



Anotación de los genes MADS-box en *Dalea cuatrecasii* (Fabaceae) y su comparación con leguminosas modelo

Luis Fernando Canchila Alvarado

Trabajo de grado presentado para optar al título de Biólogo

Asesora

Natalia Pabón Mora, Doctor (PhD) en Biología del desarrollo

Universidad de Antioquia
Facultad de Ciencias Exactas y Naturales
Biología
Caucasia, Antioquia, Colombia
2025

Cita

(Canchila Alvarado, 2025)

Referencia

Canchila Alvarado, L. F. (2025). *Anotación de los genes MADS-box en Dalea cuatrecasasii (Fabaceae) y su comparación con leguminosas modelo* [Trabajo de grado profesional]. Universidad de Antioquia, Cauca, Colombia.

Estilo APA 7 (2020)



Biblioteca Seccional Bajo Cauca (Caucasia)

Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - www.udea.edu.co

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

Dedicatoria

A mis padres, Ludis Alvarado y Medardo Canchila, cuyo apoyo incondicional y presencia constante fueron la base sobre la que pude construir este camino. A mis hermanos, Carolina Andrea, José Medardo y Daniel Eduardo, por ser mi refugio emocional y por darme la mano en cada momento de duda. A mi compañero felino, que, con su presencia en la virtualidad, convirtió largas horas de estudio en momentos más llevaderos. A mis amigas de carrera, Neira Arias, Yalidys Cerpas y Sara Cantero, quienes me acompañaron en toda esta aventura, haciéndola aún más interesante.

Agradecimientos

A la Universidad de Antioquia y al campus Caucaasia, por ser el escenario donde pude desarrollar mi pasión por la biología. Al grupo de investigación Evo-Devo en Plantas, por acogerme y permitirme crecer en este fascinante mundo del estudio vegetal.

A mi asesora, la profesora Natalia Pabón-Mora, por su guía constante, su paciencia y su presencia inquebrantable a lo largo de esta travesía, aun en la virtualidad. A la profesora Yesenia Madrigal Bedoya, por abrirme las puertas al increíble universo de las plantas y sus genes.

Finalmente, un agradecimiento especial a los colaboradores que hicieron posible el estudio de los transcriptomas de *Dalea cuatrecasasii*, Juan Fernando Álzate y Favio González de la Universidad Nacional de Colombia. Su trabajo y conocimientos fueron fundamentales en este proceso.

A todos ustedes, mi más sincera gratitud.

Tabla de contenido

Resumen	8
Abstract	9
Introducción	10
1. Metodología	14
1.1 Criterios de Selección y Curación de Secuencias	15
1.2. Construcción de Árbol Filogenético	16
2. Resultados	17
2.1 Genes MADS-box de <i>Dalea cuatrecasasii</i> identificados en este análisis	18
2.2 Evolución de los genes MADS-box de leguminosas	20
3. Discusión	30
3.1 Funciones de genes MADS-box asociados con cambios fenotípicos conocidos en leguminosas	31
3.2 Genes exclusivos de las inflorescencias, del tallo y compartidos entre los dos puntos en <i>Dalea cuatrecasasii</i>	35
Referencias	38

Lista de figuras

- Figura 1.** Inflorescencias de *Dalea cuatrecasasii* vistas de perfil (A), desde arriba (B) y en vista lateral durante la maduración de frutos (C)..... 11
- Figura 2.** Análisis de Máxima Verosimilitud (ML) de los genes MADS-box de tipo II, incluyendo todos los genes MIKCC; *AGL14*, *AGL15*, *AGL16*, *AGL17*, *AGL18*, *AGL24*, *AGL42*, *AGL6*, *AGL6 sister*, *AGL79*, *AG*, *ANR1*, *AP1*, *AP3*, *ABS*, *FUL*, *PI*, *SEP1*, *SEP3*, *SEP4*, *SHP*, *SOC1*, *STK*, *SVP*..... 17
- Figura 3.** Análisis ML que muestra en detalle los genes *AGL6*, *AGL6 sister*, *SEP1*, *SEP3* y *SEP4* en las leguminosas. Las estrellas amarillas indican eventos de duplicación a gran escala, las estrellas rojas representan duplicaciones especie-específicas 21
- Figura 4.** Análisis ML que muestra en detalle los genes *AP1*, *AGL79* y *FUL* en las leguminosas. Las estrellas amarillas indican eventos de duplicación a gran escala, las estrellas rojas representan duplicaciones especie-específicas 22
- Figura 5.** Análisis ML que muestra en detalle los genes *SVP*, *AGL24*, *AP3*, *TM6* y *PI* en las leguminosas. Las estrellas amarillas indican eventos de duplicación a gran escala, las estrellas rojas representan duplicaciones especie-específicas. 23
- Figura 6.** Análisis ML que muestra en detalle los genes *AGL15*, *AGL16*, *AGL17*, *AGL18* y *ANR1* en las leguminosas. Las estrellas amarillas indican eventos de duplicación a gran escala, las estrellas rojas representan duplicaciones especie-específicas 24
- Figura 7.** Análisis ML que muestra en detalle los genes *ABS* y *XAL1* en las leguminosas. Las estrellas amarillas indican eventos de duplicación a gran escala, las estrellas rojas representan duplicaciones especie-específicas 26
- Figura 8.** Análisis ML que muestra en detalle los genes *AGL42_Like*, *AGL14_Like* y *SOC1* en las leguminosas. Las estrellas amarillas indican eventos de duplicación a gran escala, las estrellas rojas representan duplicaciones especie-específicas de especies 27
- Figura 9.** Análisis ML que muestra en detalle los genes *AG*, *SHP* y *STK* en las leguminosas. Las estrellas amarillas indican eventos de duplicación a gran escala, las estrellas rojas representan duplicaciones especie-específicas 28
- Figura 10.** Alineamiento de la región C-terminal de las proteínas *SHP* identificadas en este análisis. La caja punteada señala la región SSQA identificada en *Medicago truncatula* como una de las más importantes para la lignificación y la torsión de los frutos..... 31
- Figura 11.** Alineamiento de secuencias de proteínas del clado *AGL79* en leguminosas. La copia *DacuAGL79.2* expresada en la inflorescencia termina en el aminoácido 229 y no presenta el motivo (LPPWML) 33

