UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

INSTITUTO DE ECOLOGIA

Genes MADS-box en el desarrollo de la raíz de *Arabidopsis thaliana*: Patrones de expresión *in situ* y análisis filogenético

TESIS QUE PRESENTA

CAROLINE NICOLE LAURA BURGEFF D'HONDT

PARA OBTENER EL GRADO DE
DOCTOR EN CIENCIAS BIOMEDICAS

DIRECTORA DE TESIS:
ELENA ALVAREZ - BUYLLA ROCES

MEXICO, D. F. SEPTIEMBRE 2001

INDICE

1. Genética molecular y evolución del desarrollo en plantas

1.1. Genes reguladores del desarrollo y aparición de nuevas morfologías

En los últimos quince años, el enfoque de la genética molecular ha permitido un avance importante en el estudio de los procesos y los mecanismos moleculares que dirigen el desarrollo de los seres vivos. Los genes que controlan distintos aspectos del desarrollo interactuan formando redes de regulación, cuya estructura básica parece haberse conservado durante la evolución. Ciertos tipos de factores de transcripción han demostrado tener una participación fundamental en los procesos de regulación del desarrollo. Estas proteínas tienen la capacidad de unirse a secuencias *cis*-regulatorias del ADN y, en combinación con otros interactores, modular la expresión de los llamados genes "blanco" (Schwechheimer y Bevan 1998). Los patrones espacio-temporales de expresión de los genes de ciertos factores de transcripción, han mostrado ser determinantes para la generación de los planes corporales de los organismos (e.g. Burke et al. 1995; Palopoli y Patel 1996). Modificaciones en las secuencias *cis*-regulatorias en este tipo de genes pueden alterar sus patrones espacio temporales de expresión, lo cual parece constituir un mecanismo importante para la generación de cambios en los planos corporales que rigen la aparición de nuevos fenotipos, sobre los cuales actúan las fuerzas evolutivas (Doebley y Lukens 1998; Shimeld 1999; Carroll 2000; Lee y Schiefelbein 2001). De esta manera, mutaciones en ciertos genes reguladores pueden provocar variaciones no lineales sobre el fenotipo e incidir en la aparición de nuevas morfologías durante la evolución (Palopoli y Patel 1996).

Los factores de transcripción se agrupan en familias multigénicas, delimitadas generalmente por las características estructurales de estas proteínas y, en particular, de sus dominios de unión al ADN (Schwechheimer y Bevan 1998; Liu et al. 1999). Los

análisis filogenéticos basados en la comparación de las secuencias de sus miembros, permiten tener una visión de la evolución de la familia, determinando el orden y nivel de duplicación y divergencia de cada uno de los genes, así como de los eventuales clados[1] que estos formen (Li 1997). La integración de análisis funcionales y filogenéticos de factores de transcripción que controlan aspectos del desarrollo, permite obtener información importante sobre su reclutamiento en la realización de funciones regulatorias novedosas, así como sobre los posibles paralelos existentes entre la evolución de la familia y la evolución morfológica de los seres vivos (Purugganan 1998).

1.2. Estrategias de estudio de la genética del desarrollo en plantas

Los estudios de la genética del desarrollo en plantas son relativamente recientes, comparados con la cantidad de información acumulada durante décadas de estudios de los modelos de desarrollo animal. Las plantas poseen ciertas peculiaridades tales como la presencia de una pared rígida que envuelve a las células y que impide que haya procesos de migración celular como en los animales; por otro lado, el plan corporal de base se establece durante la embriogénesis, pero la morfología que la planta adopta durante el desarrollo postembrionario resulta de una organogénesis continua, producto de la actividad de los meristemos (Jürgens et al. 1991; Meyerowitz 1997). La genética del desarrollo en plantas ha avanzado en la última década, principalmente en la comprensión de los mecanismos moleculares que dirigen el desarrollo floral. Gran parte de los trabajos se han realizado en *Arabidopsis thaliana* (L.) Heynh., la cual se ha convertido en el modelo experimental de la genética en plantas, gracias a características tales como su tamaño, su ciclo corto de vida y su genoma pequeño (de alrededor de

[1] Un clado o grupo monofilético está formado por el conjunto de genes, o especies, que descienden de un ancestro (gen o especie) común (Li 1998)

125 Mpdb distribuidos en 5 cromosomas), el cual ha sido recientemente secuenciado (Meyerowitz y Pruitt 1985; The Arabidopsis genome initiative 2000).

La estrategia clásica de estudio de la biología del desarrollo consiste en el análisis genético de mutantes (Wolpert et al. 1998). Los mutantes que presentan alteraciones fenotípicas en su desarrollo son caracterizados y analizados genéticamente. Así, la posición y la identidad del locus afectado es determinada. La clonación del gen a partir del ADN de plantas con fenotipo silvestre permite insertarlo por transformación en plantas mutantes con pérdida de función. Si las plantas resultantes de la transformación recuperan el fenotipo silvestre, se confirma que la mutación del gen en cuestión es responsable de la alteración del fenotipo observado. Las preguntas funcionales más específicas para cada caso pueden ser exploradas experimentalmente, realizando las pruebas moleculares, genéticas y/o bioquímicas que se consideren pertinentes.

Una estrategia alternativa es la de la genética inversa, la cual ha surgido a partir del auge de las técnicas moleculares y de los proyectos de secuenciación genómica de diversos organismos (Howell 1998; Riechmann y Ratcliffe 2000). En este caso, el estudio de la función de un gen inicia a partir de la obtención de su secuencia. La primera etapa de este estudio consiste en reunir y analizar los datos existentes previos a la caracterización genética y funcional propiamente dicha. En el caso de un gen que codifica para un factor de transcripción, esta primera etapa consiste en determinar su pertenencia a alguna familia génica conocida, su clonación, así como la descripción de su patrón de expresión, entre otros (Kranz et al. 1998). La segunda etapa consiste en obtener mutantes que anulen o alteren la expresión del gen y que permitan analizar las consecuencias que esto tiene sobre el fenotipo general de la planta. Esto se realiza generalmente haciendo escrutinios o tamizados [*screenings*] de colecciones de líneas

mutantes, por ejemplo las resultantes de inserciones de T-DNA o transposones (Meissner et al. 1999; Pelaz et al. 2000; Riechmann y Ratcliffe 2000).

Los análisis filogenéticos basados en las secuencias de los miembros de la familia génica estudiada constituyen, a su vez, una de las herramientas de estudio para la primera etapa de la genética inversa. La obtención de datos, relativos a la ortología existente entre genes de distintas especies y la paralogía entre genes hermanos, proporciona claves importantes sobre la posible existencia de conservación y redundancia funcionales (Purugganan 1998; Alvarez-Buylla et al. 2000a y b; Smyth 2000). Este tipo de información, integrada con el resto de los datos obtenidos en esta primera etapa del estudio -tales como la determinación de los patrones de expresión de los genes de interés-, contribuye a dirigir de manera razonada el análisis funcional, una vez obtenidos los mutantes. En efecto, la determinación del patrón de expresión define las regiones de la planta donde el ARN mensajero (ARNm) del gen está presente, lo que delimita inicialmente[2] la estructura, órgano o tejido que se vería afectado en el mutante; esto permite dirigir el tipo de observaciones a realizar durante el análisis fenotípico. La determinación de la posición filogenética, a su vez, nos puede dar claves sobre la eventual pertenencia del gen a un clado que en ocasiones delimita a un grupo funcional (Purugganan et al. 1995).

El estudio funcional de genes por genética inversa se confronta, en algunas ocasiones, con el caso de mutantes que no presentan alteración del fenotipo. Esto puede deberse a que la actividad del gen se integra exclusivamente en vías de respuesta morfogenética a ciertos estímulos ambientales particulares, y/o a que presenta cierto grado de redundancia. Esta última se puede dar por la existencia de vías o rutas regulatorias independientes que controlan el mismo proceso, o bien por la existencia de

[2] Dejando al margen casos en los que exista una relocalización pronunciada de la proteína.

distintos miembros de la familia multigénica, cuyas funciones se traslapan por lo menos parcialmente (Thomas 1993). En general, estos casos de redundancia funcional a nivel molecular[3] pueden involucrar genes provenientes de duplicaciones recientes que comparten patrones de expresión similares (e.g. Kempin et al. 1995; Meissner et al. 1999; Liljegren et al. 2000; Smyth 2000). Así, el análisis filogenético, en combinación con la determinación de los patrones de expresión, permite dirigir los estudios genéticos que lleven a revelar el fenotipo mutante gracias a realización de cruzas dirigidas para la obtención de mutantes dobles y triples.

Este trabajo está centrado en la familia de genes MADS-box que codifican para factores de transcripción cuyos miembros más estudiados están involucrados en la regulación del desarrollo floral. El estudio se sitúa en el marco de la primera etapa de una estrategia de análisis por genética inversa, cuya finalidad es la caracterización de algunos miembros de esta familia que se expresan en la raíz de *Arabidopsis thaliana*.

[3] A continuación utilizaremos el término de redundancia funcional refiriéndonos exclusivamente a este segundo caso.

2. La familia de Genes MADS-box en el desarrollo de *Arabidopsis thaliana*

Los genes MADS-box codifican para factores de transcripción que presentan una estructura modular, la cual se caracteriza por la presencia de un dominio muy conservado de alrededor de 60 aminoácidos. Este es el dominio MADS, que media la unión al ADN y que le da el nombre a la familia, el cual proviene del acrónimo formado por las iniciales de los cuatro primeros miembros caracterizados en distintas especies de eucariontes (*MCM1*, *AGAMOUS*, *DEFICIENS*, *SRF*) (Schwarz-Sommer et al. 1990; Treisman y Ammerer 1992; Shore y Sharrocks 1995). En plantas, la mayoría de estos factores de transcripción presentan también 3 dominios adicionales. El dominio intermedio I, que puede ser muy variable. El dominio K, cuya secuencia de 70 aminoácidos es relativamente conservada, está involucrado en la formación de hélices anfipáticas que median las interacciones entre proteínas. Finalmente, el extremo carboxilo terminal es también una zona muy variable y es putativamente la región de transactivación de la proteína (Purugganan et al. 1995; Riechmann y Meyerowitz 1997; Cho et al. 1999). Esta estructura modular ha sido denominada la estructura MIKC (figura1A) (Alvarez-Buylla et al. 2000a).

Los análisis genéticos y funcionales que identificaron a los primeros miembros de la familia MADS-box en *Arabidopsis* (tabla 1), pusieron de manifiesto su función durante el desarrollo floral[4], en particular en la determinación de la identidad de los meristemos (*APETALA1* (*AP1*), *CAULIFLOWER* (*CAL*), *FRUITFULL* (*FUL*)) y de los órganos florales (*AP1*, *APETALA3* (*AP3*), *PISTILLATA* (*PI*), *AGAMOUS* (*AG*))

[4] La flor de *Arabidopsis thaliana* está constituída de cuatro verticilos concéntricos. El primero y más externo está formado de cuatro sépalos, el segundo de cuatro pétalos, el tercero de seis estambres y el cuarto de dos carpelos fusionados.

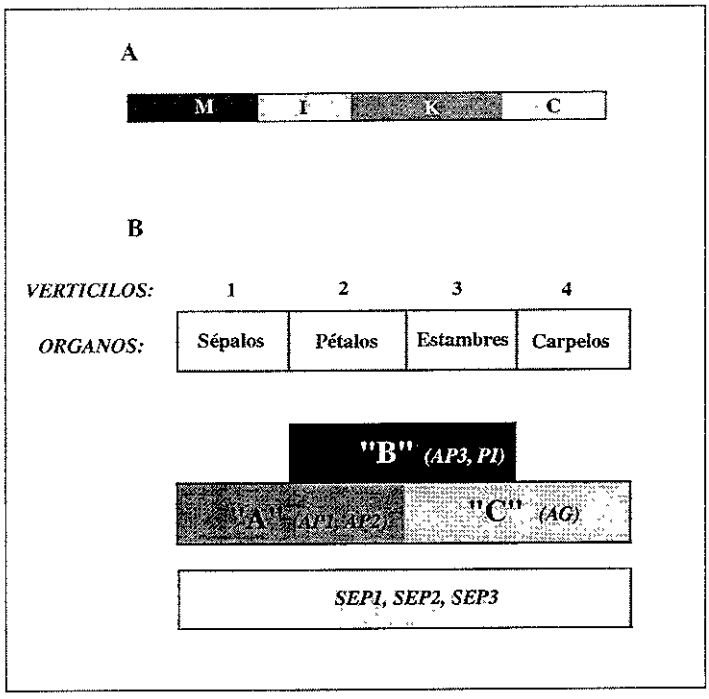

Figura 1. A: Esquema de los dominios protéicos de los factores de transcripción MADS pertenecientes al linaje tipo II (Alvarez-Buylla et al. 2000a). M: dominio MADS; I: dominio I; K: dominio K; C: dominio carboxilo terminal. B: Esquema del modelo ABC de determinación de la identidad de órganos en los distintos verticilos de la flor de *Arabidopsis*.

(Yanofsky et al. 1990; Coen y Meyerowitz 1991; Jack et al. 1992; Mandel et al. 1992; Kempin et al. 1995; Ferrándiz et al. 2000a). Líneas mutantes para estos últimos, también conocidos como genes homeóticos florales o genes ABC, presentan alteraciones en la identidad de ciertos órganos florales. Los análisis de estas mutantes permitieron postular el modelo ABC de la especificación de la identidad de los órganos florales (Figura 1B). Los genes de actividad A (*AP1* y *APETALA2* [5] (*AP2*)) se expresan en sépalos y pétalos; los de actividad B (*AP3* y *PI*) en pétalos y estambres y *AG*, el gen de actividad C, se expresa en estambres y carpelos. La actividad aislada o combinada de estos

[5] Este es el único gen de identidad de órgano que no pertenece a la familia MADS-box.

genes ABC en un verticilo dado determina la identidad de los órganos que ahí se desarrollan. Así en la planta silvestre, genes de actividad A inducen el desarrollo de sépalos, la actividad combinada de genes A y B el de pétalos, la de genes B y C el de estambres y la actividad del gen C el de carpelos (Coen y Meyerowitz 1991). Un estudio más reciente demostró que los genes de actividad B y C necesitan, además, la actividad de los genes *SEPALLATA1*, *2* y *3*, (*SEP1*, *2* y *3*, antes *AGAMOUS-LIKE* (*AGL*) *2*, *4*, y *9*). Estos tres genes MADS-box son redundantes funcionalmente y sólo una mutante triple *sep1sep2sep3* presenta una alteración importante del fenotipo, provocando que se originen sépalos en todos los verticilos de la flor (Pelaz et al. 2000). La redundancia de estos tres genes fue la causa de que su función como genes de identidad de órgano no fuese detectada antes y, por lo tanto, no se considerara en el modelo ABC inicial. Recientemente, estudios basados en líneas de sobre-expresión, demostraron que la acción combinada de los genes *SEP*, junto con los genes de actividad A y B, es suficiente para transformar las hojas en pétalos (Pelaz et al. 2001).

FUL (antes *AGL8*) es otro gen de esta familia que se expresa en hojas caulinas, inflorescencias, flores y en las valvas de la silicua (Mandel y Yanofsky 1995; Gu et al. 1998). El análisis del fenotipo mutante indica que este gen está involucrado en mantener un patrón normal de división, expansión y diferenciación celular durante la morfogénesis de fruto de *Arabidopsis*, así como en el desarrollo adecuado de las hojas caulinares (Gu et al. 1998). La presencia del ARNm de este gen en flores e inflorescencias, sugería que *FUL* podía tener, además, un papel importante durante el desarrollo de estas estructuras, a pesar de que esto no fuera observado en el fenotipo de la línea mutante. Posteriormente se confirmó que este gen participa en el control del tiempo de floración y de la identidad de los meristemos, presentando una redundancia funcional con los genes *AP1* y *CAL*, con los cuales presenta una alta similitud de

secuencia (Ferrándiz et al. 2000a). El gen *FUL* también actua reprimiendo la expresión de los genes MADS-box *SHATTERPROOF 1* y *2* (*SHP1*, *SHP2*, antes *AGL1* y *AGL5*) (Ferrándiz et al. 2000b). Estos últimos son funcionalmente redundantes, comparten patrones de expresión similares y los análisis genéticos de la doble mutante demuestran que *SHP1* y *SHP2* controlan el desarrollo de la zona de dehiscencia del fruto, que se encuentra entre las valvas y el replo de la silicua (Liljegren et al. 2000).

Nombre del gen	Proceso de desarrollo en el que se ha documentado su participación	Citas escogidas
APETALA1 (AP1) *CAULIFLOWER (CAL)* *FRUITFULL (FUL)*	Desarrollo floral: Identidad de meristemo	Mandel et al. 1992; Kempin et al. 1995; Mandel y Yanofsky 1995; Ferrándiz et al. 2000a.
	Desarrollo floral: Identidad de órgano:	Yanofsky et al. 1990; Coen y Meyerowitz 1991; Jack et al. 1992; Mandel et al. 1992; Reichmann y Meyerowitz 1997.
APETALA1 (AP1)	Actividad A	
APETALA3 (AP3)	Actividad B	
PISTILLATA (PI)	Actividad B	
AGAMOUS (AG)	Actividad C	
SEPALLATA1 (SEP1) *SEPALLATA2 (SEP2)* *SEPALLATA3 (SEP3)*	Interacción con los genes de actividad B y C	Pelaz et al. 2000; Pelaz et al. 2001.
FRUITFULL (FUL)	Desarrollo del fruto y hojas caulinares	Gu et al. 1998; Ferrándiz et al. 2000b.
SHATTERPROOF1 (SHP1) *SHATTERPROOF2 (SHP2)*	Desarrollo del fruto	Liljegren et al. 1998; Ferrándiz et al. 2000b; Liljegren et al. 2000.
FLOWERING LOCUS C (FLC) *SHORT VEGETATIVE PHASE (SVP)* *SOC*	Tiempo de floración	Hartmann et al. 2000; Lee et al. 2000; Sheldon et al. 2000.
ANR1	Desarrollo de raíz: Elongación de raíces laterales en respuesta a fuentes localizadas de nitratos	Zhang y Forde 1998.

Tabla 1. Principales genes MADS-box que han sido caracterizados funcionalmente en *Arabidopsis*, con base en análisis de mutantes y líneas transgénicas.

Otros miembros de la familia, como *FLOWERING LOCUS C* (*FLC*, antes *AGL25*), *SUPPRESOR OF OVEREXPRESSION OF CONSTANS* (*SOC* antes *AGL20*), y *SHORT VEGETATIVE PHASE* (*SVP*) participan en distintos niveles de las vías de control del tiempo de floración en *Arabidopsis* (Hartmann et al. 2000; Lee et al. 2000; Sheldon et al. 2000)

Estos ejemplos muestran que los genes de la familia MADS-box juegan papeles preponderantes en las redes de regulación del desarrollo en plantas. Para la gran mayoría de los miembros de la familia que han sido caracterizados hasta ahora, la función que desempeñan durante el desarrollo muestra una correlación muy cercana con sus patrones espacio-temporales de expresión de ARNm (Yanofsky et al. 1990; Coen y Meyerowitz 1991; Jack et al. 1992; Mandel et al. 1992; Reichmann y Meyerowitz 1997). En muchos casos, los patrones de expresión fueron utilizados para emitir hipótesis funcionales sobre su desempeño durante el desarrollo, las cuales se verificaron una vez que los análisis genéticos de las líneas mutantes pudieron ser realizados. (Mandel y Yanofsky 1995; Gu et al. 1998; Liljegren et al. 1998; Ferrándiz et al. 2000a; Liljegren et al. 2000).

Los primeros análisis filogenéticos realizados con las secuencias de los miembros de la familia MADS-box en diversas especies de plantas, muestran que los genes se agrupan en distintos clados monofiléticos, cuyos integrantes comparten funciones y patrones de expresión similares (Purugganan et al. 1995; Theissen et al. 1996). Además, las filogenias de la familia génica en *Arabidopsis*, muestran que dentro de estos clados existen, con gran frecuencia, genes que comparten secuencias muy parecidas (con un porcentaje de identidad de aminoácidos superior al 70% en la región MIK). Estos genes son probablemente el resultado de duplicaciones recientes (Ver Alvarez-Buylla et al. 2000a y b en apéndice) y poseen, en general, patrones de

expresión muy similares, lo que apoya la existencia de redundancias funcionales entre genes hermanos. Como lo describimos anteriormente, esto ha sido confirmado experimentalmente para grupos de genes como los *AP1/CAL*, *SHP1/5*, o *SEP1/2/3*, cuya función sólo pudo ser estudiada por análisis genético de los mutantes dobles o triples, ya que las mutantes sencillas no presentan alteración del fenotipo (eg. Liljegren et al. 2000; Pelaz et al. 2000; Smyth 2000). Esto confirma la alta incidencia de la redundancia funcional en la familia MADS-box.

Por otro lado, se han descrito miembros de esta familia que se expresan preferencialmente en estructuras vegetativas de la planta, los cuales se agrupan en determinados clados de la filogenia (Rounsley et al. 1995; Zhang y Forde 1998; Ver artículos Alvarez-Buylla et al. 2000a y b en apéndice). Estos datos sugieren que ciertos genes MADS-box podrían también tener funciones importantes durante el desarrollo vegetativo, y en particular, en el desarrollo de la raíz.

Completar el estudio funcional y filogenético de esta familia génica dará información valiosa para entender la evolución de la familia, el reclutamiento de sus miembros en la realización de funciones cada vez más específicas y su impacto en la evolución de los mecanismos de desarrollo en plantas, es decir, la participación de estos factores de transcripción en la aparición de nuevas morfologías vegetales.

3. La raíz de *Arabidopsis thaliana*

3.1. Descripción y desarrollo

La raíz de *Arabidopsis* presenta un crecimiento simple, continuo e indeterminado que la ha convertido en el sistema modelo para el estudio de la organogénesis en plantas (Benfey y Schiefelbein 1994; Scheres y Wolkenfelt 1998). El meristemo apical de la raíz da origen a un órgano que presenta una estructura regular y predecible a nivel espacial. En la zona apical de la raíz (ápice radical) se distinguen tres zonas a lo largo del eje ápico-basal[6], comenzando por el meristemo o zona proliferativa, donde las células originadas por las iniciales se dividen activamente formando filas celulares (figura 2A). En la zona de elongación celular, estas dejan de dividirse y empiezan a elongarse a lo largo del eje. Estas dos zonas conforman la región donde tiene lugar el crecimiento del órgano. Finalmente la zona de diferenciación se ha definido como la región donde las células adquieren las características definitivas de una celula madura o diferenciada, que dependerán del tipo celular al que pertenezcan (generalmente se considera que la zona de diferenciación inicia donde se observan las primeras protuberancias de los pelos radicales) (Dolan et al. 1993).

Esta sucesión zonal del ápice radical a lo largo del eje principal está perfectamente acoplada tridimensionalmente con un patrón radial de diferenciación tisular definido (figura 2B). Un corte transversal a nivel del meristemo de una raíz de *Arabidopsis* muestra que está formada de capas unicelulares de tejidos dispuestos en forma de anillos concéntricos. Estos son, del exterior al interior, la cofia lateral, la epidermis, el córtex y la endodermis, que rodean el cilindro central compuesto por el

[6] Con el fin de simplificar la redacción y exclusivamente al hablar de la raíz en su etapa de desarrollo postembrionario, invertiremos el sentido del eje ápico-basal de este órgano con respecto al eje ápico-basal de la planta: la zona más basal de la raíz es la que se une con el hipocótilo y el ápice es la zona más extrema o distal del órgano donde se encuentra el meristemo primario.

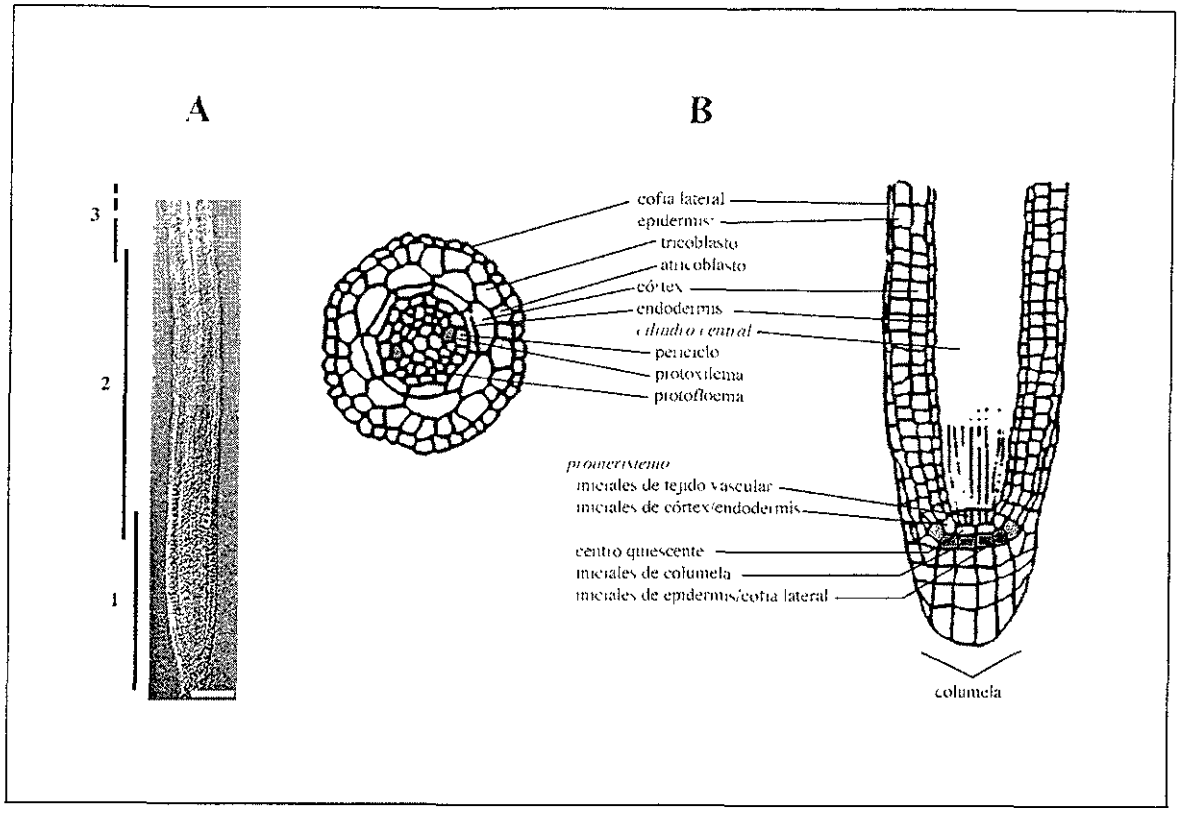

Figura 2. A: Esquema del ápice de la raíz de *Arabidopsis thaliana,* 1: meristemo apical o zona proliferativa; 2: zona de elongación celular; 3: zona de diferenciación (Escala: 100μm). B: Estructura del meristemo apical de la raíz; cortes transversal y longitudinal.

pericíclo y el tejido vascular, formado por los polos de floema y xilema en disposición diarca (Benfey y Schiefelbein 1994; Dolan et al. 1993).

A nivel de la epidermis existe un proceso adicional de diferenciación subcelular que da origen a filas con 2 tipos celulares distintos, en función de su posición relativa a las células de córtex subyacentes. Las células que se encuentran en contacto exclusivo sobre la pared periclinal de una célula de córtex adquieren identidad de atricoblastos, mientras que aquellas que están sobre la pared anticlinal (es decir, en contacto simultáneo con dos células de córtex adyacentes) adquieren la identidad de tricoblastos,

células que una vez en la zona de diferenciación forman los pelos radicales (Dolan et al. 1993).

Las filas celulares que conforman los diferentes tejidos se originan de un grupo de células iniciales, localizadas en la región subapical de la raíz, las cuales conforman el promeristemo. Las iniciales se distribuyen en tres capas "horizontales" rodeando al centro quiescente (CQ), el cual está formado por un grupo de cuatro células adyacentes cuya tasa de división es muy baja. En el primer nivel sobre el CQ se encuentran las iniciales del periciclo y tejidos vasculares, que originan el cilindro central del cuerpo de la raíz. En el segundo nivel se encuentra el CQ, rodeado de un anillo formado por las iniciales de córtex y endodermis y, finalmente, el tercer nivel por debajo del CQ se encuentran las iniciales de la cofia (columela), rodeadas por un anillo de iniciales de la epidermis y la cofia lateral. La columela constituye el tejido más apical de la raíz y junto con la cofia lateral forma una cubierta que recubre la zona proliferativa de la raíz y que se desgarra progresivamente a medida que la raíz crece (Dolan et al. 1993; Steeves y Sussex 1989).

La raíz primaria de *Arabidopsis* se origina durante la embriogénesis. El cigoto se divide en una célula basal y otra apical. La primera da lugar al suspensor y a la célula hipófisis, que origina al CQ y la región de la columela, mientras que la célula apical da origen al resto del embrión y, por lo tanto, origina la parte del meristemo de la radícula que formará el cuerpo de la raíz (Benfey y Schiefelbein 1994). La arquitectura típica del meristemo radicular se forma a partir del estadio de corazón (Scheres et al. 1994; Schiefelbein et al. 1997).

Durante la fase de desarrollo postembrionario, una raíz primaria puede producir a raíces laterales que se originan a partir de células de periciclo adyacentes a los polos de xilema de la raíz madura. Estas células proliferan formando un primordio que se

abre paso a través de los tejidos externos de la raíz, hasta que emerge de la raíz primaria. De esta manera se genera un meristemo de raíz lateral, cuya activación inicia el crecimiento de la raíz lateral, la cual posee la misma organización tisular que la raíz primaria (Malamy y Benfey 1997).

3.2. Bases moleculares del desarrollo de la raíz

Los esfuerzos emprendidos hasta ahora para desentrañar los mecanismos básicos que rigen la organogénesis de la raíz de *Arabidopsis* han seguido distintos enfoques experimentales, desde el análisis de mutantes acoplado al uso de herramientas moleculares y de micromanipulación, hasta el estudio y descripción de cinéticas celulares, entre otros (eg. Scheres et al. 1994; Hauser et al. 1995; van den Berg et al. 1995; Rost et al. 1996, Beemster y Baskin 1998; Berger et al. 1998a). Sin embargo, los datos obtenidos hasta ahora siguen siendo escasos. A continuación revisaremos brevemente los resultados y las reflexiones más significativas que se han generado sobre este tema.

La raíz primaria se origina durante la formación del embrión, cuya estructura tridimensional ha sido disectada en dos componentes espaciales. El primero corre a lo largo de un eje ápico-basal, en el cual se desarrollan las estructuras u órganos que componen al embrión (cotiledones, meristemo apical, hipocótilo, raíz y meristemo radicular). El segundo se define en un plano transversal, en el cual se distribuyen los tres tipos tisulares principales del embrión en dirección radial (el protodermis (epidermis) y el tejido cortical [*ground tissue*] formando anillos sucesivos alrededor del procambium (tejido vascular)) (Scheres et al. 1994).

