heath et al. (2014) observed that as the age increases, the 95% HPD intervals become wider, we did not observe this effect in our analyses, as the oldest families, Malvaceae s.l. and Thymelaeaceae, have narrower intervals than the youngest clades, for example, Bixaceae and Sarcolaenaceae (Figs. 2 and 3). This trend is probably related with the interaction of the root secondary calibration and the relative old age of the fossils of these groups, resulting in a short 95% HPD interval.

The width of the 95% HPD intervals varied among the six different analyses, but generally stem age estimates yielded shorter intervals overall and their width varied regardless of the number and (stem or crown) placement of fossils. The widest intervals were found among crown ages of the FBD14, which might be correlated with the low number of fossils (Fig. 2; Table S4; Heath et al., 2014). Another interesting outcome is that as more extant taxa are sampled, the 95% HPD intervals become narrower, and this appears to be independent from

the number of included fossils, as exemplified by Malvoideae and Malvaceae *s.l.*, with four and 56 fossils, respectively (Fig. 2) and both with short intervals. Groups with few tips and one or no fossils give similar results. The narrowest 95% HPD interval in the four analyses was estimated for Malvoideae, which, in this study, is the most densely sampled clade (91 species representing 79.8% of the genera, and 35.2% of the species). We argue that the comparatively dense extant taxon sampling in this clade underlies the narrow HPD intervals, resulting in more precise age estimates in comparison with the rest of the clades (Soares and Schrago 2015). The same is observed if we compare only the intervals of the malvacean families, where Malvaceae *s.l.* has both the highest taxon sampling and fossil sampling. Thus, an important observation is that precision in our estimates derived from FBD and ND are influenced more by the density of extant taxon sampling than by fossil sampling. It would be important to evaluate directly the individual and joint effect of taxon sampling, phylogenetic resolution, and missing data in the age estimates.

Importantly, as the number of fossils increased, stem ages became older and crown ages younger (observed in 67% of the clades), implying longer periods of time between the origin (stem age) and the diversification (crown age) of clades. This effect probably results from the combination of the root calibration (a hard limit) and the internal estimated nodes, such that as more fossils are included in a clade, the interval of time between its stem and crown “expands”. Thus, both observations suggest that there might be diminishing returns in FBD analyses associated with the inclusion of fossils, but this needs to be explicitly investigated. As described in section 3.3, the turnover rate estimated by FBD model increases as more fossils are included (Table S5). This is probably caused by the interaction of many parameters but specially the turnover parameter is affected by the distribution of fossils (Heath et al., 2014). This increase in turnover rates may underlie the larger difference between stem and crown age. Note that this trend results mostly from increasing the number of crown group fossils (Table S3). Turnover rates and age estimation have a strong influence on independently estimated diversification rates (see Section 4.5).

The time period between stem and crown ages (phylogenetic fuse; Cooper and Forsey, 1998) can explain the evolutionary history and diversification in the early stages of clades. In a phylogeny only containing extant species the phylogenetic fuse is observed as a single branch from origin to diversification. The length of the fuse may involve origin and extinction of lineages, or a time with no lineage origination that preceded the diversification of the crown group.

Our results suggest that including few extant species and few or no fossils may have an effect on estimating a long fuse between stem and crown ages (Fig. 4), while a short fuse might be related to dense taxon sampling (e.g. Malvoideae, see Section 4.2) or relatively high molecular rates (e.g. long branches of Cytinaceae; Figs. 1, 3 and 4). For example, a long fuse was detected for families Sphaerosepalaceae, Sarcolaenaceae, Neuradaceae and Bixaceae (Figs. 3 and 4; Table S4). These families have low species richness and a scarce or non-existent fossil representation. A short fuse was found in Malvaceae *s.l.*, Cytinaceae, Dipterocarpaceae and Malvoideae (Figs. 3 and 4; Table S4), which do not share a common pattern of taxon sampling or fossilization.

A long fuse could be related to low stem lineage diversity but might also indicate high extinction rates (Nagalingum et al., 2011), or low fossil preservation potential. For example, Sarcolaenaceae and Sphaerosepalaceae had low extinction and/or fossilization rates, and only after many million years they generated lineages that survived to the present (i.e., crown diversification). It is necessary to investigate whether the long fuse in some malvacean lineages is due to extinction rates, poor taxon sampling or to a bias against fossil preservation. A short fuse indicates that lineages that differentiated shortly after the origin of the group survived to the present. For example, a relatively short time (ca. 9 Myr) elapsed between the differentiation of Malvaceae *s.l.* from its sister clade and the onset of its diversification, which coincides with its relatively high diversification rates (Fig. 5).

4.3. Divergence time of malvacean lineages and their evolutionary implications

In this study we focused on estimation of divergence times for major lineages within order Malvales. The stem age of Malvales (divergence from Brassicales [Sun et al., 2016]) has been estimated from 78 to 89 (Bell et al., 2010) to 89.44–103.2 Ma (Magallón et al., 2015), and crown age is younger or overlapping with the HPD intervals of the estimates in our study for the families (62.9–93.27 Ma [Magallón et al., 2015]; 80.9 Ma [Tank et al., 2015]; 100.51 Ma [Zanne et al., 2014]). One of the differences with previous studies is that we explicitly focused on obtaining dates within Malvales, therefore, the taxon sampling is denser, and we included an extensive fossil sampling. Commonly, large-scale dating studies of angiosperms use fossils considered to be members of Malvaceae *s.l.* to calibrate the whole order Malvales (Smith et al., 2009; Bell et al., 2010; Magallón et al., 2015), usually following a conservative approach to calibrate phylogenies (Parham et al., 2012), yielding younger ages compared to our results.

So far, few studies have applied the FBD model to estimate ages of plant groups. In a study focused on genus *Thespesia* (Malvoideae), Areces-Berazain and Ackerman (2016) estimated age of core Malvoideae as 81.52 Ma. Our analyses estimated ages between 60 and 69 Ma for this clade in both FBD50 and FBD80 (Figs. S5–S6). We argue that the large difference is caused by the inclusion of the fossil *Hibiscoxylon nyloticum* Kräusel in the previous study. We did not include this fossil, as it has been suggested that its affinity with Malvoideae requires re-evaluation (Friis et al., 2011), and its age is uncertain (Wheeler et al., 1994).

Our estimates show that Malvaceae *s.l.* and Thymelaeaceae, and the most recent common ancestor of the families Dipterocarpaceae, Sarcolaenaceae, Cistaceae, Bixaceae and Sphaerosepalaceae originated around 99–89 Ma (Figs. 3 and 4). At this time “super greenhouse” climate (Cenomanian-Turonian; Huber et al., 2018) probably favored the origin and range expansion of these three clades. Most of these lineages have currently disjunct tropical distribution. One possibility is that the origin of the three clades took place in a tropical area of Gondwanan separated continents based on current and fossil distribution. The separation of Africa and South America was already complete, however the separated landmasses were still close to each other: Madagascar and India were still proximate, and specially so Antarctica and Australia (Boucot et al., 2013), thus dispersal among the Gondwanan continents possibly played a key role in the distribution of these, and probably, all other malvacean lineages (Morley, 2003). This hypothesis is more likely than a Laurasian origin, because it would involve many subsequent and independent colonization of South America, Africa, and Australia.

It was during the Campanian-Early Maastrichtian (ca. 83–69 Ma; Fig. 4) when Dipterocarpaceae, Sarcolaenaceae, Bixaceae, and Sphaerosepalaceae, and the nine subfamilies of Malvaceae *s.l.* originated (although in FBDs2-3 stem age Bombacoideae is at ca. 42 Ma, during Eocene), and when Thymelaeaceae diversified. At that time, warm temperate forests were present in north and south separated by arid areas and a tropical equatorial area (Boucot et al., 2013). Notably, all these groups are predominantly tropical, except for Tiliaceae, which is mainly temperate, and Thymelaeaceae has temperate and dry-tolerant members in the Old World. Thus, probably climate change, combined with zonal differences (i.e., arid vs tropical/warm temperate), favored lineage differentiation, for example in the case of Tiliaceae or Thymelaeaceae.

At the end of the Cretaceous and beginning of the Paleogene temperatures further decreased, transforming the climate into a “cool greenhouse” (Huber et al., 2018). At this time the diversification of Dipterocarpaceae took place. During this time boreotropical forests were established in North America and Eurasia (Herold et al., 2014), so considering the fossil *Parashorea pseudogoldiana* (Holl.) Wolfe (Dipterocarpaceae; Wolfe, 1977), dated to the Eocene of Alaska, it would suggest that Dipterocarpaceae dispersed towards North America, from

Gondwanan landmasses, and became part of the boreotropical forest by the Eocene. However, a direct analysis needs to be performed to corroborate the assignment of this fossil. Even if this fossil is not considered, there is no evidence that dipterocarpacean lineages dispersed to Eurasia from Alaska, and the group may have reached Asia via India sometime between the Middle and Late Eocene (47–33 Ma), as suggested by Dutta et al. (2011). However, it is also possible that the origin of Dipterocarpaceae and Sarcolaenaceae took place in the Paleocene, and the diversification of Dipterocarpaceae was at the Early Eocene (56–47.8 Ma), given the estimates obtained when only highly-scored fossils were included (Fig. 4; Figs. 9 and 10).

The early Eocene was characterized by a great increase in global temperatures (Paleocene-Eocene Thermal Maximum; Zachos et al., 2008), and a low equator-to-pole thermal gradient (Herold et al., 2014). This period of high temperatures promoted tropical and subtropical environments worldwide, evidenced by the origin and diversification of plant clades in the fossil record (Jaramillo et al., 2006), and the dispersal of thermophilic groups among plates (Boucot et al., 2013; Herold et al., 2014; Morley, 2003). Notably, most of the fossils in our study belong to this time (Supplementary File 1). This time also coincides with the diversification of Cistaceae and Cytinaceae, the former distributed in temperate and subtropical areas but with its highest species richness found in the Mediterranean region. The latter is distributed in Africa, Madagascar, Mesoamerica, and the Mediterranean where they are parasites on Cistaceae species. Probably, the radiation of Mediterranean species took place more recently.

After this, from late Eocene through the middle Miocene (37.9–13.82 Ma) a continuous decrease in temperatures marked the beginning of the modern “icehouse” climate with the addition of an increment in seasonality and aridity (Eldrett et al., 2009; Pound and Salzmann, 2017). For some tropical groups the distribution range was probably reduced (e.g., Sarcolaenaceae, Dipterocarpaceae, Bombacoideae), whereas others maintained their restricted tropical area (e.g. Muntingiaceae, Brownlowioideae). In turn, Neuradaceae, a drought-tolerant family, diversified during this time, probably favored by the extension of arid regions in Africa.

The fossil record of Malvales provides evidence that the group was widespread at the beginning of the Tertiary. How the malvacean lineages contributed to the establishment of modern biomes is a question that can be addressed in analyses combining their geographic distribution and fossil abundance (e.g. Meseguer et al., 2018).

4.4. Malvales diversification dynamics

The taxonomic sampling in this study was designed to achieve representation at the genus level, which is suboptimal for an accurate estimation of diversification parameters of the FBD model. Therefore, our estimates of FBD-estimated diversification parameters should be considered as tentative. Additionally, we estimated the net diversification rate of Malvales with the m-o-m applied to the ND14 (0.078–0.062 spp./Myr), FBD80, and FBDs2-3 (0.06 spp./Myr; Fig. 5; Table S6). The latter is higher than the one estimated with the FBD model (0.04 spp./Myr; Table S5) and was obtained considering only high relative extinction (section 2.5). This difference is the result of considering the entire species richness in the m-o-m.

Magallón and Castillo (2009) estimated a very high net diversification rate for crown Malvales, in fact, the rate estimated by them was the highest among any eudicot group (0.18–0.23 spp./Myr). The difference between the previous and the current diversification rates derives from the difference between crown ages used in the previous study and in this one (33.9 and 102 Ma, respectively). The young age estimated by them was influenced by the use of a Late Eocene fossil as the oldest calibration for the group, which estimated a very young age, and in turn, resulted in a high diversification rate estimate. The crown age obtained in the present study resulted in a rate similar to the stem diversification rate (0.069 [$\epsilon = 0.9$]) estimated by Magallón and Castillo

(2009).

Families with the highest diversification rates are Sarcolaenaceae and Malvaceae s.l., when using crown and stem age, respectively, and the lowest rates were found for Cytinaceae and Muntingiaceae. The high net diversification rate of crown Sarcolaenaceae (Fig. 5; Table S6) and its long fuse (Fig. 4) may indicate that there was low diversification, probably due to low speciation and low extinction, at the time of origin but experienced a substantial increase in the rate at the point of diversification. Notably, considering low extinction, Sarcolaenaceae and Sphaerosepalaceae have the highest diversification rate, this is interesting since these families are endemic to Madagascar. We speculate that the opportunity to colonize this region might be associated to the high diversification rate in these families. In turn, Neuradaceae and Bixaceae are characterized by a long fuse and a low diversification rate (Figs. 4 and 5; Table S6). An interesting question for these depauperate families would be whether the low rate is due to high extinction or to low speciation.

Malvaceae s.l. is characterized by short fuse and a relatively high diversification rate, obtained from stem and crown ages (Figs. 4 and 5; Table S6). Among malvacean families, Malvaceae s.l. has the highest species richness, it includes a wide variety of growth forms, flower and fruit morphology, and biome occupation, thus possibly there is an association between diversification rates and the morphological and ecological evolution of this family. Particularly, Malvoideae have relatively high stem and crown diversification rates (Fig. 5), probably facilitated by the herbaceous habit. Similar to Malvaceae s.l., Dipterocarpaceae has a short fuse (Fig. 4), but comparatively low diversification rate (Fig. 5; Table S6). This may be associated with the restricted distribution of the family in one biome—the Dipterocarpaceae forest—and the low opportunity they have to colonize other environments, also related to the strong mycorrhizae association this family has. It would be interesting to investigate how these two families took markedly different evolutionary paths, and how these different routes have been affected by morphological and ecological evolution.

4.5. The fossil record of Malvales as a clue for macroevolutionary processes

Observing our fossil database, the sample representing Malvaceae s.l., Dipterocarpaceae and Thymelaeaceae is substantially denser than the sampling of the remaining seven families. No fossil record is reported for Sphaerosepalaceae, Neuradaceae and Cytinaceae, which are characterized by a general low species richness compared to other malvacean families. In addition to species richness, according to the comparative study of Xing et al. (2016), there is a correlation between families with no fossils and herbaceous habit, which is the case of Neuradaceae and Cytinaceae, and African distribution, which is the case of the three malvacean families with fewer or no fossils (i.e., Sarcolaenaceae, Sphaerosepalaceae, Neuradaceae) and Cytinaceae, which has African and Mediterranean members.

Simply because of their longer prevalence, the probability of preservation is higher for older groups than for younger groups. It is therefore not surprising that the oldest family, Malvaceae s.l., is most highly represented in the fossil record. Moreover, most of the fossils of Malvaceae s.l. occur in Paleocene North American sediments, which is a shared pattern with other plant families (Xing et al., 2016).

All these reasons explain the absence of fossils in particular families (i.e., crown group fossils), but cannot be applied to explain the absence of fossils belonging to the earliest branches within crown Malvales, before the origin of families. This may be explained by the short interval of time elapsed between the deepest phylogenetic split within Malvales, and the differentiation of the families (Fig. 3), leaving a short period for preservation, and of morphological evolution to allow identifiability. Leaving aside the biases in the fossil record, the non-uniformly distributed fossils among the families of Malvales could be an indicator of macroevolutionary processes, such as diversification patterns (e.g. species richness of Sphaerosepalaceae vs Malvaceae s.l.),

origin and diversification age (e.g. old vs young clades), and rapid origination of families (e.g. no fossils of stem or crown Malvales).

5. Conclusions

We estimated phylogenetic relationships within Malvales with the largest molecular data matrix assembled so far, which resulted in mostly well supported relationships among the families. We used this phylogenetic hypothesis to investigate the timing and rate of diversification of major malvacean clades. This phylogenetic hypothesis represents a useful framework for macroevolutionary studies of the order, which so far, have been difficult to implement because of lack of resolution. Nevertheless, we highlight the need to use additional molecular markers to address important unresolved relationships, specifically, the position of Muntingiaceae and Cytinaceae.

Differences in age estimates among the six analyses implemented in this study are not pronounced, hence we are confident that the results obtained here regarding the origin of malvacean lineages provide a realistic picture of the evolutionary timing within the order. By implementing these six analyses, we found that the number of included fossils has a stronger influence on crown age estimates than on stem ages; that there is greater precision associated to the estimates of older ages; and a greater precision in FBD estimates compared to ND. We also found that age precision increases in clades that are more densely sampled for extant and fossil taxa.

An interesting outcome from our analyses is the observed temporal distance between stem and crown ages, particularly among poorly sampled groups. A long fuse may be a consequence of high extinction rate, a bias in the fossil sampling, or due to low taxon and fossil sampling. The (macroevolutionary) causes and effects of the fuse remains to be investigated.

The origin of most of the malvacean clades, including the subfamilies of Malvaceae *s.l.*, took place between the Upper Cretaceous and the Early Paleocene (ca. 100–62 Ma), which is an older time than previously estimated. By the early Miocene (23.03–20.44 Ma) all the clades had diversified. This marks a period of ca. 80 Myr of origin and diversification of the order, during which several important geological and environmental changes took place. The timing of origin and diversification here estimated will be critical to investigate the role of Malvales played in the establishment and assembly of particular modern tropical biomes (e.g., seasonally dry tropical forests) of which, members of Malvales are significant components. Here, we highlighted several different avenues of research to better understand the evolution of the group and its development into a significant component of different biomes at the global level.

Funding

This study benefitted from the financial support of PAPIIT UNAM (number IG200316). RHG received a doctoral scholarship from CONACyT (number 507682/288658). RHG also received financial support from the Posgrado en Ciencias Biológicas, UNAM.

Acknowledgements

Computational work was partially developed in the BioLinux Server of the Instituto de Biología-UNAM. We thank Georgina Ortega-Leite, Alfredo Wong, Luis Eguarte, Adriana Benítez-Villaseñor, Luna L. Sánchez-Reyes, Santiago Ramírez-Barahona and Andrea López-Martínez for critical help and comments during the different stages of this study. RHG thanks the Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México, PAPIIT UNAM, and CONACyT.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ympev.2019.106606>.

References

- Alverson, W.S., Karol, K.G., Baum, D.A., Chase, M.W., Swensen, S.M., McCourt, R., Syste, K.J., 1998. Circumscription of the Malvales and relationships to other Rosidae: evidence from *rbcl* sequence data. *Am. J. Bot.* 85 (6), 876–887.
- Alverson, W.S., Whitlock, B.A., Nyffeler, R., Bayer, C., Baum, D.A., 1999. Phylogeny of the core Malvales: evidence from *ndhF* sequence data. *Am. J. Bot.* 86 (10), 1474–1486.
- Aparicio, A., Martín-Hernanz, S., Parejo-Farnés, C., Arroyo, J., Lavergne, S., Yesilyurt, E.B., Zhang, M.-L., Rubio, E., Albaladejo, R.G., 2017. Phylogenetic reconstruction of the genus *Helianthemum* (Cistaceae) using plastid and nuclear DNA-sequences: systematic and evolutionary inferences. *Taxon* 66 (4), 868–885.
- Arcila, D., Pyron, R.A., Tyler, J.C., Orti, G., Betancur-R, R., 2015. An evaluation of fossil tip-dating versus node-age calibrations in tetraodontiform fishes (Teleostei: Percomorphaceae). *Mol. Phylogenet. Evol.* 82, 131–145.
- Areces-Berzain, F., Akerman, J.D., 2016. Phylogenetics, delimitation and historical biogeography of the pantropical tree genus *Thespesia* (Malvaceae, Gossypieae). *Bot. J. Linn. Soc.* 181, 171–198.
- Baum, D.A., Alverson, W.S., Nyffeler, R., 1998. A Durian by any other name: taxonomy and nomenclature of the core Malvales. *Harvard Papers Bot.* 3 (2), 315–330.
- Baum, D.A., Smith, S.D., Yen, A., Alverson, W.S., Nyffeler, R., Whitlock, B.A., Oldham, R.L., 2004. Phylogenetic relationships of Malvaceae (Bombacoideae and Malvoideae; Malvaceae *sensu lato*) as inferred from plastid DNA sequences. *Am. J. Bot.* 91 (11), 1863–1871.
- Bayer, C., Chase, M.W., Fay, M.F., 1998. Muntingiaceae, a new family of dicotyledons with malvacean affinities. *Taxon* 74 (1), 37–42.
- Bayer, C., Fay, M.F., De Bruijn, A.Y., Savolainen, V., Morton, C.M., Kubitzki, K., Alverson, W.S., Chase, M.W., 1999. Support for an expanded family concept of Malvaceae within a recircumscribed order Malvales: a combined analysis of plastid *atpB* and *rbcl* DNA sequences. *Bot. J. Linn. Soc.* 129, 267–303.
- Bell, C.D., Soltis, D.E., Soltis, P.S., 2010. The age and diversification of the angiosperms re-visited. *Am. J. Bot.* 97 (8), 1296–1303.
- 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 Comput. Biol.* 10 (4), e1003537.
- Boucot, A.J., Chen, X., Scotese, C.R., 2013. Phanerozoic Paleoclimate: An Atlas of Lithologic Indicators of Climate. Society for Sedimentary Geology.
- Bromham, L., Cowman, P.F., Lanfear, R., 2013. Parasitic plants have increased rate of molecular evolution across all three genomes. *BMC Evol. Biol.* 13, 126.
- Brunk, U., Mueller, A.N., 2012. A new tribal classification of Grewioideae (Malvaceae) based on Morphological and molecular phylogenetic evidence. *Syst. Bot.* 37 (3), 699–711.
- Cohen, K.M., Finney, S.C., Gibbard, P.L., Fan, J.-X., 2015. The ICS international chronostratigraphic chart. *Episodes* 36, 199–204.
- Conover, J.L., Karimi, N., Stenz, N., Ané, C., Grover, C.E., Skema, C., Tate, J.A., Wolff, K., Logan, S.A., Wendel, J.F., Baum, D.A., 2018. A Malvaceae mystery: a mallow maelstrom of genome multiplications and maybe misleading methods? *J. Integr. Plant Biol.* 61 (1), 12–31.
- Cooper, A., Fortey, R., 1998. Evolutionary explosions and the phylogenetic fuse. *Trends Ecol. Evol.* 13, 151–156.
- Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9, 772.
- Ducousso, M., Béna, G., Courgeois, C., Buyck, B., Eyssartier, G., Vincelette, M., Rabevohitra, R., Randrihasipara, L., Dreyfus, B., Prin, Y., 2004. The last common ancestor of Sarcocaulaceae and Asian dipterocarpaceae trees was ectomycorrhizal before the India-Madagascar separation, about 88 million years ago. *Mol. Ecol.* 13, 231–236.
- Dutta, S., Tripathi, S.M., Mallik, M., Mathews, R.P., Greenwood, P.F., Rao, M.R., Summons, R.E., 2011. Eocene out-of-India dispersal of Asian dipterocarpaceae. *Rev. Palaeobot. Palynol.* 166, 63–68.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* 32 (5), 1792–1797.
- Eldred, J.S., Greenwood, D.R., Harding, I.C., Huber, M., 2009. Increased seasonality through the Eocene to Oligocene transition in northern high latitudes. *Nature* 459, 969–974.
- Foster, C.S.P., Sauquet, H., van Der Merwe, M., McPherson, H., Rossetto, M., Ho, S.Y.W., 2016. Evaluating the impact of genomic data and priors on Bayesian estimates of the angiosperm evolutionary timescale. *Syst. Biol.* 66 (3), 338–351.
- Friis, E.M., Crane, P.R., Pedersen, K.R., 2011. Early Flowers and Angiosperm Evolution. Cambridge University Press, pp. 585.
- Gavryushkina, A., Welch, D., Stadler, T., Drummond, A.J., 2014. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLoS Comput. Biol.* 10 (12), e1003919.
- 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. *Syst. Biol.* 66 (1), 57–73.
- Guzmán, B., Vargas, P., 2009. Historical biogeography and character evolution of Cistaceae (Malvales) based on analysis of plastid *rbcl* and *trnL-trnF* sequences. *Org. Divers. Evol.* 9, 83–99.
- Heath, T.A., Huelsenbeck, J.P., Stadler, T., 2014. The fossilized birth-death process for

- coherent calibration of divergence-time estimates. PNAS E2957–E2966.
- Heath, T., 2017. Divergence Time Estimation using BEAST v2.* Dating Species Divergences with the Fossilized Birth-Death Process. Source URL: < <https://taming-the-beast.github.io/tutorials/FBD-tutorial/> > .
- Heckenhauer, J., Samuel, R., Ashton, P.S., Turner, B., Barfuss, M.H.J., Jang, T.-S., Temsch, E.M., McCan, J., Salim, K.A., Attanayake, A.M.A.S., Chase, M.W., 2017. Phylogenetic analyses of plastid DNA suggest a different interpretation of morphological evolution than those used as the basis for previous classifications of Dipterocarpaceae (Malvales). Bot. J. Linn. Soc. 185 (1), 1–26.
- Herold, N., Buzan, J., Seton, M., Goldner, A., Green, J.A., Müller, R.D., Markwick, P., Huber, M., 2014. A suit of early Eocene (–55 Ma) climate model boundary conditions. Geosci. Model Dev. 7, 2077–2090.
- Huber, B.T., MacLeod, K.G., Watkins, D.K., Coffin, M.F., 2018. The rise and fall of the Cretaceous hot greenhouse climate. Global Planet. Change 167, 1–23.
- Jaramillo, C., Rueda, M.J., Mora, G., 2006. Cenozoic plant diversity in the Neotropics. Science 311, 1893–1896.
- Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software Version 7: improvements in performance and usability. Mol. Biol. Evol. 30 (4), 772–780.
- Kubitzki, K., Chase, M.W., 2003. Introduction to Malvales. In: Kubitzki, K., Bayer, C. (Eds.), Flowering plants, Dicotyledons, Malvales, Capparales and non-betulin Caryophyllales. Springer, Berlin, pp. 12–16.
- Le Pechón, T., Gigord, L.D.B., 2014. On the relevance of molecular tools for taxonomic revision in Malvales, Malvaceae s.l. and Dombeyoideae. Molecular Plant Taxonomy, Methods and Protocols. Springer, New York.
- Magallón, S., Castillo, A., 2009. Angiosperm diversification through time. Am. J. Bot. 96 (1), 349–365.
- Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L.L., Hernández-Hernández, T., 2015. A metacalibrated time-tree documents the early rise of flowering plant Phylogenetic diversity. New Phytol. 207, 437–453.
- Magallón, S., Sánchez-Reyes, L.L., Gómez-Acevedo, S., 2018. Thirty clues to the exceptional diversification of flowering plants. Ann. Bot. mcy182. <https://doi.org/10.1093/aob/mcy182>.
- Magallón, S., Sanderson, M.J., 2011. Absolute diversification rates in angiosperm clades. Evolution 55 (9), 1762–1780.
- Meseguer, A.S., Lobo, J.M., Cornuault, J., Beerling, D., Ruhfel, B.R., Davis, C.C., Jousset, E., Sanmartín, I., 2018. Reconstructing deep-time palaeoclimate legacies in the clusoid Malpighiales unveils their role in the evolution and extinction of the boreo-tropical flora. Glob. Ecol. Biogeogr. 27, 616–628.
- 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), 14 Nov. 2010, New Orleans, LA, pp. 1–8.
- Morley, R.J., 2003. Interplate dispersal paths for megathermal angiosperms. Perspectives in Plant Ecology. Evol. Syst. 6 (1,2), 5–20.
- Nagalingum, N.S., Marshall, C.R., Quental, T.B., Rai, H.S., Little, D.P., Mathews, S., 2011. Recent synchronous radiation of a living fossil. Science 334, 796–799.
- Nee, S., 2006. Birth-Death models in Macroevolution. Annu. Rev. Ecol. Syst. 37, 1–17.
- Nickrent, D.L., 2007. Cytinaceae are sister to Muntingiaceae (Malvales). Taxon 56 (4), 1129–1135.
- Oliveira-Filho, A.T., 2017. NeoTropTree, Flora arbórea da Região Neotropical: Um banco de dados envolvendo biogeografia, diversidade e conservação. Universidade Federal de Minas Gerais <http://www.neotropree.info>.
- 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.R., Patané, J.S., Smith, N.D., Tarver, J.E., van Tuinen, M., Yang, Z., Angielczyk, K.D., Greenwood, J.M., Hipsley, C.A., Jacobs, L., Makovicky, P.F., Müller, J., Smith, K.T., Theodor, J.M., Warnock, R.C., Benton, M.J., 2012. Best practice for justifying fossil calibrations. Syst. Biol. 61 (2), 346–359.
- Pennell, M.W., Eastman, J.M., Slater, G.J., Brown, J.W., Uyeda, J.C., FitzJohn, R.G., Alfaro, M.E., Harmon, L.J., 2014. geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics 30, 2216–2218.
- Pennington, R.T., Lavin, M., Oliveira-Filho, A., 2009. Woody plant diversity, evolution, and ecology in the tropics: perspectives from seasonally dry tropical forests. Annu. Rev. Ecol. Syst. 40, 437–457.
- Pound, M.J., Salzmann, U., 2017. Heterogeneity in global vegetation and terrestrial climate change during the late Eocene to early Oligocene transition. Sci. Rep. 7, 43386.
- R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. < <https://www.R-project.org/> > .
- Rambaut, A., Suchard, M.A., Xie, D., Drummond, A.J., 2014. Tracer v1.6. Available from < <http://tree.bio.ed.ac.uk/software/tracer/> > .
- Rambaut, A., 2014. Figtree, a graphical viewer of phylogenetic trees. Tree.bio.ed.ac.uk/software/figtree.
- Renner, S.S., Grimm, G.W., Kapli, P., Denk, T., 2017. Species relationships and divergence times in beeches: new insights from the inclusion of 53 young and old fossils in a birth-death clock model. Philos. Trans. Roy. Soc. B 371, 20150135.
- 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. Syst. Biol. 61 (3), 593–1542.
- Saladin, B., Leslie, A.B., Wüest, R.O., Litsios, G., Conti, E., Salamin, N., Zimmermann, N.E., 2017. Fossils matter: improved estimates of divergence times in Pinus reveal older diversification. BMC Evol. Biol. 17, 95.
- 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). Syst. Biol. 6 (2), 289–313.
- Silvestro, D., Cascales-Miñana, B., Bacon, C.D., Antonelli, A., 2015. Revisiting the origin and diversification of vascular plants through a comprehensive Bayesian analysis of the fossil record.
- Slater, G.J., Harmon, L.J., Alfaro, M.E., 2012. Integrating fossils with molecular phylogenies improves inference of trait evolution. Evolution 66 (12), 3931–3944.
- Smith, S.A., Beaulieu, J.M., Donoghue, M.J., 2009. Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches. BMC Evol. Biol. 9, 37.
- Soares, A.E.R., Schrago, C.G., 2015. The influence of taxon sampling on Bayesian divergence time inference under scenarios of rate heterogeneity among lineages. J. Theor. Biol. 364, 31–39.
- Sosef, M.S.M., Dauby, G., Blach-Overgaard, A., van der Burgt, X., Catarino, L., Damen, T., Deblauwe, V., Dessein, S., Dransfield, J., Droissart, V., Duarte, M.C., Engledow, H., Fader, G., Figueira, R., Gereau, R.E., Hardy, O.J., Harris, D.J., de Heij, J., Janssens, S., Komberg, Y., Ley, A.C., Mackinder, B.A., Meerts, P., van de Poel, J.L., Sonké, B., Stévant, T., Stoffelen, P., Svenning, J.-C., Sepulchre, P., Zais, R., Wieringa, J.J., Couvreur, T.L.P., 2017. Exploring the floristic diversity of tropical Africa. BMC Biol. 15, 15.
- Stadler, T., 2010. Sampling-through-time in birth-death trees. J. Theor. Biol. 267, 396–404.
- Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30 (9), 1312–1313.
- Sun, M., Naeem, R., Su, J.-X., Cao, Z.-Y., Burleigh, G., Soltis, P.S., Soltis, D.E., Chen, Z.-D., 2016. Phylogeny of the Rosidae: a dense taxon sampling analysis. J. Syst. Evol. 54 (4), 363–391.
- Tank, D.C., Eastman, J.M., Pennell, M.W., Soltis, P.S., Soltis, D.E., Hinchliff, C.E., Brown, J.W., Sessa, E.B., Harmon, L.J., 2015. Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. New Phytol. 207, 454–467.
- Van der Bank, M., Fay, M.F., Chase, M.W., 2002. Molecular phylogenetics of Thymelaeaceae with particular reference to African and Australian genera. Taxon 51, 329–339.
- Webb, C.O., Cannon, C.H., Davies, S.J., 2008. Ecological organization, biogeography, and the phylogenetic structure of tropical forest tree communities. In: Carson, W.P., Schnitzer, T.A. (Eds.), Tropical Forest Community Ecology. Wiley-Blackwell, New York, pp. 79–97.
- Wheeler, E.A., Lehman, T.M., Gasson, P.E., 1994. *Javelinoxylon*, an upper Cretaceous dicotyledonous tree from Big Bend National Park, Texas, with presumed malvacean affinities. Am. J. Bot. 81 (6), 703–710.
- Wikström, N., Savolainen, V., Chase, M.W., 2001. Evolution of the angiosperms: calibrating the family tree. Proc. Roy. Soc. Lond. B 268, 2211–2220.
- Wolfe, J.A., 1977. Paleogene floras from the Gulf of Alaska region. U.S. Geological Survey Professional Paper 997, pp. 1–108.
- Xing, Y., Gandolfo, M.A., Onstein, R.E., Cantrill, D.J., Jacobs, B.F., Jordan, G.J., Lee, D.E., Popova, S., Srivastava, R., Su, T., Vikulin, S.V., Yabe, A., Linder, H.P., 2016. Testing the biases in the rich Cenozoic angiosperm macrofossil record. Int. J. Plant Sci. 177 (4), 371–388.
- Zachos, J.C., Dickens, G.R., Zeebe, R.E., 2008. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. Nature 451, 279–283.
- Zanne, A.E., Tank, D.C., Cornwell, W.K., Eastman, J.M., Smith, S.A., FitzJohn, R.G., McGlenn, D.J., O'Meara, B.C., Moles, A.T., Reich, P.B., Royer, D.L., Soltis, D.E., Stevens, P.F., Westoby, M., Wright, I.J., Aarssen, L., Bertin, R.I., Calaminus, A., Govaerts, R., Hemmings, F., Leishman, M.R., Oleksyn, J., Soltis, P.S., Swenson, N.G., Warman, L., Beaulieu, J.M., 2014. Three keys to the radiation of angiosperms into freezing environments. Nature 506, 89–92.
- Zervas, A., Petersen, G., Seberg, O., 2019. Mitochondrial genome evolution in parasitic plants. BMC Evol. Biol. 19, 87.

CAPÍTULO I: MATERIAL SUPLEMENTARIO

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: 106606.

TABLES

Table S1. Taxon sampling information. GeneBank accession data.

Table S2. Sampling proportion per loci.

Table S3. Summary of fossils placed in crown and stem group of each lineage.

Table S4. Divergence time estimates from the six different analyses.

Table S5. Net diversification rate estimated with the FBD model.

Table S6. Diversification rates estimated with the method-of-moments for each family of the Malvales and for the subfamilies of Malvaceae *s.l.*

Table S7. Sampled ancestor frequencies.

FIGURES

Figure S1. Bayesian inference tree.

Figure S2. Maximum Likelihood tree.

Figure S3. Bayesian inference tree with Cytinaceae.

Figure S4. Maximum Likelihood tree with Cytinaceae.

Figure S5. ND chronogram.

Figure S6. FBD14 chronogram.

Figure S7. FBD50 chronogram.

Figure S8. FBD80 chronogram.

Figure S9. FBDs2-3 chronogram.

Figure S10. FBDs3 chronogram.

Figure S11. Net diversification rates estimated with FBD model.

Figure S12. Maximum likelihood gene trees.

Table S1. Taxon sampling. Genbank Accession numbers for each genetic region.

Species	atpB	rbcL	ITS	matK	matR	ndhF	trnH-psbA	trnL
<i>Abelmoschus manihot</i>		KX527167.1	KP222436.1					KC488172.1
<i>Abelmoschus moschatus</i>				MF349855.1				
<i>Abroma augustum</i>	AJ233072.1	AJ012208.1	AJ277462.1	HM488448.1		AF287917.1		JN676056.1
<i>Abutilon theophrasti</i>		HM849734.1	KT779105.1	HQ696683.1			MH197410.1	HQ696727.1
<i>Acropogon bullatus</i>						AY795596.1		
<i>Acaulimalva alismatifolia</i>			AY591809.1	FJ204708.1		FJ204727.1		
<i>Adansonia digitata</i>	GU981688.1	GU981721.1	KF760382.1	AY321168.1		AF111720.1	JN400285.1	AY328150.1
<i>Alcea rosea</i>			EF419544.1	EU346805.1	AY453094.1	EU346847.1	KU198270.1	EF419766.1
<i>Allosidastrum pyramidatum</i>		JQ592490.1		JQ588210.1			HG963758.1	
<i>Alyogyne huegelii</i>						AF384657.1		
<i>Althaea hirsuta</i>		KF997457.1	EF419510.1	EU346794.1		EU346808.1	EF419625.1	EF419716.1
<i>Andeimalva chilensis</i>			AY172199.1	AY213108.1		FJ204730.1	AY184305.1	
<i>Anisodonteia malvastroides</i>			EF419547.1	EU346803.1		EU346848.1		
<i>Anisodonteia triloba</i>		AM235024.1						
<i>Anisoptera marginata</i>	AF035918.1	Y15144.1		AJ581409.1	AJ581409.1	EU002197		
<i>Anotea flavida</i>			U56775.1			U55322.1		
<i>Apeiba tibourbou</i>	AJ233098.1	AJ233145.1		JQ589326.1		AF111762.1	HG963880.1	
<i>Aquilaria beccariana</i>	AJ233079.1	Y15149.1		MF443399.1				MF443413.1
<i>Aquilaria yunnanensis</i>			KU244104.1				KU244077.1	
<i>Argyrodendron peralatum</i>		KF496739.1				AY795597.1		
<i>Arnhemia cryptantha</i>		AJ297236.1		FJ379777.1				
<i>Asterotrichion discolor</i>			AY591811.1	GU045813.1				
<i>Ayenia microphylla</i>				HM488419.1		AF287921.1		HM488377.1
<i>Ayenia micrantha</i>		JQ594192.1						
<i>Azanza garckeana</i>		JF265294.1		JF270648.1				
<i>Bakeridesia yucatanana</i>			JQ753269.1					

<i>Bastardia bivalvis</i>			KT966958.1					
<i>Bastardistrum cinctum</i>			AY591814.1					
<i>Bastardiopsis densiflora</i>		MG833598.1	AY591815.1					
<i>Bastardiopsis eggersii</i>				KJ012469.1			KJ426614.1	
<i>Batesimalva violacea</i>			AY591816.1					
<i>Bdallophytum americanum</i>					AY739012.1			
<i>Bernoullia flammea</i>			HQ658366.1	HQ696685.1		AF111737.1		HQ696732.1
<i>Berrya javanica</i>	AF035896.1	AJ233146.1		AY321182.1		AF111755.1		
<i>Billieturnera helleri</i>			AY591817.1	FJ204699.1		FJ204742.1		
<i>Bixa orellana</i>	AF035897.1	AF022128.1	KY670805.1	FM179929.1	AF520136.1	EU077547.1	HG963562.1	FM179540.1
<i>Brachychiton acerifolius</i>	GU981689.1	GU981722.1		JX495761.1		AF111748.1		
<i>Brachychiton rupestris</i>			KF760360.1					
<i>Briquetia sonorae</i>			AY591818.1					
<i>Brownlowia elata</i>	AF035898.1	AJ233147.1		AY321184.1		AF111756.1		
<i>Burretiodendron esquirolii</i>		AJ233148.1	AY629205.1					AY328162.1
<i>Byttneria aculeata</i>		AF022123.1		AY321196.1		AF111775.1	HG963519.1	
<i>Byttneria filipes</i>	AJ233073.1							HM488385.1
<i>Byttneria urticifolia</i>			AJ277464.1					
<i>Callirhoe involucrata</i>			AY591819.1	KT456898.1		JF799578.1		
<i>Calyculogygas uruguayensis</i>			AY591820.1					
<i>Camptostemon schultzei</i>		AF022120.1	HQ658360.1	AY321162.1		AF111727.1		HQ696726.1
<i>Carica papaya</i>	AF035901.1	M95671.1	JX092051.1	AY483221.1	KU739682.1	AY483248.1	KC867738.1	JX091823.1
<i>Carpodiptera ameliae</i>						AF111757.1		
<i>Catostemma fragrans</i>		JQ626285.1	HQ658370.1	AY589069.1		AF111734.1		HQ696736.1
<i>Cavanillesia platanifolia</i>		GQ981691.1	HQ658371.1	HQ696686.1		AF111728.1	GQ982172.1	HQ696737.1
<i>Ceiba acuminata</i>						AF111736.1		
<i>Ceiba pentandra</i>		JX987572.1	DQ284854.1	HQ696701.1			GQ982177.1	HQ696753.1
<i>Chiranthodendron pentadactylon</i>			HQ658356.1	AY321164.1		AF111722.1		HQ696722.1

<i>Chlamydocola chlamydantha</i>		KC628448.1		KC627779.1				
<i>Christiana africana</i>		AJ233149.1						
<i>Cienfuegosia humbertianus</i>				EF207266.1				
<i>Cienfuegosia tripartita</i>			U56777.1			U55324.1		
<i>Cistus albidus</i>		FJ225860.1	KJ534141.1	DQ092974.1				
<i>Coelostegia borneensis</i>			AF233311.1			AF230258.1		
<i>Cola acuminata</i>		AY082353.1		AY321179.1		AF111759.1		JN676057.1
<i>Cola nitida</i>	AJ233074.1							
<i>Colona floribunda</i>	AJ233103.1	AJ233150.1		AB924978.1		JF809669.1	KR533628.1	
<i>Commersonia bartramia</i>		KX527074.1				AF287926.1	KM895028.1	HQ656696.1
<i>Corchoropsis crenata</i>	FJ848901.1	FJ848897.1	JF939206.1			FJ848912.1		
<i>Corchorus capsularis</i>			FJ527599.1	JQ693595.1		AF523838.1	AWWV01003905.1	JQ609258.1
<i>Corchorus hirsutus</i>		KJ082236.1						
<i>Cotylelobium malayanum</i>				AB006372.1				
<i>Craigia yunnanensis</i>	GU981687.1	AY328189.1	AF460199.1					AY328161.1
<i>Craterosiphon scandens</i>		AJ297235.1						
<i>Crocanthemum argenteum</i>		FJ492026.1						
<i>Cristaria andicola</i>			AY372999.1	FJ204705.1		FJ204746.1	AY371670.1	
<i>Cullenia ceylanica</i>			AF233316.1			AF230254.1		
<i>Cytinus ruber</i>					AY739013.1			
<i>Dais cotinifolia</i>	AJ233094.1	AJ233144.1	AJ744928.1	JX517520.1				
<i>Daphne laureola</i>		HM849946.1	GQ167536.2	JN894952.1		GQ167429.2	GQ167481.2	AY216738.1
<i>Daphnopsis philippiana</i>		HM446790.1	GQ471044.1	HM446682.1			HM446922.1	GQ471067.1
<i>Decaschistia byrnesii</i>						AY589079.1		
<i>Deltaria brachyblastophora</i>		AM398174.1						
<i>Dendrostellera lessertii</i>			AJ549494.1					
<i>Dialyceras parvifolium</i>		AJ402942.1						
<i>Diarthron vesiculosum</i>		AM398181.1	AJ549493.1					AJ308646.1

<i>Dicranolepis disticha</i>		AM398182.1						
<i>Diegodendron humbertii</i>		Y15138.1				EU077550.1		
<i>Diplodiscus paniculatus</i>						AF230252.1		
<i>Dipterocarpus costatus</i>		AB925489.1		KC708337.1				
<i>Dirca palustris</i>		U26322.2	GQ471040.1	HQ593273.1			HQ596675.1	KT165438.1
<i>Dombeya calantha</i>	GU981692.1	GU981733.1	AY083656.1					
<i>Dombeya tiliacea</i>				JX495704.1				
<i>Drapetes muscoides</i>		AJ297237.1	AJ744919.1					
<i>Dryobalanops aromatica</i>				AB006377.1				
<i>Durio zibethinus</i>	AJ233053.1	AF022119.1	MF629780.1	AY321188.1		AF111749.1		
<i>Edgeworthia chrysantha</i>		AJ297920.1	AJ744932.1					AJ308649.1
<i>Enkleia siamensis</i>		AJ297921.1						AY216763.1
<i>Entelea arborescens</i>		JQ933325.1				AF111765.1		
<i>Eremalche rotundifolia</i>			MF964095.1					
<i>Eriolaena candollei</i>			JF939230.1					
<i>Eriotheca longitubulosa</i>		JQ626270.1	HQ658396.1	HQ696717.1				HQ696771.1
<i>Eriotheca ruizii</i>						AF111729.1		
<i>Fioria vitifolia</i>						AF384640.1		
<i>Firmiana platanifolia</i>		AY082355.1	AF460185.1			AY795589.1		AY328156.1
<i>Franciscodendron laurifolium</i>		JN564275.1		JN564141.1		AY795598.1		
<i>Fremontodendron californicum</i>	GU981693.1	GU981726.1	HQ658357.1	AY321165.1		AF111721.1		HQ696723.1
<i>Fuertesimalva limensis</i>			AY172187.1	AY213098.1		FJ204720.1	AY184295.1	
<i>Fumana thymifolia</i>		FJ225850.1	KJ534190	DQ092968.1			KX498486.1	
<i>Gaya atiquipana</i>			AY591825.1	FJ204706.1		FJ204748.1		
<i>Gerrardina foliosa</i>	AY757085.1	AY757086.1		FM179924.1		AY757130.1		FM179535.1
<i>Glossostemon bruguieri</i>						AF287922.1		
<i>Glyphaea brevis</i>						AF111767.1		
<i>Glyphaea tomentosa</i>		JX572629.1		JX517593.1				
<i>Gnidia kraussiana</i>		AJ295267.1	AM159518.1	JQ024965.1				

