



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

EVOLUCIÓN DE GENES *ARP* EN
HELECHOS

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

B I Ó L O G A

PRESENTA:

BRENDA HERNÁNDEZ HERNÁNDEZ

DIRECTORAS DE TESIS:

DRA. ROSALINDA TAPIA LÓPEZ

DRA. ALEJANDRA VASCO GUTIÉRREZ



Ciudad Universitaria, Cd. Mx., 2019



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

1. Datos del alumno

Hernández
Hernández
Brenda
5573491050
Universidad Nacional Autónoma de México
Facultad de Ciencias
Biología
310051455

2. Datos del tutor

Dra.
Rosalinda
Tapia
López

3. Datos del co-tutor

Dra.
Alejandra
Vasco
Gutiérrez

4. Datos del sinodal 1

Dra.
Guadalupe Judith
Márquez
Guzmán

5. Datos del sinodal 2

Dra.
Berenice
García
Ponce de León

6. Datos del sinodal 3

Natalia Ivalú
Cacho
González

7. Datos del trabajo escrito

Evolución de genes *ARP* en helechos
90 p.
2019

Agradecimientos

A la Universidad Nacional Autónoma de México (UNAM). Por tanto y por todo. Por mantenerse pública y gratuita.

A mis asesoras las Dras. Rosalinda Tapia y Alejandra Vasco por ser pacientes, dedicarme su tiempo y confiar en que yo podría lograrlo. Por darme todo el apoyo que necesité para concluir este gran paso. Por ser mis amigas y mi inspiración, porque a su lado siempre me sentí amparada y fuerte. Ale, Rous, creo que nunca podré demostrar todo el agradecimiento y cariño que les tengo y que por siempre tendré. Ale, fue un placer trabajar con la mejor Pteridóloga del mundo y sé que aún tengo mucho que aprender de ti. Llevas un lugar muy especial en mi corazón justo al lado de los helechos, te quiero mucho Ale. Rous, me llena de honor y orgullo decir que soy tu aprendiz, eres la mejor maestra, gracias por darme un Norte cuando en mi mente todo era bruma. Te quiero Rous.

Al financiamiento del Programa UNAM-DGAPA-PAPIIT (Proyecto IA201416) y al CONACyT (convocatoria INFR-2016-01, proyecto No. 268109) (Dra. Alejandra Vasco).

Al Dr. Juan Núñez Farfán y al laboratorio de Genética Ecológica y Evolución del Instituto de Ecología de la UNAM. Por recibirme en su laboratorio y darme las facilidades necesarias para la realización de este proyecto. Dr. Juan ha sido un honor aprender tanto de usted, es una perso

A la Dra. Guadalupe Judith Márquez Guzmán, y Dra. Berenice García Ponce de León. Por enriquecer mi trabajo y expandir mi visión acerca de la evolución. Por el tiempo que me dedicaron y las dudas que resolvieron. Sus observaciones fueron de gran ayuda para poder entender mi proyecto y darlo a conocer de la mejor manera.

A la Dra. N. Ivalú Cacho por ser siempre una inspiración, por brindarme todo su apoyo y evitar que me convirtiera en una estadística más.

A la Dra. Barbara Ambrose (New York Botanical Garden) por proporcionarme las secuencias del genoma de *Ceratopteris richardii*. Por el tiempo que dedicó a escucharme y por todas las sugerencias que me dio.

Al Dr. José de Jesus García Trejo. Por el tiempo que dedico para ayudarme con los análisis en Swiss Model.

Al laboratorio de Genómica Funcional y Sistemas del Instituto de Biología de la UNAM.

A los profesores del Taller de Evolución de plantas: biodiversidad, sistemática, fisiología y mecanismos moleculares de desarrollo: Florencia García y Ulises Rosas por todos los consejos y las preguntas que me regalaron durante el desarrollo de mi proyecto.

Al Dr. Zenón Cano Santana. Por ser un gran amigo. Doc, creo que yo y muchos estudiantes están en deuda infinita con usted. Muchas gracias por todo el apoyo y todos los consejos.

Agradecimientos personales:

A mis padres: María Natividad Hernández Moctezuma, Benigno Hernández Luna, Amparo Hernández Hernández. Por ser el mejor ejemplo de fortaleza, tenacidad y determinación. Gracias por levantarse cada día y esforzarse tanto para que yo pudiera estudiar y ser una buena persona.

A mis hermanos: Ángel, Adriana y Benny. Gracias por cuidarme y ser uno de los pilares mas importantes en mi vida.

A Eve. Por ser mi mejor amiga, por no soltarme nunca, por ser el puerto en cada tempestad, por ser mi aliada y cómplice, por creer que yo podía lograrlo.

A Gloria. Porque sin ti no habría confiado en mí.

A Mark Reyes. Por estar siempre que quería hablar de árboles, genes y helechos, y también cada que se perdía Fermín. Marco, más que un hermano eres un brother.

A Pauli. Pauli Pauli Pauli... por ser un gran pilar y un gran consuelo en tiempos difíciles, por transmitirme paz cada vez que yo la perdía y por resolver tantas dudas que habitan en mí. A los hermanos se les da el titulo de amigos muchas veces, pero los amigos muy pocas veces merecen llamarse hermanos. Te quiero Pau.

A Manu, perdiste el juego. Gracias por todas las veces que intentaste enseñarle programación funcional y matemáticas a este chango. Te quiero.

A Diana Karen. Por ser mi persona y prestarme a su familia cada vez que yo extrañaba a la mía.

A Yuli. Por ser una gran amiga, por compartir conmigo sueños y estudiar conmigo la peor materia del mundo.

Julián Carax. Gracias por ayudarme con la safranina

A mis amiguitos del taller: Verito, Andy Arabidopsis, Itzel, Joel y Jony.

A todos los amigos que hice en la carrera y a las personas que forman parte de los buenos recuerdos que tengo de la Facultad de Ciencias: Iván Sandoval García, Rodrigo Niebla, El mapa, Cesar Montes, Ana Lucía, Nydia, Majito, Jimmy, Sam, Yohualli, Oscor y especialmente a Vic Piña (I follow rivers Pa').

Y finalmente, a todos los profesores que me dejaron un valioso aprendizaje y sirvieron como inspiración: Saúl Cano Colín, Antonio Lazcano Araujo, Radames Ricardo Reynoso Manriquez, César Fernández Vargas, Andrés Argüelles Moyao, Felipe Eslava Silva, Felipe Gómez Noguez, Amadeo Luis Estrada Nieto, Alberto Vladimir Cachón Guillén, Álvaro Chaos Cador, José Fernando Aceves Quesada y Arturo Becerra.

A Fermín, que no sabe leer.

Índice

Resumen.....	7
Introducción	8
Evolución de hojas	10
Evolución de hojas en helechos.....	11
Desarrollo de hojas en helechos	13
Red genética del desarrollo foliar en plantas vasculares	15
Establecimiento de los primordios foliares.....	15
Polaridad de la hoja.....	16
El gen <i>ASI</i> (<i>ASYMMETRIC LEAVES1</i>) del grupo ARP.....	17
El dominio MYB	18
Interacción antagónica entre ARP y KNOX Clase I.....	20
El gen <i>AS2</i> (<i>ASYMMETRIC LEAVES2</i>).....	22
Genética del desarrollo de hojas en helechos	23
Justificación.....	24
Preguntas de investigación	24
Objetivos	25
Objetivo General	25
Objetivos Particulares.....	25
Hipótesis.....	25
Nula.....	25
Alternativas.....	25
Materiales y métodos	27
Etapa I: Análisis exploratorios, buscando los ortólogos de los genes <i>ARP</i> en helechos.....	27
a) Búsquedas de secuencias	27
b) Alineación de secuencias.....	28
c) Filogenia Exploratoria 1	28
Etapa II: Ampliación del muestreo a la familia de genes <i>MYB</i>	29
a) Búsquedas en angiospermas	29
b) Búsquedas en helechos, gimnospermas, licofitas y musgos.....	29
Etapa III: Historia evolutiva de los genes <i>MYB</i> y <i>ARP</i> con énfasis en helechos.....	30
a) División de los datos en dos matrices.....	30
b) Análisis de similitud	30
c) Análisis filogenéticos	30

Etapa IV: Motivos y modelos por homología de las proteínas ARP.....	31
a) Análisis de motivos proteicos.....	31
b) Modelos por homología de las proteínas ARP y clado b	31
Resultados	32
Etapas I y II: Búsqueda exploratoria de secuencias y ampliación del muestreo	32
Etapa III: Filogenias finales, historia evolutiva de los genes <i>MYB</i> y <i>ARP</i> con énfasis en helechos	33
a) Alineamientos.....	33
b) Análisis de similitud.....	35
c) Análisis filogenéticos con la Matriz G	35
d) Análisis filogenéticos con la Matriz G+T (Genomas +Transcriptomas).....	43
e) Análisis de motivos proteicos	51
f) Modelo por homología de las proteínas del clado ARP y el clado b	54
Discusión.....	61
Evolución de los genes <i>MYB</i> en helechos	61
Ausencia de genes <i>ARP</i> en helechos	63
Sustitución funcional de <i>ARP</i> en helechos	64
Identificación de nuevos genes <i>ARP</i> en <i>Selaginella</i> y <i>Ginkgo</i>	67
Modelo por homología de las proteínas del clado ARP y el clado b	67
Conclusiones	68
Perspectivas.....	69
Referencias	70
Material suplementario.....	76

Resumen

Las hojas son una innovación evolutiva de suma importancia en la historia de las plantas vasculares. A pesar del impacto que han tenido en la historia de la vida en la Tierra, nuestro conocimiento sobre su evolución aún no es claro. Los helechos muestran gran diversidad en la forma de la hoja por lo que son un linaje clave para comprender la evolución de dicho órgano en las plantas vasculares y en las eufilofitas (helechos y plantas con semilla). Entender las diferencias y similitudes en los módulos genéticos de desarrollo foliar en los diferentes linajes de plantas vasculares permitirá entender mejor el origen, la evolución y el desarrollo de las hojas. El objetivo principal de este trabajo fue entender la evolución de los genes *ARP* (*ASYMMETRIC LEAVES1*, *ROUGH SHEATH2*, *PHANTASTICA*) en plantas vasculares con énfasis en helechos, con el fin de entender mejor la evolución y el origen de las hojas en eufilofitas.

Se realizaron búsquedas de genes ortólogos de *ARP* y de genes *MYB* en los principales linajes de plantas (musgos, licofitas, helechos, gimnospermas y angiospermas) en diferentes bases de datos de genomas y transcriptomas de plantas terrestres. Con las secuencias obtenidas se realizaron análisis filogenéticos con Inferencia Bayesiana y Máxima Verosimilitud. También se realizó un análisis de motivos proteicos (MEME), con el fin de determinar los motivos característicos de los genes *ARP*. Finalmente, con la finalidad de determinar su conformación tridimensional se realizó un modelo por homología de las proteínas de los genes *ARP* y de las secuencias más cercanamente relacionadas con base en una de las filogenias obtenidas.

Nuestros resultados muestran que no hay genes *ARP* en helechos, pero encontramos que otros genes que codifican para el dominio MYB son diversos en este linaje. Con las secuencias de los linajes en lo que si se recuperaron ortólogos a *ARP* (angiospermas, gimnospermas y *Selaginella*), se logró identificar un motivo característico de estos genes. A pesar de que se han realizado modelos de estructura tridimensional para diferentes proteínas MYB, hasta ahora no se contaba con uno que describiera a las proteínas ARP; este trabajo proporciona el primer modelo de ellas. La ausencia de genes ortólogos de *ARP* en helechos sugiere que la red de desarrollo foliar en este grupo de plantas podría ser diferente a la de las plantas con semilla, lo que apoya la hipótesis de que no hay homología entre las hojas de eufilofitas. Proponemos como sustitutos de la función de *ARP* en los helechos a los genes correspondientes a las secuencias de helechos del clado b, al ortólogo de *AS2* o al gen *JLO*, en el caso de encontrarse en este linaje. A su vez, otro mecanismo alternativo para la regulación de los genes *KNOX Clase I*, podría ser la señalización a través de las auxinas. De igual forma, proponemos que otros genes *MYB* pudieran ser los responsables del desarrollo de la hoja, a través de sus regiones promotoras.

Introducción

Las hojas son el órgano lateral con mayor diversidad de formas en las plantas vasculares actuales (licofitas, helechos y plantas con semilla). Se caracterizan por tener un crecimiento determinado, un sistema vascular, presentar una polaridad adaxial-abaxial y un patrón predecible de organización alrededor de un eje (filotaxia) (Harrison & Morris, 2018; Tomescu, 2009).

El primer registro fósil de plantas terrestres con diminutas excrecencias laterales semejantes a hojas data del Silúrico tardío-Devónico medio, hace aproximadamente 427 millones de años (ma) (revisado en Beerling et al., 2001). Por otro lado, las hojas con venación dicotómica y forma laminar aparecen en el registro fósil al final del periodo Devónico, hace aproximadamente 360 ma (revisado en Beerling et al., 2001) (Figura 1). Sin embargo, a pesar de que las hojas de distintos grupos de plantas vasculares (licofitas, helechos y plantas con semilla) presentan características similares, el registro fósil y los análisis filogenéticos sugieren que su origen es independiente (Corvez et al., 2012; Tomescu, 2009).

La aparición de las hojas permitió un cambio en la estructura de las plantas, convirtiéndose en el órgano dominante de las plantas vasculares. Esta estructura también es de suma importancia en la reproducción sexual de las plantas, ya que en ellas se encuentran sostenidos los esporangios. Entender el desarrollo de las hojas permite entender la evolución de estructuras que derivaron de ellas, como los carpelos en el caso de las angiospermas (Chanderbali et al., 2016; Endress & Doyle, 2009).

Las hojas también modificaron los ciclos biogeoquímicos y ecológicos de la tierra (Beerling et al., 2001; Morris et al., 2018). Por ejemplo: la evapotranspiración y respiración de las hojas permitió que el nivel de humedad en la atmósfera aumentara y los niveles de CO₂ atmosféricos disminuyeran, lo cual dio lugar al desarrollo y diversificación de nuevas formas de vida (Beerling & Berner, 2005; Beerling et al., 2001).

A pesar de la importancia que tienen las hojas en la historia de la vida, aún existen muchas interrogantes acerca de su origen y evolución. Diferentes estudios filogenéticos, moleculares, paleobotánicos, genéticos y de biología del desarrollo han intentado resolver esta pregunta, sin embargo, los resultados son aún contradictorios y no concluyentes (revisado en Vasco et al., 2013).

Recientemente, la aparición y evolución de las hojas se ha abordado desde la biología evolutiva del desarrollo (Evo-Devo, del inglés Evolutionary Developmental Biology). La comparación de los módulos genéticos de desarrollo de las hojas en diferentes linajes de plantas vasculares ha permitido hacer nuevas inferencias acerca de su evolución (Harrison et al., 2005; Vasco et al., 2013).

El principal objetivo de este trabajo es entender la evolución de los genes *ARP* (*ASYMMETRIC LEAVES1*, *ROUGH SHEATH2*, *PHANTASTICA*) en los helechos, los cuales, en angiospermas permiten que las células del meristema apical se diferencien y participan en el establecimiento de la polaridad y el desarrollo de la vasculatura en la hoja. Conocer la historia evolutiva y la presencia o ausencia de genes *ARP* en los helechos dará luz sobre la posible conservación evolutiva de un módulo de desarrollo de las hojas en las plantas vasculares y permitirá entender mejor la evolución de las hojas en el linaje que incluye helechos y plantas con semillas (eufilofitas).

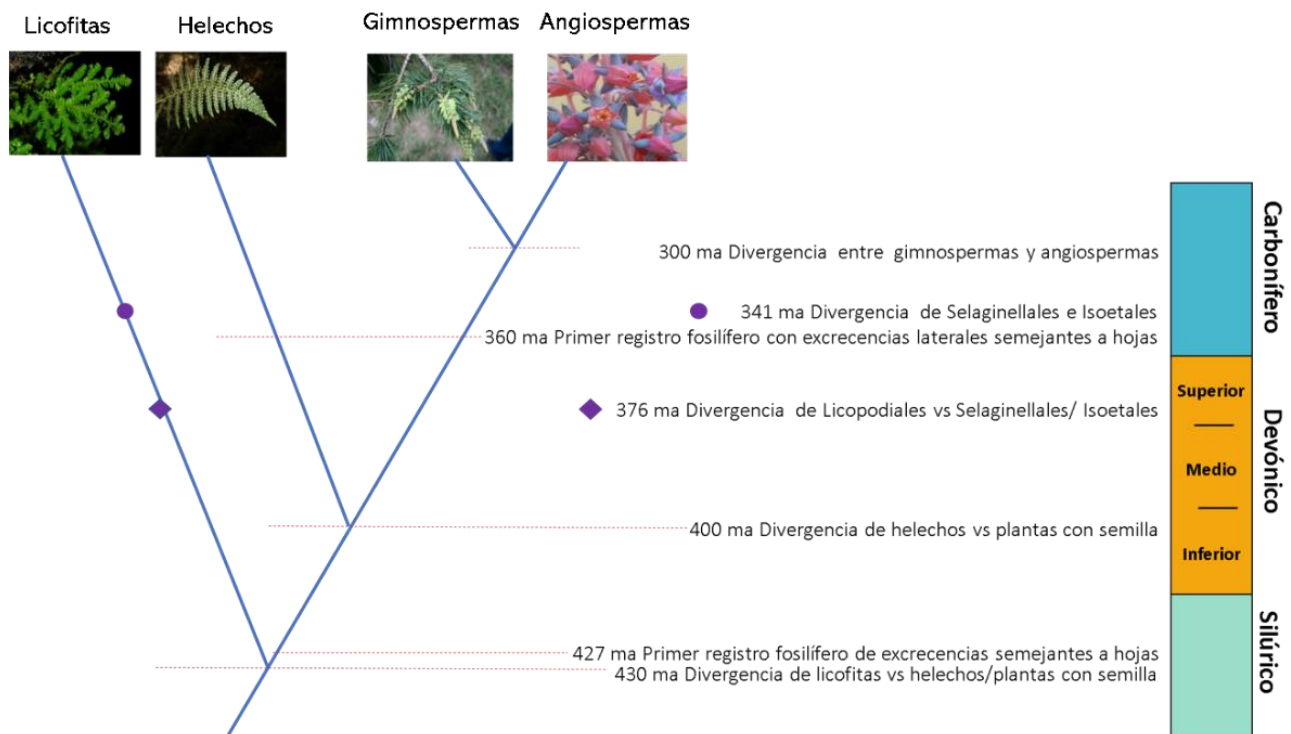


Figura 1. Edades de divergencia de los linajes de plantas terrestres (de acuerdo a Pires & Dolan, 2012) y de los primeros registros fósiles de estructuras semejantes a hojas (Revisado en Beerling et al., 2001). Imagen de acuerdo a la tabla cronoestratigráfica internacional 2015 (Tabla diseñada por K.M. Cohen, S.C. Finney y P.L. Gibbard).

Evolución de hojas

Las plantas terrestres (embriofitas) evolucionaron a partir de un ancestro de agua dulce perteneciente al grupo de las estreptofitas (Becker & Marin, 2009; Lewis & McCourt, 2004; Wickett et al., 2014). Los microfósiles de las primeras plantas terrestres indican que éstas no tenían hojas y que únicamente eran ejes ramificados desnudos, en cuyos ápices se desarrollaban esporangios (Wellman et al., 2003).

Las plantas vasculares actuales incluyen tres linajes principales: licofitas, helechos y plantas con semilla, los cuales surgieron hace 430 ma, 400 ma, y 300 ma, respectivamente (Pires & Dolan, 2012; Plackett et al., 2015) (Figura 1). Todos estos linajes poseen hojas actualmente; sin embargo, el primer registro fósil de plantas terrestres con excrecencias laterales semejantes a hojas data de finales del periodo Devónico, hace aproximadamente 360 ma (revisado en Beerling et al., 2001), es decir que, este órgano apareció en las plantas terrestres cuando las plantas vasculares ya habían tenido su primera divergencia (Tomescu, 2009) (Figura 1).

En general, se acepta que en las plantas vasculares existen dos tipos de hojas: las micrófilas que son las hojas presentes en las licofitas; y las megáfilas, que son las hojas presentes en las eufilofitas (grupo que comprende los helechos y las plantas con semilla) (Vasco et al., 2013). La principal diferencia morfológica entre las hojas micrófilas y las hojas megáfilas es el nivel de vascularización de la lámina. Las hojas megáfilas generalmente cuentan con una red compleja de haces vasculares y tienen una laguna foliar (excepto los taxones con protostelas) (Figura 2a), mientras que las hojas micrófilas generalmente tienen un solo haz vascular y carecen de laguna foliar (Figura 3a) (Dkhar & Pareek, 2014; Tomescu, 2009).

La teoría del teloma es hasta ahora la más aceptada para explicar la formación de las hojas megáfilas en eufilofitas (Zimmerman, 1952; revisado en Vasco 2013) (Figura 2). Esta teoría sostiene que las hojas megáfilas evolucionaron a partir de conjuntos de ejes, mediante tres procesos: superposición, aplanamiento y ensanchamiento (Zimmerman, 1952; revisado en Beerling & Fleming, 2007) (Figura 2). El proceso de superposición plantea que los ejes dejan de crecer a un mismo ritmo de modo que los ejes ya no tienen divisiones dicotómicas equivalentes, por lo tanto, existen ejes de diferentes tamaños a lo largo de un eje principal (Zimmerman, 1952; revisado en Beerling & Fleming, 2007). El proceso de aplanamiento explica que los ejes se orientan en un mismo plano (Zimmerman, 1952; revisado en Beerling & Fleming, 2007). Finalmente, el ensanchamiento implica que entre los ejes de la planta se forme un tejido laminar por medio de excrecencias laterales (Zimmerman, 1952; revisado en Beerling & Fleming, 2007).

Para explicar el origen de las hojas micrófilas se han propuesto tres teorías con base en el registro fósil: la teoría de reducción de telomas (Zimmerman, 1952; revisado en Vasco et al., 2013), la teoría de la esterilización (Kenrick & Crane, 1997) y la teoría de la enación (Bower, 1935; revisado en Vasco et al., 2013) (Figura 3). La teoría de la reducción de telomas propone que algunos telomas tanto fértiles como estériles se redujeron dando origen a las micrófilas (Beerling & Fleming, 2007; Kenrick & Crane, 1997) (Figura 3b). Por otro lado, la teoría de la esterilización plantea que las hojas micrófilas evolucionaron a partir de esporangios que se esterilizaron (Beerling & Fleming, 2007; Kenrick & Crane, 1997) (Figura 3c). Por su parte la teoría de la enación plantea que las hojas surgieron inicialmente como excrescencias superficiales de las ramas que posteriormente se vascularizaron (Bower, 1935; revisado en Vasco et al., 2013) (Figura 3d). Esta última teoría es la más ampliamente citada para explicar el origen de las hojas en licofitas (ver Vasco et al. 2016 para una discusión al respecto).

La evolución independiente de las hojas de licofitas y eufilofitas ha quedado bien establecida, sin embargo, aún existe un gran debate acerca de la homología de las hojas en eufilofitas, principalmente en helechos (Tomescu, 2009; Vasco et al., 2013).

Evolución de hojas en helechos

Los helechos comprenden 14 órdenes y más de 12 mil especies (PPG I, 2016). Las hipótesis filogenéticas moleculares actuales los ubican como el grupo hermano de las plantas con semilla (Pryer et al., 2001; Raubenson & Jansen, 1992; Sessa et al., 2014; Smith et al., 2006) (Figura 4). Sin embargo, las relaciones evolutivas dentro del linaje de los helechos han sido ampliamente debatidas, principalmente la inclusión de los órdenes Equisetales, Psilotales y Ophioglossales ya que su morfología es bastante diferente al resto de los helechos (Pryer et al., 2001; Vasco et al., 2013).

Las filogenias con base en datos moleculares recuperan a los órdenes Equisetales, Psilotales, Ophioglossales, Marattiales, y a los helechos leptosporangiados como monofiléticos, sugiriendo que las hojas de los helechos podrían tener un origen común (Pryer et al., 2001). Por otro lado, las filogenias que incluyen evidencia fósil ubican a cada uno de estos grupos en diferentes posiciones dentro de las plantas vasculares y recuperan a los helechos como un grupo parafilético (Rothwell & Nixon, 2006; Tomescu, 2009). Esta última hipótesis implica que las hojas en los helechos actuales pudieron tener diferentes orígenes.

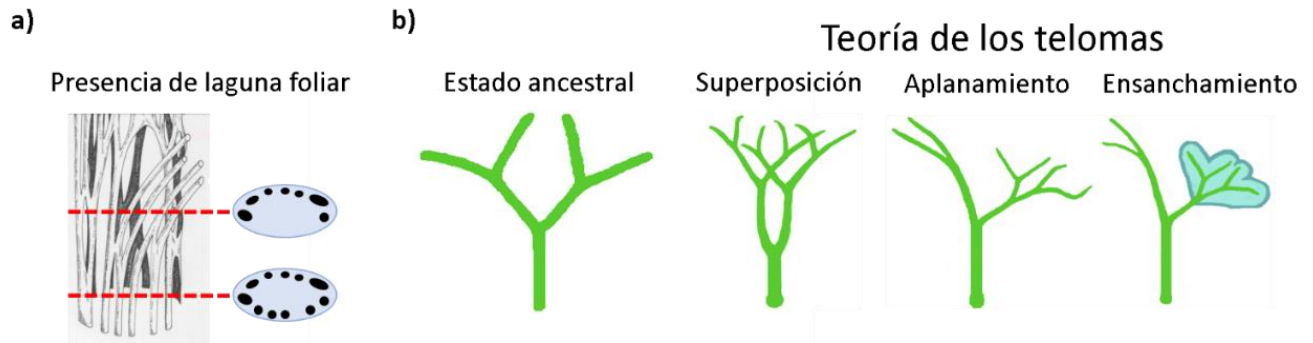


Figura 2. Características y teoría de la evolución de las hojas megáfilas. a) Presencia de laguna foliar. b) Teoría de los telomas. Esta teoría involucra tres procesos: superposición, aplanamiento y ensanchamiento de las ramas. (Modificado de Beerling & Fleming 2007; Srivastava & Chandra 2009).

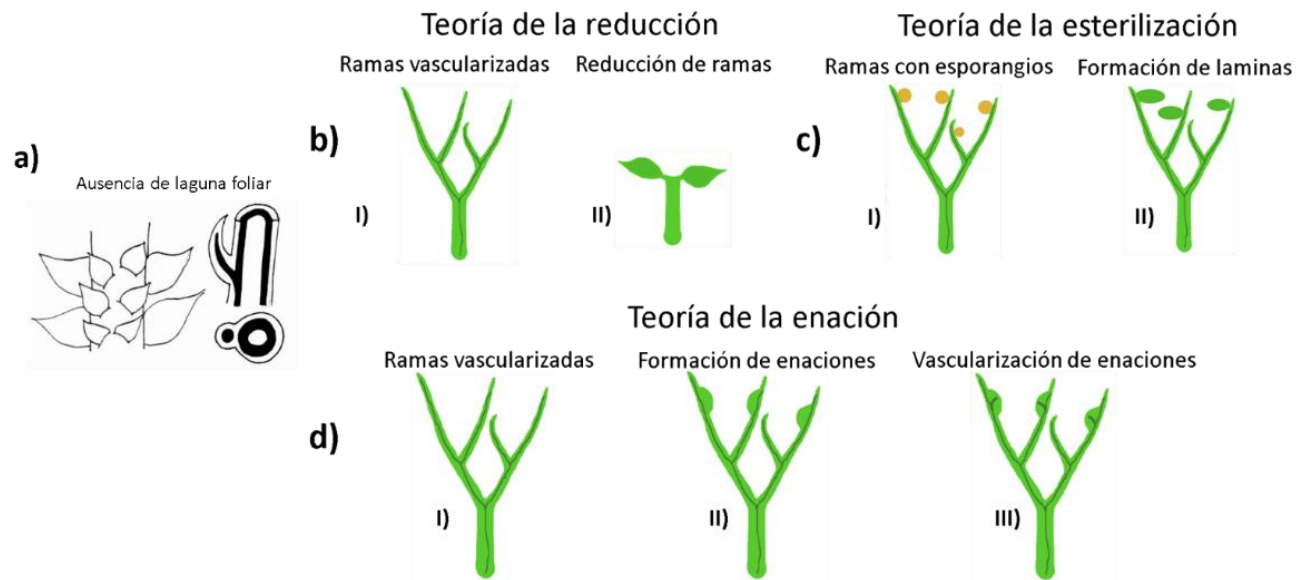


Figura 3. Características y teorías de la evolución de las hojas micrófilas. a) Ausencia de laguna foliar. b-d) Teorías propuestas para explicar la evolución de las hojas micrófilas. b) Teoría de la reducción, c) Teoría de la esterilización y d) Teoría de la enación. (Modificado de Beerling & Fleming 2007).

Desarrollo de hojas en helechos

Las hojas son órganos laterales que se desarrollan a partir de un grupo de células en los flancos del meristemo apical aéreo o SAM (Shoot Apical Meristem, por sus siglas en inglés), se acomodan de manera ordenada y específica alrededor del tallo (filotaxia) y tienen crecimiento determinado (Vasco et al., 2013). Contrario a lo ocurrido en las plantas con semilla, en donde por lo general el desarrollo de las hojas se da desde el ápice hacia la base, las hojas de los helechos tienen un desarrollo acropétalo (el ápice es la última parte que se diferencia). Este desarrollo acropétalo se debe a que sus hojas, a diferencia de las de angiospermas, poseen una célula apical que permanece indiferenciada hasta muy tarde en el desarrollo (Bower, 1952; revisado en Vasco et al., 2013).

Tanto las hojas de los helechos como las de las angiospermas poseen además un meristemo alrededor del primordio foliar denominado meristemo marginal. En angiospermas las hojas se forman principalmente por divisiones celulares intercalares, mientras que en los helechos las hojas se desarrollan principalmente a partir de divisiones en este meristemo marginal y de la célula apical inicial (Vasco et al., 2013). En las hojas compuestas de los helechos, cada una de las pinnas se desarrolla a partir de la división celular regulada del meristemo marginal y en las células submarginales de los primordios (Vasco et al., 2013; Wardlaw, 1963).

A pesar de la gran diversidad de formas que pueden tener las hojas de los helechos, la red de regulación genética que explica el desarrollo foliar ha sido poco estudiada (Byrne, 2012). Entender las diferencias y similitudes en los módulos genéticos de desarrollo foliar en los diferentes linajes de plantas vasculares, permitirá entender mejor el origen, la evolución y el desarrollo de las hojas. En los últimos 15 años se han empezado a realizar estudios comparativos de esta red genética en linajes de licofitas y helechos (Harrison et al., 2005; Plackett et al., 2015; Vasco et al., 2016).

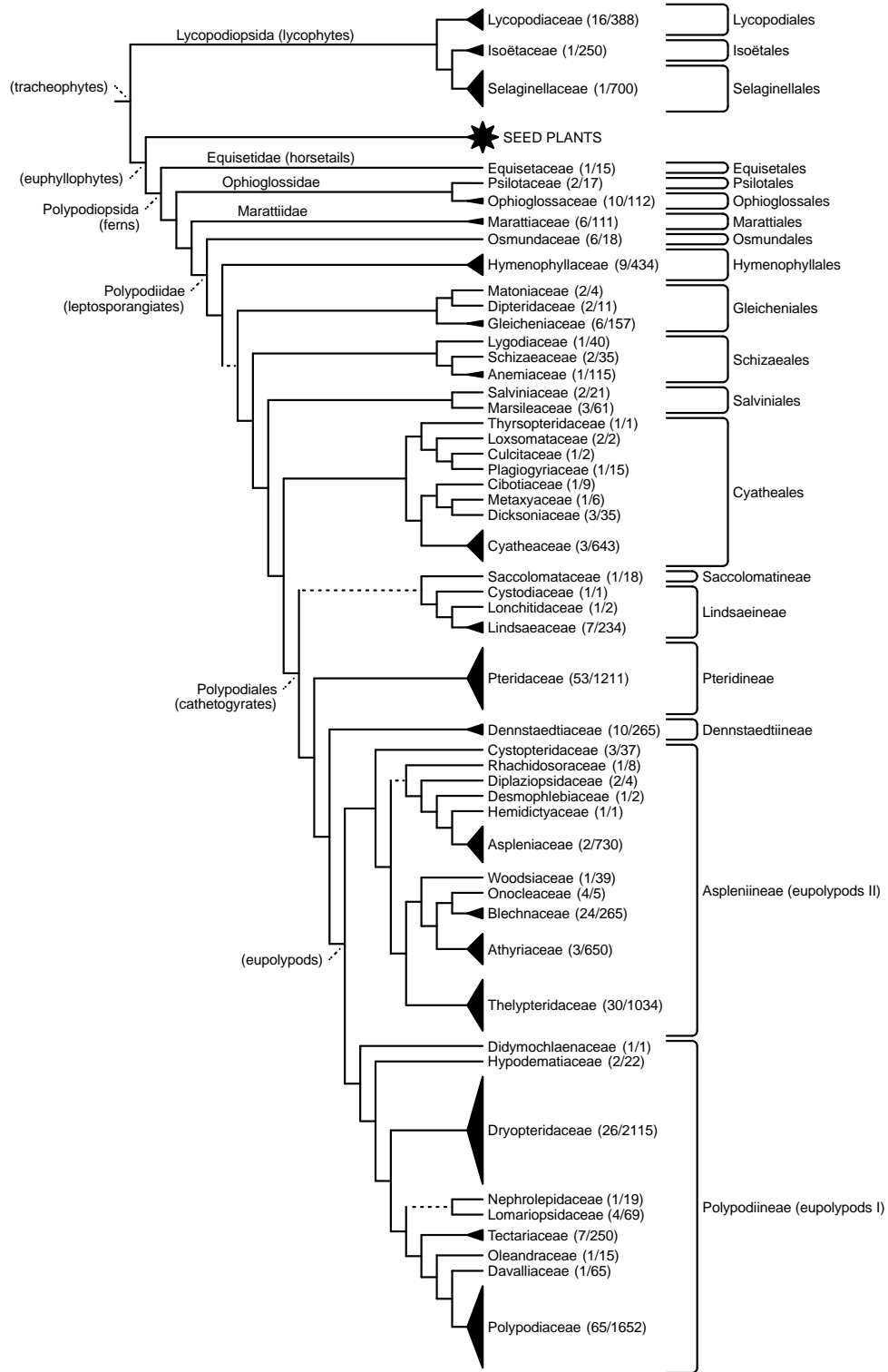


Figura 4. Filogenia consenso de las familias de los helechos y licofitas. Tomado de PPGI, 2016.

Red genética del desarrollo foliar en plantas vasculares

Establecimiento de los primordios foliares

La mayoría de los estudios acerca de la red genética de desarrollo foliar se han realizado en angiospermas, particularmente en *Arabidopsis thaliana*. A partir de estos estudios sabemos que la división constante y altamente regulada del SAM requiere de la actividad antagónica coordinada de la citoquinina, que mantiene en constante división al SAM, y de las proteínas KNOTTED1-like Homeobox (KNOX Clase I), que permiten a las células del SAM mantenerse en un estado indiferenciado (Kurakawa et al., 2007; Long et al., 1996) (Figura 5).

El paso de un órgano con crecimiento indeterminado, como el tallo, a uno con crecimiento determinado, como la hoja, requiere de la diferenciación de las células del SAM. En *Arabidopsis* la luz actúa como señal morfogénica sobre el SAM y a través de una cascada de transducción de señales que altera la expresión del gen *TRIPTOFANO AMINOTRANSFERASA DE ARABIDOPSIS1 (TAA1)*, un gen importante en la biosíntesis de auxinas (Byrne, 2012). La exposición del SAM a la luz también puede aumentar los niveles de citoquinina, ya sea al promover su biosíntesis o al regular negativamente su degradación (Carabelli et al., 2007; Yoshida, et al., 2011) (Figura 5).

Otro grupo importante de genes en el desarrollo de las hojas son el grupo de genes *ARP*. A este grupo se le dio ese nombre debido a los tres genes ortólogos que se han caracterizado en diferentes especies de angiospermas: *ASYMMETRIC LEAVES1 (Arabidopsis)*, *ROUGH SHEATH2 (Zea mays)* y *PHANTASTICA (Antirrhinum majus)* (Sun et al., 2002). Las proteínas *ASYMMETRIC LEAVES1*, *ROUGH SHEATH2* y *PHANTASTICA* reprimen a los genes *KNOX Clase I* para permitir el desarrollo de los primordios foliares a partir del SAM (Lodha et al., 2013), y participa en el establecimiento de la polaridad de la hoja (Sun et al., 2002) (Figura 5).

Las proteínas PIN promueven el establecimiento de máximos de la fitohormona auxina en los flancos del SAM (Reinhardt et al., 2003). Esta fitohormona promueve el inicio de la diferenciación celular, a través de diferentes mecanismos de activación y represión de las vías de señalización y de regulación de hormonas. Por ejemplo, en el caso de los genes *ARR7*, *ARR15*, importantes reguladores negativos de la señalización de las citoquininas, las auxinas actúan regulando negativamente a estos genes, por medio del factor de transcripción *AUXIN RESPONSE FACTOR5/ MONOPTEROS (MP)* (Zhao et al., 2010). Las auxinas también funcionan como reguladores negativos de los genes *KNOX Clase I* (Scanlon, 2003). El gen *PIN-FORMED1 (PIN1)* es regulado por las proteínas *PLETHORA (PLT3, PLT5 y PLT7)*, las cuales tienen actividad redundante y también están involucrados en el establecimiento de la filotaxia (Prasad et al., 2011) (Figura 5).

Polaridad de la hoja

La polaridad de la hoja se da mediante el establecimiento de tres ejes (proximal-distal, medio-lateral y adaxial-abaxial), esta regionalización permite que exista una especialización de las actividades que va a llevar a cabo cada célula (revisado en Byrne et al., 2000 y Byrne, 2012). En las hojas de angiospermas, se ha observado que las células de cada una de las superficies tienen forma, anatomía y función diferentes (Byrne et al., 2000; Semiarti et al., 2001): la superficie adaxial se especializa en la captación de luz y todo el proceso fotosintético subsecuente, mientras que la superficie abaxial se encarga del intercambio de gases (Bowman et al., 2002).