A lo largo del eje ápico-basal se han caracterizado mutantes deficientes en la formación del dominio basal del embrión y que presentan anomalías en el desarrollo del hipocótilo y la raíz, como *monopteros* (*mp*) y *bodenlos* (*bdl*), o del meristemo radicular, como *hobbit* (*hbt*). La mutante *bdl* presenta alteraciones en la respuesta a auxinas y *mp* tiene deficiencias en la axialización de sus células vasculares (Berleth y Jürgens 1993; Hamann et al. 1999). El gen *MP* codifica para un factor de transcripción con un motivo ARF de respuesta a auxinas que regula la transcripción de genes inducibles por esta hormona (Ulmasov et al. 1997; Hardtke y Berleth 1998). La mutante *hbt* afecta las divisiones de la célula hipófisis, que origina al CQ y la columela, y posteriormente afecta la formación del resto del meristemo radicular, lo cual revela la importancia de las interacciones celulares durante la embriogénesis (Willemsen et al. 1998). Por otro lado, aún se desconocen los procesos que rigen la iniciación y el establecimiento del patrón radial durante la embriogénesis, y de hecho, ningún gen involucrado en la especificación de los tipos tisulares principales ha podido ser identificado por análisis de fenotipo (Scheres y Heidstra 1999).

Durante el desarrollo postembrionario de la raíz, las células que se producen en el meristemo se incorporan al órgano, manteniendo el patrón radial establecido durante la embriogénesis. Una de las preguntas por resolver en el desarrollo de la raíz es la de definir cuáles son los mecanismos encargados de perpetuar este patrón radial. La expresión de genes marcadores de ciertos tipos celulares en el meristemo apical indican la existencia de un prepatrón de especificación de los tipos celulares desde etapas muy tempranas. De hecho, ciertos genes cuyas mutantes afectan el patrón radial, en particular la subespecificación de ciertos tipos celulares (ver más adelante), se expresan desde la embriogénesis (Di Laurenzio et al. 1996; Berger et al. 1998a).

Análisis clonales y experimentos de ablación celular han demostrado que el destino de una célula se rige por mecanismos de señalización posicional y no por su pertenencia a un linaje celular dado (Scheres et al. 1994; van den Berg et al. 1995). De manera que el destino final de una célula del meristemo puede modificarse si ésta es desplazada hacia una nueva posición, lo que puede suceder en ciertos procesos del desarrollo, como en las divisiones longitudinales ocasionales de los tricoblastos, o bien puede ser inducido por vías experimentales (van den Berg et al. 1995; 1997; Berger et al. 1998b). Aunque no se conoce aún la identidad molecular de este tipo de señales, las evidencias revelan que éstas se transmiten de las células maduras a las jóvenes, es decir, en dirección del ápice de la raíz. Scheres y Heidstra (1999) proponen dos hipótesis sobre los posibles procesos que perpetuan el patrón radial de la raíz, el cual se origina en el embrión y prosigue durante el desarrollo postembrionario. La primera postula que mecanismos de señalización posicional dirigen a los genes de "control de patrón" [*patterning genes*] que actúan constantemente, desde la embriogénesis y a lo largo de la fase de desarrollo postembrionario. Su hipótesis alternativa es que durante la embriogénesis se genera un prepatrón celular que, subsecuentemente, induce un mecanismo adicional de refuerzo del patrón de especificación "posicional" de las células, el cual funciona únicamente durante el desarrollo postembrionario.

Actualmente se cuenta con más información sobre los mecanismos moleculares involucrados en la subespecificación de ciertos tipos celulares de los principales tejidos de la raíz. Tal es el caso, por ejemplo, de la especificación subcelular de la epidermis en tricoblastos y atricoblastos. Los análisis de mutantes han permitido caracterizar varios de los componentes moleculares involucrados en la especificación de estos tipos celulares, así como en la morfogénesis del pelo radicular. Se han identificado moléculas que intervienen en procesos de señalización celular, y factores de transcripción que

reprimen (*TRANSPARENT TESTA GLABRA* (*TTG*), *GLABRA2* (*GL2*), *WEREWOLF* (*WER*)) o promueven (*CAPRICE* (*CPC*)) la especificación de células con pelos radicales, así como componentes de las vías de acción de hormonas vegetales como el etileno y las auxinas (eg. Galway et al. 1994; Masucci y Schiefelbein 1996; Masucci et al. 1996; Wada et al. 1997; Lee y Schiefelbein 1999; Walker et al. 1999). Estos estudios han permitido la elaboración de modelos de interacción (esquemáticos o dinámicos) entre los distintos componentes del proceso (Mendoza y Alvarez-Buylla 2000; para una revisión reciente ver Schiefelbein 2000). A nivel del tejido cortical de la raíz, existe también un proceso de especificación subcelular que da origen al córtex propiamente dicho y a la endodermis. Estas capas surgen de una división periclinal asimétrica de las células hijas de las iniciales de córtex y endodermis. Esta división origina una célula de endodermis al interior y otra de córtex al exterior. En los mutantes *short root* (*shr*) y *scarecrow* (*scr*), esta división asimétrica es inexistente en el embrión y durante el desarrollo postembrionario, por lo que sólo se forma una capa de tejido cortical, la cual presenta una identidad de córtex en *shr* y una identidad mixta en el caso de *scr*. Estos datos integrados con evidencias adicionales como el patrón de expresión y el fenotipo en tejido aéreo de mutantes de *scr*, revelan el requerimiento de estos genes para la division asimétrica y para la especificación de la endodermis (Benfey et al. 1993; Di Laurenzio et al. 1996).

El meristemo de la raíz es la zona donde se generan las células y donde se originan muchos de los procesos de su desarrollo. Esto ha fomentado el estudio del papel de la división celular en relación con aspectos específicos del desarrollo, como el funcionamiento del meristemo, la formación de patrones, el crecimiento y la morfogénesis.

Se ha considerado que el funcionamiento de un meristemo depende de su activación y su mantenimiento (Scheres y Heidstra 1999). La activación se refiere a los mecanismos que promueven la división de las células, es decir que mantienen al meristemo "activo". En el mutante *rootmeristemless1* (*rml1*) el meristemo primario de la raíz no presenta divisiones celulares después de la germinación. Lo mismo pasa en las raíces laterales de la mutante *aberrant lateral root formation3* (*alf3*), es decir que el meristemo se forma pero no se activa y las células no proliferan. De esta manera *RML1* y *ALF3* son necesarios para la activación del meristemo y la continuación del crecimiento de la raíz (Cheng et al. 1995; Celenza et al. 1995). Asimismo, variaciones en las condiciones ambientales pueden influenciar el control de la división celular. Los factores de estrés ambiental inciden en la formación de especies reactivas de oxígeno. La célula se adapta modificando su balance oxidorreductivo, o redox, haciendo uso de moleculas antioxidantes como el ácido ascórbico y la glutationa (GSH). Estas moléculas parecen estar involucradas en ciertos procesos de control de la división celular (Sánchez-Fernandez et al. 1997). Recientemente se clonó el gen *RML1*, el cual codifica para la γ-glutamylcysteinesintetasa, la primera enzima de la biosíntesis de la GSH. La ausencia de división celular en la raíz de *rml1* resulta de una deficiencia en GSH en las células del meristemo (Vernoux et al. 2000). Las modificaciones del balance redox en células proliferativas parece tener un papel primordial en los procesos de modulación del crecimiento en respuesta al medio ambiente (Sánchez-Fernández et al. 1997).

El mantenimiento de los meristemos se refiere a la conservación de un conjunto o *pool* de células madre [*stem cells*] que aseguren la existencia continua del meristemo. En el meristemo apical del vástago [*shoot apical meristem*], los genes *CLAVATA* (*CLV*), *WIGGUM* (*WIG*) y *SHOOTMERISTEMLESS* (*STM*) están involucrados en mantener un equilibrio entre la población de células madre en el centro del meristemo y aquellas que

participan en la formación de nuevos órganos en la periferia (Clark et al. 1996; Running et al. 1998). En el meristemo de la raíz, las células iniciales funcionan como células madre y rodean al CQ formado por células con baja tasa de divisón, en analogía con las células de la zona central del meristemo aéreo. La ablación de células del CQ provoca una progresión en el grado de diferenciación de las células iniciales adyacentes, ya que las iniciales de la columela acumulan amiloplastos y las iniciales del córtex/endodermis se dividen asimétricamente. El CQ inhibe así la diferenciación de las iniciales por contacto directo, manteniéndolas como células madre (van den Berg et al. 1997). Por otro lado, el fenotipo del mutante *hbt* parece indicar que el mecanismo que mantiene a las iniciales indiferenciadas se origina en la embriogénesis (Willemsen et al. 1998). El mantenimiento de los meristemos de la planta parece necesitar un balance de señales que regulen el grado de diferenciación de las células madre y sus hijas (Scheres y Heidstra 1999).

La formación de patrones, a su vez, parece ser independiente del número de células y de los planos de división. Mutantes como *fass* (*fs*) y *tonneau* (*ton*) presentan alteraciones en la secuencia y la orientación de las divisiones en el embrión, pero el patrón radial permanece intacto (Torres-Ruíz et al. 1994; Traas et al. 1995). Otra evidencia es que el patrón radial de distribución tisular está globalmente conservado en raíces de diversas especies, cuyos meristemos presentan organizaciones y secuencias de divisiones muy diversas (Steeves y Sussex 1989). Por su parte, el proceso de subespecificación de ciertos tipos celulares parece, en algunos casos, regular la tasa de división celular y la orientación de las divisiones (Di Laurenzio et al. 1996; Berger et al. 1998b).

Muchos investigadores han estudiado la relación de la división celular con el crecimiento y la morfogénesis[7]. El crecimiento y la generación de formas en una planta están asociados a la división y la expansión de las células. En muchos casos, sin embargo, se ha logrado demostrar que tanto el crecimiento de un órgano como su morfología parecen ser independientes de un control exacto de la división celular. Además, las alteraciones en la orientación de los planos de división y en la expansión celular no parecen afectar la forma final de un órgano. Estos datos han sido puestos en evidencia por el análisis de líneas de animales (*Drosophila melanogaster*) y plantas en las que la alteración de este tipo de parámetros celulares no afectan mayormente el crecimiento y la morfología del órgano considerado, ya que una disminución del número de células está generalmente acompañada de un aumento en el tamaño de las mismas (Hemerly et al. 1995, Smith et al. 1996; Weigmann et al. 1997).

En el caso de la raíz, los datos que se tienen en relación al efecto de la división y la elongación celulares sobre el crecimiento y la morfología deben ser interpretados con cautela, ya que se trata de un órgano de crecimiento indeterminado, cuya elongación y forma (por ejemplo, la densidad de pelos radiculares) dependen estrechamente de las condiciones ambientales y del efecto de ciertas hormonas vegetales. La tasa de crecimiento de una raíz depende de la cantidad de células producidas y la elongación de las mismas. La tasa de producción celular depende del número de células que se dividen y de su tasa de división (Beemster y Baskin 1998). Cuando una plántula de *Arabidopsis* crece en condiciones constantes sobre un medio de cultivo estándar, la tasa de elongación de la raíz aumenta en función del tiempo. Estudios cinéticos demuestran que

[7] En este campo se ha generado una controversia entre los adeptos de las teorías organísmica y celular. La primera considera al organismo como una unidad, y a las divisiones celulares exclusivamente como un proceso de subdivisión del volumen total del mismo. La teoría celular por su parte, sostiene que las células son la unidad básica del organismo, y que las divisiones celulares son el factor primordial de su desarrollo (Kaplan y Hagemann 1991)

la aceleración del crecimiento de la raíz de *Arabidopsis* está directamente relacionada con un aumento en el número de células que se dividen (es decir, con un mayor tamaño del meristemo) sin que su tasa de división varíe. Esto genera, a su vez, un aumento en el número de células en elongación (y un aumento del tamaño de la zona de elongación), pero no altera la longitud final de las mismas (Beemster y Baskin 1998).

Como lo señalamos anteriormente, en condiciones naturales las modificaciones del balance redox de las células parecen mediar procesos de modulación del crecimiento en respuesta al medio ambiente. La aplicación de GSH al medio de cultivo de raíces de *Arabidopsis* provoca un aumento en la producción de células y una disminución en la elongación de las mismas, lo que en conjunto provoca solo un ligero aumento en el crecimiento de la raíz, comparado con el control. Por su parte, la aplicación de un inhibidor específico de la síntesis de GSH disminuye la cantidad de células producidas, pero aumenta la elongación de las mismas lo que mantiene un nivel de crecimiento de la raíz similar a la del control (Sánchez-Fernández et al. 1997). Estos datos parecen coincidir globalmente con las evidencias sobre la aparente independencia entre el crecimiento de un órgano y el control exacto de la division celular, siendo que en este caso, una mayor producción celular es compensada con una disminución en la elongación de las células y viceversa. Con base en estos datos, algunos autores han propuesto que los tamaños de las zonas de división y elongación celular están acoplados, y que en realidad constituyen una región única en términos de desarrollo (Scheres y Heidstra 1999).

La arquitectura del sistema radical depende, en gran medida, de la producción de raíces laterales. La formación de los primordios es inducida por condiciones ambientales, así como por la aplicación de auxinas, pero aún se desconocen los procesos moleculares que dirigen su formación y su distribución espacial a lo largo de la raíz

4. Genes MADS-box de raíz: Antecedentes

Los esfuerzos iniciales dirigidos a identificar a los distintos miembros de la familia MADS-box en *Arabidopsis*, llevaron al descubrimiento de genes como *AGL12*, *AGL14* y *AGL17*, que se expresan preferencialmente en la raíz, de acuerdo con los primeros ensayos tipo Northern (Rounsley et al. 1995). Por otro lado se ha detectado la inducción de los genes MADS-box *NMH7* y *NMHC5* en las células infectadas con *Rhizobium* durante el desarrollo de nódulos en raíces de alfalfa (Heard y Dunn 1995; Heard et al. 1997). Otro estudio dirigido a identificar genes inducibles por nitratos, reveló que un gen MADS-box, denominado *ANR1*, está involucrado en estimular la elongación de las raíces laterales de *Arabidopsis* en respuesta a concentraciones locales elevadas en nitratos. Las raíces de líneas antisentido y de cosupresión para este gen no presentan el aumento en la tasa de elongación que presentan las raíces laterales de las plantas silvestres en medios adicionados localmente con nitratos (Zhang y Forde 1998). Estos datos indican que genes MADS-box juegan papeles importantes durante el desarrollo de la raíz de *Arabidopsis*.

Al inicio de este trabajo se tenían datos preliminares acerca de otros genes de la familia MADS-box que parecían expresarse preferencialmente en raíces de *Arabidopsis*. Estos genes son *AGL16*, *AGL19* y *AGL21*. Los dos últimos son parte central de este estudio.

5. Objetivos

Este trabajo se sitúa dentro de la primera etapa de un análisis por genética inversa, dirigido a estudiar funcionalmente a los genes de la familia MADS-box expresados en la raíz de *Arabidopsis thaliana*.

El objetivo general es caracterizar los patrones de expresión de genes MADS-box que se expresan en raíz, y de integrarlos con los resultados de los análisis filogenéticos de los miembros de esta familia, para detectar posibles grupos de genes redundantes en la raíz, y emitir las primeras hipótesis sobre la función de estos genes, que puedan servir de guía para la segunda etapa del análisis por genética inversa.

Los objetivos particulares son:

- Determinar la secuencia del ARNm de los genes *AGL21* y *AGL19* y clonarlos.
- Incorporar sus secuencias a los análisis filogenéticos de la familia MADS-box de *Arabidopsis* y determinar su posición filogenética.
- Determinar la posición genómica de *AGL21*.
- Corroborar el patrón de expresión de ambos genes por ensayos tipo Northern.
- Estudiar los patrones de expresión de los genes *AGL12*, *AGL16*, *AGL17*, *AGL19*, y *AGL21* por hibridación *in situ* del ARNm en cortes histológicos de raíz y de embrión de *Arabidopsis*.
- A partir de la integración de los análisis filogenéticos y de los patrones de expresión, determinar los grupos de genes potencialmente redundantes y emitir las primeras hipótesis funcionales.

6. Metodología experimental

Este apartado pretende sólo hacer una breve síntesis de los principios y las técnicas moleculares más utilizadas durante el desarrollo de este trabajo, que permitieron obtener la mayor parte de los resultados experimentales reportados en las publicaciones. Mayores detalles de los materiales y los métodos utilizados pueden ser consultados directamente en las publicaciones y referencias citadas y/o manuales de los fabricantes de los reactivos utilizados así como los manuales de referencia para técnicas moleculares (eg. Sambrook et al. 1989; ver artículos en apéndices I, II y III).

Para la determinación de la secuencia y la clonación del ADN complementario (ADNc) de los genes *AGL19* y *AGL21*, se siguieron protocolos experimentales moleculares clásicos, tales como la extracción de ARN de raíz de *Arabidopsis* y síntesis del ADNc correspondiente; el diseño de oligonucleótidos o primers para realizar diversos experimentos basados en la reacción en cadena de la polimerasa (PCR), tales como RTPCRs, siguiendo protocolos de RACE3' y RACE5' (Boehringer); electroforesis en geles de agarosa de las amplificaciones obtenidas; clonación de los fragmentos obtenidos por ligación en plásmidos tipo PGEMT, transformación de *Escherichia coli* y purificación de los plásmidos recombinantes. La secuenciación de los fragmentos clonados se realizó en un laboratorio externo; para la edición, el análisis y la alineación de las secuencias obtenidas se usaron programas como GeneWorks, Macvector, EditView, ClustalW.

El estudio de los patrones de expresión de *AGL19* y *AGL21* por ensayos de tipo Northern, se hizo siguiendo el protocolo basado en la electroforesis desnaturalizante de ARN proveniente de distintas estructuras vegetales, su transferencia a una membrana de

nylon; hibridación de la membrana con la sonda marcada con ^{32}P, lavados, exposición sobre una placa de autoradiografía y revelado de la misma.

El protocolo de localización *in situ* de ARNm sobre cortes histológicos de raíz se divide en tres etapas principales. La finalidad de la primera es la obtención de cortes histológicos y consiste en crecer las plántulas en condiciones axénicas, fijar, deshidratar e incluir su raíz en parafina. Posteriormente éstas son orientantadas y seccionadas al microtomo y montadas en portaobjetos. Esta etapa es la más crítica y laboriosa, debido a que las características intrínsecas de la raíz de *Arabidopsis*, tales como su fragilidad y tamaño, dificultan su manipulación, orientación y corte al microtomo. El diámetro de una raíz es de 100-150 μm, lo que aunado a su transparencia, imposibilita poder localizarlas visualmente en los bloques de parafina y dificulta enormemente su orientación con respecto a la cuchilla del microtomo. Por otro lado, una célula de córtex o epidermis puede elongarse hasta 15-20 veces desde que sale del meristemo hasta que se diferencia completamente, lo que debilita muchísimo la estructura de la raíz para la obtención de cortes longitudinales o transversales a partir de la zona de elongación celular, cuando la parafina es usada como matríz de inclusión.

La segunda etapa consiste en la síntesis de la sonda de ARN antisentido marcada con digoxigenina (en todos los casos una sonda de ARN sentido es utilizada como control negativo), partiendo de la linearización del plásmido que contiene el fragmento del gen utilizado como templado para la síntesis. En el caso particular de genes MADS-box, las sondas utilizadas se limitan al extremo 3' del ARNm del gen, para evitar que la secuencia de la caja MADS, la cual está muy conservada, pueda hibridar inespecíficamente con mensajeros de otros miembros de la familia.

La tercera etapa es la hibridación *in situ* propiamente dicha y consiste en una prehibridación que permeabiliza el tejido de los cortes, la hibridación con la sonda

marcada, los lavados para eliminar la sonda que no hibridó específicamente y el protocolo de detección y revelado de la sonda marcada en los cortes. Al término del experimento, se procede al análisis microscópico y la toma de microfotografías de las muestras.

Entre otras de las técnicas utilizadas se encuentran, Southern blot, hibridación *in situ* de plántulas enteras [*whole mount in situ*], mapeo genómico por PCR con la biblioteca CIC de YACs indexados (Creusot et al. 1995).

7. Genes MADS-box de raíz y análisis filogenéticos

7.1. Una duplicación ancestral tuvo lugar en la familia de genes MADS-box antes de la divergencia de plantas y animales

Resumen del artículo de Alvarez-Buylla et al. (2000a), que puede ser consultado en el apéndice I.

Las modificaciones en los genes que codifican para reguladores transcripcionales pueden alterar el desarrollo y son componentes importantes de los mecanismos moleculares de la evolución morfológica. Los genes de la familia MADS-box codifican para reguladores transcripcionales que tienen funciones biológicas importantes y diversas. Los resultados obtenidos en plantas, indican que genes de esta familia regulan el desarrollo de estructuras tales como la flor, el fruto, la hoja y la raíz. Los avances en la secuenciación del genoma de *Arabidopsis*, al momento de realizar este trabajo, permitieron obtener un muestreo significativo[8] de los miembros de la familia MADS-box en una sola especie. Para probar el paralelo que se sospecha existe entre la evolución de la familia génica MADS-box y la evolución de la morfología en plantas, era necesaria una filogenia polarizada de genes. En este trabajo sugerimos que una duplicación génica ancestral a la divergencia de plantas y animales dió origen a 2 linajes principales de genes MADS-box: el Tipo I y el Tipo II. Localizamos la raíz de la familia de genes MADS-box de eucariontes entre estos dos linajes. Un nuevo grupo filogenético de dominios MADS de plantas (AGL34-like) parece estar cercanamente relacionado a los dominios MADS SRF-like de animales previamente identificados, para formar el linaje de Tipo I (Ver fig. 3 en Alvarez-Buylla et al. 2000a, apéndice I).

[8] Cuando se terminó de secuenciar el genoma de *Arabidopsis*, se pudo establecer que este muestreo incorporó alrededor del 50% de los miembros de la familia presentes en esta planta (Martínez-Castilla y Alvarez-Buylla, datos no publicados)

Sólo los miembros de plantas del linaje Tipo II poseen un dominio K que se encuentra corriente abajo del dominio MADS en los miembros de plantas previamente identificados. Esto sugiere que el dominio K evolucionó después de la duplicación que originó a los dos linajes. Finalmente, un grupo de secuencias intermediarias de plantas puede ser el resultado de eventos de recombinación. Estos análisis podrían guiar la búsqueda de secuencias MADS-box en eucariontes basales, así como la posición filogenética de nuevos genes provenientes de otras especies de plantas.

7.2. La evolución de la familia de genes MADS-box fuera de las flores: expresión de algunos genes en polen, endospermo, raíz y tricomas

Resumen del artículo de Alvarez-Buylla et al. (2000b), que puede ser consultado en el apéndice II.

Los genes MADS-box codifican para reguladores transcripcionales, involucrados en diversos aspectos del desarrollo en plantas. En este trabajo describimos la clonación y los patrones espacio-temporales de expresión del ARNm de cinco genes MADS-box de *Arabidopsis*: *AGL16*, *AGL18*, *AGL19*, *AGL27* y *AGL31*. Estos genes probablemente podrán usarse como herramientas moleculares importantes para el análisis evolutivo y funcional de estructuras vegetativas. Hemos mapeado nuestros datos y patrones de expresión previamente obtenidos sobre una nueva filogenia de los genes MADS-box (Ver fig.6 en Alvarez-Buylla et al. 2000b, apéndice II). Estos análisis sugieren que la evolución de la familia MADS-box ha involucrado una diversificación funcional rápida y simultánea, tanto en estructuras vegetativas como reproductivas. Los genes ancestrales hipotéticos tenían patrones de expresión más generalizados que los genes más derivados que han sido cooptados para funciones putativas más especializadas, según lo sugieren sus patrones de expresión. *AGL27* y *AGL31*, que están

cercanamente relacionados al gen de tiempo de floración *FLC* (previamente *AGL25*) recientemente descrito, se expresan en la mayoría de los tejidos de la planta. *AGL19* se expresa específicamente en las capas externas del meristemo de raíz (cofia lateral y epidermis) y en el cilindro central de la raíz madura. *AGL18*, que es muy similar en secuencia al gen *AGL15* expresado en embrión, se expresa en el endospermo y durante el desarrollo de los gametofitos femenino y masculino, sugiriendo que el papel de *AGL18* sería distinto al de los genes MADS-box previamente caracterizados. Finalmente, el ARNm de *AGL16* se acumula en las células guarda de los estomas y en tricomas. Nuestra nueva filogenia revela siete nuevos clados monofiléticos de secuencias MADS-box que no son específicas de flores; esto sugiere que redes de regulación complejas involucrando a varios genes MADS-box, similares a aquellas que controlan el desarrollo de la flor, dirigen el desarrollo de estructuras vegetativas.

7.3. Posición de los genes de raíz en la filogenia de los genes MADS-box de *Arabidopsis*

Los genes MADS-box *AGL12*, *AGL14*, *AGL17*, *ANR1*, *AGL16*, *AGL19* y *AGL21* (figura 3) se expresan en la raíz de *Arabidopsis* (Rounsley et al. 1995; Zhang y Forde 1998; Alvarez-Buylla et al. 2000b; Burgeff et al. (en prensa)). Los análisis filogenéticos realizados con las secuencias de aminoácidos de miembros de la familia revelan que todos los genes arriba mencionados pertenecen al linaje de tipo II. AGL12 se sitúa en una rama independiente de otros clados y los programas de predicción de estructuras protéicas no reconocen en este gen una estructura tipo *coiled-coil*, característica del dominio K. Aunque inicialmente se le describió como un gen específico de raíz, experimentos adicionales de *in situ* y líneas transformadas con su promotor fusionado a

GUS, indican que también se expresa en flor y embrión (Acevedo y Alvarez-Buylla, datos no publicados; Burgeff et al. (en prensa)).

Los demás miembros arriba mencionados poseen la secuencia de dominios MIKC, que presentan la mayoría de estos factores de transcripción de plantas pertenecientes al linaje de tipo II. El árbol consenso obtenido por análisis de máxima parsimonia a partir de las alineaciones de las secuencias de aminoácidos de los dominios MIK (Ver fig. 6a en Alvarez-Buylla et al. 2000b), muestra que estos miembros se distribuyen en 2 clados distintos. AGL19, AGL14, SOC1 (antes AGL20) y AGL42 conforman un clado en el que los dos primeros y los dos últimos comparten, cada par, una alta similitud de secuencia. Los Northern blots muestran que *AGL19* y *AGL14* se expresan exclusivamente en raíz, mientras que *SOC1* se expresa además en hojas y flores (Samach et al. 2000); *AGL42* aún no ha sido estudiado.

AGL17, AGL21, ANR1 y AGL16, a su vez conforman un clado en el que AGL17 y AGL21 comparten una alta similitud de secuencia. Los ensayos tipo Northern indican que este clado agrupa genes que se expresan casi preferencialmente en raíz, aunque *AGL16* se expresa también en hojas y otros tejidos vegetativos y experimentos de *in situ* muestran que *AGL21* y *ANR1* también se expresan en embrión.

El mapeo de los patrones de expresión en la filogenia parece indicar que los genes ancestrales de la familia tenían un patrón de expresión generalizado y que éstos fueron divirgiendo, dando lugar a patrones de expresión más específicos y delimitados en estructuras tanto reproductivas como vegetativas. Los valores de bootstrap obtenidos en esta reconstrucción no permiten definir el orden estricto de divergencia de estos clados con respecto a los clados florales A, B, C. Probablemente la evolución de la familia fue muy rápida y los clados de raíz evolucionaron al mismo tiempo que los clados florales. De ser cierto, esto sugeriría que en la evolución morfológica de las

plantas, las estructuras reproductivas y vegetativas especializadas pudieron haber evolucionado simultáneamente.

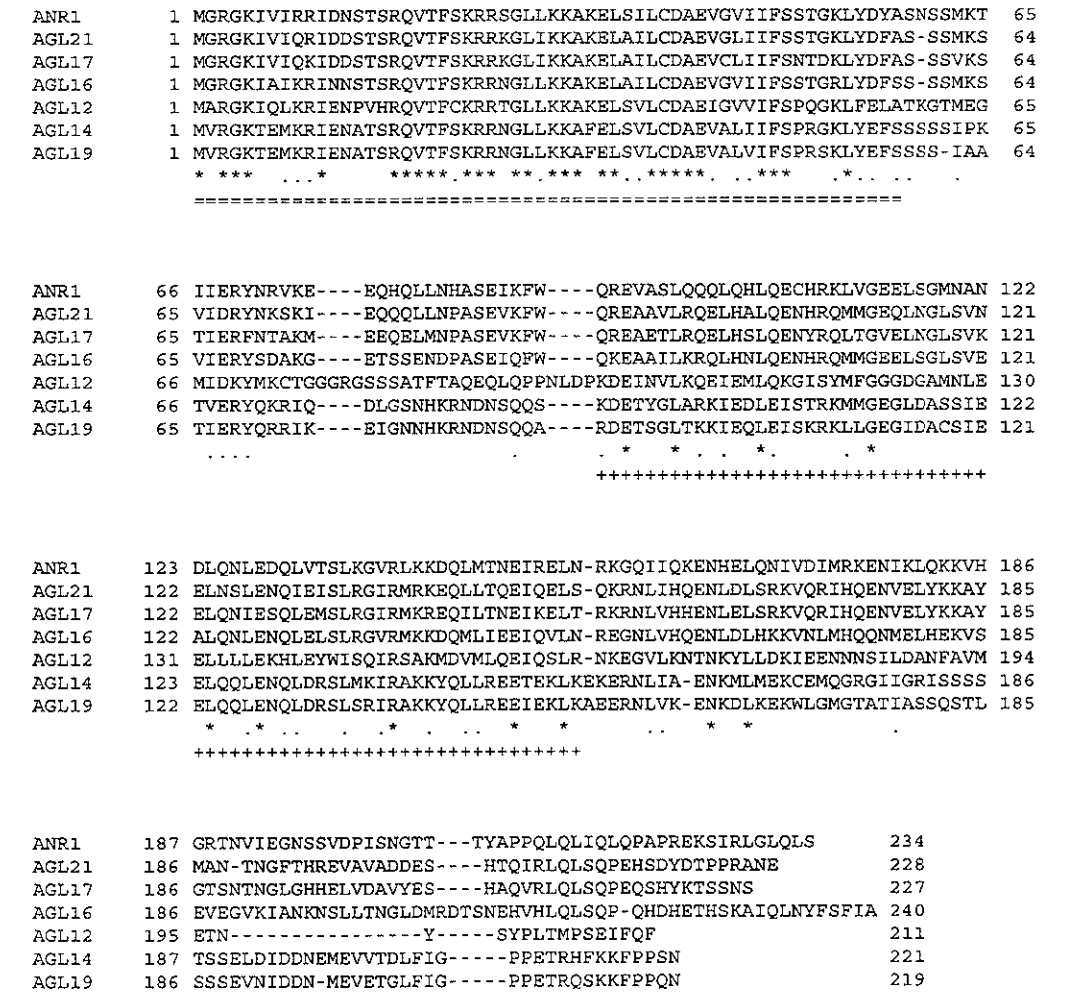

Figura 3: Alineación de las secuencias de aminoácidos de los genes MADS-box de raíz (===: dominio MADS; ++++: dominio K; *: aminoácidos idénticos; . : aminoácidos altamente conservados; números de acceso al Genbank para las secuencias del ADNc: *ANR1* (Z97057); *AGL21* (AF336979); *AGL17* (U20186), *AGL16* (AF312662), *AGL12* (U20193), *AGL14* (U20184), *AGL19* (AF312664) (ClustalW alignment).