<i>Goethalsia meiantha</i>		AJ233151.1				JF809670.1		
<i>Gonystylus macrophyllus</i>	AJ233095.1	Y15150.1		MH332506.1				
<i>Gonystylus bancanus</i>								AY216759.1
<i>Gossypoides kirkii</i>		JQ034251.1	U56783.1	AF403563.1		U55329.1		
<i>Gossypium hirsutum</i>	AJ233063.1	X15886.1	KC404827.1	AY321158.1	JX065074.1	U55340.1	HM437903.1	AF031434.1
<i>Grewia kakothamnus</i>							KR735488.1	KR737985.1
<i>Grewia occidentalis</i>	AJ233105.1	AJ233152.1		JX517699.1				
<i>Grewia biloba</i>						AF111769.1		
<i>Grielum humifusum</i>		AJ403023.1						
<i>Guazuma crinita</i>	GU981694.1			JQ589293.1				
<i>Guazuma ulmifolia</i>		GQ981753.1	AY083658.1			AF111776.1	HM446939.1	HQ656780.1
<i>Guichenotia ledifolia</i>				HM488450.1		AF287938.1		HM488408.1
<i>Gyranthera caribensis</i>			HQ658368.1			AF111738.1		HQ696734.1
<i>Gyrinops walla</i>		AM398173.1	MH134151.1					AY216758.1
<i>Halimium calycinum</i>		FJ492039.1	KY651262.1	DQ092973.1				
<i>Hampea appendiculata</i>		GQ981758.1	KT966932.1	AY589062.1		U55327.1	GQ982241.1	
<i>Hannafordia shanesii</i>						AF111778.1		HQ656771.1
<i>Helianthemum nummularium</i>			KY651277.1				FR865094.1	
<i>Helianthemum sqamatum</i>		FJ225851.1		DQ092969.1		EU077551.1		
<i>Helicteres baruensis</i>	AJ233078.1	AJ233127.1						
<i>Helicteres guazumifolia</i>			AF233300.1	AY321186.1		AF111746.1		
<i>Helicteres jamaicensis</i>							KJ426772.1	
<i>Helicteropsis microsiphon</i>				EF207264.1				
<i>Heliocarpus americanus</i>	AJ233106.1	AJ233153.1						
<i>Heliocarpus appendiculatus</i>				JQ589331		AF111772.1		
<i>Helmiopsiella madagascariensis</i>		AJ233129.1	JF939242.1					
<i>Helmiopsis pseudopopulus</i>			JF939239.1					

<i>Heritiera littoralis</i>		AY082358.1	KF032295.1	AY321181.1	AY289659.1	AF111744.1	KJ559427.1	AY328154.1
<i>Hermannia erodioides</i>	AJ233080.1	AJ233130.1						
<i>Hermannia grandiflora</i>						AF287930.1		
<i>Hermannia incana</i>			AJ277467.1					
<i>Hermannia uhligii</i>								KR738175.1
<i>Herrania purpurea</i>		GQ981762.1		GQ982011.1		AF287920.1	GQ982245.1	
<i>Hibiscus rosa-sinensis</i>		AY082359.1	AF460187.1	AY321160.1	DQ110343.1	AY589075.1	JX856893.1	AY328142.1
<i>Hildegardia barteri</i>	AJ233081.1	AJ233131.1		AY321180.1		AF111754.1		
<i>Hoheria angustifolia</i>			AY944585.1	AY944608.1		FJ204722.1		
<i>Hoheria ovata</i>		JQ933363.1						
<i>Hopea hainanensis</i>		AY328198.1		KJ611240.1	AF520138.1			
<i>Howittia trilocularis</i>			AY591832.1			AY589085.1		
<i>Huberodendron patinoi</i>			HQ658367.1			AF111739.1		HQ696733.1
<i>Hudsonia tomentosa</i>		FJ492019.1			AY453097.1			
<i>Humbertiella decaryi</i>				EF207279.1				
<i>Iliamna bakeri</i>			AF271168.1	AF542603.1				
<i>Jarandersonia clemensiae</i>						AF230253.1		
<i>Kearnemalvastrum lacteum</i>			AY591833.1	JF799577.1		JF799580.1		
<i>Kelleria dieffenbachii</i>			FJ572746.1	FJ572793.1				
<i>Keraudrenia hermanniifolia</i>	AJ233082.1	AJ233132.1						
<i>Keraudrenia hillii</i>						AF287939.1		HQ656772.1
<i>Kitaibela vitifolia</i>				EU346804.1		EU346849.1		
<i>Kokia drynarioides</i>			U56784.1	AF403564.1		KY569597.1		
<i>Kosteletzkya velutina</i>				EF207282.1				
<i>Kostermansia malayana</i>			AF233319.1			AF230257.1		
<i>Kydia calycina</i>				EF207261.1				
<i>Lachnaea grandiflora</i>		AJ697768.1	AJ745730.1					
<i>Lagunaria patersonia</i>		JQ412376.1		AY589064.1		AY589084.1		

<i>Lavatera bryoniifolia</i>			EF419440.1	EU346768.1		EU346815.1	EF419550.1	EF419666.1
<i>Lasiopetalum molle</i>						AF287936.1		
<i>Lasiopetalum parviflorum</i>								HQ656773.1
<i>Lawrenzia glomerata</i>			AY591835.1	GU045815.1				
<i>Lebronnecia kokioides</i>			U56778.1			U55325.1		
<i>Lecanophora chubutensis</i>			AY591837.1	FJ204709.1		FJ204744.1	AY371675.1	
<i>Lechea intermedia</i>						KX498362.1		
<i>Lechea tripetala</i>		FJ492022.1						
<i>Leptolaena multiflora</i>		AY157716.1						
<i>Leptonychia pallida</i>		AJ233134.1		KC627708.1		AF287927.1	KC688752.1	
<i>Lethedon cernua</i>		AM398177.1		FJ379778.1				
<i>Luehea divaricata</i>	GU981696.1	GU981728.1					KF421092.1	
<i>Luehea seemannii</i>				JQ589342.1		AF111770.1		
<i>Lueheopsis rugosa</i>		JQ626279.1						
<i>Lysiosepalum involucreatum</i>						HQ656683.1		HQ656774.1
<i>Macrostelia laurina</i>				EF207267.1		EF207299.1		
<i>Malachra fasciata</i>		JQ592510.1		JQ588230.1				
<i>Malachra radiata</i>						KT967031.1	HG963726.1	
<i>Malacothamnus fasciculatus</i>			AY591839.1					
<i>Malacothamnus palmeri</i>				AF542604.1			EF679751.1	
<i>Malope trifida</i>			EF679732.1	AY589060.1		AF111717.1		EF679772.1
<i>Malva canariensis</i>	AJ233065.1	AJ233122.1	EF419459.1	EU346778.1		EU346820.1	EF419577.1	EF419687.1
<i>Malvalthaea transcaucasica</i>			EF419511.1				EF419637.1	
<i>Malvastrum coromandelianum</i>			MH050250	FJ204716.1		FJ204741.1		
<i>Malvastrum guatemalense</i>							HG963497.1	
<i>Malvaviscus arboreus</i>		JQ592520.1		AY589061.1		AF111718.1	HG963543.1	
<i>Malvella sherardiana</i>			EF419546.1	EU346806.1			EF419663.1	EF419768.1

<i>Malvella leprosa</i>		MF963074.1						
<i>Mansonina gagei</i>						AY861681.1		
<i>Matisia cordata</i>	AJ233054.1	AJ233117.1				AF111724.1		HQ696728.1
<i>Matisia palenquiana</i>			HQ658362.1	HQ696684.1				
<i>Megistostegium perrieri</i>				EF207277.1		EF207308.1		
<i>Melhanina ovata</i>								KR738134.1
<i>Melochia tomentosa</i>				JQ589304.1			KJ426820.1	
<i>Melochia corchorifolia</i>		JQ933406.1				AF287946.1		
<i>Microcos hirsuta</i>						JF809668.1		
<i>Microcos latistipulata</i>	AJ233107.1	AJ233154.1						
<i>Microcos paniculata</i>			KP092995.1	HQ415312.1			JN407066.3	AY328196.1
<i>Modiola caroliniana</i>		KJ773689.1	AY172190.1	EF207256.1		EF207287.1	AY184293.1	
<i>Modiolastrum lateritum</i>			AY172191.1	AY213095.1		JF799582.1	AY184292.1	
<i>Monotes glaber</i>		JX572785.1		JX517931.1				
<i>Monotes adenophyllus</i>						EU077548.1		
<i>Monteiroa glomerata</i>			AY591843.1					
<i>Mortoniadendron guatemalense</i>		AJ233155.1		AY321190.1		AY326479.1		
<i>Muntingia calabura</i>	AJ233068.1	Y15146.1	AY328170.1	JQ589349.1		AF111781.1		AY328166.1
<i>Napaea dioica</i>			AJ304940.1					
<i>Nayariophyton zizyphifolium</i>		JQ933373.1						
<i>Neesia strigosa</i>			AF233315.1	AY321189.1				
<i>Neobalanocarpus heimii</i>				AB006383.1				
<i>Neobrittonia acerifolia</i>			AY591844.1					
<i>Neobuchia paulinae</i>				HQ696707.1				HQ696760.1
<i>Nesogordonia tricarpeolata</i>			JF939203.1					
<i>Neurada procumbens</i>	AJ233069.1	U06814.1	KJ004316.1			EU002245.1		
<i>Neuradopsis austroafricana</i>						EU077552.1		
<i>Nitraria retusa</i>	GQ497651.1	U39278.2	KP087772.1	EU002185.1	GU351262.1	EU002246.1	KP087764.1	KP087729.1

<i>Nototriche pedicularifolia</i>			AY217209.1	AY213106.1		FJ204729.1	AY184303.1	KP087728.1
<i>Ochroma pyramidale</i>	AF035910.1	AJ233118.1	AY635547.1	AY321172.1		AF111740.1	JX997370.1	HQ696729.1
<i>Octolepis casearia</i>		AM398178.1				EU077549.1		
<i>Ovidia andina</i>		AJ297222.1	AM159530.1					
<i>Pachira aquatica</i>	AJ233056.1	AJ233119.1	AF028522.1	AY321170.1		AF111732.1		AY328147.1
<i>Pachira quinata</i>		JQ592534.1		HQ696692.1			GQ982310.1	HQ696745.1
<i>Palaua trispala</i>			FJ204683.1	FJ204695.1		FJ204721.1	DQ156322.1	
<i>Papuodendron lepidotum</i>						AF230247.1		
<i>Paramelhantha decaryana</i>	AJ233085.1	AJ233135.1						
<i>Parashorea chinensis</i>		KM267143.1		AY305717.1			KR534161.1	
<i>Passerina filiformis</i>		AJ297225.1	AJ744915.1	JX518022.1				
<i>Pavonia multiflora</i>	AF035916.1	AJ233123.1				AF111719.1		
<i>Peddiea africana</i>		AJ297227.1	AJ744921.1	JX518167.1				
<i>Peddiea involucrata</i>						EU077553.1		
<i>Peltaea ovata</i>		JQ592551.1						
<i>Pentace polyantha</i>	AJ233109.1	AJ233156.1		AY321183.1		AF111758.1		
<i>Pentapetes phoenicea</i>			JF939204.1					
<i>Pentaplaris doroteae</i>	AJ233110.1	AJ233157.1	HQ658358.1	AY321163.1		AY326476.1		HQ696724.1
<i>Periptera punicea</i>			FJ204693.1	FJ204715.1		FJ204751.1		
<i>Perrierophytum humbertii</i>				EF207284.1		EF207315.1		
<i>Phaleria capitata</i>	AJ233096.1	AJ297228.1						
<i>Phaleria chermisideana</i>							KM895257.1	
<i>Phragmotheca ecuadorensis</i>						AF111725.1		
<i>Phymosia umbellata</i>			AY591848.1					
<i>Pimelea physodes</i>		FJ687343.1	FJ687339.1	FJ687341.1				FJ687345.1
<i>Plagianthus regius</i>		JQ933448.1	AY944607.1	AY944630.1		FJ204723.1		
<i>Pseudocorchorus alatus</i>			FR875076.1					
<i>Pseudomonotes tropenbosii</i>		AF030238.1						

<i>Pterocymbium beccarii</i>		FJ976169.1						
<i>Pterocymbium tinctorium</i>						AY795590.1		
<i>Pterygota alata</i>		JX856755.1				AY795599.1	JX856935.1	
<i>Quararibea costaricensis</i>				AY321166.1		AF111723.1		
<i>Quararibea duckei</i>		JQ626033.1	FJ037834.1				FJ038987.2	
<i>Radyera farragei</i>						AF384663.1		
<i>Rayleya bahiensis</i>				HM488447.1		AF287932.1		HM488405.1
<i>Reevesia thyrsoidea</i>	AJ233086.1	AJ233137.1	KP093000.1	AY321187.1		AF111745.1	KP095701.1	JN676074.1
<i>Rhodognaphalon schumannianum</i>		JX572336.1	HQ658380.1	HQ696695.1				HQ696747.1
<i>Rhopalocarpus lucidus</i>		AF022130.1				AF111780.1		
<i>Robinsonella lindeniana</i>			AY591851.1	FJ204711.1		FJ204750.1		
<i>Ruizia cordata</i>	AJ233087.1	AJ233138.1	JF939263.1					
<i>Rulingia magniflora</i>				HM488451.1		AF287933.1		HQ656760.1
<i>Sarcolaena eriophora</i>		KX588771.1				EU081848.1		
<i>Scaphium scaphigerum</i>		GQ436702.1	FJ980420.1				GQ435393.1	
<i>Scaphium linearicarpum</i>						AY795592.1		
<i>Scaphopetalum blackii</i>		KC628038.1		KC627855.1			KC667675.1	
<i>Scaphopetalum amoenum</i>						AF287925.1		
<i>Schoutenia glomerata</i>	AJ233111.1	AJ233159.1		AY321175.1		AY326478.1		
<i>Scleronema micranthum</i>			HQ658369.1			AF111735.1		HQ696735.1
<i>Septotheca tessmannii</i>			HQ658365.1			AY589082.1		HQ696731.1
<i>Seringia arborescens</i>						AF287947.1		HQ656776.1
<i>Shorea obtusa</i>				AB924838.1				
<i>Shorea robusta</i>		KY973249.1	KM514673.1				JX856942.1	
<i>Sida acuta</i>			JN542429.1			AF384661.1	MG601501.1	
<i>Sida ciliaris</i>				JQ588263.1				
<i>Sida cordata</i>		MH767637.1			AF520139.1			
<i>Sida tenuicarpa</i>								KR738073.1
<i>Sidalcea diploscypha</i>			AJ304884.1					

<i>Solmsia calophylla</i>		AJ295261.1		FJ379776.1				
<i>Stellera chamaejasme</i>		KT280161.1	MG516523.1					GU736357.1
<i>Stephanodaphne oblongifolia</i>		AJ745177.1	AJ744924.1	GQ248205.1			GQ248394.1	
<i>Sterculia apetala</i>							HG963798.1	
<i>Sterculia balanghas</i>					DQ401402.1			
<i>Sterculia tragacantha</i>	GU981701.1	AF022126.1		AY321178.1		AF111747.1		
<i>Sterculia foetida</i>								JN676076.1
<i>Struthiola ciliata</i>		AJ297241.1	AJ745756	FJ572803.1				
<i>Synaptolepis alternifolia</i>		AJ297239.1		JX518008.1				
<i>Talipariti tiliaceum</i>		KX909585.1	MF063781	JX517796.1	AY289657.1	KX984270.1	GU135393.2	AB817483.1
<i>Tapiscia sinensis</i>	AF209685.1	AF206825.1		KF224978.1	AF520103.1	EU002266.1	KJ881422.1	FM179541.1
<i>Tarasa capitata</i>			AY172206.1	AY232632.1		FJ204740.1	AY232639.1	
<i>Tetrasida weberbaueri</i>			AY591855.1	FJ204712.1		FJ204749.1		
<i>Thecanthes punicea</i>		AM167540.1	AM162502.1	FJ572788.1				
<i>Theobroma cacao</i>	AJ233090.1	AJ233090.1	AY074729.1	AY321195.1	DQ110362.1	AF287916.1	MF348567.1	HM488410.1
<i>Thomasia solanacea</i>	AJ233091.1	AJ233141.1						
<i>Thomasia angustifolia</i>						HQ656685.1		HQ656777.1
<i>Thymelaea hirsuta</i>	AJ233097.1	Y15151.1	AJ744930.	EU002191.1	GU351317.1			GU736402.1
<i>Tilia americana</i>	GU981686.1	AF022127.1	KF694729.1	AY321191.1		AF111760.1	HQ596866.1	
<i>Tilia mandshurica</i>					AF520107.1			
<i>Tilia paucicostata</i>								AY328159.1
<i>Triplochiton zambesiacus</i>	AJ233092.1	AJ233142.1	AF233301.1	AY321185.1		AF230256.1		
<i>Triumfetta lappula</i>		JQ594277.1		JQ589357.1			HG963658.1	
<i>Triumfetta semitriloba</i>						AF111768.1		
<i>Trochetia uniflora</i>			JF939276.1					
<i>Trochetiopsis ebenus</i>				AY321174.1		AY326477.1		
<i>Tropaeolum majus</i>		L14706.2		AY483224.1	AF520137.1	EU002270.1		AB043665.1
<i>Tropaeolum tricolor</i>	AF035917.1		DQ007295.1					
<i>Tuberaria guttata</i>	AJ233059.1	AJ233120.1	DQ092929.1	DQ092971.1				

<i>Uladendron codesuri</i>						AY589080.1		
<i>Ungeria floribunda</i>			AF233302.1			AF230248.1		
<i>Upuna borneensis</i>				AB006374.1				
<i>Urena lobata</i>		MH767648.1	MH768244.1	EF207260.1		EF207291.1	GU135323.2	
<i>Urocarpidium albiflorum</i>			AY172189.1	AY213097.1		FJ204739.1		
<i>Vatica chevalieri</i>				KJ611243.1			AY184294.1	
<i>Vatica mangachapoi</i>		AY328199.1			AF520132.1			
<i>Waltheria indica</i>		MH767736.1	MH768331.1	MF159472.1			MF143727.1	KR738059.1
<i>Wikstroemia indica</i>			KP093007.1	HQ415322.1			KM895178.1	LC164909.1
<i>Waltheria madagascariensis</i>						AF111773.1		
<i>Wikstroemia canescens</i>		AM398186.1						
<i>Wissadula periplocifolia</i>			AY591858.1			FJ204717.1		

Table S2. Taxon sampling proportion for each genetic region.

Genetic region	Base pairs	Taxa in alignment	% included	Missing taxa	% missing
<i>atpB</i>	1482	55	21.32	203	78.68
<i>ndhF</i>	2186	150	58.14	108	41.86
<i>psbA</i>	695	74	28.68	184	71.32
<i>rbcL</i>	1431	157	60.85	101	39.15
<i>trnK-matK</i>	2784	159	61.63	99	38.37
<i>trnL-F</i>	1364	81	31.39	177	68.6
<i>matR</i>	1795	21	8.14	237	91.86
ITS	855	149	57.75	109	42.25
TOTAL	12592	258			

Table S3. Summary of the number of fossils included for each lineage in the four dating analyses and their placement in crown or stem group.

	FBD14	Crown	Stem	FBD50	Crown	Stem	FBD80	Crown	Stem	FBDs2-3	Crown	Stem	FBDs3	Crown	Stem
Bombacoideae	2	1	1	11	5	6	14	6	8	2	1	1	0	0	0
Brownlowioideae	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0
Byttnerioideae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Dombeyoideae	1	1	0	2	1	1	4	2	2	4	2	2	3	2	1
Grewioideae	1	1	0	2	2	0	4	3	1	4	3	1	0	0	0
Helicterioideae	1	1	0	4	2	2	8	5	3	7	3	4	7	3	4
Malvoideae	1	1	0	3	3	0	4	4	0	3	3	0	1	1	0
Sterculioideae	1	1	0	3	1	2	7	2	5	6	2	4	0	0	0
Tilioideae	1	1	0	2	1	1	5	4	1	3	3	0	2	2	0
TOTAL	8	7	1	27	15	12	47	26	21	29	17	12	13	8	5
Bixaceae	1	0	1	1	0	1	1	0	1	1	0	1	0	0	0
Cistaceae	1	1	0	3	2	1	4	2	2	2	0	2	0	0	0
Cytinaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Dipterocarpaceae	2	1	1	9	7	2	12	9	3	6	3	3	1	0	1
Malvaceae <i>s.l.</i>	9	8	1	32	27	5	56	47	9	38	29	9	21	13	8
Muntingiaceae	0	0	0	1	0	1	1	0	1	1	0	1	0	0	0
Neuradaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sarcolaenaceae	0	0	0	1	0	1	1	0	1	0	0	0	0	0	0
Sphaerosepalaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Thymelaeaceae	1	1	0	3	2	1	5	3	2	2	1	1	0	0	0
TOTAL	14	11	3	50	38	12	80	61	19	50	33	17	22	13	9

Table S4. Estimated ages of the major lineages of the order Malvales. Upper table corresponds to the families of Malvales, the lower table corresponds to the subfamilies of Malvaceae *s.l.* Fossilized Birth-Death process (FBD). Node-dating (ND). Highest Posterior Density (HPD). 95% HPD Max: Maximum bound of the 95% HPD interval. 95% HPD Min: Minimum bound of the 95% HPD interval.

Family	ND14						FBD14						FBD50						FBD80						FBDs2-3						FBDs3					
	Crown Age			Stem age			Crown Age			Stem age			Crown Age			Stem age			Crown Age			Stem age			Crown Age			Stem age								
	Mean	95% HPD Max	95% HPD Min	Mean	95% HPD Max	95% HPD Min	Mean	95% HPD Max	95% HPD Min	Mean	95% HPD Max	95% HPD Min	Mean	95% HPD Max	95% HPD Min	Mean	95% HPD Max	95% HPD Min	Mean	95% HPD Max	95% HPD Min	Mean	95% HPD Max	95% HPD Min	Mean	95% HPD Max	95% HPD Min	Mean	95% HPD Max	95% HPD Min	Mean	95% HPD Max	95% HPD Min			
Bixaceae	36.2	57.4	16.3	74.9	91.1	56.9	32.2	51.3	7.8	74.9	88.7	60.4	37.1	58.7	17.8	77.0	91.0	62.3	34.3	56.9	16.2	77.2	91.7	62.0	31.7	50.5	15.0	73.6	87.1	60.2	32.9	51.9	14.6	68.9	85.7	50.0
Cistaceae	55.5	69.8	41.6	75.9	86.6	65.9	58.6	71.8	46.0	78.3	87.0	69.8	57.4	68.9	45.5	80.2	88.3	72.9	56.1	69.6	42.3	83.2	91.0	75.7	49.8	61.4	38.5	73.6	83.0	64.2	52.2	64.0	40.5	72.1	82.0	62.1
Cytinaceae	66.4	80.8	51.3	73.5	88.5	57.0	63.2	76.6	50.3	69.6	83.8	55.2	62.0	76.0	48.2	68.8	84.9	50.9	60.9	76.6	45.3	69.7	91.4	49.3												
Dipterocarpaceae	52.3	65.1	40.7	59.3	69.3	47.8	60.0	69.5	51.1	66.7	76.7	57.1	65.6	74.0	57.8	71.6	80.2	63.1	69.0	77.1	61.1	74.8	83.2	66.1	53.7	63.9	44.5	62.6	71.6	54.5	51.5	61.1	42.9	59.3	67.1	51.9
Malvaceae s.l.	88.6	93.7	83.5	97.4	102.4	92.5	89.5	95.5	84.0	98.5	102.9	94.3	89.9	95.0	84.7	98.6	102.7	94.3	90.0	95.4	84.7	98.8	102.8	94.6	88.5	94.5	82.1	98.9	103.0	94.3	86.5	93.4	79.8	98.2	102.9	93.1
Muntingiaceae*				97.4	102.4	92.5				98.5	102.9	94.3				98.6	102.7	94.3				98.8	102.8	94.6				98.9	103.0	94.3				98.2	102.9	93.1
Neuradaceae	25.5	52.0	6.7	73.5	88.5	57.0	31.8	56.0	10.2	69.6	83.8	55.2	28.9	51.6	8.1	68.8	84.9	50.9	26.3	48.3	7.6	69.7	91.4	49.3	22.0	40.0	6.5	92.8	101.4	83.8	22.9	41.0	8.5	95.9	102.7	88.5
Sarcolaenaceae	21.5	43.4	2.5	59.3	69.3	47.8	27.0	50.3	5.5	66.7	76.7	57.1	22.3	42.0	4.4	71.6	80.2	63.1	21.3	41.6	3.9	74.8	83.2	66.1	23.6	45.6	4.6	62.6	71.6	54.5	23.5	44.6	5.2	59.3	67.1	51.9
Sphaerosepalaceae	21.9	42.7	5.6	74.9	91.1	56.9	26.2	48.7	7.0	74.9	88.7	60.4	24.9	47.9	6.3	77.0	91.0	62.3	23.2	43.9	5.4	77.2	91.7	62.0	22.6	42.2	6.4	73.6	87.1	60.2	22.2	40.4	6.0	68.9	85.7	50.0
Thymelaeaceae	78.4	90.2	65.9	96.1	102.4	89.2	77.7	86.7	67.5	95.2	101.0	88.1	71.8	86.4	60.2	93.9	101.4	86.4	74.5	85.5	64.1	95.1	102.0	87.6	70.1	81.5	58.4	92.8	101.4	83.8	72.5	85.6	59.1	94.2	102.1	86.1
Malvales (root)	102.0	103.2	99.8				102.3	103.2	100.5				102.3	103.2	100.6				102.4	103.2	100.9				102.3	103.2	100.5				102.1	103.2	100.0			
Subfamily	ND14						FBD14						FBD50						FBD80						FBDs2-3						FBDs3					
	Crown Age			Stem age			Crown Age			Stem age			Crown Age			Stem age			Crown Age			Stem age			Crown Age			Stem age			Crown Age			Stem age		
	Mean	95% HPD Max	95% HPD Min	Mean	95% HPD Max	95% HPD Min	Mean	95% HPD Max	95% HPD Min	Mean	95% HPD Max	95% HPD Min	Mean	95% HPD Max	95% HPD Min	Mean	95% HPD Max	95% HPD Min	Mean	95% HPD Max	95% HPD Min	Mean	95% HPD Max	95% HPD Min	Mean	95% HPD Max	95% HPD Min	Mean	95% HPD Max	95% HPD Min	Mean	95% HPD Max	95% HPD Min	Mean	95% HPD Max	95% HPD Min
Bombacoideae	43.1	53.0	35.9	53.8	67.0	41.5	48.8	60.2	37.6	70.4	72.8	69.0	59.7	68.2	51.6	72.1	76.0	69.0	59.7	68.3	51.4	72.5	76.3	69.0	42.4	49.3	37.6	53.4	66.3	42.6	34.9	46.2	24.8	45.5	62.9	30.7
Brownlowioideae	25.1	42.7	10.2	72.9	82.0	59.0	33.0	51.7	17.7	73.7	84.2	61.2	29.6	47.7	13.4	72.2	83.9	58.3	30.1	49.4	12.4	74.3	83.9	64.9	25.4	41.9	12.2	69.2	79.5	59.0	26.7	42.8	12.9	68.7	79.9	56.9
Byttnerioideae	72.9	83.6	62.5	81.5	89.6	73.4	71.6	82.7	58.7	81.8	91.1	71.5	72.4	83.6	60.1	82.7	91.0	71.4	70.3	82.3	56.4	82.3	90.8	73.3	69.9	80.9	58.4	81.1	89.4	72.5	68.3	79.9	56.3	77.9	87.8	68.2
Dombeyoideae	51.7	65.4	38.7	72.9	82.0	59.0	61.8	73.1	51.9	73.7	84.2	61.2	50.3	62.2	36.9	72.2	83.9	58.3	49.3	63.0	35.9	74.3	83.9	64.9	47.1	57.7	36.3	69.2	79.5	59.0	47.3	59.3	36.1	68.7	79.9	56.9
Grewioideae	62.8	74.5	50.3	81.5	89.6	73.4	65.3	77.0	52.9	81.8	91.1	71.5	67.6	78.6	57.0	82.7	91.0	71.4	64.4	74.4	54.5	82.3	90.8	73.3	64.6	73.2	56.0	81.1	89.4	72.5	56.6	68.9	43.5	77.9	87.8	68.2
Helicterioideae	67.2	83.3	49.3	82.4	88.4	76.6	69.0	82.2	55.1	83.5	89.0	78.6	63.9	77.1	50.2	74.3	82.4	66.1	66.3	79.0	54.3	75.2	82.4	67.9	61.9	74.3	50.8	72.8	80.2	64.9	62.2	74.6	50.5	70.8	78.4	62.3
Malvoideae	65.7	69.5	62.1	68.6	72.2	66.0	68.6	72.7	64.7	72.4	75.7	69.5	71.0	76.0	66.3	74.7	79.2	70.7	71.1	76.1	66.2	74.9	79.4	70.5	68.3	73.5	63.8	71.0	76.2	66.2	66.9	71.0	63.2	69.3	74.1	65.1
Sterculioideae	38.0	52.5	23.6	75.7	81.1	70.7	48.7	63.0	34.3	80.2	84.7	75.8	44.3	58.5	32.2	81.3	87.1	75.7	41.9	57.3	28.5	81.8	87.2	76.5	40.4	53.7	27.9	79.4	85.4	74.2	38.3	51.0	26.3	76.3	82.4	70.7
Tilioideae	58.4	78.2	30.8	78.2	83.6	72.8	58.1	78.5	38.1	78.0	82.2	74.1	72.8	80.3	65.7	79.0	86.1	71.9	72.7	80.4	65.4	79.2	85.6	72.6	72.5	79.5	65.5	76.9	83.7	70.3	71.3	77.8	64.9	75.2	81.6	68.8

Table S5. Diversification and turnover rates estimated with the FBD model for the three different analyses. Speciation (λ) and extinction (μ) rate were obtained from the birth-death model (Nee 2006; see Materials and Methods).

	Diversification rate (r)		Turnover rate (ε)		Speciation rate (λ)	Extinction rate (μ)
	Mean	95% HPD	Mean	95% HPD Interval		
FBD14	0.060	[0.048, 0.071]	0.795	[0.671, 0.897]	0.294	0.233
FBD50	0.04	[0.038, 0.057]	0.880	[0.820, 0.933]	0.400	0.352
FBD80	0.041	[0.033, 0.049]	0.919	[0.883, 0.952]	0.509	0.468
FBDs3-2	0.047	[0.038, 0.057]	0.906	[0.863, 0.946]	0.514	0.466
FBDs3	0.054	[0.043, 0.066]	0.881	[0.818, 0.938]	0.463	0.408

Table S6. Net diversification rates of the families of Malvales estimated with the method-of-moments from ND and FBD80 crown and stem ages.

taxon	ND crown $\varepsilon=0.01$	ND crown $\varepsilon=0.9$	ND crown age	ND stem ε $=0.01$	ND stem $\varepsilon=0.9$	ND stem age	FBD80 crown $\varepsilon=0.91^*$	FBD80 crown age	FBD80 stem $\varepsilon=0.91^*$	FBD80 stem age	FBDs2-3 crown $\varepsilon=0.9^*$	FBDs2-3 crown age	FBDs2-3 stem $\varepsilon=0.9^*$	FBDs2-3 stem age
Bixaceae	0.07	0.03	36.21	0.03	0.02	74.85	0.03	34.33	0.01	77.20	0.04	31.69	0.02	73.61
Cistaceae	0.09	0.06	55.51	0.06	0.04	75.93	0.05	56.09	0.04	83.20	0.06	49.83	0.04	73.62
Cytinaceae	0.03	0.01	66.40	0.02	0.01	73.52	0.01	60.92	0.01	69.69			0.00	98.89
Dipterocarpaceae	0.09	0.06	52.34	0.08	0.05	59.28	0.04	69.02	0.04	74.83	0.05	53.68	0.05	62.62
Malvaceae s.l.	0.09	0.07	88.61	0.08	0.06	97.36	0.07	90.00	0.06	98.84	0.07	88.49	0.06	98.89
Malvales	0.08	0.06	102.01				0.06	102.43			0.06	102.26		
Muntingiaceae				0.00	0.00	97.36			0.00	98.84				
Neuradaceae	0.05	0.02	25.48	0.02	0.01	73.52	0.02	26.34	0.01	69.69	0.02	21.99	0.01	92.77
Sarcolaenaceae	0.16	0.09	21.54	0.06	0.03	59.28	0.08	21.27	0.02	74.83	0.08	23.62	0.03	62.62
Sphaerosepalaceae	0.10	0.05	21.93	0.03	0.01	74.85	0.04	23.24	0.01	77.20	0.04	22.55	0.01	73.61
Thymelaeaceae	0.08	0.06	78.41	0.06	0.05	96.14	0.06	74.54	0.05	95.13	0.07	70.09	0.05	92.77

* The relative extinction fraction (ε) was set with two extreme values for ND, whereas ε for the FBD ages was obtained from the model posterior estimates (see Materials and Methods).

taxon	ND crown $\varepsilon=0.01$	ND crown $\varepsilon=0.9$	ND crown age	ND stem $\varepsilon=0.01$	ND stem $\varepsilon=0.9$	ND stem age	FBD80 crown $\varepsilon=0.91^*$	FBD80 crown age	FBD80 stem $\varepsilon=0.91^*$	FBD80 stem age	FBDs2-3 crown $\varepsilon=0.9^*$	FBDs2-3 crown age	FBDs2-3 stem $\varepsilon=0.9^*$	FBDs2-3 stem age
Malvoideae	0.11	0.08	65.66	0.10	0.08	68.64	0.07	71.13	0.07	74.9	0.08	68.3	0.07	71.0
Bombacoideae**	0.10	0.07	43.11	0.08	0.05	53.83	0.04	59.65	0.04	72.46	0.07	42.4	0.05	53.4
Grewioideae	0.09	0.07	62.76	0.07	0.05	81.52	0.06	64.41	0.05	82.31	0.07	64.6	0.05	81.1
Dombeyoideae	0.10	0.07	51.66	0.07	0.05	72.87	0.07	49.33	0.04	74.31	0.07	47.1	0.05	69.2
Byttnerioideae	0.08	0.06	72.93	0.07	0.05	81.52	0.06	70.25	0.05	82.31	0.06	69.9	0.05	81.1
Sterculioideae	0.13	0.09	38.03	0.07	0.05	75.68	0.08	41.87	0.04	81.76	0.09	40.4	0.04	79.4
Brownlowioideae	0.15	0.09	25.09	0.05	0.03	72.87	0.07	30.13	0.03	74.31	0.09	25.4	0.03	69.2
Helicterioideae	0.06	0.04	67.21	0.05	0.03	82.4	0.04	66.29	0.03	75.2	0.04	61.9	0.04	72.8
Tilioideae	0.06	0.03	58.37	0.04	0.02	78.2	0.02	72.67	0.02	79.2	0.03	72.5	0.02	76.9

*The relative extinction fraction (ε) was set with two extreme values for ND, whereas ε for the FBD ages was obtained from the model posterior estimates (see Materials and Methods).

**Includes only palmately compound leaves (and derivatives; see section 3.1) and excludes *Chiranthodendron*, *Fremontodendron*, *Ochroma*, *Patinoa* and *Septotheca*. These last genera were considered only for the counting to family level.

Table S7. Frequency of sampled ancestors (SA) obtained from the three different FBD analyses.

FBD14			FBD50			FBD80		
Count	Percent	SA	Count	Percent	SA	Count	Percent	SA
300	66.52	<i>Malvaciphyllum macondicus</i>	174	48.47	<i>Malvaciphyllum macondicus</i>	301	39.97	<i>Bombax</i> -type
251	55.65	<i>Bombax</i> -type	158	44.01	<i>Bombax</i> -type	274	36.39	<i>Retimultiporopollenites qiongbeiensis</i>
226	50.11	<i>Shorea albida</i> -type	117	32.59	<i>Jandufouria seamrogiformis</i>	219	29.08	<i>Shorea albida</i> -type
205	45.45	<i>Triumfetta ovata</i>	109	30.36	<i>Dipterocarpuspollenites retipilatus</i>	206	27.36	<i>Malvaciphyllum macondicus</i>
149	33.04	<i>Tripochitioxylon oregonensis</i>	97	27.02	<i>Shorea albida</i> -type	190	25.23	<i>Craigia oregonensis</i>
FBDs2-3			FBDs3					
Count	Percent	SA	Count	Percent	SA			
1464	49.23	<i>Malvaciphyllum macondicus</i>	1489	56.81	<i>Malvaciphyllum macondicus</i>			
827	27.81	<i>Hopenium pondicherriense</i>	858	32.74	<i>Craigia oregonensis</i>			
708	23.81	<i>Luehea divaricatiformis</i>	524	19.99	<i>Triplochitioxylon oregonensis</i>			
626	21.05	<i>Craigia oregonensis</i>	321	12.25	<i>Javelinoxylon weberi</i>			
572	19.23	<i>Apeiba</i> sp.	280	10.68	<i>Shorea maomingensis</i>			

FIGURES

Figure S1. Bayesian inference tree of the eight molecular markers. For methods see section 2.1. Node numbers represent Posterior Probability.

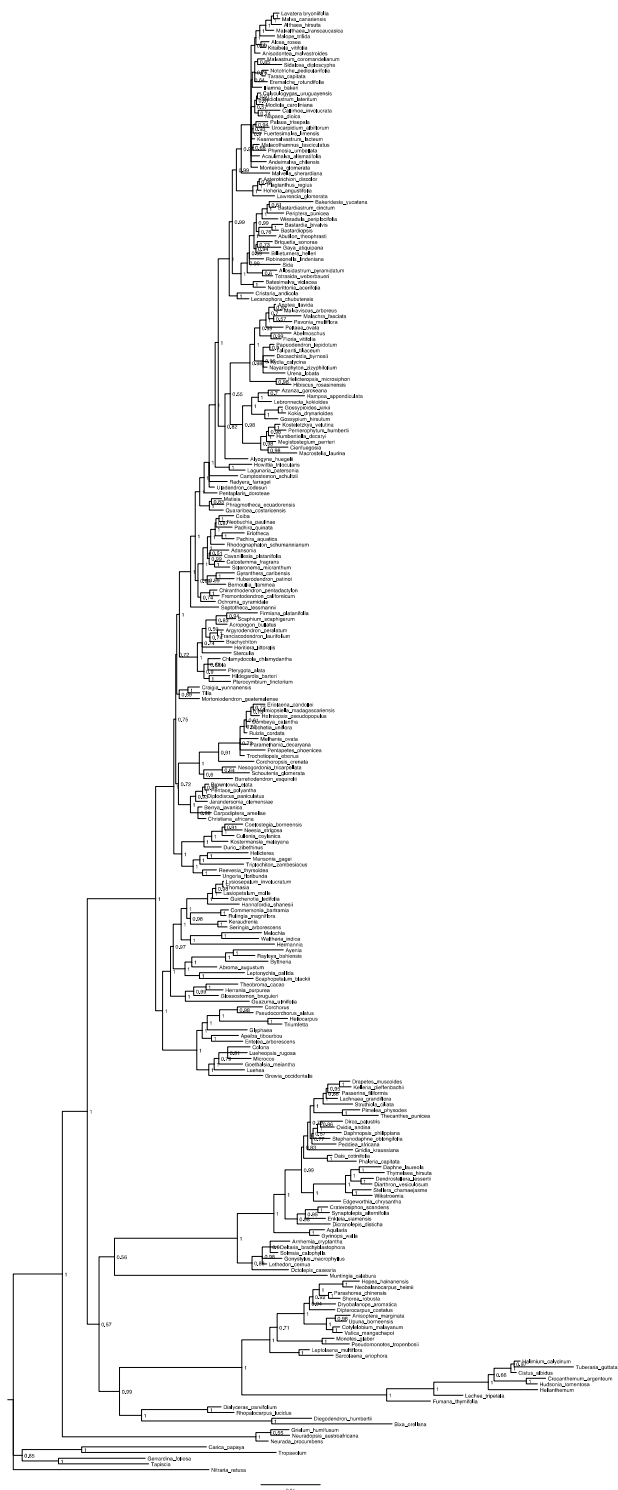


Figure S2. Maximum Likelihood tree of the eight molecular markers concatenated. For methods see section 2.1. Node numbers represent Bootstrap support.

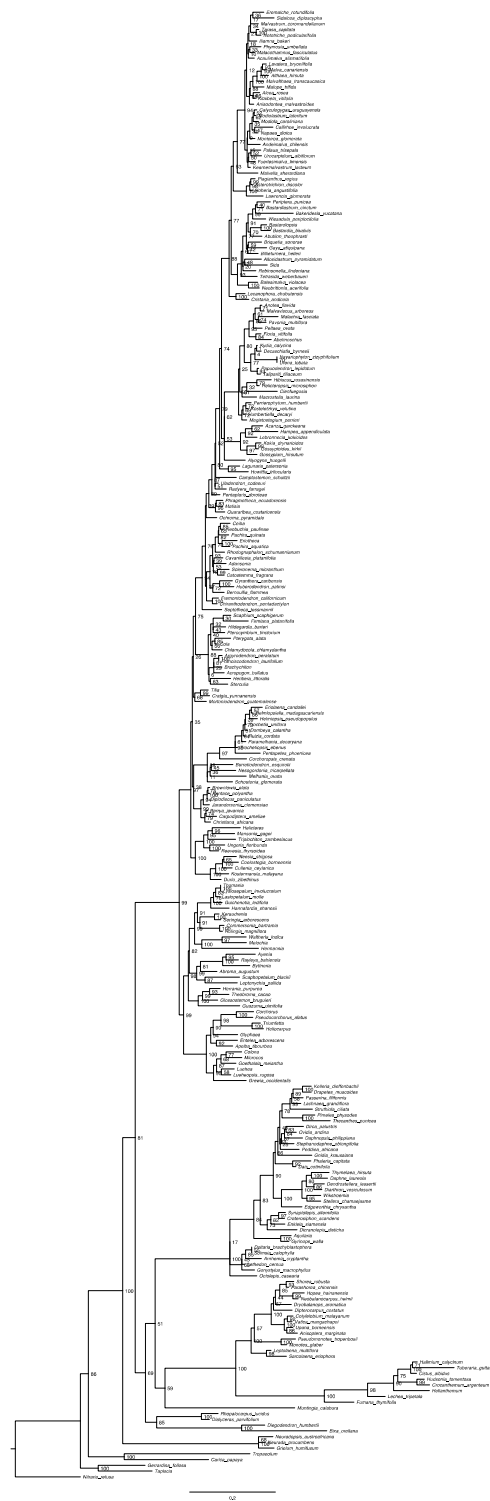


Figure S3. Bayesian inference tree of the eight molecular markers including Cytinaceae. For methods see section 2.1. Node numbers represent Posterior Probability.

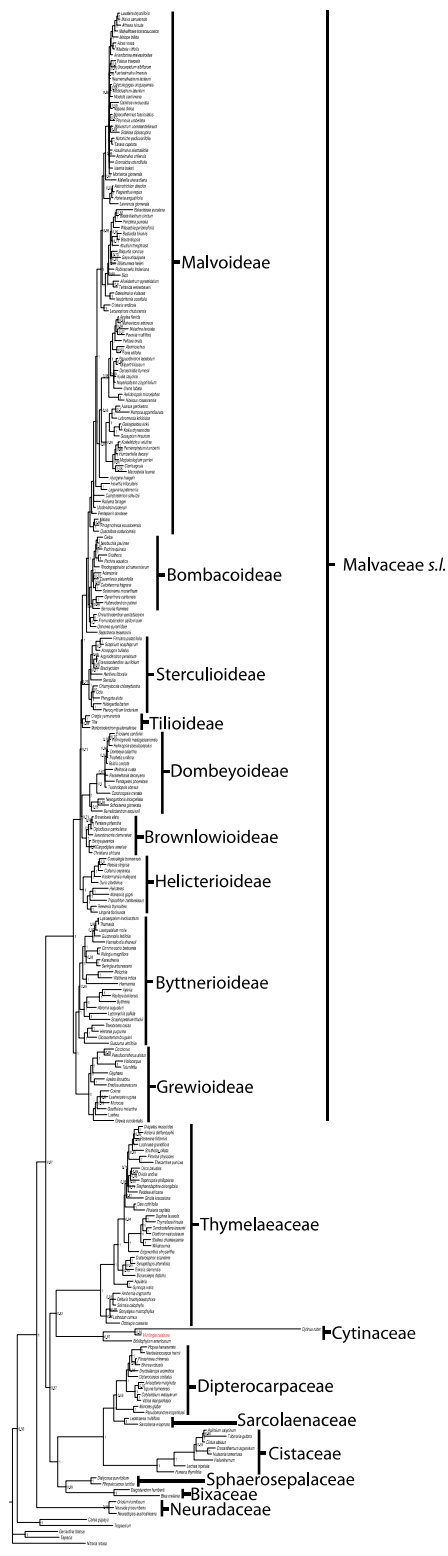


Figure S4. Maximum Likelihood tree of the eight molecular markers concatenated including Cytinaceae. For methods see section 2.1. Node numbers represent Bootstrap support.