Figura 12. Alineación múltiple de secuencias de proteínas de la familia *AG* en varias especies de leguminosas36

Siglas, acrónimos y abreviaturas

MADS-box - MCM1	Procedente de <i>Saccharomyces cerevisiae</i>
AGAMOUS	De <i>Arabidopsis thaliana</i> e implicado en su desarrollo floral
DEFICIENS	De <i>Antirrhinum majus</i> , donde interviene también en el desarrollo floral
SRF	De <i>Homo sapiens</i>
MICK - M (MADS-box)	Dominio de unión al ADN
I (Intervining)	Región que influye en la especificidad de la dimerización
K (Keratin-like)	Dominio implicado en la estabilidad y formación de dímeros
C (C-termina)	Región que contribuye a la función reguladora y a la interacción con otras proteínas
ML	Máxima Verosimilitud
BLASTN	Basic Local Alignment Search Tool for Nucleotides
API	APETALA1
AP3	APETALA3
PI	PISTILLATA
AG	AGAMOUS
SHP	SHATTERPROOF
STK	SEEDSTICK
SEP1, SEP3, SEP4	SEPALLATA1, 3 y 4
SOC1	SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1
FUL	FRUITFULL
SVP	SHORT VEGETATIVE PHASE
ABS	ABERRANT SEPAL AND PETAL
XALI	XANTHANA-LIKE 1
ANRI	ARABIDOPSIS NITRATE REGULATED 1

Resumen

La familia Fabaceae incluye seis subfamilias, alrededor de 770 géneros y más de 19,500 especies. *Dalea cuatrecasasii*, originaria de Colombia, es el único hospedero sudamericano de la planta parásita *Pilostyles boyacensis*. Los genes MADS-box regulan procesos clave del desarrollo vegetal y se dividen en tipo I y tipo II, siendo los MIKCC parte del segundo grupo. Este estudio identificó por primera vez los genes MADS-box de *D. cuatrecasasii* a partir de transcriptomas de tallo e inflorescencia. Mediante análisis filogenéticos y BLASTN con una base de datos de referencia de *Arabidopsis thaliana*, se identificaron genes expresados en tallo, inflorescencia o en ambas estructuras. Los genes activos en ambas regiones incluyen *AGL14*, *SOC1.1*, *SOC1.2*, *AP3*, *AGL6.1*, *AGL6.2*, *AGL24*, *AGL79.2*, *FUL2*, *SEP4.1* y *SHP1*. En el tallo se expresan exclusivamente *AGL24b*, *AGL42 Like1*, *AGL42 Like2*, *FUL1*, *FUL3* y *SVP*, mientras que en la inflorescencia destacan *AG1*, *AG2*, *AGL6 sister*, *AGL24.1*, *AGL24.2*, *AGL79.1* y varios *SEP* y *PI*. Los resultados sugieren que algunos integradores de la floración podrían ser compartidos entre *D. cuatrecasasii* y su parásita *Pilostyles*, mientras que los genes de identidad floral son exclusivos del hospedero. Además, los análisis filogenéticos indican duplicaciones en Papilionoideae y posibles procesos de subfuncionalización o neofuncionalización, especialmente en los genes *AGAMOUS Like 79* y *SHATTERPROOF*, los cuales presentan diferencias notables con respecto a otras Fabaceae.

Palabras clave: *Dalea cuatrecasasii*, MADS-box, Angiospermas, Fabaceae, Desarrollo floral.

Abstract

The Fabaceae family includes six subfamilies, about 770 genera, and more than 19,500 species. *Dalea cuatrecasasii*, native to Colombia, is the only South American host of the parasitic plant *Pilostyles boyacensis*. MADS-box genes regulate key processes of plant development and are divided into type I and type II, with MIKCC genes being part of the second group. This study identified for the first time MADS-box genes of *D. cuatrecasasii* from stem and inflorescence transcriptomes. Using phylogenetic and BLASTN analyses with a reference database of *Arabidopsis thaliana*, genes expressed in stem, inflorescence, or both structures were identified. Active genes in both regions include *AGL14*, *SOC1.1*, *SOC1.2*, *AP3*, *AGL6.1*, *AGL6.2*, *AGL24*, *AGL79.2*, *FUL2*, *SEP4.1* and *SHP1*. In the stem, *AGL24b*, *AGL42 Like1*, *AGL42 Like2*, *FUL1*, *FUL3* and *SVP* are exclusively expressed, while in the inflorescence, *AG1*, *AG2*, *AGL6 sister*, *AGL24.1*, *AGL24.2*, *AGL79.1* and several *SEPs* and *PIs* stand out. The results suggest that some flowering integrators could be shared between *D. cuatrecasasii* and its parasite *Pilostyles*, while floral identity genes are exclusive to the host. Furthermore, phylogenetic analyses indicate duplications in Papilionoideae and possible subfunctionalization or neofunctionalization processes, especially in the *AGAMOUS Like 79* and *SHATTERPROOF* genes, which present notable differences with respect to other Fabaceae.

Keywords: *Dalea cuatrecasasii*, MADS-box, Angiosperms, Fabaceae, Floral development.