8. Patrones de expresión del ARNm de genes MADS-box de raíz

8.1. Expresión de genes MADS-box en los primordios laterales, meristemos y tejidos diferenciados de raíces de *Arabidopsis thaliana*

Resumen del manuscrito de Burgeff et al. (en prensa), que puede ser consultado en el apéndice III.

A pesar de que los genes MADS-box involucrados en el desarrollo de la flor y del fruto han sido caracterizados, la función de este tipo de genes expresados en estructuras vegetativas aún debe ser explorado. Al menos siete miembros de esta familia se agrupan en clados de genes que se expresan preferencialmente en la raíz de *Arabidopsis thaliana* (L.) Heynh.. En este trabajo reportamos la clonación del gen MADS-box *AGL21*, que pertenece al clado *ANR1*, así como el patrón de expresión por hibridación *in situ* de ARNm de éste y otros dos genes MADS-box de raíz. *AGL17* parece ser un marcador de cofia lateral en el ápice de la raíz y hacia la zona de elongación este gen se expresa en las células de la epidermis. *AGL21* se expresa intensamente en los primordios de raíz lateral y tiene un patrón de expresión punteado en el meristemo apical similar al de *AGL12*. Ambos genes se expresan tambien en el cilindro central de la zona diferenciada de la raíz, así como durante la embriogénesis. Este estudio, combinado con análisis filogenéticos previos, indica que estos genes MADS-box podrían jugar distintos papeles regulatorios durante el desarrollo de la raíz.

8.2. Patrones de expresión de *AGL12, AGL17, AGL21, AGL19, AGL16 y ANR1*

AGL12

En el ápice de la raíz este gen se expresa de manera contínua en las células externas de la columela y en la cofia lateral. También se puede apreciar un patrón de tinción punteado, en que sólo ciertas células del corte se tiñen de manera aparentemente aleatoria. Sin embargo, en los cortes transversales de la zona cercana a la base del meristemo y a nivel de la epidermis, el patrón punteado muestra una tendencia de tinción preferencial en los atricoblastos. *AGL12* se expresa en el cilindro central en la zona diferenciada, aunque la intensidad de la tinción puede ser variable a lo largo de la raíz. Durante la embriogénesis, *AGL12* se expresa también siguiendo un patrón punteado, desde el estadio globular hasta el torpedo por lo menos. La frecuencia de células teñidas aumenta conforme se desarrolla el embrión (figura 4) (ver también fig. 4 en Burgeff et al. (en prensa), apéndice III).

AGL17

Los transcritos de este gen se encuentran localizados en la capa de la cofia lateral que recubre el meristemo apical. El gen se expresa en las células de la epidermis cerca del límite basal del meristemo y en el inicio de la zona de elongación. En la zona diferenciada, *AGL17* se expresa en células del periciclo y con menor intensidad en otras células del cilindro central (figura 4) (ver también fig. 3 en Burgeff et al. (en prensa), apéndice III).

AGL21

Este gen se expresa durante la formación de las raíces laterales, desde los estadíos tempranos, en que los primordios tienen de 3 a 4 capas celulares, hasta que la raíz lateral emerge de la raíz primaria y la marca se situa principalmente en el ápice de la raíz lateral. A medida que la raíz lateral crece, el gen se expresa en el meristemo, en donde adopta un patrón punteado, similar al que se observa en los meristemos primarios de la raíz donde sólo algunas células del corte están teñidas. En la zona diferenciada de la raíz, *AGL21* se expresa en células del cilindro central y, aunque la intensidad de la tinción puede variar a lo largo de la raíz, en algunos casos se observa una tinción intensa en células del periciclo (figura 4) (ver también fig. 2 en Burgeff et al. (en prensa), apéndice III).

Durante la embriogénesis se detectó la presencia del transcrito de *AGL21* en embriones desde el estadío globular hasta el torpedo. En muchas ocasiones la tinción sigue un gradiente de concentración que aumenta progresivamente desde la base hasta el ápice del embrión (ver fig. 2 en Burgeff et al. (en prensa), apéndice III).

AGL19

Este gen se expresa en las capas más externas del meristemo de la raíz primaria y lateral, que son la cofia (células externas de la columela y cofia lateral) y la epidermis. En la zona diferenciada el transcrito está presente en las células del cilindro central (figura 4) (ver también fig.5 en Alvarez-Buylla et al., 2000b, en el apéndice II).

AGL16

Los *in situ* realizados con distintas sondas antisentido para este gen, muestran tinción en la cofia (capas externas de la columela y cofia lateral) y en la epidermis, con un patrón

muy parecido al de *AGL17* (figuras 4 y 5). Sin embargo, los controles negativos realizados con distintas sondas sentido también se tiñen en algunas ocasiones, por lo que estos resultados no han sido publicados y deben ser corroborados con líneas transformantes con el promotor de *AGL16* fusionado a un gen marcador.

ANR1

Los primeros análisis muestran un patrón de expresión punteado, muy similar al de *AGL21* en meristemos primarios (figuras 4 y 5). La tinción se observa también en primordios de raíz lateral, sin embargo ésta también parece ser punteada, mientras que el gen *AGL21* se expresa de manera continua en estas estructuras. *ANR1* se expresa también durante la embriogénesis.

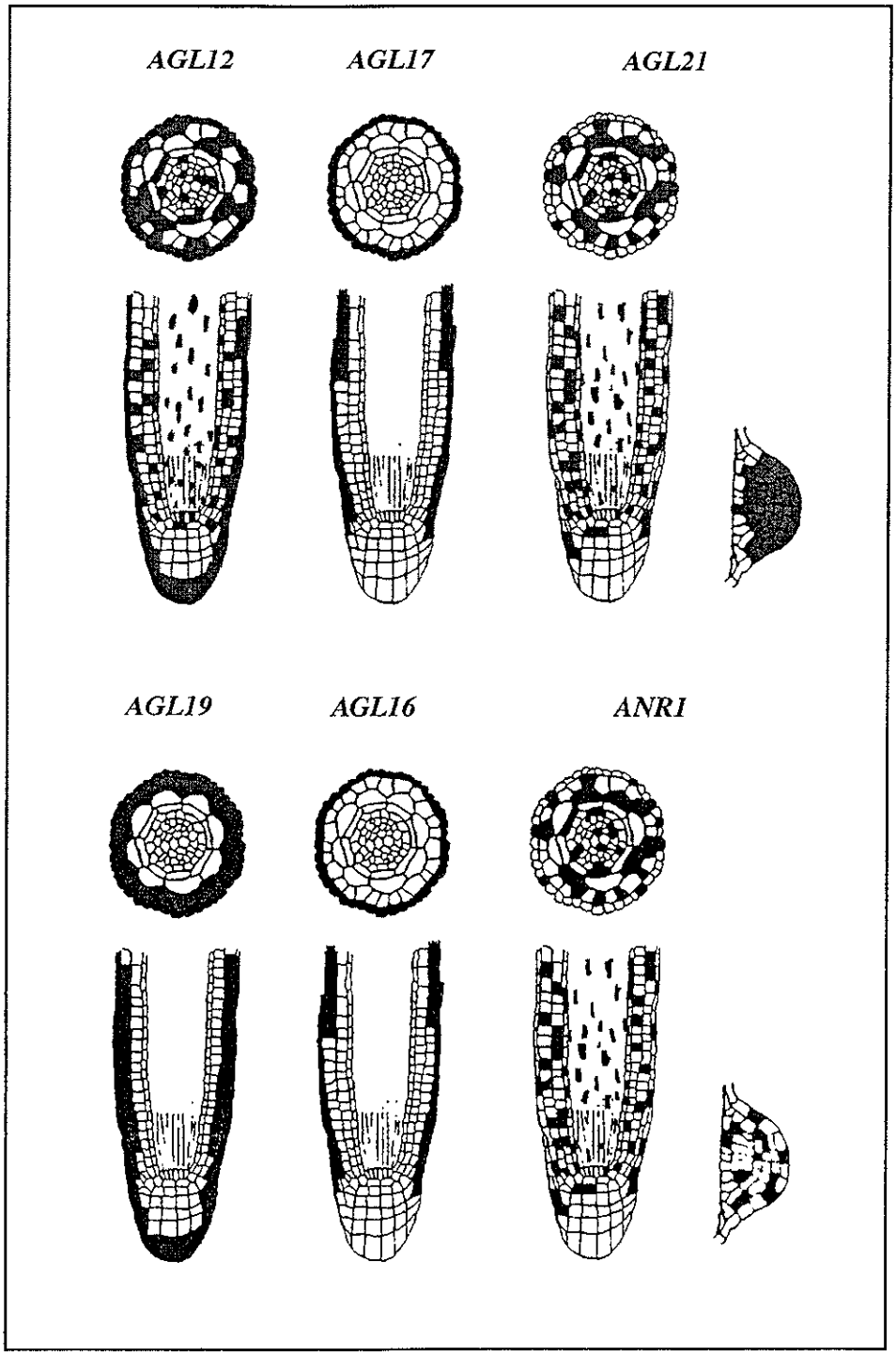

Figura 4: Esquema de los patrones de expresión de ARNm de los genes *AGL12*, *AGL17*, *AGL21*, *AGL19*, *AGL16* y *ANR1* en meristemo apical y primordio de raíz lateral de *Arabidopsis*.

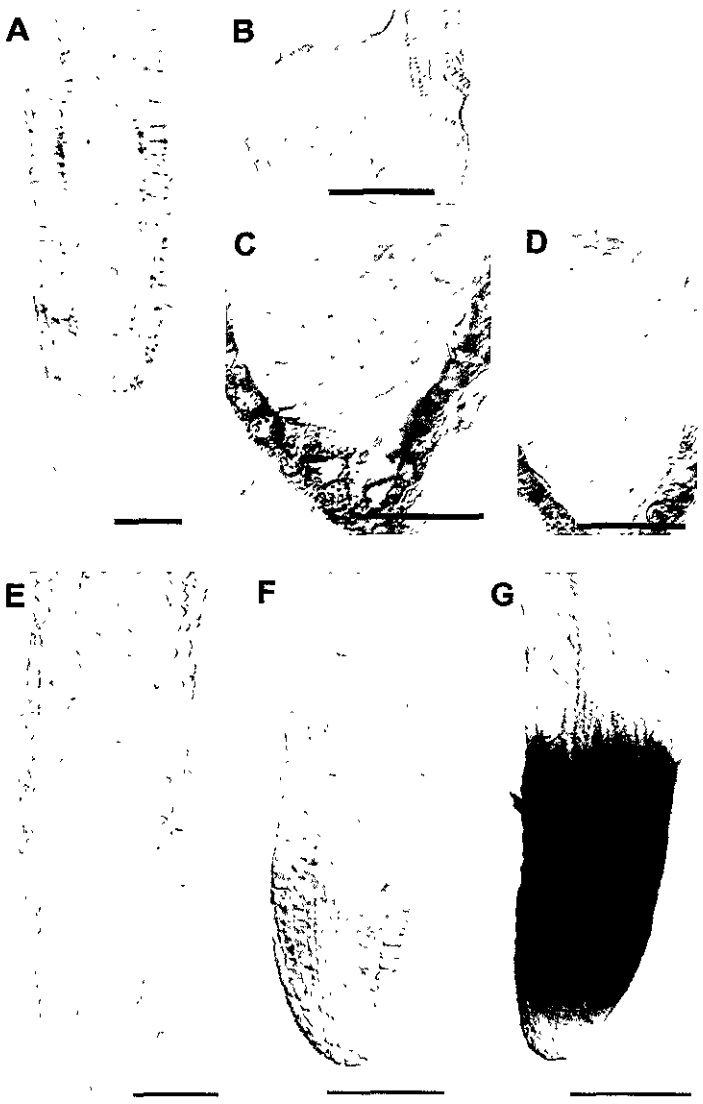

Figura 5: (A-D) Patrón de expresión del ARNm de *ANR1* en raíz y embrión de *Arabidopsis*. Hibridación *in situ* con sonda marcada con digoxigenina. Hibridación de una sonda antisentido en un meristemo apical (A), primordio lateral (B) y embrión (C). D Hibridación de una sonda sentido en embrión. (E-G) Patrón de expresión del ARNm de *AGL16* en raíz de *Arabidopsis*. (E) Hibridación *in situ* con sonda antisentido en corte longitudinal en meristemo apical. F-G *Whole mount in situ* con sonda sentido (F) y antisentido(G) del apex radical. Escala 50um.

9. Discusión y conclusiones

Los genes MADS-box que han sido estudiados presentan patrones de expresión diversos y algunos se expresan simultáneamente en diferentes zonas de la raíz. Los genes se pueden agrupar en función de la región en que se expresan y de la distribución de su transcrito (figura 4).

Apice radical

A nivel del ápice de la raíz, los genes MADS-box siguen dos tendencias. Genes como *AGL17, (AGL16)* y *AGL19* tienen una expresión contínua, asociada exclusivamente a ciertos tipos celulares, como la epidermis, la cofia lateral y la columela, por lo que se pueden considerar como marcadores de estos tejidos externos y algunos de ellos podrían estar involucrados en la regulación de procesos moleculares específicos de estos tejidos. En la cofia lateral y la epidermis tienen lugar procesos tales como la detección de estímulos externos o la redistribución basipétala de auxinas, que pueden afectar la morfogénesis y el crecimiento de la raíz (Müller et al. 1998; Tsugeki y Fedoroff 1999; Rashotte et al. 2000). La detección de los estímulos externos son primordiales para guiar la respuesta de la planta a las variaciones ambientales, es decir, para modular su plasticidad fenotípica. *AGL17* se expresa a lo largo de la cofia lateral y al final de la zona proliferativa se expresa también en la epidermis. Este tipo de patrón no parece ser inusual en raíz, ya que se asemeja al de ciertas lineas *GAL4-GFP enhancer trap* como la J2601[9] (Haseloff et al. 1997; Springer 2000). En la zona de la epidermis donde se expresa *AGL17* tienen lugar procesos morfogenéticos importantes tales como la fijación de la determinación de los tipos celulares de la epidermis o la elongación diferencial de las células que ocurre durante la respuesta a estímulos de gravedad (Masson 1995;

[9] Datos del catálogo de Jim Haseloff (http://www.plantsci.cam.ac.uk/Haseloff/IndexCatalogue.html)

Berger et al. 1998b). En esta zona, además, las células dejan de dividirse, definiendo así el límite basal del meristemo, donde probablemente tienen lugar mecanismos de control que modulan crecimiento del órgano, ya que el tamaño del meristemo tiene un impacto directo sobre la tasa de crecimiento de la raíz (Beemster y Baskin 1998; Baskin 2000). Así, estos patrones de expresión continuos en los tejidos externos de la raíz indican que las observaciones que hay que realizar en las lineas mutantes[10] *agl17* y *agl19* deben explorar varios aspectos, desde el estudio de las características celulares intrínsecas de estos tejidos, hasta eventualmente el estudio de los parámetros de crecimiento de las raíces mutantes y su respuesta a modificaciones ambientales.

Un segundo grupo de genes como *AGL21*, *AGL12* y *ANR1* se expresa de manera discontinua en los diferentes tipos celulares del meristemo, dando lugar a un patrón punteado de expresión. El patrón discontínuo que muestra *AGL12* tiene además la particularidad de que en la epidermis, los atricoblastos presentan una mayor frecuencia de tinción que los tricoblastos en la zona del meristemo más alejada de las iniciales. Esta característica parece denotar una diferencia entre los tricoblastos y el resto de las células del meristemo y se une a las diferencias ya descritas que existen entre los dos tipos celulares de la epidermis[11] a esta altura del meristemo (Galway et al. 1994, Berger et al. 1998b). En general, este tipo de patrón discontinuo es difícil de interpretar, pero la

[10] En el laboratorio de genética molecular y evolución se están analizando actualmente distintas líneas de mutantes por inserción para los genes MADS-box aquí estudiados (Alvarez-Buylla com. pers.).

[11] Los tricoblastos son más cortos que los atricoblastos, por lo que el número de células que conforman una fila celular de tricoblastos -contada desde la célula inicial hasta el comienzo de la zona de diferenciación- es 1.3 veces mayor que en una fila de atricoblastos en el mismo segmento. Según Berger et al. (1998b) esto demuestra que los tricoblastos se dividen más rápido que los atricoblastos. Mediciones más detalladas de la proporción tricoblastos/atricoblastos en distintas subregiones de la zona de crecimiento de la raíz, muestran que esta proporción aumenta a partir de la zona del meristemo más alejada de las iniciales, lo que parece coincidir con la zona en la epidermis, donde *AGL12* se expresa preferencialmente en los atricoblastos. Por otro lado, Baskin (2000) postula que la diferencia en el número de células que componen a estas filas epidérmicas deriva de una disminución inicial y momentánea de la tasa de división en atricoblastos, lo que es suficiente para provocar un aumento en la longitud de la célula, que después se divide a la misma tasa de división de cualquier célula en el meristemo, manteniendo las diferencias de longitud entre ambos tipos celulares. Probablemente se carece aún de una metodología que posea un nivel de resolución suficiente para determinar si ligeras modificaciones en las tasas de división celular están involucradas en los procesos de desarrollo de la raíz.

presencia del transcrito en todos los tejidos podría estar indicando la participación de estos genes en procesos generales del funcionamiento del meristemo. Un patrón de expresión punteado podría deberse a una transcripción preferencial en ciertas fases del ciclo celular. Tal es el caso de los patrones de expresión de ciertos genes de histonas, ciclinas, cinasas dependientes de ciclinas (cdc) y otras proteínas involucradas en la división celular (Fobert et al. 1994; Colón-Carmona et al. 1999). Por otro lado, este patrón podría más bien reflejar una distribución heterogénea de los estímulos que inducen la transcripción de estos genes, lo cual podría suceder si esta inducción fuera elicitada, por señales ambientales externas, como es el caso de algunas ciclinas D (De Veylder et al. 1999). Estos indicios muestran que no podemos descartar la posibilidad de que la expresión de este tipo de genes a nivel del meristemo apical pueda incrementarse en condiciones ambientales (e.g. nutrientes) particulares, distintas a las condiciones "estándar" que se usaron durante la realización de este trabajo. La exploración de esta posibilidad podría realizarse monitoreando la existencia de variaciones en el patrón y/o la intensidad de expresión de los genes en distintas condiciones de crecimiento (e.g. medios de cultivo con distintas concentraciones de nutrientes, etc.), lo cual puede realizarse con líneas de plantas transformadas con el promotor del gen fusionado a un gen reportero[12]. Paralelamente se tendría que analizar el fenotipo de las mutantes en las mismas condiciones. Ambas líneas de investigación se están desarrollando en el laboratorio.

Primordio de raíz lateral

Durante el desarrollo de los primordios laterales se detectó la expresión de los genes *AGL21* y *ANR1*. Es de notar la relación que existe entre el patrón de expresión de este primer gen y las zonas de la raíz que tienen un potencial proliferativo. Los datos funcionales que se han obtenido para *ANR1* demuestran que su expresión es inducida

[12] La realización de *in situ*, en este caso, resultaría demasiado laboriosa.

directamente por un aumento localizado en las concentraciones en nitratos, lo que provoca un estímulo en la elongación de las raíces laterales, resultado de un aumento en la tasa de producción celular (Zhang y Forde 1998; Zhang et al. 1999). En nuestras condiciones estándar de crecimiento –es decir de no inducción-, este gen presenta generalmente un patrón discontinuo punteado en los primordios laterales, mientras que *AGL21* se expresa de manera contínua. Ambos genes pertenecen al mismo clado filogenético y podrían tener cierto grado de redundancia funcional. Las variaciones que observamos en la distribución del transcrito durante el desarrollo del primordio lateral, podrían indicar que estos genes juegan papeles similares, pero en respuesta a distintas señales inductoras. Así, la integración de los datos funcionales de *ANR1* con los patrones de expresión de *AGL21* y *ANR1* en primordio lateral sugieren, nuevamente, que este tipo de genes podrían mediar las respuestas morfogenéticas de la raíz en respuesta a condiciones ambientales (e.g. nutrientes) particulares. En el caso de *ANR1* parece ser las fuentes localizadas de nitrato. Alternativamente, estos genes podrían estar participando en distintos momentos del proceso de formación y crecimiento de las raíces laterales. El análisis del fenotipo de las dobles mutantes *agl21 anr1* cultivadas en distintos medios de cultivo, seguramente aportará información para el análisis de ambas hipótesis.

Embrión

Al iniciar este trabajo se decidió analizar también los patrones de expresión de los distintos genes durante la embriogénesis, con la finalidad de detectar si su mARN podía asociarse a los procesos tempranos de formación del meristemo radicular. *AGL12* y *AGL21* se expresan en embrión, sin embargo, no parece haber una correlación clara entre su expresión y la formación de la radícula. *AGL21* parece en ocasiones expresarse en un gradiente decreciente desde el ápice hacia la base del embrión. Este gradiente

podría considerarse como un marcador del establecimiento de un patrón ápico-basal, en el cual el gen podría eventualmente estar involucrado. Por otro lado el patrón podría estar relacionado con el grado de actividad proliferativa de las células, la cual aumenta en el ápice del embrión durante el desarrollo de los cotiledones (B. Scheres comunicación personal). *AGL12*, a su vez, presenta un patrón discontínuo punteado, donde la frecuencia de células teñidas aumenta a medida que el embrión se desarrolla.

Raíz diferenciada

La mayoría de los genes MADS-box de raíz tales como *AGL19*, *AGL12*, *AGL21* y *AGL17* se expresan en la raíz madura, y fueron detectados a nivel de las células del cilindro central, a lo largo del cual el nivel de tinción puede variar, siendo en general más intenso en ciertas células del pericicio. Este patrón es difícil de interpretar debido, entre otras cosas, a que los *in situ* en raíz de *Arabidopsis* no proporcionan datos suficientemente precisos sobre lo que sucede a lo largo de las zonas de elongación y diferenciación. Esto impide definir la posición de las células a lo largo del eje en donde se inicia la acumulación de los transcritos. Esto permitiría, *a priori*, situarlas con respecto a los procesos de maduración o adquisición de las características finales de los tipos celulares (es decir, establecer su posición relativa a referencias estructurales o morfológicas del grado de diferenciación de las células de la raíz, por ejemplo, la aparición de la primera protuberancia en una fila de tricoblastos o el inicio de la lignificación del protoxilema). En todo caso, la expresión en células de pericicio podría indicar una relación con la formación de primordios de raíz lateral; sin embargo la técnica no permite un mantenimiento óptimo de la estructura tisular que facilite caracterizar sistemáticamene la posición de estas células teñidas con respecto a los vasos de protoxilema. No podemos descartar, a su vez, una relación de la expresión en estos tejidos con el proceso de desarrollo o engrosamiento secundario del cilindro

central de la raíz, el cual deriva de la formación de cambium o meristemos secundarios, procesos que hasta ahora han sido poco estudiados (Steeves y Sussex 1989). Las líneas de *Arabidopsis* transformadas con construcciones de genes marcadores regulados por los promotores de los distintos genes ayudarán a caracterizar con mucho mas detalle estos patrones de expresión en raíz madura.

Duplicación y redundancia

Los estudios filogenéticos de los miembros de la familia MADS-box muestran que los dos genes que se clonaron, *AGL19* y *AGL21*, pertenecen a 2 clados distintos y que cada uno presenta una similitud de secuencia muy alta con otro miembro de su clado respectivo. AGL19 y AGL14 tienen el 76.0% de los aminoácidos idénticos, lo que sugeriría que ambos se originaron de la duplicación de un gen ancestral y que probablemente comparten funciones similares. Sin embargo, para poder concretar mejor una hipótesis de redundancia, sería necesario también poder comparar los patrones de expresión de *AGL14*[13] y *AGL19*.

AGL17 y AGL21 a su vez, tienen el 73.4% de aminoácidos idénticos, lo que sugeriría que también provienen de la duplicación más reciente dentro del clado ANR1. Estos datos filogenéticos, que a primera vista sugieren la existencia de una posible redundancia funcional entre estos dos genes, contrastan con las evidencias obtenidas por el estudio de los patrones de expresión. Ambos genes se expresan de manera bastante distinta, lo que debilita la posibilidad de que exista un traslape real entre las funciones de ambos genes. Sin embargo, al interior del clado ANR1 existen miembros que comparten patrones de expresión similares. Tal es el caso de *ANR1* y *AGL21* y los primeros datos que se obtuvieron para *AGL16* indican que en raíz este gen se expresa en

[13] La caracterización del patrón de expresión de *AGL14* por in *situ* se realiza actualmente en el laboratorio (Ortiz-Moreno, Burgeff, Alvarez-Buylla, datos no publicados).

adquisición de una nueva función por alguno de los 2 genes. Estudios evolutivos de la genética del desarrollo animal han demostrado que los genes reguladores y los genes específicos de ciertos tipos celulares pueden adquirir nuevos dominios de expresión (y por ende nuevas funciones), mediadas por cambios en sus promotores o enhancers, sin necesidad aparente de un cambio inicial en la secuencia codificadora, es decir, en la proteína (Shimeld et al. 1999). *WEREWOLF* (*WER*) y *GLABRA1* (*GL1*) codifican para factores de transcripción que intervienen en la especificación de tipos celulares de la epidermis en raíz y hoja de *Arabidopsis* respectivamente, y presentan patrones de expresión que no coinciden espacialmente. Estudios de complementación recíproca entre estos genes parálogos de la familia MYB de factores de transcripción, indican que ambos codifican proteínas funcionalmente equivalentes y que la diferencia en los papeles que juegan durante el desarrollo se debe exclusivamente a cambios en sus secuencias *cis*-regulatorias (Lee y Schiefelbein 2001). Esto demuestra que cambios a nivel de las regiones *cis*-regulatorias de reguladores transcripcionales pueden ser primordiales en la evolución de la morfología en plantas (Doebley y Lukens 1998; Wang et al. 1999; Carroll 2000; Lee y Schiefelbein 2001). Si interpretamos desde esta óptica los resultados obtenidos para *AGL21* y *AGL17*, podríamos pensar que estamos frente a un caso ejemplar para estudiar, desde un punto de vista evolutivo, el efecto que las modificaciones en las secuencias *cis*-regulatorias de genes reguladores pueden tener sobre el desarrollo. Tal vez los mecanismos de evolución molecular que originaron a ambos genes y provocaron la divergencia de sus patrones de expresión sean de alguna manera similares a aquellos que, durante la duplicación y divergencia de los genes MADS-box del linaje de tipo II, dieron origen a los distintos clados florales y vegetativos actuales. Los datos funcionales de estos genes, así como el estudio a nivel

molecular y de *phylogenetic footprinting*[15] de estas secuencias, serán indispensables para poder profundizar en la comprensión de este caso (Fickett y Wasserman 2000).

Los patrones de expresión de los genes MADS-box aquí estudiados, sugieren que estos tienen funciones diversas y contrastantes durante el desarrollo de la raíz de *Arabidopsis*, y servirán de guía en futuros estudios funcionales. Estos contribuirán al entendimiento del vínculo evolutivo entre la familia MADS-box y la morfología en plantas.

[15]Estos estudios se realizan actualmente en el laboratorio (Martínez-Castilla y Alvarez-Buylla, datos no publicados)

10. Bibliografía

Alvarez-Buylla ER, Liljegren SJ, Pelaz S, Gold SJ, Burgeff C, Ditta GS, Vergara-Silva F, Yanofsky MF (2000b) MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells roots and trichomes. Plant J 24: 457-466

Alvarez-Buylla ER, Pelaz S, Liljegren SJ, Gold SE, Burgeff C, Ditta GS, Ribas de Pouplana L, Martínez-Castilla L, Yanofsky MF (2000a) An ancestral MADS-box duplication ocurred before the divergence of plants and animals. Proc Natl Acad Sci USA 97: 5328-5333

Baskin TI (2000) On the constancy of cell division rate in the root meristem. Plant Mol Biol 43: 545-554

Beemster GTS, Baskin TI (1998) Analysis of cell division and elongation underlying the developmental acceleration of root growth in *Arabidopsis thaliana*. Plant Physiol 116: 1515-1526

Benfey P, Schiefelbein JW (1994) Getting to the root of plant development: the genetics of *Arabidopsis* root formation. TIG 10: 84-88

Benfey P, Linstead PJ, Roberts K, Schiefelbein JW, Hauser MT, Aeschbacher RA (1993) Root development in *Arabidopsis*: four mutants with dramatically altered root morphogenesis. Development 119: 57-70.

