

00377



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

**POSGRADO EN CIENCIAS
BILÓGICAS
INSTITUTO DE ECOLOGIA**

**Evolución molecular de genes MADS-Box tipo
B y su papel en el origen y la evolución del
desarrollo floral**

T E S I S
QUE PARA OBTENER EL GRADO ACADÉMICO DE
MAESTRA EN CIENCIAS BIOLÓGICAS
(SISTEMÁTICA)
P R E S E N T A
BIOL. TANIA HERNÁNDEZ HERNÁNDEZ

DIRECTORA DE TESIS: DRA. ELENA ÁLVAREZ-BUYLLA ROCES

MEXICO, D. F.



SEPTIEMBRE 2005

COORDINACIÓN

m347771



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.



Autorizo a la Dirección General de Bibliotecas de la UNAM a difundir en formato electrónico e impreso el contenido de mi trabajo recepcional.
NOMBRE: Tania Hernández Hernández
FECHA: 13-jun-2005
FIRMA: Tania Hernández

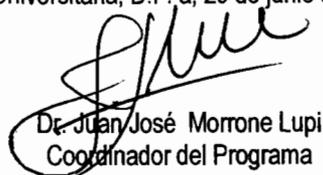
Ing. Leopoldo Silva Gutiérrez
Director General de Administración Escolar, UNAM
Presente

Por medio de la presente me permito informar a usted que en la reunión ordinaria del Comité Académico del Posgrado en Ciencias Biológicas, celebrada el día 13 de junio del 2005, se acordó poner a su consideración el siguiente jurado para el examen de grado de Maestría en Ciencias Biológicas (Sistemática) del(a) alumno(a) **Hernández Hernández Tania** con número de cuenta **95318417** con la tesis titulada: **"Evolución molecular de genes MADS-Box tipo B y su papel en el origen y evolución del desarrollo floral"**, bajo la dirección del(a) **Dra. Elena Álvarez-Buylla Roces**.

Presidente:	Dr. Daniel Piñero Dalmau
Vocal:	Dr. Lorenzo Segovia Forcella
Secretario:	Dra. Elena Álvarez-Buylla Roces
Suplente:	Dr. Luis Eguiarte Fruns
Suplente:	Dra. Susana Magallón Puebla

Sin otro particular, quedo de usted.

Atentamente
"POR MI RAZA HABLARA EL ESPIRITU"
Cd. Universitaria, D.F. a, 29 de junio del 2005


Dr. Juan José Morrone Lupi
Coordinador del Programa

c.c.p. Expediente del interesado

Este trabajo se llevó a cabo en el Laboratorio de Genética Molecular y Evolución del Desarrollo en Plantas del Instituto de Ecología de la UNAM, bajo la tutoría de la Dra. Elena Álvarez-Buylla Roces.

La investigación fue financiada con el apoyo de una beca para estudios de Maestría CONACyT (registro 172239), beca complemento DGEP-UNAM y beca Fundación TELMEX-HombreNaturaleza.

Los análisis presentados en este trabajo fueron realizados gracias al tiempo computacional facilitado por el Dr. Lorenzo Segovia en su cluster del Instituto de Biotecnología de la UNAM, y durante una estancia de investigación en su laboratorio se llevaron a cabo las etapas finales de este estudio y trabajo escrito.

El Comité Tutorial estuvo conformado por los doctores Daniel Piñero Dalmau, Susana Magallón Puebla y Elena Álvarez-Buylla Roces.

Los doctores Lorenzo Segovia Forcella, Luis Eguiarte Fruns, Daniel Piñero Dalmau, Susana Magallón Puebla y Elena Álvarez-Buylla Roces revisaron críticamente este documento y conformaron el Jurado de Examen de Grado.

ÍNDICE

Agradecimientos.....5

Abreviaturas.....9

Resumen.....10

Resumen (Inglés).....12

Introducción.....14

- I. Morfología floral e importancia de la aparición de la flor en la radiación de las angiospermas... 14
- II. Papel de los genes MADS-box tipo B en las bases genéticas del desarrollo floral... 17
- III. Características bioquímicas, función y evolución de Genes MADS-box tipo B... 20
- IV. Justificación del presente estudio.....24
Hipótesis y objetivos.

Secuencias y Métodos.....26

- I. Obtención de las secuencias de ortólogos MADS-Box tipo B de plantas.....26
- II. Alineamiento.....29
- III. Análisis filogenético.....29
- IV. Detección de heterogeneidad de tasas de sustitución entre sitios, entre linajes y presencia de selección positiva.....30
 - IV. a. Detección de heterogeneidad de tasas de sustitución (ω) entre sitios y selección positiva en sitios específicos.....32
 - IV. b. Detección de heterogeneidad de tasas de sustitución (ω) entre linajes.....33
 - IV. c. Detección de selección positiva en linajes particulares.....34
 - IV. d. Detección de sitios bajo selección positiva.....35
- V. Estimación de tiempos de divergencia.....36

Resultados.....37

- I. Selección del modelo de sustitución.....37
- II. Análisis filogenético.....37
 - II.a. Análisis filogenético de los genes MADS Box Tipo B en angiospermas y gimnospermas.....37

II.b. Análisis filogenético genes MADS tipo <i>PISTILLATA</i> (<i>PI</i>).....	39
II.c. Análisis filogenético genes tipo <i>APETALA3</i> (<i>AP3</i>).....	41
III. Inferencia de Selección Positiva (SP) y Diferencias significativas en las Tasas de Sustitución (DTS) durante la evolución de los genes MADS-box tipo B.....	44
III.a. Pruebas de detección de selección sitio-por-sitio.....	44
IIIb. Pruebas de detección de selección linaje-específicas.....	48
Papel de la selección adaptativa en la divergencia funcional de los MADS-Box Tipo B	
<i>Resultados Modelos A, B y D, genes MADS Tipo B...</i>	55
<i>Resultados Modelos A, B y D, genes MADS Tipo APETALA3...</i>	56
<i>Resultados Modelos A, B y D, genes MADS Tipo PISTILLATA...</i>	56
IV. Estimaciones de tiempos de divergencia.....	57
Discusión.....	61
Relaciones filogenéticas de genes MADS-box tipo B en plantas con semilla....	61
Diversificación funcional y selección positiva en los MADS Tipo B de gimnospermas.....	65
DTS y SP en Genes MADS Tipo B del clado ‘B-sister’.....	66
DTS y PS en Genes MADS Tipo B en angiospermas... ..	67
Importancia de la duplicación AP3-PI en la diversificación funcional de los MADS Tipo B de angiospermas... ..	67
Selección positiva en el dominio K de genes tipo PI y AP3; origen de la heterodimerización... ..	68
Importancia de la duplicación euAP3-TM6 e implicaciones en el origen y diversificación de las eudicotiledóneas superiores... ..	69
Selección positiva en el clado de los genes tipo PI... ..	70
Conclusiones.....	74
Perspectivas.....	76
Apéndice.....	78
Referencias.....	81

Agradecimientos

Por fin estoy aquí tratando de pensar en la mejor manera de agradecer esta tesis, pero resulta difícil lograr una manera poética después de estar trabajando en esto a marchas forzadas desde hace varias semanas porque a uno le urge titularse.

Evidentemente cuando pongo punto final a éste trabajo concluyo también una etapa de mi vida, quizás la más crucial y determinante, difícil y adversa, y la más importante hasta ahora; en la que he aprendido muchas cosas, tantas, que sólo puedo en resumen decir que en este tiempo me fui dando cuenta que la vida no es ni funciona como yo pensaba sino como es y funciona.

Se me empiezan a ocurrir entonces algunas ideas de cómo y a quién agradecer, y lógicamente el punto de partida está en donde termine mi tesis de licenciatura. Pasaron tres años y según el plan original -tramado hace muchos más por una muchacha ingenua que comenzaba sus estudios de biología-, ahora yo debería estar terminando un exitoso doctorado, habiendo publicado ya mi interesantísimo trabajo. La vida se interpuso en mi exitoso camino y resulta que ahora me encuentro en la ciudad de Cuernavaca, aún sin el exitoso doctorado siquiera empezado y tratando al menos de que el lector de estos agradecimientos reconozca como acto de fe que ésta es una buena tesis. Y no es que Cuernavaca sea feo, ni que mi tesis no me guste; por el contrario y para ser sincera, en mi andar hasta aquí he conocido gente maravillosa, he aprendido muchas cosas nuevas e interesantes y me he llegado a conocer y a concebir más y mejor a mí misma en mi lugar en este mundo, superando la adversidad que implica el haber llegado hasta aquí.

No tengo más que agradecer pues, en primer lugar sin duda alguna y con todo mi corazón a Fer, que me puso en este camino y segura estoy de que aún me cuida y lleva por donde más conviene a mi persona. Me ha costado trabajo asimilar todo lo que me enseñó. Por ahora una de esas cosas es dejarme caer con alegría, y tener confianza y fe siempre en que las cosas buenas pasarán en algún momento... aunque no pasen. Perico, tu valor y tu fuerza aún nos inspiran y tu sonrisa sigue iluminando nuestras vidas.

Enseguida todos los demás agradecimientos, personales también porque después vienen los académicos. A mi mamá y mi papá, siendo que indudablemente en mis actividades cotidianas percibo siempre la vocación y el gusto por el conocimiento que me transmitieron ellos. Enseguida, claro, a mi hermano Miguel, por que lo quiero y querré muchísimo siempre.

A mis queridas tías, porque de verdad que sin su apoyo ni yo ni mi familia hubiéramos salido adelante. Ojalá algún día pueda yo hacer por ellas al menos un poquito para compensar todo lo que ellas hicieron por nosotros.

También quiero insertar aquí en este preciso lugar un pequeño párrafo para Chate, por que me enseñó muchas cosas y me ayudó mucho en su momento. Todavía a veces me deja recargarme en él cuando lo necesito.

Y ahora a toda la gente del Instituto de Ecología, pues durante mi estancia ahí llegué a sentir su afecto y su aprecio. En especial y con mucho cariño a la chata Andresita, pues un día pasaba yo por ahí cuando de pronto y sin conocerla apareció y me extendió la mano para darme de regalo, gratis y sin compromiso su amistad; así de fácil y además, como un extra de promoción especial, me presentó a Xitlali. También, por supuesto, a Rosalinda, que para escuchar penas es la mejor. A Carlos y Anidia, Alicia Gamboa, Mitzi, Esther, Arturo, León, Rigo, Aida, Enrique, y

todos los del Laboratorio de Elena Álvarez que por ahí han pasado en coincidencia con mi paso. También a Caro Espinosa que hace que uno padezca un poco menos la burocracia.

Los siguientes agradecimientos no se si ponerlos en los personales o los académicos, y por eso entonces los pongo aquí en medio. Muy especialmente y con mucho cariño al trío de güeritos. Arturo Becerra y Luis Delaye por que ellos me llevaron de la mano al oscuro túnel de la evolución molecular y me incitaron al vicio de dedicarme a la biología computacional y olvidarme de los experimentos. Me han dado consejos académicos de cuates (que son diferentes de los de profes) muchas veces, y por culpa de ellos me han pasado muchas cosas buenas. Además, nada más de ver sus padeceres me han hecho darme cuenta que los míos son normales, como los de todos los biólogos. En general y muchas veces los tres güeritos me han dado mucha orientación, pero al tercer güerito le agradezco mucho en particular el cariño que me dio y la facilidad con que se puso a escuchar mis historias. El hecho de que el tercer güerito crea en mi y piense lo que piensa de mi persona le da un gran respiro de alivio a mi confianza: Alexander de Luna.

Académicamente empiezo por el principio. A Elena Álvarez-Buylla por todo su apoyo, orientación y por que ha representado siempre para mí un ejemplo a seguir en cuanto a su desempeño académico y sus capacidades intelectuales. Ojalá algún día pudiera yo tener la lucidez y claridad que ella tiene. Usualmente digo que mi formación puedo atribuirla más o menos una tercera parte a la Facultad de Ciencias y lo demás a su laboratorio.

Como miembros de mi comité tutorial a Daniel Piñero y especialmente a Susana Magallón. Resulta fascinante escucharla hablar con tanta pasión y conocimiento de las plantas. Sin su ayuda esta tesis carecería de toda interpretación y sentido biológico, sería escueta, aburrida y fea, y no me habría dado cuenta de la relevancia evolutiva de lo que en mi trabajo encontré (o que creo que encontré). Si al final resulta que me dedico a la botánica será sin duda un poco por culpa de Susana Magallón y Judith Márquez.

A León Patricio Martínez por que amablemente me dijo por dónde sin tener por qué hacerlo. Su orientación fue fundamental para que no me perdiera yo en un laberinto infinito y no terminará mi maestría en mucho más tiempo. Creo que León Patricio Martínez y Francisco Vergara Silva fueron elementos importantísimos para que en el transcurso de mi formación y la de otros descubriéramos y discutiéramos conceptos relevantes de biología evolutiva que hasta el momento siguen determinando mi línea de pensamiento (en particular pertinentes a la evolución del desarrollo), a ellos les agradezco haber compartido sus conocimientos con los estudiantes más jóvenes.

Al Consorcio de las Américas por la Ciencia Interdisciplinaria, en especial al Profesor Nitant Kenkre, pues me llevo a pasar una muy bonita estancia a Nuevo México y gracias a eso me pude comprar a Lachula y tener por fin un lugar estable donde trabajar. De Nitant también aprendí muchas cosas, por que es una persona admirable y un líder nato.

A Lorenzo Segovia porque me rescató de la miseria que implica encontrarse de pronto como un biólogo desempleado y evitó que empezara a considerar bailar en calzones, pero más importante que eso, porque me dio la oportunidad de aprender cosas nuevas y de darme cuenta de que lo que he aprendido es valioso; por que con confianza me abrió las puertas con su equipo de trabajo y con ellos me di cuenta de cómo el jefe de grupo puede hacer que la cotidianidad de la investigación en un laboratorio sea bien padre. Además le agradezco que me haya prestado a su noble Burra, por que he de confesar que con ella se llevó a cabo la fase terminal de esta tesis.

Y ya de su laboratorio pues a todos, pero en especial a Iliana, Alejandro, Arcadio, Vivis, Pavel, Daniela, y los demás, por que hicieron de mi estancia en el IBT mucho menos traumática; incluso bastante agradable.

A las amigas del Atillo, que marcaron junto con ese barrio esta etapa de mi vida. En especial a Raquel Quezada, a Marie Petronille y a la señora Mary Carmen. También a mi amigo Era, por que frecuentemente me recuerda que la vida hay que disfrutarla.

Al final quiero decir una cosa que he estado pensando, pero que ya me habían dicho antes. La carrera de un científico puede llegar a ser como la de un monje penitente en un camino escarpado, de subida siempre, sin dinero y con pocos aplausos, aunque eventualmente desemboca en una vida sencilla, cómoda y agradable. Mientras tanto, un científico joven necesita mucho apoyo para no saltar por la borda y abandonar el barco de la Investigación. Entonces pues, agradezco especialmente a todos los que me han apoyado de muchas maneras distintas para llegar hasta aquí y que no puse en mis agradecimientos.

Gracias,
Tania Hernández
Abril, 2005

Abreviaturas

PP : probabilidad posterior

SP = selección positiva

DTS = diferencias en las tasas de sustitución

MV = máxima verosimilitud

PTV = Pruebas de Tasas de Verosimilitud

Arabidopsis thaliana = *A.thaliana* = *Arabidopsis*

Antirrhinum majus = *A.majus* = *Antirrhinum*

AP3 = proteína APETALA3

PI = proteína PISTILLATA

AP3 = gen *APETALA3* o clado de ortólogos correspondientes

PI = gen *PISTILLATA* o clado de ortólogos correspondientes

ap3 = mutante nula para el gen *AP3* en *Arabidopsis thaliana*

Resumen

Los genes MADS-Box tipo B tienen un papel fundamental en los mecanismos moleculares de determinación de órganos florales. Estos genes son parte del modelo ABC de desarrollo floral y en *Arabidopsis thaliana* son *APETALA3* y *PISTILLATA*, cuya expresión conjunta con la de los genes tipo A (*APETALA1* y *APETALA2* en *Arabidopsis*) en el meristemo floral determinan la formación de los pétalos, mientras que en conjunto con *AGAMOUS*, que es el único gen de la función C descrito en *Arabidopsis*, determinan la formación de los estambres. Un enfoque posible para explorar la evolución de los mecanismos de desarrollo que subyacen los cambios morfológicos es estudiar los patrones de evolución molecular de los genes que, como los tipo B, juegan papeles clave durante el desarrollo.

La evolución de los genes tipo B se caracteriza por complejos patrones de eventos de duplicación. Además de la duplicación principal AP3-PI hay muchas otras duplicaciones en diferentes niveles filogenéticos, en particular en los genes tipo AP3. Hay dos clases de genes tipo AP3 de acuerdo a motivos específicos en su C-terminal. El primero y más antiguo es el grupo de los genes tipo paleoAP3, que son característicos de las monocotiledóneas, magnoliales y eudicotiledóneas basales. Agrupados en el mismo clado que los genes euAP3, el segundo tipo de genes AP3, encontramos como producto de una duplicación reciente al clado de los TM6, que poseen en su C-terminal el dominio paleoAP3. Estos parálogos han sido encontrados solo en las eudicotiledóneas superiores y su función aún no es bien conocida. Se ha hipotetizado que esta duplicación pudo haber tenido un papel importante en el origen independiente de pétalos en este clado.

En *Arabidopsis* la función B requiere de la heterodimerización de las proteínas AP3 con PI y otras proteínas MADS, y se piensa que la heterodimerización obligada tuvo sus orígenes en proteínas B capaces de homodimerizar, ya que los genes B en gimnospermas se unen a DNA como homodímeros. Las duplicaciones en la familia de genes tipo B debieron haber jugado un papel crucial en sus procesos de diversificación funcional, que pudo haber ocurrido por mutaciones estratégicas en los dominios de unión a otras proteínas. La formación de multímeros en distintas combinaciones puede ser una de las formas en que se han originado nuevos complejos transcripcionales con funciones regulatorias específicas, y puede ser una de las bases mecanísticas para el origen y diversificación de estructuras morfológicas.

En este estudio exploramos los patrones de variación de las secuencias disponibles para los genes B, en una búsqueda de entender mejor sus procesos de diversificación y las fuerzas evolutivas que lo han dirigido en relación con la evolución de la flor. En específico se hicieron pruebas para encontrar diferencias significativas en las tasas de sustitución (DTS) en linajes de genes, que resultan de eventos importantes de duplicación o que caracterizan a grupos taxonómicos particulares. Posteriormente se probó si las diferencias eran debidas a selección positiva (SP) en sitios particulares de la proteína. Nos concentramos en los siguientes eventos: (1) la separación de los genes tipo B de gimnospermas del clado de los genes tipo AGL34 (o más tarde llamados 'B-sister'), (2) la duplicación que dio lugar a los linajes tipo B de angiospermas y gimnospermas y que coincidió con el origen de la flor, (3) el origen de los linajes AP3 y PI en los clados de las angiospermas basales y eudicotiledóneas que es el momento clave para hacer inferencias acerca de las fuerzas evolutivas importantes durante el origen de la heterodimerización, (4) el origen de los linajes de genes AP3, paleoAP3 y euAP3, y finalmente (5) la duplicación de los genes euAP3-TM6 en las eudicotiledóneas superiores para plantear

hipótesis acerca de su papel durante la evolución de pétalos diferenciados en este grupo. Este estudio además es útil para plantear hipótesis acerca del papel funcional de los distintos residuos de las proteínas MADS tipo B.

Los resultados presentados aquí sugieren fuertemente que después de la duplicación que dio origen a AP3 y PI se dio una diversificación funcional de estos dos linajes primero por selección positiva en diferentes sitios del dominio K, cuyo papel putativo es promover la heterodimerización con otras proteínas; y posteriormente en diferentes sitios de la proteína incluyendo la caja MADS, y las regiones I y el C-terminal. También encontramos que la fecha de la duplicación fue a los 290 ma, que es una fecha mucho más cercana que la reportada anteriormente, justo después de la separación de las angiospermas y gimnospermas.

Abstract

B type MADS-box genes play a fundamental role in the molecular mechanisms of floral organ determination. Described in the ABC model of flower development, B genes are *APETALA3* and *PISTILLATA* in *Arabidopsis thaliana*, and their expression in the floral meristem together with A gene expression (*APETALA1* and *APETALA2* in *Arabidopsis*) determine petal formation, while together with C gene expression (*AGAMOUS* is the unique C type gene described so far in *Arabidopsis*) determine stamen formation. A possible way to address the evolution of developmental mechanisms that underlie morphological changes is to study the molecular evolutionary processes of genes that, like B type genes, play key roles during development.

B MADS-box gene evolution is characterized by a complex pattern of duplication events. Besides the main AP3-PI duplication there are many other ones, particularly in the AP3 lineage. There are two kinds of AP3 genes according to particular motifs found in their C-terminal region. The first and most basal one is the paleoAP3 lineage, found in AP3 genes from monocots, magnoliales and basal eudicots. In the same clade of euAP3 genes -the second main AP3 lineage-, and product of a recent duplication we can find the group of TM6-like sequences, with the paleoAP3 motif in their C-terminal region. These paralogs have been only found in core eudicots and their function is not completely understood. It has been hypothesized that the euAP3-TM6 duplication could have an important role in the independent origin of defined petals in this plant group.

In *Arabidopsis* B function requires AP3 and PI to heterodimerize between them and with other MADS proteins, and it has been proposed that obligate heterodimerization could have its origins in proteins with homodimerizing capabilities, because B proteins in gymnosperms can bind DNA as homodimers. The duplications in B genes must have had a determinant role in their functional diversification processes that could be driven by strategic mutations in protein-protein interaction domains. Multimer formation in different combinations can be one way to originate new transcriptional complexes with different specific regulatory functions, and can be one mechanistic base for the origin of morphological novelties.

In this study we explore the variation patterns of available B gene sequences, searching for a better understanding of the diversification patterns and evolutionary processes driving them in relation to the possible impact in flower evolution. We made statistical tests to detect significant differences in substitution rates (DTS) in gene lineages product of duplication events or lineages characterizing taxonomical groups. After that we tested if the DTS found were due to positive selection (SP) in particular protein residues. We concentrated in the following events: (1) the split of B genes and the AGL34 (lately called 'B-sister') clade, (2) the duplication that lead to angiosperm and gymnosperm B gene lineages that coincides with flower origins, (3) the origin of *AP3* and *PI* genes in angiosperms to explore the evolutionary forces at important sites driving heterodimerization origins, (4) the origin of the AP3 lineages paleoAP3 and euAP3 and finally (5) the euAP3-TM6 duplication in the core-eudicots clade given the possible importance of this duplication in the evolution of defined petals in this group. This study is also useful to formulate hypothesis concernig the functional role of different residues in B MADS-box proteins.

Results presented here strongly suggest that after the duplication originating AP3 and PI, there occurred functional diversification on both lineages given first by positive selection in different sites of the K domain, whose main role is promoting heterodimerization with other

proteins; and later in different sites of the protein, including the MADS box, and the I and C-terminal region. Among other results we also estimated a date of 290 MYA for the AP3-PI duplication, closer to estimated dates reported previously, just after the split of angiosperms and gymnosperms.

Introducción

I. Morfología floral e importancia de la aparición de la flor en la radiación de las angiospermas.

Las plantas con flores o angiospermas aparecieron hace aproximadamente 132 ma durante el Cretácico temprano, en una fecha muy reciente si se la compara con los tiempos de origen de las otras plantas con semilla; y en poco tiempo llegaron a dominar las comunidades vegetales de la tierra (Ligard y Crane 1988; Lupia et al. 1999 citados en Magallón y Sanderson 2001). Algunas estimaciones y conteos han mostrado que hay al menos 234, 000 especies de plantas con flores descritas en cerca de 437 familias (Thorne, 1992), lo cual excede en número a la diversidad en conjunto de todos los otros grupos de plantas (Sporne, 1974; Prance, 1977). A pesar de que al interior de las angiospermas podemos encontrar una enorme variación en su morfología, se puede reconocer a la presencia de la flor cómo su elemento distintivo, siendo que las estructuras florales proveen muchos de los caracteres empleados para su clasificación. Sin embargo, es importante destacar que también la característica flor de las angiospermas exhibe una gran variación, lo cual ha representado un desafío para encontrar homologías claras entre los caracteres florales y una circunscripción no ambigua del grupo (Friis et al, 1987).

La filogenia propuesta para las angiospermas reconstruida utilizando secuencias de los genes *rbcL*, *18S* y *atpB* (Figura 1) (APG III, 2003) permite hacer una distinción de dos clados monofiléticos: las eudicotiledóneas y las monocotiledóneas; y las magnoliales se ubicaron en un mismo clado junto con éstas últimas pero con una posición no resuelta (Chase, et al. 1993; Crane, et al. 1995; Qiu, et al. 1993). El clado de las eudicotiledóneas puede subdividirse en eudicotiledóneas basales, que comprende las Ranunculales y Proteales entre otras, y eudicotiledóneas superiores (core-eudicots), en donde se distinguen los grupos Rosidae y Asteridae que incluyen a la mayoría de especies de plantas con flores (APGII 2003). Además, se logra identificar a Amborellaceae y Nymphaeaceae como angiospermas basales, lo cual sugiere que el ancestro común de las angiospermas existentes ya tenía flores hermafroditas con perianto (algún tipo de estructuras estériles rodeando las reproductivas).

Los elementos estructurales que conforman una flor típica son los órganos florales, y en las eudicotiledóneas superiores pueden distinguirse órganos estériles (sépalos y pétalos) y órganos reproductivos (estambres y carpelos) organizados en verticilos arreglados concéntricamente surgiendo de un solo nodo en el eje de un meristemo floral determinado. Podemos ejemplificar la disposición de los órganos florales en las eudicotiledóneas superiores con el esquema que se muestra en la Figura 2. Aunque el número de órganos y verticilos puede cambiar, el primer y segundo verticilo conforman el perianto y contienen los sépalos y pétalos como apéndices estériles formando el cáliz y la corola respectivamente. El tercer verticilo contiene los estambres y el cuarto verticilo los carpelos, ambos órganos reproductivos masculinos y femeninos.

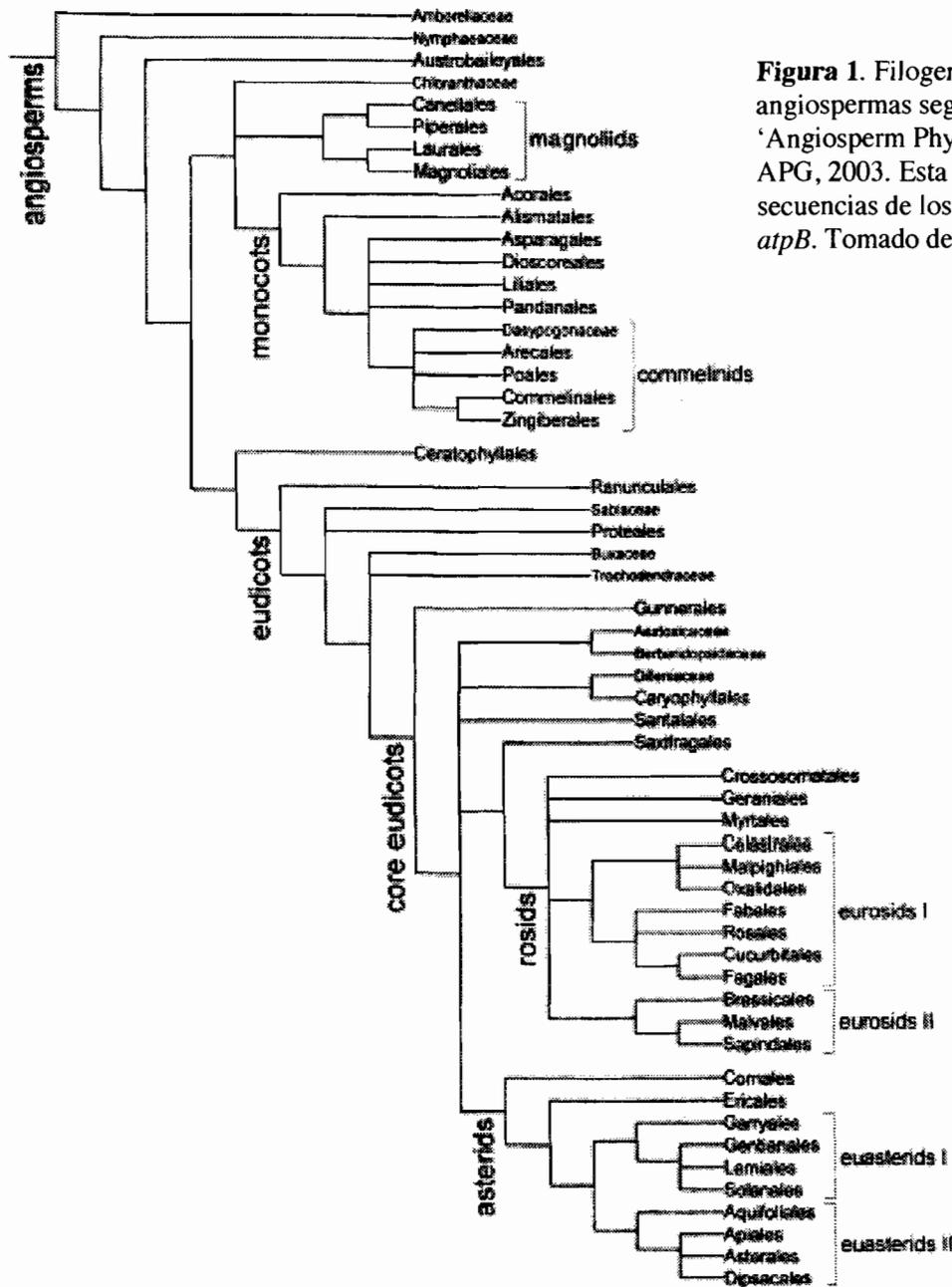


Figura 1. Filogenia de las angiospermas según el trabajo del 'Angiosperm Phylogeny Group', APG, 2003. Esta basada en secuencias de los genes *18S*, *rbcl* y *atpB*. Tomado de APGII 2003.

Un perianto con sépalos y pétalos es característico de las eudicotiledóneas superiores, mientras que en los grupos de angiospermas basales y en las monocotiledóneas la distinción en la mayoría de los casos es problemática (Endress, 2001). Los sépalos en el primer verticilo son usualmente verdes y por lo general protegen a la flor durante su desarrollo, mientras que los pétalos son coloridos, con una función general de atracción visual, aunque con frecuencia intervienen en la producción de néctar y compuestos aromáticos. En muchos casos la morfología de los sépalos y pétalos refleja una especialización hacia la polinización por varios grupos de insectos (Proctor & Yeo, 1973; Faegri & Van der Oijil, 1979 en Friis et al., 1987).

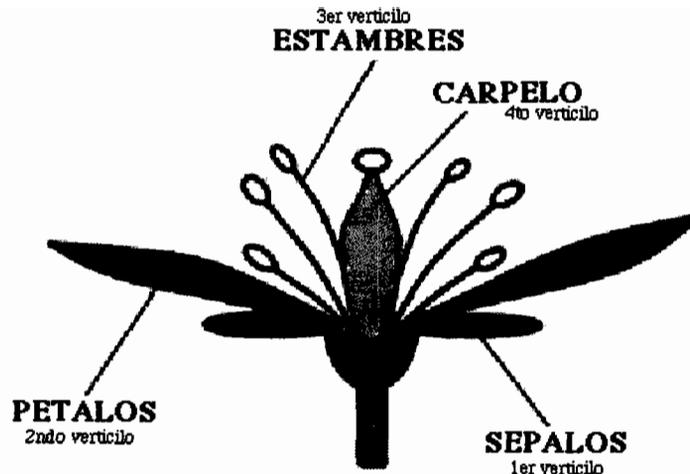


Figura 2. Diagrama de una flor de *Arabidopsis thaliana*. El primer y segundo verticilo conforman el perianto, con los órganos estériles (sépalos y pétalos). Enseguida al centro surgen del tercer y cuarto verticilo los estambres y carpelos respectivamente.

Se piensa que los estambres y carpelos han aparecido una sola vez en las angiospermas, pero en el caso de los órganos estériles (sépalos y pétalos) es aceptada la idea de que han aparecido en múltiples ocasiones, aunque los detalles de estos eventos no están del todo resueltos (Cronquist, 1988; Takhtajan, 1991; Endress, 1994 citados en Kramer et al., 1998). Existen pétalos estaminoides llamados andropétalos, que son derivados de los estambres y que han evolucionado muchas veces en las eudicotiledóneas basales y al menos una vez en la base de las eudicotiledóneas superiores y en las monocotiledóneas (Kramer et al; 1998). Hay también un segundo tipo de pétalos: los bracteopétalos, derivados de sépalos u otros órganos estériles. Las designaciones de pétalos ya sea como andropétalos o bracteopétalos se han hecho con base principalmente en caracteres morfológicos y gran parte de la controversia con respecto a la evolución del perianto se ha centrado en el número y naturaleza de los eventos de derivación de pétalos dentro de los diferentes linajes de angiospermas. Las inconsistencias en cuanto a la reconstrucción de la historia evolutiva de los diferentes caracteres dentro del grupo y la carencia, hasta hace poco, de una filogenia consensada de las angiospermas han dificultado la determinación del origen y la historia evolutiva de estos órganos florales. Estudios evolutivos de los genes que regulan el desarrollo de estos órganos pueden ayudar a resolver algunas de las preguntas pendientes en cuanto a su origen y diversificación.

En un contexto evolutivo, los elementos estructurales de las flores pueden ser definidos como el resultado de innovaciones clave (Endress, 2001). Las innovaciones clave son nuevos caracteres que adquieren un papel biológico indispensable que hace que se conserven a pesar de los cambios en las presiones adaptativas (Muller y Wagner, 1991 citados en Endress, 2001). Si pensamos en términos de la distribución de caracteres en grandes clados, el resultado de la aparición de una innovación clave es la presencia de dicho carácter de manera estable en los miembros del grupo, siendo que no se observan o se observan muy pocas reversiones al estado previo. Dicho de otra manera, mientras menos se revierta una sinapomorfía al interior de un gran clado y mientras más grande sea el clado, esa sinapomorfía es mejor interpretada como una innovación clave (Endress, 2001) que puede correlacionarse con la radiación o incremento en el número de taxa en dicho clado. Las angiospermas como grupo son un ejemplo inusual de

diversidad taxonómica (Heywood 1993 citado en Magallón y Sanderson, 2001) que aparentemente emergió en muy poco tiempo (Magallón y Sanderson 2001), y el estudio de las posibles causas de esta radiación deben dirigirse hacia el origen y evolución de sinapomorfías del grupo, como la flor, ya que muchas características relacionadas con las estructuras reproductivas florales pueden interpretarse como innovaciones clave de las angiospermas.

El éxito de la flor como órgano reproductivo e incluso su origen mismo ha sido interpretado como el resultado de una co-evolución con insectos u otro tipo de polinizadores. (Grant, 1950; Baker & Hurd, 1968; Burguer, 1981; Stebbins, 1981 citados en Crepet y Friis, 1987). La polinización por insectos confiere ventajas como la promoción del intracruzamiento, competencia gametofítica y mejoras en la eficiencia energética (Cruden, 1977; Mulcahy, 1979 citados en Crepet y Friis, 1987). Los taxa de angiospermas más diversos pueden asociarse con polinizadores específicos, y las comunidades vegetales mas complejas tienen un número mucho mayor de taxa de angiospermas con polinizadores específicos que de aquellos polinizados por viento (Stebbins, 1981; Kevan y Baker, 1983 citados en Crepet y Friis, 1987). Hay evidencia basada en el registro fósil de los insectos y en la forma de varios órganos reproductivos de gimnospermas que sugiere que los insectos fueron evolucionando hacia una especialización para alimentarse de las estructuras reproductivas de las plantas durante el intervalo entre el carbonífero-mesozoico (Crepet y Friis, 1987). La polinización específica es aparentemente un factor importante en la diversificación de las angiospermas (Friis y Crepet, 1987), de manera tal que la variación morfológica en caracteres como los pétalos en el perianto floral, que en muchos casos funcionan como atractores de polinizadores, representan un buen ejemplo de caracteres que pudieron haber evolucionado en respuesta a este tipo de presiones adaptativas. El estudio de la evolución de los mecanismos de desarrollo de los pétalos puede ayudar a esclarecer su historia evolutiva y a entender los orígenes de la diversidad de estas estructuras en las angiospermas.

