UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS BIOLÓGICAS

INSTITUTO DE BIOLOGÍA

BIOLOGÍA EVOLUTIVA

EVALUACIÓN DE MÉTODOS PARA LA RECONSTRUCCIÓN DEL PROCESO DE DIVERSIFICACIÓN Y SU RELACIÓN CON LA HISTORIA MACROEVOLUTIVA DE TAXA VIVIENTES

TESIS

QUE PARA OPTAR POR EL GRADO DE:

DOCTORA EN CIENCIAS

PRESENTA:

LUNA LUISA SÁNCHEZ REYES

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

COMITÉ TUTOR: **DRA. ELIANE REGINA RODRIGUES**
INSTITUTO DE MATEMÁTICAS, UNAM

DR. LUIS ENRIQUE EGUIARTE FRUNS
INSTITUTO DE ECOLOGÍA, UNAM

MÉXICO, CD. MX. **DICIEMBRE 2016.**

UNAM POSGRADO
Ciencias Biológicas

COORDINACIÓN

Lic. 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 y Manejo Integral de Ecosistemas del Posgrado en Ciencias Biológicas, celebrada el día 3 de octubre de 2016, se aprobó el siguiente jurado para el examen de grado de **DOCTORA EN CIENCIAS** de la alumna **SÁNCHEZ REYES LUNA LUISA** con número de cuenta **404055121** con la tesis titulada: **"Evaluación de métodos para la reconstrucción del proceso de diversificación y su relación con la historia macroevolutiva de taxa vivientes"**, realizada bajo la dirección de la **DRA. SUSANA AURORA MAGALLÓN PUEBLA:**

Presidente:	DR. DANIEL IGNACIO PIÑERO DALMAU
Vocal:	DR. ALEJANDRO GONZÁLEZ VOYER
Secretario:	DR. LUIS ENRIQUE EGUIARTE FRUNS
Suplente:	DR. NATALIA IVALÚ CACHO GONZÁLEZ
Suplente	DRA. CLAUDIA PATRICIA ORNELAS GARCÍA

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

A T E N T A M E N T E
"POR MI RAZA HABLARA EL ESPIRITU"
Cd. Universitaria, Cd. Mx, a 14 de noviembre de 2016.

DRA. MARÍA DEL CORO ARIZMENDI ARRIAGA
COORDINADORA DEL PROGRAMA

POSGRADO EN CIENCIAS BIOLOGICAS
COORDINACIÓN

c.c.p. Expediente del (la) interesado (a).

AGRADECIMIENTOS INSTITUCIONALES

En primer lugar, agradezco al Posgrado en Ciencias Biológicas de la Universidad Nacional Autónoma de México (UNAM), por las facilidades brindadas durante mis estudios de doctorado para formarme profesionalmente.

Agradezco Al Consejo Nacional de Ciencia y Tecnología (CONACYT), por la beca 262540 para estudios de posgrado otorgada durante el tiempo que duraron mis estudios. A los apoyos CONACYT 2004-C01-46475 otorgado al Dr. Luis E. Eguiarte Fruns y PAPIIT-UNAM 202310 otorgado a la Dra. Susana A. Magallón Puebla.

Agradezco a mi tutora principal, la Dra. Susana A. Magallón Puebla, y a los miembros del comité tutor, la Dra. Eliane R. Rodrigues y el Dr. Luis E. Eguiarte Fruns, por la valiosa guía para formarme como investigadora, y por su tiempo, dedicación, críticas y sugerencias certeras para la realización exitosa de este trabajo de investigación.

AGRADECIMIENTOS PERSONALES

A mi asesora, la Dra. Magallón. Por ser un ejemplo de vida y de trabajo, y una guía excelente en la investigación. Por apoyarme de todas las maneras posibles y siempre creer en mi. Gracias infinitas por dedicar estos años a mi formación.

A mis profesores, el Dr. Luis Eguiarte, la Dra. Valeria Souza y el Dr. Daniel Piñero, cuya pasión por la genética de poblaciones y la evolución en general alimentó mi interés por este extraordinario campo de investigación. A la Dra. Eliane Rodrigues y a la Dra. Hélène Morlon, por ayudarme generosamente a encarrilar esta visión matemática, y mostrarme nuevos y apasionantes caminos en la investigación.

Al Programa de Apoyo para Estudios de Posgrado (PAEP), por el apoyo brindado para la realización de estancias de investigación y para asistir a cursos y conferencias internacionales, los cuales fueron cruciales en la adquisición y desarrollo de habilidades que utilicé para la realización de esta tesis y que continuaré aplicando en mi carrera.

Al Centro de Matemáticas Aplicadas de Palaiseau, París, Francia (*Centre de Mathématiques Appliquées Palaiseau* CMAP), por facilitarme el uso de los recursos computacionales que se usaron para la realización de esta tesis.

A mis queridos padres, Héctor y Georgina, y a mi querida hermana, María. Gracias por acompañarme y apoyarme de cerca y de lejos en todas mis aventuras, por compartir su gran amor y sabiduría conmigo siempre y recordarme constantemente dónde están para poder compartirlos también. ¡Que la maravillosa y gran aventura siga y siga!

A mi familia, en particular a mi amorosa abuela Chata y a mis primos hermanos, María Devaki, Radha y Jos, y a mis queridos tíos, Javier y Cecilia. Por sus sabios consejos, la música, la alegría, las historias y el canto. Gracias por quererme tanto.

A *Botanical Nautilus*, en orden de aparición (relativa): Rebe Hernández, Adriana Benitez, Andrea López, Itzi Fragoso y Paty Rivera. Por su amistad única y más que enriquecedora, que hizo de esta travesía por el doctorado una de las experiencias que más he disfrutado en la vida.

A los amigos y compañeros en la ciencia Andrea González, Sandra Gómez, Tania Hernández, José Antonio Barba, Julián Velasco, Jéremie Bardin, Paul Simion, Roberto Trejo, Sergio Ramos Castro, Orlando Schwery, Alex Hall, y Santiago Barahona. Por todas las pláticas para tratar de entender mejor esta realidad que no deja de sorprendernos y retarnos.

A los amigos y profesores del Paleobiology Database (PBDB) Workshop 2012: Kirstin Brink, John Clarke, Jérémie Bardin, Laura Soul, Anneke Madern, Sahale Casebolt, Rafal Nawrot, David Nicholson, Rooland Sookias, Celeste Pérez, Martín Ezcurra, John Alroy, Gene Hunt, David Polly, Peter, Wagner y Tom Olszweski.

A los compañeros y profesores del lab Morlon y aliados. En particular a Dan Moen, Jonathan Rolland, Fabien Condamine, Jana Smrckova, Fanny Gascuel, Miraine Dávila y Amaury Lambert.

A mis queridos amigos, que estuvieron siempre muy cerquita (aunque estuvieran lejos) y que me acompañaron en algunas o todas las fases de este proceso, manteniéndome con los pies en la tierra (bailando), y en particular a los que llegaron conmigo hasta el final: Erick Castillo, Neda Jaramillo, Nicolás Utrilla, Emilio Fernández, Jesús Chico, Rafael Serrano, Daniel Sepúlveda, Rocío Luna, Natalia Martínez, Amilú DV, Aimée Suárez, Dominic Sowa, Karl García y Claudia Sosa. Los adoro.

DEDICATORIA

A mis padres, Héctor y Georgina

A mi hermana, María

Quién ha estado aquí
mirando el fin de la calle
sobre la cual cuelga
tan cercana la luna roja
tan enorme la roja luna

Quién ha estado
solitario en este mismo lugar
hace cien años
en quién pensaba el solitario
en qué pensaba el solitario
o simplemente miraba
un vacío rodeado por la noche

No había casas
no había sino un ruido
pero no era un ruido
sino el ruido de un río
y quién estará
en cien años más
en el lugar que ahora llamo yo mi casa
cuando yo no sea sino el silencio
quien estará en un vacío rodeado por la noche
sin saber nunca si aquí hubo casas o calles
y nadie sino el ruido de un río silencioso
podría recordarlo.

Jorge Teillier

ÍNDICE

Lista de Figuras y Tablas *ix*

Abreviaturas y variables *ix*

RESUMEN **1**

ABSTRACT **3**

INTRODUCCIÓN GENERAL **5**

Medidas de la diversificación de especies *5*

Métodos para reconstruir el proceso de diversificación *7*

Causas primarias de la diversificación: procesos de diversidad ilimitada vs. límites a la diversidad *8*

Evidencias de procesos con diversidad ilimitada *9*

Evidencias de procesos con límites a la diversidad *10*

Estadísticos descriptivos: un complemento de los métodos para reconstruir el proceso de diversificación *12*

Un estadístico descriptivo popular: la relación entre la edad de los clados y su riqueza de especies *13*

¿Qué tan informativa sobre el proceso de diversificación es la relación entre la edad de los clados y la riqueza de especies? *14*

Objetivos *16*

CAPÍTULO 1. CAMBIOS MACROEVOLUTIVOS EN LA DIVERSIFICACIÓN REVELAN UNA COMPLEJIDAD DE RUTAS HACIA LA MEGADIVERSIDAD DE LAS ANGIOSPERMAS. **21**

Manuscrito con formato para Nature **22**

Material Suplementario **51**

CAPÍTULO 2. DIVERSIFICACIÓN DE *FOUQUIERIA* (FOUQUIERIACEAE, ERICALES) EN LOS DESIERTOS NORTEAMERICANOS **64**

Manuscrito con formato para New Phytologist **65**

Material Suplementario **108**

CAPÍTULO 3. DESCUBRIENDO LA DINÁMICA DE DIVERSIFICACIÓN DE TAXA SUPERIORES CON DATOS DE EDAD Y RIQUEZA DE ESPECIES DE LOS CLADOS **121**

Artículo publicado en Systematic Biology **122**

Material Suplementario **160**

DISCUSIÓN GENERAL **182**

Utilidad de la relación entre la edad de los clados y su riqueza de especies para estudiar el proceso de diversificación *182*

Informatividad de la diferencia entre ARC corona y ARC troncal *183*

Informatividad de la diferencia de la ARC entre rangos taxonómicos *184*
Conclusiones sobre la informatividad de la ARC *186*
Estudiar la relación entre la edad de los clados y su riqueza de especies en un contexto filogenético *187*
La relación entre la edad de los clados y la riqueza de especies como complemento de métodos para reconstruir el proceso de diversificación *190*
Perspectivas *192*

CONCLUSIONES GENERALES **194**

REFERENCIAS GENERALES **195**

APÉNDICE 1. UN ÁRBOL DE TIEMPO METACALIBRADO DOCUMENTA LA APARICIÓN TEMPRANA DE LA DIVERSIDAD FILOGENÉTICA DE LAS PLANTAS CON FLOR **203**

Artículo publicado en New Phytologist **204**

Material Suplementario **221**

FIGURAS Y TABLAS

Figura 1. Modelos de cambio en la diversidad a lo largo del tiempo.

Figura 2. Diferencia entre la edad corona y la edad troncal de un clado.

ABREVIATURAS Y VARIABLES

ML: máxima versosimilitud (por su acrónimo en inglés *Maximum Likelihood*)

modelo BD: modelo de nacimiento y muerte de linajes (por sus acrónimo en inglés *Birth-Death model*)

modelo hL: modelo de vida media de los clados (por su acrónimo en inglés *halfLife of clades model*)

myr: millones de años (por su equivalente en inglés *million years*)

ARC: Correlación entre la edad de los clados y su riqueza de especies (por su acrónimo en inglés *Age-Richness Correlation*)

PAPS: Simulaciones predictivas *a posteriori* (por su acrónimo en inglés *Predictive A Posteriori Simulations*)

r: tasa de diversificación de especies

λ: tasa de especiación

μ: tasa de extinción de especies

ε: tasa de extinción de especies relativa

λ_0: tasa inicial de especiación (valor de λ al inicio del proceso de diversificación)

λ_t: tasa de especiación en el tiempo t

Φ: tasa de origen de taxa superiores

α: tasa de cambio entre clados de las tasas de diversificación

N: número de especies

K: capacidad de carga de un clado

Λ: lambda de Pagel (grado de estructura filogenética de un carácter)

RESUMEN

La relación entre la edad de los clados y su riqueza de especies es usada en biología evolutiva para distinguir entre procesos de diversificación dependientes del tiempo y dependientes de la ecología. La teoría sugiere que la falta de independencia estadística debido a una historia evolutiva compartida (estructura filogenética) en los datos de edad y riqueza de los clados, el tipo de edad (corona o troncal) usada para estimar la relación y la resolución de la clasificación taxonómica (niveles taxonómicos más inclusivos o altos –e.g., órdenes– o menos inclusivos o bajos –e.g., géneros), pueden sesgar los valores estimados de la correlación edad-riqueza de especies (ARC, por sus siglas en inglés *Age-Richness Correlation*). En esta tesis se caracterizó el comportamiento de la ARC en filogenias fechadas (cronogramas) empíricas y simuladas bajo diferentes procesos de diversificación, considerando la estructura filogenética de los datos, edades corona y troncales, y diferentes niveles taxonómicos, con el objetivo de determinar mejor la capacidad de la ARC de informarnos sobre los procesos de diversificación. Primero, para determinar la ARC en la naturaleza, se realizó una compilación de cronogramas de diferentes grupos de organismos (sistemas biológicos) provenientes de estudios publicados. Escogimos sólo aquellos cronogramas obtenidos con métodos de reloj molecular, cuyo muestreo de especies comprendiera al menos dos niveles taxonómicos del sistema biológico en cuestión. Esto incluyó cronogramas a nivel de género de anfibios (Lissamphibia), reptiles escamados (Squamata) y aves (Aves), y a nivel de familia de mamíferos (Mammalia). Para ampliar la muestra de cronogramas empíricos, se reconstruyeron las relaciones filogenéticas de las familias de plantas con flor (Angiospermae), así como su edad con un método de metacalibración (Apéndice 1). En

segundo lugar, se hizo una revisión de métodos de estimación de parámetros de diversificación (i.e., especiación y extinción), con el fin de aplicarlos a los cronogramas empíricos y obtener un rango de valores de especiación y extinción para la implementación de simulaciones. Para su estandarización, estos métodos se aplicaron en diferentes sistemas biológicos de estudio (Capítulos 1 y 2). Finalmente, se desarrolló e implementó un modelo de nacimiento y muerte de dos estados para simular cronogramas con eventos de origen de taxa superiores y una jerarquía taxonómica, con el fin de determinar la ARC esperada en modelos de diversificación dependiente del tiempo y de la ecología, y así establecer las expectativas de comportamiento de la ARC en dos hipótesis de diversificación que en general se consideran alternativas. Las ARCs empíricas y simuladas se estimaron usando edades corona y troncal en diferentes rangos taxonómicos con métodos de estimación de correlación que no consideran estructura filogenética y con métodos que sí la toman en cuenta. El modelo propuesto reproduce de manera general los patrones empíricos de ARC, a pesar de las particularidades de clasificación taxonómica inherentes a cada sistema biológico. Las ARCs estimadas con edades corona son positivas en todos los escenarios de diversificación considerados, incluyendo procesos de diversificación dependiente de la ecología. Las ARCs estimadas con edades troncales sólo son negativas en rangos taxonómicos bajos en modelos de variación de tasas de diversificación entre clados, tanto dependientes de tiempo como de la ecología (Capítulo 3). Esto implica que sólo las ARC troncales pueden distinguir entre procesos de diversificación de tasas constantes y variables entre clados, pero no pueden usarse como evidencia de variación en la diversificación dependiente de la ecología.

ABSTRACT

The relationship between clade age and species richness is used in evolutionary biology studies to distinguish among time and ecologically dependent diversification processes. Theory suggests that statistical non-independence of the data due to shared evolutionary history (phylogenetic structure) in clade age and richness data, type of clade age used to estimate the relationship (crown or stem) and resolution of taxonomic classification (more inclusive ranks –e.g., orders– or less inclusive –e.g., genera), might bias age-richness correlation (ARC) estimates. In this dissertation, behavior of the ARC was characterized with data from dated phylogenies (chronograms) of different biological systems and from chronograms simulated under different diversification processes, taking into account phylogenetic structure, crown or stem ages and different taxonomic ranks, to determine ARC informative breadth on diversification processes. First, dated molecular phylogenies sampling at least two different taxonomic levels were compiled from the literature. To increase the number of sampled phylogenies, phylogenetic relationships among flowering plant (Angiospermae) families were reconstructed and dated using a metacalibration approach (Appendix 1). The final sample comprises chronograms of amphibians (Lissamphibia), squamate reptiles (Squamata), and birds (Aves), sampled at a generic level; and of mammals (Mammalia) and flowering plants, sampled at a familial level. Then, a review of methods to estimate diversification parameters from dated phylogenies was performed. The goal of this was to apply them to the empirical phylogenies to obtain real parameters to simulate phylogenies under different time and ecologically dependent diversification scenarios, and determine ARC expectations. For standardization, some of the methods were applied to different biological systems

(Chapters 1 and 2). Finally, a two-state birth-death model to simulate chronograms including the origin of higher taxa and a hierarchical taxonomy is developed and implemented, to determine expected ARC behavior under two diversification hypotheses that are generally considered alternative. Empirical and simulated ARCs were estimated using crown and stem ages from different taxonomic ranks using methods that take phylogenetic structure into account and methods that do not consider it. The simulation model proposed here reproduces the general ARC trends from a wide range of biological systems despite the particularities of taxonomic practice within each. ARCs estimated with crown ages are positive in all the scenarios studied, including ecologically dependent processes. ARCs estimated with stem ages are only negative at less inclusive taxonomic ranks, in models of among-clade rate variation, both time- and ecologically dependent (Chapter 3). These results suggest that only stem-ARCs can distinguish between rate constancy and among-clade variation in rates, but they cannot evidence ecologically dependent rate variation.

INTRODUCCIÓN GENERAL

La variación en número de especies (diversidad o riqueza) que se observa entre grupos de organismos (sistemas biológicos), entre comunidades y regiones, y a lo largo del tiempo geológico, es un patrón que ha interesado a los naturalistas por mucho tiempo (Darwin 1859; Wallace 1878; Bokma et al. 2014). Entender las causas y determinar los procesos que dan lugar a los patrones de variación en la diversidad a diferentes escalas, a lo largo de gradientes temporales, filogenéticos, ecológicos y espaciales de diversidad, es uno de los retos actuales de la biología evolutiva.

A escala temporal y filogenética, el número de especies es resultado del proceso evolutivo de diversificación, el cual a su vez resulta de la interacción y del balance entre la especiación y la extinción. Independientemente de los mecanismos que ocurren a nivel poblacional –como la selección natural, la deriva génica, el aislamiento reproductivo, por mencionar algunos– y que hayan dado lugar a eventos particulares de especiación y extinción, se sabe que el patrón temporal de ocurrencia de estos eventos se puede describir de manera general con un modelo aleatorio en el cual los eventos de especiación y extinción se acumulan exponencialmente a lo largo del tiempo (Yule 1925; Kendall 1948; Nee 2006). Esto nos permite estudiar patrones y procesos de diversificación de especies a una escala por arriba de los mecanismos que ocurren a nivel poblacional, es decir, a escala macroevolutiva.

Medidas de la diversificación de especies

Para describir el proceso de diversificación en general y la manera en que la diversidad varía a diferentes escalas en particular, una medida comúnmente utilizada es la tasa de

diversificación (*r;* Stanley 1975), que representa el cambio en diversidad por unidad de tiempo (Sepkoski 1978), esto es, la velocidad a la cual surgen nuevas especies que sobrevivirán hasta un determinado tiempo. Para estimar *r* se requieren mínimamente dos datos: la riqueza de especies y el tiempo transcurrido (i.e., el tiempo durante el que se acumuló dicha riqueza de especies), el cual puede ser considerado desde la edad corona o desde la edad troncal de un grupo (Fig. 1). Si contamos con datos adicionales, como un conocimiento detallado de la ocurrencia de eventos de aparición y desaparición de especies, podemos calcular las tasas de especiación, *lambda* (λ; Yule 1925) y de extinción, *mu* (μ; Kendal 1948), respectivamente (Sepkoski 1998). La relación entre estos parámetros es: $r = \lambda - \mu$ (Stanley 1975). Otro parámetro que nos informa sobre la naturaleza del proceso de diversificación es la proporción entre la tasa de extinción y la de especiación, μ / λ (Nee et al. 1994a), conocida como tasa de extinción relativa, *epsilon* (ε; Magallón y Sanderson 2001), que describe la velocidad a la cual las especies son reemplazadas por nuevas especies (tasa de recambio de especies; Alfaro et al. 2009).

Estos parámetros de diversificación parecen sencillos de estimar, sin embargo, los procesos de especiación y extinción son, salvo raras excepciones, desconocidos, y deben ser inferidos. Además, la diversificación de especies es un proceso que ocurre en una escala de tiempo del orden de millones de años en la mayoría de los organismos, lo que dificulta la implementación de experimentos de laboratorio para estudiarla. Una alternativa para investigar los procesos de diversificación ocurriendo a grandes escalas temporales es la implementación de modelos probabilísticos, los cuales permiten estimar con incertidumbre estadística diferentes parámetros evolutivos potencialmente

implicados en el proceso de diversificación, mínimamente, las tasas de especiación y extinción.

Métodos para reconstruir el proceso de diversificación

Tradicionalmente, la diversificación fue estudiada por paleontólogos en organismos extintos o vivientes con un amplio registro fósil (Jepsen et al. 1949; Simpson 1953; Van Valen 1973; Gould y Eldredge 1977; Stanley 1985), con contadas excepciones (e.g., Yule, 1925). Recientemente, la disponibilidad de filogenias moleculares, que reconstruyen las relaciones filogenéticas de las especies y que permiten estimar la edad de ocurrencia de los eventos de divergencia, promovió el uso de modelos probabilísticos para la implementación de métodos estadísticos robustos, como máxima verosimilitud (ML) o estadística Bayesiana, para estimar tasas de especiación y extinción de organismos vivientes con poco o nulo registro fósil (Hey 1992; Nee et al. 1992, 1994a, 1994b; Harvey et al. 1994). Esto amplió considerablemente la gama de sistemas biológicos para estudiar los procesos de diversificación, así como el tipo y la profundidad de las preguntas evolutivas que pudieron ser planteadas.

En los últimos años, las herramientas para estudiar los procesos de diversificación en filogenias se han afinado y aumentado en número (para una revisión, ver Morlon 2014). Usando métodos de ajuste de modelos es posible evaluar una serie de hipótesis de diversificación alternativas y determinar cuál es el que se ajusta mejor a los patrones de diversidad que observamos. Con estos métodos se han podido evaluar diversas hipótesis de variación en la diversificación asociada a caracteres morfológicos (Maddison et al. 2007; Fitzjohn et al. 2009; Fitzjohn 2010; Magnuson-Ford y Otto 2012), a la distribución

geográfica (Goldberg et al. 2011; Rolland et al. 2014), a condiciones paleoclimáticas (e.g., la temperatura paleoambiental Condamine et al. 2013) y a la diversidad (Rabosky y Lovette 2008a; Etienne et al. 2012), entre otros.

Causas primerias de la diversificación: procesos de diversidad ilimitada vs. límites a la diversidad

Los modelos de diversificación se pueden clasificar de manera general en dos categorías: aquellos que predicen patrones en los que la diversidad de especies se acumula ilimitadamente y aquellos que indican que la diversidad es limitada (Fig. 2). Los modelos de diversidad ilimitada indican que sólo el tiempo determina la diversidad, y que las diferencias en las tasas de diversificación son intrínsecas a cada sistema biológico (Fig. 2a). Los modelos de diversidad limitada indican que, después de un periodo de aumento en la diversidad, ésta llega a un estado de equilibrio en que la especiación y la extinción son aproximadamente iguales. El equilibrio puede ser momentáneo, para dar paso a un decremento en la diversidad (Fig. 2b); o duradero, manteniendo una diversidad constante por un tiempo indefinido (Fig. 2c). En estos modelos las tasas de diversificación están determinadas por factores extrínsecos como la disponibilidad de nichos, el área de distribución, o el número de especies. En la actualidad existe un debate sobre la prevalencia de estos dos tipos de procesos en datos empíricos (Rabosky 2009; Wiens 2011; Cornell 2013), debate que es crucial resolver para poder abordar adecuadamente otras preguntas de gran interés biológico, como la relación entre la diversificación y la evolución morfológica (Rabosky y Adams 2012).

Evidencias de procesos con diversidad ilimitada

En 1925, el matemático Yule propuso el primer modelo de diversificación usando las distribuciones de frecuencia de número de especies vivientes y de tiempo de origen (tomado del registro fósil) de géneros de escarabajos, lagartijas, serpientes y plantas con flor. Yule (1925) determinó que las especies surgen al azar y se acumulan exponencialmente en el tiempo. Esto tiene como consecuencia que la diversidad actual de especies depende únicamente del tiempo transcurrido, del número inicial de especies y del parámetro λ –el cual no cambia en el tiempo ni entre linajes dentro de un mismo grupo, aunque sí entre grupos (Yule 1925). El modelo de Yule (también llamado de nacimiento puro) es un caso particular del modelo de nacimiento y muerte (modelo BD por sus siglas en inglés *Birth-Death*), que fue derivado años más tarde, y ejemplificado con el decaimiento de partículas en el espacio y los procesos epidémicos (Kendall 1948). El modelo BD considera además de especiación, a la extinción (Nee 2006). Este modelo es la base de la mayoría de los métodos que se han desarrollado para estudiar diversificación en los últimos años (Nee 2006).

El registro fósil también ha aportado evidencia de que la diversidad se acumula sin límites a lo largo del tiempo. Por ejemplo, los datos fósiles de los últimos 540 millones de años (myr) de vida en la Tierra parecen mostrar que la biodiversidad se ha acumulado de manera constante, hasta hace 75 myr, cuando presenta un incremento (Cornette y Lieberman 2004). Curvas de supervivencia basadas en el registro fósil de especies, géneros y familias de organismos unicelulares y multicelulares dieron lugar a la hipótesis de la Reina Roja, que explica que la extinción ocurre de manera aleatoria y constante en el tiempo (Van Valen 1973). Más adelante se probó que para que la

extinción sea constante, la especiación también debería de serlo (Stenseth y Smith 1984), lo que se ha confirmado en 101 filogenias moleculares a nivel de especie de géneros de plantas, animales y hongos (Venditti et al. 2010).

Para que la diversidad sea ilimitada, las tasas de especiación y extinción no tienen que ser constantes. La especiación puede disminuir en el tiempo, y mientras no sea igual o inferior a la tasa de extinción (resultando en una tasa de diversificación nula o negativa), la diversidad seguirá acumulándose (Morlon et al. 2011; Maruvka et al. 2013). Usando métodos de coalescencia se encontró que de 289 filogenias muestreadas a nivel de especie de clados de rangos taxonómicos bajos –casi en su totalidad géneros– de anfibios, artrópodos, aves, mamíferos, moluscos y plantas con flor, aproximadamente el 80% se encuentran en proceso de seguir acumulando diversidad, de los cuales más de la mitad presentan cambios temporales en las tasas de especiación y extinción, y el resto presenta especiación constante (Morlon *et al.* 2010). Estudios en filogenias de sistemas biológicos pertenecientes a niveles taxonómicos más altos o más inclusivos (clados de angiospermas, vertebrados, mamíferos y aves), muestran que las tasas de especiación y extinción también pueden cambiar entre linajes dentro de un mismo grupo y que la diversidad sigue en expansión (Magallón y Sanderson 2001; Alfaro et al. 2009; Stadler 2011; Jetz et al. 2012).

Evidencias de procesos con límites a la diversidad

En el registro fósil prevalecen patrones en los que la diversidad tiende a un límite, el cual es particular a cada clado y se alcanza a diferentes ritmos (Raup et al. 1973; Gould et al. 1977; Sepkoski 1978; Foote 2007; Foote et al. 2007; Benton 2009). Estos patrones

pueden ser explicados por una variante del modelo BD, el modelo de Moran, originalmente desarrollado para estudiar procesos de genética de poblaciones (Moran 1958; Nee 2006; Crawford et al. 2014). En el modelo de Moran cada evento de extinción es seguido de un evento de especiación en cualquier linaje sobreviviente, por lo que también se conoce como “modelo de recambio”. Trasladado a un fenómeno macroevolutivo, este proceso comienza con una tasa de diversificación positiva, en el que la diversidad se acumula exponencialmente. Una vez alcanzado el límite a la diversidad, la tasa de extinción aumenta para hacerse equivalente o superior a la tasa de especiación, resultando en una tasa de diversificación nula o negativa, y en un estancamiento en la diversidad (Raup et al. 1973; Gould et al. 1977; Sepkoski 1978; Sepkoski y Kendrick 1993). La implementación de pruebas de ajuste de modelos en filogenias moleculares, ha revelado que una fracción no tan insignificante de éstas (~23%) presenta evidencia de tasas de diversificación en equilibrio (Morlon et al. 2010).

Los modelos dependientes de la ecología son variantes del modelo BD que proponen que las tasas de diversificación dependen de parámetros ecológicos como el área de distribución, el número de nichos disponibles o el número de especies existentes en un determinado tiempo (Rabosky y Lovette 2008a; Rabosky 2009; Etienne et al. 2012). Estos modelos se han ajustado con éxito a niveles taxonómicos bajos en filogenias de los géneros *Dendroica* (aves Rabosky y Lovette 2008a; Etienne et al. 2012), *Plethodon* (reptiles) y *Heliconius* (mariposas), y a nivel taxonómico más alto en la familia de los cetáceos, en el filo de los foraminíferos planctónicos (Etienne et al. 2012) y en la clase de las aves (Nee et al. 1992).

Un problema con las pruebas de ajuste de modelos es que no pueden asegurar que el mejor modelo encontrado sea el modelo verdadero, i.e., el modelo que generó la diversidad observada. Además, dada una serie de modelos por evaluar, estas pruebas necesariamente identifican al menos a uno (pueden identificar varios) de entre todos los modelos considerados como el mejor. Si el modelo verdadero no está en el conjunto de modelos evaluados, entonces el mejor modelo nunca corresponderá con el verdadero. Una evidencia independiente de la filogenia son los estadísticos sumarios o descriptivos, los cuales nos pueden ayudar a distinguir entre hipótesis alternativas de diversificación cuando carecemos de filogenias, éstas son muy incompletas o los métodos de ajuste de modelos no son concluyentes. Para utilizar los estadísticos descriptivos, es necesario implementar simulaciones para determinar su comportamiento esperado bajo diferentes modelos de diversificación.

Estadísticos descriptivos: un complemento de los métodos para reconstruir el proceso de diversificación

Uno de los primeros estadísticos descriptivos que fueron utilizados para describir la naturaleza del proceso de diversificación fue la forma de los clados, definida con el centro de gravedad de los patrones de diversidad observados en el registro fósil, el cual sugirió que hay límites a la diversidad pero que estos no están determinados ecológica o determinísticamente, sino que ocurren al azar (Gould et al. 1977). Más recientemente, otro estudio que utilizó estadísticos descriptivos como la riqueza máxima del clado, la desviación estándar de la diversidad, y la relación entre la edad y la riqueza de especies de los clados, mostraron en familias y subfamilias de serpientes que la extinción es el

factor que limita la diversidad (Pyron y Burbrink 2012). En este estudio se utilizó el "modelo de vida media de los clados" (modelo hL, por sus siglas en inglés *halfLife of clades*), otra variante del modelo BD, que describe un proceso que empieza con una tasa de diversificación positiva que se mantiene hasta alcanzar un límite después del cual se hace negativa. El proceso descrito por el modelo hL genera un patrón muy similar al descrito por el modelo utilizado en las simulaciones de Gould et al. (1977), el cual muestra no sólo la interrupción de la acumulación de especies, sino una disminución de la diversidad (Pyron y Burbrink 2012; Fig. 2b).

Un estadístico descriptivo popular: la relación entre la edad de los clados y su riqueza de especies

La correlación entre la edad de los clados y su riqueza de especies (ARC por sus siglas en ingles *Age-Richness Correlation*) es un estadístico descriptivo particularmente útil para estudiar la diversificación, porque permite medir directamente el efecto del tiempo –medido como la edad del clado– sobre la diversidad. Intuitivamente, una ARC positiva indica que la edad de los clados determina su diversidad: la variación en la edad de los clados explica las variaciones en la diversidad observadas entre clados. Una ARC no significativa o negativa indica que otros factores diferentes del tiempo explican los patrones de diversidad. Por ejemplo, una ARC negativa puede surgir si diferencias en las tasas de diversificación, y no en el tiempo, determinan la variación en la diversidad entre clados (Ricklefs y Renner 1994; Magallón y Sanderson 2001).

Evidencia más reciente a favor de la hipótesis de límites a la diversificación proviene de la ARC obtenida en grupos determinados filogenéticamente. Implementando simulaciones predictivas *a posteriori* (PAPS, por sus siglas en inglés *Predictive A*

Posteriori Simulations), que usan parámetros de diversificación estimados de datos empíricos para simular la diversidad y obtener valores esperados de estadísticos descriptivos de interés bajo diferentes modelos de diversificación, se encontró que la ARC debe ser inexistente o negativa bajo modelos de diversidad limitada y positiva bajo modelos de diversidad ilimitada (Rabosky 2009; Pyron y Burbrink 2012).

Estudios posteriores han encontrado que cuando las ARCs son estimadas utilizando la edades troncales y una clasificación taxonómica basada en distancia temporal o genética, las ARCs pueden ser nulas o negativas en procesos de diversificación ilimitada (Stadler et al. 2014). En el mismo estudio no se detectó un patrón similar en ARCs estimadas con edades corona, revelando la existencia de un conflicto entre los estimados de ARC obtenidos usando la edad corona (ARC corona) y los obtenidos usando la edad troncal (ARC troncal). En este trabajo también fue sugerido que sólo las ARC corona podían usarse como evidencia del proceso de diversificación, ya que estas no estaban afectadas por las clasificaciones taxonómicas. Sin embargo, esta aseveración no ha sido evaluada explícitamente.

¿Qué tan informativa sobre el proceso de diversificación es la relación entre la edad de los clados y la riqueza de especies?

Ya que el conflicto ARC corona/ARC troncal se ha demostrado solamente en ciertos tipos de delimitaciones taxonómicas poco implementadas, nos preguntamos si se manifestará también en delimitaciones taxonómicas más comúnmente utilizadas, como la clasificación jerárquica por rangos taxonómicos, la cual se basa no sólo en características descriptivas de los organismos sino también en la información de sus relaciones

filogenéticas. En este caso, diferentes rangos taxonómicos podrían afectar los estimados de ARC de maneras distintas o incluso dar evidencia inconsistente sobre la relación entre la edad de los clados y su riqueza de especies en un mismo sistema biológico. De ser así, ¿habrá rangos taxonómicos más informativos que otros? Por ejemplo, ¿podría el nivel de género, que ha sido descrito como un rango taxonómico natural (porque surge de procesos evolutivos, Humphreys y Barraclough 2014), ser más informativo que rangos taxonómicos más inclusivos, como órdenes?

Adicionalmente, se ha encontrado evidencia de que los datos de edad y riqueza de especies podrían estar afectados por una historia evolutiva compartida y presentar autocorrelación (estructura filogenética; Pyron y Burbrink 2012; Rabosky et al. 2012), la cual debe ser tomada en cuenta para evitar falsos positivos que, en el caso de las ARCs, corresponden a una ARC positiva significativa cuando en realidad es inexistente o negativa (Rabosky et al. 2012). Sin embargo, la presencia de estructura filogenética en datos de edad y riqueza de especies no ha sido probada rigurosamente. Por una parte, la estructura filogenética ha sido evaluada principalmente en filogenias provenientes del árbol de tiempo de la vida (Hedges y Kumar 2009). Estos árboles están construidos con un método que genera árboles con politomías, las cuales son resueltas al azar, muy probablemente alterando la señal filogenética. Por otra parte, los pocos casos en los que la estructura filogenética de los datos de edad y riqueza de especies de los clados ha sido evaluada en árboles obtenidos con métodos tradicionales, ésta ha sido estimada con el método de contrastes filogenético independientes (Felsenstein 1985), método que podría no ser adecuado para datos de edad y riqueza de especies (Agapow y Isaac 2002; Isaac et al. 2003). En este sentido, ¿los datos empíricos de edad de los clados y riqueza de

especies están estructurados filogenéticamente? Y, si es así, ¿cómo es afectada la ARC por la estructura filogenética?

En este contexto el Objetivo Principal de esta tesis es:

Establecer la condición de la relación entre le edad y la riqueza de especies en la naturaleza y determinar el efecto de la estructura filogenética, el tipo de edad (corona o troncal) y la categoría taxonómica sobre ella.

Los Objetivos Particulares son:

1. Documentar la relación entre la edad de los clados y su riqueza de especies, usando datos provenientes de estudios empíricos, en un rango amplio de sistemas biológicos, evaluando estructura filogenética, y considerando los dos diferentes tipos de edad y diferentes rangos taxonómicos.
2. Determinar los parámetros de diversificación (especiación y extinción) que generaron los patrones de diversidad en los sistemas biológicos de estudio elegidos.
3. Desarrollar e implementar un modelo de simulación adecuado para establecer los procesos de diversificación que generaron las relaciones entre edad y riqueza de los clados documentadas empíricamente, considerando estructura filogenética, tipo de edad y rango taxonómico.

Para el primer objetivo particular, se realizó un estudio de fechamiento de las familias de plantas con flor (Angiospermae) utilizando diferentes métodos de reloj molecular, presentado en el Apéndice 1. Este trabajo representa el muestreo más exhaustivo de linajes de plantas con flor hasta la fecha, y su realización amplió la muestra

de sistemas biológicos en los que se documentó de forma empírica la relación entre la edad de los clados y la riqueza de especies.

Posteriormente, se realizó una serie de estudios en los que se analizó la dinámica de diversificación de diferentes sistemas biológicos. La intención de estos estudios era estimar los parámetros empíricos de diversificación de los sistemas biológicos para implementar simulaciones utilizando los parámetros obtenidos de los datos empíricos (PAPS), y así obtener distribuciones esperadas del estadístico descriptivo a evaluar –la relación entre la edad y la riqueza de los clados (ARC)– que fueran directamente comparables con los estimados empíricos de ARC. Al final no fue factible usar este método de simulación porque:

- actualmente, carecemos de métodos para estimar parámetros de diversificación de algunos de los modelos que necesitábamos evaluar (i.e., variación de las tasas entre clados con cambio dependiente de ecología a lo largo del tiempo), y
- sólo existe un método para estimar el parámetro de tasa de origen de taxa superiores, el cual asume un modelo de diversificación de tasas constantes entre clados y en el tiempo (Maruvka et al. 2013).

Adicionalmente, el error asociado a los valores estimados de parámetros de diversificación podría obscurecer la determinación de la causa principal de un desajuste entre las relaciones esperadas y las observadas. Por tal razón, se decidió usar un rango de valores de diversificación para implementar las simulaciones, y así poder explorar el comportamiento de las ARCs de manera general.

Algunos de los resultados de la implementación de estos análisis de diversificación constituyen el primero y segundo capítulo de esta tesis, en los cuales se

estudió respectivamente la diversificación de las familias de plantas con flor (Angiospermae) y de una familia endémica de los desiertos norteamericanos, los ocotillos (Fouquieriaceae, Ericales). Estos capítulos quedan como un testimonio de lo que se puede hacer con métodos de inferencia de modelos.

En el tercer capítulo se presentan los resultados de las relaciones entre edad (obtenida con métodos de reloj molecular) y riqueza de especies en diferentes rangos taxonómicos de cinco sistemas biológicos (anfibios, reptiles escamados, mamíferos, aves y plantas con flor), considerando la edad desde el tiempo troncal y desde el tiempo corona –las cuales pueden ser sustancialmente distintas (Fig. 1)– y estimando las correlaciones con métodos que toman en cuenta la estructura filogenética y métodos que no la toman en cuenta. En ese capítulo también se presentan las relaciones edad-riqueza simuladas con un modelo de origen jerárquico de taxa superiores propuesto en esta tesis, y esperadas bajo diferentes modelos de diversificación de diversidad limitada e ilimitada.

Finalmente, se presenta una comparación informal entre los patrones de ARC esperados en los diferentes modelos evaluados y los observados en las filogenias empíricas, y se discuten los resultados más relevantes.

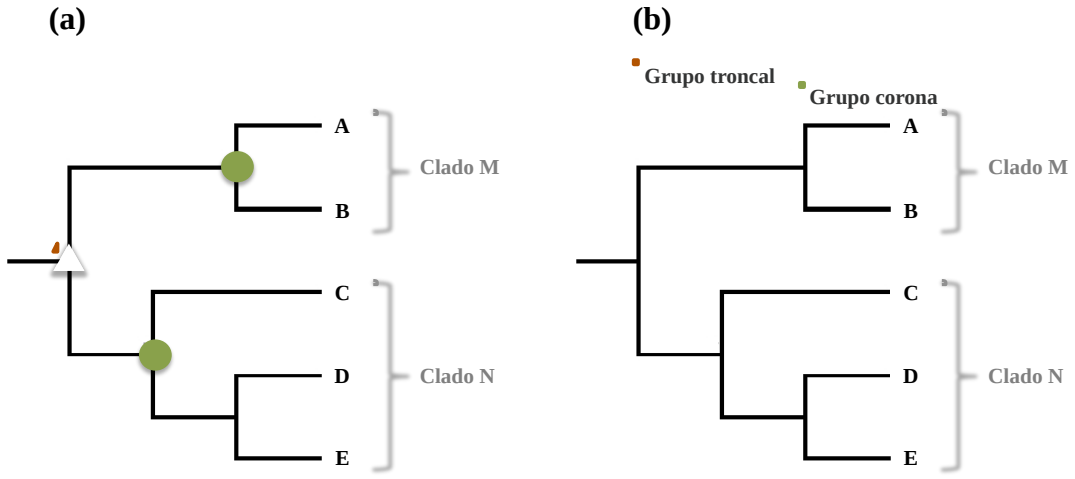


Figura 1. Edad corona y edad troncal de un clado. (a) La edad troncal de un clado (triángulo vacío) corresponde al tiempo del último ancestro común con su clado hermano. La edad corona (círculo sólido) corresponde a la edad del último ancestro común a todas las especies vivientes de un clado. Dos clados hermanos, M y N en este caso, siempre comparten la misma edad troncal pero no necesariamente la misma edad corona. (b) Estas edades sirven para definir a los grupos troncal y corona respectivamente.

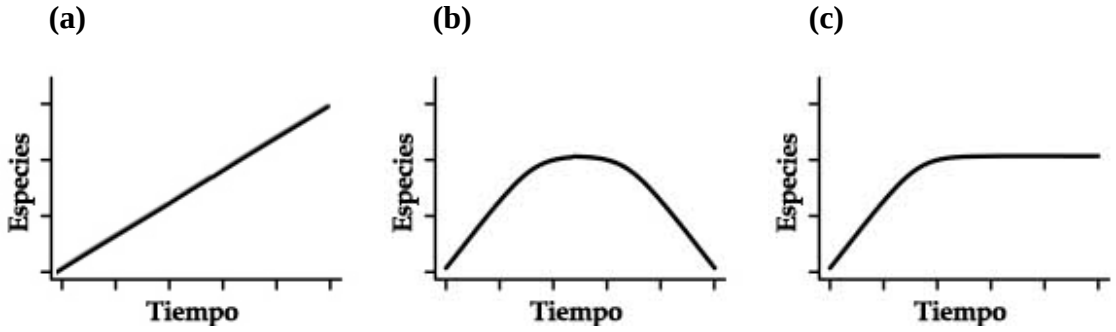


Figura 2. Modelos de cambio en la diversidad a lo largo del tiempo. La diversidad puede acumularse ilimitadamente (a) o tener un límite a la diversidad (b-c). Una vez alcanzado el límite, la diversidad puede disminuir (b) o mantenerse constante (c).

CAPÍTULO 1

CAMBIOS MACROEVOLUTIVOS EN LA DIVERSIFICACIÓN REVELAN UNA COMPLEJIDAD DE RUTAS HACIA LA MEGADIVERSIDAD DE LAS ANGIOSPERMAS

Manuscrito con formato para Nature

Macroevolutionary Diversification Shifts Reveal Complex Routes to Angiosperm Megadiversity

Susana Magallón[1], Luna L. Sánchez-Reyes[2], and Sandra L. Gómez-Acevedo[1]

[1] Instituto de Biología, Universidad Nacional Autónoma de México, Mexico City, México

[2] Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México, Mexico City, México

Summary Paragraph

Angiosperms, the last major lineage of plants to evolve, represent an unprecedented evolutionary success. As they rose to become one of the megadiverse groups of macroscopic eukaryotes, they forged modern ecosystems and promoted the evolution of extant terrestrial biota. The unequal distribution of species among lineages within angiosperms suggests that the process of diversification, which ultimately determines species richness, has acted differentially through their evolution. Here we investigate how angiosperms became megadiverse by identifying the placement of exceptional radiations, combining a densely fossil-calibrated molecular-clock phylogeny[1] with a model that identifies diversification shifts among evolutionary lineages and through time[2]. We find that angiosperm diversification shifts initiated after significant phylogenetic differentiation and morphological elaboration had occurred; took place over

100 million years (Ma) between the Early Cretaceous and the Neogene, and; are phylogenetically dispersed. Angiosperm long-term diversification trajectory is nearly constant, but is underlain by increasing speciation and extinction, and results from independent, temporally overlapping radiations and depletions in component lineages. This unmitigated diversification trajectory implies continued species accumulation, albeit increasing turnover. We discovered that major lineages within angiosperms became megadiverse through fundamentally different diversification modes, including clades with an early diversification increase shared by its evolutionary descendants, to clades encompassing several independent and unlinked radiations. The identified diversification shifts and the macroevolutionary modes they reveal provide a framework to rigorously investigate interactions among intrinsic attributes, biotic and physical variables that have driven independent radiations, and that combined, underlie angiosperm's evolutionary success.

Main Text

Flowering plants (Angiospermae) are an extraordinary evolutionary phenomenon. They represent the most recent radiation of embryophytes, a lineage that occupied land at least 470 million years ago (Mya)[3] and diverged from their closest living relatives 300-350 Mya. Since their appearance in the fossil record, angiosperms diversified exceptionally, surpassing all other plants not only in sheer species richness, but to become ecologically predominant forming the structural and energetic basis of nearly all extant terrestrial biomes. Through their evolution and ecological expansion, angiosperms promoted the diversification of other plants[4], animals[5], fungi[6,] and bacteria[7]. Human nutrition, culture and well-being inextricably depend on angiosperms.

With ca. 280,000 species described and an estimated total of 300-350,000 species[8, 9], angiosperms are among the megadiverse groups of macroscopic eukaryotes. This exceptional diversity is distributed unequally among evolutionary lines, some including tens of thousands of species (e.g., orchids, composites) and others fewer than ten (e.g., lotus, London planes), indicating that the process of diversification, the balance between speciation and extinction which determines species richness[10] (hereafter diversity), has acted differentially through angiosperm evolution.

Many studies have attempted to identify the factors that underlie angiosperm exceptional megadiversity, including intrinsic attributes[11], ecological interactions[12], extrinsic opportunity[13] or complex interactions among them[14]. Nevertheless, little is known about the diversification mechanisms that underlie the acquisition of angiosperm megadiversity through time, and its unequal distribution among phylogenetic branches. A long tradition considered that the myriad of angiosperm unique vegetative and reproductive attributes made them competitively superior[15]. Studies based on current phylogenetic understanding have shown that it is unlikely that increased phylogenetic branching characterizes angiosperms as a whole[16]. Different groups with unexpectedly high or low diversity, given a time-homogeneous diversification rate, have been identified[17]; and diversification tests based on tree asymmetry or model selection have found that diversification shifts do not always correspond to named taxonomic entities[18], and that some occurred downstream major genomic duplications[19].

The fossil record unequivocally documents an explosive angiosperm radiation in the Early Cretaceous, shortly following the appearance between the Valanginian and the Hauterivian (140-130 Mya) of pollen grains with detailed microstructural attributes found

only in some angiosperms[20]. Immediately younger sediments contain a drastically increasing diversity of pollen types, and vegetative and reproductive remains representing the earliest-diverging angiosperm branches[21] and their major evolutionary lineages[22, 23].

In this study we investigate the dynamics of angiosperm macroevolutionary diversification using a molecular Bayesian phylogenetic tree in which ca. 90% of angiosperm families are represented. This tree was dated[1] with a relaxed molecular clock[24] calibrated with 136 critically selected fossil-derived constraints, and a confidence interval on the crown node age, derived from a quantitative paleobiology method[25]. Using this comprehensive time tree, we applied a method[2] that, through a Compound Poisson Process, identifies major shifts in the rate of diversification among phylogenetic branches and through time (see Methods). The identified macroevolutionary shifts inform about the origins of angiosperm megadiversity and its long-term diversification trajectory. For the first time, we document the mechanisms underlying the acquisition of megadiversity, and discover that different macroevolutionary modes predominate in distinct phylogenetic regions.

The onset of angiosperm diversification into extant representatives in the Early Cretaceous was soon followed by the differentiation of its main evolutionary lines, including the Monocotlyedoneae (monocots) and Eudicotyledoneae (eudicots). During their earliest evolution, each evolutionary line acquired distinctive morphological groundplans. Monocots are characterized by a distinctive stem anatomy, 3-parted flowers, and embryos with a single cotyledon. Eudicots are characterized by pollen grains with three longitudinal apertures; and a vast clade within them, Pentapetalae, is characterized by 5-parted flowers with distinct calyx and corolla. The earliest major

diversification shifts within angiosperms took place after its main evolutionary lines were differentiated, and after the establishment of their respective morphological groundplans. These shifts are associated to three clades that together contain the overwhelming majority of angiosperm's diversity, morphological variety, and ecological breadth. The first and second shifts took place within Pentapetalae, subtending Superasteridae[26], and a clade that includes Rosidae[27] plus Vitales (hereafter Rosidae+), which respectively contain 40.4% and 29.5% of extant angiosperm diversity. The third shift took place within monocots, subtending a clade that includes Dioscoreales, Pandanales, Liliales, Asparagales, and the large Commelinidae[27]. This unnamed clade, here referred to as Dioscoreidae, includes 21.7% of angiosperm diversity. These three diversification shifts took place during the Early Cretaceous, between 123.7 and 119.8 Ma (Fig. 1). Most subsequent major diversification shifts are nested within Superasteridae, Rosidae+ or Dioscoreidae; they lack any phylogenetic sorting, and took place between the Cenomanian (Late Cretaceous) and the Miocene (Neogene; Fig. 1, Extended Data Figures 1-2, Extended Data Table 1). The temporal distribution and phylogenetic placement of major diversification shifts indicates that angiosperm megadiversity cannot be traced to a few radiations or to a restricted time interval, but rather, results from many independent diversification increases through the temporal extent of its evolutionary history and across its phylogenetic spectrum.

The temporal trajectory of angiosperm diversification lacks pronounced increases or decreases (Fig. 2). During approximately the first half of their history, angiosperms underwent a moderate diversification decrease, but their subsequent long-term trajectory remained flat or increased slightly. Angiosperm speciation and extinction rates underwent

early decreases, followed by increasing trends towards the present (Fig. 2). The observed sustained diversification trajectory implies that angiosperm diversity will continue to accumulate, but with a higher species turnover as a consequence of the incremental rates of speciation and extinction. Lineages within angiosperms exhibit variable diversification trajectories revealing independent radiations and depletions that overlap over different extents of their duration, masking each other (Fig. 2).

We document for the first time that different phylogenetic regions within angiosperms acquired megadiversity through fundamentally different macroevolutionary mechanisms. Superasteridae, Rosidae+ and Dioscoreidae display distinct evolutionary tendencies in their vegetative and reproductive morphology, and in their interactions with the environment. We found that they also differ in the way in which they became megadiverse. Each underwent an early diversification increase (Fig. 1), but with different ultimate consequences on their respective extant diversity. In Dioscoreidae, the early increase affected diversification patterns of its evolutionary descendants, and almost all of its diversity (>80%) -- including emblematically rich groups such as orchids, palms and bromeliads -- resulted from this ancestral (i.e., plesiomorphic) shift. Only one, relatively late shift, uniquely characterizes Poaceae (grasses). In sharp contrast, the early shift in Superasteridae played a marginal role generating the extant megadiversity of this clade. Rather, ca. 75% of Superasteridae derives from five nested drastic increases, which gave rise to Lamiidae pp, Dipsacales, Apiales pp, Asteraceae+ and Campanulaceae (Fig. 1). Without these five independent and derived (i.e., autapomorphic) increases, Superasteridae would be a rather small clade. In Rosidae+, megadiversity was acquired by combining the contribution of an early plesiomorphic shift, which gave rise to 62% of

the diversity in the clade, with four independent, nested autapomorphic radiations corresponding to Fabaceae, Euphorbiaceae, Celastraceae and Malvaceae, which account for the remainder of Rosidae+ diversity (Fig. 1).

Angiosperms are today the most prominent group of plants in terrestrial biomes, where in addition to exceptional diversity, they display astounding morphological, functional and ecological complexity and innovation. Yet, here we show that their evolutionary expansion remains unmitigated. Our results cannot indicate if overall angiosperm diversification is limited (e.g., by time, area, diversity or ecology[28]), but suggest that, if limits to angiosperm diversification exist, those bounds have not yet been reached.

Our analyses detected few diversification decreases, but it seems unlikely that this reflects the paucity of lineages in decline. Rather, as the natural outcome of decreasing diversification is extinction, lineages undergoing an evolutionary depletion persist shortly. Detection of recent diversification decreases is more likely, appearing as depauperate lineages on their way to extinction. Plocospermataceae, the sister group to the remainder of Lamiales, is a possible example. Lineages that underwent an ancient diversification decrease but survive to the present are unexpected and difficult to explain[29]. These lineages might be in decline from former megadiversity, and their ultimate demise is taking longer, or they might have recovered after a drastic decrease. Montiniaceae+ (Figs. 1-2) is a possible example. Our diversification analysis provides elements to differentiate depauperate lineages that owe their low diversity to decreasing diversification from former higher richness, from lineages that have maintained a low diversification rate through their history.

The finding that the earliest increases in angiosperm diversification took place after significant morphological elaboration which underlies the vegetative and reproductive groundplans of monocots and pentapetalous eudicots, but before their increase in morphological and functional variety, and their ecological expansion, suggests that angiosperm diversification is not a direct consequence of the numerous attributes that collectively distinguish this group from other plants[16, 17]. Rather, the placement of the earliest and subsequent major diversification shifts are congruent with a scenario of incremental acquisition of complexity[30], in which fundamental attributes, which represent the basis of integrated structures and functions, evolved in phylogenetically early branches, but did not lead directly to a substantial diversification increase. Nevertheless, these early attributes were *sine qua non* precursors for subsequent evolution of complex structures and functions, which possibly underwent a piecemeal integration on different phylogenetic branches[30], and more immediately underlie nested diversification increases. Specifically, we propose that diversification increases subtending Superasteridae, Rosidae+ and Dioscoreidae responded to the evolutionary assembly of major groundplans within angiosperms. In turn, each represented a fundamental precursor to the evolution of more specialised structures and functions, canalised along different phylogenetic branches, and which could have more immediately detonated nested radiations.

Identifying the phylogenetic and temporal placement of macroevolutionary diversification shifts considerably circumscribes the investigation of possible causes and ultimate drivers of megadiversity, which most likely result from the combination of intrinsic attributes, ecological functions and abiotic conditions. A further improved understanding of angiosperm radiations, and of the variables and factors that drive them,

should necessarily be based on a more precise phylogenetic location of incremental and decremental diversification shifts, derived from a much denser taxonomic sampling, and ideally, direct integration of fossil information.