Introducción

Las leguminosas: una familia de angiospermas de diversidad extraordinaria

La familia Fabaceae o Leguminosae consta de seis subfamilias, alrededor de 770 géneros y más de 19.500 especies (Lewis et al., 2005, 2013; LPWG, 2013a). Ésta es la tercera familia de angiospermas en términos de número de especies, después de Asteraceae y Orchidaceae. Esta diversidad no solo se refleja en su amplia distribución geográfica, que abarca desde las zonas templadas hasta las tropicales de todo el mundo, sino también en su importancia ecológica, especialmente en ecosistemas áridos y semiáridos. Actualmente la familia está circunscrita en seis subfamilias: Cercidoioideae (12 géneros y ca. 335 especies), Detarioideae (84 géneros y ca. 760 especies), Duparquetioideae (1 género y 1 especie), Dialioideae (17 géneros y ca. 85 especies), Caesalpinioideae (148 géneros y ca. 4400 especies) y finalmente la subfamilia Papilionoideae (503 géneros y ca. 14,000 especies) (Azani et al., 2017). Dentro de la subfamilia Papilionoideae, la tribu Amorphaeae es notable por su particularidad morfológica y ecológica, compuesta por ocho géneros y aproximadamente 248 especies (Lock, 2005). El género *Dalea*, el más grande de esta tribu con aproximadamente 173 especies, presenta una notable diversidad morfológica y ecológica, siendo especialmente abundante en hábitats secos de las Américas, desde Canadá hasta Argentina y Chile (Barneby, 1977).

Con aproximadamente 173 especies, *Dalea* es el género más grande de la tribu Amorphaeae (Barneby 1977, 1988; Estrada et al., 2011; Piñeros-U & González 2019). El género se distribuye principalmente en hábitats secos de las Américas, desde el sur de Canadá hasta el norte de Chile y el noroeste de Argentina, con algunas especies llegando a las Antillas y una reportada en las Islas Galápagos (Barneby, 1977). Las especies de *Dalea* de América del Norte se mantienen bajo evaluación y protección ecológica permanente por su papel en la restauración del hábitat y la protección de los santuarios de supervivencia de las abejas (Piñeros Urrego et al., 2022), pero es probable que se desempeñen funciones similares en otros lugares. Las especies se distribuyen principalmente en entornos secos a lo largo de América, con la mayor concentración de diversidad biológica en México y la región Andina, que se extiende desde Colombia hasta el norte de Argentina y Chile (Barneby, 1977).

En Colombia, cinco especies de *Dalea* se encuentran en los valles interandinos secos y la región del Caribe. Dos de ellas (*D. cuatrecasasii* Barneby y *D. wilsonii* Piñeros-U. & F. González)

son endémicas (Piñeros-U & González 2019, 2020). Dos especies de *Dalea*, a saber, *D. coerulea* L. fil. y *D. cuatrecasasii*, son parcialmente simpátricas en los Andes centrales colombianos, donde florecen profusamente (Piñeros-U & González 2019, 2020). Ambas tienen vistosas flores papilionáceas y decandras que brindan abundantes recompensas florales (a través del néctar y las secreciones glandulares) principalmente a varias especies de himenópteros (Fig.1). Los análisis filogenéticos preliminares sugieren que estas dos especies no están estrechamente relacionadas, ya que *D. cuatrecasasii* pertenece a un clado con especies de América Central y del Norte, mientras que *D. coerulea* se agrupa con otras especies de América del Sur (Piñeros Urrego et al., 2022). *D. cuatrecasasii* se encuentra en matorrales secos de los departamentos de Boyacá y Santander, y es abundante en pastizales y zonas intervenidas en la cuenca del río Chicamocha. Recientemente, se ha reportado que *Dalea cuatrecasasii* es el único hospedero de otra planta con flor, *Pilostyles boyacensis* (Apodanthaceae) en América del Sur (González & Pabón-Mora, 2014a, 2014b). Esta asociación de parasitismo obligada se ha encontrado en poblaciones cerca de Gámeza, Soatá, Tasco y Villa de Leyva (Boyacá), y en Nemocón (Cundinamarca). Esta interacción es también una oportunidad única para estudiar las adaptaciones moleculares y evolutivas que subyacen a esta relación, particularmente en lo que respecta a los genes que regulan el desarrollo floral.

Figura 1. Inflorescencias de *Dalea cuatrecasasii* vistas de perfil (A), desde arriba (B) y en vista lateral durante la maduración de frutos (C).



Nota. Pabón-Mora (2021). Inflorescencias de *Dalea cuatrecasasii* vistas de perfil (A), desde arriba (B) y en vista lateral durante la maduración de frutos (C).

Los genes MADS-box: importantes en todos los procesos de desarrollo en plantas

Los genes MADS-box son factores de transcripción presentes en plantas, animales y hongos, lo que indica que son muy antiguos evolutivamente. Todos actúan reconociendo e