El establecimiento de la superficie adaxial se establece por los factores de transcripción Homeodomain-Leucine-Zipper de Clase III (HD-ZIP Clase III) y los RNA de interferencia derivados del gen *TAS3*, los cuales actúan para reprimir la expresión de los *FACTORES DE RESPUESTA A AUXINA* (*ARF3* y *ARF4*) (Byrne, 2012; Kidner & Timmermans, 2010). Los factores de transcripción KANADI y microRNAs, miRNA 165/166 reprimen la expresión de los factores de transcripción HD-ZIP Clase III, y junto a las proteínas ARF3 y ARF4 determinan la superficie abaxial de la hoja (Byrne, 2012; Kidner & Timmermans, 2010). Los genes de la familia *YABBY* también son importantes para la abaxialización de las hojas: *YABBY-FILAMENTOUS FLOWER* (*FIL*), *YABBY2* (*YAB2*), *YAB3* y *YAB5*. Estos genes también permiten delimitar el margen y el crecimiento de la lámina foliar (Byrne, 2012; Siegfried et al., 1999).

La superficie adaxial se establece entre otros, por la proteína AS1 (*ASYMMETRIC LEAVES1*) perteneciente al grupo de genes *ARP*, y por AS2 (*ASYMMETRIC LEAVES2*) que pertenece a su vez a la familia de genes *LOB* (*LATERAL ORGAN BOUNDARIES*) (Iwakawa et al., 2007; Kidner & Timmermans, 2010). *JLO* (*JAGGED LATERAL ORGAN*) permite que AS1 y AS2 interactúen físicamente y lleven a cabo su función (Majer & Hochholdinger, 2011). Las interacciones heterodiméricas entre las proteínas de AS1, AS2 y *JLO* permiten también la regulación negativa con los genes *KNOX Clase I* (Grimplet et al., 2017; Rast & Simon, 2012) (Figura 5). En este proyecto nos enfocaremos al estudio del gen *ASI*.

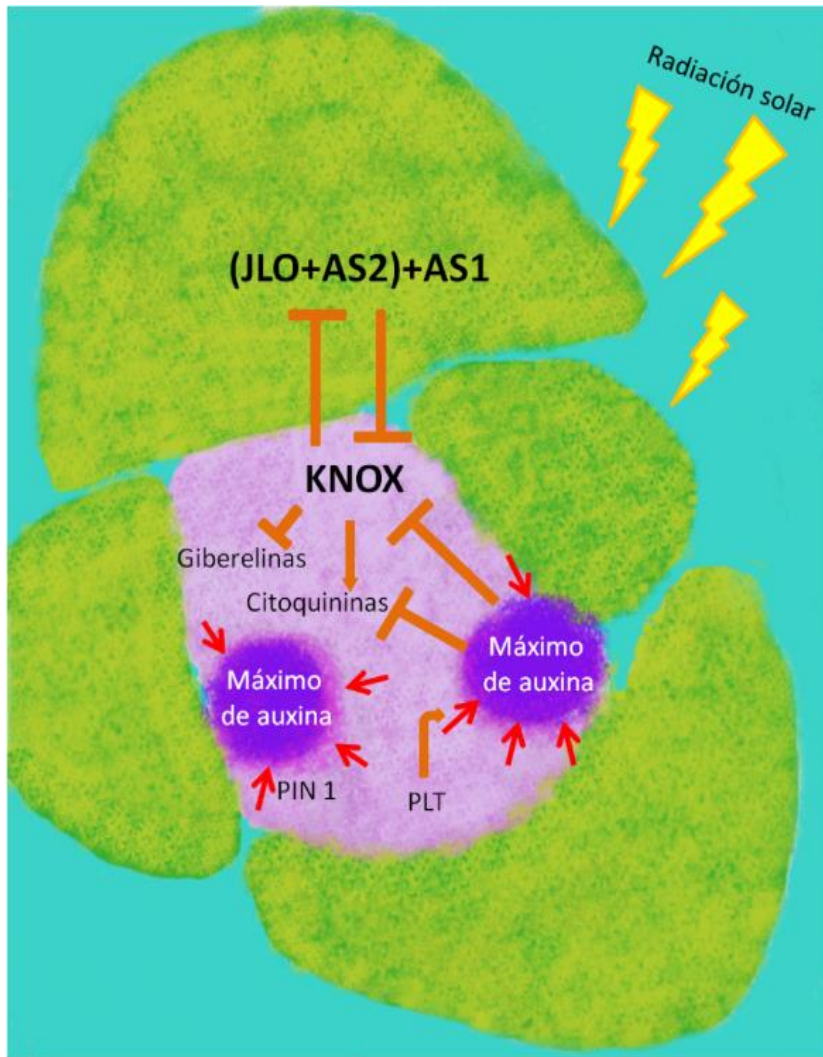


Figura 5. Modelo de la red génica de desarrollo foliar en angiospermas. La citoquinina está involucrada en una vía de señalización dependiente de luz. Esta fitohormona está encargada de mantener la división de las células del meristemo. El transporte de auxinas por medio de las proteínas PIN1 da lugar a máximos de auxinas; los genes *PIN1* están regulados por las proteínas PLT. La auxina inhibe a la citoquinina y mediante una red de transducción de señales, también regula negativamente a los genes *KNOX Clase I*, los cuales se encargan de mantener a las células del meristemo indiferenciadas. A su vez *KNOX Clase I* regulan de manera positiva a la citoquinina y de manera negativa a las giberelinas y a los genes *ASI* y *AS2*. La proteína *AS2* forma un dímero con el gen *JLO*, el cual recluta a *AS1*; esta proteína trimérica inhibe la expresión de los genes *KNOX Clase I*. Modificado de Byrne 2012.

El gen *ASI* (*ASYMMETRIC LEAVES1*) del grupo ARP

El gen *ASI* (*ASYMMETRIC LEAVES1*), descrito en la planta modelo *Arabidopsis*, pertenece al grupo de genes denominado *ARP*. *ASI* y sus ortólogos *PHAN* (*Antirrhinum majus*) y *ROUGH SHEATH2* (*Zea mays*) codifican para factores de transcripción que cuentan con un dominio MYB (Stracke, et al., 2001) (Figura 6). Se ha determinado que *ASI* está involucrada en la diferenciación de las células del meristemo apical que dan origen a las hojas en angiospermas (Byrne, 2012; Sun et al., 2002). También se ha reportado que *ASI* se encarga de establecer la polaridad en las hojas durante su desarrollo, mediante el establecimiento de los ejes próximo-distal y medio-lateral (Byrne et al., 2000; Sun et al., 2002).

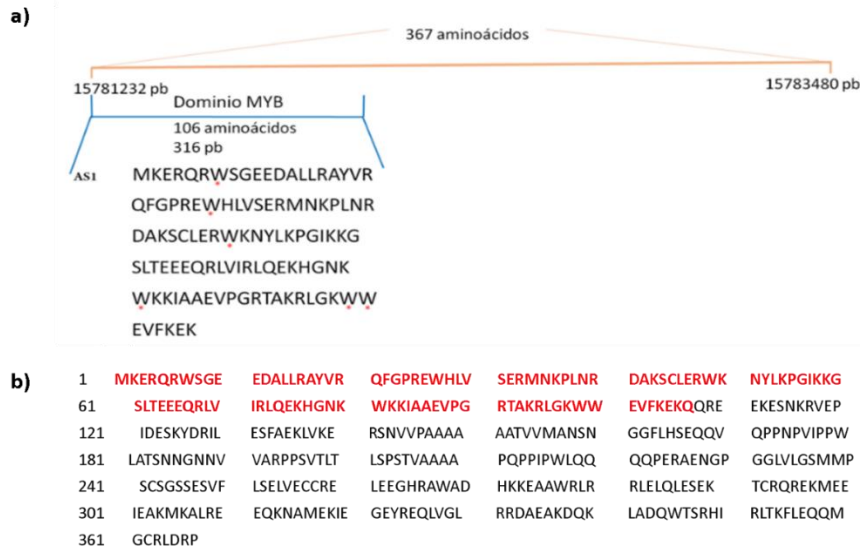


Figura 6. Representación de la región codificante del gen *AS1* en *Arabidopsis*. a) La línea anaranjada muestra la ubicación del gen *AS1* dentro del cromosoma 2. La línea azul representa la región que codifica para el dominio MYB. También se presenta la secuencia y posición de los aminoácidos dentro de este dominio. El dominio MYB cuenta con las repeticiones R2 y R3. La repetición R2 está compuesta por tres triptófanos conservados mientras que en la repetición R3 el primer triptófanos es sustituido por una leucina (posición 182), marcados con (*). Modificado de Sun *et al.*, 2002. b) Se muestra la secuencia de residuos de aminoácidos que conforma a la proteína *AS1* en *Arabidopsis*; se marca en rojo la región empleada para realizar los alineamientos utilizados en la construcción de las filogenias.

El dominio MYB

La familia de los genes MYB es una de las familias más grandes de genes que codifican para factores de transcripción en plantas (Stracke *et al.*, 2001). Por ejemplo, el genoma de arroz cuenta con 88 genes *MYB* y para *Arabidopsis* se estiman de 125 a 138 (Katiyar *et al.*, 2012; Stracke *et al.*, 2001).

El dominio MYB comprende tres motivos o repeticiones (R1, R2 y R3) que funcionan como dominios de unión al DNA (Du *et al.*, 2012; Myrset *et al.*, 1993). En general los genes que codifican para este dominio están relacionados con el control del ciclo celular en animales, plantas y la mayoría de los organismos eucariontes (Stracke *et al.*, 2001). En plantas, el dominio MYB más común es de tipo R2R3 (Stracke *et al.*, 2001), sin embargo, existen algunas proteínas en *Arabidopsis* que conservan la repetición R1 (Dubos *et al.*, 2010; Stracke *et al.*, 2001).

En las proteínas MYB de tipo R2R3, cada repetición R del dominio, forma tres alfa hélices y está formada por aproximadamente 53 aminoácidos característicos, entre ellos tres triptófanos que se encuentran altamente conservados (Stracke *et al.*, 2001). Un alineamiento realizado en *Arabidopsis* con secuencias que contaban con el dominio MYB, permitió establecer una secuencia consenso en la que se determinó que en la tercera repetición el primer triptófanos es reemplazado por una fenilalanina (Stracke *et al.*, 2001) (Figura 7). La segunda y tercera hélices de cada motivo constituyen una estructura

tridimensional de hélice-giro-hélice (Helix-Turn-Helix; HTH) (Figura 8). Cuando la proteína adquiere la conformación HTH, los residuos de triptófano quedan juntos conformando un núcleo hidrofóbico. A la tercera hélice de cada repetición se le denomina “hélice de reconocimiento” y su función principal es establecer contacto directo con el DNA (Dubos et al., 2010; Jin & Martin, 1999; Ogata et al., 1992).

Las proteínas de la familia MYB-R2R3 tienen diversas funciones durante el desarrollo de la planta; participan en la síntesis de metabolitos y el control de identidad celular y destino de las células meristemáticas (Borevitz et al., 2000; Bowman et al., 2017; Jin et al., 2000; Oppenheimer et al., 1991) (Cuadro 1).

R2	L	K	K	G	P	W	T	P	E	E	D	E	K	L	I	N	I	L	K	H	G	E	G	N	W	R	S	L	P	K	K	A	G	L	K	R	C	G	K	S	C	R	L	R	W	T	N	Y	L	R	P	D	
%	36	48	74	96	55	100	62	44	81	99	99	22	48	97	30	23	57	61	14	30	49	98	28	57	38	99	39	33	40	58	53	23	53	75	80	21	97	63	88	97	93	99	96	90	99	99	22	98	76	98	66	97	54
R3	I	K	R	G	N	F	T	E	E	E	E	E	L	I	I	E	L	H	A	L	L	G	N	R	*	W	*	S	K	I	A	K	R	L	P	G	R	T	D	N	E	I	K	N	Y	W	N	T	H	L	K	K	K
%	48	74	73	79	28	69	53	23	55	99	67	21	35	70	50	24	73	87	41	32	52	99	91	53	96	*	69	23	92	94	21	22	82	87	88	96	95	91	97	60	77	96	96	52	97	71	66	46	49	57	70	57	

Figura 7. Secuencia consenso de los motivos R2 y R3 de una proteína MYB en *Arabidopsis*. Los tres triptófanos (W) están altamente conservados en el motivo R2. En el motivo R3 una fenilalanina (F) sustituye al primer triptófano del motivo. En el alineamiento de los motivos R2 y R3, el número debajo de los aminoácidos corresponde al porcentaje de veces que el aminoácido ocupa esa posición en los diferentes MYB de *Arabidopsis*. Modificado de Stracke *et al.*, 2001.

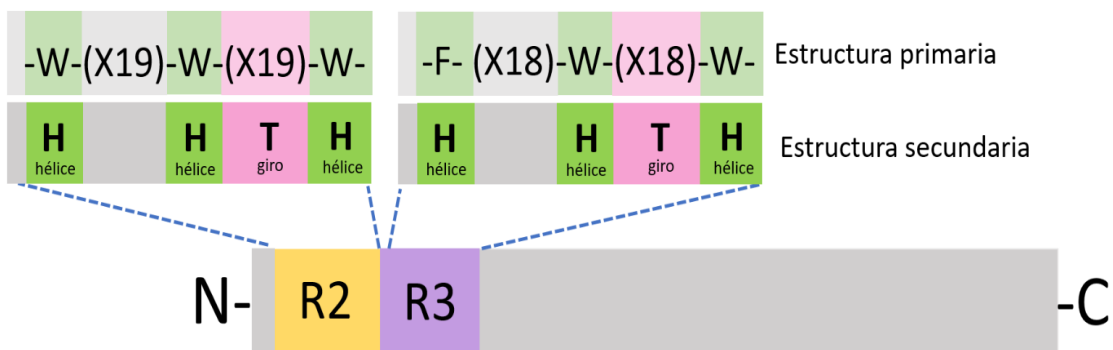


Figura 8. Estructura del dominio MYB de tipo R2R3. Cada repetición adquiere una estructura tridimensional con tres hélices, únicamente entre las dos últimas hélices de cada repetición se forma un giro. La tercera hélice de cada motivo es importante para su unión con el DNA. Modificado de Dubos *et al.*, 2010.

Interacción antagonica entre ARP y KNOX Clase I

La interacción antagonica entre las proteínas KNOX Clase I y ARP es de suma importancia para la formación de los primordios foliares (Byrne, 2012; Langdale & Harrison, 2008). Las proteínas ARP se encargan de regular negativamente a los genes *KNOX Clase I*, lo cual es importante para que las células del SAM pueden comenzar a diferenciarse en células foliares (Langdale, 2008).

Los genes *KNOX* (*KNOTTED HOMEODOMAIN*) pueden clasificarse en tres grupos: *Clase I*, *Clase II* y *KNATM*, en función de la similitud de secuencia dentro del homeodominio, posición de los intrones y patrón de expresión (Gao et al., 2015). Las proteínas KNOX Clase I se expresan principalmente en el SAM y se encargan de mantener al grupo de células que lo conforman en constante división y en un estado indiferenciado (Gao et al., 2015).

En angiospermas, los genes *KNOX Clase I* comprenden a los genes *STM* (*SHOOTMERISTEMLESS*), *KNAT1/BP* (*BREVIPEDECELLUS*), *KNAT2* y *KNAT6* (Gao et al., 2015). En *Arabidopsis*, se ha observado que el producto de estos genes se encargan de promover la proliferación de las células meristemáticas a través de la biosíntesis de citoquininas mediante la represión de las giberelinas y de factores de transcripción como los genes *AS1* y *AS2* (Byrne et al., 2002; Jasinski et al., 2005; Semiarti et al., 2001).

Se ha demostrado que los dobles mutantes de *as1* y *as2* tienen fenotipos similares a las mutantes sencillas de cada uno de los genes, lo que sugiere que éstos están involucrados en una misma vía de desarrollo (Sun et al., 2012). Las proteínas AS1 y AS2 forman un complejo represor que se une a las regiones reguladoras CWGTTD y KMKTGGAHW de los promotores de los genes *BP* y *KNAT2* formando un bucle en el DNA; en este complejo también está involucrado el factor de remodelación de la cromatina HIRA (Guo et al., 2008).

Sin embargo, para que las proteínas AS1 y AS2 puedan interactuar entre sí, es necesaria la participación de otra proteína más, codificada por el gen *JLO* (*JAGGED LATERAL ORGANS*) (Majer & Hochholdinger, 2011). *JLO* se une a AS2, y cuando este complejo proteico se encuentra ensamblado, se recluta a su vez a AS1, para formar una proteína trimérica funcional (Majer & Hochholdinger, 2011; Rast & Simon, 2012; Xu, Luo, & Hochholdinger, 2016) (Figura 5).

Otro complejo proteico importante para la regulación de los genes *KNOX Clase I* son los complejos de compactación de cromatina (PRC). Se sabe que el complejo Polimorrepresivos, PRC2 actúa después de la iniciación de la hoja y permite que los reguladores de expresión genética sean expresados en la zona correcta y en un momento de desarrollo del organismo específico (Lodha et al., 2013).

El complejo que forma AS1 con AS2 mantiene interacciones físicas con el complejo PRC2. Estas interacciones permiten la unión de este último a los promotores de los genes *KNOX Clase I: STM, BP* y *KNAT2* (Lodha et al., 2013). Además de reprimir a los genes *KNOX Clase I*, el complejo AS1-AS2 también regula negativamente a *ARF3* y *KANADI2 (KAN2)*, que son determinantes abaxiales (Lodha et al., 2013; Xu & Shen, 2008).

Resulta interesante que los genes *KNOX Clase I* y *ARP* estén presentes en licofitas y angiospermas ya que son dos linajes que divergieron hace cerca de 430 ma (Pires & Dolan, 2012; Plackett et al., 2015) (Figura 1). En helechos, sabemos que hay genes *KNOX Clase I* que se expresan en el SAM (Ambrose & Vasco, 2016; Harrison et al., 2005; Ruiz-Estévez et al., 2017; Sano et al., 2005), pero aún no se sabe si los genes *ARP* están presentes en este linaje.

El gen *AS2* (*ASYMMETRIC LEAVES2*)

El gen *ASYMMETRIC LEAVES2* (*AS2*), codifica para un dominio LOB, el cual pertenece a la familia de genes *LATERAL ORGAN BOUNDARIES DOMAIN* (*LBD*). La región N-terminal de *AS2* se caracteriza por un motivo de cuatro cisteínas separadas por un número variable de residuos de aminoácidos no conservados y actúa como sitio de unión al DNA (Majer & Hochholdinger, 2011; Xu et al., 2016). Seguido de esta región se encuentra un residuo de glicina invariable (Majer & Hochholdinger, 2011; Xu et al., 2016). En la región C-terminal hay un zipper de leucina con cinco aminoácidos hidrófobos, que pueden ser valina, isoleucina y leucina, y que a su vez se encuentran separados por seis residuos de aminoácidos no conservados (Majer & Hochholdinger, 2011; Xu et al., 2016). Este motivo probablemente actúa como un dominio de dimerización de proteínas (Majer & Hochholdinger, 2011) (Figura 9).

Al igual que *AS1*, *AS2* está implicada en el desarrollo foliar. En *Arabidopsis* se ha sugerido que *AS2* es el responsable de la adaxialización de las células de la hoja (Iwakawa et al., 2007). De manera interesante, se ha reportado que en la especie de angiosperma *Celosia cristata* (Amaranthaceae), esta proteína está involucrada en la determinación de la polaridad abaxial (Sun et al., 2012). La función de *AS2* también se ha asociado con el metabolismo de antocianinas y nitrógeno (Majer & Hochholdinger, 2011).

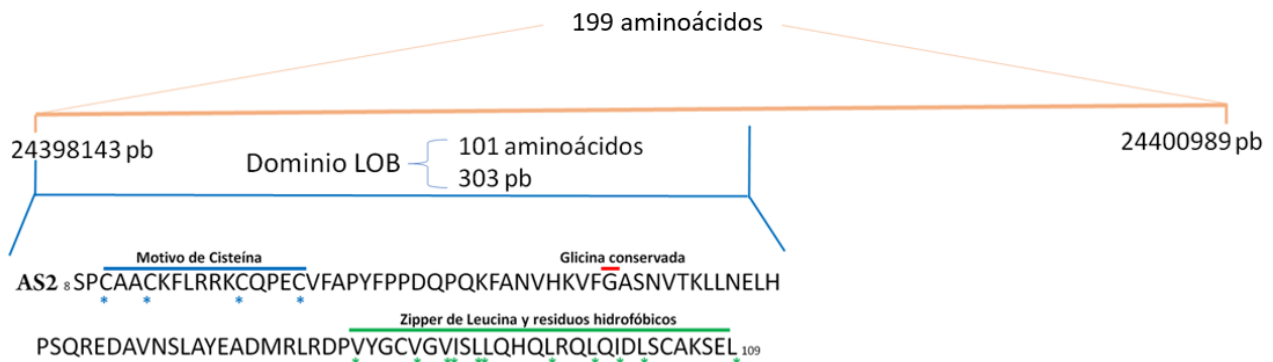


Figura 9. Representación de la región codificante del gen *AS2* y del dominio LOB en *Arabidopsis*. La línea anaranjada muestra la ubicación del gen *AS2* en el cromosoma 2. El dominio LOB se caracteriza por un motivo de cisteína (azul), una glicina conservada (rojo) y un motivo de cierre de leucina (verde). La región del cierre de leucina puede tener otros aminoácidos hidrofóbicos conservados, como valina e isoleucina. Modificado de Matsumura *et al.*, 2009.

Genética del desarrollo de hojas en helechos

El estudio del desarrollo de las hojas en los helechos (incluyendo grupos con morfologías foliares bastante peculiares como son las especies de Equisetales, Psilotales y Ophioglossales) y de las redes genéticas que lo controlan, son importantes para entender el origen y evolución de las hojas, tanto en el clado de los helechos, como en el linaje de las plantas vasculares. La genética del desarrollo de las hojas de los helechos ha sido menos estudiada que la de las angiospermas; sin embargo, recientemente se han realizado investigaciones enfocadas en entender la red genética que controla el desarrollo foliar en los helechos (Bharathan et al., 2002; Harrison et al., 2005).

Estas investigaciones han encontrado que los helechos y las angiospermas comparten algunos genes involucrados en el desarrollo de la hoja, como por ejemplo los factores de transcripción : *HD-ZIP Clase III* (Vasco et al., 2016) y *KNOX Clase I y II* (Bharathan et al., 2002; Harrison et al., 2005; Ruiz-Estévez et al., 2017; Sano et al., 2005). Los experimentos de expresión de estos dos factores de transcripción en helechos han mostrado que la expresión de estos tiene un patrón similar al reportado para angiospermas. Por ejemplo, el estudio de *HD-ZIP Clase III* en varias especies de helechos, incluyendo Equisetales y Psilotales, mostró que estos genes se expresan en la superficie adaxial de los primordios foliares en helechos, similar a lo encontrado en angiospermas (Vasco et al., 2016). También los experimentos de expresión de genes *KNOX Clase I* realizados en los helechos con hojas compuestas *Anogramma chaerophylla* (Bharathan et al., 2002), *Ceratopteris richardii* (Sano et al., 2005) y *Vandenboschia speciosa* (Ruiz-Estévez et al., 2017), mostraron que estos genes tienen un patrón de expresión similar al encontrado en angiospermas con hojas compuestas.

Justificación

Se ha propuesto que la regulación negativa entre las proteínas KNOX Clase I y ARP en plantas vasculares es la responsable de que se genere un grupo de células con destino determinado (hojas) a partir de células meristemáticas (SAM) (Lodha et al., 2013). Se han reportado genes ortólogos de *KNOX Clase I* y *ARP* en dos linajes de plantas vasculares lejanamente relacionados: licofitas y angiospermas (Harrison et al., 2005). Esto sugiere que, a pesar de que las microfílas y las megáfílas tienen un origen evolutivo diferente, podrían compartir algunos genes involucrados en regular su desarrollo. Actualmente, se sabe que los genes *KNOX Clase I* están presentes en algunos helechos como: *Anogramma chaerophylla*, *Ceratopteris richardii*, *Osmunda regalis* y *Elaphoglossum peltatum* (Ambrose & Vasco, 2016; Bharathan et al., 2002; Sano et al., 2005). Sin embargo, aún se desconoce si existen genes ortólogos a *ARP* en helechos y por lo tanto si el módulo genético *KNOX Clase I-ARP* está presente en este linaje de plantas.

Conocer si existen genes ortólogos de *ARP* en helechos es importante ya que es el primer acercamiento para establecer la evolución de este grupo de genes en las plantas vasculares. Inferir si estos genes se encontraban o no en el ancestro de las plantas vasculares nos permitirá entender mejor la evolución de las hojas a lo largo de la historia evolutiva de las plantas terrestres.

Preguntas de investigación

- ¿Existen genes ortólogos de *ARP* en helechos?
- ¿Cuál es la historia evolutiva de los genes *ARP* en plantas terrestres?
- ¿Qué nos puede decir la evolución de los genes *ARP* en plantas terrestres acerca de la evolución de las hojas megáfílas en eufilófitas?

Objetivos

Objetivo General

- Conocer la historia evolutiva de los genes *ARP* en plantas vasculares con énfasis en helechos.

Objetivos Particulares

- Determinar si existen genes ortólogos de *ARP* en helechos.
- Inferir la historia evolutiva de los genes *ARP* en plantas terrestres a través de una filogenia.
- A través de la filogenia de los genes *ARP* y de los genes *MYB* de plantas terrestres, generar hipótesis acerca de la evolución de los módulos genéticos de desarrollo de las hojas megáfilas en eufilofitas.
- Identificar los motivos característicos de los genes *ARP*.
- Inferir las posibles funciones de las secuencias de genes *MYB* en helechos a partir de una filogenia.
- Determinar y comparar modelos de conformación tridimensional de las proteínas *ARP* y las secuencias cercanamente relacionadas.

Hipótesis

Nula

Los genes *ARP* están presentes en helechos. Esto sugiere que los genes *ARP* y el módulo de desarrollo de hojas *KNOX Clase I-ARP* se encontraban en el ancestro común de las plantas vasculares, sugiriendo que existe un mecanismo común de desarrollo de hojas megáfilas (Figura 10).

Alternativas

Los genes *ARP* no están presentes en helechos, esto sugiere que: a) Los genes *ARP* se perdieron en los helechos. Indicando que existen diferentes mecanismos de desarrollo de hojas megáfilas (Figura 11 a); o b) Los genes *ARP* y el módulo de desarrollo de hojas *KNOX Clase I-ARP* no se encontraban en el ancestro común de las eufilofitas; por lo que los genes *ARP* se transfirieron horizontalmente de un linaje a otro.

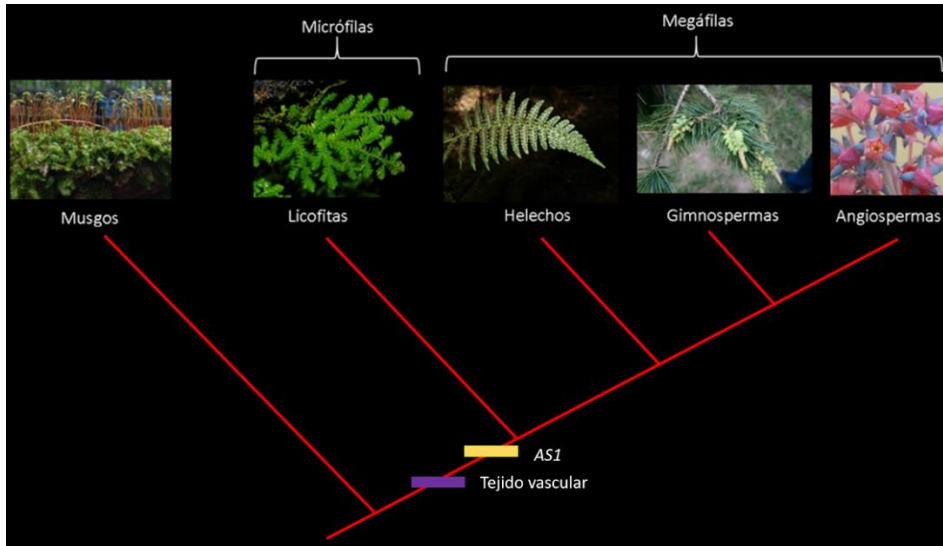


Figura 10. Hipótesis nula. Los genes *ARP* están presentes en helechos. Esto sugiere que los genes *ARP* y el módulo de desarrollo de hojas *KNOX Clase I-ARP* se encontraban en el ancestro común de las plantas vasculares, indicando que existe un mecanismo común de desarrollo de hojas megáfilas.

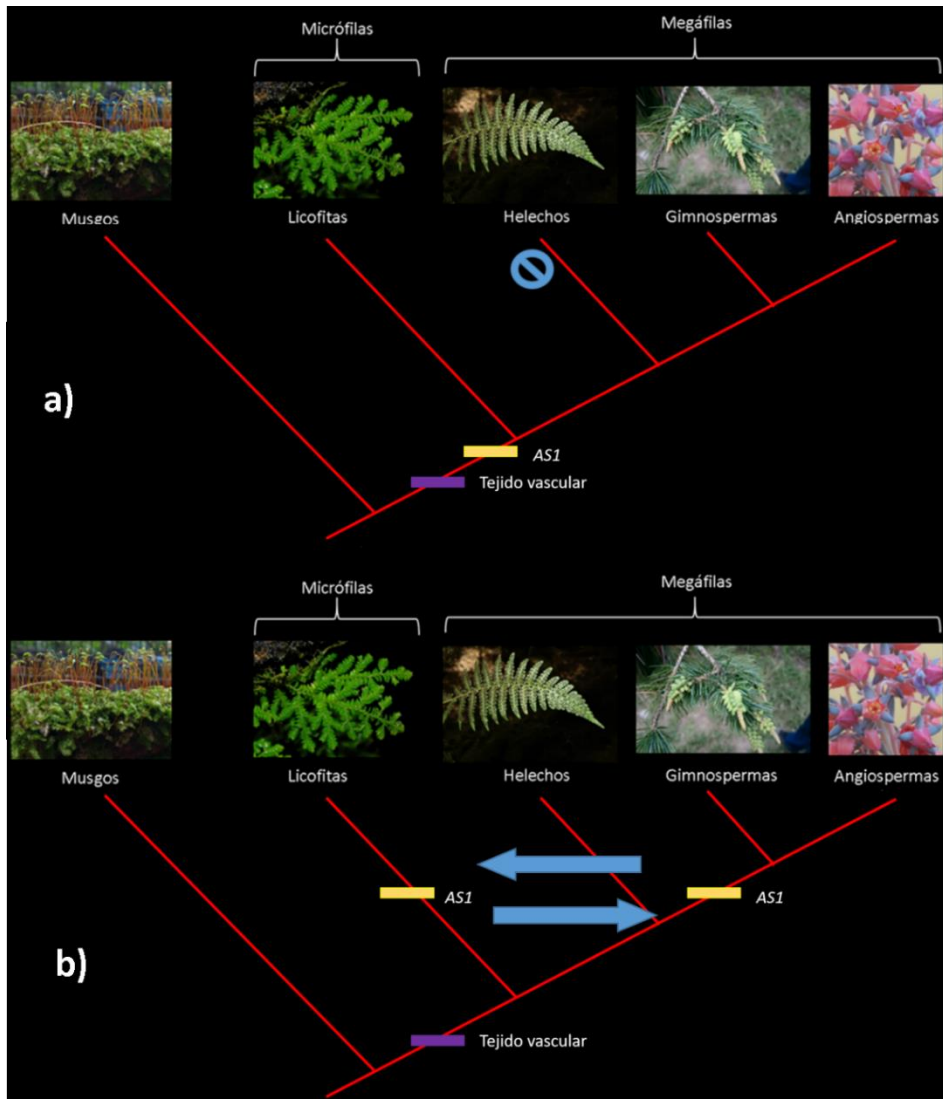


Figura 11. Hipótesis alternativas. Los genes *ARP* no están presentes en helechos, esto sugiere que: a) Los genes *ARP* se perdieron en los helechos. Indicando que existen diferentes mecanismos de desarrollo de hojas megáfilas; o b) Los genes *ARP* y el módulo de desarrollo de hojas *KNOX Clase I-ARP* no se encontraban en el ancestro común de las eufilofitas; por lo que los genes *ARP* se transfirieron horizontalmente de un linaje a otro.

Materiales y métodos

Etapa I: Análisis exploratorios, buscando los ortólogos de los genes *ARP* en helechos

a) Búsquedas de secuencias

Con la finalidad de reconstruir la filogenia de los genes *ARP* y explorar si existen genes ortólogos a *ASI* en los diferentes linajes de plantas terrestres (musgos, licofitas, helechos y gimnospermas), se realizaron diferentes búsquedas de secuencias por similitud mediante el algoritmo de BLAST (tBLASTx) (Altschul et al., 1990), con el gen *ASI* (at2g37630) de *Arabidopsis*. Estas búsquedas se realizaron en las siguientes bases de datos de genomas y transcriptomas disponibles en repositorios públicos: TAIR (Lamesch et al., 2012) (<https://www.arabidopsis.org/>), OneKP (N et al., 2015) (<https://db.cngb.org/onekp/home/>), Congenie (Nystedt et al., 2013) (<http://congenie.org/>), GenBank (Benson et al., 2017) (<https://www.ncbi.nlm.nih.gov/genbank/>), FernBase (Li et al., 2018) (<https://fernbase.org/>) Phytozome (Goodstein et al., 2012) (<https://phytozome.jgi.doe.gov/pz/portal.html>). Además, se realizaron búsquedas en el genoma de *Ceratopteris richardii* (Barbara Ambrose. Datos aún no publicados). El criterio de selección de secuencias se fijó en un valor de E de menor a 1×10^{-10} . A continuación, se mencionan las características de cada una de las búsquedas.

Angiospermas: Las secuencias de angiospermas incluidas inicialmente en el análisis fueron: *ASYMMETRIC LEAVES1* (At2g37630) de *Arabidopsis thaliana*, *ROUGH SHEATH2* (NM_001112039.2) correspondiente a *Zea mays*, y *PHANTASTICA* (AJ005586.1) perteneciente a *Antirrhinum majus*. Estas secuencias fueron obtenidas de la base de datos de GenBank (Supl. 1).

Gimnospermas: Se incluyeron secuencias de tres especies de gimnospermas: *Ginkgo biloba*, *Picea abies* y *Pinus taeda*. Las secuencias de *Ginkgo biloba* fueron obtenidas de la base de datos de transcriptomas de OneKP. Las secuencias de *Picea abies* y *Pinus taeda* fueron obtenidas de los genomas disponibles en Congenie (Supl. 1).

Helechos: Para obtener las secuencias de helechos similares a *ARP* se hizo una búsqueda en la base de datos de OneKP. La búsqueda incluyó al menos una especie que correspondiera a cada uno de los órdenes de helechos (PPG I, 2016) (Figura 4, Cuadro 2). De igual forma se realizaron búsquedas en los genomas de los helechos *Azolla filiculoides* y *Salvinia cuculata* disponibles en FernBase, así como en el genoma aún no publicado de *Ceratopteris richardii* (Supl. 1).

Licofitas: La búsqueda de genes ortólogos de *ARP* en licofitas se realizó en la base de datos de OneKP. Además de la secuencia del gen *ASI* (at2g37630), se utilizó la secuencia reportada para *Selaginella kraussiana* (*SkARP*) (Harrison et al., 2005). Con el gen *ASI* únicamente se realizaron búsquedas en: *Isoetes tegetiformans*, *Lycopodium annotinum* y *Selaginella kraussiana*. Con el gen *SkARP* se realizaron búsquedas en todas las especies de licofitas con transcriptoma disponible en OneKP y en el genoma de *Selaginella moellendorffii*, disponible en Phytozome. (Cuadro 3 y Supl. 1).

Musgos: Las secuencias de musgos se realizaron en el genoma de *Physcomitrella patens* disponible en Phytozome (Supl. 1).

b) Alineación de secuencias

Las secuencias recuperadas de Congenie, FernBase, GenBank, y Phytozome incluyen la secuencia codificante completa; por otro lado, las secuencias de OneKP, dado que son provenientes de transcriptomas, son generalmente secuencias codificantes parciales. En el caso de las secuencias provenientes de transcriptomas, solo fueron incluidas aquellas con dominios reconocibles completos. Posteriormente las secuencias fueron depuradas para eliminar las que no presentaran marco de lectura o fuesen muy cortas, según las características del alineamiento.

Los alineamientos se realizaron con la matriz de nucleótidos, en la plataforma de MAFFT (Kato et al., 2017), con una penalización de 3 para la apertura de Gap, valor de compensación de 1 y sin efecto en la puntuación de la alineación para los nucleótidos N. Posteriormente, los alineamientos se refinaron a mano con el programa de Mesquite v3.5 (Maddison, W. P. & D.R. Maddison. 2018) (<http://www.mesquiteproject.org>). Para alinear las secuencias se tomó como referencia la posición del dominio MYB en *Arabidopsis* y se verificó que todas las secuencias estuvieran leídas con la dirección 5'-3'. Finalmente se eliminaron las secuencias repetidas o con el dominio MYB incompleto.

c) Filogenia Exploratoria 1

Con la matriz alineada se realizó un análisis de Máxima Verosimilitud (MV) usando el software público RaxML BlackBox (Stamatakis, 2014), en la plataforma pública CIPRES (<https://www.phylo.org/>) (Miller et al., 2010). RaxML BlackBox realiza un análisis que permite obtener simultáneamente el árbol más verosímil y los soportes de las ramas (Stamatakis, 2014). Estos últimos se calculan a través de un análisis de Bootstrap rápido (rapid bootstrap).