Figures and Figure Legends

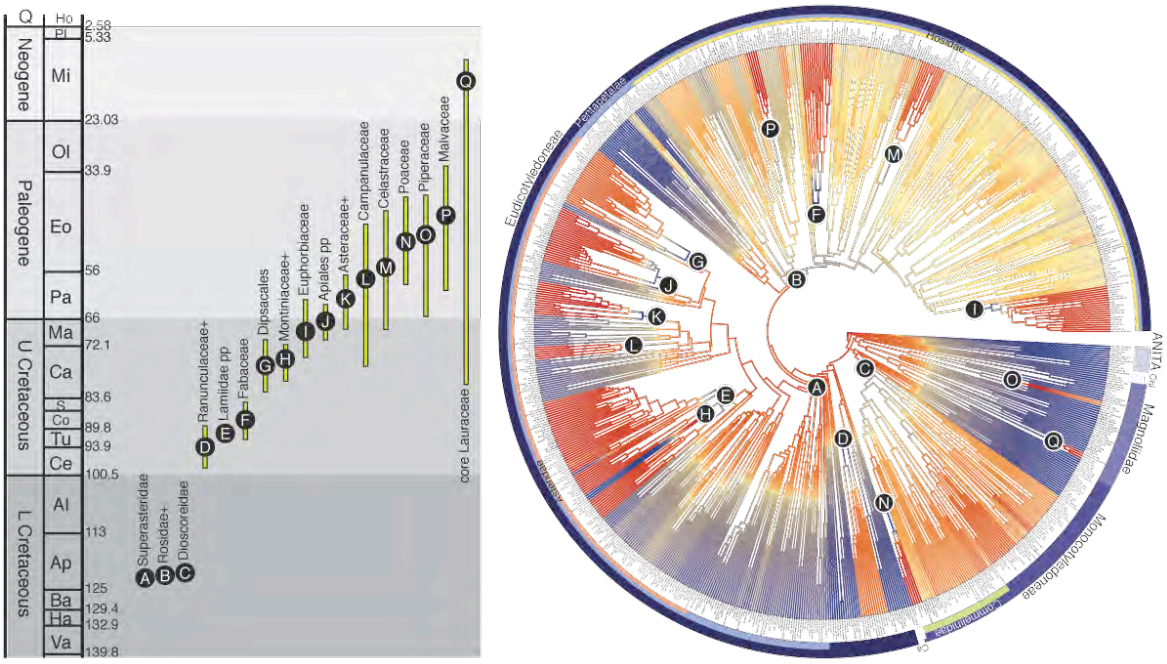
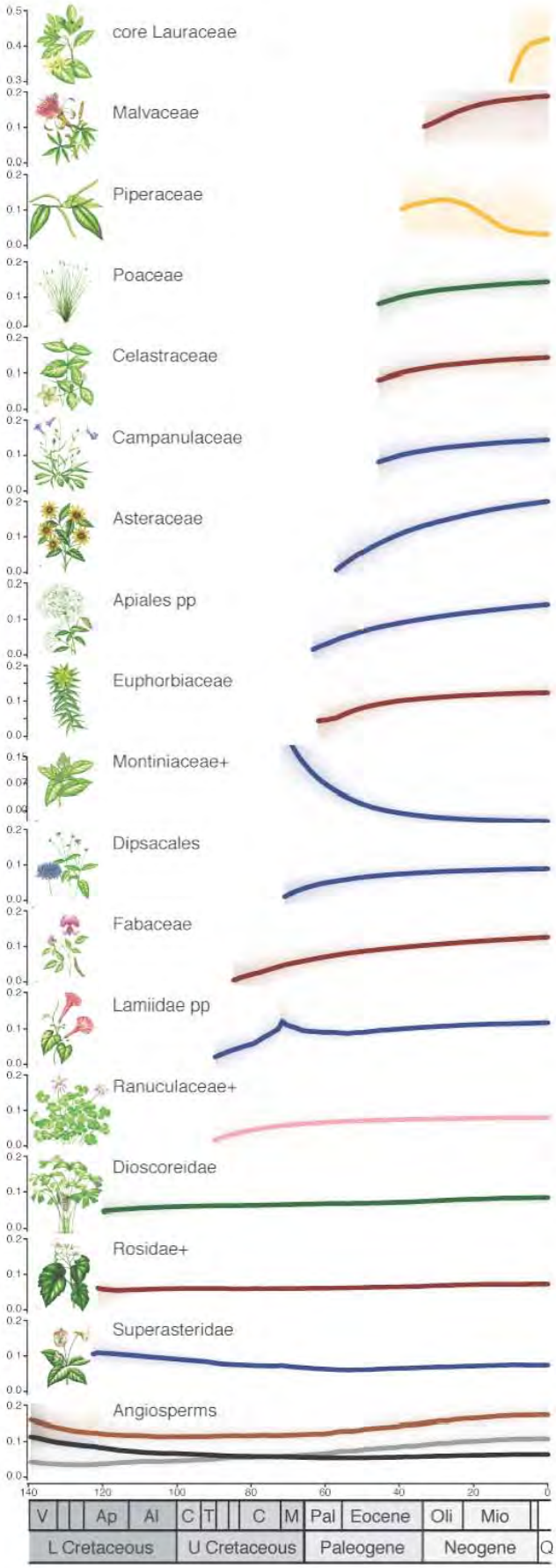


Figure 1. Average diversification rate and major diversification shifts on angiosperm phylogenetic tree. The phylorate plot (right) shows the mean, model-averaged diversification rate at each 3 Ma interval on every branch of the time-calibrated angiosperm phylogeny, model-averaged across all shift configurations in the posterior distribution estimated in the macroevolutionary diversification analysis[2] (cool colours = low rates; warm colours = high rates). Circles around the phylorate plot indicate major angiosperm groups (Ch = Chloranthales, Ce = Ceratophyllales). Diversification shifts identified in ≥50% of the configurations in the posterior distribution are indicated as black circles on the phylorate plot. These shifts correspond almost exactly to those found in the configuration with highest posterior probability, and in the configuration that maximizes the marginal probability of rate shifts along individual branches (see Extended Data Figures 1-2). The temporal distribution of each shift (left), including its range (green bar) and mean age (black circle), were summarized from the configurations in the

posterior probability. Collectively, rate shifts range from the Lower Cretaceous to the Neogene. The three earliest shifts took place in the Lower Cretaceous, subtending (A) Superasteridae sensu Soltis et al. (2011)[26] (in Pentapetalae, including Dilleniales, Santalales, Berberidopsidales, Caryophyllales and Asteridae); (B) Rosidae+ (in Pentapetalae, including Vitales and Rosidae); and (C) Dioscoreidae (in monocots, including Dioscoreales, Liliales, Asparagales, and Commelinidae). Diversifications shifts in the Upper Cretaceous subtend (D) Ranuculaceae+ (in eudicots, including Ranunculaceae, Berberidaceae and Menispermaceae; e.g., columbine, hellebores) with median age in the Cenomanian; (E) Lamiidae *pro parte* (pp, in Superasteridae, including Gentianales, Solanales, Boraginales and Lamiales; e.g., gentian, tomato, mint) with median age in the Turonian; (F) Fabaceae (legumes, in Rosidae+) with median age in the Coniacian; (G) Dipsacales (in Superasteridae; e.g., valerian) and (H) Montiniaceae+ (in Superasteridae, including Montiniaceae, Sphenocleaceae and Hydroelaceae) with median ages in the Campanian; and (I) Euphorbiaceae (spurges, in Rosidae+) and (J) Apiales pp (in Superasteridae, including Pittosporaceae, Araliaceae, Myodocarpaceae and Apiaceae; e.g., ivy, carrot) with median ages in the Maastrichtian. Paleogene diversifications subtend (K) Asteraceae+ (in Superasteridae, including Goodeniaceae, Calyceraceae and Asteraceae; e.g., composites), and (L) Campanulaceae (in Superasteridae; e.g., lobelia) with median ages in the Paleocene; and (M) Celastraceae (in Rosidae+; e.g., staff vine), (N) Poaceae (grasses, in Dioscoreidae), (O) Piperaceae (in Magnoliidae, e.g., black pepper) and (P) Malvaceae (in Rosidae+; e.g., cotton), with median ages in the Eocene. The last diversification shift took place during the Neogene, subtending (Q) core

Lauraceae (in Magnoliidae; e.g., laurel) with median age in the Miocene. Unless otherwise indicated, clade names are based on Cantino et al. 2007[27] and Stevens 2013[8].


core Lauraceae
Malvaceae
Piperaceae
Poaceae
Celastraceae
Campanulaceae
Asteraceae
Apiales pp
Euphorbiaceae
Montiniaceae+
Dipsacales
Fabaceae
Lamiidae pp
Ranuculaceae+
Dioscoreidae
Rosidae+
Superasteridae
Angiosperms
140
120
100
80
60
40
20
0
V
Ap
Al
C
T
C
M
Pal
Eocene
Oli
Mio
L Cretaceous
U Cretaceous
Paleogene
Neogene
Q

226

Figure 2. Diversification through time (DTT) plots. Each graph indicates the mean diversification rate of a clade averaged across all samples in the posterior distribution, and the 95% highest posterior density. The bottom pannel shows the diversification (black), speciation (brown) and extinction (gray) through time plots of angiosperms as a whole. Angiosperm diversification underwent a slight early decrease, followed by a nearly constant or very slightly increasing trajectory towards the present. Speciation and extinction show initial decreases, followed by increasing trends towards the present. The remaining pannels show DTT plots of clades (indicated in each pannel, defined in Figure 1) resulting from each of the diversification shifts identified in >50% of the configurations in the posterior density of the macroevoultionary diversification analysis[2], in chronological sequence (oldest at the bottom). Collectively, these DTT plots show a variety of increasing and decreasing trajectories that started at different times and mask each other. Plots in blue are nested in Superasteridae; red, in Rosidae+; green, in Dioscoreidae; pink, in eudicots (outside Pentapetalae); and yellow, in Magnoliidae.

Methods

Taxonomic sampling, molecular data and phylogenetic analyses

The diversification study is based on a previously published, temporally-calibrated phylogenetic tree of angiosperms[1]. The taxonomic sample includes 792 angiosperms; six gymnosperms representing Cycadophyta, Gnetophyta and Coniferae; and a fern belonging to Ophioglossaceae. The angiosperms belong to 374 families, representing 87% of those recognized by the Angiosperm Phylogeny Website[8] in April 2013, and encompassing 99% of angiosperm total species-richness. The molecular data are the concatenated sequences of three plastid protein-coding genes (*atpB*, *rbcL* and *matK*) and two nuclear markers (18S and 26S nuclear ribosomal DNA), which together form an alignment of 9089 base pairs (bp). We attempted to maximize data completeness by including sequences of species of the same genus when sequences of the same species were not available. When unavailable at the genus level, the sequence was left as missing data. Separate alignments of the sequences of different markers were conducted with MUSCLE v3.7[31], followed by manual refinements with BIOEDIT v7.0.9.0[32]. The molecular data set is available in the DRYAD Digital Repository (http://dx.doi.org/10.____/dryad._____).

Phylogenetic estimation was conducted with maximum likelihood (ML) in RAxML v7.2.8[33]. The data was divided into four partitions: first and second codon positions of *atpB* plus *rbcL*; third codon positions of *atpB* plus *rbcL*; *matK;* and 18S plus 26S. Substitution parameters were estimated independently for each, using unlinked GTRCAT models. Topological constraints were implemented to specify phylogenetic relationships among major clades derived from an analysis based on a larger data

matrix[26]. Trees were rooted by using the fern *Ophioglossum* as outgroup. One thousand bootstrap replicates were implemented[34, 35].

Dating analyses

The ML phylogenetic tree was dated by combining a method derived from quantitative paleobiology to constrain the age of the angiosperm crown node[25], with an uncorrelated relaxed molecular clock to estimate dates within angiosperms[24]. Whereas the angiosperm fossil record suggests the onset of a rapid radiation of angiosperms in the Lower Cretaceous, previous molecular clock studies provide contradictory estimates of the age of their crown node[1]. The increasing diversity, abundance and geographical distribution of angiosperm fossils in Lower Cretaceous geological sequences, the order of appearance of vegetative and reproductive morphological types in the fossil record, and the agreement between the order of appearance of major clades in stratigraphic sequences and their branching position in molecular phylogenetic trees together indicate that, as a whole, the fossil record provides reliable information about the timing of crown angiosperm origin and diversification. Relaxed clocks are powerful tools to estimate divergence times using molecular phylogenies, but, as with any model-based method, model misspecification will derive in incorrect estimates. It has been shown that relaxed clocks can substantially overestimate node ages when the amount of molecular substitution rate heterogeneity is insufficiently accounted for[36, 37]. A recent study evaluating estimates of angiosperm Triassic (and older) ages[38] suggests that these ages may be a consequence of complex interactions among substantial rate heterogeneity around the onset the angiosperm crown group diversification, and the amount and choice of sampled taxa.

Given the observed uncertain molecular estimates of crown angiosperm age, we calculated a confidence interval on this age based on its fossil record[39]. To calculate this confidence interval, we implemented a method that aims to date a molecular phylogenetic tree with an absolute time scale extrapolated from a confidence interval of the age of the lineage in the tree that has the most temporally-complete fossil record[25]. This method consists of three steps: (1) identifying the calibration lineage, this is, the phylogenetic branch with the greatest overlap between its oldest fossil age and its length in an ultrametric tree estimated without fossil information; (2) calculating a confidence interval that contains the true age of the calibration lineage; and (3) date the ultrametric tree with an absolute time scale derived from the confidence interval calculated in the previous step. Because our goal was to obtain an interval around the age of crown angiosperms, we only applied the second step of this method. The minimal (youngest) limit of the confidence interval is directly taken from the age of the oldest fossils of the lineage, which, for angiosperms, are morphologically and ultrastructurally distinctive pollen grains from Valanginian to Hauteriviain-age sediments (Early Cretaceous). We used the age of the Valanginian-Hauterivian boundary (136 Ma[40, 41, 42, 43]) as the minimal age of the interval (FA_{cal}). The maximal (oldest) limit of the interval (FA_c) is calculated with Equation 14 in Marshall, 2008[25],

$$FAc = \frac{FAcal}{\sqrt[nH]{1-C}},$$

where FA_{cal} is the minimum age of the interval (above), n is the number of branches in the phylogenetic tree represented in the fossil record, H is the average number of fossil localities from which each branch represented in the fossil record is known, and C is the desired level of confidence[25]. To calculate n, we pruned the phylogenetic tree estimated

with RAxML (above) to leave a single terminal per angiosperm family, for a total of 374 terminals. This tree was made ultrametric with the uncalibrated lognormal method in BEAST v1.7.5[24], assigning 100 time units to the angiosperm crown node. The BEAST analysis consisted of four independent Markov chain Monte Carlo (MCMC) runs for a total of 65 x 10^6 steps, sampling one every 5000 steps. Convergence of the MCMCs was evaluated with Tracer v1.5[44] and sampled trees and estimated parameters were summarized with LogCombiner v1.7.5 and TreeAnnotator v1.7.5[24]. Through examination the primary literature, we compiled a fossil data base of angiosperm fossils, sorted by their affinity to phylogenetically-recognized families. Using this data base – which includes >3500 fossil entries with geographical, geological and stratigraphical information – we identified the angiosperm families that are reliably represented in the fossil record. As a result, n = 123 branches in the ultrametric tree described above were identified as being represented in the fossil record. The number of localities from which each lineage represented in the fossil record is known is difficult to obtain, hence, estimating H is not trivial. Assuming that each lineage represented in the fossil record is known from a single locality (H = 1) represents a conservative baseline for this parameter[25], as any higher magnitude for H would lead to younger age estimates for the maximum age of the confidence interval. The interval was estimated at a 95% level of confidence (C = 0.95). The calculated 95% confidence interval for crown angiosperm age ranged from 136 to 139.35 Ma[1].

Ages within angiosperms were estimated with the uncorrelated lognormal method in BEAST v1.7.5[24]. The data were the alignment used in phylogenetic estimation (above) divided into plastid and nuclear partitions. Unlinked GTR+I+G models, and independent

uncorrelated relaxed clocks were applied to each partition. A Birth-Death model was used as a tree prior. The ML tree was made ultrametric with penalized likelihood[45] using r8s v1.7.1[46] and TreePL[47], and specified as starting tree. The root node, corresponding to the seed plant crown node, was calibrated with a uniform distribution between 314 and 350 Ma, corresponding to the credibility interval of the age estimated for this node in a previous analysis[48]. A prior uniform distribution between 136 and 139.35 Ma was assigned to the angiosperm crown node, corresponding to the 95% confidence interval estimated for this node (above). The angiosperm fossil data base (above) was screened to identify and select fossils that could provide reliable minimum ages within angiosperms. One hundred and thirty six fossils were selected on the basis of reliable affinity to a particular clade[1]. Assignment of fossils to nodes on the phylogenetic tree was derived from morphological distinctive attributes, or, if available, their phylogenetic position. The prior ages of the 136 calibrated nodes were obtained from lognormal distributions in which the mean was equal to the fossil age plus 10%, and a standard deviation of 1. The calibration fossils, the attributes supporting clade affinity, their assignment to the stem or the crown node of the clade, their stratigraphic position and age are discussed in detail in the original dating study[1], are available in the DRYAD digital repository (http://dx.doi.org/10.____/dryad._____). The analysis consisted of eight independent MCMC runs for a total of 170 x 10^6 steps, sampling one every 5000. The initial 600 trees from every run were excluded as burn-in. The outputs of the runs were analyzed with Tracer v1.5[44], and the estimated parameters were extracted with LogCombiner v1.7.5 and TreeAnnotator v1.7.5[24].

Diversification Analyses

The macroevolutionary diversification dynamics of angiosperms were investigated with the C++ programme Bayesian Analysis of Macroevolutionary Mixtures (BAMM) v2.0[2], which, through a compound Poisson Process implemented as a reversible jump Markov Chain Monte Carlo (rjMCMC), estimates major shifts in the rates of speciation, extinction and diversification among the branches of a phylogenetic tree, and through time. Post-run analyses were conducted with the R package BAMMtools v2.0[49].

BAMM simulates a posterior distribution of shift configurations, each corresponding to a particular combination of a number shifts (increases and decreases in diversification, speciation or extinction), and their phylogenetic and temporal placement among the branches of the tree. Given the posterior sample of shift configurations derived from the rjMCMC, it is possible to evaluate those contained in the 95% credible set, the one with the overall highest posterior probability (PP, i.e., the best configuration), the one that maximizes the marginal probability of rate shifts along individual branches (i.e., the MSC configuration), or obtain a phylorate plot in which the rate of speciation, extinction or diversification averaged across all the configurations in the posterior distribution is plotted on each time unit on each branch[50]. BAMM uses Bayes Factor as a robust measure to distinguish between core and non-core shifts. Whereas non-core shifts are expected given the prior distribution (which ultimately depends on branch lengths), core shifts represent meaningful changes in the rate of speciation, extinction or diversification that are independent of prior probability[50].

The angiosperm dated tree described above was used as input phylogeny, in which branch lengths correspond to absolute time in million-year units. BAMM was set

to conduct a speciation-extinction analysis. Priors on rate parameters were scaled to our dated tree using the setBAMMpriors function in BAMMtools. The exponential prior on the rate parameter of the Poisson process (poissonRatePrior), which determines the number and placement of shifts, was set to 0.2, following indications in BAMMtools documentation, and after empirically noticing that a higher value (1.0, recommended for <500 terminals) resulted in a very small number of shifts. The rate parameter of the exponential priors for the initial speciation and extinction values (lambdaInitPrior and muInitPrior) were both set to 2.35509851913498. The prior for the standard deviation of the normal distribution (mean fixed at zero) of the speciation shift parameter for rate regimes (lambdaShiftPrior) was set to 0.00825917607422974. Constant diversification rate branch segments were made of 3 Ma by setting the segLength parameter at 0.02152148, given a crown node age of 139.3956 Ma. Rates were allowed to vary through time (lambdaIsTimeVariablePrior = 1). The prior distribution of rate shifts was simulated (simulatePriorShifts = 1), to allow estimation of Bayes Factors associated with rate shifts.

Non-random incomplete taxon sampling of full angiosperm diversity was accounted for by indicating that clade-specific sampling probabilities would be used, and by specifying the sampled fraction of clades in the tree. Most of the clades correspond to angiosperm families recognized in the Angiosperm Phylogeny Website in April, 2013[8], with the total number of species in each family obtained from this same source. Families not represented in the dated tree were accounted for by aggregating their species-richness to that of their sister clade, according to relationships in the Angiosperm Phylogeny Website. Following BAMMtools documentation, for each terminal in the tree, we

specified the represented fraction of the clade to which it belongs by dividing the total species-richness of the clade (i.e., a family or a family plus non-represented sister families) by the number of terminals belonging to that clade. The backbone of the phylogeny was fully sampled. The total species richness of clades in the tree, their composition and the sampling fraction, indicated on each terminal, are available in the DRYAD Digital Repository (http://dx.doi.org/10.____/dryad._____).

The MCMC simulation consisted of 300 x 10^6 steps, in which one shift configuration was sampled every 200,000 steps. The configurations sampled in the initial 10% of the MCMC were discarded as burn-in, hence the total number of analysed posterior samples is 1,351. The input tree and control file, and output event data of the BAMM analysis are available in the DRYAD Digital Repository (http://dx.doi.org/10.____/dryad._____). The MCMC achieved convergence, and the effective sample size was 904.24. The posterior 95% credible set contains 1,253 configurations. While each of them has a low PP, the PP of the best is twice as high as that of the second-best (0.0059 and 0.0030, respectively; Extended Data Figure 3).

We calculated the phylorate plot (Fig. 1), and identified the best and the MSC configurations (Extended Data Figures 1-2). We identified the shifts found in all the configurations in the posterior distribution, and sorted them by their frequency of occurrence. The frequency distribution of shifts among configurations has the shape of a hollow curve, in which few shifts appear in many configurations, and the great majority appear in very few (Extended Data Figure 4). We chose to discuss those shifts that occur in 50% or more of the configurations (shown in Fig. 1) because they correspond almost exactly with those found in the best configuration and in the MSC configuration

(Extended Data Figures 1-2); and because all are identified as core shifts according to their associated Bayes Factor (Extended Data Figure 5). The total number of shifts detected in the posterior set of configurations is 1030, but only 17 occur in ≥50% of them (Extended Data Table 2). For every branch in which a shift was detected we extracted the age of the shift in all the configurations in which it was detected, and obtained the mean, minimal and maximal shift age (Fig. 1, Extended Data Tables 1, 2).

References

1. Magallón, S., Gómez-Acevedo, S. Sánchez-Reyes, L. L. & Hernández-Hernández, T. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytol.* **207**, 437-453 (2015).
2. Rabosky, D. L. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS ONE*, e89543 (2014).
3. Rubinstein, C. V., Gerrienne, P., de la Puente, G. S., Astini, R. A. & Steemans, P. Early Middle Ordovician evidence for land plants in Argentina (eastern Gondwana). *New Phytol.* **188**, 365-369 (2010).
4. Schneider, H. *et al.* Ferns diversified in the shadow of angiosperms. *Nature* **428**, 553-557 (2004).
5. Cardinal, S. & Danforth, B. N. Bees diversified in the age of eudicots. *Proc. R. Soc. Lond. B* **280**, http://dx.doi.org/10.1098/rspb.2012.2686 (2013).
6. Guzmán, B., Lachance, M.-A. & Herrera, C. M. Phylogenetic analysis of the angiosperm-floricolous insect-yeast association: have yeast and angiosperm lineages co-diversified? *Mol. Phylog. Evol.* **68**, 161-175 (2013).

7. Lemaire, B., Vandamme, P., Merckx, V., Smets, E. & Dessein, S. Bacterial leaf symbiosis in angiosperms: host specificity without co-speciation. *PLoS ONE* **6**, e24430 (2011).
8. Stevens, P.F. Angiosperm Phylogeny Website. http://www.mobot.org/MOBOT/research/APweb/ (2013).
9. Joppa, L. N., Roberts, D. L. & Pimm, S. L. How many species of flowering plants are there? *Proc. R. Soc. Lond. B* doi:10.1098/rspb.2010.1004 (2010).
10. Wiens, J. J. The causes of species richness patterns across space, time, and clades and the role of "ecological limits". *Q. Rev. Biol.* **86**, 75-96 (2011).
11. Sargent, R. D. Floral Symmetry affects speciation rates in angiosperms. *Proc. R. Soc. Lond. B* **271**, 603-608 (2004).
12. Weber, M. G. & A. A. Agrawal. Defense mutualisms enhance plant diversification. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 16442-16447 (2014).
13. Hughes, C.E. & Atchison, G.W. The ubiquity of alpine plant radiations: from the Andes to the Hengduan Mountains. *New Phytol*. **207**, 275-282 (2015).
14. Marazzi, B. *et al.* Locating evolutionary precursors on a phylogenetic tree. *Evolution* **66**, 3918-3930 (2010).
15. Stebbins, G. L. *Flowering Plants: Evolution above the Species Level.* (Belknap Press of Harvard University Press, Cambridge, 1974).
16. Sanderson, M. J. & Donoghue, M. J. Shifts in diversification rate with the origin of angiosperms. *Science* **264**, 1590-1593 (1994).
17. Magallón, S. & Sanderson, M. J. Angiosperm diversificatiopn rates in angiosperm clades. *Evolution* **55**, 1762-1780 (2001).

18. Smith, S. A., Beaulieu, J. M., Stamatakis, A. & Donoghue, M. J. Understanding angiosperm diversification using small and large phylogenetic trees. *Amer J. Bot.* **98**, 404-414 (2011).

19. Tank, D. C. *et al.* Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. *New Phytol.* **207**, 454-467 (2015).

20. Doyle, J. A. Molecular and fossil evidence on the origin of angiosperms. *Annu. Rev. Earth Planet. Sci.* **40**, 301-326 (2012).

21. Friis, E. M., Pedersen, K. R., von Balthazar, M., Grimm, G. W. & Crane, P. R. *Monetianthus mirus* gen. et sp. nov., a nymphaealean flower from the Early Cretaceous of Portugal. *Int. J. Plant Sci.* **170**, 1086–1101 (2009).

22. Friis, E. M., Pedersen, K. R. & Crane, P. R. Araceae from the Early Cretaceous of Portugal: evidence on the emergence of monocotyledons. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16565-16570 (2004).

23. Doyle, J., Biens, P., Doerenkamp, A., & Jardiné, S. Angiosperm pollen from the pre-Albian Lower Cretaceous of equatorial Africa. *B. Cent. Rech. Expl.* **1**, 451–473 (1977).

24. Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88 (2006).

25. Marshall, C.R. A simple method for bracketing absolute divergence times on molecular phylogenies using multiple fossil calibration points. *Am. Nat.* **171**, 726-742 (2008).

26. Soltis, D. E. *et al.* Angiosperm phylogeny: 17 genes, 640 taxa. *Am. J. Bot.* **98**, 704-730 (2011).

27. Cantino, P. D. *et al.* Towards a phylogenetic nomenclature of Tracheophyta. *Taxon* **56**, 822-846 (2007).

28. Rabosky, D. L. Ecological limits and diversification rate: alternative paradigms to explain the variation in species richness among clades and regions. *Ecol. Lett.* **12**, 735–743 (2009).

29. Strathmann, R. R. & Slatkin, M. The improbability of animal phyla with few species. *Paleobiology* **9**, 97-106 (1983).

30. Donoghue, M. J. & Sanderson, M. J. Confluence, synnovation, and depauperons in plant diversification. *New Phytol.* **207**, 260-274 (2015).

Methods References

31. Edgar, R. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

32. Hall, T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Sym. Ser.* **41**, 95–98 (1999).

33. Stamatakis, A. Phylogenetic models of rate heterogeneity: a high performance computing perspective. *Proceedings of 20th IEEE International Parallel and Distributed Processing Symposium (IPDPS2006), Rhodos, Greece*. (2006).

34. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).

35. Stamatakis, A. A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.* **57**, 758–771 (2008).

36. Dornburg, A., Brandley, M., McGowen, M & Near, T. Relaxed clocks and inferences of heterogeneous patterns of nucleotide substitution and divergence time estimates across whales and dolphins (Mammalia: Cetacea). *Mol. Biol. Evol.* **29**, 721–736 (2012).

37. Wertheim, J., Fourment, M., & Kosakovsky Pond, S. Inconsistencies in estimating the age of HIV-1 subtypes due to heterotachy. *Mol. Biol. Evol.* **29**, 451–456 (2012).

38. Bealieu, J. M., O'Meara, B. C., Crane, P. & Donoghue, M. J. Heterogeneous rates of molecular evolution and diversification could explain the Triassic age estimate for angiosperms. *Syst. Biol.* **64**, 869-878 (2015).

39. Sanderson, M. J. Back to the past: a new take on the timing of flowering plant diversification. *New Phytol.* **207**, 257-259 (2015).

40. Hughes, N. & McDougall, A. 1987. Records of angiospermid pollen entry into the English Early Cretaceous succession. *Rev. Paleobot. Palynol.* **50**, 255–272 (1987).

41. Hughes, N., McDougall, A. & Chapman, J. Exceptional new record of Cretaceous Hauterivian angiospermid pollen from southern England. *J. Micropalaeontol.* **10**, 75–82 (**1991**).

42. Brenner, G. in *Flowering Plant Origin, Evolution and Phylogeny* (eds. Taylor, D. & Hickey, L.) 91–115 (Chapman & Hall, New York, 1996).

43. Walker, J. & Geissman, J. Geologic time scale. *GSA Today* **19**, 60–61 (2009).

44. Rambaut, A. & Drummond A. Tracer version 1.5.0. http://tree.bio.ed.ac.uk/software/tracer/ (2008).

45. Sanderson, M.J. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol. Biol. Evol.* **19**, 101–109 (2002).

46. Sanderson, M. J. r8s, version 1.70 user's manual. http://loco.biosci.arizona.edu/r8s/ (2004).

47. Smith, S. A. & O'Meara B. C. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28**, 2689–2690 (2012).

48. Magallón, S., Hilu, K. W.; & Quandt, D. Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rate. *Am. J. Bot.* **100**, 556-573 (2013).

49. Rabosky, D. L., Donnellan, S. C., Grundler, M & Lovette, I. J. Analysis and visualization of complex macroevolutionary dynamics: an example from Australian scincid lizards. *Syst. Biol.* **64**, 610-627 (2014).

50. Rabosky, D. L. BAMM – Bayesian Analysis of Macroevolutionary Mixtures. Documentation. http://bamm-project.org/documentation.html (2014).

End Notes

Acknowledgements: We thank H. Sauquet and L. E. Eguiarte for comments, A. Luna for preparation of plant drawings, and G. Ortega-Leite for assistance. LLS-R thanks the Consejo Nacional de Ciencia y Tecnología, México for a scholarship, and the Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México (UNAM), for support. SLG-A thanks the Dirección General de Asuntos del Personal Académico, UNAM, for postdoctoral funding.

Author Contributions: This study was designed and coordinated by SM. SLG-A compiled and organised fossil information to conduct molecular clock analyses, and provided input in molecular clock and diversification analyses. LLS-R conducted phylogenetic analyses and most of the diversification analyses. She contributed to all the phases of the project, from data compilation to final discussions. SM conducted molecular clock analyses, and contributed to diversification analyses. She wrote the manuscript and associated materials, and prepared the figures, with contributions of all co-authors.

Author Information: The molecular data set used in this study and input files used in diversification analysis are deposited in the DRYAD Digital Repository (http://dx.doi.org/10.____/dryad._____). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no financial competing interests. Correspondence and requests for materials should be addressed to s.magallon@ib.unam.mx.

Material Suplementario

Extended Data Table 1. Major rate shifts within angiosperms. Seventeen significant diversification shifts detected in ≥50% of configurations in the posterior distribution of the macroevolutionary diversification analysis conducted with BAMM*, in chronological order. Angiosperms as a whole are also included (top row). The total diversity of each clade is indicated, as well as the frequency (%) with which it was detected among the configurations in the posterior distribution. The mean age and temporal range of each shift were obtained by extracting and summarizing individual ages from all the configurations in which a given shift was detected. The mean speciation and extinction rates for shift clades and for angiosperms as a whole, and their 95% highest posterior density (HPD), were estimated with BAMM. Unless otherwise indicated, clade names are according to Cantino et al. 2007[27] or Stevens 2013[8]. Superasteridae is in the sense of Soltis et al. 2011[26]. Rosidae+ includes Rosidae and Vitales. The clade here named Dioscoreidae includes Dioscoreales, Pandanales, Liliales, Asparagales and Commelinidae. Ranunculaceae+ includes Ranunculaceae, Berberidaceae and Menispermaceae. Lamiidae pp (*pro parte*) includes Gentianales, Solanales, Boraginales and Lamiales. Montiniaceae+ includes Montiniaceae, Sphenocleaceae and Hydroleaceae. Euphorbiaceae includes Rafflesiaceae. Apiales pp includes Pittosporaceae, Araliaceae, Myodocarpaceae and Apiaceae. Asteraceae+ includes Goodeniaceae, Calyceraceae and Asteraceae. Celastraceae excludes *Parnassia*. Core Lauraceae corresponds to the clade that includes *Cinnamomum*, *Laurus*, and *Sassafras*.

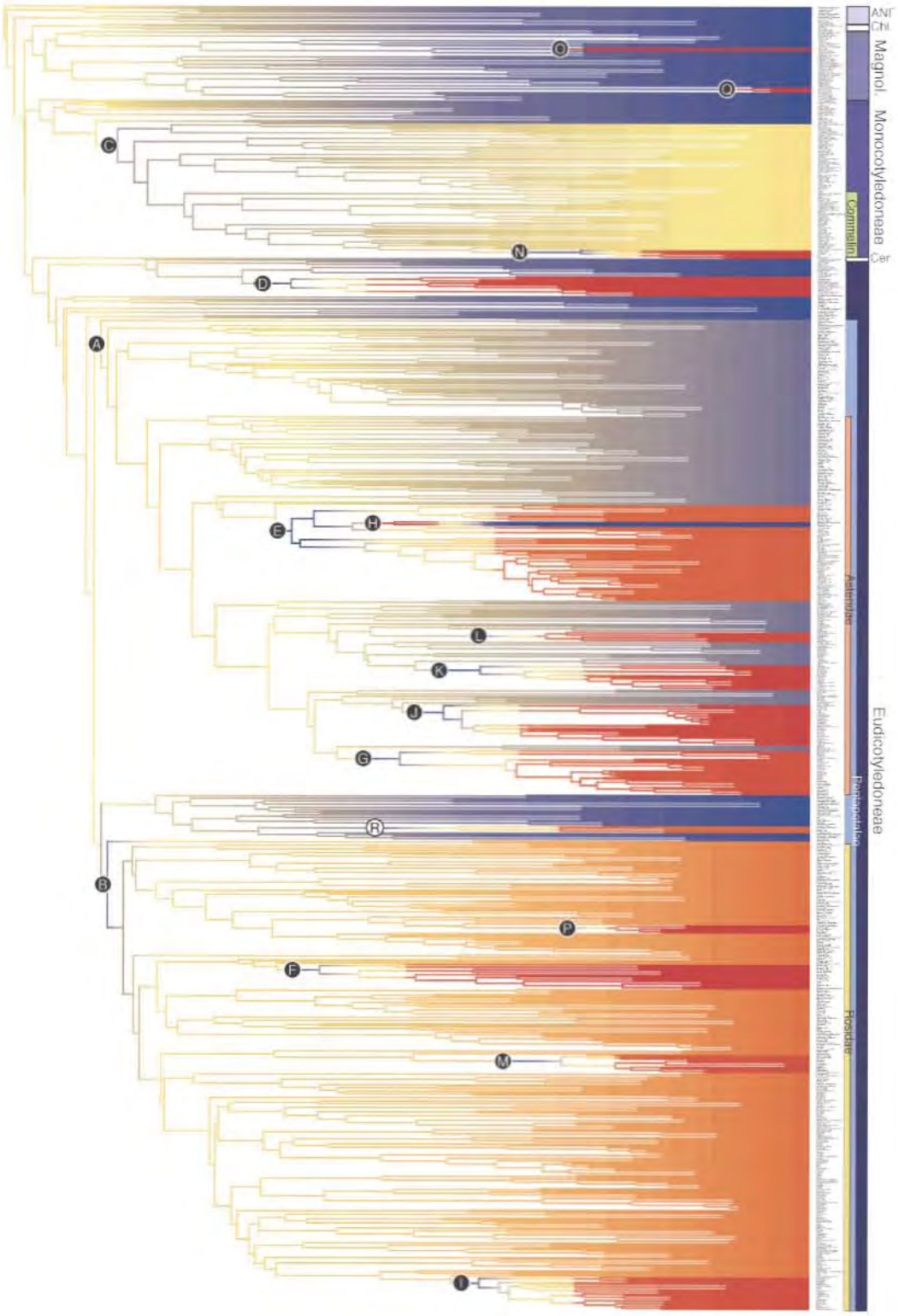
Clade	**Diversity**	**Freq. (%)**	**Shift Mean Age (max-min) in Ma**	**Speciation (95% HPD)**	**Extinction (95% HPD)**
Angiospermae	276,776			0.1505 (0.1205-0.1840)	0.0886 (0.0564-0.1253)

Superasteridae	111,758	56.81	123.16 (123.73-122.58)	0.1696 (0.1275-0.2338)	0.1000 (0.0541-0.1699)
Rosidae+	81,498	73.11	121.88 (122.40-121.29)	0.1760 (0.1172-0.2549)	0.1087 (0.0422-0.1946)
Dioscoreidae	59,994	77.48	121.68 (123.52-119.77)	0.1432 (0.0937-0.2269)	0.0715 (0.0155-0.1620)
Ranunculaceae+	3,668	91.85	94.25 (98.19-89.94)	0.2381 (0.0969-0.5086)	0.1639 (0.0114-0.4473)
Lamiidae pp	50,437	98.07	91.24 (92.68-89.75)	0.2673 (0.1458-0.5000)	0.1588 (0.0277-0.3990)
Fabaceae	19,500	83.33	88.34 (92.14-84.77)	0.3975 (0.1711-0.8193)	0.2955 (0.0529-0.7279)
Dipsacales	1,090	91.41	76.93 (81.84-70.95)	0.2902 (0.1245-0.5880)	0.2073 (0.0228-0.5317)
Montiniaceae+	19	99.26	75.50 (79.24-72.05)	0.0689 (0.0194-0.2136)	0.0503 (0.0030-0.2020)
Euphorbiaceae	5,755	79.56	69.03 (74.21-61.92)	0.3997 (0.1573-0.8410)	0.2873 (0.0224-0.7589)
Apiales pp	5,449	94.30	66.91 (70.53-63.38)	0.4160 (0.2036-0.7967)	0.2998 (0.0695-0.6955)
Asteraceae+	24,090	95.48	61.97 (68.26-57.05)	0.5597 (0.2642-0.9485)	0.4001 (0.0945-0.7999)
Campanulaceae	2,380	98.30	57.56 (76.02-45.59)	0.3273 (0.1346-0.6959)	0.2077 (0.2077-0.5945)
Celastraceae	1,349	57.48	55.83 (68.52-42.86)	0.3442 (0.1146-0.8192)	0.2497 (0.0323-0.7420)
Poaceae	11,337	84.67	50.07 (58.57-39.79)	0.7178 (0.1061-1.4618)	0.5784 (0.0239-1.3169)
Piperaceae	3,615	62.07	48.06 (65.41-39.35)	0.1826 (0.0363- 0.5210)	0.0878 (0.0056-0.4033)
Malvaceae	4,225	67.19	43.94 (59.74-33.31)	0.4245 (0.1161- 0.9868)	0.2813 (0.0225-0.8370)
Core Lauraceae	2,500	72.15	14.65 (79.98-10.02)	0.2549 (0.0495- 0.6172)	0.1288 (0.0119-0.4744)

Extended Data Table 2. Rate shifts present in ≥10% of sampled configurations. Rate shifts found in ≥10% of configurations in the posterior distribution, including the mean, minimum and maximum age of each shift. The frequency indicates the number of trees in which the shift was detected (PP freq), out of a total of 1351 sampled configurations, and the percentage that these configurations represent (% configurations). The total number of shifts detected in all sampled configurations is 1,030, but only 33 (shown) are found in ≥10% of the trees, and only 17 (shaded) are found in ≥50% of the trees. Unless otherwise indicated, clade names are according to Cantino et al. (2007)[27] and Stevens (2013)[8]. Superasteridae is in the sense of Soltis et al. (2011)[26]. Rosidae+ includes Rosidae and Vitales. The clade here named Dioscoreidae includes Dioscoreales, Pandanales, Liliales, Asparagales and Commelinidae. Ranunculaceae+ includes Ranunculaceae, Berberidaceae and Menispermaceae. Lamiidae *pro parte* (pp) includes Gentianales, Solanales, Boraginales and Lamiales. Montiniaceae+ includes Montiniaceae, Sphenocleaceae and Hydroleaceae. Euphorbiaceae includes Rafflesiaceae (unsampled). Apiales pp includes Pittosporaceae, Araliaceae, Myodocarpaceae and Apiaceae. Asteraceae+ includes Goodeniaceae, Calyceraceae and Asteraceae. Celastraceae excludes *Parnassia*. Core Lauraceae corresponds to the clade that includes *Cinnamomum*, *Laurus*, and *Sassafras*.

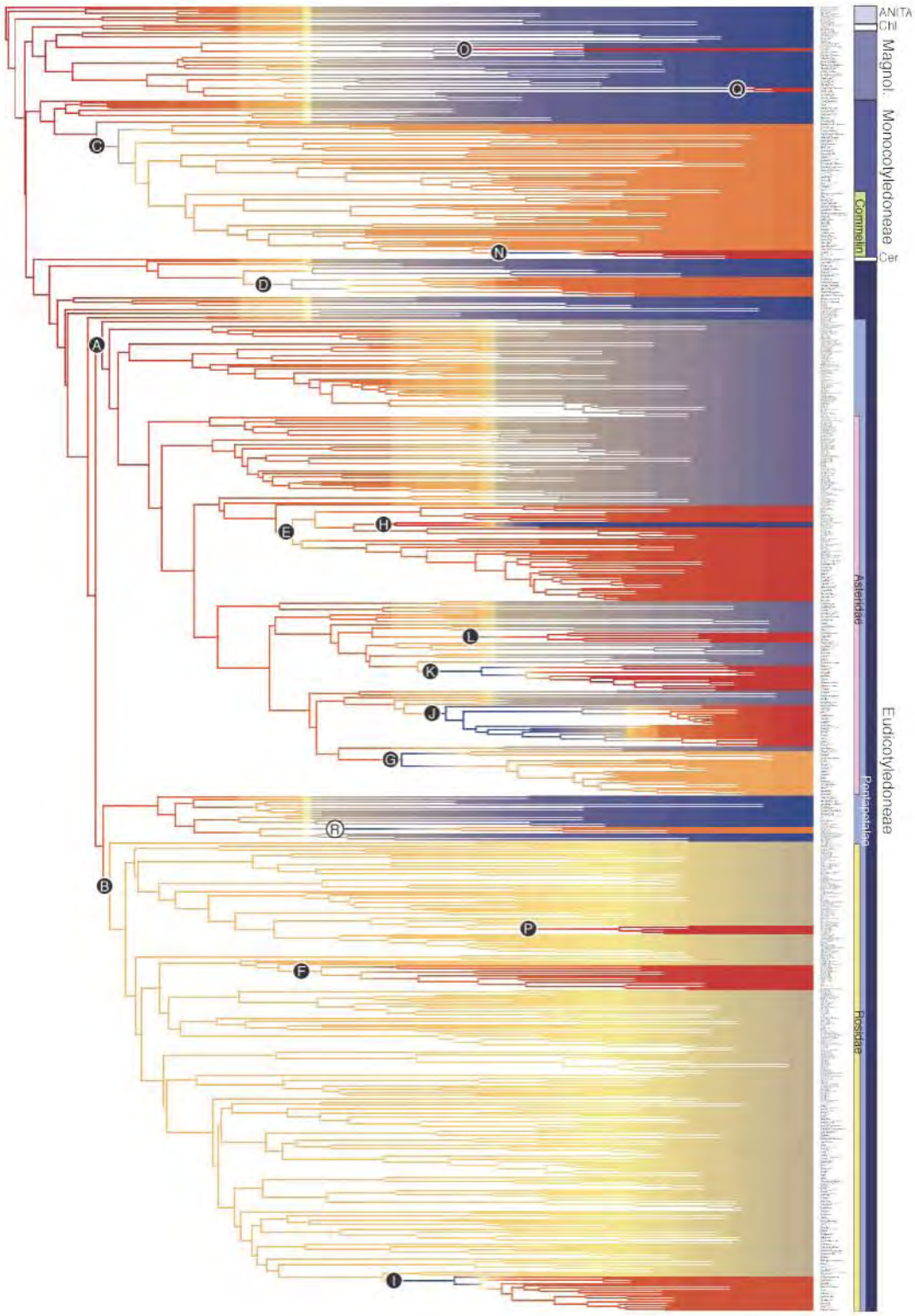
Node Number	Shift Clade	Mean Age	Min Age	Max Age	PP Freq	% Configurations
1289	Montiniaceae+	75.50	72.05	79.24	1340	99.26
1220	Campanulaceae	57.56	45.59	76.02	1327	98.30
1243	Lamiidae pp	91.24	89.75	92.68	1324	98.07
1194	Asteraceae+	61.97	57.05	68.26	1289	95.48
1159	Apiales pp	66.91	63.38	70.53	1273	94.30
1418	Ranunculaceae+	94.25	89.94	98.19	1240	91.85
1128	Dipsacales	76.93	70.95	81.84	1234	91.41
1450	Poaceae	50.07	39.79	58.57	1143	84.67
1007	Fabaceae	88.34	84.77	92.14	1125	83.33

821	Euphorbiaceae	69.03	61.92	74.21	1074	79.56
1439	Dioscoreidae	121.68	119.77	123.52	1046	77.48
806	Rosidae+	121.88	121.29	122.40	987	73.11
1542	Core Lauraceae	14.65	10.02	79.98	974	72.15
1045	Malvaceae	43.94	33.31	59.74	907	67.19
1564	Piperaceae	48.06	39.35	65.41	838	62.07
956	Celastraceae	55.83	42.86	68.52	776	57.48
1117	Superasteridae	123.16	122.58	123.73	767	56.81
1098	Crassulaceae	74.20	60.69	95.37	613	45.41
466	Plocospermataceae	38.91	2.64	76.90	503	37.26
1484	Asparagales pp	84.81	80.70	88.94	432	32.00
1502	Orchidaceae	74.92	59.96	108.17	421	31.19
1303	Ericales	108.08	103.59	112.34	380	28.15
796	Mesangiospermae	136.78	135.92	137.69	296	21.93
807	Rosidae	120.04	118.59	121.26	269	19.93
822	Euphorbiaceae minus *Neoscotechinia*	59.56	57.13	61.86	269	19.93
794	All angiosperms except *Amborella*	139.18	138.97	139.39	248	18.37
1476	Arecales	64.03	49.42	97.68	197	14.59
792	Amborellaceae	52.12	4.09	138.31	172	12.74
1008	Fabaceae pp	81.19	77.96	84.76	171	12.67
1118	Santalales-Asteridae	121.95	121.39	122.58	168	12.44
1379	Amaranthaceae	51.23	43.74	63.82	150	11.11
462	Tetrachondraceae	31.94	2.07	66.51	138	10.22
1323	Ebenaceae-Primulaceae	90.38	87.00	94.34	135	10.00



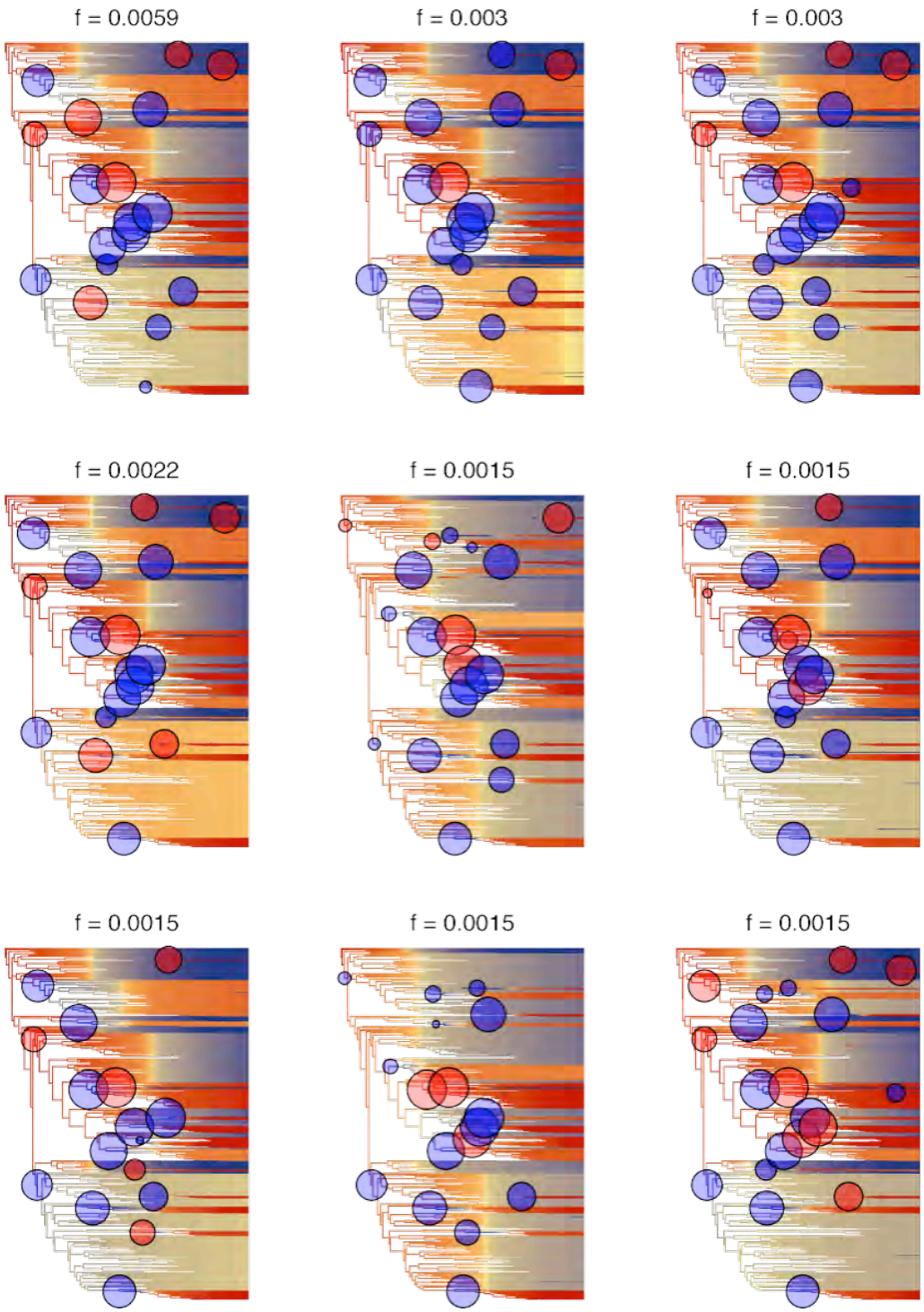
Extended Data Figure 1. Best rate shift configuration. Rate shift configuration with maximum a posteriori (MAP) probability, corresponding to the overall best set of rate

shifts. It includes 18 rate shifts. Seventeen of these shifts, (A-Q; white letters inside black circles), are those detected in ≥50% of configurations in the posterior distribution (Figs. 1-2; Extended Data Table 1). An additional shift (R; black letter inside a white circle) was found in <50% of configurations in the posterior distribution. Branch colours indicate diversification rate (cold colours= low rates; warm colours=high rates). Shift clade are as follows: (A) Superasteridae[26] (including Dilleniales, Santalales, Berberidopsidales, Caryophyllales and Asteridae); (B) Rosidae+ (including Vitales and Rosidae); (C) Dioscoreidae (including Dioscoreales, Liliales, Asparagales, and Commelinidae); (D) Ranuculaceae+ (including Ranunculaceae, Berberidaceae and Menispermaceae); (E) Lamiidae *pro parte* (pp, including Gentianales, Solanales, Boraginales and Lamiales); (F) Fabaceae; (G) Dipsacales; (H) Montiniaceae+ (including Montiniaceae, Sphenocleaceae and Hydroleaceae); (I) Euphorbiaceae; (J) Apiales pp (including Pittosporaceae, Araliaceae, Myodocarpaceae and Apiaceae); (K) Asteraceae+ (including Goodeniaceae, Calyceraceae and Asteraceae); (L) Campanulaceae; (M) Celastraceae; (N) Poaceae; (O) Piperaceae; (P) Malvaceae; (Q) core Lauraceae; and (R) Crassulaceae. Unless otherwise indicated, clade names are based on Cantino et al. 2007[27] and Stevens 2013[8].



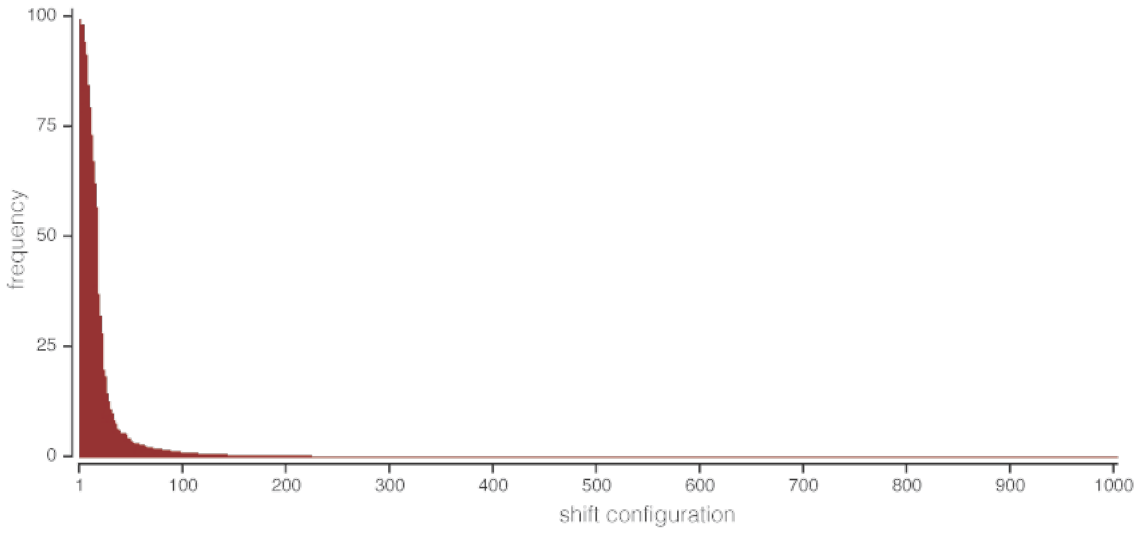
Extended Data Figure 2. Maximum shift credibility (MSC) configuration. Rate shift configuration that maximizes the marginal probability of rate shifts along individual

branches. This configuration contains seventeen shifts. Sixteen of these shifts (A-L, N-Q; white letters in black circles) are found in ≥50% of configurations in the posterior distribution. One shift (R; black letter inside a white circle) was found in <50% of configurations in the posterior distribution. Branch colours indicate diversification rate (cold colours= low rates; warm colours=high rates). Shift clade are as follows: (A) Superasteridae[26] (including Dilleniales, Santalales, Berberidopsidales, Caryophyllales and Asteridae); (B) Rosidae+ (including Vitales and Rosidae); (C) Dioscoreidae (including Dioscoreales, Liliales, Asparagales, and Commelinidae); (D) Ranuculaceae+ (including Ranunculaceae, Berberidaceae and Menispermaceae); (E) Lamiidae *pro parte* (pp, including Gentianales, Solanales, Boraginales and Lamiales); (F) Fabaceae; (G) Dipsacales; (H) Montiniaceae+ (including Montiniaceae, Sphenocleaceae and Hydroelaceae); (I) Euphorbiaceae; (J) Apiales pp (including Pittosporaceae, Araliaceae, Myodocarpaceae and Apiaceae); (K) Asteraceae+ (including Goodeniaceae, Calyceraceae and Asteraceae); (L) Campanulaceae; (N) Poaceae; (O) Piperaceae; (P) Malvaceae; (Q) core Lauraceae; and (R) Crassulaceae. Unless otherwise indicated, clade names are based on Cantino et al. 2007[27] and Stevens 2013[8].