interactuando con los dominios CArG en elementos en *cis* de los genes a los que regulan. Los genes MADS-box se clasifican en dos grandes grupos: tipo I y tipo II. Los tipos I se subdividen en genes M α , M β , Mr, mientras que los tipos II se subdividen en genes MIKCC y MIKC* (Riechmann et al., 1996). Los genes MADS-box controlan una gran cantidad de importantes transiciones y procesos del desarrollo en plantas, entre los que se encuentran la germinación, el mantenimiento del meristema vegetativo, la transición floral, la identidad de órganos florales, la histogénesis del fruto y la dispersión de las semillas, por mencionar algunas (Smaczniak et al., 2012). Pese a los avances en la comprensión de la función y evolución de los genes MADS-box en diversos linajes de plantas, existe una brecha significativa en el conocimiento de estos genes en la familia Fabaceae. Las leguminosas, a pesar de su importancia ecológica, económica y su diversidad morfológica, han sido menos estudiadas en términos de la evolución y diversificación de sus genes MADS-box, especialmente aquellos relacionados con la evolución de características únicas y procesos de domesticación en este grupo de angiospermas. Este vacío en el conocimiento es particularmente relevante dado el papel crucial que juegan estos genes en la determinación de la morfología floral, un aspecto clave en la adaptación ecológica y la especiación en plantas con flores. A la fecha los genes MADS-box se han aislado de forma sistemática de *Arachis hypogaea* (el maní) (Mou et al., 2022), *Cajanus cajan* (el guandúl) (Kumar et al., 2021), *Glycine max* (la soya) (Shu et al., 2013), *Medicago sativa* (la alfalfa) (Dong et al., 2021), y *Phaseolus vulgaris* (el frijol) (Okay et al., 2024). Así mismo, los genes MADS-box han sido estudiados en referencia con la simbiosis entre leguminosas y bacterias fijadoras de nitrógeno (Ayra et al., 2021). Con frecuencia, la actividad de algunos genes MADS-box ha estado incluso asociada a procesos de domesticación, lo que es particularmente interesante de analizar en las leguminosas (Schilling et al., 2018). Finalmente, algunos MADS-box han sido incluso vinculados con resistencia al virus del mosaico de la soya (Ren et al., 2022).

En este contexto, la presente investigación se propone: 1) identificar y caracterizar los genes MADS-box en los transcriptomas disponibles de *Dalea cuatrecasasii*, que provienen de tallos y de inflorescencias, y 2) evaluar su evolución en el contexto de los genes MADS-box disponibles de genomas de Fabaceae, como *Arachis hypogaea* (maní), *Aeschynomene evenia*, *Cercis canadensis* (árbol del amor), *Cicer arietinum* (garbanzo), *Glycine max* (soya), *Lens culinaris* (lenteja), *Lotus japonicus*, *Lupinus albus* (lupino), *Medicago truncatula*, *Phaseolus lunatus* (frijol lima), *Phaseolus vulgaris* (frijol común) y *Trifolium pratense* (trébol rojo). Estos análisis filogenéticos permitirán

identificar posibles eventos de duplicación y diversificación de los genes MADS-box en diferentes linajes de leguminosas, proporcionando una visión más completa de la evolución de los genes que pueden estar controlando numerosos procesos del desarrollo. Finalmente se evaluarán en detalle los genes presentes en el tallo de *Dalea cuatrecasasii*, y se discutirá la posible co-opción de los mismos por parte de *Pilostyles boyacensis*, lo que podría ofrecer nuevas perspectivas sobre las interacciones moleculares entre plantas parásitas y sus hospederas.

1. Metodología

Secuencias originales de genes MADS-box y búsquedas dirigidas en genomas y transcriptomas de interés

Con la intención de aislar genes MADS-box de la especie de interés, *Dalea cuatrecasasii* y de leguminosas que pudieran servir como referencia, es decir de aquellas con el genoma secuenciado se construyó una base de datos inicial de genes objetivo de *Arabidopsis thaliana*, que incluye todos los genes MIKC* con los siguientes genes: *AGAMOUS* (*AG*), *AGAMOUS Like6* (*AGL6*), *AGAMOUS Like15* (*AGL15*), *AGAMOUS Like16* (*AGL16*), *AGAMOUS Like17* (*AGL17*), *AGAMOUS Like18* (*AGL18*), *AGAMOUS Like19* (*AGL19*), *AGAMOUS Like23* (*AGL23*), *AGAMOUS Like24* (*AGL24*), *AGAMOUS Like28* (*AGL28*), *AGAMOUS Like37* (*PHERES1*), *AGAMOUS Like42* (*AGL42*), *AGAMOUS Like61* (*DIANA*, *DIA*), *AGAMOUS Like62* (*AGL62*), *AGAMOUS Like65* (*AGL65*), *AGAMOUS Like66* (*AGL66*), *AGAMOUS Like71* (*AGL71*), *AGAMOUS Like72* (*AGL72*), *AGAMOUS Like80* (*AGL80*), *APETALA1* (*AP1*), *APETALA3* (*AP3*), *ARABIDOPSIS BSISTER* (*ABS*), *ARABIDOPSIS NITRATE REGULATED 1* (*ANR1*), *CAULIFLOWER* (*CAL*), *DIANA* (*AGAMOUS Like61*) (*DIA*), *FLOWERING LOCUS C* (*FLC*), *FRUITFULL* (*FUL*), *GORDITA* (*GOA*), *MADS AFFECTING FLOWERING 1-4* (*MAF1-4*), *MADS AFFECTING FLOWERING 5* (*MAF5*), *PISTILLATA* (*PI*), *SEPALLATA1-4* (*SEP1-4*), *SHATTERPROOF 1* (*SHP1*), *SHATTERPROOF 2* (*SHP2*), *SHORT VEGETATIVE PHASE* (*SVP*), *SEEDSTICK* (*STK*), *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOCI*), *XAANTAL1* (*XAL1*).

Estas secuencias sirvieron para consultar las bases de datos de especies de leguminosas selectas, con énfasis en aquellas que disponen de información en repositorios públicos como Phytozome (<https://phytozome-next.jgi.doe.gov>) y Legume Information System (LIS - <https://www.legumeinfo.org>). Las especies seleccionadas fueron: *Arachis hypogaea* (maní), *Aeschynomene evenia*, *Cercis canadensis* (árbol del amor), *Cicer arietinum* (garbanzo), *Glycine max* (soja), *Lens culinaris* (lenteja), *Lotus japonicus*, *Lupinus albus* (lupino), *Medicago truncatula*, *Phaseolus lunatus* (frijol lima), *Phaseolus vulgaris* (frijol común) y *Trifolium pratense* (trébol rojo). Así, se realizó una búsqueda de los homólogos de estos genes en las leguminosas con genomas secuenciados disponibles en Phytozome y en el Legume Information System (LIS). Todas las secuencias resultantes fueron verificadas en NCBI (<https://www.ncbi.nlm.nih.gov>) para

confirmar que correspondieran a genes MADS-box y realizar una anotación preliminar de las secuencias aisladas.