La filogenia obtenida fue visualizada y editada con el software de FigTree v1.4.3 (Rambaut Research Group, n.d.) (<http://tree.bio.ed.ac.uk/software/figtree/>).

Etapa II: Ampliación del muestreo a la familia de genes *MYB*

Con base en los resultados de la Filogenia Exploratoria 1, se decidió ampliar las búsquedas de secuencias en los diferentes linajes de plantas vasculares para incluir a los genes *MYB*. Esta segunda búsqueda tuvo como objetivo entender mejor las relaciones filogenéticas entre las secuencias incluidas en la Filogenia Exploratoria 1 y entender las posibles funciones de las secuencias de helechos recuperadas. Las búsquedas se realizaron con el algoritmo de BLAST (tBLASTx), esta vez se seleccionaron las secuencias con un valor de $e \leq 0$.

a) Búsquedas en angiospermas

Para realizar las búsquedas adicionales de genes *MYB* de *Arabidopsis* se seleccionaron algunas secuencias de helechos de los clados principales de la Filogenia Exploratoria 1, bajo el siguiente criterio: tamaño de la secuencia superior a 200 nucleótidos y pocos caracteres ambiguos. Las búsquedas se hicieron en la base de datos de TAIR y solo se seleccionaron secuencias que estuvieran anotadas como genes *MYB* con un valor de $e \leq 0$. Las nuevas secuencias obtenidas se compararon con dos filogenias de genes *MYB* previamente publicadas (Bowman et al., 2017; Stracke et al., 2001). Las secuencias que se encontraban en las filogenias publicadas, pero que no se recuperaron en nuestras búsquedas de BLAST, fueron añadidas a la matriz inicial. La nueva matriz se alineó y analizó utilizando los mismos métodos descritos en la Etapa I, incisos b y c. De este análisis se obtuvo la Filogenia Exploratoria 2.

b) Búsquedas en helechos, gimnospermas, licofitas y musgos

La Filogenia Exploratoria 2 se utilizó para detectar los clados en donde solo algunos de los linajes de plantas vasculares estuvieran representados. De estos clados se seleccionaron diferentes secuencias de *Arabidopsis* y con ellas se realizaron nuevas búsquedas bajo el algoritmo de tBLASTx en los diferentes linajes, como se especifica a continuación.

Gimnospermas: Las búsquedas se realizaron únicamente en las especies que contaban con genoma: *Picea abies* y *Pinus taeda*, de la base de datos de Congenie (<http://congenie.org/>) (Supl. 1).

Helechos: Se realizaron búsquedas en la base de datos de genomas FernBase (<https://fernbase.org/>) de las especies *Azolla filiculoides* y *Salvinia cuculata*. Adicionalmente, usando nuevamente el gen *ASI* (at2g37630), se hicieron búsquedas en todos los transcriptomas de helechos disponibles en la base de datos de OneKP (Supl. 1).

Licofitas y musgos: Se realizaron búsquedas en *Selaginella moellendorffii*, única especie de licofitas que cuenta con un genoma reportado y en el genoma del musgo *Physcomitrella patens*. Las secuencias de ambas especies se obtuvieron de la base de datos Phytozome (Supl. 1).

Todas las nuevas secuencias obtenidas fueron integradas a la matriz de la Filogenia Exploratoria 2, que fue punto de partida para la Etapa III, como se describe a continuación.

Etapa III: Historia evolutiva de los genes *MYB* y *ARP* con énfasis en helechos

a) División de los datos en dos matrices

Con las secuencias obtenidas en las Etapas I y II, se construyeron dos matrices, que se alinearon con los métodos detallados en la Etapa I. La primera matriz incluyó todas las secuencias recuperadas de genomas y de transcriptomas (matriz G+T). La segunda matriz incluyó únicamente las secuencias provenientes de genomas (matriz G) (Cuadro 4 y Supl. 1). Esto con la finalidad de reducir el número de terminales para que estas no excedieran el número de caracteres informativos, debido a que en ocasiones cuando el número de terminales supera el número de caracteres, las filogenias no están resueltas y los soportes de las ramas son malos.

b) Análisis de similitud

Con la finalidad de determinar el grado de similitud existente entre los genes *ARP*, las secuencias con dominio *MYB* de *Arabidopsis* y las secuencias afines encontradas en todas las búsquedas, se realizó un análisis de similitud con la matriz G. De esta manera también se definió al grupo externo en el que se enraizarían las filogenias obtenidas con los diferentes análisis (RaxML, IQtree y MrBayes). Este análisis fue realizado bajo el método de agrupamiento Neighbor joining, con la matriz G de aminoácidos en la plataforma de MAFFT (Katoh et al., 2017).

c) Análisis filogenéticos

Para entender la evolución de los genes *MYB* y *ARP* en el contexto de las plantas vasculares se realizaron análisis filogenéticos con las secuencias de nucleótidos de las matrices G y G+T, bajo dos enfoques estadísticos diferentes: Máxima Verosimilitud (MV) e Inferencia Bayesiana (IB).

Los análisis de MV se realizaron con dos softwares diferentes: IQtree (Hoang et al., 2007) y RaxML BlackBox (Stamatakis, 2014) en la plataforma Cipres (Miller, et al., 2010). RaxML BlackBox emplea el modelo GTR-CAT de forma predeterminada y realiza un análisis que permite obtener simultáneamente el árbol más verosímil y los soportes de las ramas (Stamatakis, 2014). Estos últimos se calculan a través de un análisis de bootstrap rápido (rapid bootstrap).

Con IQtree se empleó el modelo de evolución molecular GTR+G+I obtenido mediante un análisis de concientes de verosimilitud jerárquico (hLRT) en MrModeltest 2.4 (Nylander, 2004). Los soportes de ramas se calcularon a través de un análisis de bootstrap ultra rápido (Ultrafast bootstrap) (Hoang et al., 2007).

Los análisis de IB fueron realizados con el programa MrBayes (Ronquist et al., 2012) bajo el modelo de evolución molecular obtenido con con MrModeltest 2.4 (GTR+G+I). Los análisis se corrieron durante 10 millones de generaciones con cuatro cadenas de Markov (tres calientes y una fría), y un muestreo de la distribución de probabilidad posterior cada 1000 generaciones. Las muestras correspondientes a la fase inicial de las cadenas de Markov (25%) se descartaron. La aplicabilidad de este valor se determinó mediante la inspección de los “likelihood scores” con el software Tracer 1.5 (Drummond & Rambaut, 2015). Los árboles obtenidos, después de descartar el 25%, se combinaron para obtener un árbol de consenso de mayoría y las respectivas probabilidades posteriores (PP) de los nodos. Todos los árboles se visualizaron utilizando FigTree v1.4.0 (Rambaut Research Group).

Etapas IV: Motivos y modelos por homología de las proteínas ARP

a) Análisis de motivos proteicos

Para detectar motivos proteicos característicos de los genes *ARP* y de las secuencias cercanamente relacionadas a ellos de acuerdo con las filogenias obtenida en la Etapa III, se realizó un análisis de MEME (Multiple Em for Motif Elicitation) (<http://meme-suite.org/tools/meme>) (Bailey & Elkan, 1994). Para este análisis se utilizó el alineamiento de aminoácidos de la matriz G+T bajo los siguientes parámetros: modo clásico de descubrimiento de motivos, cero a una ocurrencia por secuencia, y un número máximo de 40 motivos.

El análisis de MEME permite determinar motivos característicos de un grupo de secuencias. Este análisis muestra la longitud del motivo localizado y el nivel de conservación de los aminoácidos que conforman el motivo dentro del grupo de secuencias que lo comparten, mediante un análisis de similitud de las secuencias traducidas a aminoácidos. El programa MEME arroja motivos usando modelamiento estadístico para elegir automáticamente las secuencias que ocurren repetidamente dentro de un grupo de secuencias relacionadas; teniendo en cuenta la longitud, número de ocurrencias y la descripción de cada motivo. El programa también arroja un valor de *e* que describe la probabilidad de encontrar un patrón similar en secuencias aleatorias.

b) Modelos por homología de las proteínas ARP y clado b

Con la finalidad de determinar la estructura proteica de los genes *ARP* y de los genes *MYB* cercanamente relacionados (de acuerdo con las filogenias obtenidas en la Etapa III), se realizó un análisis de estructura terciaria de las proteínas en la plataforma de SWISS-MODEL (Waterhouse et al., 2018) (<https://swissmodel.expasy.org/>). Este análisis consiste en realizar un BLAST con la secuencia

traducida a aminoácidos, para compararla con otras secuencias depositadas en el repositorio de SWISS-MODEL. El software construye un modelo con base en la proteína que se haya seleccionado de acuerdo a los mejores valores de identidad (GMQE y QSQE), y finalmente el modelo se evalúa mediante la función de puntuación QMEAN, la cual debe de ser QMEAN <1, para poder considerar que el modelo representa a la estructura terciaria de la proteína en cuestión.

Para la construcción del modelo tridimensional de las proteínas de los clados ARP y b (de acuerdo con la filogenia G+T obtenida de RaxML; Figura 14) se usó como plantilla la estructura **3zqc.3.A** correspondiente a la proteína MYB3 de *Arabidopsis* reportada en esta misma plataforma. Para la realización de los modelos se utilizaron las secuencias completas de residuos de aminoácidos de las proteínas AS1, PHAN, RH2 y SkARP, mientras que para las secuencias ortólogas y las agrupadas en el clado b únicamente se tomaron en cuenta los aminoácidos que se encontraban a partir de la primera metionina hasta el primer codón de termino (Figura 12).

Resultados

Etapas I y II: Búsqueda exploratoria de secuencias y ampliación del muestreo

La Filogenia Exploratoria 1, nos permitió visualizar y entender el nivel y cobertura de las búsquedas realizadas. En esta filogenia, ninguna de las secuencias de helechos incluidas se recuperó dentro del clado que contenía las secuencias previamente publicadas de *ARP* en angiospermas, gimnospermas y licofitas (Harrison et al., 2005). La Filogenia Exploratoria 2, incluyó nuevas secuencias de genes *MYB* de todos los linajes de plantas vasculares, principalmente de *Arabidopsis*, con el objetivo de entender mejor la relación filogenética de las secuencias de la Filogenia Exploratoria 1 y entender la posible función de las secuencias de helechos, su relación con los genes *MYB* de *Arabidopsis* y con ello sus posibles funciones.

A partir de las búsquedas realizadas en las Etapas I y II y tras una depuración que incluyó la eliminación de secuencias cortas, con el dominio incompleto o duplicadas, se obtuvo un total de 405 secuencias, de las cuales 30 pertenecen a musgos (todas del genoma de *Physcomitrella patens*), 49 a licofitas (28 del genoma de *Selaginella moellendorffii* y 21 de transcriptomas de otras licofitas); 151 a helechos (99 de los genomas de *Azolla filiculoides*, *Ceratopteris richardii* y *Salvinia cucullata* y 52 de transcriptomas de otras especies de helechos); 68 a gimnospermas (66 de los genomas de *Picea abies* y *Pinus taeda* y dos del transcriptoma de *Ginkgo biloba*) y 107 a angiospermas (una de *Antirrhinum majus*, una *Zea mays* y 105 de *Arabidopsis*) (Cuadro 4 y Supl. 1).

Etapa III: Filogenias finales, historia evolutiva de los genes *MYB* y *ARP* con énfasis en helechos

a) Alineamientos

El alineamiento de la matriz G+T cuenta con 405 terminales, la región alineada corresponde a 297 nucleótidos (99 residuos de aminoácidos) del dominio MYB. Nuestros alineamientos muestran que, además de la fenilalanina, el primer triptófano de la repetición R3, posición 82 (Figura 12) es reemplazado ya sea por leucina, isoleucina, fenilalanina o, en menor medida, por valina y metionina, aunque, en algunos casos se conserva el triptófano, todos estos aminoácidos son hidrofóbicos y no parece existir un reemplazo característico por linaje.

En el alineamiento de aminoácidos de la matriz G+T (Figura 12, secuencias 1-14), se puede notar que las secuencias correspondientes al clado ARP tienen únicamente un residuo de leucina en la posición 54 en lugar del primer triptófano de la repetición R3. Además, las secuencias del clado ARP se diferencian notoriamente del resto de las secuencias del alineamiento G+T porque presentan un residuo de triptófano extra en la posición 91.

Las secuencias 15-23, que corresponden a las agrupadas en el clado b (Figuras 11 y 15) también muestran diferencias significativas en comparación con las secuencias del clado ARP y las que conforman el resto de la matriz. Por ejemplo, la presencia de dos serinas continuas en las posiciones 28 y 29, dos argininas continuas en las posiciones 31 y 32 y dos valinas consecutivas en las posiciones 62 y 63, excepto en la secuencia 15 correspondiente a la proteína MYB125 de *Arabidopsis*.

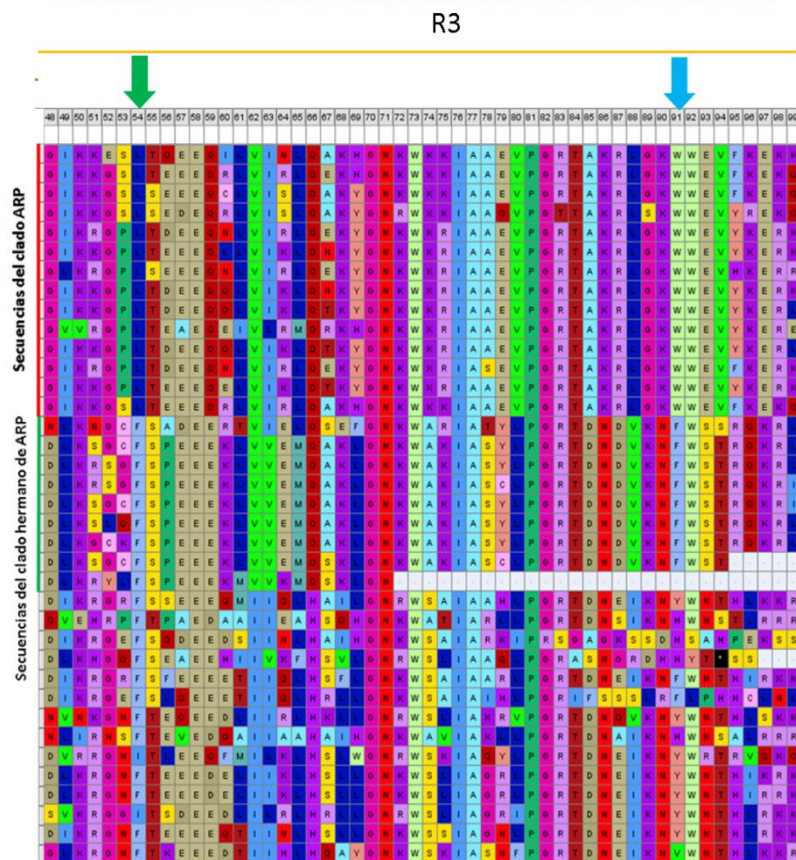
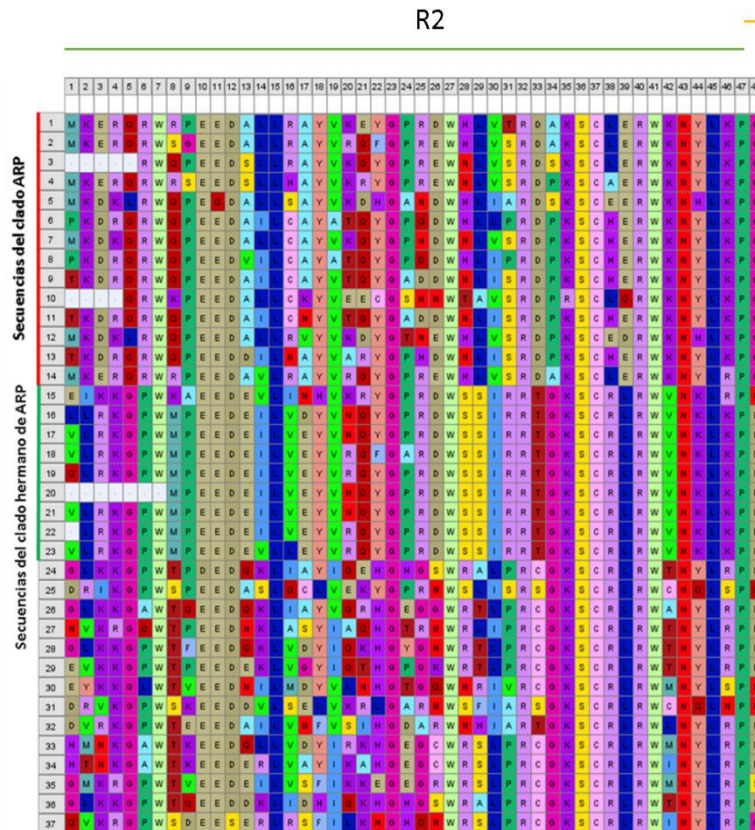


Figura 12. Alineamiento de la matriz G+T (aminoácidos). Únicamente se muestran las secuencias del clado ARP, clado b y otras secuencias de genes *MYB* de *Arabidopsis*. La región del dominio MYB empleada para el alineamiento cuenta con 99 aminoácidos y 405 terminales (en esta figura únicamente se muestran 37 secuencias). De acuerdo con Stracke *et al.*, 2001, en este alineamiento la repetición R2 abarca las posiciones 1- 48, mientras que la repetición R3 las posiciones 49-99, aproximadamente. Las primeras 14 secuencias de este alineamiento se recuperan en el clado ARP de los análisis de Máxima Verosimilitud e Inferencia Bayesiana realizados con la matriz G+T (Figura 16). Se puede observar que estas secuencias tienen un triptófano extra en la posición 91; en la posición 54, es en donde se supone que está el primer triptófano de la repetición R3, hay una leucina; en el resto de las secuencias el aminoácido que sustituye al triptófano es una fenilalanina o isoleucina. Las secuencias 15-23 corresponden a las agrupadas en el clado b. Se pueden observar ciertas diferencias entre las secuencias del clado ARP y las secuencias del clado b, por ejemplo: dos serinas continuas en las posiciones 28 y 29, dos argininas continuas en las posiciones 31 y 32 y dos valinas consecutivas en las posiciones 62 y 63, excepto en la secuencia 15 correspondiente a la proteína MYB125 de *Arabidopsis*.

b) Análisis de similitud

El análisis de similitud arrojó como resultado que las secuencias *Selaginella_moe_113921* y *Selaginella_moe_179572* son las que más difieren del resto de secuencias con dominio MYB. Por esta razón, ambas secuencias fueron utilizadas como grupo externo en los análisis filogenéticos posteriores. Cabe mencionar que el uso de estas secuencias como grupo externo no significa que su divergencia haya sido anterior a las del resto de las secuencias (material suplementario VI).

También logramos determinar que las secuencias de los genes *ARP* se agruparon todas juntas y es ese grupo el más diferente a todas las otras secuencias, sin tomar en cuenta a las secuencias *Selaginella_moe_113921* y *Selaginella_moe_179572*, usadas para enraizar el árbol filogenético.

El análisis de similitud mostró que las secuencias incluidas en la matriz G se dividen en dos grupos principales (Supl. 6).

c) Análisis filogenéticos con la Matriz G

Con la matriz G, que incluye únicamente secuencias provenientes de genomas, se realizó un análisis de Inferencia Bayesiana (IB) y dos análisis filogenéticos de Máxima Verosimilitud (MV) con diferentes programas: RaxML en BlackBox y IQtree. Las tres filogenias obtenidas a partir de estos análisis recuperaron 22 grandes clados con soportes de moderados a buenos. Estos clados se conforman por las mismas secuencias en todos los análisis, sin embargo, las relaciones entre clados se modifican dependiendo del método (Figura 13 y Supl. 3 y 5). Para la descripción de resultados nos acotaremos a la filogenia obtenida con los análisis de IB.

De los 22 clados identificados, los clados b, c, o, p, q, r agrupan secuencias de los diferentes linajes de plantas terrestres. El clado ñ agrupa únicamente a secuencias de plantas con semilla. Los clados f, h, i, k, l, están compuestos exclusivamente de secuencias de *Arabidopsis*. En la Figura 13, se indican las funciones reportadas para los diferentes genes *MYB* de *Arabidopsis* agrupadas en cada uno de los clados. En general, todas las funciones anotadas para los genes *MYB* de *Arabidopsis*, se relacionan con la respuesta a algún tipo de estrés, o con la síntesis de diferentes metabolitos, polímeros y fitohormonas. Únicamente los clados i, o y u cuentan con genes de *Arabidopsis* que se relacionan con la diferenciación de células meristemáticas o el desarrollo de órganos laterales.

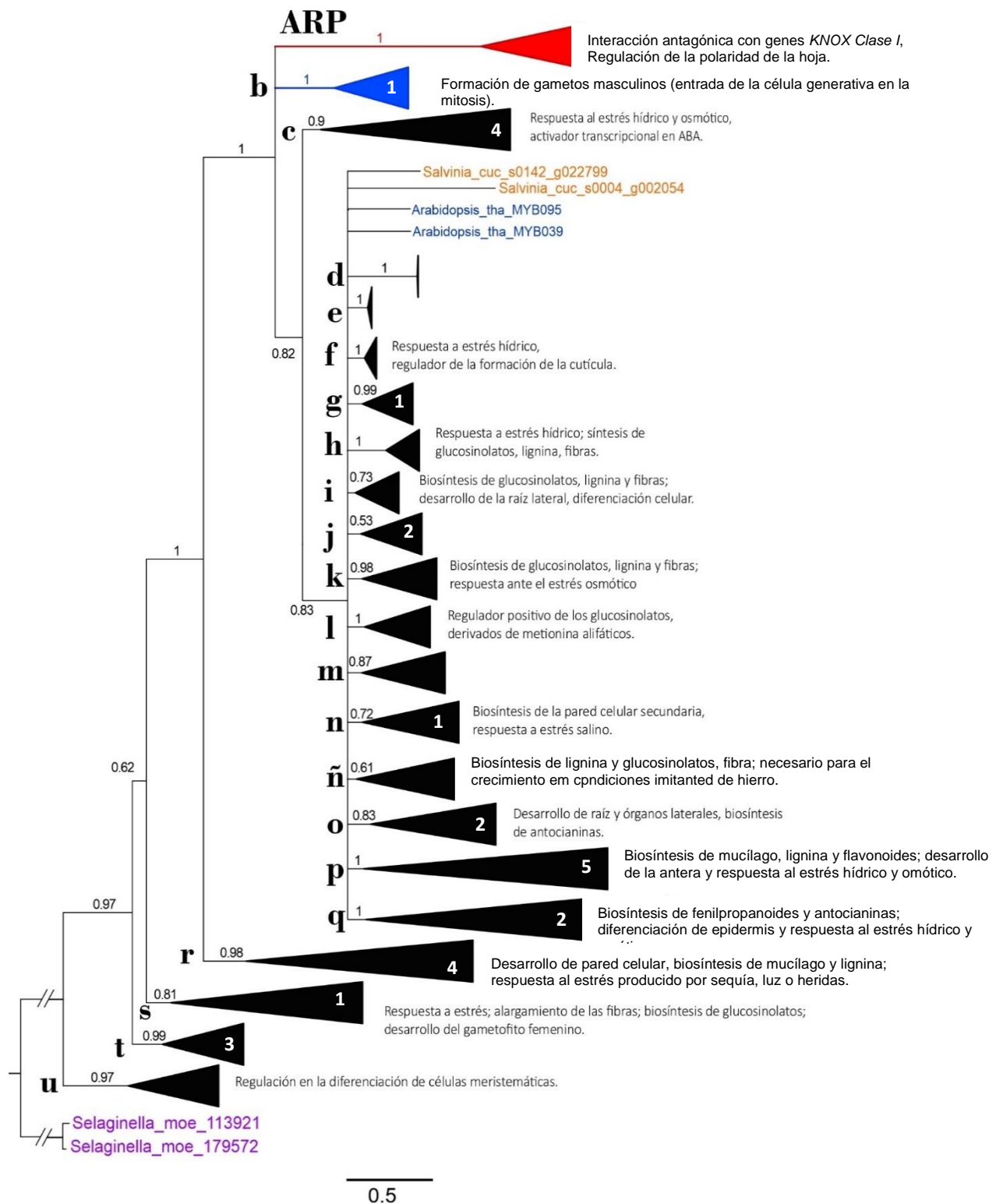
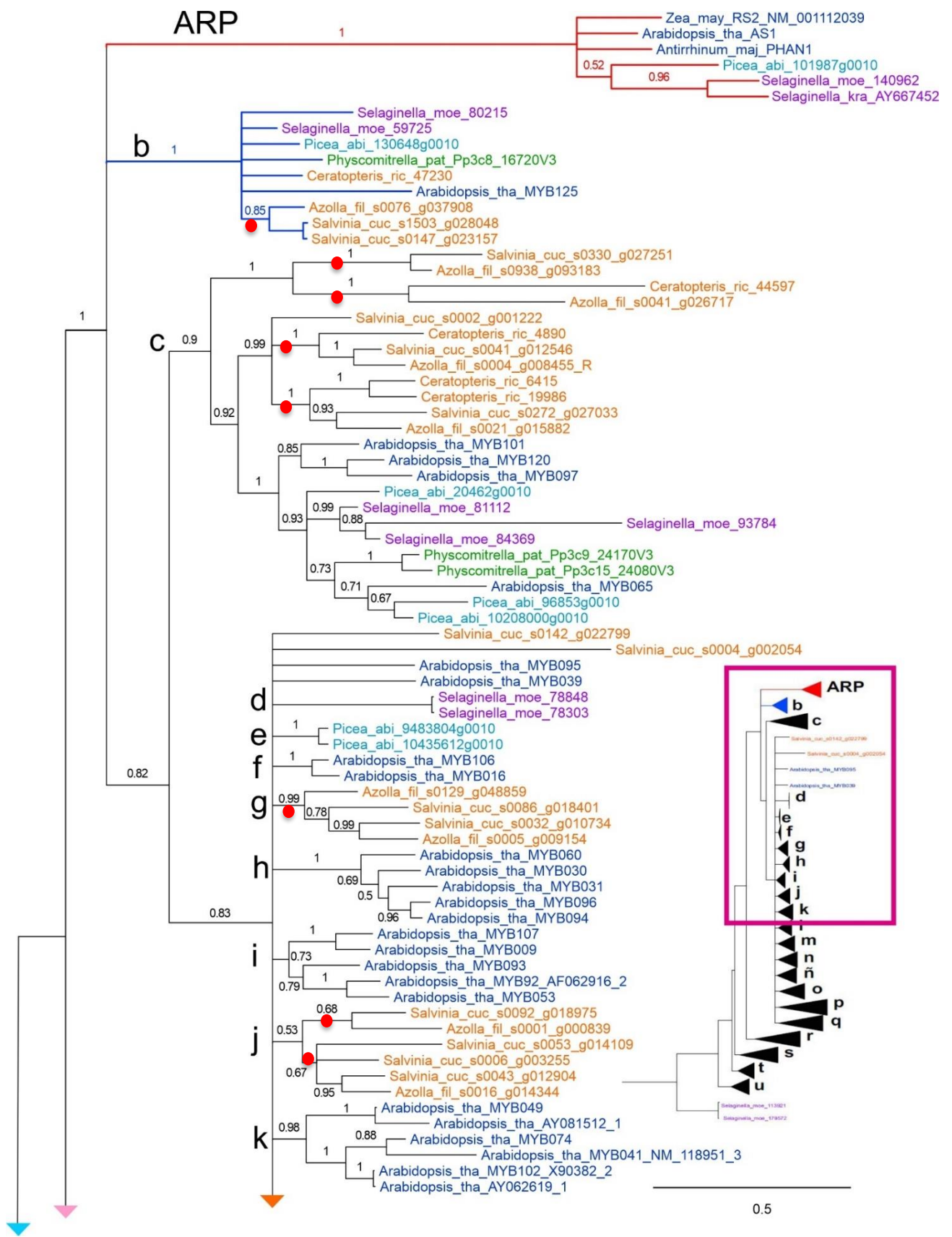
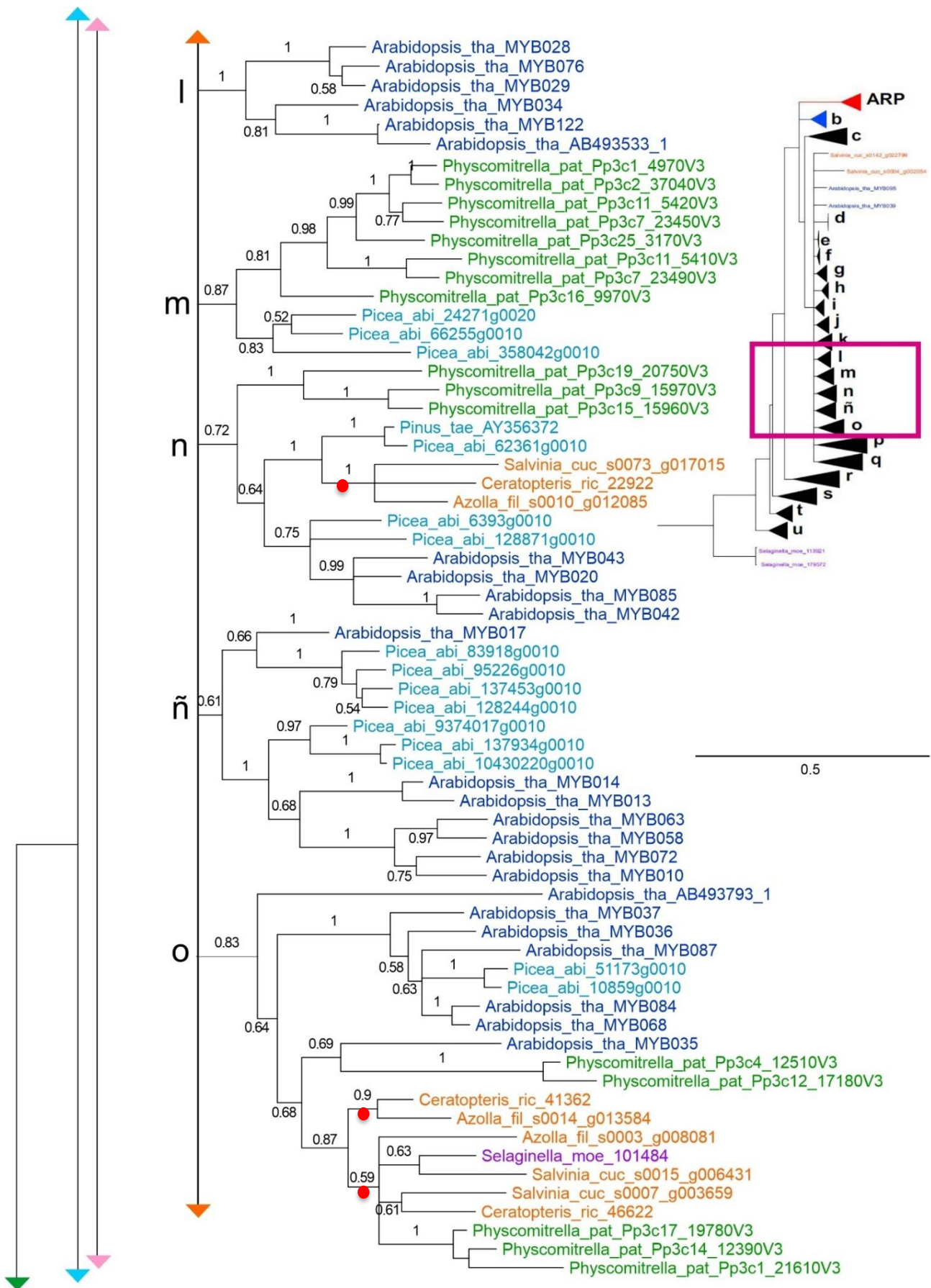
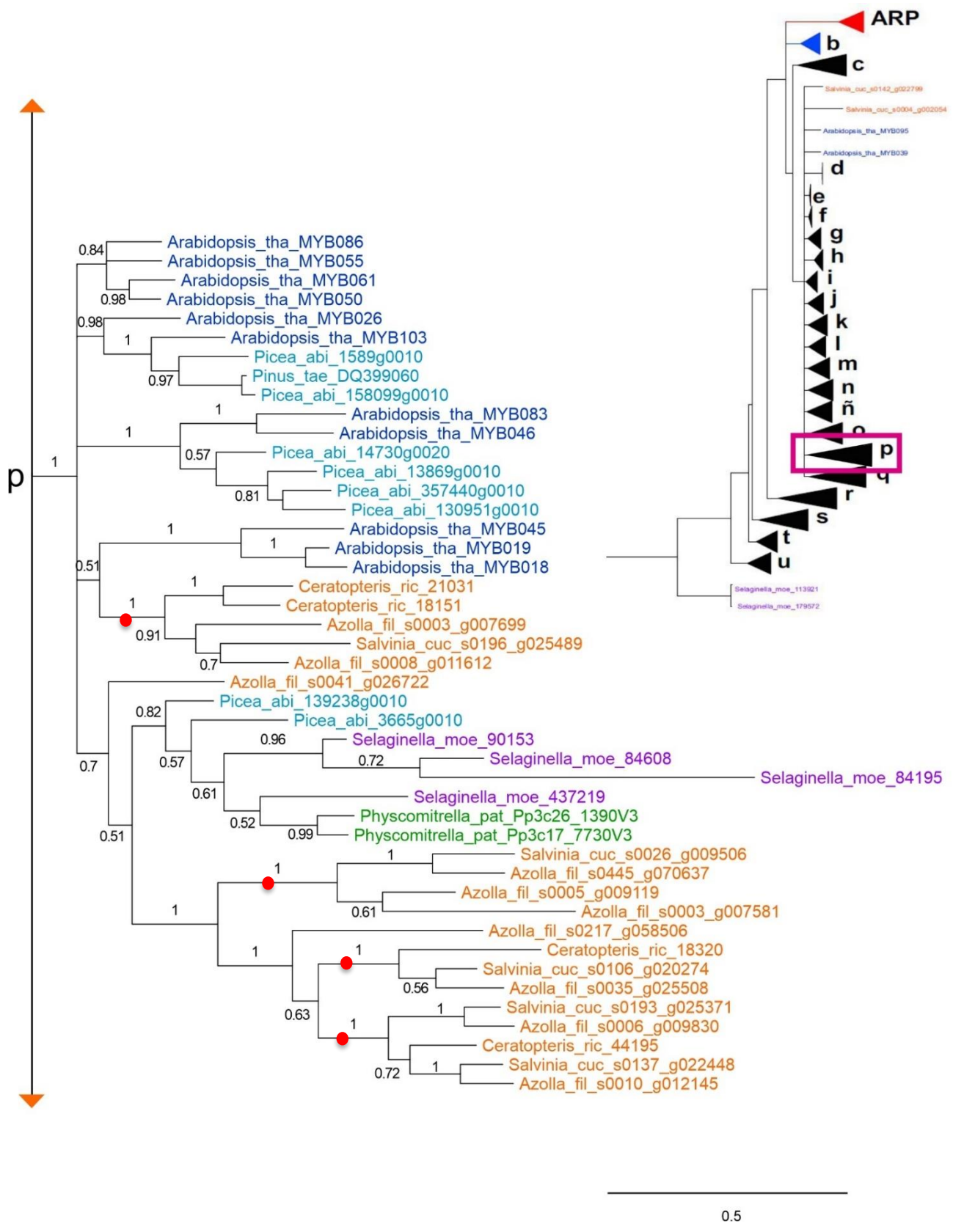
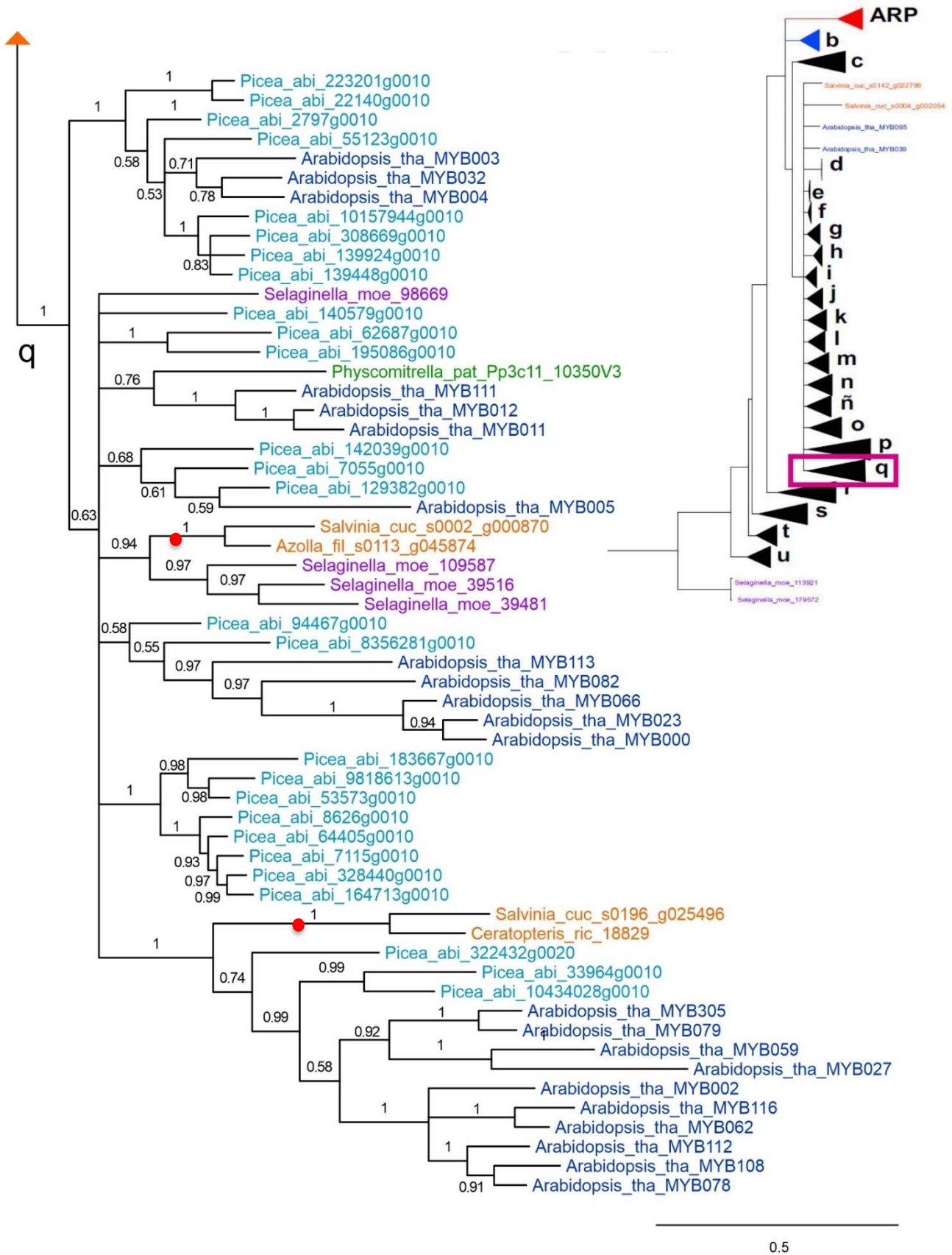


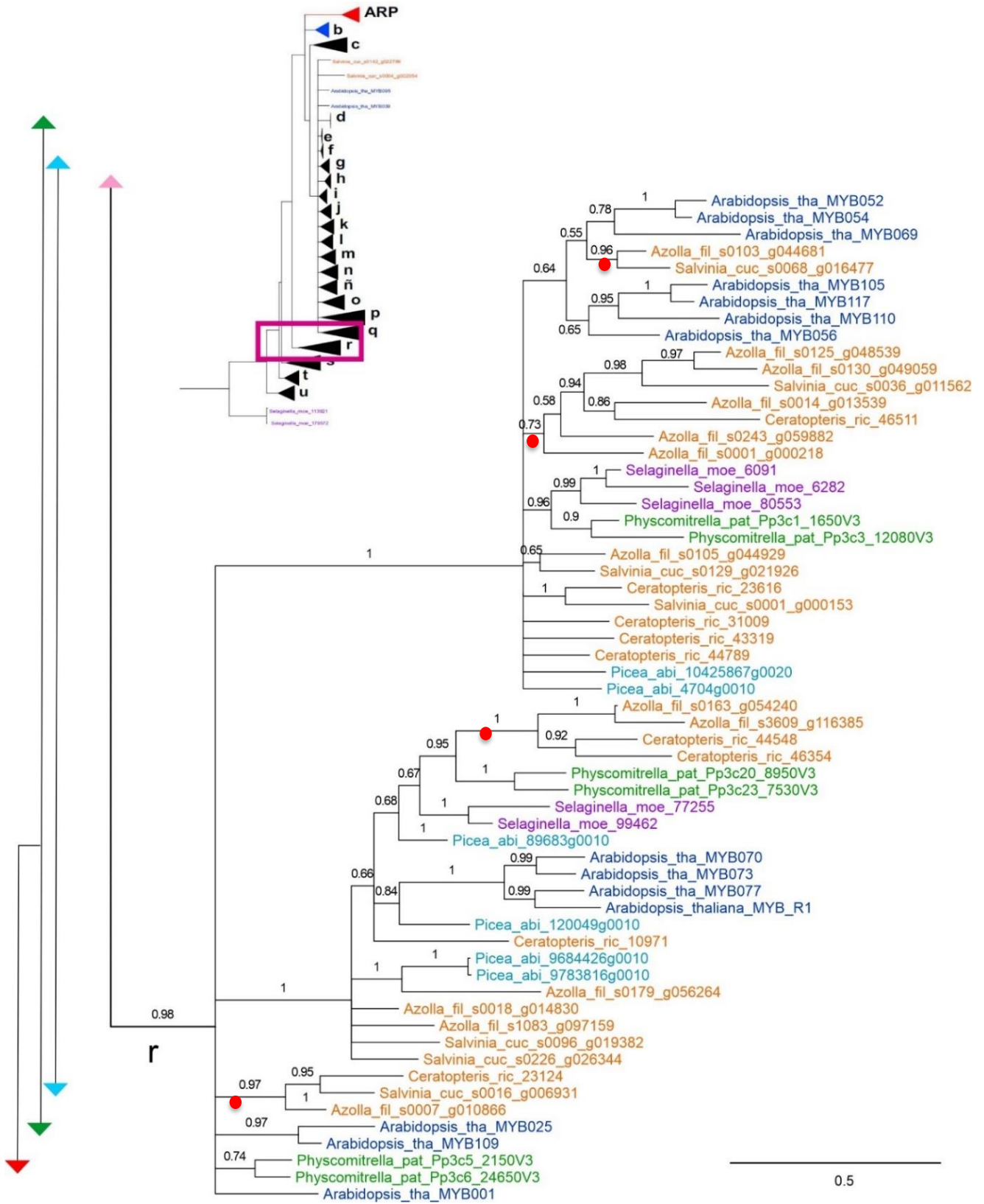
Figura 13. Resumen del análisis bayesiano de la matriz G. Esta filogenia fue realizada con Inferencia Bayesiana con el programa MrBayes (Ronquist et al., 2012) bajo el modelo de evolución molecular GTR+G+I, durante 10 millones de generaciones con cuatro cadenas de Markov (tres calientes y una fría), y un muestreo de la distribución de probabilidad posterior cada 1000 generaciones. Se muestran 22 clados, en cada uno se señalan el número de copias de genes *MYB* identificadas para helechos. Con base en los genes *MYB* de *Arabidopsis* agrupados en cada clado se determinó una posible función. Para más detalle de las secuencias que conforma a cada clado, ver Figura 14.