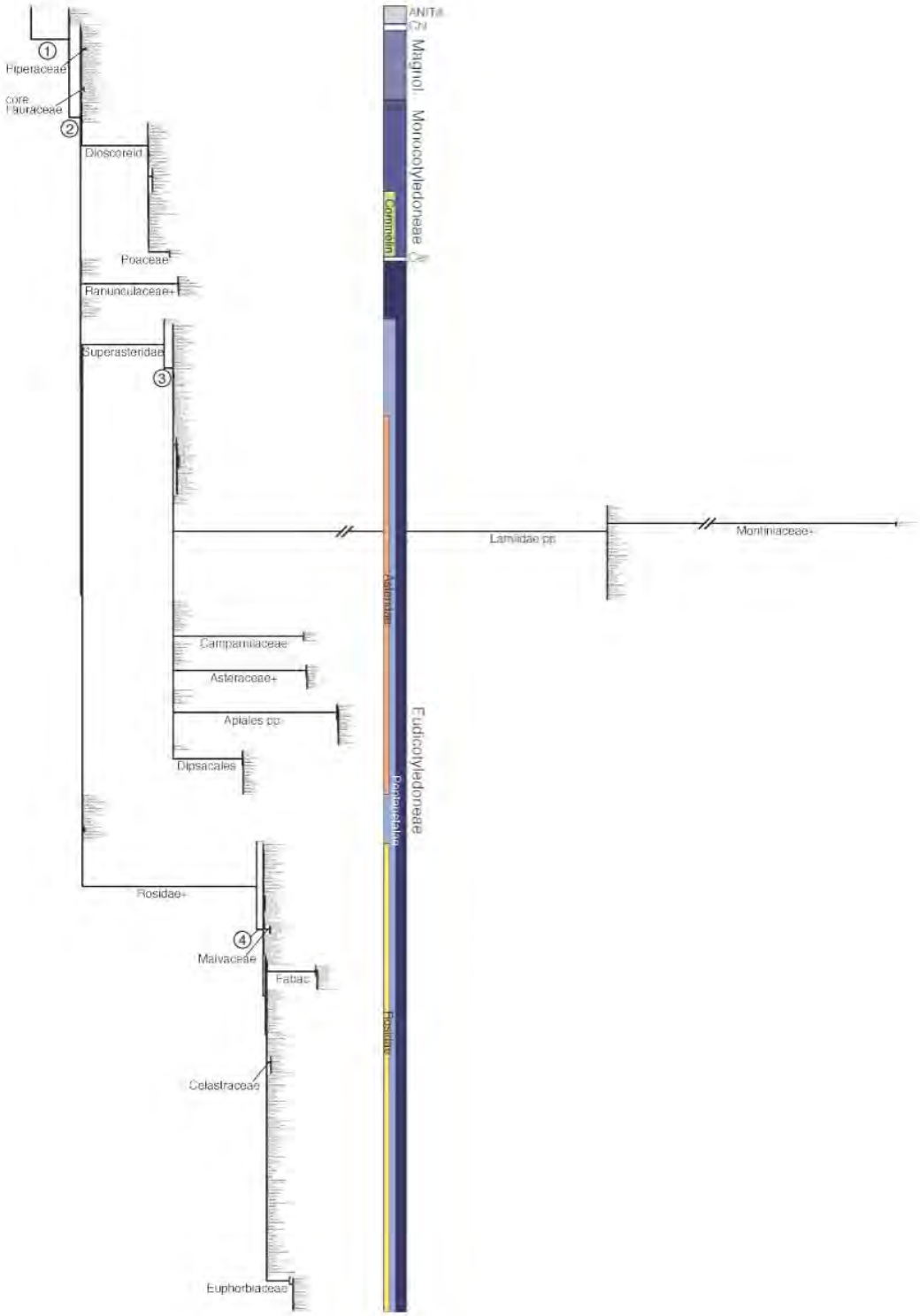


Extended Data Figure 3. Nine best shift configurations. The macroevoutionary diversification analysis conducted with BAMM simulated a posterior distribution of shift configurations, each corresponding to a particular combination of number, phylogenetic

placement and temporal placement of diversification shifts. The 95% credible set of the posterior distribution of angiosperm diversification shifts contains 1253 distinct configurations. The nine configurations with the highest posterior probability are shown, including their associated posterior probability (f). In each configuration, red circles indicate diversification acceleration, and blue circles indicate diversification decreases. The diameter of the circle is proportional to the marginal probability that a shift occurs on that specific branch. Branch colours indicate diversification rate (cold colours= low rates; warm colours=high rates). The posterior probability of the best configuration (top left) is approximately twice as high as that of the second-best configuration (top center).



Extended Data Figure 4. Frequency distribution of shifts in configurations. The frequency distribution of shifts in all configurations has the shape of a hollow curve, in which few shifts appear in many configurations, and the great majority appear in very few. The total number of shifts detected in the posterior set of configurations is 1013, but only 17 occur in ≥50% of them (Extended Data Table 1).



Extended Data Figure 5. Bayes Factor tree. Phylogenetic tree with branch lengths proportional to Bayes Factor associated with a rate shift on that branch. Bayes Factors provide a measure of the occurrence of a rate shift on a particular branch that is

independent of prior rate shift probabilities. Named or numbered clades are subtended by branches with high Bayes Factors: (1) all angiosperms except *Amborella*; (2) Mesangiospermae; (3) Superasteridae excluding Dilleniales; (4) Rosidae. Note that branches subtending Lamiidae pp and Montiniaceae+ were graphically shortened to fit the page.

CAPÍTULO 2

DIVERSIFICACIÓN DE FOUQUIERIA (FOUQUIERIACEAE, ERICALES) EN LOS DESIERTOS NORTEAMERICANOS

Manuscrito con formato para New Phytologist

Research Paper

Short Title: *Diversification of Fouquieria*

Diversification of *Fouquieria* (Fouquieriaceae, Ericales) in North American Deserts.

José Arturo De-Nova[1,2], Luna L. Sánchez Reyes[3], Luis E. Eguiarte[4], Susana Magallón[1]

[1]Departamento de Botánica, Instituto de Biología, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico; [2]Instituto de Investigación de Zonas Desérticas – Facultad de Agronomía y Veterinaria, Universidad Autónoma de San Luis Potosí, San Luis Potosí 78377, Mexico; [3]Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico. [4]Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico.

Author's e-mail addresses: arturo.denova@gmail.com, sanchez.reyes.luna@gmail.com, fruns@servidor.unam.mx, s.magallon@ib.unam.mx

Main body of text: 5257 words (Introduction: 1011; Materials and Methods: 1500; Results: 1114; Discussion: 1378; Conclusions: 167; Acknowledgements: 86).

Number of Figures: 4; number of tables: 4;

Supporting online Figures: 4.

Supporting online Tables: 7.

Summary

· Arid biomes are particularly prominent in the Neotropics providing some of its most emblematic landscapes and a substantial part of its species diversity. To understand the evolutionary processes underlying speciation in Mexican Deserts, we studied the diversification of *Fouquieria*, which includes eleven species, all endemic to the warm deserts and dry subtropical regions of North America.

· Using a phylogeny from plastid DNA sequences with samples of individuals from populations of all recognized *Fouquieria* species, we estimate divergence times, test for heterogeneity in temporal diversification, geographical structure, and conduct ancestral area reconstruction.

· *Fouquieria* is an ancient lineage that diverged from Polemoniaceae ca. 86 Ma. A Mio-Pliocene diversification of *Fouquieria*, associated with Neogene orogenesis underlying the early development of regional deserts is strongly supported.

· Tests of temporal diversification heterogeneity indicate that during most of its evolutionary history, *Fouquieria* maintained a negative diversification rate involving higher extinction than speciation. Explanations for the lack of fossil record during this period are discussed.

· From the late Miocene onwards, *Fouquieria* underwent substantial diversification change, involving high speciation decreasing to the present and negligible extinction, which is congruent with the scant fossil record. Geographic phylogenetic structure and the pattern of most sister species inhabiting different desert nuclei suggest that isolation by distance could be the main driver of speciation.

Key words: ancestral area reconstruction, dated phylogenies, diversification rate, *Fouquieria*, geographical structure, Neotropics, North American deserts.

Introduction

Arid biomes, from dry xerophytic scrubs to seasonally dry forests, are particularly prominent in North America, providing some of its most emblematic landscapes and a substantial part of its species diversity. North American regional deserts have a complex biogeographic history. It is thought they have been modelled by the influence of several vicariant events that have affected the assembly and diversification of the xerophytic biota (Findley, 1969; Hubbard, 1973; Morafka, 1977; Ortega-Gutiérrez & Guerrero-García, 1982; Murphy, 1983; Grismer, 1994; Riddle, 1995; Upton & Murphy, 1997; Murphy & Aguirre León, 2002; Riddle & Hafner, 2006; Hafner & Riddle, 2011). Given their actual distribution, biotic composition, endemism and the underlying dynamic climatic and geological history, North American deserts provide an excellent system to test hypotheses about macroevolutionary processes and biogeographical patterns that characterize the biological diversity that inhabit them.

Mexico accounts for two of the warm desert regions recognized by Shreve (1942) – the Sonoran Desert and the Chihuahuan Desert – the latter including associated regions, mainly in the Mesquital Valley in the state of Hidalgo, and in the Tehuacán-Cuicatlán Valley in the states of Puebla and Oaxaca. Also, several semidesert relicts designed as arid tropical scrub by Leopold (1950) occur in the states of Jalisco, Guerrero, Michoacán, Morelos, Puebla and Oaxaca. Mexican deserts are characterized by a high number of endemic plant and animal species (e.g., Rzedowski, 1993; Krings, 2000; Hernández *et al*., 2001; McCain 2003; Hafner & Riddle, 2005, 2011; Riddle & Hafner, 2006; Wilson & Pitts, 2010a; Hernández & Gómez-Hinostrosa, 2011; Sosa & De-Nova, 2012).

There has been profound interest in understanding how North American deserts originated. Axelrod (1979, 1983) suggested these biomes developed in local dry sites during the Tertiary aridification trend, drawing species preadapted to aridity from boreal shrub-steppes, Great Plain grasslands, Mexican highlands, Sinaloan thornscrub, and Californian chaparral. According to the fossil record, taxa preadapted to arid conditions existed in floras as old as the Middle Eocene (e.g., Green River Flora of Colorado and Utah), and are abundant in the late Miocene Mint Canyon Flora (Graham, 2010). As arid conditions expanded after the middle of the Miocene, these older elements could represent a reservoir from which the tropical dry forest and desert vegetation were assembled (eg. *Bursera*: De-Nova *et al*., 2012; *Tiquilia*: Moore & Jansen, 2006; *Leucophyllum*: Gándara & Sosa, 2014; Cactaceae: Hernández *et al.*, 2014).

Graham (2010) indicates that the modern version of the North America dry vegetation appeared primarily approximately around the Miocene-Pliocene boundary, and became modernized during the dry intervals of the Quaternary period. The author also considered that, to trace the origin of the Mexican dry vegetation, the following are important factors: 1) the availability of lineages preadapted to seasonally dry environments; 2) the spread of these environments beginning at approximately the end of the early Miocene; 3) continentality, promoted by the lowering sea levels and rising landmasses; 4) rain shadows, promoted by the orogeny that augments trends toward stationally dry climates; and 5) increasing edaphic aridity through generation of coarse soils.

Vicariant analyses (e.g., Riddle & Hafner, 2006; Wilson & Pitts, 2010b; Hafner & Riddle, 2011) trace the biotic history of North American deserts to the late Miocene or

Pliocene, associated to tectonic events that provided physical conditions for the development of regional deserts (e.g., mountain and plateau uplifting, rifting along major fault systems). Alternatively, North American desert biota may derive from Pleistocene climatic oscillations that caused disjunct desert refugia during pluvial periods. Full regional deserts apparently formed during interglacial periods, but with different times of area reduction and separation. Thus, sister taxa that occupy different regional deserts might share common ancestry at one of at least two different ages, each defined by a qualitatively different set of potential isolating processes: either older, in late Miocene to late Pliocene vicariant events, in response to Neogene orogenesis; or more recently, in Pleistocene vicariance and dispersion events, associated to Quaternary climate cycles.

Here we study the diversification of *Fouquieria*, the sole genus in Fouquieriaceae (Ericales), which includes eleven species, all endemic to the warm deserts and dry subtropical regions of North America (Henrikson, 1972). Only *Fouquieria splendens* is broadly distributed, occupying the Peninsular Baja Californian, Sonoran, and Chihuahuan Deserts. Four species have a relatively wide distribution but are restricted to a single regional desert: *F. diguetii* and *F. columnaris* in the Peninsular Baja California Desert, *F. macdougalii* in the Sonoran Desert, and *F. formosa* broadly distributed in the semidesert relicts in Southwestern Mexico. Six species are narrow endemics: *F. burragei*, found in a few restricted sites of the Peninsular Baja California Desert; *F. fasciculata* found in the drier parts of the state of Hidalgo; *F. leonilae*, and *F. ochoterenae* in some of the relictual semideserts in SW Mexico in the states of Guerrero, Morelos, and Puebla; *F. purpusii* restricted to the driest portion of the Tehuacán-Cuicatlán Desert; and the gypsophilic *F. shrevei*, in a few localities of the Chihuahuan Desert in Coahuila.

Previous phylogenetic studies revealed that family Fouquieriaceae is sister to Polemoniaceae (Bremer *et al*., 2002; Schönenberger *et al.*, 2005, 2010; Sytsma *et al*., 2006). Relationships inside the family were analysed by Schultheis and Baldwin (1999) using the ribosomal DNA (rDNA) intergenic transcribed spacer (ITS) region, but they did not find enough variation, and the estimated phylogeny was incompletely resolved.

Our study addresses several questions relating to the diversification and biogeographic history of the genus *Fouquieria*. We combine extensive sampling, which is analysed to estimate a robust hypothesis of phylogenetic relationships among the species in *Fouquieria*. To document the diversification process of *Fouquieria*, we estimated the speciation and extinction dynamics of the lineage since its separation from Polemoniaceae to the present. In addition, we conducted ancestral area reconstruction and we evaluated geographic phylogenetic structure.

We investigate whether the diversification underlying the living species of *Fouquieria* is a recent radiation, or a process that extends over a long time. In particular, we contrast the hypotheses of a) a Mio-Pliocene diversification associated with Neogene orogenesis related to the early development of regional deserts, vs. b) Pleistocene diversification promoting vicariance and dispersion associated to Quaternary climate cycles.

Materials and Methods

Taxa and Data

We collected 79 samples that represent the 11 species of *Fouquieria* throughout their distributions, mainly in Mexico. We used representatives of genera *Camellia* (Theaceae), *Ilex* (Aquifoliaceae), and *Vaccinum* (Ericaceae) within Ericales as outgroup.

Data for phylogenetic analyses are the plastid DNA sequences of the *ndhF-rpl32* intergenic spacer, *rpl14-rps8-infA-rpl36* region, *rpl32-trnL* intergenic spacer, and the *3'rps16–5'trnK* intergenic spacer. DNA was extracted from leaf tissue using the DNeasy Plant Mini Kit (Qiagen, Valencia, California, USA). Polymerase chain reaction (PCR) and sequencing followed Shaw *et al.* (2007). Previously published and newly generated sequences were combined (see Supporting Information, Table S1). Sequence alignments are available in TreeBase (http://purl.org/phylo/treebase/phylows/study/TB2:S19546?x-access-code=88aefbe25cef3fdac73142b51fa12613&format=html).

Phylogenetic Analyses

The sequences of each locus were aligned with MUSCLE (Edgar, 2004), and subsequently adjusted by eye using Se-Al v.2.0a11 Carbon (Rambaut, 2002). The best-fit substitution model for each locus was identified with the Akaike Information Criterion (AIC) implemented in ModelTest v3.06 (Posada & Crandall, 1998).

Maximum likelihood analyses were performed in RAxML v.7.0.4 (Stamatakis, 2006). We implemented an independent general time reversible model (GTR) and a gamma distribution to account for among-site rate variation for each data partition, and 1,000 nonparametric bootstrap replicates to assess nodal support. We set 25 rate categories for the gamma distribution for each locus in the single locus analyses and for each partition in the concatenated matrix analyses because an exploratory analysis in RAxML showed

this number of categories lead to improved likelihood values. The ML tree was selected from the set of resulting trees on each search. We considered those nodes with ≥70% bootstrap support as being strongly supported (Hillis & Bull, 1993; Felsenstein, 2004).

Divergence Time Estimation

Estimation of divergence times among *Fouquieria* species was done in two steps. First, the age of steem and crown group in *Fouquieria* was estimated by using a phylogenetic tree derived from ML analysis of plastid DNA sequences (*atpB*, *matK*, and *rbcL* regions), and 18S and 26S nuclear ribosomal DNA for 83 species representing the main order-level clades within eudicots, with special emphasis on Ericales. *Ceratophyllum* (Ceratophyllaceae) was used as outgroup (Table S2). The crown node of *Fouquieria* was sampled by including *Fouquieria columnaris*, *F. fasciculata*, and *F. splendens*. Eudicot clades were constrained with fossil-derived minimum ages, and the tree was calibrated using the earliest fossil occurrence of tricolpate pollen grains (Hughes & McDougall, 1990; Doyle & Hotton, 1991) to constrain the eudicot stem node. Ages were estimated with the uncorrelated lognormal relaxed clock available in BEAST v1.4.8 (Drummond & Rambaut, 2007).

In the second step, the credibility interval of the *Fouquieria* crown node estimated in the first analysis was implemented as a prior for the root height of the *Fouquieria* phylogram (see above), and used to estimate divergence times within *Fouquieria*. Two internal nodes were constrained with fossil-derived minimum ages. First, a megafossil assignable to *Fouquieria splendens* from the Late Miocene Mint Canyon California formation (Graham, 2010), was used to assign a 5.4 Ma minimum age to a clade that

includes the woody species of *Fouquieria* (*F. burragei, F. diguetii, F. formosa, F. leonilae, F. macdougalii, F. ochoterenae, F. shrevei,* and *F. splendens*), which we hereafter refer to as the woody clade. Second, Van Devender (1990a, 1990b) and Van Devender et al. (1994) recognized plant fragments of *F. columnaris* in Quaternary (ca. 2.6 Ma) packrat middens, which we use to assign a minimum age to the stem node of this species.

Diversification Rate Analyses

We investigated *Fouquieria*'s long-term diversification dynamics using a likelihood-based model-fitting method from an R package for Phylogenetic Analyses of DiversificAtion (RPANDA; Morlon et al., 2011, 2016), which simultaneously accommodates undersampling of extant taxa, declining or increasing diversification rate variation through time and among groups and potential periods of negative diversification, i.e., when extinction events are more common than speciation ones. The method can evaluate diversification dynamics of an entire phylogenetic tree, of *a priori* selected clades, or of paraphyletic groups. Nevertheless, the method is not designed to automatically identify significant diversification rate changes among clades. To facilitate this analysis, it is necessary to obtain independent evidence of clades that might have undergone important diversification rate shifts. To accomplish this, we used the Birth-Death Likelihood method (BDL; Rabosky, 2006a) and Bayesian Analysis of Macroevolutionary Mixtures software (BAMM; Rabosky, 2014, 2016).

BDL is a likelihood method that can identify significant diversification shifts through a lineage's evolution. It selects among temporally constant or variable models, and models

that include or exclude extinction (i.e., birth-death or pure birth). It estimates the likelihood of a rate change through time, and provides estimates of diversification rates during each period. We conducted this test on all *Fouquieria*, this is, from the separation of *Fouquieria*'s stem lineage from its extant sister group, Polemoniaceae, 85.48 Ma, to the present. We selected the best-fitting model among several that differ by (1) being rate-constant or rate-variable through time; (2) allowing one or two temporal rate changes; and (3) encompassing extinction or not (Rabosky 2006a, 2006b). Model selection is based on the difference in AIC scores between the best-fitting rate-constant and rate-variable models (ΔAIC_{RC}). This test was conducted using yuleSim to determine the significance of the observed ΔAIC_{RC} statistic by simulating 1,000 trees with the same number of taxa as in the input tree and the speciation rate obtained under the pure-birth model. The BDL method was implemented with LASER v2.3 (Rabosky, 2006b), using a lineage through time (LTT) plot of *Fouquieria* species mean ages derived from the dated phylogeny obtained with BEAST.

BAMM uses a compound Poisson process (implemented through a reversible jump Markov Chain Monte Carlo, rjMCMC) to identify significant diversification (speciation and extinction) shifts among branches in a phylogenetic tree and along each branch through time. It generates a credible set of shift configurations, each representing a particular number of shifts in the tree on particular phylogenetic branches. A phylorate plot shows diversification rate values averaged across all the configurations in the posterior distribution for each time unit along each branch of the phylogeny. To distinguish between significant and negligible shifts, BAMM estimates the marginal odds ratio of a rate shift on each branch. We ran 10 million generations in BAMM 2.5.0

(Rabosky, 2016). Priors were obtained by implementing the setBAMM priors function in the BAMMtools v2.5.0 package (Rabosky, 2016) as follows: expectedNumberOfShifts = 1.0; lambdaInitPrior = 10.0291415949617; lambdaShiftPrior = 0.0134676466318751; and muInitPrior = 10.0291415949617. We repeated the analysis by doubling the prior on the expected number of shifts, but the results were equal. The phylogenetic tree was the *Fouquieria* chronogram modified to include a single terminal per species, and inserting a branch for Polemoniaceae, with the divergence time between the two families estimated in the eudicot-wide analysis in BEAST (above). Polemoniaceae species-richness was obtained from The Plant List (http://www.theplantlist.org/). The sampling probability of each *Fouquieria* species was specified as 1, and for Polemoniaceae, 1/455. The segLength parameter was adjusted so that each unit time was 3 Ma.

Considering the diversification rate shifts detected by BDL and BAMM (see Results), we implemented the birth-death models from RPANDA using the fit_bd function (Morlon et al. 2016) to estimate long-term diversification dynamics, and particularly to evaluate the presence of extinction along *Fouquieria*'s stem lineage. We conducted two RPANDA analyses: one applied to all *Fouquieria* and another applied only to its crown group. In both analyses, we evaluated a diversification dynamics shift associated to the woody clade (see Results) and compared four different diversification dynamics: (1) constant speciation and extinction, (2) variable speciation and constant extinction, (3) constant speciation and variable extinction, and (4) variable speciation and extinction, considering linear and exponential variation in rates in variable models. with the second order (corrected) Akaike information criterion (AICc) for small sample sizes (Burnham &

Anderson, 2002). All analyses were performed in R (R Core Development Team, 2011), with scripts provided by F. Condamine (CNRS, France).

Ancestral Area Reconstruction

The geographical distribution range of *Fouquieria* was divided in six areas: (1) Sonoran Desert, (2) Baja California Peninsular Desert, (3) Chihuahuan Desert, (4) Hidalguense Desert, (5) Tehuacán-Cuicatlán Desert and (6) Southwestern semidesert relicts. To infer ancestral areas, the Bayesian Binary MCMC (BBM) analysis approach was followed (implemented in Reconstruct Ancestral States in Phylogenies [RASP 2.1]; Yu *et al.*, 2014) using the 20,000 post-burn-in trees from the Bayesian inference analyses run in BEAST to estimate the >0.50 posterior probability of each node. The number of areas was kept at 5.

The possible ancestral ranges at each node were estimated on the chronogram derived from the maximum clade credibility tree estimated with the uncorrelated lognormal method in BEAST. Four MCMC chains were run simultaneously for 5 million generations and the reconstructed state was sampled every 1,000 generations. The JC + G (Jukes-Cantor + Gamma) model was fixed in BBM analysis with a null root distribution.

Geographic Phylogenetic Structure

The degree of geographic structure in the phylogeny was estimated using the isolation by distance approach, taken from population genetics (e.g., Grefen *et al.*, 2004), but applied to clades, as in Schrire *et al.* (2009). The isolation by distance hypothesis predicts a

positive relationship between geographic and phylogenetic distances when other ecological determinants do not structure genetic variation.

The geographical structure of the *Fouquieria* phylogeny was quantified using a Mantel regression approach with Euclidean geographical distances between pairwise comparisons of terminal taxa as the predictor variable and phylogenetic distance measured as the branch lengths in the ML phylogram as the response variable.

Results

Sequence Characteristics and Phylogeny Estimation

The combined molecular dataset consisted of the concatenated sequences of four molecular markers (4,157 aligned nucleotide positions) of 79 individuals belonging to the eleven recognized species of *Fouquieria*.

We identified ten well-supported phylogroups within *Fouquieria* (Fig.1, Supporting Information Fig. S1), nine of them corresponding to *Fouquieria columnaris, F. fasciculata, F. formosa, F. leonilae, F. macdougalii, F. ochoterenae, F. purpusii, F. shrevei,* and *F. splendens* and the remaining one corresponding to a lineage that included individuals from the putative species *F. diguetii* and *F. burragei*, with unresolved relationships among them. In the resulting phylogenetic tree, the earliest-diverging branch leads to *F. columnaris* and the second branch, to the sister pair *F. fasciculata* and *F. purpusii*. We refer to these branches as the “early succulent grade”. Then, there is a woody clade that contains *F. burragei, F. diguetii, F. formosa, F. leonilae, F. macdougalii, F. ochoterenae, F. shrevei,* and *F. splendens* (Fig.1).

Divergence Times Estimation

The uncorrelated lognormal dating analyses in BEAST based on plastid and nuclear sequences for a representative sample of eudicot lineages resulted in an effective sample size (ESS) >300 for all estimated parameters indicating adequate estimations. The mean age of the split between extant *Fouquieria* and its sister lineage Polemoniaceae is estimated as 85.48 Ma, with a 95% highest posterior density (HPD) ranging from 72.74 to 94.95 Ma (Supporting Information Fig. S2, Table S3). The mean age for the crown *Fouquieria* was substantially younger, estimated as 6.61 Ma (95% HPD 2.96-11.38; Supporting Information Fig. S2, Table S3).

The credibility interval on the age of crown *Fouquieria* estimated in the previous analysis was then used to calibrate the same node in the subsequent species-level analysis, using it as a uniform distribution with hard bounds. The ESS of estimated parameters from the combined MCMCs of the uncorrelated lognormal dating analyses for the *Fouquieria* species-level tree are >300. The mean age of crown *Fouquieria* was estimated as 7.21 Ma (95% HPD 9.58-5.12; Fig. 1, Supporting Information Fig. S3, Table S4). A summary of estimated ages of the species and major clades within *Fouquieria* is available in Table 1. The average of mean ages of *Fouquieria* species is 4.38 Ma (SD = 1.58).

Diversification Rate Analyses

The BDL analysis rejected the null hypothesis of temporally homogeneous diversification rates for all *Fouquieria* (ΔAIC_{RC} = 16.3600, P = 0.006, Table 2). The yule3rate model was identified as having the best fit to the data (AIC = 30.2867), including two diversification rate shifts and no extinction (Table 2). According to the scenario

suggested by the yule3rate model, *Fouquieria* started to diversify ca. 85.48 Myr ago, with an initial low rate (r_1 = 0.0063 spp. Myr^{-1}). A diversification shift took place at ts_1 = 7.21 Ma, increasing to a substantially higher rate (r_2 = 0.3332 spp. Myr^{-1}), followed by a third shift at ts_2 = 3.03 Ma, decreasing to a lower rate (r_3 = 0.0304 spp. Myr^{-1}) (Table 2, Fig. 2a,b).

The 95% credibility interval of the posterior distribution of the BAMM analysis contains a single shift configuration. This configuration includes zero significant diversification dynamics shifts, but a strong temporal increase in the rate of speciation (and of diversification) immediately before the *Fouquieria* crown group, and a gradual increase in the rate of extinction along the stem lineage, which abruptly decreases to a moderate rate immediately before the crown group (Fig. 2c). According to BAMM, *Fouquieria*'s stem lineage and its crown group *Fouquieria* are thus characterized by different diversification rates (Fig. 2c).

To encompass the diversification changes detected by BDL and BAMM through *Fouquieria*'s evolution, we implemented two RPANDA analyses to investigate the diversification dynamics of *Fouquieria*: one including *Fouquieria*'s stem lineage (all *Fouquieria* analyses) and another only from its crown group (crown *Fouquieria* analyses). In each analysis, we compared the case in which no change in diversification dynamics was allowed with the case in which we allowed distinct diversification dynamics for the stem lineage plus the early succulent grade (i.e., *F. columnaris*, *F. fasciculata* and *F. purpusii*), and for the woody clade in the all *Fouquieria* analysis; and for the early succulent grade, and for the woody clade in the crown *Fouquieria* analysis. Models with a shift in diversification dynamics in the woody clade (Fig. 3) are supported

against models with no changes in diversification along *Fouquieria*'s evolutionary history (Supporting Information Fig. S4) with AICc and ML (Table 3, Supporting Information Table S5)

All and crown *Fouquieria* analyses indicates different diversification dynamics between the stem lineage and the crown group (Tables 3 and 4; Fig. 3a and 3b; Supporting Information Fig. S4). Diversification rate estimates for *Fouquieria*'s stem lineage only would be desirable. However, the scarcity of data from this period of *Fouquieria*'s evolution does not allow RPANDA (or other methods that we know of) to appropriately infer its diversification dynamics. Nevertheless, we consider that the RPANDA all *Fouquieria* analysis provides an adequate proxy of *Fouquieria*'s stem lineage diversification dynamics (Fig. 3c; Supporting Information Fig. S4a).

The best diversification models estimated for the full evolutionary history of *Fouquieria* are constant speciation rate (lambda, λ= 0.22) and linearly decreasing extinction rate (mu at the origin of the stem lineage, μ= 2.7 and mu at present, μ_p= 0.01) with associated linearly increasing diversification (r at the origin, r_o= -2.6 and r at present, r_p= -0.001) for the stem lineage plus the early succulent grade (Fig. 3c; Supporting Information Table S5); exponentially decreasing speciation (lambda at the origin, λ_o= 3.2 and lambda at present, λ_p= 0.014) and constant extinction that is very close to zero (mu, μ= 1.1e-8) with associated exponentially decreasing diversification for the woody clade (Fig. 3e; Supporting Information Table S6).

For crown *Fouquieria*, the best diversification models are exponentially decreasing speciation (λ_o= 0.19, λ_p= 0.042) and constant extinction that is very close to zero (mu, μ= 1.0e-9) with associated exponentially decreasing diversification for the early succulent

grade (Fig. 3d; Supporting Information Table S7) and the same diversification dynamics obtained in the all *Fouquieria* analysis for the woody clade, albeit with different absolute rate magnitudes (Fig. 3e; Supporting Information Table S6).

Ancestral Area Reconstruction

Results of the Bayesian Binary Analysis MCMC (BBM) suggest that vicariance played a substantial role in the biogeographic history of *Fouquieria* (Fig. 1). The BBM analysis indicates the Baja California Peninsular Desert as the ancestral area for *Fouquieria* as a whole (P = 0.99), which is retained in *F. columnaris* (P = 0.99). An initial vicariance event at 6.6 Ma is estimated to have separated the clade that includes *F. fasciculata* plus *F. purpusii*, which occupy continental deserts, from the woody clade, which retained an ancestral Peninsular Desert distribution with moderate probability (P = 0.50).

A second vicariance event is estimated at ca. 3.37 Ma, separating *F. purpusii* and *F. fasciculata*, which occupy the Tehuacán-Cuicatlán Desert and the Hidalguense Desert, respectively. The woody clade retained a Baja California Peninsular Desert distribution until ca. 4.79 Ma, after which several vicariance events took place, presumably underlying speciation. Within the woody clade, the lineage including *F. burragei* and *F. diguetii* preserved a Baja California Peninsular Desert distribution (P = 0.99), but its sister group, which includes *F. leonilae* and *F. ochoterenae*, occupied the SW semidesert relicts at ca. 4.02 Ma. The lineage that includes *F. formosa*, *F. macdougalii*, *F. shrevei* and *F. splendens* occupied continental deserts at 4.27 Ma, with highest probability of initial occupation of the SW relict deserts (P = 0.99), and *F. formosa* retaining this distribution (P = 0.61).

A subsequent vicariance event is estimated at 3.56 Ma, separating *F. macdougalii*, which occupies the Sonoran Desert, from a clade with an ancestral distribution in the Chihuahuan Desert (P = 0.68) including *F. shrevei* and *F. splendens*. The latter species underwent dispersal to the Sonoran, Baja California Peninsular, and Hidalguense Deserts.

Geographic Phylogenetic Structure

The Mantel regression found a positive relationship between geographical distance and clade ages (r = 0.4746, R^2 = 0.2252; p = 0.001; Fig. 3). The isolation by distance hypothesis is corroborated by a positive relationship between geographic and phylogenetic distances.

Discussion

Time and Mode of Diversification

The origin of *Fouquieria* is here estimated in the late Cretaceous (Coniacian; 85.48 Ma). This ancient age agrees with previous results that estimated origin of the main lineages of Ericales in the Mid-Late Cretaceous (Bremer *et al.*, 2004; Sytsma *et al.*, 2006). However, the diversification of their extant species occurred much later, at the end of Miocene, ca. 7.21 Ma (9.58-5.12, 95% HPD; Fig. 1, Fig. 2, Fig. S2). The lineages in the early succulent grade, which includes insect-pollinated species, diverged between 7.21 and 6.60 Ma. The woody clade started to diversify at the end of the Miocene, ca. 6.6 myr ago. It includes mostly hummingbird-pollinated species (Henrickson, 1972).

According to our results, *Fouquieria* is an ancient lineage possibly preadapted to seasonally dry environments that underwent a maximum speciation burst starting

approximately during the transition between the Miocene and the Pliocene. BDL analyses detected a drastic diversification rate shift at ca. 7.2 Ma (Messinian, late Miocene), from a very low stem rate of 0.006 sp myr^{-1} to a higher rate of 0.333 sp myr^{-1}. Concordantly, the BAMM analysis indicated a strong speciation (and diversification) increase shortly before *Fouquieria*'s crown radiation, at ca. 7.35 Ma. This agrees with hypotheses that some of the lineages in North American deserts diversified as early as the late Miocene to Pliocene, and not during the Pleistocene (Riddle & Hafner, 2006; Wilson & Pitts, 2010b; Hafner & Riddle, 2011). Long-term diversification dynamics estimated with RPANDA suggest that extinction also played a significant role in *Fouquieria*'s evolution, with a very high rate at the onset of the process. A possible explanation for the lack of fossil record in *Fouquieria* previous to the late Miocene is the low probability of fossil preservation in dry environments, where the lineage occurred.

Neogene Vicariant Speciation

Our results suggest that vicariance in the Neogene was the primary driver of speciation among lineages in *Fouquieria*, rather than an isolation event in the Pleistocene. According to Wilson and Pitts (2010b), paleobiological evidence suggests that Neogene uplift events created a rain-shadow effect over most of western North America, therefore leading to the formation of the different desert regions, representing the major driving factor in the diversification of a unique North American arid-adapted biota. Divergence time estimates, ancestral area reconstructions, and geographic phylogenetic structure, all are consistent with a scenario in which Neogene orogenesis played an important role in *Fouquieria* diversification, along with tectonic events related to the early development of

regional deserts (e.g., mountain and plateau uplifting, rifting along major fault systems). The northern extension of the Gulf of California known as the Bouse Sea or Bouse Embayment, during Miocene-Pliocene (~8–4 Ma) created a barrier isolating the Baja California peninsula and the westernmost part of southern California from the remainder of the Sonoran Desert (Metzger, 1968; Lucchitta, 1972; Blair, 1978; Eberly and Stanley, 1978; Boehm, 1984; Ingle, 1987; Buising, 1990; McDougal et al., 1999; Carreño and Helenes, 2002). The elevation in the central and south Mexico changed the Sierra Madre Occidental and formed the Trans-Mexican Volcanic Belt during Oligocene-Miocene, events that separated the central-north part of Mexico (Ferrari et al., 1999; Cevallos et al., 2012), and promoted the differentiation of Chihuahuan Desert and the SW semidesert relicts. Speciation in *Fouquieria* seems to follow the intricate history associated to the main desert areas in Mexico.

BBM analyses point to the Baja California Peninsular Desert as the ancestral area of the genus *Fouquieria* during the late Miocene 7.21 Ma. Numerous lines of evidence now support the evolutionary differentiation of the Peninsular from the Sonoran Desert and further indicate that the Peninsular Desert underwent an independent evolutionary trajectory before the separation of the Sonoran and Chihuahuan Deserts (Hafner & Riddle, 2011). The Baja California peninsula was formerly connected to the Mexican mainland, and is generally accepted to have split from the mainland around 6 myr ago (Oskin & Stock, 2003a, 2003b, 2003c; Wilson & Pitts, 2010b). It is likely that several extant species belonging to lineages that currently occupy the southern Baja California Peninsula are relicts of populations that occupied the former distribution before the separation of the Peninsular region from the Mexican mainland 7-10 myr ago, including

plants (Roberts, 1989), reptiles (Murphy, 1983; Grismer, 2002; Murphy & Aguirre-León, 2002), birds (Cody, 1983; Cody & Velarde, 2002), insects (Truxal, 1960), spiders (Chamberlin, 1924), and scorpions (Williams, 1980; Gantenbein *et al*., 2001; Sissom & Hendrixson, 2005).

The ancestral *Fouquieria* could represent an ancient element preadapted to aridity derived from the California chaparral and other communities, as postulated by Axelrod (1979, 1983), as the possible Tertiary vegetation from which North American deserts originated. The first vicariant event in *Fouquieria* took place during the Late Miocene (6.6 Ma), when the aridification trend reached its peak (5-8 Ma), resulting from decreasing precipitation (e.g., Wilson & Pitts, 2010b). The vicariance event that separated the succulent *F. fasciculata* plus *F. purpusii* clade from the woody clade could be associated with the secondary uplift of the Sierra Madre Occidental, which took place between the Late Miocene and the Early Pliocene. It has been widely documented that this orogenesis drove the divergence of several taxa in rodents (Riddle & Hafner, 2006; Hafner & Riddle, 2011; Bell *et al*., 2012), reptiles (Jaeger *et al*., 2005; Bryson *et al*., 2011, 2012b; Anderson & Greenbaum, 2012; Bryson & Riddle, 2012), and plants (Moore & Jansen, 2006; Reberning *et al*., 2010; Loera *et al*., 2012; Gándara & Sosa, 2014).

Three major vicariant events in *Fouquieria* seem to be associated with the second volcanic episode in the Late Miocene (7.5–3 Ma), promoting speciation of sister lineages, north and south of the Trans-Mexican Volcanic Belt, namely, (1) *F. formosa* in the SW semidesert relicts, diverging at 4.27 Ma from the northern clade containing *F. macdougalii*, *F. shrevei* and *F. splendens*; (2) *F. fasciculata*, in the Hidalguense Desert, diverging at 3.37 Ma from the southern *F. purpusii* in the Tehuacán-Cuicatlán Desert;

and (3) the clade formed by the narrow endemics *F. ochoterenae* and *F. leonilae*, from SW semidesert relicts, diverging at 4.02 Ma from the northern clade of *F. diguetii* and *F. burragei* in the Baja California Peninsular Desert. Recently, an increasing number biogeographic and phylogeographic studies of taxa that are co-distributed both south and north of the Trans-Mexican Volcanic Belt have identified similar spatio-temporal divergence patterns, suggesting that this mountain chain has been an important barrier influencing the diversification of several lineages of fish (Hulsey *et al*., 2004), reptiles (Bryson, 2011a, 2012a, 2012b, 2012c), and plants (Sosa *et al*., 2009; Ruíz-Sanchez *et al*., 2012; Ruíz-Sanchez & Specht, 2013; Gándara & Sosa, 2014).

The final vicariant event took place during the middle Pliocene at 3.56 Ma, when *Fouquieria macdougalii*, which actually occupies the Sonoran Desert, diverged from the clade formed by *F. shrevei* and *F. splendens*, distributed in the Chihuahuan Desert.

An interesting speciation event that apparently does not imply vicariance underlies the separation between *Fouquieria shrevei* and *F. splendens* during the late Pliocene (3.03 Ma) in the Chihuahuan Desert, forming a sympatric sister pair. It has been recently argued that *F. shrevei*, as a narrow endemic, underwent intense genetic drift and reduced gene flow, in which differentiating populations followed the island pattern of gypsum deposits in the Chihuahuan Desert (Aguirre-Liguori *et al.*, 2014), suggesting that some narrowly endemic species of *Fouquieria* result from edaphic adaptation. The current distribution of *F. splendens*, as the only broadly distributed species, implies dispersal processes from an ancestral distribution in the Chihuahuan Desert, to the Peninsular (0.78 Ma), Hidalguense (1.21 Ma), and Sonoran (0.78 Ma) Deserts. These dispersals could be associated to the pluvial and inter-pluvial cycles of the Middle Pleistocene. During the

pluvial periods, the xeric flora of the Chihuahuan Desert contracted and remained in southern refugia, while its area was covered by paleolakes, and expanded during the inter-pluvial periods (Van Devender, 1990a; Riddle & Hafner, 2006; Hafner & Riddle, 2011).

Regardless of age, a high degree of phylogenetic geographic structure with sister species tending to occupy the same nucleus of distribution is indicative of highly limited historical dispersal between distribution nuclei (Pennington *et al.*, 2006; Pennington *et al.*, 2009). The high levels of geographic phylogenetic structure found (Fig. 4) also agree with the previously argued ideas, where isolation by distance is a strong driver of genetic variation structure, indicating high levels of dispersal limitation in most *Fouquieria* species.

Conclusions

A Mio-Pliocene diversification in *Fouquieria* associated with Neogene orogenesis during the early development of regional deserts is strongly supported. Our results show that *Fouquieria* is an ancient lineage preadapted to seasonally dry environments, having a maximum burst of diversification during the Pliocene.

The reconstructed evolutionary history of *Fouquieria* indicates that it underwent very limited diversification until the Late Miocene, involving very low extinction. This pattern agrees with the scant fossil record during this time. The processes implied in the floristic assembly of the deserts in North America could have accelerated its diversification during the Neogene, resulting in the distribution of sister clades on different desert areas, several of which are separated by the Trans-Mexican Volcanic Belt.

Geographic phylogenetic structure shows isolation by distance as the main driver structuring the genetic variation in the species. Ancestral area reconstruction shows the Peninsular Desert as the ancestral area for *Fouquieria*, and a vicariant-dispersal history where most restricted or endemic species are older than widespread species.

Acknowledgements

The Coordinación de la Investigación Científica, Universidad Nacional Autónoma de México, provided postdoctoral funding to JADN. LLSR thanks the Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México and the Consejo Nacional de Ciencia y Tecnología, México, CONACYT scholarship 262540 for providing funding during her PhD studies. This research was partially funded by grants CONACyT 2004-C01-46475 to LEE; and PAPIIT-UNAM 202310 to SM. We thank J. Aguirre-Liguori for providing *Fouquieria shrevei* samples, M. Vásquez-Cruz for technical support in the lab and F. Condamine for providing R code for diversification analyses.

Literature Cited

Aguirre-Liguori JA, Scheinvar E, Eguiarte LE. 2014. Gypsum soil restriction drives genetic differentiation in Fouquieria shrevei (Fouquieriaceae). American Journal of Botany 101: 730-736.

Anderson CG, Greenbaum E. 2012. Phylogeography of the northern populations of black-tailed (Crotalus molossus Baird and Girard, 1853), with the revelation of C. ornatus Hallowell 1854. Herpetological Monographs 26: 19-57.

Axelrod DI. 1979. Age and origin of Sonoran Desert vegetation. Occasional Papers of the California Academy of Sciences 132: 1-74.

Axelrod DI. 1983. Paleobotanical history of the western deserts. In: Wells SG, Haragan DR, eds. Origin and Evolution of Deserts. Albuquerque, USA: University of New Mexico Press, 113-129.

Bell KC, Hafner DJ, Leitner P, Matocq MD. 2012. Phylogeography of the ground squirrel subgenus Xerospermophilus and assembly of the Mojave Desert biota. Journal of Biogeography 37: 363-378.

Blair W.N. 1978. Gulf of California in Lake Mead area in Arizona and Nevada during the late Miocene time. Bulletin of the American Association of Petroleum Geologists 62: 1159-70.

Boehm MC. 1984. An overview of the lithostratigraphy, biostratigraphy, and paleoenvironments of the Late Neogene San Felipe marine sequence, Baja California, Mexico. In: Frizzell VA ed. Geology of the Baja California peninsula, volume 39, Los Angeles, CA: Pacific Section, Society of Economic Paleontologists and Mineralogists, 253-65.

Bremer B, Bremer K, Heidari N, Erixon P, Olmstead RG, Anderberg AA, Källersjö M, Barkhordarian E. 2002. Phylogenetics of asterids based on three coding and three noncoding chloroplast DNA markers and the utility of noncoding DNA at higher taxonomic levels. Molecular Phylogenetics and Evolution 24: 274-301.

Bremer K, Friis EM, Bremer B. 2004. Molecular phylogenetic dating of asterid flowering plants shows early Cretaceous diversification. Systematic Biology 53: 496-505.

Bryson Jr. RW, Riddle BR. 2012. Tracing the origins of widespread highland species: a case of Neogene diversification across the Mexican sierras in an endemic lizard. Biological Journal Linnean Society 105: 382-394.

Bryson Jr. RW, Murphy RW, Lathrop A, Lazcano-Villareal D. 2011. Evolutionary drivers of phylogeographical diversity in the highlands of Mexico: a case study of the Crotalus triseriatus species group of montane rattlesnakes. Journal of Biogeography 38: 697-710.

Bryson Jr. RW, García-Vázquez UO, Riddle BR. 2012a. Diversification in the Mexican horned lizard Phrynosoma orbiculare across a dynamic landscape. Molecular Phylogenetics and Evolution 62: 87-96.

Bryson Jr. RW, García-Vázquez UO, Riddle BR. 2012b. Relative roles of Neogene vicariance and Quaternary climate change on the historical diversification of bunchgrass lizards (Sceloporus scalaris group) in Mexico. Molecular Phylogenetic and Evolution 62: 447-457.

Bryson Jr. RW, Jaeger JR, Lemos-Espina JA, Lazcano D. 2012c. A multilocus perspective on the speciation history of a North American aridland toad (Anaxyrus punctatus). Molecular Phylogenetic and Evolution 64: 393-400.

Buising AV. 1990. The Bouse formation and bracketing units, southeastern California and western Arizona – implications for the evolution of the proto-gulf of California and the lower Colorado River. Journal of Geophysical Research – Solid Earth and Planets 95: 20111–20132.

Burnham KP, DR Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York.

Carreño AL, J Helenes. 2002. Geology and ages of the islands. In: Case TJ, ML Cody, E Ezcurra, eds. A new island biogeography of the Sea of Cortés. New York: Oxford University Press, 14–40.

Cevallos-Ferriz SRS, E.A. González-Torres, L. Calvillo-Canadell. 2012. Perspectiva paleobotánica y geológica de la biodiversidad en México. Acta Botanica Mexicana 100: 317-350.

Chamberlin RV. 1924. Expedition of the California Academy of Sciences to the Gulf of California in 1921 - The spider fauna of the shores and islands of the Gulf of California. Proceedings of the California Academy of Sciences 12: 561-694.

Cody ML. 1983. The land birds. In: Case TJ, Cody ML, eds. Island Biogeography in the Sea of Cortéz. Berkeley USA: University of California Press, 210–245.

Cody ML, Velarde E. 2002. Land birds. In: Case TJ, Cody ML, Ezcurra E, eds. A New Island Biogeography of the Sea of Cortéz. New York, USA: Oxford University Press, 271–312.

De-Nova JA, Medina R, Montero JC, Weeks A, Rosell JA, Olson ME, Eguiarte LE & Magallón S. 2012. Insights into the historical construction of species- rich Mesoamerican seasonally dry tropical forests: the diversification of Bursera (Burseraceae, Sapindales). New Phytologist 193: 276-276.

Doyle J, Hotton C. 1991. Diversification of early angiosperm pollen in a cladistic context. In: Blackmore S, Barnes S, eds. Pollen and Spores: Pattern of Diversification. Oxford: Clarendon Press, 169–195.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7:214.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792-1797.

Eberly LD, TB Jr Stanley. 1978. Cenozoic stratigraphy and geologic history of southeastern Arizona. Geological Society of America Bulletin 89: 921-40.

Felsenstein J. 2004. Inferring phylogenies. Sunderland, MA: Sinauer Associates.

Ferrari L, M. López-Martinez, G Aguirre-Díaz, G Carrasco-Nuñez. 1999. Space time patterns of Cenozoic arc volcanism in central Mexico: from the Sierra Madre Occidental to Mexican Volcanic Belt. Geology 27: 303-306.

Findley JS. 1969. Biogeography of southwestern boreal and desert mammals. University of Kansas, Miscellaneous Publications of the Museum of Natural History 51: 113–128.

Gantenbein B, Fet V, Barker MD. 2001. Mitochondrial DNA reveals a deep, divergent phylogeny in Centruroides exilicauda (Wood, 1863) (Scorpiones: Buthidae). In: Fet V, Selden PA, eds. Scorpions 2002. In Memoriam Gary A. Polis. Burnham Beeches, UK: British Arachnological Society, 235-244.

Gándara E, Sosa V. 2014. Spatio-temporal evolution of Leucophyllum pringlei and allies (Scrophulariaceae): A group endemic to North American xeric regions. Molecular Phylogenetics and Evolution 76: 931-101.

Graham A. 2010. Late Cretaceous and Cenozoic history of Latin American vegetation and terrestrial environments. St. Louis, Missouri: Missouri Botanical Garden Press.

Grefen E, Anderson MJ, Wayne RK. 2004. Climate and habitat barriers to dispersal in the highly mobile gray wolf. Molecular Ecology 13: 2481-2490

Grismer LL. 1994. The origin and evolution of the peninsular herpetofauna of Baja California, Mexico. Herpetological Natural History 2: 51-106.

Grismer LL. 2002. A re-evaluation of the evidence for a mid-Pleistocene mid-peninsular seaway in Baja California: A reply to Riddle et al. Herpetological Review 33: 15-16.

Hafner DJ, Riddle BR. 2005. Mammalian phylogeography and evolutionary history of northern Mexico's deserts. In: Cartron JLE, Ceballos G, Felger RS eds. Biodiversity, Ecosystems, and Conservation in Northern Mexico. New York, USA: Oxford University Press, 225-245.

Hafner DJ, Riddle BR. 2011. Boundaries and barriers of North American warm deserts: an evolutionary perspective. In: McGowan A, Upchurch A, Slater PC, eds. Palaeogeography and Palaeobiogeography Biodiversity in Space and Time. Boca Raton: Systematics Association Special Volume Series, CRC Press, 75-114.

Henrickson J. 1972. A taxonomic revision of the Fouquieriaceae. Aliso 7: 439-537.

Hernández HM, Gómez-Hinostrosa C. 2011. Areas of endemism of Cactaceae and the effectiveness of the protected area network in the Chihuahuan Desert. Oxyx 45: 191-200.

Hernández HM, Gómez-Hinostrosa C, and Bárcenas R. 2001. Diversity, spatial arrangement, and endemism of Cactaceae in the Huizache area, a hot-spot in the Chihuahuan Desert. Biodiversity and Conservation 10: 1097-1112.

Hernández-Hernández T, Brown JW, Schlumpberger BO, Eguiarte LE, Magallón S. 2014. Beyond aridification: multiple explanations for the elevated diversification of cacti in the New World Succulent Biome. New Phytologist 202: 1382-397.

Hillis DM, Bull JJ. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Systematic Biology 42: 182-192.

Hubbard JP. 1973. Avian evolution in the aridlands of North America. Living Bird 12: 155-196.

Hughes NF, McDougall AB. 1990. Barremian-Aptian angiospermoid pollen records from southern England. Review of Palaeobotany and Palynology 65: 145–151.

Hulsey CD, García de León FJ, Johnson YS, Hendrickson DA, Near TJ. 2004. Temporal diversification of Mesoamerican cichlid fishes across a majorboundary. Molecular Phylogenetics and Evolution 31: 754-764.

Ingle JC Jr. 1987. Paleoceanographic evolution of the Gulf of California: foraminiferal and lithofacies evidence. Abstracts with Programs, Geological Society of America 19: 721.

Jaeger JR, Riddle BR, Bradford DF. 2005. Cryptic Neogene vicariance and Quaternary dispersal of the red-spotted toad (Bufo punctatus): insights on the evolution of North American warm deserts. Molecular Ecology 14: 3033-3048.

Krings A. 2000. A phytogeographical characterization of the vine flora of the Sonoran and Chihuahuan deserts. Journal of Biogeography 27: 1311-1319.

Leopold AS. 1950. Vegetation zones of México. Ecology 31: 507-518.

Loera I, Sosa V, Ickert-Bond SM. 2012. Diversification in North American arid lands: niche conservatism, divergence and expansion of habitat explain speciation in the genus Ephedra. Molecular Phylogenetics and Evolution 65: 437-450.

Lucchitta I. 1972. Early history of the Colorado River in the Basin and Range Province. Geological Society of America Bulletin 83: 1933-1948.

Magallón S, Sanderson MJ. 2005. Angiosperm divergence times: the effect of genes, codon positions, and time constraints. Evolution 59: 1653-1670

McCain CM. 2003. North American desert rodents: a test of the mid-domain effect in species richness. Journal of Mammalogy 84: 967-980.

McDougall KA, RZ Poore, JC Matti. 1999. Age and paleoenvironment of the Imperial Formation near San Gorgonio Pass, Southern California. The Journal of Foraminiferal Research 29: 4-25.

Metzger DG. 1968. The Bouse Formation (Pliocene) of the Parker-Blythe-Cibola area, Arizona and California. Geological Survey Research 1968: US Geological Survey Professional Paper 600-D, D126-36.

Morlon H, Parsons TL, Plotkin JB. 2011. Reconciling molecular phylogenies with the fossil record. Proceedings of the National Academy of Sciences USA 108: 16327-16332.

Morlon H, Lewitus E, Condamine FL, Manceau M, Clavel J, Drury J. 2016. RPANDA: An R package for macroevolutionary analyses on phylogenetic trees. Methods in Ecology and Evolution.

Moore MJ, Jansen RK. 2006. Molecular evidence for the age, origin, and evolutionary history of the American desert plant genus Tiquilia (Boraginaceae). Molecular Phylogenetics Evolution 39: 668-687.

Morafka DJ. 1977. A Biogeographical Analysis of the Chihuahuan Desert Through its Herpetofauna. USA: Dr. W. Junk B.V. Publishers, The Hague.

Murphy RW. 1983. Paleobiogeography and genetic differentiation of the Baja California herpetofauna. Occasional Papers of the California Academy of Sciences 137: 1-48.

Murphy RW, Aguirre-Leon G. 2002. Nonavian reptiles; origins and evolution. In: Case TJ, Cody ML, Ezcurra E, eds. A New Island Biogeography of the Sea of Cortes. New York, USA: Oxford University Press, 181-220.