Los mismos genes canónicos de *Arabidopsis thaliana* fueron empleados como punto de consulta para buscar genes homólogos en los transcriptomas de *Dalea cuatrecasasii* utilizando BLASTN. Se analizaron dos transcriptomas: uno de flor ("DaleaFLO") y otro de tallo no infectado ("DNI"), recolectados en ausencia del holoparásito *Pilostyles boyacensis*. Las búsquedas se realizaron con valores de corte de *e-value* de 1×10^{-5} y 1×10^{-30} . Las secuencias obtenidas fueron curadas para asegurar la presencia de dominios proteicos MADS-box y limpiar las secuencias eliminando las regiones no transcritas del 5' y del 3'.

1.1 Criterios de Selección y Curación de Secuencias

Una vez obtenidas las secuencias de las leguminosas seleccionadas y de *Dalea cuatrecasasii*, se llevó a cabo un proceso de curación detallado con el objetivo de asegurar la calidad y relevancia de las secuencias para los análisis filogenéticos subsecuentes. Este proceso se diseñó para depurar secuencias redundantes, incompletas o no representativas, manteniendo sólo aquellas que cumplieran con criterios estrictos de calidad. Los pasos seguidos fueron:

- **Prioridad de Secuencias:** En caso de encontrar secuencias redundantes, es decir, secuencias idénticas o casi idénticas en una misma especie (que a veces provenían de bases de datos diferentes), se seleccionó una única secuencia. Se priorizaron las secuencias obtenidas del repositorio Phytozome sobre las de [LIS](#) en caso de tener contigs repetidos, ya que Phytozome tiende a tener secuencias mejor anotadas, con posiciones cromosómicas disponibles y en varios casos más completas.
- **Longitud de Secuencias:** Se estableció un umbral de longitud mínimo de 150 a 200 aminoácidos para las secuencias de proteínas (450-600 nucleótidos), con el fin de asegurar que sólo se incluyeran secuencias completas y funcionalmente relevantes. Secuencias más cortas fueron eliminadas, ya que podrían representar fragmentos o errores de ensamblaje.
- **Integridad estructural:** La selección de secuencias también priorizó las secuencias que mostraran la mayor integridad estructural. Por ejemplo, se revisó que la secuencia aminoacídica comenzando preferentemente con los motivos "MGRG" o "MVRG", lo cual es indicativo de secuencias completas y funcionales.

- **Eliminación de Secuencias Truncadas:** Secuencias incompletas o que carecían de dominios funcionales importantes reportados fueron eliminadas si se disponía de versiones más largas y completas. Esto fue especialmente importante para evitar errores en el alineamiento y en la construcción de árboles filogenéticos.
- **Curación de Secuencias de *Dalea cuatrecasasii*:** Se aplicaron los mismos criterios de curación a las secuencias de *Dalea cuatrecasasii*, tanto de los transcriptomas de flor ("DaleaFLO") como de tallo no infectado ("DNI"). Además, se verificó la conservación de los dominios MADS-box, esenciales para la correcta clasificación funcional de los genes. Las secuencias fueron curadas para eliminar las regiones UTR.