II. Papel de los genes MADS-box tipo B en las bases genéticas del desarrollo floral.

El aislamiento de mutantes florales homeóticos en las plantas modelo *Arabidopsis thaliana* y *Antirrhinum majus* ha sido la base para empezar a desentrañar los mecanismos genéticos que subyacen el desarrollo floral. Los análisis de estas mutantes han permitido el desarrollo de un modelo mecanístico de determinación de órganos florales que incluye tres clases de genes conocidos como genes tipo A, B y C; que interactúan de manera combinatoria para especificar la identidad de los órganos (Bowman et al. 1991; Coen y Meyerowitz 1991; Meyerowitz et al. 1991). Se han descrito dos genes parálogos de la clase B en *Arabidopsis thaliana*: *APETALA3 (AP3)* y *PISTILLATA (PI)*. En particular, las mutaciones ocurridas en los genes tipo B dan lugar a fenotipos florales característicos de pétalos transformados a sépalos y estambres transformados en carpelos (ver Figura 3), lo cual indica que los productos de estos genes en los tipos silvestres son requeridos para especificar la identidad de estambres y pétalos (Jack et al. 1992; Goto y Meyerowitz 1994). Como la mayoría de los genes comprendidos en el modelo ABC, los genes tipo B en *Arabidopsis* codifican para factores transcripcionales y pertenecen a la familia de genes MADS-box.

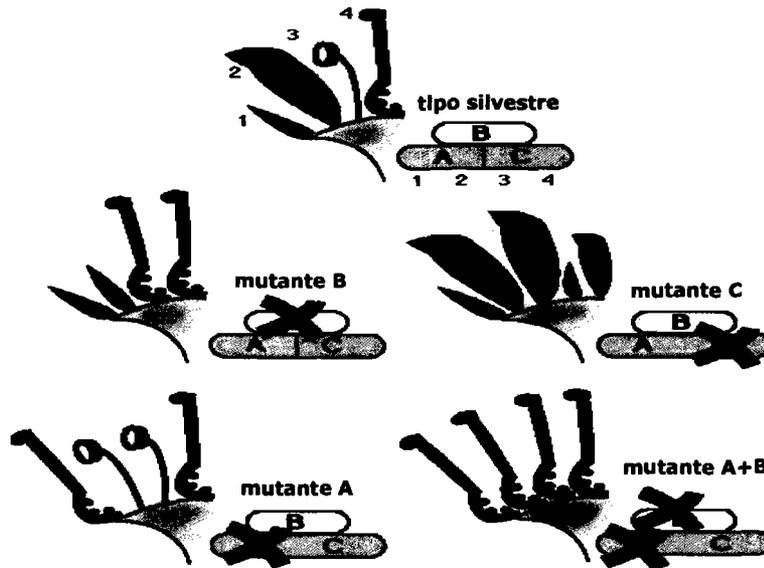


Figura 3. Esquemización del modelo ABC de desarrollo floral y fenotipos resultantes de las mutaciones nulas en los tres tipos de genes en *Arabidopsis*. En el tipo silvestre los genes tipo A se expresan en el primer verticilo del meristemo floral, dando lugar a la determinación de los sépalos. En el segundo verticilo la expresión conjunta de los genes tipo A y B da lugar a la determinación de pétalos; en el tercer verticilo, la expresión conjunta de los genes tipo B y C determinará los estambres y por último, en el cuarto verticilo la expresión de los genes tipo A determinará el desarrollo de carpelos. Las mutaciones nulas de los diferentes genes comprendidos en el modelo ABC da lugar a los mutantes homeóticos dibujados en la figura. Tomado de: <http://www.jic.bbsrc.ac.uk/staff/>

A partir de la postulación de este modelo se han descubierto muchos otros genes que interactúan con los genes ABC. Recientemente se ha propuesto un modelo dinámico de red de interacciones genéticas que integra la acción concertada de los genes descritos hasta ahora y que sustenta una explicación mecanístico-dinámica al modelo ABC (Espinosa, Padilla y Álvarez-Buylla 2004, ver Figura 4). Cabe destacar que en éste modelo de red, los genes tipo B representan nodos altamente conectados, lo cual podría justificar su importancia en el proceso de desarrollo floral y su historia evolutiva.

Cabe destacar que las dos especies modelo, tanto *A. thaliana* como *A. majus*, pertenecen al grupo de las eudicotiledóneas, que es el grupo de angiospermas más numeroso y que por si mismo representa una radiación al interior de éstas. Aquí se encuentran cerca del 75% de todas las especies de plantas con flores (Drinnan et al. 1994). En particular *A. thaliana* y *A. majus* están incluidas en el subgrupo de las eudicotiledoneas superiores, que a su vez es el más abundante dentro de las eudicotiledoneas. Las eudicotiledoneas superiores se caracterizan por tener flores cíclicas y la presencia de órganos internos y externos del perianto -pétalos y sépalos o cáliz y corola- bien diferenciados (Magallón et al., 1999; Judd y Olmstead, 2004). Hacia la base de las angiospermas los órganos florales estériles que rodean a los estambres y carpelos están menos diferenciados, siendo que en muchos casos no es posible hacer una distinción entre sépalos y pétalos y entonces se les denomina tépalos.

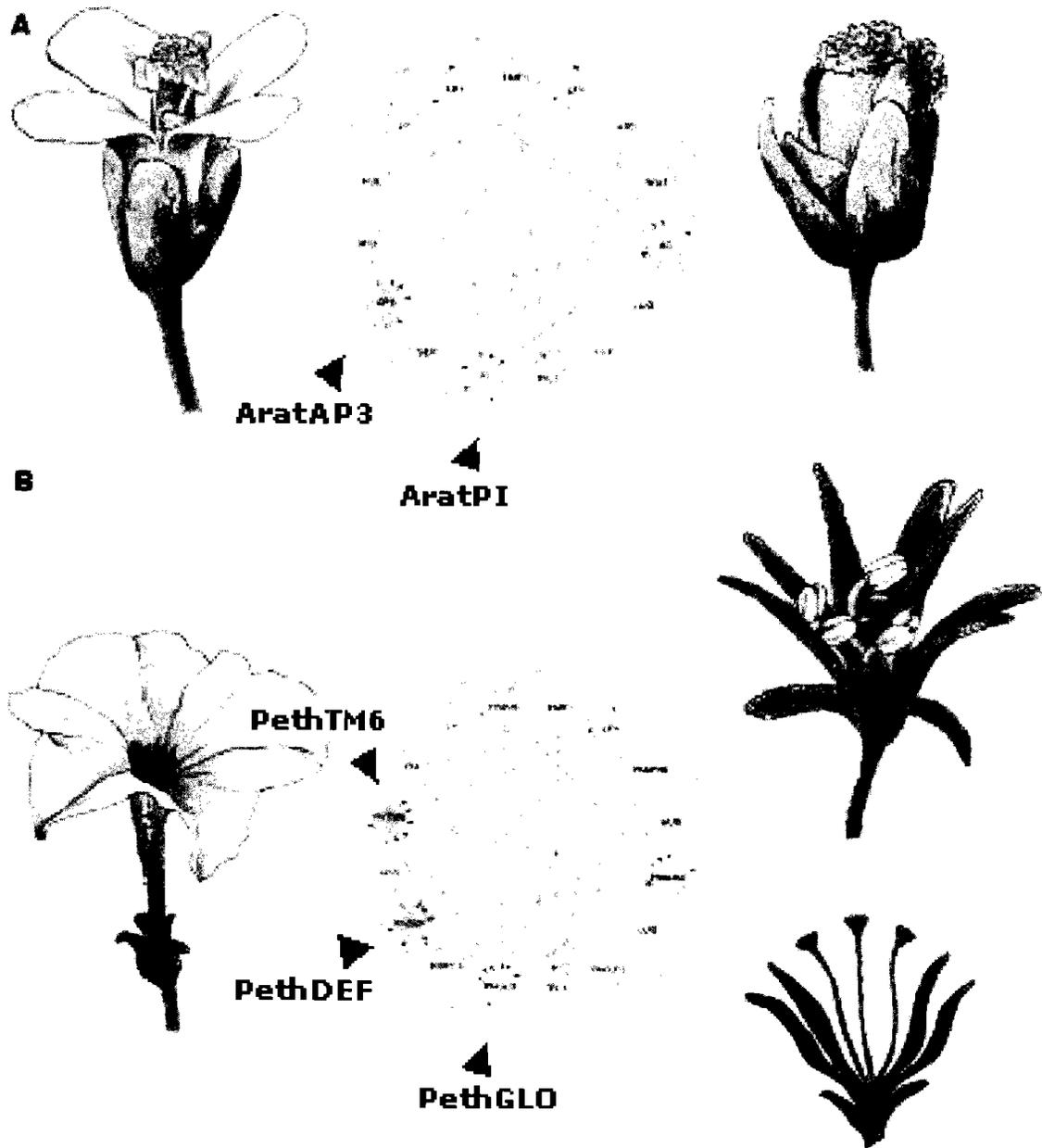


Figura 6. Modelos de redes para el desarrollo floral de *Arabidopsis thaliana* (A) y *Petunia hybrida* (B) con dibujos de la flor tipo silvestre para cada especie (extrema izquierda) y del fenotipo floral resultante para mutantes nulas *ap3* (extrema derecha). Se indican los genes tipo B en cada red y en amarillo se muestran los nodos “apagados” que corresponden a las mutaciones nulas al hacer la simulación de las mutantes. Tomado y modificado de Espinosa-Soto, Padilla-Longoria y Álvarez-Buylla, 2004.

Se ha evaluado la conservación de la función de los genes tipo B con estudios funcionales de homólogos de *AP3* y *PI* en diferentes especies de eudicotiledóneas superiores, y en general puede decirse que sus patrones de expresión son similares a los patrones encontrados en *Arabidopsis* (Theissen y Saedler, 1995; Irish y Kramer 1998; citados en Kramer, et al. 1998). Además, los fenotipos de los mutantes en las especies analizadas son generalmente consistentes con la conservación de la función de promoción del establecimiento de la identidad de pétalos y

estambres que tienen *AP3* y *PI* en *A. thaliana* y *A. majus* (Sommer, et al. 1990; Trobner et al. 1992; Angenent et al. 1993; Van der Krol y Chua 1993; citados en Kramer, et al. 1998). Como ha sido ampliamente aceptado, los pétalos en las eudicotiledóneas superiores son homólogos, y por tanto no es extraño pensar que las eudicotiledóneas superiores estudiadas hasta ahora tengan un mecanismo genético de desarrollo de pétalos conservado (Kramer, et al. 1998). Algunos estudios han ido más allá de las especies modelo incluidas en este grupo hacia clados de eudicotiledóneas basales como las Ranunculales (Kramer, et al. 1998; Kramer, E.M. y V. Irish. 1999) e incluso hacia angiospermas basales representantes de las Magnoliales y a las monocotiledóneas (Kramer e Irish, 2000); y se ha visto que en estos clados -donde los pétalos aún no están bien definidos- hay divergencia en los patrones de expresión de los genes B. Aún no se han hecho estudios detallados en angiospermas basales como Nymphales y Austrobaileyales, o incluso en gimnospermas (a excepción de *Picea abies*). Los estudios de la función B en estos grupos ayudarían a afinar las hipótesis sobre evolución del perianto en las plantas con flores y encontrar homologías con los órganos reproductivos de las gimnospermas. Por medio de análisis funcionales se ha mostrado que existe una relativa conservación de la función B en las especies modelo de monocotiledóneas maíz y arroz (Whipple, et al. 2004; Ambrose, et al. 2000), sin embargo aún no son claras las hipótesis de homología que deben de hacerse entre los órganos del perianto de monocotiledóneas y eudicotiledóneas.

Los estudios evolutivos de genes MADS tipo B en angiospermas basales han mostrado que la evolución de éstos es sumamente dinámica, ya que durante su historia han ocurrido diversas duplicaciones en los distintos niveles filogenéticos de los grupos taxonómicos basales, antes y después de la separación de los linajes PI y AP3 (Stellari, et al. 2004). Los estudios de la evolución molecular y funcional de los genes MADS-box tipo B en plantas, que son fundamentales en el desarrollo de los pétalos y estambres, pueden aportar elementos importantes para plantear hipótesis más claras acerca del origen y evolución de estos órganos, de importancia crucial durante la historia evolutiva de las angiospermas.

III. Características bioquímicas, función y evolución de Genes MADS-box tipo B.

Los genes que confieren la función B pertenecen a la superfamilia de genes de eucariontes conocida como MADS-box, que codifica para factores de transcripción que funcionan como proteínas diméricas o multiméricas de unión a DNA. Aunque los MADS-box se encuentran también en animales y en hongos, sólo en plantas han adquirido la función de determinantes de la diferenciación celular (Theissen y Saedler, 1995).

En plantas, los genes MADS-box codifican para proteínas con una estructura conocida como MIKC (Figura 7), con dos dominios distintivos: en el extremo amino terminal el dominio MADS, de aproximadamente 180 pares de bases (pb) o 60 amino ácidos (aa) cuya función presumiblemente es la unión al DNA y posiblemente tiene un papel en la dimerización con otras proteínas MADS (Shore y Sharrocks, 1995), y el dominio K (por 'keratin-like') de aproximadamente 237-259 pb (79-86 aas) que solo está presente en plantas, y probablemente tenga un papel importante en la dimerización (Ma et al., 1991). Según las estructuras de rayos-X obtenidas de los dominios MADS de las proteínas de mamíferos SRF y MEF2, y MCM1 de levadura, los ~60 aminoácidos en el dominio MADS poseen funciones tanto de dimerización como

de unión a DNA. El extremo amino terminal del dominio MADS hace contacto con el DNA, mientras que el extremo carboxilo terminal contiene dos estructuras β plegadas antiparalelas que funcionan como una interfase de dimerización (Pellegrini et al, 1995; Tan y Richmond, 1998; Santelli y Richmond, 2000; Hassler y Richmond, 2001 citados en Yang et al, 2003). Los ~80 aminoácidos del dominio K contienen muchas (abcdefg)_n hepta-repeticiones en las cuales las posiciones a y d están ocupadas por aminoácidos hidrofóbicos, lo cual sugiere que el dominio K forma una serie de α -hélices anfipáticas (Ma et al., 1991; Riechmann y Meyerowitz, 1997). En el caso de los genes tipo *AP3* y *PI*, el dominio K juega un papel determinante en la heterodimerización entre ellas (Yang et al, 2003).

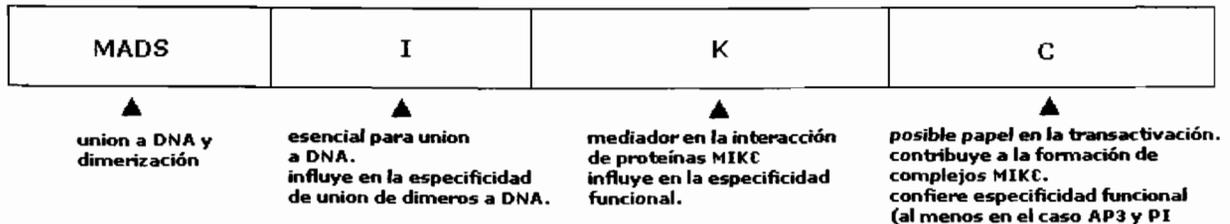


Figura 7. Esquematación de la estructura de dominios de los genes MIKC. Se indica en resumen la actividad que estos genes pueden tener con base en evidencia experimental de diferentes fuentes. La región más conservada corresponde al dominio MADS, seguida por el dominio I y el K. La menor conservación se encuentra en el dominio C del extremo C-terminal.

Entre el dominio MADS y el dominio K se encuentra una región intermedia (el dominio I) de longitud variable. Los resultados de estudios de intercambio de dominios sugieren que el dominio I es necesario para la dimerización y la especificidad funcional de las proteínas MADS *APETALA3* (*AP3*), *PISTILLATA* (*PI*), *AGAMOUS* (*AG*) y *APETALA1* (*AP1*) (Krizek y Meyerowitz, 1996; Riechmann y Meyerowitz, 1997; Riechmann et al., 1996). Por último encontramos en las proteínas MADS el dominio C-terminal, que es sumamente variable en cuanto a su longitud y secuencia entre las diferentes proteínas MADS de plantas. En algunas ocasiones, pero no en todos los casos, el dominio C incluye un dominio de activación transcripcional (Cho et al., 1999; Honma y Goto, 2001; Moon et al., 1999 citados en Yang, et al. 2003), y en algunas proteínas MADS, se ha postulado que el dominio C media la formación de multímeros de proteínas MADS (Egea-Cortines et al., 1999; Honma y Goto, 2001).

Las proteínas MADS en *Arabidopsis* tienen especificidades similares de unión a DNA (Riechmann et al., 1996), siendo que la especificidad funcional de las proteínas MADS de plantas *in vivo* no depende de la especificidad de unión a DNA, lo cual se ha demostrado con experimentos de intercambio de dominios entre *AP3*, *PI*, *AG*, *AP1*, *SRF* y *MEF2* (Krizek y Meyerowitz, 1996; Riechmann y Meyerowitz, 1997). La independencia de la especificidad funcional y la especificidad de unión a DNA sugiere la importancia potencial de la formación de multímeros entre estas proteínas. En *Arabidopsis* las proteínas MADS de determinación de identidad de órganos florales como *AG* y *AP1* son típicas en cuanto a que se pueden unir al DNA ya sea como homo o heterodímeros (Huang et al., 1993; 1996; Mizukami et al., 1996 citados en Yang, 2003). Sin embargo, *AP3* y *PI* no pueden unirse al DNA como homodímeros, sino solamente como heterodímeros *AP3/PI* (Davies et al., 1996; Riechmann et al., 1996b). Se ha mostrado que la expresión de ambas proteínas es necesaria para conferir un fenotipo ectópico en *Arabidopsis* (Krizek y Meyerowitz, 1996 citado en Lamb e Irish, 2003).

Siguiendo las filogenias más recientes reportadas para los genes MADS en *Arabidopsis* (Martínez-Castilla y Álvarez-Buylla, 2003; Nam, et al., 2003), podemos encontrar que los genes *AP3* y *PI* son parálogos inmediatos, evolutivamente muy cercanos al interior de su genoma. De ahí puede inferirse que la duplicación génica y la heterodimerización tienen una aparición relativamente reciente y que ha evolucionado a partir de la homodimerización (Winter, et al. 2002). Estudios de dimerización y unión a DNA *in vitro* con proteínas B ancestrales en la gimnosperma *Gnetum gnemon* mostraron que su proteína tipo B GGM2 es capaz de unirse al DNA como homodímero (Winter, et al. 2002). Para el caso de los ortólogos tipo B en la monocotiledónea *Lilium regale* llamados LRGLOA y LRGLOB estudios similares mostraron que estas proteínas son capaces de unirse al DNA tanto como homodímeros como heterodímeros (Winter, et al. 2002). Se piensa que la heterodimerización funcional obligada en eudicotiledóneas superiores (*Arabidopsis*) evolucionó a partir de la homodimerización presente en gimnospermas. Es posible que cambios estratégicos en los dominios de unión a proteínas hayan jugado un papel importante durante los procesos de diversificación funcional de las proteínas tipo B, otorgando las propiedades de heterodimerización en las eudicotiledóneas superiores. Probablemente el origen de la heterodimerización obligada en las eudicotiledóneas resultó importante durante el origen de los pétalos bien diferenciados en este grupo.

Para todos los genes B de angiospermas estudiados al momento, las filogenias muestran dos linajes parálogos, llamados tipo AP3 y tipo PI por sus respectivas proteínas en *Arabidopsis*. Las proteínas tipo PI y AP3 presentan regiones conservadas muy características en el extremo carboxilo terminal (ver Figura 8) (Kramer, et al. 1998). Los genes tipo PI tienen una región característica denominada 'motivo-PI'; y en los parálogos tipo AP3 existe un motivo equivalente alineable a éste (al motivo-PI) pero que ha divergido y está menos conservado que en los PI. Esta región se ha llamado 'motivo PI-derivado'. En los parálogos tipo AP3 de angiospermas podemos encontrar además una región extra en el extremo C-terminal, que hace que sean un poco más largos que los tipo PI. Con base en esta región extra se puede dividir a los genes tipo AP3 en dos grupos: los que tienen esta región terminal del tipo llamado euAP3 (con el 'motivo euAP3') y los que tienen el tipo llamado paleoAP3 (por el 'motivo paleoAP3') (ver Figura 8) (Kramer, et al. 1998). Las angiospermas no-eudicotiledóneas y eudicotiledóneas basales poseen genes AP3 con el motivo paleoAP3, mientras que los AP3 de las eudicotiledóneas superiores tienen el motivo euAP3 en su carboxilo terminal. En el caso de las eudicotiledóneas superiores también podemos encontrar parálogos tipo paleoAP3 particulares del grupo, que en los análisis filogenéticos se agrupan en un clado hermano de los tipo euAP3 y que se han llamado tipo TM6.

En gimnospermas se han identificado varios genes tipo B, aunque ha sido difícil su caracterización ya sea como parálogos tipo PI, parálogos tipo AP3 o hermanos de los tipo B de angiospermas en general, pues su alineamiento con respecto a los ortólogos de angiospermas es muy ambiguo. Los análisis filogenéticos y funcionales de genes MADS-box tipo B en *Picea abies*, *Gnetum gnemon* y *Cryptomeria japonica* (coníferas) sugieren que a pesar de las diferencias morfológicas y la distancia evolutiva –refiriéndonos particularmente a su morfología- entre coníferas y angiospermas, ambos grupos de plantas comparten características importantes en los mecanismos moleculares de regulación del desarrollo de sus órganos reproductivos (Sundström et al. 1999; Fukui, et al. 2001; Becker, et al. 2000). Por medio de estudios experimentales con el sistema de doble híbrido se ha mostrado que los productos de algunos genes MADS tipo B de gimnospermas son capaces de formar homodímeros y los genes tipo B DAL11 y DAL13 de *Picea abies* pueden formar heterodímeros entre ellos y con AP3, pero no con PI (Sundström y Engström, 2002). Por estas razones y siguiendo los análisis de los patrones de expresión de los genes B en

gimnospermas se ha argumentado que la función B en este grupo ha tenido una divergencia funcional y patrones evolutivos diferentes a los patrones encontrados en angiospermas; lo cual puede verse reflejado en una agrupación de estos genes como un clado hermano de los tipo B de angiospermas en algunos análisis filogenéticos reportados (Sundström y Engström, 2002).

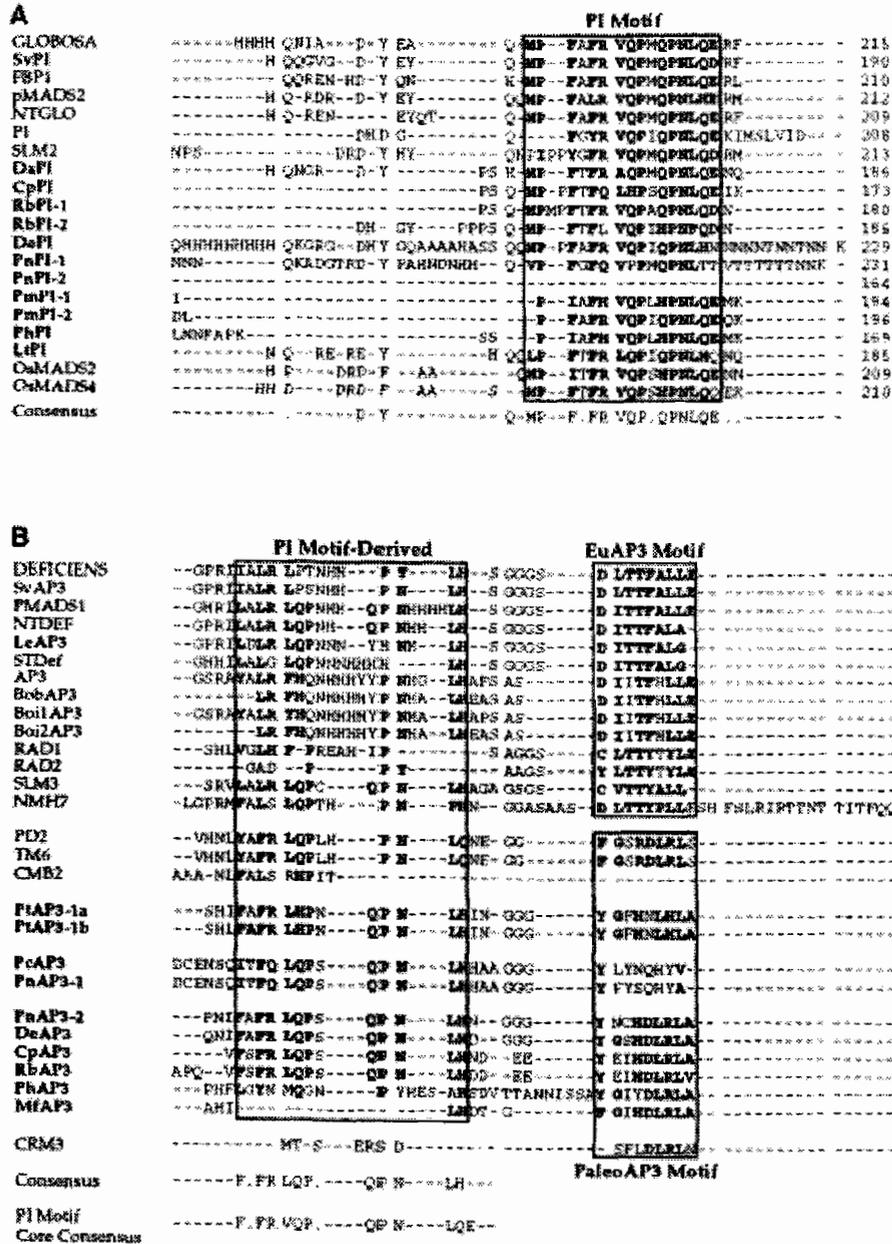


Figura 8. Alineamientos del C-terminal para algunos parálogos de genes tipo AP3 y PI. Se muestran encerrados en recuadros los motivos característicos de los genes tipo PI (motivo PI) y su derivado en los tipo AP3 (motivo PI-derivado) y los característicos de los dos diferentes tipos de AP3, motivo euAP3 y motivo paleoAP3. Tomado de Kramer, et al. 1998.

IV. Justificación del presente estudio. Hipótesis y objetivos.

En este trabajo llevamos a cabo un análisis detallado de la evolución molecular de secuencias de genes MADS-Box tipo B disponibles en bases de datos públicas, tanto de angiospermas como de gimnospermas, con la finalidad de entender mejor sus patrones evolutivos y de detallar los eventos de diferenciación funcional relacionados con las duplicaciones más importantes que han ocurrido en estos genes. La diversificación funcional de los genes B puede estar dada por cambios en sitios estratégicos de la proteína en los dominios de unión a DNA o de unión a proteínas, que provoquen cambios en sus especificidades de unión; en especial después de duplicaciones y en linajes particulares. Las posibilidades de unirse a diferentes secuencias de DNA o a formar heterodímeros con otras proteínas MADS pueden representar cambios en los mecanismos de desarrollo que eventualmente den lugar a la aparición de novedades morfológicas. En particular la evolución de la función B en el contexto del programa de desarrollo floral pudo tener un impacto importante en la evolución del perianto de las angiospermas.

Para entender los procesos de diversificación funcional, probaremos estadísticamente la hipótesis de diferencias significativas en las tasas de sustitución de los genes tipo B en etapas tempranas después de duplicaciones importantes y posteriormente durante sus eventos de diversificación. Para determinar las hipótesis se seleccionaran linajes particulares dada su paralogía (linajes de genes) y dada su ortología (linajes de taxa que pertenecen al mismo grupo taxonómico). Además se probaron otra serie de hipótesis para determinar si las diferencias en las tasas de sustitución están dadas por selección adaptativa y para detectar cuáles son los sitios específicos de la proteína que estuvieron sujetos a presiones selectivas y fueron fijados por selección positiva.

En una primer etapa del trabajo y como primer objetivo realizaremos un análisis filogenético cuidadoso para poder contar con una hipótesis de la evolución de estos genes, que por sí misma contribuya al conocimiento de sus historia evolutiva y que sirva de base para los posteriores análisis de tasas de sustitución, que conforman la segunda etapa del trabajo, así como para llevar a cabo un análisis de estimación para las fechas de duplicaciones más representativas en el grupo, que es la tercer etapa del trabajo. Las fechas estimadas se cotejarán con la información obtenida a partir del registro fósil.

El análisis de tasas de sustitución consistirá en la realización de pruebas de hipótesis estadísticas en un contexto de máxima verosimilitud, utilizando las filogenias obtenidas en el paso anterior y modelos explícitos de sustitución por codones que permiten a los linajes seleccionados tener diferentes tasas de sustitución y otros que permiten una clase de sitios bajo selección positiva en esos linajes. Los sitios o codones pertenecientes a esta clase se determinarán con métodos de estadística bayesiana. El análisis para la estimación de fechas de divergencia se llevará a cabo también con métodos basados en máxima verosimilitud que permiten introducir diferencias en las tasas de sustitución en diferentes ramas del árbol, con la finalidad de fechar los principales eventos de cladogénesis y de hacer una correlación de las fechas estimadas con aquellas obtenidas a partir del registro fósil.

Los métodos estadísticos de máxima verosimilitud han sido útiles anteriormente para detectar eventos de selección positiva durante la evolución de los genes homeóticos *Hox* en animales (Ali Fares, et al. 2003), y para los genes MADS de *Arabidopsis* los métodos para detectar

SP en linajes y sitios particulares fueron útiles para detectar algunos sitios en el dominio K de los genes del linaje de tiempo de floración (Martínez-Castilla y Álvarez-Buylla, 2003). Otros esfuerzos han tratado de correlacionar SP detectada en linajes particulares de los genes regulatorios *LEGCYC* con la evolución de la morfología floral en Leguminosas (Ree, et al. 2004).

Para los genes B, nosotros utilizaremos tres diferentes conjuntos de secuencias y sus alineamientos respectivos. Un alineamiento incluirá genes tipo B en general, otro exclusivo para genes tipo PI y otro para genes tipo AP3; para evaluar diferentes niveles de detalle en cuanto a su historia de duplicaciones. La filogenia obtenida para los genes tipo B servirá para plantear hipótesis más generales en cuanto al origen y evolución de la función B. En particular con esta filogenia nos concentraremos en: (1) La separación de los genes tipo B de gimnospermas del clado de los genes tipo AGL34 –posteriormente conocidos como “B-sister”-, (2) la duplicación que dio lugar a los linajes de genes tipo B de gimnospermas y angiospermas y que coincide con la aparición de la flor, (3) el origen de la duplicación AP3-PI y el posible papel de la selección positiva en el origen de la heterodimerización obligada a partir de la homodimerización y (4) en el análisis de las diferencias en las tasas evolutivas como evidencia de diversificación funcional en los genes tipo AP3 y los PI.

La filogenia obtenida para los genes tipo AP3 nos permitirá analizar detalladamente: (1) el origen de los linajes paleoAP3 y euAP3, que caracterizan a las angiospermas no-eudicotiledóneas (incluyendo angiospermas basales y eudicotiledóneas basales) y a las eudicotiledóneas superiores respectivamente (2) la duplicación que generó los linajes euAP3 y TM6 en el clado de las eudicotiledóneas superiores, para evaluar hipótesis acerca de su importancia en la evolución de los pétalos establecidos en una corola propiamente definida que es característica del grupo.

Las diferencias en las tasas de sustitución y la búsqueda de selección positiva se llevará a cabo también en los genes tipo PI para cotejar los resultados con los obtenidos para los AP3.

Secuencias y Métodos

I. Obtención de las secuencias de ortólogos MADS-Box tipo B de plantas.

Las secuencias de ortólogos de los genes MADS-Box tipo B en las diferentes familias de angiospermas y gimnospermas se obtuvieron de la base de datos pública GenBank (<http://www.ncbi.nlm.nih.gov/Genbank/index.html>). Primero se hizo una búsqueda usando palabras clave con los nombres y abreviaciones de los genes MADS-Box tipo B en *Arabidopsis thaliana* (*APETALA3* y *PISTILLATA*), y *Antirrhinum majus* (*GLOBOSA* y *DEFICIENS*) junto con los nombres de las diferentes familias de angiospermas reportadas por el APGII, 2003. Posteriormente se hizo una búsqueda con BLASTN y BLASTP con las secuencias mencionadas (*AP3*, *PI*, *DEF* y *GLO*) seleccionando secuencias bajo criterios de valor E relajados. Muchas secuencias fueron descartadas *a posteriori* de los análisis filogenéticos por su ubicación incongruente con respecto a los ortólogos MADS tipo B y las secuencias seleccionadas como grupos externos. De las secuencias obtenidas se hizo una selección, buscando que el muestreo taxonómico fuera homogéneo a todo lo largo de la filogenia de las angiospermas, es decir, buscando que no hubiera familias sobre representadas o poco representadas, siguiendo la filogenia más reciente consultada al momento basada en datos moleculares (APGII, 2003). Las secuencias utilizadas finalmente en este estudio, con su número de acceso y cita bibliográfica se presentan en la Tabla 1, y en el Apéndice I se hace un listado de los nombres en orden alfabético para facilitar su identificación. Se consideró la secuencia completa de los genes, excepto en el alineamiento de los MADS Tipo B, eliminando algunos aminoácidos iniciales en la caja MADS, ya que en más de la mitad de las secuencias ésta no estaba reportada.