Ortega-Gutiérrez F, Guerrero-Garcia JC. 1982. The geologic regions of Mexico. In: Palmer AR ed. Perspectives in Regional Geological Synthesis. Boulder: CO: Geological Society of America, 99-104.

Oskin M, Stock J. 2003a. Marine incursion synchronous with plate boundary localization in the Gulf of California. Geology 31: 23-26.

Oskin M, Stock J. 2003b. Pacific-North America plate motion and opening of the Upper Delfín basin, northern Gulf of California, Mexico. Geological Society of America Bulletin 115: 1173-1190.

Oskin M, Stock J. 2003c. Cenozoic volcanism and tectonics of the continental margins of the Delfin basin, northern Gulf of California, Mexico. In Johnson SE, Paterson SR, Fletcher JM, Girty GH, Kimbrough DL, Martin-Barajas A, eds. Tectonic evolution of northwestern México and the southwestern USA. Geological Society of America Special Paper 374, Boulder, CO: Geological Society of America, 421-38.

Pennington RT, Richardson JA, Lavin M. 2006. Insights into the historical construction of species-rich biomes from dated plant phylogenies, phylogenetic community structure and neutral ecological theory. New Phytologist 172:605-616

Pennington RT, Lavin M, Oliveira-Filho A. 2009. Woody plant diversity, evolution and ecology in the tropics: perspectives from seasonally dry tropical forests. The Annual Review of Ecology, Evolution, and Systematics 40: 437-457.

Posada D, Crandall KA. 1998. Modeltest: Testing the model of DNA substitution. Bioinformatics 14:817-818.

Rabosky DL. 2006a. Likelihood methods for inferring temporal shifts in diversification rates. Evolution 60:1152-1164.

Rabosky DL. 2006b. LASER: a maximum likelihood toolkit for detecting temporal shifts in diversification rates from molecular phylogenies. Evolutionary Bioinformatics Online 2: 247–250.

Rambaut A. 2002. SE-AL v.2.0a11: sequence alignment editor.Oxford, UK: University of Oxford. http://tree.bio.ed.ac.uk/software/seal.

Reberning CA, Schneeweiss GM, Bardy KE, Schönswetter P, Villaseñor JL, Obermayer R, Stuessy TF. 2010. Multiple Pleistocene refugia and Holocene range expansion of an abundant southwestern American desert plant species (Melampodium leucanthum, Asteraceae). Molecular Ecology 19: 3421-3443.

Riddle BR. 1995. Molecular biogeography in the pocket mice (Perognathus and Chaetodipus) and grasshopper mice (Onychomys)-the Late Cenozoic development of a North American aridlands rodent guild. Journal of Mammalogy 76: 283-301.

Riddle BR, Hafner DJ. 2006. A step-wise approach to integrating phylogeographic and phylogenetic biogeographic perspectives on the history of a core North American warm deserts biota. Journal of Arid Environments 66: 435-461.

Roberts NC. 1989. Baja California Plant Field Guide. La Jolla, CA: Natural History.

Ruíz-Sanchez E, Specht CD. 2013. Influence of the geological history of the Trans-Mexican Volcanic Belt on the diversification of Nolina parviflora (Asparagaceae: Nolinoideae). Journal of Biogeography 40: 1336-1347.

Ruíz-Sanchez E, Rodriguez-Gomez F, Sosa V. 2012. Refugia and geographic barriers of populations of the desert poppy, Hunnemannia fumariifolia (Papaveraceae). Organisms Diversity & Evolution 12: 133-143.

Rzedowski J. 1993. Diversity and origins of the phanerogamic flora of Mexico. In: Ramamoorthy TP, Bye R, Lot A, Fa J, eds. Biological diversity of Mexico: origins and distribution. Oxford, UK.: Oxford University Press 129-148.

Schrire BD, Lavin M, Barker NP, Forest F. 2009. Phylogeny of the tribe Indigofereae (Leguminosae-Papilionoideae): geographically structured more in succulent-rich and temperate settings than in grass rich environments. American Journal of Botany 96: 816-852.

Schönenberger J, Anderberg AA, Sytsma KJ. 2005. Molecular phylogenetics and patterns of floral evolution in the Ericales. International Journal of Plant Sciences 166: 265-288.

Schönenberger J, von Balthazar M, Sytsma KJ. 2010. Diversity and evolution of floral structure among early diverging lineages in the Ericales. Philosophical Transactions of the Royal Society B: Biological Sciences 365: 437-448.

Schultheis LM, Baldwin BG. 1999. Molecular phylogenetics of Fouquieriaceae: evidence from nuclear rDNA ITS studies. American Journal of Botany 86: 578-589.

Shaw J, Lickey E, Schilling E, Small R. 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in Angiosperms: The tortoise and the hare III. American Journal of Botany 94: 275-288.

Shreve F. 1942. The desert vegetation of North America. The Botanical Review 8: 195-246.

Sissom WD, Hendrixson BE. 2005. Scorpion biodiversity and patterns of endemism in northern Mexico. In: Cartron JLE, Ceballos G, Felger RS, eds. Biodiversity, Ecosystems,

and Conservation in Northern Mexico. New York, USA: Oxford University Press, 122-137.

Sosa V, De-Nova JA. 2012. Endemic angiosperm lineages in Mexico: hotspots for conservation. Acta Botánica Mexicana 100: 293-315.

Sosa V, Ruiz-Sanchez W, Rodriguez-Gomez F. 2009. Hidden phylogeographic complexity in the Sierra Madre Oriental: the case of the Mexican tulip poppy Hunnemannia fumariifolia (Papaveraceae). Journal of Biogeography 36:18-37.

Stamatakis A. 2006. RAxML-VI-HPC: Maximum Likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688-2690.

Sytsma KJ, Walker JB, Schönenberger J, Anderberg AA. 2006. Phylogenetics, biogeography, and radiation of Ericales. Annual BSA Meeting, Botany 2006 Chico, CA: Abstract 71.

Truxal FS. 1960. The entomofauna with special reference to its origin and affinities. Symposium: The biogeography of Baja California and adjacent seas. Systematic Zoology 9:165-170.

Upton DE, Murphy RW. 1997. Phylogeny of the side-blotched lizards (Phrynosomatidae: Uta) based on mtDNA sequences: support for a midpeninsular seaway in Baja California. Molecular Phylogenetics and Evolution 8: 104-113.

Van Devender T.R. 1990a. Late Quaternary vegetation and climate of the Chihuahuan Desert, United States and Mexico. In: Betancourt JL, Van Devender TR, Martin PS, eds. Packrat middens: the last 40,000 years of biotic change. Tucson, Arizona, USA: The University of Arizona Press, 104-133.

Van Devender TR. 1990b. Late Quaternary vegetation and climate of the Sonoran Desert, United States and Mexico. In: Betancourt JL, Van Devender TR, Martin PS, eds. Packrat middens: the last 40,000 years of biotic change. Tucson, Arizona, USA: The University of Arizona Press, 134-165.

Van Devender TR, Burgess TL, Piper JC, Turner RM. 1994. Paleoclimatic Implications of Holocene Plant Remains from the Sierra Bacha, Sonora, Mexico. Quaternary Research 41: 99-108.

Williams SC. 1980. Scorpions of Baja California, Mexico, and adjacent islands. Occasional Papers of the California Academy of Sciences 135: 1-127.

Wilson JS, Pitts JP. 2010a. Phylogeographic analysis of the nocturnal velvet ant genus Dilophotopsis (Hymenoptera: Mutillidae) provides insights into diversification in the Neartic deserts. Biological Journal of the Linnean Society 101: 360-375.

Wilson JS, Pitts JP. 2010b. Illuminating the lack of consensus among descriptions of earth history data in the North American deserts: a resource for biologists. Progress in Physical Geography 34: 419-441.

Yu Y, Harris AJ, He XJ. 2014. RASP (Reconstruct Ancestral State in 748 Phylogenies) 3.0. Available at: <http://mnh.scu.edu.cn/soft/blog/RASP>.

Table 1. Summary of estimated ages for phylogroups and major clades within *Fouquieria.*

	Stem group mean age (95% HPD)	**Crown group mean age (95% HPD)**
Fouquieria	85.48 (94.95-72.74)	7.21 (9.58-5.13)
F. columnaris	7.21 (9.58-5.13)	2.51 (3.82-1.36)
F. fasciculata	3.37 (4.97-2.00)	2.98 (4.44-1.66)
F. purpusii	3.37 (4.97-2.00)	2.53 (3.78-1.43)
Woody clade	6.6 (8.68-4.77)	4.79 (5.89-4.06)
F. burragei-F. diguetii	4.02 (5.41-2.67)	2.84 (4.16-1.66)
F. splendens	3.03 (4.23-1.80)	2.03 (3.12-1.01)
F. leonilae	2.58 (3.89-1.38)	1.95 (3.14-0.87)
F. ochoterenae	2.58 (3.89-1.38)	1.82 (2.92-0.80)
F. shrevei	3.03 (4.23-1.80)	1.29 (2.20-0.48)
F. formosa	4.27 (5.61-3.01)	1.1 (2.43-0.12)
F. macdougalii	3.56 (4.82-2.31)	1.05 (2.20-0-14)

Table 2. Results of fitting five birth-death models with the BDL method for stem group in *Fouquieria* with constant net diversification rates "*r*" or variable at a time "*st*", and including an extinction fraction "*a*" or not. ΔAIC is the difference in AIC scores between each model and the overall best-fit model. Best model selected by AIC and ML in bold.

Description of model	* Rate-constant Pure Birth, $a = 0$	Rate-constant Birth-Death, $a \geq 0$	Rate-variable Pure Birth, $a = 0$	Rate-variable Birth-Death, $a \geq 0$	**Rate-variable Pure Birth, $a = 0$**
Parameters in model	$r = 0.04$	$r = 2.01$, $a = 0.99$	$r1 = 0.01$, $st = 6.60$, $r2 = 0.13$	$r1 = 2.93e^{-8}$, $st = 3.36$, $r2 = 2.40e^{-9}$, $a = 1$	**$r1 = 0.01$, $st1 = 7.21$, $r2 = 0.33$, $st2 = 3.02$, $r3 = 0.03$**
Log-likelihood	-22.32	-16.17	-16.70	-11.76	**-10.14**
AIC	46.65	36.34	39.40	31.52	**30.28**
ΔAIC_{RC}					**p=0.006**

Table 3. Diversification dynamics of all *Fouquieria* identified with RPANDA evaluating 1 shift in diversification dynamics in the woody clade. N is the number of parameters in the model; Bcons is constant speciation; Dlin is linear variable extinction; Bexp is exponential variable speciation; Dcons is constant extinction and **DAICc** is the difference in corrected Akaike Information Criterion (AICc).

All *Fouquieria*		**Best Model**	**N**	**Log-likelihood**	**AICc**	**DAICc**
No shifts		Bcons Dlin	3	-23.662	57.323	6.174
1 shift	Stem lineage + succulent grade Woody clade	Bcons Dlin Bexp Dcons	6	-16.774	51.149	0

Table 4. Diversification dynamics of crown *Fouquieria* identified with RPANDA evaluating 1 shift in diversification dynamics in the woody clade. N is the number of parameters in the model; Bcons is constant speciation; Dlin is linear variable extinction; Bexp is exponential variable speciation; Dcons is constant extinction and **DAICc** is the difference in corrected Akaike Information Criterion (AICc).

Crown	**group *Fouquieria***	**Best Model**	**N**	**Log-likelihood**	**AICc**	**DAICc**
No shifts		Bcons Dcons	2	-20.583	46.880	0
1 shift	Succulent grade Woody clade	Bexp Dcons Bexp Dcons	6	-14.012	47.661	0.781

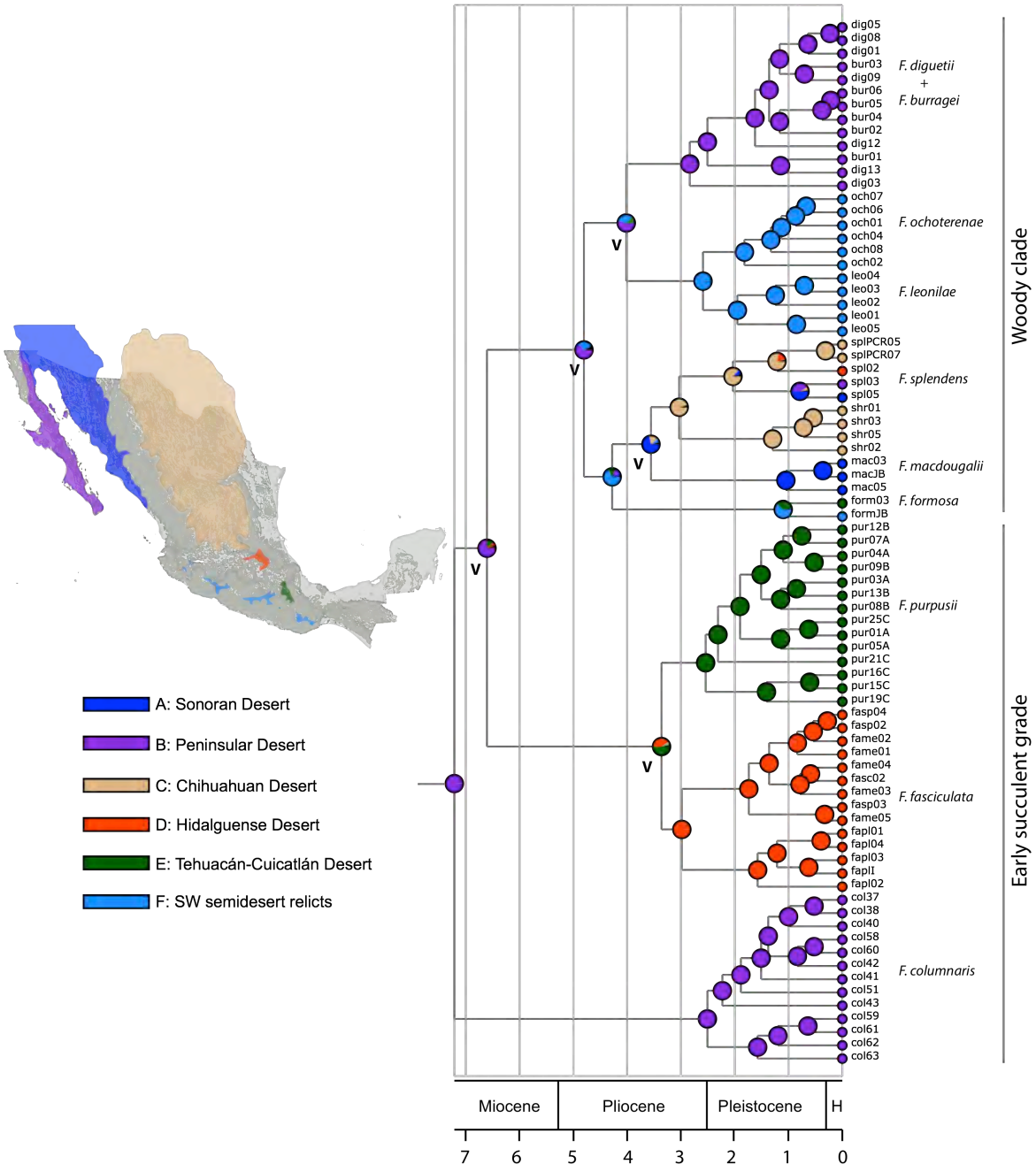


Figure 1. *Fouquieria* chronogram resulting from the dating analyses in BEAST and results of the biogeographic analyses with BMM in RASP. Pie charts show the probability values of the ancestral areas reconstructed at each node. “V” indicates Vicariance events.

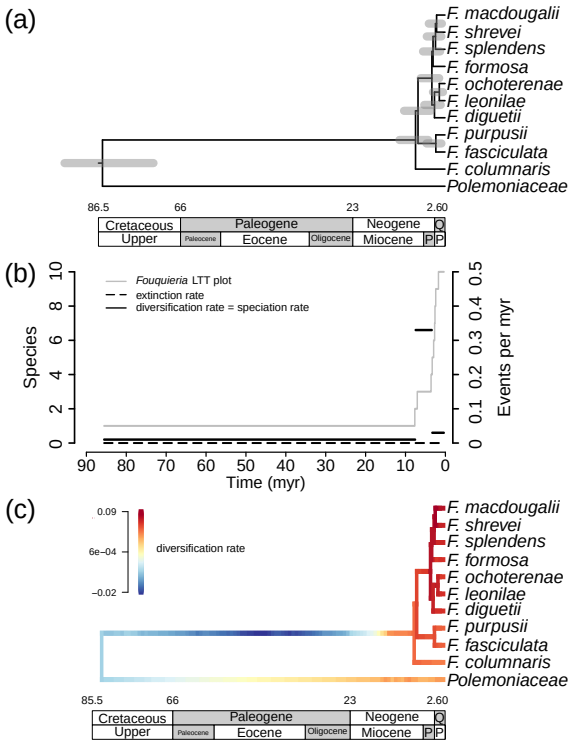


Figure 2. Fouquieria diversification dynamics. a) *Fouquieria* chronogram resulting from the two BEAST analyses. Bars upon nodes represent the 95% highest posterior density (HPD) intervals. b) Best model inferred by BDL. c) Best diversification dynamics configuration inferred by BAMM.

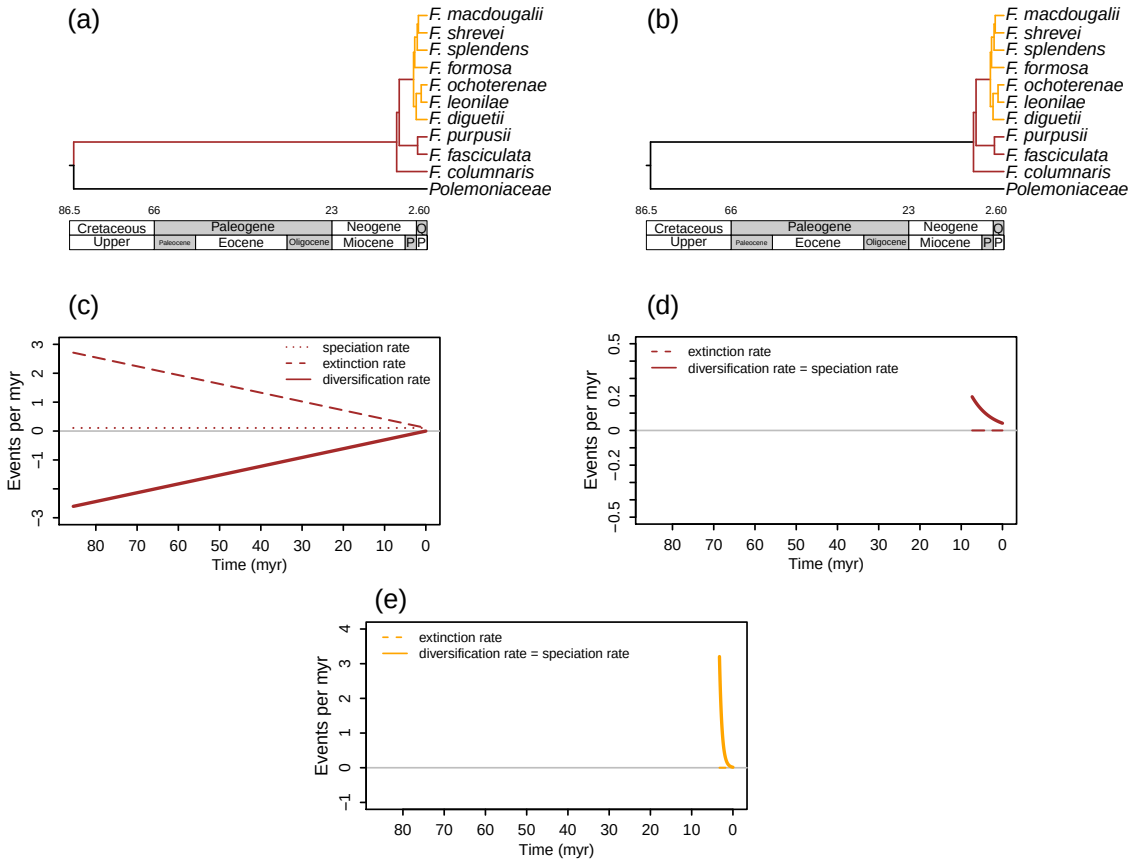


Figure 3. *Fouquieria* diversification dynamics inferred by RPANDA. a) All *Fouquieria* analysis best model. b) Crown group *Fouquieria* analysis best model. c) Rate variation through time according to the best model inferred for the stem lineage + succulent grade. d) Rate variation through time according to the best model inferred for the succulent grade. e) Rate variation through time according to the best model inferred for the woody clade.

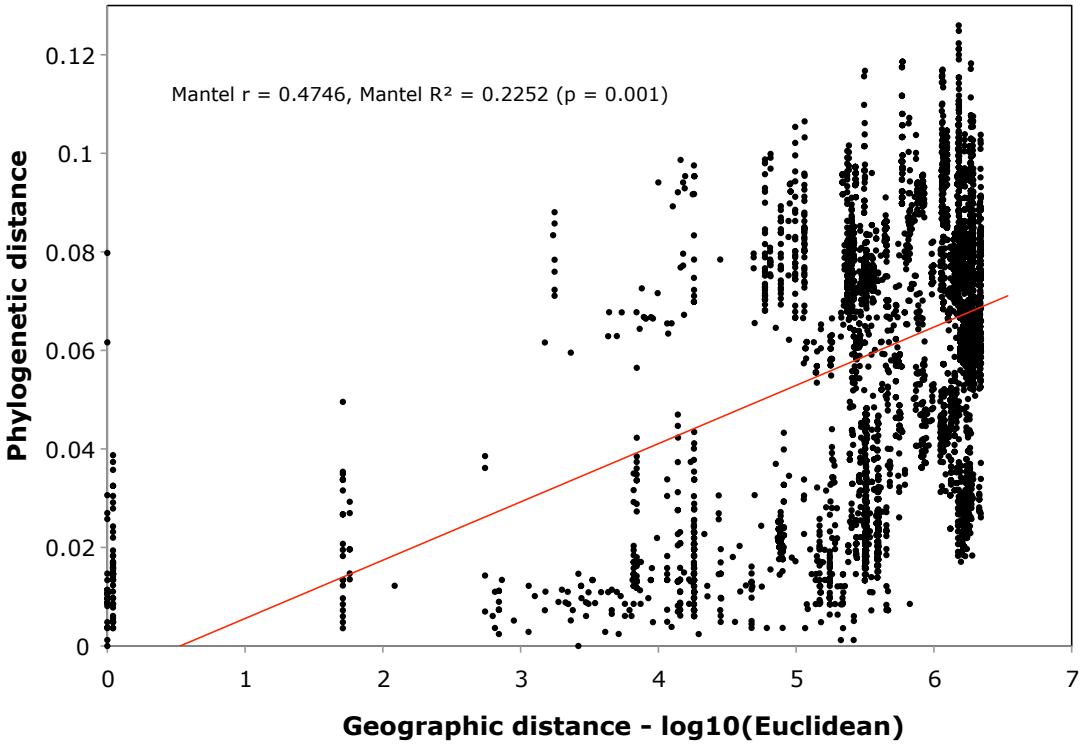


Figure 4. The relationship between geographical and phylogenetic distance in *Fouquieria*. Phylogenetic distance was measured with the GTR+I+G nucleotide substitution model through the branch lengths in the ML phylogram.

Material Suplementario

Supporting Information Table S1. List of species and GenBank accessions in *Fouquieria* analyses.

Species	**Individual Id.**	***rpl32-trnL***	***rps16-trnK***	***ndhF-rpl32***	***rpl14-rps8-infA-rpl36***
Fouquieria burragei	bur01	KX516875	KX516950	KX517088	KX517089
	bur02	KX516876	KX516951	KX517087	KX517090
	bur03	KX516877	KX516952	KX517086	KX517091
	bur04	KX516878	KX516953	KX517085	KX517092
	bur05	KX516879	KX516954	KX517084	KX517093
	bur06	KX516880	KX516955	KX517083	KX517094
Fouquieria columnaris	col37	KX516881	KX516956	KX517082	KX517095
	col38	KX516882	KX516957	KX517081	KX517096
	col40	KX516883	KX516958	-	KX517097
	col41	KX516884	KX516959	KX517080	KX517098
	col42	KX516885	KX516960	KX517079	KX517099
	col43	KX516886	KX516961	KX517078	KX517100
	col51	KX516887	KX516962	KX517077	KX517101
	col58	KX516888	KX516963	KX517076	KX517102
	col59	KX516889	KX516964	KX517075	KX517103
	col60	KX516890	KX516965	-	KX517104
	col61	-	KX516966	KX517074	KX517105
	col62	KX516891	KX516967	KX517073	KX517106
	col63	KX516892	KX516968	KX517072	KX517107
Fouquieria diguetii	dig01	KX516893	KX516969	KX517071	KX517108
	dig03	KX516894	KX516970	KX517070	KX517109
	dig05	KX516895	-	KX517069	KX517110
	dig08	KX516896	KX516971	KX517068	KX517111
	dig09	KX516897	KX516972	KX517067	KX517112
	dig12	KX516898	KX516973	KX517066	KX517113
	dig13	KX516899	KX516974	KX517065	KX517114
Fouquieria fasciculata	fame01	KX516900	KX516975	KX517064	KX517115
	fame02	KX516901	KX516976	-	KX517116
	fame03	KX516902	KX516977	KX517063	KX517117
	fame04	KX516903	KX516978	KX517062	KX517118
	fame05	KX516904	KX516979	KX517061	KX517119
	fapl01	KX516905	KX516980	KX517060	KX517120
	fapl02	KX516906	KX516981	-	KX517121

	fapl03	KX516907	KX516982	-	KX517122
	fapl04	KX516908	KX516983	KX517059	KX517123
	faplI	KX516909	KX516984	KX517058	KX517124
	fasc02	KX516910	KX516985	KX517057	KX517125
	fasp02	KX516911	KX516986	KX517056	KX517126
	fasp03	KX516912	KX516987	KX517055	-
	fasp04	KX516913	KX516988	KX517054	KX517127
Fouquieria formosa	form03	KX516914	KX516989	-	KX517128
	formJB	-	KX516990	KX517053	KX517129
Fouquieria leonilae	leo01	KX516915	KX516991	KX517052	KX517130
	leo02	KX516916	KX516992	KX517051	KX517131
	leo03	KX516948	KX516993	KX517050	KX517132
	leo04	KX516949	KX516994	KX517049	KX517133
	leo05	KX516917	KX516995	-	KX517134
Fouquieria macdougalii	mac03	KX516918	KX516996	KX517048	KX517135
	mac05	-	-	-	KX517136
	macJB	-	-	-	KX517137
Fouquieria ochoterenae	och01	KX516919	KX516997	-	KX517138
	och02	KX516920	KX516998	KX517047	-
	och04	KX516921	KX516999	KX517046	KX517139
	och06	KX516922	KX517000	-	KX517140
	och07	KX516923	KX517001	KX517045	-
	och08	KX516924	KX517002	-	KX517141
Fouquieria purpusii	pur01A	KX516925	KX517003	KX517044	KX517142
	pur03A	KX516926	KX517004	KX517043	KX517143
	pur04A	KX516927	KX517005	KX517042	KX517144
	pur05A	KX516928	KX517006	KX517041	KX517145
	pur07A	KX516929	KX517007	KX517040	KX517146
	pur08B	KX516930	KX517008	KX517039	KX517147
	pur09B	KX516931	KX517009	KX517038	KX517148
	pur12B	KX516932	KX517010	-	KX517149
	pur13B	KX516933	KX517011	-	KX517150
	pur15C	KX516934	KX517012	KX517037	KX517151
	pur16C	KX516935	KX517013	KX517036	KX517152
	pur19C	KX516936	KX517014	KX517035	KX517153
	pur21C	KX516937	KX517015	KX517034	KX517154
	pur25C	KX516938	KX517016	KX517033	KX517155
Fouquieria shrevei	shr01	KX516939	KX517017	KX517032	KX517156
	shr02	KX516940	KX517018	KX517031	KX517157

	shr03	KX516941	KX517019	KX517030	KX517158
	shr05	KX516942	KX517020	KX517029	KX517159
Fouquieria splendens	spl02	KX516943	KX517021	KX517028	KX517160
	spl03	KX516944	KX517022	KX517027	KX517161
	spl05	KX516945	KX517023	KX517026	KX517162
	splPCR05	KX516946	KX517024	-	KX517163
	splPCR07	KX516947	KX517025	-	KX517164
Ilex paraguariensis		FR849983.1	-	FR849978.1	-
Camellia sinensis		KC143082.1	KC143082.1	KC143082.1	KC143082.1
Vaccinium macrocarpon		JQ248601.1	JQ248601.1	JQ248601.1	JQ248601.1

Supporting Information Table S2. List of species and GenBank accessions for Ericales in Eudicots.

Species	***rbcL***	***atpB***	***matK***	**18s**	**26s**
Acorus gramineus	D28866.1	AF197616.1	DQ182341. 1	AF197584.1	AF036490.1
Actinidia spp.	L01882.2	AJ235382.2	U61324.1	AF419792.1	AY727964.1
Ailanthus altissima	U02726	AF035895	EF489111	AF206842	-
Alangium spp.	L11209.2	AJ235386.2	U96880.1	AF206843.1	AY260009.1
Altingia spp.	AF206732.1	AF092103.1	AF304520.1	U42552.1	AF479208.1
Anagallis spp.	M88343.1	AJ235390.2	AJ581446.1	AF206845.1	AF479149.1
Androsace spp.	AF395004.1	AF213775.1	AY647535.1	AF206847.1	AF479150.1
Arbutus spp.	L12597	AF266738	U61345	AF206853	DQ067894
Barringtonia spp.	EU980812.1	AJ235407.2	DQ924095.1	AY289647.1	AY727949.1
Buxus sempervirens	DQ182333.1	AF092110.1	AF543728.1	X16599.1	AF389243.1
Camellia sinensis	AF380037.1	AY725933.1	AF380077.1	AB120309.1	AY727975.1
Carica papaya	M95671.1	AF035901.1	AY042564. 1, 45775521	U42514.1	AF479145. 1
Catalpa spp.	L11679.1	HQ384724	HQ384519	AF107579	-
Ceratophyllum demersum	D89473.1	AJ235430.2	AF543732.1	U42517.1	AY095456.1
Cercis canadensis	U74188	DQ401328	EU361912	-	-
Clavija	AF213818	AJ235437	JQ588865.1	AJ235998	AF479151
Clethra spp.	AF421089.1	AF420966.1	AB697681.1	AF419793.1	AY727968
Clusia gundlachii	Z75673	AY788209	EF135520	AY674584	-
Cobaea scandens	Z83143.1	AJ235440.2	L48568.1	L49277.1	AY727944.1
Couroupita	Z80181	AJ236224	-	AJ235993	AY727950
Cyrilla racemiflora	L01900.2	AJ235449.2	AF380080.1	U43294.1	AY727969.1
Datisca spp.	L21939.1	AJ235450.2	AB016467.1	AF008952.1	AY968411
Davidsonia spp.	AF291934.1	AF209574.1	U92846.1	AF206897.1	AY935812.1
Diospyros spp.	EU980774.1	DQ923957.1	DQ924064.1	U43295.1	AY727957.1
Dipelta	AJ420876	GQ983629	20530893	GQ983567	GQ983584
Dipsacus spp.	L13864.1	AF209577.1	22795882	U43150.1	AF479231.1
Drosera spp.	L01909.2	AY096110.1	AY096122	U42532.1	AF389248.1
Enkianthus campanulatus	L12616.2	AF420968.1	U61344.2	AF419802.1	AY727970.1
Euclea crispa	EU980789.1	DQ923966.1	DQ924073.1	GU476413.1	-
Eucnide spp.	U17874.1	AJ236227.1	AF503315.1	AJ235988.1	AY260031.1
Euptelea polyandra	L12645.2	AF528850.1	DQ401348.1	L75831.1	AF389249.1
Eurya spp.	AF089714.1	AF420969.1	AF380081.1	AJ235995.1	AY727953.1
Fagus spp.	AY263936.1	AY147105.1	AB046507.1	AF206910.1	AY935813.1

Fendlera rupicola	AF206766.1	AJ236234.1	AY254063.1	AJ235986.1	AY260041.1
Fouquieria columnaris	AY725861.1	AY725923.1	EU628508.1	AF003961.1	AF479159.1
Fouquieria fasciculata	AY725862.1	AY725924.1	-	-	AY727940.1
Fouquieria splendens	L11675.1	-	EU628509.1	L49280	-
Galax spp.	Z80184.1	AY725936.1	L48576.1	L49281.1	AY727983.1
Gilia spp.	Z83144.1	AJ236220.1	L34198.1	DQ080013.1	AF479155. 1
Gossypium spp.	X15886.1	AJ233063.1	AF403561	U42827.1	-
Gunnera spp.	AF093724.1	AF093374.1	AY042596.1	U43787.1	AF389250.1
Halesia spp.	Z80190.1	DQ923988.1	DQ924097.1	L49284.1	AY727981.1
Helianthus annuus	AF097517.1	AJ236205.1	DQ383815.1	AF107577.1	AF479183.1
Ilex spp.	GQ997347.1	GQ997300.1	GQ248140. 1, EF590403	AF161010. 1	AF479203. 1
Impatiens spp.	AB043508.1	AY725922.1	AJ429280.1	L49285.1	AF479154.1
Ipomoea spp.	AY100963.1	AY100754.1	AJ429355.1	U38310.1	AF146016. 1
Itea spp.	AF190435.1	AF093383.1	EF456732.1	U42545.1	AF479216.1
Juglans spp.	AY263932.1	AY263952. 1	AF118036	AF206943.1	AF479105.1
Lamium spp.	AB266225	AJ236165	AJ429332	L49287	-
Leea guineensis	AJ235783.1	AJ235520.2	AF274621.1	AF206951.1	AF274653.1
Lissocarpa benthamii	EU980793.1	DQ923969.1	DQ924077.1	-	AY727956.1
Maesa spp.	Z80203	AF213781	AJ429288.1	-	AY727959
Manilkara zapota	AF213793.1	AJ235528.1	DQ924092.1	L49288.1	AY727946.1
Marcgravia spp.	Z83148.1	AJ235529.1	AJ429289.1	-	AY727937.1
Menispermum spp.	FJ026493.1	AF093384.1	GU266604.1	L75834.1	AF389257.1
Norantea spp.	JQ625952.1	AF420978.1	JQ626401.1	-	AY727938.1
Nyssa spp.	L11228.2	AJ235545.2	U96886.1	U52032.1	AF297545. 1
Parnassia spp.	AY935729	AJ235552.2	AY935908	-	AF036496.1
Petrophile spp.	U79181.1	AF060401.1	EU169655.1	AF293761.1	DQ008610.1
Phlox spp.	AF206809.1	AJ236221.1	L34203.1	L49293.1	AF148281.1
Phytolacca americana	M62567.1	AF528855.1	DQ401362.1	AF094557.1	HQ843459.1
Pisum sativum	FN435842.1	X03852.1	AY386961.1	U43011.1	-
Platanus occidentalis	AF081073.1	AF528858.1	AF543747.1	U42794.1	AF274662.1
Polemonium spp.	L11687.1	AY725925.1	AJ429292.1	L49294.1	AY727941.1
Polygonum spp.	AF297127.1	AJ235569.2	EF438020	AF206996.1	AF479085.1
Populus spp.	M58392.1	AF209658.1	AB038186.1	AF206999.1	AF479118.1
Primula spp.	AF213801.1	AF213785.1	AY647489.1	L49295.1	AY727960.1
Quintinia spp.	AF299092	AJ318983	AJ429366	GQ983576	GQ983590
Rhamnus spp.	AJ390070.1	AJ235579.2	AY257533.1	AJ235979.1	JF317393.1
Rhododendron spp.	L01949.2	AY725932.1	EU087361.1	AF419807.1	AY727973.1

Roridula gorgonias	L01950.2	AJ236180.1	AJ429294.1	AF207010.1	AY727965.1
Sarracenia spp.	L01952.2	AJ235594.2	JQ218257.1	U42804.1	JQ519380.1
Sladenia spp.	AF320784.1	AF420988.1	AJ429297.1	AF320782.1	-
Sloanea spp.	AF022131.1	AJ235603.2	AY935938.1	U42826.1	AY935812.1
Spathiphyllum wallisii	AJ235807.1	AJ235606.2	209417664	AF207023.1	AY095473.1
Stellaria media	AF206823.1	AF209680.1	AY936299	AF207027.1	AF479084.1
Styrax spp.	Z80189.1	AJ235615.2	DQ924099.1	L49296.1	AF479156.1
Symplocos spp.	AY725865.1	AY725934.1	AY336340.1	U43297.1	AY727978.1
Terminalia spp.	U26338	AF209686	GU135057	AF207037	AF479147
Ternstroemia spp.	Z80199.1	AJ235623.2	AJ429304.1	AF207039.1	AF479153.1
Tetramerista spp.	Z80199.1	AJ235623.2	AJ429304.1	AF207039.1	AF479153.1
Vaccinium spp.	AF421107.1	AF420987.1	AF419716.1	L49297.2	AY727974.1
Vitis spp.	DQ424856	DQ424856	DQ424856	AF207053.1	AF479207. 1

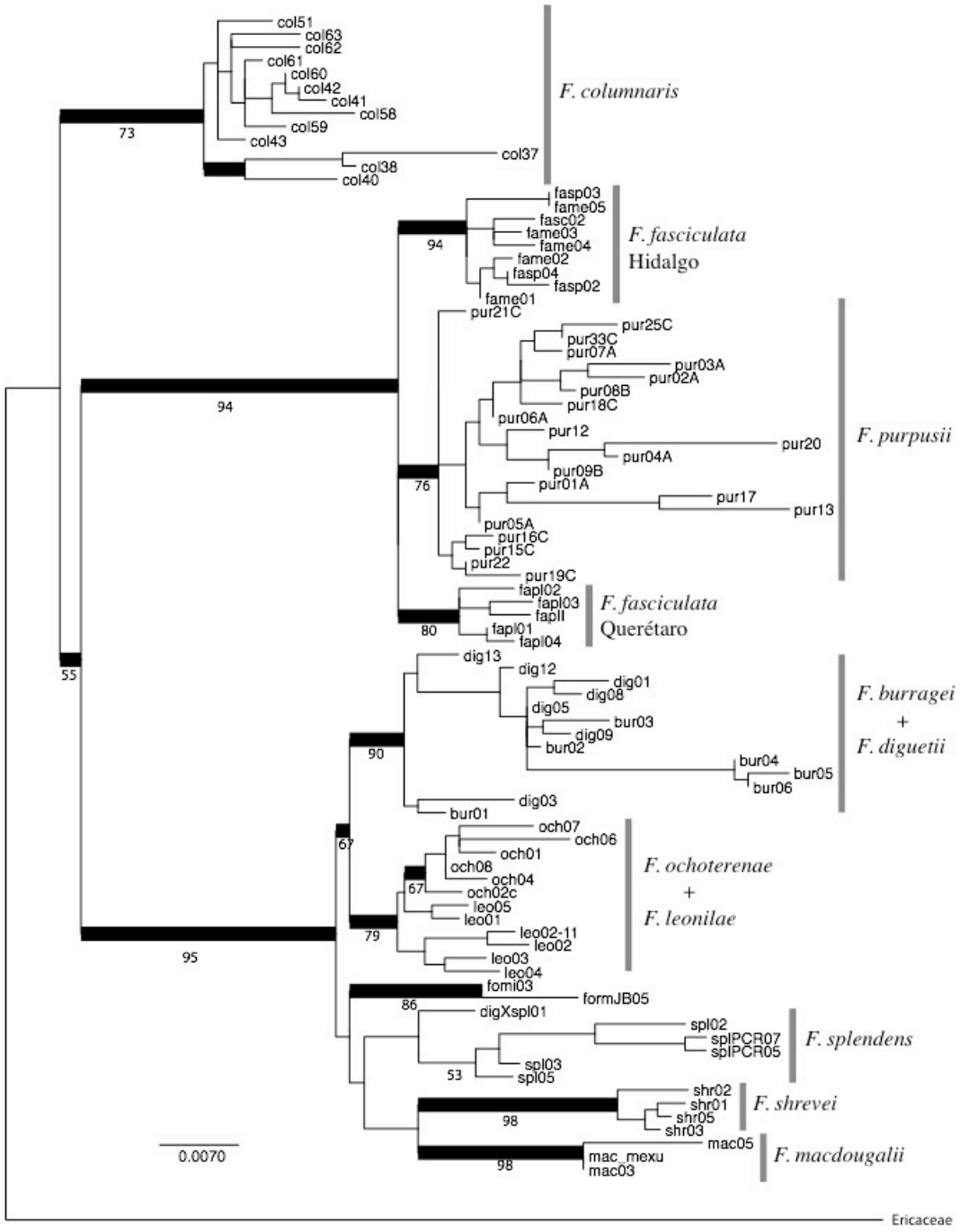
Supporting Information Table S3. Polemoniceae-Fouquieraceae and crown *Fouquieria* divergence dates credibility intervals from the first BEAST analysis.

Divergence	**Polemoniaceae, Fouquieriaceae**	**Crown *Fouquieria***
node	111	112
height_95%_HPD_MIN	72.74	2.97
height_95%_HPD_MAX	94.95	11.38
length_range_MIN	1.45	42.5
length_range_MAX	36.81	95.18
length_95%_HPD_MIN	2.68	65.16
length_95%_HPD_MAX	21.25	89.87
matK.rate_95%_HPD_MIN	0	0
matK.rate_95%_HPD_MAX	0	0
height_range_MIN	57	2.08
height_range_MAX	99.57	21.21
nuclear.rate_range_MIN	0	0
nuclear.rate_range_MAX	0	0
rbcLatpB.rate	0	0
height_median	85.42	6.55
height	84.76	6.87
matK.rate_range_MIN	0	0
matK.rate_range_MAX	0	0
rbcLatpB.rate_median	0	0
rbcLatpB.rate_95%_HPD_MIN	0	0
rbcLatpB.rate_95%_HPD_MAX	0	0
matK.rate_median	0	0
nuclear.rate_95%_HPD_MIN	0	0
nuclear.rate_95%_HPD_MAX	0	0
length	10.94	77.95
nuclear.rate	0	0
matK.rate	0	0
length_median	10.09	78.51
rbcLatpB.rate_range_MIN	0	0
rbcLatpB.rate_range_MAX	0	0
nuclear.rate_median	0	0
posterior	0.99	1

Supporting Information Table S4. *Fouquieria* main phylogroups divergence dates credibility intervals from the second BEAST analysis.

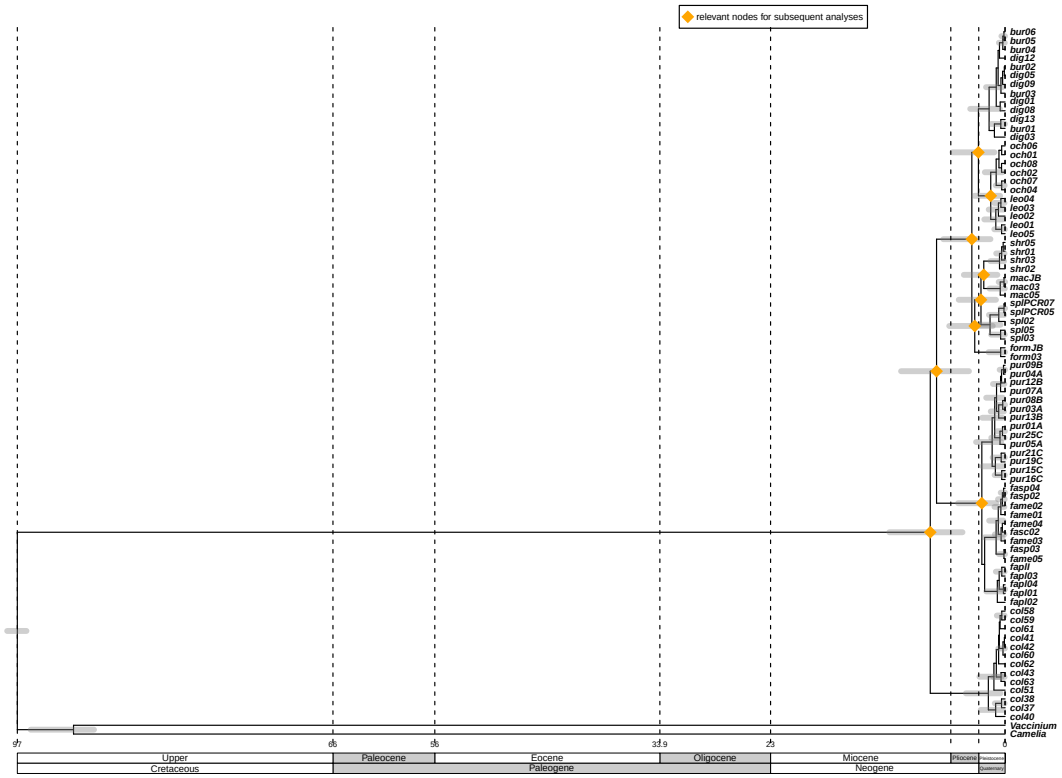
Divergence	**Crown *Fouquieria***	***F. columnaris*, 85**	***F. fasciculata*, *F. purpusii***	**Crown woody clade**	***F. formosa*, 116**	***F. splendens*, 117**	***F. macdougalii*, *F. shrevei***	***F. diguetti* + *F. burragei*, 140**	***F. leonilae*, *F. ochoterenae***
node	84	85	86	113	114	116	117	127	140
ndhFrpl32.rate	0	0	0	0	0	0	0	0	0
height_95%_HPD_MIN	4.18	3.55	0.88	1.44	1.21	0.9	0.75	1.04	0.49
height_95%_HPD_MAX	11.37	10.22	4.62	6.07	5.49	4.55	4.05	5.18	3.01
length_range_MIN	78.66	0	0.83	0.07	0	0	0	0	0.02
length_range_MAX	95.1	5.69	9.91	10.61	4	4.04	2.87	4.59	5.76
length_95%_HPD_MIN	85.56	0	1.71	0.81	0	0.01	0	0	0.11
length_95%_HPD_MAX	93.26	2.19	7.18	5.44	1.14	1.36	1.19	1.63	2.72
height_range_MIN	2.34	1.86	0.51	0.98	0.77	0.4	0.35	0.49	0.26
height_range_MAX	17.9	11.38	8.23	9.61	8.78	8.46	6.5	9.48	7.52
rps16trnK.rate_median	0.01	0	0	0	0	0.01	0	0	0
rpl14rpl36.rate	0	0	0	0	0.01	0	0.01	0	0
rpl32trnL.rate	0	0	0	0	0	0	0	0	0
rpl32trnL.rate_95%_HPD_MIN	0	0	0	0	0	0	0	0	0
rpl32trnL.rate_95%_HPD_MAX	0	0	0	0	0	0	0	0	0
ndhFrpl32.rate_95%_HPD_MIN	0	0	0	0	0	0	0	0	0
ndhFrpl32.rate_95%_HPD_MAX	0	0	0	0.01	0	0.01	0.01	0	0.01
rpl14rpl36.rate_range_MIN	0	0	0	0	0	0	0	0	0
rpl14rpl36.rate_range_MAX	0.05	0.26	0.01	0.02	0.56	0.12	0.5	0.54	0.06
rpl14rpl36.rate_95%_HPD_MIN	0	0	0	0	0	0	0	0	0
rpl14rpl36.rate_95%_HPD_MAX	0.01	0.01	0	0	0.02	0.01	0.02	0.01	0.01
rpl32trnL.rate_range_MIN	0	0	0	0	0	0	0	0	0
rpl32trnL.rate_range_MAX	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
rps16trnK.rate_range_MIN	0	0	0	0	0	0	0	0	0
rps16trnK.rate_range_MAX	1.9	0.08	0.03	0.04	0.2	1.27	0.29	0.36	0.05
height_median	7.35	6.73	2.28	3.27	2.98	2.37	2.09	2.62	1.41
height	7.46	6.77	2.5	3.46	3.14	2.53	2.23	2.82	1.56
ndhFrpl32.rate_median	0	0	0	0	0	0	0	0	0
rpl14rpl36.rate_median	0	0	0	0	0	0	0	0	0

ndhFrpl32.rate_range_MIN	0	0	0	0	0	0	0	0	0
ndhFrpl32.rate_range_MAX	0.01	0.05	0.01	0.04	0.09	0.08	0.06	0.09	0.06
rps16trnK.rate	0.01	0	0	0	0	0.01	0	0.01	0
length	89.54	0.87	4.36	2.97	0.42	0.56	0.45	0.6	1.28
rpl32trnL.rate_median	0	0	0	0	0	0	0	0	0
length_median	89.64	0.7	4.23	2.81	0.31	0.45	0.35	0.46	1.13
rps16trnK.rate_95%_HPD_MIN	0	0	0	0	0	0	0	0	0
rps16trnK.rate_95%_HPD_MAX	0.04	0.01	0.01	0.01	0.01	0.04	0.01	0.02	0.01
posterior	1	0.6	1	1	0.82	0.95	0.71	0.94	1

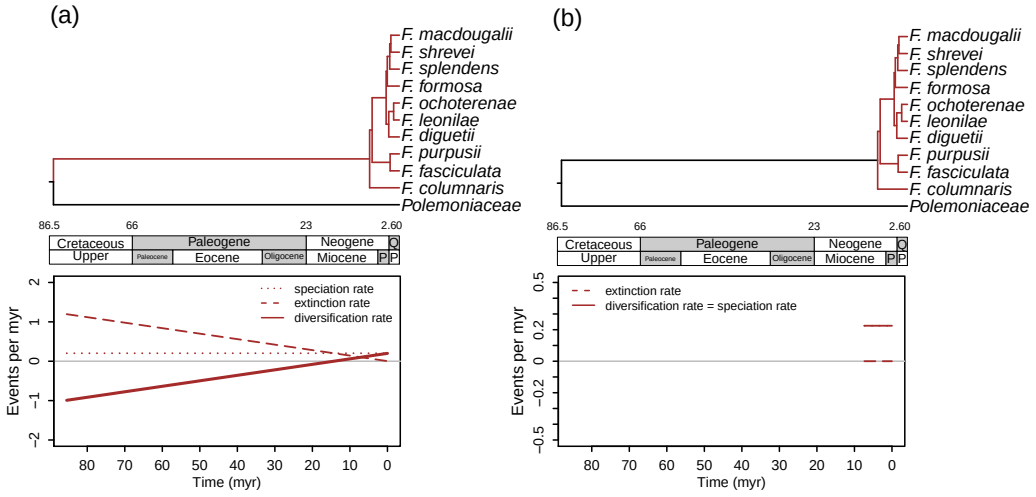


Supporting Information Figure S1. Best ML tree for *Fouquieria* using *ndhF-rpl32, rpl14-rps8-infA-rpl36, rpl32-trnL, and rps16-trnK* regions.Number bellow branches indicates ML Bootstrap Support. Bold lines indicates PP > 0.95, calculated with Bayesian inference (two runs with 10 million generations).

Supporting Information Figure S2. Timing of stem and crown *Fouquieria* divergence. Chronogram derived from the maximum clade credibility tree estimated with the uncorrelated lognormal method in the first BEAST analysis. Bars upon nodes represent the 95% highest posterior density (HPD) intervals.



Supporting Information Figure S3. Timing of divergence of crown *Fouquieria* and its main lineages. Chronogram derived from the maximum clade credibility tree estimated with the uncorrelated lognormal method in the second BEAST analysis. Bars upon nodes represent the 95% highest posterior density (HPD) intervals.



Supporting Information Figure S4. Diversification dynamics results from the two RPANDA analyses when no changes among clades were allowed. a) All *Fouquieria* best model. b) Crown *Fouquieria* best model.

CAPÍTULO 3

DESCUBRIENDO LA DINÁMICA DE DIVERSIFICACIÓN DE TAXA SUPERIORES CON DATOS DE EDAD Y RIQUEZA DE ESPECIES DE LOS CLADOS

Artículo publicado en Systematic Biology

Regular research paper

Uncovering higher-taxon diversification dynamics from clade age and species-richness data

Running title: Diversification in hierarchically organized taxa

5 Figures

4 Supplementary Tables

9 Supplementary Figures

Luna L. Sánchez-Reyes[1,2]

Hélène Morlon[3]

Susana Magallón[1]

[1]Instituto de Biología, Universidad Nacional Autónoma de México, 3er Circuito de Ciudad Universitaria, Coyoacán, Ciudad de México 04510, México

[2]Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México, Av. Universidad 3000, Coyoacán, Ciudad de México 04510, México

[3]École Normale Supérieure, UMR 8197 CNRS, 46 rue d'Ulm, 75005, Paris, France

Author's e-mail addresses: sanchez.reyes.luna@gmail.com, morlon@biologie.ens.fr, s.magallon@ib.unam.mx.

Corresponding author: Luna L. Sánchez-Reyes[1,2]

Abstract. — The relationship between clade age and species richness has been increasingly used in macroevolutionary studies as evidence for ecologically versus time-dependent diversification processes. However, theory suggests that phylogenetic structure, age type (crown or stem age) and taxonomic delimitation can affect estimates of the age-richness correlation (ARC) considerably. We currently lack an integrative understanding of how these different factors affect ARCs, which in turn, obscures further interpretations. To assess its informative breadth, we characterise ARC behaviour with simulated and empirical phylogenies, considering phylogenetic structure and both crown and stem ages. First, we develop a two-state birth-death model to simulate phylogenies including the origin of higher taxa and a hierarchical taxonomy to determine ARC expectations under ecologically and time-dependent diversification processes. Then, we estimate ARCs across various taxonomic ranks of extant amphibians, squamate reptiles, mammals, birds and flowering plants. We find that our model reproduces the general ARC trends of a wide range of biological systems despite the particularities of taxonomic practice within each, suggesting that the model is adequate to establish a framework of ARC null expectations for different diversification processes when taxa are defined with a hierarchical taxonomy. ARCs estimated with crown ages were positive in all the scenarios we studied, including ecologically dependent processes. Negative ARCs were only found at less inclusive taxonomic ranks, when considering stem age, and when rates varied among clades. This was the case both in ecologically and time-dependent processes. Together, our results warn against direct interpretations of single ARC estimates and advocate for a more integrative use of ARCs across age types and taxonomic ranks in diversification studies.

Key words: Birth-Death models; Crown age; Diversity dependence; Extinction; Phylogenetic structure; Speciation; Stem age; Taxonomy; Time dependence; Tree simulations.

Acronyms:

ARC: Age-richness correlation

crown-ARC: Correlation between crown age and richness

stem-ARC: Correlation between stem age and richness

PGLS: Phylogenetic generalized least squares

ML: Maximum Likelihood

Variables:

λ: speciation rate

λ_0: initial speciation rate

λ_t: speciation rate at time t

μ: extinction rate

Φ: origination rate

α: rate of change in diversification rates among taxa

N: species number

K: carrying capacity

Λ: Pagel's lambda

Explaining the causes of the heterogeneous distribution in species numbers (richness) among taxa, communities and regions, and across time, is a fascinating and long-standing problem in evolutionary biology (Darwin 1859; Wallace 1878; Bokma et al. 2014). While many factors have been postulated to explain the heterogeneous distribution of richness we observe in nature, species richness is ultimately the result of the interplay between speciation and extinction, i.e., the diversification process. Diversification processes can be represented by various models (Morlon 2014), including the classical 'time-dependent' and the 'ecologically dependent' models. In time-dependent processes, the rate at which new species are generated (speciation rate, λ) and the rate at which species become extinct (extinction rate, μ) vary as a function of time. These models predict that richness is either expanding unbounded through time and space (Yule 1925; Nee et al. 1992) or that it is controlled by increases in μ (Raup et al. 1973; Pyron and Burbrink 2012) or by decreases in λ through time (Morlon et al. 2011). In ecologically dependent processes, the speciation and extinction rates vary according to ecological factors such as the number of species alive at a given time (diversity-dependent process). In these models, richness is expected to be bounded by a macroevolutionary carrying capacity or by extrinsic ecological factors (Mittelbach et al. 2007; Rabosky 2009).