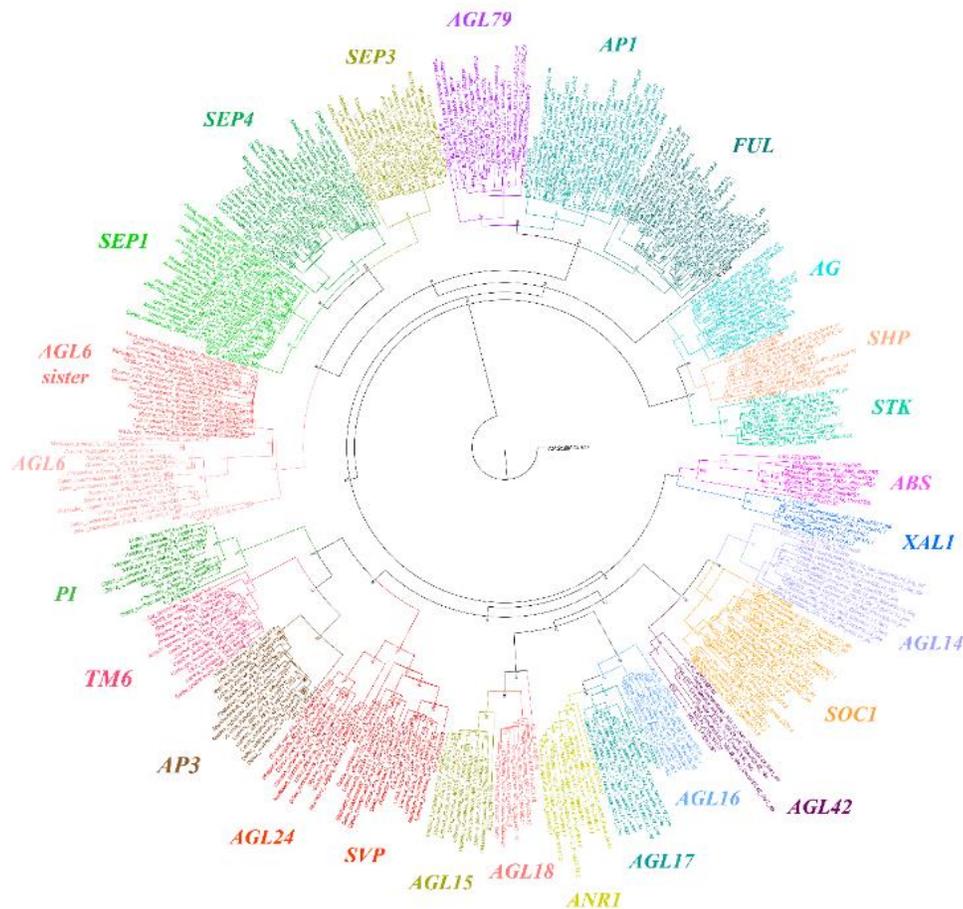
1.2. Construcción de Árbol Filogenético

Todos los homólogos de los genes MADS-box de *Arabidopsis thaliana*, las especies modelo de leguminosas, y *Dalea cuatrecasasii* fueron incluidas en una misma matriz y alineadas utilizando [MAFFT](https://mafft.cbrc.jp/alignment/server/index.html) (https://mafft.cbrc.jp/alignment/server/index.html) (Katoh y Standley, 2013), aplicando una penalización por apertura de brechas de 5.0 y un valor de desplazamiento de 1.0. La hipótesis filogenética se construyó mediante el método de Máxima Verosimilitud (ML) utilizando [IQ-TREE](http://www.iqtree.org) (http://www.iqtree.org) (Trifinopoulos et al., 2016). Se seleccionó el modelo de evolución molecular más adecuado para los datos con ModelFinder, integrado en IQ-TREE (Kalyaanamoorthy et al., 2017). El soporte de las ramas se determinó mediante Ultrafast Bootstrap (UFBoot) con 1000 pseudo-réplicas (Hoang et al., 2018). Los árboles filogenéticos resultantes fueron visualizados, editados y coloreados con [FigTree v1.4.3](http://tree.bio.ed.ac.uk/software/figtree/) (http://tree.bio.ed.ac.uk/software/figtree/), utilizando *Arabidopsis thaliana* (*AtAGL66*) como grupo externo.

2. Resultados

Se realizó una búsqueda exhaustiva en bases de datos de referencia disponibles, basada en genes MIKCC de *Arabidopsis thaliana*, y se llevó a cabo un análisis BLASTN en los transcriptomas, lo que permitió aislar 367 secuencias. De estas, 36 correspondieron a genes MADS-box de tipo II del subgrupo MIKCC, mientras que las 331 secuencias restantes pertenecieron a diversas leguminosas, distribuidas de la siguiente manera: *Aeschynomene evenia* (10), *Arachis hypogaea* (21), *Cercis canadensis* (26), *Cicer arietinum* (5), *Dalea cuatrecasasii* (47), *Glycine max* (55), *Lens culinaris* (21), *Lotus japonicus* (24), *Lupinus albus* (38), *Medicago truncatula* (31), *Phaseolus vulgaris* (35) y *Trifolium pratense* (18) (Figura 2).

Figura 2. Análisis de Máxima Verosimilitud (ML) de los genes MADS-box de tipo II, incluyendo todos los genes MIKCC; AGL14, AGL15, AGL16, AGL17, AGL18, AGL24, AGL42, AGL6, AGL6 sister, AGL79, AG, ANR1, AP1, AP3, ABS, FUL, PI, SEP1, SEP3, SEP4, SHP, SOC1, STK, SVP



Nota. Los números sobre las ramas indican los valores de Ultrafast Bootstrap (UFBS).

2.1 Genes MADS box de *Dalea cuatrecasasii* identificados en este análisis

De la búsqueda dirigida de los genes MADS-box en los dos transcriptomas disponibles, de tallo no infectado e inflorescencia de *Dalea cuatrecasasii*, se rescatan los siguientes homólogos.

Genes integradores de la floración: De los genes MADS-box que integran las señales de la floración (i.e. vernalización, fotoperiodo, edad, ruta hormonal) se encontraron homólogos de *AGAMOUS-like 24 (AGL24)*, *SHORT VEGETATIVE PHASE (SVP)*, y *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS (SOC1)*. De los genes homólogos a *AGL24*, que han sido vinculados en especies modelo a funciones de promoción de la floración, se encontraron en *Dalea cuatrecasasii* tres copias, *DacuAGL24*, *DacuAGL24.1* y *DacuAGL24.2*. La copia *DacuAGL24* se encontró en el transcriptoma del tallo, mientras que las otras dos copias se aislaron del transcriptoma de inflorescencia. De los genes homólogos a *SVP*, que han sido asociados a represión de la floración, se aisló un solo homólogo, *DacuSVP*, del transcriptoma del tallo. Finalmente, de los genes homólogos a *SOC1*, posibles promotores de la floración, se aislaron dos copias, *DacuSOC1.1* y *DacuSOC1.2*, las dos expresadas activamente en tallos e inflorescencias (Figura 2, 5, 8).

De los demás genes MADS-box integradores de la ruta de floración que incluyen *AGAMOUS Like17 (AGL17)*, *AGAMOUS Like18 (AGL18)* y *AGAMOUS Like42 (AGL42)* solamente fueron aisladas dos copias del clado *AGL42*, *DacuAGL42 Like1* y *DacuAGL42 Like2* del transcriptoma del tallo. No se encontraron homólogos de los clados *AGL17* ni *AGL18* (Figura 2, 6, 8).