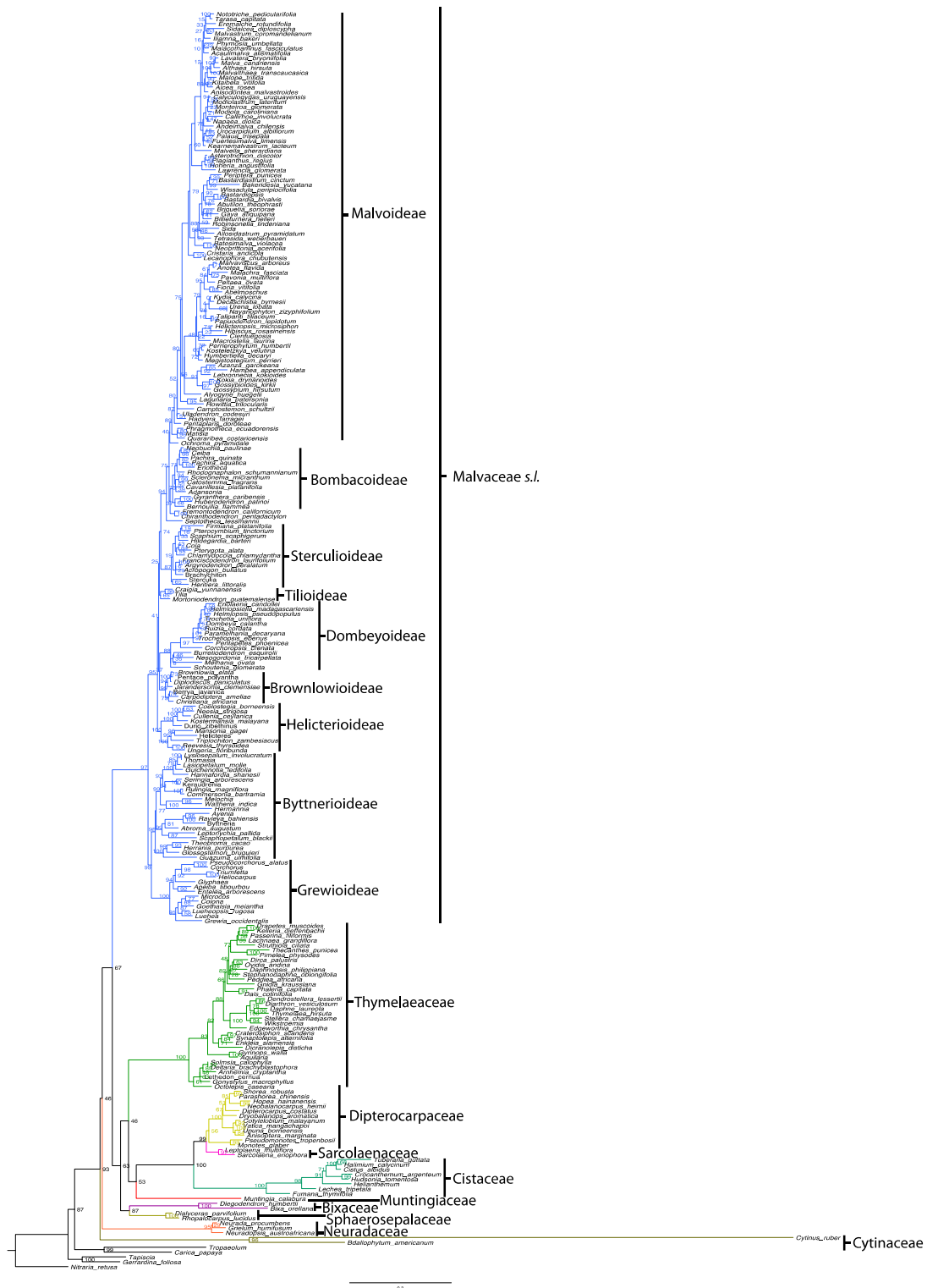


Figure S5. ND chronogram. Age estimation is shown in node associated numbers. Posterior probabilities are shown in numbers associated to branches. Bars represent the 95% HPD.

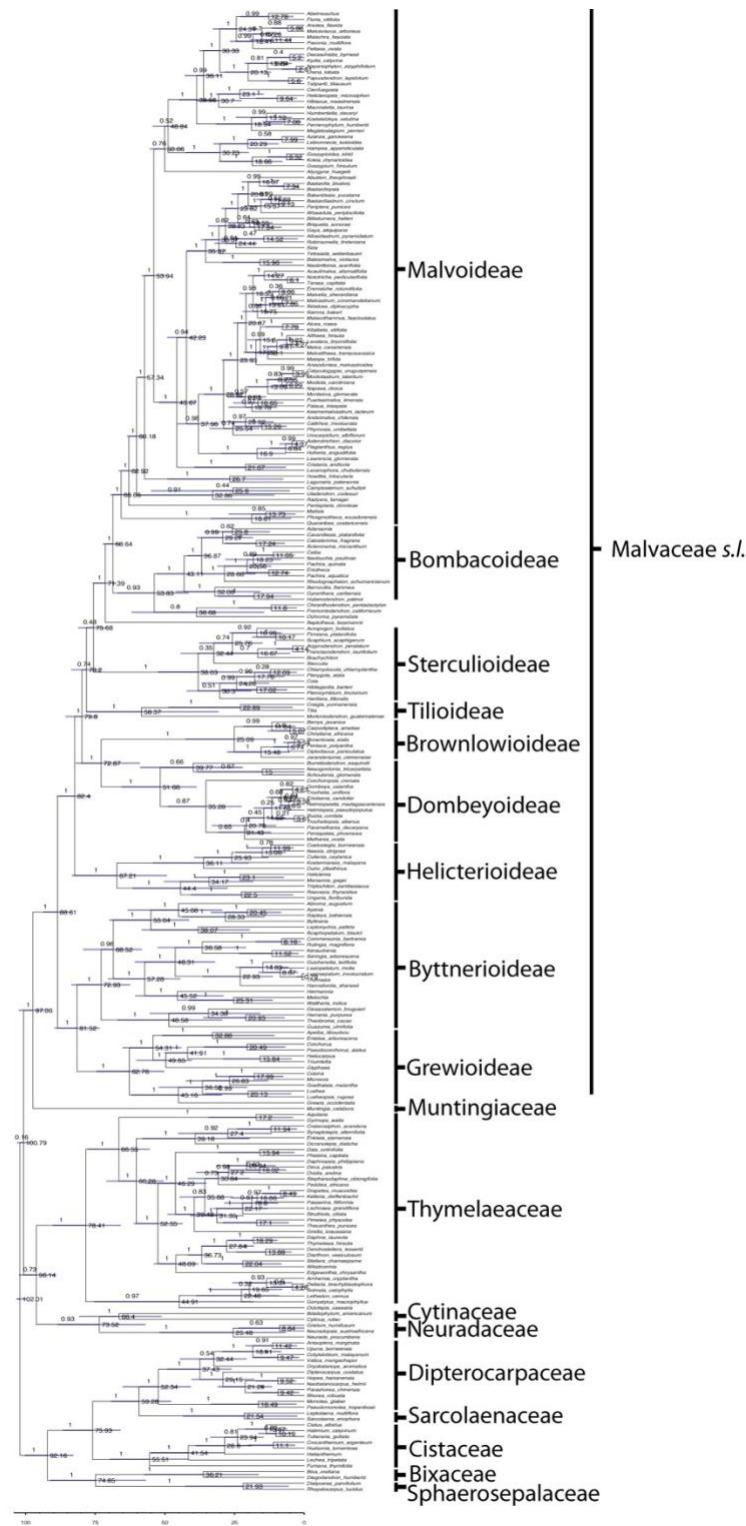


Figure S6. FBD14 chronogram. Age estimation is shown in node associated numbers. Posterior probabilities are shown in numbers associated to branches. Bars represent the 95% HPD.

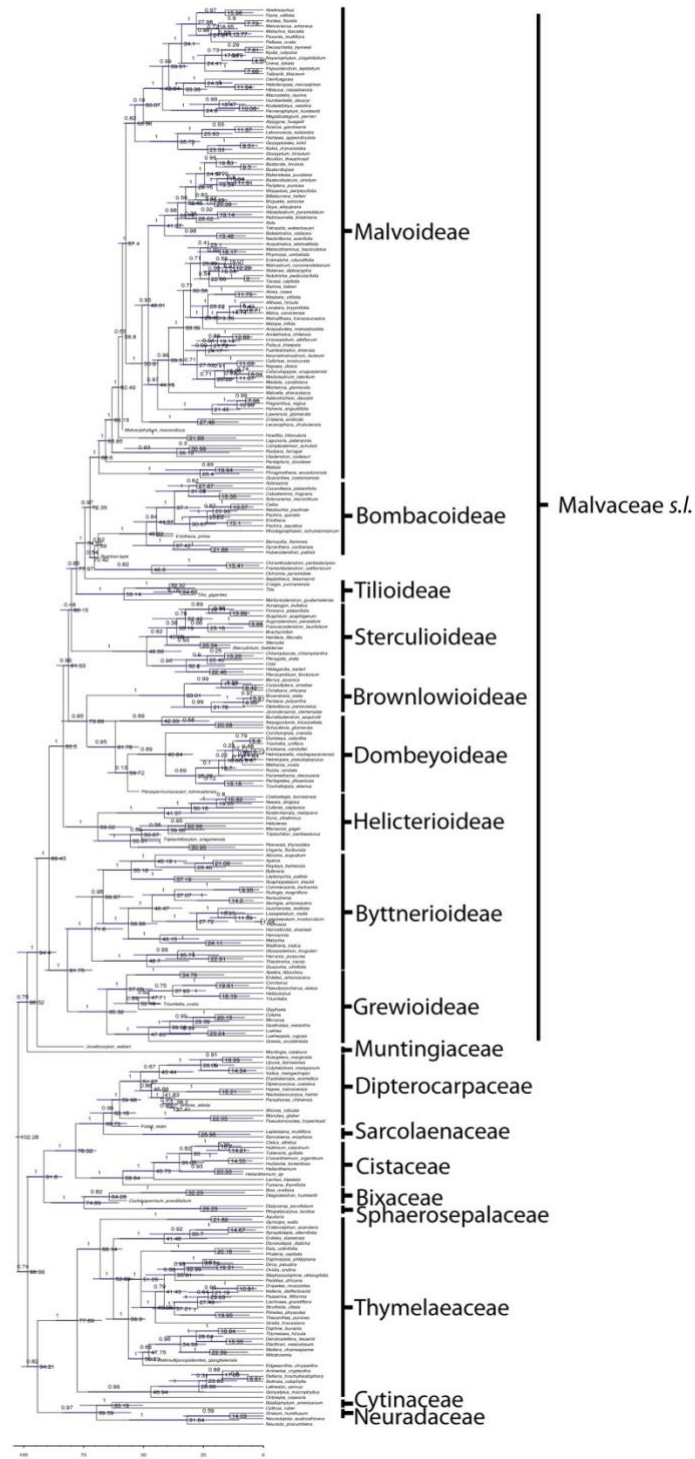


Figure S7. FBD50 chronogram. Age estimation is shown in node associated numbers. Posterior probabilities are shown in numbers associated to branches. Bars represent the 95% HPD.

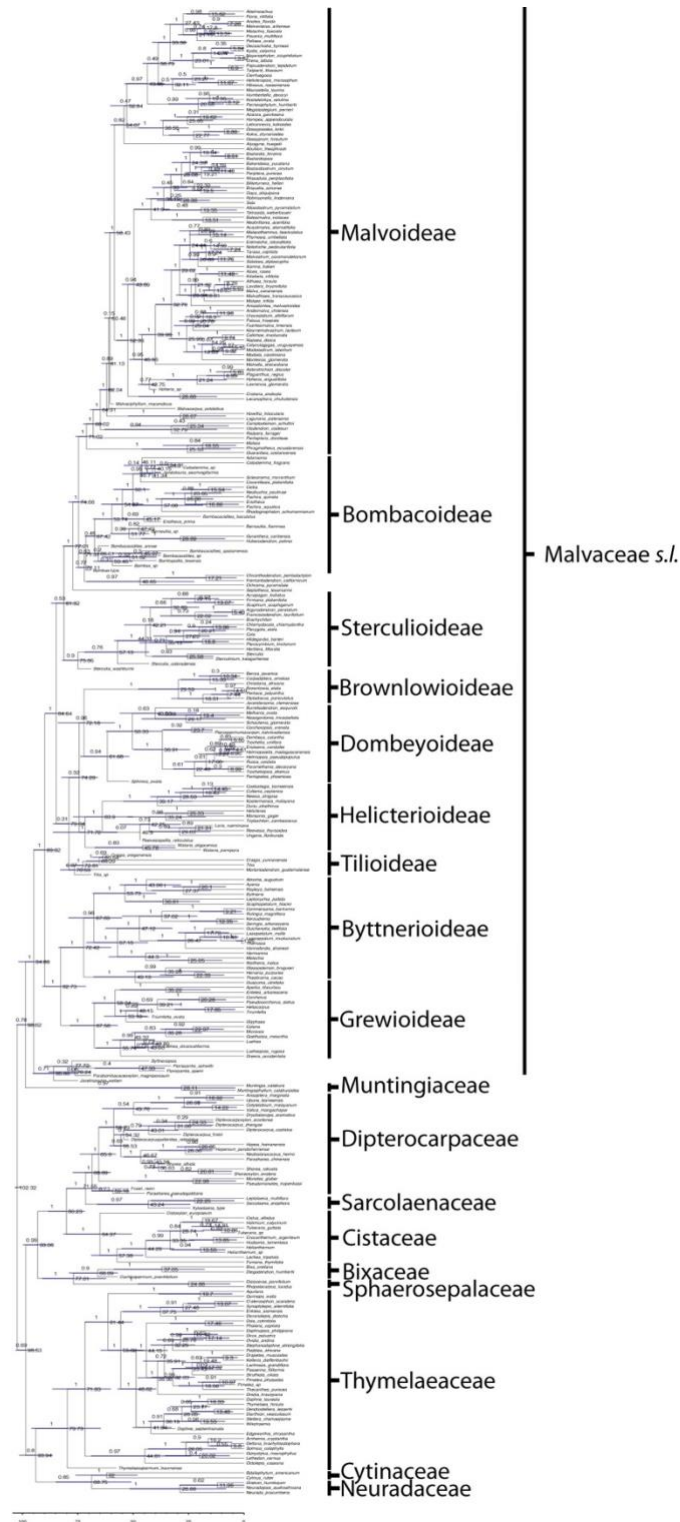


Figure S8. FBD80 chronogram. Age estimation is shown in node associated numbers. Posterior probabilities are shown in numbers associated to branches. Bars represent the 95% HPD.

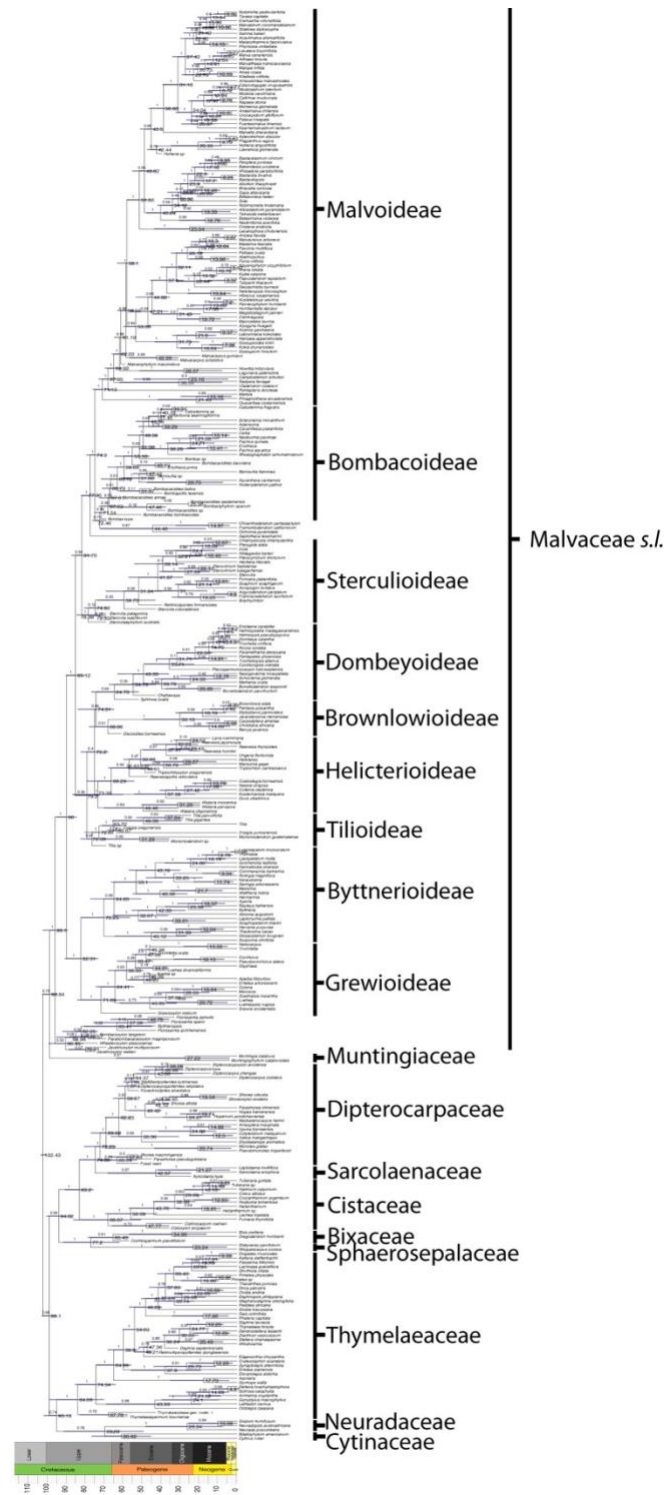


Figure 39. FBDs2-3 chronogram. Age estimation is shown in node associated numbers. Posterior probabilities are shown in numbers associated to branches. Bars represent the 95% HPD.

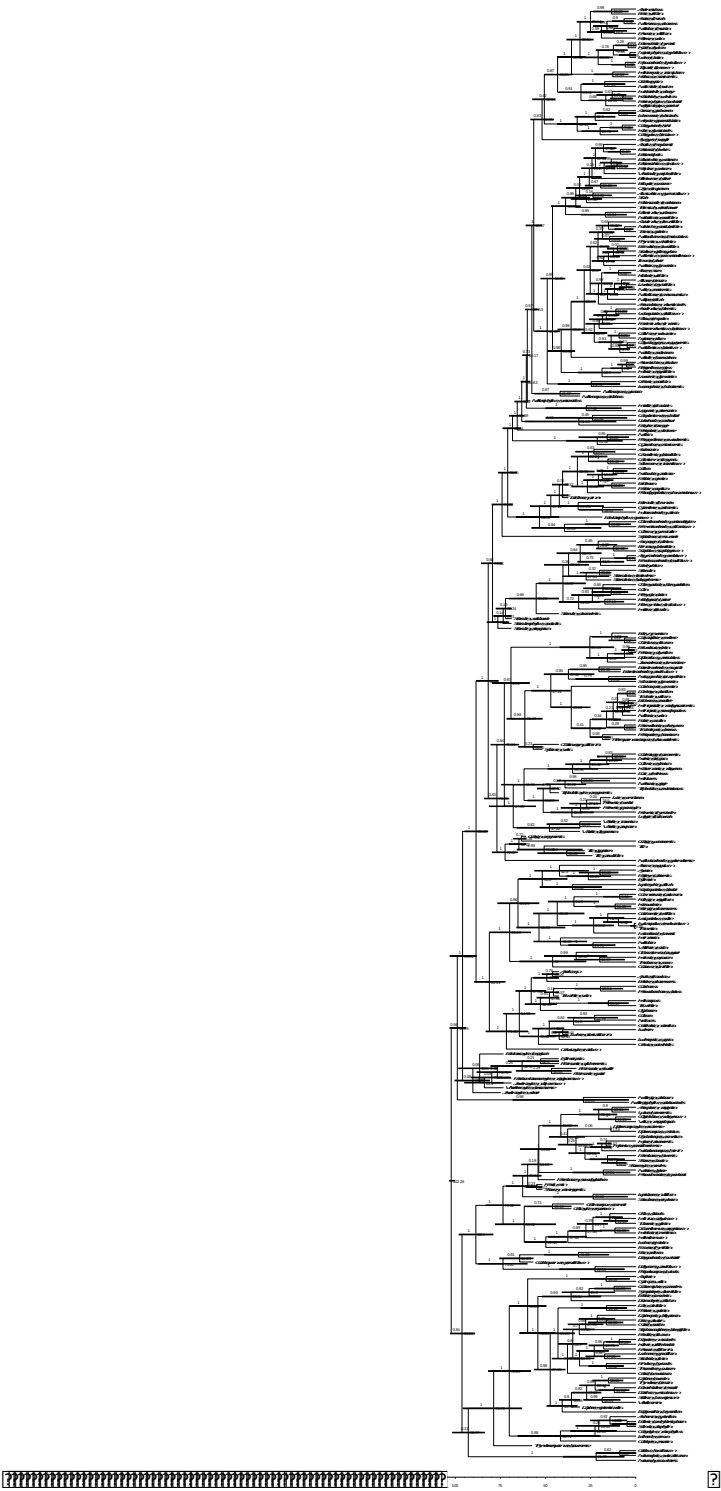


Figure S10. FBDs3 chronogram. Age estimation is shown in node associated numbers. Posterior probabilities are shown in numbers associated to branches. Bars represent the 95% HPD.

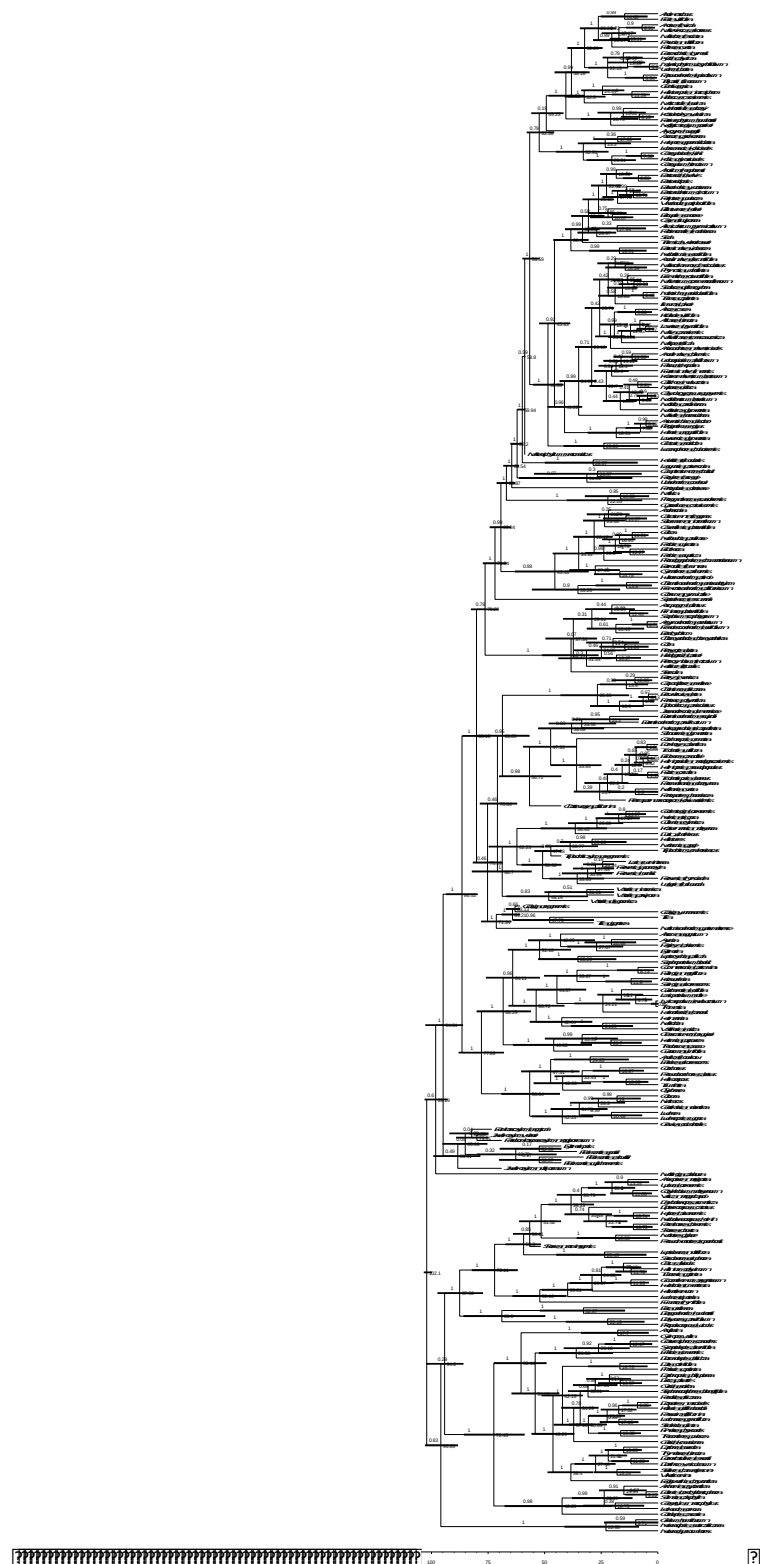


Figure S11. Net diversification rate (r) and turnover rate (ϵ) estimated with the FBD model. Bars indicate 95% HPD intervals.

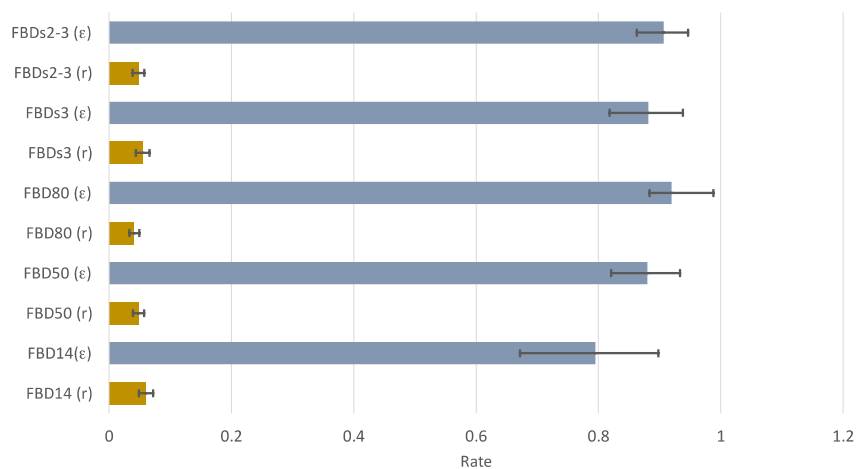
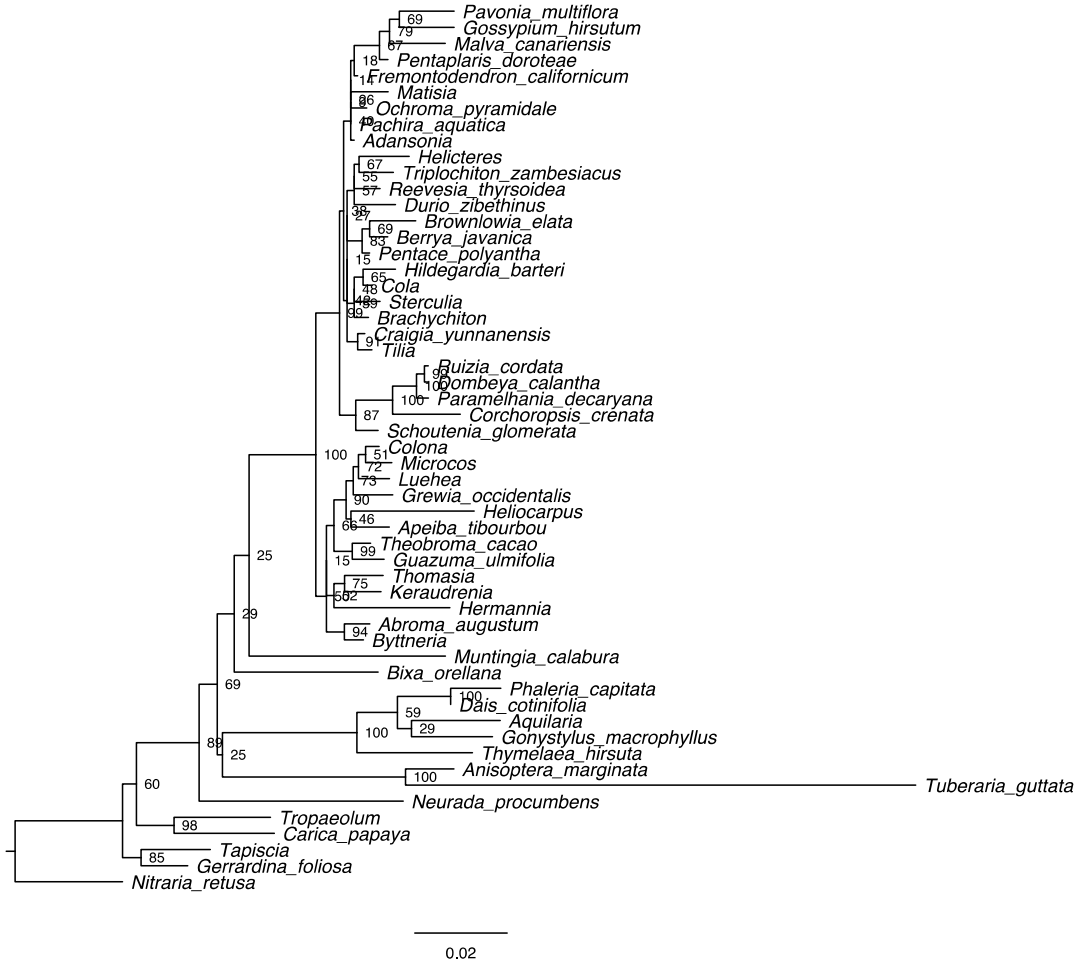
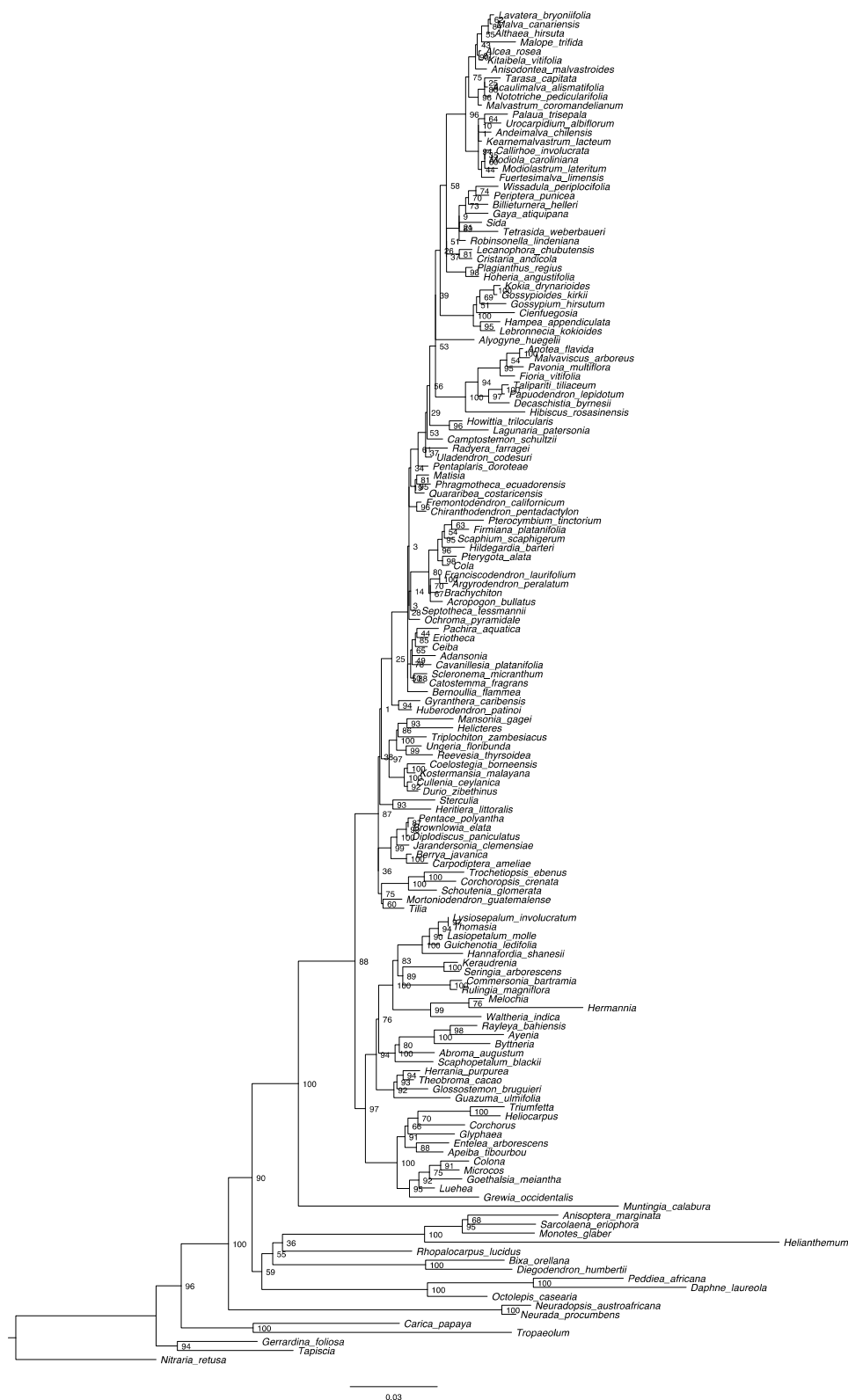


Figure S12. Maximum Likelihood tree of each individual locus. Implemented in RAxML v. 8.2.10 (Stamatakis 2014). 500 Bootstrap iterations. Node numbers represent Bootstrap support.

atpB



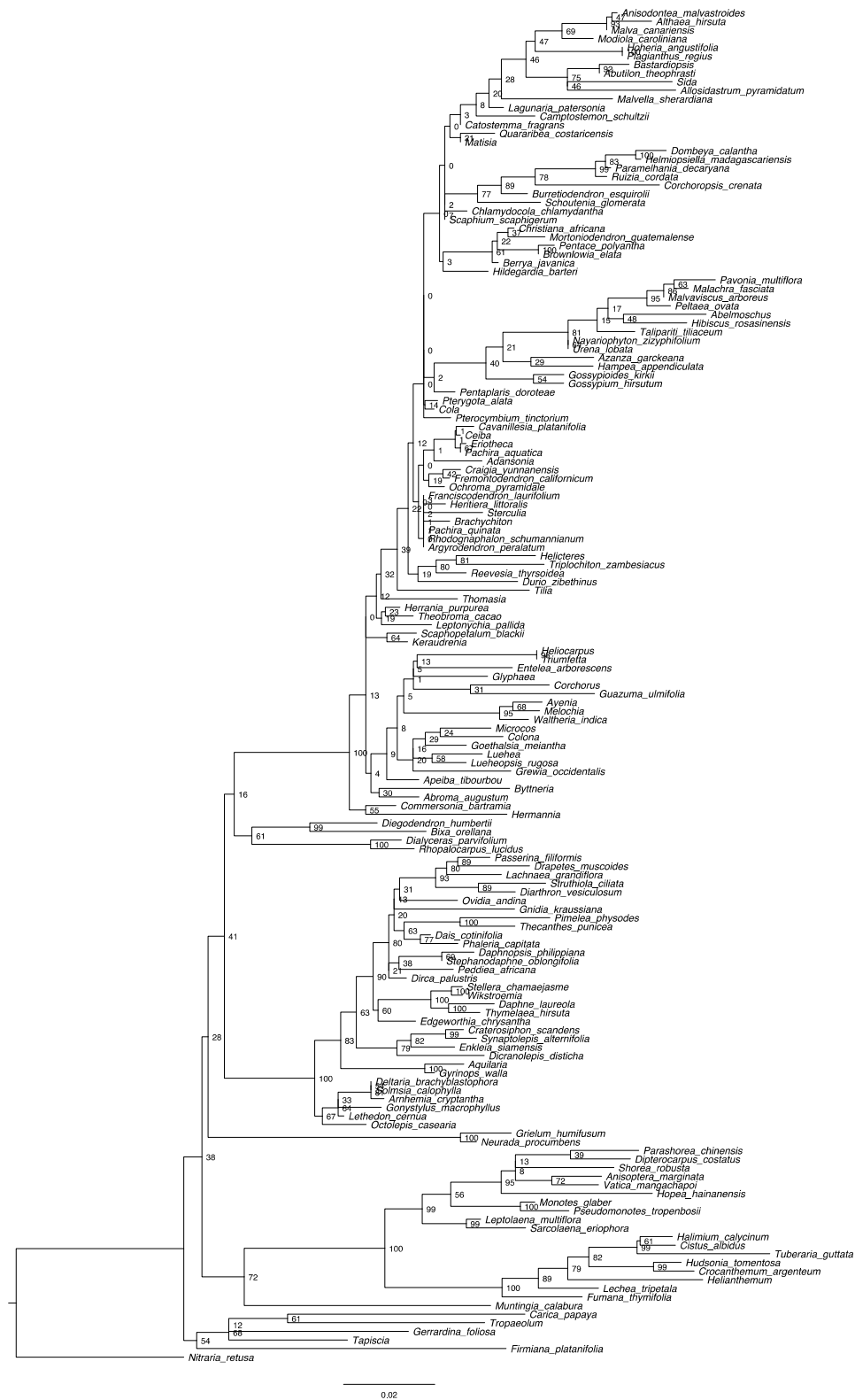
ndhF



psbA



rbcl

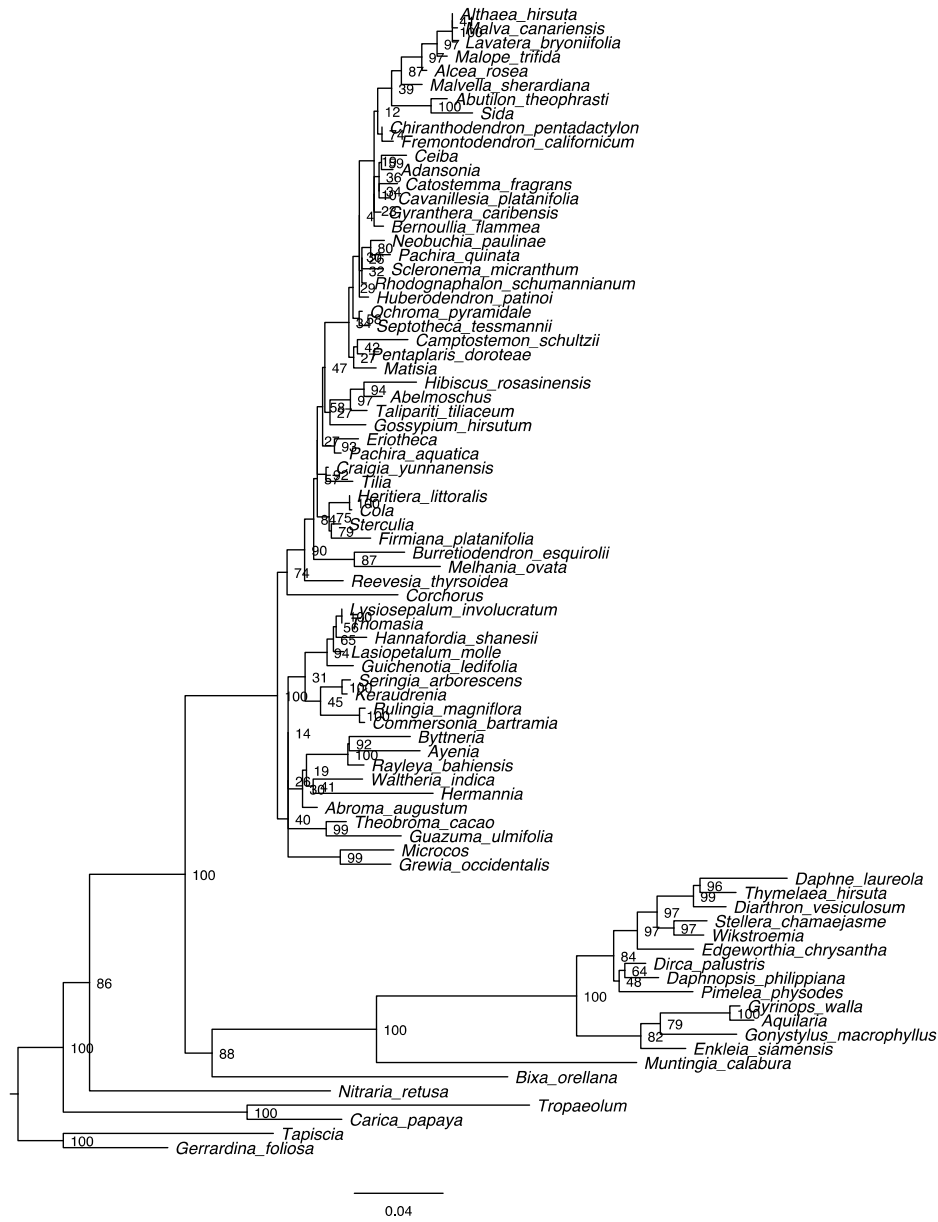


Phylogenetic tree of the order Rubiales, showing relationships between various plant families and genera. The tree is rooted at *Nitraria retusa* and branches out to include families like Malvaceae, Sterculiaceae, and others. Bootstrap values are provided at the nodes.

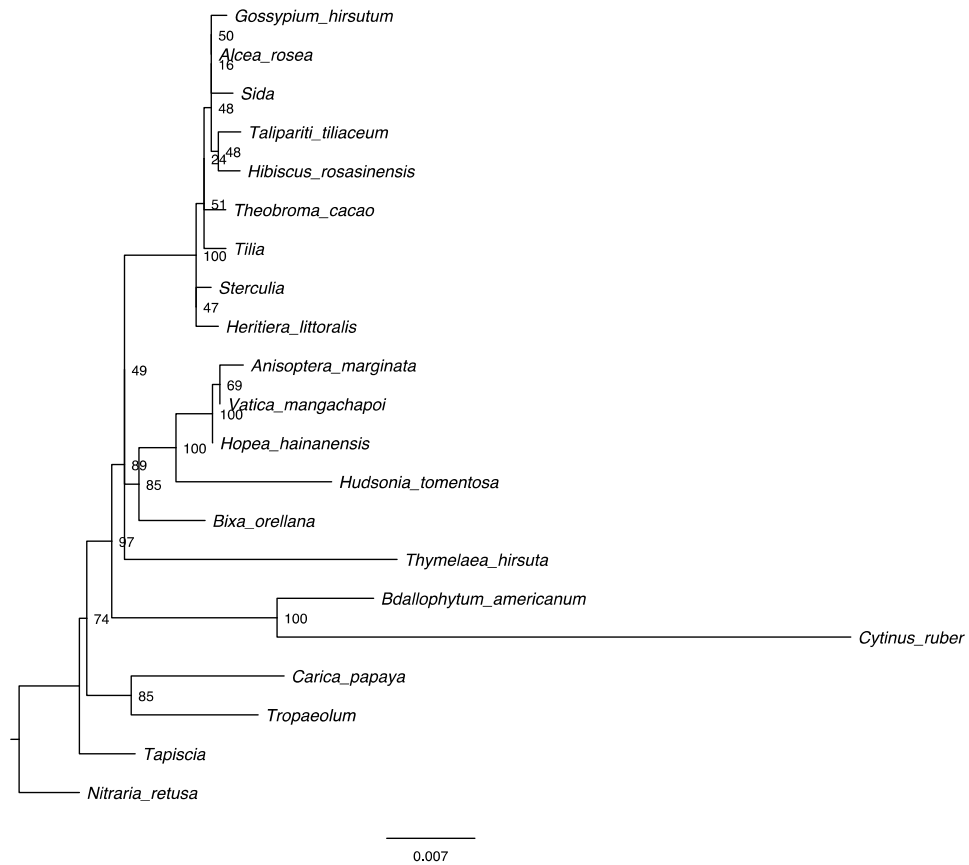
Key taxa and bootstrap values (from left to right):

- Nitraria retusa* (97)
- Tapiscia* (99)
- Carica papaya* (97)
- Guttaria folicosa* (99)
- Gonyostylus macrophyllus* (99)
- Heliconia carnea* (99)
- Solmsia calophylla* (99)
- Azadirachta indica* (99)
- Synaptolepis alternifolia* (99)
- Dais crotifolia* (99)
- Thymelaea hirsuta* (99)
- Wikstroemia* (99)
- Daphne laureola* (99)
- Gnidia kraussiana* (99)
- Podocarpus neriifolia* (99)
- Dirca palustris* (99)
- Stephanodaphne oblongifolia* (99)
- Pimelea physodes* (99)
- Thecaphys paphiophylla* (99)
- Struthiola ciliata* (99)
- Passerina nitens* (99)
- Kelleria diellenbachii* (99)
- Muntingia calabura* (99)
- Bixa orellana* (99)
- Fumaria thymifolia* (99)
- Halimium calycinum* (99)
- Cistus albidus* (99)
- Helianthemum* (99)
- Monotes glaber* (99)
- Anisoptera marginata* (99)
- Urena bomeensis* (99)
- Coleleobium malayanum* (99)
- Yucca mangachapoi* (99)
- Shorea robusta* (99)
- Parashorea chinensis* (99)
- Shorea hainanensis* (99)
- Chytalanoos aromatica* (99)
- Dipterocarpus costatus* (99)
- Thymelaea hirsuta* (99)
- Wikstroemia* (99)
- Daphne laureola* (99)
- Gnidia kraussiana* (99)
- Podocarpus neriifolia* (99)
- Dirca palustris* (99)
- Stephanodaphne oblongifolia* (99)
- Pimelea physodes* (99)
- Thecaphys paphiophylla* (99)
- Struthiola ciliata* (99)
- Passerina nitens* (99)
- Kelleria diellenbachii* (99)
- Muntingia calabura* (99)
- Bixa orellana* (99)
- Fumaria thymifolia* (99)
- Halimium calycinum* (99)
- Cistus albidus* (99)
- Helianthemum* (99)
- Monotes glaber* (99)
- Anisoptera marginata* (99)
- Urena bomeensis* (99)
- Coleleobium malayanum* (99)
- Yucca mangachapoi* (99)
- Shorea robusta* (99)
- Parashorea chinensis* (99)
- Shorea hainanensis* (99)
- Chytalanoos aromatica* (99)
- Dipterocarpus costatus* (99)
- Thymelaea hirsuta* (99)
- Wikstroemia* (99)
- Daphne laureola* (99)
- Gnidia kraussiana* (99)
- Podocarpus neriifolia* (99)
- Dirca palustris* (99)
- Stephanodaphne oblongifolia* (99)
- Pimelea physodes* (99)
- Thecaphys paphiophylla* (99)
- Struthiola ciliata* (99)
- Passerina nitens* (99)
- Kelleria diellenbachii* (99)
- Muntingia calabura* (99)
- Bixa orellana* (99)
- Fumaria thymifolia* (99)
- Halimium calycinum* (99)
- Cistus albidus* (99)
- Helianthemum* (99)
- Monotes glaber* (99)
- Anisoptera marginata* (99)
- Urena bomeensis* (99)
- Coleleobium malayanum* (99)
- Yucca mangachapoi* (99)
- Shorea robusta* (99)
- Parashorea chinensis* (99)
- Shorea hainanensis* (99)
- Chytalanoos aromatica* (99)
- Dipterocarpus costatus* (99)
- Thymelaea hirsuta* (99)
- Wikstroemia* (99)
- Daphne laureola* (99)
- Gnidia kraussiana* (99)
- Podocarpus neriifolia* (99)
- Dirca palustris* (99)
- Stephanodaphne oblongifolia* (99)
- Pimelea physodes* (99)
- Thecaphys paphiophylla* (99)
- Struthiola ciliata* (99)
- Passerina nitens* (99)
- Kelleria diellenbachii* (99)
- Muntingia calabura* (99)
- Bixa orellana* (99)
- Fumaria thymifolia* (99)
- Halimium calycinum* (99)
- Cistus albidus* (99)
- Helianthemum* (99)
- Monotes glaber* (99)
- Anisoptera marginata* (99)
- Urena bomeensis* (99)
- Coleleobium malayanum* (99)
- Yucca mangachapoi* (99)
- Shorea robusta* (99)
- Parashorea chinensis* (99)
- Shorea hainanensis* (99)
- Chytalanoos aromatica* (99)
- Dipterocarpus costatus* (99)
- Thymelaea hirsuta* (99)
- Wikstroemia* (99)
- Daphne laureola* (99)
- Gnidia kraussiana* (99)
- Podocarpus neriifolia* (99)
- Dirca palustris* (99)
- Stephanodaphne oblongifolia* (99)
- Pimelea physodes* (99)
- Thecaphys paphiophylla* (99)
- Struthiola ciliata* (99)
- Passerina nitens* (99)
- Kelleria diellenbachii* (99)
- Muntingia calabura* (99)
- Bixa orellana* (99)
- Fumaria thymifolia* (99)
- Halimium calycinum* (99)
- Cistus albidus* (99)
- Helianthemum* (99)
- Monotes glaber* (99)
- Anisoptera marginata* (99)
- Urena bomeensis* (99)
- Coleleobium malayanum* (99)
- Yucca mangachapoi* (99)
- Shorea robusta* (99)
- Parashorea chinensis* (99)
- Shorea hainanensis* (99)
- Chytalanoos aromatica* (99)
- Dipterocarpus costatus* (99)
- Thymelaea hirsuta* (99)
- Wikstroemia* (99)
- Daphne laureola* (99)
- Gnidia kraussiana* (99)
- Podocarpus neriifolia* (99)
- Dirca palustris* (99)
- Stephanodaphne oblongifolia* (99)
- Pimelea physodes* (99)
- Thecaphys paphiophylla* (99)
- Struthiola ciliata* (99)
- Passerina nitens* (99)
- Kelleria diellenbachii* (99)
- Muntingia calabura* (99)
- Bixa orellana* (99)
- Fumaria thymifolia* (99)
- Halimium calycinum* (99)
- Cistus albidus* (99)
- Helianthemum* (99)
- Monotes glaber* (99)
- Anisoptera marginata* (99)
- Urena bomeensis* (99)
- Coleleobium malayanum* (99)
- Yucca mangachapoi* (99)
- Shorea robusta* (99)
- Parashorea chinensis* (99)
- Shorea hainanensis* (99)
- Chytalanoos aromatica* (99)
- Dipterocarpus costatus* (99)
- Thymelaea hirsuta* (99)
- Wikstroemia* (99)
- Daphne laureola* (99)
- Gnidia kraussiana* (99)
- Podocarpus neriifolia* (99)
- Dirca palustris* (99)
- Stephanodaphne oblongifolia* (99)
- Pimelea physodes* (99)
- Thecaphys paphiophylla* (99)
- Struthiola ciliata* (99)
- Passerina nitens* (99)
- Kelleria diellenbachii* (99)
- Muntingia calabura* (99)
- Bixa orellana* (99)
- Fumaria thymifolia* (99)
- Halimium calycinum* (99)
- Cistus albidus* (99)
- Helianthemum* (99)
- Monotes glaber* (99)
- Anisoptera marginata* (99)
- Urena bomeensis* (99)
- Coleleobium malayanum* (99)
- Yucca mangachapoi* (99)
- Shorea robusta* (99)</

trnL-F



matR



[illegible]

References

Stamatakis A. 2014. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9): 1312-1313.