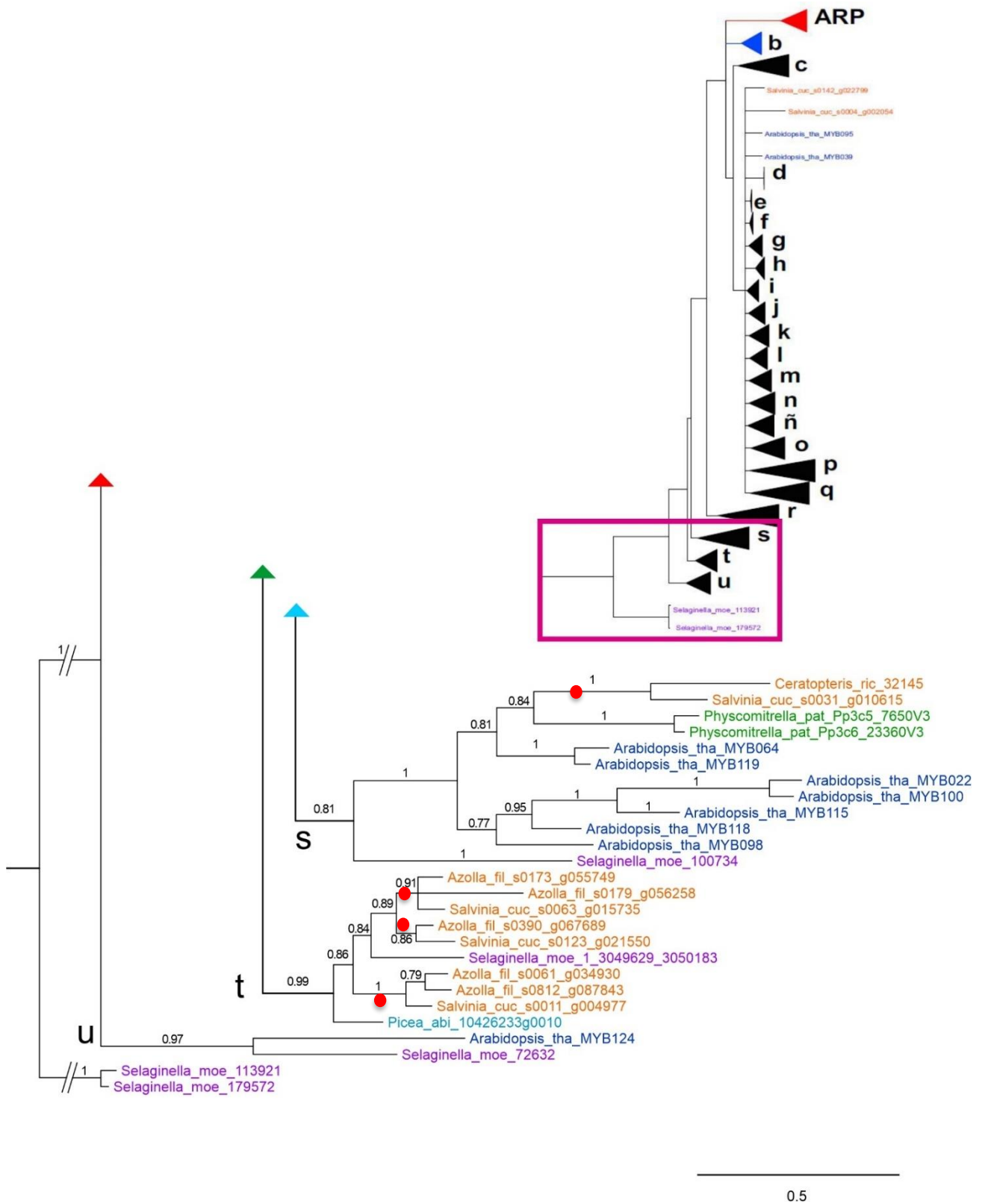


Figura 14. Filogenia de genes MYB de plantas terrestres. Esta filogenia fue realizada con Inferencia Bayesiana con el programa MrBayes (Ronquist et al., 2012) bajo el modelo de evolución molecular GTR+G+I. Con los puntos rojos se marcan las copias de genes MYB identificadas para helechos (26 copias). Cada linaje se identifica por un color diferente: musgos- verde, licofitas-morado, helechos-anaranjado, gimnospermas-azul claro, angiospermas-azul marino. Se muestran todas las secuencias de la matriz G agrupadas en los 22 clados resumidos en la Figura 13.

d) Análisis filogenéticos con la Matriz G+T (Genomas + Transcriptomas)

Con el fin de determinar si existen genes ortólogos de *ARP* en helechos, se realizaron tres análisis filogenéticos con la matriz G+T: un análisis basado en Inferencia Bayesiana (IB) (Figura 16) y dos análisis de Máxima Verosimilitud (MV), con diferentes programas: RaxML en BlackBox y IQtree (Supl 2 y 4).

Al comparar las filogenias que se obtuvieron con la matriz G+T y la matriz G, encontramos que los 22 clados recuperados con la matriz G también se recuperan con la matriz G+T. A pesar de que algunas relaciones entre las terminales se modifican los clados se conforman por las mismas secuencias, la única diferencia es que en las filogenias de la matriz G+T se integran las secuencias provenientes de los transcriptomas de diferentes linajes de plantas.

Los análisis realizados con IB y MV en ambas matrices (G y G+T) nos muestran que un número menor de terminales, en este caso, no modificó considerablemente la resolución del árbol ni mejora en gran medida los valores de probabilidad posterior (PP) y Bootstrap.

En las tres filogenias realizadas con la matriz G+T, se recuperó un clado, que llamamos clado ARP, el cual agrupó las secuencias previamente reportadas como *ARP* (*ASI*, *PHAN*, *RH2*, *SkARP*) y diez secuencias más, 8 de *Selaginella* y 2 de gimnospermas (Figura 16). Todas las filogenias realizadas con la matriz G+T muestran que el gen ortólogo de *ARP* en este linaje, es de copia única. Ninguna de las tres filogenias recupera dentro del clado ARP secuencias de helechos. También se recupera un clado (Figura 16, clado b) conformado por una secuencia de musgos, dos de *Selaginella*, cuatro de helechos, una de gimnospermas y una de angiospermas.

A pesar de que el clado b se recupera en las tres filogenias (Figura 15, Supl. 2 y 4), su relación con el clado ARP cambia. En la filogenia obtenida con RaxML el clado b es hermano del clado ARP. En la filogenia obtenida con MrBayes es parte de una tricotomía. En la filogenia obtenida con IQtree, el clado b se recupera como grupo hermano al clado c-q. En la Figura 17 se muestra un resumen de las diferentes relaciones del clado b con el clado ARP y clado c-q.

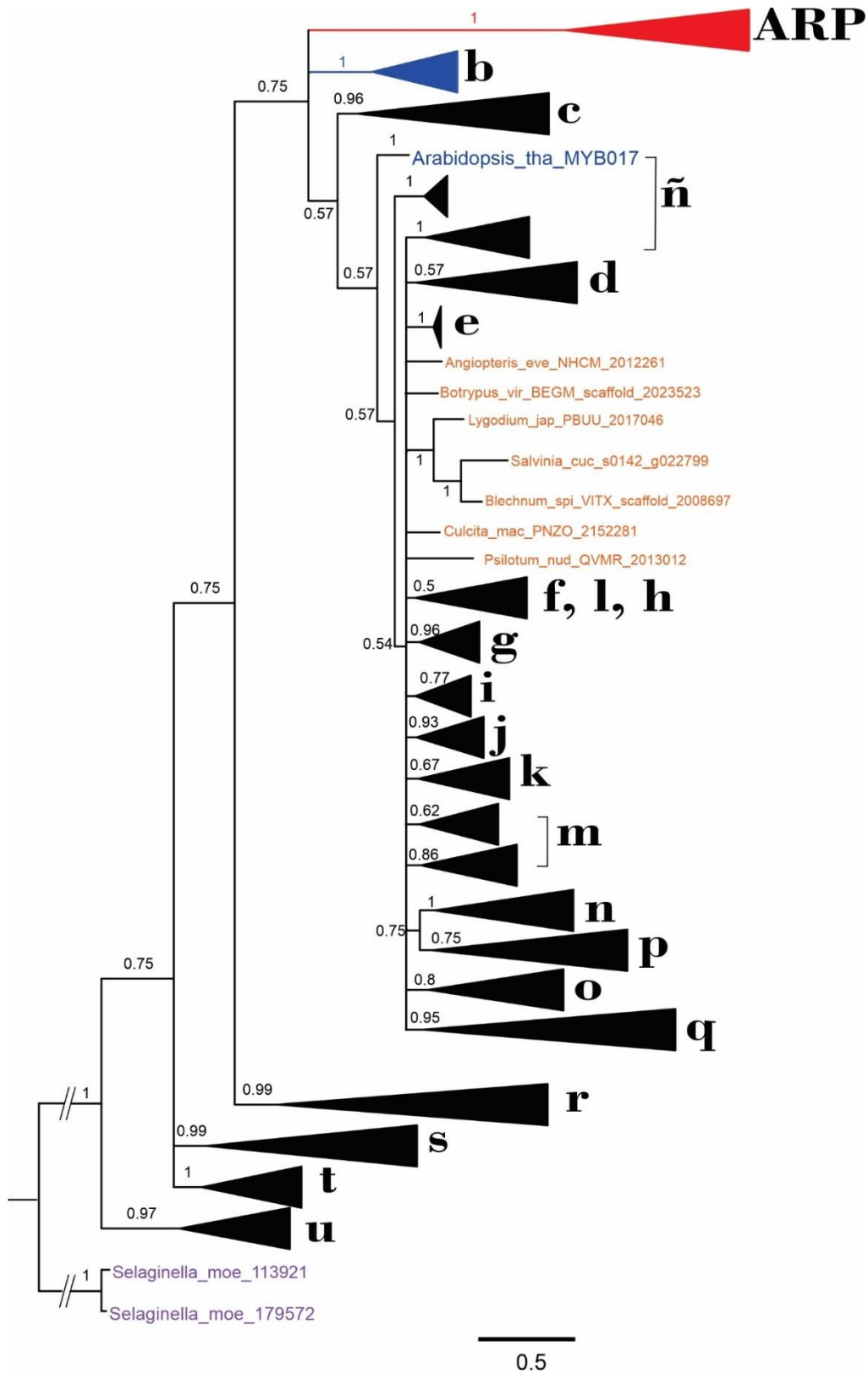
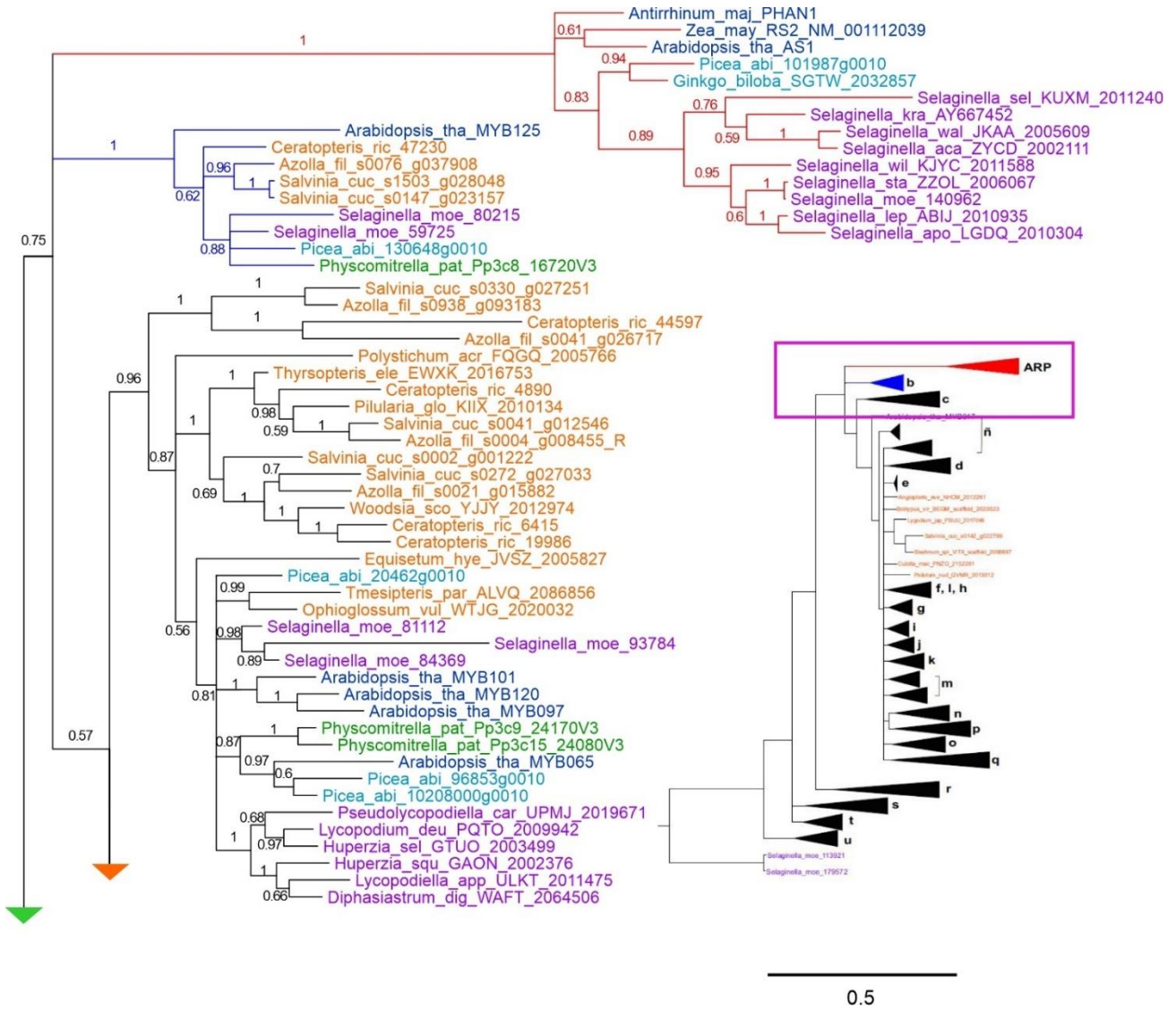
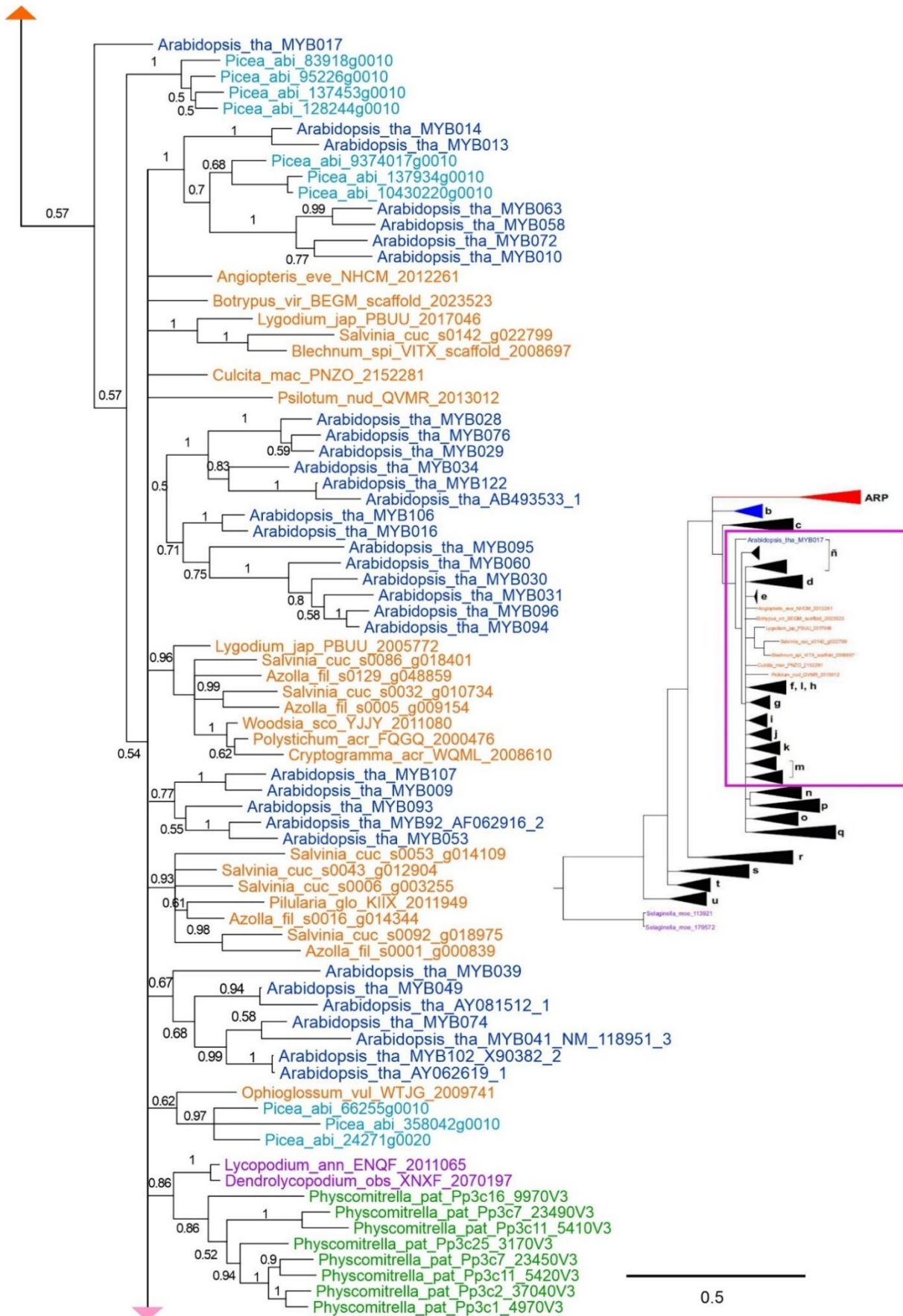
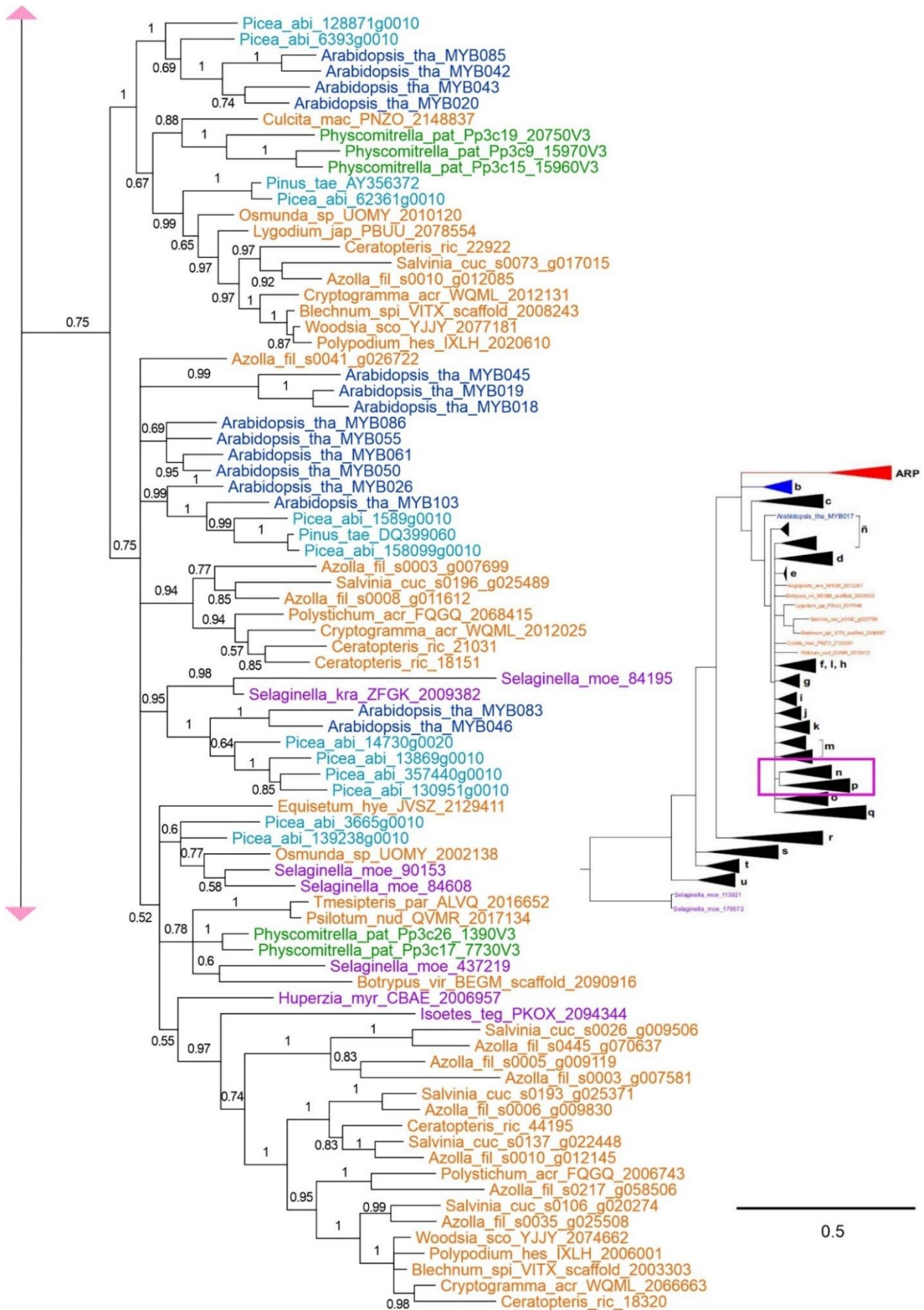
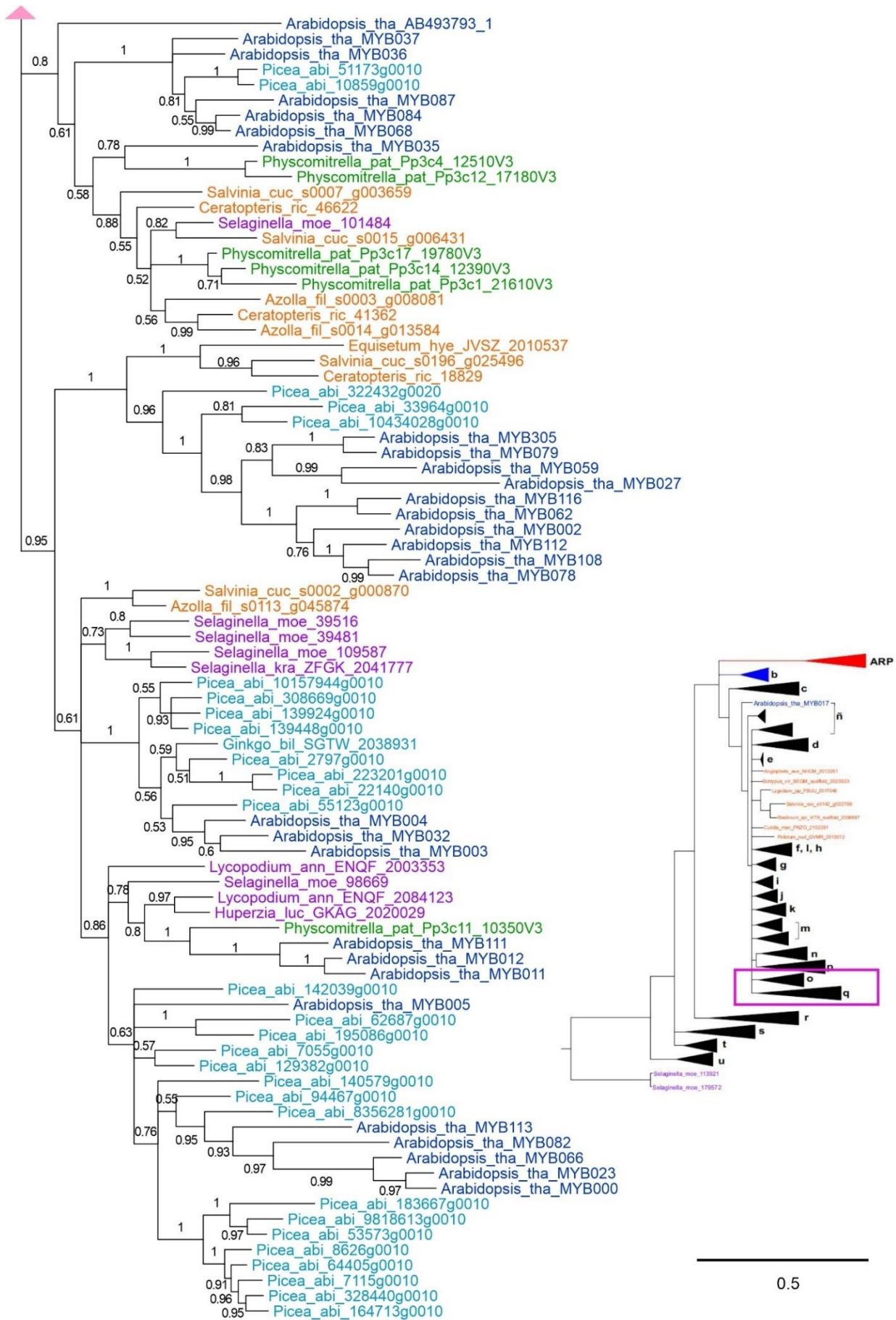


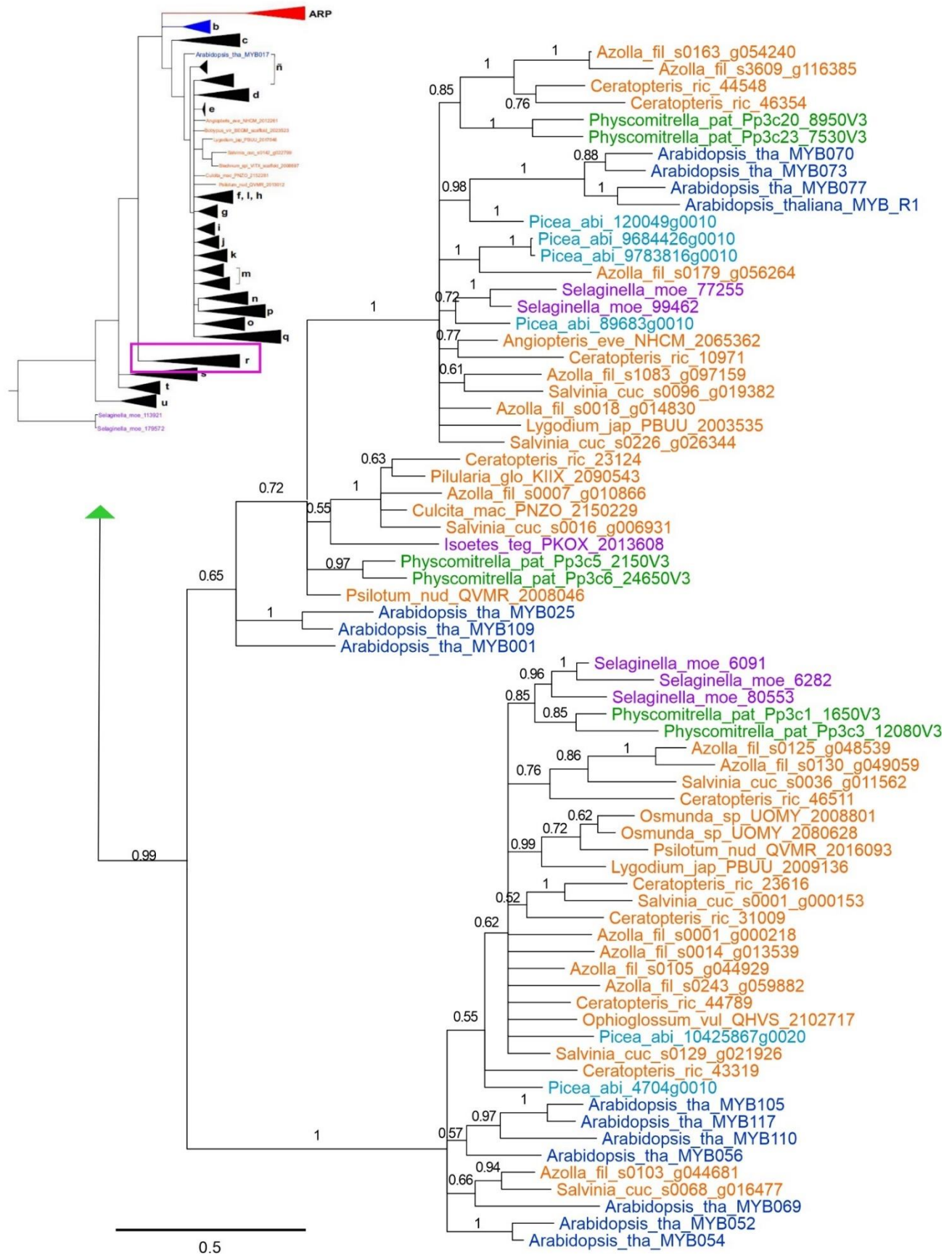
Figura 15. Resumen del análisis bayesiano de la matriz G+T. Esta filogenia fue realizada con Inferencia Bayesiana con el programa MrBayes (Ronquist et al., 2012) bajo el modelo de evolución molecular GTR+G+I. En esta filogenia se recuperan los 22 clados previamente identificados con los análisis filogenéticos de la matriz G. Las secuencias de helechos que no se agrupan en ningún clado provienen de transcriptomas.











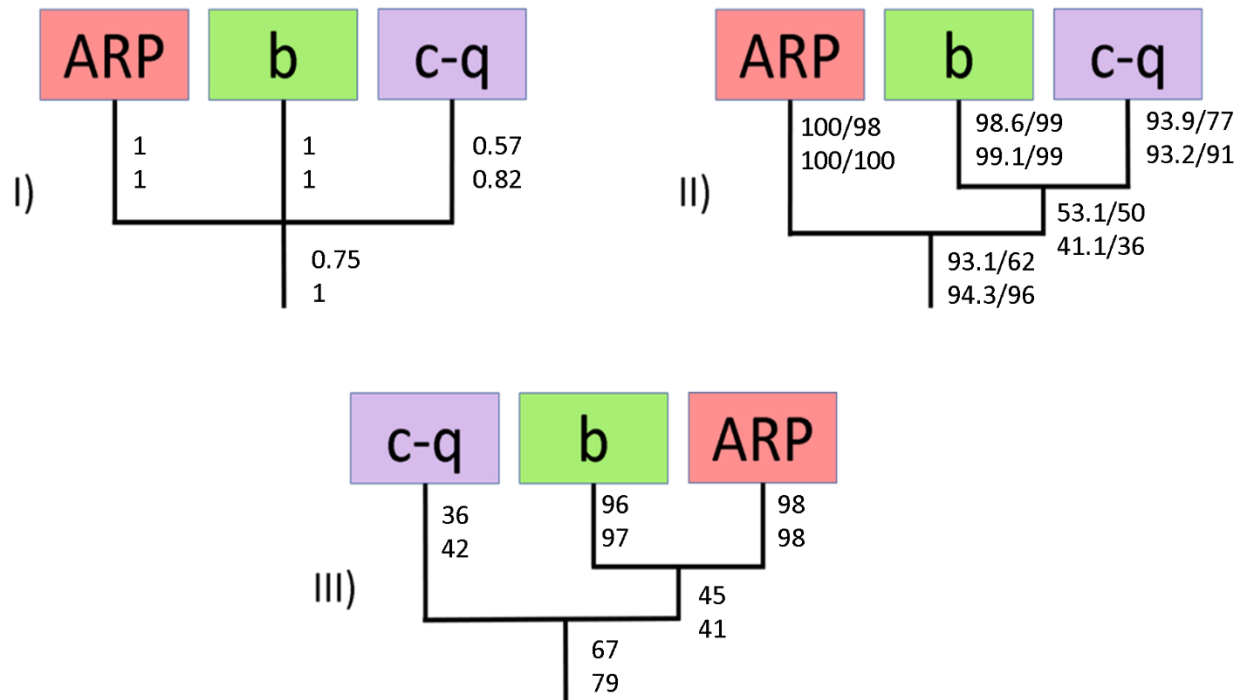


Figura 17. Resumen de las filogenias realizadas con Inferencia Bayesiana y Máxima Verosimilitud con las matrices G y G+T. Se muestran las relaciones entre los clados ARP, b y c-q y los soportes de las ramas, dependiendo del análisis empleado. Los valores mas altos en cada una de las ramas representan los soportes de probabilidad posterior (IB) y bootstrap (MV) obtenidos con la matriz G; los valores mas bajos de cada rama corresponden a los obtenidos con la matriz G+T. I) Topología obtenida con inferencia Bayesiana (Figuras 13 y 15). II) Topología obtenida con el análisis de Máxima Verosimilitud (IQtree) (Sulp. 4 y 5); III) Topología obtenida con el análisis de Máxima Verosimilitud (RaxML) (Sulp. 2 y 3).

e) Análisis de motivos proteicos

Con el fin de identificar los motivos proteicos que caracterizaban a las secuencias del clado ARP y del clado b en la matriz G+T (Figura 12), se realizó un análisis de motivos proteicos (MEME), utilizando el dominio MYB. Este análisis identificó 40 motivos (Supl. 7); los motivos proteicos que tienen el mejor valor de E se muestran en la Figura 18 (motivos a-j).