Genes responsables de la identidad de órganos florales: De los genes *APETALA1 (API)*, que han sido asociados a la identidad de los sépalos y pétalos en *Arabidopsis*, se aislaron dos copias activamente expresadas en la inflorescencia, estas son *DacuAPI Like1* y *DacuAPI Like2* (Figura 2, 4). Por otra parte, de los genes *APETALA3 (AP3)*, *PISTILLATA (PI)* y *TOMATO MADS BOX GENE 6 (TM6)*, que son todos genes MADS-box, clase B, y que han sido asociados a la formación de los pétalos y estambres en *Arabidopsis* y en *Solanum* fueron aislados varios homólogos. A saber, de los genes homólogos a *AP3*, se encontraron en *Dalea cuatrecasasii* una copia, *DacuAP3*, tanto en el transcriptoma de tallo e inflorescencia. De los genes homólogos a *PI*, se aislaron tres copias, *DacuPI1*, *DacuPI2* y *DacuPI3*. Las tres copias de *DacuPI* se encontraron en el transcriptoma de inflorescencia. Finalmente, de los genes homólogos a *TM6*, se aisló un solo homólogo, *DacuTM6*, del transcriptoma de inflorescencia (Figura 2, 5).

De los genes *AGAMOUS* (*AG*), que son genes MADS-box clase C, y que han sido asociados a la identidad de los estambres y carpelos en *Arabidopsis*, se aislaron dos copias activamente expresadas en la inflorescencia, estas son *DacuAG1* y *DacuAG2*. Por su parte de genes homólogos a *SEEDSTICK* (*STK*), genes clase D, asociados con el desarrollo de óvulos, fue identificado un solo homólogo, *DacuSTK*, del transcriptoma de inflorescencia (Figura 2, 8).

Continuando con la búsqueda de los genes MADS-box vinculados a la correcta identidad de todos los órganos florales, a saber, sépalos, pétalos, estambres y carpelos en *Arabidopsis thaliana*, se identificaron homólogos *SEPALLATA 1, 2, 3 y 4* y *AGAMOUS Like 6* (*AGL6*). De los genes *SEP1*, se encontraron en *D. cuatrecasasii* dos copias, *DacuSEP1 Like1* y *DacuSEP1 Like2*, las cuales se aislaron del transcriptoma de inflorescencia. De los genes homólogos a *SEP3*, se aisló un solo ortólogo, *DacuSEP3*, del transcriptoma de inflorescencia. En cuanto a los genes homólogos de *SEP4*, se encontraron tres copias, *DacuSEP4.1*, *DacuSEP4.2* y *DacuSEP4.3*. La copia *DacuSEP4.1* se encontró en el transcriptoma de tallo e inflorescencia, mientras que las otras dos copias se aislaron del transcriptoma de inflorescencia. Mientras, de los genes homólogos a *AGL6*, se aislaron dos copias activamente expresadas en el tallo e inflorescencia, estas son *DacuAGL6.1* y *DacuAGL6.2*. Finalmente, de los genes homólogos a *AGL6 sister*, se aisló un solo homólogo, *DacuAGL6 sister*, del transcriptoma de inflorescencia (Figura 2, 3).

Genes de desarrollo de frutos: De los genes MADS-box que integran el desarrollo de órganos reproductivos y la formación de frutos en las plantas, se encontraron homólogos de *FRUITFULL* (*FUL*) y *SHATTERPROOF* (*SHP*). De los homólogos a *FUL*, se identificaron cuatro copias en *D. cuatrecasasii*: *DacuFUL1* y *DacuFUL3*, ambas presentes en el transcriptoma de tallo, mientras que una variante de *DacuFUL1* está en el transcriptoma de inflorescencia. Finalmente, la copia *DacuFUL2* se encontró activamente expresada tanto en el transcriptoma de tallo como en el de inflorescencia. De los genes homólogos a *SHP*, se encontraron dos copias, una activamente expresada en tallo y en inflorescencia la cual es *DacuSH1*, En cambio, solamente expresada en el transcriptoma de inflorescencia *DacuSHP2* (Figura 2, 4, 8).

Genes responsables de otros procesos del desarrollo: De los genes *AGAMOUS Like79* (*AGL79*), los cuales están implicados en la regulación del desarrollo de la flor y arquitectura de la inflorescencia, se aislaron dos copias de *D. cuatrecasasii*. *DacuAGL79.2*, la cual representa el transcriptoma de tallo e inflorescencia. La copia de *DacuAGL79.1* se encontró en el transcriptoma de inflorescencia (Figura 2, 4). De los genes *ARABIDOPSIS B SISTER* (*ABS*), los cuales han sido

asociados a la pigmentación de las semillas y desarrollo del endotelio, no se encontró ningún homólogo (Figura 2, 7).

De los genes *XAANTALI* (*XALI*), que han sido asociados a la regulación del ciclo celular del desarrollo radicular; transición a la floración (activador), se aisló una copia activamente expresada en el tallo, nombrada *DacuXALI* (Figura 2, 7).

De los genes *AGAMOUS Like14* (*AGL14*), los cuales se expresan preferentemente en los tejidos radiculares, por tanto, representan los únicos genes MADS-box caracterizados que se expresan en las raíces. Se encontró una sola copia activamente expresada en tallo y en inflorescencia la cual es *DacuAGL14 Like* (Figura 2, 8).

De los genes *AGAMOUS Like15* (*AGL15*), *AGAMOUS Like16* (*AGL16*) y *ARABIDOPSIS NITRATE REGULATED1* (*ANRI*), no se encontraron homólogos en *Dalea cuatrecasasii* (Figura 2, 6). Estos genes han sido vinculados a diversos procesos del desarrollo. Por ejemplo, los genes *AGL16* está relacionado con funciones en número y distribución de estomas (Kutter et al., 2007). Finalmente, *AGL15* tiene función relacionada a embriogénesis y transición a la floración (Heck et al., 1995; Fernandez et al., 2000; Harding et al., 2003).

2.2 Evolución de los genes MADS-box de leguminosas

El análisis aquí realizado permite no solo identificar los homólogos de los genes MIKCC de *Dalea cuatrecasasii*, sino también explorar la evolución de los distintos clados de genes en la familia de las leguminosas. A continuación, presentaremos las duplicaciones identificadas clado por clado siguiendo la filogenia de Zhao et al (2021).