Tabla 1. Secuencias utilizadas en el análisis filogenético. Las secuencias de los MADS tipo PI se muestran en negritas

Subclase	Familia	Especie	Gen	No. Acceso	Cita
Cucurbitales	Cucurbitaceae	<i>Cucumis sativus</i>	CucsCUM26	AF043255	<i>Kater, M.M., et al. Unpublished</i>
Rosales	Rosaceae	<i>Rosa rugosa</i>	RrMASAKOBP RrMASAKOeuB3	AB038462 AB099875	<i>Kihara, K. et al. 2002.</i>
		<i>Malus domestica</i>	MaldPI MaldTM6 MaldMADS13	AB081092	<i>Matsumoto, S., Ohtsubo, T. and Soejima, J. Unpublished</i>
Fabales	Fabaceae	<i>Medicago sativa</i>	MedsNMH7 MedsNGL9	AF042068 AF335473	<i>Heard, J. and Dunn, K. 1995</i> <i>Zucchero, J.C., Caspi, M. and Dunn, K. 2001</i>
Fagales	Betulaceae	<i>Betula pendula</i>	BetpMADS2	BPE488589	<i>Jarvinen, P. et al., 2003</i>
Brassicales	Brassicaceae	<i>Arabidopsis thaliana</i>	AratAP3 AratPI	M86357 D30807	<i>Jack, T., et al. 1992</i> <i>Goto, K. and Meyerowitz, E.M. 1994</i>
		<i>Arabidopsis lyrata</i>	AralPI	AF143382	<i>Lawton-Rauh, A.L., Buckler, E.S. IV and Purugganan, M.D. 1999</i>
		<i>Brassica napus</i>	BranAP3	AF124814	<i>Pylatuik, J.D., Davis, A.R. and Bonham-Smith, P.C. 1999</i>
Myrtales	Myrtaceae	<i>Eucalyptus grandis</i>	EucgEGM2	AF029976	<i>Southerton et al., 1998</i>
Lamiales	Veronicaceae	<i>Antirrhinum majus</i>	AntmDEF	X52023	<i>Sommer, H., et al. 1990</i>

			AntmGLO	X68831	<i>Trobner, W., 1992</i>
		<i>Syringa vulgaris</i>	SyrvPI SyrvAP3	AF052861 AF052869	<i>Kramer, E.M., Dorit, R.L. and Irish, V.F. 1998</i>
Solanales	Solanaceae	<i>Lycopersicon esculentum</i>	LyceTDR6 LyceAP3	X60759 AF052868	<i>Pnueli, L. et al., 1991</i>
		<i>Petunia hybrida</i>	PethGP PethMADS2 PethTM6	X69946 X69947 AF230704	<i>Kramer, E.M., 1998</i> <i>Kush, A. et al., 1993</i> <i>Kramer, E.M. and Irish, V.F. 2000</i>
		<i>Nicotiana tabacum</i>	NtGLO	NTGLOBOSA	<i>Hansen, G. et al., 1993</i>
Asterales	Asteraceae	<i>Gerbera hybrida</i>	GerhDEF1 GerhDEF2 GerhGLO1	AJ009724 AJ009725 AJ009726	<i>Yu, D. et al., 1999</i> " "
		<i>Hieracium piloselloides</i>	HiepDEF1	AF180364	<i>Guerin, J. et al., Unpublished.</i>
		<i>Chrysanthemum moriflorum</i>	ChrmCDM19 ChrmCDM115 ChrmCDM86	AY173064 AY173060 AY173061	<i>Shchennikova, A. et al., Unpublished</i> " "
Apiales	Apiaceae	<i>Daucus carota</i>	DaucMADS2 DaucMADS3	AJ271148 AJ271149	<i>Linke, B., Nothnagel, T. and Borner, T. 2003</i>
Cornales	Hydrangeaceae	<i>Hydrangea macrophylla</i>	HydmAP3 HydmPI HydmTM6	AF230702 AF230711 AF230703	<i>Kramer, E.M. and Irish, V.F. 2000</i> " "
Caryophyllales	Polygonaceae	<i>Rumex acetosa</i>	RumaRaD1 RumaRaD2	U28483 U28482	<i>Ainsworth, C. et al., Unpublished.</i> "
	Caryophyllaceae	<i>Silene latifolia</i>	SillSLM2 SillSLM3	X80489 X80490	<i>Hardenack, S. et al., 1994</i> "
	Buxaceae	<i>Pachysandra terminalis</i>	PactAP3	AF052870	<i>Kramer, E.M., Dorit, R.L. and Irish, V.F. 1998</i>
Ranunculales	Papaveraceae	<i>Papaver nudicaule</i>	PapnAP3-2 PapnPI-1	AF052873 AF052856	<i>Kramer, E.M., Dorit, R.L. and Irish, V.F. 1998</i> "
		<i>Dicentra eximia</i>	DiceAP3	AF052875	"
	Ranunculaceae	<i>Ranunculus ficaria</i>	RanfAP3 RanfPI	AF052858 AF052854	<i>Kramer, E.M. and Irish, V.F. 1999</i> "
		<i>Trautvetteria carolinensis</i>	TracAP3 TracPI-1 TracPI-2	AY162904 AY162905 AY162906	" " "
		<i>Berberis gilgiana</i>	BergPI-1 BergPI-2	AY162860 AY162861	" "
		<i>Akebia quintata</i>	AkeqAP3-1 AkeqAP3-2 AkeqPI-1	AY162835 AY162839 AY162837	" " "
		<i>Sanguinaria canadensis</i>	SancAP3	AF130868	"
nd	Sabiaceae	<i>Meliosma dilleniiifolia</i>	MeldAP3-1 MeldAP3-2 MeldAP3-3 MeldPI	AY436709 AY436710 AY436711 AY436712	<i>Stellari G.M., et al. 2004</i> " " "
Laurales	Calycanthaceae	<i>Calycanthus floridus</i>	CalfAP3 CalfPI	AF230699 AF230700	<i>Kramer, E.M. and Irish, V.F. 2000</i> "
Magnoliales	Magnoliaceae	<i>Liriodendron tulipifera</i>	LirtAP3 LirtPI	AF052878 AF052864	<i>Kramer, E.M., Dorit, R.L. and Irish, V.F. 1998</i>
		<i>Magnolia praecocissima</i>	MagpMADS8 MagpMADS7	AB050650 AB050649	<i>Ito, M. et al., Unpublished.</i> "
		<i>Michelia figo</i>			

			MicfAP3	AF052877	"
			MicfPI	AF052863	"
Chloranthales	Chloranthaceae	<i>Chloranthus spicatus</i>	ChlsAP3	AF230710	Kramer,E.M. and Irish,V.F. 2000
			ChlsPI	AF230701	"
Piperales	Piperaceae	<i>Piper magnificum</i>	PipmPI-1	AF052867	Kramer,E.M., Dorit,R.L. and Irish,V.F. 1998.
			PipmPI-2	AF052866	"
	Aristolochiaceae	<i>Peperomia hirta</i>	PepPhPI	AF052865	"
			PepPhAP3	AF052879	"
		<i>Asarum europaeum</i>	AsaePI	AF230707	"
			AsaeAP3	AF230698	"
Canellales	Winteraceae	<i>Drimys winterii</i>	DriwAP3-1	AY436722	Stellari G.M., et al. 2004
			DriwAP3-2	AY436723	"
			DriwAP3-3	AY436724	"
			DriwAP3-4	AY436725	"
			DriwPI-1	AY436727	"
			DriwPI-2	AY436728	"
			DriwB_sister	AY436726	"
LILIOPSIDA					
Poales	Poaceae	<i>Zea mays</i>	ZeamSilky1	AF181479	Ambrose, et al., 2000
			ZeamZPI1bm29	AJ292961	Munster,T. et al., 2001
			Zeam m16	ZMA292959	"
			Zeam m18	ZMA292960	"
		<i>Oryza sativa</i>	OrysMADS16	AF077760	Becker,A. et al., 2002
			OsMADS2	L37526	Moon,Y. et al., 1999
			OsMADS4	L37527	"
			OrysAB0033	AB003323	Chung,Y. Et al., 1995
		<i>Triticum aestivum</i>	TriamADS51	AB007506	Shinozuka,Y. et al., 2002
Asparagales	Agapanthaceae	<i>Agapanthus praecox</i>	AgapMADS1	AB079259	Nakamura,T. and Nakano,M. 2003
	Orchidaceae	<i>Oncidium cv. 'Gower Ramsey'</i>	OrciMADS3	AY196350	Hsu,H.F. and Yang,C.H. 2002
		<i>Orchis italica</i>	OrciPI	AB094985	Park,J. et al., 2002
	Hyacinthaceae	<i>Hyacinthus orientalis</i>	HyaoHPI 2	AF134115	Quanz,L., Xingguo,L. and Xiansheng,Z. Unpublished.
		<i>Asparagus officinalis</i>	HyaoHPI 1	AF134114	Montieri,S., Gaudio,L. and Aceto,S. Unpublished.
	Asparagaceae		AspoDEF	AB094964	"
Alismatales	Alismataceae	<i>Sagittaria montevidensis</i>	SmAP3	AF230705	Kramer,E.M. and Irish,V.F. 2000
Liliales	Liliaceae	<i>Lilium regale</i>	LilrDEF	AB071378	Winter,K.U. et al., 2002
			LilrGLOA	AB071379	Tzeng,T.Y. and Yang,C.H. 2001
		<i>Lilium longiflorum</i>	LilIMADS1		
Dioscoreales	Dioscoreaceae	<i>Tacca chantieri</i>	TaccPI	AF230713	Kramer,E.M. and Irish,V.F. 2000
			TaccAP3	AF230706	"
Austrobaileales	Illiciaceae	<i>Illicium henryii</i>	IlhAP3-1	AY436729	Stellari G.M., et al. 2004
			IlhAP3-2	AY436731	"
			IlhAP3-3	AY436733	"
			IlhPI-1	AY436734	"
			IlhPI-2	AY436735	"
nd	Nymphaeaceae	<i>Nymphaea sp.</i>	NympAP3	AY436748	Stellari G.M., et al. 2004
			NympPI	AY436744	"
GIMNOSPERMAS					
Coniferales	Pinaceae	<i>Pinus radiata</i>	PrDGL	AF120097	Sundstrom,J. et al., 1999
		<i>Picea abies</i>	PicaDAL11	AF158539	Fukui,M. et al., 2001
			PicaDAL12	AF158541	Winter,K.U. et al., 1999
			PicaDAL13	AF158543	"
	Cupressaceae	<i>Cryptomeria japonica</i>	CryMADS2	AF097747	Becker,A. et al., 2000
			CryMADS1	AF097746	"
Gnetales	Gnetaceae	<i>Gnetum gneon</i>	GnegGGM12	AJ132219	Becker,A. et al., 2000
			GnegGGM13	AJ132216	"
			GnegGGM15	AJ132209	"
		<i>Physcomitrella patens</i>	PpMADS2	AB067689	Henschel,K. et al. 2002

II. Alineamiento

Las secuencias fueron alineadas de manera automática usando el programa CLUSTALW (Thompson, et al., 1994) y posteriormente el alineamiento se refinó manualmente con el programa BIOEDIT (Hall, 1999) siguiendo los alineamientos reportados en la literatura (Kramer, et al. 1998; 2000), por codones siguiendo siempre las secuencias en aminoácidos. Los análisis filogenéticos se llevaron a cabo usando tres matrices de alineamiento de nucleótidos por codones: una para los genes tipo PISTILLATA, otra para los tipo APETALA3 y otra para los MADS tipo-B. En todas las matrices se incorporaron las secuencias de gimnospermas para la evaluación de su posición en la filogenia resultante. Los análisis filogenéticos fueron hechos con base en nucleótidos, pero estos se alinearon observando la secuencia de aminoácidos.

III. Análisis filogenético

Para cada alineamiento se llevaron a cabo análisis filogenéticos de distancias con el método de Neighbour Joining y Máxima Parsimonia utilizando el programa de cómputo PAUP 4b10 (Swofford, 1998), pero con éstos métodos solo pudieron recuperarse filogenias parcialmente resueltas (los resultados no se muestran), es decir, con politomías. Posteriormente se hicieron análisis con métodos bayesianos utilizando el programa MrBayes 3.0 (Huelsenbeck y Ronquist, 1998). El modelo de sustitución fue seleccionado haciendo pruebas de tasas de verosimilitud (PTV) siguiendo comparaciones anidadas según el algoritmo implementado en el programa Modeltest (Posada y Crandall, 1998; Huelsenbeck y Crandall, 1997). Estas evaluaciones del mejor modelo de sustitución también sirvieron para hacer una interpretación preliminar de los patrones evolutivos de las secuencias.

Las búsquedas en MrBayes se iniciaron con un árbol aleatorio en cuatro diferentes cadenas de Markov. Se corrieron 2, 500 000 generaciones con muestreos de árboles cada 100 generaciones. Se descartaron las primeras 30, 000 generaciones (300 árboles) como “burnin” después de hacer una confirmación visual de la convergencia en la grafica de valores de verosimilitud contra el numero de generaciones. Con los árboles restantes se construyó un consenso por regla de mayoría. Los valores que corresponden al porcentaje de consenso equivalen a la probabilidad posterior para los clados recuperados (Huelsenbeck y Ronquist, 1998). Estas probabilidades se usaron para evaluar el apoyo o soporte de los clados.

En la matriz del alineamiento de los MADS tipo-B se utilizó como grupo externo definido *a posteriori* solo para la visualización del enraizamiento el gen *AGL15* y el gen *PpMADS1* de *Physcomitrella patens subsp. patens*. En los alineamientos de los genes tipo PI y tipo AP3 además se utilizó el parólogo correspondiente descrito en *Arabidopsis thaliana*.

IV. Detección de heterogeneidad de tasas de sustitución entre sitios, entre linajes y presencia de selección positiva

Un criterio estricto y robusto para detectar evolución adaptativa (o selección positiva, a veces también llamada darwiniana) a nivel molecular en un gen codificante es identificar tasas de sustitución no sinónimas (d_N , en las cuales el aminoácido cambia) más elevadas en relación con las tasas de sustitución sinónimas (d_S , mutaciones 'silenciosas' que no generan cambio a nivel de aminoácidos). A esta tasa se le representa como ω , siendo que $\omega = d_N/d_S$.

Las diferencias en las tasas de fijación de ambos tipos de sustitución facilitan una forma de medir las presiones de selección sobre una proteína. Bajo las diferentes hipótesis evolutivas, las tasa ω esperada se representa en el cuadro siguiente.

HIPOTESIS	TASA ω ESPERADA
Neutralidad	$\omega \cong 1$
Selección Positiva (diversificadora, adaptativa)	$\omega > 1$
Selección Purificadora	$\omega < 1$

Es relativamente sencillo detectar selección purificadora en un alineamiento de DNA o de aminoácidos, ya que se observan como secuencias, dominios o sitios altamente conservados; sin embargo es más complicado estadísticamente diferenciar los sitios que se encuentran bajo selección positiva y neutralidad de manera significativa.

Al momento se han sugerido dos clases de métodos para la estimación de d_N y d_S , (Yang y Bielawski, 2000). En la primera clase podemos agrupar los métodos tradicionales, que incluyen cerca de una docena de métodos intuitivos desarrollados desde los años 80's (Miyata y Yasunaga 1980; Nei y Gojobori, 1986; Li et al., 1985, 1993; Pamilo y Bianchi, 1993; Comeron, 1995; Ina, 1995; Yang y Nielsen, 2000 citados en Yang y Bielawski, 2000). Estos métodos son empíricos y solucionan el problema haciendo una comparación por pares entre dos secuencias bajo el supuesto de homología. Estos métodos siguen en general la siguiente metodología: se cuentan los sitios sinónimos (S) y no sinónimos (N) en las dos secuencias, posteriormente se cuentan las diferencias sinónimas y no sinónimas entre las dos secuencias y después se hace una corrección de la estimación para sustituciones múltiples en un mismo sitio (Yang y Bielawski, 2000). En estos métodos se prueba si $d_N - d_S$ es significativamente mayor que 0 para cada par de secuencias comparadas. El principal problema con ésta metodología es que se promedian las tasas de sustitución entre todos los sitios en cada secuencia, además de que se hace un promedio también sobre el intervalo de tiempo que las separa, pero es razonable esperar que en una proteína funcional la mayoría de los aminoácidos estén altamente conservados, y que la evolución adaptativa afecte solo unos cuantos sitios en espacios temporales breves (Gillespie 1991 citado en Yang, 2002), y al momento de hacer un promedio la señal de selección positiva o neutralidad se pierde, por lo que estos métodos son demasiado restrictivos y no son capaces de detectar selección positiva en la mayoría de los casos (Yang, 2002). Además, la mayoría de los métodos tradicionales hacen suposiciones simplistas acerca de los procesos de sustitución nucleotídica y por tanto son métodos aproximados (Yang y Bielawski, 2000).

Los métodos de Miyata y Yasunaga (Miyata y Yasunaga 1980, citado en Yang y Bielawski, 2000) y Nei y Gojobori (Nei y Gojobori 1986, citado en Yang y Bielawski, 2000), asumen una tasa igual para las transiciones ($T \leftrightarrow C$ y $A \leftrightarrow G$) y transversiones ($T, C \leftrightarrow A, G$), así como un uso uniforme de codones. Debido a que las transiciones en la tercera posición de los codones resultan en un cambio sinónimo con mayor probabilidad que los cambios dados por transversiones, el ignorar la tasa de transiciones/transversiones da lugar a una sub-estimación de S y sobre-estimación de N (Li et al., 1993 citado en Yang y Bielawski, 2000). Se han hecho esfuerzos recientes para incorporar estos parámetros en la estimación de d_N y d_S (Yang y Bielawski, 2000) en tales métodos. Existen dentro de este grupo de métodos tradicionales los métodos basados en la reconstrucción de codones ancestrales en cada nodo siguiendo el principio de la máxima parsimonia (por ejemplo Creevey y McInerney, 2002), sin embargo es claro que el principio de parsimonia no es el más adecuado en cuanto a principio de inferencia de historias evolutivas a nivel molecular, especialmente cuando se trata de secuencias alejadas evolutivamente (por ejemplo, más allá del nivel de género) (Felsenstein, 2004).

La segunda clase de métodos, los aplicados en éste trabajo de tesis, son los métodos de Máxima Verosimilitud (MV), basados en modelos explícitos de sustitución de codones. Los parámetros en el modelo (i.e. divergencia de las secuencias t , tasa de transiciones/transversiones κ y la tasa d_N/d_S o ω) son estimados a partir de los datos por MV, y se usan para calcular d_N y d_S de acuerdo a sus definiciones (Yang y Bielawski, 2000).

En los modelos de sustitución por codones usados en este trabajo (Goldman y Yang, 1994), la tasa instantánea de sustitución para pasar del codon i al codon j ($i \neq j$) esta dada por:

$$Q_{ij} = \begin{cases} 0, & \text{si los dos codones difieren en mas de una posición,} \\ \pi_j, & \text{para transversiones sinónimas,} \\ \kappa\pi_j, & \text{para transiciones sinónimas,} \\ \omega\pi_j, & \text{para transversiones no-sinónimas,} \\ \omega\kappa\pi_j, & \text{para transiciones no-sinónimas,} \end{cases}$$

donde κ es la tasa de transiciones/transversiones, ω es la tasa de sustituciones sinónimas/no-sinónimas, y π_j es la frecuencia en equilibrio del codon j , calculada usando las frecuencias de los nucleótidos en cada una de las tres posiciones (Yang, 2002). Como se observa, los modelos de codones son muy completos en cuanto a la descripción de los procesos de sustitución en las secuencias codificantes. Una de las principales características de este método (MV) es que se formula a nivel de tasas instantáneas (dónde no hay posibilidad de sustituciones múltiples) y la teoría de probabilidad lleva a cabo todas las tareas difíciles en un solo paso: se estiman los parámetros mutacionales (como κ), se corrigen los cálculos para sustituciones múltiples, y se da peso a los diferentes tipos de cambio entre codones (Yang y Bielawski, 2000). Como se verá más adelante, por medio de la aplicación de diferentes variantes de este modelo básico de sustitución de codones junto con métodos basados en estadística bayesiana, es posible detectar selección positiva a nivel molecular en sitios particulares de la proteína y en linajes particulares dada una filogenia, o de detectar diferencias en las tasas de sustitución posteriores a eventos de duplicación como indicio para la justificación de diversificación funcional en familias de proteínas (Bielawski y Yang, 2004).

IV. a. Detección de heterogeneidad de tasas de sustitución (ω) entre sitios y selección positiva en sitios específicos

Los métodos tradicionales de detección de selección positiva asumen que todos los sitios dentro de una proteína están sujetos a la misma presión de selección y tienen la misma tasa ω . Los análisis hacen un promedio de la tasa ω en todos los sitios y solo detectan selección positiva si el promedio es > 1 . Esta es una prueba muy conservadora, ya que dentro de una misma proteína, muchos sitios pueden estar bajo una fuerte selección purificadora debido a restricciones funcionales, con un ω cercana a 0, y el promedio impediría la detección de algunos pocos sitios que se encontrasen bajo presiones de selección diversificadora.

Los métodos basados en verosimilitud permiten considerar dichas diferencias de presión de selección entre sitios. De manera práctica, las pruebas de detección de selección positiva comienzan con el cálculo de la verosimilitud (logaritmo de la verosimilitud) de los datos (matriz de las secuencias de nucleótidos alineadas) dado el modelo de sustitución de codones diseñado para probar hipótesis particulares y la topología de la filogenia o hipótesis acerca de la historia evolutiva de las secuencias analizadas. Para evaluar diferencias en la tasa ω se cambian las restricciones de dicho parámetro dentro del modelo de codones, creando de esa manera diferentes modelos a comparar. Algunos de los modelos desarrollados por Ziheng Yang y colaboradores (Yang et al., 2000) para el posterior cálculo de la verosimilitud están resumidos en la Tabla 2. Todos ellos se encuentran definidos con su código en el programa PAML (<http://abacus.gene.ucl.ac.uk/software/paml.html>) y son los que utilizamos en éste trabajo. Todos los modelos incluyen como parámetros la longitud de las ramas en la filogenia, la tasa de transición/transversión κ , y las frecuencias de las bases en las tres posiciones del codon.

Tabla 2. Modelos de sustitución de codones para detectar heterogeneidad de tasas de sustitución ω y selección positiva entre sitios en un alineamiento de nucleótidos. p es el número de categorías para la distribución de ω definida *a priori* en el modelo.

MODELO (y nombre código en PAML)	p	Parámetros	Descripción
M0 (one-ratio)	1	ω	Asume una sola ω para todos los sitios.
M1 (neutral)	1	P_0	$p_1 = 1 - p_0$, $\omega_0 = 0$, $\omega_1 = 1$; El modelo neutral. Asume una proporción de sitios conservados p_0 con $\omega_0 = 0$ y una proporción de sitios neutrales con $p_1 = 1 - p_0$ con $\omega_1 = 1$.
M2 (selección)	3	p_0, p_1, ω_2	$p_2 = 1 - p_0 - p_1$, $\omega_0 = 0$, $\omega_1 = 1$; El modelo de selección. Adiciona una clase más de sitios con frecuencia $p_2 = 1 - p_0 - p_1$ y con ω_2 estimada a partir de los datos.
M3 (discreto)	$2K - 1$ ($K = 3$)	p_0, p_1, \dots, p_{K-2} $\omega_0, \omega_1, \dots, \omega_{K-1}$	$p_{K-1} = 1 - p_0 - p_1 - \dots - p_{K-2}$; El modelo discreto. Usa una distribución discreta no restringida para modelar tasas ω de manera heterogénea entre sitios. Los valores de ω_k se arreglan en orden ascendente para su identificación única. El modelo con K clases tiene $K-1$ parámetros de frecuencia y K parámetros ω_k .
M7 (beta)	2	p, q	Para $\beta(p, q)$; La distribución beta tiene una densidad $\beta(p, q)$: $f(\omega) = \omega^{p-1} (1-\omega)^{q-1} / \beta(p, q)$, $0 \leq \omega \leq 1$, donde $\beta(p, q)$ es la función beta. Esta distribución

			puede tomar diferentes formas (\lfloor , \lceil , $\lfloor \rfloor$, $\lceil \rceil$, etc) en el intervalo (0,1). Este modelo no permite sitios bajo selección positiva (con $\omega > 1$)
M8 (beta& ω)	4	p_0, p, q, ω	Adiciona una clase extra de sitios bajo selección positiva al modelo beta. Una proporción p_0 de sitios tiene ω dibujada de la distribución beta $B(p, q)$ y los sitios restantes (con proporción $p_1 = 1-p_0$) tienen la misma tasa ω_1

Como se observa, algunos de estos modelos son nulos, ya que solo permiten tasas ω iguales o menores que 1 (ver M0 o M1), y éstos funcionan como hipótesis nulas que se comparan con los modelos que permiten tasas ω mayores que 1 (o sitios bajo selección positiva), que representan las hipótesis alternativas. Los métodos estadísticos empleados para hacer dichas comparaciones se denominan **pruebas de tasas de verosimilitud (PTV)**. Estas pruebas requieren que los modelos o hipótesis a comparar estén anidados, es decir, que el modelo que describe la hipótesis alternativa incluya un mayor número de parámetros, y las PTV se utilizan para evaluar si la mejora en los valores de verosimilitud bajo el modelo alternativo es significativa.

El estadístico empleado (estadístico PTV) es el valor igual al doble de la diferencia del logaritmo de la verosimilitud entre los dos modelos ($2\Delta l = 2(l_2 - l_1)$), y aunque su distribución no es muy conocida, según los estadísticos teóricos ésta puede aproximarse con una distribución χ^2 , por lo que para evaluar la significancia se hace una prueba de bondad de ajuste, donde los grados de libertad son los parámetros extra del modelo para la hipótesis alternativa (Anisimova, Bielawsky y Yang, 2001). Es importante considerar que las distribuciones teóricas del estadístico $2\Delta l$ no son bien conocidas, lo que ha dificultado la completa aceptación de las PTVs en la detección de SP.

IV. b. Detección de heterogeneidad de tasas de sustitución (ω) entre linajes

Se ha postulado que la evolución de nuevas funciones en proteínas duplicadas se lleva a cabo por medio de el reclutamiento genético de uno de los genes duplicados hacia nuevas funciones mientras que el gen “original” mantiene la función inicial (Ohta, 1993). Una manera de estudiar este “reclutamiento” genético es por la identificación de los sitios o aminoácidos particulares que experimentaron cambios en las presiones selectivas enseguida de un evento de duplicación seguido de divergencia funcional. Recientemente, Bielawski y Yang (2004) desarrollaron modelos de codones diseñados para la detección de presiones de selección divergentes en linajes particulares en una filogenia (modelos C y D), sin la necesidad de el supuesto de divergencia por selección positiva. Este modelo permite que haya una variación de ω entre sitios, con una fracción de éstos evolucionando bajo regímenes divergentes de presiones selectivas entre linajes (Bielawski y Yang, 2004), en particular se sugiere la prueba para clados parálogos en una familia multigénica. En este trabajo aplicamos el modelo D descrito en el trabajo mencionado, que es una extensión del modelo discreto (M3) de Yang et al. 2000 por que permite que las presiones selectivas en una clase de sitios varíen en diferentes partes de la filogenia, es decir, que ésta clase de sitios tiene dos parámetros ω independientes, uno para cada clado de parálogos que surgió de un evento de duplicación (ω_{1A} y ω_{1B}). Las PTVs implementadas en éste

estudio se seleccionaron siguiendo las recomendaciones descritas en Bielawski y Yang, 2004 para la detección de heterogeneidad de tasas estadísticamente significativa con el modelo D.

IV. c. Detección de selección positiva en linajes particulares

Cuando se hacen las pruebas de detección de selección positiva en sitios particulares en un alineamiento de nucleótidos por codones se corre un riesgo razonable al hacer un promedio de la tasa ω para cada sitio, lo cual implica que las presiones selectivas en dicho sitio son homogéneas a todo lo largo de la filogenia, o a lo largo de la historia evolutiva de las secuencias estudiadas, y entonces perder la señal de selección positiva ocurriendo en momentos o linajes particulares. Con la finalidad de detectar selección positiva en linajes y sitios particulares Yang y Nielsen (2002) desarrollaron los modelos de codones A y B, que permiten una proporción de sitios con una ω permitida con valores >1 (selección positiva) en sitios particulares de un linaje definido. En estos modelos, como los anteriores descritos para las pruebas sitio por sitio, se asume que la filogenia es conocida o estimada de manera independiente (en este caso nosotros la estimamos con métodos bayesianos), pero las ramas que esperamos que estén bajo selección positiva se especifican *a priori*, igual que en los modelos para heterogeneidad de tasas. Las ramas seleccionadas serán las llamadas “foreground branches”, y las demás ramas serán “background”. En el modelo A se fijan $\omega_0 < 1$ y $\omega_1 = 1$, siendo que éste modelo es una extensión del M1 (neutral) descrito por Nielsen y Yang en 1998 y que asume dos clases de sitios con $\omega_0 < 1$ y $\omega_1 = 1$ en todos los linajes. En el modelo B los valores para ω_0 y ω_1 se estiman a partir de los datos por máxima verosimilitud como parámetros libres. Este es una extensión del modelo 3 (M3, ‘discreto’) de Yang, et al. (2000) con $K=2$ clases de sitios. En ambos modelos A y B las proporciones p_0 y p_1 así como ω_2 se estiman a partir de los datos.

En la Tabla 3 se muestra una explicación de estos modelos así como de los parámetros calculados en cada caso.

Tabla 3. En el caso del modelo A se fijan $\omega_0 < 1$ y $\omega_1 = 1$ en el linaje ‘background’ mientras que en el ‘foreground’ se estiman pero restringidos a ser sitios neutros o seleccionados negativamente. En el caso del modelo B los sitios en el linaje ‘background’ se fijan en $\omega_0 < 1$ y $\omega_1 = 1$ pero en el linaje ‘foreground’ se estiman permitiendo la $\omega_2 = 2$ para algunos sitios.

Parámetros para los modelos rama-por-sitio Yang y Nielsen, 2002			
Clase de sitio	Proporción	ω ramas ‘background’	ω ramas ‘foreground’
0	p_0	ω_0	ω_0
1	p_1	ω_1	ω_1
2	$p_2 = (1 - p_0 - p_1)p_0 / (p_0 + p_1)$	ω_0	ω_2
3	$p_3 = (1 - p_0 - p_1)p_1 / (p_0 + p_1)$	ω_1	ω_2

Las pruebas PTV en el caso de los modelos A y B rama-por-sitio para evaluar la significancia estadística de los resultados se seleccionaron siguiendo las recomendaciones descritas en Yang y Nielsen, 2002.

IV. d. Detección de sitios bajo selección positiva

Los cálculos de verosimilitud con los modelos de codones y las comparaciones con las pruebas de tasas de verosimilitud se llevan a cabo para detectar con significancia estadística la presencia de heterogeneidad de tasas ω entre sitios y la presencia de sitios con $\omega > 1$, sin embargo estos análisis no nos indican qué sitios están bajo selección positiva. Hay que considerar que la estimación de una tasa ω independiente para cada sitio daría lugar a una sobreparametrización y un aumento excesivo en los niveles de error, por lo cual los modelos de codones implementan una categorización de los valores de ω para los sitios siguiendo en algunos casos una distribución particular (por ejemplo la distribución beta en los modelos sitio-por-sitio 7 y 8).

Después de éstos análisis los sitios particulares se detectan utilizando una aproximación bayesiana empírica, con la que se busca inferir a que clase de sitio es mas probable que pertenezca un sitio observado en el alineamiento (siendo que las clases de sitios están definidas en el modelo de codones empleado) (Nielsen y Yang 1998), dada la distribución anterior representada en este caso por los parámetros calculados por máxima verosimilitud para el modelo de codones. La probabilidad posterior de que el sitio h con un vector de datos \mathbf{x}_h observados pertenece a la clase de sitio k esta dada por:

$$f(y_h = k | \mathbf{x}_h) = p_k f(\mathbf{x}_h | y_h = k) / f(\mathbf{x}_h) = p_k f(\mathbf{x}_h | y_h = k) / \sum p_j f(\mathbf{x}_h | y_h = j)$$

El resultado para cada sitio va a ser la serie de probabilidades de su pertenencia a cada clase de sitio (por ejemplo en el modelo 2 'selección' las clases serian $\omega = 1$, $\omega < 1$ y $\omega > 1$) y se decide a qué clase pertenece dicho sitio si tiene una mayor probabilidad de pertenecer a esa clase. Los sitios detectados bajo selección positiva son aquellos con una mayor probabilidad de pertenecer a una clase con $\omega > 1$.

Los métodos basados en modelos de codones que utilizan el criterio estadístico de la máxima verosimilitud han arrojado buenos resultados para la detección de selección positiva en sitios particulares (Yang, et al. 2000; Yang y Swanson, 2002) y recientemente se han logrado desarrollar modelos de codones con los que se logra detectar selección positiva en sitios específicos de linajes particulares (Yang y Nielsen, 2002; Bielawski y Yang, 2004), que son muy adecuados para los objetivos del presente trabajo dadas las características de nuestros datos y de las hipótesis que queremos probar. Estos métodos acomodan diferentes distribuciones para el cálculo de las tasas ω heterogéneas entre sitios con la finalidad de evitar el cálculo de una ω para cada sitio, pues daría lugar a un problema de sobreparametrización y aumento del error (Yang y Bielawski, 2000). Sin embargo, los métodos basados en verosimilitud con modelos de codones tienen algunas limitantes que deben considerarse. Al igual que con cualquier método de análisis de secuencias, se requiere de un alineamiento debe confiable y con el menor número de ambigüedades posibles, ya que en éstos datos estará basado el calculo de las tasas ω . Los métodos desarrollados por Yang y colaboradores, y todos los que se basan en el criterio de máxima verosimilitud para pruebas filogenéticas siguiendo el algoritmo de Felsenstein (1989) tienen problemas para considerar los "gaps" en un alineamiento, y en particular los modelos de codones no son capaces de considerar sitios con cambios en más de una posición, lo cual puede arrojar muchos falsos positivos. Para lidiar con estos problemas decidimos seguir el análisis con alineamientos minuciosamente trabajados de manera manual, y de tener mucho cuidado en la

interpretación de los sitios resultantes bajo SP y solo considerar aquellos que se encuentran en las partes menos ambiguas del alineamiento. De cualquier manera trabajar con los genes MADS implica dificultades, en particular considerando el dominio C-terminal, que es el menos conservado y por tanto el más difícil de alinear. Se requiere una cautelosa consideración de los resultados obtenidos.

V. Estimación de tiempos de divergencia.

La estimación de tiempos de divergencia se llevó a cabo también usando el software PAML (<http://abacus.gene.ucl.ac.uk/software/paml.html>), siguiendo el método descrito en Yang, Z. 2004 que implementa el algoritmo llamado “A heuristic rate smoothing algorithm” o AHRS por sus siglas en inglés, y que consta de tres pasos:

1. Estima las longitudes de las ramas por máxima verosimilitud dada una topología que puede ser estimada de manera independiente siguiendo el algoritmo “pruning” de Felsenstein, 1981. Se asume que las sustituciones no siguen un modelo tipo reloj molecular (homogeneidad de las tasas de sustitución en todos los linajes). En este caso se utilizaron las topologías encontradas con métodos Bayesianos.
2. Siguiendo el algoritmo AHRS se estiman las tasas de sustitución de las ramas en el árbol así como los tiempos de divergencia como valores preliminares. En este paso se ajusta un modelo estocástico de cambio en las tasas de sustitución con los valores de las longitudes de las ramas estimadas mientras se busca una minimización de los cambios de tasas al interior de los linajes. Los valores obtenidos se utilizan para hacer una clasificación de las ramas en grupos con diferentes tasas, siguiendo una maximización por máxima verosimilitud.
3. Se estiman los tiempos de divergencia y las tasas de sustitución para cada grupo.

Se aplicó este método de estimación de tiempos de divergencia calibrando un reloj molecular para los genes MADS Tipo B y para los genes MADS tipo APETALA3 con fechas obtenidas a partir de datos reportados de registro fósil con 132 ma en el origen de las angiospermas por granos de polen fósiles descritos en Brenner, 1996 y de 121 ma para la edad de las eudicotiledóneas, con base en granos de polen tricolpados fósiles descritos en Doyle y Hotton, 1991 respectivamente, según datos tomados de Magallón y Sanderson, 2001.

Resultados

I. Selección del modelo de sustitución.

Las pruebas de tasas de verosimilitud (PTV) realizadas con Modeltest arrojaron resultados diferentes para los clados de genes parálogos tipo *PISTILLATA* (*PI*) y tipo *APETALA3* (*AP3*). Para el linaje de los MADS tipo *AP3* el modelo que dio el mayor valor de verosimilitud fue el “**general reversible en el tiempo GTR**” (-lnL = 24582.6289); mientras que para el linaje de los tipo *PI* fue el modelo “**Tamura-Nei TN**” (-lnL = 17887.3066). En ambos casos las PTVs indicaron una mejora significativa en el valor de verosimilitud al adicionar el parámetro para la distribución Gamma a las tasas de sustitución y una proporción de sitios invariantes. El modelo de Tamura-Nei es mas sencillo, y permite frecuencias variables de las bases, pero igualdad en las frecuencias de transición y transversión (Tamura-Nei 1993). El modelo GTR permite también frecuencias variables de las bases, con una matriz de sustitución simétrica en la que cada tipo de cambio es diferente (Tavare, 1986). Para la matriz de todos los genes MADS tipo B de plantas se utilizó también el modelo general reversible en el tiempo, GTR.

II. Análisis filogenético

II.a. Análisis filogenético de los genes MADS-box Tipo B en angiospermas y gimnospermas

Obtuvimos reconstrucciones filogenéticas para los genes B de gimnospermas y angiospermas mediante el uso de métodos Bayesianos. Los análisis filogenéticos con todos los ortólogos de genes MADS-Box tipo B de angiospermas y gimnospermas fueron útiles para: 1) discutir la posición de los genes B de gimnospermas, 2) corroborar la ortología de las secuencias obtenidas de GenBank, 3) realizar los análisis de inferencia del papel de la selección positiva en los patrones de sustitución observados, y 4) obtener calibraciones del reloj molecular con el fin de fechar la duplicación que dio lugar a los linajes de los parálogos *AP3* y *PI*.

Como grupo externo para todos los análisis filogenéticos realizados en este estudio se utilizaron las secuencias AGL19, AGL18 y AGL15, que son los parálogos mas cercanos al clado de los genes tipo B (*APETALA3* y *PISTILLATA*) en las filogenias más recientes reportadas para los MADS-Box de *Arabidopsis* (Martínez-Castilla y Álvarez-Buylla, 2003, Nei, M., et al, 2003) además de la secuencia tipo MADS *PpMADS1* del musgo *Physcomitrella patens*. La filogenia resultante del análisis por métodos bayesianos se muestra en la Figura 9.