To evaluate different diversification dynamics one can use empirical phylogenetic and/or taxonomic data and apply a maximum likelihood or Bayesian approach to determine the diversification model that best explains the data (e.g., Hey 1992; Nee et al. 1992; Rabosky 2006a, 2014; Alfaro et al. 2009; Morlon et al. 2010, 2011; Etienne et al. 2011). When phylogenetic data are not available or phylogenies are incomplete, it is possible to use summary statistics—such as the maximum clade richness or the

correlation between clade age and species richness (the Age-Richness Correlation, ARC)— to distinguish among alternative hypotheses of diversification (Ricklefs and Renner 1994; Ricklefs et al. 2007; Rabosky 2009; Pyron and Burbrink 2012). The ARC is particularly useful because it allows direct evaluation of the effect of time –measured as clade age– on species richness. Intuitively, a positive relationship should indicate that clade age determines richness distribution patterns. A non-significant or negative relationship indicates that other factors explain these patterns. For example, a negative ARC could arise if differences in diversification rates rather than time explain the heterogeneous distribution of richness among taxa (Ricklefs and Renner 1994; Magallón and Sanderson 2001).

Evaluations of previous intuitions with simulations have indicated that ARCs are negative when richness is bounded (by ecological factors, Rabosky 2009, 2010; or by extinction, Pyron and Burbrink 2012). A more recent study has shown that when ARCs are estimated using the stem age of clades (stem-ARCs), a taxonomic classification based on time can also generate null or negative ARCs in diversification processes of unbounded richness (Stadler et al. 2014). This was not observed with ARCs estimated using the crown age of clades (crown-ARCs), revealing a conflict between ARCs estimated using crown or stem ages within the same diversification process. Accordingly, it was suggested that only crown-ARCs could be informative about the underlying diversification process (Stadler et al. 2014), with positive relationships expected from processes of unbounded richness and null or negative relationships expected from processes of bounded richness. However, this has not been explicitly evaluated.

Since the crown-ARC/stem-ARC conflict seems to arise only in some types of taxonomic delimitation (Stadler et al. 2014), it is natural to wonder if different taxonomic ranks could affect ARC estimations differently or even provide inconsistent evidence about the relationship between the age and richness of clades. If this is the case, can a particular taxonomic rank inform us better than others about the true ARC? For example, could the genus level, which has been considered to be more natural than other taxonomic ranks (Humphreys and Barraclough 2014), be more informative than more inclusive ranks, such as orders or families? Thus, accounting for the hierarchical nature of taxonomy is fundamental to determine the informative potential of ARCs in diversification studies. In this work, we first develop and implement a two-state birth-death model based on Rabosky *et al.* (2012), to simulate phylogenies incorporating the origin of higher taxa as the process of hierarchical emergence of attributes (any morphological, ecological, etc. trait inherited to all the descendants of a species) that can be used to assign monophyletic named groups (i.e., taxa) to nested categories (i.e., ranks) in a hierarchical arrangement equivalent to common taxonomical practice. To assess the informative breadth of the simulated ARC trends, we characterise empirical ARCs from different taxonomic ranks (ranging from more inclusive –e.g., orders– to less inclusive –e.g. genera) obtained from extant amphibians (Lissamphibia), birds (Aves), mammals (Mammalia), squamate reptiles (Squamata) and flowering plants (Angiospermae), using stem and crown ages. We find that our model reproduces the empirical ARC trends, suggesting that it captures relevant features of taxonomic practice that might be affecting ARCs in the biological systems analysed. This allows us to establish crown- and stem-ARC expectations in different ecologically and time-dependent diversification models

when richness is categorized into nested taxonomic ranks and when phylogenetic structure is considered. We discuss the validity of current ARC interpretations and ways forward in the study of diversification processes using the relationship between clade age and richness.

MATERIALS AND METHODS

Expected Diversification Processes: Modelling a Hierarchical Taxonomy

As an approximation to modelling a hierarchical taxonomy, we extended the birth-death model implemented by Rabosky et al. (2012) to allow a hierarchical origin of taxa. The model incorporates an origination rate parameter (Φ) that controls the frequency at which higher taxa arise. Briefly, the process starts with a branch that represents a single species. This branch can speciate through cladogenesis and generate two branches with rate λ; it can go extinct with rate μ; or it can undergo an origination event with rate Φ, and become a higher taxon. The origin of a higher taxon is not a branching event; it should be understood as an event such as the acquisition (or loss) of a distinctive attribute (*sensu lato*, e.g., morphology, function) that is inherited by all the descendants of the branch on which it occurred. Because each origination event occurs only once during the process, the associated distinctive properties would correspond to synapomorphies of the resulting clade, which are used to identify clade membership. The first node descending from the branch on which the origination event occurs represents the crown node of the higher taxon. Descending branches have the same three possibilities: they can speciate, go extinct or become a higher taxon of the next rank. The first origination event to occur along a lineage will correspond to the most inclusive taxon, belonging to the highest rank (i.e., first rank); the next event to occur within that lineage

will delimit groups of second-highest rank (i.e., second rank), and so on (Fig. 1). Therefore, small-numbered ranks correspond to highly inclusive taxa closer to the root of the tree, such as classes and orders, whereas ranks with high numbers correspond to least inclusive taxa, closer to the tips of the tree, such as families and genera.

Two or more origination events can occur by chance on the same branch, but this will not change its rank; rather, the rank level of any given branch is determined by the number of branches towards the root of the tree that have undergone origination events (Fig. 1). Our model does not simulate morphological change, and branch lengths are neither proportional to any measure of morphological distance (e.g., distinctiveness) nor to the number of origination events that occur on the same branch. Speciation and extinction probabilities can be constant or variable among clades by introducing the parameter α as the probability of a diversification rate shift occurring on any given branch. The timing of highertaxon origination events are drawn from an exponential distribution with parameter $\beta = \lambda + \mu + \Phi + \alpha$. The probability of each type of event is proportional to β. Origination probability is constant in all models evaluated in this study, but it could be modified to vary through time, and in a diversity-dependent manner.

Simulations

We conducted tree simulations under the general model described above by developing R code available in the package RPANDA (Morlon et al. 2016). Tree simulations were run for 50 time units or until 5000 extant tips were reached. Initial λ values for each model were set to generate large trees (> 500 extant tips; Supplementary Table S1 available on Dryad at http://dx.doi.org/10.5061/dryad.2b7d5). Since small trees could still be generated by chance, we ran simulations until 500 large trees were obtained

for each diversification model described below. Analyses were performed on large trees only.

As a null model of a time-dependent process with unbounded richness, we first considered constant rates of speciation and extinction through time and among clades. We evaluated the effect of extinction by constraining μ to be a constant proportion of λ at any point in time, so that we could explore the effect of different turnover regimes: absent ($\mu = 0$), low ($\mu = 0.2\lambda$), moderate ($\mu = 0.5\lambda$) and high turnover ($\mu = 0.9\lambda$). To assess the effect of a hierarchical taxonomy, we considered three different origination rate regimes: low ($\Phi = 0.5\lambda$), moderate ($\Phi = \lambda$) and high origination ($\Phi = 1.3\lambda$). When origination rate is low, few higher taxa arise and diversity is partitioned among few taxonomic ranks. As origination rate increases, eventually exceeding speciation rate, more higher taxa arise and diversity becomes more partitioned, generally among less inclusive taxonomic ranks. An example can be extracted from flowering plants. Order Malvales (e.g., cotton, chocolate) and order Piperales (e.g., black pepper) have roughly similar species richness (~6000 and ~4000, respectively). However, Malvales includes a much higher number of families and genera (10 and 338, respectively) than Piperales (4 and 17, respectively). This indicates that similar species richness is partitioned in a larger number of higher taxa in Malvales than in Piperales, suggesting a higher rate of origin of ecological, morphological, etc. descriptive attributes in the former group than in the latter.

Next, we allowed speciation and extinction rates to vary among clades with a rate shift probability of $\alpha = 0.05$ and an initial λ class (Supplementary Table S1 available on Dryad). Given a rate shift event occurring on a randomly chosen branch, a new λ class was drawn from a gamma distribution with shape parameter $k = 0.1$, generating a

distribution of new λ classes that are generally small but that can occasionally take high values.

As the null model of an ecologically dependent process with bounded richness, we implemented linear diversity-dependent change in speciation rate, allowing carrying capacity (K) and initial speciation rate (λ_0) to shift among clades with a probability of α = 0.01. The process starts with a given initial value of K and λ_0 (Supplementary Table S1 available on Dryad). Speciation rate decreases following the equation described by Nee et al. (1992) and used by Rabosky (2006b) and Rabosky and Lovette (2008), in which the speciation rate at any point in time (λ_t) is bounded by the number of species at that time (N_t) and by K:

$$\lambda_t = \lambda_0 (1 - N_t / K)$$

Given a rate shift event, a randomly chosen branch is assigned new K and λ_0 classes drawn uniformly from the intervals (50, 500) and (e^{-6}, 1), respectively. This branch and its descendants no longer belong to the initial K and λ_0 classes and will continue to evolve under the new class until another shift event occurs.

In both models of among-clade rate variation, we implemented Φ values that resulted in approximately the same number of ranks generated in the constant-rates models under high origination (Supplementary Table S1 available on Dryad). The effect of extinction was evaluated with the same four turnover regimes described above.

Empirical Diversification Processes: Age and Richness Data from Extant Taxa

We selected published molecular dating studies of various biological systems from which we could gather data on stem and crown ages from more than 80% of taxa within at least two different taxonomic ranks. The final data set comprised a dated tree

sampling >80% amphibian species (De Lisle and Rowe 2015); a dated tree encompassing >85% squamate genera (Pyron and Burbrink 2014); a mammal family-level dated tree generated with a supermatrix approach (Meredith et al. 2011); a maximum clade credibility tree obtained with TreeAnnotator (Rambaut and Drummond 2007) from a species-level dating study of birds (Hackett backbone; Jetz et al. 2012); and a family-level dated tree encompassing >87% flowering plant families (Magallón et al. 2015).

Richness counts were obtained following the taxonomy from the specialized databases of the Amphibian Species of the World 6.0 (Frost 2015); The Reptile Database (Uetz and Josek 1995); The Mammals Species of the World, 3rd edition (Wilson and Reeder 2005); and the Angiosperm Phylogeny Website (Stevens 2001). Richness counts of bird taxa were obtained following the master taxonomy from the Jetz *et al*. (2012) study, available in their electronic supplementary material.

ARC Estimation

To accurately estimate correlations between richness and any biological attribute within a phylogenetic framework and to avoid false positive ARC estimates (Rabosky et al. 2012), non-independence (autocorrelation or phylogenetic structure) of the data must be assessed and corrected. As a null model for the presence of phylogenetic structure, most tests use a Brownian motion model of association of normally distributed characters (or Brownian motion scaled to a link function if characters follow a distribution different from a Gaussian; Hadfield and Nakagawa 2010). This model is used to account for variation and covariation between characters, which are then used to estimate variance-corrected contrasts at each node as the average of derived character values (Felsenstein 1985). Ideally, the nature of the distribution of age and richness data should be taken into

account. However, the methods usually applied to ARC estimation assume a normal distribution of characters. Concerns have also been raised about how contrasts are estimated, which might be unsuitable for species-richness data (Agapow & Isaac 2002; Isaac et al. 2003). These concerns have promoted the development of alternative methods (see Agapow & Isaac 2002), that, nonetheless, seem to present other problems (Freckleton et al. 2008). For the objectives of this study and for comparative purposes with previous research (e.g., Rabosky 2009; Rabosky et al. 2012; Stadler et al. 2014; Hedges et al. 2015), we selected two approaches that have been commonly used in recent years to estimate age-richness relationships. First, to account for phylogenetic structure we employed Pagel's (1999) lambda (Λ) test within a phylogenetic generalized least squares (PGLS) framework (Freckleton et al. 2002), performed with the pgls function in the R package CAPER (Orme et al. 2013). Briefly, Λ is estimated with maximum likelihood (ML) and 95% confidence intervals are obtained to infer the model that best explains covariation between clade age and log-transformed richness data (i.e., residuals of the relationship of log-transformed richness regressed on clade age). A Λ value of 0 indicates phylogenetic independence of the data (residuals are randomly distributed across the phylogeny) while a value of 1 indicates that traits covary following a Brownian motion model and that there is phylogenetic structure (residuals are more similar among closely related lineages). Λ can also take intermediate values, indicating a weak phylogenetic structure explained by a different, unknown model of character association (Freckleton et al. 2002) or by a non-Gaussian character distribution (Hadfield and Nakagawa 2010). This makes the Λ test particularly useful to detect phylogenetic structure in characters that might not behave following Brownian motion. Second, we

estimated ARCs with the Spearman's rank test, a method that does not account for phylogenetic structure, using the cor.test function in the R base package STATS (R Core Team 2014).

ARCs were estimated with these two methods at all taxonomic ranks available from the simulated and empirical phylogenies. PGLS analyses were performed using higher-level phylogenies. To produce these, the original phylogenies were pruned to leave only one representative per taxon at the corresponding taxonomic rank. Except for flowering plants and mammal orders and families, non-monophyletic taxa were found in the empirical data sets. To evaluate the effect of non-monophyletic taxa, we estimated ARCs including and excluding them.

RESULTS

Expected ARC Patterns

Analyses of data on age and richness simulated with the model of hierarchical taxonomy show that ARCs vary across taxonomic ranks (Figs. 2 and 3). This was detected with both PGLS and Spearman's rank test. ARC variation across simulated ranks seems to be controlled mainly by origination and extinction rate parameters. In constant rate models, age and richness data appear to lack phylogenetic structure, since Λ were null with crown and stem ages and across all simulated ranks (Supplementary Fig. S1 available on Dryad). Accordingly, ARCs display very similar trends with PGLS and Spearman's test. In general, ARC values decrease as taxonomic rank becomes less inclusive. Decrease in ARCs is more pronounced when either origination or extinction rates are moderate to high (Figs. 2a and 3a). When estimated with PGLS and when extinction is moderate to high (Fig. 2a), ARCs can increase in high taxonomic ranks and

subsequently decrease. ARCs tend to stabilize and become almost equal across lower taxonomic ranks. ARC stabilization is reached more slowly as extinction increases (Figs. 2a and 3a; Supplementary Fig S2 available on Dryad). Under the constant rate models, crown- and stem-ARCs are always positive.

When rates are allowed to vary among clades, Λ take a range of intermediate values up to 1 among higher simulated ranks. Λ values decrease until they are null at lower simulated ranks (Supplementary Fig. S1 available on Dryad). Despite the presence of phylogenetic structure at higher simulated ranks, PGLS and Spearman's ARC estimates display the same general trends. We observe that both crown- and stem-ARCs decrease with taxonomic rank. However, crown-ARCs remain non-negative across taxonomic ranks, even in the model of diversity-dependent decrease in rates (Figs. 2b and 3b). In contrast, stem-ARCs are positive at high taxonomic ranks but take negative values at low taxonomic ranks, when extinction is moderate to high, in both models of among-clade rate variation (Figs. 2b and 3b).

Empirical ARC Trends

Empirical ARCs resulting from the analyses including and excluding non-monophyletic taxa were very similar (Supplementary Tables S2 and S3 available on Dryad). Here, we only present and discuss results from the analyses including non-monophyletic taxa (Figs. 4 and 5). Pagel's (1999) test applied to empirical age and richness data reveals different degrees of phylogenetic structure among biological systems (Supplementary Fig. S3 and Supplementary Table S4 available on Dryad). Low to almost null Λ estimates that are significantly different from 1 indicate that age and richness covary independently of phylogeny. This was observed in stem age-richness data

of amphibian families and subfamilies, squamate infraorders, superfamilies and genera, and flowering plant orders and families. It was also observed in crown age-richness data of amphibian subfamilies and of squamate infraorders and superfamilies. Squamate families and crown age-richness data of squamate genera also have almost null Λ, albeit not significantly different from 1 and with large confidence intervals, preventing to confidently assess phylogenetic independence in these data sets. This was also the case for mammal superorders and orders, and for bird orders. In contrast, Λ estimates are significantly different from 0 in amphibian genera and in bird families and genera, revealing the presence of phylogenetic structure at lower taxonomic ranks in these groups. Since Λ was also significantly different from 1 in these data sets, it indicates that a Brownian motion model does not account for the detected phylogenetic autocorrelation, and that a different character association model should explain phylogenetic structure. This is also the case for crown age-richness data of flowering plant families. Crown age-richness data of flowering plant orders and stem age-richness data of mammal families display high and intermediate Λ estimates, respectively, suggesting the presence of some degree of phylogenetic structure in these data. However, the test was not significantly different from 0 and confidence intervals were wide, preventing to confidently assert the presence of phylogenetic structure. Intermediate levels of phylogenetic structure can also occur when character distributions are non-Gaussian (Hadfield and Nakagawa 2010). A Brownian motion model of character association was only detected in the amphibial families crown age-richness data set.

Despite the presence of phylogenetic structure, PGLS and Spearman's test provide very similar ARC estimates, both in terms of the sign of the correlation and the

significance of the test (Figs. 4 and 5). Considering only the sign of ARC estimates, the empirical relationship between age and richness across taxonomic ranks holds within the same age type in all biological systems studied, except for mammal stem-ARCs estimated with Spearman's test, which are negative within superorders and families but positive within orders (Fig. 5b). The significance of the correlation test is also consistent across taxonomic ranks in amphibians, flowering plants, mammal stem-ARCs and squamate and bird crown-ARCs. In mammal crown-ARCs and squamate and bird stem-ARCs significance is achieved at lower taxonomic ranks, but not always so at higher ranks (Fig. 5).

Similar to ARC patterns simulated with the model of hierarchical taxonomy proposed here (Figs. 2 and 3), empirical ARC estimates tend to decrease at lower taxonomic ranks (Fig. 4). The only exceptions appear in amphibians and in stem-ARCs from flowering plants and mammals, in which we detect a moderate increase at lower ranks. In these three groups, ARC estimates obtained with stem and crown ages are decoupled: crown-ARCs are positive while stem-ARCs are negative. In birds and squamates, stem-ARCs are weaker than crown-ARCs, but both are always positive, and in most cases, significant. In amphibians, ARC estimates are almost equal with both age types and across taxonomic ranks, being positive but not significantly different from 0.

DISCUSSION

Simulations represent a fundamental tool to provide expectations from null models for hypothesis testing at macroevolutionary timescales. A good simulation framework must incorporate all elements potentially affecting the behaviour of the parameters used to describe the null model of interest in order to allow us to reach

accurate interpretations. In diversification studies, taxonomy appears to be an element that must be considered to understand comprehensively biodiversity patterns and processes. Paleobiologists have investigated the effect and validity of using higher taxa to study biodiversity patterns in the fossil record, with encouraging results (e.g., Raup et al. 1973; Wagner 1995; Robeck et al. 2000; Soul and Friedman 2015). In the study of the relationship between age and richness of extant clades, some forms of taxonomic delimitation have been demonstrated to affect ARC estimates (Stadler *et al*. 2014). However, the different implementations of taxonomic criteria have hampered the establishment of a null taxonomic model applicable to different organisms (Stadler et al. 2014). The model of hierarchical origin of higher taxa developed here reproduces different ARC trends observed in common taxonomic ranks of a wide variety of biological systems, appearing as a potential good tool to study the expected behaviour of ARCs under alternative diversification scenarios.

When diversification rates are constant through time and among clades, time is the only factor expected to explain variation in species richness (Rabosky 2009; Stadler et al. 2014). Our simulation framework is consistent with this expectation (Figs. 2a and 3a): the relationship between age and richness of clades is positive when time is considered since crown or stem age in constant rates diversification models. Nevertheless, we observe that stem-ARCs are weaker than crown-ARCs. The difference in ARC estimates generated by the use of crown or stem ages cannot be explained by unusually long stem branches generated by extinction (Pyron and Burbrink 2012), since it also arises in the absence of extinction (Figs. 2a and 3a). Hence, the simple fact that stem ages are older than crown ages should be weakening stem-ARCs compared to crown-

ARCs. Furthermore, the strength of the relationship between time and richness varies with taxonomic rank, weakening as rank becomes less inclusive. ARC variation across taxonomic ranks has not been documented before and is difficult to explain. It might be related to the presence of monotypic taxa, which are usually represented in more than one taxonomic rank. If monotypic taxa appear early in the tree, they could affect the age and richness variance structure of less inclusive ranks. More tests are needed to evaluate if monotypic taxa can effectively weaken the relationship between age and richness across taxonomic ranks.

When diversification rates vary among clades but are constant through time, richness is positively correlated with crown ages but negatively or not correlated with stem ages, only with certain forms of taxonomic delimitation (Stadler et al. 2014). In our simulations, less inclusive taxonomic ranks display this ARC pattern (Figs. 2b and 3b, top row), which also emerges in diversity-dependent diversification processes (Figs. 2b and 3b, bottom row). Hence, earlier work reporting positive relationships between crown age and richness should not necessarily be interpreted as time explaining richness patterns (Stephens and Wiens 2003; Wiens et al. 2006; McPeek and Brown 2007; Escudero and Hipp 2013), since ecologically dependent processes can generate that relationship too. Positive crown-ARCs and null or negative stem-ARCs have been obtained using PGLS among tribes of sedges (Escudero and Hipp 2013) and families and genera of mammals and birds (Hedges et al. 2015), and have been interpreted as evidence for time dependency and rate constancy. Of the simple models considered in this study, those that allowed among-clade varying rates (either constant through time or varying in

a diversity-dependent manner) and a high extinction are the only ones producing ARC patterns consistent with the ones observed in those groups.

We note that stem-ARCs previously reported in Rabosky et al. (2012) for birds differ from the ones obtained here. This might be a consequence of the use of different dated phylogenies in each study, which differ in topology and in dating methodology, likely resulting in differences in estimated clade ages. The effect of phylogenetic uncertainty and of error in age estimation on descriptive statistics of the diversification process is an important issue that should be addressed in future work. Our simulated ARCs show the ideal case in which we know clade ages precisely. In this context, our results suggest that stem-ARCs are more useful than crown-ARCs to reveal characteristics of the underlying diversification process. Considering both crown and stem ages to estimate correlations might prove useful to reveal the presence of extinction. However, neither crown- nor stem-ARCs can be used to distinguish between bounded and unbounded diversity scenarios, as previously proposed (Rabosky 2009; Pyron and Burbrink 2012; Rabosky et al. 2012).

Phylogenetic structure is another factor that has been posited to affect ARC behaviour (Rabosky et al. 2012). When we assume Brownian evolution ($\Lambda = 1$; Supplementary Fig. S4 available on Dryad) to estimate ARCs applying PGLS to data with no phylogenetic structure, correlations appear overestimated as compared to those obtained using the phylogenetic structure model inferred from the data (ML Λ values; Fig. 2; Supplementary Fig. S1 available on Dryad) and using a method that does not account for phylogenetic structure (Spearman's test; Fig. 3). A similar bias appears in the complementary case when richness evolves with a high degree of phylogenetic structure

and a non-phylogenetic method is used to estimate ARCs (Rabosky et al. 2012). This raises the need for more work to correctly account for phylogenetic structure before attempting to estimate correlations. For the moment, the results presented here suggest that when phylogenetic structure is weak, non-phylogenetic methods are as effective and valid as PGLS to estimate ARCs.

The models of among-clade variation reproduce to some degree the empirical lineage-through-time plots of the biological systems presented here (Supplementary Figs. S5 and S6 available on Dryad). In the models of rate variation among clades, we also detect some negative values of Pybus and Harvey (2000) gamma statistic (Supplementary Fig. S7 available on Dryad). However, the distribution of simulated gamma values tends towards null or positive estimates, showing that our model does not reproduce all properties of real phylogenies. The simulations also produced more taxonomic ranks than those generally established by common taxonomic practice, preventing the establishment of a formal correspondence between empirical and simulated taxonomic ranks. Since we aimed to generate complete trees (including extinct branches; Nee et al. 1992), simulation of very large trees (with > 5 000 extant tips) required substantial computational resources, hindering simulations with certain parameter values (e.g., high diversification rate). Hence, we only used a small range of parameter values not estimated from the empirical data to conduct the simulations. To implement a formal comparison between observed and expected properties of higher taxa, our model needs to be computationally optimised to perform predictive *a posteriori* simulations using parameters estimated from empirical data. At present, there are many methods to estimate speciation and extinction rates (Morlon 2014), but there are very few methods to estimate origination rate from extant

taxa. A framework to estimate origination rate in a constant birth-death scenario has been developed (Maruvka et al. 2013), and it could be extended to accomodate diversification rate variation. It is also important to highlight that we only considered a model of constant origination through time, which implies that more origination events occur as species number increases, resulting in more higher taxa accumulating towards the present. However, it is unknown if we should expect something similar in real higher taxa and if it is not the case, how it can affect ARCs. Simulated clde richnesss frequency distributions (Supplementary Fig. S8 available on Dryad) are in some ways similar to the empirical ones (Supplementary Fig. S9 available on Dryad). Given the apparent sensitivity of the results to the origination rate parameter, exploration of the effect of different forms of highertaxon accumulation –such as protracted or diversity-dependent origination– should be evaluated.

In the last decade, the correlation between age and species richness of higher taxa has extensively and increasingly been used to explore the effect of time versus ecological effects on the heterogeneous distribution of species among taxa and regions. Overall, these studies have reported contradictory conclusions regarding the relationship between age and richness (e.g., McPeek and Brown 2007; Rabosky et al. 2012; Hedges et al. 2015). In this study, we show that the conflict probably comes from the fact that ARCs have been estimated using different taxonomic ranks, age types and correlation estimation methods. Moreover, previous interpretations have relied on ARCs estimated from one taxonomic rank only. In the study of the latitudinal diversity gradients, it has been observed that ecologically and time-dependent diversification models cannot be differentiated with a simple estimate of the relationship between age and richness of

clades (Hurlbert and Stegen 2014), conforming to our concerns. ARCs should be used with caution as supporting evidence in diversification analyses. Whenever possible, we encourage researchers to estimate ARCs considering different taxonomic ranks available in their data sets, and to carefully analyse their trends. A null or negative ARC should not be taken as evidence in favour of diversification processes that bound species accumulation, since processes of unbounded diversity can generate such relationships between age and richness as well. However, it can be used to document among-clade rate variability in general. It is important to note that we only considered few diversification models from the wide variety available in the literature. The evaluation of other models such as diversification limited by area or by extinction might change ARC interpretations. Moreover, ARC patterns described here might not hold for subgroups evolving within a general diversification process. For example, Passerine birds have consistently portrayed negative ARCs (Ricklefs 2006; Rabosky 2009), differing from the ARC patterns portrayed by the whole class reported here and elsewhere (Stadler et al. 2014). Finally, we emphasize the importance of incorporating a higher-taxon origination parameter in simulation frameworks, as it appears to have a relevant effect on the behaviour of descriptive statistics of diversification dynamics, whose importance for the study of richness patterns should not be neglected.

SUPPLEMENTARY MATERIAL

Supplementary material, including empirical and simulated data on age and richness and online-only figures and tables, is available in the Dryad digital repository at http://dx.doi.org/10.5061/dryad.2b7d5.

ACKNOWLEDGEMENTS

The Centre de Mathématiques Appliquées at Palaiseau, Paris, France, provided the computational resources to run simulations and perform analyses. LLSR thanks the Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México and the Consejo Nacional de Ciencia y Tecnología, México, for granting scholarship 262540. This article is part of L.L.S.R. PhD research. She thanks her working group at the Instituto de Biología, UNAM: T. Hernández, S. Gómez-Acevedo, J.A. Barba, R. Hernández, A. Benítez, A. López, I. Fragoso, P. Rivera and her professors E.R. Rodrigues, L.E. Eguiarte, D. Piñero, G. Salazar and I. Cacho for support, discussions and feedback. She also thanks J. Rolland, F. Condamine, D. Moen, J. Smrckova, J. Green, F. Gascuel, A. Lambert, H. Sauquet, J. Bardin, and P. Simion for discussions and support during research visits to Morlon's Lab. The authors thank the Editor-in-chief, F. Anderson; the Associate Editor, S. Renner; and A. Phillimore and D. Rabosky for valuable comments that greatly improved this manuscript.

CONFLICT OF INTEREST: None declared.

REFERENCES

Alfaro M.E., Santini F., Brock C., Alamillo H., Dornburg A., Rabosky D.L., Carnevale G., Harmon L.J. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proc. Natl. Acad. Sci. U. S. A. 106:13410–4.

Agapow P., Isaac N. 2002. Macrocaic: revealing correlates of species richness by comparative analysis. Divers. Distrib. 8:41–43.

Bokma F., Baek S.K., Minnhagen P. 2014. 50 Years of Inordinate Fondness. Syst. Biol. 63:251–256.

Darwin C. 1859. On the origin of species by means of natural selection. London: Murray.

Escudero M., Hipp A. 2013. Shifts in diversification rates and clade ages explain species richness in higher-level sedge taxa (Cyperaceae). Am. J. Bot. 100:2403–2411.

Etienne R.S.R., Haegeman B., Stadler T., Aze T., Pearson P.N., Purvis A., Phillimore A.B. 2011. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. Proc. R. Soc. B. 279:1300–1309.

Felsenstein J. 1985. Phylogenies and the Comparative Method. Am. Nat. 125:1–15.

Freckleton R.P., Harvey P.H., Pagel M.D. 2002. Phylogenetic Analysis and Comparative Data: a test and review of evidence. Am. Nat. 160:712–726.

Freckleton R.P., Phillimore A.B., Pagel M. 2008. Relating Traits to Diversification: A Simple Test. Am. Nat. 172:102–115.

Frost D.R. 2015. Amphibian Species of the World: an Online Reference. Version 6.0 (Accessed February 2015). Electronic Database accessible at http://research.amnh.org/herpetology/amphibia/index.html. American Museum of Natural History, New York, USA.

Hadfield J.D., Nakagawa S. 2010. General quantitative genetic methods for comparative biology: Phylogenies, taxonomies and multi-trait models for continuous and categorical characters. J. Evol. Biol. 23:494–508.

Hedges S.B., Marin J., Suleski M., Paymer M., Kumar S. 2015. Tree of life reveals clock-like speciation and diversification. Mol. Biol. Evol. msv037.

Hey J. 1992. Using Phylogenetic Trees to Study Speciation and Extinction. Evolution. 46:627–640.

Humphreys A.M., Barraclough T.G. 2014. The evolutionary reality of higher taxa in mammals. Proc. R. Soc. B. 281(1783):20132750.

Hurlbert A.H., Stegen J.C. 2014. On the processes generating latitudinal richness gradients: identifying diagnostic patterns and predictions. Front. Genet. 5:420.

Isaac N.J.B., Agapow P.M., Harvey P.H., Purvis A. 2003. Phylogenetically nested comparisons for testing correlates of species richness: a simulation study of continuous variables. Evolution. 57:18–26.

Jetz W., Thomas G., Joy J., Hartmann K., Mooers A. 2012. The global diversity of birds I n space and time. Nature. 491:1–5.

De Lisle S.P., Rowe L. 2015. Independent evolution of the sexes promotes amphibian diversification. Proceeding R. Soc. B. 282:20142213.

Magallón S., Gómez-Acevedo S., Sánchez-Reyes L., Hernández-Hernández T. 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. New Phytol. 207:437–543.

Magallón S., Sanderson M.J. 2001. Absolute diversification rates in angiosperm clades. Evolution. 55:1762–1780.

Maruvka Y.E., Shnerb N.M., Kessler D. a, Ricklefs R.E. 2013. Model for macroevolutionary dynamics. Proc. Natl. Acad. Sci. U. S. A. 110:E2460-9.

McPeek M. a, Brown J.M. 2007. Clade age and not diversification rate explains species richness among animal taxa. Am. Nat. 169:E97–E106.

Meredith R.W., Janečka J.E., Gatesy J., Ryder O.A., Fisher C.A., Teeling E.C., Goodbla A., Eizirik E., Simão T.L.L., Stadler T., Rabosky D.L., Honeycutt R.L., Flynn J.J., Ingram C.M., Steiner C., Williams T.L., Robinson T.J., Burk-Herrick A.,

Westerman M., Ayoub N.A., Springer M.S., Murphy W.J. 2011. Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. Science. 334(6055):521–524.

Mittelbach G.G., Schemske D.W., Cornell H. V, Allen A.P., Brown J.M., Bush M.B., Harrison S.P., Hurlbert A.H., Knowlton N., Lessios H.A., McCain C.M., McCune A.R., McDade L.A., McPeek M.A., Near T.J., Price T.D., Ricklefs R.E., Roy K., Sax D.F., Schluter D., Sobel J.M., Turelli M. 2007. Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. Ecol. Lett. 10:315–331.

Morlon H. 2014. Phylogenetic approaches for studying diversification. Ecol. Lett. 17:508–525.

Morlon H., Lewitus E., Condamine F.L., Manceau M., Clavel J., Drury J. 2016. RPANDA: an R package for macroevolutionary analyses on phylogenetic trees. Methods Ecol. Evol. 7:589-597.

Morlon H., Parsons T., Plotkin J. 2011. Reconciling molecular phylogenies with the fossil record. Proc. Natl. Acad. Sci. U. S. A. 108:16327–16332.

Morlon H., Potts M.D., Plotkin J.B. 2010. Inferring the dynamics of diversification: A c oalescent approach. PLoS Biol. 8:e1000493.

Nee S., Mooers A., Harvey P. 1992. Tempo and mode of evolution revealed from molecular phylogenies. Proc. Natl. Acad. Sci. U. S. A. 48:523–529.

Orme D., Freckleton R., Thomas G., Petzoldt T., Fritz S., Isaac N., Pearse W. 2013. caper: Comparative Analyses of Phylogenetics and Evolution in R. R package version 0.5.2. Available from URL https://cran.r-project.org/web/packages/caper/

Pagel M. 1999. Inferring the historical patterns of biological evolution. Nature. 401:877–884.

Pybus O.G., Harvey P.H. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. Proc. R. Soc. B. 267:2267–2272.

Pyron R.A., Burbrink F.T. 2012. Extinction, ecological opportunity, and the origins of global snake diversity. Evolution. 66:163–178.

Pyron R.A., Burbrink F.T. 2014. Early origin of viviparity and multiple reversions to oviparity in squamate reptiles. Ecol. Lett. 17:13–21.

Rabosky D.L. 2006a. Likelihood methods for detecting temporal shifts in diversification rates. Evolution. 60:1152–1164.

Rabosky D.L. 2006b. LASER: a maximum likelihood toolkit for detecting temporal shifts In diversification rates from molecular phylogenies. Evol. Bioinform. Online. 2:247-250.

Rabosky D.L. 2009a. Ecological limits and diversification rate: alternative paradigms to explain the variation in species richness among clades and regions. Ecol. Lett. 12:735–743.

Rabosky D.L. 2009b. Ecological limits on clade diversification in higher taxa. Am. Nat. 173:662–674.

Rabosky D.L. 2010. Primary controls on species richness in higher taxa. Syst. Biol. 59:634–645.

Rabosky D.L. 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. PLoS One. 9:e89543.

Rabosky D.L., Lovette I.J. 2008. Density-dependent diversification in North American wood warblers. Proc. R. Soc. B. 275:2363–2371.

Rabosky D.L., Slater G.J., Alfaro M.E. 2012. Clade age and species richness are decoupled across the eukaryotic tree of life. PLoS Biol. 10:e1001381.

R Core Team. 2014. R: A Language and Environment for Statistical Computing. Available from URL https://cran.r-project.org.

Rambaut A., Drummond A. 2007. TreeAnnotator.

Raup D.M., Gould S.J., Schopf T.J.M., Simberloff D.S. 1973. Stochastic models of phylogeny and the evolution of diversity. J. Geol. 81:525–542.

Ricklefs R.E. 2006. Global Variation in the Diversification Rate of Passerine Birds. Ecology. 87:2468–2478.

Ricklefs R.E., Losos J.B., Townsend T.M. 2007. Evolutionary diversification of clades of squamate reptiles. J. Evol. Biol. 20:1751–1762.

Ricklefs R.E., Renner S.S. 1994. Species Richness Within Families of Flowering Plants. Evolution. 48:1619–1636.

Robeck H.E., Maley C.C., Donoghue M.J. 2000. Taxonomy and temporal diversity patterns. Paleobiology. 26:171–187.

Soul L.C., Friedman M. 2015. Taxonomy and phylogeny can yield comparable results in comparative paleontological analyses. Syst. Biol. 64:608–620.

Stadler T., Rabosky D.L., Ricklefs R.E., Bokma F. 2014. On Age and Species Richness of Higher Taxa. Am. Nat. 184:447–455.

Stephens P.R., Wiens J.J. 2003. Explaining species richness from continents to communities: the time-for-speciation effect in emydid turtles. Am. Nat. 161:112–128.

Stevens P. 2001. Angiosperm Phylogeny Website. Available from http://www.mobot.org/MOBOT/research/APweb/ (accessed July 2014).

Uetz P., Josek J. 1995. The Reptile Database. Available from http://reptile-database.reptarium.cz/ (accessed December 2014).

Wagner P.J. 1995. Diversity Patterns Among Early Gastropods: Contrasting Taxonomic and Phylogenetic Descriptions. Paleobiology. 21:410–439.

Wallace A.R. 1878. Tropical Nature and Other Essays. New York: McMillan Press.

Wiens J.J., Graham C.H., Moen D.S., Smith S. a, Reeder T.W. 2006. Evolutionary and ecological causes of the latitudinal diversity gradient in hylid frogs: treefrog trees unearth the roots of high tropical diversity. Am. Nat. 168:579–596.

Wilson D.E., Reeder D.M. 2005. Mammals species of the world. A taxonomic and geographic reference. Baltimore: Johns Hopkins University Press.

Yule G. 1925. A mathematical theory of Evolution, Based on the Conclusions of Dr. J. C. Willis, F.R.S. Philos. Trans. R. Soc. London. Ser. B. Contain. Pap. a Biol. Character 213:21–87.

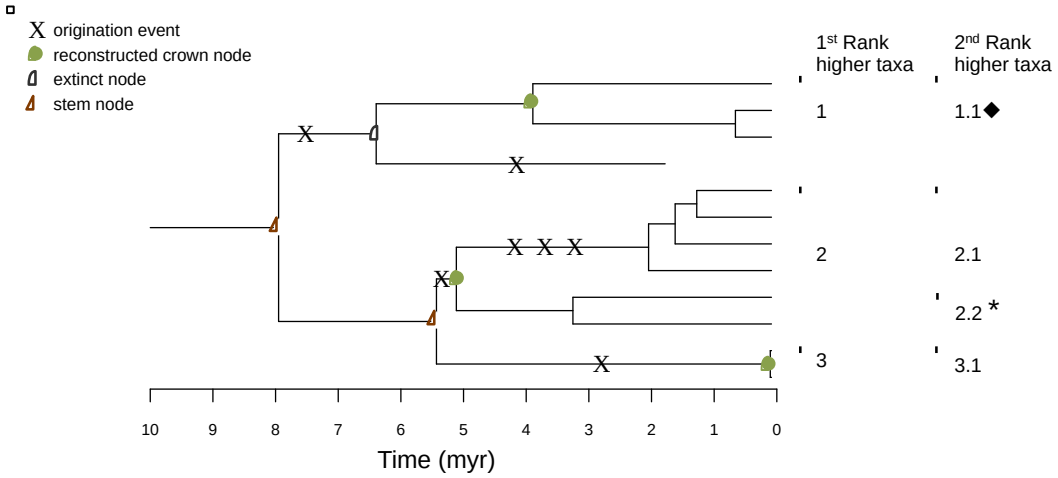


FIGURE 1. Example tree simulated with the model of hierarchical origin of higher taxa proposed in here. Highertaxon origination events (X) can occur at any branch. The node descending from the higher-taxon origination event deepest in the tree corresponds to the crown node of a 1st rank higher taxon (filled circles). The node subtending the branch where the origination event occurred represents the stem node of the higher taxon (open triangles). Nodes in which any of the two daughter branches went extinct cannot define higher taxa (empty circle). To preserve monophyly, a clade that lacks origination events, but is sister to a higher taxon, is defined as a higher taxon of the same rank as its sister, as exemplified by taxon 2.2 (asterisks). Clades with fewer origination events are assigned to all less inclusive taxonomic ranks generated during the simulation, as exemplified by taxon 1.1 (filled diamonds). This *a posteriori* highertaxon delimitation guarantees the inclusion of all lineages of a clade in all ranks.

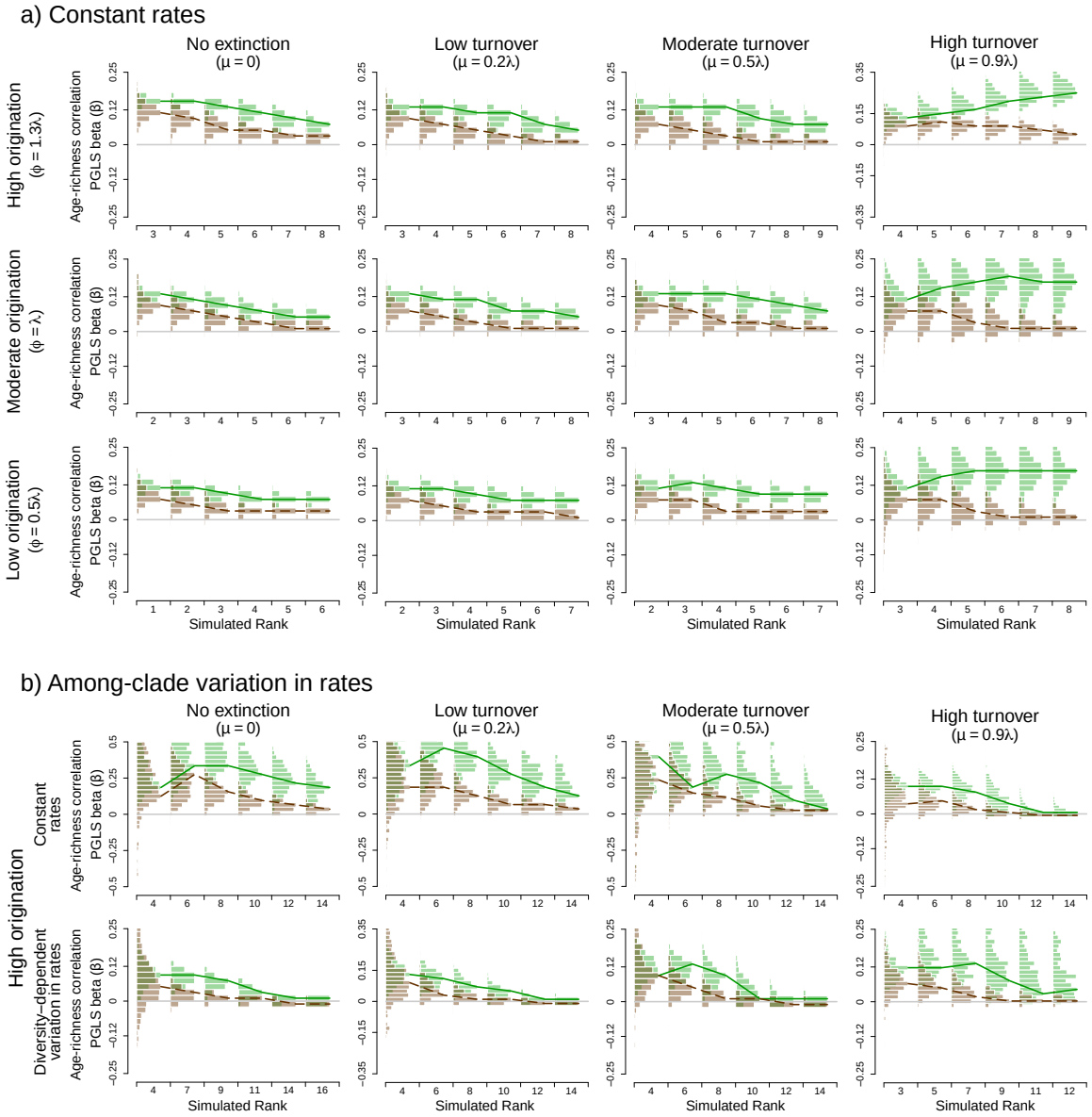


FIGURE 2. Expected relationship between clade age and species richness inferred with phylogenetic generalized least squares (PGLS). X-axis numbers denote simulated taxonomic ranks, from the most to the least inclusive. At each simulated taxonomic rank, ARCs are estimated using crown ages (crown-ARCs) and stem ages (stem-ARCs). A solid line across distributions connects crown-ARC modes of each rank while a dashed line connects stem-ARC modes. PGLS was applied to data simulated under models of

constant speciation and extinction rates (a) and of among-clade variation in speciation and extinction (b) with no rate change through time (b – top row), and with diversity-dependent rate variation (b – bottom row). Different extinction (increasing μ from left to right) and origination rates (increasing $\mathbf{\Phi}$ from bottom to top in (a); for (b) we only show results obtained with a high origination) were considered. To facilitate comparisons, results are shown for the six ranks displaying the most extreme ARC values of each model. The horizontal grey line indicates a correlation of zero.

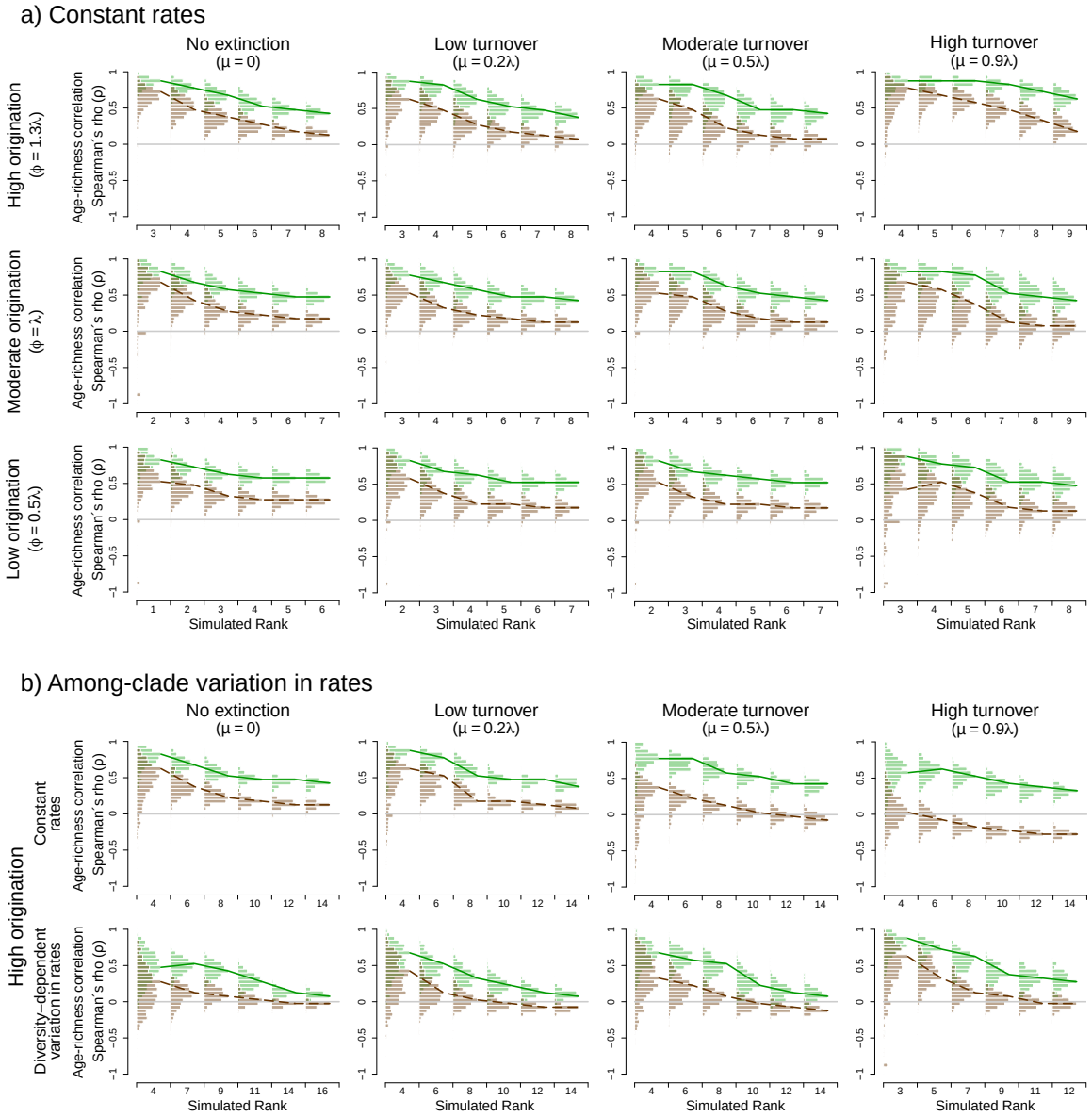


FIGURE 3. Expected relationship between clade age and species richness inferred with Spearman's rank test. X-axis numbers denote simulated taxonomic ranks, from the most to the least inclusive. At each simulated taxonomic rank, ARCs are estimated using crown ages (crown-ARCs) and stem ages (stem-ARCs). A solid line across distributions connects crown-ARC modes of each rank while a dashed line connects stem-ARC modes. Spearman's test was applied to data simulated under models of constant

speciation and extinction rates (a) and of among-clade variation in speciation and extinction (b) with no rate change through time (b – top row), and with diversity-dependent rate variation (b – bottom row). Different extinction (increasing μ from left to right) and origination rates (increasing $\mathbf{\Phi}$ from bottom to top in (a); for (b) we only show results obtained with a high origination) were considered. To facilitate comparisons, results are shown for the six ranks displaying the most extreme ARC values of each model. The horizontal grey line indicates a correlation of zero.

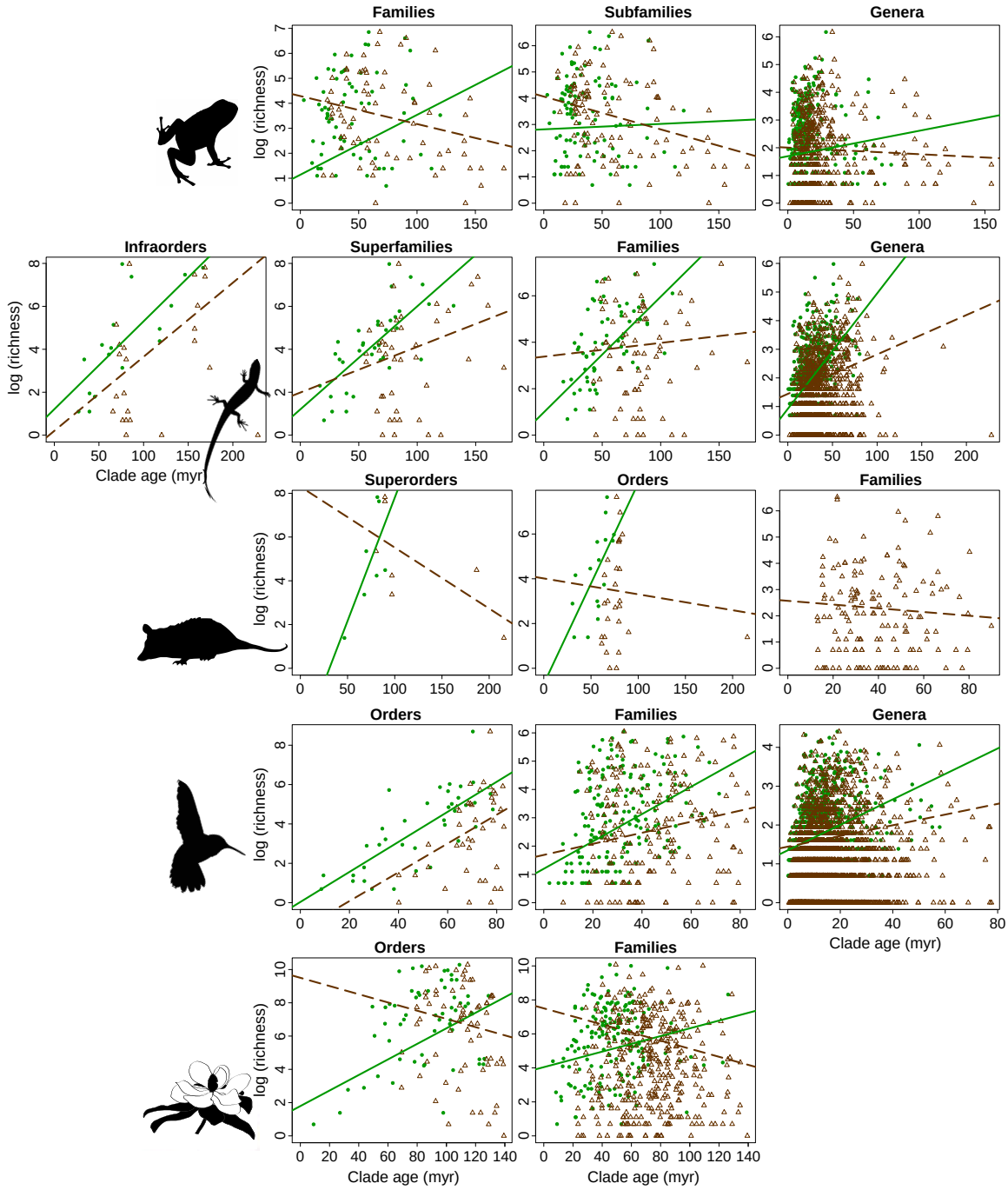


FIGURE 4. Empirical relationship between clade age (filled circles for crown and open triangles for stem age) and richness of higher taxa from different taxonomic ranks of amphibians, scaled reptiles, mammals, birds, and flowering plants, estimated with phylogenetic generalized least squares (PGLS). Lines represent the fitted relationship

between crown age and richness (solid) and between stem age and richness (dashed). Organismal silhouettes are available from www.phylopic.org. Material is presented unmodified. Mammal material was provided by Sarah Werning under a CC by 3.0 license (creativecommons.org/licenses/by/3.0).