CAPÍTULO I: ARCHIVO SUPLEMENTARIO

LISTA DE FÓSILES

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: 106606.

Supporting File 1: Fossil information

The following is a list of the 80 fossils used in this study with information about structure, age and reference. Moreover, we indicate the analysis for which each fossil was used and the assignment (constraint) in the phylogeny. The criterion applied to assign a fossil to the crown, or the stem group of a certain clade is based on the synapomorphies of the fossilized structures. If the fossil has the synapomorphy(s) of extant species of a clade, the fossil is assigned to the crown group. In turn, if the fossil resembles extant species but it does not have the synapomorphy of this clade the assignment is in the stem group. Additionally, we assign a “score” to the fossils based on the type of study from which each fossil was obtained. From the highest ranked to the least, these are the categories: (3) morphological comparison with extant taxa; (2) morphological description and assumption of taxonomic affinities; (1) palynoflora. There would be a fourth rank corresponding to studies that investigate the position of the fossil with phylogenetic analysis of morphological characters, but in our fossil sampling there are no such cases.

- | | |
|--|---|
| <p>Bixaceae</p> <ol style="list-style-type: none">1. <i>Cochlospermum previtifolium</i> Berry
Structure: leaf
Locality: Rio Pichileufu, Patagonia, Chile
Stratigraphy: Paleocene
Absolute time: 66-56 Ma.
Reference: Berry 1935, 1938; Wilf et al. 2005; González 2009.
Analysis: ND14; FBD14; FBD50; FBD80.
Assignment: ND: clade including Bixaceae and Sphaerosepalaceae; FBD: Stem Bixaceae
Score: 2 | <p>Reference: Gottwald 1992; Gregory et al. 2009.
Analysis: FBD50; FBD80.
Assignment: Stem Cistaceae.
Score: 2</p> |
| <p>Cistaceae</p> <ol style="list-style-type: none">2. <i>Cistinocarpum roemeri</i>
Structure: fruit, flower
Locality: Germany
Stratigraphy: Middle Oligocene
Absolute time: 33.9-23.03 Ma.
Reference: Conwentz 1886.
Analysis: FBD80.
Assignment: Stem Cistaceae
Score: 23. <i>Cistoxydon europaeum</i>
Structure: wood
Locality: Helmstedt, Lower Saxony, Germany
Stratigraphy: Upper Eocene
Absolute time: 37.8-33.9 Ma. | <ol style="list-style-type: none">4. <i>Helianthemum</i> sp.
Structure: pollen
Locality: Mirabel, Coiron, Ardèche, France.
Stratigraphy: Upper Miocene
Absolute time: 11.63-5.33 Ma.
Reference: Naud & Suc 1975.
Analysis: ND14; FBD14; FBD50; FBD80.
Assignment: ND and FBD: clade including <i>Lechea</i>, <i>Helianthemum</i>, <i>Hudsonia</i>, <i>Crocanthemum</i>, <i>Cistus</i>, <i>Halmium</i> and <i>Tuberaria</i>.
Score: 15. <i>Tuberaria</i> sp.
Structure: pollen
Locality: Schleswig-Holstein, Germany
Stratigraphy: Pliocene
Absolute time: 5.33-2.58 Ma.
Reference: Menke 1976; Muller 1981; Hooghiemstra & van Geel 1998.
Analysis: FBD50; FBD80.
Assignment: Sister group of <i>Tuberaria</i>.
Score: 1 |

Dipterocarpaceae

6. *Albertipollenites kutchensis* Mandal & Rao
Structure: pollen
Locality: Matanomadh Formation, Kutch Basin, Gujarat, India
Stratigraphy: Early Eocene
Absolute time: 56-47.8 Ma.
Reference: Mandal & Rao 2001; Prasad et al. 2009; Mathews et al. 2013.
Analysis: FBD80.
Assignment: Clade including *Dipterocarpus costatus*.

Score: 1

7. *Dipterocarpoxydon arcotense* Awasthi
Structure: wood
Locality: Nungarh Sot, Kalagarh, Pauri Garhwal District, Uttar Pradesh, India
Stratigraphy: Middle Miocene
Absolute time: 15.97-11.63 Ma.
Reference: Prasad 1993
Analysis: FBD50; FBD80.
Assignment: Clade including *Dipterocarpus costatus*.

Score: 2

8. *Dipterocarpus*-type
Structure: pollen
Locality: Northwest Borneo
Stratigraphy: Oligocene
Absolute time: 33.9-23.03 Ma.
Reference: Muller 1970
Analysis: FBD50; FBD80.
Assignment: Clade including *Dipterocarpus costatus*.

Score: 1

9. *Dipterocarpus zhengae* H. M. Li & G. L. Shi
Structure: fruit wing
Locality: Fotan Group, Zhangpu County, Fujian Province, China.
Stratigraphy: Middle - Late Miocene

Absolute time: 15.97-5.33 Ma.
Reference: Shi & Li 2010.
Analysis: FBD50; FBD80.
Assignment: Clade including *Dipterocarpus costatus*.

Score: 1

10. *Dipterocarpuipollenites retipilatus* Kar & Jain

Structure: pollen

Locality: Matanomadh Formation, Kutch Basin, Gujarat, India
Stratigraphy: Early Eocene
Absolute time: 56-47.8 Ma.
Reference: Mandal & Rao 2001; Prasad et al. 2009; Mathews et al. 2013.
Analysis: FBD50; FBD80.
Assignment: Clade including *Dipterocarpus costatus*.
Score: 1

11. Fossil resin

Structure: resin, wood, pollen
Locality: Vastan mine, Cambay Shale Formation, Gujarat, India
Stratigraphy: Early Eocene (Ypresian)
Absolute time: 56-47.8 Ma.
Reference: Rust et al. 2010; Dutta et al. 2011.
Analysis: ND14; FBD14; FBD50; FBD80.
Assignment: ND: Clade including Dipterocarpaceae and Sarcolaenaceae; FBD: Stem Dipterocarpaceae.
Score: 2

12. *Foveotricolpites alveolatus* Mandal & Rao
Structure: pollen
Locality: Matanomadh Formation, Kutch Basin, Gujarat, India
Stratigraphy: Early Eocene
Absolute time: 56-47.8 Ma.
Reference: Mandal & Rao 2001; Prasad et al. 2009; Mathews et al. 2013.
Analysis: FBD80.
Assignment: Clade including *Dipterocarpus costatus*.
Score: 1

13. *Hopenium pondicherriense* Awasthi
Structure: wood
Locality: Nungarh Sot, Kalagarh, Pauri Garhwal District, Uttar Pradesh, India
Stratigraphy: Middle Miocene
Absolute time: 15.97-11.63 Ma.
Reference: Prasad 1993.
Analysis: FBD50; FBD80.
Assignment: Sister group of *Hopea hainanensis*.
Score: 2

14. *Parashorea pseudogoldiana* (Holl.) Wolfe

- Structure: leaf
 Locality: Kushtaka and Kulthieth, Alaska, USA; Ketavik Formation, Alaska, USA; Chumstick Formation, Washington State, USA.
 Stratigraphy: Eocene
 Absolute time: 56-33.9 Ma.
 Reference: Wolfe 1977; Evans 1991; Parrish et al. 2010.
 Analysis: FBD50; FBD80.
 Assignment: Stem Dipterocarpaceae.
 Score: 2
15. *Shorea albida*-type
 Structure: pollen
 Locality: Belait Formation, Berakas, Brunei.
 Stratigraphy: Miocene
 Absolute time: 23.03-5.33 Ma.
 Reference: Anderson & Muller 1975.
 Analysis: ND14; FBD14; FBD50; FBD80.
 Assignment: ND: Clade including *Shorea robusta* and *Parashorea chinensis*; FBD: Clade including *Shorea robusta*.
 Score: 1
16. *Shorea maomingensis* Feng, Kodrul & Jin
 Structure: fruit wing
 Locality: Maoming Basin, Guangdong Province, Shouth China.
 Stratigraphy: Late Eocene
 Absolute time: 41.2-33.9 Ma.
 Reference: Feng et al. 2013
 Analysis: FBD80.
 Assignment: Stem Dipterocarpaceae.
 Score: 3
17. *Shoreoxylon evidens* Eyde
 Structure: wood
 Locality: Ghogra River Section, Siang District, Arunachal Pradesh, India.
 Stratigraphy: Late Miocene - Early Pliocene
 Absolute time: 7.24-3.6 Ma.
 Reference: Mehrotra et al. 1999.
 Analysis: FBD50; FBD80.
 Assignment: Sister group of *Shorea robusta*.
 Score: 2
- Malvaceae s.l.
18. *Apeiba* sp.
 Structure: leaf
 Locality: Aycross Formation, Wind River Basin, Wyoming, USA
- Stratigraphy: Middle Eocene
 Absolute time: 47.8-37.8 Ma.
 Reference: MacGinitie 1974; Taylor 1990; Graham 1999.
 Analysis: FBD80.
 Assignment: Sister to *Apeiba tibourbou*.
 Score: 2
19. *Bernoullia* sp.
 Structure: pollen
 Locality: Collazo Quebrada, San Sebastian Formation, Puerto Rico
 Stratigraphy: Middle Oligocene
 Absolute time: 33.9-23.03 Ma.
 Reference: Graham & Jarzen 1969
 Analysis: FBD50; FBD80.
 Assignment: Sister to *Bernoullia flammea*
 Score: 1
20. *Bombacacidites annae* (Van der Hammen) Germeraad
 Structure: pollen
 Locality: Arcillas de El Limbo Formation, Piñalerita section, Colombia
 Stratigraphy: Paleocene
 Absolute time: 66-56 Ma.
 Reference: Jaramillo & Dilcher 2001; Graham 2010.
 Analysis: FBD50; FBD80.
 Assignment: Stem Bombacoideae.
 Score: 1
21. *Bombacacidites baculatus* Muller
 Structure: pollen
 Locality: Santa Sofía, Mocagua and Los Chorros sections, Solimoes Formation, Colombia
 Stratigraphy: Miocene
 Absolute time: 23.03-5.33 Ma.
 Reference: Hoorn 1994
 Analysis: FBD50; FBD80.
 Assignment: Stem tribe Adansonieae.
 Score: 1
22. *Bombacacidites bellus* Colmenares
 Structure: pollen
 Locality: Misoa Formation, Lake Maracaibo Basin, Venezuela
 Stratigraphy: Middle Eocene
 Absolute time: 47.8-37.8 Ma.
 Reference: Colmenares 1988; Graham 2010.
 Analysis: FBD80.

- Assignment: Stem Bombacoideae.
Score: 1
23. *Bombacidites bombaxoides* Couper
Structure: pollen
Locality: Kapuni Formation, Taranaki Basin, New Zealand
Stratigraphy: Early to Middle Eocene
Absolute time: 56-41.2 Ma.
Reference: Pocknall 1989, 1990.
Analysis: FBD50; FBD80.
Assignment: Stem Bombacoideae.
Score: 1
24. *Bombacidites qaidamensis* Zhu, Wu, Xi, Song & Zhang
Structure: pollen
Locality: Upper Ganchaigou Formation, China
Stratigraphy: Oligocene
Absolute time: 33.9-23.03 Ma.
Reference: Song et al. 1999 cited in Song et al. 2004
Analysis: FBD50; FBD80.
Assignment: Stem Bombacoideae.
Score: 1
25. *Bombacidites* sp.
Structure: pollen
Locality: Manning Formation, Jackson Group, Texas, USA
Stratigraphy: Upper Eocene
Absolute time: 37.8-33.9 Ma.
Reference: Elsik & Yancey 2000
Analysis: FBD50; FBD80.
Assignment: Stem Bombacoideae.
Score: 1
26. *Bombaciphyllum opacum* Engelhardt
Structure: leaf
Locality: Curanilahue Formation, Arauco-Concepción, Chile
Stratigraphy: Miocene
Absolute time: 23.03-5.33 Ma.
Reference: Berry 1922; Graham 2010.
Analysis: FBD80.
Assignment: Stem Bombacoideae.
Score: 2
27. *Bombacoxylon langstoni* Wheeler & Lehman
Structure: wood
- Locality: Aguja Formation, Big Bend, Texas, USA
Stratigraphy: Late Campanian
Absolute time: 83.6-72.1 Ma.
Reference: Wheeler & Lehman 2000
Analysis: FBD80.
Assignment: Stem Malvaceae s.l.
Score: 3
28. *Bombapollis texensis* Elsik
Structure: pollen
Locality: Crockett and Yegua formations, Claiborne group, Texas, USA
Stratigraphy: Middle Eocene
Absolute time: 47.8-37.8 Ma.
Reference: Elsik & Yancey 2000
Analysis: FBD50; FBD80.
Assignment: Stem Bombacoideae.
Score: 1
29. *Bombax* sp. (fossil 1)
Structure: pollen
Locality: Collazo Quebrada, San Sebastian Formation, Puerto Rico
Stratigraphy: Middle Oligocene
Absolute time: 33.9-23.03 Ma.
Reference: Graham & Jarzen 1969
Analysis: FBD80.
Assignment: Stem tribe Adansonieae (*Catostemma*, *Scleronema*, *Adansonia*, *Cavanillesia*, *Ceiba*, *Neobuchia*, *Pachira*, *Rhodognaphalon*).
Score: 1
30. *Bombax*-type
Structure: pollen
Locality: Navesink Formation, Atlantic Highlands, New Jersey USA
Stratigraphy: Maastrichtian
Absolute time: 72.1-66 Ma.
Reference: Wolfe 1975; Muller 1981; Taylor 1990; Graham 2010.
Analysis: ND14; FBD14; FBD50; FBD80.
Assignment: ND: Clade including Bombacoideae, Malvoideae and *Chiranthodendron*, *Fremontodendron* and *Septotheca*; Stem Bombacoideae.
Score: 1
31. *Burretiodendron parvifructum* Anberrée et al.
Structure: fruit

- Locality: Xiaolongtan Formation, Wnshan and Maguan, China
Stratigraphy: Upper Miocene
Absolute time: 11.63-5.33 Ma.
Reference: Anberrée et al. 2015.
Analysis: FBD80.
Assignment: Sister to *Burretiodendron esquirolii*.
Score: 3
32. *Byttneriopsis* Kvacek et Wilde
Structure: leaf
Locality: Messel, Germany
Stratigraphy: Middle Eocene
Absolute time: 47.8-37.8 Ma.
Reference: Kvacek & Wilde 2010
Analysis: FBD50; FBD80.
Assignment: Stem Malvaceae *s.l.*
Score: 3
33. *Catostemma* sp.
Structure: pollen
Locality: Collazo Quebrada, San Sebastian Formation, Puerto Rico
Stratigraphy: Middle Oligocene
Absolute time: 33.9-23.03 Ma.
Reference: Graham & Jarzen 1969
Analysis: FBD50; FBD80.
Assignment: Clade including *Catostemma fragans* and the fossil *Jandufouria seamrogiformis*.
Score: 1
34. *Chattawayia paliformis* Manchester
Structure: wood
Locality: Nut Beds, Clarno Formation, Oregon, USA
Stratigraphy: Middle Eocene
Absolute time: 47.8-37.8 Ma.
Reference: Manchester 1980
Analysis: FBD80.
Assignment: Stem Dombeyoideae.
Score: 3
35. *Craigia oregonensis* (Arnold) Kvacek, Buzek et Manchester
Structure: fruit valves
Locality: Anadyrka River, Napana Formation, Kamchatka, Russia
Stratigraphy: Late Paleocene – Early Eocene
Absolute time: 61.6-47.8 Ma.
Reference: Kvacek et al. 2005
Analysis: FBD50; FBD80.
Assignment: Sister to *Craigia yunnanensis*.
Score: 3
36. *Discoidites borneensis* Muller
Structure: pollen
Locality: Sarawak, Bungo area, Plateau Sandstone Formation, Malaysia
Stratigraphy: Paleocene
Absolute time: 66-56 Ma.
Reference: Muller 1968; Mandal & Rao 2001; Mathews et al. 2013
Analysis: FBD80.
Assignment: Stem Brownlowioideae
Score: 1
37. *Eriotheca prima* Duarte
Structure: flower
Locality: Fonseca Formation, Quadrilátero Ferrífero, Minas Gerais, Brazil
Stratigraphy: Late Eocene
Absolute time: 41.2-33.9 Ma.
Reference: Duarte 1974; Lima & Salard-Chebouldaeff 1981.
Analysis: ND14; FBD14; FBD50; FBD80.
Assignment: ND: Crown tribe Adansonieae; FBD: Stem tribe Adansonieae
Score: 2
38. *Florissantia ashwillii* Manchester
Structure: flowers and fruits
Locality: Sheep Rock Creek Oregon, USA; Sumner Spring, Gray Butte, Oregon, USA; Goshen flora, Oregon, USA.
Stratigraphy: Late Eocene – early Oligocene
Absolute time: 37.8-27.82 Ma.
Reference: Manchester 1992
Analysis: FBD50; FBD80.
Assignment: Stem Malvaceae *s.l.*
Score: 3
39. *Florissantia quilchenensis* (Mathews & Booke) Manchester
Structure: flowers and fruits
Locality: Quilchena, British Columbia, Canada; Green River Formation, Wyoming, USA; Republic flora, Washington, USA.
Stratigraphy: Middle Eocene
Absolute time: 47.8-37.8 Ma.
Reference: Manchester 1992
Analysis: FBD80.
Assignment: Stem Malvaceae *s.l.*

- Score: 3
40. *Florissantia speirii* (Lesquereux) Manchester
 Structure: flowers and fruits
 Locality: Florissant Formation, Colorado, USA; John Day Formation, Dugout Gulch, Oregon, USA; Clarno Formation, West Branch Creek, Oregon, USA.
 Stratigraphy: Late Eocene – early Oligocene
 Absolute time: 37.8-27.82 Ma.
 Reference: Manchester 1992
 Analysis: FBD50; FBD80.
 Assignment: Stem Malvaceae s.l.
 Score: 3
41. *Grewioxylon indicum* Prakash & Dayal
 Structure: wood
 Locality: Mahurzari, Deccan Intertrappean beds, Nagpur, Maharashtra, India
 Stratigraphy: Eocene
 Absolute time: 56-33.9 Ma.
 Reference: Prakash & Dayal 1963
 Analysis: FBD80.
 Assignment: Stem Grewioideae.
 Score: 2
42. *Hoheria* sp.
 Structure: pollen
 Locality: Crockett and Yegua Formations, Claiborne group, Texas, USA
 Stratigraphy: Late middle Eocene
 Absolute time: 41.2-37.8 Ma.
 Reference: Elsik & Yancey 2000
 Analysis: FBD50; FBD80.
 Assignment: Clade including *Asterotrichion*, *Plagianthus*, *Hoheria* and *Lawrenzia*.
 Score: 1
43. *Jandufouria seamrogiformis* Germeraad
 Structure: pollen
 Locality: La Capacha, Venezuela.
 Stratigraphy: Late Eocene
 Absolute time: 41.2-33.9 Ma.
 Reference: Colmenares et al. 1993
 Analysis: FBD50; FBD80.
 Assignment: Clade including *Catostemma fragans*.
 Score: 1
44. *Javelinoxylon multiporosum* Wheeler, Lehman & Gasson
 Structure: wood
 Locality: Javelina Formation, Texas, USA
 Stratigraphy: Maastrichtian
 Absolute time: 72.1-66 Ma.
 Reference: Wheeler et al. 1994
 Analysis: FBD80.
 Assignment: Stem Malvaceae s.l.
 Score: 3
45. *Javelinoxylon weberi* Estrada-Ruiz, Martínez-Cabrera & Cevallos-Ferriz
 Structure: wood
 Locality: Olmos Formation, Coahuila, Mexico
 Stratigraphy: Late Campanian – early Maastrichtian
 Absolute time: 77.85-69.05 Ma.
 Reference: Estrada-Ruiz et al. 2007
 Analysis: ND14; FBD14; FBD50; FBD80.
 Assignment: Stem Malvaceae s.l.
 Score: 3
46. *Laria rueminiana* (Heer) G. Worobiec & Kvacek
 Structure: leaf
 Locality: Belchatow Lignite Mine, Kleszczow Graben, Poland
 Stratigraphy: Late Miocene
 Absolute time: 11.63-5.33 Ma.
 Reference: Worobiec et al. 2012
 Analysis: FBD50; FBD80.
 Assignment: Clade including *Reevesia thyrsoidea*.
 Score: 3
47. *Luehea divaricatiformis* Fittipaldi, Simoes, Diulietti & Pirani
 Structure: leaf
 Locality: Itaquaquecetuba Formation, Sao Paulo Basin, Brazil
 Stratigraphy: Late Eocene
 Absolute time: 41.2-33.9 Ma.
 Reference: Fittipaldi et al. 1989
 Analysis: FBD50; FBD80.
 Assignment: Sister to *Luehea*.
 Score: 2
48. *Malvacarpus guinazui* Berry
 Structure: fruit
 Locality: Ventana Formation, Río Pichileufu, Argentina
 Stratigraphy: Early Miocene
 Absolute time: 23.03-15.97 Ma.
 Reference: Berry 1938

- Analysis: FBD80.
Assignment: Stem Eumalvoideae (excluding *Howittia*, *Lagunaria*, *Camptostemon*, *Radyera*, *Uladendron*, *Pentaplaris*, *Matisia*, *Phragmothea* and *Quararibea*).
Score: 2
49. *Malvacarpus octolobus* Berry
Structure: fruit
Locality: Belen Flora, Peru
Stratigraphy: Early Oligocene
Absolute time: 33.9-27.82 Ma.
Reference: Berry 1929 cited in Graham 2010; Manchester et al. 2012.
Analysis: FBD50; FBD80.
Assignment: Stem Eumalvoideae.
Score: 2
50. *Malvaciphyllum macondicus* M. Carvalho
Structure: leaf
Locality: Cerrejón Formation, Ranchería Basin, Colombia
Stratigraphy: Middle – Late Paleocene
Absolute time: 61.6-56 Ma.
Reference: Carvalho et al. 2011
Analysis: ND14; FBD14; FBD50; FBD80.
Assignment: ND: Clade including Eumalvoideae and *Howittia* and *Lagunaria*;
FBD: Stem Eumalvoideae.
Score: 3
51. *Mortoniodendron* sp.
Structure: pollen
Locality: Gatuncillo Formation, Alcalde Díaz, Panama
Stratigraphy: Upper Eocene
Absolute time: 37.8-33.9 Ma.
Reference: Graham 1985
Analysis: FBD80.
Assignment: Sister to *Mortoniodendron guatemalense*.
Score: 1
52. *Parabombacaceoxylon magniporosum* Wheeler, Lee & Matten
Structure: wood
Locality: McNairy Formation, Black Powder Hollow, Alexander County, Illinois, USA
Stratigraphy: Maastrichtian
Absolute time: 72.1-66 Ma.
Reference: Wheeler et al. 1987
Analysis: FBD50; FBD80.
- Assignment: Stem Malvaceae s.l.
Score: 3
53. *Pterospermumocarpon kalviwadiensis* R. Srivast., RK Saxena & Gaurav Srivast.
Structure: fruit
Locality: Sindhudurg Formation, Kalviwadi Village, Sindhudurg District, Maharashtra, India
Stratigraphy: Miocene
Absolute time: 23.03-5.33 Ma.
Reference: Srivastava et al. 2012
Analysis: ND14; FBD14; FBD50; FBD80.
Assignment: Clade including *Corchoropsis*, *Dombeya*, *Trochetia*, *Eriolaena*, *Helmiopsiella*, *Helmiopsis*, *Melhania*, *Ruizia*, *Paramelhania*, *Trochetiopsis* and *Pentapetes*.
Score: 3
54. *Reevesia japonoxyla* K. Terada & M. Suzuki
Structure: wood
Locality: Hachiya Formation, Hida River, Kawabe Dam, Kawabe-machi, Kamo-gun, Gifu Prefecture, Japan
Stratigraphy: Lower Miocene
Absolute time: 23.03-15.97 Ma.
Reference: Terada & Suzuki 1998
Analysis: FBD80.
Assignment: Clade including *Reevesia thyrsoidea*.
Score: 3
55. *Reevesia hurnikii* Kvacek
Structure: fruits, leaves, winged seeds
Locality: Most Basin, Bilina, North Bohemia, Czech Republic
Stratigraphy: Miocene
Absolute time: 23.03-5.33 Ma.
Reference: Kvacek 2006
Analysis: FBD80.
Assignment: Clade including *Reevesia thyrsoidea*.
Score: 3
56. *Reevesiapollis reticulatus* (*Ungeria*) (Couper) Krutzsch
Structure: pollen
Locality: Papakaio Formation, Hakataramea Valley, Otago, New Zealand
Stratigraphy: Early Eocene
Absolute time: 56-47.8 Ma.

- Reference: Pocknall 1990; Raine et al. 2011
 Analysis: FBD50; FBD80.
 Assignment: Clade including *Reevesia*,
Ungeria, *Helicteres*, *Mansonia* and
Triplochiton.
 Score: 1
57. *Retitricolporites firmianoides* Ke et Shi.
 Structure: pollen
 Locality: Shahejie Formation, Dongying Sag,
 China
 Stratigraphy: Eocene
 Absolute time: 56-33.9 Ma.
 Reference: Song et al. 1999 cited in Song et
 al. 2004
 Analysis: FBD80.
 Assignment: Stem Sterculioideae.
 Score: 1
58. *Sphinxia ovalis* Reid & Chandler
 Structure: fruit, seeds
 Locality: Sheppey and Harne Bay, London
 Basin, London Clay, England
 Stratigraphy: Early Eocene
 Absolute time: 56-47.8 Ma.
 Reference: Reid & Chandler 1933
 Analysis: FBD50; FBD80.
 Assignment: Stem Dombeyoideae.
 Score: 2
59. *Sterculia patagónica* Berry
 Structure: leaf
 Locality: Río Pichileufu, Carilaufquen Basin,
 Argentina
 Stratigraphy: Late Cretaceous
 Absolute time: 72.1-66 Ma.
 Reference: Berry 1938
 Analysis: FBD80.
 Assignment: Stem Sterculioideae.
 Score: 2
60. *Sterculiaephyllum australis* Dutra
 Structure: leaf
 Locality: Zamek Formation, Baranowski
 Glacier Group, Zamek Hill, Admiralty Bay,
 King George Island, Antarctica
 Stratigraphy: Late Cretaceous
 Absolute time: 72.1-66 Ma.
 Reference: Dutra & Batten
 Analysis: FBD80.
 Assignment: Stem Sterculioideae.
 Score: 2
61. *Sterculia coloradensis* Brown
 Structure: leaf
 Locality: Green River Shale, Colorado, USA
 Stratigraphy: Middle Eocene
 Absolute time: 47.8-37.8 Ma.
 Reference: MacGinitie 1969
 Analysis: FBD50; FBD80.
 Assignment: Stem Sterculioideae.
 Score: 2
62. *Sterculia washburnii* Berry
 Structure: leaf
 Locality: Río Pichileufu, Carilaufquen Basin,
 Argentina
 Stratigraphy: Late Cretaceous
 Absolute time: 72.1-66 Ma.
 Reference: Berry 1938
 Analysis: FBD50; FBD80.
 Assignment: Stem Sterculioideae.
 Score: 2
63. *Sterculinium foetidense* Prakash
 Structure: wood
 Locality: Siwalik beds, Kanandoo sot,
 Kalagarh, Pauri Garhwal District, Uttar
 Pradesh, India
 Stratigraphy: Middle Miocene
 Absolute time: 15.97-11.63 Ma.
 Reference: Prasad 1993
 Analysis: ND14; FBD14; FBD80.
 Assignment: ND: Crown Sterculioideae; FBD:
 Clade including *Sterculia*.
 Score: 2
64. *Sterculinium kalagarhense* (Trivedi & Ahuja)
 Guleria
 Structure: wood
 Locality: Siwalik beds, Hathia sot, Kalagarh,
 Pauri Garhwal District, Uttar Pradesh, India
 Stratigraphy: Middle Miocene
 Absolute time: 15.97-11.63 Ma.
 Reference: Prasad 1993
 Analysis: FBD50; FBD80.
 Assignment: Clade including *Sterculia*.
 Score: 2
65. *Tilia* sp.
 Structure: pollen
 Locality: Rynolds Creek, Gabriola Formation,
 British Columbia, Canada
 Stratigraphy: Maastrichtian

- Absolute time: 72.1-66 Ma.
Reference: Rouse et al. 1970
Analysis: FBD50; FBD80.
Assignment: Stem Tilioideae.
Score: 1
66. *Tilia parvulifolia* H. V. Smith
Structure: leaves, bracts, flowers, fruit and pollen
Locality: Geertson Formation, Lemhi River, Idaho, USA
Stratigraphy: Late Oligocene – Early Miocene
Absolute time: 27.82-20.44 Ma.
Reference: Hall & Swain 1971
Analysis: FBD80.
Assignment: Clade including *Tilia*.
Score: 2
67. *Tilia gigantea* Ettingshausen
Structure: leaf
Locality: Bechlejovice, Usti Formation, Ceske stredohori Mountains, Czech Republic
Stratigraphy: Oligocene
Absolute time: 33.9-23.03 Ma.
Reference: Kvacek & Walther 2004
Analysis: ND14; FBD14; FBD80.
Assignment: ND: Crown Tilioideae; FBD: Clade including *Tilia*.
Score: 3
68. *Triplochitoxylon oregonensis* Manchester
Structure: wood
Locality: Nut Beds, Clarno Formation, Wheeler County, Oregon, USA
Stratigraphy: Middle Eocene
Absolute time: 47.8-37.8 Ma.
Reference: Manchester 1979
Analysis: ND14; FBD14; FBD80.
Assignment: ND and FBD: Clade including *Triplochiton*, *Mansonina* and *Helicteres*.
Score: 3
69. *Triumfetta ovata* MacGinitie
Structure: leaf
Locality: Green River Formation, northwestern Colorado and northeastern Utah, USA
Stratigraphy: Middle Eocene
Absolute time: 47.8-37.8 Ma.
Reference: MacGinitie 1969 cited in Graham 1999.
- Analysis: ND14; FBD14; FBD50; FBD80.
Assignment: ND and FBD: Clade including *Corchorus*, *Pseudocorchorus*, *Heliocarpus* and *Triumfetta*.
Score: 2
70. *Wataria miocenica* (Watari) K. Terada & M. Suzuki
Structure: wood
Locality: Atsumi Formation, Iragawa, Atsumi-machi, Nishitagawa-gun, Yamagata Prefecture, Japan
Stratigraphy: Lower Miocene
Absolute time: 23.03-15.97 Ma.
Reference: Terada & Suzuki 1998
Analysis: FBD80.
Assignment: Stem Helicterioideae.
Score: 3
71. *Wataria oligocenica* (M. Suzuki) K. Terada & M. Suzuki
Structure: wood
Locality: Tsuyazaki Formation, Koinoura, Tsuyazaki-machi, Fukuoka Prefecture, northern Kyushu, Japan
Stratigraphy: Lower Oligocene
Absolute time: 33.9-27.82 Ma.
Reference: Terada & Suzuki 1998
Analysis: FBD50; FBD80.
Assignment: Stem Helicterioideae.
Score: 3
72. *Wataria parvipora* K. Terada & M. Suzuki
Structure: wood
Locality: Hachiya Formation, Minamihora, Kawabe-machi, Kamogun, Gifu Prefecture, Japan
Stratigraphy: Lower Miocene
Absolute time: 23.03-15.97 Ma.
Reference: Terada & Suzuki 1998
Analysis: FBD50; FBD80.
Assignment: Stem Helicterioideae.
Score: 3
73. *Wheeleroxylon atascosense* Estrada-Ruiz, Martínez-Cabrera & Cevallos-Ferriz
Structure: wood
Locality: Olmos Formation, Atascoso Ranch, Múzquiz Municipality, Coahuila, Mexico
Stratigraphy: Late Campanian – early Maastrichtian
Absolute time: 77.85-69.05 Ma.

Reference: Estrada-Ruiz et al. 2010
 Analysis: FBD80.
 Assignment: Stem Malvaceae s.l.
 Score: 2

Muntingiaceae

74. *Muntingiophyllum calabuioides* Pons
 Structure: leaf
 Locality: Mesa Formation, Department of Tolima, Colombia
 Stratigraphy: Pliocene
 Absolute time: 5.33-2.58 Ma.
 Reference: Pons 1977 cited in Graham 2010.
 Note: Paratype located at Collection de Paléobotanique de l'Université Pierre et Marie Curie, Paris, France.
 Analysis: FBD50; FBD80.
 Assignment: Sister to *Muntingia calabura*.
 Score: 2

Thymelaeaceae

75. *Daphne septentrionalis* (Lesquereux) MacGinitie
 Structure: leaf
 Locality: Florissant Formation, Colorado, USA
 Stratigraphy: Early Oligocene
 Absolute time: 33.9-27.82 Ma.
 Reference: MacGinitie 1953 cited in Graham 1999.
 Analysis: FBD50; FBD80.
 Assignment: Clade including *Daphne*, *Thymelaea*, *Dendrostellera*, *Diarthron*, *Stellera* and *Wikstroemia*.
 Score: 2
76. *Pimelea* sp.
 Structure: pollen
 Locality: Murray Basin, Australia
 Stratigraphy: Pliocene
 Absolute time: 5.33-2.58 Ma.
 Reference: Macphail 1999; Raine et al. 2011.
 Analysis: FBD50; FBD80.
 Assignment: Sister to *Pimelea physodes*.
 Score: 1
77. *Retimultiporopollenites qiongbeiensis*
 Structure: pollen
 Locality: Liushagang Formation, Fushan Sag, Beibuwan Basin, China.
 Stratigraphy: Eocene
 Absolute time: 56-33.9 Ma.

Reference: Song et al. 1999 cited in Song et al. 2004
 Analysis: ND14; FBD14; FBD80.
 Assignment: ND and FBD: Clade including *Daphne*, *Thymelaea*, *Dendrostellera*, *Diarthron*, *Stellera* and *Wikstroemia*.
 Score: 1

78. *Thymelaeaspermum bournense* Chandler
 Structure: seeds
 Locality: Bournemouth Marine Beds, Boscombe Sands, Highcliff Sands, Hampshire Basin, England
 Stratigraphy: Eocene
 Absolute time: 56-33.9 Ma.
 Reference: Chandler 1964
 Analysis: FBD50; FBD80.
 Assignment: Stem Thymelaeaceae.
 Score: 2

79. Thymelaeoideae gen. indet. 1
 Structure: pollen
 Locality: Florissant Fossil Beds, Colorado, USA.
 Stratigraphy: Late Eocene
 Absolute time: 41.2-33.9 Ma.
 Reference: Bouchal et al. 2016
 Analysis: FBD80.
 Assignment: Stem Thymelaeaceae.
 Score: 1

Sarcolaenaceae

80. *Xyloolena*-type
 Structure: pollen tetrads
 Locality: Elandsfontyn Formation, Noordhoek, Southwestern Cape, South Africa
 Stratigraphy: Early Miocene
 Absolute time: 23.03-15.97 Ma.
 Reference: Coetzee & Muller 1984
 Analysis: FBD50; FBD80.
 Assignment: Stem Sarcolaenaceae.
 Score: 1

References

- Anberrée JL, Manchester SR, Huang J, Li S, Wang Y, Zhou Z-K. 2015. First fossil fruits and leaves of *Burretiodendron* s.l. (Malvaceae s.l.) in Southeast Asia: Implications for taxonomy, biogeography and paleoclimate. *International Journal of Plant Sciences* 176 (7): 682-696.
- Anderson JA & Muller J. 1975. Palynological study of a Holocene peat and a Miocene coal deposit from NW Borneo. *Review of Palaeobotany and Palynology* 19: 291-351.
- Berry EW. 1922. Contribution to the paleobotany of Peru, Bolivia and Chile. *Johns Hopkins University studies in Geology* 4: 1-221.
- Berry EW. 1929. Tertiary fruits and seeds from northwestern Peru. *Johns Hopkins University studies in Geology* 10: 137-180.
- Berry EW. 1935. A fossil *Cochlospermum* from northern Patagonia. *Bulletin of the Torrey Botanical Club* 62: 65-67., Berry, E.W. 1938. Tertiary flora from the Río Pichileufú. *Geological Society of America. Special Paper* 12: 1-198.
- Berry EW. 1938. Tertiary flora from the Río Pichileufú, Argentina. *Special Papers of the Geological Society of America* 12, VII, 149 pp.
- Bouchal JM, Zetter R, Denk T. 2016. Pollen and spores of the uppermost Eocene Florissant Formation, Colorado: a combined light and scanning electron microscopy study. *Grana* 55(3): 179-245.
- Carvalho MR, Herrera FA, Jaramillo CA, Wing SL, Callejas R. 2011. Paleocene Malvaceae from Northern South America and their biogeographical implications. *American Journal of Botany* 98(8): 1337-1355.
- Chandler MEJ. 1964. The lower Tertiary floras of Southern England. Vol 4. A summary and survey of findings in the light of recent botanical observations. *British Museum (Natural History), London*.
- Coetzee JA, Muller J. 1984. The phytogeographic significance of some extinct Gondwana pollen types from the Tertiary of the Southwestern Cape (South Africa). *Annals of the Missouri Botanical Garden* 71(4): 1088-1099.
- Colmenares-C OA. 1988. A palynological study of samples from three wells of the Boscan Dield, Venezuela. *Revista Técnica Intevep* 8: 83-97.
- Colmenares OA, Teran L, Intevep SA. 1993. A biostratigraphic study of Paleogene sequences in Southwestern Venezuela. *Palynology* 17: 67-89.
- Conwentz H. 1886. Die Angiospermen des Bernsteins. Cistaceae. In: Goeppert HR & Menge A. *Die Flora des Bernsteins und ihre Beziehungen zur Flora der Tertiärformation und der Gegenwart*. V.2. Wilhelm Engelmann, Leipzig. 59-60 pp.
- Duarte L. 1974. Sobre uma flor de Bombacaceae, da Bacia Terciária de Fonseca, MG. *Anais da Academia Brasileira de Ciências* 46: 407-411.
- Dutra TL, Batten DJ. 2000. Upper Cretaceous floras of King George Island, West Antarctica, and their palaeoenvironmental and phytogeographic implications. *Cretaceous Research* 21: 181-209.
- Dutta S, Tripathi SM, Mallick M, Mathews RP, Greenwood PF, Rao MR, Summons RE. 2011. Eocene out-of-India dispersal of Asian dipterocarps. *Review of Palaeobotany and Palynology* 166: 63-68.
- Elsik WC, Yancey TE. 2000. Palynomorph biozones in the context of changing paleoclimate, middle Eocene to Lower Oligocene of the Northwest Gulf of Mexico. *Palynology* 24: 177-186.
- Estrada-Ruiz E, Martínez-Cabrera HI, Cevallos-Ferriz SRS. 2007. Fossil Woods from the late Campanian – early Maastrichtian, Olmos Formation, Coahuila, Mexico. *Review of Palaeobotany and Palynology* 145: 123-133.
- Estrada-Ruiz E, Martínez-Cabrera HI, Cevallos-Ferriz SRS. 2010. Upper Cretaceous woods from the Olmos formation (late Campanian – early Maastrichtian), Coahuila, Mexico. *American Journal of Botany* 97(7): 1179-1194.
- Evans JE. 1991. Paleoclimatology and paleobotany of the Eocene Chumstick Formation, Cascade Range, Washington (USA): a rapidly subsiding alluvial basin. *Palaeogeography, Palaeoclimatology, Palaeoecology* 88: 239-264.
- Feng X, Tang B, Kodrul TM, Jin J. 2013. Winged fruits and associated leaves of *Shorea* (Dipterocarpaceae) from the Late Eocene of South China and their phytogeographic and paleoclimatic implications. *American Journal of Botany* 100(3): 574-581.
- Fittipaldi FC, Simoes MG, Giulietti AM, Pirani JR. 1989. Fossil plants from the Itaquaquecetuba Formation (Cenozoic of the São Paulo basin) and their possible paleoclimatic significance. *Boletim do Instituto de Geociências, Universidade de São Paulo, Publicação Especial* 7: 183-203.

- González CC. 2009. Revisión taxonómica y biogeográfica de las familias de angiospermas dominantes de la "Flora del Hunco" (Eoceno Temprano), Chubut, Argentina. Ph.D. dissertation, Universidad de Buenos Aires, Buenos Aires, Argentina.
- Gottwald H. 1992. Holz aus marinen Sanden des oberen Eozän von Helmstedt. *Palaeontographica Abteilung B* 225: 27-103.
- Graham A. 1985. Studies in Neotropical paleobotany. IV. The Eocene communities of Panama. *Annals of the Missouri Botanical Garden* 72: 504-534.
- Graham A. 1999. Late Cretaceous and Cenozoic history of North American vegetation. Oxford University Press, Oxford.
- Graham A. 2010. Late Cretaceous and Cenozoic history of Latin American vegetation and terrestrial environments. Missouri Botanical Garden Press, St. Louis.
- Graham A, Jarzen DM. 1969. Studies in Neotropical Paleobotany. I. The Oligocene communities of Puerto Rico. *Annals of the Missouri Botanical Garden* 56(3): 308-357.
- Gregory M, Poole I, Wheeler EA. 2009. Fossil dicot wood names, an annotated list with full bibliography. *International Association of Wood Anatomists Journal, Supplement* 6.
- Hall JW, Swain AM. 1971. Pedunculate bracts of *Tilia* from the Tertiary of Western United States. *Bulletin of the Torrey Botanical Club* 98(2): 95-100.
- Hooghiemstra H, van Geel B. 1998. World list of Quaternary pollen and spore atlases. *Review of Paleobotany and Palynology* 104: 157-182.
- Hoorn MC. 1994. An environmental reconstruction of the palaeo-Amazon river system (Middle to late Miocene, NW Amazonia). *Palaeogeography, Palaeoclimatology, Palaeoecology* 112: 187-238.
- Kvacek Z. 2006. Fossil fruits of *Reevesia* (Malvaceae, subfam. Helicterioideae) and associated plant organs (seeds, foliage) from the Lower Miocene of North Bohemia (Czech Republic). *Neues Jahrbuch für Geologie und Paläontologie* 7: 431-448.
- Kvacek Z, Walther H. 2004. Oligocene flora of Bechlejovice at Decín from the neovolcanic area of the Český Středohoří Mountains, Czech Republic. *Acta Musei Nationalis Pragae, Series B, Natural History* 60(1-2): 9-60.
- Kvacek Z, Manchester SR, Akhmetiev MA. 2005. Review of the fossil history of *Craigia* (Malvaceae s.l.) in the northern hemisphere based on fruits and co-occurring foliage. *Transactions of the International Palaeobotanical Conference* V. 1: 114-140.
- Kvacek Z, Wilde V. 2010. Foliage and seeds of malvacean plants from the Eocene of Europe. *Bulletin of Geosciences* 85(1): 163-182.
- Jaramillo CA, Dilcher DL. 2001. Middle Paleogene palynology of central Colombia, South America: A study of pollen and spores from tropical latitudes. *Palaeontographica, Abteilung B. Paläophytologie*: 258: 87-213.
- Lima MR, Salard-Cheboldaef M. 1981. Palynologie des bassins de Gandarela et Fonseca (Eocene de l'état de Minas Gerais, Brésil). *Boletim IG, Instituto de Geociências, Universidade de São Paulo* V. 12: 33-54.
- MacGinitie HD. 1953. Fossil plants of the Florissant Beds, Colorado. Carnegie Institution of Washington publication 599: 1-197.
- MacGinitie HD. 1969. The Eocene Green River flora of northwestern Colorado and northeastern Utah. *University of California Publications in Geological Sciences* 83: 1-140.
- MacGinitie HD. 1974. An early middle Eocene flora from the Yellowstone-Absaroka volcanic province, northwestern Wind River Basin, Wyoming. *University of California Publications in Geological Sciences* 108: 1-103.
- Macphail MK. 1999. Palynostratigraphy of the Murray basin, inland Southeastern Australia. *Palynology* 23(1): 197-240.
- Manchester SR. 1979. *Triplochitioxylon* (Sterculiaceae): A new genus of wood from the Eocene of Oregon and its bearing on xylem evolution in the extant genus *Triplochiton*. *American Journal of Botany* 66(6): 699-708.
- Manchester SR. 1980. *Chattawayia* (Sterculiaceae): A new genus of wood from the Eocene of Oregon and its implications for xylem evolution of the extant genus *Pterospermum*. *American Journal of Botany* 67(1): 59-67.
- Manchester SR. 1992. Flowers, fruits, and pollen of *Florissantia*, an extinct Malvacean genus from the Eocene and Oligocene of Western North America. *American Journal of Botany* 79(9): 996-1008.