Todas las secuencias incluidas en la matriz G+T comparten dos motivos proteicos que son los de mayor longitud y tienen 41 y 29 aminoácidos respectivamente (Figura 18a y b). Ambos motivos cuentan con pocos aminoácidos conservados en todas las secuencias; sin embargo, en ambos motivos se identifican los residuos de triptófanos conservados característicos de cada repetición (Figura 18a y b).

El análisis de motivos proteicos reveló que existen dos motivos que son exclusivos del clado ARP (Figura 18 e y i). Uno de ellos se localiza en la región amino terminal, y en lo sucesivo lo llamaremos motivo ARP inicial (Figura 18 i), y el otro se localiza en la región próxima al carboxilo terminal, por lo que en lo sucesivo lo llamaremos motivo ARP final (Figura 18 e). El motivo ARP

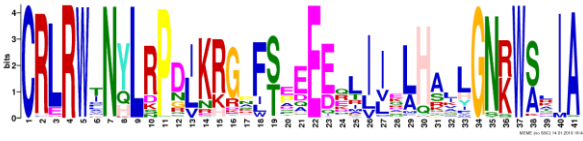
inicial, está conformado únicamente por seis residuos de aminoácidos. El motivo ARP final cuenta con un residuo de triptófano extra junto al tercer triptófano de la repetición R3 y tiene 21 aminoácidos de los cuales 14 están conservados en todas las secuencias del clado ARP.

Las filogenias obtenidas con la matriz G+T y los análisis de MEME realizados con esta misma matriz (Figura 18), muestran la presencia de genes ortólogos de *ARP* en ocho especies de *Selaginella*: *Selaginella acanthonota*, *S. apoda*, *S. selaginoides*, *S. stauntoniana*, *S. wallacei*, *S. willdenowii*, *S. lepidophylla* y *S. moellendorffii*. Hasta ahora, únicamente se habían reportado genes ortólogos de *ARP* en *S. kraussiana* (Harrison et al., 2005). También se identificaron dos secuencias de gimnospermas ortólogos de *ARP*: *Picea_abi_101987g0010* (*Picea abies*), *Ginkgo_biloba_SGTW_2032857* (*Ginkgo biloba*). Los genes que detectamos como ortólogos de *ARP*, comparten con *AS1*, *PHAN* y *RS2* el motivo ARP (Figura 18e).

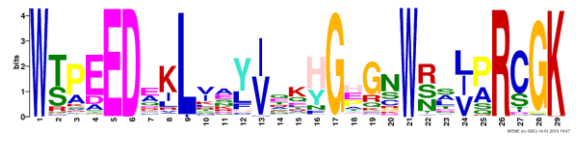
Para las secuencias del clado b, el análisis de MEME no identificó motivos exclusivos. Sin embargo, como se menciona en los resultados del alineamiento, la visualización directa permite identificar algunos residuos de aminoácidos conservados en posiciones específicas y que son exclusivos de las secuencias correspondientes a este clado: Por ejemplo, dos serinas continuas en las posiciones 28 y 29, dos argininas continuas en las posiciones 31 y 32 y dos valinas consecutivas en las posiciones 62 y 63, excepto en la secuencia 15 correspondiente a la proteína MYB125 de *Arabidopsis* (Figura 12).

Las secuencias *Selaginella_moe_113921* y *Selaginella_moe_179572*, que fueron identificadas en el análisis de similitud como las secuencias que más difieren, también tuvieron dos motivos exclusivos (Figura 18 j, k).

a) $E = 2.5e-10885$



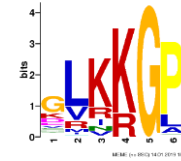
b) $E = 3.5e-6884$



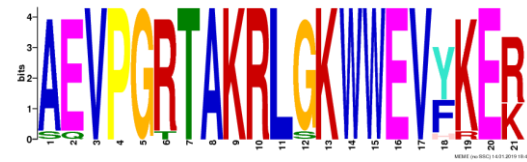
c) $E = 1.4e-4801$



d) $E = 1.0e-436$



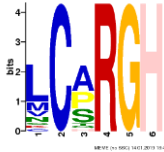
e) $E = 2.2e-187$



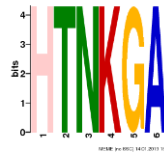
f) $E = 6.6e-123$



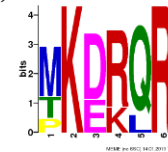
g) $E = 9.5e-033$



h) $E = 1.4e-007$



i) $E = 1.2e-003$



j) $E = 8.4e-003$



k) $E = 4.0e+003$

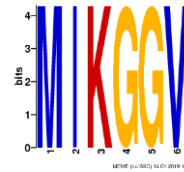


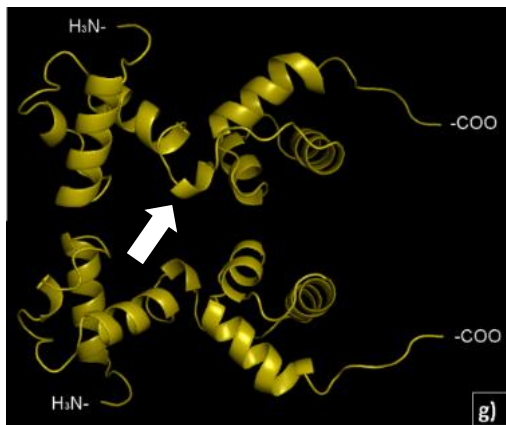
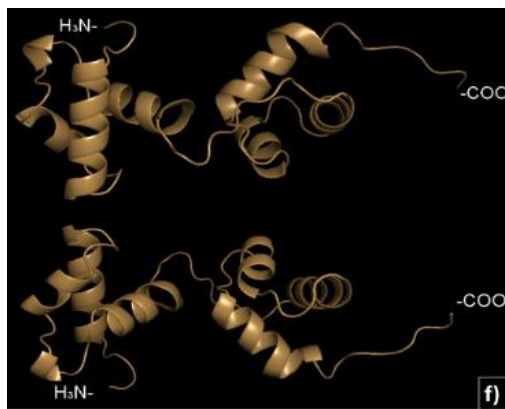
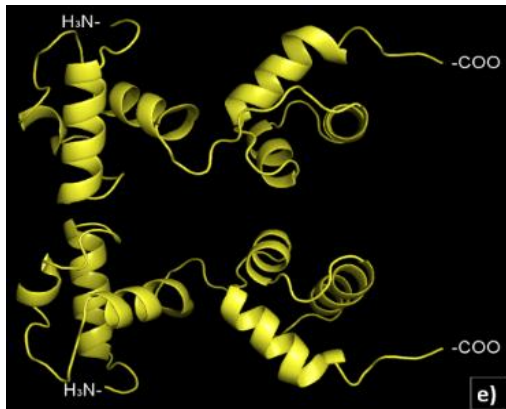
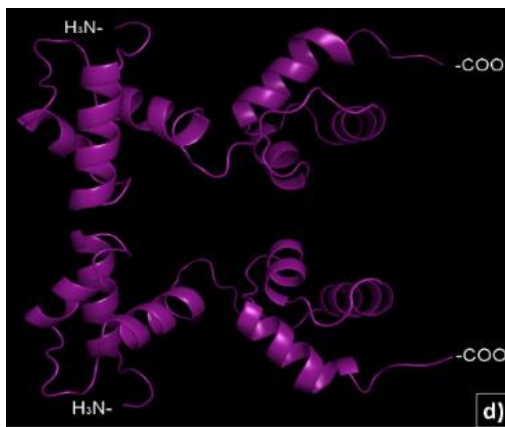
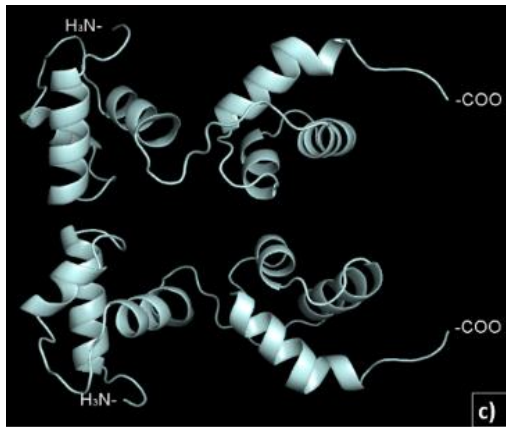
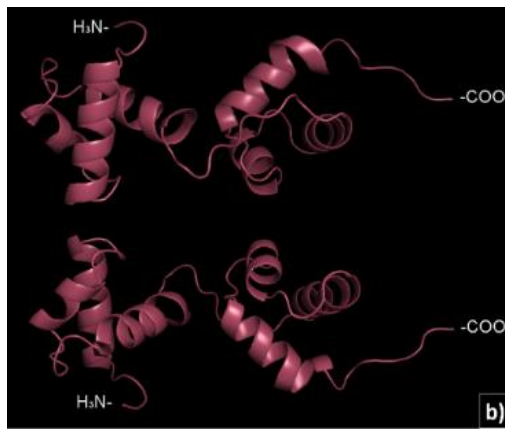
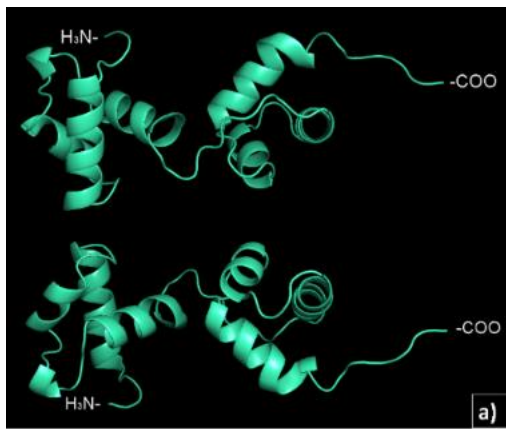
Figura 18. Análisis de motivos proteicos. Se muestran los esquemas que representan a cada uno de los motivos proteicos identificados mediante el análisis de MEME. El tamaño de la letra que de cada aminoácido representa la frecuencia con la que se repite dicho residuo en las secuencias en las que se identificó el motivo. La cantidad de aminoácidos que conforman cada dominio está marcada por el eje inferior. El valor de E indica la probabilidad que existe de que esa secuencia de aminoácidos se haya encontrado por azar en el alineamiento, los primeros diez motivos son los que obtuvieron el mejor valor de E (a-j), estos motivos son compartidos por la mayoría de las secuencias incluidas en este análisis. Los motivos de mayor interés para este estudio son los siguientes: e) Motivo proteico identificado como exclusivo de las secuencias ARP; j), k) motivos característicos del grupo externo.

f) Modelo por homología de las proteínas del clado ARP y el clado b

Con el fin de evaluar y comparar la conformación tridimensional que adquieren las proteínas *ARP* y las secuencias del clado b (grupo hermano del clado ARP en los análisis de MV), se realizaron modelos de estructura tridimensional en la plataforma de SWISS-MODEL.

La mayoría de los modelos proteicos obtenidos con el análisis en SWISS-MODEL tiene dos grupos de alfa hélices: el grupo cercano a la región amino terminal que cuenta con dos o tres alfa hélices y el grupo cercano a la región carboxilo terminal que en todos los casos cuenta con tres alfa hélices (Figura 19). La única excepción fue el modelo correspondiente a *Selaginella moellendorffii* (Selaginella_moe_80215) en donde no es posible diferenciar el grupo cercano a la región carboxilo terminal (Figura 19 ñ).

Cada uno de los grupos de alfa hélices corresponde a una repetición, es decir, el grupo cercano a la región amino terminal corresponde a la repetición R2 y el grupo de alfa hélices cercano a la región carboxilo terminal corresponde a la repetición R3. Estas repeticiones, en todos los modelos, están unidas por un asa. Dos de los modelos correspondientes a las secuencias provenientes de helechos cuentan con una pequeña hélice que une a ambas repeticiones (Figura 19 g, i). Algunos de los modelos tienen una hélice incompleta en la repetición R2 (Figura 19 a, e, f, g).



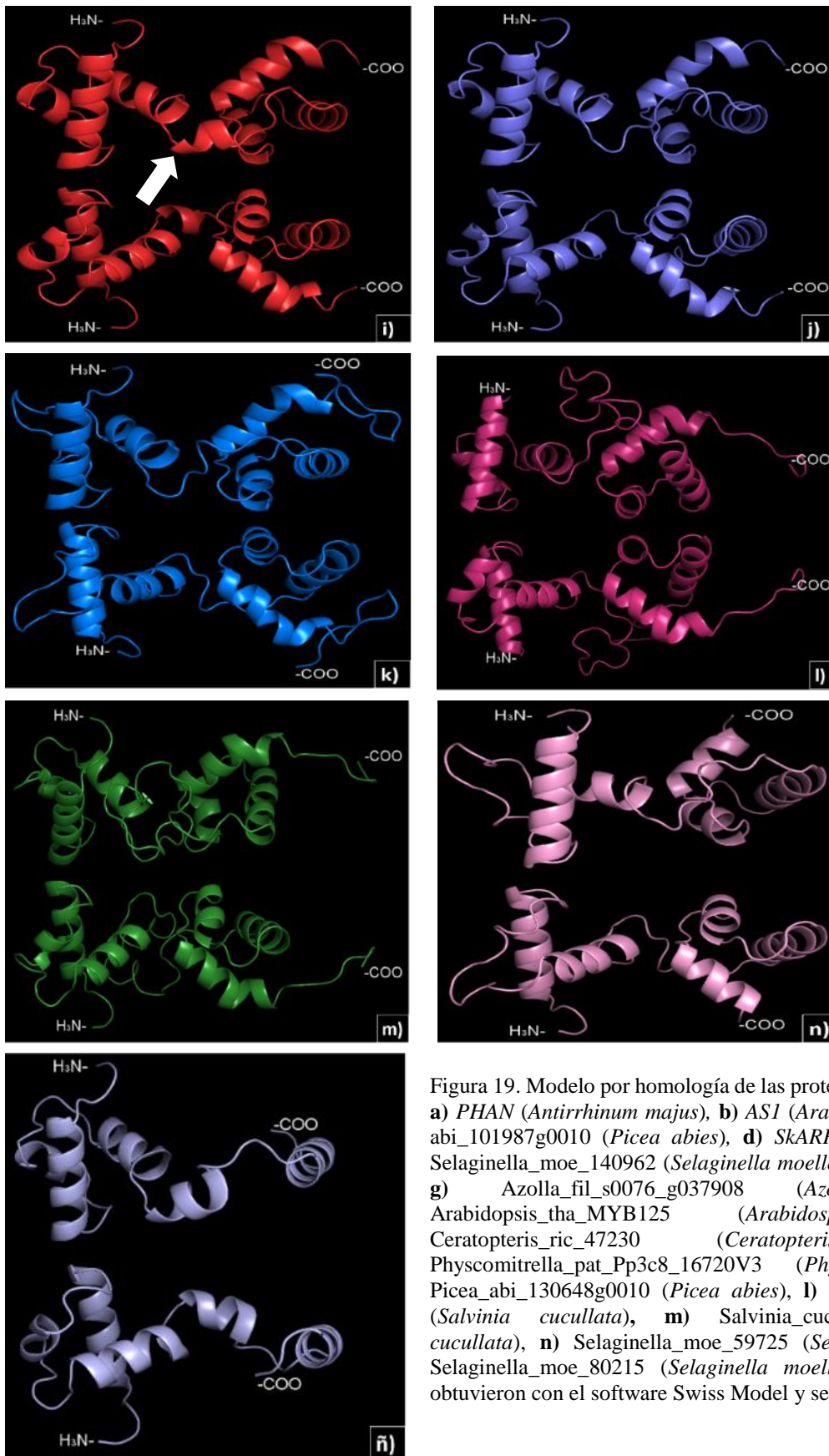


Figura 19. Modelo por homología de las proteínas del clado ARP y clado b. **a)** PHAN (*Antirrhinum majus*), **b)** ASI (*Arabidopsis thaliana*), **c)** Picea_abi_101987g0010 (*Picea abies*), **d)** SkARP (*Selaginella krausiana*), **e)** Selaginella_moe_140962 (*Selaginella moellendorffii*), **f)** RS2 (*Zea mays*), **g)** Azolla_fil_s0076_g037908 (*Azolla filiculoides*), **h)** Arabidopsis_tha_MYB125 (*Arabidopsis thaliana*), **i)** Ceratopteris_ric_47230 (*Ceratopteris richardii*), **j)** Physcomitrella_pat_Pp3c8_16720V3 (*Physcomitrella patens*), **k)** Picea_abi_130648g0010 (*Picea abies*), **l)** Salvinia_cuc_s0147_g023157 (*Salvinia cucullata*), **m)** Salvinia_cuc_s1503_g028048 (*Salvinia cucullata*), **n)** Selaginella_moe_59725 (*Selaginella moellendorffii*), **ñ)** Selaginella_moe_80215 (*Selaginella moellendorffii*). Los modelos se obtuvieron con el software Swiss Model y se visualizaron en PyMol.

Cuadro 1. Genes MYB y sus funciones en plantas. En este cuadro se muestran los genes *MYB* de *Arabidopsis* incluidos en los análisis filogenéticos y sus funciones reportadas. Los genes *MYB* relacionados con desarrollo de órganos laterales o diferenciación celular se resatan en negritas.

Nombre de la secuencia en el análisis	Función	Referencia
Arabidopsis_tha_AB493533_1	Proteína hipotética reportada como MYB, no hay prueba funcional.	---
Arabidopsis_tha_AB493793_1	Proteína hipotética reportada como MYB, no hay prueba funcional.	---
Arabidopsis_tha_AS1	Proteína MYB. Especificación próximo-distal de la hoja.	Byrne et al., 2000
Arabidopsis_tha_AY059820_1	Proteína hipotética reportada como MYB, no hay prueba funcional.	---
Arabidopsis_tha_AY062619_1	Proteína hipotética reportada como MYB, no hay prueba funcional.	---
Arabidopsis_tha_MYB000	Desarrollo de tricomas y raíz.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB001	Respuesta a estrés.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB002	Respuesta al estrés hídrico y osmótico. Activador transcripcional en ABA.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB003	Biosíntesis de fenilpropanoides.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB004	Biosíntesis de fenilpropanoides.	Jin et al., 2000
Arabidopsis_tha_MYB005	Biosíntesis de flavonoides.	Parish et al., 2009
Arabidopsis_tha_MYB009	Biosíntesis de lignina, glucosinolato, fibra.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB010	Necesario para el crecimiento en condiciones limitantes de hierro.	Palmer et al., 2013
Arabidopsis_tha_MYB011	Biosíntesis de flavonoides.	Pandey, Misra, & Trivedi, 2015
Arabidopsis_tha_MYB012	Biosíntesis de flavonoides	Wang et al., 2016
Arabidopsis_tha_MYB013	Biosíntesis de lignina, glucosinolato, fibra.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB014	Respuesta de la planta al estrés por frío.	Chen et al., 2013
Arabidopsis_tha_MYB016	Regulador de la formación de la cutícula.	Oshima & Mitsuda, 2013
Arabidopsis_tha_MYB017	Biosíntesis de lignina, glucosinolato, fibra	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB018	Desarrollo, biosíntesis de mucilago y lignina.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB019	Desarrollo, biosíntesis de mucilago y lignina.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB020	Respuesta a estrés salino.	Cui et al., 2013
Arabidopsis_tha_MYB022	Desarrollo del gametofito femenino, biosíntesis de glucosinolato.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB023	Control de diferenciación de células epidérmicas.	Nam et al., 2009
Arabidopsis_tha_MYB025	Respuesta a estrés.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB026	Engrosamiento secundario en la antera.	Simpson et al., 2017
Arabidopsis_tha_MYB027	Respuesta a estrés hídrico y a altas condiciones de sal.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB028	Regulador positivo de los glucosinolatos derivados de metionina alifáticos.	Gigolashvili et al., 2007
Arabidopsis_tha_MYB029	Síntesis de glucosinolato.	Vaneechoutte et al., 2017
Arabidopsis_tha_MYB030	Síntesis de glucosinolato, lignina, fibras.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB031	Síntesis de glucosinolato, lignina, fibras.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB032	Síntesis de fenilpropanoides.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB034	Síntesis de glucosinolato, lignina, fibras.	Revisado en Ambawat et al., 2013
Arabidopsis_tha_MYB035	Desarrollo.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB036	Desarrollo.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB037	Regulación del meristema axilar y formación lateral de órganos laterales.	Revisado en Ambawat et al., 2013
Arabidopsis_tha_MYB039	Síntesis de glucosinolato, lignina, fibras.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB042	Biosíntesis de la pared celular secundaria.	Zhong et al., 2008
Arabidopsis_tha_MYB043	Biosíntesis de la pared celular secundaria.	Zhong et al., 2008
Arabidopsis_tha_MYB045	Desarrollo, biosíntesis de mucilago y lignina.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB046	Expresión en silicua y capullos de flores.	Revisado en Ambawat et al., 2013
Arabidopsis_tha_MYB049	Respuesta a estrés por altas condiciones de sal.	Yanhui et al., 2006
Arabidopsis_tha_MYB050	Desarrollo, biosíntesis de mucilago y lignina.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB052	Engrosamiento de la pared celular.	Revisado en Ambawat et al., 2013
Arabidopsis_tha_MYB053	Desarrollo, biosíntesis de mucilago y lignina.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB054	Engrosamiento de la pared celular.	Revisado en Ambawat et al., 2013
Arabidopsis_tha_MYB055	Desarrollo, biosíntesis de mucilago y lignina.	Revisado en Bowman et al., 2017

Arabidopsis_tha_MYB056	Desarrollo, biosíntesis de mucilago y lignina.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB058	Biosíntesis de lignina.	Revisado en Ambawat et al., 2013
Arabidopsis_tha_MYB059	Respuesta a estrés hídrico y a altas condiciones de sal.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB060	Respuesta a estrés hídrico.	Revisado en Ambawat et al., 2013
Arabidopsis_tha_MYB061	Desarrollo, biosíntesis de mucilago y lignina.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB062	Respuesta a estrés hídrico y a altas condiciones de sal.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB063	Biosíntesis de lignina.	Revisado en Ambawat et al., 2013
Arabidopsis_tha_MYB064	Biosíntesis de glucosinolatos.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB065	Desarrollo del estambre y de la antera.	Revisado en Ambawat et al., 2013
Arabidopsis_tha_MYB066	Desarrollo de pelos radiculares.	Revisado en Ambawat et al., 2013
Arabidopsis_tha_MYB068	Alargamiento de la raíz, formación de órganos laterales.	Revisado en Ambawat et al., 2013
Arabidopsis_tha_MYB069	Engrosamiento de la pared celular.	Revisado en Ambawat et al., 2013
Arabidopsis_tha_MYB070	Respuesta al estrés producido por sequía, luz o heridas.	Revisado en Ambawat et al., 2013
Arabidopsis_tha_MYB072	Necesario para el crecimiento en condiciones limitantes de hierro.	Palmer et al., 2013
Arabidopsis_tha_MYB073	Respuesta al estrés producido por sequía, luz o heridas.	Revisado en Ambawat et al., 2013
Arabidopsis_tha_MYB074	Biosíntesis de glucosinolatos, lignina y fibras.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB076	Biosíntesis de glucosinolatos.	Gigolashvili et al., 2007
Arabidopsis_tha_MYB077	Respuesta al estrés producido por sequía, luz o heridas.	Revisado en Ambawat et al., 2013
Arabidopsis_tha_MYB078	Respuesta a estrés hídrico y a altas condiciones de sal.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB079	Respuesta a estrés hídrico y a altas condiciones de sal.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB082	Biosíntesis de fenilpropanoides y desarrollo de tricomas.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB083	Desarrollo, biosíntesis de mucilago y lignina.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB084	Alargamiento de la raíz, formación de órganos laterales.	Revisado en Ambawat et al., 2013
Arabidopsis_tha_MYB085	Biosíntesis de la pared celular secundaria.	Zhong et al., 2008
Arabidopsis_tha_MYB086	Biosíntesis de flavonoides.	Zhang et al., 2015
Arabidopsis_tha_MYB087	Biosíntesis de antocianinas.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB093	Regulador negativo del desarrollo de la raíz lateral.	Gibbs et al., 2014
Arabidopsis_tha_MYB094	Respuesta a estrés por sequía.	Revisado en Ambawat et al., 2013
Arabidopsis_tha_MYB095	Biosíntesis de glucosinolatos, lignina y fibras.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB096	Respuesta a estrés por sequía.	Revisado en Ambawat et al., 2013
Arabidopsis_tha_MYB097	Control de la diferenciación del tubo polínico.	Liang et al., 2013
Arabidopsis_tha_MYB098	Biosíntesis de glucosinolatos.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB100	Biosíntesis de glucosinolatos.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB101	Respuesta hormonal (señalización Giberelinas).	Revisado en Ambawat et al., 2013
Arabidopsis_tha_MYB103	Respuesta a estrés hídrico y a altas condiciones de sal.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB105	Desarrollo pared celular.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB106	Biosíntesis de glucosinolatos, lignina y fibras.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB107	Biosíntesis de glucosinolatos, lignina y fibras.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB108	Respuesta a estrés hídrico y a altas condiciones de sal.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB109	Alargamiento de las fibras.	Revisado en Ambawat et al., 2013
Arabidopsis_tha_MYB110	Desarrollo pared celular.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB111	Biosíntesis de flavonoides.	Revisado en Ambawat et al., 2013
Arabidopsis_tha_MYB112	Respuesta a estrés hídrico y a altas condiciones de sal.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB113	Biosíntesis de antocianinas.	Revisado en Ambawat et al., 2013
Arabidopsis_tha_MYB115	Embriogénesis.	Revisado en Ambawat et al., 2013
Arabidopsis_tha_MYB116	Respuesta a estrés hídrico y a altas condiciones de sal.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB117	Desarrollo pared celular.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB118	Embriogénesis.	Revisado en Ambawat et al., 2013
Arabidopsis_tha_MYB119	Biosíntesis de glucosinolatos.	Revisado en Bowman et al., 2017
Arabidopsis_tha_MYB120	Control de la diferenciación del tubo polínico.	Liang et al., 2013
Arabidopsis_tha_MYB122	Biosíntesis de glucosinolatos, lignina y fibras.	Revisado en Ambawat et al., 2013
Arabidopsis_tha_MYB124	Regulación en la diferenciación de células meristemáticas.	arabidopsis.org
Arabidopsis_tha_MYB125	Formación de gametos masculinos (entrada de la célula generativa en la mitosis).	(Peters et al., 2016)
Arabidopsis_tha_MYB305	Respuesta a Luz Ultravioleta	Revisado en Ambawat et al., 2013

Arabidopsis_tha_MYB92_AF062916_2	Diferenciación celular, regulación de la transcripción.	arabidopsis.org
Arabidopsis_thaliana_MYB_R1	Proteína hipotética reportada como MYB, no hay prueba funcional.	arabidopsis.org
Arabidopsis_tha_MYB041_NM_118951_3	Respuesta ante el estrés osmótico.	Schlereth et al., 2009
Arabidopsis_tha_MYB102_X90382_2	Biosíntesis de glucosinolatos, lignina y fibras.	Revisado en Bowman et al., 2017

Cuadro 2. Especies de helechos muestreadas con *ASI* (at2g37630) en OneKP en el análisis exploratorio (Etapa I). En la última columna se muestra el número de secuencias recuperadas por cada especie.

	Especie	Número de secuencias obtenidas
Helechos Eusporangiados	<i>Angiopteris evecta</i>	2
	<i>Botrypus virginianum</i>	2
	<i>Equisetum hyemale</i>	3
	<i>Equisetum diffusum</i>	1
	<i>Ophioglossum vulgatum</i>	3
	<i>Psilotum nudum</i>	4
	<i>Tmesipteris parva</i>	2
Helechos Leptosporangiados	<i>Blechnum spicant</i>	3
	<i>Cryptogramma acrostichoides</i>	4
	<i>Culcita macrocarpa</i>	3
	<i>Lygodium japonicum</i>	5
	<i>Osmunda sp (UOMY)</i>	4
	<i>Pilularia globulifera</i>	3
	<i>Ploystichum acrostichoides</i>	4
	<i>Polypodium hesperium</i>	2
	<i>Thyrsopteris elegans</i>	1
	<i>Woodsia scopulina</i>	4

Cuadro 3. Especies de licofitas muestreadas con *AS1* (at2g37630) y *SkARP* (Harrison et al., 2005) en OneKP. La segunda columna indica en número de secuencias recuperadas por cada especie. El símbolo * indica con que gen se recuperaron esas secuencias.

Especie	Numero de secuencias obtenidas	Secuencias obtenidas con el gen <i>AS1</i>	Secuencias obtenidas con el gen <i>SkARP</i>
<i>Dendrolycopodium obscurum</i>	1		*
<i>Diphasiastrum digitatum</i>	1		*
<i>Huperzia lucidula</i>	1		*
<i>Huperzia myrsinites</i>	1		*
<i>Huperzia selago</i>	1		*
<i>Huperzia squarrosa</i>	1		*
<i>Isoetes tegetiformans</i>	2	*	
<i>Lycopodiella appressa</i>	1		*
<i>Lycopodium annotinum</i>	3	*	
<i>Lycopodium deuterodensum</i>	1		*
<i>Pseudolycopodiella caroliniana</i>	1		*
<i>Selaginella acanthonota</i>	1		*
<i>Selaginella apoda</i>	1		*
<i>Selaginella kraussiana</i>	2	*	
<i>Selaginella lepidophylla</i>	1		*
<i>Selaginella selaginoides</i>	1		*
<i>Selaginella stauntoniana</i>	1		*
<i>Selaginella wallacei</i>	1		*
<i>Selaginella willdenowii</i>	1		*

Cuadro 4. Total de secuencias muestreadas por linaje. En la tercera columna se muestra únicamente el número de secuencias provenientes de genomas recuperadas por linaje y en la última columna se muestra el número de secuencias provenientes de transcriptomas recuperadas por linaje.

Linaje	Total de secuencias obtenidas (genomas y transcriptomas)	Secuencias de genomas	Secuencias de transcriptomas
Musgos	30	30	0
Licofitas	49	28	21
Helechos	151	99	52
Gimnospermas	68	66	2
Angiospermas	107	107	0

Cuadro 5. Características de los alineamientos y valores de $-\ln L$ con cada análisis para cada matriz .

Matriz	Número de Terminales	Longitud del alineamiento (nucleótidos)	% de datos faltantes	Número de sitios constantes	Log verosimilitud del árbol consenso (RaxML)	Log verosimilitud del árbol consenso (IQtree)
G+T	405	297	1.04%	15	-67570.054561	-67540.697
G	330	297	1.04%	15	-56946.663590	-56906.900

Discusión

Las hojas son un órgano de gran importancia en la historia de la vida en la tierra. Su aparición dio lugar a nuevas rutas evolutivas en la forma de las plantas, además de ocasionar grandes cambios ambientales (Beerling et al., 2001; Chanderbali et al., 2016; Endress & Doyle, 2009; Morris et al., 2018). El estudio comparativo de las redes de interacciones genéticas que controlan el desarrollo foliar es un enfoque que pretende entender mejor el origen y evolución de las hojas. El principal objetivo de este proyecto fue entender la evolución de los genes *ARP* en los helechos, con el fin de aportar evidencia que nos permita comprender mejor la evolución de las hojas en las eufilofitas.

Para poder determinar si existen ortólogos de los genes *ARP* en los helechos y entender la evolución de los genes *MYB* y *ARP* en las plantas vasculares, se realizaron análisis de IB y MV con la matriz G+T y la matriz G. Debido a que en los árboles de IB y MV de ambas matrices se recuperan los mismos clados pero las relaciones entre ellos cambian, decidimos tomar la filogenia obtenida con IB de la matriz G para guiar la discusión.

Evolución de los genes *MYB* en helechos

Los árboles filogenéticos obtenidos a partir de la matriz G con los análisis de IB y MV muestran las mismas relaciones internas de los clados a-u (Supl. 3, 5). El árbol obtenido con los análisis de IB tiene una politomía que abarca más de la mitad de los clados recuperados (Figura 13), esto puede deberse a que la longitud del dominio MYB es muy pequeña (297 nucleótidos) comparada con el número de terminales (330 terminales). Además, al ser un dominio se encuentra altamente conservado (96 caracteres conservados) (Stracke et al., 2001; Sun et al., 2002). A pesar de la poca resolución, nuestra filogenia recupera relaciones similares a las reportadas anteriormente entre los genes *MYB* de *Arabidopsis* (Bowman et al., 2017; Stracke et al., 2001).

A partir de la filogenia obtenida con IB con la matriz G, inferimos que existen al menos 26 copias de genes *MYB* en helechos (Figura 13 y 14); debido a la falta de resolución en los clados que forman la politomía c-q no es posible determinar copias adicionales en este linaje. Estos resultados nos muestran que la familia de genes *MYB* es también muy diversa en los helechos, similar a lo reportado previamente para angiospermas (Stracke et al., 2001).

Algunos clados están conformados exclusivamente por secuencias de helechos (por ejemplo, clados g, j), lo que podría estar indicando que estas copias de genes *MYB* son exclusivas de este grupo. Adicionar secuencias de otros linajes afines a las secuencias de estos clados de helechos ayudará a

entender mejor la historia evolutiva de estas copias y resolver si son realmente únicas a helechos o si se encontraban en el ancestro común de las plantas vasculares o de las plantas terrestres. Esto también permitirá generar hipótesis acerca de las posibles funciones de los genes *MYB* codificados por esas secuencias.

La filogenia de genes *MYB* muestra que existen clados que agrupan únicamente secuencias de plantas con semilla, como es el caso del clado ñ (Figura 14), esto podría indicar que la copia que dio lugar a esas secuencias apareció después de la divergencia de las eufilofitas. No es posible establecer con claridad que estas copias sean únicas de plantas con semilla ya que el clado ñ hace parte una politomía. Esto mismo sucede con los clados f, h, i, k, l, los cuales están compuestos exclusivamente de secuencias de *Arabidopsis*. Las filogenias realizadas con MV, al tener mayor resolución, podrían ayudar a tomar decisiones respecto a la localización de las duplicaciones que dieron lugar a estas copias de genes *MYB*, (Supl. 3 y 5).

Finalmente, los clados b, c, n, o, p, q, r, s están conformados por secuencias de musgos, licofitas, helechos y plantas con semilla (con excepción de los clados n, s; en los que solamente hay cuatro de estos cinco linajes), lo que indica que estas copias posiblemente ya estaban presentes en el ancestro de plantas terrestres. Las secuencias que conforman estos clados pueden ser más fácilmente relacionadas con alguna función ya que en ellos se agrupan secuencias de genes *MYB* de *Arabidopsis* cuya función ya ha sido reportada (Figura 13, Cuadro 1). Por ejemplo: podemos asumir que los genes correspondientes a las secuencias de helechos del clado c podrían estar regulando la producción de ácido abscísico y participan en la respuesta a estrés hídrico y osmótico. Hasta donde sabemos las filogenias de genes *MYB* no habían integrado secuencias de helechos, por lo que la comprensión de la evolución de esta familia de genes provenía de un panorama limitado. Hasta antes de este trabajo se desconocía la diversidad y las posibles funciones de los genes *MYB* en el linaje de los helechos.

Ausencia de genes *ARP* en helechos

Las diferentes filogenias realizadas con la matriz G+T permitieron determinar que no hay genes ortólogos de *ARP* en los helechos (Figura 14, Supl. 2 y 4 clado ARP). Este resultado sugiere que el módulo de desarrollo *KNOX Clase I-ARP* no es posible en helechos y que por lo tanto la red genética de desarrollo foliar en los helechos es diferente a la descrita en angiospermas. Lo anterior permite rechazar la hipótesis nula planteada en este trabajo (Figura 10), además de que concuerda con los datos paleobotánicos que indican que las hojas de eufilofitas no son homólogas (Corvez et al., 2012; Galtier, 2010).

Para explicar la ausencia de los genes *ARP* en los helechos proponemos dos hipótesis, la primera es que el ancestro de las plantas vasculares presentaba el gen *ARP* y que éste se perdió en el linaje de los helechos. Esto implicaría que las hojas de eufilofitas comparten un módulo genético de desarrollo y que en ese sentido son homólogas al menos a nivel molecular. Esto apoya trabajos previos realizados con otros genes, en los que se propone que existe un mecanismo molecular compartido en el desarrollo de hojas en todas las plantas vasculares, por ejemplo: *HD-ZIP Clase III* (Vasco et al., 2016) y *KNOX Clase I* (Harrison et al., 2005; Ruiz-Estévez et al., 2017; Sano et al., 2005).

Nuestros análisis filogenéticos con la matriz G+T además de demostrar la ausencia de genes *ARP* en helechos permiten inferir que estos genes también están ausentes en otros linajes de licofitas, ya que en el clado ARP de las filogenias construidas con la matriz G+T únicamente se recuperan secuencias de Selaginellales (Figura 14, Supl. 2 y 4 clado ARP).

Otra hipótesis para explicar la presencia de genes *ARP* en dos linajes lejanamente relacionados como son *Selaginella* y plantas con semilla, es que el gen *ARP* fue adquirido en alguno de estos dos linajes por medio de transferencia horizontal de genes (THG) desde un linaje a otro. Los estudios publicados por Davis et al. (2005) y Li et al. (2014) muestran que es posible una THG de un gen completo entre linajes profundamente divergentes. Sin embargo, actualmente no hay pruebas suficientes que apoyen esta última hipótesis. Tampoco resulta fácil aclarar la dirección en la que ocurrió la THG.