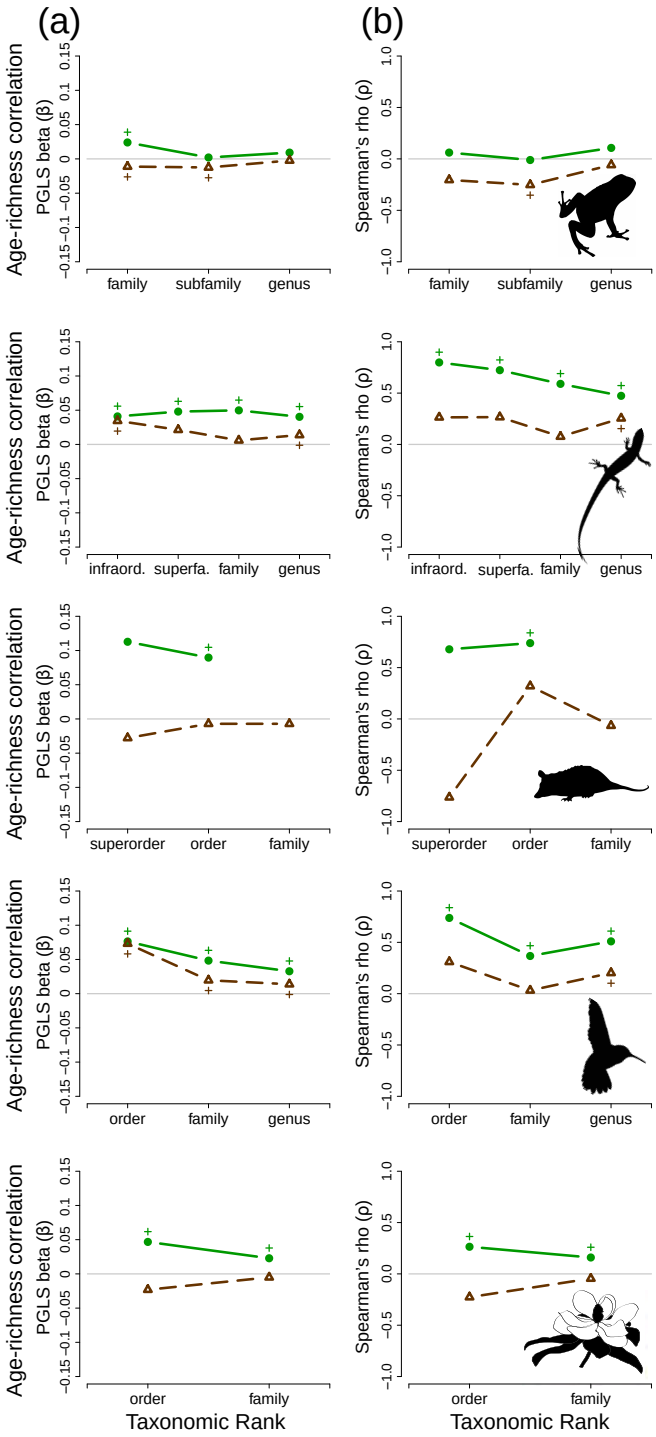


FIGURE 5. Relationship between crown age and richness (filled circles, solid lines) and between stem age and richness (open triangles, dashed lines) of higher taxa from different taxonomic ranks of amphibians, scaled reptiles, mammals, birds and flowering plants, as

estimated with a) phylogenetic generalized least squares (PGLS) and b) Spearman´s rank test. A horizontal grey line is drawn at β (or ϱ) = 0; + significant correlation ($p \leq 0.05$). Exact values are shown in Supplementary Table S2 available on Dryad. Organismal silhouettes are available from www.phylopic.org. Material is presented unmodified. Mammal material was provided by Sarah Werning under a CC by 3.0 license (creativecommons.org/licenses/by/3.0).

Material Suplementario

Los datos de riqueza y edad de clados empíricos y simulados usados en los análisis presentados en este Capítulo 3, así como las tablas y figuras suplementarias que se presentan a continuación, se encuentran disponibles en el repositorio digital Dryad http://dx.doi.org/10.5061/dryad.2b7d5

.

SUPPLEMENTARY TABLES

Sánchez-Reyes L.L., Morlon H., Magallón S. 2016. Uncovering higher-taxon diversification dynamics from clade age and species-richness data. Systematic Biology.

TABLE S1. Parameter values used to simulate trees under the diversification models evaluated in this study.

Diversification Model	Extinction rate (μ) regime[a]	Origination rate (Φ) regime[b]	Speciation rate (λ; initial speciation rate λ_0)
Constant rates	absent	Low	λ = 0.125
		Moderate	λ = 0.13
		High	λ = 0.16
	low	Low	λ = 0.3
		Moderate	λ = 0.3
		High	λ = 0.3
	moderate	Low	λ = 0.6
		Moderate	λ = 0.6
		High	λ = 0.6
	high	Low	λ = 0.8
		Moderate	λ = 0.8
		High	λ = 0.8
Among-clade rate change & constant rates through time	absent	Φ = 0.65	λ_0 = 0.1
	low	Φ = 0.65	λ_0 = 0.1
	moderate	Φ = 0.65	λ_0 = 0.1
	high	Φ = 0.65	λ_0 = 1.5
Among-clade rate change & diversity-dependent decrease	absent	Φ = 0.65	λ_0 = 0.25
	low	Φ = 0.65	λ_0 = 0.2
	moderate	Φ = 0.65	λ_0 = 0.3
	high	Φ = 0.65	λ_0 = 5

a) Extinction rate regimes correspond to =0 (absent), =0.2*λ (low), =0.5*λ (moderate) and =0.9*λ (high).

b) Origination rate regimes correspond to =0.5*λ (low), = λ (moderate) and =1.3*λ (high).

TABLE S2. Summary of ARC estimates resulting from the analyses of data sets including non-monophyletic taxa.

Biological System	Taxonomic Rank (df[a] crown, df[a] stem*)	Mean age in Myrs (min, max, sd)[b]		PGLS β (p value)		Spearman's ρ (p value)	
		crown	stem*	crown	stem*	crown	stem*
	Family (65,67)	45.75 (2.97,119.9,27.2)	72.23 (19.87,174.4,39.5)	0.024 (0.009)	-0.011 (0.024)	0.061 (0.621)	-0.205 (0.091)
Amphibians	Subfamily (97,100)	36.74 (2.97,119.9,24.2)	58.8 (12.11,174.4,37.1)	0.002 (0.764)	-0.013 (0.001)	-0.011 (0.912)	-0.253 (0.010)
	Genus (292,433)	15.82 (0.293,78.66,13.9)	27.03 (1.979,154.9,24.6)	0.009 (0.085)	0.002 (0.365)	0.107 (0.067)	-0.059 (0.260)
	Infraorder (15,18)	79.31 (27.21,166.6,42.2)	81.43 (65.63,174.1,42.9)	0.041 (4.6e-4)	0.035 (0.006)	0.811 (7.74e-5)	0.281 (0.87)
Squamates	Superfamily (33,36)	62.6 (16.74,131.1,27.4)	91.81 (56.35,174.1,30.0)	0.048 (7.77e-7)	0.021 (0.053)	0.737 (4.54e-7)	0.281 (0.087)
	Family (57,62)	53.52 (16.74,110.1,22.1)	81.56 (42.62,174.1,26.8)	0.05 (3.55e-7)	0.059 (0.564)	0.579 (1.56e-6)	0.061 (0.63)
	Genus (498,690)	25.27 (0.585,80.4,14.7)	37 (5.036,174.1,19.9)	0.033 (<2.2e-16)	0.014 (2.64e-13)	0.482 (<2.2e-16)	0.262 (2.63e-12)
	Superorder (5,5)	74.26 (46.73,89.86,14.4)	122.2 (80.21,215.6,54.8)	0.113 (0.075)	-0.028 (0.102)	0.679 (0.09)	-0.764 (0.061)
Mammals	Order (14,23)	55.51 (30.38,74.25,13.8)	79.99 (61.41,215.6,29.3)	0.09 (0.007)	-0.007 (0.624)	0.739 (0.001)	0.32 (0.119)
	Family (NA,117)	NA	39.41 (13.01,89.86,17.6)	NA	-0.007 (0.391)	NA	-0.066 (0.477)
	Order (35,35)	48.84 (8.58,78.4,19.3)	70.34 (40.11,82.9,10.1)	0.076 (5.83e-8)	0.073 (0.02)	0.738 (1.87e-7)	0.31 (0.062)
Birds	Family (171,171)	31.28 (2.307,71.42,14.2)	44.15 (14.76,82.9,17.0)	0.048 (3.47e-10)	0.02 (0.017)	0.367 (6.71e-7)	0.031 (0.682)

	Genus (1232,1232)	10.4 (0.159,57.91,8.0)	16.45 (0.822,77.02,11.0)	0.06 (<2.2e-16)	0.016 (4.3e-11)	0.51 (<2.2e-16)	0.202 (8.42e-13)
Angiosperms	Order (57,63)	86.35 (9.08,128.9,26.2)	109.8 (69.35,139,17.8)	0.047 (1.7e-4)	-0.023 (0.219)	0.264 (0.043)	-0.227 (0.079)
	Family (165,355)	47.2 (6.253,126.2,21.1)	73.13 (22.02,134.6,23.2)	0.023 (0.004)	-0.005 (0.349)	0.16 (0.039)	-0.046 (0.39)

(*) monotypic lineages excluded from the analyses

(a) df = Degrees of freedom (N-2)

(b) min = minimum age, max = maximum age, sd = standard deviation

TABLE S3. Summary of ARC estimates resulting from the analyses of data sets excluding non-monophyletic taxa.

Biological System	Taxonomic Rank (df[a] crown, df[a] stem*)	Mean age in Myrs (min, max, sd)[b]		PGLS β (p value)		Spearman's ρ (p value)	
		crown	stem*	crown	stem*	crown	stem*
Amphibians	Family (60,62)	45.54 (2.97,119.9,27.4)	73.59 (19.87,174.4,40.1)	0.019 (0.055)	-0.013 (0.011)	-0.0008 (0.996)	-0.256 (0.042)
	Subfamily (89,92)	36.98 (2.97,119.9, 24.3)	59.44 (12.11,174.4,38.1)	-0.008 (0.193)	-0.013 (0.0007)	-0.032 (0.762)	-0.271 (0.008)
	Genus (244,316)	15.46 (0.293,78.66,14.3)	27.08 (1.979,154.9,25.1)	0.011 (0.056)	-0.002 (0.419)	0.092 (0.152)	-0.057 (0.312)
Squamates	Infraorder (10,13)	61.55 (27.21,146.4,32.1)	94.11 (65.63,174.1,38.1)	0.055 (8.7e-4)	0.035 (0.02)	0.799 (0.002)	0.102 (0.712)
	Superfamily (31,34)	60.11 (16.74,110.1,25.3)	90.12 (56.35,174.1,28.5)	0.052 (6.06e-7)	0.018 (0.131)	0.723 (1.98e-6)	0.235 (0.168)
	Family (57,62)	53.62 (16.74,110.1,22.5)	82.45 (42.62,174.1,26.8)	0.044 (1.01e-6)	0.010 (0.222)	0.590 (1.37e-6)	0.076 (0.573)
	Genus (498,690)	22.92 (0.585,80.4,13.0)	36.07 (5.036,11.1,18.9)	0.039 (<2.2e-16)	0.012 (1.77e-9)	0.473 (<2.2e-16)	0.254 (2.05e-7)
Mammals	Superorder (5,5)	71.66 (46.73,83.31,13.9)	111.5 (81.21,215.6,51.4)	0.154 (0.032)	-0.036 (0.097)	0.886 (0.018)	-0.794 (0.090)
	Order	-	-	-	-	-	-
	Family	-	-	-	-	-	-
Birds	Order (35,35)	48.02 (8.58,77.45,18.9)	70.03 (40.11,82.9,10.1)	0.079 (5.92e-8)	0.073 (0.025)	0.753 (1.18e-7)	0.304 (0.072)
	Family (171,171)	28.80 (2.307,71.42,14.4)	45.28 (14.76,82.9,18.4)	0.042 (6.97e-7)	0.013 (0.069)	0.439 (5.37e-7)	0.126 (0.169)

	Genus (1232,1232)	9.12 (0.159,55.09,7.1)	15.84 (0.822,77.02,10.8)	0.045 (<2.2e-16)	0.013 (1.34e-7)	0.505 (<2.2e-16)	0.186 (3.22e-9)
Angiosperms	Order	-	-	-	-	-	-
	Family	-	-	-	-	-	-

(*) monotypic lineages excluded from the analyses

(a) df = Degrees of freedom (N-2)

(b) min = minimum age, max = maximum age, sd = standard deviation

(-) taxonomic ranks without non-monophyletic taxa in the analysed data sets.

TABLE S4. Phylogenetic structure of covariation between age and richness data, estimated with Pagels' (1999) lambda test.

Biological System	**Taxonomic Rank**	**Pagel's (1999) lambda ML estimate [95% CI]**		**logLikelihood**		**Significantly different from 0 (phylogenetically structured in some way)**		**Significantly different from 1 (various explanations, see main text)**	
		crown	**stem***	**crown**	**stem***	**crown**	**stem***	**crown**	**stem***
Amphibians	Family	0.882 [0.475,1]	0 [0, 0.650]	-120.973	-128.554	yes p = 0.0006	no p = 1	no p = 0.115	yes p = 0.0002
	Subfamily	0.220 [0, 0.695]	1e-6 [0,0.215]	-176.009	-177.656	no p = 0.254	no p = 1	yes p = 1.27e-7	yes p = 1.19e-10
	Genus	0.413 [0.102, 0.716]	0.144 [0.014,0.396]	-437.401	-552.526	yes p = 5.83e-5	yes p = 0.012	yes p=0	yes p = 0
Squamates	Infraorder	0.000001 [0,0.912]	0.000001 [0,0.755]	-30.48887	-34.56699	no p = 1	no p = 1	yes p = 0.0309	yes p = 0.0110
	Superfamily	0.000001 [0,0.859]	0.0492	-56.81650	-68.23715	no p = 1	no p = 0.866	yes p = 0.0242	yes p = 0.0134
	Family	0.000001 [0,1]	0.000001 [0,1]	-17.28827	-20.21616	no p = 1	no p = 1	no p = 0.094	no p = 0.0711
	Genus	0.000001 [0,1]	0.000001 [0,0.749]	-43.93969	-46.80911	no p = 1	no p = 1	no p = 0.107	yes p = 0.0211
Mammals	Superorder	0.000001 [0,1]	0.000001 [0,1]	-12.76832	-13.14386	no p = 1	no p = 1	no p = 0.0883	no p = 0.0905
	Order	0.000001 [0,1]	0.000001 [0,1]	-28.25410	-52.41119	no p = 1	no p = 1	no p = 0.199	no p = 0.117
	Family	NA	0.486	NA	-207.6775	NA	no	NA	yes

							p = 0.508		p = 1.01e-4
Birds	Order	0.000001 [0,1]	0.000001 [0,1]	-60.91298	-73.77465	no p = 1	no p = 1	no p = 0.0748	no p = 0.207
	Family	0.750 [0.404,0.979]	0.723 [0.296,0.982]	-291.9804	-308.9319	yes p = 5.37e-6	yes p = 9.33e-4	yes p = 2.68e-2	yes p = 0.0307
	Genus	0.344 [0.205,0.488]	0.0512	-1382.543	-1510.098	yes p = 1.78e-15	yes p = 9.82e-4	yes p < 2.2e-16	yes p < 2.2e-16
Angiosperms	Order	1 [0,1]	0.000001 [0, 0.963]	-129.4544	-134.4796	no p = 0.111	no p = 1	no p = 1	yes p = 0.0463
	Family	0.503 [0.0871,0.917]	0.000001 [0,0.606]	-355.5117	-355.7116	yes p = 0.0145	no p = 1	yes p = 0.0141	yes p = 0.000371

(*) monotypic lineages excluded from the analyses

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, **401**, 877–884.

SUPPLEMENTARY FIGURES

Sánchez-Reyes L.L., Morlon H., Magallón S. 2016. Uncovering higher-taxon diversification dynamics from clade age and species-richness data. Systematic Biology.

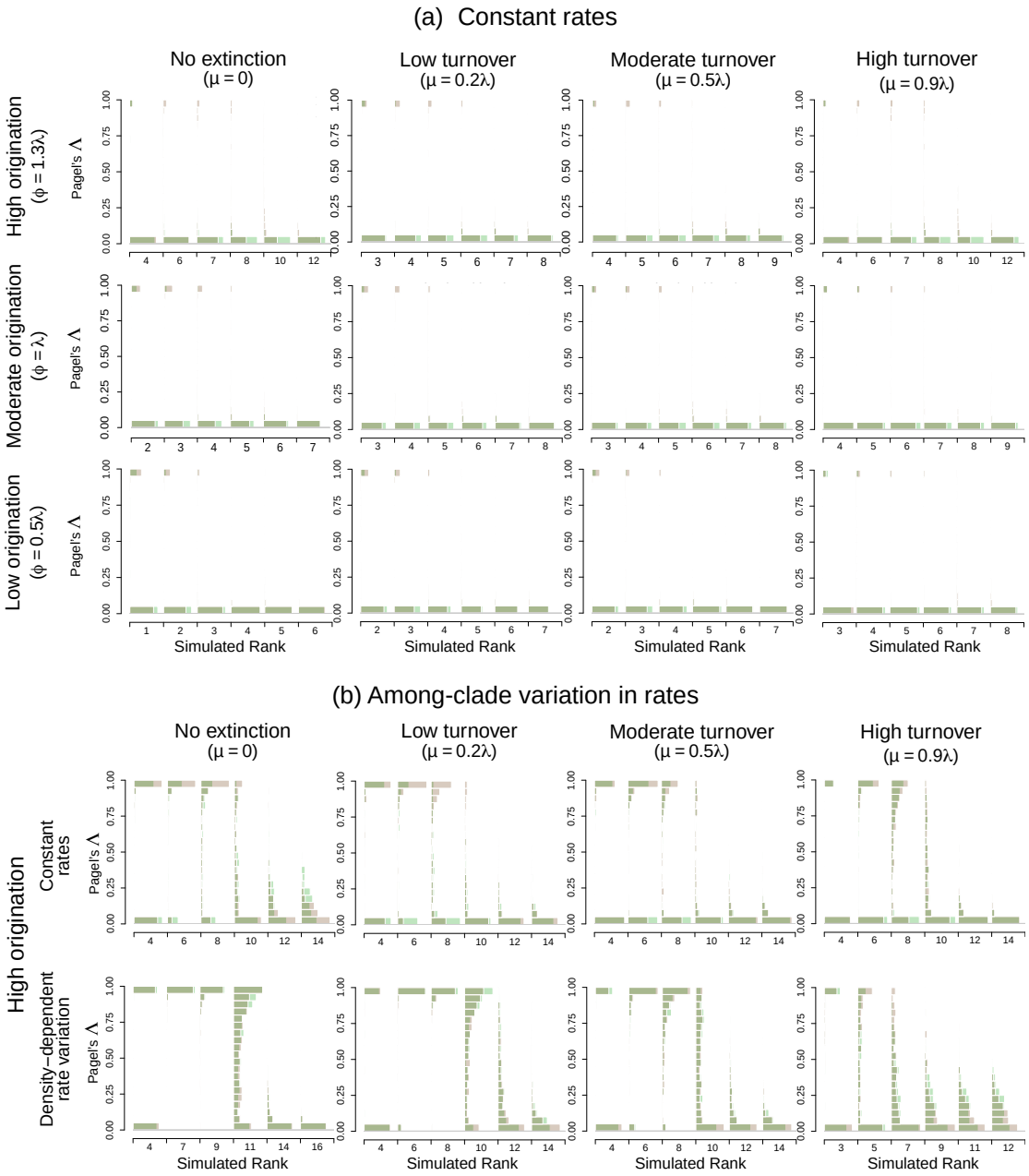


FIGURE S1. Expected phylogenetic structure of covariation between clade age and richness, as estimated with Pagel's (1999) lambda (Λ) test. X-axis numbers denote

simulated taxonomic ranks, from the most to the least inclusive. At each simulated taxonomic rank, Λ is estimated using crown ages (green) and stem ages (brown). Data was simulated under models of constant speciation and extinction rates (a) and of among-clade variation in speciation and extinction (b) with no rate change through time (b – top row) and with diversity-dependent rate variation (b – bottom row). Different extinction (increasing μ from left to right) and origination rates (increasing from bottom to top in a; for b we only show results obtained with a high origination) were considered. To facilitate comparisons, results are shown for six ranks displaying the most extreme values of Λ of each model.

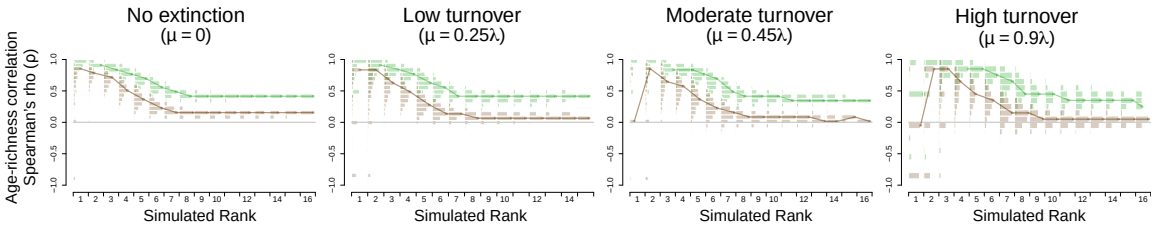


FIGURE S2. Expected relationship between clade age and species richness inferred with Spearman's rank test in all simulated taxonomic ranks. X-axis numbers denote simulated taxonomic ranks, from the most to the least inclusive. Distributions show crown age-richness correlations (crown-ARCs) in green and stem age-richness correlations (stem-ARCs) in brown. ARCs are estimated with data simulated under models of constant speciation and extinction rates, different values of extinction (increasing μ from left to right) and a high origination rate. To estimate correlations, at least three clades are needed. Correspondingly, taxonomic ranks shown are represented by at least three clades in each simulated tree. Lines across distributions connect ARC modes of each rank.

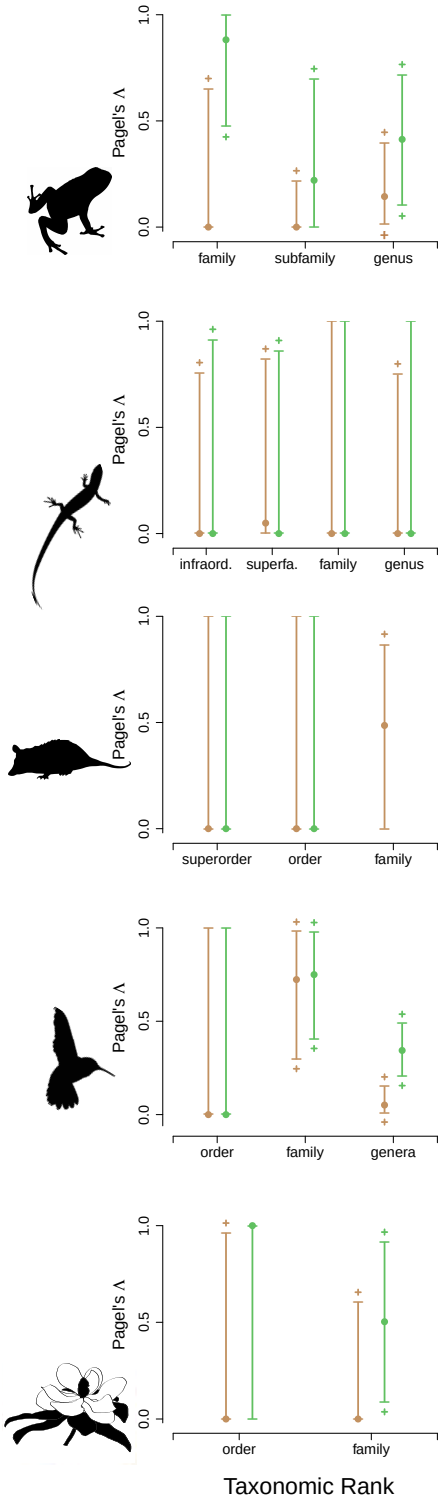


FIGURE S3. Empirical phylogenetic structure of covariation between clade age and richness, as estimated with Pagel's (1999) lambda (Λ) test evaluated at different

taxonomic ranks of amphibians, scaled reptiles, mammals, birds, and flowering plants. Organismal silhouettes are available from www.phylopic.org. Material is presented unmodified. Mammal material was provided by Sarah Werning under a CC by 3.0 license (creativecommons.org/licenses/by/3.0).

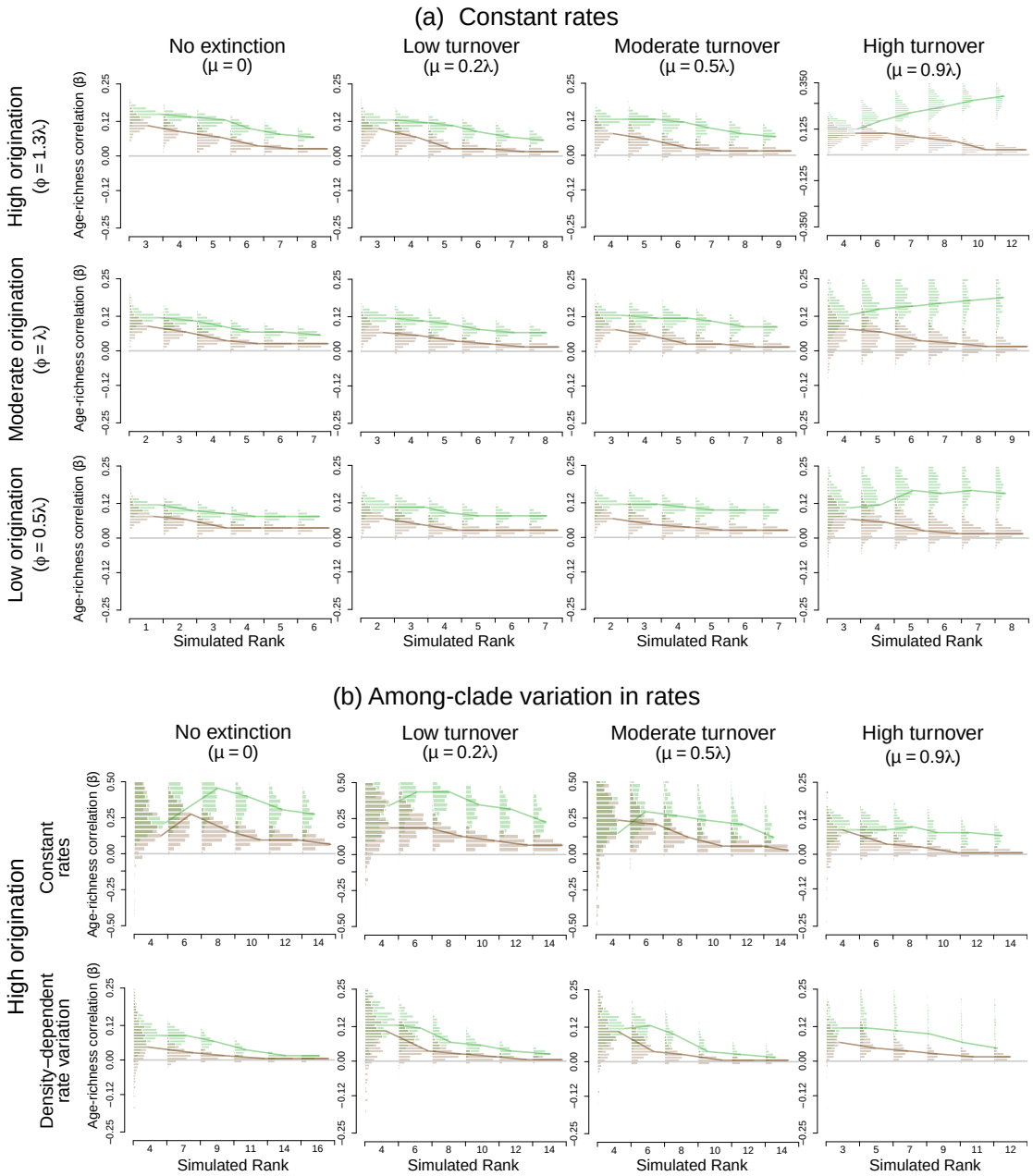


FIGURE S4. Expected relationship between clade age and species richness inferred with phylogenetic generalized least squares (PGLS) when assuming complete phylogenetic structure (Pagel's Λ = 1). Distributions show crown age-richness correlations (crown-ARCs) in green, and stem age-richness correlations (stem-ARCs) in brown. ARCs are estimated across different taxonomic ranks simulated under models of constant speciation

and extinction rates (a) and of among-clade variation in speciation and extinction (b) with no rate change through time (b – top row), and with diversity-dependent rate variation (b – bottom row). X-axis numbers denote simulated taxonomic ranks, from the most to the least inclusive. Data shown were simulated with different extinction (increasing μ from left to right) and origination rates (increasing from bottom to top in a; for b we only show results obtained with a high origination). To facilitate comparisons, results are shown for the six ranks displaying the most extreme ARC values from each model. Lines across plots connect ARC modes of each rank.

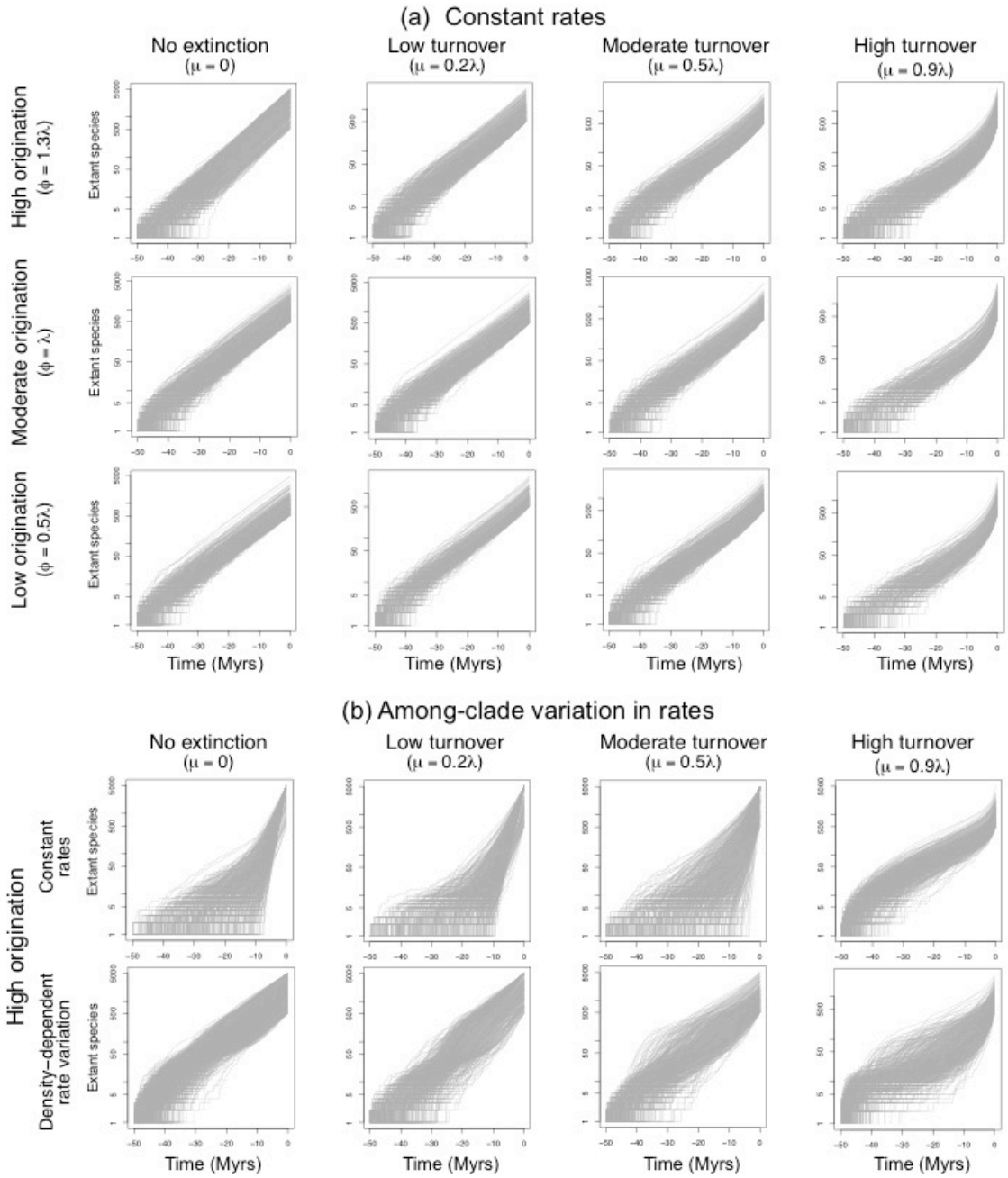


FIGURE S5. Expected lineage through time (LTT) plots obtained from trees simulated under diversification models of (a) constant speciation and extinction, and (b) among-clade variation in speciation and extinction with no change in rates through time (b – upper row), and with density-dependent change in rates (b – bottom row). Data shown were simulated with different extinction (increasing μ from left to right) and origination

rates (increasing from bottom to top in a; for b we only show results obtained with a high origination).

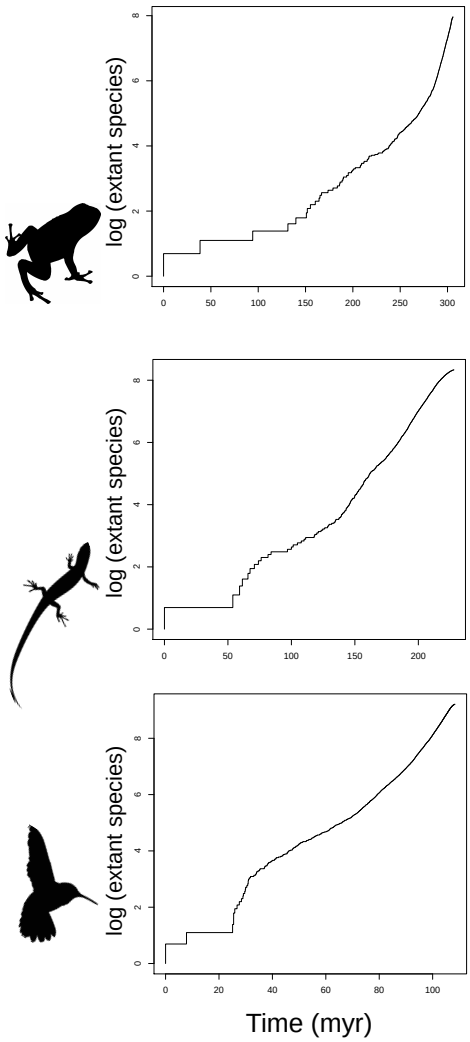


FIGURE S6. Empirical lineage through time (LTT) plots from dated phylogenies of amphibians, scaled reptiles, and birds used to extract clade age and richness data. Organismal silhouettes are available from www.phylopic.org. Material is presented unmodified.

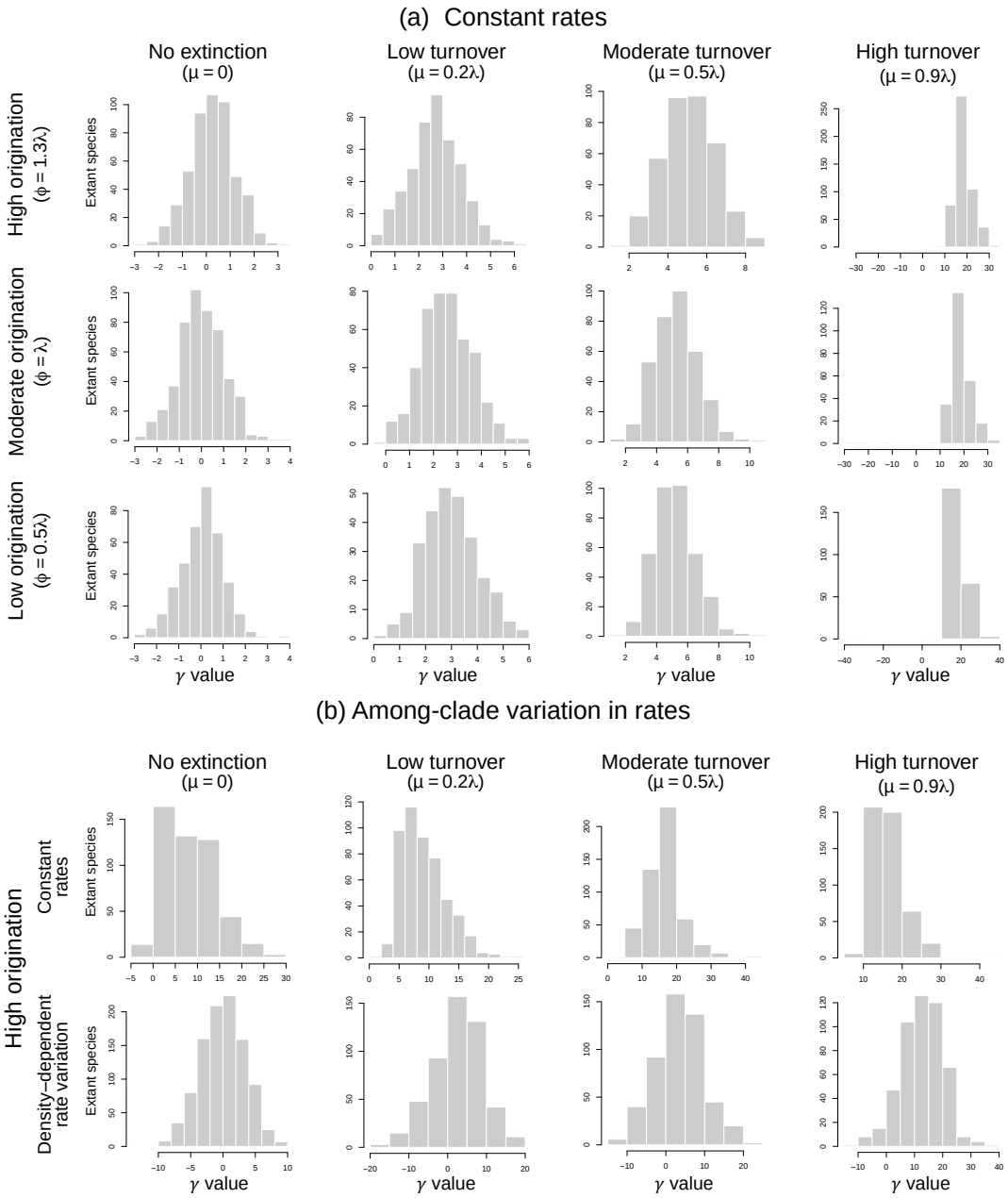


FIGURE S7. Expected distribution of Pybus and Harvey (2000) gamma test (γ) values from trees simulated under diversification models of (a) constant speciation and extinction, and (b) variable among clades speciation and extinction, with no change in rates through time (b – upper row), and with density-dependent change in rates (b – lower row).

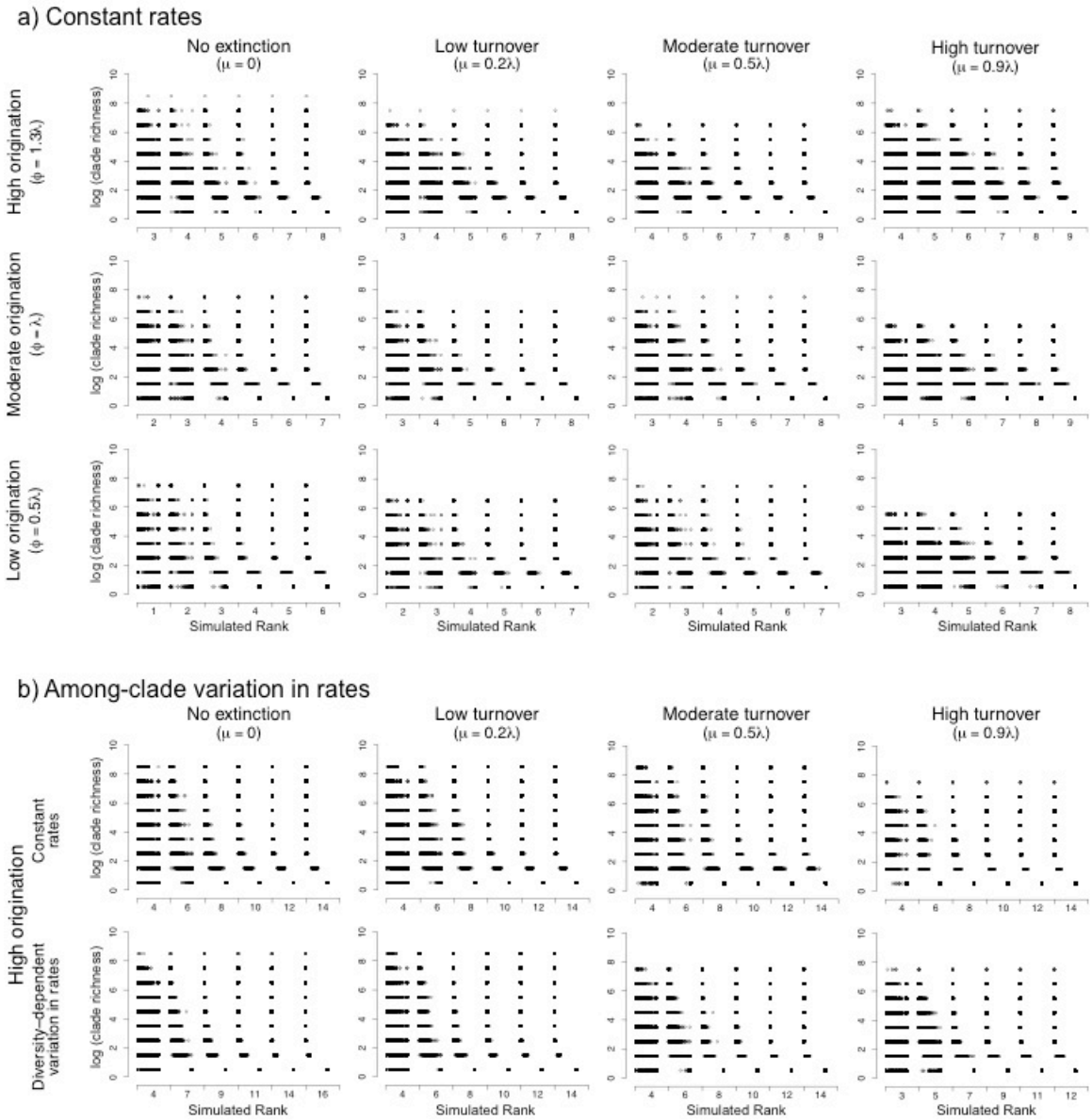


FIGURE S8. Expected clade richness frequency distributions. X-axis numbers denote simulated taxonomic ranks, from the most to the least inclusive. Clade richness data was simulated under models of constant speciation and extinction rates (a) and of among-clade variation in speciation and extinction (b) with no rate change through time (b – top row), and with diversity-dependent rate variation (b – bottom row). Different extinction

(increasing μ from left to right) and origination rates (increasing from bottom to top in a; for b we only show results obtained with a high origination) were considered.

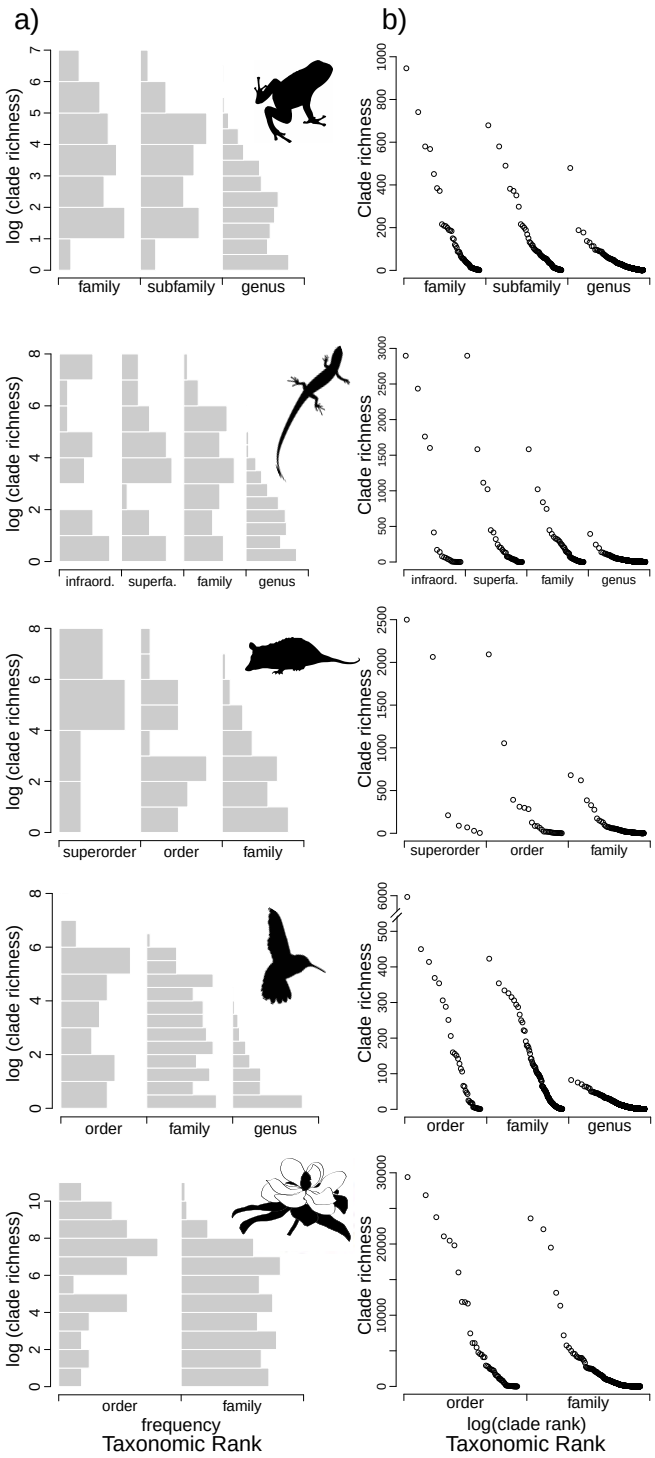


FIGURE S9. (a) Logarithmic (natural) clade richness frequency distributions and (b) logarithmic (natural) rank ordered clade richness of higher taxa from different taxonomic ranks of amphibians, scaled reptiles, mammals, birds, and flowering plants. Organismal

silhouettes are available from www.phylopic.org. Material is presented unmodified. Mammal material was provided by Sarah Werning under a CC by 3.0 license (creativecommons.org/licenses/by/3.0).

SUPPLEMENTARY FIGURES REFERENCES

Pagel M. 1999. Inferring the historical patterns of biological evolution. Nature. 401:877–884.

Pybus O.G., Harvey P.H. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. Proc. R. Soc. B. 267:2267–2272.

DISCUSIÓN GENERAL

El modelo de origen jerárquico de taxa superiores propuesto en este trabajo (Fig. 1, Capítulo 3) permitió la caracterización de la relación entre la edad de los clados y su riqueza de especies (la ARC), un estadístico descriptivo comúnmente utilizado para distinguir entre diferentes hipótesis en estudios de diversificación a nivel macroevolutivo. Esta caracterización aparece como clave para determinar la manera en que la ARC se puede usar para complementar los resultados de métodos de ajuste de modelos de diversificación (Figs. 1 y 2, Capítulo 1). Los métodos de ajuste de modelos con ML y estadística Bayesiana permiten reconstruir el proceso de diversificación con gran detalle usando diferentes tipos de datos, tomados de filogenias fechadas y/o del registro fósil, y pueden ser aplicados en una variedad de sistemas biológicos. Sin embargo, distinguir entre modelos alternativos puede ser complicado y la falta de modelos generales dificulta la evaluación de hipótesis alternativos de diversificación en un mismo grupo de datos (Figs. 2 y 3, Capítulo 2).

Utilidad de la relación entre la edad de los clados y su riqueza de especies para estudiar el proceso de diversificación

Se sabe que, aún proviniendo del mismo proceso de diversificación, existe una discrepancia entre la relación de la edad corona de los clados y su riqueza de especies (ARC corona) y la relación de la edad troncal de los mismos clados con su riqueza de especies (ARC troncal), cuando los clados son delimitados por distancia temporal o genética (Stadler et al. 2014), como es el caso de la clasificación de las aves de Sibley y Ahlquist (1990), la cual está basada en datos de hibridación ADN-ADN. Los resultados

principales de esta tesis muestran que la diferencia entre ARC corona y troncal también puede surgir en una clasificación similar a la implementada en la práctica taxonómica común, en la que los clados están delimitados jerárquicamente, de acuerdo a características descriptivas y, en últimos años, en base a información de sus relaciones filogenéticas (Figs. 2 y 3, Capítulo 3).

La existencia de estas diferencias entre ARCs corona y troncales en rangos taxonómicos como órdenes, familias y géneros, se corroboró en ARCs de cinco sistemas biológicos distintos: anfibios, reptiles escamados, mamíferos, aves y angiospermas (Figs. 4 y 5, Capítulo 3). Los datos simulados y empíricos muestran también que las ARCs no sólo varían dependiendo de la edad usada para estimarlas, sino del rango taxonómico utilizado (Figs. 2-4, Capítulo 3). A pesar de la variación registrada en la ARC, los resultados de las simulaciones permitieron establecer patrones generales en su comportamiento y señalar aspectos del proceso de diversificación sobre los cuales la ARC puede ser informativa.

Informatividad de la diferencia entre ARC corona y ARC troncal

En todos los modelos de diversificación evaluados, las ARC troncales son 1) más débiles que las ARC corona, es decir, con valores inferiores, más cercanos a 0 –la ausencia de correlación–, ó 2) opuestas a las ARC corona, siendo negativas en algunos casos (Figs. 2 y 3, Capítulo 3). La diferencia entre ARCs corona y ARCs troncales se observa incluso cuando no se considera extinción o cuando ésta es muy baja, sugiriendo que la discrepancia no puede ser explicada únicamente por la presencia de ramas troncales inusualmente largas generadas por el proceso de extinción, como había sido propuesto

(Pyron y Burbrink 2012). Es posible que el simple hecho de que las edades troncales sean más viejas que las edades corona pueda estar debilitando los estimados de ARCs troncales en comparación con los de ARCs corona. Cabe notar que esto no implica necesariamente que la extinción no tenga ningún efecto sobre las ARC. De hecho, los resultados de las simulaciones parecen mostrar que la extinción puede afectar tanto la ARC corona –haciéndola más positiva– como la ARC troncal –haciéndola más débil o negativa (Fig. 3, Capítulo 3). La diferencia entre ARCs corona y troncales se hace más pronunciada en presencia de extinción (Figs. 2 y 3, Capítulo 3), por lo que la relación de ambos tipos de ARC podría llevar alguna señal del proceso de extinción. Para usar la relación ARC corona/ARC troncal para cuantificar el proceso de extinción se tendría que optimizar el modelo de origen jerárquico de taxa superiores aquí propuesto para que los taxa simulados correspondieran con los observados.

Informatividad de la diferencia de la ARC entre rangos taxonómicos

Otro resultado relevante de este estudio es que las ARCs, tanto corona como troncales, varían con el rango taxonómico. La pendiente de ambos tipos de ARC disminuye hacia rangos taxonómicos menos inclusivos, pero las ARCs corona nunca llegan a ser negativas. En contraste, las ARCs troncales sí pueden ser negativas en rangos taxonómicos menos inclusivos. Esto sólo ocurre en modelos de tasas variables entre clados, tanto de tasas constantes en el tiempo como variables dependientes de la diversidad (Figs. 2b y 3b, Capítulo 3), indicando que una ARC troncal negativa es evidencia de que hay un proceso en el que las tasas varían entre clados, ya sea con diversidad ilimitada o con límites ecológicos a la diversidad.

La variación entre rangos taxonómicos en la ARC de datos empíricos es en general menos pronunciada que la obtenida con las simulaciones, ya que, en la mayoría de los casos, las ARCs mantienen el mismo signo entre rangos taxonómicos (Figs. 4 y 5, Capítulo 3). En los sistemas biológicos en los que hay evidencia de que las ARCs troncales son negativas (anfibios, mamíferos y plantas con flor), estas aparecen en todos los rangos taxonómicos considerados en cada grupo, incluyendo rangos que consideramos más inclusivos, como órdenes. Dado que el modelo de simulación implementado no está optimizado para corresponder exactamente con los rangos taxonómicos de los sistemas biológicos de estudio, una posible explicación a esto es que los clados que están asignados formalmente a un rango taxonómico en estos sistemas biológicos en realidad son equivalentes a los rangos menos inclusivos resultantes de las simulaciones. Los clados simulados de rangos más inclusivos podrían corresponder con clados que no están asignados formalmente a ningún rango taxonómico. Tomando un ejemplo de las plantas con flor, esto correspondería a clados determinados por la APG III como Magnólidas, Monocotiledóneas, y Rósidas y Astéridas dentro de Eudicotiledóneas, los cuales son clados más inclusivos que órdenes, pero no están asignados formalmente a ningún rango taxonómico (Bremer et al. 2009). Un siguiente paso de investigación sería estimar en diferentes sistemas biológicos las ARCs con este tipo de clados que no corresponden a rangos taxonómicos formales, para determinar si el modelo describe adecuadamente el comportamiento de estos datos.

Independientemente del signo de las correlaciones, en los sistemas biológicos en los que se pudieron estimar las ARCs desde el nivel de género hasta el rango taxonómico más inclusivo de cada grupo (en anfibios, reptiles escamados y aves), las correlaciones

parecen ser significativas más frecuentemente en rangos menos inclusivos, como en géneros y familias (Fig. 5, Capítulo 3). Una prueba estadísticamente significativa en rangos menos inclusivos puede ser explicado por el efecto del tamaño de muestra (Royall 1986). Ya que el número de clados en rangos menos inclusivos siempre es mayor que el de rangos más inclusivos, por azar, es más probable que los rangos más inclusivos resulten en relaciones no significativas, aún cuando el tiempo sí esté relacionado con la riqueza de especies a este nivel, ya sea negativa o positivamente.

Conclusiones sobre la informatividad de la ARC

A pesar de la variación entre rangos taxonómicos y entre edades corona y troncal, los resultados de las simulaciones muestran que cuando las tasas de diversificación son constantes en el tiempo y entre clados, el tiempo es un factor que siempre se relaciona positivamente con la variación en la diversidad, corroborando resultados de estudios previos (Rabosky 2009; Stadler et al. 2014). Además, los resultados aquí presentados extienden esta idea, mostrando que la relación entre la edad de los clados y la riqueza de especies también puede ser positiva en modelos de variación en las tasas y modelos de diversidad limitada (Figs. 2 y 3, Capítulo 3). Una ARC positiva se consideraba frecuentemente como evidencia de procesos en los que el tiempo es la principal explicación de los patrones de riqueza. Sin embargo, este resultado sugiere que la ARC no debe ser utilizada como evidencia del efecto del tiempo para la especiación (Stephens y Wiens 2003) ni como evidencia de procesos de diversidad ilimitada (Hedges et al. 2015). Además, plantea dos preguntas relevantes: i) ¿por qué las correlaciones siempre son positivas en edades corona? y ii) ¿por qué las ARCs siempre son positivas en rangos

más inclusivos? Dado que todos los procesos de diversificación tienen una primera fase de expansión, una posible respuesta a la primera pregunta es que el grupo corona siempre representa esta primera parte de los procesos, mientras que la edad troncal incluye otra parte de la historia de los clados. En el caso de los clados más inclusivos, es fácil imaginar cómo su diversidad es resultante de varios "subprocesos" de diversificación, cada uno con su propia tasa de diversificación, las cuales por azar pueden tomar valores similares y también valores completamente opuestos. A gran escala temporal, la suma del efecto de estas tasas contrastantes podría "anularse" o "enmascararse", como se detecta en el estudio de diversificación de las plantas con flor (Fig. 2, Capítulo 1) y en el de los ocotillos (Fig. 3, Capítulo 2) generando un patrón de diversidad semejante al producido por un proceso de tasas constantes. Esto también explicaría porqué estudios de diversificación que evalúan la distribución de la diversidad en grupos muy inclusivos, frecuentemente encuentran evidencia de constancia en los parámetros de diversificación (Van Valen 1973; McPeek y Brown 2007; Hedges et al. 2015).