- Manchester SR, Herrera F, Fourtanier E, Barron J, Martinez J-N. 2012. Oligocene age of the classic Belén fruit and seed assemblage of North Coastal Peru based on diatom biostratigraphy. *Journal of Geology* 120(4): 467-476.
- Mandal J, Rao MR. 2001. Taxonomic revision of tricolpate pollen from Indian Tertiary. *Palaeobotanist* 50: 341-368.
- Mathews RP, Tripathi SM, Banerjee S, Dutta S. 2013. Palynology, Palaeoecology and Palaeodepositional environment of Eocene lignites and associated sediments from Matanomadh mine, Kutch Basin, Western India. *Journal Geological Society of India* 82: 236-248.
- Mehrotra RC, Awasthi N, Dutta SK. 1999. Study of fossil wood from the upper Tertiary sediments (Siwalik) of Arunachal Pradesh, India and its implication in palaeoecological and phytogeographical interpretations. *Review of Palaeobotany and Palynology* 107: 223-247.
- Menke B. 1976. Pliozäne und ältestquartäre Sporen- und Pollenflora von Schleswig-Holstein. *Geologisches Jahrbuch. Reihe A.* 32: 3-197.
- Muller J. 1968. Palynology of the Pedawan and Plateau Sandstone Formations (Cretaceous-Eocene) in Sarawak, Malaysia. *Micropaleontology* 14 (1): 1-37.
- Muller J. 1970. Palynological evidence on early differentiation of angiosperms. *Biological Reviews* 45: 417-450.
- Muller J. 1981. Fossil pollen records of extant Angiosperms. *The Botanical Review* 47 (1): 1-142.
- Naud G, Suc JP. 1975. Contribution a l'étude paléofloristique des Coirons (Ardeche). Première analyses polliniques dans les alluvios sous-basaltique et interbasaltiques de Mirabel (Miocene supérieur). *Bulletin de la Société Géologique de France* 17: 820-827.
- Parrish JT, Fiorillo AR, Jacobs BF, Currano ED, Wheeler EA. 2010. The Ketavik Formation: A new stratigraphic unit and its implications for the Paleogene paleogeography and paleoclimate of southwestern Alaska. *Palaeogeography, Palaeoclimatology, Palaeoecology* 295: 348-362.
- Pocknall DT. 1989. Late Eocene to early Miocene vegetation and climate history of New Zealand. *Journal of the Royal Society of New Zealand* 19(1): 1-18.
- Pocknall DT. 1990. Palynological evidence for the early to middle Eocene vegetation and climate history of New Zealand. *Review of Palaeobotany and Palynology* 65: 57-69.
- Pons D. 1977. *Muntingiophyllum calaburoides* nov. gen., nov. sp. Elaeocarpaceae fossile de la Formation Mesa (Colombie). 102e Congrès National des Sociétés Savantes, Limoges, France 1: 171-185.
- Prakash U, Dayal R. 1963. Fossil woods of *Grewia* from the Deccan Intertrappean series, India. *Palaeobotanist* 13(1): 17-24.
- Prasad M. 1993. Siwalik (Middle Miocene) woods from the Kalagarh area in the Himalayan foot hills and their bearing on palaeoclimate and phytogeography. *Review of Palaeobotany and Palynology* 76: 49-82.
- Prasad V, Farooqui A, Tripathi SK, Garg R, Thakur B. 2009. Evidence of Late Palaeocene-Early Eocene equatorial rain forest refugia in southern Western Ghats, India.
- Raine JJ, Mildenhall DC, Kennedy EM. 2011. New Zealand fossil spores and pollen: an illustrated catalogue. 4th edition. GNS Science miscellaneous series no. 4 <http://data.gns.cri.nz/sporepollen/index.htm>
- Reid ME, Chandler MEJ. 1933. The London Clay Flora. British Museum (Natural History). London, 561 pp.
- Rouse GE, Hopkins WS Jr, Piel KM. 1970. Palynology of some Late Cretaceous and Early Tertiary deposits in British Columbia and adjacent Alberta. In: RM Kosanke & AT Cross (eds.). *Symposium on palynology of the Late Cretaceous and Early Tertiary*. Geological Society of America vol. 27: 213-246.
- Rust J, Singh H, Rana RS, McCann T, Singh L, Anderson K, Sarkar N, Nascimbene PC, Stebner F, Thomas JC, Solórzano Kraemer M, Williams CJ, Engel MS, Sahni A, Grimaldi D. 2010. Biogeographic and evolutionary implications of a diverse paleobiota in amber from the early Eocene of India. *PNAS* 107(43): 18360-18365.
- Shi G, Li H. 2010. A fossil fruit of *Dipterocarpus* from the middle Miocene of Fujian, China and its palaeoclimatic significance. *Review of Palaeobotany and Palynology* 162: 599-606.
- Song Z-C, Zheng Y-H, Li M-Y, Zhang Y-Y, Wang W-M, Wang D, Zhao C-B, Zhou S-F, Zhu Z-N, Zhao Y-N. 1999. Fossil spores and pollen of China. Vol. 1. The late Cretaceous and Tertiary Spores and Pollen. Beijing: Science Press.
- Song Z-C, Wang W-M, Huang F. 2004. Fossil pollen records of extant angiosperms in China. *Botanical Review* 70: 425-458.
- Srivastava R, Saxena RK, Srivastava G. 2012. Pterospermumocarpon, a new malvacean fruit from the Sindhudurg Formation (Miocene) of Maharashtra, India, and its phytogeographical significance. *Journal of Earth System Science* 121(1): 183-193.

- Taylor DW. 1990. Paleobiogeographic relationships of angiosperms from the Cretaceous and early Tertiary of the North American area. *The Botanical Review* 56: 279-417.
- Terada K, Suzuki M. 1998. Revision of the so-called 'Reevesia' fossil woods from the Tertiary in Japan – a proposal of the new genus *Wataria* (Sterculiaceae). *Review of Palaeobotany and Palynology* 103: 235-251.
- Wheeler EF, Lee M, Matten LC. 1987. Dicotyledonous woods from the Upper Cretaceous of southern Illinois. *Botanical Journal of the Linnean Society* 95 (2): 77-100.
- Wheeler EA, Lehman TM. 2000. Late Cretaceous woody dicots from the Aguja and Javelina Formations, Big Bend National Park, Texas, USA. *IAWA Journal* 21 (1): 83-120.
- Wheeler EA, Lehman TM, Gasson PE. 1994. Javelinoxylon, an upper Cretaceous Dicotyledonous tree from Big Bend national park, Texas, with presumed Malvacean affinities. *American Journal of Botany* 81(6): 703-710.
- Wilf P, Johnson DR, Cúneo NR, Smith ME, Singer BS, Gandolfo MA. 2005. Eocene plant diversity at Laguna del Hunco and Río Pichileufú, Patagonia, Argentina. *American Naturalist* 165: 634-650.
- Wing SL, Bown TM, Obradovich JD. 1991. Early Eocene biotic and climatic change in interior western North America. *Geology* 19: 1189-1192.
- Wolfe JA. 1975. Some aspects of plant geography of the Northern Hemisphere during the late Cretaceous and Tertiary. *Annals of the Missouri Botanical Garden* 62(2): 264-279.
- Wolfe JA. 1977. Paleogene floras from the Gulf of Alaska region. U.S. Geological Survey Professional Paper 997: 1-108.
- Worobiec G, Worobiec E, Szyrkiewicz A. 2012. Plant assemblage from the Upper Miocene deposits of the Belchatow Lignite Mine (Central Poland). *Acta Palaeobotanica* 52(2): 369-413.

CAPÍTULO II

SINGLE-COPY NUCLEAR GENES REVEAL NEW EVIDENCE OF INCONGRUENCE IN PHYLOGENETIC RELATIONSHIPS WITHIN MALVACEAE S.L.¹

RESUMEN

La familia Malvaceae *sensu lato* es un clado que comprende nueve subfamilias, cuyas relaciones filogenéticas no están completamente resueltas y hay inconsistencias entre estudios previos, probablemente debido a que se han utilizado marcadores moleculares convencionales con baja informatividad filogenética. Marcadores convencionales para plantas han sido ampliamente utilizados para estimar las relaciones filogenéticas entre rangos taxonómicos profundos (p.ej. a nivel angiospermas) pero usualmente el poder que tienen para resolver relaciones de grupos más exclusivos es limitado. El presente estudio presenta nueva evidencia de incongruencia en genes para estimar las relaciones filogenéticas de Malvaceae s.l. usando marcadores específicos para el grupo. Con información de *One Thousand Plant Project* (1KP) y el genoma del algodón (*Gossypium raimondii*), el chocolate (*Theobroma cacao*) y *Arabidopsis thaliana*, se aplicó una técnica de minado para obtener genes nucleares de una sola copia. A partir del minado, se seleccionaron cuatro marcadores moleculares para construir la filogenia de Malvaceae s.l. con representantes de las nueve subfamilias. Se comparó la informatividad filogenética de estos cuatro marcadores con marcadores convencionales, dos del cloroplasto (*matk* y *rbcl*) y la región nuclear ITS. Los resultados muestran que los genes minados en conjunto pueden resolver las relaciones filogenéticas de Malvaceae s.l. en diferentes niveles dentro de la filogenia. El análisis de concordancia mostró que uno de los genes minados no es congruente con el resto y que es necesario este tipo de análisis para corroborar las relaciones filogenéticas.

ABSTRACT

The family Malvaceae *sensu lato* is a clade that comprises nine subfamilies, which phylogenetic relationships are not completely solved and are inconsistent, probably due to the low phylogenetic informativeness of conventional molecular markers. Conventional molecular

¹ Esta investigación se realizó con la colaboración de Carolina Granados Mendoza (Instituto de Biología, UNAM, México).

markers for plants are widely used to analyze the relationships among high-rank taxonomic groups, but they frequently provide limited information. In the present study, we provide new evidence of incongruence in the phylogeny for the Malvaceae *s.l.* using group-specific nuclear markers. By applying genome/transcriptome mining with data from the One Thousand Plant Project (1KP) and genome information from cotton, cacao and *Arabidopsis* we obtained single-copy nuclear genes. Four markers were selected for the reconstruction of the phylogeny. To compare the phylogenetic potential of these regions and conventional markers we calculated the phylogenetic informativeness of each nuclear marker, two chloroplast markers (*matk* and *rbcl*) and the nuclear regions ITS. The results show that when the mined nuclear regions are joined it is possible to resolve most of the relationships at different taxonomic levels within the phylogeny, and that incongruence among genes is frequent in the group, causing the prevalence of unresolved phylogenetic relationships.

Key words- Genome-mining, Malvaceae *s.l.*, Molecular Systematics, transcriptome-mining.

Malvaceae *sensu lato* (*s.l.*) includes the traditional families Sterculiaceae, Tiliaceae, Bombacaceae and Malvaceae *sensu stricto* (*s.s.*) (Alverson et al. 1999; Bayer et al. 1999) and comprises ca. 4,550 species and 238 genera (The Plant List, 2017[www.theplantlist.org]). Among the features that characterize the family, androecial morphology is particularly relevant given its high structural and morphological diversity. In addition to the androecium, Malvaceae *s.l.* contains several other important floral and vegetative traits that need further exploration, for example, floral symmetry, structure and placentation of ovaries, floral nectaries and inflorescence pattern (Bayer et al. 1999).

Fundamental work has been conducted to resolve relationships within Malvaceae *s.l.*, which has served as a base for the current understanding of its systematics and evolution (Alverson et al. 1998; Bayer et al. 1999; Baum et al. 2004; Nyffeler et al. 2005). Numerous studies have provided substantial phylogenetic information within sub-families (Whitlock et al. 2001; Tate et al. 2005; Wilkie et al. 2006; Duarte et al. 2011; Brunken & Muellner, 2012), although relationships among the families have remained difficult to estimate, probably due to

the complex genome duplication history in the family (Conover et al. 2018). Importantly, Conover et al. (2018) performed a series of analyses to investigate whole genome duplications in the family and evidenced that Malvaceae *s.l.* has had a complex genome evolution and concluded that there are two possible scenarios, both including incomplete lineage sorting and whole genome duplications in Helicterioideae and Malvatheca.

Plastid molecular markers are widely used to generate phylogenies. In contrast to plastid loci, nuclear markers are less used because they may experience duplication effects resulting in non-homologous sequences. Nevertheless, nuclear information has several properties that make its use appropriate for phylogenetics (Álvarez & Wendel, 2003). Conventional molecular markers have been used to generate phylogenies of high taxonomic ranks (e.g. orders or families) (Soltis et al. 2000), although in some cases these markers have little information to resolve the relationships within such large groups. An alternative to this is Next Generation Sequencing (NGS), which provides a high number of loci capable to resolve phylogenetic relationships. However, NGS application can be very expensive and bioinformatic processing of sequences intricate, especially if the study group has high species richness or numerous genera.

Another alternative to conventional markers is using group-specific nuclear markers. Genomic and transcriptomic information for many plant species is now available in numerous resources (Soltis et al. 2013). This molecular information can be useful for systematic purposes by directly extract phylogenetic informative regions (i.e., genes useful to resolve phylogenetic relationships) with a mining procedure (Granados-Mendoza et al. 2015). The aims of this research are to reconstruct relationships among sub-families of the Malvaceae *s.l.* with information of nuclear single-copy regions and to propose phylogenetic informative loci that can be used to study particularly Malvaceae *s.l.*

MATERIALS AND METHODS

Transcriptomic/genomic mining— The mining was performed using information from the transcriptome database The One Thousand Plant Project (1KP) and the genomes of *Arabidopsis thaliana* (Brassicaceae), a diploid species of cotton (*Gossypium raimondii*, Malvaceae *s.l.*: Malvoideae; CottonGen: <https://www.cottongen.org>), and cacao (*Theobroma cacao*, Malvaceae

s./.: Byttnerioideae; Cacao Genome Database: <https://www.cacaogenomedb.org>). We downloaded transcriptomes (1KP) of three malvacean species, six outgroups representing different families of Malvales, and one species from Moringaceae (Brassicales). Table 1 shows the taxon sampling from the databases.

We performed a transcriptomic/genomic mining to obtain group-specific nuclear genes (Fig. 1). We assembled the transcriptome with *A. thaliana* genes using the program GENEIOUS v. 5.4 (www.geneious.com). We selected those genes that were shared among *A. thaliana* and at least five species with transcriptomic data (1KP) from the order Malvales, resulting in 250 shared genes. From this first set of genes, we selected those that were shared among 1KP taxa and the genomes of *A. thaliana*, *G. raimondii* and *T. cacao*, obtaining 75 genes. To determine whether these genes have only one copy in cotton and cacao, we searched each gene in the Plaza database (<https://bioinformatics.psb.ugent.be/plaza>). This resulted in 12 single-copy genes shared among 1KP taxa, cotton, cacao and *A. thaliana*. The sequences were aligned with respect to the cotton, cacao and *A. thaliana* genomes to design oligonucleotides for each region to amplify by PCR. The primers were designed to amplify a region of the gene; thus, the primers were designed to include the sequence from exon to exon, including one intron. We are going to refer to these regions as loci.

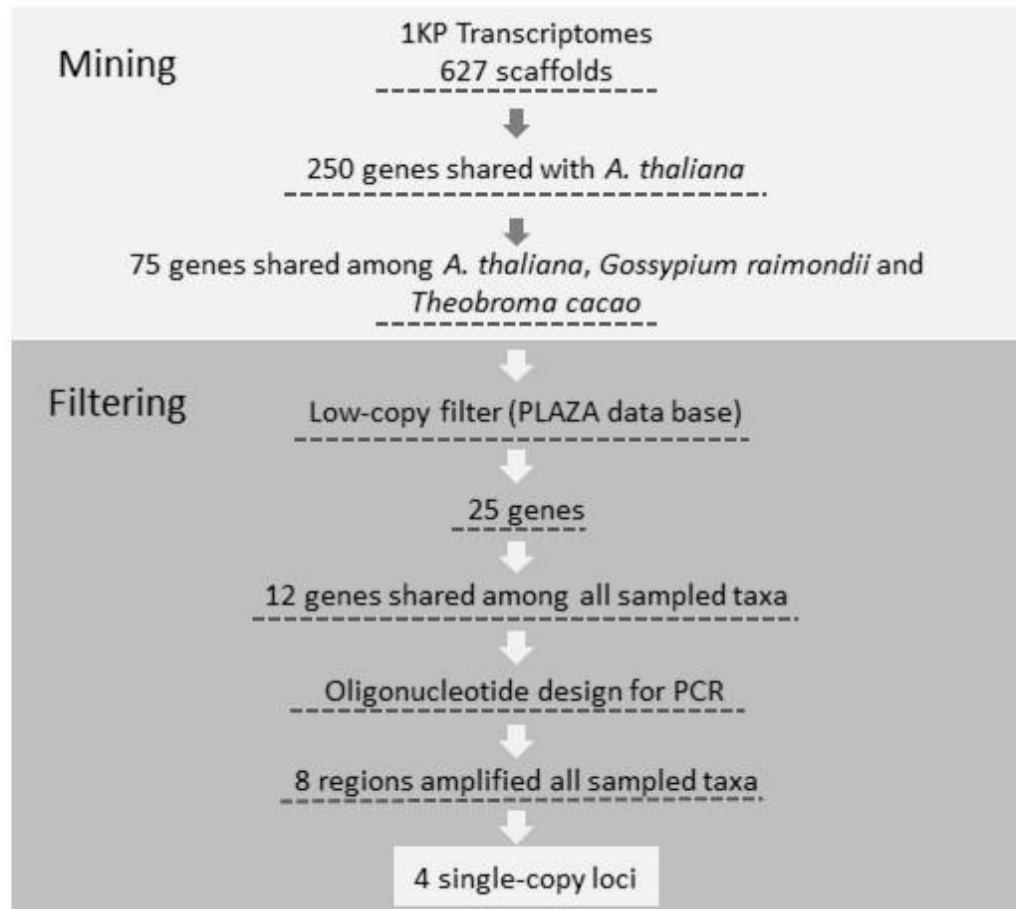


Figure 1. Genome-transcriptome mining. Schematic workflow to obtain single-copy nuclear genes, starting with the mining of shared genes and later the filtering of homologous genes.

Taxon sampling— We applied a CTAB protocol for DNA extraction. We applied PCR protocol for 22 species representing the nine sub-families of Malvaceae s.l. (Table 1). The PCR technique was carried out to amplify the 12 loci. After visual corroboration by electrophoresis, only eight loci were shared by all the species included (Fig. 1). To ensure that we were obtaining one copy of the genes, the PCR products were submitted to a 3 h electrophoresis, after which we purified the gel band with the QIAquick Gel Extraction Kit (Cat. #28706). After this, another electrophoresis was performed to confirm the single-band of the amplified nuclear region. Later, the PCR products were sequenced by Sanger technique at the Laboratorio de Secuenciación Genómica de la Biodiversidad y de la Salud of the Instituto de Biología at UNAM. From the eight shared loci, four regions had only one copy in most of the taxa sampled. Forward

and reverse sequences were assembled in Sequencher 5.2 (Gene Codes Corporation). We edited the raw consensus sequence of the loci in order to remove the extremes that had low density capture and more uncertainty in nucleotide signal. We assembled a molecular matrix in Geneious and built a multiple sequence alignment for each locus with MAFFT v.2.1 (Lanfear et al. 2016).

Table 1. Taxon sampling according to origin of sequences.

#	Origin	Family or Sub-family	Species
1	Lab	Bixaceae	<i>Cochlospermum vitifolium</i>
2	Lab	Bombacoideae	<i>Ceiba pentandra</i>
3	Lab	Bombacoideae	<i>Pseudobombax ellipticum</i>
4	Lab	Bombacoideae	<i>Bernoullia flammea</i>
5	Lab	Bombacoideae	<i>Adansonia digitata</i>
6	Lab	Brownlowioideae	<i>Berrya cubensis</i>
7	Lab	Byttnerioideae	<i>Guazuma ulmifolia</i>
8	Lab	Dombeyoideae	<i>Dombeya wallichii</i>
9	Lab	Grewioideae	<i>Triumfetta lappula</i>
10	Lab	Grewioideae	<i>Heliocarpus appendiculatus</i>
11	Lab	Helicterioideae	<i>Durio zibethinus</i>
12	Lab	Helicterioideae	<i>Helicteres guazumifolia</i>
13	Lab	Malvoideae	<i>Quararibea funebris</i>
14	Lab	Malvoideae	<i>Sida abutifolia</i>
15	Lab	Malvoideae	<i>Chiranthodendron pentadactylon</i>
16	Lab	Malvoideae	<i>Ochroma pyramidale</i>
17	Lab	Malvoideae	<i>Napaea dioica</i>
18	Lab	Sterculioideae	<i>Sterculia apetala</i>
19	Lab	Sterculioideae	<i>Firmiana simplex</i>
20	Lab	Thymelaeaceae	<i>Daphnopsis megacarpa</i>
21	Lab	Tilioideae	<i>Mortoniodendron guatemalense</i>
22	Lab	Tilioideae	<i>Tilia americana</i>
23	1KP	Bixaceae	<i>Bixa orellana</i>
24	1KP	Cistaceae	<i>Cistus inflatus</i>
25	1KP	Dombeyoideae	<i>Dombeya burgessiae</i>

26	1KP	Malvoideae	<i>Hibiscus cannabinus</i>
27	1KP	Malvoideae	<i>Hoheria angustifolia</i>
28	1KP	Moringaceae (Brassicales)	<i>Moringa oleifera</i>
29	1KP	Muntingiaceae	<i>Muntingia calabura</i>
30	1KP	Sarcocaulaceae	<i>Schizolaena sp.</i>
31	1KP	Thymelaeaceae	<i>Edgeworthia papyrifera</i>
32	Genome	Malvoideae	<i>Gossypium raimondii</i>
33	Genome	Byttnerioideae	<i>Theobroma cacao</i>
34	Genome	Brassicaceae (Brassicales)	<i>Arabidopsis thaliana</i>

Phylogenetic estimation— Sequences from the four nuclear loci were aligned with MUSCLE v.3.7 (Edgar 2004). We estimated a partition scheme for all the four mined loci with PartitionFinder 2.1.1 (Lanfear et al. 2016) based on introns and exons, and evaluated substitution model fit for each partition. Subsequently, phylogenetic relationships were estimated with maximum likelihood estimation implemented in RAxML v.8.2.12 (Stamatakis 2014) and Bayesian inference in MrBayes v.3.2.6 (Ronquist et al. 2012). For the Maximum Likelihood analyses, we applied the GTRCAT substitution model and performed 1000 Bootstrap replicates. For Bayesian inference we chose the models estimated with partition finder, which were in all cases the GTR+I+ Γ , except in the first partition from locus M1, which was GTR+I. The MCMC ran 100 million generations, sampling every 1000 steps. All phylogenetic analyses were performed in the CIPRES Science Gateway (Miller et al. 2010). We carried out a series of phylogenetic analyses for each locus without partitions, each locus partitioned in exons and intron, first and second codon positions and third position, a concatenated matrix of the four loci with four partitions representing each locus, and another with nine partitions (two for three loci and three for one locus, based on results from PartitionFinder).

To estimate the degree of concordance among the nuclear loci trees we performed a Bayesian concordance analysis (BCA; Ané et al. 2007) implemented in BUCKy v.1.4.4 (Larget et al. 2010). BUCKy infers the primary concordance tree, which has concordance factor (CF) that represent the proportion of genes that have a specific clade, along with credibility intervals from the MCMC posterior density. For this, we used all the trees in the posterior distribution

obtained from the Bayesian phylogenetic analysis of each individual locus. The parameter α , the *a priori* level of discordance, was set to 1 and ran 100,000 generations.

Phylogenetic informativeness—It is necessary to estimate the phylogenetic informativeness of each locus in order to avoid potential sources of non-phylogenetic signal (PhyInformR: Dornburg et al. 2016). This method detects which loci are phylogenetically informative for the level of phylogenetic resolution proposed in a given phylogeny. To evaluate whether the four nuclear markers have potential to solve phylogenetic relationships within Malvaceae *s.l.* and to what extent, we compared their phylogenetic informativeness with the plastid conventional molecular markers *rbcL* and *trnK-matK*, and the nuclear region ITS. We obtained this information from GenBank. The phylogenetic informativeness was estimated in PhyDesign (López-Giráldez & Townsend 2011), which uses an ultrametric tree as reference and the concatenated matrix of the evaluated sequences. We generated the ultrametric phylogenetic tree with MrBayes v. 3.2.6 (Ronquist et al. 2012) with an alignment including the four nuclear loci, the plastid regions, and the ITS nuclear region and applied a model of strict molecular clock to generate a phylogeny with branch lengths corresponding to relative time. The resulting tree and the alignment served as the input files for PhyDesign.

We examined the obtained relative substitution rates for each locus from PhyDesign, where we detected some sites that had extremely high substitution rates compared to other sites. We removed from the alignment those sites with high substitution sites and performed additionally phylogenetic analyses to evaluate whether these sites affect phylogenetic inference.

RESULTS

Genome-Transcriptome Mining— The transcriptomic mining resulted in 12 nuclear loci with potential to be present in all the species of Malvaceae *s.l.* in a single copy (Fig. 1). After the laboratory procedure, eight loci were shared among most of the sampling, and only four loci appeared to have one copy. For some species it was difficult to obtain the four loci, probably due to multiple copies in the specific taxon. The range of molecular size of each locus is 800-

1000 bp. The four loci recovered are transcription factors (for simplicity, we will refer to the loci with a letter and a number; GTF2H2 = M1; GR07G17490 = M2; GR04G17560 = M3) and one gene that codes for a methylase (MTA = M8).

Phylogenetic inference— The phylogeny derived from the concatenated four nuclear loci is shown in Fig. 2. Malvaceae *s.l.* is a monophyletic group, its sister group is *Muntingia calabura* (Muntingiaceae; Fig. 2). The phylogenetic relationships show a clade conformed by the original Malvadendrina subfamilies and Grewioideae, with high support (1 PP), leaving Byttnerioideae as the sister group of the rest of the subfamilies. Grewioideae is the sister clade of the rest of the subfamilies, although with low support (0.56 PP). A clade formed by the original Malvadendrina is highly supported (1 PP) and is divided in two groups, one conformed by *Chiranthodendron*, *Ochroma*, and Malvoideae + Bombacoideae (Malvatheca) (1 PP); and other, poorly resolved, formed by the remaining subfamilies (0.6 PP). In this last group, all subfamilies are monophyletic, except Tilioideae, whose relationships are not resolved, forming a polytomy.

The Bayesian Concordance analyses evaluate the extent to which each locus is concordant with the rest of the loci by examining the number of clades that each locus support. Given that for some taxa we were unable to amplify some of the loci, the BUCKy analysis has a reduced taxon sampling. We performed two analyses, one with the four loci, where there were two clusters, one of which included only the locus M2; then, we performed another analysis excluding M2 and obtained one cluster. The results from BUCKy analyses indicate that the locus M2 has a contradictory history with the other loci in general, this is particularly notable at the divergence of Malvadendrina and Byttneriina (0.5 CF; Fig. 3). With these results, we performed an additional Bayesian phylogenetic analysis, with only the three concordant loci, M1, M3, and M8; we specified the same parameters as described in the previous section. We noticed that including species with less than two loci sampled influenced the phylogenetic relationships (data not shown but see Fig. 1 for complete taxon sampling). Thus, Fig. 4 was generated only with species with three and four sampled loci. From this analysis, we recovered Malvadendrina and Byttneriina with high support (0.9-1 PP). Tilioideae, represented by *Mortonioidendron*, is sister to Sterculioideae (0.7 PP), and there is a polytomy including these two subfamilies,

Dombeyoideae, and Helicterioideae. *Ochroma* sister to *Malvatheca* was also recovered with high support (1 PP).

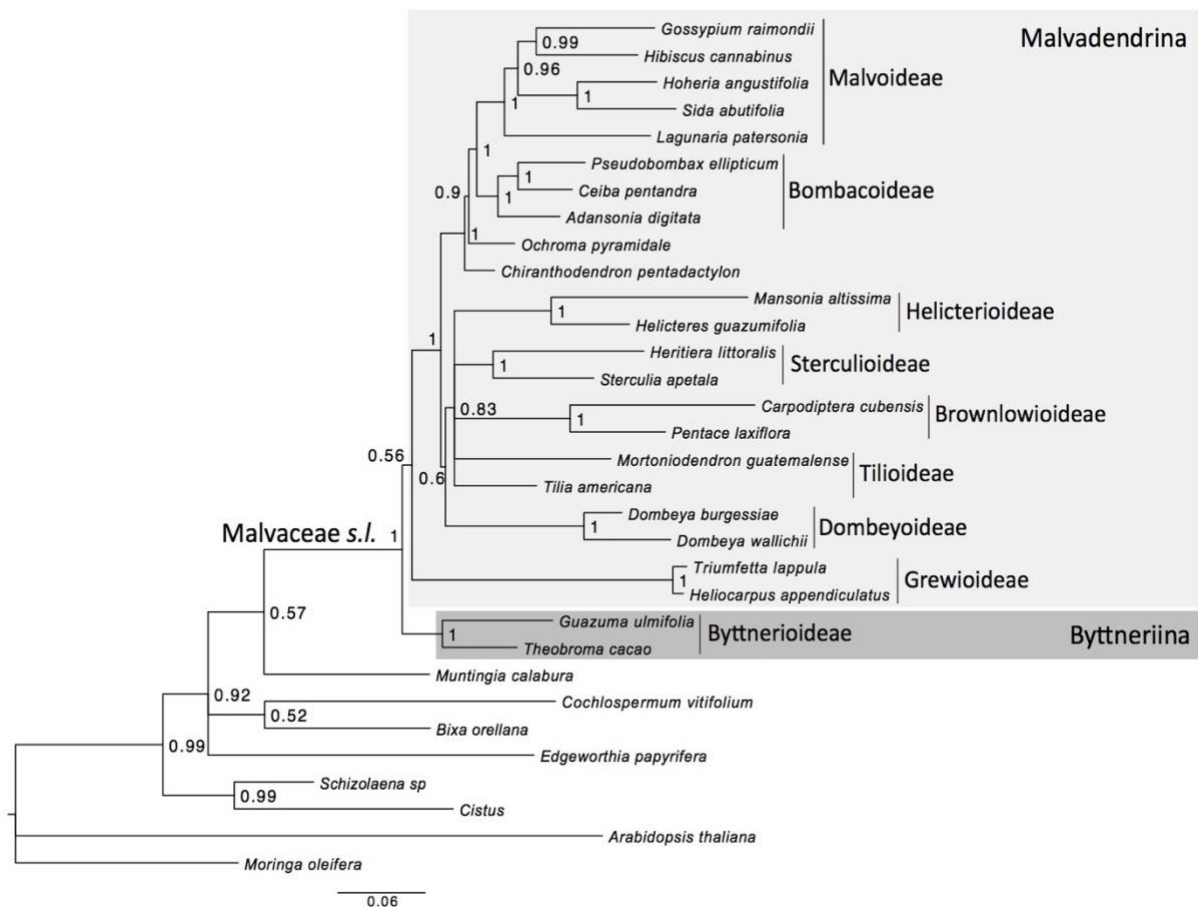


Figure 2. Phylogeny of Malvaceae s.l. Bayesian inference of the four concatenated nuclear loci obtained from the mining procedure. Values associated to nodes represent Posterior Probabilities.

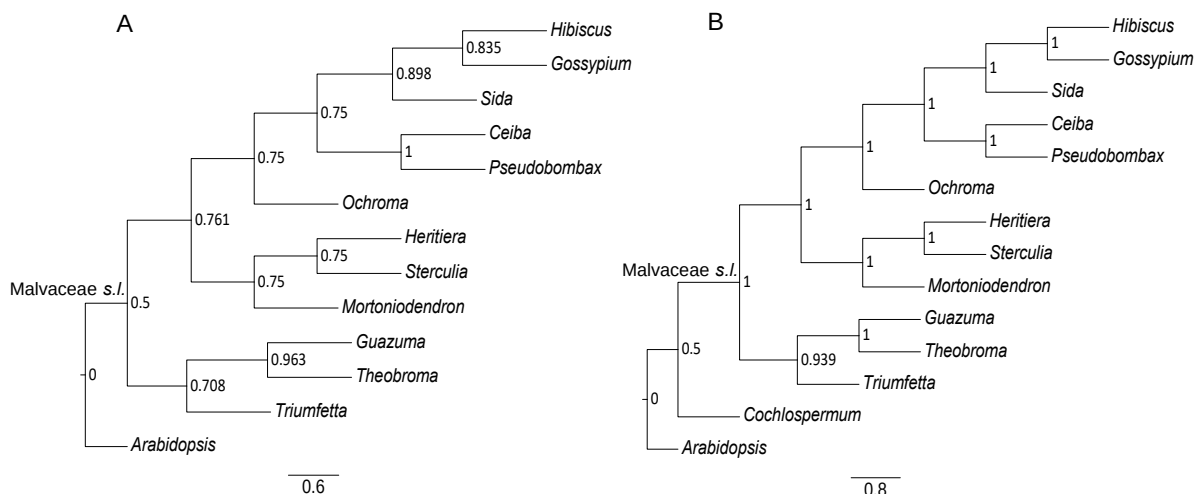


Figure 3. Bayesian concordance analysis. Values associated to nodes correspond to Concordance Factors. A: four loci. B: three loci (M1, M3 and M8).

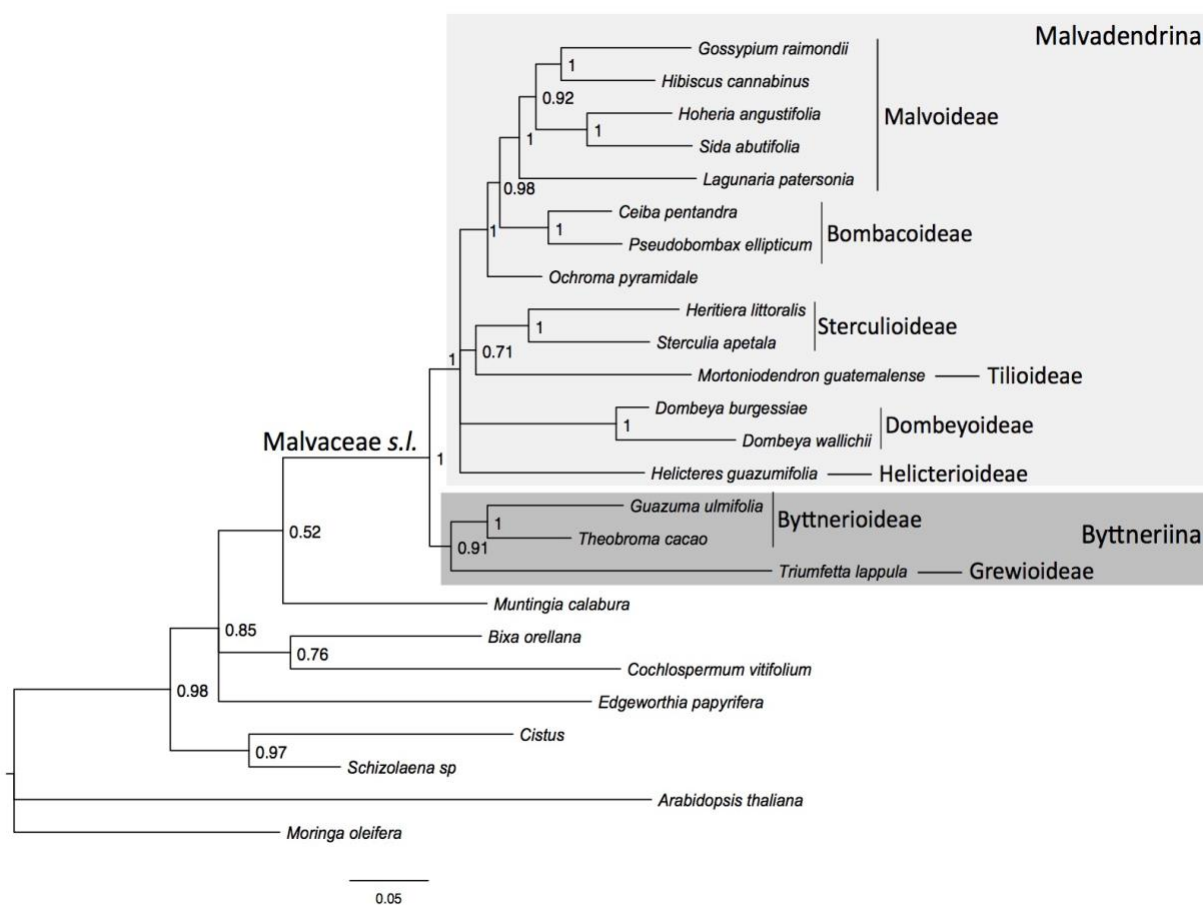


Figure 4. Bayesian phylogram generated with three concatenated nuclear loci (M1, M3 and M8) and a reduced taxon sampling. Values associated to nodes correspond to Posterior Probabilities.

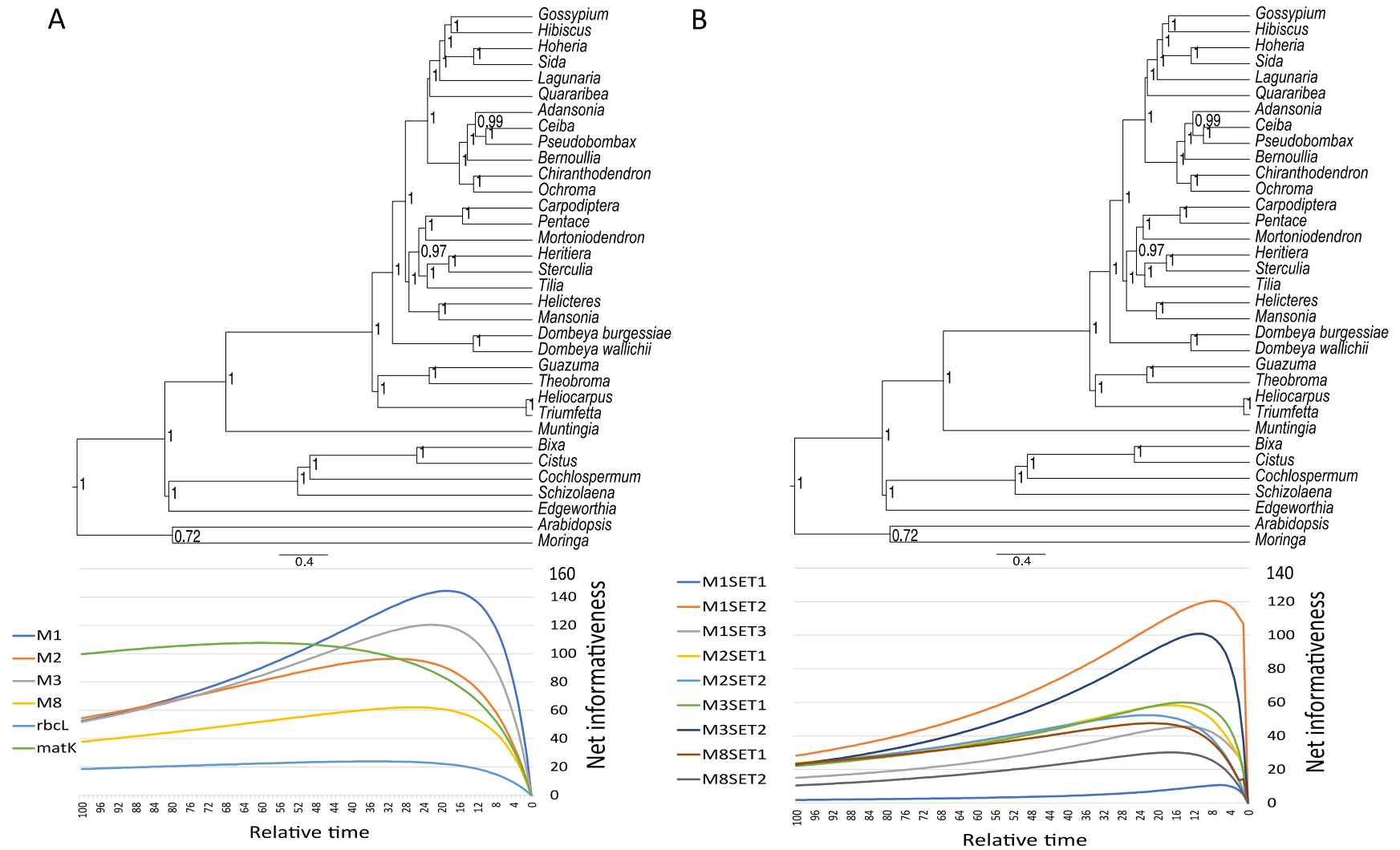


Figure 5. Phylogenetic informativeness. A: four nuclear loci compared to conventional chloroplast molecular markers (*rbcL* and *matK*). B: four nuclear loci divided by partition (exons-introns).

Phylogenetic informativeness— PhyDesign yields the substitution rates of each site in the alignment and the informativeness profile of the contribution of each site to the resolution of phylogenetic relationships at a specific moment in the ultrametric tree. Figure 5 shows the contribution of each of the seven regions to the resolution of Malvacean phylogenetic relationships. The four loci have high informativeness at different phylogenetic levels. The results show that the *rbcL* molecular marker has poor informativeness in general (Fig. 5A). In turn, *matK* has poor informativeness in recent nodes but moderate in deep nodes of the phylogeny (Fig. 5A). We removed the ITS region from the graphics because it has an extremely high substitution rate. The regions M1 and M3 have high values of informativeness in superficial levels of the phylogeny, whereas M2 has moderate values. Note that the M8 region has very low values to solve phylogenetic relationships, although the exon and intron information yield different results (Fig. 5B).

DISCUSSION

Concordance and Phylogenetic informativeness— We obtained four single-copy nuclear markers specific to family Malvaceae *s.l.* Analyzed together, these four loci can provide insights into the phylogenetic relationships of the family. We found that one of these loci is incongruent with the others, the M2 locus. This was observed in almost all nodes (Fig. 3), but markedly at the early diverging lineages of Malvadendrina and Byttneriina, which is interesting because without this analysis we would accept the phylogenetic hypothesis of the four concatenated loci (Fig. 2), which shows, with high support, that Byttnerioideae is the sister group of the rest of the subfamilies. Conover et al. (2018) used 1KP entire transcriptomes for a smaller taxon sampling of Malvaceae *s.l.*, and they found that there was incongruence in numerous genes. This is a first attempt to propose new molecular markers for the entire family, so extensive analyses are needed, for example, with the other mined loci, in order to corroborate the frequency of discordance among genes and their potential to resolve the phylogenetic relationships and the early evolution of the subfamilies.

Each of the four loci have different informative profiles, especially when we analyze them by partition (exons and intron). Thus, together they can resolve most of the relationships

of the family but individually each locus has lesser power. Importantly, the partitioning of the sequences can change the phylogenetic signal of the loci. This was observed when we analyzed the exons and intron of each locus separately (Fig. 5).

Phylogenetic relationships within Malvaceae s.l. —Discordant information was obtained with the nuclear loci, especially observed in the groups Malvadendrina and Byttneriina (Fig. 1). These two clades, originally proposed by Alverson et al. (1999) with chloroplast information, were here inferred in a different arrangement with the concatenated matrix of the four nuclear loci. Byttnerioideae was highly supported as the sister group of the rest of the families (Fig. 1). However, the Bayesian Concordance Analysis detected incongruence among the loci and the phylogenetic analysis with only the loci that produced a cluster recovered the two clades (Fig. 4). Concordance analyses are relevant because concatenation of incongruent genes may lead to high supported clades in a wrong species tree (Chung & Ané 2011), which is probably the cause of the highly supported arrangement of Byttnerioideae as sister to the rest of the malvacean subfamilies.

Malvoideae and Bombacoideae (Malvatheca) are highly supported clades; however, we found *Ochroma* to be outside Malvatheca, similar to its position with plastid sequences (Baum et al., 2004), and not closely related to Malvoideae nor Bombacoideae as in a previous study (Hernández-Gutiérrez & Magallón, 2019). This genus has no circumscription to a particular subfamily, as well as *Chiranthodendron*. Having a high confidence in their position may be of great relevance, given that the evolution of androecium in Malvatheca is characterized by the presence of monothecate anthers, which are bithecate in these two genera and genera allied to them, and it has been hypothesized that the monothecate condition appeared twice independently (von Balthazar et al., 2006; Janka et al., 2008).

The relationships among Dombeyoideae, Helicterioideae, and Brownlowioideae, are unclear with our dataset. The resulting polytomy may be caused by some of the phenomena highlighted by Conover et al. (2018), who inferred that in the history of the family, one or two events of hybridization occurred, probably associated with a member of Helicterioideae (they analyzed the durian genome) and a member of Dombeyoideae. Additionally, Conover et al.

postulate that the short branches subtending these subfamilies, and Tilioideae and Sterculioideae, may be the result of incomplete lineage sorting at different moments of the history of these clades. They conclude that the remaining unresolved relationships in these clades may be due to all these events.