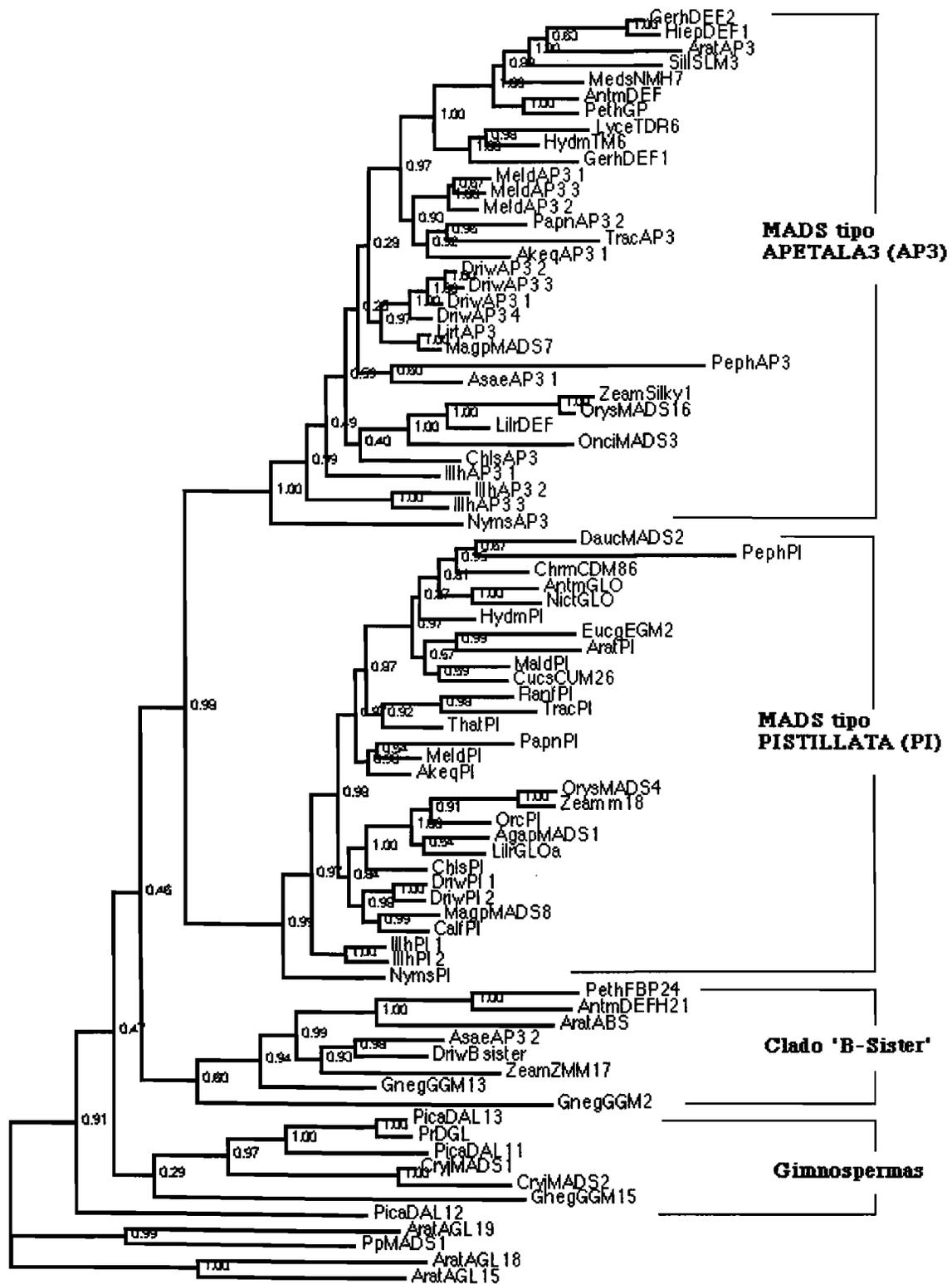


Figura 9. Filogenia resultante del análisis por métodos Bayesianos de los genes MADS tipo B de plantas. Los números en los nodos indican los valores de probabilidad posterior para cada clado, que se emplean para considerar su soporte. Se corrieron 2,500,000 generaciones, salvando un árbol cada 100 y descartando los primeros 300 como 'burnin'.

Los genes tipo *AP3* y *PI* se agruparon en dos clados monofiléticos hermanos con altos valores de probabilidad posterior o pp (0.98 pp). Como hermanos a estos dos grupos se ubicó con baja probabilidad posterior (0.46 pp) el clado de los genes conocidos como ‘**B-Sister**’, que incluye a GGM2 y GGM13 de las *Gnetales*. La monofilia de este clado también tiene bajo apoyo (0.60 pp). Los genes tipo B de gimnospermas se agruparon de manera externa al grupo de los genes *AP3-PI* y **B-Sister**, la mayoría en un solo clado a excepción de *PicaDAL12* de *Picea abies*, que se ubicó de manera externa a todas las secuencias, sin embargo esta posición también está apoyada con bajos valores de probabilidad posterior (0.47 pp).

II.b. Análisis filogenético genes MADS tipo *PISTILLATA* (*PI*)

La filogenia resultante por métodos bayesianos para los genes tipo *PISTILLATA* se muestra en la Figura 10. Tanto en este como en el árbol de los *AP3* se incluyeron las secuencias de gimnospermas. Los genes MADS tipo B de gimnospermas se ubican como secuencias hermanas de los genes tipo *PI* de angiospermas, con altos valores de probabilidad posterior. Sin embargo, parece que la posición de estos genes es muy inestable porque resulta muy difícil alinearlos sin ambigüedad con las secuencias de angiospermas y su posición cambia mucho dependiendo de pequeños ajustes en su alineamiento (estos resultados no se muestran).

Los análisis realizados con máxima parsimonia arrojaron filogenias con politomías (resultados no mostrados) y decidimos recurrir a métodos Bayesianos para obtener una filogenia totalmente resuelta que es necesaria para los análisis de tasas de sustitución. Es importante notar, sin embargo, que los valores de probabilidad posterior para los clados resueltos en el árbol de la Figura 10 son bajos. Pero las secuencias tipo *PI* de angiospermas se agruparon en un clado monofilético con 1.0 pp, y al interior de éste grupo se obtuvieron recurrentemente dos principales clados monofiléticos, que identificamos como los clados PI-I y PI-II.

Clado PI-I, eudicotiledóneas basales y angiospermas basales

En este gran grupo se distinguieron tres subclados, a los cuales identificamos como PI-Ia, PI-Ib y PI-Ic (ver Figura 10).

Subclado PI-Ia. Piperales

Dentro del clado PI-Ia (0.57 pp) encontramos a las secuencias de las angiospermas basales representantes de las **Piperales** (*Peperomia hirta* y *Piper magnificum*). Su posición como hermanas de las demás secuencias de angiospermas basales está pobremente soportada, si se consideran los valores en conjunto de apoyo a la monofilia del clado PI-I que tiene 0.36pp y la monofilia del clado interno siguiente (PI-Ib+PI-Ic), que tiene solo 0.76pp.

En repeticiones de este análisis con la misma matriz de nucleótidos pero con pequeñas variaciones en la manera de alinearlos se observó que la posición de este clado a veces cambia a una posición de hermano de los genes del clado PI-II, pero también con bajos valores de pp (estos resultados no se muestran).

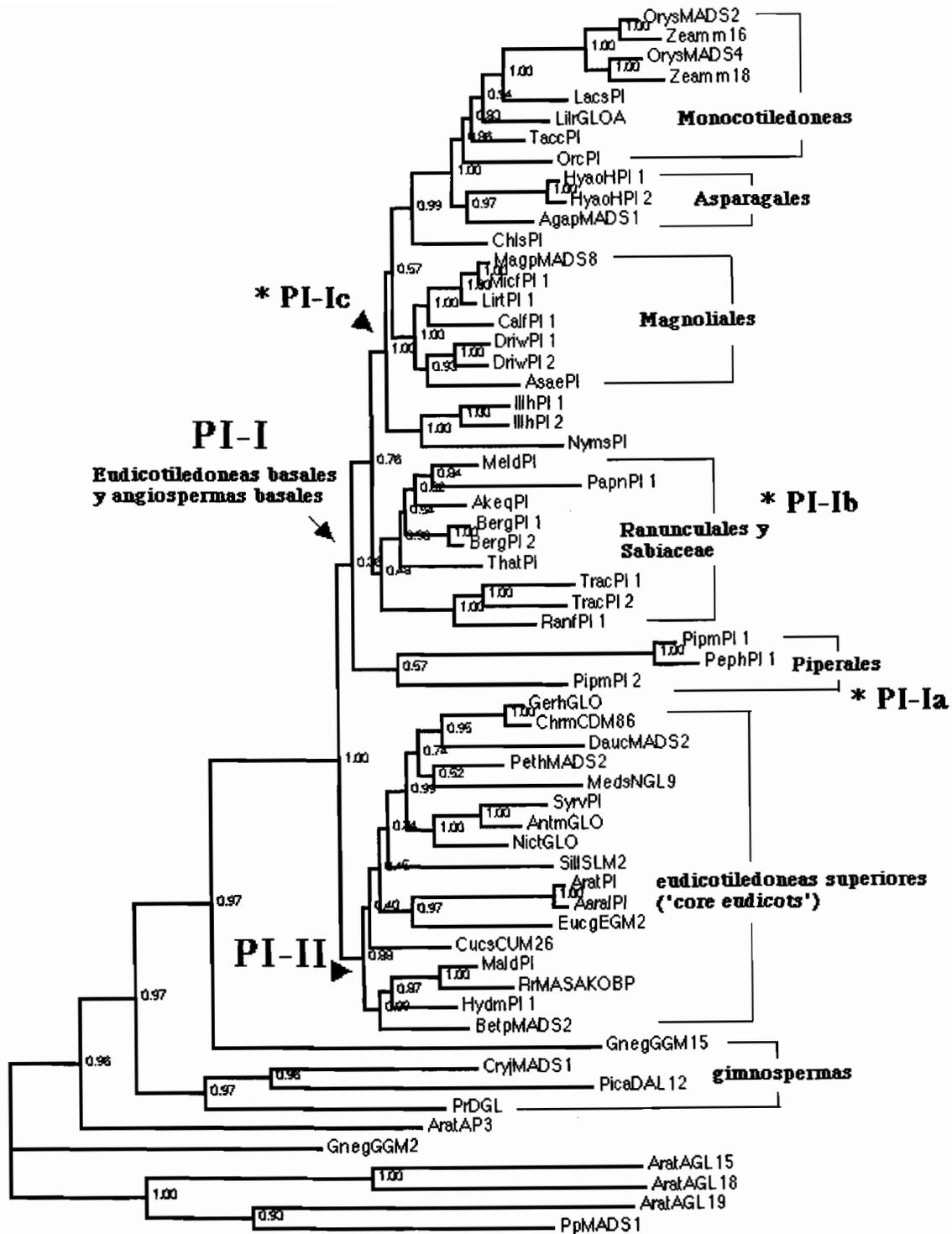


Figura 10. Filogenia resultante del análisis por métodos Bayesianos de los genes tipo PSITILLATA. Los números en los nodos indican los valores de probabilidad posterior para cada clado, que se emplean para considerar su soporte. Se corrieron 2,500,000 generaciones, salvando un árbol cada 100 y descartando los primeros 300 como 'burnin'.

Subclado PI-Ib. Ranunculales y Sabiaceae.

En este subclado, con una probabilidad posterior de 0.48 quedan incluidas las secuencias representantes de las eudicotiledóneas basales **Ranunculales y Sabiaceae**.

Subclado PI-Ic. Angiospermas basales, Chlorantales, Magnoliales y monocotiledóneas.

En el clado PI-Ic (1.00 pp) encontramos en primer lugar en posición de grupo hermano con respecto al resto de las secuencias a *NymaPI*, *IllhPI_1* e *IllhPI_2* (1.00 pp) correspondientes a las angiospermas basales **Nymphaceae y Austrobaileyales**. El clado siguiente (1.00 pp) corresponde a las secuencias representantes de las **magnoliales**. Enseguida encontramos la secuencia *ChlsPI*, ortólogo *PI* de *Chlorantus spicatus* (**Chlorantales**) como hermana del resto de las secuencias correspondientes a las **Asparagales**, representadas por *Agapanthus praecox* (*AgapMADS1*) y *Hyacinthus orientalis*, *HyaoHPI_1* e *HyaoHPI_2* (0.97 pp) y al clado de las **monocotiledóneas** (0.86). La posición de la secuencias de *Chlorantus spicatus* es muy variable dependiendo de pequeños cambios en la manera de alinearla.

Clado PI-II. Eudicotiledóneas ('core-eudicots')

En este clado, con un apoyo por probabilidad posterior de 0.88 encontramos a las secuencias tipo PI representantes de las eudicotiledóneas superiores (core-eudicots).

II.c. Análisis filogenético genes tipo APETALA3

La filogenia para los AP3 se muestra en la Figura 11. A diferencia de lo ocurrido en el análisis de los genes tipo PI, la mayoría de los genes tipo B de gimnospermas se ubicaron formando un grado parafilético externo al grupo de los AP3 de angiospermas, en el cual el parálogo *PISTILLATA* (*AratPI*) esta embebido. Una única excepción la encontramos en la ubicación del gen GGM2 de *Gnetales*, que se colocó como hermana de las AP3 de gimnospermas. No logramos recuperar un clado monofilético para las gimnospermas.

Posteriormente a los genes de gimnospermas y al parálogo *AratPI* encontramos a los ortólogos de AP3 de angiospermas agrupados en un clado monofilético con un apoyo por probabilidad posterior de 0.99 pp. Las secuencias correspondientes a angiospermas basales **Illiciaceae y Nymphaceae** se ubican en un grupo hermano con respecto al resto, en concordancia con la filogenia de las angiospermas basada en datos moleculares (APG II, 2003). Posteriormente logramos recuperar dos grandes clados, que hemos llamado AP3-I (0.42 pp) y AP3-II (0.99).

Clado AP3-I. Monocotiledóneas y angiospermas basales.

En este clado encontramos a las secuencias correspondientes a las monocotiledóneas y angiospermas basales en dos subclados diferentes (ver Figura 11). En primer lugar a las **Magnoliaceas** (0.91 pp) en el subclado llamado AP3-Ia.

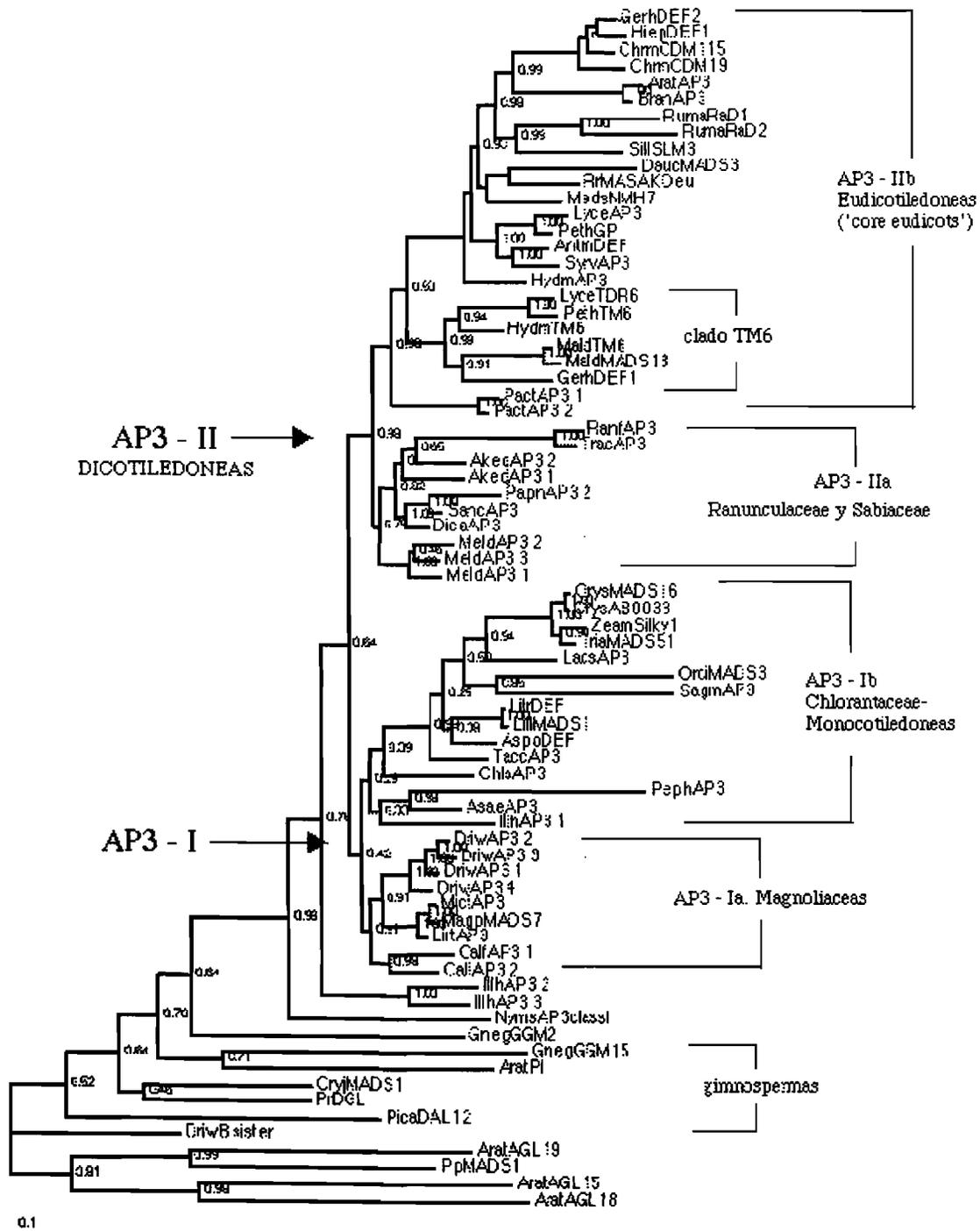


Figura 11. Filogenia resultante del análisis por métodos Bayesianos de los genes tipo APETALA3. Los números en los nodos indican los valores de probabilidad posterior para cada clado, que se emplean para considerar su soporte. Se corrieron 2,500,000 generaciones, salvando un árbol cada 100 y descartando los primeros 300 como 'burnin'.

Subclado AP3-Ib. Chlorantaceae-Monocotiledóneas.

Este clado recupera de manera basal a *ChlsAP3* de *Chloranthus spicatus* (0.39) como secuencia hermana al clado de las **Monocotiledóneas** (0.98). El clado Chlorantaceae-Monocotiledóneas se ubicó como hermano de un clado no basal que incluye a las angiospermas basales **Illiciaceae y Piperales** (*PephAP3*, *AsaeAP3* y *IllhAP3_1*) sin embargo, esta ubicación esta apoyada con bajos valores de pp (0.29 pp). La alineación de estas secuencias debe revisarse para corroborar esta posición, en especial de *IllhAP3-1*.

Clado AP3-II. Eudicotiledóneas.

Este clado comprende al gran grupo de los *AP3* de eudicotiledóneas. En este grupo encontramos primero en posición basal a las secuencias tipo *AP3* de las eudicotiledóneas basales **Ranunculaceae y Sabiaceae** formando un grupo monofilético (0.75 pp) en el clado que denominamos AP3-IIa. Estas secuencias se ubican como hermanas de las **eudicotiledóneas superiores o core-eudicots** (0.99 pp) en el subclado AP3-IIb, a pesar de que comparten con las secuencias del clado AP3-I el motivo paleoAP3 en el carboxilo terminal.

Subclado AP3-IIb. Eudicotiledóneas

En este grupo encontramos en primer lugar a las secuencias *PactAP3_1* y *PactAP3_2* de *Pachysandra terminalis*, representantes de la familia de las Buxaceae, con apoyo de 0.98 pp. Como hermanas de estas secuencias encontramos dos clados: el clado de las **eudicotiledóneas centrales o superiores** (core-eudicots) con un soporte por probabilidad posterior de 0.53, y por último, como hermanas a estas recuperamos el clado de los genes **TM6** (0.99), parálogos de los tipo-AP3 dentro de las eudicotiledóneas superiores.

III. Inferencia de Selección Positiva (SP) y Diferencias Significativas en las Tasas de Sustitución (DTS) en la evolución de los genes MADS-box tipo B

III.a. Pruebas de detección de selección sitio-por-sitio. Tasas de sustitución heterogéneas entre sitios

Utilizamos las topologías recuperadas para las secuencias MADS tipo B, MADS tipo AP3 y MADS tipo PI con métodos bayesianos descritos anteriormente junto con modelos explícitos de sustitución por codones para llevar a cabo los análisis de detección de heterogeneidad en las tasas de sustitución para cada codon, y las pruebas de selección positiva por medio de pruebas de tasas de verosimilitud (PTV). Los valores de verosimilitud así como los sitios bajo selección positiva detectados por probabilidades posteriores se estimaron con el programa PAML (ver sección de Secuencias y Métodos).

Las pruebas de detección de heterogeneidad en las tasas de sustitución por codones y selección positiva en sitios específicos se llevaron a cabo haciendo las estimaciones de verosimilitud bajo los modelos 0, 1, 2, 3, 7 y 8 descritos por Yang et al, 2000. Las comparaciones así como los detalles sobre cada modelo se explican en la parte de Secuencias y Métodos.

A continuación se muestran las tablas con los parámetros estimados bajo cada modelo y los sitios detectados. Enseguida de cada una de estas tablas se muestra una pequeña tabla que resume los resultados de las pruebas estadísticas por tasas de verosimilitud.

a) Resultados. Genes MADS-Box Tipo B.

Tabla 4a. Resumen de resultados obtenidos con PAML para los modelos de tasas de sustitución heterogéneas entre sitios aplicados a los genes MADS-Box tipo B. p es el número de parámetros para cada modelo, ι son los valores de $-\ln l$ de verosimilitud estimados con PAML bajo cada modelo. Se reportan solo los sitios encontrados bajo selección positiva con una probabilidad posterior > 0.8 .

MODELO	Breve descripción	p	ι	valores de p estimados	Sitios detectados bajo SP
Modelo 0 (una tasa)	Los valores de ω para cada sitio se distribuyen de manera homogénea	392	-37531.987393	$\omega = 0.17357$	Ninguno
Modelo 1 (neutral)	Hay sitios bajo selección negativa y sitios neutros $\omega=1$	392	-37144.100266	$\omega_1 = 0.17454$ $\omega_2 = 1.00$ $p_1 = 0.57899$ $p_2 = 0.42101$	No se permiten
Modelo 2 (selección)	Además de sitios bajo selección negativa y sitios neutros se permite una categoría de sitios bajo SP	394	-37144.100266	$\omega_1 = 0.17454$ $\omega_2 = 1.00$ $\omega_3 = 1.00$ $p_1 = 0.57899$ $p_2 = 0.21248$ $p_3 = 0.20853$	Ninguno
Modelo 3 (discreto)	Tres categorías de sitios con tasas ω estimadas libremente por verosimilitud	396	-36340.498513	$\omega_1 = 0.01409$ $\omega_2 = 0.10227$ $\omega_3 = 0.32991$	Ninguno

				$p_1 = 0.13329$ $p_2 = 0.32547$ $p_3 = 0.54124$	
Modelo 3 (discreto) k=2	categorías de sitios con tasas ω estimadas libremente por verosimilitud. En este caso k=2 categorías	394	-36517.928936	$\omega_1 = 0.06327$ $\omega_2 = 0.30601$ $p_1 = 0.40205$ $p_2 = 0.59795$	Ninguno
Modelo 7 (beta)	Una distribución beta para las categorías de sitios según sus valores de ω . En este caso utilizamos K=10 categorías	393	-36244.496767	$p = 0.76594$ $q = 2.43782$	No se permiten
Modelo 8 (beta & ω)	Igual que el modelo 7, pero se añade una categoría de sitios bajo SP	395	-36244.496846	$p_0 = 1.00$ $p = 0.76594$ $q = 2.43783$ $p_1 = 0.00000$ $\omega = 2.59906$	Ninguno

Tabla 4b. Resultados de las pruebas de tasas de verosimilitud para los modelos de codones aplicados a las secuencias MADS-Box Tipo B. Se muestran los modelos comparados, los grados de libertad considerados (gl) para cada comparación, el estadístico de la prueba tasas de verosimilitud ($2\Delta\iota$) y la P calculada en el resultado.

Modelos a comparar	gl	$2\Delta\iota$	resultado
M1 vs M2	2	0	$P=1$; no se puede rechazar M1
M0 vs M3 (k=3)	4	2382.9777	$P<0.0001$; se rechaza M0
M7 vs M8	2	0.00016	$P=0.999$; no se puede rechazar M7
Nuevas comparaciones para ω heterogéneas entre sitios (Bielawski & Yang, 2004)			
M0 vs M3 (k=2)	2	2028.1116914	$P<0.0001$; se rechaza M0
M3 (k=2) vs M3 (k=3)	2	117.430423	$P<0.0001$; se rechaza M3 k=2

b) Resultados. Genes tipo *PISTILLATA*

Tabla 5a. Resumen de resultados obtenidos con PAML para los modelos de tasas de sustitución heterogéneas entre sitios aplicados a los genes MADS tipo *PISTILLATA*. p es el número de parámetros para cada modelo, ι son los valores de $-\ln l$ de verosimilitud estimados con PAML bajo cada modelo. Se reportan solo los sitios encontrados bajo selección positiva con una probabilidad posterior > 0.8 .

MODELO	Breve descripción	p	ι	valores de p estimados	Sitios detectados bajo SP
Modelo 0 (una tasa)	Los valores de ω para cada sitio se distribuyen de manera homogénea	392	-19734.250582	$\omega = 0.16528$	Ninguno
Modelo 1 (neutral)	Hay sitios bajo selección negativa y sitios neutros $\omega=1$	392	-19345.353886	$\omega_1 = 0.13837$ $\omega_2 = 1.00$ $p_1 = 0.64632$ $p_2 = 0.35368$	No se permiten
Modelo 2 (selección)	Además de sitios bajo selección negativa y sitios neutros se permite una categoría de sitios bajo SP	394	-19345.353886	$\omega_1 = 0.13837$ $\omega_2 = 1.00$ $\omega_3 = 1.00$ $p_1 = 0.64632$ $p_2 = 0.20151$ $p_3 = 0.15217$	70* (0.863), 76* (0.981), 79* (0.978).
Modelo 3 (discreto)	categorías de sitios con tasas ω estimadas libremente por verosimilitud. En este caso k = 3 categorías	396	-18956.576517	$\omega_1 = 0.01553$ $\omega_2 = 0.15497$ $\omega_3 = 0.47055$ $p_1 = 0.26847$ $p_2 = 0.37690$ $p_3 = 0.35463$	Ninguno
Modelo 3 (discreto) k=2	Tres categorías de sitios con tasas ω estimadas libremente por verosimilitud	394	-19119.473886	$\omega_1 = 0.05818$ $\omega_2 = 0.34924$ $p_1 = 0.47687$	Ninguno

				$p_2 = 0.52313$	
Modelo 7 (beta)	Una distribución beta para las categorías de sitios según sus valores de ω . En este caso utilizamos $K=10$ categorías	393	-18922.823639	$p = 0.53975$ $q = 1.53080$	No se permiten
Modelo 8 (beta & ω)	Igual que el modelo 7, pero se añade una categoría de sitios bajo SP	395	-18907.255505	$p_0 = 0.92922$ $p = 0.59572$ $q = 2.11293$ $p_1 = 0.07078$ $\omega = 4.83230$	70* (0.919), 72* (0.876), 76* (0.984), 79* (0.986), 300N (0.857)

Tabla 5b. Resultados de las pruebas de tasas de verosimilitud para los modelos de codones aplicados a las secuencias *PISTILLATA*. Se muestran los modelos comparados, los grados de libertad considerados (gl) para cada comparación, el estadístico de la prueba tasas de verosimilitud ($2\Delta\iota$) y la P calculada en el resultado.

Modelos a comparar	gl	$2\Delta\iota$	resultado
M1 vs M2	2	0	$P=1$; no se puede rechazar M1
M0 vs M3 (k=3)	4	1555.348	$P<0.0001$; se rechaza M0
M7 vs M8	2	31.136	$P=0.000017$; se rechaza M7
Nuevas comparaciones para ω heterogéneas entre sitios (Bielawski & Yang, 2004)			
M3 (k=2) vs M0	2	1229.5534	$P<0.0001$; se rechaza M0
M3 (k=3) vs M3 (k=2)	2	325.79474	$P<0.0001$; se rechaza M3 (k=2)

b) Resultados. Genes tipo *APETALA3*

Tabla 6a. Resumen de resultados obtenidos con PAML para los modelos de tasas de sustitución heterogéneas entre sitios aplicados a los genes MADS tipo *APETALA3*. p es el número de parámetros para cada modelo, ι son los valores de $-\ln l$ de verosimilitud estimados con PAML bajo cada modelo. Se reportan solo los sitios encontrados bajo selección positiva con una probabilidad posterior > 0.8 .

MODELO	Breve descripción	p	ι	valores de p estimados	Sitios detectados bajo SP
Modelo 0 (una tasa)	Los valores de ω para cada sitio se distribuyen de manera homogénea	392	-29732.280255	$\omega = 0.16528$	Ninguno
Modelo 1 (neutral)	Hay sitios bajo selección negativa y sitios neutros $\omega=1$	392	-29470.611411	$\omega_1 = 0.16797$ $\omega_2 = 1.00$ $p_1 = 0.72125$ $p_2 = 0.27875$	No se permiten
Modelo 2 (selección)	Además de sitios bajo selección negativa y sitios neutros se permite una categoría de sitios bajo SP	394	-29470.611411	$\omega_1 = 0.16798$ $\omega_2 = 1.00$ $\omega_3 = 1.00$ $p_1 = 0.72124$ $p_2 = 0.13936$ $p_3 = 0.13940$	Ninguno
Modelo 3 (discreto)	categorías de sitios con tasas ω estimadas libremente por verosimilitud. En este caso $k=3$ categorías	396	-28778.516647	$\omega_1 = 0.01231$ $\omega_2 = 0.14087$ $\omega_3 = 0.37161$ $p_1 = 0.21283$ $p_2 = 0.33249$ $p_3 = 0.45468$	Ninguno
Modelo 3 (discreto) $k=2$	categorías de sitios con tasas ω estimadas libremente por verosimilitud. En este caso $k=2$ categorías	394	-28929.892464	$\omega_1 = 0.03903$ $\omega_2 = 0.28832$ $p_1 = 0.33854$ $p_2 = 0.66146$	
Modelo 7 (beta)	Una distribución beta para las categorías de sitios según sus valores de ω . En este caso utilizamos $K=10$ categorías	393	-28739.551098	$P = 0.69480$ $q = 2.23147$	No se permiten

Modelo 8 (beta & ω)	Igual que el modelo 7, pero se añade una categoría de sitios bajo SP	395	-28738.648626	$p_0 = 0.98825$ $p = 0.70878$ $q = 2.38142$ $p_1 = 0.01175$ $\omega = 3.42369$	46* (0.809)
--------------------------------	--	-----	---------------	--	--------------------

Tabla 6b. Resultados de las pruebas de tasas de verosimilitud para los modelos de codones aplicados a las secuencias *APETALA3*. Se muestran los modelos comparados, los grados de libertad considerados (gl) para cada comparación, el estadístico de la prueba tasas de verosimilitud ($2\Delta\iota$) y la P calculada en el resultado.

Modelos a comparar	gl	$2\Delta\iota$	resultado
M1 vs M2	2	0	$P=1$; no se puede rechazar M1
M0 vs M3 (k=3)	4	1907.527	$P<0.0001$; se rechaza M0
M7 vs M8	2	1.804	$P=0.4057$; no se puede rechazar M7
Nuevas comparaciones para ω heterogéneas entre sitios (Bielawski & Yang, 2004)			
M0 vs M3 (k=2)	2	1604.77558	$P<0.0001$; se rechaza M0
M3 (k=2) vs M3 (k=3)	2	302.75164	$P<0.0001$; se rechaza M3 (k=2)

El modelo 0, que calcula un solo valor para la tasa ω de manera general en todos los sitios y en todas las ramas del árbol puede ser útil para hacer una comparación de las tasas de sustitución generales en los genes tipo B. La tasa ω calculada bajo este modelo para los genes Tipo B es de **0.173**, para los genes tipo PI es de **0.165** y para los tipo AP3 es también de **0.165**, que son valores muy similares y que indican una fuerte conservación (por tener valores cercanos a 0) en la mayoría de sitios en estos genes. El hacer un promedio para el cálculo de las tasas de sustitución entre todos los sitios y ramas del árbol reduce el poder de detectar SP y DTS, ya que se pierde la señal proveniente de eventos específicos de diversificación funcional en sitios y linajes particulares. Con los modelos para detección de selección sitio por sitio de manera general utilizados en esta parte del análisis no conseguimos detectar sino unos cuantos sitios en los genes tipo PI y sólo uno en los genes AP3 bajo selección positiva, que son sitios con alineamientos ambiguos y por tanto de difícil interpretación. Las pruebas estadísticas de tasas de verosimilitud que se detallan en las tablas 4b, 5b y 6b indican el poco poder para detectar selección positiva con éstos modelos (ver comparaciones M1 vs M2, M0 vs M3 (k=3) y M7 vs M8), pues no todas las PTV resultaron en un rechazo de la hipótesis nula. Sin embargo, las pruebas realizadas para detectar heterogeneidad en las tasas de sustitución entre sitios (M0 vs M3 (k=2) y M3 (k=2) vs M3 (k=3)), que no implican necesariamente selección positiva, resultaron en todos los casos en el rechazo de la hipótesis nula de homogeneidad de tasas. A pesar de que logramos detectar ésta heterogeneidad, los modelos generales de detección de selección sitio por sitio no nos permitieron detectar selección positiva de manera significativa ni sitios particulares bajo selección positiva de manera considerable. Esto puede indicar que los eventos de diversificación funcional de las proteínas MADS tipo B responden a eventos particulares de cambios en las tasas de sustitución de sitios específicos con implicaciones funcionales importantes y en linajes particulares. Para probar esta hipótesis y para explicar de manera más detallada la relación entre eventos de duplicación y diversificación funcional empleamos los modelos de detección de SP y DTS en sitios y linajes particulares que se explican en la sección de Secuencias y Métodos, y cuyos resultados se presentan a continuación.

IIIb. Pruebas de detección de selección linaje-específicas. Papel de la selección adaptativa en la divergencia funcional de los MADS-Box Tipo B

Además de hacer las pruebas para detectar diferencias en las tasas de sustitución y selección positiva en la filogenia de todos los genes MADS tipo B se hicieron de manera más específica para los parálogos AP3 y PI por separado, con la finalidad de lograr que los alineamientos fueran más compactos y hacer un análisis más detallado. El alineamiento de los MADS tipo B incluye secuencias más divergentes entre si, por lo que tiene un alineamiento con mayor número de ambigüedades y requiere de una interpretación más cuidadosa de los resultados.

Primero se probaron los modelos de codones diseñados para evaluar la presencia de tasas de sustitución heterogéneas entre clados hermanos producto de una duplicación, es decir, para probar divergencia funcional en la familia de los MADS Tipo B, de los AP3 y de los PI con el modelo D; y después se utilizaron los modelos de codones A y B para probar si esas diferencias eran debidas a selección positiva.

Las hipótesis planteadas en cada caso se muestran en la Tabla 7, así como una breve justificación de los motivos biológicos de su evaluación. En la Tabla 8 se muestran los resultados de las pruebas para detectar diferencias significativas en las tasas de sustitución y en la Tabla 9 para las pruebas de detección de selección positiva. Los resultados se encuentran resumidos y esquematizados en las figuras 12 a 14, dónde se indican las hipótesis planteadas en los árboles filogenéticos correspondientes. En ellos se puede observar que la definición de las hipótesis o selección de ramas y clados se planteó de manera tal que se pudiera seguir la evolución de los genes B, buscando eventos de selección positiva y cambios en las tasas de sustitución, desde la rama que da lugar a los MADS-Box tipo B y los separó de los MADS tipo B de gimnospermas, hasta llegar a las puntas de las ramas en los clados de AP3 y PI; poniendo especial atención en los eventos de duplicación que caracterizan a esta familia. Una adecuada interpretación de los resultados obtenidos, no solo de los sitios detectados bajo selección positiva, sino también de los diferentes valores de ω con las proporciones de sitios correspondientes para cada caso puede aportar importantes elementos para comprender como ha sido la evolución molecular de estos factores de transcripción.