Estudiar la relación entre la edad de los clados y su riqueza de especies en un contexto filogenético

Hay evidencia en la literatura de que los datos de riqueza de especies y edad presentan estructura filogenética la cual produce falsos positivos en las ARCs cuando no es tomada en cuenta, indicando que el tiempo está relacionado con la diversidad cuando en realidad no lo está (Rabosky et al. 2012). En este trabajo, los datos empíricos y simulados de riqueza de especies y de edad de los clados de diferentes rangos taxonómicos, así como los de edades corona y troncal, muestran diferentes grados de estructura filogenética y, en

la mayoría de los casos, no es detectada (Figs. S1 y S3, Capítulo 3). Aún cuando la estructura filogenética no fue simulada explícitamente en este trabajo, ésta es observada en los modelos de variación de las tasas entre clados (Fig. S1b, Capítulo 2). Además, la estructura filogenética aparece más frecuentemente en rangos taxonómicos más inclusivos y su influencia se va perdiendo en rangos menos inclusivos (Fig. S1b, Capítulo 2). Estos resultados sugieren que los modelos de nacimiento y muerte de tasas variables tienen el potencial de generar una señal filogenética en los datos de edad y riqueza de especies, por lo que un modelo de movimiento Browniano no sería adecuado para detectar dicha señal en este tipo de datos.

A pesar de la presencia de estructura filogenética, las correlaciones estimadas con el método de mínimos cuadrados filogenéticos (PGLS) y las estimadas con la prueba de Spearman que no toma en cuenta la señal filogenética son muy similares (Figs. 2 y 3, Capítulo 3). Además, cuando se supone movimiento Browniano con el método de PGLS (lambda (Λ) de Pagel (1999) =1; Fig. S4, Capítulo 3), las ARCs aparecen sobrestimadas en comparación a las obtenidas cuando se usa la estructura filogenética inferida de los datos (Λ estimada con ML; Fig. 2, Capítulo 3) y a las obtenidas con el método de Spearman que no considera estructura filogenética (Fig. 3, Capítulo 3). Esto muestra que el uso de un modelo de asociación de caracteres inadecuado para los datos de edad y riqueza de especies puede resultar en estimados de ARC erróneos, como se ha observado en otros caracteres (Price 1997). Así, los métodos que suponen movimiento Browniano, como los contrastes independientes filogenéticos (Felsenstein 1985; Pagel 1992), pueden resultar en falsos positivos al usarse para estimar las ARCs.

La estructura filogenética de los datos de edad y riqueza provenientes de estudios empíricos concuerda con los patrones obtenidos en las simulaciones (Fig. S3, Capítulo 3). En general, los rangos taxonómicos más inclusivos no presentan evidencia de estructura filogenética mientras que los rangos menos inclusivos muestran algún grado de señal filogenética, diferente de movimiento Browniano. Una excepción notable la constituyen los datos de edad corona de familias de anfibios, en los que se encontró evidencia de movimiento Browniano, y los reptiles escamados, en los que no se detectó estructura filogenética en ningún rango taxonómico (Fig. S3 y Tabla S4, Capítulo 3).

La similitud entre los patrones de estructura filogenética simulados y los observados refuerza la idea de que la edad y la riqueza de especies en la naturaleza no deben ser evaluados con modelos que suponen movimiento Browniano, y realzan la necesidad de investigar a fondo e incluso desarrollar nuevos métodos que sean adecuados para detectar estructura filogenética en datos de edad y riqueza de especies con el fin de tomarla en cuenta correctamente antes de evaluar la existencia de correlaciones. Mientras tanto, métodos que no corrigen estructura filogenética, como la prueba de Spearman y métodos que sí la toman en cuenta, como el estadístico Λ parecen ser de utilidad. En especial este último, ya que permite evaluar modelos de asociación de caracteres diferentes del modelo de movimiento Browniano y corregir la presencia de estructura filogenética con estos modelos alternativos.

La relación entre la edad de los clados y la riqueza de especies como complemento de métodos para reconstruir el proceso de diversificación

De acuerdo a los resultados de esta tesis, la ARC es un estadístico descriptivo menos informativo de lo que se esperaba, ya que sólo permite distinguir modelos de variación en las tasas entre clados. Estos modelos se relacionan conceptualmente con el modelo de innovaciones clave expuesto por Simpson (1953). Ambos modelos de tasas variables asumen que la invasión de un nuevo espacio adaptativo (el cambio en la tasa) depende de la interacción entre características intrínsecas de los organismos y factores extrínsecos del ambiente, es decir, que ocurre al azar. La manera en que estos nuevos espacios serán invadidos es sustancialmente diferente en ambos modelos explorados aquí. El modelo de tasas constantes indica que el llenado del nuevo espacio no tiene límites. El de denso-dependencia indica que el llenado del espacio adaptativo depende de características extrínsecas, ecológicas. Sin embargo, la ARC no permite distinguir entre esos dos procesos. Por lo tanto, para poder diferenciarlos se requerirá de la implementación de otros métodos y más estudios sobre los estadísticos descriptivos. El estadístico beta (Shao y Sokal 1998) –que describe la simetría de los árboles filogenéticos– y el estadístico gamma (Pybus y Harvey 2000) –que describe la temporalidad de los eventos de divergencia– han sido usados con éxito para revelar procesos ecológicos en datos simulados (Gascuel et al. 2015). Sin embargo, el estadístico gamma parece estar afectado por la presencia de extinción (Rabosky y Lovette 2008b) y por el nivel taxonómico de muestreo de los árboles (Cusimano y Renner 2010), por lo que evaluar el comportamiento de estos estadísticos bajo el modelo de origen jerárquico de taxa superiores para determinar su informatividad podría ser importante.

En este sentido, los métodos de ajuste de modelos pueden proveernos de información más detallada. Por ejemplo, en los ocotillos se observó que la tasa de diversificación está disminuyendo (Fig. 3, Capítulo 2), lo que podría indicar que el grupo ha llegado a un límite. En contraste, las angiospermas como un todo, no muestran evidencia de que hayan alcanzado un límite a la diversidad (Fig. 2, Capítulo 1). Aún con su gran nivel de detalle, estos métodos no siempre son lo suficientemente satisfactorios, principalmente porque no consideran todos los modelos posibles en la evaluación. Un avance importante sería desarrollar un marco general de análisis con máxima verosimilitud o estadística Bayesiana, que considere todos los modelos de diversificación posibles para permitir directamente su comparación. Por ejemplo, Matzke (2013, 2014) ha desarrollado recientemente un marco general de ML para evaluar modelos biogeográficos, mostrando que efectivamente el mecanismo de evento fundador el cual – si bien siempre se había considerado como un mecanismo fundamental para la especiación (Mayr 1954; Carson y Templeton 1984)– nunca se había evaluado formalmente con métodos de inferencia de modelos, es el que explica la especiación en una gran cantidad de clados isleños.

Un elemento igualmente valioso, que va asociado al uso de estadísticos descriptivos, son las simulaciones. Para la correcta implementación de estadísticos descriptivos del proceso de diversificación, las simulaciones son necesarias porque proporcionan rangos de valores esperados bajo diferentes modelos nulos, los cuales son la base para contrastar hipótesis alternativas de diversificación (Raup et al. 1973). Un método de simulación adecuado debe usar modelos en los que los elementos esenciales – de importancia para el fenómeno de estudio– estén bien representados (Wooley y Lin

2005). Durante muchos años se ignoró el efecto de la clasificación taxonómica en las ARCs, lo cual ha llevado a interpretaciones incorrectas de los datos, como se discute en el Capítulo 3. Así, en el contexto particular de este trabajo, es importante incluir en las simulaciones todos los elementos que afecten el comportamiento de los estadísticos descriptivos de interés para que la interpretación de resultados y las inferencias que se hacen a partir de ellos sea adecuada.

Perspectivas

Las simulaciones implementadas aquí permitieron llegar a conclusiones que pueden estar afectadas también por otros factores como la incertidumbre filogenética y los errores en el fechamiento, por lo que el comportamiento de la ARC en relación a estos factores debe de ser evaluado explícitamente en estudios futuros. Además, las diferencias en las distribuciones simuladas de la ARC entre los diferentes modelos no fue evaluada de manera formal. Aunque no es una práctica común en estudios de simulación, sería importante determinar si estas diferencias son estadísticamente significativas. En este sentido, una extensión clave sería poder comparar formalmente las ARCs empíricas con las simuladas, en un marco de simulaciones predictivas a posteriori (PAPS). Para esto, es necesario desarrollar una función de verosimilitud del modelo de origen jerárquico de taxa superiores, para poder estimar el parámetro de origen en diferentes grupos, y extenderlo para permitir variación en las tasas entre clados y denso-dependientemente.

Finalmente, sería relativamente fácil desarrollar un método para detectar estructura filogenética que usara un modelo acorde a la naturaleza de los datos de edad y riqueza de especies. Esto sería particularmente útil, ya que una práctica muy común es

evaluar el efecto de variables ambientales (e.g., área, temperatura) o intrínsecas de los organismos (e.g., tamaño corporal, dieta) sobre la diversidad usando métodos de correlación que asumen movimiento Browniano para corregir la estructura filogenética de los datos. Así, desarrollar un método que considere adecuadamente la estructura de varianza de los datos de riqueza sería un aporte significativo al área de investigación de la biología comparada.

CONCLUSIONES GENERALES

(1) En este estudio se muestra que las contradicciones en ARCs reportadas en estudios previos probablemente se deben, cuando menos en parte, a la falta de criterios homogéneos en la metodología de estimación de las ARC.

(2) Las ARCs deben ser usadas con cautela en los estudios de diversificación. Cuando sea posible, deben usarse varios estimados de ARC a diferentes niveles taxonómicos y analizar los patrones cuidadosamente.

(3) Una ARC nula o negativa no es evidencia de diversificación limitada, ya que procesos de diversificación ilimitada también generan ese tipo de ARCs. Una ARC nula o negativa es evidencia de variación de las tasas entre clados.

(4) Una ARC positiva no es evidencia de constancia de tasas, ya que procesos de cambio en las tasas entre clados y de variación dependiente de la diversidad también presentan relaciones positivas entre la edad y la riqueza de especies.

(5) Cuando la estructura filogenética es débil en los datos de edad y riqueza de especies, los métodos no filogenéticos para estimación de correlaciones pueden ser tan válidos y efectivos como el método de PGLS y aún mejores que métodos que asumen una estructura de movimiento Browniano en la evolución de los caracteres para estimar la ARC.

(6) Los métodos de simulación son una herramienta aliada en los estudios de diversificación, y en un escenario ideal deben de incluir todos los factores que potencialmente afecten los parámetros de estudio.

REFERENCIAS GENERALES

Agapow P., Isaac N. 2002. Macrocaicerge : revealing correlates of species richness by comparative analysis. Divers. Distrib. 8:41–43.

Alfaro M.E., Santini F., Brock C., Alamillo H., Dornburg A., Rabosky D.L., Carnevale G., Harmon L.J. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proc. Natl. Acad. Sci. U. S. A. 106:13410–4.

Benton M.J. 2009. The Red Queen and the Court Jester: Species Diversity and the Role of Biotic and Abiotic Factors Through Time. Science. 323:728–732.

Bokma F., Baek S.K., Minnhagen P. 2014. 50 Years of Inordinate Fondness. Syst. Biol. 63:251–256.

Bremer B., Bremer K., Chase M.W., Fay M.F., Reveal J.L., Soltis D.E., Soltis P.S., Stevens P.F., Anderberg A.A., Moore M.J. et al. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Bot. J. Linn. Soc. 161:105–121.

Carson H.L., Templeton A.R. 1984. Genetic Revolutions in Relation to Speciation Phenomena: The Founding of New Populations. Annu. Rev. Ecol. Syst. 15:97–131.

Condamine F., Rolland J., Morlon H. 2013. Macroevolutionary perspectives to environmental change. Ecol. Lett. 16:72–85.

Cornell H.V. 2013. Is regional species diversity bounded or unbounded? Biol. Rev. Camb. Philos. Soc. 88:140–65.

Cornette J.L., Lieberman B.S. 2004. Random walks in the history of life. Proc. Natl. Acad. Sci. U. S. A. 101:187–91.

Crawford F.W., Minin V.N., Suchard M.A. 2014. Estimation for General Birth-Death Processes. J. Am. Stat. Assoc. 109506:730–747.

Cusimano N., Renner S.S. 2010. Slowdowns in diversification rates from real phylogenies may not be real. Syst. Biol. 59:458–464.

Darwin C. 1859. On the origin of species by means of natural selection. London: Murray.

Etienne R.S., Haegeman B., Stadler T., Aze T., Pearson P.N., Purvis A., Phillimore A.B. 2012. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. Proc. Biol. Sci. 279:1300–9.

Felsenstein J. 1985. Phylogenies and the Comparative Method. Am. Nat. 125:1–15.

Fitzjohn R.G. 2010. Quantitative traits and diversification. Syst. Biol. 59:619–633.

Fitzjohn R.G., Maddison W.P., Otto S.P. 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. Syst. Biol. 58:595–611.

Foote M. 2007. Symmetric waxing and waning of marine invertebrate genera. Paleobiology. 33:517–529.

Foote M., Crampton J.S., Beu A.G., Marshall B.A., Cooper R.A., Maxwell P.A., Matcham I. 2007. Rise and fall of species occupancy in Cenozoic fossil mollusks. Science. 318:1131–1134.

Gascuel F., Ferrière R., Aguilée R., Lambert A. 2015. How Ecology and Landscape Dynamics Shape Phylogenetic Trees. Syst. Biol. 64:590–607.

Goldberg E.E., Lancaster L.T., Ree R.H. 2011. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. Syst. Biol. 60:451–465.

Gould S.J., Eldredge N. 1977. Paleontological Society Punctuated Equilibria : The Tempo and Mode of Evolution Reconsidered Punctuated equilibria : the tempo and mode of evolution reconsidered. Paleobiology. 3:115–151.

Gould S.J., Raup D.M., Sepkoski J.J., Schopf T.J.M., Simberloff D.S. 1977. The shape of evolution: a comparison of real and random clades. Paleobiology. 3:23–40.

Harvey P., May R., Nee S. 1994. Phylogenies Without Fossils. Evolution. 48:523–529.

Hedges S., Kumar S. 2009. The Timetree of Life. New York: Oxford University Press.

Hedges S., Marin J., Suleski M., Paymer M., Kumar S. 2015. Tree of life reveals clock-like speciation and diversification. Mol. Biol. Evol.:1–24.

Hey J. 1992. Using Phylogenetic Trees to Study Speciation and Extinction. Evolution. 46:627–640.

Humphreys A.M., Barraclough T.G. 2014. The evolutionary reality of higher taxa in mammals. Proc. R. Soc. B. 281:20132750.

Isaac N.J., Agapow P., Harvey P.H., Purvis A. 2003. Phylogenetically Nested Comparisons for Testing Correlates of Species Richness : A Simulation Study of Continuous Variables. Evolution. 57:18–26.

Jepsen G., Mayr E., Simpson G. 1949. Genetics, paleontology and evolution. London: Princeton University Press.

Jetz W., Thomas G., Joy J., Hartmann K., Mooers A. 2012. The global diversity of birds in space and time. Nature. 491:1–5.

Kendall D. 1948. On the Generalized “ Birth-and-Death ” Process. Ann. Math. Stat. 19:1–15.

Maddison W.P., Midford P.E., Otto S.P. 2007. Estimating a binary character’s effect on speciation and extinction. Syst. Biol. 56:701–10.

Magallón S., Sanderson M.J. 2001. Absolute diversification rates in angiosperm clades. Evolution. 55:1762–1780.

Magnuson-Ford K., Otto S.P. 2012. Linking the Investigations of Character Evolution and Species Diversification. Am. Nat. 180:225–245.

Maruvka Y.E., Shnerb N.M., Kessler D.A, Ricklefs R.E. 2013. Model for macroevolutionary dynamics. Proc. Natl. Acad. Sci. U. S. A. 110:E2460-9.

Matzke N.J. 2013. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model testing. Front. Biogeogr. 5:242–248.

Matzke N.J. 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. Syst. Biol. 63:951–970.

Mayr E. 1954. Change of genetic environment and evolution. In: Huxley J., Hardy A., Ford E., editors. London: Allen & Unwin. p. 157–180.

McPeek M.A, Brown J.J.M. 2007. Clade age and not diversification rate explains species richness among animal taxa. Am. Nat. 169:E97–E106.

Moran P. 1958. Random processes in genetics. Math. Proc. Cambridge Philos. Soc. 54:60–71.

Morlon H. 2014. Phylogenetic approaches for studying diversification. Ecol. Lett. 17:508–525.

Morlon H., Parsons T., Plotkin J. 2011. Reconciling molecular phylogenies with the fossil record. Proc. Natl. Acad. Sci. U. S. A. 108:16327–16332.

Morlon H., Potts M.D., Plotkin J.B. 2010. Inferring the dynamics of diversification: A coalescent approach. PLoS Biol. 8:e1000493.

Nee S. 2006. Birth-Death Models in Macroevolution. Annu. Rev. Ecol. Evol. Syst. 37:1–17.

Nee S., Holmes E.C., May R.M., Harvey P.H. 1994a. Extinction rates can be estimated from molecular phylogenies. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 344:77–82.

Nee S., May R.M., Harvey P.H. 1994b. The reconstructed evolutionary process. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 344:305–311.

Nee S., Mooers A., Harvey P. 1992. Tempo and mode of evolution revealed from molecular phylogenies. Proc. Natl. Acad. Sci. U. S. A. 48:523–529.

Pagel M. 1999. Inferring the historical patterns of biological evolution. Nature. 401:877–884.

Pagel M.D. 1992. A method for the analysis of comparative data. J. Theor. Biol. 156:431–442.

Price T. 1997. Correlated evolution and independent contrasts. Philos. Trans. R. Soc. B-Biological Sci. 352:519–529.

Pybus O.G., Harvey P.H. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. Proc. R. Soc. B. 267:2267–72.

Pyron R.A., Burbrink F.T. 2012. Extinction, ecological opportunity, and the origins of g lobal snake diversity. Evolution. 66:163–178.

Rabosky D.L. 2009. Ecological limits on clade diversification in higher taxa. Am. Nat. 173:662–674.

Rabosky D.L., Adams D.C. 2012. Rates of morphological evolution are correlated with species richness in salamanders. Evolution. 66:1807–18.

Rabosky D.L., Lovette I.J. 2008a. Density-dependent diversification in North American wood warblers. Proc. R. Soc. B. 275:2363–2371.

Rabosky D.L., Lovette I.J. 2008b. Explosive evolutionary radiations: decreasing speciation or increasing extinction through time? Evolution (N. Y). 62:1866–1875.

Rabosky D.L., Slater G.G.J., Alfaro M.E. 2012. Clade Age and Species Richness Are Decoupled Across the Eukaryotic Tree of Life. PLoS Biol. 10:e1001381.

Raup D.M., Gould S.J., Schopf T.J.M., Simberloff D.S. 1973. Stochastic Models of Phylogeny and the Evolution of Diversity. J. Geol. 81:525–542.

Ricklefs R.E., Renner S.S. 1994. Species Richness Within Families of Flowering Plants. Evolution. 48:1619–1636.

Rolland J., Condamine F.L., Jiguet F., Morlon H. 2014. Faster Speciation and Reduced Extinction in the Tropics Contribute to the Mammalian Latitudinal Diversity Gradient. PLoS Biol. 12.

Royall R.M. 1986. The Effect of Sample Size on the Meaning of Significance Tests. Am. Stat. 40:313–315.

Sepkoski J. 1978. A Kinetic Model of Phanerozoic Taxonomic Diversity I. Analysis of Marine Orders. Paleobiology. 4:223–251.

Sepkoski J.J. 1998. Rates of speciation in the fossil record. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 353:315–326.

Sepkoski J.J., Kendrick D.C. 1993. Numerical Experiments with Model Monophyletic and Paraphyletic Taxa. Palaeobiology. 19:168–184.

Shao K.T., Sokal R.R. 1990. Tree Balance. Syst. Biol. 39:266–276.

Sibley C.G., Ahlquist J.E. 1990. Phylogeny and Classification of Birds: A Study in Molecular Evolution. Yale University Press.

Simpson G. 1953. The Major Features of Evolution. New York, USA: Columbia University Press.

Stadler T. 2011. Mammalian phylogeny reveals recent diversification rate shifts. Proc. Natl. Acad. Sci. U. S. A. 108:6187–92.

Stadler T., Rabosky D.L., Ricklefs R.E., Bokma F. 2014. On Age and Species Richness of Higher Taxa. Am. Nat. 184:447–455.

Stanley S. 1975. A theory of evolution above the species level. Proc. Natl. Acad. Sci. U. S. A. 72:646–50.

Stanley S. 1985. Rates of evolution. Paleobiology. 11:13–26.

Stenseth N.C., Smith J.M. 1984. Coevolution in ecosystems: Red Queen evolution or stasis? Evolution. 38:870–880.

Stephens P.R., Wiens J.J. 2003. Explaining species richness from continents to communities: the time-for-speciation effect in emydid turtles. Am. Nat. 161:112–28.

Van Valen L. 1973. A new evolutionary law. Evol. Theory. 1:1–30.

Venditti C., Meade A., Pagel M. 2010. Phylogenies reveal new interpretation of speciation and the Red Queen. Nature. 463:349–352.

Wallace A.R. 1878. Tropical Nature and Other Essays. New York: McMillan Press.

Wiens J.J. 2011. The causes of species richness patterns across space, time, and clades and the role of “ecological limits”. Q. Rev. Biol. 86:75–96.

Wooley J., Lin H. 2005. Computational Modeling and Simulation as Enablers for Biological Discovery. In: Wooley J., Lin H., editors. Catalyzing Inquiry at the Interface of Computing and Biology. Washington D.C.: National Academies Press US. p. 117–205.

Yule G. 1925a. A mathematical theory of Evolution, Based on the Conclusions of Dr. J. C. Willis, F.R.S. Philos. Trans. R. Soc. London. Ser. B, Contain. Pap. a Biol. Character. 213:21–87.

APÉNDICE 1

UN ÁRBOL DE TIEMPO METACALIBRADO DOCUMENTA LA APARICIÓN TEMPRANA DE LA DIVERSIDAD FILOGENÉTICA DE LAS PLANTAS CON FLOR

Artículo publicado en New Phytologist

New Phytologist

Research

A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity

Susana Magallón[1], Sandra Gómez-Acevedo[1], Luna L. Sánchez-Reyes[1,2] and Tania Hernández-Hernández[3]

[1]Instituto de Biología, Universidad Nacional Autónoma de México, Mexico City, Mexico; [2]Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México, Mexico City, Mexico; [3]Departamento de Biología Evolutiva, Instituto de Ecología A.C., Xalapa, Veracruz, México

Author for correspondence:
Susana Magallón
Tel: +52 55 5622 9087
Email: s.magallon@ib.unam.mx

Received: *1 August 2014*
Accepted: *21 November 2014*

New Phytologist (2015)
doi: 10.1111/nph.13264

Key words: calibrations, constraints, diversification, fossil record, maximum age, radiations, relaxed molecular clocks.

Summary

- The establishment of modern terrestrial life is indissociable from angiosperm evolution. While available molecular clock estimates of angiosperm age range from the Paleozoic to the Late Cretaceous, the fossil record is consistent with angiosperm diversification in the Early Cretaceous.
- The time-frame of angiosperm evolution is here estimated using a sample representing 87% of families and sequences of five plastid and nuclear markers, implementing penalized likelihood and Bayesian relaxed clocks. A literature-based review of the palaeontological record yielded calibrations for 137 phylogenetic nodes. The angiosperm crown age was bound within a confidence interval calculated with a method that considers the fossil record of the group.
- An Early Cretaceous crown angiosperm age was estimated with high confidence. Magnoliidae, Monocotyledoneae and Eudicotyledoneae diversified synchronously 135–130 million yr ago (Ma); Pentapetalae is 126–121 Ma; and Rosidae (123–115 Ma) preceded Asteridae (119–110 Ma). Family stem ages are continuously distributed between *c.* 140 and 20 Ma.
- This time-frame documents an early phylogenetic proliferation that led to the establishment of major angiosperm lineages, and the origin of over half of extant families, in the Cretaceous. While substantial amounts of angiosperm morphological and functional diversity have deep evolutionary roots, extant species richness was probably acquired later.

Introduction

Life on Earth today is critically linked to flowering plants (angiosperms). Angiosperms are primary producers and fundamental structural components in modern terrestrial ecosystems, and contribute vast diversity in terms of species richness and functional innovations. Many biological lineages have flourished in association with angiosperms, or the biomes they created, depending on them for food, shelter, or symbiotic partnering (e.g. Wikström & Kenrick, 2001; Schneider *et al.*, 2004; McKenna *et al.*, 2009; Cardinal & Danforth, 2013). An understanding of the evolutionary establishment of modern terrestrial biomes is indissociable from angiosperm origin, diversification and rise to ecological predominance.

The timing of angiosperm diversification is among the classic questions in evolutionary biology (Friedman, 2009). Several elements to investigate it effectively have recently become available. Increasingly powerful relaxed clock methods have been incorporated into the general toolbox of modern phylogenetic biology (e.g. Baum & Smith, 2013). The ubiquitous application of relaxed clocks to all major branches of the tree of life (e.g. Hunt *et al.*, 2007; Hibbett & Matheny, 2009; Jetz *et al.*, 2012; Wahlberg *et al.*, 2013; Ericson *et al.*, 2014) has highlighted critical issues that affect the accuracy of age estimation. The crucial relevance of independent calibrations in relaxed clock analyses has been recognized (e.g. Aris-Brosou & Yang, 2003; Smith *et al.*, 2006; Yang & Rannala, 2006; Donoghue & Benton, 2007; Ho, 2007; Rannala & Yang, 2007; Wilkinson *et al.*, 2011), and also that relaxed clock models may insufficiently capture the degree of molecular variation in empirical phylogenies, leading to incorrect estimation of absolute rates and divergence times (e.g. Dornburg *et al.*, 2012; Wertheim *et al.*, 2012). Simultaneously, promising new avenues are being developed, for example, renewed implementations of local clocks (e.g. Drummond & Suchard, 2010; Ronquist *et al.*, 2012a), and highly parametric approaches to implement prior distributions (e.g. Heath, 2012).

On another front, significant fossil findings – including structurally preserved reproductive organs from Early Cretaceous sediments (Heimhofer *et al.*, 2007; Friis *et al.*, 2011) – document the minimum time of lineage origin and morphological evolution. These findings are especially relevant in the context of the increasingly solid molecular-based picture of angiosperm relationships at all phylogenetic levels (e.g. Soltis *et al.*, 2011), and the as yet few, but increasing numbers of studies explicitly investigating their phylogenetic relationships (e.g. Doyle & Endress, 2000, 2010; Magallón, 2007; Martínez-Millán *et al.*, 2009;

Sauquet *et al.*, 2012). Currently available methods allow one to estimate the phylogenetic position of fossils using maximum likelihood (ML) and Bayesian approaches, and to include fossils as terminals in phylogenetic dating analyses (Pyron, 2011; Ronquist *et al.*, 2012b).

Relaxed clock analyses applied to angiosperms have combined these methodological and palaeontological advances. However, most of them lack a firm temporal constraint on the onset of extant angiosperm diversification and, consequently, provide different estimates of the age of angiosperms as a whole (Fig. 1) and of the clades within them. Molecular clock estimates of crown angiosperm age range between 300 Ma or older (e.g. Ramshaw *et al.*, 1972; Brandl *et al.*, 1992; Magallón, 2010) and 86 Ma (Sanderson & Doyle, 2001), with many recent estimates lying

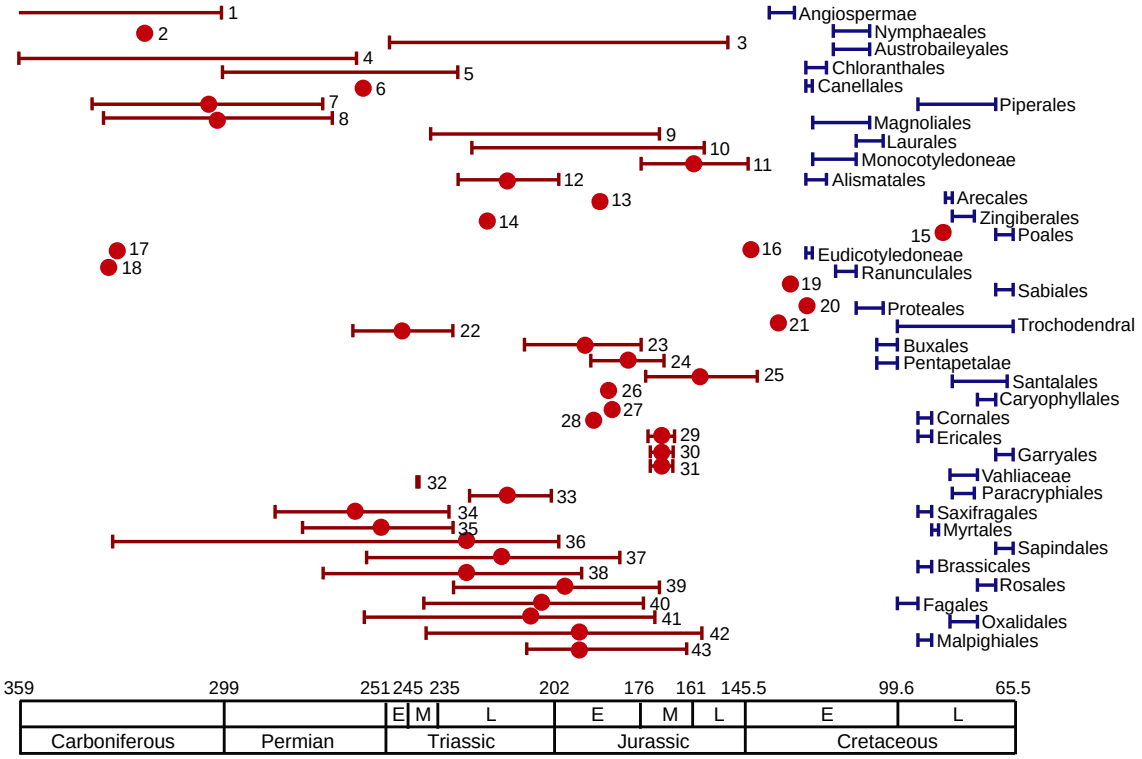


Fig. 1 Molecular and fossil-based estimates of angiosperm age. Blue bars indicate the stratigraphic interval from which the oldest fossil belonging to angiosperm orders and major clades is known, obtained by selecting the oldest representative fossil in Supporting Information Methods S1. Orders and major clades with oldest Tertiary fossils are not shown. The name of the order or major clade is shown next to each bar. Red dots and bars indicate the age and/or range of angiosperm crown age estimated in molecular clock studies. The numbers next to each red bar or dot correspond to estimates in published analyses indicated below. While the fossil record is consistent with an onset of angiosperm crown diversification in the Early Cretaceous, molecular estimates provide a generally older, but disparate picture of the age of the angiosperm crown group. 1, Ramshaw *et al.* (1972); 2, Martin *et al.* (1989); 3, Wolfe *et al.* (1989); 4, Brandl *et al.* (1992). Maize–wheat divergence calibration. 5, Brandl *et al.* (1992). Bryophyte–tracheophyte divergence calibration. 6, Brandl *et al.* (1992). Plant–animal divergence calibration. 7, Martin *et al.* (1993); method of Li & Graur (1991); 8, Martin *et al.* (1993); method of Li & Tanimura (1987); 9, Laroche *et al.* (1995). Maize–wheat divergence calibration. 10, Laroche *et al.* (1995). Vicieae–Phaseolinae divergence calibration. 11, Goremykin *et al.* (1997); 12, Sanderson (1997); 13, Sanderson & Doyle (2001). 18S. 14, Sanderson & Doyle (2001). *rbcL* 1st + 2nd. 15, Sanderson & Doyle (2001). *rbcL* 3rd. 16, Sanderson & Doyle (2001). *rbcL* all. 17, Soltis *et al.* (2002). Calibration 12: *Dicksonia*/*Plagiogyria*/*Cyathea*. 18, Soltis *et al.* (2002). Calibration 19: *Angiopteris*/*Marattia*. 19, Soltis *et al.* (2002). Calibration 25: gymnosperms. 20, Soltis *et al.* (2002). Calibration 29: lycopsids at 377.4. 21, Soltis *et al.* (2002). Calibration 19: lycopsids at 400. 22, Schneider *et al.* (2004); 23, Bell *et al.* (2005). MD with three minimum age constraints. 24, Bell *et al.* (2005). BLs estimated separately for each data partition. 25, Bell *et al.* (2005). Penalized likelihood (PL) with three minimum age constraints. 26, Magallón & Sanderson (2005). Four genes, 1st + 2nd. 27, Magallón & Sanderson (2005). Four genes, 3rd. 28, Magallón & Sanderson (2005). Four genes, all. 29, Moore *et al.* (2007). Unconstrained. 30. Moore *et al.* (2007). One hundred and twenty-five million years ago (Ma) minimum age to stem eudicots. 31, Moore *et al.* (2007). One hundred and twenty-five Ma minimum age to crown eudicots. 32, Magallón & Castillo (2009). Unconstrained. 33, Magallón (2010). PLBP. 34, Magallón (2010). PLFB. 35, Magallón (2010). MD. 36, Magallón (2010). Uncorrelated lognormal (UCLN). 37, Smith *et al.* (2010). With eudicot calibration. 38, Smith *et al.* (2010). Without eudicot calibration. 39, Clarke *et al.* (2011). Embryophyta at 509 Ma. 40, Clarke *et al.* (2011). Embryophyta at 1042 Ma. 41, Magallón *et al.* (2013). UCLN *atpB*, *psaA*, *psbB* and *rbcL*. 42, Magallón *et al.* (2013). UCLN *matK*. 43, Magallón *et al.* (2013). UCLN all genes.

between *c.* 190 and 150 Ma (e.g. Magallón, 2010; Smith *et al.*, 2010; Clarke *et al.*, 2011; Magallón *et al.*, 2013; Fig. 1).

The fossil record is consistent with an onset of angiosperm crown diversification in the Early Cretaceous, as shown by the increasing diversity and abundance of angiosperms in local outcrops and at a global level; increasing morphological complexity of leaves, pollen and flowers in agreement with expectations based on plant morphology; and congruence between the appearance of lineages in stratigraphic sequences and the branching order in molecular phylogenies (Doyle, 2012; Magallón, 2014). The fossil record is consistent with a proliferation of angiosperm lineages during the Cretaceous (145.5–65.5 million yr ago (Ma)), as indicated by the first stratigraphic occurrence of 35 orders and major clades during this period (Fig. 1; Supporting Information Methods S1). The observed distribution of the angiosperm fossil record as a whole may provide consistent landmarks to aid relaxed clock estimations, and for calculation of the age of the group as a whole.

The goal of this study is to estimate a time-frame of angiosperm evolution, including the origin and diversification of its major clades. It is based on a comprehensive sample of nearly 800 placeholders for 87% of angiosperm families, and the sequences of five molecular markers from the plastid and nuclear genomes. Divergence times were estimated with penalized likelihood and an uncorrelated Bayesian method. Two factors set this study apart from previous attempts. First, relaxed clock analyses incorporate a large number of fossil-derived age calibrations across the angiosperms, which provide landmarks for molecular rates and minimum ages across the tree. The phylogenetic assignment and age of each calibrating fossil are critically justified. Secondly, a confidence interval on the age of the angiosperm crown node, calculated with a method that considers the overall fossil record of the group (Marshall, 2008), was introduced in relaxed clock analyses. The resulting time-trees provide a solid framework for investigating angiosperm evolution. We discuss technical aspects, including the differential reliability of estimated family stem and crown ages; and biological implications, including the timing of angiosperm crown diversification and the origin of its major clades. The obtained time-frame documents profuse phylogenetic branching early in angiosperm history leading to the establishment of clades that contain major proportions of extant diversity, and, particularly, the early rise of many angiosperm families.

Materials and Methods

Taxonomic sample, molecular data and phylogenetic analyses

Phylogenetic, dating and diversification analyses were based on a data set including 792 angiosperm, six gymnosperm and one fern species. The gymnosperms belong to six families representing Cycadophyta, Ginkgophyta, Gnetophyta and Coniferae (*sensu* Cantino *et al.*, 2007). The angiosperms belong to 374 family-level clades corresponding to 87% of those recognized on the Angiosperm Phylogeny Website in April 2013 (Stevens, 2013). The represented families encompass 99% of angiosperm species. All angiosperm orders (plus four families unassigned at order level) are represented.

The data are the nucleotide sequences of three protein-coding plastid genes (*atpB*, *rbcL* and *matK*), and two nuclear markers (18S and 26S nuclear ribosomal DNA). Sequences were obtained through searches in GenBank, aiming to represent all angiosperm families with a complete sample of the five molecular markers. As first choice, species for which the five markers were available were selected. When one or more markers were unavailable, the missing markers were sampled from different species within the same genus. Markers that were unavailable at the genus level were left as missing data. Sequences of each marker were aligned with MUSCLE v3.7 (Edgar, 2004) followed by manual refinements using BioEdit v7.0.9.0 (Hall, 1999). The sampled species, represented families and orders, and GenBank accessions are listed in Table S1. The molecular data set is available in the DRYAD Digital Repository (http://doi.org/10.5061/dryad.k4227).

Phylogenetic analyses were conducted with maximum likelihood (ML) using RAxML v7.2.8. (Stamatakis, 2006a), implementing a GTRCAT model with 1000 bootstrap replicates (Stamatakis, 2006b, 2008). Substitution parameters were estimated independently for four data partitions: first plus second codon positions of *atpB* plus *rbcL*; third codon positions of *atpB* plus *rbcL*; *matK*; and 18S plus 26S. The fern *Ophioglossum* was specified as the outgroup. A topological constraint specifying relationships among major angiosperm clades according to a recent angiosperm-wide phylogenetic analysis based on a larger molecular data set (Soltis *et al.*, 2011) was implemented. Subsequently, a RAxML analysis to obtain 100 bootstrap trees with optimized branch lengths and model parameters was conducted with the same data partitions and constraints as above, but implementing the GTRGAMMA model.

Fossil calibrations

To obtain temporal calibrations for relaxed clock analyses, we conducted an intensive literature-based search of angiosperm fossils. Several search approaches were combined, including reviewing the palaeobotanical articles published during the last 20 yr, and extracting records from palaeofloristic works (e.g. Collinson, 1983; Knobloch & Mai, 1986) or from palaeobotanical compilations (e.g. Manchester, 1999; Martínez-Millán, 2010; Friis *et al.*, 2011) from which primary references were traced. We assembled a large data set in which fossils were grouped by their presumed affinity with angiosperm family-level or order-level clades, by combining the taxonomic assignment of the authors of the palaeobotanical description and our current knowledge of angiosperm phylogenetic relationships (Stevens, 2013). This data set includes over 3500 entries.

From this large data set, we selected fossils that could reliably represent the oldest record of angiosperm clades down to family (or lower) level. Accurate clade recognition was favoured over older ages (Sauquet *et al.*, 2012) by selecting only those fossils that could be identified as members of a clade with certainty. We gave preference to fossils of flowers, fruits and seeds over vegetative remains or pollen. In several instances we relegated the

putative oldest fossil of a clade because it was a vegetative structure of insecure affinity in favour of the oldest reproductive structure of the same clade. We also preferred whole-plant reconstructions and structurally preserved fossils over compression/impression remains.

To identify the relationship of selected fossils with the taxa in the tree and postulate calibration nodes, we combined intuitive, apomorphy-based and phylogenetic approaches. We relied as much as possible on the description and illustrations, and on the discussion of their taxonomic assignment in original publications. When available, we gave preference to a phylogenetic result (e.g. Doyle & Endress, 2010) over intuitive or apomorphy-based assignments. The absolute age assigned to each fossil was equal to the uppermost boundary of the narrowest stratigraphic interval to which the fossil could be assigned. Ages of stratigraphic boundaries were obtained from Walker & Geissman (2009). The assignment of each fossil to a particular node, as well as the absolute minimum age calibration, are discussed in detail in Methods S1.

Confidence interval on angiosperm age

A confidence interval that contains the true age of the angiosperm crown node was calculated with a method that considers the number of branches in a phylogenetic tree that are represented in the fossil record (Marshall, 2008). This method derives from quantitative palaeobiology approaches to calculate confidence intervals that contain the true time of origin and extinction of lineages based on local or global stratigraphic sequences (Strauss & Sadler, 1989; Marshall, 1990, 1994, 1997). The goal of Marshall's (2008) method is to date a molecular phylogenetic tree by using an absolute time-scale extrapolated from a confidence interval that contains the true age of the lineage in the tree that has the most temporally complete fossil record (i.e. the calibration lineage). In brief, the method has three components. The first is to identify the calibration lineage, which is achieved by finding the branch with the greatest overlap between the age of its oldest fossil and its node-to-tip length in an ultrametric tree estimated without any reference to the fossil record. The second step is to calculate a confidence interval that contains the true age of the calibration lineage. This calculation only requires the number of branches in the tree that are represented in the fossil record, the average number of fossil localities from which each branch represented in the fossil record is known, and the age of the oldest fossil of the lineage. The third step is to date the ultrametric tree by directly transforming its branch lengths into time by using the confidence interval of the calibration lineage as an absolute time-scale.

Because our aim is to calculate a confidence interval that contains the true age of crown angiosperms, we only implemented the second step of Marshall's (2008) method in the angiosperms as a whole. In a collateral study, described in Methods S2, we conducted step 1 and identified the calibration lineage of the angiosperms. This application of the method is justifiable because our intention is to estimate the maximal age of angiosperms, and not to temporally calibrate an ultrametric tree (C. R. Marshall, pers. comm.). The minimum (i.e. youngest) age of the confidence interval is directly given by the oldest known fossil(s) of the lineage. We consider angiosperm pollen grains from Valanginian to Hauterivian sediments (Early Cretaceous; Hughes & McDougall, 1987; Hughes *et al.*, 1991; Brenner, 1996) as the oldest fossils of the angiosperm crown group on the basis of their morphological and ultrastructural attributes; the increasing abundance, diversity and geographical distribution of angiosperm fossils starting in immediately younger sediments; and the congruence in morphological evolution and sequence of lineage appearance between the fossil record and expectations derived from morphological studies and molecular phylogenies, respectively (Methods S1). We used the Valanginian–Hauterivian boundary, corresponding to 136 Ma (Walker & Geissman, 2009), as the minimum age.

The maximum (i.e. oldest) age of the confidence interval is calculated with eqn 14 from Marshall (2008):

$$FA_c = \frac{FA_{cal}}{\sqrt[nH]{1 - C}}.$$

FA_c is the maximum age of the confidence interval; FA_{cal} is the age of the oldest fossil of the lineage (corresponding to the minimum age of the interval), in this case, 136 Ma; n is the number of branches in the phylogenetic tree represented in the fossil record; H is the average number of fossil localities from which each branch represented in the fossil record is known; and C is the desired confidence level associated with the interval. To calculate n, we used a tree in which each angiosperm family included in the main phylogenetic and dating analyses was represented by a single terminal (Methods S2). Based on the fossils used to calibrate internal nodes in the main dating analyses (Methods S1), we counted the number of branches in the family-level angiosperm tree represented in the fossil record. The average number of fossil localities from which each branch in the tree represented in the fossil record is known (H) is difficult to calculate. We chose to consider $H = 1$, implying that each branch with a fossil record is known from a single locality. Because the maximum age of the confidence interval decreases as H increases (Marshall, 2008), assuming $H = 1$ is a conservative approach that will bias the maximum age of angiosperms towards older ages. FA_c was calculated with confidence levels (C) of 0.5, 0.95 and 0.99. The calculated confidence interval was then implemented as a constraint in relaxed clock analyses.

Relaxed clock analyses

Dating analyses were conducted with two relaxed clock methods: penalized likelihood (PL; Sanderson, 2002) and the uncorrelated lognormal (UCLN) Bayesian method available in BEAST (Drummond *et al.*, 2006). Penalized likelihood analyses were conducted combining the softwares r8s (Sanderson, 1997, 2004) and TREE-PL (Smith & O'Meara, 2012). PL analyses were based on the ML phylogram obtained with RAxML after excluding the

outgroup (*Ophioglossum*); hence the seed plant crown node became the new root. To identify the appropriate level of rate heterogeneity in the phylogram, a data-driven cross-validation was conducted with treePL. The cross-validation tested nine smoothing values (λ) separated by one order of magnitude, starting at 1×10^{-7}. The age of the root was fixed at 330 Ma, based on previous estimates for crown seed plants (Magallón *et al.*, 2013). The angiosperm crown node was bracketed between 136 and 140 Ma (see the section 'Confidence interval on angiosperm age'), and 136 nodes within angiosperms were constrained with fossil-derived minimum ages (see the Materials and Methods section, and Methods S1).

Penalized likelihood age estimation was conducted with treePL and with r8s on the ML phylogram. The identified optimal smoothing value, the root node calibration, the bracket on crown angiosperm age (136–139.35 Ma in r8s; see the Results section), and the 136 minimum age constraints were implemented as indicated above. One hundred ML bootstrap phylograms were also dated with r8s, using the optimal smoothing magnitude identified with treePL. Age statistics of internal nodes were summarized with TreeAnnotator v1.7.5 (Drummond *et al.*, 2006).

Bayesian age estimation was conducted with the UCLN model in beast v1.7.5 (Drummond *et al.*, 2006). The data were the nucleotide sequences of the five molecular markers used in phylogeny estimation concatenated in a single alignment, including only seed plants. Data were partitioned into plastid (*atpB*, *rbcL* and *matK*) and nuclear components (18S and 26S nrDNA). Nucleotide substitution was under a GTR+I+Γ model, allowing independent estimation of parameters for each partition. Independent uncorrelated relaxed clock models were allowed between partitions, and the tree prior was under a Birth-Death model. The root was calibrated with a uniform distribution between 314 and 350 Ma, corresponding to the credibility interval (95% highest posterior density (HPD)) of the age of this node estimated in an independent study (Magallón *et al.*, 2013). The angiosperm root node was calibrated with a uniform distribution between 136 and 139.35 Ma (see the Results section). The prior ages of 136 nodes within angiosperms were obtained from lognormal distributions with mean equal to the fossil age plus 10%, to place the bulk of the distribution at ages older than the fossil, and a standard deviation of 1 (Methods S1). We considered assigning different standard deviation magnitudes depending on our confidence on each calibration, but, because we were unable to rigorously quantify our perception, we decided to assign the same magnitude to all. The chronogram obtained in the r8s analysis on the ML tree was used as a starting tree, and estimators of tree topology were unselected. Eight independent Markov Chain Monte Carlo (MCMC) runs of different lengths, but under the same estimation conditions, were conducted, for a total of 170×10^6 generations. Each MCMC was sampled every 5000 steps. The initial 600 trees sampled in each run were removed as burn-in and, in all cases, the post-burn-in trees were in the stable part of the chain. Analyses were conducted in the CIPRES Science Gateway (Miller *et al.*, 2010). Log outputs of the beast analyses were jointly evaluated with tracer v1.5 (Rambaut & Drummond, 2009). Effective sample sizes of estimated parameters were in most cases > 200, and always > 100. As a consequence of usage restrictions in CIPRES, we were unable to conduct further beast analyses. Files containing the sampled trees of each MCMC run were combined using LogCombiner v1.7.5, annotated using TreeAnnotator v1.7.5 (Drummond *et al.*, 2006), and visualized using FigTree v1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/).

Results

Molecular data set and phylogenetic tree

The concatenated alignment of nucleotide sequences of the five molecular markers (cp *atpB*, *rbcL* and *matK*; nu 18S and 26S) is 9089 base pairs long. From the total number of five markers for 799 taxa, 432 are missing; hence the molecular data set is 89% complete. The ML tree is consistent with current understanding of angiosperm relationships and most branches are supported with bootstrap values ≥95%. Phylogenetic relationships are shown in Figs 2 and 3.

Fossil calibrations

Based on the large data set of angiosperm fossils, 151 fossils that can reliably indicate the oldest occurrence of angiosperm clades down to family (sometimes lower) level were selected. Because of sister group relationships or sampling density, 20 nodes were each calibrated by two fossils; and two nodes by three fossils of the same age. In total, 136 internal nodes were calibrated with fossil-derived minimum ages (calibrations 2–137; Methods S1). An additional record, corresponding to the oldest remains of angiosperms, was used as the minimum bound of the confidence interval on crown angiosperm age (calibration 1; Methods S1). Of the selected fossils, eight provide calibrations for genera, seven for intrafamilial clades, 112 for families, seven for clades between families and orders, 12 for orders, and five for clades above the order level (Stevens, 2013). Twenty calibrations were assigned to nodes based on phylogenetic results, and the rest were placed intuitively or based on apomorphies. Except in one case (Martínez-Millán *et al.*, 2009), the phylogenetic position of fossils matched previous intuitive assignments. Detailed discussions of the 137 fossil-based calibrations, including formal names, authors and references, stratigraphic ranges or radiometric dates, justification of node assignment, and absolute age, are provided in Methods S1.

Confidence interval on angiosperm age

The confidence interval of angiosperm crown age was calculated considering the fossil record of the entire group (Marshall, 2008) in the context of a family-level tree. The number of branches represented in the fossil record (n) is 123. The minimum bound of the confidence interval (FA_{cal}) is given by the age of the oldest fossil of the clade, that is, 136 Ma (Hughes

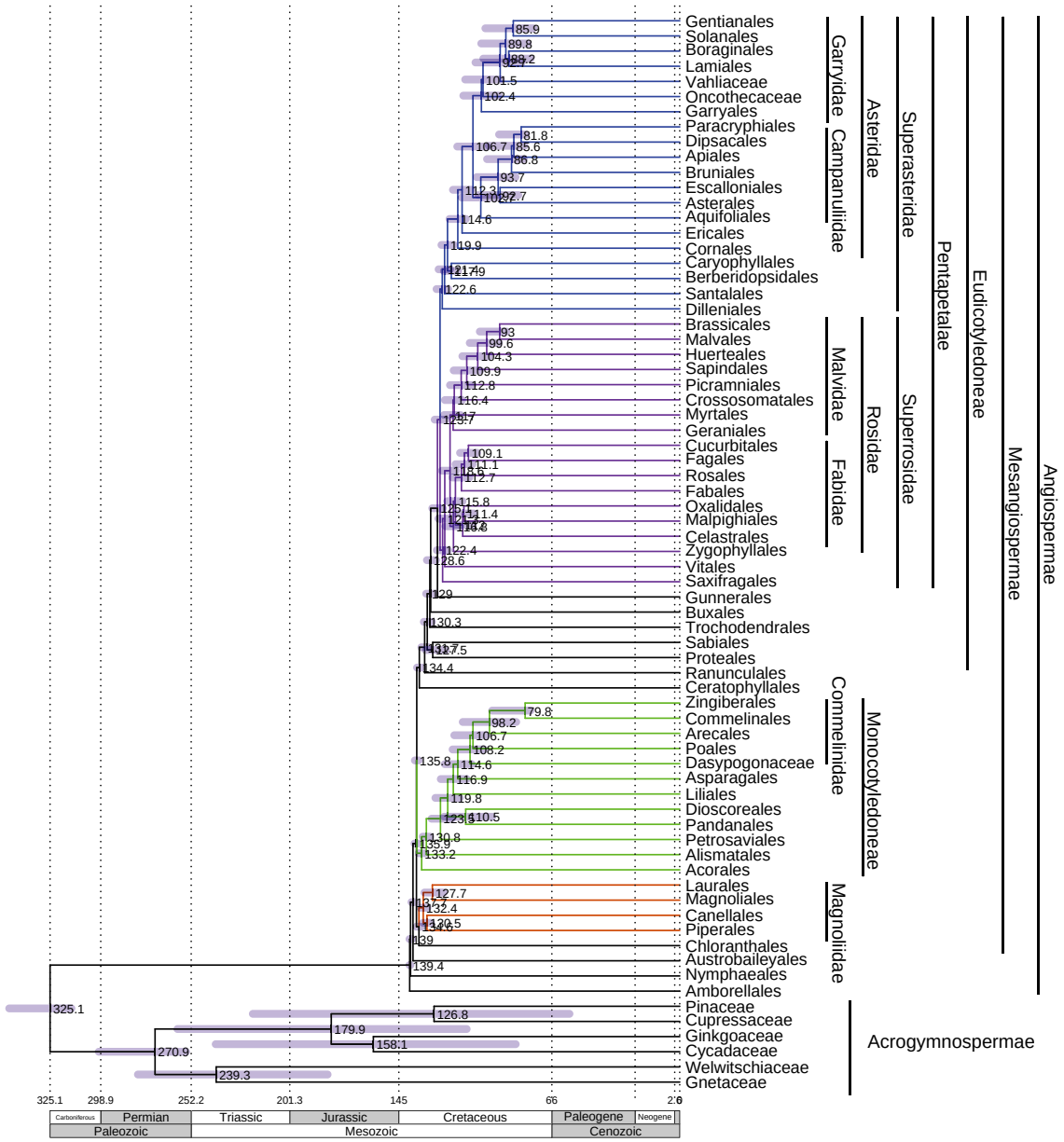


Fig. 2 Angiosperm time-tree estimated using the uncorrelated lognormal method in BEAST, with terminals collapsed to represent orders. Numbers next to nodes indicate the median age, and blue bars correspond to the 95% highest posterior density (HPD).

& McDougall, 1987; Hughes *et al.*, 1991; Brenner, 1996; Methods S1). The maximum bound of the confidence interval (FA_c), calculated with confidence levels (C) of 0.5, 0.95 and 0.99, is 136.77, 139.35 and 141.19 Ma, respectively. Hence, the 95% confidence interval of angiosperm age is between 136 and 139.35 Ma.

Relaxed clock analyses

Stem and crown ages of 14 major angiosperm clades and 62 orders, and stem ages of 374 families are provided in Table S2. Fig. 2 shows the UNCL time-tree at the level of angiosperm orders, and Fig. 3(a–e) shows the family-level UNCL angiosperm

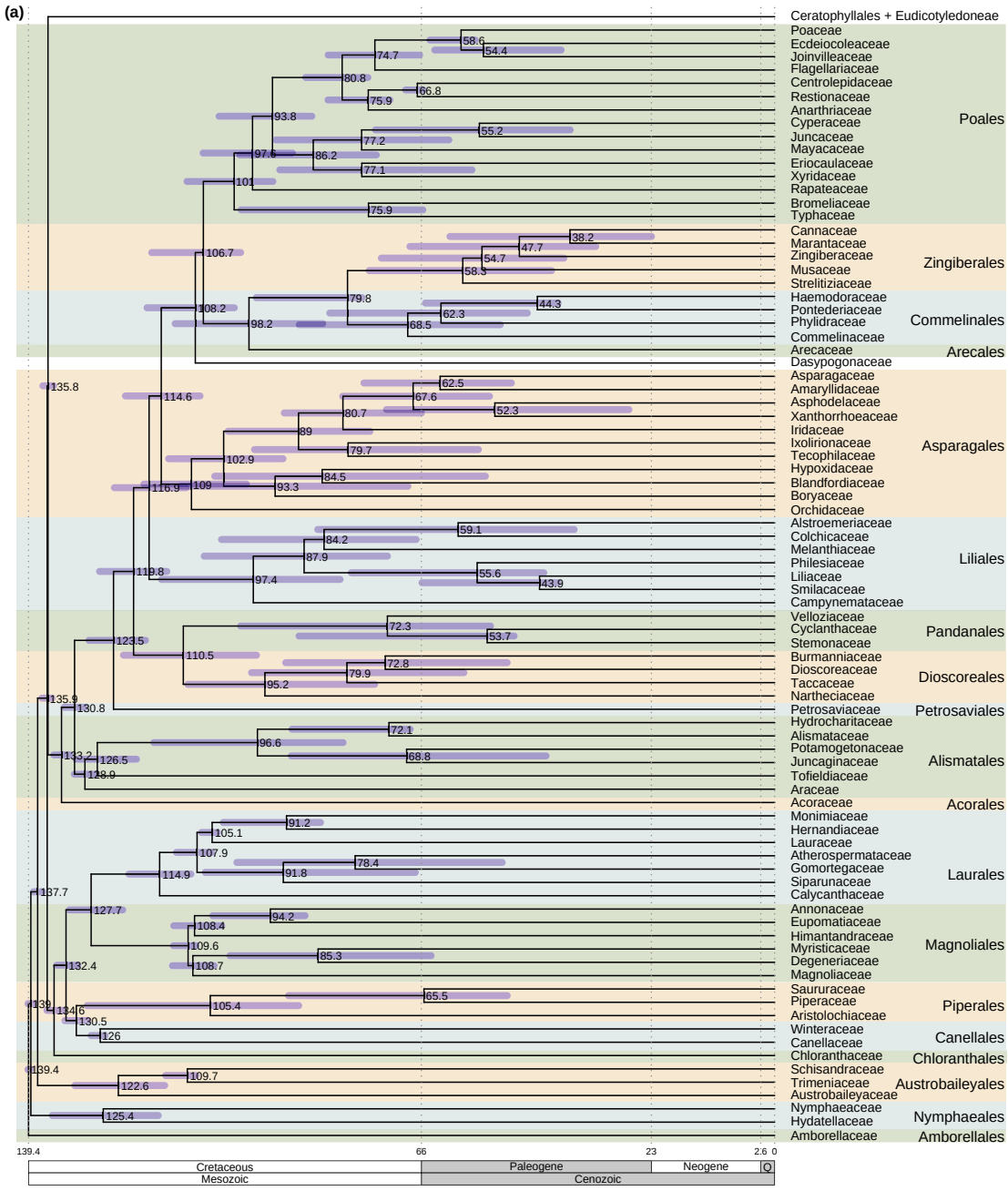


Fig. 3 Angiosperm time-tree estimated using the uncorrelated lognormal method in BEAST, with terminals collapsed to represent families. Numbers next to nodes indicate the median age, and blue bars correspond to the 95% highest posterior density (HPD). (a) Amborellales to Poales (Monocotyledoneae). (b) Ceratophyllales to Ericaceae (Asteridae, Eudicotyledoneae). (c) Garryidae to Campanulidae (Asteridae, Eudicotyledoneae). (d) Saxifragales to Brassicales (Malvidae, Rosidae, Eudicotyledoneae). (e) Zygophyllales to Malpighiales (Fabidae, Rosidae, Eudicotyledoneae). Ages of nodes are provided in Supporting Information Table S2.