Para explicar la ausencia de genes ortólogos de *ARP* en los otros dos órdenes de licofitas, Lycopodiales e Isoetales, proponemos que el gen *ARP* apareció de manera independiente en las plantas con semilla y en Selaginellales. Esta hipótesis es menos probable y más difícil de probar, sin embargo, un estudio publicado por Weng & Noel (2013) señala que existió una evolución convergente de ciertos metabolitos entre Selaginellales y plantas con semilla, lo cual indica que es posible una evolución convergente de ciertos caracteres moleculares.

Es importante señalar que como no contamos con un genoma secuenciado de Licopodiales e Isoetales, todas las búsquedas en estos linajes se hicieron en transcriptomas. Lo que también podría explicar la ausencia de genes *ARP* para estos linajes en nuestras búsquedas. El estudio de los genomas de estas especies permitirá esclarecer la evolución de los genes *ARP* en licofitas.

Sustitución funcional de *ARP* en helechos

Puesto que la interacción antagónica KNOX-ARP no es posible en los helechos, resulta interesante conocer de qué manera los genes *KNOX Clase I* están regulados y cómo se lleva a cabo el establecimiento de los primordios foliares en este grupo. Las filogenias realizadas con la matriz G+T (Figura 13, Supl. 2 y 4) nos permitieron ubicar algunos genes que podrían estar cumpliendo la función de los genes *ARP* en el linaje de los helechos.

El clado b resultó ser el clado más cercanamente relacionado al clado ARP en los análisis de MV en las Etapas I, II y III (Supl. 2 y 4). De acuerdo con nuestra filogenia (Figura 13) hubo al menos una duplicación que dio lugar al gen *MYB125* y *ARP*, la cual tuvo lugar en la rama que soporta la politomía de los clados ARP, b y c-q. En el clado b se encuentra una secuencia de *Arabidopsis* correspondiente al gen *MYB125*, un gen también conocido como *DUO1* (Alves-Junior et al., 2009; Durbarry, 2005). Hasta el momento, la única función reportada para el gen *DUO1* es la formación de gametos masculinos en plantas (Figura 20) (Durbarry, 2005; Tair: Klepikova et al., 2016), por lo que la posible función de las secuencias de helechos agrupadas en este clado no es clara.

Por lo tanto, a pesar de que el clado b es el más cercanamente relacionado al clado ARP, las secuencias de helechos agrupadas en el clado b no son las principales candidatas para sustituir la función de ARP en helechos. Además, los análisis de MEME muestran que las secuencias del clado b no comparten el motivo identificado para los genes *ARP* (Figura 18 e). Por lo tanto, podríamos suponer que las proteínas correspondientes a cada uno de los clados no son funcionalmente similares, a pesar de que los modelos de motivos proteicos (Figura 19) no muestran diferencias notorias en la estructura tridimensional.

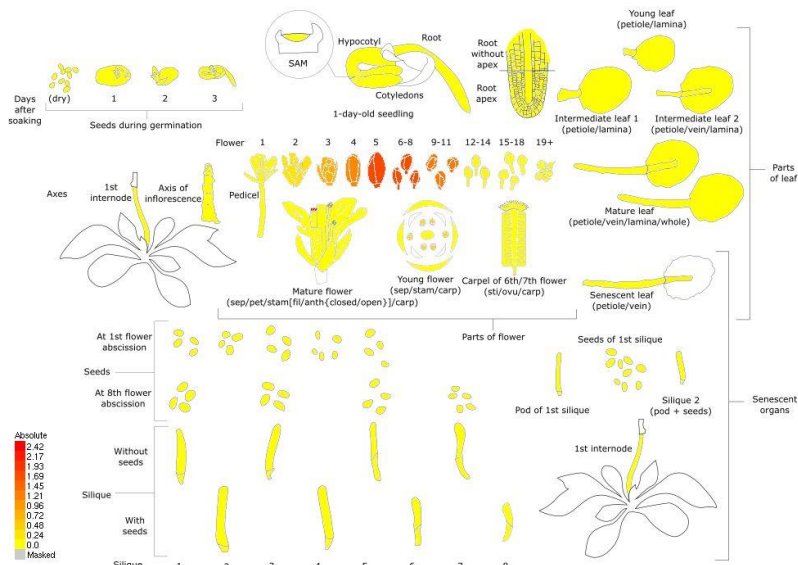


Figura 20. Expresión del gen *MYB125/DUO* durante la formación de los gametos masculinos en *Arabidopsis*. Este gen se agrupa en el clado b de todas las filogenias. Se ha reportado que este gen se expresa durante la formación de los gametos masculinos (Durbarry, 2005). Imagen tomada de Tair (Klepikova, Kasianov, Gerasimov, Logacheva, & Penin, 2016).

A pesar de que las secuencias del clado b no comparten el motivo característico de las proteínas ARP, es importante destacar que en estudios previos se ha determinado que en algunos eventos de duplicación y neofuncionalización genética las distintas funciones adquiridas en los genes duplicados, pueden deberse a cambios en las regiones regulatorias de los genes (Moore & Purugganan, 2005; Flagel & Wendel 2009).

Otros posibles genes candidatos para cumplir la función de los genes *ARP* en los helechos son los correspondientes a las secuencias de helechos agrupadas en los clados i, y o (Figura 13). Estos clados contienen secuencias de *Arabidopsis* que han sido anotadas como genes *MYB* involucrados en el desarrollo de órganos laterales o actividad meristemática (revisado en Ambawat et al., 2013; arabidopsis.org): *Arabidopsis_tha_MYB037*, *Arabidopsis_tha_MYB084* (clado o) y *Arabidopsis_tha_MYB92_AF062916_2* (clado i). Con base en las filogenias realizadas con MV (Supl. 3 y 5), determinamos que las secuencias de helechos más cercanamente relacionadas al clado i son las agrupadas en el clado g (conformado exclusivamente por secuencias helechos) y el clado o.

También resulta importante la posibilidad de que la función de *AS1* en los helechos esté cubierta por proteínas codificadas por otros genes de la red genética del desarrollo foliar (Figura 5). Estudios anteriores han mostrado que el gen *AS2* está involucrado en la misma vía de desarrollo que *AS1* en *Arabidopsis* (Sun et al., 2012). Por lo tanto, *AS2* podría sustituir la función de *AS1* en los helechos. Actualmente, no existen estudios que exploren si este gen se encuentra presente en helechos.

Las auxinas juegan un papel integral, especificando la filotaxia y la polaridad de las hojas entre algunos otros procesos de desarrollo en angiospermas (Byrne 2002). Se sabe que estas fitohormonas, actúan mediante una red de transducción, en la cual como reguladores negativos de los genes *KNOX*

Clase I en el SAM (Reinhardt et al., 2003; Zhao et al., 2010). La evidencia experimental indica que en las hojas de los helechos se produce una señalización similar entre el SAM y los primordios foliares, pero se conoce poco sobre la naturaleza molecular de estas señales en los helechos (revisado en Vasco et al. 2013). Se ha demostrado que las auxinas afectan la complejidad de la hoja en *Marsilea*, un helecho acuático, así como el desarrollo del primordio foliar (Allsopp, 1952; Steeves & Briggs, 1960). Por lo tanto las auxinas son un candidato probable para la regulación negativa de los genes *KNOX Clase I* en el SAM de los helechos.

Sería importante realizar estudios en helechos sobre la historia evolutiva y las funciones de algunos de los genes que promueven la biosíntesis de auxinas, como *YUCCA* y *TAA1* (Byrne, 2012), así como de los que motivan la expresión de *PINI* para formar los máximos de auxina, como es el caso de *PTL* (Prasad et al., 2011). Además de determinar cómo son los flujos de auxina y como se transporta.

El estudio de la formación de órganos laterales, y en especial de las hojas, ha sido ampliamente estudiado en angiospermas, principalmente en *Arabidopsis*. Sin embargo, recientemente el estudio de los genes involucrados en el desarrollo de hojas ha sido abordado en grupos tempranamente divergentes (Plackett et al., 2018). En el helecho *Ceratopteris richardii* (Pteridaceae), se ha demostrado recientemente que los genes *LFY* (Plackett et al., 2018) y *KNOX Clase II* se expresan en el SAM (Bharathan et al., 2002; Sano et al., 2005; Ambrose & Vasco, 2016). Además es importante mencionar que estudios realizados en otras especies de angiospermas *Pisum sativum* y *Medicago truncatula*, se ha mostrado que *KNOX Clase I* está excluida en los primordios foliares y en el desarrollo posterior de las hojas, y su papel ha sido remplazado por la proteína ortóloga de *LFY* (revisado en Wang et al., 2008). Respecto a la regulación por auxinas, en *Pisum sativum* y *Medicago truncatula* los genes *LFY*, son regulados negativamente por auxinas mediante una cascada de señalización (revisado en DeMason & Chawla, 2004), a diferencia de lo que ocurre en *Arabidopsis*. Por lo que el establecimiento de los primordios foliares y su desarrollo posterior podría depender principalmente de la interacción entre *LFY* y la señalización de las auxinas.

Esta evidencia nos permite pensar en la posibilidad de que los helechos hayan conservado la función superpuesta de los genes *KNOX Clase I* y *LFY* y que al igual que sucede en *Pisum sativum* y *Medicago truncatula*, donde el paso de un órgano indeterminado a un órgano determinado (tallo-hoja) pudiera estar regulado por la interacción por *LFY* y Auxinas y no por *KNOX Clase I*- ARP.

Identificación de nuevos genes *ARP* en *Selaginella* y *Ginkgo*

Las secuencias que se recuperaron en el clado ARP comparten el motivo ARP (Figura 18e), lo cual sugiere que esta región está involucrada en que las proteínas ARP cumplan su función. Hasta ahora se habían realizado alineamientos de aminoácidos los genes *ARP* de diferentes especies, y se habían marcado las regiones compartidas (Evkaikina et al., 2017), pero no se había reportado un motivo característico de proteínas ARP.

Aunque en el trabajo realizado por Evkaikina *et al.*, (2017) se había reportado una secuencia de *Pinus taeda* similar a una proteína ARP, en el clado ARP de nuestras filogenias no se agrupó ninguna secuencia de *Pinus taeda*. Esto puede deberse al esfuerzo de muestreo, pues nuestras búsquedas para encontrar genes ortólogos de genes *ARP* en gimnospermas fueron enfocadas principalmente en *Picea abies*, debido a que intentábamos reducir el número de terminales.

Modelo por homología de las proteínas del clado ARP y el clado b

La forma y función de las proteínas están inevitablemente ligadas, y la actividad de cada una de las proteínas existentes depende de la capacidad que esta tenga para unirse específicamente a otras proteínas o moléculas (Alberts et al., 2004). Esto resalta la relevancia de conocer si la estructura tridimensional de las proteínas ARP y de las proteínas del clado b eran iguales.

La diferencia en la secuencia de aminoácidos de las proteínas del clado ARP y el clado b dada por el motivo ARP (Figura 12 y 18), sugería que la estructura terciaria de ambas proteínas sería diferente. Sin embargo, los modelos de estructura terciaria de las proteínas de ambos clados (ARP y b), nos muestran que no es así (Figura 19). Estos análisis nos permitieron observar que no hay grandes cambios en las hélices de reconocimiento (tercera hélice de cada repetición) entre los genes del clado ARP y los genes del clado b.

La estructura consenso de la repetición R3 de *Arabidopsis* tiene una fenilalanina en el lugar del primer triptófano (Stracke et al., 2001). En el alineamiento de la matriz G+T (Figura 12), se observa que además de fenilalanina, este triptófano puede ser reemplazado por una leucina, isoleucina o valina. Esta alteración en la secuencia primaria de la proteína podría modificar su estructura terciaria. Sin embargo, los modelos de estructura terciaria de las proteínas de los clados ARP y b, no mostraron diferencias notorias en la repetición R3 (Figura 19). Esto puede deberse a que los aminoácidos que sustituyen al triptófano como leucina e isoleucina en la repetición R3 son de la misma naturaleza

que el triptófano (no polar), por lo tanto, las interacciones químicas entre las tres alfa hélices de esta repetición no se alteren.

Los modelos de las proteínas ARP obtenidos en este trabajo, son el primer registro que se tiene de la estructura terciaria de *PHAN*, *ASI* y *RS2*. En la base de datos de SWISS- MODEL existen modelos de otras proteínas MYB, pero no de las pertenecientes al clado ARP.

Conclusiones

- No se encontraron genes ortólogos de *ARP* en helechos en las bases de datos revisadas en 151 secuencias progenientes de 18 especies, por lo que proponemos que los helechos y angiospermas no comparten el módulo de desarrollo foliar *KNOX Clase I- ARP*.
- Encontramos que existen al menos 27 copias de genes *MYB* en helechos.
- Los genes *ARP* presentan el motivo característico A(S)E(Q)VPGR(T)TAKRLG(S)KWWEVY(F)KER(K) (motivo ARP)
- En las proteínas ARP, la sustitución del primer triptófano (W) en la repetición R3 por otro aminoácido hidrofóbico no altera la estructura tridimensional característica de una proteína con dominio MYB de *Arabidopsis*, según los análisis de modelado realizados.
- Se identificaron nuevos ortólogos de genes *ARP* en ocho especies de *Selaginella* y en dos de gimnospermas (*Picea abies* y *Ginkgo biloba*).

El presente trabajo constituye la primera aproximación al estudio evolutivo de los genes *ARP* en los helechos. Este trabajo demuestra que el módulo de desarrollo foliar *KNOX Clase I- ARP*, reportado en angiospermas y en *Selaginella*, no existe en helechos. Por esta razón este módulo pudo haber sido reclutado o coptado de forma independiente a lo largo de la evolución de las plantas terrestres en dos linajes no relacionados. O bien, el módulo genético *KNOX Clase I- ARP* pudo haber estado presente en el ancestro común de plantas vasculares y posteriormente fue perdido en el linaje de los helechos. Ambas posibilidades nos permiten inferir que el desarrollo de las hojas megáfilas no es homólogo. Las filogenias elaboradas con el dominio MYB, muestran que existen otros genes *MYB* en helechos, los cuales hacen parte de clados que incluyen secuencias de *Arabidopsis* para las cuales se han reportado funciones que tienen que ver con desarrollo de hoja. Esto abre la posibilidad de explorar más a fondo la historia evolutiva de estas copias y probarlas funcionalmente en helechos, para entender si estas copias están involucradas en la red de desarrollo de hojas en helechos.

Perspectivas

- El gen *AS2*, y los genes de helechos del clado b, o, i son candidatos para cumplir la función de *ARP* en los helechos. Otro candidato más para cumplir esta función sería el otro miembro del módulo, el gen *JLO*. Por lo que sería importante realizar una búsqueda similar a la que se hizo en el presente trabajo, para estos dos genes.
- Las auxinas son las fitohormonas candidatas para regular la actividad de *KNOX Clase I* o *LFY*. Por lo que también se podrían realizar pruebas histológicas o funcionales que probaran los niveles de auxinas en los primordios foliares de helechos.
- Las genes *MYB* de helechos agrupados en clados diferentes al *ARP*, pudieran cumplir su función a través de sus regiones regulatorias. Por esta razón resulta importante analizar si las regiones regulatorias de los diferentes *MYB* se encuentran conservadas.
- Realizar pruebas funcionales a partir de los genes candidatos encontrados, en los que el gen candidato se utilizaría para complementar a mutantes de *Arabidopsis* .

Referencias

- Alberts, B., Bray, D., Hopkin, K., Johnson, A., Lewis J., Raff, M., Roberts, K., Walter, P., (2004). *Introducción a la Biología Celular*, Madrid, España: Editorial Médica Panamericana.
- Allsopp, A. (1952). Experimental and Analytical Studies of Pteridophytes XXIV. Investigations on Marsilea. *Annals of Botany*, 18(4), 449–450. <https://doi.org/10.1093/oxfordjournals.aob.a083408>
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Alves-Junior, L., Niemeier, S., Hauenschield, A., Rehmsmeier, M., & Merkle, T. (2009). Comprehensive prediction of novel microRNA targets in *Arabidopsis thaliana*. *Nucleic Acids Research*, 37(12), 4010–4021. <https://doi.org/10.1093/nar/gkp272>
- Ambrose, B. A., & Vasco, A. (2016). Bringing the multicellular fern meristem into focus. *New Phytologist*, 210(3), 790–793. <https://doi.org/10.1111/nph.13825>
- Becker, B., & Marin, B. (2009). Streptophyte algae and the origin of embryophytes. *Annals of Botany*, 103(7), 999–1004. <https://doi.org/10.1093/aob/mcp044>
- Beerling, D. J., & Berner, R. A. (2005). Feedbacks and the coevolution of plants and atmospheric CO₂. *Proceedings of the National Academy of Sciences*, 102(5), 1302–1305. <https://doi.org/10.1073/pnas.0408724102>
- Beerling, D. J., & Fleming, A. J. (2007). Zimmermann's telome theory of megaphyll leaf evolution: a molecular and cellular critique. *Current Opinion in Plant Biology*, 10(1), 4–12. <https://doi.org/10.1016/J.PBI.2006.11.006>
- Beerling, D. J., Osborne, C. P., & Chaloner, W. G. (2001). Evolution of leaf-form in land plants linked to atmospheric CO₂ decline in the Late Palaeozoic era. *Nature*, 410(6826), 352–354. <https://doi.org/10.1038/35066546>
- Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2017). GenBank. *Nucleic Acids Research*, 45(D1), D37–D42. <https://doi.org/10.1093/nar/gkw1070>
- Bharathan, G., Goliber, T. E., Moore, C., Kessler, S., Pham, T., & Sinha, N. R. (2002). Homologies in leaf form inferred from *KNOX1* gene expression during development. *Science*, 296(June), 1858–1860. <https://doi.org/10.1126/science.1070343>
- Borevitz, J. O., Xia, Y., Blount, J., Dixon, R. A., & Lamb, C. (2000). Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *The Plant Cell*, 12(12), 2383–2394. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11148285>
- Bowman, J. L., Eshed, Y., & Baum, S. F. (2002). Establishment of polarity in angiosperm lateral organs. *Trends in Genetics*, 18(3), 134–141. [https://doi.org/10.1016/S0168-9525\(01\)02601-4](https://doi.org/10.1016/S0168-9525(01)02601-4)
- Bowman, J. L., Kohchi, T., Yamato, K. T., Jenkins, J., Shu, S., Ishizaki, K., ... Schmutz, J. (2017). Insights into Land Plant Evolution Garnered from the *Marchantia polymorpha* Genome. *Cell*, 171(2), 287–304.e15. <https://doi.org/10.1016/j.cell.2017.09.030>
- Byrne, M. E. (2012). Making leaves. *Current Opinion in Plant Biology*, 15(1), 24–30. <https://doi.org/10.1016/j.pbi.2011.10.009>
- Byrne, M. E., Barley, R., Curtis, M., Arroyo, J. M., Dunham, M., Hudson, A., & Martienssen, R. A. (2000). Asymmetric leaves1 mediates leaf patterning and stem cell function in *Arabidopsis*. *Nature*, 408(6815), 967–971. <https://doi.org/10.1038/35050091>
- Byrne, M. E., Simorowski, J., & Martienssen, R. A. (2002). ASYMMETRIC LEAVES1 reveals knox gene redundancy in *Arabidopsis*. *Development (Cambridge, England)*, 129(8), 1957–1965. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11934861>
- Carabelli, M., Possenti, M., Sessa, G., Ciolfi, A., Sassi, M., Morelli, G., & Ruberti, I. (2007). Canopy shade causes a rapid and transient arrest in leaf development through auxin-induced cytokinin oxidase activity. *Genes & Development*, 21(15), 1863–1868. <https://doi.org/10.1101/gad.432607>
- Chanderbali, A. S., Berger, B. A., Howarth, D. G., Soltis, P. S., & Soltis, D. E. (2016). Evolving ideas on the origin and evolution of flowers: New perspectives in the genomic era. *Genetics*, 202(4), 1255–1265. <https://doi.org/10.1534/genetics.115.182964>
- Cohen K. M., Finney S.C. y Gibbard P.L. 2015. International Commission on Stratigraphy. Consultado en: <<http://www.stratigraphy.org/ICSChart/ChronostratChart2015-01Spanish.pdf> > enero 2019

- Corvez, A., Barriel, V., & Dubuisson, J. Y. (2012). Diversity and evolution of the megaphyll in Euphyllophytes: Phylogenetic hypotheses and the problem of foliar organ definition. *Comptes Rendus - Palevol*, 11(6), 403–418. <https://doi.org/10.1016/j.crvp.2012.05.003>
- Davis, C. C., Anderson, W. R., & Wurdack, K. J. (2005). Gene transfer from a parasitic flowering plant to a fern. *Proceedings of the Royal Society B: Biological Sciences*, 272(1578), 2237–2242. <https://doi.org/10.1098/rspb.2005.3226>
- DeMason, D. A., & Chawla, R. (2004). Roles for auxin during morphogenesis of the compound leaves of pea (*Pisum sativum*). *Planta*, 218(3), 435–448. <https://doi.org/10.1007/s00425-003-1100-x>
- Dkhar, J., & Pareek, A. (2014). What determines a leaf's shape? *EvoDevo*, 5(1), 47. <https://doi.org/10.1186/2041-9139-5-47>
- Drummond, A. J., & Rambaut, A. (2015). Bayesian evolutionary analysis by sampling trees. *Bayesian Evolutionary Analysis with BEAST*, 8, 79–96. <https://doi.org/10.1017/CBO9781139095112.007>
- Du, H., Feng, B.-R., Yang, S.-S., Huang, Y.-B., & Tang, Y.-X. (2012). The R2R3-MYB Transcription Factor Gene Family in Maize. *PLoS ONE*, 7(6), e37463. <https://doi.org/10.1371/journal.pone.0037463>
- Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., & Lepiniec, L. (2010). MYB transcription factors in Arabidopsis. *Trends in Plant Science*, 15(10), 573–581. <https://doi.org/10.1016/j.tplants.2010.06.005>
- Durbarry, A. (2005). Male Germ Line Development in Arabidopsis. duo pollen Mutants Reveal Gametophytic Regulators of Generative Cell Cycle Progression. *Plant Physiology*, 137(1), 297–307. <https://doi.org/10.1104/pp.104.053165>
- Endress, P. K., & Doyle, J. A. (2009). Reconstructing the ancestral angiosperm flower and its initial specializations. *American Journal of Botany*, 96(1), 22–66. <https://doi.org/10.3732/ajb.0800047>
- Evkaikina, A. I., Berke, L., Romanova, M. A., Proux-Wéra, E., Ivanova, A. N., Rydin, C., ... Voitsekhovskaja, O. V. (2017). The *Huperzia selago* shoot tip transcriptome sheds new light on the evolution of leaves. *Genome Biology and Evolution*, 9(9), 2444–2460. <https://doi.org/10.1093/gbe/evx169>
- Flagel L. E., & Wendel J. F. (2009). Gene duplication and evolutionary novelty in plants. *New Phytologist*, 183(3), 557–564.
- Galtier, J. (2010). The Origins and Early Evolution of the Megaphyllous Leaf. *International Journal of Plant Sciences*, 171(6), 641–661. <https://doi.org/10.1086/653130>
- Gao, J., Yang, X., Zhao, W., Lang, T., & Samuelsson, T. (2015). Evolution, diversification, and expression of KNOX proteins in plants. *Frontiers in Plant Science*, 6, 882. <https://doi.org/10.3389/fpls.2015.00882>
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., ... Rokhsar, D. S. (2012). Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Research*, 40(D1), 1178–1186. <https://doi.org/10.1093/nar/gkr944>
- Grimplet, J., Pimentel, D., Agudelo-Romero, P., Martinez-Zapater, J. M., & Fortes, A. M. (2017). The LATERAL ORGAN BOUNDARIES Domain gene family in grapevine: genome-wide characterization and expression analyses during developmental processes and stress responses. *Scientific Reports*, 7(1), 15968. <https://doi.org/10.1038/s41598-017-16240-5>
- Guo, M., Thomas, J., Collins, G., & Timmermans, M. C. P. (2008). Direct Repression of KNOX Loci by the ASYMMETRIC LEAVES1 Complex of Arabidopsis. *THE PLANT CELL ONLINE*, 20(1), 48–58. <https://doi.org/10.1105/tpc.107.056127>
- Harrison, C. J., & Morris, J. L. (2018). The origin and early evolution of vascular plant shoots and leaves. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 373(1739). <https://doi.org/10.1098/rstb.2016.0496>
- Harrison, C. J., Coriey, S. B., Moylan, E. C., Alexander, D. L., Scotland, R. W., & Langdale, J. A. (2005). Independent recruitment of a conserved developmental mechanism during leaf evolution. *Nature*, 434(7032), 509–514. <https://doi.org/10.1038/nature03410>
- Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Le, S. V. (2007). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, 35(2), msx281. <https://doi.org/10.5281/zenodo.854445>
- Iwakawa, H., Iwasaki, M., Kojima, S., Ueno, Y., Soma, T., Tanaka, H., ... Machida, C. (2007). Expression of the ASYMMETRIC LEAVES2 gene in the adaxial domain of Arabidopsis leaves represses cell proliferation in this domain and is critical for the development of properly expanded leaves. *The Plant Journal*, 51(2), 173–184. <https://doi.org/10.1111/j.1365-313X.2007.03132.x>

- Jasinski, S., Piazza, P., Craft, J., Hay, A., Woolley, L., Rieu, I., ... Tsiantis, M. (2005). KNOX Action in Arabidopsis Is Mediated by Coordinate Regulation of Cytokinin and Gibberellin Activities. *Current Biology*, 15(17), 1560–1565. <https://doi.org/10.1016/j.cub.2005.07.023>
- Jin, H., & Martin, C. (1999). Multifunctionality and diversity within the plant MYB-gene family. *Plant Molecular Biology*, 41(5), 577–585. <https://doi.org/10.1023/A:1006319732410>
- Jin, H., Cominelli, E., Bailey, P., PaBeerling, D. J., & Berner, R. A. (2005). Feedbacks and the coevolution of plants and atmospheric CO₂. *Proceedings of the National Academy of Sciences*, 102(5), 1302–1305. <https://doi.org/10.1073/pnas.0408724102>
- Katiyar, A., Smita, S., Lenka, S. K., Rajwanshi, R., Chinnusamy, V., & Bansal, K. C. (2012). Genome-wide classification and expression analysis of MYB transcription factor families in rice and Arabidopsis. *BMC Genomics*, 13, 544. <https://doi.org/10.1186/1471-2164-13-544>
- Katoh, K., Rozewicki, J., & Yamada, K. D. (2017). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*. <https://doi.org/10.1093/bib/bbx108>
- Kenrick, P., & Crane, P. R. (1997). The origin and early evolution of plants on land. *Nature*, 389(6646), 33–39. <https://doi.org/10.1038/37918>
- Kidner, C. A., & Timmermans, M. C. P. (2010). Signaling sides. Adaxial-abaxial patterning in leaves. *Current Topics in Developmental Biology*, 91(C), 141–168. [https://doi.org/10.1016/S0070-2153\(10\)91005-3](https://doi.org/10.1016/S0070-2153(10)91005-3)
- Klepikova, A. V., Kasianov, A. S., Gerasimov, E. S., Logacheva, M. D., & Penin, A. A. (2016). A high resolution map of the Arabidopsis thaliana developmental transcriptome based on RNA-seq profiling. *Plant Journal*, 88(6), 1058–1070. <https://doi.org/10.1111/tpj.13312>
- Kurakawa, T., Ueda, N., Maekawa, M., Kobayashi, K., Kojima, M., Nagato, Y., ... Kyojuka, J. (2007). Direct control of shoot meristem activity by a cytokinin-activating enzyme. *Nature*, 445(7128), 652–655. <https://doi.org/10.1038/nature05504>
- Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., ... Huala, E. (2012). The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Research*, 40(D1), 1202–1210. <https://doi.org/10.1093/nar/gkr1090>
- Langdale, J. A. (2008). Evolution of developmental mechanisms in plants. *Current Opinion in Genetics and Development*, 18(4), 368–373. <https://doi.org/10.1016/j.gde.2008.05.003>
- Langdale, J. A., & Harrison, C. J. (2008). Developmental transitions during the evolution of plant form. *Evolving Pathways: Key Themes in Evolutionary Developmental Biology*, 299–315. <https://doi.org/10.1017/CBO9780511541582.021>
- Lewis, L. A., & McCourt, R. M. (2004). Green algae and the origin of land plants. *American Journal of Botany*, 91(10), 1535–1556. <https://doi.org/10.3732/ajb.91.10.1535>
- Li, F. W., Brouwer, P., Carretero-Paulet, L., Cheng, S., de Vries, J., Delaux, P. M., ... Pryer, K. M. (2018). Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nature Plants*, 4(July), 1–13. <https://doi.org/10.1038/s41477-018-0188-8>
- Li, F. W., Villarreal, J. C., Kelly, S., Rothfels, C. J., Melkonian, M., Frangedakis, E., ... Pryer, K. M. (2014). Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns. *Proceedings of the National Academy of Sciences*, 111(18), 6672–6677. <https://doi.org/10.1073/pnas.1319929111>
- Lodha, M., Marco, C. F., & Timmermans, M. C. P. (2013). The ASYMMETRIC LEAVES complex maintains repression of KNOX homeobox genes via direct recruitment of Polycomb-repressive complex2. *Genes & Development*, 27(6), 596–601. <https://doi.org/10.1101/gad.211425.112>
- Long, J. A., Moan, E. I., Medford, J. I., & Barton, M. K. (1996). A member of the KNOTTED class of homeodomain proteins encoded by the STM gene of Arabidopsis. *Nature*, 379(6560), 66–69. <https://doi.org/10.1038/379066a0>
- Maddison, W. P. & D.R. Maddison. 2018. Mesquite: a modular system for evolutionary analysis. Version 3.51 <http://www.mesquiteproject.org>
- Majer, C., & Hochholdinger, F. (2011). Defining the boundaries: Structure and function of LOB domain proteins. *Trends in Plant Science*, 16(1), 47–52. <https://doi.org/10.1016/j.tplants.2010.09.009>
- Matsumura, Y., Iwakawa, H., MacHida, Y., & MacHida, C. (2009). Characterization of genes in the ASYMMETRIC LEAVES2-LATERAL ORGAN BOUNDARIES (AS2-LOB) family in Arabidopsis thaliana, and functional and molecular comparisons between AS2 and other family members. *Plant Journal*, 58(3), 525–537. <https://doi.org/10.1111/j.1365-313X.2009.03797.x>

- Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. 2010 Gateway Computing Environments Workshop, GCE 2010. <https://doi.org/10.1109/GCE.2010.5676129>
- Minh, B. Q., Nguyen, M. A. T., & Von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*, 30(5), 1188–1195. <https://doi.org/10.1093/molbev/mst024>
- Moore, R. C., & Purugganan, M. D. (2005). The evolutionary dynamics of plant duplicate genes. *Current Opinion in Plant Biology*, 8(2), 122–128. <https://doi.org/10.1016/j.pbi.2004.12.001>
- Morris, J. L., Puttick, M. N., Clark, J. W., Edwards, D., Kenrick, P., Pressel, S., ... Donoghue, P. C. J. (2018). The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences*, 201719588. <https://doi.org/10.1073/pnas.1719588115>
- Myrset, A. H., Bostad, A., Jamin, N., Lirsac, P. N., Toma, F., & Gabrielsen, O. S. (1993). DNA and redox state induced conformational changes in the DNA-binding domain of the Myb oncoprotein. *The EMBO Journal*, 12(12), 4625–4633. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8223472>
- N, M., LH, H., Z, Y., EJ, C., NJ, W., S, M., ... GK, W. (2015). Data access for the 1,000 Plants (1KP) project. *GigaScience*, 3, 17. <https://doi.org/10.1186/2047-217X-3-17>
- Nylander, J. A. A., 2004. MrModeltest v2. Programa distribuido por el autor. Centro de Biología Evolutiva, Universidad de Uppsala. Disponible en: https://www.researchgate.net/publication/285805344_MrModeltest_V2_Program_Distributed_by_the_Author
- Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y. C., Scofield, D. G., ... Jansson, S. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature*, 497(7451), 579–584. <https://doi.org/10.1038/nature12211>
- Ogata, K., Hojo, H., Aimoto, S., Nakai, T., Nakamura, H., Sarai, A., ... Nishimura, Y. (1992). Solution structure of a DNA-binding unit of Myb: a helix-turn-helix-related motif with conserved tryptophans forming a hydrophobic core. *Proceedings of the National Academy of Sciences of the United States of America*, 89(14), 6428–6432. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1631139>
- Oppenheimer, D. G., Herman, P. L., Sivakumaran, S., Esch, J., & Marks, M. D. (1991). A myb gene required for leaf trichome differentiation in *Arabidopsis* is expressed in stipules. *Cell*, 67(3), 483–493. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1934056>
- Peters, B., Casey, J., Twell, D., Zohrab, S., Brownfield, L., Borg, M., & Aidley, J. (2016). A Conserved cis-Regulatory Module Determines Germline Fate through Activation of the Transcription Factor DUO1 Promoter. *Plant Physiology*, 173(1), 280–293. <https://doi.org/10.1104/pp.16.01192>
- Pires, N. D., & Dolan, L. (2012). Morphological evolution in land plants: New designs with old genes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1588), 508–518. <https://doi.org/10.1098/rstb.2011.0252>
- Plackett, A. R. G., Conway, S. J., Hazelton, K. D. H., Rabbinowitsch, E. H., Langdale, J. A., & Di Stilio, V. S. (2018). LEAFY maintains apical stem cell activity during shoot development in the fern *Ceratopteris richardii*. *eLife*, 7, 1–34. <https://doi.org/10.7554/eLife.39625>
- Plackett, A. R. G., Di Stilio, V. S., & Langdale, J. A. (2015). Ferns: the missing link in shoot evolution and development. *Frontiers in Plant Science*, 6, 972. <https://doi.org/10.3389/fpls.2015.00972>
- PPG I. (2016). A community-derived classification for extant lycophytes and ferns. *Journal of Systematics and Evolution*, 54(6), 563–603. <https://doi.org/10.1111/jse.12229>
- Prasad, K., Grigg, S. P., Barkoulas, M., Yadav, R. K., Sanchez-Perez, G. F., Pinon, V., ... Scheres, B. (2011). *Arabidopsis* PLETHORA transcription factors control phyllotaxis. *Current Biology*, 21(13), 1123–1128. <https://doi.org/10.1016/j.cub.2011.05.009>
- Pryer, K. M., Schneider, H., Smith, A. R., Cranfill, R., Wolf, P. G., Hunt, J. S., & Sipes, S. D. (2001). Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature*, 409(6820), 618–621. <https://doi.org/10.1038/35054555>
- Rambaut Research Group. (n.d.). FigTree. Retrieved August 25, 2018, from <http://tree.bio.ed.ac.uk/software/figtree/>
- Rast, M. I., & Simon, R. (2012). *Arabidopsis* JAGGED LATERAL ORGANS Acts with ASYMMETRIC LEAVES2 to Coordinate KNOX and PIN Expression in Shoot and Root Meristems. *The Plant Cell*, 24(7), 2917–2933. <https://doi.org/10.1105/tpc.112.099978>

- Raubenson, L. A., & Jansen, R. K. (1992). Chloroplast DNA Evidence on the Ancient Evolutionary Split in Vascular Land Plants. *Science*, 255(5052), 1697–1699. <https://doi.org/10.1126/science.255.5052.1697>
- Reinhardt, D., Pesce, E. R., Stieger, P., Mandel, T., Baltensperger, K., Bennett, M., ... Kuhlemeier, C. (2003). Regulation of phyllotaxis by polar auxin transport. *Nature*, 426(6964), 255–260. <https://doi.org/10.1038/nature02081>
- Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., ... Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61(3), 539–542. <https://doi.org/10.1093/sysbio/sys029>
- Rothwell, G. W., & Nixon, K. C. (2006). How Does the Inclusion of Fossil Data Change Our Conclusions about the Phylogenetic History of Euphyllophytes? *International Journal of Plant Sciences*, 167(3), 737–749. <https://doi.org/10.1086/503298>
- rr, A., Mehrtens, F., Jones, J., ... Martin, C. (2000). Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in Arabidopsis. *The EMBO Journal*, 19(22), 6150–6161. <https://doi.org/10.1093/emboj/19.22.6150>
- Ruiz-Estévez, M., Bakkali, M., Martín-Blázquez, R., & Garrido-Ramos, M. A. (2017). Identification and Characterization of TALE Homeobox Genes in the Endangered Fern *Vandenboschia speciosa*. *Genes*, 8(10). <https://doi.org/10.3390/genes8100275>
- Sano, R., Juarez, C. M., Hass, B., Sakakibara, K., Ito, M., Banks, J. A., & Hasebe, M. (2005). KNOX homeobox genes potentially have similar function in both diploid unicellular and multicellular meristems, but not in haploid meristems. *Evolution and Development*, 7(1), 69–78. <https://doi.org/10.1111/j.1525-142X.2005.05008.x>
- Semiarti, E., Ueno, Y., Tsukaya, H., Iwakawa, H., Machida, C., & Machida, Y. (2001). The ASYMMETRIC LEAVES2 gene of *Arabidopsis thaliana* regulates formation of a symmetric lamina, establishment of venation and repression of meristem-related homeobox genes in leaves. *Development (Cambridge, England)*, 128(10), 1771–1783.
- Sessa, E. B., Banks, J. A., Barker, M. S., Der, J. P., Duffy, A. M., Graham, S. W., ... Wolf, P. G. (2014). Between Two Fern Genomes. *GigaScience*, 3(1), 1–7. <https://doi.org/10.1186/2047-217X-3-15>
- Siegfried, K. R., Eshed, Y., Baum, S. F., Otsuga, D., Drews, G. N., & Bowman, J. L. (1999). Members of the YABBY gene family specify abaxial cell fate in Arabidopsis. *Development (Cambridge, England)*, 126(18), 4117–4128. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10457020>
- Smith, A. R., Pryer, K. M., Schuettpelz, E., Korall, P., Schneider, H., & Wolf, P. G. (2006). A classification for extant ferns. *Taxon*, 55(3), 705–731. <https://doi.org/10.2307/25065646>
- Srivastava A., Chandra S. (2009). Structure and organization of the rhizome vascular system of four Polypodium species. *American Fern Journal*. 99 (3): 182-193
- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Steeves, T. A., & Briggs, W. R. (1960). Morphogenetic Studies on *Osmunda cinnamomea* L. *Journal of Experimental Botany*, 11(1), 45–67. <https://doi.org/10.1093/jxb/11.1.45>
- Stracke, R., Werber, M., & Weisshaar, B. (2001). The R2R3-MYB gene family in *Arabidopsis thaliana*. *Current Opinion in Plant Biology*, 4(5), 447–456. [https://doi.org/10.1016/S1369-5266\(00\)00199-0](https://doi.org/10.1016/S1369-5266(00)00199-0)
- Sun, S. B., Song, J. P., & Meng, L. S. (2012). ASYMMETRIC LEAVES2 gene, a member of LOB/AS2 family of *Arabidopsis thaliana*, causes an abaxializing leaves in transgenic cockscomb. *Molecular Biology Reports*, 39(4), 4927–4935. <https://doi.org/10.1007/s11033-011-1288-y>
- Sun, Y., Zhou, Q., Zhang, W., Fu, Y., & Huang, H. (2002). ASYMMETRIC LEAVES1, an Arabidopsis gene that is involved in the control of cell differentiation in leaves. *Planta*, 214(5), 694–702. <https://doi.org/10.1007/s004250100673>
- Timothy L. Bailey, & Charles Elkan. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, 28–36. Retrieved from <http://www.sdsc.edu/~tbailey/papers/ismb94.pdf>
- Tomescu, A. M. F. (2009). Megaphylls, microphylls and the evolution of leaf development. *Trends in Plant Science*, 14(1), 5–12. <https://doi.org/10.1016/j.tplants.2008.10.008>
- Vasco, A., Moran, R. C., & Ambrose, B. A. (2013). The evolution, morphology, and development of fern leaves. *Frontiers in Plant Science*, 4(September), 1–16. <https://doi.org/10.3389/fpls.2013.00345>

- Vasco, A., Smalls, T. L., Graham, S. W., Cooper, E. D., Wong, G. K. S., Stevenson, D. W., ... Ambrose, B. A. (2016). Challenging the paradigms of leaf evolution: Class III HD-Zips in ferns and lycophytes. *New Phytologist*, 212(3), 745–758. <https://doi.org/10.1111/nph.14075>
- Wang, H., Chen, J., Wen, J., Tadege, M., Li, G., Liu, Y., ... Chen, R. (2008). Control of Compound Leaf Development by FLORICAULA/LEAFY Ortholog SINGLE LEAFLET1 in *Medicago truncatula*. *Plant Physiology*, 146(4), 1759–1772. <https://doi.org/10.1104/pp.108.117044>
- Wardlaw, C. W. (1963). Apical organization and differential growth in ferns. *Journal of the Linnean Society of London, Botany*, 58(373), 385–400. <https://doi.org/10.1111/j.1095-8339.1990.tb00909.x>
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., ... Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46(W1), W296–W303. <https://doi.org/10.1093/nar/gky427>
- Wellman, C. H., Osterloff, P. L., & Mohiuddin, U. (2003). Fragments of the earliest land plants. *Nature*, 425(6955), 282–285. <https://doi.org/10.1038/nature01884>
- Weng, J.-K., & Noel, J. P. (2013). Chemodiversity in *Selaginella*: a reference system for parallel and convergent metabolic evolution in terrestrial plants. *Frontiers in Plant Science*, 4(May), 1–17. <https://doi.org/10.3389/fpls.2013.00119>
- Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., ... Leebens-Mack, J. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences*, 111(45), E4859–E4868. <https://doi.org/10.1073/pnas.1323926111>
- Xu, C., Luo, F., & Hochholdinger, F. (2016). LOB Domain Proteins: Beyond Lateral Organ Boundaries. *Trends in Plant Science*, 21(2), 159–167. <https://doi.org/10.1016/J.TPLANTS.2015.10.010>
- Xu, L., & Shen, W.-H. (2008). Polycomb Silencing of KNOX Genes Confines Shoot Stem Cell Niches in *Arabidopsis*. *Current Biology*, 18(24), 1966–1971. <https://doi.org/10.1016/j.cub.2008.11.019>
- Yoshida, S., Mandel, T., & Kuhlemeier, C. (2011). Stem cell activation by light guides plant organogenesis. *Genes & Development*, 25(13), 1439–1450. <https://doi.org/10.1101/gad.631211>
- Zhao, Z., Andersen, S. U., Ljung, K., Dolezal, K., Miotk, A., Schultheiss, S. J., & Lohmann, J. U. (2010). Hormonal control of the shoot stem-cell niche. *Nature*, 465(7301), 1089–1092. <https://doi.org/10.1038/nature09126>

Material suplementario

Suplementario 1. Tabla de todas las secuencias empleadas en los análisis filogenéticos. Columna uno, especie a la que corresponde la secuencia. Columna dos, nombre asignado en los análisis filogenéticos. Columna tres, código con el que puede ser encontrada esa secuencia en la base de datos de donde fue tomada. Columna cuatro, base de datos de donde se obtuvo la secuencia. Columna cinco, origen de la secuencia: genoma o transcriptoma. Columna seis, etapa de las búsquedas en la que se obtuvo la secuencia.