Tabla 7 Hipótesis que se probaron para divergencia en las tasas de sustitución (modelos C y D) y selección positiva linaje-específica (modelos A y B) para los genes Tipo B, los tipo AP3 y los tipo PI. Las hipótesis se indican también en las filogenias mostradas en las figuras 12 a 14.

Genes MADS Tipo B		
HIPOTESIS	DESCRIPCION	JUSTIFICACION
B - H1	En la rama que da lugar a las gimnospermas	Diferencias tempranas en los patrones de diversificación funcional de los MADS-Tipo B de gimnospermas.
B - H2	En la rama que da lugar a los MADS-Tipo B, incluyendo el clado de los B-sister	Diferencias tempranas en los patrones de diversificación funcional de los MADS-Tipo B de angiospermas y genes "B-sister". Homogeneidad en las tasas de sustitución de genes MADS tipo B y genes "B-sister".
B - H3	En la rama que da lugar a las secuencias del clado de los 'B-Sister'.	Evaluación de diferencias tempranas en las tasas de sustitución de estos genes que puedan indicar diversificación funcional en este grupo con funciones poco estudiadas. Búsqueda de posibles sitios característicos de este clado bajo selección positiva que puedan aportar pistas para el entendimiento de su función.
B - H4	En la rama que da lugar a los parálogos AP3 y PI	Evaluación de diferencias tempranas en las tasas de sustitución y en las presiones selectivas de los genes AP3-PI en conjunto en comparación con los tipo B de gimnospermas y los 'B-sister'.
B - H5	En la rama que da lugar a los PI	Detección de divergencia funcional temprana inmediata de la duplicación AP3-PI en los genes PI en relación con el origen y

		evolución de la heterodimerización AP3/PI.
B - H6	En la rama que da lugar a los AP3	Detección de divergencia funcional temprana inmediata de la duplicación AP3-PI en los genes AP3 en relación con el origen y evolución de la heterodimerización AP3/PI.
B - H7	En el clado de los genes AP3 (en todas las ramas que integran este clado)	Detección de divergencia funcional posterior a la duplicación y diferenciación, y detección de sitios bajo selección positiva característicos del clado. Detección de sitios que soportan la divergencia funcional de estas secuencias.
B - H8	En el clado de los genes PI (en todas las ramas que integran este clado)	Detección de divergencia funcional posterior a la duplicación y diferenciación, y detección de sitios bajo selección positiva característicos del clado. Detección de sitios que soportan la divergencia funcional de estas secuencias.
B - H9	En el clado de los genes tipo B de gimnospermas (en todas las ramas que integran este clado)	Divergencia funcional posterior a la duplicación que separó a los genes tipo B de angiospermas y gimnospermas y divergencia funcional posterior, en los genes de gimnospermas. Detección de sitios bajo selección positiva particulares en este grupo.
Genes tipo APETALA3		
AP3 - H1	En la rama que da lugar a los AP3 excluyendo a las angiospermas basales Nymphales e Illiciaceae	Divergencia funcional temprana en los AP3 en angiospermas, posterior a la diversificación después de la divergencia de las angiospermas basales Nymphales e Illiciaceae. Búsqueda de sitios distintivos bajo selección positiva de manera temprana en la aparición de este clado, justo después de la separación de las angiospermas basales como punto de partida para las radiaciones adaptativas en las familias de angiospermas.
AP3 - H2	En la rama que da lugar a los cladros euAP3-TM6	Diferencias tempranas en las fuerzas evolutivas que anteceden a la duplicación AP3-TM6 en el contexto de la diferenciación posterior de ese grupo de genes característicos de las eudicotiledóneas superiores o core-eudicots.
AP3 - H3	En el clado de los genes euAP3-TM6 (en todas las ramas que integran este clado)	Eventos de selección positiva y tasas de sustitución homogéneas en estos dos cladros, para hacer una comparación con las hipótesis euAP3 y TM6 por separado.
AP3 - H4	En la rama que da lugar al clado de los genes euAP3	Evaluación de eventos tempranos de divergencia funcional y selección positiva en estos genes, particularmente en la parte del C-terminal. Detección de sitios bajo selección positiva que puedan tener un papel importante en su diversificación funcional.
AP3 - H5	En la rama que da lugar a los TM6	Evaluación de eventos tempranos de divergencia funcional y selección positiva en estos genes, particularmente en la parte del C-terminal. Detección de sitios bajo selección positiva que puedan tener un papel importante en su diversificación funcional.
AP3 - H6	En el clado los euAP3 (en todas las ramas que integran este clado)	Evaluación del papel de la selección positiva en la diversificación funcional de los genes euAP3, en etapas posteriores a la duplicación que le dio origen.
AP3 - H7	En el clado de los TM6 (en todas las ramas que integran este clado)	Evaluación del papel de la selección positiva en la diversificación funcional de los genes TM6, en etapas posteriores a la duplicación que le dio origen.
Genes tipo PISTILLATA		
PI - H1	En la rama que da lugar a todos los PI	Divergencia funcional temprana en los PI en angiospermas. Búsqueda de sitios distintivos bajo selección positiva de manera temprana en la aparición de este clado como punto de partida para su diferenciación funcional. Hacer un contraste de la evolución de este clado con el de los AP3.
PI - H2	En la rama que da lugar a los genes en eudicotiledóneas superiores (clado PI-II)	Básicamente para comparar con los patrones de divergencia funcional y selección positiva temprana encontrados en el clado de los euAP3 y TM6 en los genes tipo AP3. Evaluar si también en este clado hay importantes diferencias en las tasas de sustitución.
PI - H3	En el clado de los genes tipo PI de eudicotiledóneas superiores (en todas las ramas que integran el clado PI-II)	Persistencia de las diferencias en las tasas de sustitución y selección positiva en los genes tipo PI de eudicotiledóneas superiores o core-eudicots y su comparación con los resultados obtenidos para los respectivos parálogos tipo AP3

Tabla 8. Resultados de las Pruebas de Tasas de verosimilitud (PTV) para detectar Diferencias Significativas en las Tasas de Sustitución de los genes MADS-Box tipo B, APETALA3 y PISTILLATA. Se muestra cada hipótesis, los valores de verosimilitud calculados con el modelo MD, los valores del estadístico $2\Delta\iota$ para las PTV, las comparaciones realizadas, la P calculada y el resultado de la prueba de hipótesis.

Hipótesis	Valores de Verosimilitud MD k=2 / MD k=3	$2\Delta\iota$ M3 k=3 (nula) vs MD k=3 gl=2 / M3 k=2 (nula) vs MD k=2 gl=2		P calculada
		Genes MADS-box tipo B		
B-H1 (rama gimnospermas)	-36517.469778 / -36339.968877	1.059 / 0.4591		$P=0.5888$ / $P=0.7948$
B-H9 (clado gimnospermas)	-36502.406724 / -36328.234678	24.527 / 31.0444		$P<0.0001$, rechazo hip. nula / $P<0.0001$, rechazo hip. nula
B-H2 (rama AP3-PI-Bsister)	-36515.787032 / -36338.353288	4.290 / 4.2838		$P=0.1170$ / $P=0.1174$
B-H4 (rama AP3-PI)	-36516.587715 / -36337.690483	5.61606 / 2.682		$P=0.0603$, rechazo hip. nula, 94% confianza / $P=0.2615$
B-H5 (rama PI)	-36510.263966 / -36333.569965	13.857 / 15.3299		$P=0.0009$, rechazo hip. nula / $P=0.0004$, rechazo hip. nula
B-H6 (rama AP3)	-36515.798940 / -36338.102042	4.792 / 4.25998		$P=0.0910$, rechazo hip. nula, 91% confianza / $P=0.1188$, rechazo hip. nula, 90% confianza
B-H8 (clado PI)	-36513.276162 / -36336.935312	7.1264 / 9.3055		$P=0.0283$, rechazo hip. nula, 98% confianza / $P=0.0095$, rechazo hip. nula
B-H7 (clado AP3)	-36517.887640 / -36340.497412	0.0022 / 0.0825		$P=0.9989$ / $P=0.9595$
Genes tipo APETALA3				
AP3-H1 (rama angio-AP3)	-28928.903407 / -28778.505377	0.0225 / 1.978		$P=0.9888$ / $P=0.3719$
AP3-H2 (rama euAP3-TM6)	-28926.699583 / -28774.952135	7.129 / 6.3857		$P=0.0283$, rechazo hip. nula, 97% confianza / $P=0.0410$, rechazo hip. nula, 96% confianza
AP3-H3 (clado euAP3-TM6)	-28929.509343 / -28778.356306	0.3206 / 0.7662		$P=0.8518$ / $P=0.6817$
AP3-H5 (rama TM6)	-28926.048034 / -28773.914434	9.2044 / 7.6888		$P=0.0100$, rechazo hip. nula / $P=0.0213$, rechazo hip. nula, 98% confianza
AP3-H4 (rama euAP3)	-28778.516452 / -28929.642544	0.00038 / 0.4998		$P=0.9998$ / $P=0.7788$
AP3-H7 (clado TM6)	-28926.387701 / -28775.20362	6.626 / 7.00952		$P=0.0364$, rechazo hip. nula, 97% confianza / $P=0.0300$, rechazo hip. nula, 97% confianza
AP3-H6 (clado euAP3)	-28925.910702 / -28777.49861	2.036 / 7.96352		$P=0.3613$ / $P=0.0186$, rechazo hip. nula, 98% confianza
Genes tipo PISTILLATA				
PI-H1 (rama angio-PI)	-19116.518336 / -18954.720403	3.71 / 5.91		$P=0.1564$ / $P=0.0520$, rechazo hip. nula, 95% confianza
PI-H2 (rama core-eudicots)	-19117.945728 / -18956.203328	0.746 / 3.056		$P=0.6886$ / $P=0.2169$
PI-H3 (clado core-eudicots)	-19112.264620 / -18945.295406	22.56 / 14.418		$P<0.0001$, rechazo hip. nula / $P=0.0007$, rechazo hip. nula

Tabla 9. Resultados de las Pruebas de Tasas de Verosimilitud (PTV) para detectar Selección Positiva en linajes particulares. Se muestra la hipótesis, los valores de verosimilitud calculados con los modelos MA y MB, el valor del estadístico $2\Delta\ln$ en las comparaciones evaluadas, la P calculada y el resultado de la prueba de hipótesis. Cuando el resultado es significativo se muestran los sitios detectados bajo SP o neutros así como los parámetros estimados de su proporción y valores de ω .

Hipótesis	Valores de Verosimilitud MA / MB	$2\Delta\ln$		P calculada	Sitios bajo SP & parámetros estimados (PP)
		MA vs M1 (nula)/ MAconservative*	vs M1 (nula) / MB vs M3 k=2 (nula)		
Genes MADS-box tipo B					
B-H1 (rama gimnospermas)	-37143.765671 / -36517.351161	0.66918 / 0.56116 / 1.15554		$P=0.7156 / P=0.7553 / P=0.5611$	
B-H9 (clado gimnospermas)	-37099.348727 / -36514.213087	89.503 / 89.503 / 7.4317		$P<0.0001$, rechazo hip. nula/ $P<0.0001$, rechazo hip. nula/ $P<0.0243$ rechazo hip. nula, 98% confianza	ninguno
B-H3 (rama Bsister)	-37143.734546 / -36517.928903	0.73144 / 0.35164 / 0.64806		$P=0.6936 / P=0.8387 / P=0.7232$	
B-H2 (rama AP3-PI-Bsister)	-37144.100266 / -36517.928936	0 / 0 / 0		$P=1 / P=1 / P=1$	
B-H4 (rama AP3-PI)	-37143.69819 / -36517.928936	0.80414 / 0.2605 / 0		$P=0.6689 / P=0.8778 / P=1$	
B-H5 (rama PI)	-37135.64899 / -36517.148328	16.9025 / 9.715 / 1.561		$P=0.0002$, rechazo hip. nula/ $P=0.0077$, rechazo hip. nula/ $P=0.4581$	Modelo A: 91 H (0.867), 100 R (0.948), 115 L (0.893), 119 I (0.885), 142 H (0.982), 165 D (0.992), 225 A (0.973), 255 * (0.921), 297 H (0.800) $p_{25} = 0.12631$, $p_{25} = 0.09249$; $\omega_{25} = 6.24653$, $\omega_{25} = 6.24653$
B-H6 (rama AP3)	-37137.801316 / -36515.203888	12.5979 / 5.089 / 5.450		$P=0.0018$, rechazo hip. nula/ $P=0.0785$, rechazo hip. nula, 93% confianza / $P=0.0655$, rechazo hip. nula, 94% confianza	Modelo A: 102K (0.862), 119I (0.934), 120R (0.961), 121Q (0.927), 130G (0.989) $p_{25} = 0.07385$, $p_{25} = 0.05507$, $\omega_{25} = 8.53434$, $\omega_{25} = 8.53434$ Model B: 119I (0.896), 120R (0.884), 122R (0.962) $p_{25} = 0.03439$, $p_{25} = 0.05140$; $\omega_{25} = 3.98661$, $\omega_{25} = 3.98661$
B-H8 (clado PI)	-37016.315092 / -36431.389952	255.570 / 255.570 / 173.077		$P<0.0001$, rechazo hip. nula/ $P<0.0001$, rechazo hip. nula/ $P<0.0001$, rechazo hip. nula	Modelo A: 36T (0.851), 42K (0.931), 51T (1.000), 55H (0.995), 70T (1.000), 72K (1.000), 80S (0.869), 141Q (0.991), 142H (0.835), 148T (1.000), 152E (0.995), 157V (1.000), 158I (1.000), 173N (1.000), 194D (0.935), 219D (0.953), 225A (1.000), 286H (0.962) $p_{25} = 0.14184$, $p_{25} = 0.08568$; $\omega_{25} = 1.00$, $\omega_{25} = 1.00$
B-H7 (clado AP3)	-37034.990261 / -36385.989919	218.22 / 218.22 / 263.878		$P<0.0001$, rechazo hip. nula/ $P<0.0001$, rechazo hip. nula/ $P<0.0001$, rechazo hip. nula	Modelo A: 33H (0.999), 54F (0.950), 82V (0.974), 104T (1.000), 111T (1.000), 144Q (0.983), 145D (1.000), 167C (1.000), 174L (1.000), 175E (0.980), 177R (1.000), 179G (0.976) $p_{25} = 0.14675$, $p_{25} = 0.08515$; $\omega_{25} = 1.00$, $\omega_{25} = 1.00$
Genes tipo APETALA3					
AP3-H1 (rama angio-AP3)	-29466.927302 / -28929.892464	7.368 / 3.837 / 0		$P=0.0251$, rechazo hip. nula, 98% confianza / $P=0.1468 / P=1$	Modelo A: 153N (0.939), 195Y (0.813), 197V (0.973) $p_{25} = 0.08382$, $p_{25} = 0.03254$; $\omega_{25} = 16.10785$, $\omega_{25} = 16.10785$
AP3-H2 (rama euAP3-TM6)	-29469.262162 / -28929.778895	2.698 / 1.607 / 0.227		$P=0.2594 / P=0.4477 / P=0.8927$	
AP3-H3 (clado euAP3-TM6)	-29363.283866 / -28859.301144	214.6551 / 214.655 / 141.182		$P<0.0001$, rechazo hip. nula / $P<0.0001$, rechazo hip. nula / $P<0.0001$, rechazo hip. nul	Modelo A: 99G (0.874), 110T (0.999), 123R (0.999), 128S (0.967), 135N (1.000), 136S (0.986), 177R (0.964), 178G (1.000), 188K (1.000), 205F (0.975), 230P (1.000), 260R (0.984), 282P (0.923), 283S (0.999), 319R (0.860) $p_{25} = 0.15122$, $p_{25} = 0.03304$; $\omega_{25} = 1.00$, $\omega_{25} = 1.00$
AP3-H5 (rama TM6)	-29470.611332 / -28929.892464	0.00016 / 0.00016 / 0		$P=0.9999 / P=0.9999 / P=1$	
AP3-H4 (rama euAP3)	-29457.045013 / -28905.64694	27.1328 / 1.607 / 48.4910		$P<0.0001$, rechazo hip. nula/ $P=0.4477 / P<0.0001$, rechazo hip. nul	Model A: 75I (0.842), 153N (0.969), 167D (0.969), 182N (0.942), 196H (0.913), 316H (0.986), 317E (0.978), 318L (0.989), 319R (0.909), 321G (0.989), 322* (0.828) $p_{25} = 0.11463$, $p_{25} = 0.04319$; $\omega_{25} = 1.00$, $\omega_{25} = 6.28$ Model B: 49* (0.843), 75I (0.940), 92L (0.839), 153N (0.846), 167D (0.997), 196H (0.996), 199N (0.865), 315T (0.866), 316H (0.999), 317E (0.998), 318L (0.999), 319R (0.976), 321G (0.948) $p_{25} = 0.04897$, $p_{25} = 0.09612$; $\omega_{25} = 6.8097$, $\omega_{25} = 6.8097$
AP3-H7 (clado TM6)	-29423.59212 / -28921.886755	94.038 / 94.038 / 16.011		$P<0.0001$, rechazo hip. nula/ $P<0.0001$, rechazo hip. nula/ $P=0.0003$, rechazo hip. nul	Modelo A: 99G (0.976), 126H (0.948), 164V (0.981), 173E (0.953), 177R (0.988), 178G (0.998), 184V (1.000), 188K (0.945), 221S (0.801), 230P (0.999), 233* (0.855), 257F (0.815), 279R (0.851), 297* (0.989), 311V (0.848), 312Y (0.995), 313S (0.993) $p_{25} = 0.15257$, $p_{25} = 0.05124$; $\omega_{25} = 1.00$, $\omega_{25} = 1.00$
AP3-H6 (clado euAP3)	-29385.703861 / -28880.656293	169.815 / 169.815 / 98.4723		$P<0.0001$, rechazo hip. nula/ $P<0.0001$, rechazo hip. nula/ $P<0.0001$, rechazo hip. nul	Modelo A: 110T (1.000), 120I (0.985), 123R (0.887), 135N (1.000), 157R (0.950), 178G (0.991), 188K (1.000), 197V (0.986), 205F (0.924), 218E (0.931), 260R (0.939), 261N (0.974), 264* (0.849), 276F (0.985), 283S (0.992), 319R (0.985) $p_{25} = 0.13419$, $p_{25} = 0.03295$; $\omega_{25} = 1.00$, $\omega_{25} = 1.00$
Genes tipo PISTILLATA					
PI-H1 (rama angio-PI)	-19343.071419 / -19113.681345	4.56 / 4.318 / 11.58		$P=0.1022 / P=0.1154 / P=0.0030$, rechazo hip. nul	Modelo B: 138N (0.967), 165A (0.956), 192A (0.855), 207E (0.886), 320H (0.806) $p_{25} = 0.31109$, $p_{25} = 0.04466$; $\omega_{25} = 6.32576$, $\omega_{25} = 6.32576$
PI-H2 (rama core-eudicots)	-19344.672084 / -19117.320476	1.363 / 1.3636 / 4.306		$P=0.5058 / P=0.5057 / P=0.1161$	
PI-H3 (clado core-eudicots)	-19303.315267 / -19057.271589	84.077 / 84.077 / 124.4046		$P<0.0001$, rechazo hip. nula/ $P<0.0001$, rechazo hip. nula/ $P<0.0001$, rechazo hip. nul	Model A: 125H (1.000), 132S (0.897), 203N (0.864), 243K (1.000), 286R (0.991), 291D (0.804), 299N (0.994), 317D (0.949), 318N (0.937), 321Q (1.000), 322V (0.999), 323 P (0.877) $p_{25} = 0.12306$, $p_{25} = 0.04450$; $\omega_{25} = 1.00$, $\omega_{25} = 1.00$

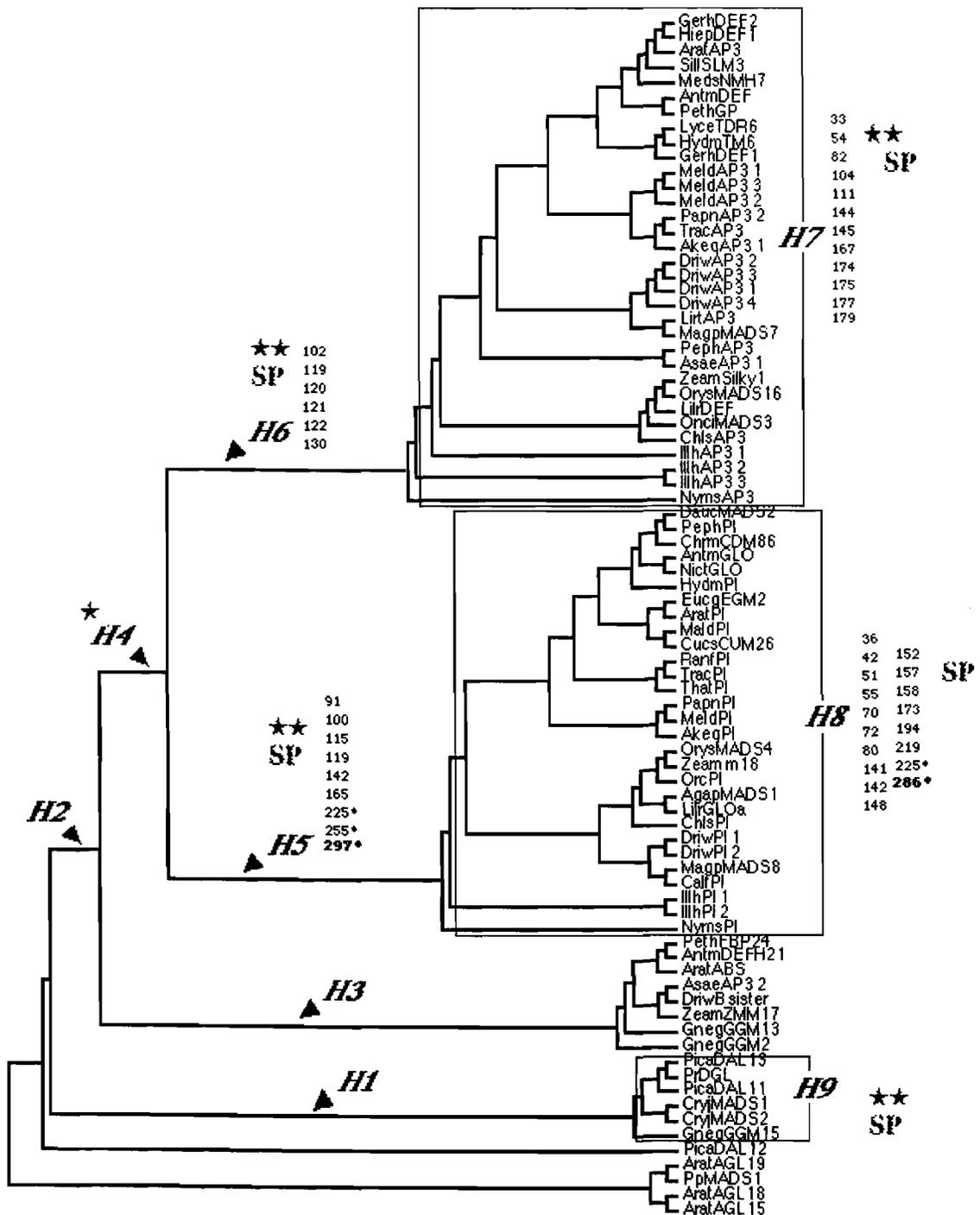


Figura 12. Cladograma de la topología obtenida por métodos Bayesianos para los genes MADS Tipo B que muestra las diferentes hipótesis (H1 a 9) probadas para diferencias en las tasas de verosimilitud y selección positiva en linajes particulares. ★ y ★★ es el nivel de significancia estadística que soporta diferencias en las tasas ω en ese linaje en particular, considerada con base en los resultados de la tercera y cuarta PTV con el modelo D que se utilizan para evaluar dicha significancia. SP indica que hubo soporte estadístico para la presencia de selección positiva en ese linaje. En caso de encontrados, se muestran los sitios detectados bajo selección positiva con $pp > 0.8$ para cada hipótesis. Seguidos de un * los sitios que se ubican en regiones del alineamiento muy ambiguas debido a el elevado número de inserciones y deleciones. En negritas los sitios que se ubican en el carboxilo terminal de la proteína. Los resultados detallados de las PTV así como los valores de los parámetros calculados se muestran en las tablas 8 a 16.

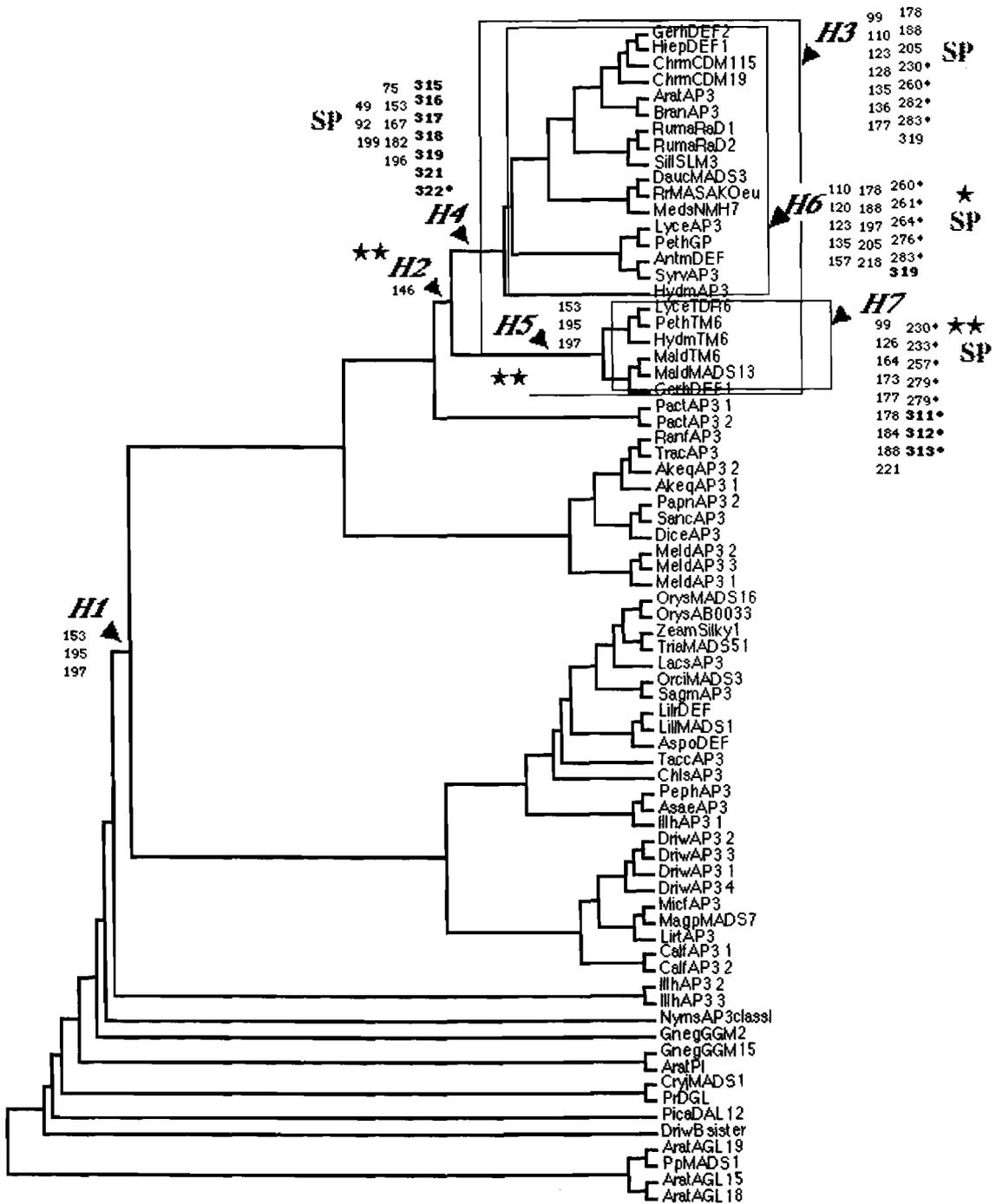


Figura 13. Cladograma de la topología obtenida por métodos Bayesianos para los genes tipo APETALA3 que muestra las diferentes hipótesis (H1 a 7) probadas para diferencias en las tasas de verosimilitud y selección positiva en linajes particulares. ★ y ★★ es el nivel de significancia estadística que soporta diferencias en las tasas ω en ese linaje en particular, considerada con base en los resultados de la tercera y cuarta PTV con el modelo D que se utilizan para evaluar dicha significancia. SP indica que hubo soporte estadístico para la presencia de selección positiva en ese linaje. En caso de encontrados, se muestran los sitios detectados bajo selección positiva con $pp > 0.8$ para cada hipótesis. Seguidos de un * los sitios que se ubican en regiones del alineamiento muy ambiguas debido a el elevado número de inserciones y deleciones. En negritas los sitios que se ubican en el carboxilo terminal de la proteína. Los resultados detallados de las PTV así como los valores de los parámetros calculados se muestran en las tablas 17 a 23.

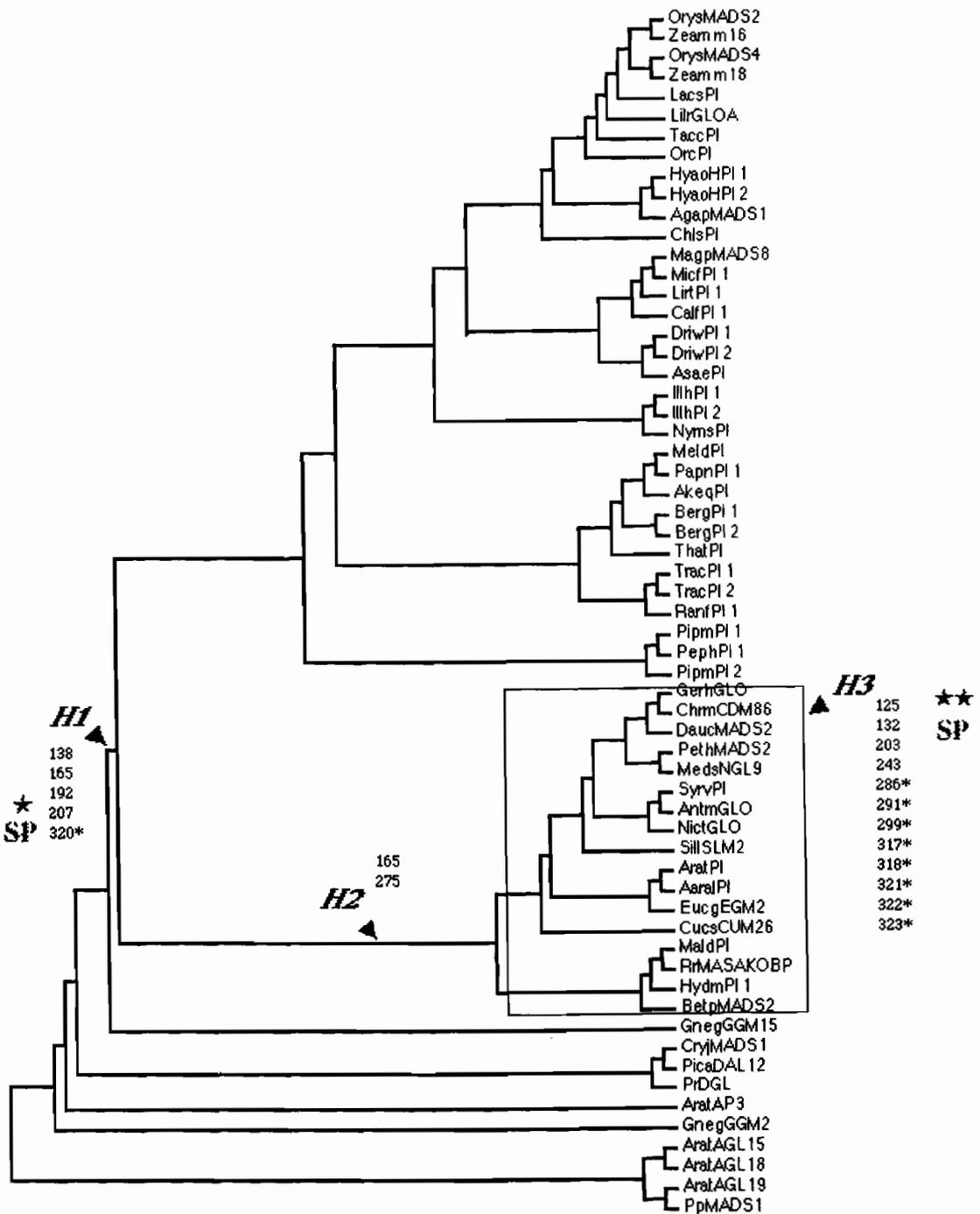


Figura 14. Cladograma de la topología obtenida por métodos Bayesianos para los genes tipo PISTILLATA que muestra las diferentes hipótesis (H1 a 3) probadas para diferencias en las tasas de verosimilitud y selección positiva en linajes particulares. ★ y ★★ es el nivel de significancia estadística que soporta diferencias en las tasas ω en ese linaje en particular, considerada con base en los resultados de la tercera y cuarta PTV con el modelo D que se utilizan para evaluar dicha significancia. SP indica que hubo soporte estadístico para la presencia de selección positiva en ese linaje. En caso de encontrados, se muestran los sitios detectados bajo selección positiva con $pp > 0.8$ para cada hipótesis. Seguidos de un * los sitios que se ubican en regiones del alineamiento muy ambiguas debido a el elevado número de inserciones y deleciones. En negritas los sitios que se ubican en el carboxilo terminal de la proteína. Los resultados detallados de las PTV así como los valores de los parámetros calculados se muestran en las tablas 24 a 26

Resultados Modelos A, B y D, genes MADS Tipo B

A pesar de que los resultados obtenidos con los modelos sitio por sitio indicaron que los genes MADS tipo B y los parálogos AP3 y PI se encontraban bajo fuertes regímenes de selección purificadora, o bien con muchos sitios neutros, las PTV realizadas con los valores de verosimilitud obtenidos con los modelos A, B y D nos permitieron, además de detectar diferencias en las tasas de sustitución en diferentes momentos de la evolución de los MADS tipo B, detectar SP en una pequeña proporción de sitios en algunos de los linajes seleccionados. Estos resultados pueden servir para describir mejor el proceso de diversificación funcional en los genes tipo B.

Para las gimnospermas se detectaron DTS significativas durante su diversificación, indicada por la hipótesis B-H9, pero no así en la rama que da lugar al clado, justo después que se separaron de las angiospermas. Lo mismo ocurrió en la rama que da lugar a los genes B-sister, indicada como hipótesis B-H3. En el clado de las gimnospermas, las PTV favorecen los modelos que aceptan SP, sin embargo ninguno de los parámetros ω calculados tienen valores > 1 .