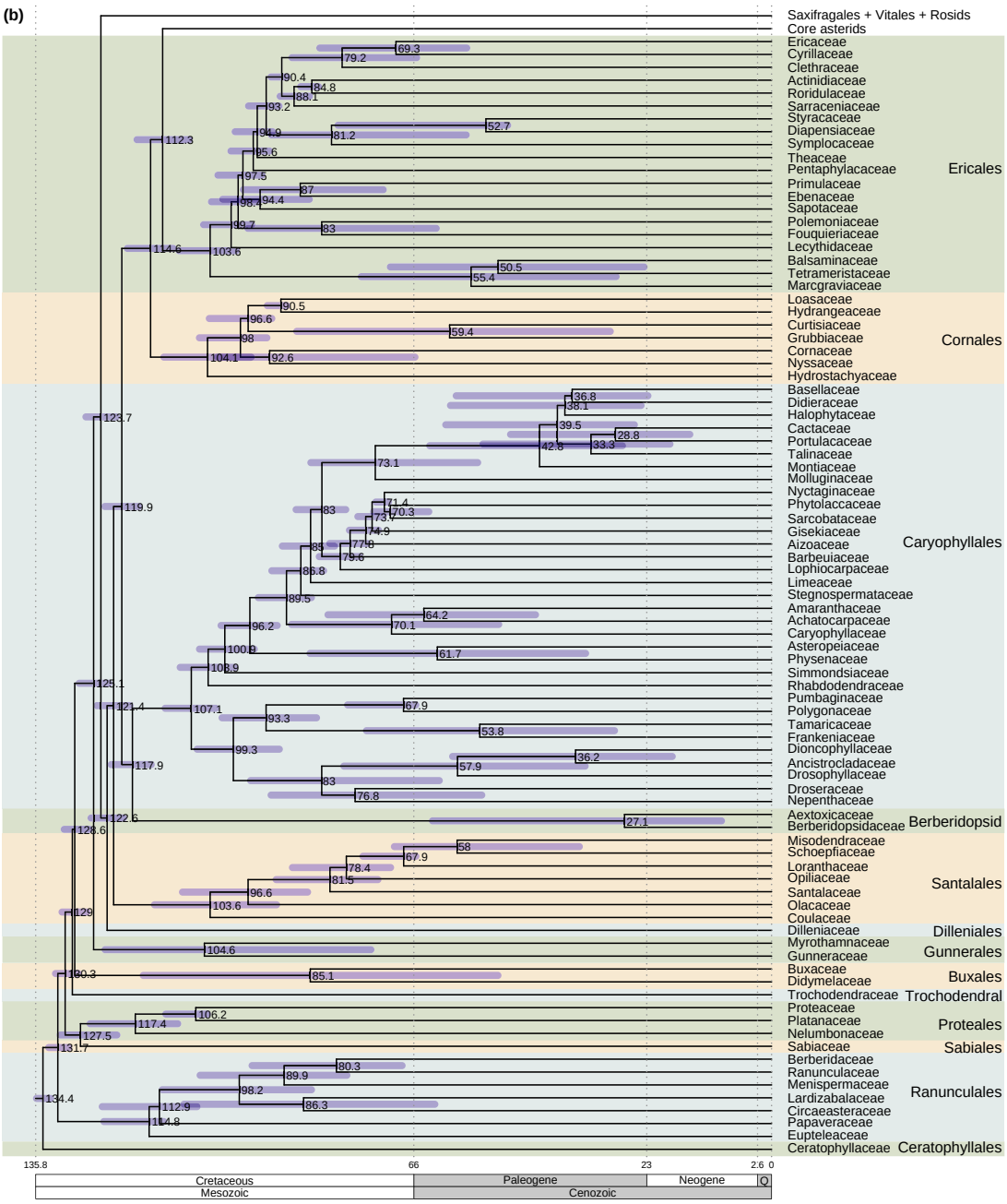


Fig. 3 Continued

time-tree. The full dated trees in NEXUS format are available in the DRYAD Digital Repository (http://doi.org/10.5061/dryad.k4227). Ages estimated with PL tend to be older than those estimated with the UCLN method (Table S2). As expected, associated errors on ages are narrower in the PL time-tree (i.e. average range of values in bootstrap ML trees = 9.25 Ma) than in the UNCL time-tree (i.e. average magnitude of 95% HPD = 27.03 Ma).

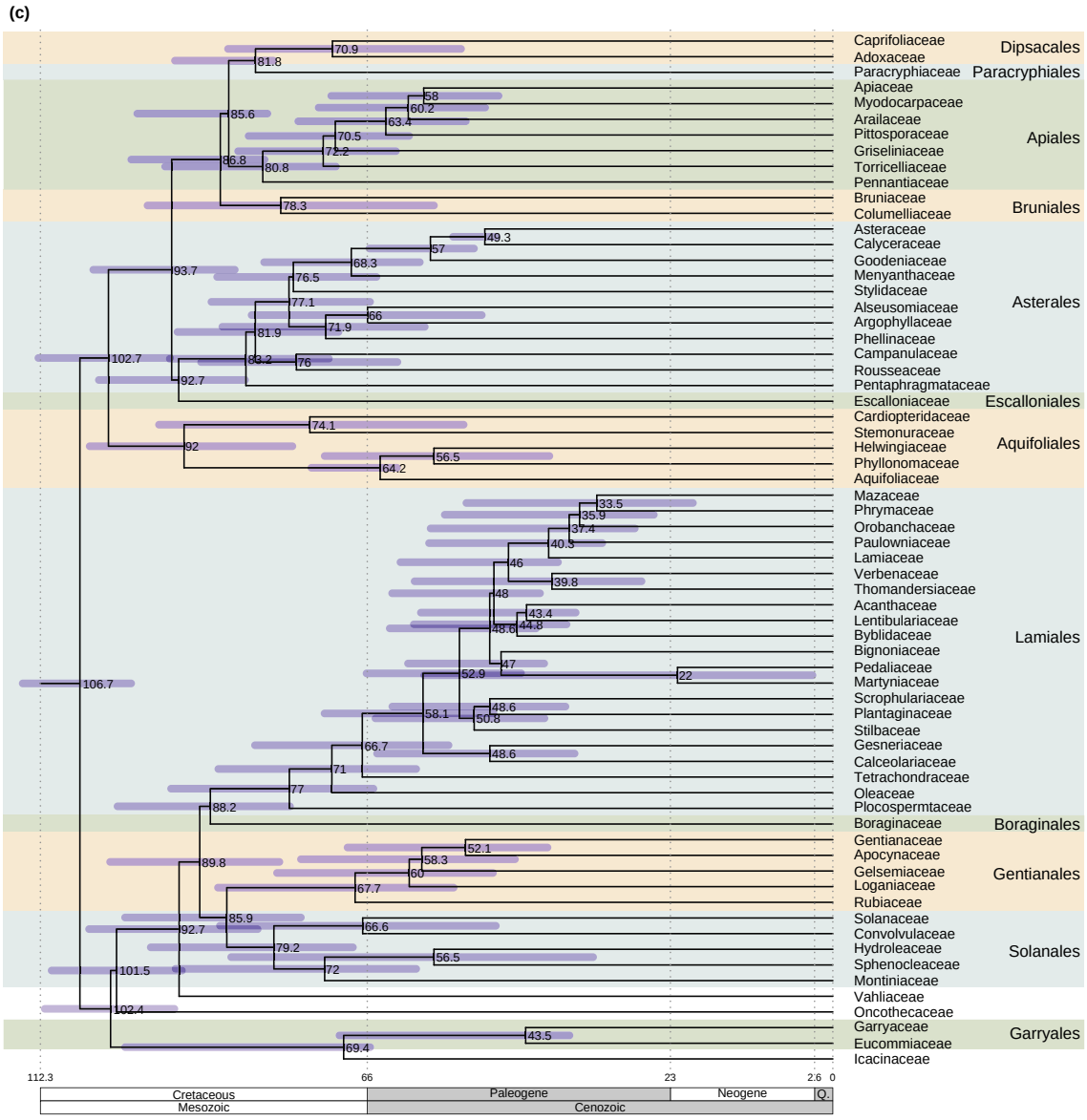


Fig. 3 Continued

Discussion

Estimated dates

The dates obtained depend on the correctness of relationships and branch lengths estimated in phylogenetic and dating analyses. Estimates of some family crown ages are probably too young for the following reasons. (1) The taxonomic sample may not represent the crown node of each family. The crown age of a clade can only be estimated if at least one member of each of the two sister branches derived from the deepest phylogenetic split in the clade is included. Although we aimed to sample representatives from both sides of the deepest split of each family, this was not consistently achieved because of insufficient knowledge of intrafamilial phylogenetic relationships, or unavailability of molecular markers for the required taxa. (2) Fossils selected as calibrations may not be the oldest fossil members of a clade. To calibrate, we selected fossils with greater chances of correctly

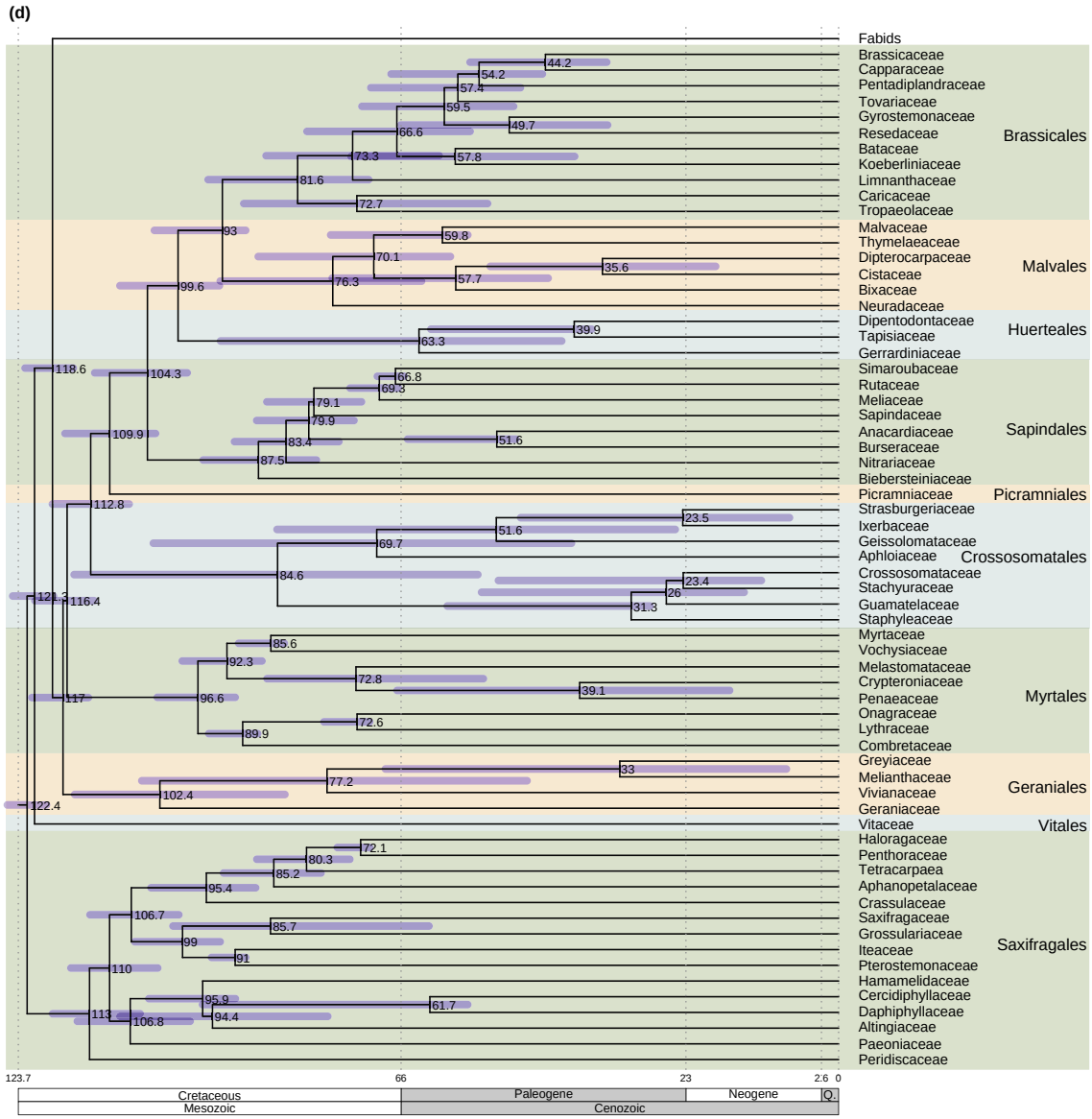


Fig. 3 Continued

reflecting clade membership over fossils whose membership in the clade is equivocal, thus favouring 'safe but late' over 'early but risky' fossils (Sauquet *et al.*, 2012). Consequently, the minimum ages provided by some calibrations may be too young. (3) The assignment of calibrations to nodes on the tree was done conservatively. Most assignments were based on the presence of an apomorphy, or intuitive criteria (Sauquet *et al.*, 2012). In these cases, a fossil was assigned to a more inclusive clade that could securely contain it (e.g. a stem group) than to a less inclusive clade where its membership was dubious (e.g. a crown group). Consequently, minimum ages that are too young (thus uninformative) might have been applied to stem nodes, rather than minimum ages that are too old (thus incorrect and misguiding) to crown nodes. The taxonomic sample in this study represents angiosperm major clades and their relationships; hence the factors mentioned above are unlikely to affect age estimation at deeper phylogenetic levels.

Because previous molecular estimates have provided disparate estimates of angiosperm age (Fig. 1), we implemented a

(e)

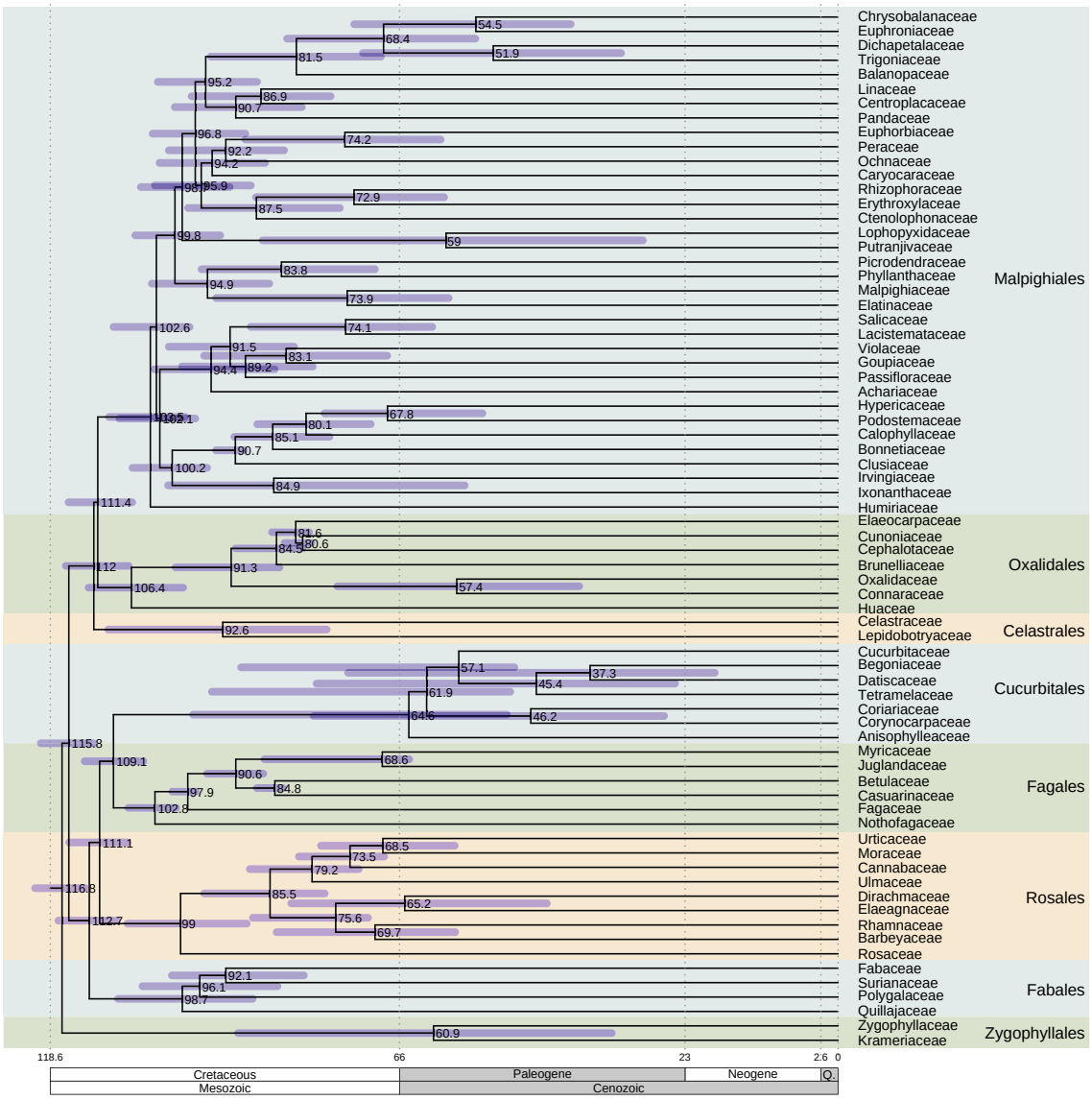


Fig. 3 Continued

confidence interval on the age of the angiosperm crown node. The minimum age of the interval can be obtained from the oldest fossils of the group, but the maximum age is difficult to obtain. Marshall's (2008) method allows calculation of the maximum age by quantifying the density of the fossil record of the group as a whole. Marshall's method has one caveat – reliance on a single calibration lineage, albeit selected from the complete fossil record of the group – and relies on several assumptions: that the affinity and absolute age of fossils are known with certainty; that the topology and relative branch lengths of the uncalibrated, ultrametric tree are accurate; that fossilization is random; and that the value of H (eqn 14) is close to the true average number of fossil localities from which each lineage represented in the fossil record is known (Marshall, 2008). However, only the last two assumptions are relevant to calculating the confidence interval around the true age of a lineage.

Nonrandom fossilization and, specifically, a lower fossilization probability early in the history of a lineage may underestimate its true time of origin. However, only if all the lineages within the group suffer similarly from a long period of dramatically decreased initial preservation potential will the method fail to bracket the true time of origin (Marshall, 2008). We considered reticulate semitectate pollen grains with columellate sexine from the Valanginian–Hauterivian (Hughes & McDougall, 1987; Brenner, 1996) as belonging, or being close, to the angiosperm crown group, based on their morphological attributes (Doyle, 2012) and, importantly, because their stratigraphic appearance is followed by a continuous and increasingly dense and diverse angiosperm record in immediately younger sediments (Fig. 1). Nevertheless, the possibility of a cryptic pre-Cretaceous angiosperm history has been discussed (e.g. Axelrod, 1952, 1970). Specifically, angiosperm-like pollen grains similar to those from Early Cretaceous sediments have been reported from the Middle Triassic Germanic Basin and other pre-Cretaceous localities (Hochuli & Feist-Burkhardt, 2013). These Triassic pollen grains resemble the *Retimonocolpites* morphotypes, known from late Hauterivian and younger sediments (Phases 1–4 of Hughes, 1994), and not the *Clavatipollenites* morphotypes, to which some of the early Hauterivian angiosperm pollen grains belong (Phase 0 of Hughes, 1994; Hochuli & Feist-Burkhardt, 2013). We agree with Hochuli & Feist-Burkhardt (2013, p. 11) in interpreting the pre-Cretaceous angiosperm-like pollen grains as possible angiosperm stem relatives, based on morphological differences from the earliest Early Cretaceous pollen grains, and a 100 Myr gap in the fossil record before the angiosperm radiation.

If the true average number of fossil localities from which each lineage with a fossil record is known is much larger than the value of H (Marshall, 2008), the calculated maximum bound of the confidence interval will substantially overestimate the true age of the lineage. We considered $H = 1$, which, in the case of angiosperms, is a strong underestimate. Any value $H > 1$ would have resulted in a younger maximum bound of the confidence interval of angiosperm age. The angiosperm age interval was therefore calculated under two biases with opposite effects. It is unknown if the potential underestimation caused by nonrandom fossilization and the overestimation caused by assuming $H = 1$ cancel each other.

We emphasize that the relaxed clock-estimated ages are strongly contingent on the confidence interval placed on the age of the angiosperm crown node. Preliminary analyses conducted under the same conditions as described in the UCLN analysis above, but excluding the confidence interval on the angiosperm crown node, estimated a substantially older age for the angiosperm crown node (219.9 Ma; 160.0–255.8 95% HPD; results available from the authors). Interestingly, internal nodes were not much older than when the confidence interval was used.

Origin of major angiosperm clades

The estimated time-trees indicate the timing of phylogenetic branching that gave rise to major angiosperm clades, and provide a reliable basis for understanding the onset and dynamics of accumulation of different components of extant angiosperm diversity. PL-estimated ages are typically older than UCLN ages, but their associated intervals are overlapping, and provide congruent timeframes of the stem ages of families. The following discussion is based on 95% HPDs for crown ages obtained in the UCLN analysis. Mesangiospermae – which contains the vast majority of angiosperm diversity (*c.* 99.96% of extant species richness) – began to diversify between 137 and 135 Ma, soon after the crown diversification of angiosperms. The clades that include most angiosperm diversity started to diversify almost simultaneously: Magnoliidae (3.61% of extant richness) between 134.1 and 130.2 Ma; Monocotyledoneae (monocots; 23.32% of extant richness) between 134.7 and 131.6 Ma; and Eudicotyledoneae (eudicots; 73% of extant richness) between 133.4 and 129.7 Ma.

The finding that crown eudicots are approximately contemporaneous with crown Magnoliidae and monocots is a noteworthy difference from most previous molecular clock studies (e.g. Soltis *et al.*, 2002; Bell *et al.*, 2005, 2010; Magallón & Sanderson, 2005; Magallón & Castillo, 2009; Magallón *et al.*, 2013), where they were estimated to be younger. Eudicots are morphologically characterized by tricolpate pollen (or derived from this condition; Walker & Walker, 1984; Donoghue & Doyle, 1989), which probably evolved on the stem lineage of this clade. Tricolpate pollen grains are morphologically distinctive, can easily become preserved as fossils, and unequivocally indicate membership to a single clade, thus providing an exceptionally good calibration. Tricolpate pollen has been previously used to calibrate eudicots with a fixed or maximum age of *c.* 125 Ma, derived from the Barremian–early Aptian age of its oldest fossils (Doyle *et al.*, 1977; Hughes & McDougall, 1990; Doyle & Hotton, 1991). While we recognize the superior potential of tricolpate pollen grains for calibration, here we treated them in the same way as any other fossil used for calibration, and applied their oldest stratigraphic age as a minimum constraint on the eudicot stem node. The eudicots are here estimated as being nearly contemporaneous with Magnoliidae and monocots; therefore, the major components of angiosperm extant diversity began to diversify by the Hauterivian, between 136 and 130 Ma.

Within eudicots, Pentapetalae (70.7% of extant angiosperm richness), characterized by flowers with a five-part organization and distinct calyx and corolla, began to diversify between 126.5 and 120.9 Ma. A very large proportion of species diversity within Pentapetalae is contained in two large clades, Rosidae and Asteridae (29.15% and 35.16% of extant angiosperm richness, respectively), both of which are important components of modern terrestrial biomes. The initial diversification of Rosidae took place between 122.7 and 115.4 Ma, and apparently preceded that of Asteridae, which is estimated to have taken place between 118.8 and 110.4 Ma. Nevertheless, there is a substantial overlap between the two.

Current phylogenetic reconstructions (e.g. Wang *et al.*, 2009; Soltis *et al.*, 2011; but see Qiu *et al.*, 2010 and Zhang *et al.*, 2012) indicate that Rosidae consists of a pair of sister clades, Malvidae and Fabidae (10.68% and 18.47% of extant richness, respectively). The onset of Malvidae diversification is estimated to have occurred between 121.7 and 113.2 Ma, and Fabidae began to diversify between 120.9 and 113.3 Ma. The crown

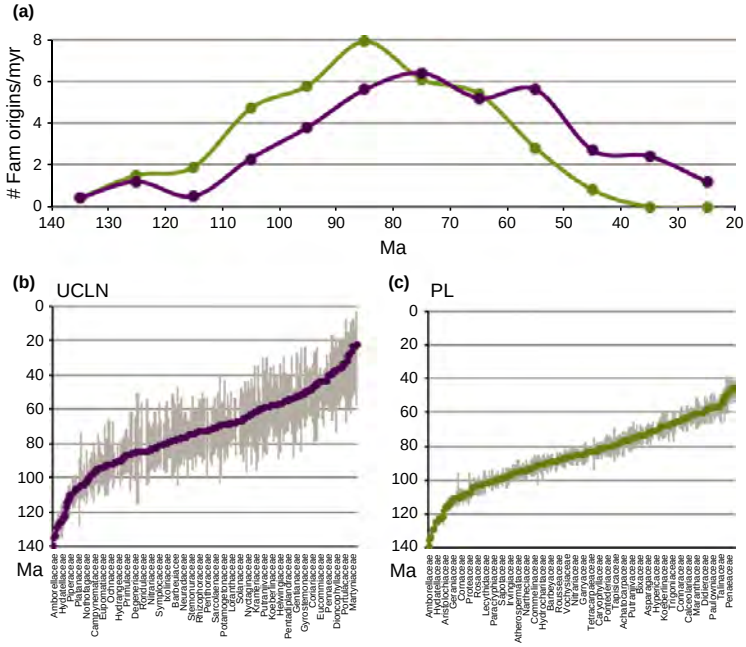


Fig. 4 Temporal distribution of family origins. (a) Number of family origins per million year in 10-Myr sliding windows. Green dots: penalized likelihood (PL) estimates; purple dots: Bayesian uncorrelated lognormal (UCLN) method estimates. (b) UCLN-estimated stem ages of families sorted from oldest to youngest. Grey bars correspond to 95% highest posterior density associated with each estimate. Family origins range from *c.* 140 to 20 million yr ago (Ma). (c) PL-estimated stem ages of families sorted from oldest to youngest. Grey bars correspond to confidence interval derived from a sample of dated maximum likelihood bootstrap trees associated with each estimate. Family origins range from *c.* 140 to 40 Ma. According to both UCLN and PL, family origins are constantly distributed, and periods in which family origins are markedly concentrated are not observed. There are fewer family origins at the beginning and end of the respective ranges, and they are more abundant between *c.* 100 and 50–40 Ma. These results are congruent with the number of family origins per Myr shown in (a), where the highest rates are also found between *c.* 100 and 50 Ma.

diversifications of Fabidae and Malvidae took place almost simultaneously, soon after their differentiation within Rosidae.

Asteridae contains a 'core asterid' clade consisting of the sister pair Garryidae (18.29% of extant richness) and Campanulidae (12.37% of extant richness). The diversifications of Garryidae and Campanulidae took place almost simultaneously, between 111.7 and 93.4 Ma, and 112.6 and 93.9 Ma, respectively. Hence, the two major clades within Rosidae and within Asteridae each diversified synchronously, but the first pair did so *c.* 10 Myr earlier than the second pair.

The estimated time-frame indicates an early proliferation of major clades in angiosperm history (Figs 2, 3). The major lineages Magnoliidae, Monocotyledoneae and Eudicotlyledoneae had originated and started to diversify by the Hauterivian. By the Aptian, the clades that contain a substantial proportion of extant angiosperm species richness, and are major components of extant biomes, had started to radiate: Malvidae and Fabidae in the early Aptian, and Garryidae and Campanulidae in the late Aptian, all during the Early Cretaceous.

The early rise of extant angiosperm families

According to the UNCL time-tree, angiosperm families originated between the Valanginian (Early Cretaceous) and the Miocene (Tertiary), but in the PL time-tree the range is shorter: from the Hauterivian to the Middle Eocene (Tertiary). Considering the oldest and youngest families in the UCLN and PL time-trees, their respective average number of family origins per Myr is 3.18 and 3.96. The number of family origins per Myr calculated in 10-Myr sliding windows indicates the highest rates between 100 and 60 Myr for PL, and between 90 and 50 Myr for UCLN (Fig. 4a).

Plots of UCLN and PL stem ages sorted from oldest to youngest (Fig. 4b,c) show a continuous origin of families from the Early Cretaceous (Hauterivian) to the Tertiary (early Miocene according to UCLN, and Middle Eocene according to PL), with fewer family origins immediately after the onset of angiosperm diversification (between *c.* 140 and 100 Ma), and as the present is approached (between *c.* 55 and 20 Ma in UCLN, and 60–40 Ma in PL). The time between *c.* 100 and 50–40 Ma shows the highest accumulation (Fig. 4b,c). These findings are congruent with the number of family origins per Myr in 10-Myr windows (Fig. 4a), which show lower rates at the beginning and end of the range and higher rates in the middle. The UCLN and PL time-trees show that the number of Cretaceous family origins substantially exceeds the Tertiary number (62.9% and 82.3%, respectively), showing that well over half of extant families have deep evolutionary roots.

Do families originate? The evolutionary significance of clades above the species level is currently being investigated (e.g. Barraclough, 2010; Humphreys & Barraclough, 2014). Processes that influence the generation of new species have been shown to be relevant above the species level, and to lead to higher evolutionarily significant units (Humphreys & Barraclough, 2014). Here, we consider that the origin of a family corresponds to a speciation event in which at least one of the descendants has acquired (or will acquire through its evolutionary trajectory) some type of distinctiveness (genetic, phenotypic, functional, or ecological) that (in hindsight) will allow taxonomists to postulate that species or its descendants as a family. Distinctiveness may be associated with

the phylogenetic differentiation and early stem evolution of the lineage, involving the establishment of new niches. Species richness (including the crown group) would be acquired subsequently. This implies that an adaptive radiation took place early within angiosperms, leading to the establishment of the major morphological and functional attributes that characterize its major lineages. The rapid construction of morphospace early in evolutionary radiations has been documented (Erwin, 2007). Alternatively, distinctiveness may be associated with the acquisition of species richness within a lineage, possibly associated with the diversification of its crown group. This scenario implies independent radiations in separate angiosperm lineages, possibly taking place at different times. These alternatives are not mutually exclusive, specifically considering an early adaptive radiation within angiosperms involving phylogenetic branching associated with large-level differentiation, and subsequent differentiation within separate lineages involving the generation of species richness.

The lineages that contain a very substantial amount of today's angiosperm species richness and their morphological, functional, ecological and genetic diversity were distinct very early in angiosperm history. Nevertheless, because of sampling density, the time-trees cannot show the time of acquisition of species richness. Species richness may be dissociated from the proliferation of major phylogenetic branches. Species that exist in the present (i.e. crown groups) may have originated soon after the differentiation of the family that contains them (short fuse) or substantially postdate it (long fuse), including the possibility that extant species originated recently.

The period in which most family origins are concentrated roughly corresponds to the onset of the Late Cretaceous to the end of the Middle Eocene. We note that this period was a time of pronounced tectonic and geological activity, and high global temperatures (Zachos *et al.*, 2001; Willis & McElwain, 2002). Is there a link between these global events and the increased origin of angiosperm families? Some studies have discussed the possible relationship between high environmental energy and high species richness (Davies *et al.*, 2004; Jansson & Davies, 2008). However, as previously discussed, the origin of a major clade (e.g. a family) implies differences from the increase in speciation, namely, the association with some type of substantial distinctiveness that will allow taxonomists to recognize that lineage as evolutionarily cohesive. There is no conclusive evidence of a relationship between the time of global tectonic change and high temperatures and the major concentration of family origins, and it is not clear what (if any) are the links between higher temperatures and enhanced morphological and functional evolution.

Conclusions

A large number of fossil-derived calibrations and a confidence interval on angiosperm age have been combined in relaxed clock analyses to provide a time-frame of angiosperm evolution. The maximum age of the onset of diversification of angiosperms into their living diversity has been calculated with high confidence to lie in the Early Cretaceous. This estimate was obtained in the context of a bias to estimate an old maximum age, derived from a strong underestimation of the average number of fossil localities from which each angiosperm family is known. An independent evaluation of the numerically estimated maximum angiosperm age provided here could be conducted with a recently available method that integrates the fossil record of the group in the context of a birth–death process (Heath *et al.*, 2014).

Why many molecular clocks have estimated substantially older ages for the angiosperm crown node (Fig. 1) requires to be investigated. Whereas relaxed clocks rely on increasingly powerful models to estimate divergence times, it has been shown that under complex estimation conditions or molecular model misspecification, absolute molecular rates and divergence times may be erroneously estimated (e.g. Jansa *et al.*, 2006; Hugall *et al.*, 2007; Lepage *et al.*, 2007; Brandley *et al.*, 2011). Relaxed clocks have been found to underestimate the magnitude of rate heterogeneity in trees with extensive rate variation (e.g. Wertheim *et al.*, 2012), or a single parametric distribution has been found to be insufficient to capture the variation in molecular rates found in some empirical trees (e.g. Dornburg *et al.*, 2012), in both cases leading to age overestimation.

This study documents the early rise of angiosperm phylogenetic diversity, including the early origin of more than half of extant angiosperm families. The estimated time-trees represent a solid framework for further investigating angiosperm evolution, for example, the rate of morphological and molecular change, biogeographical history, diversification dynamics, ancestral character reconstruction and state-dependent diversification; as well as coevolution with other biological lineages, correlations between diversification and the physical environment, and the evolution of modern terrestrial biomes. Nevertheless, many questions about the processes involved in early angiosperm evolution remain. To name only a few: What is the relationship between the detected early phylogenetic proliferation and the acquisition of distinctive (e.g. phenotypic or ecological) attributes? What is the relationship between the early rise of angiosperm major clades, including numerous families, and the acquisition of species richness, particularly extant species richness? Is there a relationship between global high temperatures and the origin of major angiosperm clades? If so, how do high temperatures influence the rates of phylogenetic branching and the evolution of phenotypic and functional attributes?

Acknowledgements

We thank Colin Hughes, Peter Linder and Reto Nyffeler for organizing the Radiations Conference, and inviting us to contribute an article to this special issue. We thank C. R. Marshall for guidance in calculating the confidence interval on angiosperm age; J. A. Doyle for information on angiosperm pollen; M. J. Sanderson and J. Schenk for suggestions on dating analyses; H. Sauquet for relevant observations; and P. R. Crane for helpful advice. C. Bell and two anonymous reviewers made many relevant observations. L. Eguiarte, A. Delgado-Salinas, G. Ortega-Leite, R. Hernández-Gutiérrez and James Fouzi provided comments and critical help. P. Linder provided useful comments. Dating analyses were conducted in the CIPRES Science Gateway.

The Coordinación de la Investigación Científica, Universidad Nacional Autónoma de México (UNAM) provided postdoctoral funding to S.G-A. CONACyT (grant no. 410511/262540) provides funding to L.S-R. L.S-R. thanks the Posgrado en Ciencias Biológicas, UNAM.

References

Aris-Brosou S, Yang Z. 2003. Bayesian models of episodic evolution support a late Precambrian explosive diversification of the Metazoa. *Molecular Biology and Evolution* **20**: 1947–1954.

Axelrod D. 1952. A theory of angiosperm evolution. *Evolution* **6**: 29–60.

Axelrod D. 1970. Mesozoic paleogeography and early angiosperm history. *Botanical Review* **36**: 277–319.

Barraclough T. 2010. Evolving entities: towards a unified framework for understanding diversity at the species and higher levels. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**: 1801–1813.

Baum D, Smith S. 2013. *Tree thinking: an introduction to phylogenetic biology.* Greenwood Village, CO, USA: Roberts & Company.

Bell C, Soltis D, Soltis P. 2005. The age of the angiosperms: a molecular timescale without a clock. *Evolution* **59**: 1245–1258.

Bell C, Soltis D, Soltis P. 2010. The age and diversification of the angiosperms re-revisited. *American Journal of Botany* **97**: 1296–1303.

Brandl R, Mann W, Sprinzl M. 1992. Estimation of the monocot-dicot age through tRNA sequences from the chloroplast. *Proceedings of the Royal Society B* **249**: 13–17.

Brandley M, Wang Y, Guo X, Montes-de-Oca A, Fería-Ortíz M, Hikida T, Ota H. 2011. Accommodating heterogenous rates of evolution in molecular divergence dating methods: an example using intercontinental dispersal of Plestiodon (Eumeces). *Systematic Biology* **60**: 3–15.

Brenner G. 1996. Evidence for the earliest stage of angiosperm pollen evolution: a paleoequatorial section from Israel. In: Taylor D, Hickey L, eds. *Flowering plant origin, evolution and phylogeny.* New York, NY, USA: Chapman & Hall, 91–115.

Cantino P, Doyle J, Graham S, Judd W, Olmstead R, Soltis D, Soltis P, Donoghue M. 2007. Towards a phylogenetic nomenclature of Tracheophyta. *Taxon* **56**: 1599–1613.

Cardinal S, Danforth B. 2013. Bees diversified in the age of eudicots. *Proceedings of the Royal Society B* **280**: 2012–2686.

Clarke J, Warnock R, Donoghue P. 2011. Establishing a time-scale for plant evolution. *New Phytologist* **192**: 266–301.

Collinson M. 1983. *Fossils plants of the London Clay.* London, UK: The Palaeontologial Association.

Davies TJ, Savolainen V, Chase MW, Moat J, Barraclough TG. 2004. Environmental energy and evolutionary rates in flowering plants. *Proceedings of the Royal Society B* **271**: 2195–2200.

Donoghue MJ, Doyle JA. 1989. Phylogenetic analysis of angiosperms and the relationships of Hamamelidae. In: Crane PR, Blackmore S, eds. *Evolution, systematics, and fossil history of the Hamamelidae*, vol **40A**. Oxford, UK: Clarendon Press, 17–45.

Donoghue P, Benton M. 2007. Rocks and clocks: calibrating the Tree of Life using fossils and molecules. *Trends in Ecology & Evolution* **22**: 424–431.

Dornburg A, Brandley M, McGowen M, Near T. 2012. Relaxed clocks and inferences of heterogeneous patterns of nucleotide substitution and divergence time estimates across whales and dolphins (Mammalia: Cetacea). *Molecular Biology and Evolution* **29**: 721–736.

Doyle J. 2012. Molecular and fossil evidence on the origin of angiosperms. *Annual Review of Earth and Planetary Sciences* **40**: 301–326.

Doyle J, Biens P, Doerenkamp A, Jardiné S. 1977. Angiosperm pollen from the pre-Albian Lower Cretaceous of equatorial Africa. *Bulletin des Centres de Recherches Exploration-Production Elf– Aquitaine* **1**: 451–473.

Doyle J, Endress P. 2000. Morphological phylogenetic analysis of basal angiosperms: comparison and combination with molecular data. *International Journal of Plant Sciences* **161**: S121–S153.

Doyle J, Endress P. 2010. Integrating Early Cretaceous fossils into the phylogeny of living angiosperms: Magnoliidae and eudicots. *Journal of Systematics and Evolution* **48**: 1–35.

Doyle J, Hotton C. 1991. Diversification of early angiosperm pollen in a cladistic context. In: Blackmore S, Barnes SH, eds. *Pollen and spores: patterns of diversification.* Oxford, UK: Claredon, 169–195.

Drummond A, Ho S, Phillips M, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology* **4**: e88.

Drummond A, Suchard M. 2010. Bayesian random local clocks, or one rate to rule them all. *BMC Biology* **8**: 114.

Edgar R. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792–1797.

Ericson P, Klopfstein S, Irestedt M, Nguyen J, Nylander J. 2014. Dating the diversification of the major lineages of Passeriformes (Aves). *BMC Evolutionary Biology* **14**: 8.

Erwin D. 2007. Disparity: morphological pattern and developmental context. *Palaeontology* **50**: 57–73.

Friedman W. 2009. The meaning of Darwin's "abominable mystery". *American Journal of Botany* **96**: 5–21.

Friis E, Crane P, Pedersen K. 2011. *Early flowers and angiosperm evolution.* Cambridge, UK: Cambridge University Press.

Goremykin V, Hansmann S, Martin WF. 1997. Evolutionary analysis of 58 proteins encoded in six completely sequenced chloroplast genomes: revised estimates of two seed plant divergence times. *Plant Systematics and Evolution* **206**: 337–351.

Hall T. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**: 95–98.

Heath TA. 2012. A hierarchical Bayesian model for calibrating estimates of species divergence times. *Systematic Biology* **61**: 793–809.

Heath TA, Huelsenbeck JP, Stadler T. 2014. The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences, USA* **111**: E2957–E2966.

Heimhofer U, Hochuli P, Burla S, Weissert H. 2007. New records of Early Cretaceous angiosperm pollen from Portuguese coastal deposits: implications for the timing of the early angiosperm radiation. *Review of Palaeobotany and Palynology* **144**: 39–76.

Hibbett D, Matheny P. 2009. The relative ages of ectomycorrhizal mushrooms and their plant hosts estimated using Bayesian relaxed molecular clock analyses. *BMC Biology* 7: 13.

Ho S. 2007. Calibrating molecular estimates of substitution rates and divergence times in birds. *Journal of Avian Biology* **38**: 409–414.

Hochuli PA, Feist-Burkhardt S. 2013. Angiosperm-like pollen and Afropollis from the Middle Triassic (Anisian) of the Germanic Basin (Northern Switzerland). *Frontiers in Plant Science* **4**: 344.

Hugall A, Foster R, Lee M. 2007. Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. *Systematic Biology* **56**: 543–563.

Hughes NF. 1994. *The enigma of angiosperm origins.* Cambridge, UK: Cambridge University Press.

Hughes N, McDougall A. 1987. Records of angiospermid pollen entry into the English Early Cretaceous succession. *Review of Palaeobotany and Palynology* **50**: 255–272.

Hughes N, McDougall A. 1990. Barremian-Aptian angiospermid pollen records from southern England. *Review of Palaeobotany and Palynology* **65**: 145–151.

Hughes N, McDougall A, Chapman J. 1991. Exceptional new record of Cretaceous Hauterivian angiospermid pollen from southern England. *Journal of Micropalaeontology* **10**: 75–82.

Humphreys AM, Barraclough TG. 2014. The evolutionary reality of higher taxa in mammals. *Proceedings of the Royal Society B* **281**: 20132750.

Hunt T, Bergsten J, Levkanicova Z, Papadopoulou A, St. John O, Wild R, Hammond P, Ahrens D, Balke M, Caterino M *et al.* 2007. A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. *Science* **318**: 1913–1916.

Jansa S, Barker F, Heaney L. 2006. The pattern and timing of diversification of Philippine endemic rodents: evidence from mitochondrial and nuclear gene sequences. *Systematic Biology* **55**: 73–88.

Jansson R, Davies TJ. 2008. Global variation in diversification rates of flowering plants: energy vs. climate change. *Ecology Letters* **11**: 173–183.

Jetz W, Thomas G, Joy J, Hartmann K, Mooers A. 2012. The global diversity of birds in space and time. *Nature* **491**: 444–448.

Knobloch E, Mai D. 1986. Monographie der Früchte und Samen in der Kreide von Mitteleuropa. *Rozpravy Ústředního Ústavu Geologického* **47**: 1–219.

Laroche J, Li P, Bousquet J. 1995. Mitochondrial DNA nd monocot-dicot divergence time. *Molecular Biology and Evolution* **12**: 1151–1156.

Lepage T, Bryant D, Philippe H, Latrillot N. 2007. A general comparison of relaxed molecular clock models. *Molecular Biology and Evolution* **24**: 2669–2680.

Li W-H, Graur D. 1991. *Fundamentals of molecular evolution.* Sunderland, MA: Sinauer.

Li W-H, Tanimura M. 1987. The molecular clock runs more slowly in man than in apes and monkeys. *Nature* **326**: 93–96.

Magallón S. 2007. From fossils to molecules: phylogeny and the core eudicot floral groundplan in Hamamelidoideae (Hamamelidaceae, Saxifragales). *Systematic Botany* **32**: 317–347.

Magallón S. 2010. Using fossils to break long branches in molecular dating: a comparison of relaxed clocks applied to the origin of angiosperms. *Systematic Biology* **59**: 384–399.

Magallón S. 2014. A review of the effect of relaxed clock method, long branches, genes, and calibrations in the estimation of angiosperm age. *Botanical Sciences* **92**: 1–22.

Magallón S, Castillo A. 2009. Angiosperm diversification through time. *American Journal of Botany* **96**: 349–365.

Magallón S, Hilu K, Quandt D. 2013. Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *American Journal of Botany* **100**: 556–573.

Magallón S, Sanderson M. 2005. Angiosperm divergence times: the effect of genes, codon positions, and time constraints. *Evolution* **59**: 1653–1670.

Manchester S. 1999. Biogeographical relationships of North American Tertiary Floras. *Annals of the Missouri Botanical Garden* **86**: 472–522.

Marshall C. 1990. Confidence intervals on stratigraphic ranges. *Paleobiology* **16**: 1–10.

Marshall C. 1994. Confidence intervals on stratigraphic ranges: partial relaxation of the assumption of randomly distributed fossil horizons. *Paleobiology* **20**: 459–469.

Marshall C. 1997. Confidence intervals on stratigraphic ranges with nonrandom distributions of fossil horizons. *Paleobiology* **23**: 165–173.

Marshall C. 2008. A simple method for bracketing absolute divergence times on molecular phylogenies using multiple fossil calibration points. *The American Naturalist* **171**: 726–742.

Martin W, Gierl A, Saedler H. 1989. Molecular evidence for pre-Cretaceous angiosperm origins. *Nature* **339**: 46–48.

Martin W, Lydiate D, Brinkmann H, Forkmann G, Saedler H, Cerff R. 1993. Molecular phylogenies in angiosperm evolution. *Molecular Biology and Evolution* **10**: 140–162.

Martínez-Millán M. 2010. Fossil record and age of the Asteridae. *The Botanical Review* **76**: 83–135.

Martínez-Millán M, Crepet W, Nixon K. 2009. *Pentapetalum trifasciculandricus* gen. et sp. nov., a thealean fossil flower from the Raritan Formation, New Jersey, USA (Turonian, Late Cretaceous). *American Journal of Botany* **96**: 933–949.

McKenna D, Sequeira A, Marvaldi A, Farrell B. 2009. Temporal lags and overlap in the diversification of weevils and flowering plants. *Proceedings of the National Academy of Sciences, USA* **106**: 7083–7088.

Miller M, Pfeiffer W, Schwartz T. 2010. *Creating the CIPRES Science Gateway for inference of large phylogenetic trees.* New Orleans, LA, USA: Gateway Computing Environments Workshop (GCE).

Moore M, Bell C, Soltis P, Soltis D. 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proceedings of the National Academy of Sciences, USA* **104**: 19363–19368.

Pyron R. 2011. Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. *Systematic Biology* **60**: 466–481.

Qiu Y-L, Li L, Wang B, Xue J-Y, Hendry TA, Li RQ, Brown JW, Liu Y, Hudson GT, Chen ZD. 2010. Angiosperm phylogeny inferred from sequences of four mitochondrial genes. *Journal of Systematics and Evolution* **48**: 391–425.

Rambaut A, Drummond A. 2009. *Tracer version 1.5.0.* [WWW document] URL http://tree.bio.ed.ac.uk/software/tracer/ [accessed 1 July 2013].

Ramshaw J, Richardson D, Meatyard B, Brown R, Richardson M, Thompson E, Boulter D. 1972. The time of origin of the flowering plants determined by using amino acid sequence data of cytochrome c. *New Phytologist* **71**: 773–779.

Rannala B, Yang Z. 2007. Inferring speciation times under an episodic molecular clock. *Systematic Biology* **56**: 453–466.

Ronquist F, Klopfstein S, Vilhelmsen L, Schulmeister S, Murray D, Rasnitsyn A. 2012b. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Systematic Biology* **61**: 973–999.

Ronquist F, Teslenko M, van der Mark P, Ayres D, Darling A, Höhna S, Larget B, Liu L, Suchard M, Huelsenbeck J. 2012a. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**: 539–542.

Sanderson M. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution* **14**: 1218–1231.

Sanderson M. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution* **19**: 101–109.

Sanderson M. 2004. *r8s, version 1.70 user's manual.* [WWW document] URL http://loco.biosci.arizona.edu/r8s/ [accessed 1 April 2013].

Sanderson M, Doyle J. 2001. Sources of error and confidence intervals in estimating the age of angiosperms from rbcL and 18S rDNA data. *American Journal of Botany* **88**: 1499–1516.

Sauquet H, Ho S, Gandolfo M, Jordan G, Wilf P, Cantrill D, Bayly M, Bromham L, Brown G, Carpenter R. 2012. Testing the impact of calibration on molecular divergence times using a fossil-rich group: the case of *Nothofagus* (Fagales). *Systematic Biology* **61**: 298–313.

Schneider H, Schuettpelz E, Pryer K, Cranfill R, Magallón S, Lupia R. 2004. Ferns diversified in the shadow of angiosperms. *Nature* **428**: 553–557.

Smith A, Pisani D, Mackenzie-Dodds J, Stockley B, Webster B, Littlewood D. 2006. Testing the molecular clock: molecular and paleontological estimates of divergence times in the Echinoidea (Echinodermata). *Molecular Biology and Evolution* **23**: 1832–1851.

Smith S, Beaulieu J, Donoghue M. 2010. An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proceedings of the National Academy of Sciences, USA* **107**: 5897–5902.

Smith S, O'Meara B. 2012. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28**: 2689–2690.

Soltis D, Smith S, Cellinese N, Wurdack K, Tank D, Brockington S, Refuli-Rodriguez N, Walker J, Moore M, Carlsward B *et al.* 2011. Angiosperm phylogeny: 17 genes, 640 taxa. *American Journal of Botany* **98**: 704–730.

Soltis P, Soltis D, Savolainen V, Crane P, Barraclough T. 2002. Rate heterogeneity among lineages of tracheophytes: integration of molecular and fossil data and evidence for molecular living fossils. *Proceedings of the National Academy of Sciences, USA* **99**: 4430–4435.

Stamatakis A. 2006a. Phylogenetic models of rate heterogeneity: a high performance computing perspective. *Proceedings of 20th IEEE International Parallel and Distributed Processing Symposium (IPDPS2006)*, Rhodos, Greece.

Stamatakis A. 2006b. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.

Stamatakis A. 2008. A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology* **57**: 758–771.

Stevens P. 2013. *Angiosperm phylogeny website. Version 12, July 2012 [and more or less continuously updated since]*. [WWW document] URL http://www.mobot.org/MOBOT/Research/APweb/welcome.html [accessed 1 April 2013].

Strauss D, Sadler P. 1989. Stochastic models for the completeness of stratigraphic sections. *Mathematical Geology* **21**: 37–59.

Wahlberg N, Wheat C, Peña C. 2013. Timing and patterns in the taxonomic diversification of Lepidoptera (butterflies and moths). *PLoS One* **8**: e80875.

Walker J, Geissman J. 2009. Geologic time scale. *GSA Today* **19**: 60–61.

Walker JW, Walker AG. 1984. Ultrastructure of Lower Cretaceous angiosperm pollen and the origin and early evolution of flowering plants. *Annals of the Missouri Botanical Garden* **71**: 464–521.

Wang H, Moore M, Soltis P, Bell C, Brockington S, Alexandre R, Davis C, Latvis M, Manchester S, Soltis D. 2009. Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proceedings of the National Academy of Sciences, USA* **106**: 3853–3858.

Wertheim J, Fourment M, Kosakovsky Pond S. 2012. Inconsistencies in estimating the age of HIV-1 subtypes due to heterotachy. *Molecular Biology and Evolution* **29**: 451–456.

Wikström N, Kenrick P. 2001. Evolution of Lycopodiaceae (Lycopsida): estimating divergence times from *rbcL* gene sequences by use of nonparametric rate smoothing. *Molecular Phylogenetics and Evolution* **19**: 177–186.

Wilkinson R, Steiper M, Soligo C, Martin R, Yang Z, Tavaré S. 2011. Dating primate divergences through an integrated analysis of palaeontological and molecular data. *Systematic Biology* **60**: 16–31.

Willis KJ, McElwain JC. 2002. *The evolution of plants.* Oxford, UK: Oxford University Press.

Wolfe KH, Gouy M, Yang YW, Sharp PM, Li WH. 1989. Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. *Proceedings of the National Academy of Sciences, USA* **86**: 6201–6205.

Yang Z, Rannala B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution* **23**: 212–226.

Zachos J, Pagani M, Sloan L, Thomas E, Billups K. 2001. Trends, rythms, and aberrations in global climate 65 Ma to present. *Science* **292**: 686–693.

Zhang N, Zeng L, Shan H, Ma H. 2012. Highly conserved low-copy nuclear genes as effective markers for phylogentic analyses in angiosperms. *New Phytologist* **195**: 923–937.

Supporting Information

Additional supporting information may be found in the online version of this article.

Table S1 List of species and GenBank accessions

Table S2 Ages of major angiosperm clades

Methods S1 Fossil-derived calibrations.

Methods S2 Angiosperm branches with the highest empirical scaling factor.

Please note: Wiley Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

Material Suplementario

Las tablas y métodos suplementarios, y los datos generados en este Apéndice 1, se encuentran disponibles respectivamente en la version en línea del artículo http://onlinelibrary.wiley.com/wol1/doi/10.1111/nph.13264/suppinfo, y en el repositorio digital Dryad http://datadryad.org/resource/doi:10.5061/dryad.k4227/4.