Berger F, Yung CY, Dolan L, Schiefelbein J (1998b) Control of cell division in the root epidermis of *Arabidopsis thaliana*. Dev Biol 194: 235-245

Berger F, Haseloff J, Schiefelbein J, Dolan L (1998a) Positional information in root epidermis in defined during embriogenesis and acts in domains with strict boundaries. Current Biol 8: 421-430

Berleth T, Jürgens G (1993) The role of the *monopteros* gene in organising the basal body region of the *Arabidopsis* embryo. Development 118: 575-587

Burgeff C, Liljegren SJ, Tapia López R, Yanofsky MF, Alvarez-Buylla ER (en prensa) MADS-box gene expression in lateral primordia, meristems and differentiated tissues of *Arabidopsis thaliana* roots. Planta (DOI 10.1007/s004250100637)

Burke AC, Nelson CE, Morgan BA, Tabin C (1995) Hox genes and the evolution of vertebrate axial morphology. Development 121: 333-346

Carroll SB (2000) Endless forms: The evolution of gene regulation and morphological diversity. Cell 101: 577-580

Celenza JL Jr, Grisafi PL, Fink GR (1995) A pathway for lateral root formation in *Arabidopsis thaliana*. Genes Dev 9: 2131-2142

Clark SE, Jacobsen SE, Levin JZ, Meyerowitz EM (1996) The *CLAVATA* and *SHOOTMERISTEMLESS* loci competitively regulate meristem activity in *Arabidopsis*. Development 122: 1567-1575

Coen ES, Meyerowitz EM (1991) The war of the whorls: genetic interactions controlling flower development. Nature 353: 31-37

Colón-Carmona A, You R, Haimovitch-Gal T, Doerner P (1999) Spatiotemporal analysis of mitotic activity with a labile cyclin-GUS fusion protein. Plant J 20: 503-508

Creusot F, Fouilloux E, Dron M, Lafleuriel J, Picard G, Billault A, Le Paslier D, Cohen D, Chaboue ME, Durr A, et al. (1995) The CIC library: a large insert YAC library for genome mapping in *Arabidopsis thaliana*. Plant J 8: 763-770

Cheng JC, Seeley KA, Sung ZR (1995) *RML1* and *RML2*, *Arabidopsis* genes required for cell proliferation at the root tip. Plant Physiol 107: 365-376

Cho S, Jang S, Chae S, Chung KM, Moon YH, An G, Jang SK (1999) Analysis of the C-terminal region of *Arabidopsis thaliana APETALA1* as a transcription activation domain. Plant Mol Biol 40: 419-429

De Veylder L, de Almeida Engler J, Burssens S, Manevski A, Lescure B, Van Montagu M, Engler G, Inze D (1999) A new D-type cyclin of *Arabidopsis thaliana* is expressed during lateral root primordia formation. Planta 208: 453-462

Di Laurenzio L, Wysocka-Diller J, Malamy JE, Pysh L, Helariutta Y, Freshour G, Hahn MG, Feldmann KA, Benfey PN (1996) The *SCARECROW* gene regulates an asymmetric cell division that is essential for generating the radial organization of the *Arabidopsis* root. Cell 86: 423-433

Doebley J, Lukens L (1998) Transcriptional regulators and the evolution of plant form. Plant Cell 10: 1075-1082

Dolan L, Janmaat K, Willemsen V, Linstead P, Poethig S., Roberts K, Scheres B (1993) Cellular organisation of the *Arabidopsis thaliana* root. Development 119: 71-84

Ferrándiz C, Gu Q, Martienssen R, Yanofsky MF (2000a) Redundant regulation of meristem identity and plant architecture by *FRUITFULL*, *APETALA1* and *CAULIFLOWER*. Development 127: 725-734

Ferrándiz C, Liljegren SJ, Yanofsky MF (2000b) Negative regulation of the *SHATTERPROOF* genes by *FRUITFULL* during *Arabidopsis* fruit development. Science 289: 436-438

Fickett JW, Wasserman WW (2000) Discovery and modeling of transcriptional regulatory regions. Curr Op in Biotech 11: 19-24

Fobert PR, Coen ES, Murphy GJP, Doonah JH (1994) Patterns of cell division revealed by transcriptional regulation of genes during the cell cycle in plants. EMBO J 13: 616-624

Galway ME, Masucci JD, Lloyd AM, Walbot V, Davis RW, Schiefelbein JW (1994) The *TTG* gene is required to specify epidermal cell fate and cell patterning in the *Arabidopsis* Root. Dev Biol 166: 740-754

Gu Q, Ferrándiz C, Yanofksy M., Martienssen R (1998) The *FRUITFULL* MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. Development 125: 1509-1517

Hamann T, Mayer U, Jürgens G (1999) The auxin-insensitive *bodenlos* mutation affects primary root formation and apical-basal patterning in the *Arabidopsis* embryo. Development 126: 1387-1395.

Hardtke CS, Berleth T (1998) The *Arabidopsis* gene *MONOPTEROS* encodes a transcription factor mediating embryo axis formation and vascular development. EMBO J 17: 1405-1411.

Hartmann U, Höhmann S, Nettesheim K, Wisman E, Saedler H, Huijser P (2000) Molecular cloning of *SVP*: a negative regulator of floral transition in *Arabidopsis*. Plant J 21: 351-360

Haseloff J, Siemering KR, Prasher DC, Hodge S (1997) Removal of a cryptic intron and subcellular localization fo green fluorescent protein are required tomark transgenic *Arabidopsis* plants brightly. Proc Natl Acad Sci USA 94: 2122-2127

Hauser MT, Morikami A, Benfey PN (1995) Conditional root expansion mutants of *Arabidopsis*. Development 121: 1237-1252

Heard J, Dunn K (1995) Symbiotic induction of a MADS-box gene during development of alfalfa root nodules. Proc Natl Acad Sci USA 92: 5273-5277

Heard J, Caspi M, Dunn K (1997) Evolutionary diversity of symbiotically induced nodule MADS-box genes: characterization of nmhC5, a member of a new subfamily. Mol Plant Microbe Interact 10: 665-676

Hemerly A, de Almeida Engler J, Bergounioux C, Van Montagu M, Engler G, Inze D, Ferreira P (1995) Dominant negative mutnts of the Cdc2 kinase uncouple cell division from iterative plant development. EMBO J 14: 3925-3936

Howell SH (1998) Molecular Genetics of Plant Development. Cambridge University Press, UK. 365p

Jack T, Brockman LL, Meyerowitz EM (1992) The homeotic gene *APETALA 3* of *Arabidopsis thaliana* encodes a MADS-box and is expressed in petals and stamens. Cell 68: 683-697

Jürgens G, Mayer U, Torres-Ruiz RA, Berleth T, Misera S (1991) Genetic analysis of pattern formation in the *Arabidopsis* embryo. Development supplement 1: 27-38

Kaplan DR, Hagemann W (1991) The relationship of cell and organism in vascular plants. Bioscience 41: 693-703

Kempin SA, Savidge B, Yanofsky MF (1995) Molecular basis of the *cauliflower* phenotype in *Arabidopsis*. Science 267: 522-525

Kranz HD, Denekamp M, Greco R, Jin H, et al. (1998) Towards functional characterisation of the members of the *R2R3-MYB* gene family from *Arabidopsis thaliana*. Plant Journal 16: 263-276

Lee MM, Schiefelbein J (1999) WEREWOLF, a MYB-related protein in *Arabidopsis* is a position-dependent regulator of epidermal cell patterning.Cell 99: 473-483

Lee H, Suh SS, Park E, Cho E, Ahn JH, Kim SG, Lee JS, Kwon YM, Lee I (2000) The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in Arabidopsis. Genes Dev 14: 2366-2376

Lee MM, Schiefelbein J (2001) Developmentally distinct MYB genes encode functionally equivalent proteins in *Arabidopsis*. Development 128: 1539-1546

Li WH (1997) Molecular Evolution. Sinauer Associates, Inc. Sunderland MA, USA. 487p

Liljegren SJ, Ditta GS, Eshed Y, Savidge B, Bowman JL, Yanfosky MF (2000) *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. Nature 404: 766-770.

Liljegren SJ, Ferrándiz C, Alvarez-Buylla RE, Pelaz S, Yanofsky MF (1998) *Arabidopsis* MADS-box genes involved in fruit dehiscence. Flowering Newsletter 25: 9-19

Liu L, White MJ, MacRae TH (1999) Transcription factors and their genes in higher plants. Functional domains, evolution and regulation. Eur J Biochem 262: 247-257

Malamy JE, Benfey P (1997) Organization and cell differentiation in lateral roots of *Arabidopsis thaliana*. Development 124: 33-44

Mandel MA, Gustafson-Brown C, Savidge B, Yanofsky MF (1992) Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. Nature 360: 273-277

Mandel MA, Yanofksy MF (1995) The *Arabidopsis AGL8* MADS-box gene is expressed in inflorescence meristems and is negatively regulated by *APETALA1*. Plant Cell 7: 1763-1771

Masson PH (1995) Root gravitropism. BioEssays 17: 119-127

Masucci JD, Schiefelbein JW (1996) Hormones act downstream of *TTG* and *GL2* to promote root hair outgrouth during epidermis development in the *Arabidopsis* root. Plant Cell 8: 1505-1517

Masucci JD, Rerie WG, Foreman DR, Zhang M, Galway ME, Marks MD, Schiefelbein JW (1996) The homeobox gene *GLABRA 2* is required for position-dependent cell differentiation in the root epidermis of *Arabidopsis thaliana*. Development 122: 1253-1260

Meissner RC, Hailing J, Cominelli E, Denekamp M, et al. (1999) Function search in a large transcription factor gene family in *Arabidopsis*: Assesing the potential of Reverse Genetics to identify insertional mutation in *R2R3 MYB* genes. Plant Cell 11: 1827-1840

Mendoza L, Alvarez-Buylla ER (2000) Genetic regulation of root hair development in Arabidopsis thaliana: A network model. J theor Biol 204: 311-326

Meyerowitz EM, Pruitt RE (1985) *Arabidopsis thaliana* and plant molecular genetics. Science 229: 1214-1218.

Meyerowitz EM (1997) Plants and the logic of development. Genetics 145: 5-9

Müller A, Guan C, Gälweiler L, Tänzler P, Huijser P, Marchant A, Parry G, Bennett M, Wisman E, Palme K (1998) *AtPIN2* defines a locus of *Arabidopsis* for root gravitropism control. EMBO J 17: 6903-6911

Palopoli M F, Patel NH (1996) Neo-Darwinian developmental evolution: can we bridge the gap between pattern and process? Curr Op Gen Dev 6: 502-508

Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF (2000) B and C floral organ identity functions require *SEPALLATA* MADS-box genes. Nature 405: 200-203

Pelaz S, Tapia-López R, Alvarez-Buylla ER, Yanofsky MF (2001) Conversion of leaves into petals in Arabidopsis. Curr Biol 11: 182-184

Purugganan MD, Rounsley SD, Schmidt RJ, Yanofsky MF (1995) Molecular evolution of flower development: Diversification of the plant MADS-box regulatory gene family. Genetics 140: 345-356

Purugganan MD (1998) The molecular evolution of development. BioEssays 20: 700-711

Rashotte AM, Brady SR, Reed RC, Ante SJ, Muday GK (2000) Basipetal auxin transport is required for gravitropism in roots of Arabidopsis. Plant Physiol 122: 481-490

Riechmann JL, Meyerowitz EM (1997) MADS domain proteins in plant development. Biol Chem 378: 1079-1101

Riechmann JL, Ratcliffe OJ (2000) A genomic perspective on plant transcription factors. Curr Op Plant Biol 3: 423-434

Rogg LE, Lasswell J, Bartel B (2001) A Gain-of-function mutation in *IAA28* suppresses lateral root development. Plant Cell 13: 465-480

Rost TL, Baum SF, Nichols S (1996) Root apical organization in *Arabidopsis thaliana* ecotype 'WS' and a comment on root cap structure. Plant Soil 187: 91-95

Rounsley SR, Ditta GS, Yanofsky M (1995) Diverse roles for MADS-Box genes in *Arabidopsis* development. Plant Cell 7:1259-1569

Running MP, Fletcher JC, Meyerowitz EM (1998) The *WIGGUM* gene is required for proper regulation of floral meristem size in *Arabidopsis*. Development 125: 2545-2553

Samach A, Onuchi H, Gold SE, Ditta GS, Schwarz-sommer Z, Yanofksy MF, Coupland G (2000) Distinct roles of CONSTANS target genes in reprodctive development of Arabidopsis. Science 288:1613-1616

Sambrook J, Fritisch EF, Maniatis T (1989) Molecular cloning. A laboratory manual. 2nd edition (Cold Spring Harbour, NY; Cold Spring Harbor Laboratory Press). 3 tomos

Sánchez-Fernández R, Fricker M, Corben LB, White NS, Sheard N, Leaver CJ, Van Montagu M, Inze D, May MJ (1997) Cell proliferation and hair tip growth in the *Arabidopsis* root are under mechanistically different forms of redox control. Proc Natl Acad Sci USA 94: 2745-2750

Scheres B, Wolkenfelt H (1998) The *Arabidopsis* root as a model to study plant development. Plant Physiol Biochem 36: 21-32

Scheres B, Wolkenfelt H, Willemsen V, Terlouw M, Lawson E, Dean C, Weisbeek P (1994) Embryonic origin of the *Arabidopsis* primary root and meristem initials. Development 120: 2475-2487

Scheres B, Heidstra R (1999) Digging out roots: Pattern formation, cell division, and morphogenesis in plants. Curr Top Dev Biol 45: 207-247

Schiefelbein JW, Masucci JD, Wang H (1997) Building a root: The control of patterning and morphogenesis during root development. Plant Cell 9:1089-1098

Schiefelbein JW (2000) Constructing a plant cell. The genetic control of root hair development. Plant Physiology, 124: 1525-1531

Schwarz-Sommer Z, Huijser P, Nacken W, Saedler H, Sommer H (1990) Genetic control of flower development by homeotic genes in *Antirrhinum majus*. Science 250: 931-936

Schwechheimer C, Bevan M (1998) The regulation of transcription factor activity in plants. Trends Plant Sci 3: 378-383

Sheldon CC, Burn JE, Perez PP, Metzger J, Edwards JA, Peacock WJ, Dennis ES (2000) The molecular basis of vernalization: The central role of *FLOWERING LOCUS C* (*FLC*). Proc Natl Acad Sci USA 97: 3753-3758

Shimeld SM (1999) Gene function, gene networks and the fate of duplicated genes. Sem Cell Dev Biol 10: 549-553

Shore P, Sharrocks AD (1995) The MADS-box family of transcription factors. Eur J Biochem 229: 1-13

Smith LG, Hake S, Sylvester AW (1996) The *tangled-1* mutation alters cell division orientations throuthout maize leaf development without altering leaf shape. Development 122: 481-489

Smyth D (2000) A reverse trend-MADS functions revealed. Trends Plant Sci 5: 315-317

Springer PS (2000) Gene traps: tools for plant development and genomics. Plant Cell 12: 1007-1020

Steeves TA, Sussex IM (1989) Patterns in plant development. Cambridge University Press, UK. 388p

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796-815

Theissen G, Kim JT, Saedler H (1996) Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. J Mol Evol 43: 484-516

Thomas JH (1993) Thinking about genetic redundancy. Trends Gen 9: 395-399

Torres-Ruiz RA, Jürgens G (1994) Mutations in the *FASS* gene uncouple pattern formation and morphogenesis in *Arabidopsis* development. Development 120: 2967-2978.

Traas J, Bellini C, Nacry P, Kronenberger J, Bouchez D, Caboche M (1995) Normal differentiation patterns in plants lacking microtubular preprophase bands. Nature 375: 676-677

Treisman R, Ammerer G (1992) The SRF and MCM1 transcription factors. Curr Op Gen Dev 2: 221-226

Tsugeki R, Fedoroff NV (1999) Genetic ablation of root cap cells in *Arabidopsis*. Proc Nat Acad Sci USA 96: 12941-12946

Ulmasov T, Hagen G, Guilfoyle TJ (1997) ARF1, a transcription factor that binds to auxin response elements. Science 276: 1865-1868

van den Berg C, Willemsen V, Hage W, Weisbeek, Scheres B (1995) Cell fate in the *Arabidopsis* root meristem determined by directional signalling. Nature 378: 62-65

van den Berg C, Willemsen V, Hendriks G, Weisbeek P, Scheres B (1997) Short-range control of cell differentiation in the *Arabidopsis* root meristem. Nature 390: 287-289

Vernoux T, Wilson RC, Seeley KA, Riechheld JP, Muroy S, Brown S, Maughan SC, Cobbett CS, Van Montagu M, Inze D, May MJ, Sun GZR (2000) The *ROOT MERISTEMLES1 / CADMIUM SENSITIVE 2* gene defines a glutathione-dependent pathway involved in initiation and maintenance of cell division during postembryonic root development. Plant Cell, 12 : 97-109

Wada T, Tachibana T, Shimura Y, Okada K (1997) Epidermal cell differentiation in *Arabidopsis* determined by a Myb homolog, CPC. Science 277: 1113-1116

Walker AR, Davidson PA, Bolognesi-Winfield AC, James CM, Srinivasan N, Blundel TL, Esch J, Marks MD Gray JC (1999) The *TRANSPARENT TESTA GLABRA1* locus, which regulates trichome differentiation and anthocyanin biosynthesis in *Arabidopsis*, encodes a WD40 repeat protein. Plant Cell 11: 1337-1349

Wang RL, Stec A, Hey J, Lukens L, Doebley J (1999) The limits of selection during maize domestication. Nature 398: 236-239

Weigmann K, Cohen SM, Lehner CF (1997) Cell cycle progression, growth and patterning in imaginal discs despite inhibition of cell division after inactivation of Drosophila Cdc2 kinase. Development 124: 3555-3563

Willemsen V, Wolkenfelt H, de Vrieze G, Weisbeek P, Scheres B (1998) The *HOBBIT* gene is required for formation of the root meristem in the *Arabidopsis* embryo. Development, 125:521-531

Wolpert L, Beddington R, Brockes J, Jessell T, Lawrence P, Meyerowitz E (1998) Principles of development. Current Biology LTD, Oxford University Press, 484p

Yanofsky MF, Ma H, Bowmann JL, Drews GN, Feldmann KA, Meyerowitz EM (1990) The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. Nature 346: 35-39

Zhang H, Forde BG (1998) An *Arabidopsis* MADS-Box gene that controls nutrient-induced changes in root architecture. Science 279: 407-409

Zhang H, Jennings A, Barlow PW, & Forde BG (1999) Dual pathways for regulation of root branching by nitrate. Proc. Natl Acad Sci USA 96: 6529-6534

An ancestral MADS-box gene duplication occurred before the divergence of plants and animals

Elena R. Alvarez-Buylla*†‡, Soraya Pelaz*, Sarah J. Liljegren*, Scott E. Gold*§, Caroline Burgeff†, Gary S. Ditta*, Lluis Ribas de Pouplana¶, León Martínez-Castilla†, and Martin F. Yanofsky*‡

*Department of Biology, University of California at San Diego, La Jolla, CA 92093-0116; †Instituto de Ecología, Universidad Nacional Autónoma de Mexico, AP-Postal 70-275, México D F. 04510, México; and ¶Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92117

Communicated by Elliot M. Meyerowitz, California Institute of Technology, Pasadena, CA, March 13, 2000 (received for review September 8, 1999)

Changes in genes encoding transcriptional regulators can alter development and are important components of the molecular mechanisms of morphological evolution. MADS-box genes encode transcriptional regulators of diverse and important biological functions. In plants, MADS-box genes regulate flower, fruit, leaf, and root development. Recent sequencing efforts in *Arabidopsis* have allowed a nearly complete sampling of the MADS-box gene family from a single plant, something that was lacking in previous phylogenetic studies. To test the long-suspected parallel between the evolution of the MADS-box gene family and the evolution of plant form, a polarized gene phylogeny is necessary. Here we suggest that a gene duplication ancestral to the divergence of plants and animals gave rise to two main lineages of MADS-box genes: TypeI and TypeII. We locate the root of the eukaryotic MADS-box gene family between these two lineages. A novel monophyletic group of plant MADS domains (AGL34 like) seems to be more closely related to previously identified animal SRF-like MADS domains to form TypeI lineage. Most other plant sequences form a clear monophyletic group with animal MEF2-like domains to form TypeII lineage. Only plant TypeII members have a K domain that is downstream of the MADS domain in most plant members previously identified. This suggests that the K domain evolved after the duplication that gave rise to the two lineages. Finally, a group of intermediate plant sequences could be the result of recombination events. These analyses may guide the search for MADS-box sequences in basal eukaryotes and the phylogenetic placement of new genes from other plant species.

MEF2 | SRF | homeotic genes | MADS | development

Changes in genes encoding transcriptional regulators may represent the most important determinants of morphological evolution in plants and animals (1), and phylogenetic analyses provide a historical framework to identify such changes. The MADS-box genes encode a eukaryotic family of transcriptional regulators involved in diverse and important biological functions, ranging from cardiac muscle development in animals to pheromone response in yeast (2). In plants, MADS-box genes encode the three floral homeotic functions predicted by the genetic ABC model of flower organ identity (3, 4). In addition, plant MADS-box genes regulate the timing of flower initiation and flower meristem identity, as well as various aspects of ovule, fruit, leaf, and root development (4, 5).

Previously identified plant MADS-box genes encode proteins that share a stereotypical MIKC structure (Fig. 1), with the highly conserved DNA-binding MADS domain at the amino terminus. The moderately conserved K domain in the central portion of these proteins has been shown to be important for protein–protein interactions and likely forms a coiled-coil structure. The MADS and K domains are linked to one another by a weakly conserved I domain, whereas a poorly conserved carboxyl-terminal (C) region may function as a trans-activation domain (4). In animals and fungi, two distinct types of MADS-box genes have been identified, the SRF-like and MEF2-like classes (ref. 2; see Fig. 1).

This paper provides a hypothesis on the evolutionary history of the eukaryotic MADS-box gene family. Previous studies of eukaryotic MADS-box gene evolution, which included plant and animal sequences, provided unrooted trees useful to infer the phylogenetic relationships of the MADS-box lineages (6). These previous studies suggested that at least one MADS-box gene was present in the common ancestor of plants, animals, and fungi, and that probably the duplication that gave rise to the animal MEF2- and SRF-like genes occurred after animals diverged from plants but before fungi diverged from animals (6). However, previous plant and eukaryotic studies were based on a relatively small sampling of plant MADS-box sequences for a particular species (6–9). To test whether all *Arabidopsis* MADS-box sequences group in a monophyletic clade distinct from all animal and fungal MADS-box sequences, we performed phylogenetic analyses. We used 45 *Arabidopsis* MADS domain sequences, including 26 new ones, 9 sequences representative of the MEF2-like class from animals, and 8 sequences from the animal SRF-like group.

We present a rooted phylogenetic tree of the eukaryotic MADS domain lineages and postulate new hypotheses on the evolutionary history of this gene family. Our results suggest that a duplication ancestral to the divergence of plants and animals gave rise to two lineages (herein called TypeI and TypeII MADS), and that the protein motifs that define each group were fixed in the common ancestors of plants, animals, and fungi. Our analyses also identify new monophyletic clades of plant MADS-box sequences. Most plant MADS-box genes including all of the ones that have been characterized functionally in previous studies, group with the animal MEF2-like sequences in what we have named the TypeII MADS-box lineage. But we have identified a group of *Arabidopsis* MADS-box sequences that seems to be more closely related to the animal SRF-like genes forming the group that we herein call TypeI MADS. This finding suggests that both lineages are present in plants, animals, and fungi. Finally, we show that the K domain, typical of plant MADS-domain proteins, is found only in the TypeII MADS domain sequences of plants, suggesting that this domain evolved after this lineage diverged from the TypeI MADS. These results have enabled us to put forward a model for the evolution of this important family of regulatory genes in eukaryotes (see Fig. 4).

Materials and Methods

Sequence Sources and/or Accession Numbers. Sequence sources or GenBank accession numbers are as follows: *AGAMOUS* (10),

Abbreviations MP, maximum parsimony; NJ, neighbor joining, QP, quartet puzzling, USP, Universal Stress Protein

‡To whom reprint requests should be addressed E-mail abuylla@servidor unam mx or marty@biomail.ucsd edu.

§Present address Department of Plant Pathology, University of Georgia, Athens, GA 30602-7274

The publication costs of this article were defrayed in part by page charge payment This article must therefore be hereby marked "*advertisement*" in accordance with 18 U S C §1734 solely to indicate this fact

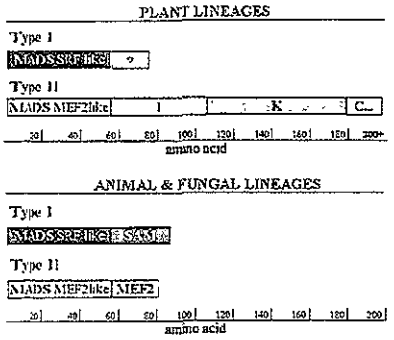

Fig. 1. Schematic representation of the protein domains of plant, animal, and fungal TypeI (SRF-like) and TypeII (MEF2-like) MADS-domain proteins. The scale indicates the number of amino acids along the protein. Plant TypeII-like proteins have carboxyl-terminal domains that go beyond 200 amino acids. In plant TypeI-like proteins the "?" indicates carboxyl-terminal domains not well defined yet and of variable lengths

Fig. 2. Amino acid alignment of the MADS-domain (amino acids 1 to 60) for some representative members of the plant, animal, and fungal TypeI (SRF-like) and TypeII (MEF2-like) lineages. We also show representative sequences of the genes that are not clearly assigned to either one (MADS-domains Type?). One gene from each monophyletic clade identified in MP and NJ was selected. Conserved amino acids within each group and not found in any (or in no more than two) of the MADS domains of the other group are in red. Green names indicate plant sequences and red names, animal or fungal ones (see *Materials and Methods*).

APETALA3 (11), *PISTILLATA* (12), *AGL1–6* (13), *APETALA1* (14), *AGL8* (15), *AGL9* (16), *CAULIFLOWER* (17), *AGL11, AGL12, AGL13, AGL14, AGL15* and *AGL17* (18), *AGL16* (AL137080, S.L. and M.Y., unpublished data), *AGL18* (AL137080, S.G. and M.Y., unpublished data), *AGL19* (AL161558, S.G. and M.Y., unpublished data), *AGL20* (AC003680, S.G. and M.Y., unpublished data), *AGL21* (ATF20D10), *AGL22* (AC006592), *AGL23* (AC004512), *AGL24* (AF005158), *AGL25* (AF116527), *AGL26* (AF007270), *AGL27* (AC002291/cDNA sequence, S.P. and M.Y., unpublished data), *AGL28* (Y12776), *AGL29* (AC004077), *AGL30* (AC004138), *AGL31* (T45787/cDNA sequence, S.P. and M.Y., unpublished data), *AGL32* (AB007648), *AGL33* (AC004484), *AGL34* (AF058914), *AGL35* (AF058914), *AGL36* (AF058914), *AGL37* (AC00451), *AGL38* (AC004512), *AGL39* (AF007271), *AGL40* (Z99708), *ANR1* (19). *AGL23, AGL26,* and *AGL28–38* were recently identified by the *Arabidopsis* Genome Sequencing project. Although we lack cDNA clones for these genes, their predicted MADS-box domain sequences, on which our analyses are based, are unequivocal, because no introns have ever been found in this region.

GenBank accession numbers for the animal and fungal sequences are as follows. The *MEF2*-like genes used are: *Homo sapiens MEF2C* (L08895), *Caenorhabditis elegans CEMEF2* (U36198), *H. sapiens MEF2A* (S25831), *H. sapiens MEF2D* (Q14814), *Halocynthia roretzi ASMEF2* (D49970), *H. sapiens MEF2B* (X68502), *Drosophila melanogaster DMEF2* (U03292), *Saccharomyces cerevisiae SMP1* (P38128), and *S. cerevisiae RLM1* (D63340). The *SRF*-like genes used are: *H. sapiens SRF* (J03161), *Xenopus laevis* SRF (S15018), *D melanogaster DSRF* (X77532), *S. cerevisiae* MCM1 (P11746), *S cerevisiae ARG80* (X05327), and *Schizosaccharomyces pombe PLN* (D78483). The bacterial Universal Stress Protein (*USP*) family sequences that served as outgroup for some of the analyses are: *Escherichia coli* EcuspA (X67639), *E coli* Ecyit (P32132), *Coxiella burnetti* Coxyfmu (P45680), and *Bacillus subtilis* Bsyxic (P42297).

Alignment and Phylogenetic Analyses. We used 65 amino acid sequences for the analyses. These cover the 57–60 amino acids that different authors (2, 6) have defined as the MADS domain plus a few additional conserved amino acids. These sequences were aligned by using CLUSTAL X; the alignment generated was unambiguous (complete alignment available from authors on request, and see Fig. 2). Phylogenetic analyses were conducted with unweighted maximum parsimony (MP), neighbor joining (NJ), and quartet puzzling (QP), by using the test version 4d64 of PAUP* (D. L. Swofford, Laboratory of Molecular Systematics, Smithsonian Institution, Washington, D.C.). For MP analyses, 100 replicates of random addition sequences keeping all optimal trees in each replicate, TBR branch swapping, and no maxtrees limit were used. Gaps were treated as missing data. NJ analyses were done by using the default factory settings and the p-distance (proportion of different amino acids between two sequences) as a distance estimator. This is the recommended distance measure when comparing distantly related sequences, because it has a smaller variance than other estimates (20).

Nonparametric bootstrap (100 pseudoreplicates) was used to assess the reliability of individual branches. Bootstrap proportions are considered here as an index of support for a particular clade and not a statement about probability or confidence limit in the statistical sense (21). QP trees were based on 1,000 replicates by using the factory default settings. The phylogenetic relationships inferred from the trees presented here do not depend on specific sequences used to estimate phylogeny; by using subsamples of protein sequences, the same relationships were inferred (data not shown). Trees were examined with TREEVIEW (22).

To study the branching order of MADS-box gene lineages and the timing of duplications relative to the divergence of the main groups of eukaryotes (plants, animals, and fungi), we need a rooted tree. An unambiguous root location depends on using an outgroup MADS-box domain sequence. We have attempted this rooting by using four bacterial sequences that belong to the USP family as outgroup. These share very few conserved amino acids with known eukaryotic MADS-box sequences but have been defined as MADS-domain homologues based on these few conserved residues and other functional criteria (23). A better outgroup could come from a taxon representative of a sister clade of plants, animals, and fungi, such as *Euglena*, but this is not yet available.

As an alternative way to objectively root the MADS-box tree, we used a parsimony-based approach from Page and Charleston (24, 25). This method reconciles the gene tree to the species tree and finds the rooted gene tree that minimizes the number of gene sorting events (which could include gene losses or insufficient sampling of genomes) and duplications. This is the MADS

EVOLUTION

domain tree that we put forward as a polarized phylogenetic hypothesis for this gene family. We used the species tree proposed by Baldauf and Palmer (26), in which animals and fungi are each other's closest relatives. We used groups of sequences that were shared by the NJ and MP trees, which were supported by high bootstrap values and were *bona fide* subfamilies, as possible outgroups to be tested. We tested seven alternative outgroups from the NJ and MP searches. The reconciled trees' method requires completely resolved trees. Therefore, one tree from each island sampled in the MP search was used. Trees from each island were very similar and differed only in some of the terminal branches. To avoid a bias because of the excessive number of possible losses and duplications found in the terminal branches where only taxa from either plants (only *Arabidopsis*) or one of the animal or fungal groups used were represented, we repeated this analysis by counting only basal duplications (i.e., those that are at the base of clades that combine sequences of plants, animals, and fungi).

Protein Structure Prediction. The predictions of coiled-coil regions within the protein sequences were performed with the programs PAIRCOIL and MULTICOIL (27, 28) and were based on the presence in the sequences of heptat-repeat signature motifs. In all cases, both programs used yielded the same result. A K domain was predicted to be present when the probability cutoff of finding coiled-coils downstream of the MADS-box domain was >0.35. The default value of 0.5 has been determined empirically to work well. However, to avoid false negatives, we decreased the cutoff value by 20%. Additionally, we predicted possible protein secondary structures using discrete state-space probability models, as implemented by the program PSA (http://bmerc-www.bu.edu/psa; ref. 29). These predictions identified α-helices for the same sequences and were used to confirm results obtained from the coiled-coil prediction programs.