Las tasas de sustitución permanecen homogéneas en la rama que da lugar a los B-sister (hipótesis B-H3) y en la rama que da lugar a los genes B de angiospermas y B sister (hipótesis B-H2) así como en la rama que da lugar a los genes B de angiospermas. Los resultados para esta hipótesis (B-H4), en la rama que da lugar al clado AP3-PI, indicaron una señal pobre para detectar tanto SP como DTS. A pesar de que se identificó un porcentaje de sitios bajo selección positiva en la rama (5% de sitios con $\omega=6.69$ con el modelo A y 54% de sitios con $\omega=1.8$ con el modelo D $K=3$) los resultados de las comparaciones por PTV no indicaron que los modelos A, B y D aporten una mejora significativa en comparación con los modelos nulos que implican homogeneidad. Sin embargo, ya en la rama que da a los genes tipo PI y la rama que da a los AP3 la señal se hace más fuerte, y las PTV en su mayoría nos permitieron rechazar las hipótesis nulas de homogeneidad de tasas. Esto se observa en los resultados para las hipótesis B-H5 para la rama que da a los PI y la B-H6 para la rama que da a los AP3. En el caso de la rama que da a los PI con el modelo D se detectaron una gran cantidad de sitios bajo SP (54% o 59.9% bajo una ω de 1.26 con el modelo D $k=2$ y $k=3$ respectivamente, 12%+9% de sitios bajo una $\omega=6.2$ y 3%+5% de sitios bajo una $\omega=2.58$ con el modelo B). La señal de SP también es fuerte en el caso de la rama que da a los AP3, pero aquí el modelo D no detectó sitios bajo SP. En esta rama, con el modelo A se encontraron 7%+5% de sitios bajos una $\omega=8.5$, y con el modelo B detectamos 3%+5% de sitios bajo una $\omega=3.9$. Esto indica que los procesos de diversificación funcional ocurrieron muy pronto después de la duplicación que dio lugar a los parálogos AP3 y PI; pero solo un pequeño porcentaje de sitios cambiaron por SP.

Las hipótesis B-H7 y B-H8 en el clado de los AP3 y PI respectivamente permiten evaluar si estas DTS y SP detectadas tempranamente en la evolución de los clados AP3 y PI, justo después de su duplicación, se mantuvo posteriormente a su diversificación al interior de cada clado. Los resultados de las PTV indicaron que hay DTS significativas en el clado de los genes PI, más no en los AP3. Sin embargo, las PTV para detectar SP resultaron todas en un rechazo de las hipótesis nulas, lo cual puede indicar que los procesos de diversificación funcional en estos clados esta dada por SP en un pequeño porcentaje de sitios. A pesar de se rechazaron los modelos nulos, en ningún caso obtuvimos valores de ω mayores a 1, pero en ambos casos se detectaron sitios que si bien pueden no pertenecer a categorías de sitios bajo selección positiva, si se pueden interpretar como pertenecientes a categorías de sitios bajo presiones selectivas relajadas, por tanto están cercanos a la neutralidad (con el modelo A, 14%+8% de sitios bajo $\omega=1$ en el clado de los AP3 y el de los PI) Es difícil dar una interpretación adecuada a estos sitios.

Resultados Modelos A, B y D, genes MADS Tipo APETALA3

Las hipótesis realizadas para los genes tipo AP3 tienen el objetivo primordial de evaluar el papel de la duplicación euAP3-TM6 en la diversificación funcional de estos genes, presentes en el clado de las eudicotiledóneas superiores. La hipótesis AP3-H1, señala una rama profunda, cerca de la base de los AP3, después de su separación de las angiospermas basales Nymphaea e Illiciaceae, y sirve como control a las hipótesis planteadas posteriormente. En este caso (AP3-H1) no pudimos rechazar con ninguna PTV los modelos nulos. La hipótesis AP3-H2 permite evaluar una diferenciación temprana de los genes B en las eudicotiledóneas superiores, al señalar la rama que les da lugar. En este caso conseguimos detectar DTS significativas, pero no así rechazar las hipótesis nulas para SP. El modelo D, tanto con K=3 como con K=2 categorías indican un gran número de sitios bajo SP (45% bajo una $\omega=2.56$ y 66% con una $\omega=2.02$ con el modelo D K=2 y K=3 respectivamente), pero con el modelo A solo se detectó un 1%+3% de sitios bajo una $\omega=5.7$ y con el modelo B un 1%+2% de sitios bajo una $\omega=2.81$. Esto puede indicar la fuerte influencia de la SP tempranamente en la evolución de estos genes.

La hipótesis AP3-H3 prueba DTS y SP en los genes AP3 de eudicotiledóneas superiores, al señalar todas las ramas en el clado. En este caso no detectamos DTS con el modelo D, pero si se rechazaron las hipótesis nulas con los modelos A y B. A pesar de esto ningún parámetro ω en estos modelos indica la presencia de selección positiva, siendo que los sitios encontrados deben interpretarse como neutros o bajo presiones selectivas relajadas.

Las hipótesis AP3-H4 a H7 se dirigen hacia las diferencias particulares de los cladogramas euAP3 y TM6. La hipótesis AP3-H4 y AP3-H5 se refieren a posibles DTS y SP como indicadores de diferenciación funcional temprana en estos genes, justo después de su duplicación, en la rama que da lugar a los euAP3 y TM6 respectivamente. En el caso de la rama que da lugar a los euAP3, las PTV no permitieron rechazar la hipótesis de DTS con el modelo D, y solo dos de tres PTV permitieron aceptar SP significativamente. A pesar de esto, tanto con el modelo A como con el modelo B encontramos sitios bajo selección positiva (4% bajo $\omega=6.28$ con el modelo A y 9%+4% bajo $\omega=6.8$ con el modelo B), lo cual da confianza en los sitios encontrados. En la rama que da lugar a los genes TM6 encontramos el caso contrario, por que con el modelo D pudimos rechazar la hipótesis nula pero no pudimos encontrar SP con el modelo A ni con el B.

Por último, y para evaluar DTS y PS persistente en etapas posteriores a la diversificación de estos genes planteamos las hipótesis AP3-H6 en el clado de los euAP3 y AP3-H7 en el clado de los TM6. Aunque en ambos casos logramos rechazar las hipótesis nulas para DTS y PS, y encontramos un gran número de sitios con altos valores de PP, ningún parámetro ω calculado nos indica presencia de SP, por lo que estos sitios encontrados también deben interpretarse como neutros o bajo presiones selectivas relajadas.

Resultados Modelos A, B y D, genes MADS Tipo PISTILLATA

Los resultados para las pruebas con los genes tipo PI nos llamaron mucho la atención, porque dada la conservación observada en sus alineamientos como la evidencia experimental que la soporta, no esperábamos encontrar señales de DTS o SP.

Las hipótesis probadas se diseñaron más bien para su comparación con los resultados obtenidos para los genes AP3. La primera hipótesis, PI-H1 indica la rama que da lugar a estos genes en angiospermas, y a pesar de que logramos rechazar las hipótesis nulas solo parcialmente, con el modelo A y B detectamos una proporción de sitios bajo SP (5%+3% con $\omega=1.66$ con el modelo A

y 3%+4% con $\omega=6.32$ con el modelo B). La hipótesis PI-H2 se refiere a una temprana presencia de DTS y SP en la rama que da lugar a los genes tipo PI de las core-eudicots, pero en este caso no pudimos rechazar ninguna hipótesis nula, a pesar de que con el modelo B encontramos un 1%+2% de sitios bajo SP. La última hipótesis, PI-H3, se concentra en el clado de los genes tipo PI de las eudicotiledoneas superiores. Aquí las PTV nos permitieron rechazar la hipótesis nula en todos los casos, indicando la presencia de DTS y SP de manera significativa en el clado, pero ningún parámetro ω indicó selección positiva, aunque si un cambio importante en las presiones selectivas con el modelo B (un 6% de sitios cambian de $\omega=0.03$ a $\omega=0.93$ en las core-eudicots, y un 11% cambian de $\omega=0.2$ a $\omega=0.93$) y con el modelo A (12% de sitios cambian de tener $\omega=0.13$ a $\omega=1.00$ en este clado).

IV. Estimaciones de tiempos de divergencia

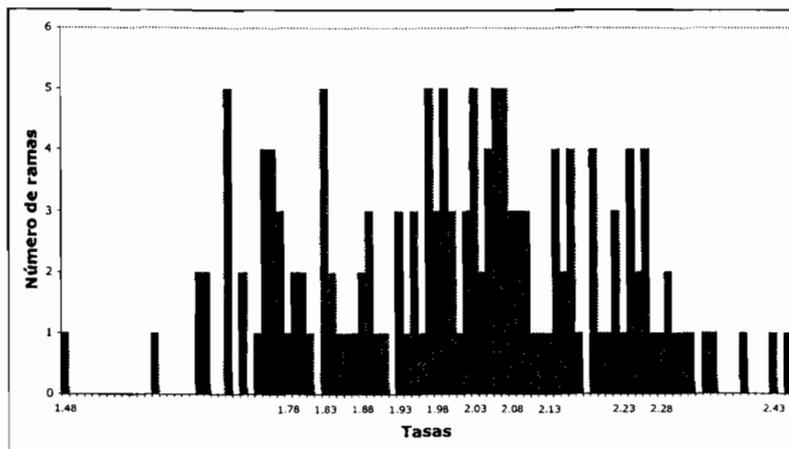
Las estimaciones de las fechas para los eventos de duplicación en los genes MADS tipo B y en los tipo *APETALA3* se llevaron a cabo también con el software PAML, que implementa el algoritmo llamado 'A Heuristic Rate Smoothing' (AHRs) descrito en Yang, Z. 2004, en un contexto de máxima verosimilitud, para considerar diferencias en las tasas de sustitución en los diferentes linajes de la filogenia al estimar las fechas de duplicaciones.

Con el objetivo principal de fechar la duplicación AP3-PI en la filogenia de todos los MADS tipo B se utilizó como fecha de calibración los 121 ma para la edad de las eudicotiledóneas, con base en granos de polen tricolpados fósiles descritos en Doyle y Hotton, 1991 según datos tomados de Magallón y Sanderson, 2001. En el caso de los MADS tipo B se distinguieron 4 grupos de ramas según sus diferentes tasas de sustitución (ver tabla 27).

Tabla 27. Resultado de la clasificación de ramas del árbol en grupos siguiendo diferencias en sus tasas de sustitución según el algoritmo "A heuristic Rate Smoothing" implementado en PAML para los genes MADS-Box tipo B.

Resultados análisis con AHRs para los MADS tipo B			
Grupo	Número de ramas en ese grupo	Tasa de sustitución	Tasa de sustitución media
0	8	<1.691833	1.633865
1	25	<1.830907	1.764163
2	57	<2.064028	1.965562
3	62	<2.064028	2.193414

La gráfica 1 muestra la distribución de las tasas de sustitución de las ramas estimadas por máxima verosimilitud. Una vez hecha la clasificación de las ramas en grupos dependiendo de estas tasas de sustitución obtenidas por máxima verosimilitud, se estiman los tiempos de divergencia y las tasas para los grupos de ramas (Yang, Z. 2004). La topología que resulto siguiendo esta clasificación, así como las principales fechas de divergencia estimadas se muestran en la topología de figura 15a, y el árbol de tasas se muestra en la figura 16b. Las fechas estimadas para cada clado se muestran en la tabla 28.



Gráfica 1. Distribución de las tasas de sustitución estimadas por máxima verosimilitud para cada rama utilizando PAML para los genes MADS-Box tipo B, tomando como restricción la topología (sin longitudes de ramas) obtenida en los análisis filogenéticos hechos con métodos Bayesianos descritos anteriormente.

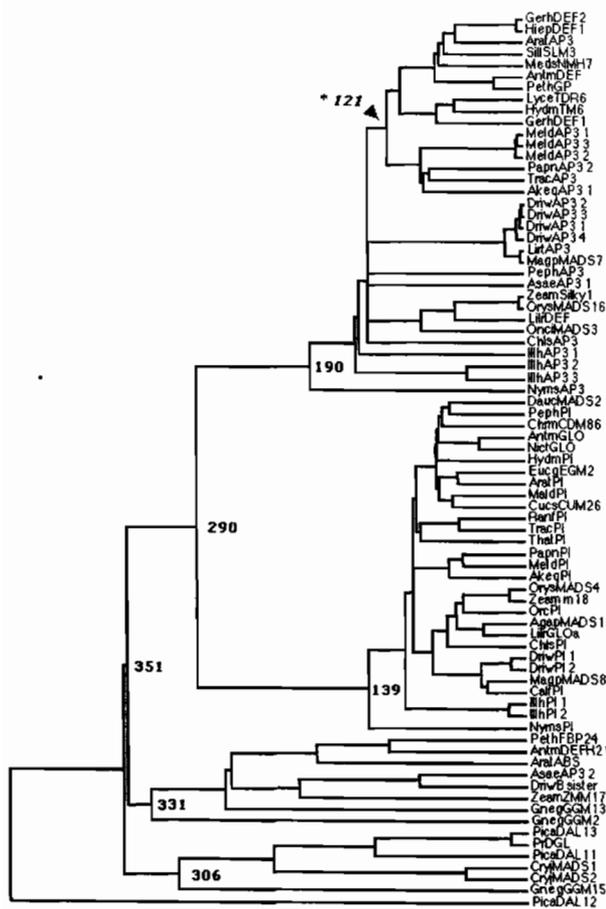


Figura 15a. Topología resultante del análisis de tasas de sustitución y estimación de fechas de divergencia con PAML para los genes MADS-Box Tipo B. Se muestran las fechas estimadas para algunas duplicaciones (en millones de años). *121 es la fecha utilizada para calibración. La longitud de las ramas es equivalente a tiempo.

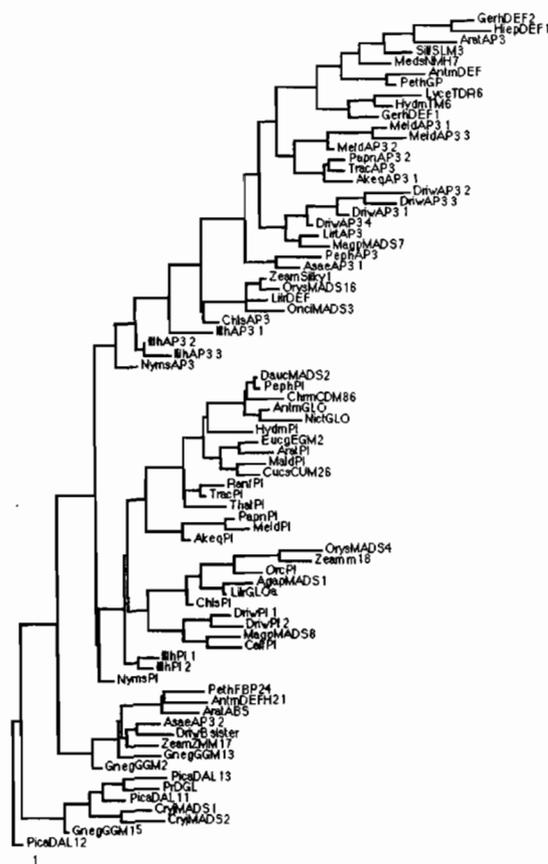


Figura 15b. Ratograma resultante de la estimación de tasas de sustitución con PAML para los genes MADS-Box Tipo B. La longitud de las ramas es equivalente a tasa de sustitución. Con base en las estimaciones de estas tasas se construyó la clasificación de ramas siguiendo el algoritmo AHRS

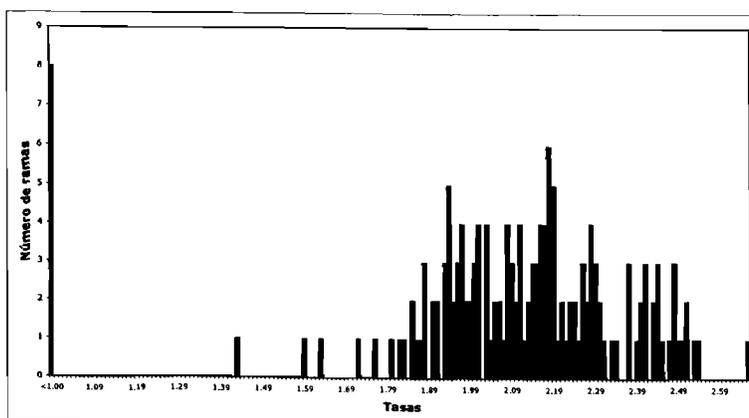
Tabla 28. Fechas de duplicación o cladogénesis estimadas con PAML y comparación con registro fósil.

Clado o duplicación genes tipo B	Fecha estimada con PAML (millones de años)	Grupo o evento en la evolución de las plantas	Fecha según registro fósil (millones de años)
Duplicación (B-gimnospermas) – (B-sister & B-angiospermas)	354.3	Origen de las plantas con semilla a partir de las pro gimnospermas	350 (Stewart y Rothwell, 1993 citados en Sundström, 1999)
Duplicación (B-sister) – (B-angiospermas)	351.1	Origen de las plantas con semilla	290-309.2 ma (Mapes y Rothwell, 1984, 1991. citados en Kim, et al. 2004)
Duplicación AP3-PI	290		
Clado B-sister	331.3		
Inicio diversificación AP3	190.2		
Inicio diversificación clado PI	139		

Posteriormente se hizo el análisis de estimación de fechas de divergencia siguiendo el mismo método pero con los genes MADS tipo APETALA3, con el principal objetivo de estimar la fecha de la duplicación que dio lugar a los genes tipo euAP3 y TM6 en las eudicotiledóneas superiores. La fecha utilizada para la calibración fue de 132 ma en el origen de las angiospermas por granos de polen fósiles descritos en Brenner, 1996 según datos que tomamos de Magallón y Sanderson, 2001. La tabla 29 muestra los grupos de ramas y sus tasas estimadas siguiendo el algoritmo AHRS; en este caso también se distinguieron 4 grupos con tasas que van de 0.001 a 2.6555. La gráfica 2 muestra la distribución de las tasas con respecto al número de ramas. Cabe destacar que también se probó como fecha de calibración alternativa la utilizada en el caso anterior, de 121 ma para las eudicotiledóneas, sin embargo, con esta calibración, las fechas estimadas para los clados son incongruentes con las fechas del registro fósil para los clados de las plantas de las que provienen las secuencias de ortólogos AP3, por lo que esos resultados fueron descartados. La figura 16a muestra la topología con las longitudes de las ramas clasificadas según los 4 grupos así como las principales fechaciones estimadas para los genes APETALA3 y la tabla 30 muestra las fechas resultantes para los principales clados. La figura 16b muestra el árbol de tasas calculado para estos genes.

Tabla 29. Resultado de la clasificación de ramas del árbol en grupos siguiendo diferencias en sus tasas de sustitución según el algoritmo “A heuristic Rate Smoothing” implementado en PAML para los genes MADS-Box tipo AP3.

Resultados análisis con AHRS para los MADS tipo B			
Grupo	Número de ramas en ese grupo	Tasa de sustitución	Tasa de sustitución media
0	5	< 0.564597	0.092909
1	3	< 0.945724	0.817800
2	1	< 1.584583	1.430139
3	149	< 2.655463	2.140649



Gráfica 2. Distribución de las tasas de sustitución estimadas por máxima verosimilitud para cada rama utilizando PAML para los genes tipo APETALA3, tomando como restricción la topología (sin longitudes de ramas) obtenida en los análisis filogenéticos hechos con métodos Bayesianos descritos anteriormente.

Discusión

Relaciones filogenéticas de genes MADS-box tipo B en plantas con semilla.

El objetivo principal de esta tesis fue hacer un estudio completo de evolución molecular de los genes MADS tipo B para entender mejor sus procesos de diversificación. La comparación de las tasas de sustitución de clados particulares que corresponden a grupos taxonómicos de angiospermas y en clados producto de duplicaciones es útil en la evaluación de los procesos de diferenciación funcional génica, y el análisis comparativo de las tasas de sustitución sinónimas y no sinónimas puede ayudar a evaluar que presiones selectivas están dirigiendo la diferenciación funcional. Para poder hacer estas inferencias es necesario contar con filogenias resueltas, por ello como primer objetivo obtuvimos hipótesis acerca de las relaciones filogenéticas para los genes B disponibles mediante métodos Bayesianos. Los árboles fueron usados en los análisis de heterogeneidad de tasas de sustitución y análisis de sitios bajo SP, pero además nos han servido para reevaluar las discusiones acerca de las relaciones filogenéticas de los genes B en plantas con semilla y la evolución de su función. Así mismo aprovechamos nuestros datos para calibrar un reloj molecular y con él proveer estimaciones nuevas de las fechas de divergencia de los grupos principales de las secuencias MADS tipo B y su correlación con eventos importantes durante la evolución de las plantas.

Los grupos existentes de plantas con semilla son las angiospermas, con aproximadamente 250,000 especies, y las gimnospermas, que incluyen a las coníferas, con 550 especies, las cicadas, con 150 especies, las ginknoales, que incluye solo a *Ginkgo biloba* y las Gnetales, con 70 especies distribuidas en tres grupos distintos, *Ephedra*, *Welwitschia* y *Gnetum* (Raven, Evert y Eichorn, 1999). La determinación de las relaciones entre estos grupos es complicada porque los linajes radiaron muy rápido y hace mucho tiempo: las plantas con semilla aparecieron en el Devónico Tardío, hace aproximadamente 370 millones de años (ma), y al menos tres de los cinco linajes probablemente se separaron durante el Carbonífero Tardío, hace 290-320 ma (Donoghue y Doyle, 2000). Además, una complicación adicional es el hecho de que los grupos vivientes están separados por grandes “gaps” o saltos morfológicos.

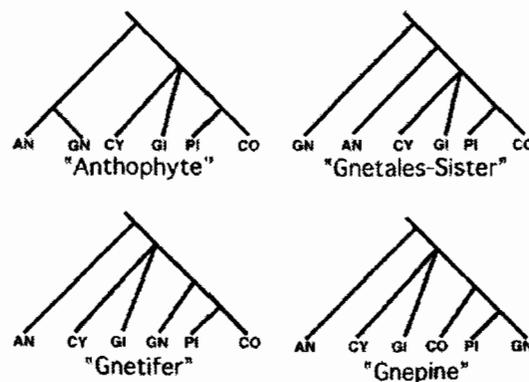


Figura 17. Las cuatro principales hipótesis acerca de las relaciones existentes entre plantas con semilla. AN=angiospermas, GN=gnetales, CY=cicadas, GI=*Ginkgo*, PI=pinaceae, CO=coníferas no-pinaceas. Tomado de Burleigh y Mathews, 2004.

A inicios de los años 1900's se pensaba que las Gnetales eran el grupo hermano de las angiospermas, y se interpretaba a sus órganos reproductivos como estructuras simples, unisexuales y parecidas a flores como los precursores de las flores de angiospermas (Donoghue y Doyle, 2000). Para mediados de los 80's muchos análisis filogenéticos basados en datos morfológicos dieron soporte a esta idea, y se agrupó a las Gnetales, las angiospermas y el grupo fósil Bennetitales en el clado llamado 'anthophyta'; hipótesis que tuvo mucho auge y se volvió casi un dogma (Doyle y Donoghue, 1986; Crane 1985, Donoghue y Doyle, 2000). Con los análisis filogenéticos basados en datos moleculares se puso en duda este dogma ya que se ubicó a las Gnetales como hermanas de las coníferas e incluso a las gimnospermas como un clado (Chaw et al; 2000), y el grupo basal podría estar conformado por las cicadas, enseguida *Gynkgo* y todas las coníferas excepto Pinnaceae como hermanas del clado Gnetales+Pinnaceae (Chaw, et al. 2000) (ver Figura 17); lo cual requirió una reinterpretación de la evolución de los caracteres reproductivos (Donoghue y Doyle, 2000). En este caso las gnetales debieran verse como coníferas extremadamente divergentes, y las coincidencias morfológicas con angiospermas como homoplasias. En 1999 se dieron a conocer los resultados de un análisis filogenético de genes MADS-Box de Gnetales en relación con los de angiospermas (Winter et al. 1999). En este estudio el único grupo de Gnetales que se muestreó fue *Gnetum*, y no reportaron secuencias para cicadas o para *Ginkgo*. De los 13 genes que reportaron para gnetales, cinco cayeron en clados monofiléticos con secuencias de coníferas y una relación de grupo hermano con respecto a secuencias de angiospermas. Entre éstos destacan ortólogos de los genes florales de función B y C. Los autores interpretaron estos datos como evidencia de que las gnetofitas están más cercanamente relacionadas a las coníferas que a las angiospermas (Winter, et al. 1999).

Nuestros análisis de los genes MADS tipo B nos permiten retomar la discusión acerca de las relaciones filogenéticas entre las gimnospermas y angiospermas. En el caso de la filogenia de todos los MADS tipo B encontramos que el gen *GGM15* de *Gnetum* se ubicó como hermano de las demás secuencias de gimnospermas, básicamente coníferas (*PicaDAL11* y *PicaDAL13* de *Picea abies*, *PrDGL* de *Pinus radiata*, y *CryjMADS1* y *CryjMADS2* de *Cryptomeria japonica*), como hermanos de los 'B-sister' y los Tipo B de angiospermas, sin embargo los valores de apoyo por probabilidad posterior para este clado (0.29PP) así como de su posición con respecto a los genes tipo B-sister y tipo B de angiospermas (0.47PP) son muy bajas. Dado nuestro análisis filogenético y los anteriores reportados en la bibliografía, no es claro que ésta posición sea la correcta para los tipo B de gimnospermas, sin embargo es la posición que da lugar a interpretaciones biológicas más consistentes dados los conocimientos sobre las relaciones evolutivas de angiospermas y gimnospermas.

Para evaluar más detalladamente su posición con respecto a los genes AP3 y PI de angiospermas, incluimos en el análisis de éstos genes por separado a las secuencias tipo B de gimnospermas. En el caso de los genes tipo PI, el gen *GGM15* de *Gnetum gnemon* se ubicó como hermano inmediato de los PI, en un clado con bastante buen apoyo por PP (0.97) siendo que los genes de coníferas formaron un clado monofilético. En el caso de la filogenia para los AP3, encontramos a *GGM2*, también de *Gnetum gnemon*, como secuencia basal de los genes AP3 de angiospermas, siendo que los demás genes de coníferas y gnetales (*GGM15*) formaron un grado basal con el parálogo *PISTILLATA* de *Arabidopsis* embebido en él.

En un análisis filogenético básico por métodos de distancia con genes MADS Box incluyendo la secuencia *GGM2* de gnetales, se observó que éstas se agrupaban junto con los genes del clado AP3-PI de angiospermas, pero con una posición no resuelta (Mouradov et al. 1999). Durante las etapas iniciales del presente trabajo, realizamos análisis filogenéticos exploratorios

preliminares con métodos de distancia y parsimonia, y encontramos que la posición de los genes B de gimnospermas, las cuales se agrupaban casi siempre en un solo clado, era variable: a veces se agrupaban como hermanas del clado AP3-PI, a veces como hermanas de los PI y a veces como hermanas de los AP3. Nos dimos cuenta que esta inconsistencia respondía a la ambigüedad del alineamiento de las secuencias de gimnospermas, sobre todo en la región del carboxilo terminal, ya que podía ser alineada, con igual incertidumbre, ya sea como C-terminal tipo paleoAP3 o como tipo PI con algunos aminoácidos extras en el C-terminal. Hicimos un análisis filogenético siguiendo la misma metodología descrita en la sección de Métodos, con el mismo alineamiento de los MADS Tipo B pero quitando la región del C-terminal. La filogenia resultante se observa en la Figura 18. Como se ve, los genes de gimnospermas en este caso caen también como grupo externo al clado AP3-PI, pero los valores de apoyo por PP para su monofilia y para sus posiciones son muy pobres. Por ahora, podemos decir que nuestros datos evidencian que el emplear los genes MADS tipo B para evaluar las relaciones entre gimnospermas y angiospermas no es recomendable, dada la ambigüedad en el alineamiento de los genes en ambos grupos como en los bajos valores de soporte que se obtiene de sus posiciones.

El muestreo taxonómico en nuestro análisis filogenético, al igual que los antecedentes hechos con los MADS de gimnospermas, es deficiente, ya que solo incluimos las secuencias disponibles de Gnetales y de Coníferas. La integración de más secuencias de estos grupos así como de las correspondientes de Cicadas y *Ginkgo* podrían mejorar los soportes para los clados de estos genes y sus relaciones entre ellos, a menos de que el gran “gap” o salto morfológico entre gimnospermas y angiospermas también este reflejado en la evolución de estos genes y no sea posible resolver confiablemente sus relaciones filogenéticas.

Recientemente se publicó un estudio en el que se calibró un reloj molecular para estimar la fecha de la duplicación que dio lugar a los linajes AP3 y PI, y se sugirió una fecha de ~260 millones de años, con un intervalo de 230-290 ma (Kim et al, 2004), lo cual la ubica justo después de la separación de las angiospermas y gimnospermas y en la base de las plantas con flores (Kim, et al. 2004). En este estudio utilizaron el método NPRS (‘Non Parametric Rate Smoothing’), que acomoda la heterogeneidad de las tasas de sustitución permitiendo que las tasas varíen en las diferentes ramas (Sanderson, 1997). Nosotros utilizamos un método diferente, basado en verosimilitud pero que introduce la heterogeneidad en las tasas de sustitución con base en el método de Sanderson (1997, 2002) y que se llama AHRS (‘A Heuristic Rate Smoothing’ Algorithm) desarrollado por Ziheng Yang (2004). Nuestros resultados indican que esta duplicación ocurrió hace 290 ma, casi de manera inmediata a la fecha de origen de las plantas con semilla, hace 290-309.2 ma (Mapes y Rothwell, 1984 y 1991 citados en Kim, et al. 2004). La mayoría de la evidencia indica que hubo una separación muy abrupta entre las gimnospermas existentes y el linaje que dio lugar a las angiospermas (P. Soltis, et al. 2002 citado en Kim, et al. 2004). Nuestra estimación apoya esta propuesta y esto explicaría porque las relaciones de paralogía de los genes tipo B de gimnospermas y los parálogos tipo AP3 o los PI de angiospermas no es clara, y que su ubicación en la filogenia es muy difícil debido, precisamente, a su temprana separación y la casi inmediata duplicación de los AP3-PI en las angiospermas. La duplicación AP3-PI habría ocurrido entonces mucho antes de las fechas dadas a los fósiles más viejos disponibles para angiospermas, de 131.8 ma (Hughes, 1994; P. Soltis et al., 2002 citados en Kim, et al. 2004), siendo que las fechas que obtuvimos para el punto de origen de los clados AP3 y PI son de 190 y 139 ma respectivamente, lo cual implicaría que sus procesos de diversificación ocurrieron alrededor de 40 ma después de la duplicación que les dio origen. Sería interesante evaluar estas fechas para otros

genes de desarrollo floral, y evaluar si hay una correlación de las fechas de diversificación para los genes tipo B y otros genes.

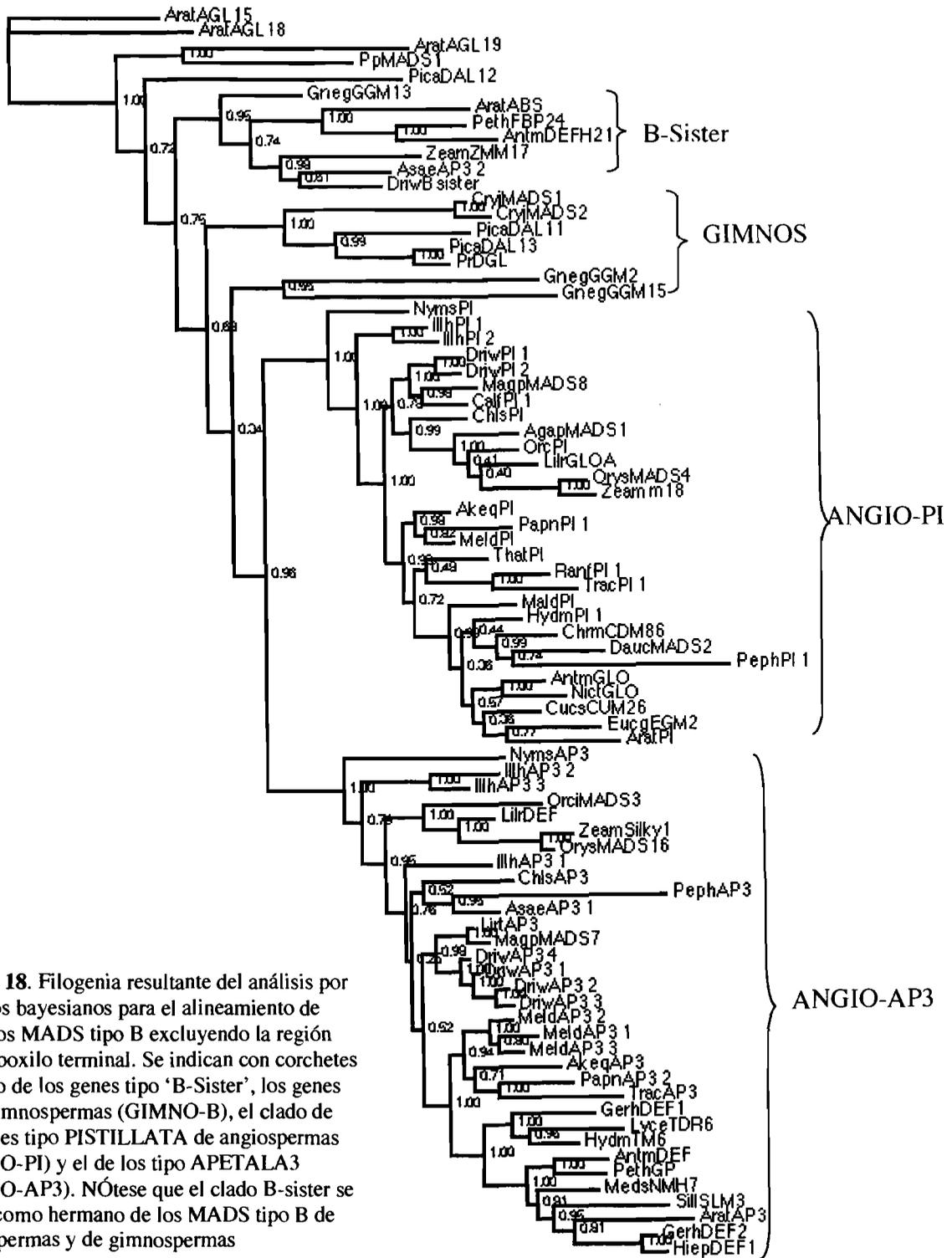


Figura 18. Filogenia resultante del análisis por métodos bayesianos para el alineamiento de todos los MADS tipo B excluyendo la región del carboxilo terminal. Se indican con corchetes el clado de los genes tipo 'B-Sister', los genes B de gimnospermas (GIMNO-B), el clado de los genes tipo PISTILLATA de angiospermas (ANGIO-PI) y el de los tipo APETALA3 (ANGIO-AP3). NÓtese que el clado B-sister se ubicó como hermano de los MADS tipo B de angiospermas y de gimnospermas

En el caso de que supongamos que la posición de los genes tipo B de gimnospermas como hermanos de los de angiospermas es la correcta (a pesar de los bajos soportes de probabilidad posterior que obtuvimos), podemos discutir la fecha obtenida para la duplicación que dio lugar a los genes tipo B en ambos linajes en el contexto de su evolución morfológica. La fecha obtenida en este caso fue de 354.3 ma. De acuerdo al registro fósil, las plantas con semilla evolucionaron a partir de las progimnospermas en el Devónico Tardío hace aproximadamente 350 millones de años (Stewart y Rothwell, 1993 citados en Sundström et al., 1999), lo cual ubicaría el origen de la función B y su diversificación muy temprano, en el origen de las plantas con semilla; dato que es congruente con resultados anteriores que se han obtenido con análisis de reloj molecular de la familia MADS que ubican el origen de los genes florales homeóticos como anteriores o cercanos al origen de las plantas con semilla (Purugganan, 1997).

Diversificación funcional y selección positiva en los MADS Tipo B de gimnospermas.