Especie	Nombre en el análisis	Código de identificación	Base de datos	genoma/ transcriptoma	Etapa
<i>Angiopteris evecta</i>	Angiopteris_eve_NHCM_2012261	NHCM_scaffold_2012261	One KP	transcriptoma	Etapa I
	Angiopteris_eve_NHCM_2065362	NHCM_scaffold_2065362	One KP	transcriptoma	Etapa I
	<i>Antirrhinum majus</i>	Antirrhinum_maj_PHAN1	AJ005586.1	NCBI / TAIR	genoma
<i>Arabidopsis thaliana</i>	Arabidopsis_tha_AB493533_1	AB493533.1	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_AB493793_1	AB493793.1	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_AS1	At2g37630	NCBI / TAIR	genoma	Etapa I
	Arabidopsis_tha_AY062619_1	AY062619.1	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_AY081512_1	AY081512.1	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB000	AY519590.1	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB001	NM_111757.4	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB002	NM_130287.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB003	NM_111525.4	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB004	NM_120023.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB005	NM_112200.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB009	NM_121683.4	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB010	NM_112118.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB011	NM_116126.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB012	NM_130314.4	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB013	NM_100499.2	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB014	NM_128674.4	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB016	NM_121535.4	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB017	NM_115989.4	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB018	NM_118688.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB019	NM_124605.4	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB020	NM_105294.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB022	NM_123407.1	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB023	NM_123397.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB025	NM_129546.2	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB026	NM_001084678.2	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB027	NM_124340.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB028	NM_130287.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB029	NM_113337.4	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB030	NM_001084678.2	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB031	NM_106123.4	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB032	NM_119665.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB034	NM_125482.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB035	NM_113767.2	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB036	NM_125143.4	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB037	NM_122206.4	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB039	NM_117888.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB041_NM_118951_3	NM_118951.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB042	NM_117305.5	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB043	NM_121666.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB045	NM_114751.2	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB046	NM_121290.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB049	NM_124803.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB050	NM_104552.2	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB052	AF469468.1	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB053	NM_125921.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB054	NM_106001.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB055	NM_001125441.1	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB056	NM_121786.4	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB058	NM_101514.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB059	NM_125370.4	NCBI / TAIR	genoma	Etapa II

<i>Arabidopsis thaliana</i>	Arabidopsis_tha_MYB060	NM_100755.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB061	NM_100825.5	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB062	NM_105503.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB063	NM_106569.4	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB064	NM_12114	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB065	NM_001337925.1	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB066	NM_121479.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB068	NM_125976.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB069	NM_119499.4	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB070	NM_127893.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB072	NM_104495.1	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB073	NM_119889.2	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB074	NM_116749.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB076	NM_001342956.1	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB077	NM_114866.4	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB078	NM_124340.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB079	NM_117422.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB082	NM_148121.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB083	NM_111685.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB084	NM_114829.4	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB085	NM_118394.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB086	NM_180548.4	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB087	NM_119940.5	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB093	NM_103190.2	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB094	NM_114628.4	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB095	NM_106103.4	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB096	NM_125641.4	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB097	NM_118827.2	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB098	NM_117993.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB100	NM_128080.1	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB101	NM_128805.4	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB102_X90382_2	X90382.2	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB103	NM_105065.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB105	NM_001334415.1	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB106	NM_001337322.1	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB107	NM_001337450.1	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB108	NM_111525.4	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB109	NM_115431.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB110	NM_113823.1	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB111	NM_124310.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB112	NM_103696.3	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB113	NM_105308.2	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB115	NM_123400.2	NCBI / TAIR	genoma	Etapa II
	Arabidopsis_tha_MYB116	NM_001332675.1	NCBI / TAIR	genoma	Etapa II
Arabidopsis_tha_MYB117	NM_102441.3	NCBI / TAIR	genoma	Etapa II	
Arabidopsis_tha_MYB118	NM_113694.2	NCBI / TAIR	genoma	Etapa II	
Arabidopsis_tha_MYB119	NM_125275.2	NCBI / TAIR	genoma	Etapa II	
Arabidopsis_tha_MYB120	NM_124884.2	NCBI / TAIR	genoma	Etapa II	
Arabidopsis_tha_MYB122	NM_106068.3	NCBI / TAIR	genoma	Etapa II	
Arabidopsis_tha_MYB124	NM_001084065.3	NCBI / TAIR	genoma	Etapa II	
Arabidopsis_tha_MYB125	NM_115910.2	NCBI / TAIR	genoma	Etapa II	
Arabidopsis_tha_MYB305	NM_113337.4	NCBI / TAIR	genoma	Etapa II	
Arabidopsis_tha_MYB92_AF062916_2	AF062916.2	NCBI / TAIR	genoma	Etapa II	
Arabidopsis_thaliana_MYB_R1	CP002688.1	NCBI / TAIR	genoma	Etapa II	
<i>Azolla filiculoides</i>	Azolla_fil_s0001_g000218	Azfi_s0001.g000218	FernBase	genoma	Etapa II
	Azolla_fil_s0001_g000839	Azfi_s0001.g000839	FernBase	genoma	Etapa II
	Azolla_fil_s0003_g007581	Azfi_s0003.g007581	FernBase	genoma	Etapa II
	Azolla_fil_s0003_g007699	Azfi_s0003.g007699	FernBase	genoma	Etapa I
	Azolla_fil_s0003_g008081	Azfi_s0003.g008081	FernBase	genoma	Etapa II
	Azolla_fil_s0004_g008455_R	Azfi_s0004.g008455_R	FernBase	genoma	Etapa I
	Azolla_fil_s0005_g009119	Azfi_s0005.g009119	FernBase	genoma	Etapa II
	Azolla_fil_s0005_g009154	Azfi_s0005.g009154	FernBase	genoma	Etapa I
	Azolla_fil_s0006_g009830	Azfi_s0006.g009830	FernBase	genoma	Etapa I
	Azolla_fil_s0007_g010866	Azfi_s0007.g010866	FernBase	genoma	Etapa I
	Azolla_fil_s0008_g011612	Azfi_s0008.g011612	FernBase	genoma	Etapa I
	Azolla_fil_s0010_g012085	Azfi_s0010.g012085	FernBase	genoma	Etapa I
	Azolla_fil_s0010_g012145	Azfi_s0010.g012145	FernBase	genoma	Etapa I

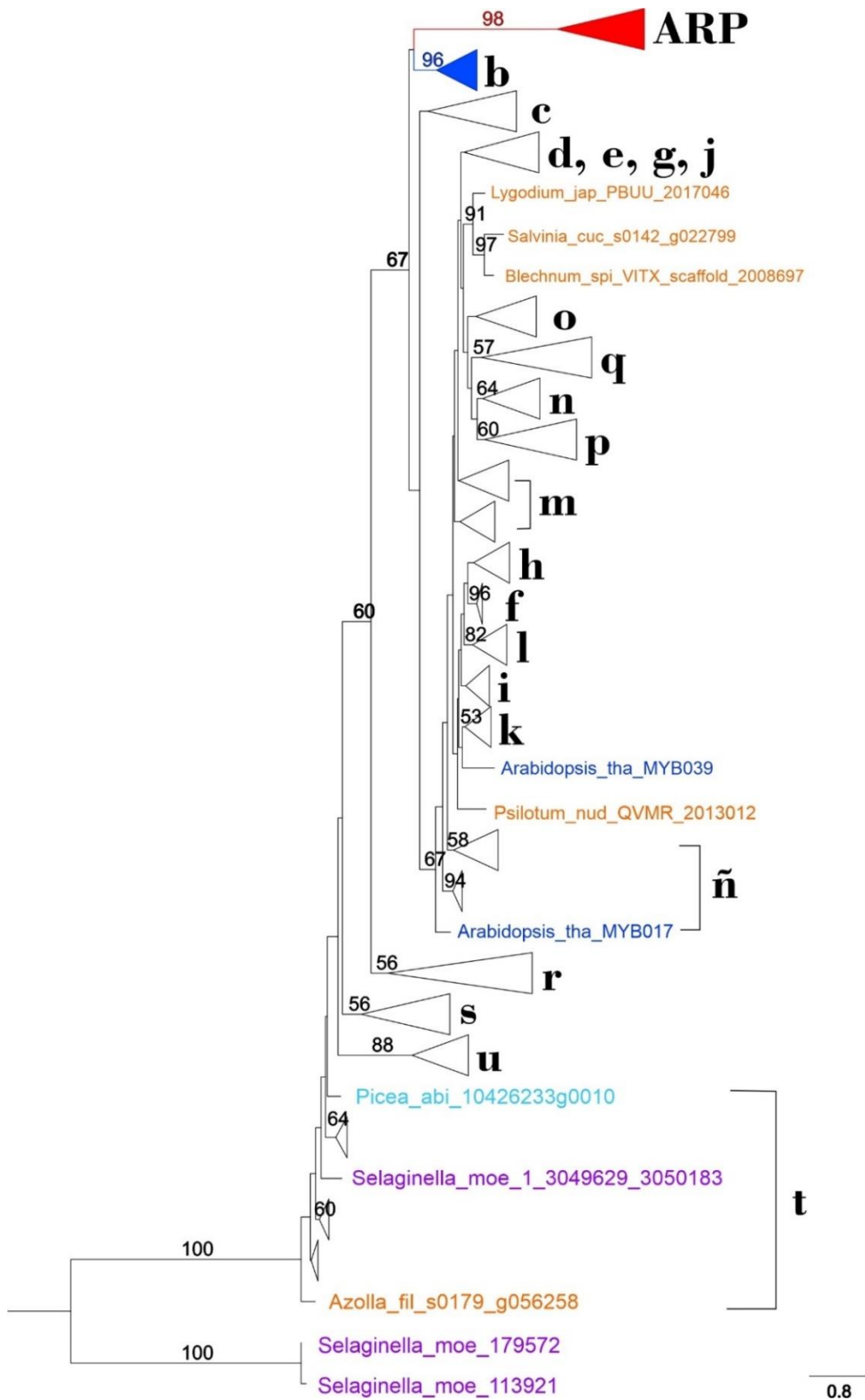
<i>Azolla filiculoides</i>	Azolla_fil_s0014_g013539	Azfi_s0014.g013539	FernBase	genoma	Etapa II
	Azolla_fil_s0014_g013584	Azfi_s0014.g013584	FernBase	genoma	Etapa I
	Azolla_fil_s0016_g014344	Azfi_s0016.g014344	FernBase	genoma	Etapa II
	Azolla_fil_s0018_g014830	Azfi_s0018.g014830	FernBase	genoma	Etapa II
	Azolla_fil_s0021_g015882	Azfi_s0021.g015882	FernBase	genoma	Etapa I
	Azolla_fil_s0035_g025508	Azfi_s0035.g025508	FernBase	genoma	Etapa I
	Azolla_fil_s0041_g026717	Azfi_s0041.g026717	FernBase	genoma	Etapa I
	Azolla_fil_s0041_g026722	Azfi_s0041.g026722	FernBase	genoma	Etapa I
	Azolla_fil_s0061_g034930	Azfi_s0061.g034930	FernBase	genoma	Etapa II
	Azolla_fil_s0076_g037908	Azfi_s0076.g037908	FernBase	genoma	Etapa I
	Azolla_fil_s0103_g044681	Azfi_s0103.g044681	FernBase	genoma	Etapa II
	Azolla_fil_s0105_g044929	Azfi_s0105.g044929	FernBase	genoma	Etapa II
	Azolla_fil_s0113_g045874	Azfi_s0113.g045874	FernBase	genoma	Etapa I
	Azolla_fil_s0125_g048539	Azfi_s0125.g048539	FernBase	genoma	Etapa II
	Azolla_fil_s0129_g048859	Azfi_s0129.g048859	FernBase	genoma	Etapa I
	Azolla_fil_s0130_g049059	Azfi_s0130.g049059	FernBase	genoma	Etapa II
	Azolla_fil_s0163_g054240	Azfi_s0163.g054240	FernBase	genoma	Etapa II
	Azolla_fil_s0173_g055749	Azfi_s0173.g055749	FernBase	genoma	Etapa II
	Azolla_fil_s0179_g056258	Azfi_s0179.g056258	FernBase	genoma	Etapa II
	Azolla_fil_s0179_g056264	Azfi_s0179.g056264	FernBase	genoma	Etapa II
	Azolla_fil_s0217_g058506	Azfi_s0217.g058506	FernBase	genoma	Etapa II
	Azolla_fil_s0243_g059882	Azfi_s0243.g059882	FernBase	genoma	Etapa II
	Azolla_fil_s0390_g067689	Azfi_s0390.g067689	FernBase	genoma	Etapa II
Azolla_fil_s0445_g070637	Azofi_s0445.g070637	FernBase	genoma	Etapa II	
Azolla_fil_s0812_g087843	Azfi_s0812.g087843	FernBase	genoma	Etapa II	
Azolla_fil_s0938_g093183	Azfi_s0938.g093183	FernBase	genoma	Etapa II	
Azolla_fil_s1083_g097159	Azfi_s1083.g097159	FernBase	genoma	Etapa II	
Azolla_fil_s3609_g116385	Azfi_s3609.g116385	FernBase	genoma	Etapa II	
<i>Blechnum spicant</i>	Blechnum_spi_VITX_scaffold_2003303	VITX_scaffold_2003303	FernBase	transcriptoma	Etapa I
	Blechnum_spi_VITX_scaffold_2008243	VITX_scaffold_2008243	FernBase	transcriptoma	Etapa I
	Blechnum_spi_VITX_scaffold_2008697	VITX_scaffold_2008697	FernBase	transcriptoma	Etapa I
<i>Botrypus virginianus</i>	Botrypus_vir_BEGM_scaffold_2023523	BEGM_scaffold_2023523	FernBase	transcriptoma	Etapa I
	Botrypus_vir_BEGM_scaffold_2090916	BEGM_scaffold_2090916	FernBase	transcriptoma	Etapa I
<i>Ceratopteris richardii</i>	Ceratopteris_ric_10971	iOHQ_Cfern c123/f1p123/10971_ARP	Ambrose	genoma	Etapa I
	Ceratopteris_ric_18151	iOHQ_Cfern c123/f1p123/18151_ARP	Ambrose	genoma	Etapa I
	Ceratopteris_ric_18320	iOHQ_Cfern c123/f1p123/18320_ARP	Ambrose	genoma	Etapa I
	Ceratopteris_ric_18829	iOHQ_Cfern c123/f1p123/18829_ARP	Ambrose	genoma	Etapa I
	Ceratopteris_ric_19986	iOHQ_Cfern c123/f1p123/19986_ARP	Ambrose	genoma	Etapa I
	Ceratopteris_ric_21031	iOHQ_Cfern c123/f1p123/21031_ARP	Ambrose	genoma	Etapa I
	Ceratopteris_ric_22922	iOHQ_Cfern c123/f1p123/22922_ARP	Ambrose	genoma	Etapa I
	Ceratopteris_ric_23124	iOHQ_Cfern c123/f1p123/23124_ARP	Ambrose	genoma	Etapa I
	Ceratopteris_ric_23616	iOHQ_Cfern c123/f1p123/23616_ARP	Ambrose	genoma	Etapa I
	Ceratopteris_ric_31009	iOHQ_Cfern c123/f1p123/31009_ARP	Ambrose	genoma	Etapa I
	Ceratopteris_ric_32145	iOHQ_Cfern c123/f1p123/32145_ARP	Ambrose	genoma	Etapa I
	Ceratopteris_ric_41362	iOHQ_Cfern c123/f1p123/41362_ARP	Ambrose	genoma	Etapa I
	Ceratopteris_ric_43319	iOHQ_Cfern c123/f1p123/43319_ARP	Ambrose	genoma	Etapa I
	Ceratopteris_ric_44195	iOHQ_Cfern c123/f1p123/44195_ARP	Ambrose	genoma	Etapa I
	Ceratopteris_ric_44548	iOHQ_Cfern c123/f1p123/44548_ARP	Ambrose	genoma	Etapa I
	Ceratopteris_ric_44597	iOHQ_Cfern c123/f1p123/44597_ARP	Ambrose	genoma	Etapa I
	Ceratopteris_ric_44789	iOHQ_Cfern c123/f1p123/44789_ARP	Ambrose	genoma	Etapa I
	Ceratopteris_ric_46354	iOHQ_Cfern c123/f1p123/46354_ARP	Ambrose	genoma	Etapa I
	Ceratopteris_ric_46511	iOHQ_Cfern c123/f1p123/46511_ARP	Ambrose	genoma	Etapa I
	Ceratopteris_ric_46622	iOHQ_Cfern c123/f1p123/46622_ARP	Ambrose	genoma	Etapa I
	Ceratopteris_ric_47230	iOHQ_Cfern c123/f1p123/47230_ARP	Ambrose	genoma	Etapa I
	Ceratopteris_ric_4890	iOHQ_Cfern c123/f1p123/4890_ARP	Ambrose	genoma	Etapa I
	Ceratopteris_ric_6415	iOHQ_Cfern c123/f1p123/6415_ARP	Ambrose	genoma	Etapa I
<i>Cryptogramma acrostichoides</i>	Cryptogramma_acr_WQML_2008610	WQML_scaffold_2008610	One KP	transcriptoma	Etapa I
	Cryptogramma_acr_WQML_2012025	WQML_scaffold_2012025	One KP	transcriptoma	Etapa I
	Cryptogramma_acr_WQML_2012131	WQML_scaffold_2012131	One KP	transcriptoma	Etapa I
	Cryptogramma_acr_WQML_2066663	WQML_scaffold_2066663	One KP	transcriptoma	Etapa I
<i>Culcita macrocarpa</i>	Culcita_mac_PNZO_2148837	PNZO_scaffold_2148837	One KP	transcriptoma	Etapa I
	Culcita_mac_PNZO_2150229	PNZO_scaffold_2150229	One KP	transcriptoma	Etapa I
	Culcita_mac_PNZO_2152281	PNZO_scaffold_2152281	One KP	transcriptoma	Etapa I
<i>Dendrolycopodium obscurum</i>	Dendrolycopodium_obs_XNXF_2070197	XNXF_scaffold_2070197	One KP	transcriptoma	Etapa I
<i>Diphasiastrum digitatum</i>	Diphasiastrum_dig_WAFT_2064506	WAFT_scaffold_2064506	One KP	transcriptoma	Etapa I

<i>Equisetum diffusum</i>	Equisetum_dif_CAPN_2039541	CAPN_scaffold_2039541	One KP	transcriptoma	Etapa I
<i>Equisetum hymale</i>	Equisetum_hye_JVSZ_2005827	JVSZ_scaffold_2005827	One KP	transcriptoma	Etapa I
	Equisetum_hye_JVSZ_2010537	JVSZ_scaffold_2010537	One KP	transcriptoma	Etapa I
	Equisetum_hye_JVSZ_2129411	JVSZ_scaffold_2129411	One KP	transcriptoma	Etapa I
<i>Ginkgo biloba</i>	Ginkgo_bil_SGTW_2038931	SGTW_scaffold_2038931	One KP	transcriptoma	Etapa I
	Ginkgo_biloba_SGTW_2032857	SGTW_scaffold_2032857	One KP	transcriptoma	Etapa I
<i>Huperzia lucidula</i>	Huperzia_luc_GKAG_2020029	GKAG_scaffold_2020029	One KP	transcriptoma	Etapa I
<i>Huperzia myrsinites</i>	Huperzia_myrs_CBAE_2006957	CBAE_scaffold_2006957	One KP	transcriptoma	Etapa I
<i>Huperzia selago</i>	Huperzia_sel_GTUO_2003499	GTUO_scaffold_2003499	One KP	transcriptoma	Etapa I
<i>Huperzia squarrosa</i>	Huperzia_squ_GAON_2002376	GAON_scaffold_2002376	One KP	transcriptoma	Etapa I
<i>Isoetes</i>	Isoetes_teg_PKOX_2013608	PKOX_scaffold_2013608	One KP	transcriptoma	Etapa I
<i>tegetiformans</i>	Isoetes_teg_PKOX_2094344	PKOX_scaffold_2094344	One KP	transcriptoma	Etapa I
<i>Lycopodiella appressa</i>	Lycopodiella_app_ULKT_2011475	ULKT_scaffold_2011475	One KP	transcriptoma	Etapa I
<i>Lycopodium annotinum</i>	Lycopodium_ann_ENQF_2003353	ENQF_scaffold_2003353	One KP	transcriptoma	Etapa I
	Lycopodium_ann_ENQF_2011065	ENQF_scaffold_2011065	One KP	transcriptoma	Etapa I
	Lycopodium_ann_ENQF_2084123	ENQF_scaffold_2084123	One KP	transcriptoma	Etapa I
<i>Lycopodium deuterodensum</i>	Lycopodium_deu_PQTO_2009942	PQTO_scaffold_2009942	One KP	transcriptoma	Etapa I
<i>Lygodium japonicum</i>	Lygodium_jap_PBUU_2003535	PBUU_scaffold_2003535	One KP	transcriptoma	Etapa I
	Lygodium_jap_PBUU_2005772	PBUU_scaffold_2005772	One KP	transcriptoma	Etapa I
	Lygodium_jap_PBUU_2009136	PBUU_scaffold_2009136	One KP	transcriptoma	Etapa I
	Lygodium_jap_PBUU_2017046	PBUU_scaffold_2017046	One KP	transcriptoma	Etapa I
	Lygodium_jap_PBUU_2078554	PBUU_scaffold_2078554	One KP	transcriptoma	Etapa I
<i>Ophioglossum vulgatum</i>	Ophioglossum_vul_QHVS_2102717	QHVS_scaffold_2102717	One KP	transcriptoma	Etapa I
	Ophioglossum_vul_WTJG_2009741	WTJG_scaffold_2009741	One KP	transcriptoma	Etapa I
	Ophioglossum_vul_WTJG_2020032	WTJG_scaffold_2020032	One KP	transcriptoma	Etapa I
<i>Osmunda sp (UOMY)</i>	Osmunda_sp_UOMY_2002138	UOMY_scaffold_2002138	One KP	transcriptoma	Etapa I
	Osmunda_sp_UOMY_2008801	UOMY_scaffold_2008801	One KP	transcriptoma	Etapa I
	Osmunda_sp_UOMY_2010120	UOMY_scaffold_2010120	One KP	transcriptoma	Etapa I
	Osmunda_sp_UOMY_2080628	UOMY_scaffold_2080628	One KP	transcriptoma	Etapa I
<i>Physcomitrella patens</i>	Physcomitrella_pat_Pp3c1_1650V3	Pp3c1_1650V3	Phytozome	genoma	Etapa II
	Physcomitrella_pat_Pp3c1_21610V3	Pp3c1_21610V3	Phytozome	genoma	Etapa II
	Physcomitrella_pat_Pp3c1_4970V3	Pp3c1_4970V3.1	Phytozome	genoma	Etapa II
	Physcomitrella_pat_Pp3c11_10350V3	Pp3c11_10350V3	Phytozome	genoma	Etapa II
	Physcomitrella_pat_Pp3c11_5410V3	Pp3c11_5410V3	Phytozome	genoma	Etapa I
	Physcomitrella_pat_Pp3c11_5420V3	Pp3c11_5420V3.1	Phytozome	genoma	Etapa II
	Physcomitrella_pat_Pp3c12_17180V3	Pp3c12_17180V3.1	Phytozome	genoma	Etapa II
	Physcomitrella_pat_Pp3c14_12390V3	Pp3c14_12390V3	Phytozome	genoma	Etapa II
	Physcomitrella_pat_Pp3c15_15960V3	Pp3c15_15960V3.1	Phytozome	genoma	Etapa II
	Physcomitrella_pat_Pp3c15_24080V3	Pp3c15_24080V3.1	Phytozome	genoma	Etapa I
	Physcomitrella_pat_Pp3c16_9970V3	Pp3c16_9970V3	Phytozome	genoma	Etapa II
	Physcomitrella_pat_Pp3c17_19780V3	Pp3c17_19780V3	Phytozome	genoma	Etapa II
	Physcomitrella_pat_Pp3c17_7730V3	Pp3c17_7730V3	Phytozome	genoma	Etapa II
	Physcomitrella_pat_Pp3c19_20750V3	Pp3c19_20750V3	Phytozome	genoma	Etapa II
	Physcomitrella_pat_Pp3c2_37040V3	Pp3c2_37040V3.1	Phytozome	genoma	Etapa II
	Physcomitrella_pat_Pp3c20_8950V3	Pp3c20_8950V3.1	Phytozome	genoma	Etapa I
	Physcomitrella_pat_Pp3c23_7530V3	Pp3c23_7530V3	Phytozome	genoma	Etapa I
	Physcomitrella_pat_Pp3c25_3170V3	Pp3c25_3170V3.1	Phytozome	genoma	Etapa II
	Physcomitrella_pat_Pp3c26_1390V3	Pp3c26_1390V3	Phytozome	genoma	Etapa II
	Physcomitrella_pat_Pp3c3_12080V3	Pp3c3_12080V3	Phytozome	genoma	Etapa II
	Physcomitrella_pat_Pp3c4_12510V3	Pp3c4_12510V3	Phytozome	genoma	Etapa II
	Physcomitrella_pat_Pp3c5_2150V3	Pp3c5_2150V3	Phytozome	genoma	Etapa II
	Physcomitrella_pat_Pp3c6_24650V3	Pp3c6_24650V3	Phytozome	genoma	Etapa II
	Physcomitrella_pat_Pp3c7_23450V3	Pp3c7_23450V3	Phytozome	genoma	Etapa II
	Physcomitrella_pat_Pp3c7_23490V3	Pp3c7_23490V3	Phytozome	genoma	Etapa I
	Physcomitrella_pat_Pp3c8_16720V3	Pp3c8_16720V3.1	Phytozome	genoma	Etapa II
	<i>Physcomitrella patens</i>	Physcomitrella_pat_Pp3c9_15970V3	Pp3c9_15970V3	Phytozome	genoma
Physcomitrella_pat_Pp3c9_24170V3		Pp3c9_24170V3.1	Phytozome	genoma	Etapa I
Physcomitrella_pat_Pp3c5_7650V3		Pp3c5_7650V3	Phytozome	genoma	Etapa II
Physcomitrella_pat_Pp3c6_23360V3		Pp3c6_23360V3	Phytozome	genoma	Etapa II
Picea_abi_10157944g0010		MA_10157944g0010	CONGENIE	genoma	Etapa II
Picea_abi_101987g0010	MA_101987g0010	CONGENIE	genoma	Etapa I	
Picea_abi_10208000g0010	MA_10208000g0010	CONGENIE	genoma	Etapa II	

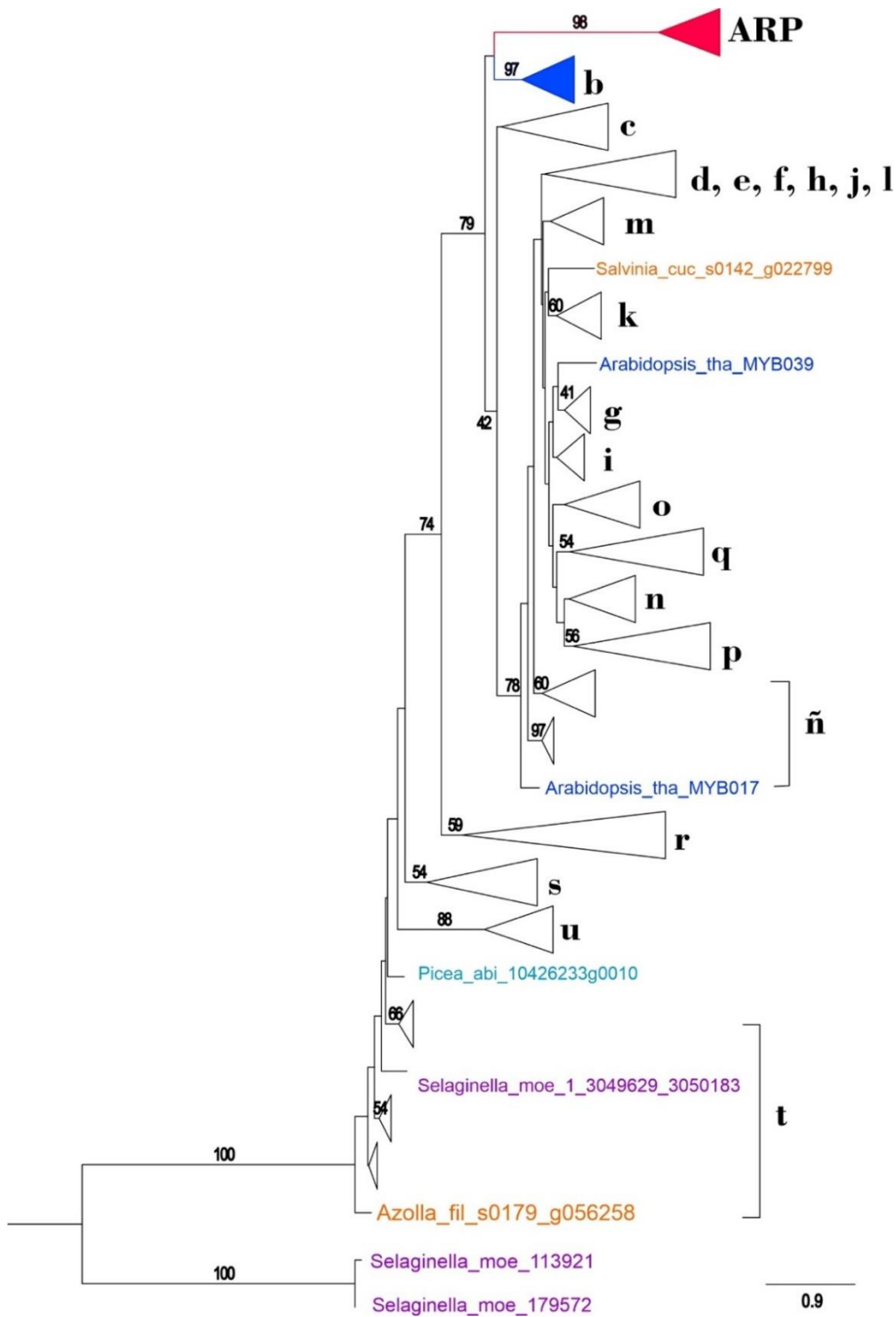
	Picea_abi_10425867g0020	MA_10425867g0020	CONGENIE	genoma	Etapa II
	Picea_abi_10426233g0010	MA_10426233g0010	CONGENIE	genoma	Etapa I
	Picea_abi_10430220g0010	MA_10430220g0010	CONGENIE	genoma	Etapa I
	Picea_abi_10434028g0010	MA_10434028g0010	CONGENIE	genoma	Etapa II
	Picea_abi_10435612g0010	MA_10435612g0010	CONGENIE	genoma	Etapa I
	Picea_abi_10859g0010	MA_10859g0010	CONGENIE	genoma	Etapa I
	Picea_abi_120049g0010	MA_120049g0010	CONGENIE	genoma	Etapa I
	Picea_abi_128244g0010	MA_128244g0010	CONGENIE	genoma	Etapa II
	Picea_abi_128871g0010	MA_128871g0010	CONGENIE	genoma	Etapa II
	Picea_abi_129382g0010	MA_129382g0010	CONGENIE	genoma	Etapa I
	Picea_abi_130648g0010	MA_130648g0010	CONGENIE	genoma	Etapa II
	Picea_abi_130951g0010	MA_130951g0010	CONGENIE	genoma	Etapa II
	Picea_abi_137453g0010	MA_137453g0010	CONGENIE	genoma	Etapa II
	Picea_abi_137934g0010	MA_137934g0010	CONGENIE	genoma	Etapa I
	Picea_abi_13869g0010	MA_13869g0010	CONGENIE	genoma	Etapa II
	Picea_abi_139238g0010	MA_139238g0010	CONGENIE	genoma	Etapa I
	Picea_abi_139448g0010	MA_139448g0010	CONGENIE	genoma	Etapa I
	Picea_abi_139924g0010	MA_139924g0010	CONGENIE	genoma	Etapa I
	Picea_abi_140579g0010	MA_140579g0010	CONGENIE	genoma	Etapa II
	Picea_abi_142039g0010	MA_142039g0010	CONGENIE	genoma	Etapa I
	Picea_abi_14730g0020	MA_14730g0020	CONGENIE	genoma	Etapa II
	Picea_abi_158099g0010	MA_158099g0010	CONGENIE	genoma	Etapa I
	Picea_abi_1589g0010	MA_1589g0010	CONGENIE	genoma	Etapa I
	Picea_abi_164713g0010	MA_164713g0010	CONGENIE	genoma	Etapa II
	Picea_abi_183667g0010	MA_183667g0010	CONGENIE	genoma	Etapa II
	Picea_abi_195086g0010	MA_195086g0010	CONGENIE	genoma	Etapa II
	Picea_abi_20462g0010	MA_20462g0010	CONGENIE	genoma	Etapa II
	Picea_abi_22140g0010	MA_22140g0010	CONGENIE	genoma	Etapa II
	Picea_abi_223201g0010	MA_223201g0010	CONGENIE	genoma	Etapa II
	Picea_abi_24271g0020	MA_24271g0020	CONGENIE	genoma	Etapa I
	Picea_abi_2797g0010	MA_2797g0010	CONGENIE	genoma	Etapa I
	Picea_abi_308669g0010	MA_308669g0010	CONGENIE	genoma	Etapa II
	Picea_abi_322432g0020	MA_322432g0020	CONGENIE	genoma	Etapa I
	Picea_abi_328440g0010	MA_328440g0010	CONGENIE	genoma	Etapa II
	Picea_abi_33964g0010	MA_33964g0010	CONGENIE	genoma	Etapa I
	Picea_abi_357440g0010	MA_357440g0010	CONGENIE	genoma	Etapa II
	Picea_abi_358042g0010	MA_358042g0010	CONGENIE	genoma	Etapa II
	Picea_abi_3665g0010	MA_3665g0010	CONGENIE	genoma	Etapa I
	Picea_abi_4704g0010	MA_4704g0010	CONGENIE	genoma	Etapa I
	Picea_abi_51173g0010	MA_51173g0010	CONGENIE	genoma	Etapa II
	Picea_abi_53573g0010	MA_53573g0010	CONGENIE	genoma	Etapa II
	Picea_abi_55123g0010	MA_55123g0010	CONGENIE	genoma	Etapa II
	Picea_abi_62361g0010	MA_62361g0010	CONGENIE	genoma	Etapa I
	Picea_abi_62687g0010	MA_62687g0010	CONGENIE	genoma	Etapa II
	Picea_abi_6393g0010	MA_6393g0010	CONGENIE	genoma	Etapa I
	Picea_abi_64405g0010	MA_64405g0010	CONGENIE	genoma	Etapa II
	Picea_abi_66255g0010	MA_66255g0010	CONGENIE	genoma	Etapa II
	Picea_abi_7055g0010	MA_7055g0010	CONGENIE	genoma	Etapa II
	Picea_abi_7115g0010	MA_7115g0010	CONGENIE	genoma	Etapa II
	Picea_abi_8356281g0010	MA_8356281g0010	CONGENIE	genoma	Etapa I
	Picea_abi_83918g0010	MA_83918g0010	CONGENIE	genoma	Etapa II
	Picea_abi_8626g0010	MA_8626g0010	CONGENIE	genoma	Etapa II
	Picea_abi_89683g0010	MA_89683g0010	CONGENIE	genoma	Etapa I
	Picea_abi_9374017g0010	MA_9374017g0010	CONGENIE	genoma	Etapa I
	Picea_abi_94467g0010	MA_94467g0010	CONGENIE	genoma	Etapa II
	Picea_abi_9483804g0010	MA_9483804g0010	CONGENIE	genoma	Etapa I
	Picea_abi_95226g0010	MA_95226g0010	CONGENIE	genoma	Etapa II
	Picea_abi_9684426g0010	MA_9684426g0010	CONGENIE	genoma	Etapa II
	Picea_abi_96853g0010	MA_96853g0010	CONGENIE	genoma	Etapa II
	Picea_abi_9783816g0010	MA_9783816g0010	CONGENIE	genoma	Etapa II
	Picea_abi_9818613g0010	MA_9818613g0010	CONGENIE	genoma	Etapa II
	Pilularia_glo_KIIX_2010134	KIIX_scaffold_2010134	One KP	transcriptoma	Etapa I
	Pilularia_glo_KIIX_2011949	KIIX_scaffold_2011949	One KP	transcriptoma	Etapa I
	Pilularia_glo_KIIX_2090543	KIIX_scaffold_2090543	One KP	transcriptoma	Etapa I
	Pinus_tae_AY356372	FJ469926	NCBI	genoma	Etapa I
	Pinus_tae_DQ399060	FJ469927	NCBI	genoma	Etapa I
	Polypodium_hes_IXLH_2006001	IXLH_scaffold_2006001	One KP	transcriptoma	Etapa I