Results and Discussion

Ancient Duplications of Eukaryotic MADS-Box Sequences. We present molecular evolutionary analyses of plant, animal, and fungal MADS-domain sequences, including 26 newly identified MADS-domain sequences from *Arabidopsis*, along with 19 previously analyzed members of this extensive gene family. The most striking result of our analyses is the discovery that animal and fungal MEF2-like sequences are more closely related to most plant MADS-domain sequences than to animal SRF-like sequences. Some conserved amino acids put the MEF2-like animal and most plant sequences in a clear monophyletic clade (hereafter referred to as TypeII MADS domains), suggesting that at least one gene-duplication event occurred before the divergence of plants and animals. In addition, a group of *Arabidopsis* MADS-domain sequences (AGL34-like) seem to share a more closely related ancestor with the SRF-like sequences of animals and fungi than with other plant MADS-domain sequences. The clade formed by these two related groups is referred to hereafter as the lineage of TypeI MADS domains. However, the monophyly of this group is not as well supported as that of the TypeII MADS domains, because it is supported by very few shared and unique amino acids (Fig. 2). Finally, we found a group of intermediate plant sequences that could be the result of recombination between TypeI and II MADS-box genes. These results are based on NJ, QP, and MP phylogenies, described below.

The NJ tree rooted with the putative MADS-domain sequences from bacteria is well resolved (Fig. 3*a*) and is similar to the one obtained by the rooting method described below (Fig. 3*b*). In the tree of Fig. 3*a*, the TypeII MADS domains that group the animal MEF2-like and most plant sequences form a well-supported monophyletic clade. However, the rest of the clades that in Fig. 3*b* are grouped into the TypeI lineage do not form a monophyletic group in Fig. 3*a*. Results in Fig. 3*a* suggest that AGL39-like sequences were lost or have not been found in animals and fungi. Both of the latter possibilities are unlikely, because yeast and *C. elegans*, whose genomes are completely sequenced, have both TypeI and TypeII MADS domains and no other types. It would be highly improbable that in both organisms the same genes were lost. We also performed MP analyses using the bacterial sequences as outgroup (not shown), but the strict consensus MP tree for these sequences does not resolve any basal branching other than that of the bacterial sequences. In the rest of the analyses, we have included only the eukaryotic MADS-domain sequences.

An alternative way to root the MADS-domain protein tree objectively is to use Page and Charleston's (24) approach to find the root position that minimizes the number of duplications and sorting events in the protein tree, when this is reconciled to the species tree (see *Materials and Methods*). We show the rooted NJ tree that minimized the reconciliation cost (49 total or 3 basal duplications and 17 sorting events) as the polarized phylogenetic hypothesis for this gene family. The bootstrap NJ tree reveals two well supported (>50%) clades. The first one is constituted by the TypeI MADS-domain sequences and groups the animal SRF-like genes with two newly identified plant lineages, AGL34- and AGL23-like, plus AGL30, AGL33, and AGL39. The second, TypeII MADS-domain sequences, includes the rest of the plant sequences and the animal MEF2-like sequences.

Using MP analyses, we obtained a total of 647 most parsimonious trees (consistency index = 0.544, retention index = 0.695, rescaled consistency index = 0.378) of a length of 700 steps. The strict consensus-rooted MP tree resolves the monophyletic clade that includes animal SRF-like and plant AGL34-like sequences plus AGL30, AGL33, and AGL39, but with a low bootstrap support (<50%). In contrast to the NJ tree, the strict consensus MP tree identifies the AGL-23 plant MADS-domain clade as a sister branch of the animal MEF2-like sequences, but with a very low bootstrap support (<20%). The MP tree also resolves the AGL25 clade as sister to the monophyletic group formed by the rest of the plant TypeII and the animal MEF2-like sequences, also with a very low bootstrap support (<20%). MP groups the animal and fungal MEF2-like sequences with the plant MADS-domain sequences in a monophyletic clade and places the animal and fungal SRF-like sequences as sister group with a good bootstrap support (>50%).

When reconciled to the species tree, the least costly MP gene tree still requires a greater number of basal gene duplications and losses (49 total or 8 basal duplications and 22 sorting events) than the NJ tree shown (Fig. 3*b*). This MP tree also defined TypeI and TypeII groups as sister to each other. These results confirm that the most parsimonious root location among all trees tested is between the TypeI and TypeII lineages that we have identified. We compared the length of the Bootstrap NJ topology with the MP strict consensus tree using MACCLADE (Ver. 3.0) and found that they are of equal length. Therefore, based on the data at hand, we propose the tree shown in Fig. 3*b* as the most parsimonious hypothesis on the polarized evolutionary history of the eukaryotic MADS-box gene family. Finally, the QP tree also resolved the same TypeI and TypeII clades formed by the same family members as in the NJ tree shown (frequency value equal to 40%).

The inconsistent placement of the AGL23 clade between the NJ/QP and MP topologies, as well as the low bootstrap value for the TypeI clade in the MP strict consensus tree, suggests that some plant sequences cannot be unambiguously associated to either the TypeI or TypeII lineages. In fact, if AGL30, AGL33, AGL39, and the AGL23-like genes are removed, NJ, MP, and QP analyses yield resolved and well supported trees (bootstrap values of >90% and 50% for both lineages in NJ and MP analyses, respectively; see Fig. 3*b*; and 89% frequency in QP). These problematic sequences could be the result of recombina-

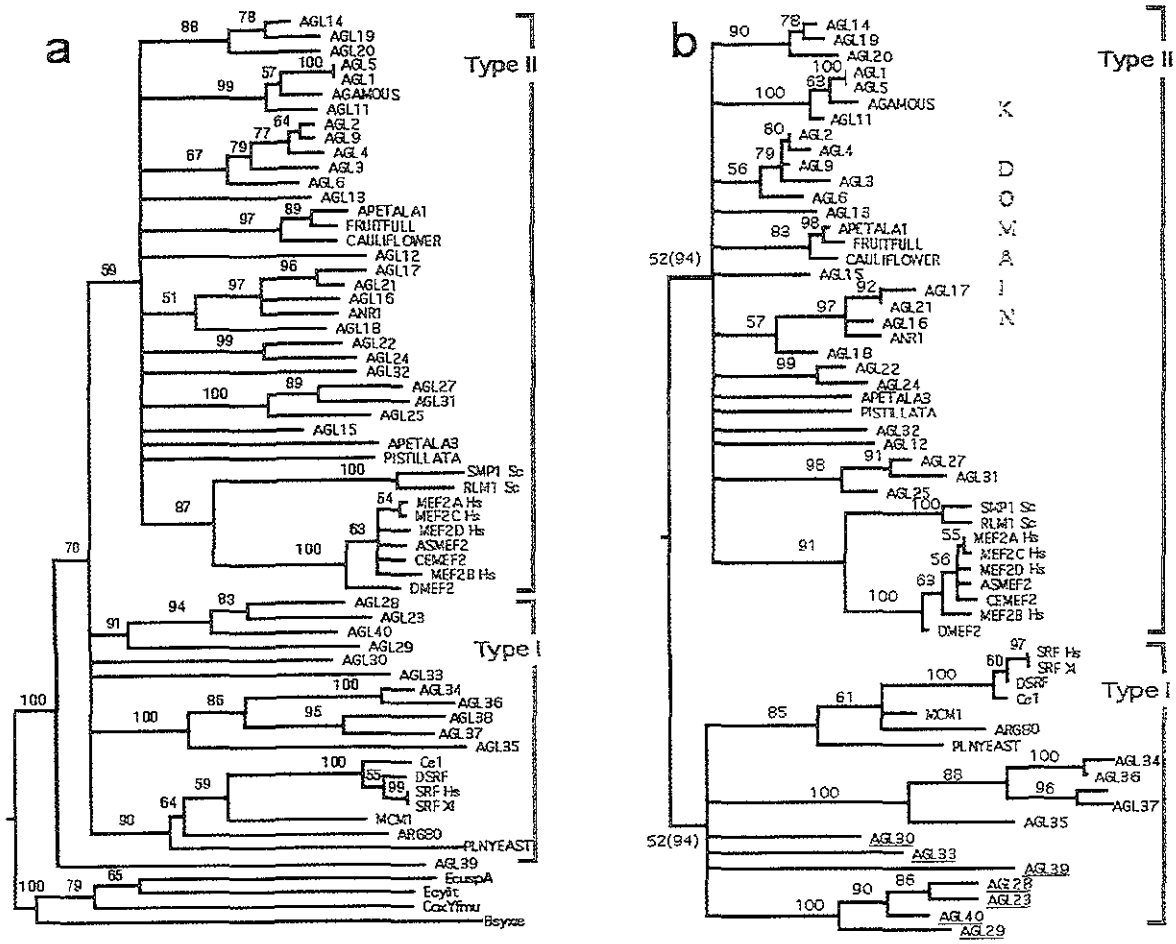

Fig. 3. Phylogeny of the eukaryotic MADS-box gene family. Animal and fungal sequences (*H. sapiens*: MEF2A_Hs, MEF2C_Hs, MEF2D_Hs, MEF2B_Hs, SRF_Hs; *X. laevis*. SRF_Xl; *C. elegans* CEMEF2, *H. roretzi*. ASMEF2; *D melanogaster*: DMEF2, DSRF; *S. cerevisiae:* SMP1_Sc, RLM1_Sc, MCM1, ARG80; *S. pombe*: PLNYEAST) are red, plant sequences (all from *A. thaliana*) are green, bacterial USP family sequences (*E. coli*: EcuspA, Ecyiit; *Coxiella burnetti*: CoxYfmu, *B. subtilis*: Bsyxie) (23) are blue. TypeI (SRF-like) and TypeII (MEF2-like) lineages are indicated by blue and pink brackets, respectively (*a*) The NJ tree rooted with the bacterial USP family (see ref. 23) is shown in *a*, and the NJ tree rooted by minimizing the reconciliation cost (see *Materials and Methods*) is shown in *b*. Branch lengths are proportional to the number of amino acid substitutions Bootstrap values shown on branches, in *b*, values in parentheses correspond to analyses done without the underlined sequences. Branches with bootstrap values <50% are collapsed. Sequences within purple square are those for which a coiled-coil structure downstream of the MADS-domain (K domain) was predicted.

tion between TypeI and TypeII sequences. This possibility is suggested because they share some of the synapomorphies that define each of the two lineages (see Fig. 2). The fact that these sequences group in a clearly monophyletic clade suggests an ancient recombination event that would have been followed by several duplications. To unambiguously resolve the origin and phylogenetic position of these genes, more information is required.

In an effort to explore further the monophyly of the TypeI groups that we propose, we did MP and NJ phylogenetic analyses of this clade by using only one sequence of the MEF-2 sequences as outgroup (not shown). In these analyses, the plant AGL34-like, plus AGL30 and AGL33, plus the animal SRF-like sequences, form a well supported (bootstrap = 63%) monophyletic group, and AGL23-like and AGL39 sequences group in a clade sister to that formed by the former sequences. Both of these clades form a monophyletic lineage with 76% of bootstrap support.

The results presented here imply that features shared by proteins within the MEF2-like and SRF-like clades were present in the ancestral eukaryotes and have remained practically unchanged during the evolution of animal, fungal and plant lineages. The TypeII MADS-domain sequences share some conserved amino acids that are found in none of the TypeI MADS domains (synapomorphies; see Fig. 2). In contrast, the TypeI MADS have only one synapomorphy that defines this clade and some that are shared by all but one or a few sequences. This suggests that there has been a stronger functional constraint within the TypeII than the TypeI MADS-domain lineages. TypeI MADS domains are conserved within animals and within plants, but they differ between these two species' lineages. MADS domains from yeast from both TypeI and TypeII lineages are the most divergent ones.

It will be interesting to determine whether the plant TypeI MADS-box sequences represent expressed genes or are instead pseudogenes. But the fact that at least one of these sequences, AGL39, is represented as an EST clone (GenBank accession no. C99890), as well as the high conservation among AGL34-like sequences, suggests that these members are indeed expressed. Future studies should be devoted to characterizing functionally these genes in *Arabidopsis*.

The conserved MADS-domain motifs within each lineage may serve as the basis of the common functional properties of all proteins within the TypeI and TypeII clades Indeed, *in vitro*

EVOLUTION

DNA-binding assays revealed that chimeric proteins with either the SRF or MEF2A amino-terminal region of the MADS domain and the rest of the AP1, AP3, PI, and AG plant proteins, acquired the respective and distinct DNA-binding specificity of SRF or MEF2A. However, *in vivo* assays did not distinguish between chimeric and full-length wild-type proteins' functions. Both results put together suggest that DNA-binding specificity, which must underlie functional specificity of MADS-domain proteins, is determined not solely by sequences within the MADS-domain but also by sequences within other domains that may affect dimerization with protein partners (30).

Additional *in vivo* experiments show that although chimeric genes with the amino terminus of either the SRF or MEF2A MADS-domain and the rest of AP1 may rescue *ap1–1* mutant plants when expressed under the wild-type *AP1* promoter, the chimera with the MEF2A MADS-domain amino terminus (i.e., within-lineage chimera) rescued mutant phenotypes more effectively than those harboring crosslineage constructs (i.e., from SRF; ref. 31). Our phylogenetic results support, as suggested by these functional analyses, that differences between TypeII and TypeI MADS domains have a role in defining function. Indeed, ectopic expression experiments of chimeric proteins suggest that the MADS and I domains define functional specificities of APETALA1 and AGAMOUS (32, 33), both TypeII (MEF2-like) plant members. However, the conservation of MADS-domain sequences within each lineage and additional functional studies (see below) also suggest that domains outside the MADS domain are important for functional specificity. The K domain, typical of previously characterized plant TypeII proteins, is one such domain.

Evolution of the Plant K Domain. The K domain is an ≈70-aa domain located downstream of the DNA-binding MADS domain, typically spanning positions 110 to 180 of plant MADS proteins. It has a regular spacing of hydrophobic amino acids, and it is assumed to adopt a coiled-coil structure (see Fig. 1). This structural motif has been described for the great majority of previously identified plant MADS-domain proteins (4). To investigate the origin and evolution of the K domain, we used protein-structure programs to predict whether the AGL34 and AGL23 clade members, as well as the other plant and animal MADS-domain sequences analyzed, contain a K domain. In Fig. 3*b*, we boxed the sequences with a predicted coiled-coil structure downstream of the MADS domain.

Coiled-coil structures were not predicted for any of the animal sequences, any of the plant AGL34 or AGL23-like, or for AGL30, AGL33, and AGL39. These sequences also lack any significant sequence similarity to other plant MADS-domain sequences outside of the MADS domain. Interestingly, whereas protein–structure prediction programs clearly identify a coiled-coil domain for most plant members of the TypeII lineage (MEF2-like), they fail to predict such a structure for a few members of this group (the AGL25-like and AGL12) that seem to lack some of the conserved hydrophobic amino acids. This result suggests that the absent amino acids might be critical for the formation of the coiled-coil structure. Both methods used here have been reported to identify positively all of the sequences that form coiled coils in Protein Data Bank structures containing this type of helical structure (27). Thus, the coiled-coil predictions presented in this work have a high level of reliability (>95%), well above standard secondary structure prediction methods.

Animal SRF- and MEF2-like proteins contain additional conserved regions, referred to as SAM and MEF2 domains (2). These and the K domain could be the regions involved in the functional divergence among members of each MADS-domain lineage. Ectopic expression experiments of chimeric proteins suggest that functional specificities of APETALA3 and PISTIL-

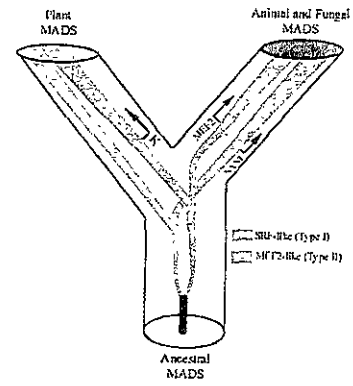

Fig. 4. Model for the evolution of the MADS-box gene family in eukaryotes. At least one duplication of the ancestral MADS-box gene is postulated to have occurred before the divergence of plants and animals. The K domain was probably added to the plant TypeII (MEF2-like) lineage. Similarly, animal MADS-domain proteins evolved specific domains (SAM and MEF2) in SRF-like and MEF2-like lineages, respectively. Pink, TypeI (SRF-like) lineage; blue, TypeII (MEF2-like) lineage.

LATA MADS-domain proteins in organ determination rely on the I and K domains of these genes (31, 32). Recent experiments for two plant MADS-domain proteins (*APETALA1* and *CAULIFLOWER*) suggest that differences between the K domains of these two recently duplicated genes explain at least part of the functional differences between these paralogous loci (E.R.A.-B. and M.F.Y., unpublished results).

Evolution of MADS-Domain Proteins in Eukaryotes: A Synthesis. The results described here suggest a hypothetical scenario for the evolution of the MADS-box gene family in eukaryotes (Fig. 4). From our analyses, it appears that at least one ancestral MADS-box gene duplicated in the common ancestor of the major eukaryotic kingdoms more than a billion years ago to give rise to the distinct TypeI (SRF-like) and TypeII (MEF2-like) lineages found in plants, fungi, and animals today. In yeast and *C. elegans* genomes, MADS-box sequences of both TypeI and TypeII have been found (several of each in yeast and one of each in *C. elegans*). These results support our proposition that eukaryotic MADS-box sequences can be assigned to either of two main lineages that are both present at least in fungi and animals. The *Arabidopsis* genome will be sequenced to completion soon, and we will then be able to test unambiguously the presence of these and additional lineages in plants. Phylogenetic analyses that include MADS domains from basal eukaryotes and TypeI sequences from other plants will help confirm the uniqueness of the ancestral duplication and the monophyly of the TypeI clade.

The evolution of additional domains beyond the MADS domain could have occurred independently along the animal and plant lineages after their divergence from each other, as suggested in our model (Fig. 4), or these could have been present in the ancestral MADS-box genes and then lost along different lineages. In plants, the K domain evolved within the TypeII (MEF2-like) lineage but not the TypeI (SRF-like) lineage. Because most of the TypeII class of plant MADS-box genes are predicted to encode a K domain, this plant-specific domain probably evolved before the extensive duplications that generated this particular lineage. Interestingly, some of the recently cloned MADS-box genes from ferns (33) are predicted to contain K domains (data not shown), indicating that this domain was present at least 395 million years ago in the common ancestors of ferns and seed plants.

We can use parsimony to argue that the K domain originated after the duplication that led to the MEF2- and SRF-like animal

MADS-box genes. However, based on the phylogeny of Fig. 3*b*, we cannot distinguish whether it evolved along the plant lineage after it diverged from the animal one, or whether it was present in the ancestral TypeII-like gene and then lost in animal and some plant lineages. A recent phylogenetic analysis of the M, I, and K domains of all plant protein sequences, (E.R.A.-B., S.L., S.P., S.G., C.B., G.D., and M.Y., unpublished work) suggests that AGL12 and the AGL25-like sequences are basal to the rest of the *Arabidopsis* TypeII AGLs. This result supports the hypothesis that the K domain evolved along the plant lineage after it diverged from animals and fungi (Fig. 4). Identification of MADS-box genes within the most basal extant green plant lineages (including green algae and the bryophytes) and in one of the extant common ancestors of plants and animals (e.g., *Euglena*) should provide experimental tests for the hypotheses postulated in this model of MADS-box gene family evolution. Animal SRF- and MEF2-like domains (see Figs. 1 and 4) may have evolved within animal lineages (as suggested in Fig. 4), or they could have been present also before the divergence of plants and animals and subsequently lost and replaced in plants.

MADS-box genes probably played key roles in the early evolution of flowering plants and in plant evolution in general, perhaps analogous to the roles played by homeobox genes in the evolution of animal form (34, 35). This scenario is suggested by the fact that MADS-box gene mutations, as those of homeobox genes in animals, also produce homeotic conversions in flowers, suggesting that they occupy similar places in the regulatory networks that control development (36). Like homeobox genes, MADS-box genes are also highly conserved among distantly related plants, and orthologous genes form monophyletic clades (6–9). To test the long-suspected parallel between the molecular evolution of the MADS-box gene family and the evolution of plant form, a polarized gene phylogeny is necessary. We have proposed here a hypothesis for the evolutionary history of the MADS-domain protein family, including the nearly complete *Arabidopsis* MADS-box sequence complement, which suggests that eukaryotic MADS-box sequences can be assigned to two main lineages and locates the root of the whole family between them. These analyses may be used to guide the search for MADS-box sequences in basal eukaryotes and the assignment of newly cloned genes from other plant species to one of the clades proposed in this study. Further phylogenetic and population genetic studies (e.g., ref. 37) as well as functional analyses of the MADS-box family and other important transcriptional regulators should lead to a better understanding of the molecular evolution of developmental mechanisms. These mechanisms underlie the morphological evolution of plants and animals, the understanding of which is still elusive to evolutionary biologists

We thank C. Ferrándiz, W. Crosby, C. Gustafson-Brown, and S. Rounsley for providing unpublished sequences. Special thanks to A. Chaos and F. Vergara for illuminating phylogenetic conversations and for help during figure preparation. Two anonymous reviewers put important effort into helping us improve this paper. Many thanks to A. Cortés for help in various tasks and to R. Salas for help in Fig. 4. C. Ferrándiz, S. Kempin, M. Ng, and A. Sessions provided useful suggestions. This work was supported by a grant from the National Science Foundation to M.F.Y. and a CONACYT (Consejo Nacional de Ciencia y Tecnología, México) grant to E.R.A.-B. Also, E.R.A.-B. was a Pew Foundation Fellow during the completion of this work. S.P. had a long-term postdoctoral fellowship from the Human Frontiers Science Program Organization, S.L. had a Lucille P. Markey predoctoral fellowship, and C.B. and L.M.C. had a Ph.D. scholarship from CONACYT and Universidad Nacional Autónoma de México.

1 Doebley, J & Lukens, L. (1998) *Plant Cell* **10**, 1075–1082
2 Shore, P. & Sharrocks, A. D. (1995) *Eur. J. Biochem.* **229**, 1–13.
3. Coen, E. S & Meyerowitz, E. M. (1991) *Nature (London)* **353**, 31–37.
4 Riechmann, J. L. & Meyerowitz, E. M. (1997) *J. Biol Chem* **378**, 1079–1101.
5 Liljegren, S. J., Ferrándiz, C., Alvarez-Buylla, E R., Pelaz, S. & Yanofsky, M. F (1998) *Flowering Newsletter* **25**, 9–19
6. Theissen, G., Kim, J T. & Saedler, H. (1996) *J. Mol. Evol.* **43**, 484–516.
7. Doyle, J. J (1994) *Syst. Biol.* **43**, 307–328.
8. Purugganan, M. D., Rounsley, S. D., Schmidt, R. J. & Yanofsky, M. F. (1995) *Genetics* **140**, 345–356.
9. Purugganan, M D. (1997) *J Mol Evol.* **45**, 392–396.
10. Yanofsky, M. F., Ma, H., Bowman, J. L., Drews, G. N , Feldmann, K A. & Meyerowitz, E M. (1990) *Nature (London)* **346**, 35–39
11. Jack, T., Brockman, L. L. & Meyerowitz, E. M. (1992) *Cell* **68**, 683–697.
12. Goto, K. & Meyerowitz, E M. (1994) *Genes Dev.* **8**, 1548–1560.
13 Ma, H., Yanofsky, M. F. & Meyerowitz, E. M. (1991) *Genes Dev* **5**, 484–495
14. Mandel, M. A., Gustafson-Brown, C, Savidge, B. & Yanofsky, M. F. (1992) *Nature (London)* **360**, 273–277.
15. Mandel, M A. & Yanofsky, M F. (1995) *Plant Cell* **7**, 1763–1771
16. Mandel, M A & Yanofsky, M. F. (1998) *Sexual Plant Reprod.* **11**, 22–28.
17. Kempin, S. A., Savidge, B. & Yanofsky, M F. (1995) *Science* **267**, 522–525
18. Rounsley, S D., Ditta, G S & Yanofsky, M. F. (1995) *Plant Cell* **7**, 1259–1269
19 Zhang, H & Forde, B G (1998) *Science* **279**, 407–409.
20 Burke, W D., Eickbush, D. G., Xiong, Y., Jacubczak, J & Eickbush, T H. (1993) *Mol Biol. Evol* **10**, 163–185.
21. Hillis, D M & Bull, J. J. (1993) *Syst Biol* **42**, 182–192.
22. Page, R. D. M. (1996) *CABIOS* **12**, 357–358
23. Mushegian, A. R. & Koonin, E. V. (1996) *Genetics* **144**, 817–828
24. Page, R. D M. & Charleston, M. A. (1997) *Mol Phyl. Evol.* **7**, 231–240.
25. Page, R. D. M. & Holmes, E. C. (1998) *Molecular Evolution. A Phylogenetic Approach* (Blackwell Scientific, Oxford).
26. Baldauf, S. L. & Palmer, J. D. (1993) *Proc Natl Acad. Sci. USA* **90**, 11558–11562.
27. Berger, B., Wilson, D. B., Wolf, E., Tonchev, T., Milla, M. & Kim, P. S (1995) *Proc. Natl Acad. Sci. USA* **92**, 8259–8263
28 Wolf, E., Kim, P. S. & Berger, B. (1997) *Protein Sci* **6**, 1179–1189.
29 Stultz, C. M., White, J. V & Smith, T. F. (1993) *Protein Sci.* **2**, 305–314
30 Riechmann, J. L. & Meyerowitz, E. M (1997) *Mol Biol Cell* **8**, 1243–1259
31. Krizek, B. A., Riechmann, J. L. & Meyerowitz, E. M (1999) *Sex Plant Reprod.* **12**, 14–26.
32. Krizek, B. A & Meyerowitz, E M (1996) *Proc. Natl. Acad Sci. USA* **93**, 4063–4070
33. Munster, T., Pahnke, J., DiRosa, A., Kim, J., Martin, W, Saedler, H. & Theissen, G. (1997) *Proc. Natl Acad, Sci USA* **94**, 2415–2420.
34 Raff, R. A. (1996) *The Shape of Life: Genes, Development, and the Evolution of Animal Form* (Univ. of Chicago Press, Chicago).
35 Carroll, S. B. (1995) *Nature (London)* **376**, 479–485.
36 Mendoza, L & Alvarez-Buylla, E. R (1998) *J Theor Biol.* **193**, 307–319.
37 Purugganan, M. D. & Suddith, J. I. (1998) *Proc Natl Acad Sci USA* **95**, 8130–8134.

The Plant Journal (2000) **24**(4), 457–466

MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells, roots and trichomes

Elena R. Alvarez-Buylla[1,2], Sarah J. Liljegren[1], Soraya Pelaz[1], Scott E. Gold[1,†], Caroline Burgeff[2], Gary S. Ditta[1], Francisco Vergara-Silva[2] and Martin F. Yanofsky[1,*]

[1] *Department of Biology, UCSD, La Jolla, CA, 92093-0116, USA, and*

[2] *Instituto de Ecología, UNAM, AP-Postal 70-275, México D.F. 04510, México*

Received 12 July 2000; revised 4 September 2000; accepted 5 September 2000

*For correspondence (e-mail marty@ucsd.edu; abuylla@servidor.unam.mx).

†Present address: Plant Pathology, University of Georgia, Athens, GA 30602-7274, USA

Summary

MADS-box genes encode transcriptional regulators involved in diverse aspects of plant development. Here we describe the cloning and mRNA spatio-temporal expression patterns of five new MADS-box genes from *Arabidopsis*: *AGL16, AGL18, AGL19, AGL27 and AGL31*. These genes will probably become important molecular tools for both evolutionary and functional analyses of vegetative structures. We mapped our data and previous expression patterns onto a new MADS-box phylogeny. These analyses suggest that the evolution of the MADS-box family has involved a rapid and simultaneous functional diversification in vegetative as well as reproductive structures. The hypothetical ancestral genes had broader expression patterns than more derived ones, which have been co-opted for putative specialized functions as suggested by their expression patterns. *AGL27* and *AGL31*, which are closely related to the recently described flowering-time gene *FLC* (previously *AGL25*), are expressed in most plant tissues. *AGL19* is specifically expressed in the outer layers of the root meristem (lateral root cap and epidermis) and in the central cylinder cells of mature roots. *AGL18*, which is most similar in sequence to the embryo-expressed *AGL15* gene, is expressed in the endosperm and in developing male and female gametophytes, suggesting a role for *AGL18* that is distinct from previously characterized MADS-box genes. Finally, *AGL16* RNA accumulates in leaf guard cells and trichomes. Our new phylogeny reveals seven new monophyletic clades of MADS-box sequences not specific to flowers, suggesting that complex regulatory networks involving several MADS-box genes, similar to those that control flower development, underlie development of vegetative structures.

Keywords: MADS-box, endosperm, guard cells, root, trichome, *Arabidopsis*.

Introduction

Transcriptional regulators play important roles in developmental pathways (Schwechheimer and Bevan, 1998), and changes in them are likely to be key molecular determinants of the morphological evolution of plants and animals (Doebley and Lukens, 1998). Phylogenies and ancestral character reconstructions of developmental regulators, such as the one presented here, provide the historical framework for studies of the evolution of developmental genetic pathways and give useful clues about the molecular basis of morphological evolution, thus linking the fields of development and evolution (Purugganan, 1998). To this end, complete phylogenies of these regulatory multigene families for model systems are fundamental for comparative analyses and interpretations of sequences from other species.

The MADS-box gene family encodes transcription factors involved in diverse biological functions in eukaryotes (Riechmann and Meyerowitz, 1997; Shore and Sharrocks, 1995). In plants these genes play central roles in flower and fruit development (Bowman *et al.*, 1999; Weigel, 1995). Other MADS-box genes are expressed in vegetative tissues, ovules and embryos, suggesting that this family of genes plays diverse roles throughout plant development (Colombo *et al.*, 1995; Ma *et al.*, 1991; Rounsley *et al.*, 1995; Zhang and Forde, 1998).