There is plenty genetic data now published and the phylogenies of big inclusive groups, such as Malvaceae *s.l.*, need to be revisited, for example the phylogenetic relationships among the 10 families of the order Malvales were addressed by Hernández-Gutiérrez & Magallón (2019) using six chloroplast loci about only one nuclear region –the ITS. The loci obtained here with the transcriptomic/genomic mining method have two relevant properties regarding systematics. First, they are single-copy regions, which means they are homologous. Second, they are specific for the Malvaceae *s.l.*, thus its resolution may be relatively high, depending on the taxon sampling and the congruence among gene histories. Additionally, they can potentially be used to reconstruct phylogenies of less-inclusive groups (e.g., tribes), as shown by their relatively high phylogenetic informativeness.

For groups where there is genome or transcriptome data, we found that a small number of loci designed specifically for a certain group can be a good option for analysis aiming to resolve its phylogenetic relationships. It will be very interesting to know whether these nuclear single-copy loci provide the same phylogenetic information as NGS data. If that is the case, then it could be enough to select specific loci to answer the same questions without spending a high amount of effort and funding. Previous analyses showed important insights into the relationships of Malvaceae *s.l.* and the other families that compose the order Malvales but failed to completely resolve the relationships among sub-clades. The results in this study are relevant, because until now only “conventional” molecular markers, from chloroplast and the nuclear region ITS, have been used to study the whole Malvaceae *s.l.*

LITERATURE CITED

Alverson W, Whitlock B, Nyffeler R, Bayer C, Baum D. 1999. Phylogeny of the core Malvales: evidence from *ndhF* sequence data. *American Journal of Botany* 86(10): 147-1486.

- Ané C, Larget B, Baum DA, Smith SD, Rokas A. 2007. Bayesian estimation of concordance among gene trees. *Molecular Biology and Evolution* 24: 412-426.
- Baum D, Smith S, Yen A, Alverson W, Nyffele R, Whitlock B, Oldham R. 2004. Phylogenetic relationships of Malvaceae (Bombacoideae and Malvoideae; Malvaceae sensu lato) as inferred from plastid DNA sequences. *American Journal of Botany* 91(11): 1863-1871
- Bayer C, Fay MF, De Bruijn AY, Savolainen V, Morton CM, Kubitzki K, Alverson WS, Chase MW. 1999. Support for an expanded family concept of Malvaceae within a circumscribed order Malvales: a combined analysis of plastid *atpB* and *rbcl* DNA sequences. *Botanical Journal of the Linnean Society* 129:267-303.
- Brunken U & Muellner AN. 2012. A new tribal classification of Grewioideae (Malvaceae) based on morphological and molecular phylogenetic evidence. *Systematic Botany* 37(3): 699-711.
- Buddenhagen CE, Lemmon AR, Lemmon EM, Bruhl J, Cappa J, Clement L, Donoghue MJ, Edwards EJ, Hipp AL, Kortyna M, Mitchell N, Moore A, Prychid CJ, Segovia-Salcedo MC, Simmons MP, Soltis PS, Wanke S, Mast A. 2016. Anchored phylogenomics of angiosperms I: assessing the robustness of phylogenetic estimates. *bioRxiv* 086298: 10.1101/086298.
- Chung Y & Ané C. 2011. Comparing two bayesian methods for gene tree/species tree reconstructions: simulations with incomplete lineage sorting and horizontal gene transfer.
- Conover JL, Karimi N, Stenz N, Ané C, Grover CE, Skema C, Tate JA, Wolff K, Logan SA, Wendel JF, Baum DA. 2018. A Malvaceae mystery: A mallow maelstrom of genome multiplications and maybe misleading methods? *Journal of Integrative Plant Biology* 61(1): 12-31.
- Dornburg A, Fisk NJ, Tamagnan J, Townsend JP. 2016. PhyloInformR: phylogenetic experimental design and phylogenomic data exploration in R. *BMC Evolutionary Biology* 16: 262.
- Duarte MC, Esteves GL, Salatino ML, Walsh KC, Baum DA. 2011. Phylogenetic Analyses of *Eriotheca* and Related Genera (Bombacoideae, Malvaceae). *Systematic Botany* 36(3): 690-701.

- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nuc Acids Res* 32:1792-1797.
- Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2016. PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution* 34(3): 772-773.
- Larget BR, Kotha SK, Dewey CN, Ané C. 2010. BUCKy: Gene tree/species tree reconciliation with Bayesian concordance analysis. *Bioinformatics Applications Note* 26(22): 2910-2911.
- Lemmon AR, Emme SA, Lemmon EM. 2012. Anchored Hybrid Enrichment for Massively HighThroughput Phylogenomics. *Systematic Biology* 0(0):1-18
- Lemmon EM. & Lemmon AR. 2013. High-Throughput Genomic Data in Systematics and Phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* 44: 99-121.
- López-Giráldez F & Townsend JP. 2011. PhyDesign: an online application for profiling phylogenetic informativeness. *BMC Evolutionary Biology* 11: 152.
- Miller MA, 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)*, 14 Nov. 2010, New Orleans, LA pp 1 - 8.
- Nyffeler R, Bayer C, Alverson WS, Yen A, Whitlock BA, Chase MK, Baum DA. 2005. Phylogenetic analysis of the malvadendrina clade (*Malvaceae s.l.*) based on plastid DNA sequences. *Organisms, Diversity & Evolution* 5: 109-123.
- Posada D & Buckley TR. 2004. Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian approaches over Likelihood Ratio Tests. *Systematic Biology* 53(5): 793-808.
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice across a Large Model Space. *Systematic Biology* 61(3): 539-542.
- Soltis DE, Gitzendanner MA, Stull G, Chester M, Chanderbali A, Chamala S. et al. 2013. The potential of genomics in plant systematics. *Taxon* 62: 886-898.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9): 1312-1313.

- Tate JA, Fuertes Aguilar J, Wagstaff SJ, Duke JC, Bodo Slotta TA, Simpson BB. 2005. Phylogenetic relationships within the tribe Malveae (Malvaceae, Subfamily Malvoideae) as inferred from ITS sequence data. *American Journal of Botany* 92(4): 584-602.
- Whitlock BA, Bayer C, Baum DA. 2001. Phylogenetic relationships and floral evolution of the Byttnerioideae ("Sterculiaceae" or Malvaceae s.l.) Based on sequences of the chloroplast gene, *ndhF*. *Systematic Botany* 26(2):420-437.
- Wilkie P, Clark A, Pennington RT, Cheek M, Bayer C, Wilcock CC. 2006. Phylogenetic relationships within the subfamily Sterculioideae (Malvaceae/Sterculiaceae) using the chloroplast gene *ndhF*. *Systematic Botany* 31(1): 160-170.

CAPÍTULO III

MEASURING HETEROGENEITY IN DIVERSIFICATION AND FLORAL EVOLUTIONARY RATES IN MALVACEAE S.L. (MALVALES, EUDICOTYLEDONEAE)^{2,3}

Resumen

Malvaceae *sensu lato* (Bombacaceae, Malvaceae *sensu stricto*, Sterculiaceae y Tiliaceae) comprende alrededor de 4,550 especies, muchas de ellas de interés económico, ecológico y cultural alrededor del mundo, y están distribuidas en nueve subfamilias. Una de las características más distintivas de la familia es la diversidad morfológica y estructural del androceo, pues presenta numerosos tipos de fusión y multiplicación de estambres y anteras, anteras sésiles y diferentes tipos de estaminodios. Las relaciones filogenéticas de las subfamilias no están consistentemente resueltas, lo cual dificulta investigar la evolución de la familia en general. El objetivo de este estudio fue investigar cómo ha evolucionado el androceo en Malvaceae s.l. a través del tiempo y evaluar si la tasa de evolución del androceo es heterogénea a través de los linajes del grupo, y si esto potencialmente está asociado con la heterogeneidad en la tasa de diversificación. Se obtuvieron 218 loci nucleares de la técnica de secuenciación de nueva generación *Anchored Hybrid Enrichment*. Se estimaron las relaciones filogenéticas con Máxima Verosimilitud e Inferencia Bayesiana. Se realizó un fechamiento molecular para estimar los tiempos de divergencia dentro de la familia con nueve calibraciones derivadas del registro fósil. Se utilizaron 12 caracteres morfológicos, la mayoría del androceo, incluyendo información del desarrollo, con los cuales se reconstruyeron los estados ancestrales con mapeo estocástico. Se estimó la heterogeneidad en las tasas de diversificación y morfológica con BAMM (*Bayesian Analysis of Macroevolutionary Mixtures*). La familia se originó en el Cretácico Superior, así como la mayoría de las subfamilias. La evolución del androceo ha tenido una historia compleja y puede estar parcialmente relacionado con la heterogeneidad en la tasa de diversificación encontrada en la familia.

²Esta investigación fue desarrollada en colaboración con Maria von Balthazar (University of Vienna, Austria) y Cássio van den Berg (Universidade Estadual de Feira de Santana, Brasil).

³Para solicitar información sobre caracteres morfológicos escribir a rebecahdezgtz@gmail.com.

ABSTRACT

Malvaceae *sensu lato* (traditionally Bombacaceae, Malvaceae *sensu stricto*, Sterculiaceae and Tiliaceae) comprises ca. 4,550 species, many of which are of economic, ecological and cultural interest worldwide, and are distributed in nine subfamilies. One of the most distinctive characteristics of the family is the morphological and structural diversity of the androecium, given that it presents numerous types of fusion and multiplication of stamens and anthers, sessile anthers and different kinds of staminodes. Phylogenetic relationships among the subfamilies are not consistently resolved, which complicates the understanding of the evolution of the family in general. The goal of this study is to assess how the androecium has evolved in Malvaceae *s.l.* considering absolute time and test whether androecial evolution has high heterogeneity in its evolutionary rate, and whether this may be potentially associated to heterogeneity in diversification rates. We obtained 218 nuclear loci from the next generation sequencing Anchored Hybrid Enrichment technique. We estimated the phylogenetic relationships with Maximum Likelihood and Bayesian Inference. We performed a molecular dating analysis to estimate divergence times within the family with nine calibration points informed from the fossil record. We used 12 morphological characters, most from the androecium, including developmental traits, and with which we reconstructed ancestral states with Stochastic Mapping. We estimated the rate heterogeneity in the species diversification and in the morphology with Bayesian Analysis of Macroevolutionary Mixtures (BAMM). The main results indicate that the family and most of the subfamilies originated in the Upper Cretaceous. The evolution of the androecium has a complex history and may be partially related with the heterogeneity in diversification rate found in the family.

KEY WORDS

Anchored Hybrid Enrichment; androecium; ancestral state reconstruction; diversification rate; floral evolution; Malvaceae *s.l.*; molecular dating.

INTRODUCTION

Species richness is obviously not distributed uniformly across the tree of life (Rabosky 2009). To describe this disparity we can start by estimating the diversification rate (i.e., the difference between speciation rate and extinction rate; Morlon, 2014), if differences in the rate are detected among lineages and/or through time (i.e., rate heterogeneity) it would partially explain the species richness disparity. The heterogeneity in diversification rate of a lineage could be –partially– caused by the acquisition of morphological traits or establishment of ecological interactions (Beaulieu & O’Meara, 2016; Maddison et al., 2007), resulting in a different diversification rate (acceleration or reduction) relative to its sister lineages (Donoghue & Sanderson, 2015).

Current methodological advances in phylogenetic comparative methods to investigate trait evolution allow to evaluate the relationship of intrinsic components of the organisms with lineage evolution. One of these methods is the direct evaluation of how character evolution promotes species diversification (e.g., with the BiSSE model [Binary-State Speciation and Extinction and posterior extensions]; Maddison et al., 2007; O’Meara, 2012); although this evaluation has statistical and conceptual obstacles. In spite of the high rate of type-I error found in the BiSSE model (FitzJohn et al., 2007; FitzJohn, 2012) and the pseudoreplication observed by the non-independence of the distribution of character states in the tips of a phylogeny (Rabosky & Goldberg, 2015; Maddison & FitzJohn, 2015), some studies estimated the magnitude of detected caveats and provide alternative strategies to consider when using character-dependent speciation and extinction methods, as well as the re-evaluation of characters as components of lineage diversification against approaches where only one character leads to or underlies species diversification (Beaulieu & O’Meara, 2016; Rabosky & Goldberg, 2017).

In flowering plants there is a large heterogeneity in diversification rates, whether estimated in a clade-specific analysis (e.g. Bacon et al. 2018) or in an angiosperm-wide analysis (e.g. Magallón et al. 2018). Some particular characters have been proposed to increase diversification rate in angiosperms, including zygomorphy, fruit and dispersal syndromes, herbaceousness, and whole-genome duplication, as well as the tolerance of some climatic variables (Vamosi et al., 2018). Importantly, there are also some studies that point to a more

integrative approach in which the estimation is focused on the extent of the contribution of a particular character to the diversification rate, instead of producing a causal relationship of character-diversification rate (Beaulieu & O'Meara, 2016; Zenil-Ferguson et al., 2019). Empirically, this array of approaches allows to enrich our understanding of how the biology of groups of plants influence their species diversification.

One relevant model to apply these methods is the family Malvaceae *s.l.* This family is the largest in order Malvales, including the traditional families Sterculiaceae, Tiliaceae, Bombacaceae and Malvaceae *s.s.* (Alverson et al. 1999; Bayer et al. 1999). It comprises ca. 4,550 species and 238 genera (The Plant List, 2017), many of which are important components of tropical rainforests and seasonally dry tropical forests around the world. Many species are economically important as well (e.g. cotton [*Gossypium hirsutum*], chocolate [*Theobroma cacao*], cola nut [*Cola acuminata*]). Morphologically, Malvaceae *s.l.* is characterized by the nectary packages in the calyx, and the presence of an epicalyx, an organ formed by bracts surrounding the calyx (Alverson et al. 1999; Bayer et al. 1999). Floral diversity in Malvaceae *s.l.* is remarkable. Two whorled perianth is the basic pattern of the family, with modifications in size, color, ornamentation, and morphology. Gynoecium is less variant, commonly it is formed by five carpels, superior or semi-inferior ovary, one style with one to five stigmas (Bayer & Kubitzki, 2003; Alverson et al., 1999). Noticeable, androecial (male reproductive system) morphology is particularly significant, given its high structural and morphological diversity, for example, stamen splitting and union to form fascicles or tubes. Moreover, the increasing number of stamens in some lineages and the relative low number in others, along with the synnorganization with other floral organs (e.g., the corolla and stamens in *Theobroma cacao*) has been associated with a variety of pollinators (Whitlock et al., 2001; Janka et al., 2008). In general, the floral evolution of Malvaceae *s.l.* has never been directly examined probably due to a lack of consistency in the phylogenetic relationships among the subfamilies.

Previous phylogenetic work has laid the foundation of our current understanding of systematics and relationships within the family (Alverson et al., 1999; Bayer et al., 1999; Baum et al., 2004; Nyffeler et al., 2005). Unfortunately, a consistent phylogenetic resolution among the malvacean subfamilies has not been achieved in part perhaps due to the use of few

molecular markers, and their limited informativeness at different phylogenetic levels (Le Péchon & Gigord, 2014), and probably because the complex molecular evolution observed with a selected taxon sampling of the family (Conover et al., 2018).

After the circumscription of the Malvaceae *s.l.* investigations in which sampling of taxa and genes is increased to corroborate or, in some cases, resolve the relationships among the subfamilies have not been conducted. On the other hand, numerous studies have provided substantial phylogenetic information at the subfamilial level (Whitlock et al., 2001; Tate et al., 2005; Wilkie et al., 2006; Duarte et al., 2011; Brunken & Muellner, 2012), although several important representatives have remained unsampled. The lack of a complete subfamilial phylogenetic tree of Malvaceae *s.l.* hinders a comprehensive study to understand the evolution of its complex floral morphology, which may lead to a better understanding of the general evolutionary history of this important group of plants.

Floral morphology is closely linked with reproductive and pollination systems, ultimately affecting the speciation process in angiosperms (Armbruster, 2017). Particularly, the interaction with pollinators operates as a selective pressure that can drive morphological evolution in flowers (Fenster et al., 2004; Fenster et al., 2015), consequently causing (jointly with other factors) the divergence of species (ecological speciation) through microevolutionary processes, such as character displacement (Muchhala & Potts, 2007; Muchhala et al., 2014).

Phylogenetically, floral evolution has been studied by analyzing separately the floral whorls (i.e., perianth, androecium, and gynoecium; O'Meara et al., 2016; Chartier et al., 2017), or combination of parts in terms of functionality or development (i.e., modules; Dellinger et al., 2019). Some of these studies have mentioned that the androecium has a high rate of evolution—heterogeneity, evolvability—related to pollination (e.g. in Ericales [Chartier et al., 2017], and in Merianieae [Melastomataceae; Dellinger et al., 2019]), and that changes in stamen (androecial organ) number, along with other floral modifications, increase the diversification rate in angiosperms (O'Meara et al., 2016). Given that the androecium is the floral part that most varies in Malvaceae *s.l.*, the hypothesis underlying this research is that floral characters, particularly those of the androecium, have experienced many changes that correspond to a high rate heterogeneity in morphological evolution (i.e., evolvability; Wiens, 2017; Onstein, 2019),

which may have affected the diversification dynamics of the family. To preliminary address this hypothesis, in this study we estimate a genus-level phylogeny of the family using data derived from massive genomic sequencing, trace the evolution of floral characters with emphasis on androecial traits, and measure the amount of evolutionary rate heterogeneity in the floral morphology, as well as measure the heterogeneity in diversification rate, and the timing of potential shifts in the rates.

MATERIAL AND METHODS

Taxon sampling and sequencing— We included 97 species of Malvaceae *s.l.* representing its nine subfamilies and one species from the Muntingiaceae (Malvales) as outgroup. DNA extraction was performed following a CTAB protocol with tissue from field-collected plants, as well as herbarium exemplars. For sequencing we implemented the Anchored Hybrid Enrichment method (AHE; Lemmon et al., 2012). This methodology consists in the capture of DNA scaffolds that hybridize with probes or anchors of targeted loci. For angiosperms, the targeted loci consist of a set of baits obtained from the genomes of different model angiosperm species (Buddenhagen et al., 2016). For a sub-sample of 48 species we applied a meta-preparation technique, which consists on mixing species from phylogenetically distant lineages (e.g., Malvales, Lamiales, Poales), where each lineage has a barcode that identifies it, proceeding with the anchored hybrid capture, and later with the separation of lineages through a bioinformatic pipeline (Granados-Mendoza et al., in prep.). After obtaining the reads for each species and loci, we built a multiple-sequence alignment with MAFFT v.7.402 (Katoh & Standley, 2013). The other 49 species were sequenced via AHE individually. We included six control species, *Theobroma cacao* (Byttnerioideae), *Robinsonella mirandae* (Malvoideae), *Pachira aquatica* (Bombacoideae), *Triumfetta lappula* (Grewioideae), *Muntingia calabura* (Muntingiaceae), and *Luehea speciosa* (Grewioideae), for which we obtained sequences with the meta-preparation and an individual anchored hybrid capture.

Phylogenetic inference and divergence time estimation— We concatenated the matrix of all captured loci that comprised 218 loci and a total length of 249,249 aligned base pairs (bp). To

obtain partitions we used Partition Finder v.2.1.1 (Lanfear et al., 2016) implementing the greedy search method. We introduced the best partition scheme into RAxML v.8.2.12 (Stamatakis, 2014) to estimate phylogenetic relationships with the concatenated matrix and specified the GTRCAT substitution model and 1000 bootstrap replicates. *Muntingia calabura* was specified as outgroup. We performed this analysis in the CIPRES Science Gateway (Miller et al., 2010).

Given that the taxon sampling in this study is relatively low, we decided to implement a node-dating method to estimate divergence times for the Malvaceae *s.l.* In this study, we applied the information of nine fossils obtained from the fossil database of Hernández-Gutiérrez & Magallón (2019), which correspond to the oldest fossils of Malvaceae *s.l.* and its subfamilies. We used uniform calibrations for the subfamilies, with the lower bound of the distribution being the younger limit of the stratigraphic range where the fossil occurred and the upper bound was set to 72 Ma, which is the lower limit of the age of the fossil used to calibrate the root. We estimated the divergence times in BEAST v.2.5.2 (Bouckaert et al., 2014). We applied an uncorrelated relaxed molecular clock, with rates obtained from a lognormal distribution (UCLN), where the prior was set to an exponential distribution with mean = 10, and standard deviation from a gamma distribution with $\alpha = 0.5396$ and $\beta = 0.3819$ (default). Two independent runs were conducted, each with 600 million Markov Chain Monte Carlo (MCMC) generations. We corroborated that the chains mixed and reached convergence in Tracer v.1.6 (Rambaut et al., 2014), and that the Effective Sample Sizes (ESS) of different parameters were larger than 200.

Morphological characters— We focused on floral characters, considering floral organization, inflorescence type, perianth, gynoecium, and specially the androecium; however, we also consider fruit characters with taxonomic importance for some groups (e.g., endocarp), and, additionally, we scored the growth form. In total, we scored 68 morphological characters for all the species included in the phylogeny. Of these characters, 27 are androecial characters. Supplementary File 1 contains a detailed description and justification for all 68 characters and their states (available upon request). We built a character matrix in the platform PROTEUS (Sauquet, 2016), where each data point has an associated reference, so all the information scored is from previously published studies; except the information of floral development of

byttnerioids from an unpublished study (von Balthazar, unpublished data). Importantly, scoring is exemplar-based, for which we only take the information of the species we are including in the taxon sampling, without considering the intra-specific variation and considering what has been published only for the species in our study.

From this database, we focused on morphological characters with the greatest variation among the included species. For example, those characters that were invariable (or nearly so) among the included species were not further examined for ancestral state reconstructions. We selected 12 characters, including the most variable ones, and others were specifically selected to address their particular evolution (e.g. the epicalyx). Table 1 shows the selected characters with the proportion of missing data.

The androecium in *Malvaceae s.l.* is morphologically diverse, as exemplified by the organization of filaments and anthers. In *Malvatheca*, the anthers are monothecate, contrasting with the rest of the subfamilies, where the anthers are bithecate. Previous studies showed that this condition is not a mere reduction of thecae, but a division of filaments in two, each one bearing a theca. In order to compare homologous characters, we modified the definition of von Balthazar et al. (2006) of *androecial unit* (AU), and applied it to the present study. Here we define AU as a structure differentiated from a secondary androecial primordium with or without further subdivisions, i.e., two thecae whether sessile or attached to a single or two filaments (Fig. 1).

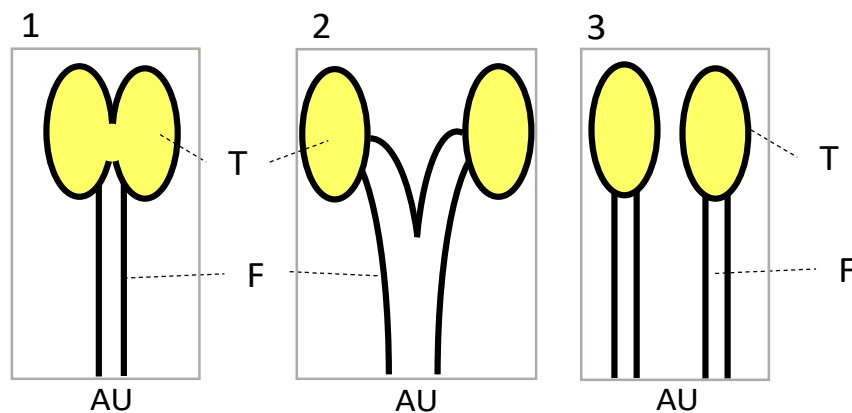


Figure 1. Androecial unit (AU). The three panels represent different forms of the AU that are present in *Malvaceae s.l.* F: filament; T: theca.

Ancestral state reconstruction—We compared two models of evolution, one in which all states have the same rate of evolution (ER: equal rates) and one where each state has its own rate of evolution (ARD: all rates different). For this, we estimated the ancestral state in all nodes of the phylogeny with a Maximum Likelihood approach with the function *rayDISC* in the package *corHMM* (Beaulieu et al., 2013), applying both models of evolution. We then compared both models with the Akaike Information Criterion (AIC) and chose the model with greatest support for all 12 morphological characters (Table 1). With the chosen models, we conducted a Bayesian ancestral state reconstruction with stochastic character mapping (Huelsenbeck et al., 2003), implemented in the R package *phytools* (Revell, 2012). We introduced a sample of 100 trees from the posterior distribution of the Bayesian dating analysis. We generated one simulation for each tree in the sample, and ran the MCMC for 1,000 generations.

The information of floral characters is scarce for some species in our dataset, resulting in a high proportion of missing data (Table 1). To include species with missing morphological data in our Stochastic Mapping analyses, we explicitly assigned a low probability (0.5) for each character state where the actual state is unknown or was not found in the literature (Revell 2016; *phytools* blog: <http://blog.phytools.org/2016/10/stochastic-mapping-discrete-character.html>). For example, for taxon A the character state “present” is scored for a given character, thus it has a probability of 1 of having the state “present”; in contrast, for taxon B there is no information on the condition of that character, so we assigned a probability of 0.5 to be “present” and a probability of 0.5 to be “absent”. This way we are introducing an uninformative prior of the probability of the states for some species in the phylogeny, and the analysis can estimate the overall probability of the character states throughout the phylogeny.

Table 1. Characters selected for ancestral state reconstruction analyses and the models estimated for each character. ER: equal rates; ARD: all rates different. Supplementary File 1 contains detailed character descriptions (upon request).

Character	Description	Character states	Proportion of missing data (%)	Estimated model
1. Habit	Growth form	Herb; shrub; tree; vine.	1.1	ER
2. Floral organization	Arrangement of flowers in the plant (i.e. solitary flowers or type of inflorescence)	Solitary flowers; cyme; raceme; panicle; fascicle; umbel; bicolor unit; corymb.	7.5	ARD
3. Epicalyx	Presence of sterile bracts located immediately externally of the calyx	Presence; absence.	4.3	ER
4. Fusion perianth-androecium	Fusion of inner perianth whorl with androecium	Presence; absence.	22.6	ER
5. Nectary	Presence of floral nectaries	Presence; absence.	44.1	ER
6. Nectary location	Placement of the nectaries in the different floral organs	Inner base of sepals; inner base of petals; between perianth-androecium; androgynophore; between androecium-gynoecium.	51.6	ER
7. Androecial primordia	Organization of the earliest developmental stage of the androecium	Single primordia; complex primordium; common primordia; ring primordium.	58.1	ER
8. Androecial organization	Number of whorls of the androecium	One whorl; two whorls; more than two whorls.	38.7	ER
9. Androecial whorl differentiation	Whorl differentiation in fertile or sterile structures	Two whorls + inner whorl sterile; both whorls fertile; only one whorl fertile; only one whorl fertile & sterile.	51.6	ARD
10. Position androecial whorl-perianth	Position of the outer androecial whorl relative to inner perianth whorl	Alternate; opposite.	63.4	ER
11. Presentation androecium	Architecture of the androecium at anthesis	Staminal tube with sessile AU; staminal tube with lateral stalked AU; Staminal tube fasciculate with stalked AU; AU on free filaments; staminal tube divided in	5.4	ARD

		androecial lobes with sessile AU; staminal tube with terminal stalked AU.		
12. Total androecial units (AU) (discrete)	Discrete categories of the total number of androecial units in a flower	5-20; 20-50; 50-100; more than 100.	14.0	ARD

Diversification rate heterogeneity— To estimate whether there is heterogeneity in diversification rate in the phylogeny of the family, and to identify the phylogenetic location of potential shifts, we implemented the Bayesian analysis of macroevolutionary mixtures (BAMM v.2.5.0; Rabosky 2014; Rabosky et al. 2014). BAMM estimates diversification and morphological rate shifts under a compound Poisson process through time and among lineages, using reversible-jump MCMC to evaluate models that vary in the number of shifts proposed (Rabosky 2014). We used the Maximum Clade Credibility (MCC) tree from the Bayesian inference dating. Initial priors were set according to an analytical procedure performed in the R package BAMMtools (Rabosky et al. 2014). We informed the proportion of species richness in each genus included in the phylogeny and the species for genera not sampled in this study were added to its closest relative according to previously published phylogenetic studies. We ran the analysis for 10 million generations. We evaluated the convergence of chains and with the package coda (Plummer et al. 2006) we corroborated that the ESS of the MCMC was 200 or above.

Morphological rate heterogeneity— We also evaluated whether there is heterogeneity in the rate of morphological (floral) evolution. With stochastic mapping we obtained the ancestral state in all nodes of the phylogeny and, additionally, the probability of a species having each character state (i.e., the states at the tips). Species for which data was available had probability of 1 of having the scored state and 0 of having alternative states; in turn, for species with missing data, we obtained posterior probabilities of having each one of the states for all 12 characters. BAMM analyses require morphological matrices to be complete (i.e., disallow missing data). Therefore, for species with missing data, we took the highest probability of character states to assemble a “probabilistically” complete dataset to evaluate morphological

rate heterogeneity in BAMM. We ran analyses for the 12 characters in BAMM; we obtained initial priors in BAMMtools for morphological evolution rate. We ran 300-500 million generations, depending on the number of character states, because more generations were needed to reach convergence of Markovian chains as more states are considered for a given character. We corroborated that the MCMC analysis converged and that the ESS for the parameters were equal or higher than 200. For diversification and morphological data, we calculated the posterior probabilities of the number of shifts and compared this to the null model of no heterogeneity (i.e., no shifts); we corroborated that the posterior probability was not overlapped with the prior probability.

RESULTS

Phylogenetic relationships— The molecular matrix for phylogenetic analyses comprised 97 species that included representatives of the nine subfamilies and 249,249 bp. All subfamilies were recovered as monophyletic except Helicterioideae, because *Durio zibethinus* is sister to a clade that comprises Tilioideae, Dombeyoideae and Brownlowioideae, whereas all remaining members of Helicterioideae form a clade that is sister to the remaining Malvadendrina (Fig. 1). Support values for all the clades are higher than 70 BS and 0.8 PP (Fig. 2-3). The phylogenetic relationships obtained with BEAST2 were similar to those obtained with Maximum Likelihood, except for Helicterioideae, which in the former case is monophyletic with high support (1 PP) (Fig. 3).

We recovered two major clades within Malvaceae *s.l.*, Byttneriina (Byttnerioideae and Grewioideae) and Malvadendrina (Fig. 2). Within Malvadendrina, Helicterioideae is sister to the remaining subfamilies, which form a clade consisting in two groups, one including Tilioideae and Sterculioideae sister to Brownlowioideae and Dombeyoideae, and the other group is the clade called Malvatheca, which includes Bombacoideae and Malvoideae.

Divergence time estimation— Table 2 shows the estimated ages and the 95% Highest Posterior Density interval (HPD). The origin of stem group of Malvaceae *s.l.*, which in this study is the divergence time from *Muntingia*, was estimated to be 128.3 Ma (100.2-170.6 95% HPD), and

the crown age was estimated at 89.5 Ma (83.7-96.3 95% HPD). The subfamilies were estimated to originate (stem group) between 78 and 65 Ma, between the Upper Cretaceous and early Paleogene. Only Bombacoideae and Malvoideae diversified (crown group age) at the end of the Upper Cretaceous, the rest of the subfamilies diversified in the Paleogene, between the Paleocene and Eocene.

Table 1. Estimated divergence times of Malvaceae *s.l.* and subfamilies. Values represent the mean estimated age and the values inside brackets represent the 95% Highest Posterior Density interval (HPD).

Group	Stem age	Crown age
Malvaceae <i>s.l.</i>	128.3 [100.2-170.6]	89.56 [83.7-96.3]
Bombacoideae	77.8 [71.7-85.2]	68.2 [66-71.3]
Brownlowioideae	67.6 [62-72]	59.5 [56-64.3]
Byttnerioideae	65.7 [59.3-72]	44.6 [35.7-62]
Dombeyoideae	67.6 [62-72]	47.8 [39-57]
Grewioideae	65.7 [59.3-72]	60.9 [55.5-65.4]
Helicterioideae	85.2 [75.5-92.9]	45.9 [37.8-61.5]
Malvoideae	77.8 [71.7-85.2]	69.6 [64.2-74.7]
Sterculioideae	68.5 [66-71.5]	38 [21-62]
Tilioideae	68.5 [66-71.5]	56.9 [47.8-64.5]

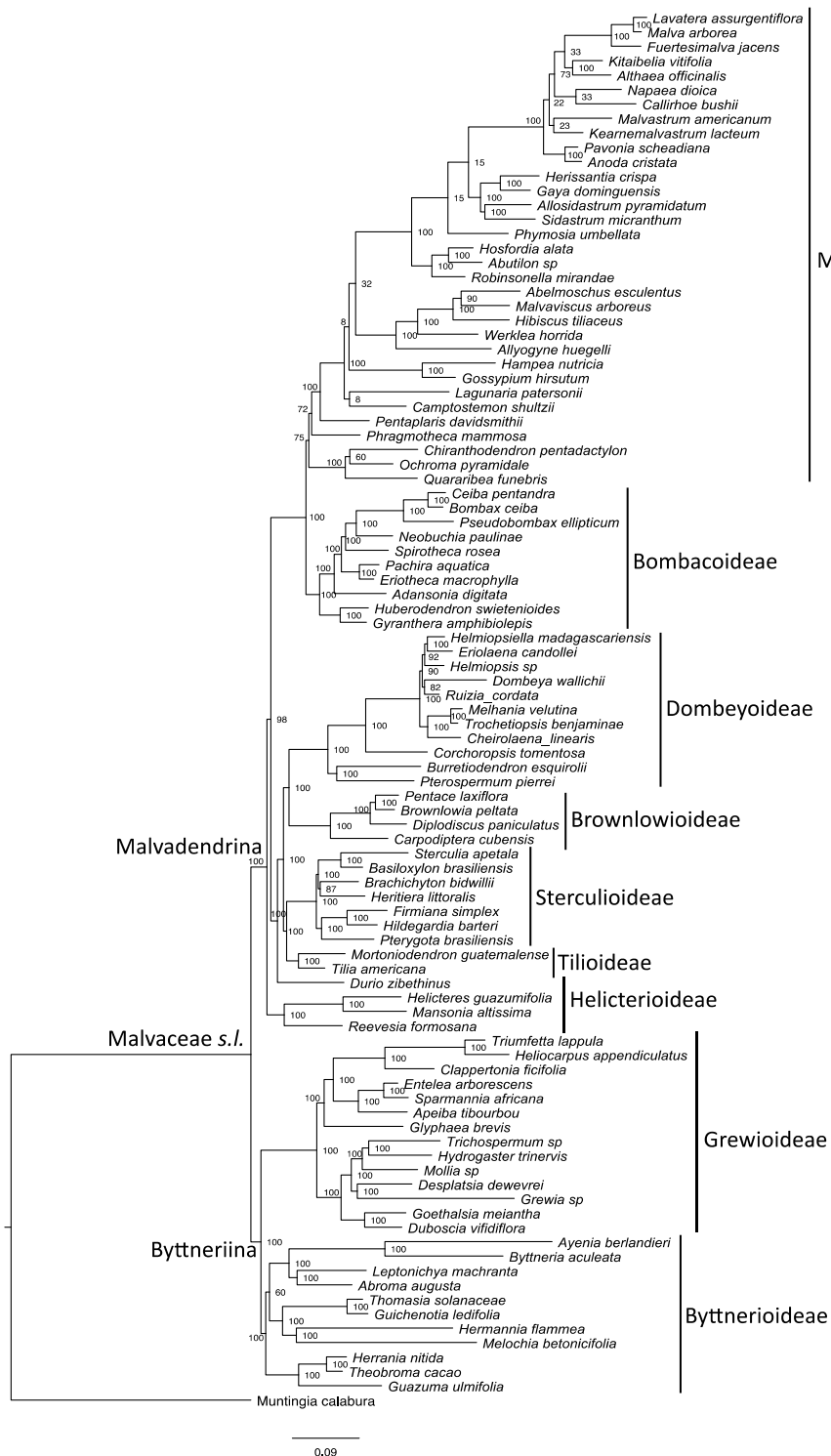


Figure 2. Phylogenetic relationships of Malvaceae s.l. Maximum Likelihood estimation with the concatenate matrix of 218 loci. Numbers associated to nodes represent Bootstrap support.

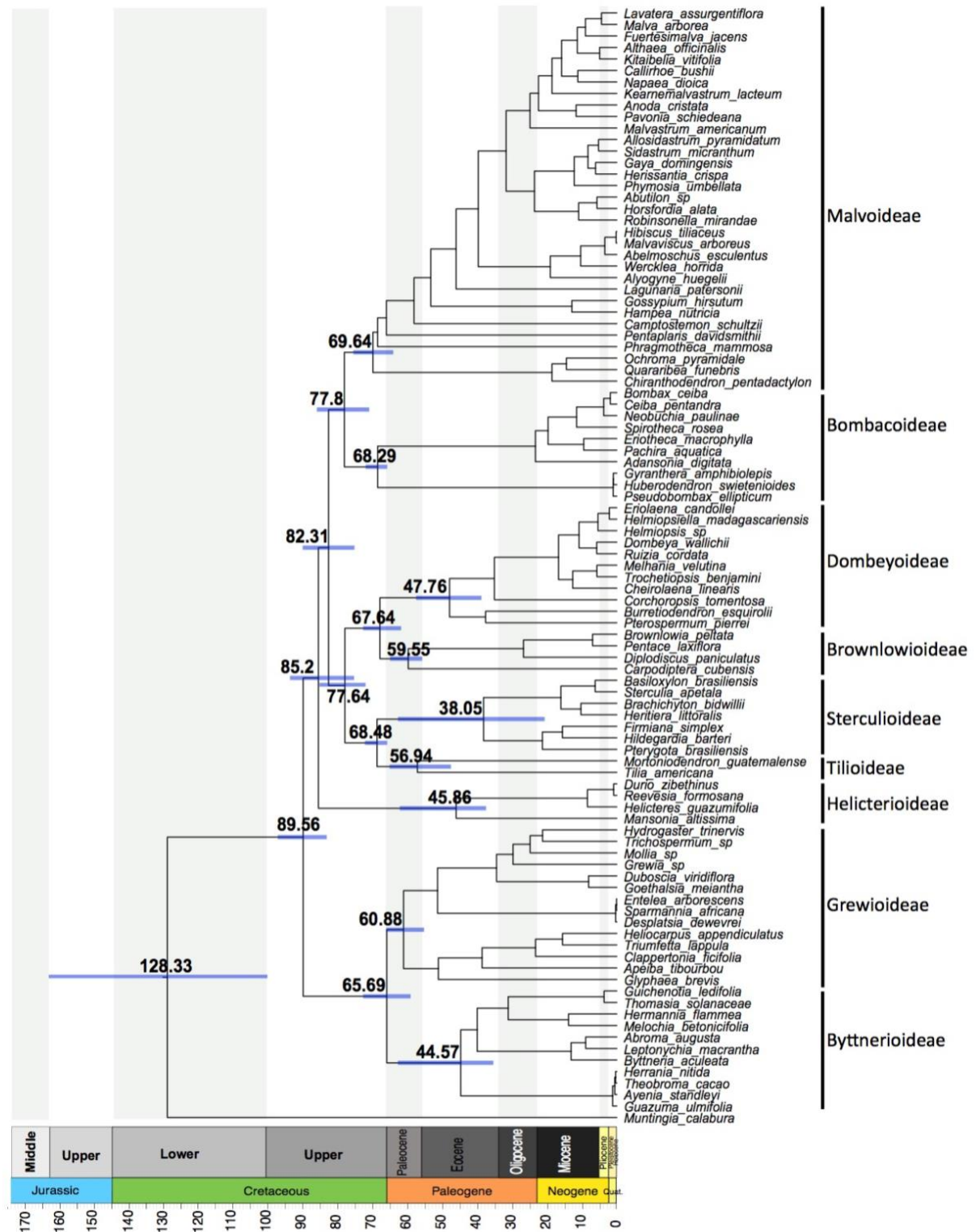


Figure 3. Malvaceae s.l. chronogram obtained with a Bayesian node-dating approach using nine calibration points. Values associated to nodes are estimated ages and bars represent the 95% Highest Posterior Density.

Ancestral state reconstruction—Table 1 shows the 12 selected characters and the proportion of missing data. The characters that have the highest proportion of missing data (50-64 %) are nectary location, androecial primordia, androecial whorl differentiation and the position of the androecial whorl relative to the perianth. We acknowledge that the high proportion of missing data for these characters impacts on the results, therefore we treat the results of these character reconstructions as tentative.

We evaluated which model of state transitions fitted each character best (Table 1). Later, we estimated ancestral states with stochastic character mapping and obtained posterior probabilities of the ancestral state at the nodes –and missing data at the tips– of the phylogeny. The results of the ancestral state reconstructions are shown in Table 3, where the state with the highest posterior probability is presented (the corresponding graphic reconstructions are shown in Supplementary Material).

Our reconstruction suggests that the most recent common ancestor (MRCA) of *Malvaceae s.l.* was a tree in which flowers were solitary or arranged in panicles, flowers lacked an epicalyx, the nectary was placed in the inner base of sepals, the perianth was probably conformed by two differentiated whorls (i.e., calyx and corolla), and was free from the androecium; the androecium was conformed by two whorls, the outer one opposite the (inner) perianth whorl (i.e., obdiplostemonous), and the inner one sterile, all androecial organs developed from a common primordia; at anthesis, the androecium was presented in a staminal tube with lateral stalked AU, which were 20-50 in number (Fig. 4). We did not reconstruct the ancestral state of the gynoecium; however, based on the morphological matrix, probably the ancestral condition was a flower with superior ovary, five united carpels, and a single style (matrix available upon request). The ancestral character states of the subfamilies are similar to the ancestral state of the whole family, with a few deviations. For example, most of the subfamilies are reconstructed to have cymes instead of solitary flowers or panicles; the androecial organization changed from inner whorl sterile to both whorls fertile in *Dombeyoideae* and *Malvoideae*, and only one whorl fertile in *Grewioideae*. The total number of AU varied in *Bombacoideae* (more than 100), *Byttnerioideae*, *Helicterioideae* and *Sterculioideae* (5-20).

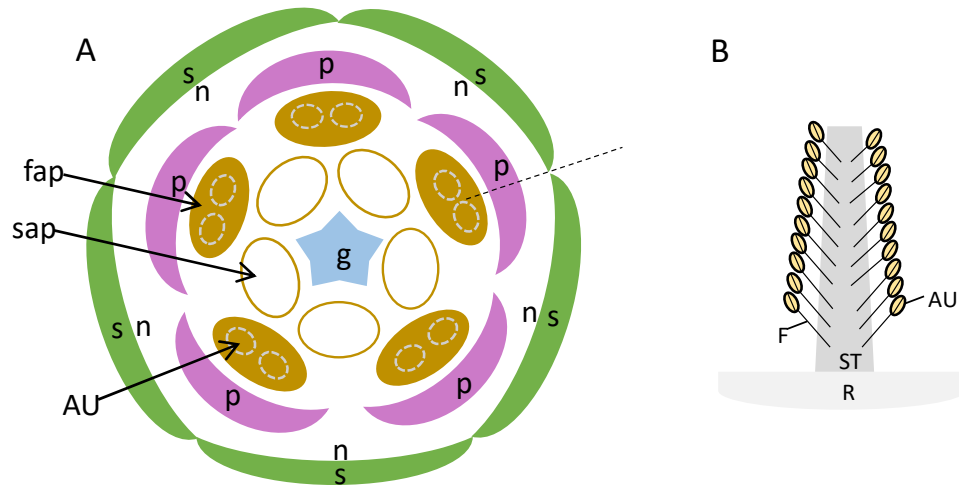


Figure 4. Graphical summary of the ancestral character reconstructions of the most recent common ancestor (MRCA) of the Malvaceae s.l. (A) Floral diagram including results of developmental traits. Colors represent the different floral whorls: green represent the calyx; pink, the corolla; yellow, the androecium; and blue, the gynoecium. Dotted line marks the opposite position of the androecium relative to the corolla (i.e., obdiplostemonous condition). s: sepal; n: nectary; p: petal; AU: androecial unit (for simplicity, two AU per primordium); fap: fertile androecial primordium; sap: sterile androecial primordium; g: gynoecium. (B) Diagram of the reconstructed androecial presentation or architecture at anthesis. AU: androecial unit; F: filament; ST: staminal tube; R: receptacle.

Table 2. Ancestral state reconstructions for the whole family and subfamilies, and rate heterogeneity of each character. Results show the state with highest posterior probability from the stochastic character mapping. Number of shifts in evolutionary rate of each character estimated with BAMM (95% posterior probability).

Character	Malvaceae s.l.	Bombacoideae	Brownlowioideae	Byttnerioideae	Dombeyoideae	Grewioideae	Helicterioideae	Malvoideae	Sterculioideae	Tilioideae	Number of shifts
1. Habit	Tree	Tree	Tree	Shrub	Tree	Tree	Tree	Tree	Tree	Tree	18
2. Floral organization	Solitary flowers/Panicle	Cyme	Cyme	Solitary flowers	Cyme	Panicle	Cyme	Cyme	Cyme	Cyme	25
3. Epicalyx	Absence	Equivocal	Absence	Absence	Equivocal	Absence	Presence	Equivocal	Absence	Absence	20
4. Fusion perianth-androecium	Absence	Absence	Absence	Absence	Absence	Absence	Absence	Presence	Absence	Absence	9
5. Nectary	Presence	Presence	Presence	Presence	Presence	Presence	Presence	Presence	Presence	Presence	5
6. Nectary location	Inner base of sepals	Inner base of sepals	Inner base of sepals	Inner base of sepals	Inner base of sepals	Inner base of sepals	Inner base of sepals	Inner base of sepals	Inner base of sepals	Inner base of sepals	7
7. Androecial primordia	Common primordia	Complex primordium	Common primordia	Common primordia	Common primordia	Common primordia	Common primordia	Complex primordium	Common primordia	Common primordia	9
8. Androecial organization	Two whorls	Two whorls	Two whorls	Two whorls	One whorl	Two whorls	Two whorls	One whorl	Two whorls	Two whorls	10
9. Androecial whorl differentiation	Inner whorl sterile	Inner whorl sterile	Inner whorl sterile	Inner whorl sterile	Both whorls fertile	Only one whorl fertile	Inner whorl sterile	Inner whorl sterile/Both whorls fertile	Inner whorl sterile	Inner whorl sterile	23
10. Position androecial whorl-perianth	Opposite	Opposite	Opposite	Opposite	Opposite	Opposite	Opposite	Opposite	Opposite	Opposite	6
11. Presentation androecium	Androecial tube with lateral stalked AU	Androecial tube with lateral stalked AU	Androecial tube with lateral stalked AU	Androecial tube with lateral stalked AU	Androecial tube with lateral stalked AU	Androecial tube with lateral stalked AU	Androecial tube with lateral stalked AU	Androecial tube with lateral stalked AU	Androecial tube with lateral stalked AU	Androecial tube with lateral stalked AU	27
12. Tota AU (discrete)	20-50	More than 100	5-20	20-50	20-50	20-50	5-20	20-50	5-20	5-20/20-50	11

Diversification rate heterogeneity— Results from the diversification rate estimation are observed in the phylorates shown in Figure 5, where the estimated rate at each segment of the branches is the mean of the marginal posterior density of the estimated rate, either speciation, extinction, or net diversification. BAMM yields the posterior probability of a set of changes in the rate (i.e., shifts). We consider the 95% credibility set of shifts (configurations) obtained from the posterior density (Fig. 6). BAMM yielded, consistently, two shifts in speciation rate (Fig. 5): one increase within Malvoideae and the other, a decrease, in the outgroup, *Muntingia calabura*, here representing the whole family Muntingiaceae (three species).