En resumen podemos decir que la evidencia experimental obtenida básicamente a partir de patrones de expresión y los análisis filogenéticos de las secuencias MADS tipo B de gimnospermas indican con claridad que la función B, aunque quizá de manera parcial, está presente en este grupo (Sundström et al., 1999; Winter, et al. 1999). En el caso de los genes *DAL11*, *DAL12* y *DAL13*, las diferencias en los patrones espacio-temporales de expresión de estos genes sugiere que, al menos en parte, han adquirido diferentes funciones. Los resultados de Sundström y Engström (2002) sugieren que la función B en coníferas es dual e interviene tanto en la determinación de identidad de meristemo como en la determinación de identidad de órganos en las estructuras reproductivas, de manera independiente a la interacción que estos tienen con genes de función C. Con base en sus resultados, postulan que incluso a pesar de que pudiera ser que la función B en gimnospermas es actuar en combinación con la función C para especificar los micro y mega esporangios (distinción entre sexos), la función B ha evolucionado de manera diferente en coníferas y angiospermas (Sundström y Engström, 2002). Con estas ideas en mente, y considerando que es muy probable que la agrupación de los genes de función B de gimnospermas sea como clado hermano de las angiospermas, planteamos dos hipótesis para evaluar diferencias en las tasas de sustitución (DTS) y presencia de selección positiva (SP): una en la rama que dio lugar a los genes B de gimnospermas, para evaluar una pronta divergencia justo después de su separación del clado de las angiospermas (hipótesis B-H1, Rama-gimnospermas, ver Resultados), y otra en el clado de los MADS tipo B en gimnospermas para evaluar las DTS que están ocurriendo en estos genes como reflejo de los procesos diversificación funcional de la función B en gimnospermas (hipótesis B-H9, Clado-gimnospermas). Las PTVs¹ para la rama-gimnospermas mostraron que no hay diferencias significativas en sus tasas de sustitución y tampoco selección positiva, sin embargo, en el clado de los genes MADS-Tipo B de gimnospermas las pruebas tanto para DTS considerables como para SP fueron positivas, en todos los casos se rechazaron los modelos que implican tasas de sustitución homogéneas y que no permiten selección positiva. Sin embargo no logramos detectar ningún sitio bajo SP.

¹ Las PTVs de MD(k=2)vsMD(k=3) y M3(k=2)vsMD(k=3) no fueron consideradas como buenas pruebas de significancia ya que en todos los casos aprobaron la hipótesis alternativa. No se les considero en ningun caso en los resultados de apoyo estadístico.

Los genes tipo B de gimnospermas están muy alejados de los demás genes Tipo B de angiospermas, lo cual se ve reflejado evidentemente en las secuencias de los genes, que son muy diferentes y hay muchas dificultades en alinearlas. Es posible que adicionando más secuencias de gimnospermas podríamos lograr obtener estimados más consistentes y detectar sitios particulares bajo SP, sin embargo con las pocas secuencias que fueron analizadas aquí y a pesar de que las pruebas de tasas de verosimilitud resultaron positivas para diversificación funcional y selección positiva en congruencia con la evidencia experimental y la historia evolutiva de la función B y las estructuras reproductivas en gimnospermas, estos resultados deben tomarse con cautela.

DTS y SP en Genes MADS tipo B del clado 'B-sister'

Como clado hermano inmediato del clado AP3-PI encontramos al clado de los genes conocidos como 'B-sister', aunque con bajos valores de soporte (0.46 pp). Este clado se recupera de manera recurrente en todas las repeticiones de los análisis filogenéticos que realizamos y también en todos los análisis reportados en la literatura que los incluyen.

En el clado de los 'B-sister' encontramos al gen GGM2, con valores de soporte de 0.6 pp. GGM2 había sido agrupado hasta ahora como gen hermano del clado AP3-PI y no del clado B-sister (Becker, et al. 2002), sin embargo aquí postulamos lo contrario.

Estudios realizados con estos genes mostraron que sus patrones de expresión están restringidos a estructuras reproductivas femeninas (óvulos y paredes de carpelos) (Becker, et al. 2002), lo cual contrasta con los patrones de expresión para los genes tipo B, que se expresan en estructuras reproductivas masculinas (estambres) y en pétalos. Con base en estos datos, Becker et al. (2002) sugirieron que el papel de los genes B-sister posiblemente es muy importante en la evolución del desarrollo de las estructuras reproductivas, y que el establecimiento de los diferentes linajes TipoB y 'B-sister' después de la duplicación de un gen ancestral pudo estar acompañado de la evolución de las microesporófilas (órganos masculinos) y las megaesporófilas (órganos femeninos) hace 400-300 ma (Becker, et al. 2002). Esta hipótesis no está del todo fundamentada, pero que de ser cierta, la posición correcta del clado B-sister en la filogenia debería ser como hermana de los genes TipoB-angiospermas+TipoB-gimnospermas, resultado que obtuvimos al hacer el análisis filogenético de las secuencias MADS tipo B excluyendo el C-terminal (ver Figura 18), pero no así en la filogenia obtenida con las secuencias completas.

Existen muy pocos estudios funcionales para los genes B-sister. Con el fin de buscar nuevas pistas acerca de la función de estos genes decidimos también analizar sus patrones de evolución molecular y hacer inferencias del papel de la SP en fijar algunos de sus residuos. Estos análisis corresponden a la hipótesis B-H3 o rama-Bsister, para detectar cambios tempranos justo después a la duplicación que dio lugar a éste linaje. En este caso los resultados fueron negativos, lo cual refleja que no ha habido heterogeneidad en las tasas de sustitución en la evolución temprana de este linaje con respecto a los demás genes en la familia y tampoco se pudo inferir que la SP fijara algunos de los residuos de las proteínas para las cuales codifican estos genes. Estos resultados pueden sugerir que estos genes son redundantes con los demás genes de función B. La falta de evidencia sobre su diferenciación funcional hace pensar que en la probabilidad de que estos genes sean redundantes con los demás de función B tanto en angiospermas como en gimnospermas.

La fecha de duplicación que dio lugar a los genes tipo B-sister estimada con el algoritmo AHRS fue de 351 ma, muy poco tiempo después de la fecha de duplicación estimada para los genes de gimnospermas (354 ma). Las fechas estimadas para el clado B-sister y por tanto para el inicio de su diversificación es de 331 ma, lo cual indica que estos genes han tenido una evolución lenta y de orígenes muy antiguos que se han conservado en los principales grupos de plantas con

semilla (al menos Gnetales y angiospermas). Sería muy interesante tener una completa caracterización funcional de estos genes.

DTS y PS en Genes MADS Tipo B en angiospermas

Hasta ahora hemos visto que, con base en análisis filogenéticos (los reportados en los antecedentes y el realizado por nosotros) y en evidencia experimental, resulta congruente pensar que los procesos evolutivos que han seguido los genes B y su función en las angiospermas y en las gimnospermas son diferentes. Para evaluar la importancia de la duplicación que dio lugar a los clados AP3 y PI en estos procesos, decidimos hacer pruebas para detectar DTS y SP en el clado de los MADS tipo B de angiospermas en general (clado AP3-PI), para detectar diferencias con respecto a los de gimnospermas (clado AP3-PI-Bsister, como hipótesis nula) y entender mejor así el impacto de la duplicación AP3-PI en los procesos de diversificación funcional que estos genes han tenido en las plantas con flores. Estas pruebas corresponden a la hipótesis B-H2, en la rama que da lugar a los MADS tipo B de angiospermas en general, incluyendo el clado B-sister (que también incluye secuencias de gimnospermas), y la hipótesis B-H4, en la rama que da lugar al clado AP3-PI anterior a la duplicación (con secuencias exclusivamente de angiospermas). En ambos casos los resultados fueron negativos, lo cual implica que de haber diversificación funcional y selección positiva en los genes Tipo B exclusivos de angiospermas (es decir, en el clado AP3-PI, con resultados positivos tanto para DTS como SP, como veremos más adelante), su diversificación funcional ocurrió como respuesta o como consecuencia (o facilitada) por la duplicación AP3-PI destacando la relevancia de esta, siendo que la duplicación B-Bsister no fue tan relevante.

Importancia de la duplicación AP3-PI en la diversificación funcional de los MADS Tipo B de angiospermas

Tanto en *A. thaliana* como en *A. majus* el producto de los ortólogos *AP3* y *PI* debe heterodimerizar para promover la identidad de pétalos y estambres (Schwarz-Sommer et al., 1992; Trobner et al, 1992; Jack, et al. 1994). Esto evidentemente habla de que hay una diversificación funcional en ambos genes de manera tal que implica, al menos, que las diferencias específicas eliminaron la posibilidad de homodimerización.

La determinación de pétalos y estambres en *A. thaliana* y *A. majus* requiere de la expresión de ambos genes durante todo el tiempo de desarrollo de los pétalos, mientras que en los estambres, la expresión puede quedar restringida a tejidos particulares conforme procede la maduración (Bowman et al., 1989; Jack et al., 1992; Zachgo et al., 1995). La coexpresión de ambos genes se mantiene por interacciones autorreguladoras positivas, pues una expresión continua y fuerte de ambos genes depende de haya una elevada concentración del heterodímero AP3/PI (Schwarz-Sommer et al. 1992; Goto y Meyerowitz 1994), además, se requiere la expresión de *AP3* y *PI* para conferir un fenotipo ectópico en *Arabidopsis* (Krizek y Meyerowitz, 1996 citado en Lamb e Irish, 2003). Los resultados de los estudios de los patrones de expresión de los genes B en angiospermas basales y los análisis filogenéticos de sus secuencias han aportado evidencia que parece indicar que el programa de desarrollo determinado por la función B era inicialmente plástico, y que quedó más restringido en etapas posteriores de la evolución de las angiospermas (Kramer & Irish 2000). El hecho de que en *Arabidopsis* la función B requiera de una heterodimerización obligada (Davies et al, 1996; Riechman et al, 1996b) de los productos AP3 y PI; que en monocotiledóneas (*Lilium regale*) las proteínas sean capaces de unirse al DNA tanto como homodímeros como

heterodímeros (Winter, et al. 2002) y que en gimnospermas (*Gnetum gnemon*, gnetales) la unión a DNA sea como homodímeros sugiere que el origen y evolución de la heterodimerización jugó un papel crucial en la evolución de la función B y el desarrollo floral. La duplicación que dio lugar a los genes tipo *AP3* y *PI* y la diversificación funcional posterior a ésta debió haber sido crucial en la evolución de la heterodimerización, y seguramente tuvo sus orígenes en cambios estratégicos en los dominios de unión a proteínas en uno o ambos parálogos.

Como análisis exploratorio, decidimos evaluar la presencia de DTS en los linajes *AP3* y *PI* en la filogenia que incluye a todos los MADS tipo B, y evaluar de manera general en cada clado la presencia de selección positiva para proponer sitios que pudieran ser importantes en las diferencias entre ambos linajes que promuevan las capacidades de heterodimerizar. Las primeras hipótesis en este caso fueron: la hipótesis B-H5, en la rama que da lugar a los genes tipo *AP3* y la hipótesis B-H6, en la rama que da lugar a los genes tipo *PI*, para evaluar diferencias tempranas y evolución de sitios particulares determinada por selección positiva.

Tanto la rama que da lugar a los genes tipo *PI* como la que da lugar a los genes tipo *AP3*, las PTVs indicaron que en éstas existen tasas de sustitución diferentes de manera significativa, lo cual refleja claramente los procesos de diversificación funcional en éstos desde etapas tempranas, justo después de la duplicación que les dio origen. Los valores de ω para los sitios con tasas cambiantes en el modelo D con $k=2$ y $k=3$ para los genes tipo *PI* indican que las diferencias están dictadas por selección positiva ($\omega_{3c} = 1.266$). Esto fue distinto para los genes tipo *AP3*, ya que el valor correspondiente fue menor que 1 ($\omega_{3c} = 0.77$ para $k=2$ y $\omega_{2b} = 0.65$ para $k=3$). En la rama que da lugar a los genes tipo *PI* se detectaron varios sitios bajo selección positiva, sin embargo de las tres PTVs evaluadas, solo dos de ellas fueron positivas. Los valores de ω resultantes para los sitios bajo SP son de 6.2 en el modelo A (MA) y de 2.5 en el modelo B (MB) con proporciones de $0.09+0.12=0.21$ y $0.03+0.05=0.08$ respectivamente (ver tabla correspondiente en Resultados). De los sitios detectados, los sitios 91H, 100R, 115L, 119L, 142H y 165D caen en la región K, que es imprescindible para la heterodimerización *AP3-PI* (Yang et al. 2003). Los sitios 225A, 255* y 297H son de difícil interpretación, ya que caen en regiones del C-terminal anteriores al motivo *PI* que son difíciles de alinear.

En la rama que da lugar a los genes tipo *AP3* también se detectaron varios sitios bajo selección positiva, siendo que en este caso las tres PTVs realizadas dieron soporte estadístico al resultado de selección positiva. Los valores de ω resultantes para los sitios bajo SP son de 8.5 en el modelo A (MA) y de 3.9 en el modelo B (MB) con una proporción de $0.07+0.05=0.11$ y $0.03+0.05=0.08$ respectivamente, lo cual nos indica que en esta rama, a pesar de que la proporción de sitios bajo selección positiva es menor que en la rama que da a los *PI*, la intensidad de la selección es mayor. Todos los sitios detectados en esta prueba caen en el dominio K. Estos resultados son los más importantes de la tesis, pues sugieren que desde muy pronto, justo después de la duplicación *AP3-PI*, empezó la diferenciación en éstos linajes, en particular por selección positiva en el dominio K, fundamental para la heterodimerización de estas proteínas.

Selección positiva en el dominio K de genes tipo *PI* y *AP3* y origen de la heterodimerización

Para evaluar diferencias en las tasas de sustitución y presencia de selección positiva que siguieron a la diversificación en los clados *AP3* y *PI* se plantearon las hipótesis H7-TipoB para las secuencias en el clado de los genes tipo *AP3* y H8-TipoB en el clado de los genes tipo *PI*.

En el caso de los genes tipo *PI*, las PTVs ya no detectan diferencias significativas en las tasas de sustitución con los modelos D con $k=2$ y $k=3$, sin embargo, las PTVs para selección

positiva favorecieron todas un rechazo de las hipótesis nulas de homogeneidad de tasas. Es importante enfatizar que ningún parámetro estimado con éstos modelos implicó selección positiva ($\omega > 1$) para éste clado, y los sitios detectados pertenecen a una categoría con valores de $\omega = 1$, por lo que deben interpretarse como neutros o bajo restricciones selectivas relajadas. De los sitios detectados, el 36T, 42K, 51T y 55H están en la caja MADS, especialmente en la parte C-terminal de la caja MADS, que otorga capacidades de heterodimerización con otras proteínas (Yang et al, 2003). Los sitios 70T, 72K y 80S caen en el dominio I, mientras que los sitios 141Q, 142H, 148T y 152E caen en el dominio K. Los sitios 173N, 194D, 219D, 225A y 286H fueron detectados también, pero su interpretación es difícil ya que caen en una posición del C-terminal con un alineamiento muy ambiguo. En el caso del clado de los AP3, las PTVs para diferencias en las tasas de sustitución y selección positiva fueron todas significativas, pero de igual manera que en el caso del clado de los PI, ningún parámetro estimado para ω implicó selección positiva, por lo que los sitios detectados aquí también pertenecen a una categoría con $\omega = 1$. De los sitios detectados, el 33H y 54F son parte de la caja MADS; el sitio 82V pertenece al dominio I y los sitios 104T, 111T, 144Q y 145D son parte del dominio K. Los sitios 174L, 175E, 177R y 179G están en el C-terminal, pero en una región con buen alineamiento, y forman parte de un pequeño motivo que en los genes tipo PI esta muy conservado, con secuencia LEEENK, misma que en los genes tipo AP3 es muy variable, y que va del aminoácido 174 al 179 en nuestro alineamiento. Este motivo puede ser funcionalmente importante para la dimerización y/o especificidad funcional en las proteínas PI. Como se observa a partir de estos datos, los procesos de diversificación funcional posteriores a la duplicación de las proteínas AP3 y PI son más generalizados, siendo que los sitios variables, que si bien ya no se encuentran bajo fuertes presiones selectivas sino más bien relajadas, ya no se restringen al dominio K. Es probable que la posibilidad de heterodimerizar haya surgido de manera previa a la diversificación funcional de las proteínas AP3 y PI por fijación de sitios por SP.

Importancia de la duplicación euAP3-TM6 e implicaciones en el origen y diversificación de las eudicotiledóneas superiores.

Una duplicación importante que ocurrió en el clado de los genes AP3 dio lugar al linaje de los genes TM6 y al linaje de los euAP3. Con base en la distribución filogenética de los taxa que presentan estos genes, se ha propuesto que esta duplicación ocurrió justo antes de la diversificación de las eudicotiledóneas superiores ('core-eudicots') (Kramer, et al., 1998; Kramer e Irish, 2000); clado que comprende cerca del 73% de todas las especies de plantas con flores (Drinnan, et al., 1994) y que representa en si mismo un evento de radiación inusual al interior de las angiospermas (Magallón y Sanderson, 2001). La mayoría de las especies incluidas en el clado de las eudicotiledóneas superiores o 'core-eudicots' tienen genes representantes del clado de los genes tipo euAP3 y de los TM6 (Kramer et al. 1998). *Arabidopsis thaliana*, perteneciente a la familia de las Brassicaceae, es inusual en este sentido ya que al parecer ha perdido la copia del gen tipo TM6 y retenido a su gen del linaje euAP3; aunque otros miembros de la misma familia (Eurosidae I) si lo presentan (Lamb e Irish, 2003).

La flor de las eudicotiledóneas superiores se caracteriza en parte por poseer un número fijo de órganos florales en cada verticilo, filotaxia floral verticilada y un perianto caracterizado por un cáliz y corola morfológicamente diferenciados (Magallón et al, 1999; Drinnan et al, 1994 citados en Lamb e Irish, 2003). En contraste, las especies no-eudicotiledóneas se caracterizan por una mayor plasticidad en su estructura floral (Lamb e Irish, 2003).

En un estudio reciente, Lamb e Irish (2003) hicieron un análisis experimental minucioso sobre la diversificación funcional de las proteínas AP3 y PI. En este estudio las autoras investigaron la posibilidad de que cambios en las funciones de las proteínas AP3 y PI hayan tenido un papel importante en la evolución de morfologías florales diferentes. Se había mostrado antes que el dominio MADS no contribuía de manera significativa a la especificidad funcional de proteínas florales homeóticas, ya que plantas de *Arabidopsis* transgénicas que contenían construcciones quiméricas en las cuales el dominio MADS de una proteína homeótica floral estaba fusionada con las regiones IKC de otra proteína homeótica MADS diferente, mostraban las características otorgadas por la fuente IKC de la construcción (Krizek y Meyerowitz 1996; Krizek et al, 1999 citados en Lamb e Irish, 2003). De manera similar, proteínas quiméricas y estudios de mutagénesis habían mostrado que los dominios I y K tanto en AP3 como en PI eran necesarios para la especificidad de la dimerización *in vitro* y para la función de la proteína *in vivo* (Krizek et al, 1999; Yang et al, 2003; McGonigle et al, 1996; Irish y Yamamoto, 1995; citados en Lamb e Irish; 2003). Para evaluar la importancia funcional del dominio C-terminal, Lamb e Irish (2003) analizaron el fenotipo de plantas de *Arabidopsis* mutantes de la función B transformadas con proteínas truncadas y proteínas quiméricas en las que intercambiaron el dominio C-terminal en AP3 por el de PI y viceversa. Sus análisis mostraron que los extremos C-terminales de AP3 y PI son fundamentales para conferir la función B, y que no se pueden reemplazar uno a otro para rescatar el fenotipo mutante; lo cual es prueba de la diversificación funcional entre ambos linajes. Además, transformaron mutantes *ap3* de *Arabidopsis* con construcciones de la proteína AP3 de *Arabidopsis* (una proteína del linaje *euAP3*) con el extremo C-terminal del gen AP3 de *Dicentra eximia* (que es tipo *paleoAP3*), con la finalidad de probar el papel funcional del dominio C-terminal *euAP3*. En este caso, sus resultados indicaron que la construcción con el dominio *paleoAP3* podía rescatar solo parcialmente el desarrollo de los estambres, pero no el de pétalos, indicando la diversificación funcional de las proteínas tipo *euAP3* (con el dominio C-terminal *euAP3*) y *TM6* (con el dominio *paleoAP3*). Se han caracterizado genes *paleoAP3* en dos monocotiledóneas: *Zea mays* (maíz) y *Oriza sativa* (arroz), y en ambos casos se ha visto que los genes correspondientes, *SILKY1* y *OsMADS16*, controlan la identidad de los estambres y los lodículos (Ambrose et al., 2000; Nagasawa et al., 2003). Lamb e Irish (2003) especulan en su trabajo que la duplicación que dio lugar a los linajes *euAP3* y *TM6* puede correlacionarse con la derivación de pétalos en las eudicotiledóneas superiores, y sugieren que nuevas secuencias en el dominio C-terminal de los genes AP3 en estos linajes pudieron estar asociadas a la evolución de los pétalos en éste grupo de angiospermas (Kramer, et al. 1998; Lamb e Irish, 2003).

En nuestro estudio, planteamos diferentes hipótesis para analizar los procesos de diversificación funcional de los MADS tipo AP3 y tipo PI probada experimentalmente a nivel de sus tasas de sustitución en linajes particulares de plantas, utilizando filogenias reconstruidas por separado para los genes tipo *PI* y tipo *AP3*. Hay que destacar que de manera general encontramos que, en contraste con lo que ocurre para los MADS tipo B de gimnospermas o el clado 'B-sister', en los genes tipo *AP3* y *PI* detectamos una fuerte señal que indica tanto DTS significativas entre linajes como presencia de selección positiva, lo cual habla de la importancia de los procesos de diversificación funcional en estos genes al interior de las angiospermas.

En primer lugar, para la filogenia de los genes *AP3* probamos las hipótesis AP3-H1 en la rama que da lugar a los genes tipo AP3 de angiospermas para detectar diversificación funcional y SP muy temprana antes de la diversificación de este linaje. En este caso las pruebas para DTS fueron negativas, y de las tres que se realizaron para probar SP solo una fue positiva, con una $P=0.0251$. Los sitios detectados: 153N, 195Y y 197V caen todos en el dominio K. Posteriormente

probamos la hipótesis AP3-H2, en la rama que da lugar al clado euAP3-TM6. En este caso las pruebas para detectar DTS fueron positivas, con $P=0.028$ y $P=0.041$, lo cual habla de la diversificación funcional temprana, antes de la duplicación de este clado. Las PTV para detectar SP fueron negativas, aunque en todos los casos y con los tres modelos los parámetros estimados para ω fueron mayores que 1, lo cual implica una fuerte señal de selección positiva, al menos en una pequeña proporción de sitios. El único sitio detectado bajo SP en este caso (146N) está también en el dominio K.

La siguiente hipótesis se realizó en todas las ramas del clado euAP3-TM6, que corresponde a la AP3-H3. A diferencia de lo que esperábamos dados los resultados con AP3-H2, las pruebas para DTS fueron negativas, pero las pruebas para SP fueron positivas, aunque ningún parámetro ω estimado bajo el modelo A o B implica SP. Los valores de P fueron <0.0001 en todos los casos. De los sitios detectados con altos valores de PP de pertenecer a categorías con $\omega=1$, uno de ellos está en el dominio MADS (99G), los sitios 110T, 123T, 128S, 135N y 136S están en la región I y los sitios 177R, 178G y 188K están en el dominio K. Los demás sitios encontrados se encuentran en el dominio C-terminal. A pesar de que el alineamiento para los genes AP3 es mucho más compacto, confiable y con menos ambigüedades que el alineamiento de todos los MADS tipo B, los sitios 230P, 260R, 282P y 283S tienen una difícil interpretación por que se encuentran en regiones de la proteína con muchos 'in-dels'. Solo el sitio 319R se encuentra en el motivo C-terminal característico de los AP3, que en este caso sería euAP3 y paleoAP3 para el clado TM6. Estos resultados nos hablan de la importancia que en la diversificación funcional que siguió a la duplicación *euAP3-TM6* han tenido también otros sitios a todo lo largo de la proteína además de los del C-terminal, sin embargo dados los valores de ω no deben interpretarse como fijados por SP sino como nulos o bajo restricciones selectivas relajadas.

Las siguientes hipótesis, AP3-H4 y AP3-H5 corresponden a las ramas que dan lugar a los clados euAP3 y TM6 respectivamente. En el caso de la rama que da a los euAP3, detectamos DTS, pero dos de las tres pruebas para detectar SP fueron significativas, con valores estimados de $\omega=6.28$ con el modelo A y $\omega=6.8$ con el modelo B. Este tipo de resultados no son contradictorios, sino que pueden estar indicando que la diversificación funcional está dada más por selección positiva en unos cuantos sitios de las proteínas que por un evento generalizado de cambios en las presiones selectivas de manera homogénea en toda la proteína. De los sitios detectados en esta rama, el 49I, 75I y 92L están en la caja MADS; el 153N, 167D, 182N, 196H y 199N están en el dominio K. Posteriormente encontramos bajo SP con altos valores de PP a los sitios que corresponden a aquellos del motivo euAP3, que se alinean muy bien (salvo el sitio 322*), y que corresponden en nuestro alineamiento a los sitios 315T, 316H, 317E, 318L, 319R, 321G y 322* esperados bajo selección positiva dada la evidencia experimental de la importancia de este motivo.

En el caso de los resultados para la rama que da al clado de los TM6 (AP3-H5), solo se detectaron unos cuantos sitios bajo SP (153N, 195Y y 197V), pero las pruebas estadísticas que los sustentan resultaron negativas. Los parámetros ω estimados bajo el modelo A y B no indican selección positiva, pues tienen valores de 1. Solo con el modelo D el valor estimado de ω resultó mayor que 1 (2.4 con el modelo D K=2 y 2.5 con K=3). Este resultado puede indicar una conservación general que persiste en el clado de los TM6 (que conservan el motivo ancestral tipo paleoAP3 en su extremo C-terminal) a diferencia de la rápida y temprana diversificación en los genes tipo euAP3 dada por SP.

Por último para los MADS tipo AP3 realizamos las pruebas para detectar SP y DTS persistentes en los clados euAP3 y TM6 y que responden a sus procesos de diversificación. Estas hipótesis corresponden a AP3-H6 para el clado de los euAP3 y AP3-H7 para el clado de los TM6.

Los resultados son consistentes con la evidencia experimental, ya que hay una señal para DTS en ambos clados. A pesar de esto, no se estimaron en ningún caso parámetros ω con valores mayores a 1. Esto puede deberse a las dificultades en detectar la señal de SP dado el pequeño número de sitios bajo este régimen selectivo o por el reducido número de secuencias analizadas. Los sitios detectados para el clado euAP3 son 110T, 120I, 123R y 135N en la región I; 157R, 178G, 188K, 197V y 205F en el dominio K y 218E, 260R, 261N, 264*, 276F y 283S en el extremo N-terminal del dominio C-terminal de la proteína, siendo en el motivo euAP3 solo se detectó el sitio 319R. Los sitios detectados en el clado TM6 son 99G en la caja MADS; 126H en el dominio I; 164V, 173E, 177R, 178G, 184V y 188K en el dominio K y 221S, 230P, 233*, 257F, 279R y 297* están en el extremo N-terminal del dominio C-terminal, y solo 311V y 312Y están en el motivo paleoAP3. En el caso de los TM6, el extremo C-terminal presenta más ambigüedades en su alineamiento y la interpretación de estos sitios es más difícil. Estos resultados indican que la diversificación funcional en el dominio C-terminal fue dada en las etapas iniciales en el clado euAP3, justo después de la duplicación que los separó de los TM6, y que posteriormente estos sitios adquiridos por SP fueron conservados al interior del clado. Este resultado es consistente con la observación de la adquisición del motivo euAP3 a partir de uno tipo paleoAP3 por un evento de deleción de un nucleótido y corrimiento del marco de lectura de Kramer (no publicado, comunicación personal). Es probable que la conservación de la función ancestral B de determinación de estambres dada por el dominio paleoAP3 (Lamb e Irish, 2003) en el ancestro de las eudicotiledóneas superiores (core-eudicots) haya permitido que el parálogo duplicado divergiera en secuencia y en función, llegando a tener un papel más importante en la determinación y establecimiento de los pétalos en este grupo.

Existen pocos y pobres estudios acerca de la posible función de los genes TM6 y sus patrones de expresión en las eudicotiledóneas superiores, pero recientemente se publicó un estudio bastante completo de los MADS Tipo B en *Petunia hybrida* (Solanaceae) que incluye al su ortólogo TM6 (Vandenbussche et al, 2004). En esta especie hay solo una copia del gen tipo AP3: *PethDEF* y dos para el tipo PI: *PethGLO1* y *PethGLO2*, que en nuestro análisis nombramos como *PethMADS2*. En el trabajo de Vandenbussche et al (2004) encontraron que *PethTM6* es capaz de heterodimerizar de manera preferencial con *PethGLO2* en levadura, pero que también es capaz de hacerlo con *PethGLO1* en menor medida, siendo que *PethDEF* interactúa con *PethGLO1* y *PethGLO2* siguiendo una conservación típica de la función B en éstos genes. Encontraron que el patrón de expresión de *PethTM6* en *Petunia* es, a diferencia del encontrado para *PethDEF*, de niveles muy débiles en pétalos, lo cual junto con el hecho de que las mutantes *pethdef* resultan en una conversión homeótica total de pétalos a sépalos, sugiere que la función de *PethTM6* no tiene importancia en el desarrollo de pétalos. Con análisis más detallados encontraron que los patrones de expresión de *PethTM6* se extiende a los estambres en desarrollo, y análisis de recuperación de fenotipos mutantes mostraron que *PethTM6* es capaz de complementar la mutación *pethdef* en cuanto al desarrollo de estambres pero no de pétalos, lo cual sugiere la conservación de su función ancestral en el desarrollo de estructuras masculinas dadas por el dominio paleoAP3 (Vandenbussche et al, 2004). Nuestros análisis detectan estos eventos de diversificación funcional con ortólogos de los linajes euAP3 y paleoAP3 en diferentes especies y sugieren sitios específicos bajo SP, además de los presentes en el motivo C-terminal que pueden estar jugando un papel importante en esta diversificación.

Calibramos un reloj molecular para obtener estimados de la fecha de la duplicación euAP3-TM6 y ver si correlacionaba con la fecha del origen de las eudicotiledóneas superiores. La fecha

obtenida para esta duplicación fue de 92.3 ma, calibrando el reloj con la fecha de 132 ma reportada como edad del fósil más viejo conocido para angiospermas (Brenner, 1996 citado en Magallón y Sanderson, 2001). Esta fecha coincide aproximadamente con la reportada para las eudicotiledóneas superiores, con fósiles de Fagales que datan de 96.2 ma que pueden considerarse como una fecha confiable para estimar la edad del grupo (Pacltová 1966, 1981 en Magallón y Sanderson, 2001; Magallón, comunicación personal). Es muy posible que la duplicación euAP3-TM6 haya tenido un papel fundamental en la evolución de pétalos en las eudicotiledóneas superiores y en la radiación del grupo.

Selección positiva en el clado de los genes tipo PI

A pesar de que en el alineamiento de los genes tipo PI se observa una mayor conservación en comparación con el alineamiento para los genes tipo AP3, y de que en los resultados arrojados en la selección del modelo de sustitución indicaron que el modelo que mejor describe los procesos de cambio en los genes tipo PI es un modelo más conservador (ver Resultados), las pruebas que realizamos de detección de SP y DTS resultaron positivas, aunque de manera menos drástica y evidente que en los genes AP3. En este caso únicamente planteamos tres hipótesis, de manera tal que pudiéramos contrastarlas con los resultados para los genes tipo AP3. La primera hipótesis corresponde a la PI-H1, en la rama que da lugar a todos los genes MADS tipo PI. En este caso una de las dos pruebas para detectar DTS fue positiva, así como una de tres pruebas para SP. Los valores de ω fueron de 1.6 con el modelo A y de 6.3 con el modelo B. Los sitios encontrados en 132N y 165A están en el dominio I, el 192A y 207E están en el dominio K y el 320H es difícil por que se encuentra en una región del C-terminal muy difícil de alinear.

La siguiente hipótesis puesta a prueba es la PI-H2, en la rama que da lugar al clado de los genes PI de eudicotiledóneas. En este caso las PTVs resultaron negativas en todos los casos, pero se detectaron dos sitios bajo SP, correspondientes al 165A en el dominio I y al 275R, en el extremo amino-terminal del C-terminal de la proteína.

La tercer hipótesis fue la PI-H3, que corresponde a todas las ramas en el clado de los PI de eudicotiledóneas para evaluar la diversificación funcional presente en este clado. De manera inquietante, encontramos que todas las pruebas, tanto para SP como para DTS fueron positivas, casi todas con valores de $P < 0.0001$, lo cual indica claramente que también los genes tipo PI en este linaje están evolucionando de manera dinámica, quizá respondiendo a la diversificación que ocurre en los genes AP3. De cualquier manera los parámetros ω estimados no implican SP. De los sitios encontrados, el 125H y 132S están en el dominio I, el 203N esta en el dominio K y todos los demás caen en el dominio C-terminal. La mayoría de estos sitios caen en una región del C-terminal de difícil interpretación por las ambigüedades en su alineamiento.

Conclusiones

En este trabajo presentamos un análisis detallado de la evolución molecular de los genes MADS-box tipo B, dada su relevancia en los mecanismos de desarrollo del perianto floral. Se buscó hacer una determinación de eventos de DTS y de SP en linajes particulares utilizando métodos basados en máxima verosimilitud, que tienen limitantes importantes, sobre todo en cuanto a las ambigüedades que puedan estar presentes en los alineamientos de las secuencias. A pesar de que fuimos muy cuidadosos tanto al hacer el alineamiento como al interpretar los sitios encontrados bajo SP, se recomienda cautela al considerar los resultados obtenidos.

No logramos resolver satisfactoriamente la posición de los genes tipo B de gimnospermas con respecto a los de angiospermas, pues su alineamiento es muy ambiguo y su posición en todos los casos se apoya con bajos valores de PP. Consideramos que no es recomendable empelar estas secuencias para evaluar las relaciones filogenéticas entre plantas con semilla. La pobre resolución puede deberse a el pequeño número de secuencias de gimnospermas disponibles y empleadas en nuestro análisis o a que el salto morfológico entre angiospermas y gimnospermas se refleja también en las secuencias de genes MADS-Box.

Según las referencias consultadas para el registro fósil, las fechas que obtuvimos para la separación de los genes tipo B de angiospermas y gimnospermas (354 ma) coincide con la fecha para el origen de las plantas con semilla (350 ma), lo cual es congruente con estudios antecedentes de reloj molecular realizados con la familia MADS. La fecha obtenida para la duplicación AP3-PI, de 290 ma, la ubica 30 ma antes que la fecha obtenida en un estudio reciente, mientras que nuestra fecha es mucho más cercana al origen y diversificación de las plantas con semillas, y la separación de angiospermas y gimnospermas. La pronta separación de los genes AP3 y PI después de la separación de angiospermas y gimnospermas dificulta la ubicación de los genes B de gimnospermas con respecto a los de angiospermas.

Se detecto DTS y PS de manera significativa en los genes de gimnospermas, pero no de manera temprana después de su origen sino durante su diversificación, lo cual refleja los diferentes patrones evolutivos que siguen estos genes con respecto a los de angiospermas. No se detectaron sitios bajo SP en éstos genes, lo cual puede hablar de una dificultad del método para detectarlos o de que sus procesos de diversificación responden a cambios en las tasas evolutivas de manera homogénea a lo largo de la proteína. En contraste, no detectamos DTS ni SP en el clado de los B-sister, cuyas fechas estimadas de origen y posterior diversificación resultaron muy antiguas, siendo que aparecen justo después de la diversificación de las plantas con semilla y permanecen conservados así tanto en las angiospermas como en las gimnospermas. Es probable que estos genes sean redundantes con la función B de angiospermas y gimnospermas y sería muy interesante realizar su caracterización funcional completa.

Según nuestros resultados, la duplicación MADS tipo B y Bsister no tuvo mucha importancia en los procesos de diversificación funcional de los genes B en plantas con semilla, pero al duplicación AP3-PI fue fundamental para la diferenciación funcional de estos genes en angiospermas, siendo que esta diferenciación estuvo dada por SP en sitios específicos del dominio K de manera temprana en el origen de los genes AP3 y PI, lo cual significaría el origen de la heterodimerización de estas proteínas en angiospermas. Posteriormente, los procesos de diversificación funcional en ambos clados se caracterizaron por una diversificación más generalizada en diferentes sitios en la proteína, ya no restringidos al dominio K, lo cual se observa

como una relajación generalizada de las restricciones selectivas más que por selección positiva en sitios particulares.