We report the isolation, chromosome mapping and mRNA *in situ* expression profiles of five new MADS-box genes from *Arabidopsis thaliana*. Our data suggest that these genes play novel regulatory roles in guard cells, trichomes, roots, pollen and endosperm, and thus may provide useful markers for further functional and evolutionary analyses. Expression data mapped onto a new MADS-box phylogeny reveal new clades that group genes with similar expression patterns, mostly in vegetative structures, and that seem to have evolved at about the same time that the major floral clades evolved (Purugganan, 1997). Mapping analyses suggest that the ancestral pattern of expression was a generalized one, and that duplications gave rise to genes with restricted spatio-temporal expression patterns probably recruited to specialized functions in either reproductive or vegetative structures. We also identify 10 new groups of potentially redundant genes that are predicted to share more than 70% amino acid identity. These results will aid future functional analyses because several closely related MADS-box genes with overlapping expression patterns encode redundant functions (Ferrándiz *et al.*, 2000; Kempin *et al.*, 1995; Liljegren *et al.*, 2000; Pelaz *et al.*, 2000). Because previous studies have found a close correlation between expression patterns of MADS-box genes and assigned functions (Gu *et al.*, 1998; Liljegren *et al.*, 2000), this study is the first step toward functional characterization of several MADS-box genes involved in vegetative development, and provides a guide to characterizing the corresponding mutant alleles.

Results and Discussion

Five new Arabidopsis *MADS-box genes*

Phylogenetic studies are most informative when all or most members of a multigene family are known. A PCR-based approach led to the identification of three new MADS-box genes: *AGL16*, *AGL18* and *AGL19*. In addition, through a yeast two-hybrid we found *AGL27* that is closely related to *AGL31*, a gene recently identified through *Arabidopsis* genome sequencing, and to *FLC* (Michaels and Amasino, 1999; Sheldon *et al.*, 1999).

AGL16, *AGL18* and *AGL19* are similar to previously characterized plant MADS-box genes. Each encodes deduced proteins of 240, 256 and 219 amino acids, respectively, with a highly conserved MADS-box at the amino-terminus, a clearly identifiable K-box region, and a divergent C-terminal region. In contrast, cDNAs for *AGL27* and *AGL31* encode deduced proteins of 192 and 178 amino acids, respectively. Two transcripts, resulting from differential splicing, were detected via RT–PCR for *AGL27* (I and II). The more abundant transcript would encode a deduced protein of 192 amino acids, and the less abundant transcript would encode a protein of 173 amino acids. *AGL27* and *AGL31* each have well defined MADS-box regions, and although their K regions are very similar to each other, they are quite different from the K regions previously described for other MIKC type MADS-box genes (Alvarez-Buylla *et al.*, 2000). As in other plant MADS-box genes, the C-terminal regions of *AGL27* and *AGL31* are also more divergent, both between themselves and with respect to other MADS-box genes.

Chromosome mapping

To discover whether any of the new MADS-box genes could be linked to known mutants, we determined the map position of each gene, shown in Figure 1 along with positions of all MADS-box sequences found in *Arabidopsis* up to now. *AGL16* maps to chromosome III between the RFLP markers 249 and g4014. *AGL18* also maps to chromosome III approximately 3.9 cM below the g2778 marker. *AGL19* maps to chromosome IV between markers g10086 and g4564a; and *AGL 27* maps to the bottom of chromosome I, approximately 2.8 cM from marker m532 and 8.6 cM from marker g17311.

Novel expression patterns of MADS-box genes

A preliminary identification of organs in which these newly isolated genes are expressed was undertaken using RNA gel-blot analyses. In addition, gene-specific probes were used in RNA *in situ* hybridization analyses to define the temporal and spatial patterns of expression of these genes.

AGL16 is expressed in trichomes and guard cells

AGL16 RNA accumulates at high levels in leaves, moderate levels in roots and stems, and at low levels in flowers and

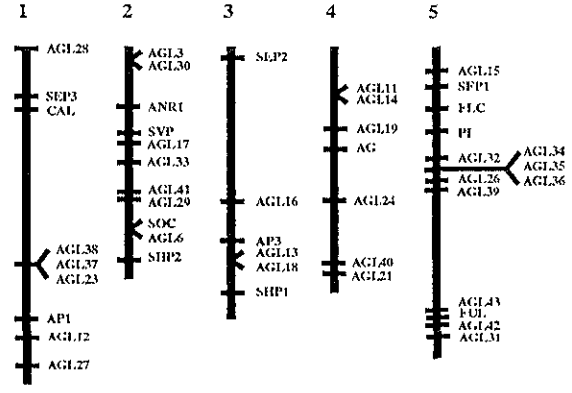

Figure 1 Map positions of *Arabidopsis* MADS-box genes.

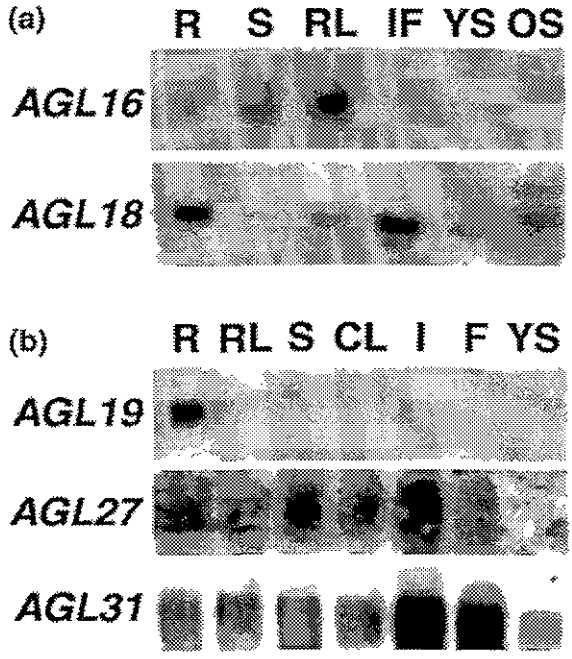

Figure 2. Expression of *AGL16, AGL18, AGL19, AGL27 and AGL31* transcripts in wild-type plants.
Total RNA (a) or poly(A) RNA (b) was used from roots (R), stems (S), rosette leaves (RL), cauline leaves (CL), inflorescences (I), mature flowers (F), inflorescences and mature flowers (IF), young siliques (YS), and older siliques (OS).

young siliques (Figure 2). Within rosette leaves *AGL16* is expressed in mature guard cells and trichomes found in both the abaxial and adaxial epidermis (Figure 3). Additional data suggest that *AGL16* is also expressed in epidermal cells of roots (C. Burgeff, S. Liljegren, M. Yanofsky and E.R. Alvarez-Buylla, unpublished results). Recent studies have begun to uncover the molecular mechanisms of cell-type specification in leaves and roots (reviewed by Larkin *et al.*, 1997; Schiefelbein *et al.*, 1997), including a number of genes that have been found to be expressed in both leaf and root epidermal cells. Furthermore, it has been shown that *GLABRA2* (*GL2*) is part of regulatory networks regulating both leaf and root epidermis development (Schiefelbein *et al.*, 1997). Further functional analyses should be performed to test if *AGL16 and* other MADS-box genes with similar sequences and expression patterns are also part of such conserved networks (see Figure 6).

AGL16 expression was also observed in guard cells of the hypocotyl, but was not detected in guard cells of the inflorescence stem, flower pedicel or sepals (data not shown). The evolution of stomata was a key event during the early evolution of land plants, yet very little is known about the regulatory pathways underlying the development of this differentiated epidermal cell type (Larkin *et al.*, 1997). Other genes have been shown to control stomatal patterning on both the hypocotyl and root epidermis (Berger *et al.*, 1998; Lee and Schiefelbein, 1999), and it will be interesting to investigate the possible regulatory role of *AGL16* in controlling these patterning processes.

AGL18 *is expressed in pollen and endosperm*

AGL18 is expressed at highest levels in roots, flowers and siliques, and significant expression was observed in stems and leaves (Figure 2). In flowers, no expression was detected in sepals or petals. Within stamens, expression was first detected during stage 9 and was localized to the sporogenous tissue of anthers, while it was absent from filaments or anther walls. *AGL18* RNA was particularly apparent in the microspores before they separate from each other (stage 9–10, data not shown), but it was still observed at high levels within pollen grains up to stage 13 of flower development when anthers dehisce (Figure 4a). *AGL18* is the first *Arabidopsis* MADS-box gene reported to show high expression levels in pollen. Previously, pollen-specific genes have been assigned either to 'early' genes, probably encoding cytoskeletal and cell wall proteins (McCormick, 1991), or 'late' genes, expressed around the time of microspore mitosis and strongly expressed during pollen maturation. *AGL18* is expressed during both stages. Another MADS-box gene, *DEF125*, was shown to be expressed 'late' in pollen from *Antirrhinum* (Zachgo *et al.*, 1997), but this gene is more similar to the *ANR1 Arabidopsis* root gene than to *AGL18*.

During carpel development, *AGL18* RNA is first detected in developing ovules (stage 10). Expression appears uniform early in ovule development (Figure 4b). After fertilization, *AGL18* RNA was detected only in globular structures or nodules of proliferating free nuclear endosperm that are important for embryo development (Mansfield and Briarty, 1991; Scott *et al.*, 1998). This gene was detected only up to the heart stage of embryo development, when very little nuclear endosperm remains, and it was not detected in developing embryos at any stage (Figure 4c,g; data not shown). In contrast, the closely related *AGL15* MADS-box gene is expressed in embryos, but not in the endosperm (Rounsley *et al.*, 1995).

Despite the key evolutionary and functional relevance of the endosperm, one of the key innovations of angiosperms (Friedman, 1992), little is known about the genetic circuitry controlling its development. *AGL18* represents one of the few reported molecular markers of endosperm in *Arabidopsis* (Drews *et al.*, 1998). Differential expression of *AGL18* within endosperm tissues correlates with the significant structural and developmental differences between the micropylar and the chalazal chambers in

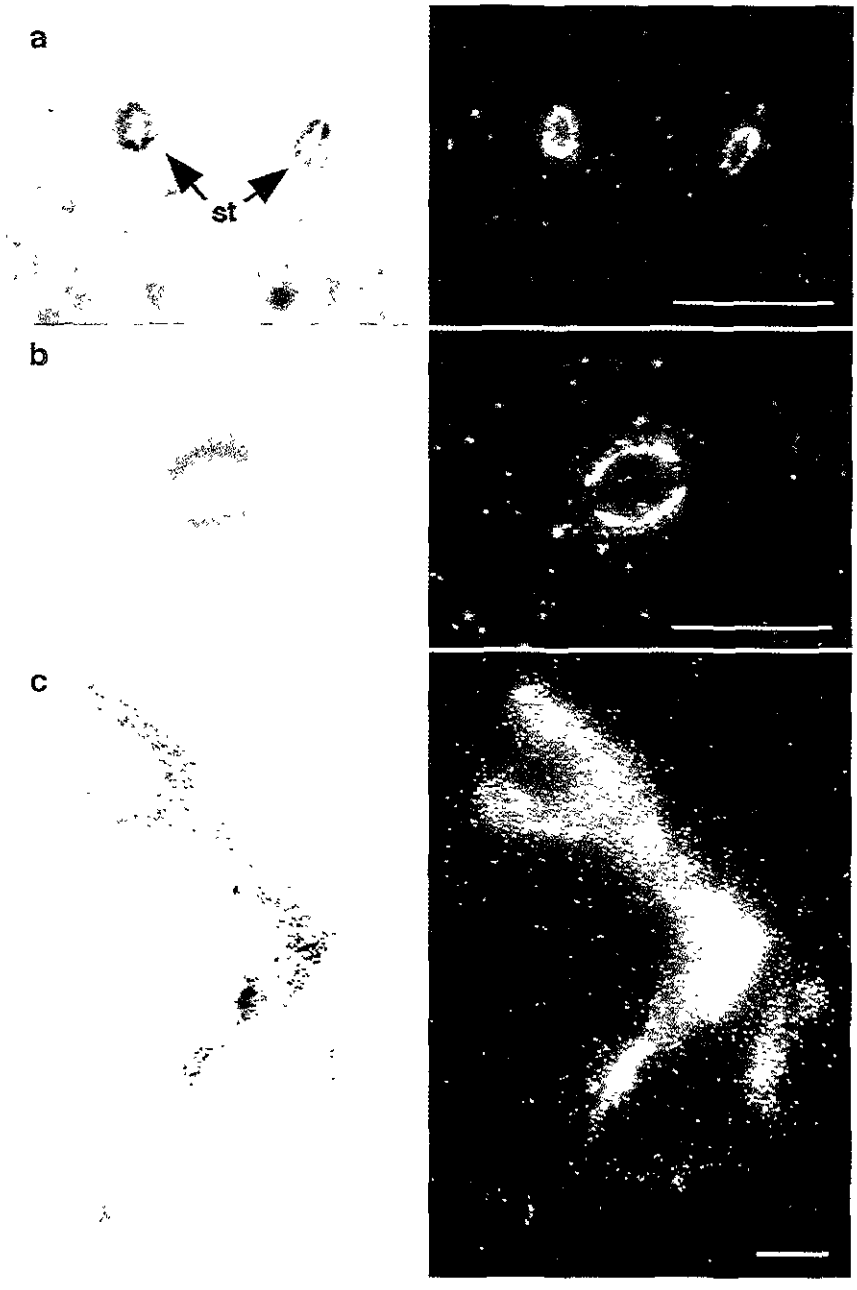

Figure 3. *AGL16* expression in leaf guard cells and trichomes.
Sections of wild-type rosette leaves were probed with *AGL16* antisense RNA, and bright-field (left) and dark-field (right) photographs of the same section are shown in panels (a–c).
(a,b) *AGL16* is expressed in stomata (st) of expanding (a) and more mature (b) leaves, along the pore-facing edge of each guard cell. (c) Stellate leaf trichome showing *AGL16* expression. Scale bars, 25 μm.

Arabidopsis and other Brassicaceae (Mansfield and Briarty, 1991).

AGL19 *is expressed in roots*

AGL19 RNA is specific to roots, and no expression was detected in leaves, stems, flowers, or siliques (Figure 2). This gene is expressed in the columella, lateral root cap and epidermal cells of the meristematic region of the primary and lateral root tips (Figure 5b–d). In the mature differentiated region of the root, *AGL19* RNA was observed in all cell types of the central cylinder from the pericycle to the inner cell types of the vascular bundles, but the endodermis, cortex and epidermis remained unlabelled (Figure 5e).

The root cap plays an important role in gravitropic sensing, and recent ablation experiments show that this tissue is important for the signaling system determining root growth rate and the suppression of lateral root formation (Tsugeki and Fedoroff, 1999). Such signaling

circuits are probably linked to transduction pathways that respond to external stimuli, as is has been proposed for the *ANR1* MADS-box gene (Zhang and Forde, 1998).

AGL27 and AGL31 are similar in sequence and expression patterns to FLC

RNA blot analyses revealed significant levels of *AGL27* and *AGL31* expression in most plant tissues, including roots, leaves, stems and flowers, and low levels of expression were detected in siliques with the *AGL31* probe (Figure 2). *In situ* data using inflorescences (data not shown) revealed that they have very similar overall expression patterns. Their mRNAs were detected in all flower organs and in early floral meristems, with the highest levels of expression observed in flower pedicels. These similar expression patterns suggest that these two closely related genes (Figure 6) probably represent functionally redundant loci.

FLC is very closely related to *AGL27* and *AGL31* (Figure 6), and they all share similar expression patterns. Therefore *FLC* could share some functions with *AGL27* and *AGL31*, but it also appears to have at least some independent roles because the single *flc* loss of function mutants have a clear phenotype of early flowering. This phenotype suggests that this gene is a repressor of flowering. Despite the generalized presence of *FLC* transcripts in all plant organs, the mutants only show alterations in flowering time (Michaels and Amasino, 1999; Sheldon *et al.*, 1999). The overlapping expression patterns of these two genes and *FLC*, and the close sequence similarity of the three genes, suggest that *AGL27* and *AGL31* may also control the transition to flowering. These two genes could also share additional redundant functions with *FLC* that might be revealed in double or triple mutants.

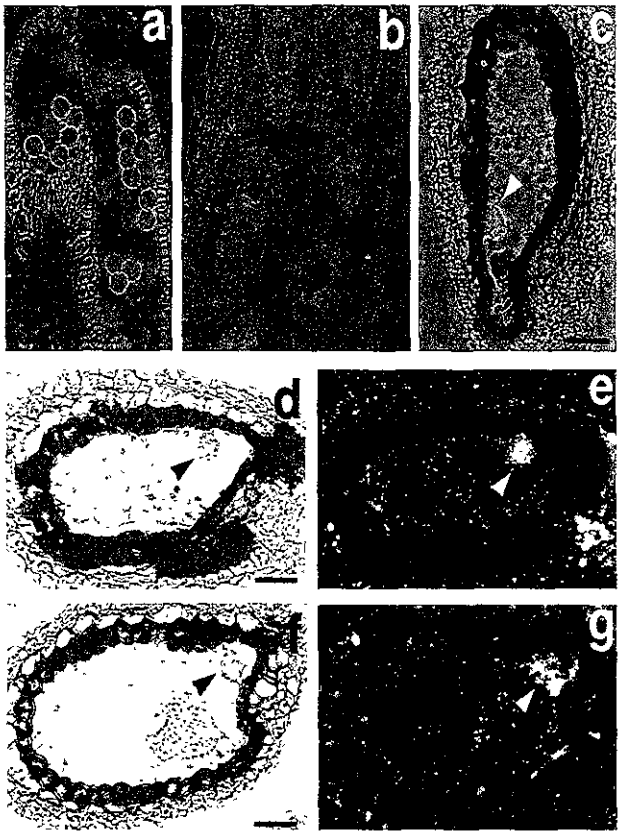

Figure 4. *AGL18* expression in wild-type flowers, ovules and developing seeds.
(a–c) Bright/dark field double exposures of *in situ* hybridizations probed with *AGL18* antisense RNA, signal in red. (a) Longitudinal section of an anther at stage 13 of flower development. (b) Longitudinal section of a carpel at stage 16 of flower development. (c) Longitudinal section of an embryo sac showing endosperm chalazal nodules. (d–g) Sections of embryo sacs were probed with *AGL18* antisense RNA, and bright-field (left) and dark-field (right) photographs of the same sections. Arrows point at the nodules of free-nuclear stage endosperm with strong signal. Scale bars, 50 μm.

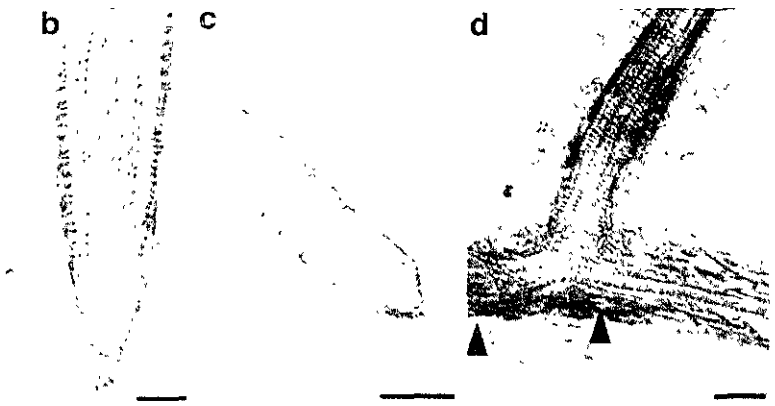

Figure 5 *AGL19* expression in wild-type roots
(a) Longitudinal section of root meristem probed with sense *AGL19* RNA. Longitudinal sections of main (b) and lateral (c) root tips probed with antisense AGL19 RNA. Note that (c) is not a medial section and a few stained epidermal cells are seen near the tip. (d) Mature root probed with antisense *AGL19* RNA showing staining in all cell types of the central cylinder (right arrow: pericycle, phloem and parenchyma cells). Left arrow points at unstained endodermis cell. Scale bars (a,b,d) 50 μm, (c) 100 μm.

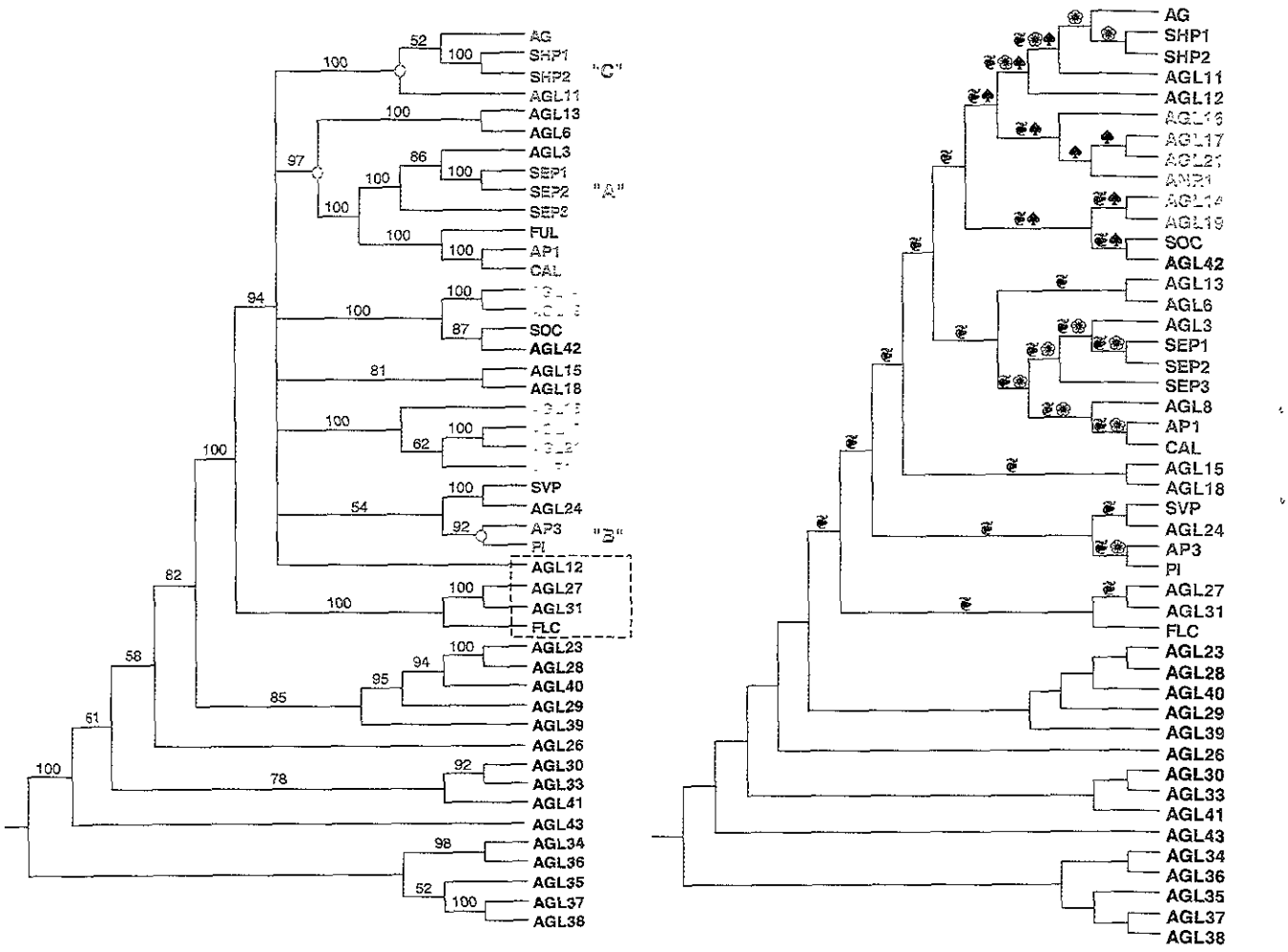

Figure 6. Phylogeny of the *Arabidopsis* MADS-box gene family.
(a) Strict consensus maximum parsimony phylogeny of the elison alignment of the MADS, I and K amino acid sequences, rooted with AGL34 clade as outgroup. Bootstrap support indicated on branches (those with <50% collapsed). Color indicates mRNA expression of genes encoding each protein as reproductive (red), vegetative (green), or both (blue). Proteins boxed in purple are those for which coiled-coils were predicted at K domain (Alvarez-Buylla *et al.*, 2000), those within dashed-line squares have conserved amino acids within the K domain but no predicted coiled-coils. ABC clades are indicated by red dots at their nodes in the tree and labeled on the right. (b) Single most parsimonious tree (CI = 0.556; RI = 0.518; RC = 0.288; length = 34345) with ancestral states of expression reconstructed under parsimony. More than one symbol indicates ambiguous reconstruction.

Phylogeny: evolution of MADS-box gene function in plants

We used MIK amino acid sequences to infer phylogenies of the newly identified MADS-box genes together with 42 previously reported MADS-box sequences (Alvarez-Buylla *et al.*, 2000 and references therein for sequence sources). The only known MADS-box sequence not included in our phylogeny was *AGL32*, as we were not able to predict its amino acid sequence reliably. We present a tree (Figure 6a) rooted with the *AGL34*-like sequences that are clear members of the Type I MADS-box lineage (Alvarez-Buylla *et al.*, 2000). It is likely that most of the different clades that define the MADS-box gene family in *Arabidopsis* are represented in this phylogeny and that the clades that include the well-characterized ABC genes are complete, while the newly uncovered clades may still lack additional members.

Sequences for which a predicted coiled-coil protein domain (K-domain) downstream of the MADS-domain was found are within the purple box outside the dash-line square (Alvarez-Buylla *et al.*, 2000). Within the dashed-square are sequences with more divergent K domains in which some of the conserved amino acids are missing, and for which the programs do not predict coiled-coils.

These and previous analyses (Alvarez-Buylla *et al.*, 2000), which showed that predicted coiled-coils are found only in plant sequences, suggest that the K domain evolved along the plant lineage after its divergence from the animal lineage.

Our analyses (Figure 6a) confirm that genes within each monophyletic clade share similar expression patterns (Doyle, 1994; Purugganan *et al.*, 1995; Theissen *et al.*, 1996). The previously identified floral-specific clades are recovered and indicated by the letters ABC (originally typified by *AP1*, *AP3/PI* and *AG*, respectively). The ABC clades are largely flower- and fruit-specific, and none appears to be expressed in roots. However, some genes in the A clade are also expressed in leaves and stems (Purugganan, 1998; Rounsley *et al.*, 1995). Within this clade it is noteworthy that the recently characterized *SEPALLATA* (*SEP*) genes are required for the canonical B and C functions (Pelaz *et al.*, 2000). As mentioned above, B-function genes, *APETALA3* (*AP3*) and *PISTILLATA* (*PI*), group with *SHORT VEGETATIVE PHASE* (*SVP*, previously *AGL22*; Hartmann *et al.*, 2000) and *AGL24*. But *SVP* and *AGL24* are also expressed in inflorescence meristems, stems and leaves (C. Gustafson-Brown, C. Ferrándiz, and M. Yanofsky, unpublished results; Hartmann *et al.*, 2000).

None of the seven newly identified clades is specific to flowers or fruits (Figure 6a). The *ANR1* clade is largely root-specific, although *AGL16* is also expressed in leaves. The *SOC1* clade is also largely root-specific, with *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*, previously *AGL20*) also expressed in leaves and flowers (Samach *et al.*, 2000), while its sister gene, *AGL42*, has not yet been studied. *AGL15* and *AGL18* group together in another clade, and they are expressed in the embryo and nuclear endosperm, respectively, as well as in other plant tissues throughout the complete life cycle of *Arabidopsis*. Within the polytomy that groups all the clades described up to now, *AGL12* is the only single-gene branch. *AGL12* is also expressed at high levels in roots, and is found in a few cell types in flowers and shoots (C. Burgeff and E.R. Alvarez-Buylla, unpublished results). *FLC*, *AGL27* and *AGL31* form a sister monophyletic clade to all the groups described above (Figure 6a). All three genes have a very similar and generalized expression pattern. The newly discovered clades of MADS-box genes, whose expression is restricted to vegetative structures, suggests that complex genetic circuits also underlie development of these structures (Scheres *et al.*, 1995).

Outside the clades described above, several monophyletic groups are well resolved (Figure 6a). None of the sequences in the latter shows clear similarity beyond the MADS-domain, and cDNA clones have not been identified for any of these sequences, although an *AGL39* EST is available (Alvarez-Buylla *et al.*, 2000). Thus the expression patterns and functions of all of these sequences have yet to be determined.

Purugganan (1998) has interpreted previous similar expression and phylogenetic analyses as an indication that vegetative MADS-box functions evolved before reproductive ones. He argues that rapid early duplications and functional diversification of these genes could have been caused by selective pressures on ancestral land plant lineages to evolve more complex reproductive structures. The mapping of the expression patterns onto the gene phylogeny enables us to revisit these arguments on the functional diversification within and among gene clades in the MADS-box family.

Mapping analyses on the bootstrap (not shown) and the most parsimonious tree (Figure 6b) suggest that the ancestral pattern of expression was a generalized one, and that duplications gave rise to genes with restricted patterns of expression probably recruited to specialized functions in either reproductive or vegetative structures. In most cases the ancestral states within clades are ambiguous. However, the putative ancestral gene of the *AG–SHP* clade appears to have had an mRNA expression already restricted to reproductive structures. Similarly, the putative ancestral gene in the *AGL17–AGL21–ANR1* clade appears to have had an expression restricted to vegetative tissue, but the putative ancestor of these two clades had a generalized pattern of expression. Therefore our data suggest that the function of the plant MADS-box genes did not progress from vegetative to reproductive. Instead, it seems that differential gene recruitment was correlated with spatio-temporal restriction of expression pattern in both reproductive and vegetative structures as gene duplication events occurred.

Traditional morphological studies (Foster and Gifford, 1991) and the analyses presented here suggest that specialized vegetative structures have apparently evolved simultaneously to reproductive structures along a species tree. Furthermore, under these assumptions selectionist arguments that would explain the diversification of reproductive structures during early evolution of land plants should also be applied to the diversification of vegetative structures. On the other hand, rapid evolution could explain the lack of resolution in the branching order of the clades within the purple box in Figure 6(a). These are appealing hypotheses that should be tested with estimations of divergence times among gene clades and analyses of other species sequences (Purugganan, 1997, Purugganan, 1998).

Finally, our phylogeny reveals new groups of closely related and possibly functionally redundant MADS-box sequences, with over 70% identity at the MIK amino acid level and supported by bootstrap values >87%: *AGL13/AGL6*, *AGL14/AGL19*, *AGL17/AGL21*, *SOC/AGL42*, *SVP/AGL24*, *AGL27/AGL31/FLC*, *AGL23/AGL28/AGL40/AGL29*,

AGL30/AGL33, AGL34/AGL36 and *AGL37/AGL38*. These data should be useful to guide further functional characterization of these genes, as it is possible that only double, triple (Pelaz *et al.*, 2000) or even quadruple mutants of these closely related sequences may show phenotypes amenable to further analysis.