Our results show heterogeneity in diversification rates in the phylogeny of Malvaceae *s.l.* Particularly, shifts in diversification rate are located, in general, in *Muntingia*, within Malvoideae, core Bombacoideae, Theobromeae (Byttnerioideae), and a clade conformed by *Desplatsia*, *Entelea* and *Sparmannia* within Grewioideae (hereafter, the DES clade). The nature of the shifts, however, is different among these groups. The shift in rate within Malvoideae was located on the branch subtending tribes Malveae + Hibisceae, or only in the tribe Malveae (Figs. 5-6), represented by an increase in speciation and maintaining a low extinction rate (Fig. 5). In core Bombacoideae is observed as a slight increase in both speciation and extinction rates, i.e. an increase in the turnover rate. In Theobromeae the shift is a speciation rate increase and an even higher extinction rate (Fig. 5). The DES clade, as well as *Muntingia*, are characterized by a shift that corresponds to an increase in extinction rate (Fig. 5); however, the former is not included in the 95% posterior density (Fig. 6).

Morphological rate heterogeneity— For morphological characters we extracted the maximum *a posteriori* probability configuration estimated by BAMM for the 12 selected, highly variable characters (Fig. 7), where each shift (Table 3) is observed in the phylogeny. The number and placement of shifts varied among the characters. Characters with low relative rate heterogeneity, measured by the number of shifts, are the position of androecium relative to perianth (character 10) and the presence of nectary and its position (characters 5 and 6). Characters with high relative rate heterogeneity are the presentation of the androecium

(character 11), androecial whorl differentiation (character 9), floral organization (character 2), and the presence of epicalyx (character 3).

The placement of the shifts (Fig. 7) coincided in particular groups. For example, on the branch subtending the DES clade (Grewioideae) there are shifts for seven characters (habit, floral organization, presence of nectary, androecia primordia, androecial whorl differentiation, presentation of the androecium, and total number of AU). Malvoideae has an increase in evolutionary rate in two characters (habit and androecial organization) and Malvoideae + Bombacoideae (Malvatheca) in three characters (presence of epicalyx, fusion of perianth to androecium, and androecial primordia). Tribe Theobromeae has a shift in six characters (habit, floral organization, presence of epicalyx, androecial primordia, presentation of androecium and total number of AU). In general, all subfamilies have shifts in morphological evolutionary rates, however Tilioideae and Brownlowioideae had comparatively fewer.

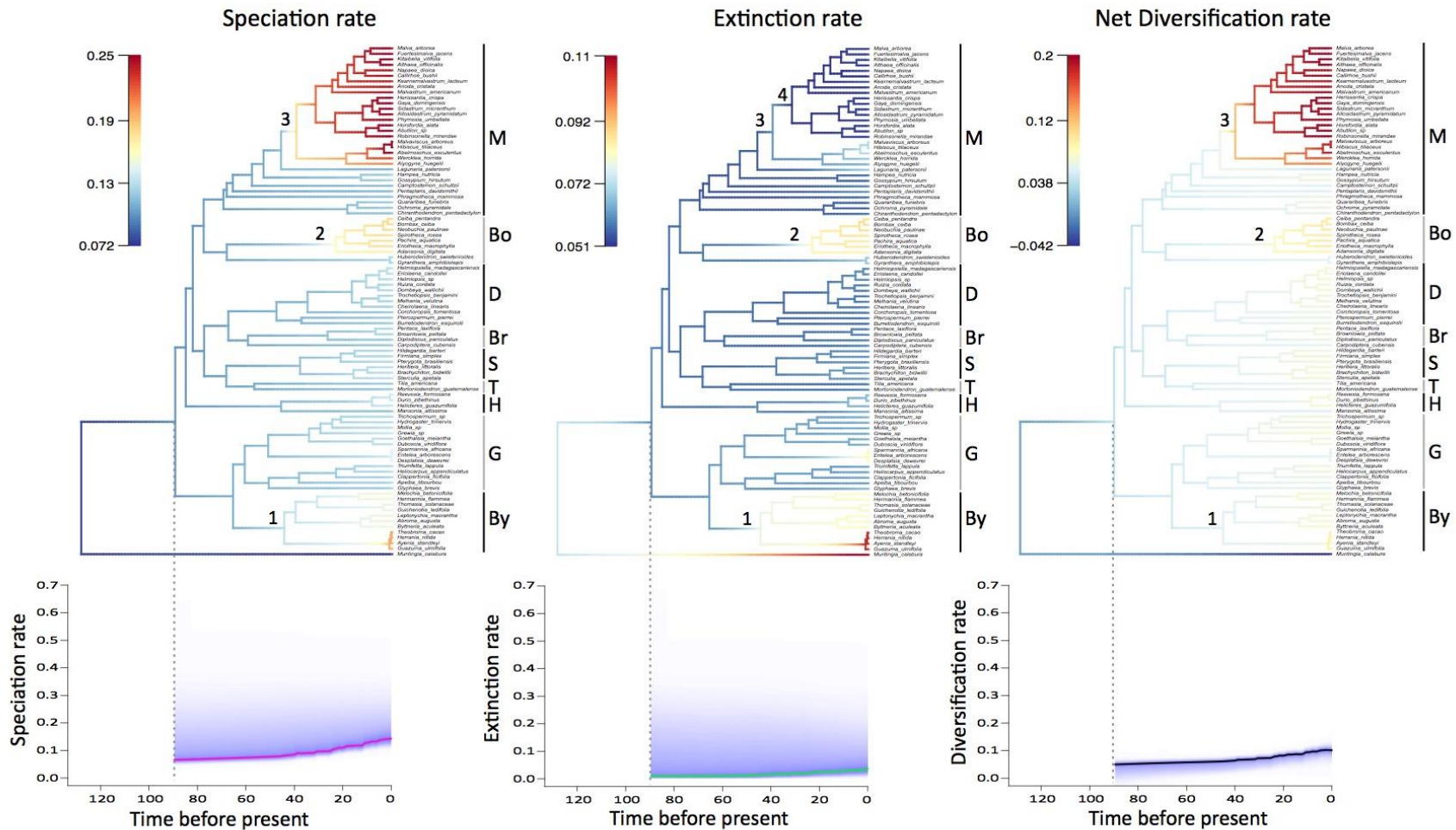


Figure 5. Phylorate of speciation, extinction and net diversification rates estimated with BAMM. Color scale is the estimated rate along the branches. 1: Byttnerioideae (abrupt increase in the tribe Theobromeae); 2: Bombacoideae; 3: tribes Malveae and Hibisce; 4: tribe Malveae. Bottom panels correspond to the rates of Malvaceae s.l. through time. M: Malvoideae; Bo: Bombacoideae; D: Dombeyoideae; Br: Brownlowioideae; S: Sterculioideae; T: Tilioideae; H: Helicterioideae; G: Grewioideae; By: Byttnerioideae.

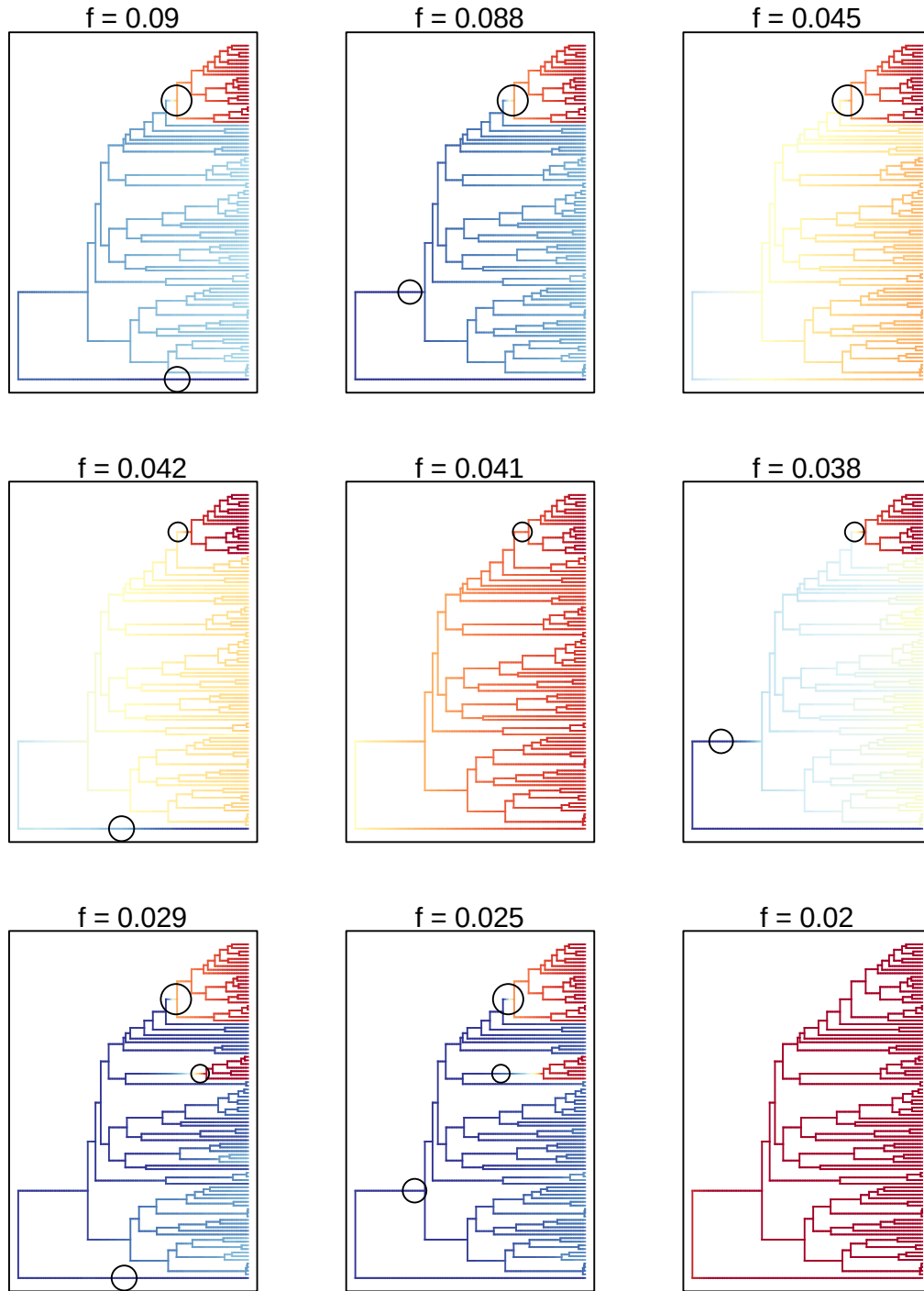


Figure 6. Speciation rate. Pannels represent the configurations obtained within the 95% credibility sample from the posterior density. Gray circles depict the place where the shift was detected and the size of the circle is the posterior probability of that shift occuring in the posterior density of each configuration. f : frequency.

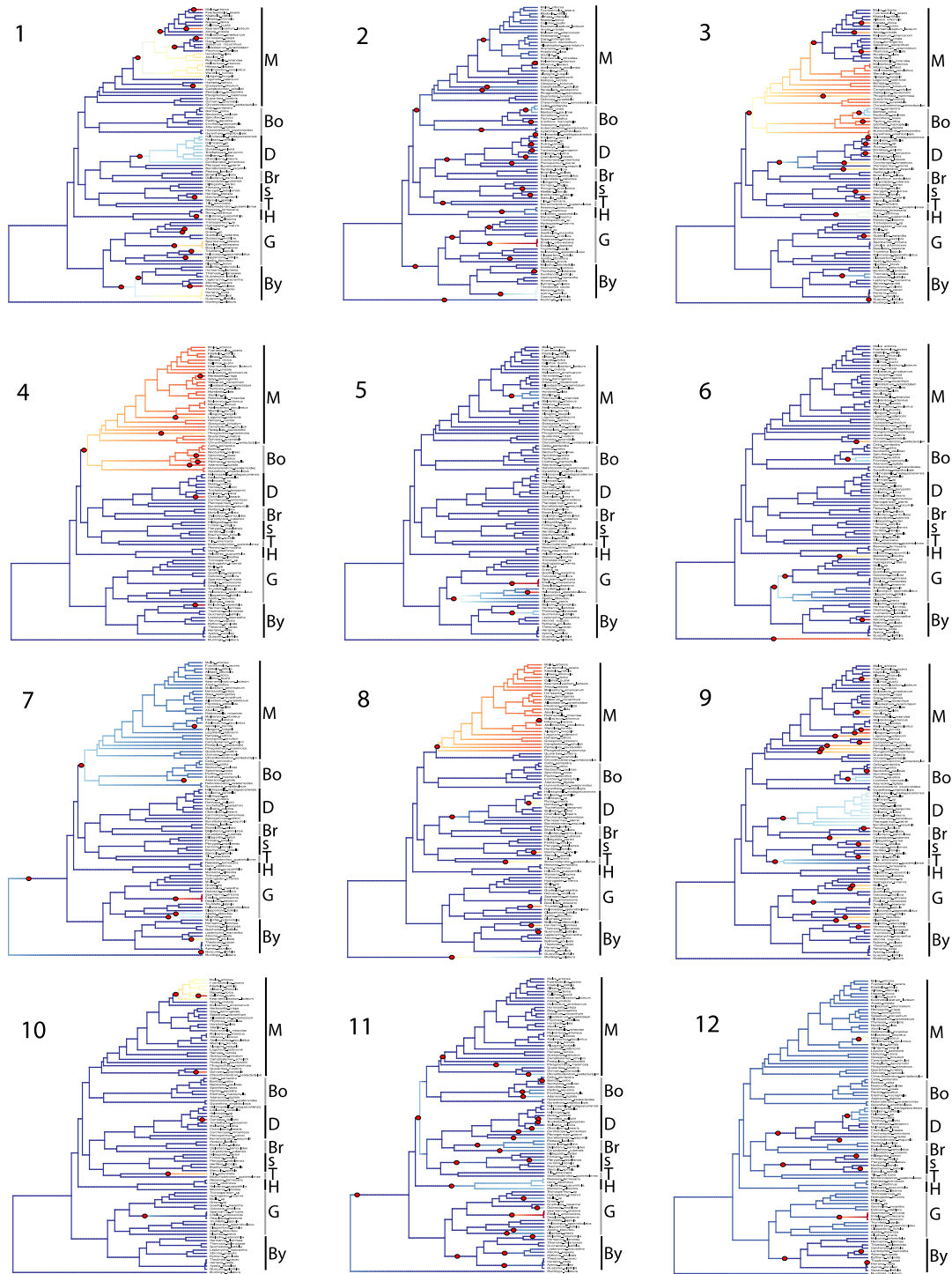


Figure 7. Maximum *a posteriori* rate configuration for each character. Red circles indicate the shifts in morphological evolutionary rate. Warm colors represent increases in the – morphological— rate. 1: habit; 2: floral organization; 3: epicalyx; 4: fusion perianth-androecium;

5: nectary; 6: nectary location; 7: androecial primordia; 8: androecial organization; 9: androecial whorl differentiation; 10: position androecium-perianth; 11: presentation of androecium; 12: total number of AU. M: Malvoideae; Bo: Bombacoideae; D: Dombeyoideae; Br: Brownlowioideae; S: Sterculioideae; T: Tilioideae; H: Helicterioideae; G: Grewioideae; By: Byttnerioideae.

DISCUSSION

Phylogenetic relationships— Previous results on the relationships among the families, especially those from Conover et al. (2018), show conflict given the different molecular information used to estimate relationships, particularly in the Helicterioideae, Sterculioideae and Tilioideae (Nyffeler et al., 2005). In turn, our results are robust and resolve the position of the subfamilies with high support (Fig. 1). We detect that the conflict is mostly in Helicterioideae and in particular given to the inclusion of the durian (*Durio zibethinus*). As discussed by Conover et al. (2018), this may be caused by incomplete lineage sorting in the past of this subfamily, and potentially also due to hybridization.

Within Byttnerioideae relationships are similar to previous hypotheses (Withlock et al., 2001), however, in the BEAST analysis *Ayenia* is placed as sister to *Theobroma* and *Herrania* but in the ML analysis to *Byttneria* as in Withlock et al. (2006). For Grewioideae, we recovered the *Apeiba*-clade as in Brunken & Muellner (2012) with ML but there is conflict with the BEAST analysis where *Sparmannia* and *Entelea* because they are placed with *Desplatsia* in the *Grewia*-clade. In Sterculioideae, relationships are congruent with the previous study on the systematics of the group (Wilkie et al. 2006), except for *Brachychiton* appears related to *Sterculia* and *Heritiera*, whereas in the previous publication appeared as an early diverging lineage within the subfamily. While *Pterygota* and *Basiloxylon* are considered as synonymous (Colli-Silva & Pirani 2019), they do not appear as sister taxa in our analyses, where instead, *Basiloxylon* is related to *Sterculia*, and *Pterygota* is more closely related to *Hildegardia* and *Firmiana*. This may need further research. For Dombeyoideae there is no previously published phylogeny for the entire subfamily, however, our phylogenetic hypotheses are congruent with previous studies regarding *Dombeya* and allied genera (e.g. Skema 2012 Le Péchon et al. 2015).

There are no systematic studies including all genera of Helicterioideae, Tilioideae and Brownlowioideae. For Helicterioideae, relationships obtained with BEAST and ML are different, the main difference is that in the latter analysis the group is monophyletic and in the former *Durio* is sister to a large clade formed by Tilioideae, Sterculioideae, Brownlowioideae and Dombeyoideae. This is particularly interesting since recently it was proposed that a lineage in Helicterioideae hybridize with a member of Malvoideae or Malvoideae + Bombacoideae and it was observed with the durian genome (Conover et al. 2018). This family needs further research regarding relationships and floral morphology, as well as Brownlowioideae, for which phylogenetic relationships and floral morphology are barely known. In contrast, Tilioideae, a family comprising three genera of the North Hemisphere, is well studied in terms of anatomy and morphology in general, and particularly its rich fossil record, but there is no study regarding its systematics and its evolution in general.

Divergence time estimation—We estimated the origin of the Malvaceae *s.l.* (stem node) at 128.3 Ma (100.2-170.6), in the Lower Cretaceous, which is considerably older than the earliest fossil occurrences around the Campanian (83.6-72.1 Ma) in the Upper Cretaceous, for example *Bombacoxylon langstoni* (Wheeler & Lehman, 2000), the fossil we used to calibrate the crown node of the family. Our taxon sampling was designed to estimate the crown node of Malvaceae *s.l.*, which our results indicate was at 89.6 Ma (83.7-963), during the Coniacian (Upper Cretaceous), which is congruent with the Fossilized Birth-Death dating results of the order Malvales from Hernández-Gutiérrez & Magallón (2019). All subfamilies originate at the end of the Upper Cretaceous (83.6-66 Ma), except Byttnerioideae and Grewioideae, which originate at the beginning of the Paleogene, in the early Paleocene (66-61.6 Ma). Bombacoideae and Malvoideae diversified in the Maastrichtian (72.1-66 Ma), Brownlowioideae, Grewioideae, and Tilioideae in the Paleocene (66-56 Ma), and Byttnerioideae, Dombeyoideae, Helicterioideae, and Sterculioideae, from the early to middle Eocene (56-37.8 Ma).

In general, the estimates in this study are younger than those estimated in a previous study focused on Malvales (Hernández-Gutiérrez & Magallón 2019), possibly due to the higher taxon sampling in that study and the incorporation of fossils in the phylogeny. Other previous

study of malvacean lineages is the timing of genus *Theobroma* by Richardson et al. (2015), where they also estimated the ages and diversification rates of the subfamilies. The age of the subfamilies in that study are considerably younger than in our study, probably due to the young secondary calibration point they applied to the root of the tree and that an additional calibration point was applied to the malvoids.

The diversification of Malvaceae *s.l.* took place after the Cretaceous Thermal Maximum (KTM), but the hot greenhouse period still characterized the global temperatures until the end of the Cretaceous, in the Maastrichtian (Huber et al., 2018), when the temperature cooled gradually maintaining a warm period during the Paleocene. Soil evidence indicates tropical climates in the equator, and spread to Southern, but mostly Northern latitudes throughout the Cretaceous (Boucot et al. 2013). It has been proposed that the origin of the tropical rainforest took place in the mid-Cretaceous, but in a wider range than its present equatorial limits, specifically in Northern latitudes (Couvreur & Baker, 2013). This is consistent with the fossil record of Malvaceae *s.l.*, given that some of the oldest fossils have been found in North America (Graham, 1999). Thus, temperature, soil, and fossil record evidence are indicating that probably tropical rainforest, or a similar biome, was established by the end of the Cretaceous, and possibly the Malvaceae *s.l.* diversified, and its subfamilies originated, in this type of vegetation.

Between the Paleocene and Eocene temperatures abruptly increased (Paleocene/Eocene Thermal Maximum [PETM]) and decreased progressively through the Eocene. This period of time was additionally marked by a widening in the arid area north to the equator (Boucot et al. 2013), probably enhancing the diversification of groups with members adapted to seasonal environments (e.g., Helicterioideae, Byttnerioideae). Marked seasonality started later at the boundary of Eocene-Oligocene (ca. 33.5 Ma; Eldrett et al. 2009), a time when extant lineages of all subfamilies were already present.

Ancestral state reconstruction— The ancestral state of the flower of Malvaceae *s.l.* had the same probability of being solitary or arranged in panicles, and did not have an epicalyx. According to Bayer (1999), the epicalyx evolved from the reduction of a cymoid inflorescence and different degrees of this organization are present in the family. He named this floral

organization a “bicolor unit”, which is observed in various species (e.g. *Theobroma bicolor*), but it would be observed in different extents among the malvacean species. We evaluated the evolution of the inflorescences and epicalyx, in an attempt to reconstruct the ancestral state of the bicolor unit, however, given that the most probable states of the ancestral flower of Malvaceae s.l. was solitary or in a panicle, our results do not have enough information to confirm the origin of the epicalyx. Nevertheless, given that the ancestral state was absence of the epicalyx and a type of inflorescence (i.e., panicle), this may indicate that the origin of the epicalyx started in an inflorescence as hypothesized by Bayer (1999).

The ancestral state of the androecium was 2-whorled, with the inner whorl sterile, the outer opposite to the inner perianth whorl, and androecial organs developing from common primordia (Fig. 4A). At anthesis, the presentation or architecture of the androecium probably was in a tube with lateral stalked AU (Fig. 4B), which deviated from this architecture inside each subfamily in many different presentations. In extant species, there are no representatives with the combination of two whorls, the inner sterile, and a tube with lateral stalked AU (van Heel 1966). In Dombeyoideae and Byttnerioideae, in early stages of development some species have two whorls and at anthesis these fused forming a staminal tube (van Heel 1966); however, they usually have stalked terminal AU. In Malvoideae –where the presentation of staminal tube with stalked lateral AU is the most common– vascular bundles are lost in sterile parts of the staminal tube, whereas the in Bombacoideae there are sterile projections of the staminal tube vascularized in an alternate position, possibly indicating the ancestral condition of two whorls (von Balthazar et al. 2004, 2006).

Noticeably, the ancestral state of most of the subfamilies of Malvaceae s.l. shared most of the states of the family as a whole, meaning that the diversity of morphologies was gained after the diversification of the subfamilies. Could this mean that the early lineages within the subfamilies were similar among them? According to the origin of tropical biomes, probably the oldest is the rainforest (Couvreur & Baker, 2013), and probably the most recent common ancestor of each subfamily inhabited this biome or a similar one, based on the ages estimated in our study and previously (Hernández-Gutiérrez & Magallón, 2019). It might be a possibility that the MRCA of each subfamily inhabited a rainforest at the early stages of their diversification,

and shared a similar floral morphology, based on our reconstructions, possibly sharing a similar pollination syndrome.

Rate heterogeneity— The causes of the estimated changes in diversification rate may be numerous. According to our results, morphological heterogeneity is high in malvoids, especially in Malveae + Hibisceae, and in the DES clade, which indicates contrasting results, possibly derived from different processes; in the former case speciation abruptly increased, and in the latter case the extinction increased. This means that for some groups, morphological change might be associated with species diversification and in some others morphological change might limit the diversification, probably due to a high specialization that reduces the opportunity to diversify. This latter pattern may correspond to a case where “confluence”, the combination of traits, influence the diversification rates to a decrease, hence the DES clade may be a “depauperon” (Donoghue & Sanderson, 2015) derived from the high morphological evolutionary rate. With these results as a background it would be possible to evaluate explicitly the effect on diversification rate of morphological specialization against morphological lability of those characters with highest morphological evolutionary rate.

In core Bombacoideae –the group of the baobab– which is characterized by brush-like flowers, the morphology of flower changed in the multiplication of AUs, in addition to complex androecial primordium. All of them have numerous AUs but some of them have fascicles (e.g., *Pachira*) and others have clusters (e.g. *Bombax*). The presentation of the androecium in this group shifted three times, and this lability could possibly be associated with the estimated high turnover rate (i.e., high speciation and extinction rates; Donoghue & Sanderson, 2015). The high extinction rate of the DES clade may be caused by the high morphological evolutionary rate, which maybe correlated with high morphological specialization, which may reduce the opportunities for reproduction. Other possibility is that, given that these genera are relatively young (0.35 Ma; Fig. 3), maybe there has been little time to diversify, compared to other clades, and thus the model estimated high extinction. In the case of Malvoideae, floral morphology may be associated to the evolution of the group, since many shifts occurred within it, but also the stasis of other characters represent high stability and maintenance of a combination of

characters that are conserved throughout the group, similar to the relationship between diversification and morphological stasis found in Malpighiaceae (Davis et al., 2014).

Interestingly, diversification rates were also found to increase in Malveae (Areces-Berazain & Ackerman 2017) and in *Theobroma–Herrania* (Richardson et al. 2015), although our age estimations vary substantially from those studies. In Malveae, Areces-Berazain & Ackerman associated the increase in diversification rate with fruit evolution, whereas in *Theobroma–Herrania*, Richardson et al. associated it with the uplift of the Andes. We obtained diversification rates for each subfamily –and sub-groups— and postulate the androecium as a facilitator of the diversification. Our results are a first evidence that diversification and morphological rates are heterogeneous within the family, which are is background for more studies to directly evaluate the contribution of each factor on the diversification rate for each case.

The heterogeneity in morphological evolutionary rate is high compared to the heterogeneity in diversification rate, meaning that probably the lability in morphological traits could maintain the background diversification rate, which is higher than the outgroup. We detected a pattern that integrates ancestral character states, morphological evolution rate shifts, and diversification. This pattern consists of characters that transitioned from the ancestral state, and have a few or no shifts in the diverging lineages (e.g. subfamilies) but the morphological evolutionary rate is high (i.e., high heterogeneity observed by the presence of several rate shifts), and the diversification rate of the groups is also high. This is the case of the presence of epicalyx, fusion of perianth to the androecium, and androecial primordia organization in Malvatheca; and the androecium organization in Eumalvoideae. Possibly this is also the case in Dombeyoideae for androecial whorl differentiation, although net diversification rate did not experience a shift. This pattern coincides with the developmental robustness hypothesis of Melzer & Theißen (2016), according to which the species richness in some groups may be due to the “robustness” in developmental programs. For example, the organization of the perianth in angiosperms (i.e., the presence of calyx and corolla) may be conserved among flowering plants given strong developmental constraints and this may give opportunity to other character states or innovations are gained (e.g., fused corolla). Additionally, the pattern is consistent with the hypothesis of the “dynamic characters” of Wiens (2017), in which the

evolutionary rate of a character is correlated with the diversification and not the presence of the character itself. However, in some cases morphological stasis may be important for the speciation of a group (Davis et al., 2014), for example, the androecium presentation in malvoids has a shift at the crown group and did not change after that.

CONCLUSIONS

The measuring of rate heterogeneity is a first step towards understanding the dynamics of the trait-dependent diversification. Here, we estimated how the diversification and morphological rate changed through time in *Malvaceae s.l.* We found that morphological rate heterogeneity is considerably higher than in diversification rate, probably meaning that shifts in morphology contribute to the maintenance of background diversification, and only a few characters correspond to places where shifts in diversification rate are observed (e.g. fusion of perianth to androecium, androecial primordia, and androecial organization).

Heterogeneity patterns are different among the subfamilies of *Malvaceae s.l.*, but probably this heterogeneity took place after the diversification of subfamilies, given that the ancestral state of the subfamilies in most of the characters resembled that of the ancestral state of the family as a whole. Floral characters may be relevant for the diversification of certain lineages, for example *Malvoideae*, *Bombacoideae* and *Byttnerioideae*, but direct testing is needed to corroborate this. More importantly, a specific hypothesis on how some floral characters influence species diversification should be made and directly tested with macroevolutionary methods, specifically by testing the contribution of that particular character on diversification.

ACKNOWLEDGMENTS

Posgrado en Ciencias Biológicas UNAM. PAPIIT UNAM IG200316. CONACyT Fronteras de la Ciencia 1602701-72. RHG received a doctoral scholarship from CONACyT 507682/288658. To develop this research RHG received the Graduate Student Award from the American Society of Plant Taxonomist (ASPT), the IAPT (International Association for Plant Taxonomy) Research Grant, and the Elizabeth Bascom Fellowship from the Missouri Botanical Garden. RHG thanks

the following institutions: Missouri Botanical Garden, Royal Botanical Gardens KEW, Royal Botanic Garden Edinburgh, and Meise Botanical garden, Herbario Nacional de Mexico (MEXU), Herbario Nacional del Ecuador (QCNE), and Universidade Estadual de Feira de Santana. RHG thanks the following people for substantial help during the development of this research or sample collection: Alfonso Delgado, Jürg Schönenberger, Alex Popovkin, Mario Ishiki, Gerardo Salazar, Itzi Fragoso-Martínez, Carolina Granados Mendoza, Marcia Peñafiel, Efraín Freire.

CITED LITERATURE

- Alverson WS, Whitlock BA, Nyffeler R, Bayer C, Baum DA. 1999. Phylogeny of the core Malvales: Evidence from *ndhF* sequence data. *American Journal of Botany* 86(10): 1474-1486.
- Areces-Berazain F & Ackerman JD. 2017. Diversification and fruit evolution in eumalvoids. *Botanical Journal of the Linnean Society* 184: 401-417.
- Armbruster W.S. 2017. The specialization continuum in pollination systems: diversity of concepts and implications for ecology, evolution and conservation. *Functional Ecology* 31: 88-100.
- Bacon CD, Velázquez-Puentes FJ, Felipe Hinojosa L, Schwartz T, Oxelman B, Pfeil B, Arroyo MTK, Wanntorp L, Antonelli A. 2018. Evolutionary persistence in Gunnera and the contribution of southern plant groups to the tropical Andes biodiversity hotspot. *PeerJ* 6:e4388; DOI 10.7717/peerj.4388.
- Baum D, Smith S, Yen A, Alverson W, Nyffeler R, Whitlock B, Oldham R. 2004. Phylogenetic relationships of Malvatheca (Bombacoideae and Malvoideae; Malvaceae sensu lato) as inferred from plastid DNA sequences. *American Journal of Botany* 91(11): 1863-1871.
- Bayer C. 1999. The bicolor unit –homology and transformation of an inflorescence structure unique to core Malvales. *Plant Systematics and Evolution* 214: 187-198.
- Bayer C, Fay MF, De Bruijn AY, Savolainen V, Morton CM, Kubitzki K, Alverson WS, Chase MW. 1999. Support for an expanded family concept of Malvaceae within a recircumscribed order Malvales: a combined analysis of plastid *atpB* and *rbcl* DNA sequences. *Botanical Journal of the Linnean Society* 129:267-303.

- Bayer C & Kubitzki K. 2003. Malvaceae. In Kubitzki K & Bayer C. Flowering plants dicotyledons. Capparales, Malvales and Non-betalain Caryophyllales. Springer.
- Beaulieu JM, O'Meara BC, Donoghue MJ. 2013. Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. *Systematic Biology* 62: 725-737.
- Beaulieu JM & O'Meara BC. 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Systematic Biology* 65(4): 583-601.
- Boucot AJ, Chen X, Scotese CR. 2013. Phanerozoic paleoclimate: an atlas of lithologic indicators of climate. Society for Sedimentary Geology.
- Brunken U. & Muellner AN. 2012. A new tribal classification of Grewioideae (Malvaceae) based on morphological and molecular phylogenetic evidence. *Systematic Botany* 37(3): 699-711.
- Buddenhagen C, Lemmon AR, Lemmon EM, Bruhl J, Cappa J, Clement WL, Donoghue M, Edwards EJ, Hipp AL, Kortyna M, Mitchell N, Moore A, Prychid DJ, Segovia-Salvedo MC, Simmons MP, Soltis PS, Wanke S, Mast A. 2016. Anchored phylogenomics of angiosperms I: Assessing the robustness of phylogenetic estimates. *bioRxiv* <http://dx.doi.org/10.1101/086298>.
- Chartier M, Löfstrand S, von Balthazar M, Gerber S, Jabbour F, Sauquet H, Schönenberger J. 2017. How (much) do flower vary? Unbalanced disparity among flower functional modules and a mosaic pattern of morphospace occupation in the order Ericales. *Proceeding of the Royal Society B* 284: 20170066.
- Colli-Silva M & Pirani JR. 2019. Notes on the circumscription of Pterygota (Malvaceae, Sterculioideae) with special reference to a Brazilian taxon with a misapplied name. *Phytotaxa* 415(2): 113-115.
- Conover JL, Karimi N, Stenz N, Ané C, Grover CE, Skema C, Tate JA, Wolff K, Logan SA, Wendel JF, Baum DA. 2018. A Malvaceae mystery: A mallow maelstrom of genome multiplications and maybe misleading methods? *Journal of Integrative Plant Biology* 61(1): 12-31.
- Couvreur TLP & Baker WJ. 2013. Tropical rain forest evolution: palms as a model group. *BMC Biology* 11: 48.

- Davis CC, Schaefer H, Xi Z, Baum DA, Donoghue MJ, Harmon LJ. 2014. Long-term morphological stasis maintained by a plant-pollinator mutualism. *Proceedings of the National Academy of Sciences* 111(16): 5914-5919.
- Dellinger AS, Artuso S, Pamperl S, Michelangeli FA, Penneys DS, Fernández-Fernández DM, Alvear M, Almeda F, Armbruster WS, Staeder Y, Schönenberger J. 2019. Modularity increases rate of floral evolution and adaptive success for functionally specialized pollination systems. *Nature Communications Biology* 2: 453.
- Donoghue MJ & Sanderson MJ. 2015. Confluence, synnovation, and depauperons in plant diversification. *New Phytologist* 207: 260-274.
- Duarte M. C., Esteves G. L., Salatino M. L., Walsh K. C., Baum D.A. 2011. Phylogenetic Analyses of *Eriotheca* and Related Genera (Bombacoideae, Malvaceae). *Systematic Botany* 36(3): 690-701.
- Eldrett JS, Greenwood DR, Harding IC, Huber M. 2009. Increased seasonality through the Eocene to Oligocene transition in northern high latitudes. *Nature Letters* 459: 969-974.
- Graham A. 1999. Late Cretaceous and Cenozoic history of North American vegetation. Oxford University Press, Oxford.
- Heath TA, Huelsenbeck JP, Stadler T. 2014. The fossilized birth-death process for coherent calibration of divergence-time estimates. *PNAS* E2957-E2966.
- Hernández-Gutiérrez R. y Magallón S. The timing of Malvales evolution: incorporating its extensive fossil record to inform about lineage diversification. *Molecular Phylogenetics and Evolution* 140: 106606.
- Huelsenbeck JP, Nielsen R, Bollback JP. 2003. Stochastic mapping of morphological characters. *Systematic Biology* 52(2): 131-158.
- Katoh K & Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772-780.
- Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2016. PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution* 34(3): 772-773.

- Lemmon AR, Emme SA, Moriarty Lemmon E. 2012. Anchored Hybrid Enrichment for massively high-throughput phylogenomics. *Systematic Biology* 61(5): 727-744.
- Le Péchon P. y Gigord LDB. 2014. On the Relevance of Molecular Tools for Taxonomic Revision in Malvales, Malvaceae *s.l.*, and Dombeyoideae. Chapt. 7. En *Methods in molecular Biology*. Ed. Besse, P. Springer, New York.
- Le Péchon T, Dai Q, Zhang L-B, Gao X-F, Sauquet H. 2015. Diversification of Dombeyoideae (Malvaceae) in the Mascarenes: old taxa on young islands? *International Journal of Plant Sciences* 176(3): 211-221.
- Maddison WP & FitzJohn RG. 2015. The unsolved challenge to phylogenetic correlation tests for categorical characters. *Systematic Biology* 64(1):127-136.
- Maddison WP, Midford PE, Otto SP. 2007. Estimating a binary character's effect on speciation and extinction. *Systematic Biology* 56(5): 701-710.
- Magallón S, Sánchez-Reyes LL, Gómez-Acevedo S. 2018. Thirty clues to the exceptional diversification of flowering plants. *Annals of Botany*, mcy182
<http://doi.org/10.1093/aob/mcy182>.
- Melzer R. y Theißen G. 2016. The significance of developmental robustness for species diversity. *Annals of Botany* 117(5): 725-732.
- Miller MA, 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)*, 14 Nov. 2010, New Orleans, LA pp 1 - 8.
- Morlon H. 2014. Phylogenetic approaches for studying diversification. *Ecology Letters* 17: 508-525.
- Nyffeler, R., Bayer, C., Alverson, W.S., Yen, A., Whitlock, B.A., Chase, M.K., Baum, D.A. 2005. Phylogenetic analysis of the malvadendrina clade (Malvaceae *s.l.*) based on plastid DNA sequences. *Organisms, Diversity & Evolution* 5: 109-123.
- O'Meara B, Smith SD, Armbruster WS, Lawrence DH, Hardy CR, Hileman LC, Hufford L, Litt A, Magallón S, Smith SA, Stevens PF, Fenster CB, Diggle PK. 2016. Non-equilibrium dynamics and floral trait interactions shape extant angiosperm diversity. *Proceedings of the Royal Society B* 283: 20152304.

- Onstein RE. 2019. Darwin's second 'abominable mystery': trait flexibility as the innovation leading to angiosperm diversity. *New Phytologist* doi:10.1111/nph.16294.
- Plummer M, Best N, Cowles K, Vines K. 2006. CODA: Convergence diagnosis and output analysis for MCMC. *R News* 1: 7-11.
- Sauquet H. 2016. PROTEUS: a database for recording morphological data and creating NEXUS matrices. Version 1.26. <http://eflower.myspecies.info/proteus>.
- Rabosky DL, Donnellan SC, Grundler M, Lovetter IJ. 2014. Analysis and visualization of complex macroevolutionary dynamics: an example from Australian scincid lizards. *Systematic Biology* 63: 610-627.
- Rabosky DL, Grundler MC, Anderson CJ, Title PO, Shi JJ, Brown JW, Huang H, Larson JG. 2014b. BAMMtools: An R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution* 5: 701-707.
- Rabosky DL. 2009. Ecological limits and diversification rate: alternative paradigms to explain the variation in species richness among clades and regions. *Ecology Letters* 12: 735-743.
- Rabosky DL. 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLOS One* 9: e89543.
- Rambaut A, Suchard MA, Xie D, Drummond AJ. 2014. Tracer v1.6. Available from <http://tree.bio.ed.ac.uk/software/tracer/>
- Revell LJ. 2012. phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3: 217-223.
- Richardson JE, Whitlock BA, Meerow AW, Madriñán S. 2015. The age of chocolate: a diversification history of *Theobroma* and Malvaceae. *Frontiers in Ecology and Evolution* 3: 120. doi: 10.3389/fevo.2015.00120.
- Skema C. 2012. Toward a new circumscription of *Dombeya* (Malvales: Dombeyaceae): A molecular phylogenetic and morphological study of *Dombeya* of Madagascar and a new segregate genus, *Andringitra*. *Taxon* 61(3): 612-628.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9): 1312-1313.
- Sun M, Naeem R, Su J-X, Cao Z-Y, Burleigh G,

- Soltis PS, Soltis DE, Chen Z-D. 2016. Phylogeny of the Rosidae: a dense taxon sampling analysis. *Journal of Systematics and Evolution* 54(4): 363-391.
- The Plant List. 2017. URL <http://www.theplantlist.org/>.
- Tate, J.A., Fuertes Aguilar, J., Wagstaff, S.J., Duke, J.C., Bodo Slotta, T.A., Simpson, B.B. 2005. Phylogenetic relationships within the tribe Malveae (Malvaceae, Subfamily Malvoideae) as inferred from ITS sequence data. *American Journal of Botany* 92(4): 584-602.
- Vamosi JC, Magallón S, Mayrose I, Otto SP, Sauquet H. 2018. Macroevolutionary patterns of flowering plant speciation and extinction. *Annual Review in plant Biology* 69: 9.1-9.22.
- van Heel WA. 1966. Morphology of the androecium in Malvales. *Blumea* 13: 177-394.
- von Balthazar M, Alverson WS, Schönenberger J, Baum DA. 2004. Comparative floral development and androecium structure in Malvoideae (Malvaceae *s.l.*). *International Journal of Plant Sciences* 165(4):445-473.
- von Balthazar M, Schönenberger J, Alverson WS, Janka H, Bayer C, Baum DA. 2006. Structure and evolution of the androecium in the Malvatheca clade (Malvaceae *s.l.*) and implications for Malvaceae and Malvales. *Plant Systematics and Evolution* 260:171-197.
- Wheeler EA, Lehman TM. 2000. Late Cretaceous woody dicots from the Aguja and Javelina Formations, Big Bend National Park, Texas, USA. *IAWA Journal* 21 (1): 83:120.
- Whitlock BA, Bayer C, Baum DA. 2001. Phylogenetic relationships and floral evolution of the Byttnerioideae ("Sterculiaceae" or Malvaceae *s.l.*) Based on sequences of the chloroplast gene *ndhF*. *Systematic Botany* 26(2):420-437.
- Wiens J.J. 2017. What explains patterns of biodiversity across the Tree of Life? *Bioessays* 38: 1600128.
- Wilkie P, Clark A, Pennington RT, Cheek M, Bayer C, Wilcock CC. 2006. Phylogenetic relationships within the subfamily Sterculioideae (Malvaceae /Sterculiaceae Sterculieae) using the chloroplast gene *ndhF*. *Systematic Botany* 31(1): 160-170.
- Zenil-Ferguson R, Burleigh JG, Freyman WA, Igic B, Mayrose I, Goldberg EE. 2019. Interaction between ploidy, breeding system, and lineage diversification. *bioRxiv* <http://dx.doi.org/10.1101/709329>.

CAPÍTULO III: MATERIAL SUPLEMENTARIO

Figures 1-12. Ancestral character reconstruction of the selected characters with stochastic mapping (See Materials and Methods for more information). Pie charts represent the posterior probability of each character state.