De manera similar a la duplicación AP3-PI, la duplicación euAP3-TM6 fue importante para la diversificación de estos genes al interior de las eudicotiledóneas superiores. Hubo DTS desde etapas tempranas en la aparición del clado y justo después de la duplicación detectamos eventos de SP y DTS en diferentes regiones de la proteína en los genes euAP3, con excelente soporte estadístico. En etapas posteriores, que corresponden a la diversificación funcional en ambos clados, encontramos también una fuerte señal de DTS y SP, aunque los sitios detectados tienen valores de omega menores a 1, lo cual implica que al igual que en el caso de los genes AP3 y PI, después de la duplicación hubo una relajación de las presiones selectivas. Nuestros análisis ayudan a hacer énfasis en la importancia de otros sitios de la proteína además de los incluidos en la región C-terminal durante los procesos de diferenciación funcional.

La fecha que estimamos para la duplicación euAP3-TM6 fue de 92.3 ma, calibrando con 132 ma al origen de las angiospermas. Esta fecha coincide con la reportada para las eudicotiledóneas superiores, con fósiles de Fagales que datan de 96.2 ma que pueden considerarse como una fecha confiable para estimar la edad del grupo.

Además de la SP encontrada en los genes tipo AP3, también encontramos, aunque en menor medida, sitios bajo SP en los genes tipo PI, que pueden estar evolucionando en respuesta a la intensa diversificación que está ocurriendo en sus parálogos AP3.

Los detalles finos de los mecanismos moleculares de la evolución de la función B que siguieron a los momentos críticos que detectamos en este estudio así como las consecuencias morfológicas permanecen sin aclararse. Nuestros resultados proponen un punto de partida para plantear hipótesis experimentales que pueden probarse en el laboratorio.

Perspectivas

Existen varias líneas que pudieran dar seguimiento a este trabajo, de las cuales algunas son más viables dados los elementos que se requieren para trabajarlas y dadas las investigaciones que otros grupos están realizando.

En primer lugar podrían realizarse estudios similares a éste con otros genes de desarrollo floral, sobre todo si se seleccionan con base en la red de desarrollo floral reportada recientemente (Espinosa-Soto et al., 2004). Una limitante de esto sería que en muchos casos el número de secuencias disponibles en las bases de datos públicas es limitado, pues los estudios se han concentrado básicamente en los genes tipo A, B y C, siendo que a los demás no se les ha puesto mucha atención. En este caso habría que extraer las secuencias directamente de tejido vegetal.

En el caso del estudio de la función B, podrían secuenciarse también genes MADS tipo B de otras gimnospermas, como de *Ginkgo*, cicadas y otras coníferas, para evaluar su posición filogenética y repetir las pruebas para detectar DTS y PS, además de que se requiere de análisis de caracterización funcional más profundos y detallados que ayuden a entender mejor el papel de estos genes en gimnospermas. Varios grupos han trabajado intensamente con la función B en gimnospermas, entre ellos el grupo de Jen Sündstrom en la Universidad de Uppsala en Suecia y de Günter Theissen, en el Instituto Max-Planck en Alemania.

Otra línea importante que podría seguirse es la experimentación con las proteínas AP3 y PI en *Arabidopsis*. En este trabajo propusimos muchos sitios además de los característicos del C-terminal cuya importancia podría corroborarse por mutagénesis dirigida y experimentos de recuperación de fenotipo. El grupo de Yingzhen Yang en Darmouth College en los EU ha publicado muchos estudios acerca del mecanismo fino de la funcionalidad de los diferentes dominios de las proteínas PI y AP3, sobre todo en el dominio K, y tal vez abordarán pronto de manera más profunda y detallada los otros dominios.

Un experimento interesante que nos llevaría al origen de la heterodimerización en angiospermas podría ser el sintetizar una proteína ancestral del clado *AP3/PI* para transformar una mutante *ap3-pi* de *Arabidopsis*. No estoy segura de que este experimento podría arrojar resultados diferentes a los ya realizados en los que se transforman *Arabidopsis* con proteínas tipo B de gimnospermas.

Dada la importancia asociada a la duplicación euAP3-TM6 en el desarrollo floral de las eudicotiledóneas superiores, el paso lógico siguiente es hacer un estudio profundo de la función de los genes TM6 en diferentes especies. Además de que podría estudiarse de manera más profunda la función B en las eudicotiledóneas basales, para elucidar el origen y las posteriores consecuencias que tuvo la duplicación euAP3-TM6 en la evolución de la función B y el establecimiento de pétalos bien diferenciados en una corola establecida en las eudicotiledóneas superiores. El grupo de la Dra. Elena Kramer en la Universidad de Harvard seguramente enfocará sus siguientes investigaciones al estudio de angiospermas basales y posiblemente de eudicotiledóneas basales.

Por último propondría un experimento de transformación de mutantes que tal vez contribuiría a entender el papel que la duplicación euAP3-TM6 tuvo en la evolución de la flor en las eudicotiledóneas superiores. Tal vez si se pudiera transformar una mutante nula de la función B en alguna angiosperma no-eudicotiledónea como arroz (*Oryza sativa*) o maíz (*Zea mays*) con construcciones que contengan proteínas AP3(euAP3) y TM6 obtenidas de alguna eudicotiledónea

superior, podríamos ver el efecto del fenotipo resultante y evaluar su importancia en la determinación de las estructuras del perianto. Lamentablemente desconozco mucho acerca de las técnicas y herramientas de biología y genética molecular disponibles para trabajar con monocotiledóneas y no se si es posible obtener una mutante nula de la función B y transformarla de manera fácil como sería hacerlo en *Arabidopsis*.

Apéndice

(Lista de secuencias empleadas por orden alfabético)

Gen	Subclase	Familia	Especie
AgapMADS1	Asparagales	Agapanthaceae	<i>Agapanthus praecox</i>
AkeqAP3-1	Ranunculales	Ranunculaceae	<i>Akebia quintata</i>
AkeqAP3-2	Ranunculales	Ranunculaceae	<i>Akebia quintata</i>
AkeqPI-1	Ranunculales	Ranunculaceae	<i>Akebia quintata</i>
AntmDEF	Lamiales	Veronicaceae	<i>Antirrhinum majus</i>
AntmGLO	Lamiales	Veronicaceae	<i>Antirrhinum majus</i>
AralPI	Brasicales	Brassicaceae	<i>Arabidopsis lyrata</i>
AratAP3	Brasicales	Brassicaceae	<i>Arabidopsis thaliana</i>
AratPI	Brasicales	Brassicaceae	<i>Arabidopsis thaliana</i>
AsacAP3	Piperales	Aristolochiaceae	<i>Asarum europaeum</i>
AsaePI	Piperales	Aristolochiaceae	<i>Asarum europaeum</i>
AspoDEF	Asparagales	Asparagaceae	<i>Asparagus officinalis</i>
BergPI-1	Ranunculales	Ranunculaceae	<i>Berberis gilgiana</i>
BergPI-2	Ranunculales	Ranunculaceae	<i>Berberis gilgiana</i>
BetpMADS2	Fagales	Betulaceae	<i>Betula pendula</i>
BranAP3	Brasicales	Brassicaceae	<i>Brassica napus</i>
CalfAP3	Laurales	Calycanthaceae	<i>Calycanthus floridus</i>
CalfPI	Laurales	Calycanthaceae	<i>Calycanthus floridus</i>
ChlsAP3	Chlorantales	Chloranthaceae	<i>Chloranthus spicatus</i>
ChlsPI	Chlorantales	Chloranthaceae	<i>Chloranthus spicatus</i>
ChrmCDM115	Asterales	Asteraceae	<i>Chrysanthemum moriflorum</i>
ChrmCDM19	Asterales	Asteraceae	<i>Chrysanthemum moriflorum</i>
ChrmCDM86	Asterales	Asteraceae	<i>Chrysanthemum moriflorum</i>
CryjMADS1	Coniferales	Cupressaceae	<i>Cryptomeria japonica</i>
CryjMADS2	Coniferales	Cupressaceae	<i>Cryptomeria japonica</i>
CucsCUM26	Cucurbitales	Cucurbitaceae	<i>Cucumis sativus</i>
DaucMADS2	Apiales	Apiaceae	<i>Daucus carota</i>
DaucMADS3	Apiales	Apiaceae	<i>Daucus carota</i>
DiceAP3	Ranunculales	Papaveraceae	<i>Dicentra cxinia</i>
DrwAP3-1	Canellales	Winteraceae	<i>Drimys winterii</i>
DrwAP3-2	Canellales	Winteraceae	<i>Drimys winterii</i>
DrwAP3-3	Canellales	Winteraceae	<i>Drimys winterii</i>
DrwAP3-4	Canellales	Winteraceae	<i>Drimys winterii</i>
DrwB_sister	Canellales	Winteraceae	<i>Drimys winterii</i>
DrwPI-1	Canellales	Winteraceae	<i>Drimys winterii</i>
DrwPI-2	Canellales	Winteraceae	<i>Drimys winterii</i>
EucgEGM2	Myrtales	Myrtaceae	<i>Eucalyptus grandis</i>
GerhDEF1	Asterales	Asteraceae	<i>Gerbera hybrida</i>
GerhDEF2	Asterales	Asteraceae	<i>Gerbera hybrida</i>

PephAP3	Piperales	Aristolochiaceae	<i>Peperomia hirta</i>
PephPI	Piperales	Aristolochiaceae	<i>Peperomia hirta</i>
PethGP	Solanales	Solanaceae	<i>Petunia hybrida</i>
PethMADS2	Solanales	Solanaceae	<i>Petunia hybrida</i>
PethTM6	Solanales	Solanaceae	<i>Petunia hybrida</i>
<i>PicaDAL11</i>	Coniferales	Pinaceae	<i>Picea abies</i>
<i>PicaDAL12</i>	Coniferales	Pinaceae	<i>Picea abies</i>
<i>PicaDAL13</i>	Coniferales	Pinaceae	<i>Picea abies</i>
PipmPI-1	Piperales	Piperaceae	<i>Piper magnificum</i>
PipmPI-2	Piperales	Piperaceae	<i>Piper magnificum</i>
PpMADS2			<i>Physcomitrella patens</i>
<i>PrDGL</i>	Coniferales	Pinaceae	<i>Pinus radiata</i>
RanfAP3	Ranunculales	Ranunculaceae	<i>Ranunculus ficaria</i>
RanfPI	Ranunculales	Ranunculaceae	<i>Ranunculus ficaria</i>
RrMASAKOBP	Rosales	Rosaceae	<i>Rosa rugosa</i>
RrMASAKOeuB3	Rosales	Rosaceae	<i>Rosa rugosa</i>
RumaRaD1	Caryophyllales	Polygonaceae	<i>Rumex acetosa</i>
RumaRaD2	Caryophyllales	Polygonaceae	<i>Rumex acetosa</i>
SancAP3	Ranunculales	Ranunculaceae	<i>Sanguinaria canadensis</i>
SillSLM2	Caryophyllales	Caryophyllaceae	<i>Silene latifolia</i>
SillSLM3	Caryophyllales	Caryophyllaceae	<i>Silene latifolia</i>
SmAP3	Alismatales	Alismataceae	<i>Sagittaria montevidensis</i>
SyrvAP3	Lamiales	Veronicaceae	<i>Syringa vulgaris</i>
SyrvPI	Lamiales	Veronicaceae	<i>Syringa vulgaris</i>
TaccAP3	Dioscoreales	Dioscoreaceae	<i>Tacca chantieri</i>
TaccPI	Dioscoreales	Dioscoreaceae	<i>Tacca chantieri</i>
TracAP3	Ranunculales	Ranunculaceae	<i>Trautvetteria carolinensis</i>
TracPI-1	Ranunculales	Ranunculaceae	<i>Trautvetteria carolinensis</i>
TracPI-2	Ranunculales	Ranunculaceae	<i>Trautvetteria carolinensis</i>
TriaMADS51	Poales	Poaceae	<i>Triticum aestivum</i>
Zeam m16	Poales	Poaceae	<i>Zea mays</i>
Zeam m18	Poales	Poaceae	<i>Zea mays</i>
ZeamSilky1	Poales	Poaceae	<i>Zea mays</i>
ZeamZPI1bm29	Poales	Poaceae	<i>Zea mays</i>

Referencias

- Álvarez-Buylla, E.R., Pelaz, S., Liljergren, S.J., Gold, S.E., Burgeff, C., Ditta, G.S., Ribas, D.P., Martínez-Castilla, L., y M.F. Yanofsky. 2000. An ancestral MADS-Box gene duplication occurred before the divergence of plants and animals. *Proc. Natl. Acad. Sci. USA*. 97(10):5328-5333
- Ambrose, B.A., Lerner, D.R., Ciceri, P., Padilla, C.M., Yanofsky, M.F., y R.J. Schmidt. (2000). Molecular and genetic analyses of the *Silki1* gene reveal conservation in floral organ specification between eudicots and monocots. *Molecular Cell* 5, 569-579
- Angiosperm Phylogeny Group II. 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APGII. *Bot. Jour. Linn. Soc.* 141: 399-436
- Anisimova, M., Bielawski, J.P. y Z. Yang. Accuracy and Power of the Likelihood Ratio Test in Detecting Adaptive Molecular Evolution. *Mol. Biol. Evol.* 18(8):1585-1592.
- Becker, A., Winter, K., Meyer, B., Saedler, H. y G. Theissen. 2000. MADS-Box Gene Diversity in Seed Plants 300 Million Years Ago. *Mol. Biol. Evol.* 17(10):1425-1434.
- Becker, A., Kaufmann, K., Freialdenhoven, A., Vincent, C., Li, M.-A., Saedler, H. y G. Theissen. 2002. A novel MADS-Box gene subfamily with a sister-group relationship to class B floral homeotic genes. *Mol. Genet. Genomics*. 266:942-950.
- Bielawski, J.P. y Z. Yang. 2004. A Maximum Likelihood Method for Detecting Functional Divergence at Individual Codon Sites, with Application to Gene Family Evolution. *J. Mol. Evol.* 59:121-132.
- Bowman, J.L., D.R. Smyth y E.M. Meyerowitz, 1991. Genetic interactions among floral homeotic genes of *Arabidopsis*. *Development*. 112: 1-20.
- Burleigh, J.G. y S. Mathews. 2004. Phylogenetic Signal in Nucleotide Data From Seed Plants: Implications for Resolving the Seed Plant Tree of Life. *Am. J. Botany*. 91(10):1599-1613.
- Bush, R.M. et al. 1999. Positive Selection on the H3 hemagglutinin gene of human influenza virus. *A. Mol. Biol. Evol.* 16, 1457-1465. (en Yang y Bielawski, 2000)
- Carroll, S.B., Grenier, J.K. y S.D. Weatherbee. (2001) From DNA to Diversity. Molecular genetics and the Evolution of Animal Design. Blackwell Science, Malden, Massachusetts, USA.
- Chase, M.W., D.E. Soltis, R.G. Olmstead, D. Morgan, D.H. Les et al., 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcl*. *Ann. Miss. Bot. Gard.* 80: 528-580.
- Coen, E.S. y E.M. Meyerowitz. 1991. The war of the whorls: genetic interactions controlling flower development. *Nature*. 353: 31-37.
- Crandall, K.A., Kelsev, C.R., Imamichi, H., Lane, H.C., y N.P. Salzman. 1999. Parallel evolution of drug resistance in HIV: failure of nonsynonymous/synonymous substitution rate ratio to detect selection. *Mol. Biol. Evol.* 16:372-382.

- Crane, P.R., E.M. Friis and K.R. Pedersen, 1995. The origin and early diversification of angiosperms. *Nature* 374: 27-33.
- Crepet, W.L. y E.M. Friis. 1987. The evolution of insect pollinators in angiosperms, en Friis, Else M., Chaloner, W.G. y Craene, P.R. The Origins of Angiosperms and their Biological Consequences. International Congress of Systematic and Evolutionary Biology (3rd, 1985, University of Sussex). Cambridge Univ. Press, 1989, c1987 New York
- Creevey, C.J. y J.O. McInerney. 2002. An algorithm for detecting directional and non-directional positive selection, neutrality and negative selection in protein coding DNA sequences. *Gene*, 300:43-51
- Cronquist, A., 1988. The Evolution and Classification of Flowering Plants. The New York Botanical Garden, New York.
- Davies , B. Y Z. Schwarz-Sommer, 1994. Control of floral organ identity by homeotic MADS-Box transcription factors. *Results Probl. Cell Differ.* 20: 235-258.
- Davies, B., Egea-Cortines, M., de Andrade Silva, E., Saedler, H. y Sommer, H. 1996. Multiple interactions amongst floral homeotic MADS box proteins. *EMBO J.* 16, 4330–4343.
- Doyle, J.A. y M.J. Donoghue. 1987. The Origin of Angiosperms: a cladistic approach. En Friis, Else M., Chaloner, W.G. y Craene, P.R. The Origins of Angiosperms and their Biological Consequences. International Congress of Systematic and Evolutionary Biology (3rd, 1985, University of Sussex). Cambridge Univ. Press, 1989, c1987 New York
- Donoghue, M.J. y J.A. Doyle. 1989. Phylogenetic analysis of angiosperms and the relationships of Hamamelidae. In P.R. Crane and S. Blackmore [eds.], Evolution, systematics, and fossil history of the Hamamelidae, vol. 1, Introduction and “lower” Hamamelidae. Systematics Association Special Volume 40A, 17-45. Clarendon Press, Oxford, UK.
- Donoghue, M.J. y Doyle, J.A. 2000. Seed plant phylogeny: Demise of the anthophyte hypothesis?. *Current Biol.* 10: R106-R109
- Doyle, J.J. y C.L. Hotton. 1991. Diversification of early angiosperm pollen in a cladistic context. In S. Blackmore and S.H. Barnes [eds.] Pollen and spores: patterns of diversification, 165-195. Clarendon Press. Oxford. UK.
- Drinnan, A.N., P.R. Crane y S.B. Hoot. 1994. Patterns of floral evolution in the early diversification of the non magnoliid dicotyledons (eudicots). *Plant Systematics and Evolution (suppl.)* 8:93-122.
- Egea-Cortines, M., Saedler, H. y H. Sommer. 1999. Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS, and GLOBOSA is involved in the control of floral architecture in *Antirrhinum majus*. *Embo j.* 18, 5370-5379
- Endo, T., Ikeo, K. y T. Gojobori. 1996. Large-scale search for genes on which positive selection may operate. *Mol. Biol. Evol.* 13: 685-690
- Endress, P. K., 1994. Floral Structure and Evolution of Primitive Angiosperms: Recent Advances. *Plant Syst. Evol.* 192: 79-97.

- Endress, P.K. y A. Igersheim. 2000. Gynoecium Structure and Evolution in Basal Angiosperms. *Int. J. Plant Sci.* 161(6 suppl.):S211-S223.
- Endress, P.K. 2001. Origins of Flower Morphology. *Journal of Experimental Zoology (Mol. Dev. Evol.)* 291:105-115.
- Espinosa-Soto, C., Padilla-Longoria, P. y E.R. Álvarez-Buylla. 2004. A Gene Regulatory Network Model for cell-Fate Determination during *Arabidopsis thaliana* Flower Development That Is Robust and Recovers Experimental Gene Expression Profiles. *The Plant Cell*. 16: 2923-2939.
- Felsenstein, J. 2004. Inferring Phylogenies. Sinauer Associates, Inc. Publishers. Sunderland, MA. USA.
- Fitch. W.M. et al., 1997. Long term trends in the evolution of H(3) HA1 human influenza type A. *Proc. Natl. Acad. Sci. USA.* 94,7712-7718. (en Y ang y Bielawski, 2000)
- Friis, E.M. y W.L. Crepet. 1987. Time of appearance of floral features. En *The Origins of Angiosperms and Their Biological Consequences*, E.M. Friis, W.G. Chaloner, P.R. Crane. Cambridge Univ. Press, Cambridge, 1987, pp. 145-179
- Friis, Else M., Chaloner, W.G. y Craene, P.R. 1989. *The Origins of Angiosperms and their Biological Consequences*. International Congress of Systematic and Evolutionary Biology (3rd, 1985, University of Sussex). Cambridge Univ. Press, 1989, c1987 New York
- Fukui, M., Futamura, N., Mukai, Y., Wang, Y., Nagao, A. y K. Shinohara. 2001. Ancestral MADS Box Genes in Sugi, *Cryptomeria japonica* D. Don (Taxodiaceae), Homologous to the B Function Genes in Angiosperms. *Plant Cell Physiol.* 42(6):566-575
- Goldman, N., y Z. Yang. 1994. A codon based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* 11: 725-736
- Goto, K. y Meyerowitz, E.M. 1994. Function and regulation of the *Arabidopsis* floral homeotic gene *PISTILLATA*. *Genes Dev.* 8:1548-1560.
- Hall, T.A. 1999. BioEdit. Sequence alignment editor. *Nucleic Acids Symp. Ser.* 41: 95-98.
- Homma, T. y K. Goto. 2001. Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature*, 409, 525-529
- Huelsenbeck, J.P. y F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics.* 17: 754-755.
- Hughes, N.F. 1994. *The enigma of angiosperm origins*. Cambridge University Press. Cambridge, UK.
- Soltis, P.S., Soltis, D.E., Savolainen, V., Crane, P.R. y T.G. Barraclough. 2002. rate heterogeneity among lineages of tracheophytes: integration of molecular and fossil data and evidence of molecular living fossils *Proc. Natl. Acad. Sci. USA.* 99: 4430-4435
- Hsu, H-F. y C-H. Yang. 2002. An Orchid (*Oncidium* Gower Ramsey) AP3-like MADS Gene Regulates Floral Formation and Initiation. *Plant Cell Physiol.* 43(10):1198-1209.

- Jack T, Brockman L.L., Meyerowitz, E.M. 1992. The homeotic gene *APETALA3* of *Arabidopsis thaliana* encodes a MADS box and is expressed in petals and stament. *Cell*, 68, 683-697
- Jack, T., Fox, G.L. y E.M. Meyerowitz. 1994. *Arabidopsis* homeotic gene *APETALA3* ectopic expression: transcriptional and posttranscriptional regulation determine floral organ identity. *Cell*. 76:703-716.
- Judd. W.S. y R.G. Olmstead. 2004. A Survey of Tricolpate (Eudicot) Phylogenetic Relationships. *American Journal of Botany*, 91(10): 1627-1644.
- Juenger, T., Purugganan, M. & T. F. Mackay. 2000. Quantitative Trait Loci for Floral Morphology in *Arabidopsis thaliana*. *Genetics* 156:1379-1392
- Kang, H-Y., Jeon, J-S., Lee, S. y G. An. 1998. Identification of class B and class C floral organ identity genes from rice plants. *Plant Mol. Biol.* 38: 1021-1029.
- Kim, S., Yoo, M-J., Albert, V.A., Farris, J.S., Soltis, P.S. y D.E. Soltis. 2004. Phylogeny and Diversification of B-function MADS-box genes in angiosperms: evolutionary and functional implications of a 260-million-year-old duplication. *Am. J. Botany*. 91(12):2102-2118.
- Kramer, E.M., Dorit, R.L. and V.F. Irish, 1998. Molecular Evolution of genes Controlling Petal and Stamen Development: Duplication and Divergence Within the *APETALA3* and *PISTILLATA* MADS-Box Gene Lineages. *Genetics*. 149: 765-783.
- Kramer, E.M. y V.F. Irish. 1999. Evolution if the genetic mechanisms controlling petal development. *Nature*. 399,144-148.
- Kramer, E.M. y V. F. Irish. 2000. Evolution of the petal and stamen developmental programs: evidence from comparative studies of the lower eudicots and basal angiosperms. *Int. J. Plant Sci.* 161(6 Suppl.):S29-S40.
- Kramer, E.M., Di Stilo, V.S. y P.M. Schlüter. 2003. Complex patterns of gene duplication in the *APETALA3* and *PISTILLATA* lineages of the Ranunculaceae. *Int. J. Plant Sci.* 164(1):1-11.
- Krizek, B.A. y E.M. Meyerowitz, 1996. Mapping the protein regions responsible for the functional specificities of the *Arabidopsis* MADS domain organ-identity proteins. *Proc. Nat. Acad. Sci. UNA*. 93: 4063-4070.
- Lamb, R.S. y V. Irish. 2004. Functional divergence within the *APETALA3/PISTILLATA* floral homeotic gene lineages. *Proc. Natl. Acad. Sci.* 100(11):6558-6563.
- Lawton-Rauh, A.L., Álvarez-Buylla, E.R. y M.D. Purugganan. 2000. Molecular Evolution of Flower Development. *TREE*. 15:144-149.
- Lupia, R., S. Lidgard, and P.C. Crane. 1998. Comparing palynological abundance and diversity: implications for biotic replacement during the Cretacious angiosperm radiation. *Paleobiology*. 25:305-340.
- Ma, H., M.F. Yanofsky y E.M. Meyerowitz. 1991. *AGL1-AGL6*, an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes. *Genes Dev*, 5: 484-495.
- Magallón, S., Crane, P. y P. S. Herendeen. 1999. Phylogenetic Pattern, Diversity and Diversification of Eudicots. *Ann. Missouri Bot. Gard.* 86:297-372.

- Magallón, S. y M.J. Sanderson. 2001. Absolute diversification rates in angiosperm clades. *Evolution*. 55(9):1762-1780.
- Mapes, G. y G.W. Rothwell. 1984. Permineralized ovulate cones of *Lebachia* from the late Paleozoic limestones of Kansas. *Paleontology*. 27:69-94.
- Mapes, G. y G.W. Rothwell. 1991. Structure and relationships of primitive conifers. *Neues Jahrbuch für Geologie and Paläontologie-Abhandlungen*. 183:269-287.
- Martinez-Castilla, L.P. y Álvarez-Buylla, E.R. 2003. Adaptive evolution in the *Arabidopsis* MADS-box gene family inferred from its complete resolved phylogeny. *Proc. Nat. Acad. Sci. USA*. 100(23):13407-13412.
- Meyerowitz, E.M., J. L. Bowman, L.L. Brockman, G. Drews, T. Jack et al., 1991. A genetic and molecular model for floral development in *Arabidopsis thaliana*. *Development*. 1: 157-167.
- Miyata, T. y T. Yasunaga. 1980. Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and aminoacid substitutions from homologous nucleotide sequences and its applications. *J. Mol. Evol.* 16:23-36
- Mouradov, A., Hamdorf, B., Teasdale, R.D., Kim, J.T., Winter, K-U y G. Theissen. 1999. A *DEF/GLO*-like MADS-Box gene from a gymnosperm: *Pinus radiata* contains an ortholog of angiosperm B class floral homeotic genes. *Developmental Genetics*, 25: 245-252.
- Münster, T., Pahnke, J., Di Rosa, A., Kim, J.T., Martin, W., Saedler, H. y G. Theissen. 1997. Floral homeotic genes were recruited from homologous MADS-box genes preexisting in the common ancestor of ferns and seed plants. *Proc. Natl. Acad. Sci.* 94: 2415-2420.
- Münster, T., Wingen, L.U., Faigl, W., Werth, S., Saedler, H. and G. Theissen. 2001. Characterization of three *GLOBOSA*-like MADS-Box genes from maize: evidence for ancient paralogy in one class of floral homeotic B-function genes of grasses. *Gene*. 262:1-13.
- Nagasawa, N., Miyoshi, M., Sano, Y., Satoh, H., Hirano, H., Sakai, H., y Y. Nagato. 2003. *SUPERWOMAN1* and *DROOPING LEAF* genes control floral organ identity in rice. *Development* 130:705-718.
- Nam, J., dePamphilis, C.W., Ma, H. y M. Nei. 2003. Antiquity and Evolution of the MADS-Box Gene Family Controlling Flower Development in Plants. *Mol. Biol. Evol.* 20(9):1435-1447.
- Nei M. y S. Kumar. *Molecular Evolution and Phylogenetics*. Oxford University Press, 2000. New York.
- Norman, A., M. Runswick, R. Pollock y R. Treisman. 1988. Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. *Cell* 55. 989-1003.
- Ohta, T. 1993. Pattern of nucleotide substitution in growth hormone-prolactin gene family: a paradigm for evolution by gene duplication. *Genetics* 134:1271-1276
- Pnueli, L., M. Abu-Abeid, D. Zamir, W. Nacken, Z. Schwarz-Sommer et al. 1991. The MADS-Box gene family in tomato: temporal expression during floral development, conserved secondary structures and homology with homeotic genes from *Antirrhinum* and *Arabidopsis*. *Plant J.* 1: 255-266.

- Pollock, R. y R. Treisman. 1991. Human SRF-related proteins: DNA- binding properties and potential regulatory targets. *Genes Dev.* 5: 2327-2341.
- Posada, D. y K.A. Crandall. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics.* 14(9): 817-818.
- Prance, G.T. 1977. Floristic inventory of the tropics: Where do we stand?. *Annals of the Missouri Botanical Garden.* 64:659-84
- Purugganan, M.D. 1997. The MADS-Box Floral Homeotic Gene Lineages Predate the Origin of Seed Plants: Phylogenetic and Molecular Clock Estimates. *J. Mol. Evol.* 45:392-396.
- Qiu, Y.-L., M.W. Chase, D.H. Les and C. R. Parks. 1993. Molecular phylogenetics of the Magnoliidae; cladistic analyses of nucleotide sequences of the plastid gene rbcL. *Ann. Miss. Bot. Gard.* 80: 587-606.
- Raven, P.H., Evert, R.F. y S.E. Eichorn. *Biology of Plants.* Sixth Edition. W.H. Freeman and Company Worth Publishers. New York, 1999.
- Riechmann, J.L., B.A. Krizek y E.M. Meyerowitz, 1996. Dimerization specificity of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLAT and AGAMOUS. *Proc. Nat. Acad. Sci. USA.* 93: 4793-4798.
- Riechmann, J.L., Wang, M. y Meyerowitz, E.M. 1996b. DNA binding properties of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS. *Nucl Acids Res.* 24, 3134-3141.
- Riechmann, J.L. y E.M. Meyerowitz. 1997. MADS domain proteins in plant development. *Biol. Chem.* 378, 1079/1101
- Riechmann, J.L. y E.M. Meyerowitz. 1997. Determination of Floral Organ Identity by *Arabidopsis* MADS Domain Homeotic Proteins AP1, AP3, PI, and AG Is Independent of Their DNA-binding Specificity. *Mol. Biol. Of the Cell.* 8, 1243-1259.
- Riense de Craene, L.P. 2004. Floral development of *Berberidopsis corallina*: a crucial link in the evolution of flowers in core-eudicots. *Annals of Botany.* 94:741-751.
- Sanderson, M.J., 1997. A nonparametric approach for estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.* 14: 1218-1232.
- Sanderson, M.J., 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol. Biol. Evol.* 19: 101-109.
- Shore , P. y Sharrocks, A.D. 1995. The MADS-box family of transcription factors. *Eur. J. Biochem.* 229, 1-13
- Schwarz-Sommer, Z., Hue, I., Huijser, P., Flor, P.J., Hansen, R., Tetens, F., W-E., Lonngig, et al. 1992. Characterization of the *Antirrhinum* floral homeotic MADS-box gene *deficiens*: evidence for DNA binding and autoregulation of its persistent expression throughout flower development. *EMBO J.* 11: 251-263
- Sporne, K.R. 1974. *The Morphology of Angiosperms.* London: Hutchinson

- Stebbins, G.L. 1981. Why are there so many species of flowering plants? *Bioscience*. 31: 573-577
- Stellari, G.M., Jaramillo, M.A. y E. M. Kramer. 2004. Evolution of the *APETALA3* and *PISTILLATA* Lineages of MADS-Box Containing Genes in the Basal Angiosperms. *Mol. Biol. Evol.* 21(3):506-519
- Sundström, J., Carlsbecker, A., Svensson, M.E., Svensson, M., Johanson, U., Theissen, G. y P. Engström. 1999. MADS-Box Genes Active in Developing Pollen Cones of Norway Spruce (*Picea abies*) Are Homologous to the B-Class Floral Homeotic Genes in Angiosperms. *Developmental Genetics*. 25:253-266.
- Sündstrom, J. y P. Engstrom. 2002. Conifer reproductive development involves B-type MADS-box genes with distinct and different activities in male organ primordia. *The Plant Journal*. 31(2): 161-169.
- Suzuki, Y. y T. Gojobori. 1999. A method for detecting positive selection at single aminoacid sites. *Mol. Biol. Evol.* 16, 1457-1465
- Suzuki, Y. 2004. New Methods for Detecting Positive Selection at Single Amino Acid Sites. *J. Mol. Evol.* 59:11-19
- Swofford, D. L. 1991. PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1 Computer program distributed by the Illinois Natural History Survey, Champaign, Illinois.
- Takhtajan, A., 1991. Evolutionary Trends in Flowering Plants. Columbia University press, New York.
- Takhtajan, A. 1997. Diversity and classification of flowering plants. Columbia Univ. Press, New York.
- Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* 10, 512-526.
- Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. In: *Some mathematical questions in biology - DNA sequence analysis* (ed. Miura RM), pp. 57-86. Amer. Math. Soc., Providence, RI.
- Theissen, G., y H. Saedler. 1995. MADS-box genes in plant ontogeny and phylogeny: Haeckel's "biogenetic law" revisited. *Current Opinion in Genetics and Development*, 5:628-639
- Thompson, J.D., Higgins, D.G. y Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22:4673-4680.
- Thorne, R.F. 1992. Classification and geography of the flowering plants. *The Botanical Review* 58:225-384.
- Vandenbussche, M., Zethof, J., Royaert, S., Weterings, K. y T. Gerats. 2004. The Duplicated B-Class Heterodimer Model: Whorl-Specific Effects and Complex Genetic Interactions in *Petunia hybrida* Flower Development. *The Plant Cell*. 16:741-754.
- Winter, K., Becker, A., Münster, T., Kim, J.T., Saedler, H. y G. Theissen. 1999. MADS-Box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. *Proc. Natl. Acad. Sci.* 96:7342-7347.

- Winter, K., Weiser, C., Kaufmann, K., Bohne, A., Kirchner, C., Kanno, A., Saedler, H. and G. Theissen. 2002. Evolution of Class B Floral Homeotic Proteins: Obligate Heterodimerization Originated from Homodimerization. *Mol. Biol. Evol.* 19(5):587-596.
- Whipple, C.J., Ciceri, P., Padilla, C.M., Ambrose, B.A., Bandong, S. L., y Schmidt, R.J. (2004) Conservation of B-class floral homeotic gene function between maize and *Arabidopsis*. *Development* 131, 6083-6091.
- Yang, Y., Fanning, L. y T. Jack. 2003. The K domain mediates heterodimerization of the *Arabidopsis* floral organ identity proteins, APETALA3 and PISTILLATA. *The Plant Journal*, 33, 47-59
- Yang, Y., Xiang, H.M. y Jack, T. 2003 *plastillata-S*, an *Arabidopsis* B class mutant with strong defects in petal but not in stamen development. *Plant J.* 33:177-188.
- Yang, Y. y T. Jack. 2004. Defining subdomains of the K domain important for protein-protein interactions of plant MADS proteins. *Plant Molecular Biology*. 55:45-49.
- Yang, Z., Nielsen, R., Goldman, N. y A.K. Pedersen. 2000. Codon-Substitution Models for Heterogeneous Selection Pressure at Amino Acid Sites. *Genetics*. 155:431-449
- Yang, Z. y J.P. Bielawski. 2000. Statistical methods for detecting molecular adaptation. *TREE*, 15(12):496-503
- Yang, Z. y W.J. Swanson. 2002. Codon-Substitution Models to Detect Adaptive Evolution that Account for Heterogeneous Selective Pressures Among Site Classes. *Mol. Biol. Evol.* 19(1):49-57
- Yang, Z. 2002. Inference of selection from multiple species alignments. *Curr. Op. Genetics & Development*. 12:688-694
- Yang, Z. y R. Nielsen. 2002. Codon-Substitution Models for Detecting Molecular Adaptation at Individual Sites Along Specific Lineages. *Mol. Biol. Evol.* 19(6): 908-917
- Yang, Z. 2004. A heuristic rate smoothing procedure for maximum likelihood estimation of species divergence times. *Acta Zoologica Sinica*. 50(4):645-656.
- Zanis, M., Soltis, P.S., Qiu, Y.L., Zimmer, E., y D.E. Soltis. 2003. Phylogenetic analyses and perianth evolution in basal angiosperms. *Annals of the Missouri Botanical Garden* 90: 129-150.