Given the crucial roles of MADS-box genes in plant development and the rapid pace at which new MADS-box genes from diverse plant species are being cloned, this family is becoming a promising paradigm for unravelling mysteries underlying the molecular basis of morphological evolution in plants (Kramer *et al.*, 1998; Theissen *et al.*, 2000). For example, *AGL15 and AGL18* may become useful molecular markers for studying the molecular evolution of endosperm development, and thus testing the theory that endosperm evolved from a supernumerary embryo that was transformed into nourishing tissue during early angiosperm evolution (Friedman, 1992). These two genes appear to share a more recent common ancestor with each other than with any other family member. Within the embryo sac, *AGL15* expression is restricted to the embryo, while that of *AGL18* is specific to the free nuclear endosperm. Probably both genes were expressed in embryos in the common ancestors of gymnosperms and angiosperms, and during evolution of the latter one of them (*AGL18*) was recruited to control endosperm development. This hypothesis could be tested with expression patterns of genes orthologous to *AGL15 and AGL18* in gymnosperms that constitute the sister group to angiosperms, and still have supernumerary embryos.

Experimental procedures

Cloning of new Arabidopsis *MADS-box genes*

AGL16 was cloned following Rounsley *et al* (1995) and its 5′ coding region was derived from a genomic clone (SL76), which was identified by screening a genomic library (J. Mulligan and R Davis, unpublished results) with a probe synthesized from a 400 bp *Eco*RI *AGL16* cDNA fragment (pSL4). *AGL18* and *AGL19* were cloned with degenerate primers based on the tomato TM8 sequence, 5′-CGGAATTCATGGG(AGCT)(CA)G(AGCT)GG(AGCT)-(AC)G(AGCT)AC-3′ and 5′-CGGGATCC(AGTC)AC(CT)TC(AGTC)-GC(GA)TC(GA)CA(AGTC)A(GA)(TGA)AT-3′. With the sequences of these novel MADS-boxes, nested oligonucleotides were designed to amplify the corresponding cDNAs by reverse transcription PCR with oligo(dT). RNA from root, from leaf and root tissues, and from whole plant tissue was used to clone *AGL19, AGL16* and *AGL18* cDNAs, respectively. Each cDNA was independently amplified, cloned, and sequenced at least twice to check for PCR-induced mutations. TAIL/PCR was used to obtain the 5′ ends (Boehringer Kit, Indianapolis, USA).

AGL27 was isolated as an interactor of AP1. The cDNA expression library was poly(T) primed (S. Pelaz *et al.*, unpublished results) and cloned into pBI771 plasmid (pPC86 plasmid from Chevray and Nathans, 1992 with minor modifications) in *Sal–Not* sites. The interactor clone was not full-length as it began 61 bases after the end of the MADS box of AGL27-I. Based on the genomic sequence that later appeared in the database, oligos were designed to isolate the entire *AGL27* cDNA using RT–PCR: YSP28.11-3′: 5′-CCAATCCGTACATTCAGACA-3′ (for RT) and AGL27-6: 5′-CCGGATCCGAAGCCATGGGAAGAAGA-3′, and AGL27-7: 5′-CCGGATCCTCAGGCTTTGAGTTTAAGG-3′. Two different full-length cDNAs were isolated: AGL27-I and AGL27-II. *AGL31* cDNA clone was deduced from the overlapping ESTs (one putative intron, by similarity with *AGL27*, present in the cDNA was removed; the ESTs are T45787 and comprise the 5′ end and H36546). cDNA sequences are in accession numbers: *AGL16* (AF312662), *AGL18* (AF312663), *AGL19* (AF312664), *AGL27-1* (AF312665), *AGL27-11* (AF312666) and *AGL31* (AF312667).

Chromosomal mapping

To map *AGL16*, a genomic clone (SL76) was used to score an *Xba*I polymorphism between the Columbia (Col) and Ler ecotypes for 36 individuals of a mapping population. *AGL18* and *AGL19* were mapped using PCR-amplified genomic clones and analyzing *Eco*RI and *Eco*RV polymorphisms, respectively, between Ler and Col for 79 individuals. To map *AGL27*, an artificial *Mae*III polymorphism between Ler and Col was created and scored for 22 individuals. The genotyping oligo designed to produce this polymorphism, 5′-CCCGGTATTTTTAATTTGTGAAATTTGTAA-3′, took advantage of two nucleotide differences between these ecotypes: AT (Col) and CA (Ler). Mapping data were analyzed with MAPMAKER MACINTOSH version 2.0 (E.I. duPont de Nemours, Wilmington, DE, USA) as described by Reiter *et al.* (1992).

RNA blot analyses

Total RNA was extracted from Landsberg *erecta* plant tissues as described by Rounsley *et al.* (1995) and used for the *AGL16, AGL18* and *AGL19* Northern blots. Poly(A)+ RNA was isolated from Columbia plant tissues using the Dynabeads Oligo(dT) kit from Dynal AS (Oslo, Norway) and used for the *AGL27* and *AGL31* RNA blots. The *AGL16* probe was synthesized from a 400 bp *Eco*RI fragment of pSL4; the *AGL18* probe from a 700 bp fragment of pAGL18–4; the *AGL19* probe from a 670 bp fragment of pAGL19-3; the *AGL27* probe from a 670 bp fragment of AGL27-I which starts at the end of the I region; and the *AGL31* probe from a 500 bp fragment of H36546 which starts at the beginning of the K-box. All probes used are gene-specific and exclude the MADS-box region. As RNA-loading controls, blots were stripped and rehybridized with *β-TUBULIN* probes (data not shown) as described by Marks *et al.* (1987).

In situ *hybridization*

Tissue fixation and sectioning were performed as described by Drews *et al.* (1991), with minor modifications. Hybridization conditions used for *AGL16* and *AGL18* were as described by Drews *et al.* (1991) and Ferrándiz *et al.* (2000) for *AGL19. AGL16* and *AGL18* ^{35}S-labeled antisense mRNAs were synthesized with SP6 RNA polymerase from *Bgl*I-digested pSL5 template and *Xba*I-digested pAGL18-3 template, respectively. *AGL19* digoxigenin-labeled antisense mRNA was synthesized with T7 RNA polymerase from *Bam*HI-digested pAGL19-3 template and *AGL19* sense mRNA with SP6 RNA polymerase from the same *Not*I-digested

plasmid, both according to the manufacturer's instructions (Boehringer).

Phylogenetic analyses

The sequences used and their accession numbers are given by Alvarez-Buylla *et al.* (2000). New sequences included are: *AGL41* (AC005168), *AGL42* (AB016880), and *AGL43* (AB016885). To test the robustness of the phylogeny under different alignments, we varied the gap and extension penalties to obtain 11 different alignments. These alignments were put in a large matrix that contained the 11 alignments placed one after the next. This is called an 'elison matrix'; if used to reconstruct phylogeny, the effect of ambiguous sites on the tree topology is minimized. The topology then largely depends on unambiguously aligned sites that retain their alignment across various gap penalties (Wheeler *et al.*, 1995). Phylogenetic analyses were conducted with unweighted parsimony using PAUP 4b2 (Swofford, 2000). Heuristic searching was used with 100 replicates of random addition sequences, TBR branch swapping, and no maxtrees limit, keeping all optimal trees and with gaps treated as missing data. The tree was rooted with the AGL34-like clade as outgroup, defining it as a monophyletic sister group to the ingroup. The non-parametric bootstrap (1000 replicates) was used to assess the reliability of branches. Expression patterns were assigned as reproductive, vegetative (root/leaves) or general (both reproductive and root/leaves), and the parsimony reconstruction of ancestral states method (Maddison and Maddison, 1992) was used to infer ancestral expression patterns of putative ancestral genes at each node (Figure 6b).

Acknowledgements

We thank Alejandra Mandel (*AGL16 in situs*), Laura Alper (*AGL16* genomic clones), Francisca Acevedo, América Castañeda, Rosalinda Tapia-López (*AGL19* cDNA cloning, RNA blot and *in situs*), and especially Cristina Ferrándiz (extensive digoxigenin *in situ* technical advice and set up) for experimental assistance. Cristina Ferrándiz, Bill Crosby, Cindy Gustafson-Brown and Steve Rounsley let us use unpublished sequences, Sonia Vázquez and Ned Friedman provided extensive assistance interpreting the *AGL18 in situ* results, and Medard Ng made comments on this manuscript. This work was supported by grants from the National Science Foundation to M.F.Y. and from CONACYT (Consejo Nacional de Ciencia y Tecnología, México) to E.R.A.B. E.R.A.B. was a PEW fellow during the completion of this work and was also supported by Instituto de Ecología (UNAM-Mexico). S.L. had a Lucille P. Markey predoctoral fellowship, S.P. had a long-term postdoctoral fellowship from the Human Frontier Science Program Organization, and C.B. and F.V. had doctoral scholarships from CONACYT and UNAM

References

Alvarez-Buylla, E.R., Pelaz, S., Liljegren, S.L. *et al.* (2000) An ancestral MADS-box gene duplication occurred prior to the divergence of plants and animals *Proc. Natl Acad. Sci. USA*, **97**, 5328–5333.

Berger, F., Lindstead, P., Dolan, L. and Haseloff, J. (1998) Stomata patterning on the hypocotyl of *Arabidopsis thaliana* is controlled by genes involved in the control of root epidermis patterning. *Dev. Biol.* **194**, 226–234.

Bowman, J.L., Baum, S.F., Eshed, Y., Putterill, J. and Alvarez, J. (1999) Molecular genetics of gynoecium development in *Arabidopsis Curr. Top. Dev. Biol.* **45**, 155–205.

Chevray, P.M. and Nathans, D. (1992) Protein interaction cloning in yeast: identification of mammalian proteins that react with the leucine zipper of Jun *Proc. Natl Acad. Sci. USA*, **89**, 5789–5793.

Colombo, L., Franken, J., Koetje, E., van Went, J., Dons, H.J.M., Angenet, G.C. and van Tunen, A.J. (1995) The petunia MADS-box gene *FBP11* determines ovule identity. *Plant Cell*, **7**, 1859–1868.

Doebley, J. and Lukens, L. (1998) Transcriptional regulators and the evolution of plant form. *Plant Cell*, **10**, 1075–1082.

Doyle, J.J. (1994) Evolution of a plant multigene family – towards connecting molecular systematics and molecular developmental genetics. *Sytematic Biol.* **43**, 307–328.

Drews, G.N., Bowman, J.L. and Meyerowitz, E.M. (1991) Negative regulation of the Arabidopsis homeotic gene *AGAMOUS* by the *APETALA2* product. *Cell*, **65**, 991–1002.

Drews, G.N., Lee, D. and Christensen, C.A. (1998) Genetic analysis of female gametophyte development. *Plant Cell*, **10**, 5–17.

Ferrándiz, C., Gu, Q., Martienssen, R. and Yanofsky, M.F. (2000) Redundant regulation of meristem identity and plant architecture by *FRUITFULL*, *APETALA1* and *CAULIFLOWER*. *Development*, **127**, 725–734.

Foster, A.A. and Gifford, E.M. (1991) *Comparative Morphology of Vascular Plants*, 3rd edn. New York: W. H Freeman.

Friedman, W.E. (1992) Evidence of pre-angiosperm origin of endosperm: implications for the evolution of flowering plants. *Science*, **255**, 336–339.

Gu, Q., Ferrándiz, C., Yanofsky, M.F. and Martienssen, R. (1998) The *FRUITFULL* MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. *Development*, **125**, 1509–1517.

Hartmann, U., Höhmann, S., Nettesheim, K., Wisman, E., Saedler, H. and Huijser, P. (2000) Molecular cloning of *SVP*: a negative regulator of the floral transition in *Arabidopsis*. *Plant J.* **21**, 351–360.

Kempin, S.A., Savidge, B. and Yanofsky, M.F. (1995) Molecular basis of the *cauliflower* phenotype in *Arabidopsis*. *Science*, **267**, 522–525.

Kramer, E.M., Dorit, R.L. and Irish, V.F. (1998) Molecular evolution of genes controlling petal and stamen development: duplication and divergence within the *APETALA 3* and *PISTILLATA* MADS-box gene lineages. *Genetics*, **149**, 765–783.

Larkin, J.C., Marks, M.D., Nadeau, J. and Sack, F. (1997) Epidermal cell fate and patterning in leaves. *Plant Cell*, **9**, 1109–1120.

Lee, M.M. and Schiefelbein, J. (1999) *WERWOLF*, a MYB-related protein in *Arabidopsis*, is a position-dependent regulator of epidermal cell patterning. *Cell*, **99**, 473–483.

Liljegren, S.J., Ditta, G.S., Eshed, Y., Savidge, B., Bowman, J.L. and Yanofsky, M.F. (2000) *SHATTERPROOF MADS-box genes control seed dispersal in* Arabidopsis. *Nature*, **404**, 766–770.

Ma, H., Yanofsky, M.F. and Meyerowitz, E.M. (1991) *AGL1–AGL6*, an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes. *Genes Dev.* **5**, 484–495.

Maddison, W.P. and Maddison, D.R. (1992) *macclade: Analysis of Phylogeny and Character Evolution, Version 3.0*. Sunderland, MA, USA: Sinauer.

Mansfield, S.G. and Briarty, L.G. (1991) Early embryogenesis in *Arabidopsis thaliana*. II. The developing embryo. *Can. J. Bot.* **69**, 461–476.

Marks, M.D., West, J.L. and Weeks, D.P. (1987) The relatively large

beta-tubulin gene family of *Arabidopsis* contains a member with an unusual transcribed 5′ noncoding sequence. *Plant Mol. Biol.* **10**, 91–104

McCormick, S. (1991) Molecular analysis of male gametogenesis in plants. *Trends Genet.* **7**, 298–303.

Michaels, S.D. and Amasino, R.M. (1999) *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering *Plant Cell*, **11**, 949–956.

Pelaz, S., Ditta, G.S., Baumann, E., Wisman, E. and Yanofsky, M.F. (2000) B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature*, **405**, 200–203.

Purugganan, M.D. (1997) The MADS-box floral homeotic gene lineages predate the origin of the seed plants: phylogenetic and molecular clock estimates. *J. Mol. Evol.* **45**, 392–396.

Purugganan, M.D. (1998) The molecular evolution of development. *Bioessays*, **20**, 700–711.

Purugganan, M.D., Rounsley, S.D., Schmidt, R.J. and Yanofsky, M.F. (1995) Molecular evolution of flower development –diversification of the plant MADS-box regulatory gene family. *Genetics*, **140**, 345–356.

Reiter, R.S., Williams, J.G.K., Feldman, K.A., Rafalski, J.A., Tingey, S.V. and Scolnik, P.A. (1992) Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs. *Proc. Natl Acad. Sci. USA*, **89**, 1477–1481.

Riechmann, J.L. and Meyerowitz, E.M. (1997) MADS domain proteins in plant development. *J. Biol. Chem.* **378**, 1079–1101.

Rounsley, S.D., Ditta, G.S. and Yanofsky, M.F. (1995) Diverse roles for MADS-box genes *Arabidopsis* development. *Plant Cell*, **7**, 1259–1269.

Samach, A., Onouchi, H., Gold, S.E., Ditta, G.S., Schwarz-Sommer, Z., Yanofsky, M.F. and Coupland, G. (2000) Distinct roles of *CONSTANS* target genes in reproductive development of *Arabidopsis*. *Science*, **288**, 1613–1616.

Scheres, B., Wolkenfelt, H., Willemsen, V., Terlouw, M., Lawson, D., Dean, C. and Weisbeek, P. (1995) Mutations affecting the radial organization of the *Arabidopsis* root display specific defects throughout the embryonic axis. *Development*, **121**, 53–62.

Schiefelbein, J.W., Massuci, J.D. and Wang, H. (1997) Building a root: the control of patterning and morphogenesis during root development. *Plant Cell*, **9**, 1089–1098.

Schwechheimer, C. and Bevan, M. (1998) The regulation of transcription factor activity in plants. *Trends Plant Sci.* **3**, 378–383.

Scott, R.J., Spielman, M., Bailey, J. and Dickinson, H.G. (1998) Parent-of-origin effects on seed development in *Arabidopsis thaliana*. *Development*, **125**, 3329–3341.

Sheldon, C.C., Burn, J.E., Perez, P.P., Metzger, J., Edwards, J.A., Peacock, W.J. and Dennis, E.S. (1999) The *FLF* MADS-box gene: a repressor of flowering in Arabidopsis regulated by vernalization and methylation. *Plant Cell*, **11**, 445–458.

Shore, P. and Sharrocks, A.D. (1995) The MADS-box family of transcription factors *Eur. J. Biochem.* **229**, 1–13.

Swofford, D.L. (2000) *Paup*. Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4.* Sunderland, MA, USA: Sinauer.

Theissen, G., Kim, J.T. and Saedler, H. (1996) Classification and phylogeny of the MADS-box multigene family suggests defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. *J. Mol. Evol.* **43**, 484–516.

Theissen, G., Becker, A., DiRosa, A., Kanno, A., Kim, J.T., Münster, T. and Saedler, H. (2000) A short history of MADS-box genes in plants. *Plant Mol. Biol.* **42**, 115–149.

Tsugeki, R. and Fedoroff, N.V. (1999) Genetic ablation of root cap cells in *Arabidopsis*. *Proc. Natl Acad. Sci. USA*, **96**, 12941–12946.

Weigel, D. (1995) The genetics of flower development – from floral induction to ovule morphogenesis. *Annu. Rev. Genet.* **29**, 19–39.

Wheeler, W.C., Gatesy, J. and DeSalle, R. (1995) ELISON: a method for accomodating multiple molecular sequence alignments with alignment-ambiguous sites. *Mol. Phylogenet. Evol.* **4**, 1–9.

Zachgo, S., Saedler, H. and Schwarz-Sommer, Z. (1997) Pollen-specific expression of DEFH125, a MADS-box transcription factor in *Antirrhinum* with unusual *Features*, **11**, 1043–1050.

Zhang, H. and Forde, B.G. (1998) An *Arabidopsis* MADS-box gene that controls nutrient-induced changes in root architecture. *Science*, **279**, 407–409.

Accession numbers AF312662, AF312663, AF312664, AF312665, AF312666, AF312667.

Planta
DOI 10.1007/s004250100637

Original Article

MADS-box gene expression in lateral primordia, meristems and differentiated tissues of *Arabidopsis thaliana* roots

Caroline Burgeff · Sarah J. Liljegren · Rosalinda Tapia-López · Martin F. Yanofsky · Elena R. Alvarez-Buylla(✉)

C. Burgeff · R. Tapia-López · E.R. Alvarez-Buylla
Instituto de Ecología, UNAM, AP-Postal 70-275, México D.F. 04510, México

S.J. Liljegren · M.F. Yanofsky
Department of Biology, UCSD, La Jolla, CA 92093-0116, USA

✉ E-mail: abuylla@servidor.unam.mx
Fax: +52-5-6161976

Received: 8 April 2001 / **Accepted:** 2 May 2001 / **Published online:** †

Abstract. Although MADS-box genes involved in flower and fruit development have been well characterized, the function of MADS-box genes expressed in vegetative structures has yet to be explored. At least seven members of this family are grouped in clades of genes that are preferentially expressed in roots of *Arabidopsis thaliana* (L.) Heynh.. We report here the cloning of the *AGL21* MADS-box gene, which belongs to the *ANR1* clade, and the mRNA in situ expression patterns of this and two other root MADS-box genes. *AGL17* appears to be a lateral root cap marker in the root tip, and towards the elongation zone this gene is expressed in the epidermal cells. *AGL21* is highly expressed in lateral root primordia and it has a punctate expression pattern in the primary root meristem. *AGL12* also has a punctate expression pattern in the primary root meristem. *AGL12* and *AGL21* are also expressed in the central cylinder of differentiated roots and both are expressed in developing embryos. This study, combined with previous phylogenetic analyses, indicates that these MADS-box genes may play distinct regulatory roles during root development.

Keywords. *Arabidopsis* - *ANR1* - Embryo development - Lateral root primordium - MADS-box - Root (MADS-box genes)

Abbreviations. LRC: lateral root cap LRP: lateral root primordium RT-PCR: reverse transcription-polymerase chain reaction

MADS-box gene expression in lateral primordia, meristems and differentiated tissues of *Arabidopsis thaliana* roots

Caroline Burgeff [1], Sarah J. Liljegren[2], Rosalinda Tapia-López[1], Martin F. Yanofsky[2], Elena R. Alvarez-Buylla[1]

[1] Instituto de Ecología, UNAM, AP-Postal 70-275, México D.F. 04510. México

[2] Department of Biology, UCSD, La Jolla CA, 92093-0116, USA

Running Title: Expression patterns of root MADS-box genes

Received: 8 April 2001 / Accepted: 2 May 2001

Correspondence to: E.R. Alvarez-Buylla; E-mail: abuylla@servidor.unam.mx

Fax: 52(5)6161976

Abstract. Although MADS-box genes involved in flower and fruit development have been well characterized, the function of MADS-box genes expressed in vegetative structures has yet to be explored. At least seven members of this family are grouped in clades of genes that are preferentially expressed in roots of *Arabidopsis thaliana* (L.) Heynh.. We report here the cloning of the *AGL21* MADS-box gene, which belongs to the *ANR1* clade, and the mRNA *in situ* expression patterns of this and two other root MADS-box genes. *AGL17* appears to be a lateral root cap marker in the the root tip and towards the elongation zone this gene is expressed in the epidermal cells. *AGL21* is highly expressed in lateral root primordia and it has a punctate expression pattern in the primary root meristem. *AGL12* also has a punctate expression pattern in the primary root meristem. *AGL12* and *AGL21* are also expressed in the central cylinder of differentiated roots and both are expressed in developing embryos. This study combined with previous phylogenetic analyses indicate that these MADS-box genes may play distinct regulatory roles during root development.

Key words: *Arabidopsis* - *ANR1* – Embryo development - lateral root primordium - MADS-box – Root (MADS-box genes)

Abbreviations: LRC = lateral root cap; LRP = lateral root primordium; RT-PCR: reverse transcription-polymerase chain reaction

Introduction

MADS-box genes play important functional roles in eukaryotes (Riechmann and Meyerowitz 1997; Black and Olson 1998; Theissen et al. 2000). In plants, members of this family of transcription factors are involved in flower and fruit development in aspects such as meristem and flower organ identity determination, silique development, as well as flowering initiation (Riechmann and Meyerowitz 1997; Ferrándiz et al. 1999; Sheldon et al. 1999; Lee et al. 2000; Pelaz et al. 2000). Phylogenetic analyses have shown that MADS-box genes form monophyletic groups of genes specifically expressed in reproductive structures (Purugganan et al. 1995; Theissen et al. 1996) and more recent analyses show that in *Arabidopsis,* MADS-box genes expressed in vegetative structures also group in distinct clades (Alvarez-Buylla et al. 2000b). These results suggest that MADS-box genes could be involved in the regulation of the development of vegetative structures. However, very little is known about the function of vegetatively expressed MADS-box genes of which several are preferentially expressed in the *Arabidopsis* root (Alvarez-Buylla et al. 2000a).

The anatomical simplicity, small size and transparency of the *Arabidopsis* root have made it a model for studies on plant organogenesis (Benfey and Schiefelbein 1994). In the root tip of *Arabidopsis,* three successive but ovelapping regions have been distinguished (Dolan et al. 1993). The root meristem or proliferative zone constists of actively dividing cells and is located at the root tip distal end. Behind the meristem, cells expand longitudinally to constitute the elongation zone and finally, cells progressively attain their fully mature stage in the differentiation zone. In cross section, the root presents a radial pattern of single concentric tissue layers. At the root meristem these are delimited by the

lateral root cap (LRC) that surrounds the epidermis, the cortex and the endodermis, which, in turn, encircle the central cylinder or stele consisting of a pericycle and a vascular bundle (Dolan et al. 1993). All root tissues are derived from a set of initial cells that surround the quiescent center and together constitute the promeristem (Clowes 1954). This is the minimal construction center of the root and is located towards the root tip, which is distally protected by actively dividing and exfoliating columella root cap cells.

Laser ablation experiments have established the importance of cell position and short-range signaling during cell-fate determination (van den Berg et al. 1995, 1997). Several developmental root mutants have been characterized and some of the genes underlying these mutations have been functionally analyzed and studied at the molecular level (e.g. Di Laurenzio et al. 1996; Masucci et al. 1996; reviewed by Scheres and Heidstra 1999). This is particularly true for the genes controlling sub-differentiation of the root epidermis (reviewed by Schiefelbein 2000). However, much more work is still needed in order to uncover the molecular mechanisms that underlie root development.

Recently identified root MADS-box genes are likely to become important molecular tools to with which continue exploring the transcriptional networks controlling root development (Alvarez-Buylla et al. 2000a). A root MADS-box gene, *ANR1*, seems to mediate nitrate-induced lateral root growth stimulation (Zhang and Forde 1998). However, no MADS-box gene has been identified in previous mutant screens. This could be due to overlapping functions among members of this family (e.g. Liljegren et al. 2000; Pelaz et al. 2000; Smyth 2000), but lethality and subtle phenotypes that are hard to score under standard growth conditions should not be excluded. In this context, we have decided to explore the role of MADS-box genes in root development with a reverse genetic approach, starting with the analysis of mRNA expression patterns.

Based on the studies of the MADS-box genes thus far characterized at the functional level (eg. Yanofsky et al. 1990; Mandel et al. 1992; Riechmann and Meyerowitz 1997; Liljegren et al. 2000) it is reasonable to propose the hypothesis that expression patterns of members of this gene family are a good guide of where they likely function. In fact, some of these genes were first characterized at the expression level, and, based on these data, it was possible to predict their function. Only later was it genetically proven that they indeed act when and where they had previously been shown to be expressed (e.g. Liljegren et al. 2000; Pelaz et al. 2000 and references therein). Therefore, expression patterns and phylogenetic analyses of these genes will likely become guidelines for identifying redundant groups and directing genetic and further functional analyses (e.g. Rounsley et al. 1995; Liljegren et al. 2000; Molin et al. 2000).

We report here the cloning of *AGL21*, a new MADS-box gene, and the mRNA *in situ* expression patterns of this and two other root MADS-box genes, *AGL17* and *AGL12,* in roots and embryos of *Arabidopsis*. We use the mRNA expression patterns reported here to advance preliminary hypotheses on their functions and, together with previous phylogenetic reconstructions, to evaluate the possible redundant groups and hence guide genetic analyses.

Materials and Methods

Plant material and growth conditions. For root *in situ* analyses, seeds of *Arabidopsis thaliana* (L) Heynh.,Columbia ecotype (Col), were surface-sterilized for 2 min in 70% ethanol, and 20 min in 5% sodium hypochlorite, 1% SDS. After rinsing and re-suspending in 0.1% tissue culture agar (TC agar; Carolina Biological Supply Co.), the seeds were vernalized for 2 days at 4°C. Seeds were put on 1X MS (SIGMA M0404) agar medium (1% saccharose, pH 5.7) in petri dishes held in vertical position in a plant growth chamber (Lab-line Bionette) with continuous light at 22 °C. Landsberg *erecta* (Ler) and Col plants grown in soil were vernalized after sowing and held in a growth chamber at 22 - 24 °C under either continuous light or a 16 h light/8 dark regime.

AGL21 cloning and sequence determination. Using Ler genomic DNA as template, the *AGL21* MADS-box region was amplified with the MADS-1 and MADS-43 degenerate PCR oligos as previously described (Rounsley et al. 1995). The following nested oligos SL3-1 (5'-TGTGATCCAAAGGATCGATG-3') AND SL3-2 (5'-GATGATTCAACGAGTAGACA-3') were then used with oligo (dT) to amplify the 3' end of the *AGL21* cDNA from root tissue, using reverse transcription-polymerase chain reaction (RT-PCR). The 5' region of the MADS-box domain was obtained through TAIL (Thermal Asymmetric Interlaced) PCR performed as previously described (Tsugeki et al. 1996). *AGL21*-specific nested oligonucleotides, (5'-GCAAAGTCATAGAGCTTTCCGGTG-3') and (5'GATGAGACCGACCTCGGCATCAC-3'), were used in successive round with the following degenerate primer, 5'-(A/G/C/T)TCGA(C/G)T(A/T)T(C/G)G(A/T)GTT-3'. The complete coding sequence of the gene was amplified by RT-PCR from root cDNA using the oligonucleotides CB15F (5'-CGGATCCAATGGGAAGAGGGAAGATTGT-3') and CB16R (5'-CGGATCCTTATTCGTTTGCTCTTGGTG-3'). The cloning and subcloning

vectors used were: Bluescript (Stratagene) and PGEMT (Promega). Sequence analysis and edition was done using GeneWorks and MacVector computer programs.

AGL21 mapping. *AGL21* was initially mapped by PCR using three-dimensional YAC pools of the CIC library (Creusot et al. 1995; *Arabidopsis* Biological Resource Center http://aims.cps.msu.edu/aims/). Afterwards, map position was confirmed by genomic sequencing information.

RNA blot analysis. Total RNA was extracted from various Ler plant tissues as previously described (Rounsley et al. 1995). The *AGL21* probe was synthesized from a 578-bp EcoRI fragment of pSL73 and is specific for the 3'-region of the gene. RNA loading controls were performed as in Rounsley et al. (1995).

In situ hybridization. The roots of plants grown on MS agar plates for 7 - 14 days were fixed in 4% paraformaldehyde, vacuum-infiltrated and kept overnight at 4°C. Siliques of soil-grown plants were fixed overnight in FAE solution (ethanol: acetic acid:formaldehyde:water, 50:5:3.5:41.5, by vol.). Samples were then dehydrated through an ethanol series, passed through ethanol:Histoclear (National Diagnostics) series and embedded in Paraplast+ (Oxford Labware). Microtome sections (9μm) of the samples were mounted on Probe-On Plus Slides (Fisher Scientific). Roots used in whole-mount *in situ* experiments were fixed as described above and kept refrigerated in 70% ethanol until *in situ* hybridization. Afterwards, the root tips were pre-embedded in agarose (1%), then embedded in Paraplast+ and transversally sectioned (technical details concerning root manipulation prior to sectioning are available upon request). Digoxigenin-labeled probes were synthesized following the manufacturer's instructions (Boehringer Mannheim) and hydrolyzed to an average length of 200bp.

Probes were chosen to avoid the presence of the MADS-box sequence. The antisense probe for *AGL12* was synthesized with the T7 RNA polymerase from an *XbaI*-linearized pSR103 plasmid, consisting of a 266-bp fragment of the 3' UTR (untranscribed region) of the cDNA. For *AGL17*, a probe consisting of 430bp of the 3' region of the cDNA was synthesized with T7 RNA polymerase over a *BamHI*-linerarized pSR192 clone that contains the *EcoRI* 3'-end fragment of the cDNA. The antisense probe for *AGL21* was obtained with T3 RNA polymerase from the *KpnI*-linearized plasmid pSL215. This plasmid contains the 578 bp *EcoRI* 3'-end fragment of the cDNA. For all the genes studied and in all experiments performed for different tissues, sense probes were used as negative controls. Pre-hybridization, hybridization, washes and detection conditions followed Ferrándiz et al. (2000), with minor modifications: The proteinase K concentration was 0.5µg/ml; hybridization and 2xSSC/50%formamide washes were done at 50 °C.