Figure 1. Habit

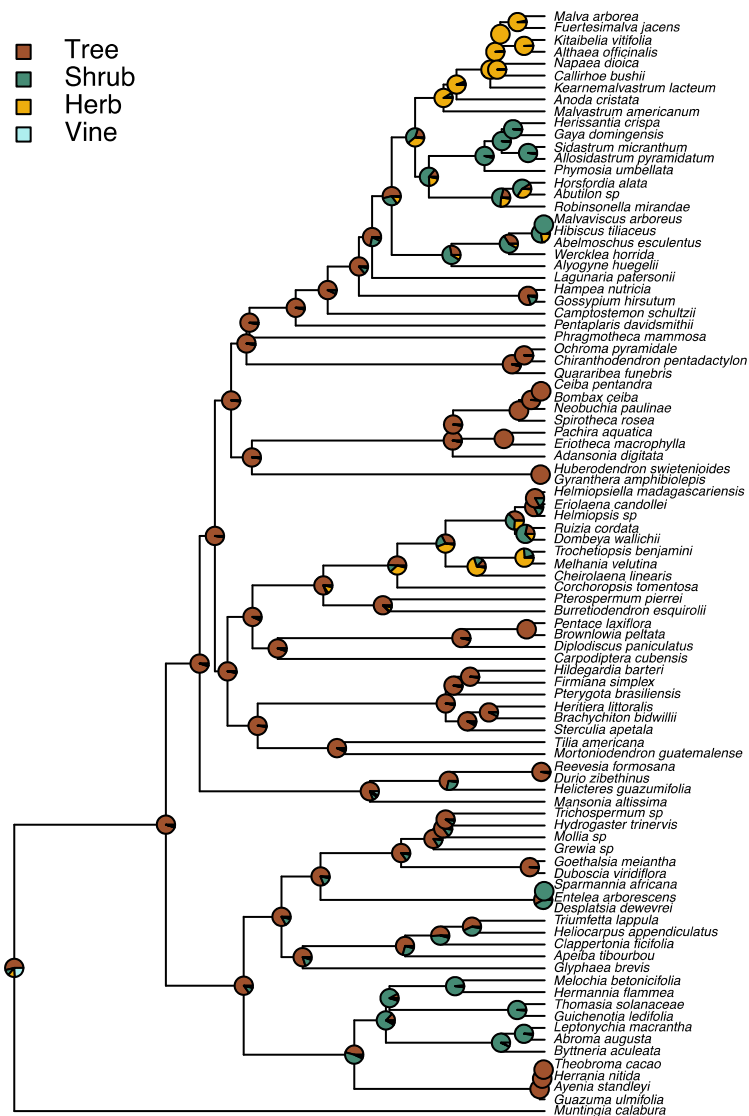


Figure 2. Floral organization

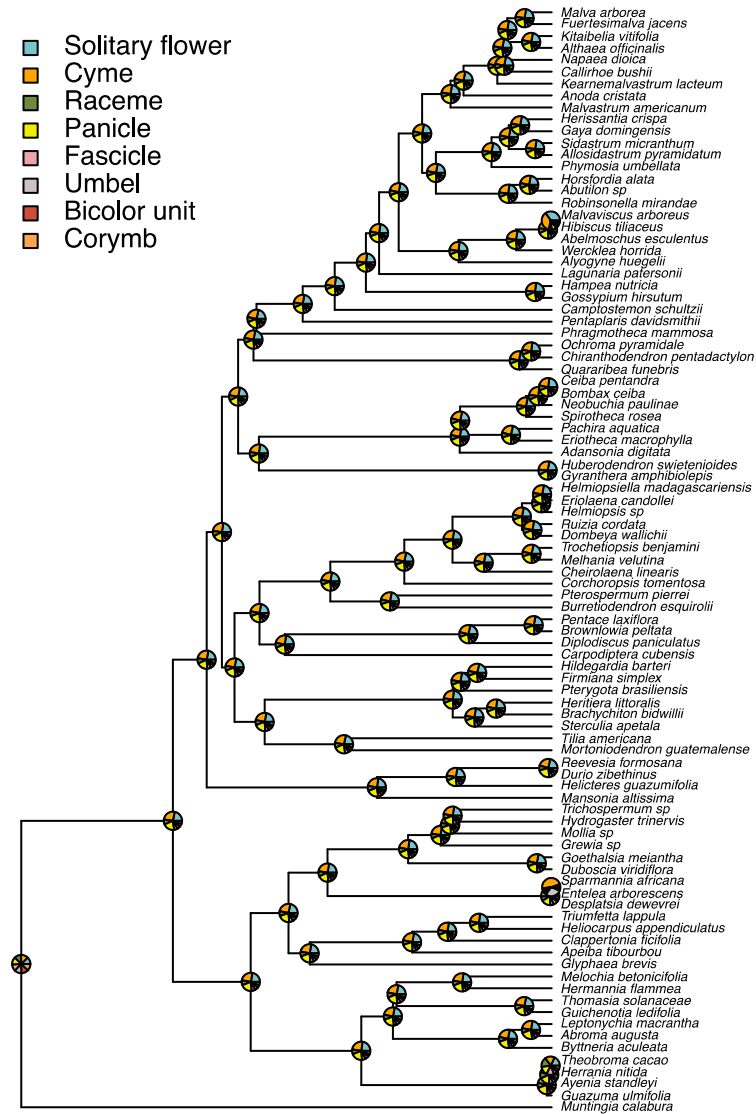


Figure 3. Epicalyx

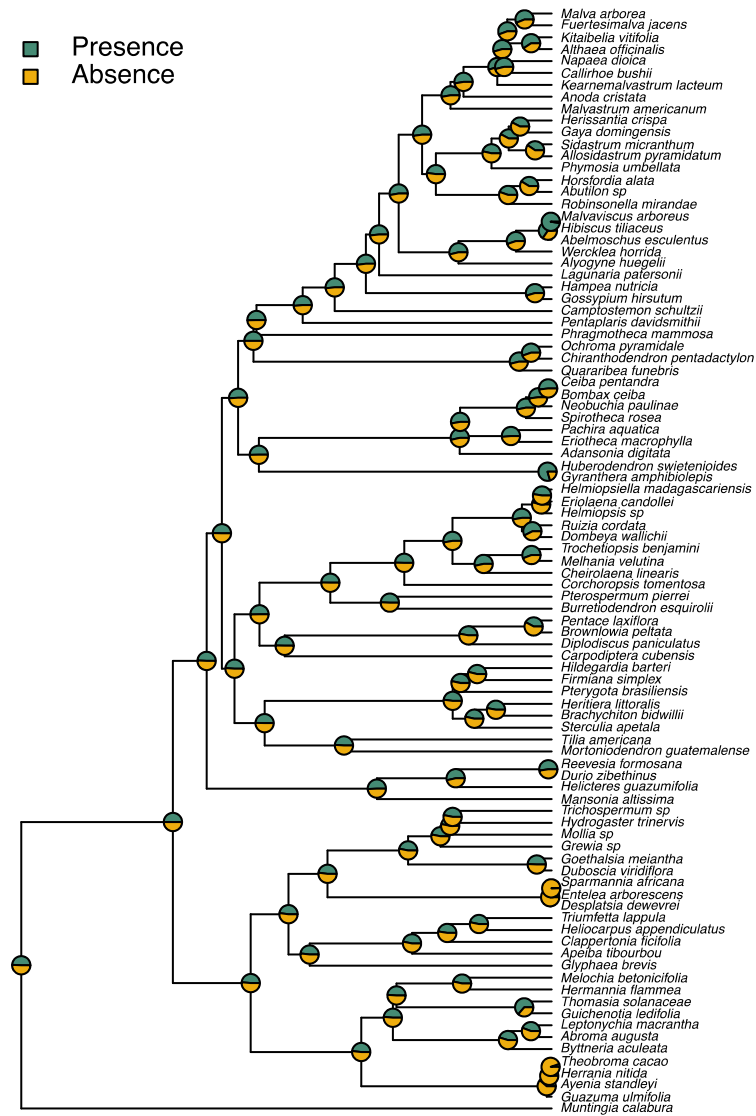


Figure 4. Fusion perianth-androecium

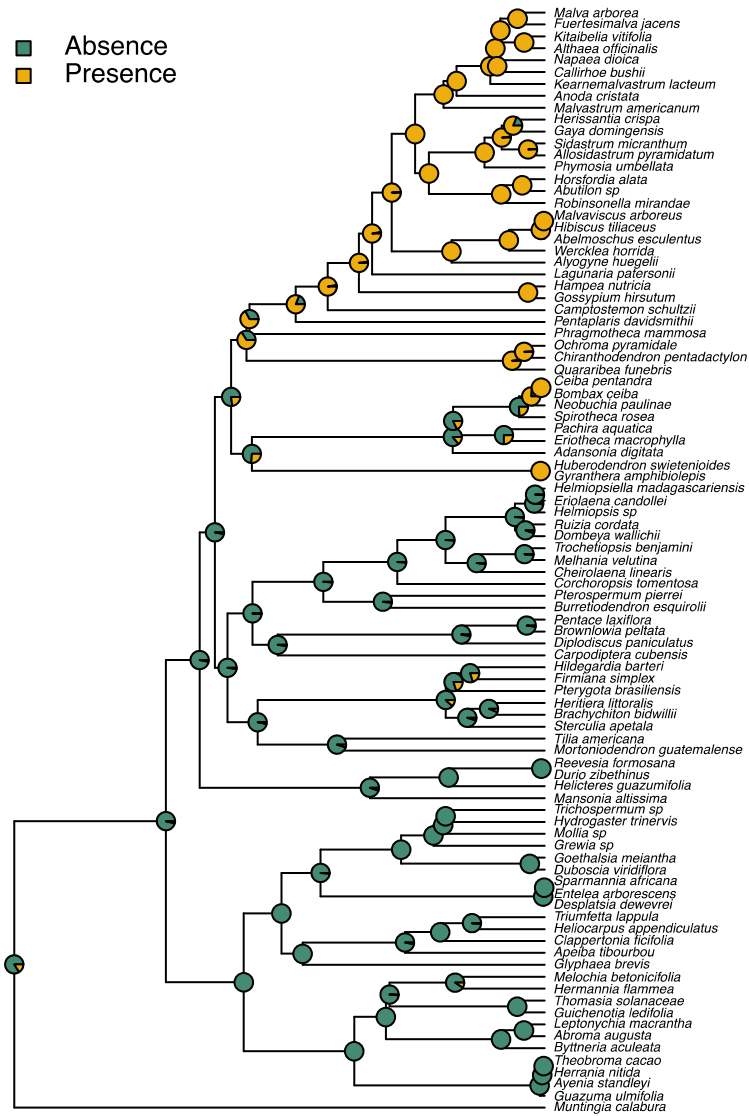


Figure 5. Nectary.

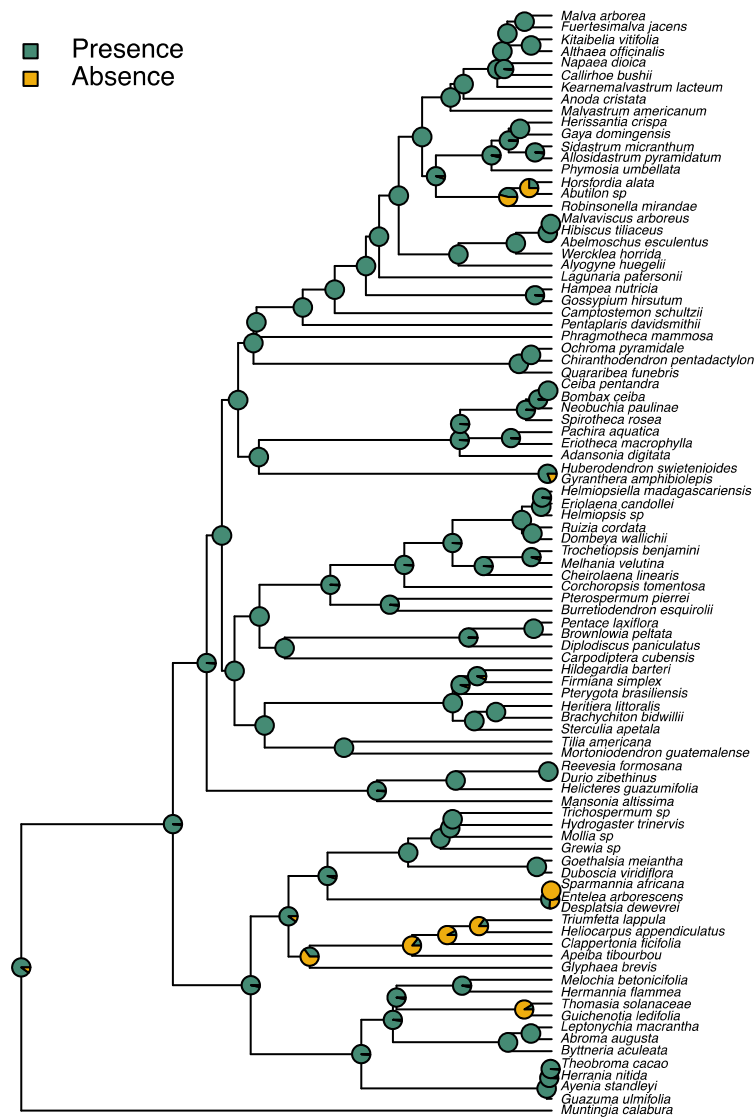


Figure 6. Nectary location.

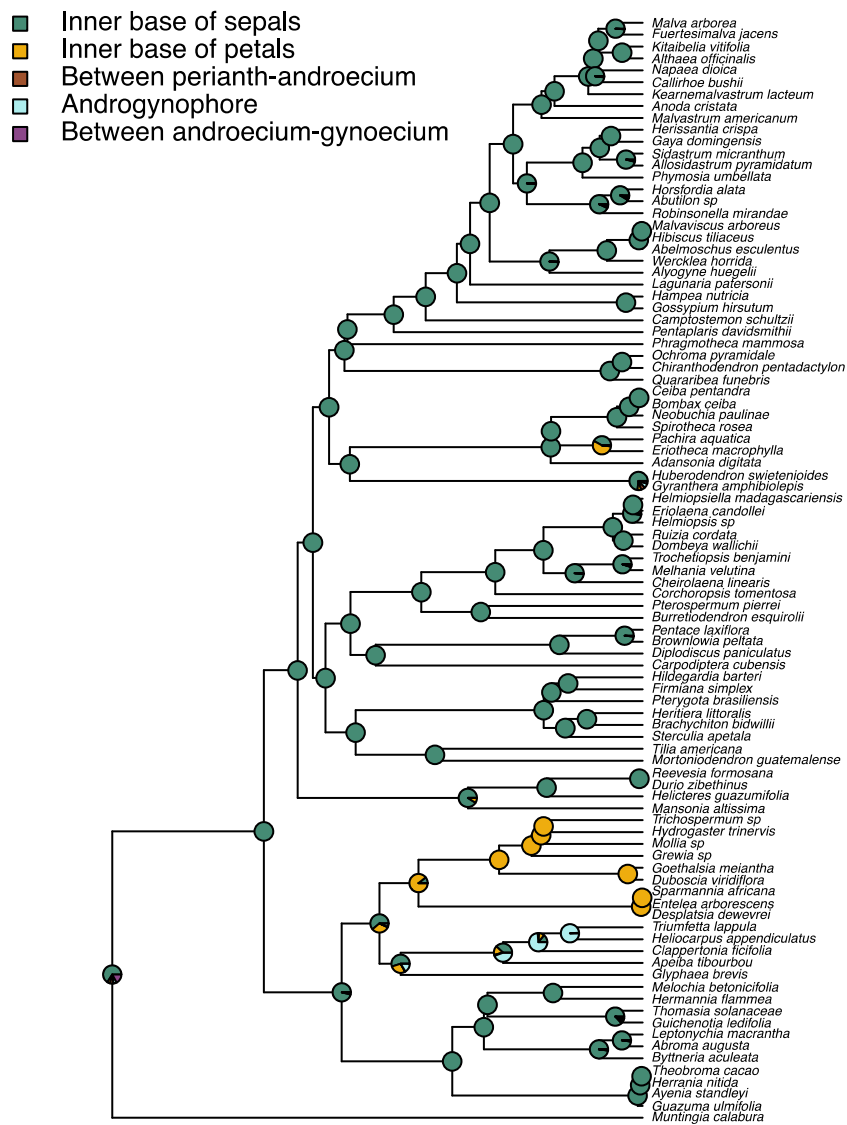


Figure 7. Androecial primordia.

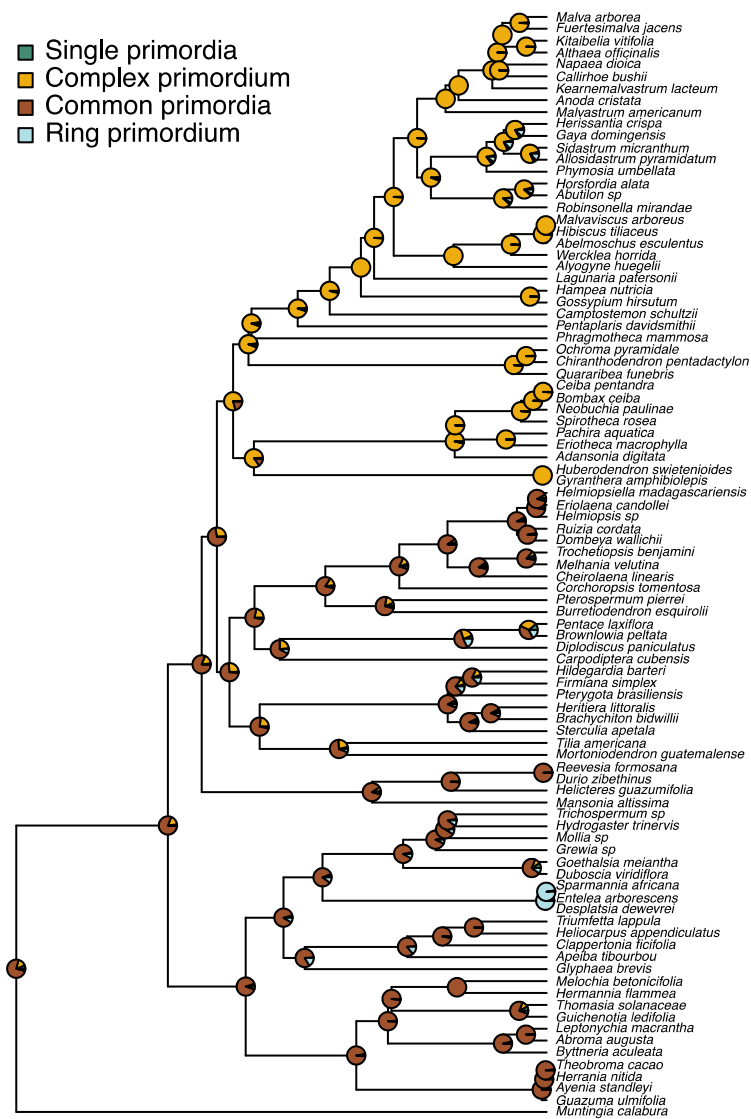


Figure 8. Androecial organization.

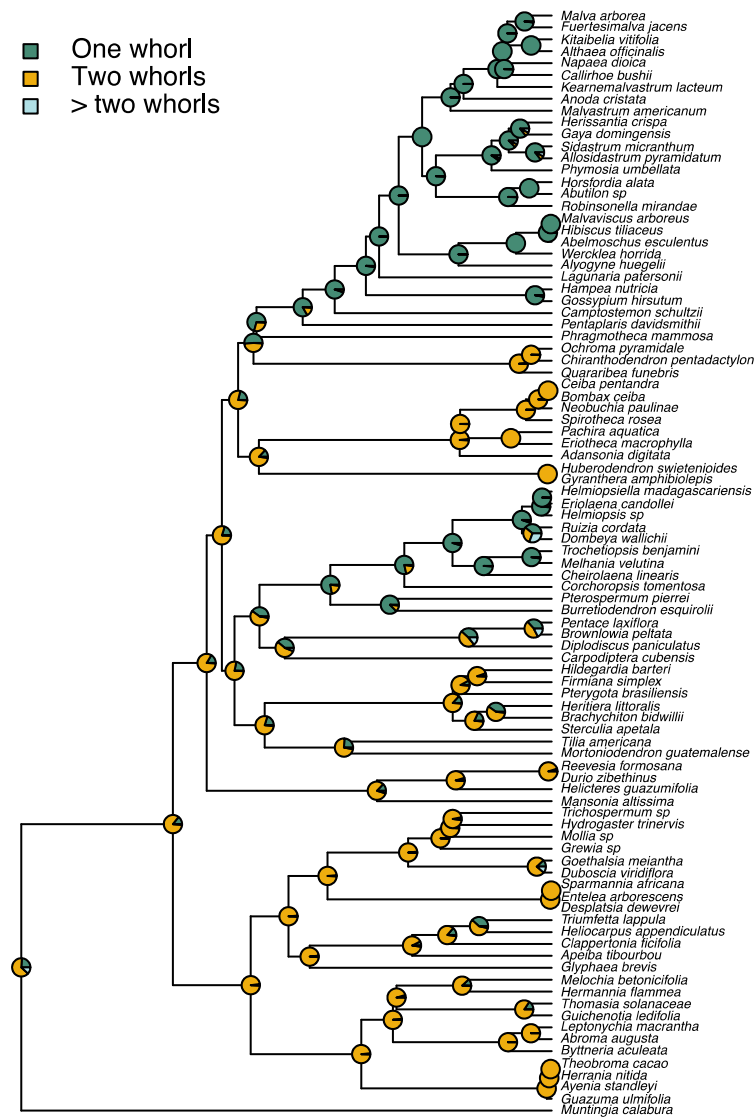


Figure 9. Androecial whorl differentiation.

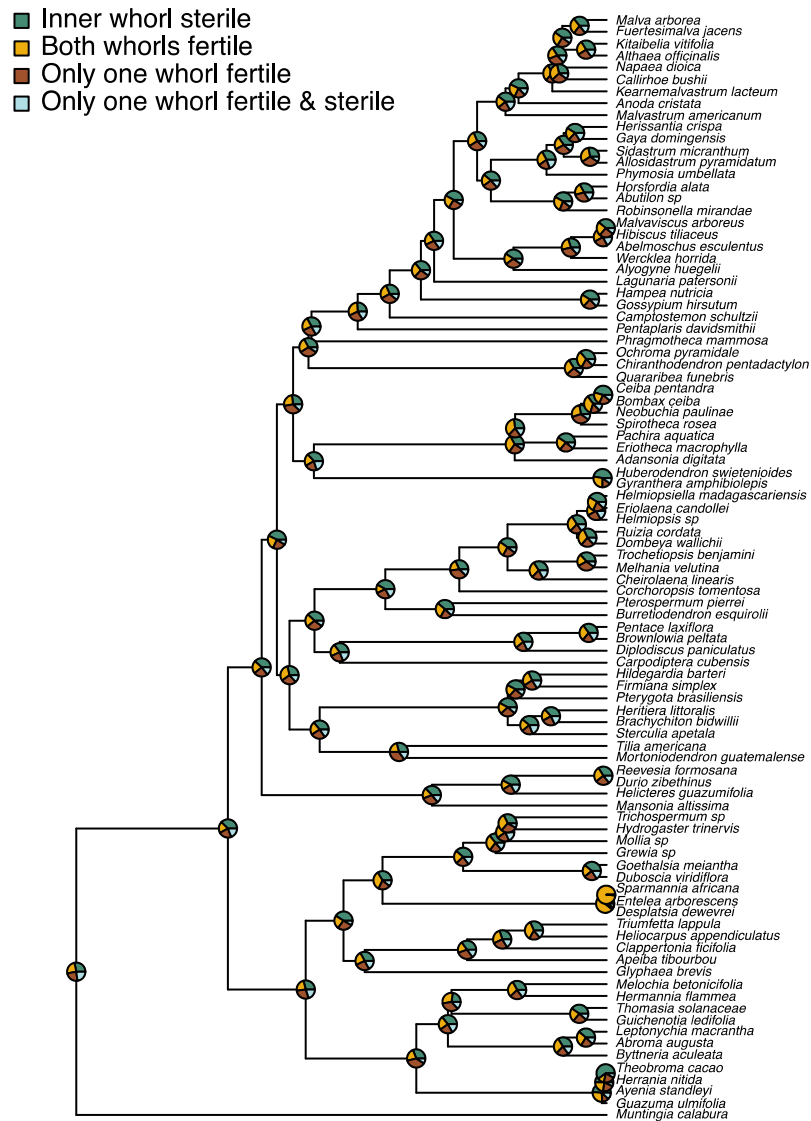


Figure 10. Position androecial whorl-perianth.

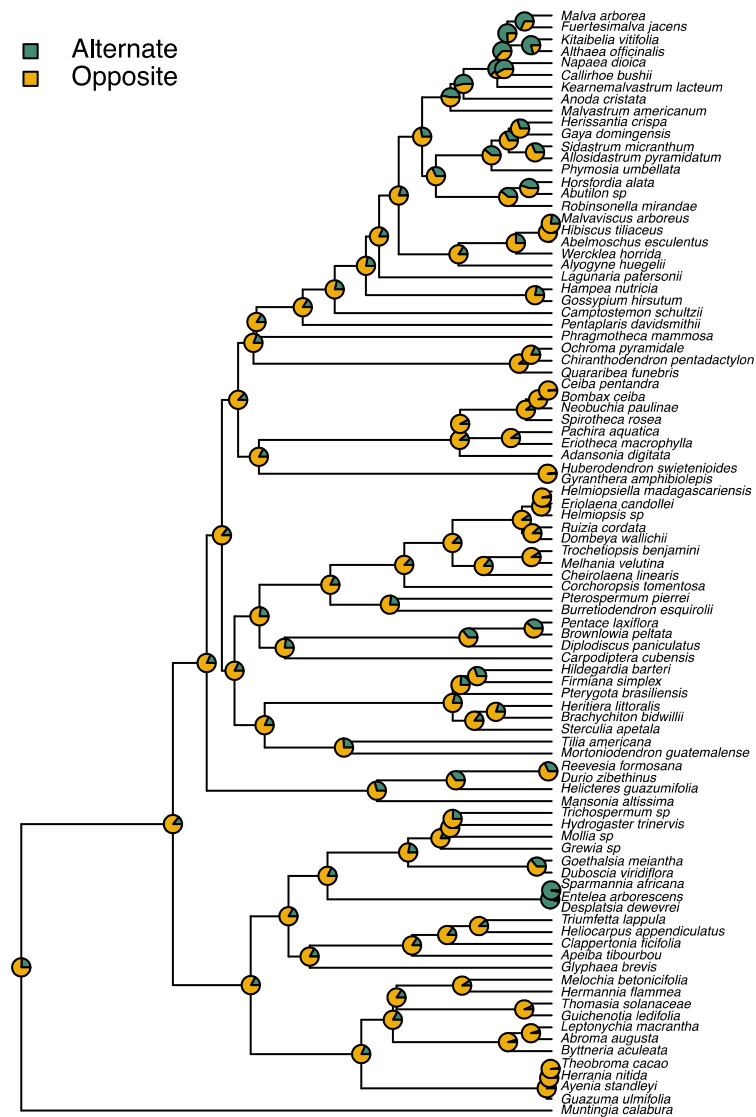


Figure 11. Presentation of androecium.

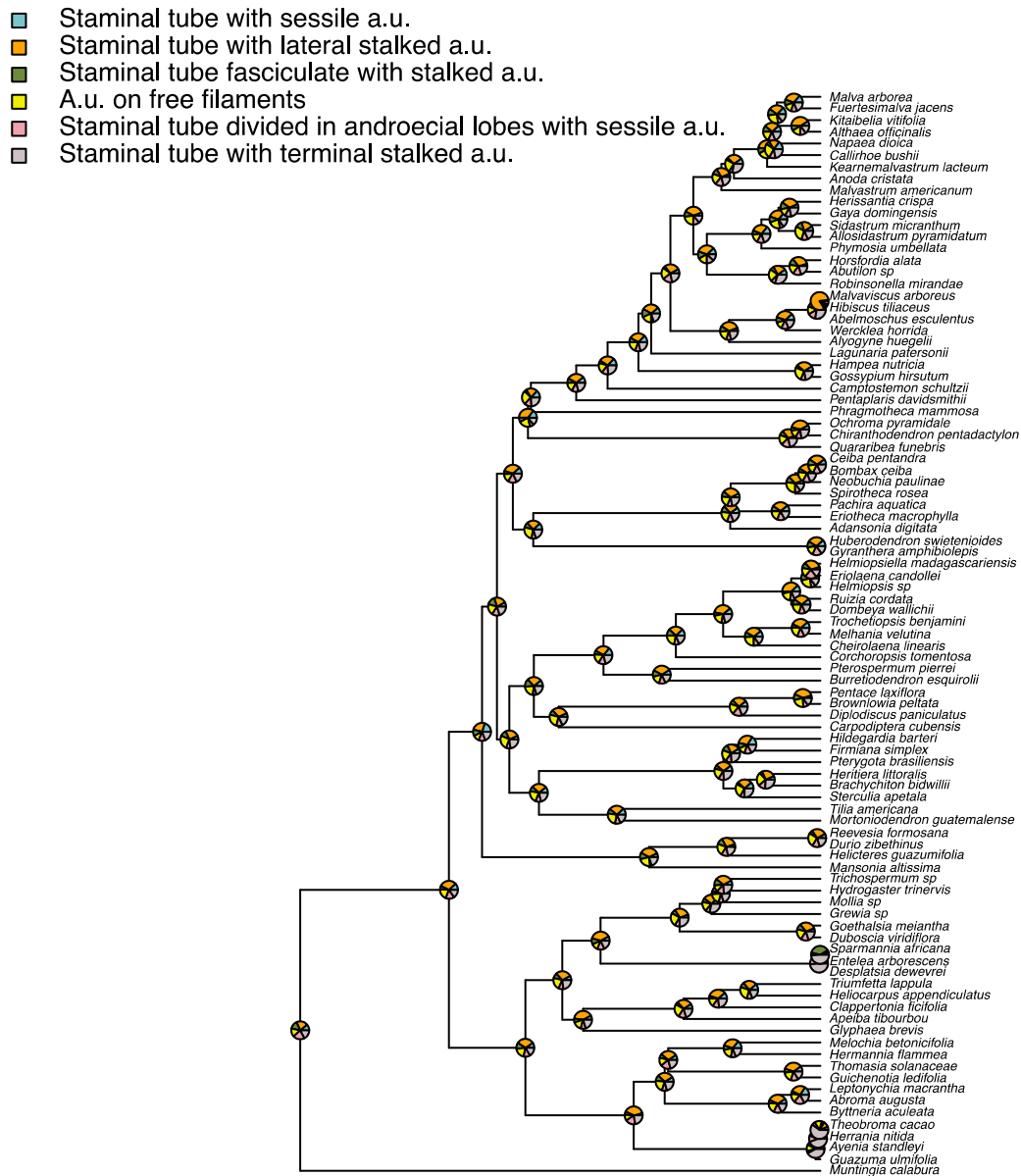
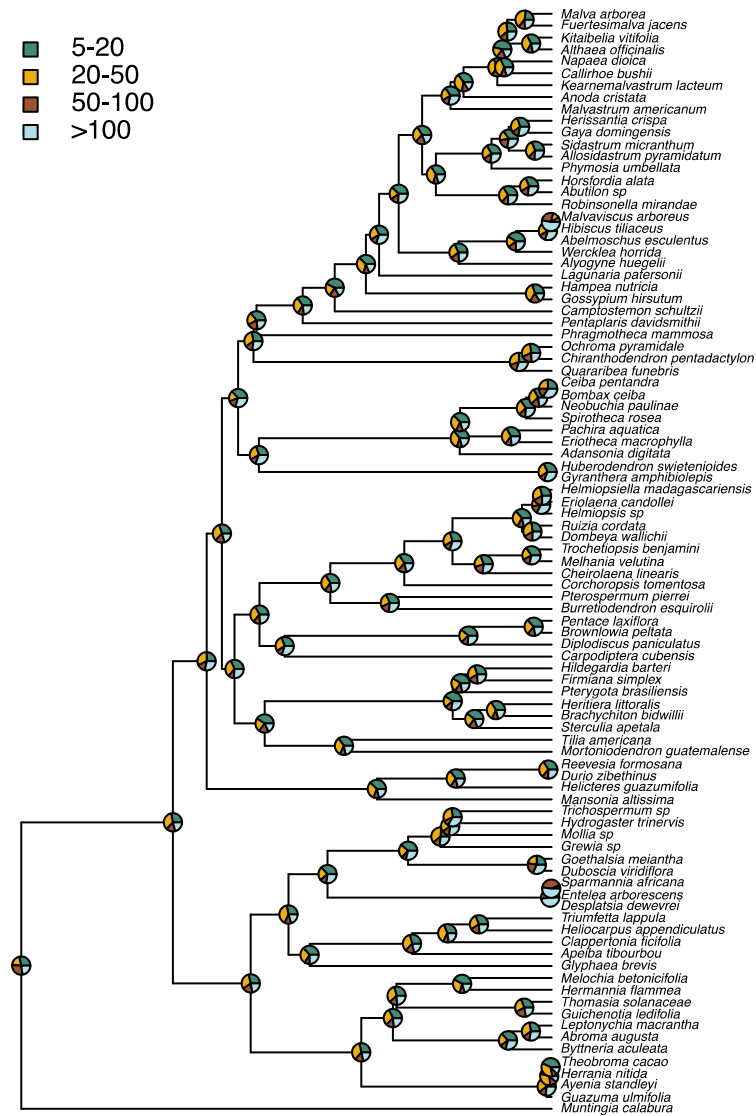


Figure 12. Total androecial units (AU).



DISCUSIÓN GENERAL

El primer paso planteado para abordar el objetivo general de esta tesis fue estimar el tiempo de origen y diversificación del orden Malvales (Cap. I). Con estas edades se estimó la tasa de diversificación neta de cada familia (Cap. I-Fig. 5). Además de obtener esta información, con el amplio muestreo de fósiles y los diferentes análisis que se llevaron a cabo, se revelaron dos aspectos interesantes que están relacionadas con la diversificación del orden. Por un lado, se encontró una relación entre el muestreo de especies vivientes y extintas y la distancia entre la edad troncal y la corona (mecha filogenética). Una larga mecha pudo deberse, además de a sesgos en el muestreo, a procesos macroevolutivos, como alta extinción o baja diversificación (Nagalingum et al., 2011). Por otro lado, la corta distancia entre la raíz del árbol y la divergencia de familias, aunada a la falta de fósiles de linajes del grupo corona de todo el orden, es decir, antes del origen de las familias, es un indicador de que la diversificación fue, probablemente, rápida en la historia evolutiva temprana de Malvales. La relevancia de este capítulo radica en que la integración del registro fósil y los métodos filogenéticos es esencial para comprender patrones macroevolutivos, particularmente aquellos relacionados con especiación y extinción (Silvestro et al. 2018).

Para alcanzar el segundo objetivo particular de la tesis fue necesaria una filogenia molecular de la familia Malvaceae s.l. Para ello, se aplicaron dos métodos de muestreo genético alternativos. Uno fue el minado de transcriptomas-genomas y el otro fue la secuenciación masiva mediante Enriquecimiento Híbrido Anclado (AHE por sus siglas en inglés; Lemmon et al., 2012). Las relaciones filogenéticas resultantes de cada método concuerdan entre sí y con estudios previos (Alverson et al., 1999; Baum et al., 2004; Conover et al., 2018) en tres grandes clados, Byttneriina, Malvadendrina, y Malvatheca (Cap. II-Fig. 4; Cap. III-Fig. 2). Los Caps. II y III coinciden en que Malvatheca es el grupo hermano del resto de las subfamilias (excepto Helicterioideae en Cap. III-Fig. 5), diferente a resultados anteriores y diferente al Cap. I (Fig. 1), probablemente debido a que estos últimos utilizaron en su mayoría información del cloroplasto y la región ITS. Además, se encontró que Sterculioideae y Tilioideae son grupos hermanos (Cap. II-Fig. 4; Cap. III-Fig. 2), lo cual coincide con Conover et al. (2018), pero ellos encontraron estas dos subfamilias más cercanamente relacionadas a Malvatheca; en cambio, en los Caps. II y III se

encontró que las subfamilias están más cercanamente relacionadas a Dombeyoideae, Brownlowioideae, y Helicterioideae (Cap. II-Figs. 4 y 5), o a Dombeyoideae y Brownlowioideae (Cap. III-Fig.2). Es precisamente en la posición de Helicterioideae donde los análisis de genes minados (Cap. II) y AHE (Cap. III) difieren, pues en este último Helicterioideae es el grupo hermano del resto de Malvadendrina. A pesar de que las relaciones filogenéticas de AHE (Cap. III) están resueltas, y en general con alto apoyo estadístico, no se realizó un análisis de concordancia (Ané et al., 2007) como el aplicado a los genes minados (Cap. II). Si se realizara este análisis es posible que los resultados difieran, pues ya se observó que el alto apoyo estadístico no significa que la relación entre dos taxa sea real (Cap. II).

No obstante, la filogenia de AHE es una base sólida para llevar a cabo los análisis macroevolutivos subsecuentes. En primer lugar, se estimaron los tiempos de origen y diversificación de cada linaje. Los resultados indican que la familia Malvaceae *s.l.* se originó en el Cretácico superior, probablemente entre el Turoniano y el Coniaciano (93.9-86.3 Ma; Cap. I y III), al igual que muchas de las otras familias de Malvales (Capítulo I). En ese momento, Gondwana ya estaba separado pero la distancia entre las masas continentales gondwánicas era corta, por lo que se sugiere que la dispersión a larga distancia pudo ser un factor importante en el intercambio de malvales a través de los continentes cercanos, pero independientes, en los cuales actualmente se distribuyen de manera abundante (Morley, 2003).

En segundo lugar, se estimó la tasa neta de diversificación, de especiación y de extinción, pero en particular el objetivo fue evaluar si ha habido cambios en estas tasas y en qué ramas del árbol filogenético ocurrieron. Se encontró que sí ha habido cambios en las tres tasas. En concreto, ha habido dos cambios con mayor probabilidad, un incremento en la tasa en la subfamilia Malvoideae, en el grupo de la jamaica (Hibisceae) y las malvas (Malveae), y un decremento en el grupo externo, *Muntingia calabura* (Cap. III-Fig. 6). Malvoideae es la subfamilia con el mayor número de especies de la familia y su tasa de diversificación neta es alta comparada con el resto de las familias (Cap. I-Fig. 5). Por su parte *Muntingia*, representante de toda la familia Muntingiaceae, tiene la tasa más baja del orden Malvales (Cap. I-Fig. 5), pues está conformada por tres géneros monotípicos. Esta diferencia en la tasa de diversificación entre grupos hermanos es un patrón común en plantas (Donoghue y Sanderson, 2015).

Muntingiaceae es evidentemente un clado depauperado (*sensu* Donoghue y Sanderson, 2015), sin embargo, las causas pueden ser diversas y poco se conoce sobre esta familia: sólo *Muntingia calabura* se ha muestreado en filogenias moleculares, mientras que los otros dos géneros, *Dicraspidia* (Centroamérica y Colombia) y *Neotessmannia* (endémica a Perú), han sido poco estudiados en general (Kubitzki, 2003).

Subsecuentemente, la filogenia fechada de AHE se utilizó para analizar la evolución floral de Malvaceae s.l. (Cap. III). Uno de los retos más grandes en este estudio fue obtener información morfológica de la literatura, pues no fue equitativa la cantidad de información que existe para las especies muestreadas. Esto es un obstáculo que se superó a través de estadísticos estimados por el mapeo estocástico (ver métodos del Cap. III). Sin embargo, el sesgo en el muestreo de caracteres morfológicos es determinante para los modelos utilizados en este estudio (Sauquet y Magallón, 2018). Por lo tanto, los resultados de los análisis del Cap. III son tentativos para los caracteres con mayor número de datos faltantes.

Uno de los resultados inesperados de este análisis fue que el estado ancestral de todas las subfamilias de Malvaceae s.l. fue similar al ancestro común más reciente de toda la familia (Cap. III-Fig. 4), lo cual indica dos cosas: que las flores de los ancestros comunes de las subfamilias eran similares entre sí y que hubo un periodo de estasis morfológica entre el nodo troncal y el corona de cada subfamilia, es decir, las subfamilias se diferenciaron entre sí manteniendo una forma floral similar. Tal vez este patrón de diferenciación a niveles microevolutivos sea similar a lo sugerido en Malpighiaceae, en donde encuentran altas tasas de diversificación pero con estasis morfológica, conservada por la especialización en la polinización (Davis et al., 2014).

Posterior a esta estasis, las subfamilias se diversificaron y la tasa de evolución floral aumentó (Cap. III-Fig. 7), es decir, la morfología floral cambió después de que cada subfamilia se diversificó. Esto señala que la diversidad de formas florales puede explicar, parcialmente, la diversificación en cada subfamilia. Naturalmente, es necesario hacer una investigación detallada para cada grupo e integrar más información (p. ej., hábitat, variables ambientales, interacciones) para entender los procesos evolutivos que pudieron ocurrir en cada caso. No obstante, la heterogeneidad en las tasas de algunos caracteres (p. ej., la organización floral, la

organización del androceo, la presentación del androceo; Cap. III-Tabla 4) y la heterogeneidad en la tasa de diversificación revelan que puede existir una asociación entre ambas tasas. Este estudio es un primer acercamiento para contestar si la evolución floral está asociada a la diversificación en *Malvaceae s.l.*

En conjunto, los tres capítulos que conforman esta tesis representan una base para poder entender con mayor detalle qué procesos macroevolutivos han originado y mantenido la diversidad biológica en *Malvales*. Por ejemplo, se determinó en qué momento del tiempo geológico se originaron y diversificaron los grandes linajes de *Malvales*, se estimó la tasa a la que se generan nuevas especies, y se encontró que la evolución floral es diferente entre las subfamilias de *Malvaceae s.l.* pero esa diferencia es por aceleraciones.

Conclusiones

- Se realizó un muestreo genético, taxonómico (especies vivientes y extintas) y morfológico denso, con el cual se obtuvieron, en general, resultados sólidos sobre la evolución de *Malvales*.
- Las relaciones filogenéticas de *Malvales* y de *Malvaceae s.l.* están resueltas en su mayoría, a excepción de algunos grupos inconsistentes como *Muntingiaceae*, *Cytinaceae*, y las subfamilias de *Malvaceae s.l.*, *Helicterioideae* y *Tilioideae*.
- La información molecular del cloroplasto y el núcleo es incongruente en algunos casos, y entre diferentes genes nucleares, corroborando la compleja historia evolutiva del grupo.
- Las edades estimadas del origen y diversificación de clados de *Malvales* están fuertemente apoyadas por la correspondencia entre dos análisis independientes.
- El número de fósiles y la mecha filogenética están asociados a procesos macroevolutivos, como especiación y extinción, y es relevante considerarlos en conjunto para formular hipótesis evolutivas.
- La tasa de diversificación es heterogénea entre los diferentes linajes de *Malvales*. Particularmente alta en algunas familias como *Malvaceae s.l.* y *Sarcolaenaceae*, y en *Malvoideae* y *Sterculioideae*.

- La evolución floral de Malvaceae s.l. esta marcada por un periodo de estasis, siguiendo un periodo de alta tasa de evolución morfológica dentro de cada subfamilia.
- La Integración de información de especies vivientes (molecular) y extintas, y morfología enriqueció el conocimiento que se tiene sobre la evolución de Malvales y Malvaceae s.l.

Perspectivas

Para continuar respondiendo la pregunta de cómo las Malvales han llegado a su diversidad actual, se necesita abordar cómo se dispersaron a lo largo de los continentes, particularmente en las masas Gondwánicas, pero en general cómo llegaron a su distribución cosmopolita. Actualmente, es posible estimar el área ancestral de un clado y trazar las potenciales rutas de dispersión de los subclados a través de la reconstrucción de áreas ancestrales incorporando especies vivas y registro fósil (Matzke, 2013). Aún más, en términos de conservación, es relevante conocer cuál ha sido la contribución histórica de las Malvales al ensamblaje de los biomas tropicales modernos. Esto podría contestarse, en principio, con la información generada en esta investigación: cuándo se originaron y diversificaron las familias de Malvales. Subsecuentemente, análisis de diversificación asociada a áreas geográficas podría responder si los clados diversificaron gracias a la colonización de dichas áreas, utilizando modelos como GeoHiSSE (Caetano et al., 2018).

Por otro lado, los resultados de esta tesis marcan una base para proponer las causas de la alta diversidad de Malvaceae s.l. Numerosos factores intrínsecos (p. ej., morfología) y extrínsecos (p. ej., variables climáticas) afectan la diversificación de los organismos simultáneamente (Donoghue y Sanderson 2015; Vamosi et al., 2018), por lo tanto, al estimar la relación entre carácter y diversificación, es importante tener en cuenta la interpretación que se da a los resultados. Dos maneras de estimar la contribución de los caracteres florales en Malvaceae s.l. se pueden llevar a cabo complementariamente. Una es directamente estimando la contribución de cada carácter en la diversificación con un modelo dependiente del carácter morfológico, más un carácter oculto (*Hidden State Speciation and Extinction*; Beaulieu & O'Meara 2016; Zenil-Ferguson et al., 2019); y la otra manera es integrando todos los caracteres

florales en una sola tasa de evolución mediante la incorporación de caracteres, y sus estados de carácter, por módulos ontogenéticos (Tarasov et al. 2019).

Literatura citada

- Alverson WS, Whitlock BA, Nyffeler R, Bayer C, Baum DA. 1999. Phylogeny of the core Malvales: Evidence from *ndhF* sequence data. *American Journal of Botany* 86(10): 1474-1486.
- Ané C, Larget B, Baum DA, Smith SD, Rokas A. 2007. Bayesian estimation of concordance among gene trees. *Molecular Biology and Evolution* 24: 412-426.
- Baum D, Smith S, Yen A, Alverson W, Nyffeler R, Whitlock B, Oldham R. 2004. Phylogenetic relationships of Malvaceae (Bombacoideae and Malvoideae; Malvaceae sensu lato) as inferred from plastid DNA sequences. *American Journal of Botany* 91(11): 1863-1871.
- Beaulieu JM y O'Meara BC. 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Systematic Biology* 65: 583-601.
- Caetano DS, O'Meara BC y Beaulieu JM. 2018. Hidden state models improve state-dependent diversification approaches, including biogeographical models. *Evolution* 72(11): 2308-2324.
- Conover JL, Karimi N, Stenz N, Ané C, Grover CE, Skema C, Tate JA, Wolff K, Logan SA, Wendel JF, Baum DA. 2018. A Malvaceae mystery: A mallow maelstrom of genome multiplications and maybe misleading methods? *Journal of Integrative Plant Biology* 61(1): 12-31.
- Davis CC, Schaefer H, Xi Z, Baum DA, Donoghue MJ, Harmon LJ. 2014. Long-term morphological stasis maintained by a plant-pollinator mutualism. *Proceedings of the National Academy of Sciences* 111(16): 5914-5919.
- Donoghue MJ y Sanderson MJ. 2015. Confluence, synnovation, and depauperons in plant diversification. *New Phytologist* 207: 206-274.
- Kubitzki K. 2003. Muntingiaceae. In Kubitzki K & Bayer C. *Flowering plants dicotyledons. Capparales, Malvales and Non-betalain Caryophyllales*. Springer.
- Lemmon AR, Emme SA, Moriarty Lemmon E. 2012. Anchored Hybrid Enrichment for massively high-throughput phylogenomics. *Systematic Biology* 61(5): 727-744.

- Matzke NJ. 2013. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers in Biogeography* 5: 242-248.
- Morley RJ. 2003. Interplate dispersal paths for megathermal angiosperms. *Perspectives in Plant Ecology, Evolution and Systematics* 6 (1,2): 5-20.
- Nagalingum NS, Marshall CR, Quental TB, Rai HS, Little DP, Mathews S. 2011. Recent synchronous radiation of a living fossil. *Science* 334: 796-799.
- Rabosky DL. 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLOS One* 9: e89543.
- Sauquet H y Magallón S. 2018. Key questions and challenges in angiosperm macroevolution. *New Phytologist* 219: 1170-1189.
- Silvestro D, Warnock RCM, Gavryushkina A, Stadler T. 2018. Closing the gap between palaeontological and neontological speciation and extinction rate estimates. *Nature Communications* 9: 5337.
- Tarasov S, Miko I, Yoder MJ, Uyeda JC. 2019. PARAMO pipeline: reconstructing ancestral anatomies using ontologies and stochastic mapping. *bioRxiv* doi: <http://dx.doi.org/10.11.01/553370>.
- Vamosi JC, Magallón S, Mayrose I, Otto SP, Sauquet H. 2018. Macroevolutionary patterns of flowering plant speciation and extinction. *Annual Review of Plant Biology* 69: 9.1-9.22.
- Davis CC, Schaefer H, Xi Z, Baum DA, Donoghue MJ, Harmon LJ. 2014. Long-term morphological stasis maintained by a plant-pollinator mutualism. *Proceedings of the National Academy of Sciences* 111(16): 5914-5919.