<i>Polypodium hesperium</i>	Polydium_hes_IXLH_2020610	IXLH_scaffold_2020610	One KP	transcriptoma	Etapal
<i>Polystichum acrostichoides</i>	Polystichum_acr_FQGQ_2000476	FQGQ_scaffold_2000476	One KP	transcriptoma	Etapal
	Polystichum_acr_FQGQ_2005766	FQGQ_scaffold_2005766	One KP	transcriptoma	Etapal
	Polystichum_acr_FQGQ_2006743	FQGQ_scaffold_2006743	One KP	transcriptoma	Etapal
	Polystichum_acr_FQGQ_2068415	FQGQ_scaffold_2068415	One KP	transcriptoma	Etapal
<i>Pseudolycopodiella caroliniana</i>	Pseudolycopodiella_car_UPMJ_2019671	UPMJ_scaffold_2019671	One KP	transcriptoma	Etapal
<i>Psilotum nudum</i>	Psilotum_nud_QVMR_2008046	QVMR_scaffold_2008046	One KP	transcriptoma	Etapal
	Psilotum_nud_QVMR_2013012	QVMR_scaffold_2013012	One KP	transcriptoma	Etapal
	Psilotum_nud_QVMR_2016093	QVMR_scaffold_2016093	One KP	transcriptoma	Etapal
	Psilotum_nud_QVMR_2017134	QVMR_scaffold_2017134	One KP	transcriptoma	Etapal
<i>Salvinia cucullata</i>	Salvinia_cuc_s0001_g000153	Sacu_v1.1_s0001.g000153	FernBase	genoma	Etapall
	Salvinia_cuc_s0002_g000870	Sacu_v1.1_s0002.g000870	FernBase	genoma	Etapall
	Salvinia_cuc_s0002_g001222	Sacu_v1.1_s0002.g001222	FernBase	genoma	Etapal
	Salvinia_cuc_s0004_g002054	Sacu_v1.1_s0004.g002054	FernBase	genoma	Etapall
	Salvinia_cuc_s0006_g003255	Sacu_v1.1_s0006.g003255	FernBase	genoma	Etapall
	Salvinia_cuc_s0007_g003659	Sacu_v1.1_s0007.g003659	FernBase	genoma	Etapal
	Salvinia_cuc_s0011_g004977	Sacu_v1.1_s0011.g004977	FernBase	genoma	Etapall
	Salvinia_cuc_s0015_g006431	Sacu_v1.1_s0015.g006431	FernBase	genoma	Etapall
	Salvinia_cuc_s0016_g006931	Sacu_v1.1_s0016.g006931	FernBase	genoma	Etapal
	Salvinia_cuc_s0026_g009506	Sacu_v1.1_s0026.g009506	FernBase	genoma	Etapall
	Salvinia_cuc_s0031_g010615	Sacu_v1.1_s0031.g010615	FernBase	genoma	Etapall
	Salvinia_cuc_s0032_g010734	Sacu_v1.1_s0032.g010734	FernBase	genoma	Etapall
	Salvinia_cuc_s0036_g011562	Sacu_v1.1_s0036.g011562	FernBase	genoma	Etapall
	Salvinia_cuc_s0041_g012546	Sacu_v1.1_s0041.g012546	FernBase	genoma	Etapal
	Salvinia_cuc_s0043_g012904	Sacu_v1.1_s0043.g012904	FernBase	genoma	Etapal
	Salvinia_cuc_s0053_g014109	Sacu_v1.1_s0053.g014109	FernBase	genoma	Etapal
	Salvinia_cuc_s0063_g015735	Sacu_v1.1_s0063.g015735	FernBase	genoma	Etapall
	Salvinia_cuc_s0068_g016477	Sacu_v1.1_s0068.g016477	FernBase	genoma	Etapall
	Salvinia_cuc_s0073_g017015	Sacu_v1.1_s0073.g017015	FernBase	genoma	Etapal
	Salvinia_cuc_s0086_g018401	Sacu_v1.1_s0086.g018401	FernBase	genoma	Etapall
	Salvinia_cuc_s0092_g018975	Sacu_v1.1_s0092.g018975	FernBase	genoma	Etapall
	Salvinia_cuc_s0096_g019382	Sacu_v1.1_s0096.g019382	FernBase	genoma	Etapall
	Salvinia_cuc_s0106_g020274	Sacu_v1.1_s0106.g020274	FernBase	genoma	Etapal
	Salvinia_cuc_s0123_g021550	Sacu_v1.1_s0123.g021550	FernBase	genoma	Etapall
	Salvinia_cuc_s0129_g021926	Sacu_v1.1_s0129.g021926	FernBase	genoma	Etapall
	Salvinia_cuc_s0137_g022448	Sacu_v1.1_s0137.g022448	FernBase	genoma	Etapal
	Salvinia_cuc_s0142_g022799	Sacu_v1.1_s0142.g022799	FernBase	genoma	Etapall
	Salvinia_cuc_s0147_g023157	Sacu_v1.1_s0147.g023157	FernBase	genoma	Etapal
	Salvinia_cuc_s0193_g025371	Sacu_v1.1_s0193.g025371	FernBase	genoma	Etapal
	Salvinia_cuc_s0196_g025489	Sacu_v1.1_s0196.g025489	FernBase	genoma	Etapal
	Salvinia_cuc_s0196_g025496	Sacu_v1.1_s0196.g025496	FernBase	genoma	Etapal
	Salvinia_cuc_s0226_g026344	Sacu_v1.1_s0226.g026344	FernBase	genoma	Etapall
	Salvinia_cuc_s0272_g027033	Sacu_v1.1_s0272.g027033	FernBase	genoma	Etapal
Salvinia_cuc_s0330_g027251	Sacu_v1.1_s0330.g027251	FernBase	genoma	Etapall	
Salvinia_cuc_s1503_g028048	Sacu_v1.1_s1503.g028048	FernBase	genoma	Etapal	
<i>Selaginella acanthonota</i>	Selaginella_aca_ZYCD_2002111	ZYCD_scaffold_2002111	One KP	transcriptoma	Etapal
<i>Selaginella apoda</i>	Selaginella_apo_LGDQ_2010304	LGDQ_scaffold_2010304	One KP	transcriptoma	Etapal
<i>Selaginella kraussiana</i>	Selaginella_kra_AY667452	AY667452	NCBI	genoma	Etapal
	Selaginella_kra_ZFGK_2009382	ZFGK_scaffold_2009382	One KP	transcriptoma	Etapal
<i>Selaginella lepidophylla</i>	Selaginella_kra_ZFGK_2041777	ZFGK_scaffold_2041777	One KP	transcriptoma	Etapal
	Selaginella_lep_ABIJ_2010935	ABIJ_scaffold_2010935	One KP	transcriptoma	Etapal
<i>Selaginella lepidophylla</i>	Selaginella_moe_1_3049629_3050183	3049629_3050183	Phytozome	genoma	Etapall
	Selaginella_moe_100734	100734	Phytozome	genoma	Etapall
	Selaginella_moe_101484	101484	Phytozome	genoma	Etapall
	Selaginella_moe_109587	109587	Phytozome	genoma	Etapall
	Selaginella_moe_113921	113921	Phytozome	genoma	Etapall
	Selaginella_moe_140962	140962	Phytozome	genoma	Etapall
	Selaginella_moe_179572	179572	Phytozome	genoma	Etapall
	Selaginella_moe_39481	39481	Phytozome	genoma	Etapall
	Selaginella_moe_39516	39516	Phytozome	genoma	Etapall
	Selaginella_moe_437219	437219	Phytozome	genoma	Etapall
	Selaginella_moe_59725	59725	Phytozome	genoma	Etapall
	Selaginella_moe_6091	6091	Phytozome	genoma	Etapall

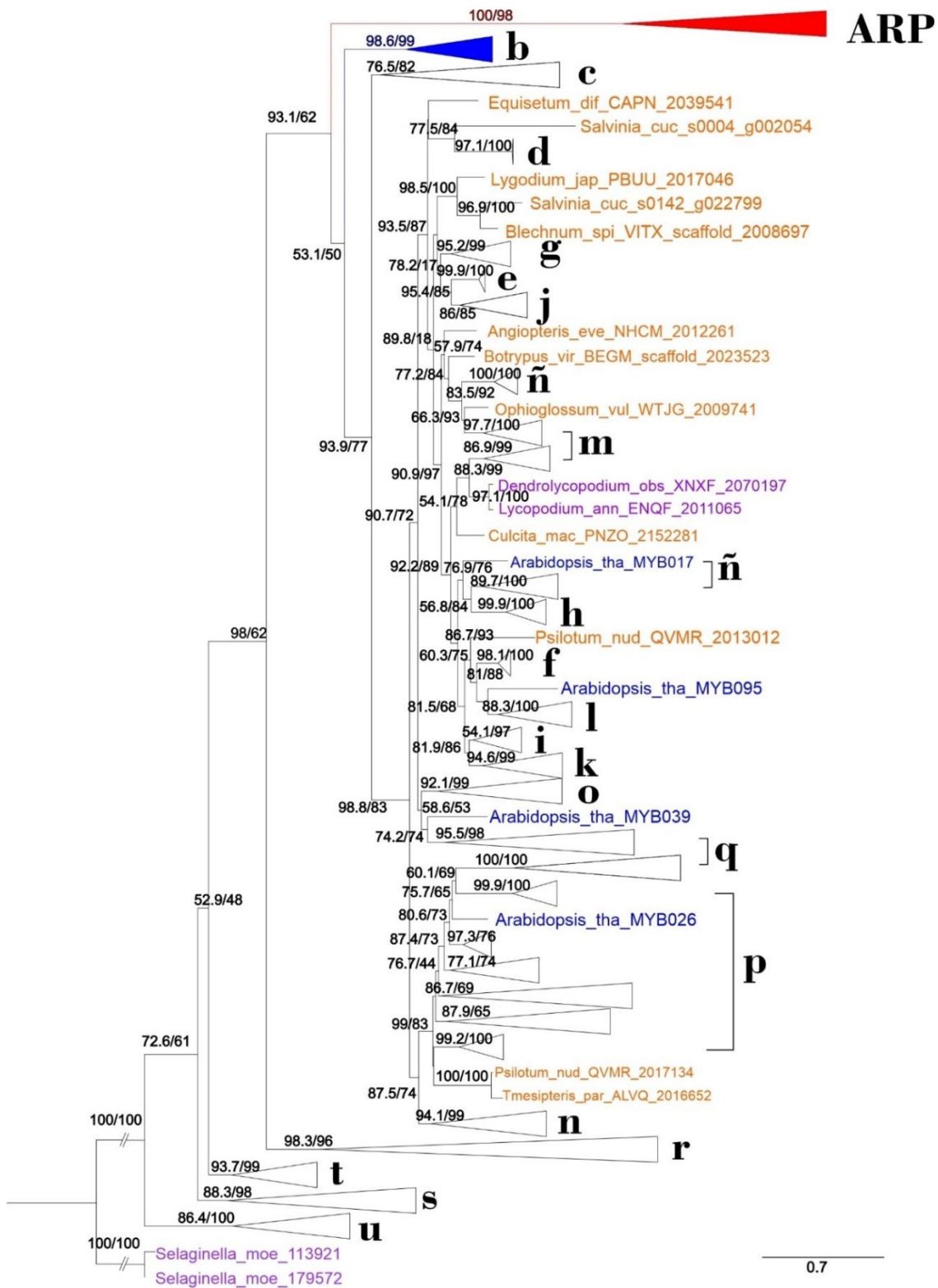
<i>Selaginella moellendorffii</i>	Selaginella_moe_6282	6282	Phytozome	genoma	Etapall
	Selaginella_moe_72632	72632	Phytozome	genoma	Etapall
	Selaginella_moe_77255	77255	Phytozome	genoma	Etapall
	Selaginella_moe_78303	78303	Phytozome	genoma	Etapall
	Selaginella_moe_78848	78848	Phytozome	genoma	Etapall
	Selaginella_moe_80215	80215	Phytozome	genoma	Etapall
	Selaginella_moe_80553	80553	Phytozome	genoma	Etapall
	Selaginella_moe_81112	81112	Phytozome	genoma	Etapall
	Selaginella_moe_84195	84195	Phytozome	genoma	Etapall
	Selaginella_moe_84369	84369	Phytozome	genoma	Etapall
	Selaginella_moe_84608	84608	Phytozome	genoma	Etapall
	Selaginella_moe_90153	90153	Phytozome	genoma	Etapall
	Selaginella_moe_93784	93784	Phytozome	genoma	Etapall
	Selaginella_moe_98669	98669	Phytozome	genoma	Etapall
	Selaginella_moe_99462	99462	Phytozome	genoma	Etapall
<i>Selaginella selaginoides</i>	Selaginella_sel_KUXM_2011240	KUXM_scaffold_2011240	One KP	transcriptoma	Etapal
<i>Selaginella stauntoniana</i>	Selaginella_sta_ZZOL_2006067	ZZOL_scaffold_2006067	One KP	transcriptoma	Etapal
<i>Selaginella wallacei</i>	Selaginella_wal_JKAA_2005609	JKAA_scaffold_2005609	One KP	transcriptoma	Etapal
<i>Selaginella willdenowii</i>	Selaginella_wil_KJYC_2011588	KJYC_scaffold_2011588	One KP	transcriptoma	Etapal
<i>Thyrsopteris elegans</i>	Thyrsopteris_ele_EW XK_2016753	EW XK_scaffold_2016753	One KP	transcriptoma	Etapal
<i>Tmesipteris parva</i>	Tmesipteris_par_ALVQ_2016652	ALVQ_scaffold_2016652	One KP	transcriptoma	Etapal
	Tmesipteris_par_ALVQ_2086856	ALVQ_scaffold_2086856	One KP	transcriptoma	Etapal
<i>Woodsia scopulina</i>	Woodsia_sco_YJY_2011080	YJY_scaffold_2011080	One KP	transcriptoma	Etapal
	Woodsia_sco_YJY_2012974	YJY_scaffold_2012974	One KP	transcriptoma	Etapal
	Woodsia_sco_YJY_2074662	YJY_scaffold_2074662	One KP	transcriptoma	Etapal
<i>Woodsia scopulina</i>	Woodsia_sco_YJY_2077181	YJY_scaffold_2077181	One KP	transcriptoma	Etapal
<i>Zea mays</i>	Zea_may_RS2_NM_001112039	NM_001112039.2	NCBI	genoma	Etapal



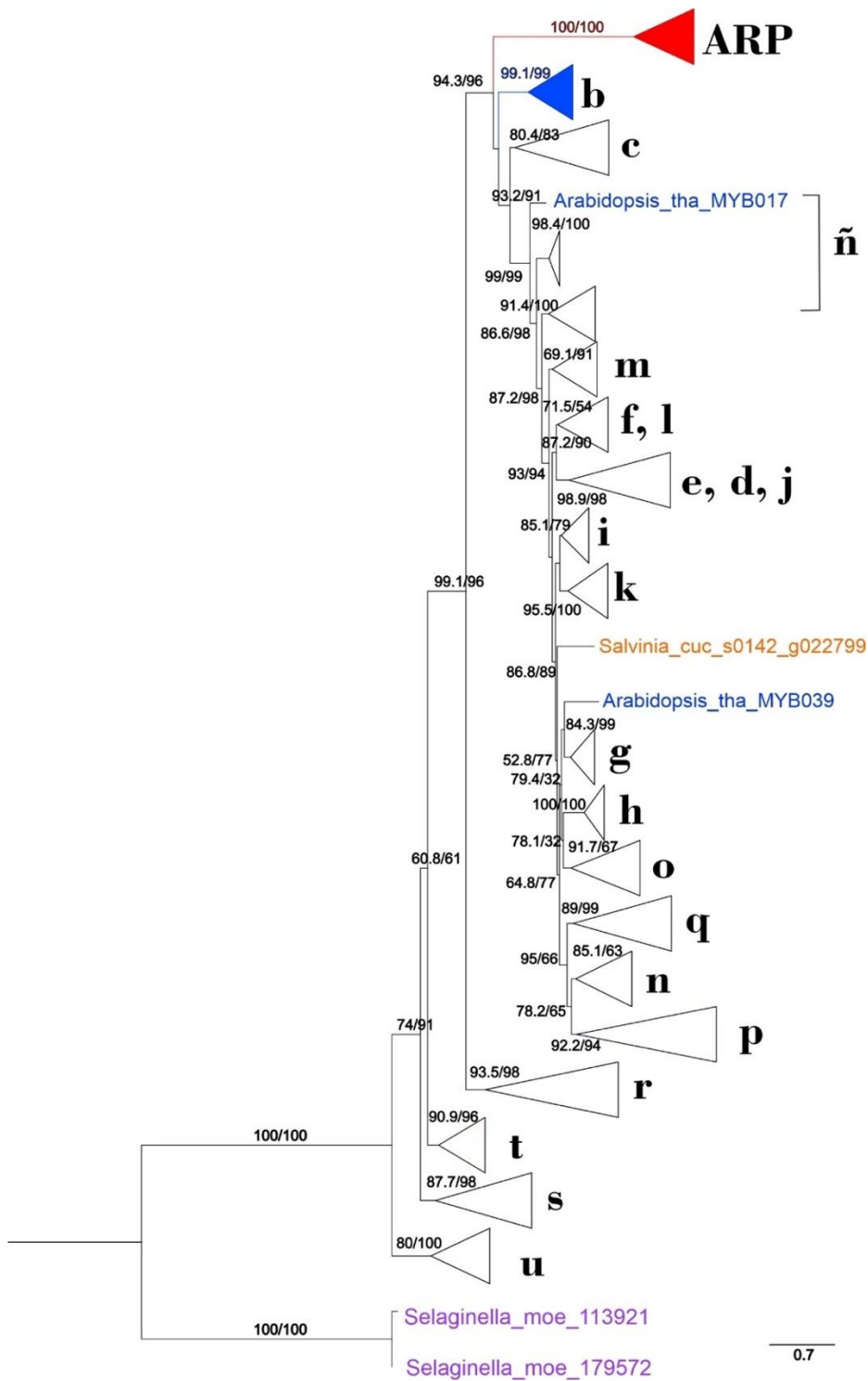
Suplementario 2. Filogenia de la Matriz G+T. Esta filogenia fue obtenida mediante un análisis de Máxima Verosimilitud en RaxML BlackBox (Stamatakis, 2014) en Cipres (Miller, et al. 2011), con el modelo de evolución molecular GTR-CAT.



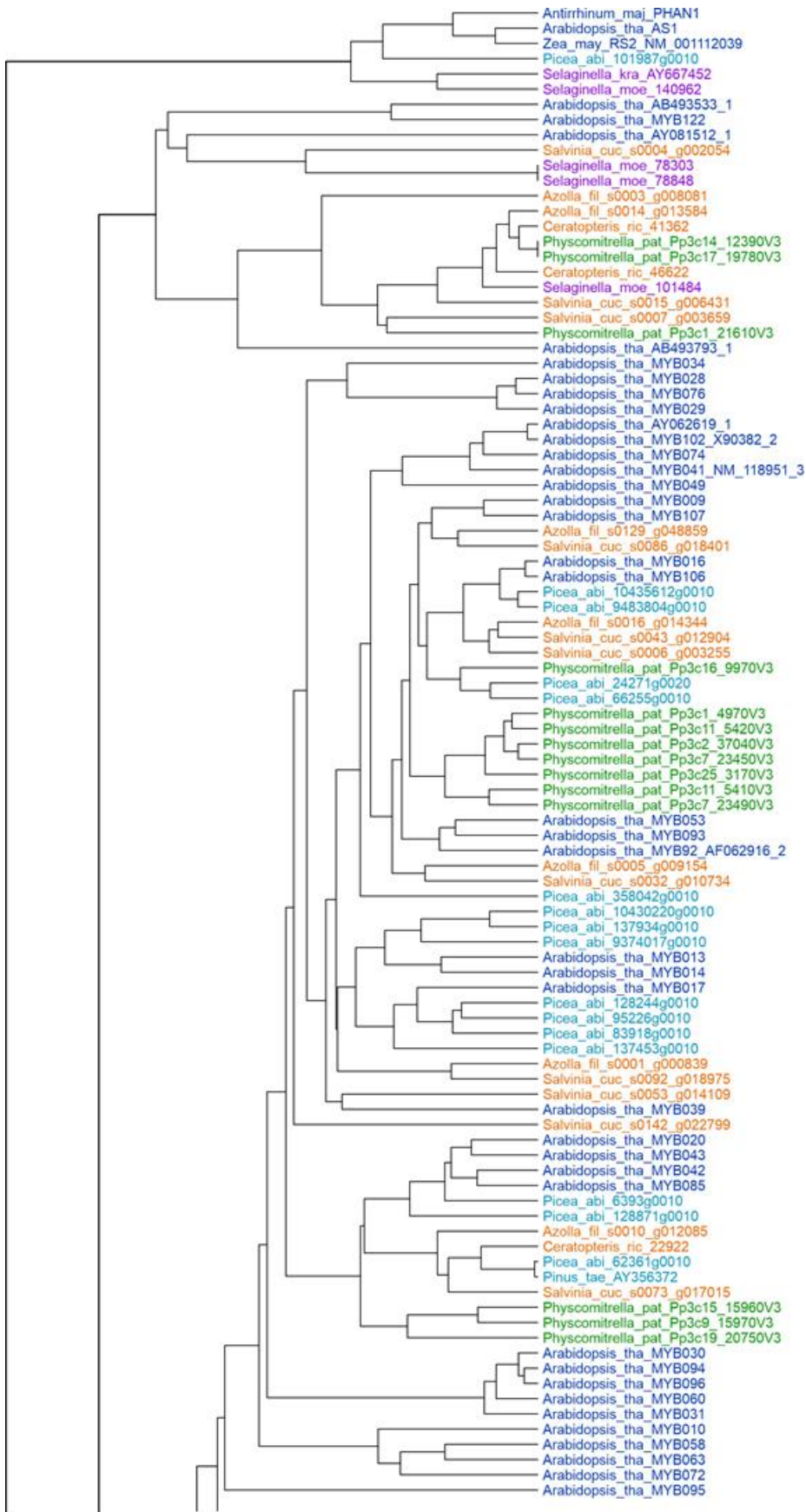
Suplementario 3. Filogenia de la Matriz G. Esta filogenia fue obtenida mediante un análisis de Máxima Verosimilitud en RaxML BlackBox (Stamatakis, 2014) en Cipres (Miller, et al. 2011), con el modelo de evolución molecular GTR-CAT.

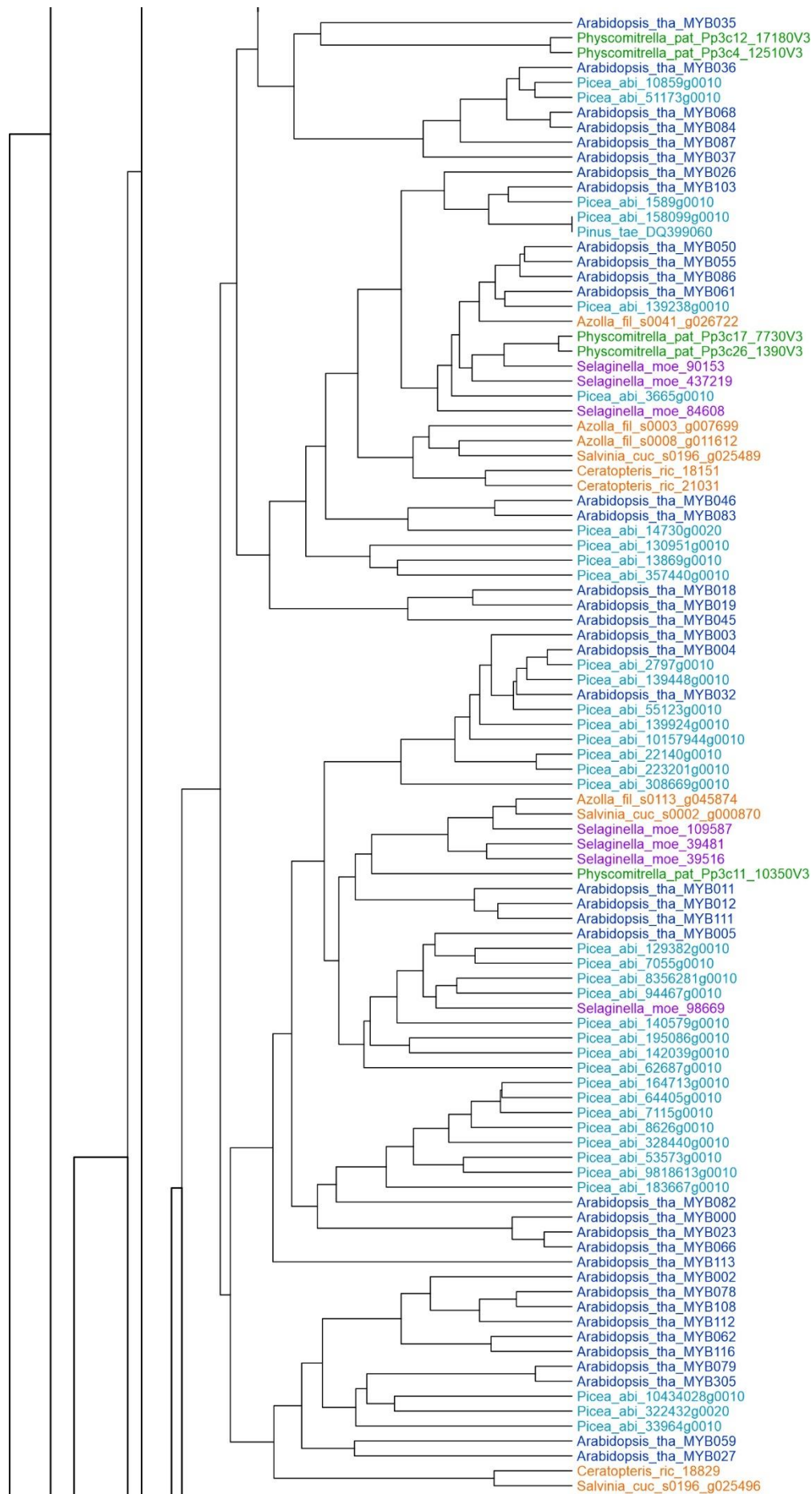


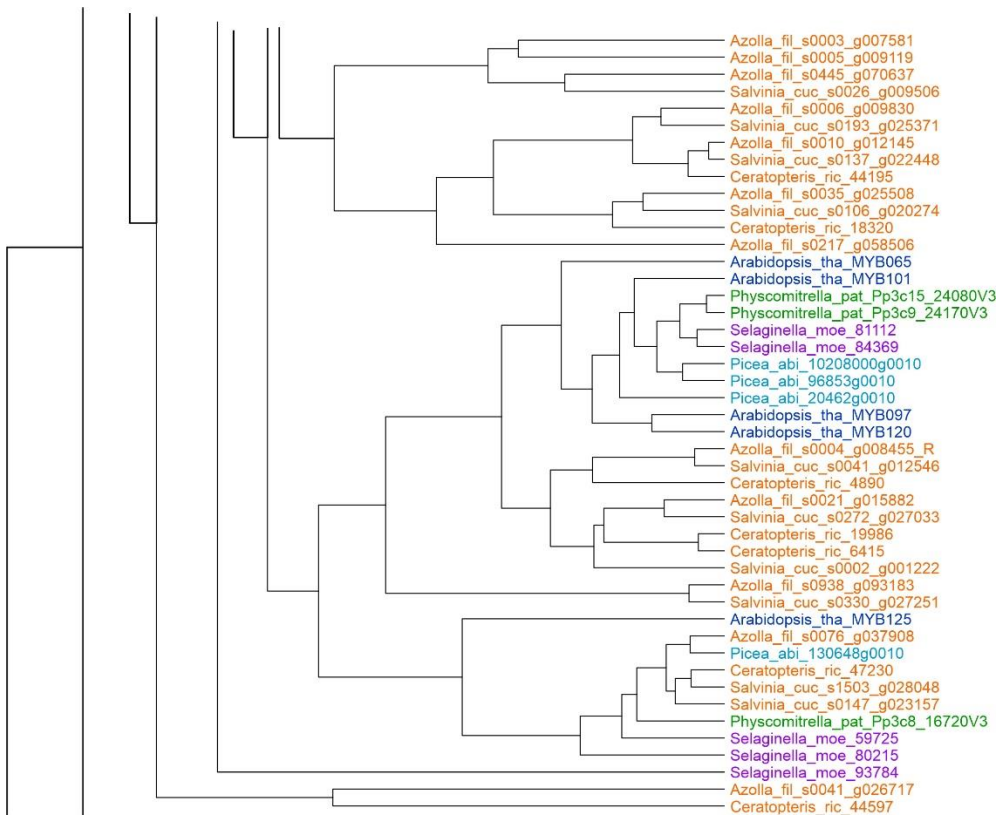
Suplementario 4. Filogenia de la Matriz G+T. Esta filogenia se obtuvo mediante un análisis de Máxima Verosimilitud en IQtree, con el modelo de evolución molecular GTR+G+I.

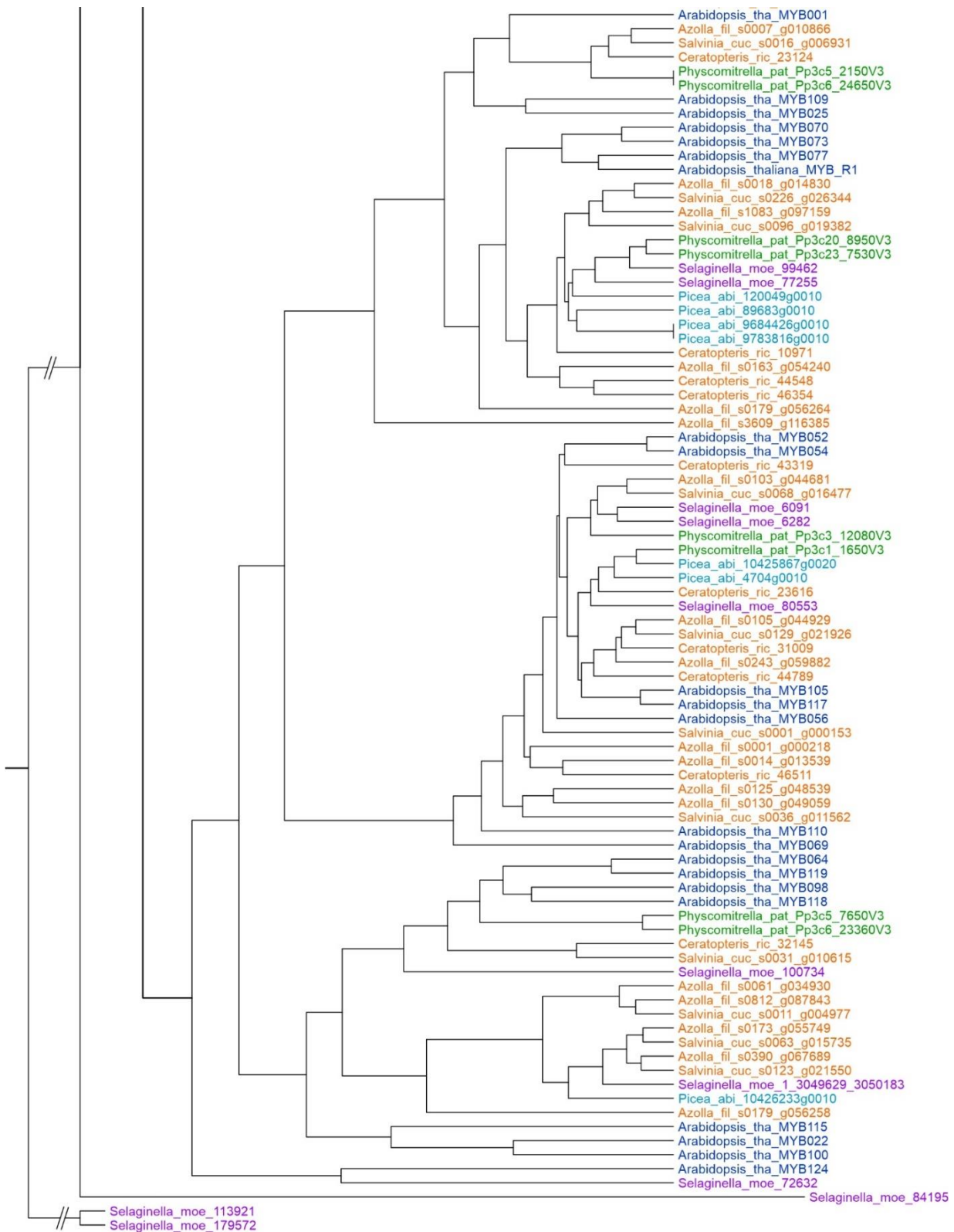


Suplementario 5. Filogenia de la Matriz G. Esta filogenia se obtuvo mediante un análisis de Máxima Verosimilitud en IQtree, con el modelo de evolución molecular GTR+G+I. Se muestran dos valores de Bootstrap, el valor de la izquierda fue obtenido con el enfoque de Aproximación Ultrarrápida Bootstrap (UFBoot)(Minh, Nguyen, & Von Haeseler, 2013). El valor de Bootstrap de la derecha muestra el soporte que tiene la rama con un enfoque tradicional de Máxima Verosimilitud.









Suplementario 6. Análisis de similitud. Este análisis fue realizado bajo el método de agrupamiento Neighbor joining, con la matriz G de aminoácidos en la plataforma de MAFFT (Kato, Rozewicki, & Yamada, 2017).



Suplementario 7. Motivos proteicos recuperados con el análisis de MEME. El tamaño de la letra que representa a cada aminoácido equivale al número de veces que ese aminoácido se encuentra en la misma posición en las secuencias del alineamiento. La cantidad de aminoácidos que conforman cada dominio está marcada por el eje inferior. El valor de E indica la probabilidad que existe de que esa secuencia de aminoácidos se haya encontrado por azar en el alineamiento. La longitud del motivo también indica el porcentaje de secuencias que comparten ese motivo