Microscopy and image analysis. Photographs were taken using a Nikon Labophot2 MT or an Olympus BX 60 microscope with Nomarski optics. Slides were scanned using Nikon SF200 scanner and Nikon Scan 2.1 software. Images were processed using Adobe Photoshop 5.0.

Results

AGL21 cloning and sequence determination. In an effort to uncover additional root-specific MADS-box genes, a PCR-based approach (Rounsley et al. 1995; Alvarez-Buylla et al. 2000b) was used to clone additional members of the MADS-box gene family. *AGL21* cDNA has an open reading frame of 228 amino acids (Genbank accession number AF336979). The putatively encoded protein contains a 58 amino acid MADS domain at the N-terminus, followed by a 31-amino-acid I-domain, a 64-amino-acid K-domain, and a 75-amino-acid C-terminus domain, with the typical structure of Type-II plant MADS-box genes (see also Alvarez-Buylla et al. 2000a, b).

AGL21 was found in the YAC CIC 6C9, positioned on chromosome 4, (http://www.Arabidopsis.org) (Accession number AL035538 for BAC F20D10).

AGL21 is primarily expressed during lateral root formation and embryogenesis. RNA blot analysis detected *AGL21* expression in roots, while the transcript was not detected in leaves, stems, flowers or siliques (Fig. 1). However, RT-PCR analyses revealed a clear expression of this gene in siliques as well as roots (data not shown). To obtain a more detailed pattern of expression, *in situ* hybridization experiments were performed.

mRNA *in situ* hybridization experiments detected *AGL21* expression at high levels during lateral root formation, from young lateral root primordia (LRPs) of about three to four cell layers, up to lateral root emergence, when labeling was mostly confined to the apex (Fig. 2A,B). We were not able to detect transcripts of this gene in younger LRPs. As the lateral root started to grow, the gene continued to be expressed mainly in the proliferative region gradually adopting a punctate pattern. *AGL21* transcripts were also found in the LRPs that developed near the hypocotyl-root junction (Fig. 2C). This gene was

expressed at relatively lower levels in the primary meristem of the roots where only some cells were labeled in a punctate pattern, although some sections showed a higher proportion of stained cells (Fig. 2D, E). In the differentiated zone of the root, *AGL21* was observed in the central cylinder. Staining intensity varied along the root length and in some cases higher labeling could be seen in the pericycle cells (results not shown).

The presence of *AGL21* mRNA during LRP development led us to investigate whether this and other root MADS-box genes are expressed during root formation in embryos. *In situ* experiments showed that *AGL21* is also present in embryos at least from the globular stage up to the torpedo stage (Fig. 2G-I). During these stages the labeling sometimes showed an increasing gradient from the basal to the apical pole of the embryo (Fig. 2G, I). Preliminary data showed that *ANR1* had an overall similar expression pattern to that of *AGL21* in roots and embryos (data not shown) in plants grown in our conditions (see Materials and methods).

AGL17 mRNA *is specific to the LRC and epidermis of root tips.* *AGL17* was specifically expressed in the LRC along the root meristem (Fig. 3A,C). Once the LRC disappeared close to the limit of the proliferative zone, this gene was expressed in the epidermis (Fig. 3A, F). Whole-mount *in situ* hybridization of root tips subsequently embedded and transversally sectioned was used to confirm the latter pattern (Fig. 3E, F). Some of the signal observed in Fig 3F might belong to collapsed LRC cells overlying the epidermis. However, careful observation of the pattern of labeled cells along successive sections (results not shown), as well as the labeling of the cells along the anticlinal walls of the epidermis support the possibility that the signal is actually present in the epidermis. In some cases, *AGL17* expression was observed in files of central cylinder cells that seemed to be

pericycle cells in the early elongation zone of longitudinal sections. In the differentiation zone, *AGL17* was observed in the pericycle and at very low levels in other central cylinder cells (results not shown). We could not detect *AGL17* mRNA during LRP formation or during embryogenesis at least up to the late heart stage.

AGL12 transcripts are present in the meristem and in the central cylinder of the mature root. In root meristems this gene was mainly expressed in the external cells of the columella and in the LRC. Most sections also showed staining in all tissue types but with a punctate pattern, similar to that observed for *AGL21* (Fig. 4A, C). However, in transverse sections towards the basal end of the proliferative zone, *AGL12* transcripts were preferentially expressed in atrichoblasts rather than in trichoblasts (Fig. 4A). In longitudinal sections of differentiated roots, *AGL12* was highly expressed in the central cylinder (Fig. 4E). Staining intensity varied along the root length and in some cases a more intense labeling was observed in pericycle cells (results not shown). We did not detect the transcript in LRPs but once lateral roots were formed, we found the same pattern of expression as in the primary root meristem. *AGL12* was also expressed in a punctate pattern from the globular stage to at least the torpedo stage of embryogenesis, with the frequency of stained cells increasing as embryogenesis proceeded (Fig. 4D). RT/PCR experiments corroborated the presence of *AGL12* transcript during silique development (results not shown).

Discussion

Two of the MADS-box genes analyzed in this paper belong to the single monophyletic *ANR1* clade (Alvarez-Buylla et al. 2000b). The newly cloned MADS-box gene, *AGL21*, is most similar to *AGL17*, and both are closely related to *AGL16* and *ANR1* (Rounsley et al. 1995; Zhang and Forde 1998; Alvarez-Buylla et al. 2000b). All these genes are almost exclusively expressed in roots, but *AGL16* is also expressed in specific cells of the leaf epidermis (Alvarez-Buylla et al. 2000a) and *AGL21* is also expressed in embryos. Based only on pair-wise sequence percentage identity at the amino acid level, we would expect that AGL17 and AGL21 have the greatest possibility of being functionally redundant. AGL21 shares with AGL17 amino acid sequence identities of 73.4% for the complete protein and 77.1% for the MADS-I and K-domains sequence. The rest of the pair-wise comparisons for the genes in the same clade yielded identity percentages below 60%, suggesting less chance of functional redundancy between genes within a pair.

In contrast to the sequence comparison data, our expression data show that *AGL17* and *AGL21* have contrasting mRNA expression patterns, suggesting that these two genes are not functionally redundant. On the other hand, preliminary results show that similar expression patterns exist in roots between *AGL17* and *AGL16*, as well as between *AGL21* and *ANR1* (C. Burgeff and E. R. Alvarez-Buylla unpublished results). These data suggest that possible functional redundancy may be present within gene pairs of this clade, although not strictly between the most recently duplicated genes. In *Caenorhabditis*, a case of functional redundancy has been recently reported between two genes of the forkhead transcription factors family that show similar expression patterns but which have more sequence divergence between each other than with respect to other family members (Molin et al. 2000).

We detected *AGL21* mRNA during lateral root formation in LRPs starting at stages III or IV (Malamy and Benfey 1997) and up to when the root has emerged. At this stage the expression was confined to the apex, presumably in the region where cells actively proliferate. We cannot rule out expression in earlier stages, because we were not able to observe younger primordia. These results suggest an association of this gene with regions of cells that maintain proliferative potential.

Once a lateral root meristem has formed, the expression pattern of *AGL21* is restricted to the proliferative zone and as the root elongates it becomes punctate and less intense. A similar punctate pattern was also observed in the primary root meristem where the number of stained cells varied among experiments as exemplified by the longitudinal and transverse sections shown in Fig. 2D-E. Similar punctate mRNA and protein patterns have been reported in plants for cell-cycle related genes, such as cyclins, histones and some cdc2-like genes (Ferreira et al. 1994; Fobert et al. 1994). In roots, *ANR1* appears to be involved in lateral root growth stimulation in response to higher local concentrations of nitrates. This suggestion was derived from antisense and co-suppression lines (Zhang and Forde 1998). Nitrate-induced lateral root growth seems to be associated with changes in the rate of cell production rather than with changes in cell elongation (Zhang et al. 1999). It will be interesting to test if *AGL21* and *ANR1* have similar roles. Interestingly, our preliminary data on mRNA pattern for *ANR1* show that this gene is also expressed in the LRPs (data not shown).

AGL21 is also expressed during embryo development, at least from the octant–globular to the late heart–torpedo stages, as assessed by *in situ* and RT/PCR experiments. The inability to detect its presence in siliques by RNA blot analysis might be due to the fact that the gene expression level in these structures could be below the detection limits of this

technique. In about 70% of the cases, a higher staining level is present in the apical region of the embryo and sometimes it is not clearly seen in the embryonic root, suggesting that *AGL21* may not be continuously expressed during embryo root formation. This pattern suggests that this gene transcription could be associated with the onset of high cell production, which, in the apical region of the embryo, leads to cotyledon formation, or it could indicate that *AGL21* is involved in different genetic circuits during embryogenesis and root post-embryonic development.

AGL17 is expressed in the LRC and the epidermis. Overall, our careful examination of successive sections overall suggests that *AGL17* is indeed expressed in the epidermis from the end of the meristem up to the elongation zone. However, in whole-mount *in situ*, the epidermal staining, which at this level is mainly seen towards the cell walls, could be interpreted as background (Fig. 3F). But no such signal was observed in the sense controls at this level (data not shown), rather suggesting then, that such cellular localization of the signal could be due to cell vacuolisation in the incipient elongating cells in which the cytoplasm is displaced towards the cell margins.

The lateral root cap and the epidermis seem to play fundamental roles in sensing external stimuli and in the basipetal redistribution of auxin respectively, affecting root growth and morphogenesis (eg. Müller et al. 1998; Tsugeki and Fedoroff 1999). From the end of the proliferative zone up to the elongation zone *AGL17* is expressed in the epidermis. In this zone, the fixation of cell-fate determination and the differential cell elongation during gravitropic response take place (Masson 1995; Berger et al. 1998). Additionally, this zone includes the region where the cells in a file cease proliferation, defining the basal end of the meristematic zone. One would expect that important controls affecting root growth plasticity occur around this zone, because the total number of

proliferating cells per file in the meristem has a direct impact on the total root growth rate (Beemster and Baskin 1998). Alternatively, it could be hypothesized that *AGL17* may be involved in LRC cell specification in the root. However, the radial organization of the root is laid down during embryogenesis (Scheres 1996), and this gene is neither expressed in the late heart stage of the embryo, at the end of which the formation of an LRC layer by the epidermal initials is often viewed as the onset of root meristematic activity, nor in the LRC of young emergent LRPs when LRC molecular markers are already expressed (Dolan et al. 1993; Scheres et al. 1994; Malamy and Benfey 1997). Our data suggest, then, that this gene is not likely to be involved in the early stages of cell fate specification in the LRC.

AGL12 is another largely root-specific gene that does not belong to the *ANR1* monophyletic clade. This gene branches off by itself in the MADS tree and up to now no MADS-box genes closely related to *AGL12* have been described (Alvarez-Buylla et al. 2000b). In the proliferative zone of the root its punctate pattern is similar to that of *AGL21* and these two genes could be partially redundant at the functional level. In the epidermis, however, *AGL12* seems to be preferentially expressed in atrichoblast rather than in trichoblast cells. This is clear in the basal end of the meristem. Cytoplasm density and the degree of cell vacuolization also change between atrichoblasts and trichoblasts at this level of the meristem (Dolan et al. 1993; Masucci et al. 1996).

In the differentiated tissues of the root, *AGL12* is found in the central cylinder cells including the pericycle, although the staining level varies along the root length. Several MADS-box genes studied here and elsewhere (see also Alvarez-Buylla et al. 2000b) seem to be expressed in a similar pattern in the differentiated root portion. *AGL12* mRNA was also detected during embryo development up to at least the torpedo stage.

Root MADS-box gene expression patterns suggest that these genes might have contrasting and diverse roles during *Arabidopsis* root development and guide further analyses. Functional studies of these genes are needed to unravel their role during root development.

Aknowledgements

We thank A. Pérez for assistance during *in situ* experiments, A. Cortés for logistical support, and C. Ferrándiz, A. Gamboa and F. Acevedo for valuable advice during laboratory work. Many thanks to J. Dubrovsky, G. Cassab and two anonymous reviewers for sharing their knowledge and for comments on earlier drafts of this manuscript. Special thanks to M. Delseny, and the "physical mapping team" in Perpignan (A.B., R.C., R.G., G.B.) for help in mapping *AGL21* and to H. Merchant for sharing with us his Nomarski Optics Microscope facilities. This work was supported by grants from CONACYT (Consejo Nacional de Ciencia y Tecnología, México), PAPIIT (UNAM) and the Human Frontiers Science Program to E.R.A.-B., and from National Science Foundation to M.F.Y.. E.R.A.-B. was a PEW fellow during part of this work; C.B. and R.T. had Ph.D.scholarships from CONACYT and UNAM. S.L. had a Lucille P. Markey pre-doctoral fellowship.

References

Alvarez-Buylla ER, Liljegren SJ, Pelaz S, Gold SJ, Burgeff C, Ditta GS, Vergara-Silva F, Yanofsky MF (2000a) MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells roots and trichomes. Plant J 24: 457-466

Alvarez-Buylla ER, Pelaz S, Liljegren SJ, Gold SE, Burgeff C, Ditta GS, Ribas de Pouplana L, Martínez-Castilla L, Yanofsky MF (2000b) An ancestral MADS-box duplication ocurred before the divergence of plants and animals. Proc Natl Acad Sci USA 97: 5328-5333

Beemster GTS, Baskin TI (1998) Analysis of cell division and elongation underlying the developmental acceleration of root growth in *Arabidopsis thaliana*. Plant Physiol 116: 1515-1526

Benfey P, Schiefelbein JW (1994) Getting to the root of plant development: the genetics of *Arabidopsis* root formation. Trends Genet 10: 84-88

Berger F, Yung CY, Dolan L, Schiefelbein J (1998) Control of cell division in the root epidermis of *Arabidopsis thaliana*. Dev Biol 194: 235-245

Black BL, Olson EN (1998) Transcriptional control of muscle development by myocyte enhancer factor-2 (MEF2) protein. Annu Rev Cell Dev Biol 14: 167-196

Clowes FAL (1954) The promeristem and the minimal constructional centre in grass root apices. New Phytol 52: 48-57

Creusot F, Fouilloux E, Dron M, Lafleuriel J, Picard G, Billault A, Le Paslier D, Cohen D, Chaboue ME, Durr A, et al (1995) The CIC library: a large insert YAC library for genome mapping in *Arabidopsis thaliana*. Plant J 8: 763-770

Di Laurenzio L, Wysocka-Diller J, Malamy JE, Pysh L, Helariutta Y, Freshour G, Hahn MG, Feldmann KA, Benfey PN (1996) The *SCARECROW* gene regulates an asymmetric cell division that is essential for generating the radial organization of the *Arabidopsis* root. Cell 86: 423-433

Dolan L, Janmaat K, Willemsen V, Linstead P, Poethig S, Roberts K, Scheres B (1993) Cellular organisation of the *Arabidopsis thaliana* root. Development 119: 71-84

Ferrándiz C, Pelaz S, Yanofsky MF (1999) Control of carpel and fruit development in *Arabidopsis*. Annu Rev Biochem 68: 321-354

Ferrándiz C, Gu Q, Martienssen R, Yanofsky MF (2000) Redundant regulation of meristem identity and plant architecture by *FRUITFULL*, *APETALA1* and *CAULIFLOWER*. Development 127: 725-734

Ferreira PCG, Hemerly AS, de Almeida Engler J, Van Montagu M, Engler G, Inze D (1994) Developmental expression of the *Arabidopsis* cyclin gene *cyc1At*. Plant Cell 6: 1763-1744

Fobert PR, Coen ES, Murphy GJP, Doonan JH (1994) Patterns of cell division revealed by transcriptional regulation of genes during the cell cycle in plants. EMBO J 13: 616-624

Lee H, Suh SS, Park E, Cho E, Ahn JH, Kim SG, Lee JS, Kwon YM, Lee I (2000) The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in *Arabidopsis*. Genes Dev 14: 2366-2376

Liljegren SJ, Ditta GS, Eshed Y, Savidge B, Bowman JL, Yanofsky MF (2000) *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. Nature 404: 766-70

Malamy JE, Benfey P (1997) Organization and cell differentiation in lateral roots of *Arabidopsis thaliana*. Development 124: 33-44

Mandel MA, Gustafson-Brown C, Savidge B, Yanofsky MF (1992) Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. Nature 360: 273-277

Masson PH (1995) Root gravitropism. BioEssays 17: 119-127

Masucci JD, Rerie WG, Foreman DR, Zhang M, Galway ME, Marks MD, Schiefelbein JW (1996) The homeobox gene *GLABRA 2* is required for position-dependent cell differentiation in the root epidermis of *Arabidopsis thaliana*. Development 122: 1253-1260

Molin L, Mounsley A, Aslam S, Bauer P, Young J, James M, Sharma-Oates A, Hope IA (2000) Evolutionary conservation of redundancy between a diverged pair of forkhead transcription factor homologues. Development 127: 4825-4835

Müller A, Guan C, Gälweiler L, Tänzler P, Huijser P, Marchant A, Parry G, Bennett M, Wisman E, Palme K (1998) *AtPIN2* defines a locus of *Arabidopsis* for root gravitropism control. EMBO J 17: 6903-6911

Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF (2000) B and C floral organ identity functions require *SEPALLATA* MADS-box genes. Nature 405: 200-203

Purugganan MD, Rounsley SD, Schmidt RJ, Yanofsky MF (1995) Molecular evolution of flower development: diversification of the plant MADS-box regulatory gene family. Genetics 140: 345-356

Riechmann JL, Meyerowitz EM (1997) MADS domain proteins in plant development. Biol Chem 378: 1079-1101

Rounsley SR, Ditta GS, Yanofsky M (1995) Diverse roles for MADS-Box genes in *Arabidopsis* development. Plant Cell 7:1259-1569

Scheres B (1996) Embryo patterning genes and reinforcement cues determine cell fate in the *Arabidopsis thaliana* root. Sem Cell Dev Biol 7: 857-865

Scheres B, Heidstra R (1999) Digging out roots: pattern formation, cell division, and morphogenesis in plants. Curr Top Dev Biol 45: 207-247

Scheres B, Wolkenfelt H, Willemsen V, Terlouw M, Lawson E, Dean C, Weisbeek P (1994) Embryonic origin of the *Arabidopsis* primary root and meristem initials. Development 120: 2475-2487

Schiefelbein JW (2000) Constructing a plant cell. The genetic control of root hair development. Plant Physiol, 124: 1525-1531

Sheldon CC, Burn JE, Perez PP, Metzger J, Edwards JA, Peacock WJ, Dennis ES (1999) The *FLF* MADS box gene: a repressor of flowering in *Arabidopsis* regulated by vernalization and methylation. Plant Cell 11: 445-458

Smyth D (2000) A reverse trend-MADS functions revealed. Trends Plant Sci 5: 315-317

Theissen G, Becker A, Di Rosa A, Kanno A, Kim JT, Münster T, Winter KU, Saedler H (2000) A short story of MADS-box genes in plants. Plant Mol Biol 42: 115-149

Theissen G, Kim JT, Saedler H (1996) Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. J Mol Evol 43: 484-516

Tsugeki R, Fedoroff NV (1999) Genetic ablation of root cap cells in *Arabidopsis*. Proc Natl Acad Sci USA 96: 12941-12946

Tsugeki R, Kochieva EZ, Fedoroff NV (1996) A transposon insertion in the *Arabidopsis* SSR16 gene causes an embryo-defective lethal mutation. Plant J 10:479-489

van den Berg C, Willemsen V, Hage W, Weisbeek, Scheres B (1995) Cell fate in the *Arabidopsis* root meristem determined by directional signalling. Nature 378: 62-65

van den Berg C, Willemsen V, Hendriks G, Weisbeek P, Scheres B (1997) Short-range control of cell differentiation in the *Arabidopsis* root meristem. Nature 390: 287-289

Yanofsky MF, Ma H, Bowman JL, Drews GN, Feldmann KA, Meyerowitz EM (1990) The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. Nature 346: 35-39

Zhang H, Forde BG (1998) An *Arabidopsis* MADS-Box gene that controls nutrient-induced changes in root architecture. Science 279: 407-409

Zhang H, Jennings A, Barlow PW, Forde BG (1999) Dual pathways for regulation of root branching by nitrate. Proc Natl Acad Sci USA 96: 6529-6534

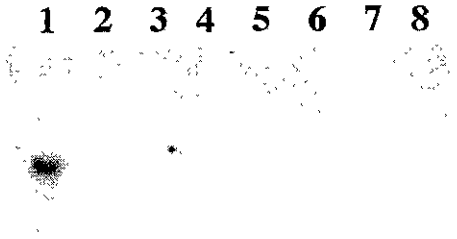

Fig. 1. mRNA gel blot hybridization of *AGL21* from *Arabidopsis thaliana*. A gene specific probe was used and equal RNA loading was tested with a tubulin probe. Lane 1=root; 2=leaves; 3=stems; 4=flowers; 5=young siliques; 6=old siliques; 7= 3 days old seedlings; 8= 8 days old seedlings.

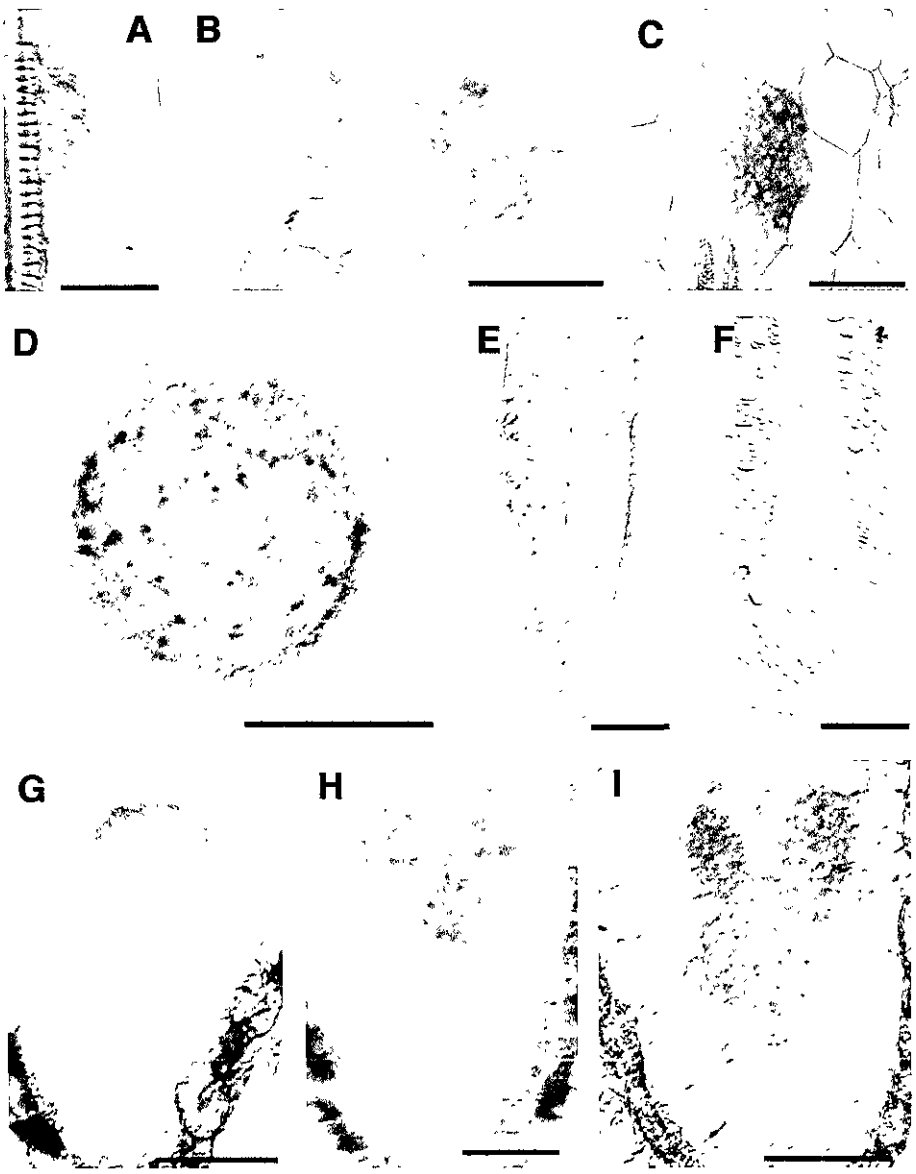

Fig. 2a-i. Pattern of mRNA expression of *AGL21* in roots and embryos of *A. thaliana*. In situ Hibridization of digoxigenin-labeled probes is shown. **a-c** Antisense probe hybridizations of longitudinal sections of a young LRP (**a**), an emergent LRP (**b**), and a LRP at a hypocotyl-root junction (**c**). **d,e** Antisense probe hybridization of transverse (**d**) and longitudinal (**e**) sections of a primary root meristem. **f** Sense probe hybridization of a longitudinal section of a primary root meristem. **g-i** antisense probe hybridizations of longitudinal sections of emryos at the globular (**g**), late triangular (**h**) and torpedo (**i**) stages. Bars= 50 um (**b-f, i**), 25um (**a,g,h**)

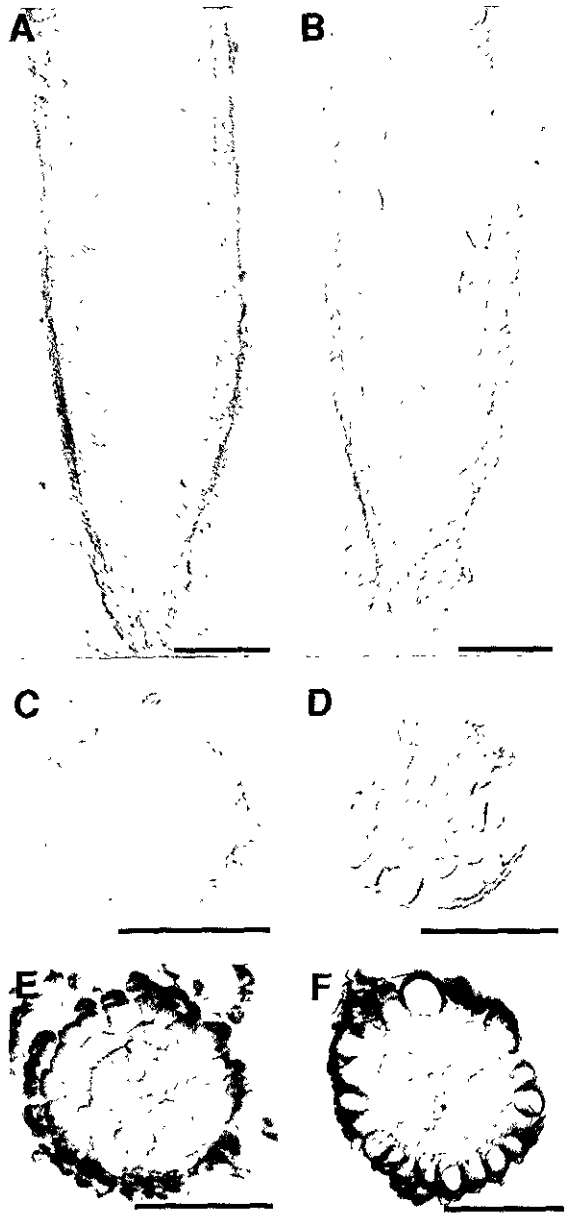

Fig. 3a-f. Pattern of mRNA expression of *AGL17* in roots of *A. thaliana*. In situ hybridizations of digoxigenin-labeled probes is shown. **a,b** Antisense (**a**) and sense (**b**) probe hybidizations in a longitudinal section of a 14-day-old root tip. **c** Antisense probe hybridization of a transverse senction in the primary root meristem. **d-f** Whole mount in situ hybridizations of 7-day-old roots, subsequently embedded and sectioned. **d** Sense probe in a transverse section at the meristematic zone near the initials. **e, f** Antisense probe hybridizatton in a transverse section at the meristematic zone near the initials (**e**), and at the incipient elongation zone (f). Bars = 50um

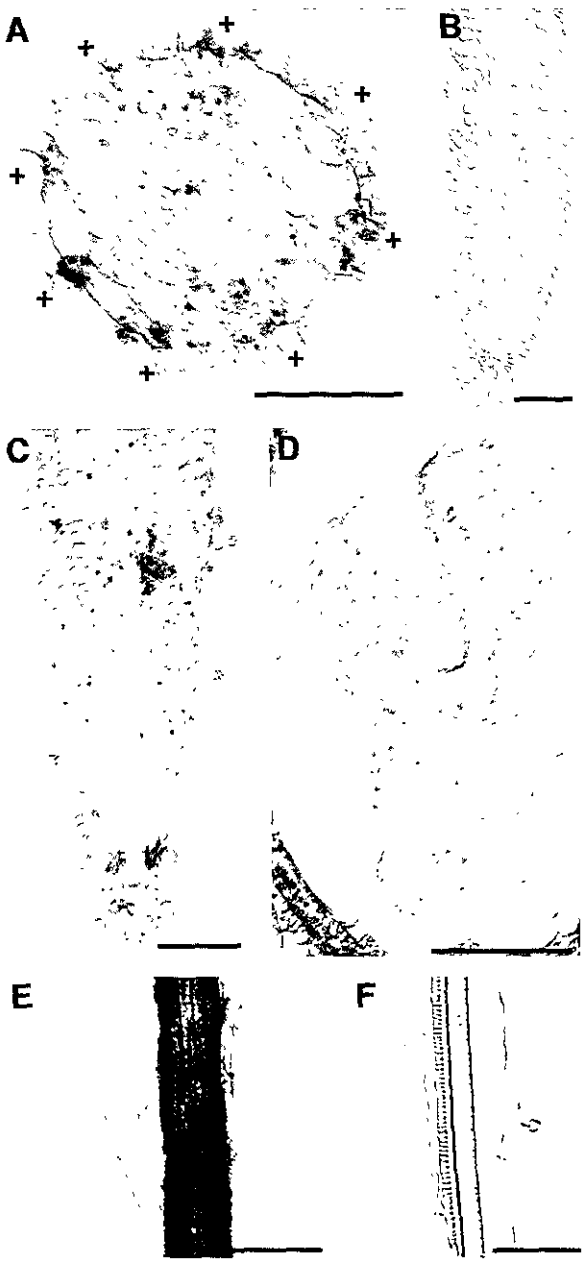

Fig 4a-f. Pattern of mRNA expression of *AGL12* in roots and embryos of *A. thaliana*. In situ hybridization of digoxigenin-labeled probes is shown. **a** Antisense probe hybridization in a transverse section at the basal end of the primary root meristem. Note punctate labeling preferentially in atrichoblasts (+) in the epidermis. **b, c** Sense (**b**) and antisense (**c**) probe hybridizations of longitudinal sections of a primary root tip. **d** Antisense probe hybridization of a longitudinal section of a torpedo-stage embryo. **e,f** Antisense (**e**), and sense (**f**) probe hybridizations of a longitudinal section of a mature root. Bars = 50um