Los genes *AGL6* y *SEP* (*SEP1* / 2, 3 y 4) se recuperan con altos soportes. Los genes homólogos de *AGL6* y *AGL13* de *A. thaliana* han sufrido una duplicación en las leguminosas (UFBS = 63), de la que surgen los clados *AGL6* (UFBS = 100) y *AGL6 sister* (UFBS = 99). En el clado *AGL6*, se observa una duplicación a gran escala (UFBS = 100), previa a la diversificación de las Papilionoideae. También se detectan dos duplicaciones intraespecíficas en *Glycine max* (Figura 3). En el clado *AGL6 sister*, se evidencia una duplicación a gran escala (UFBS = 57), que posiblemente predata la diversificación de todas las leguminosas. Además, hay duplicaciones locales en *Glycine max* y *Lupinus albus*.

En el clado de los *SEP1_Like* (UFBS = 85), se observa una duplicación exclusiva de las Papilionoideae (UFBS = 84) y dos duplicaciones locales en *Glycine max* y *Lupinus albus*. En el clado *SEP4* (UFBS = 98), se puede observar una duplicación previa a todas las leguminosas, que

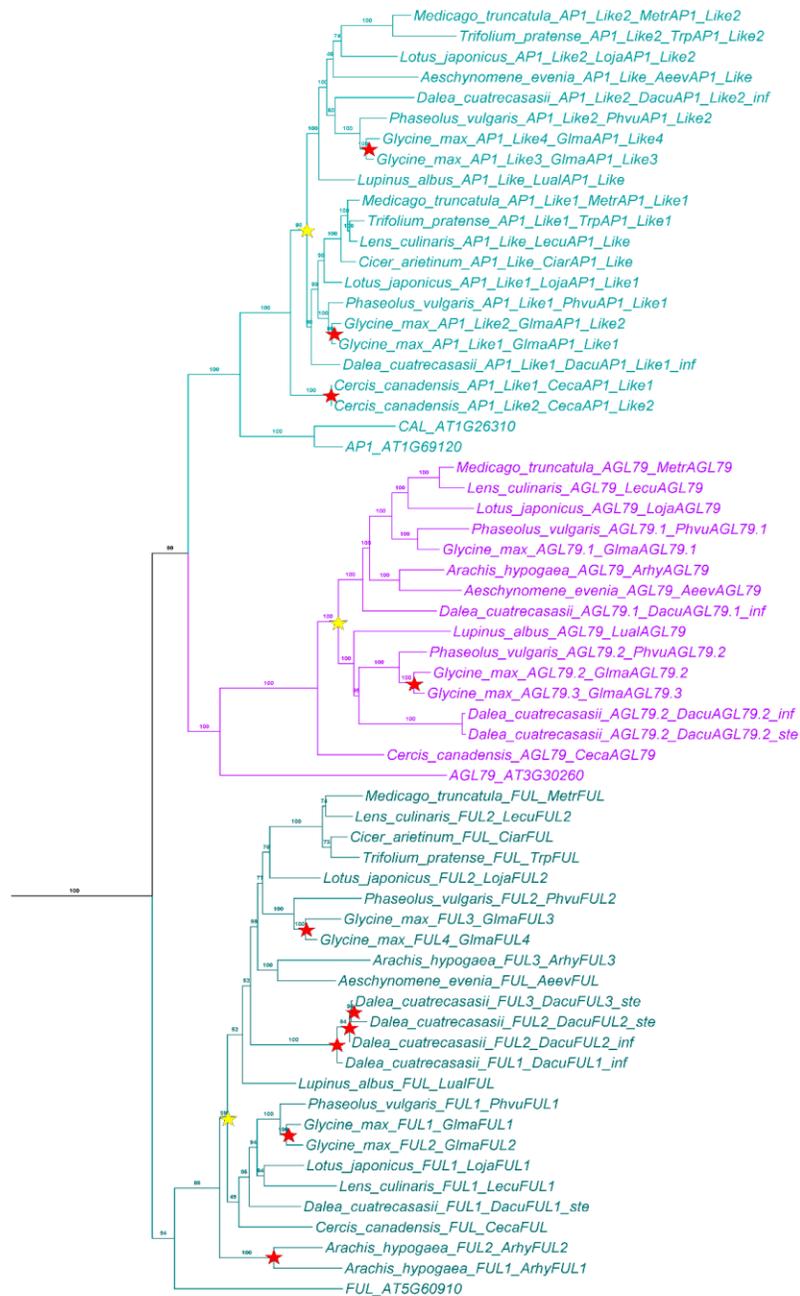
afecta tanto a la subfamilia Papilionoideae como a la subfamilia Cercidoideae. Debido a que incluso el homólogo *SEP4* de *Arabidopsis*, se anida en uno de los clados es posible que esta duplicación afecte otras eudicotiledóneas centrales. Además, se identifican cuatro duplicaciones locales en *Glycine max*, *Lupinus albus* y *Cercis canadensis*. Por último, los genes *SEP3* (UFBS = 100), también sufren una duplicación que afecta a todas las leguminosas (UFBS = 92), además de una duplicación local en *Glycine max* (Figura 3).

Figura 3. Análisis ML que muestra en detalle los genes *AGL6*, *AGL6 sister*, *SEP1*, *SEP3* y *SEP4* en las leguminosas. Las estrellas amarillas indican eventos de duplicación a gran escala, las estrellas rojas representan duplicaciones especie-específicas



Nota. Los números en cada rama indican los valores de Ultrafast Bootstrap (UFBS).

Figura 4. Análisis ML que muestra en detalle los genes *API*, *AGL79* y *FUL* en las leguminosas. Las estrellas amarillas indican eventos de duplicación a gran escala, las estrellas rojas representan duplicaciones especie-específicas

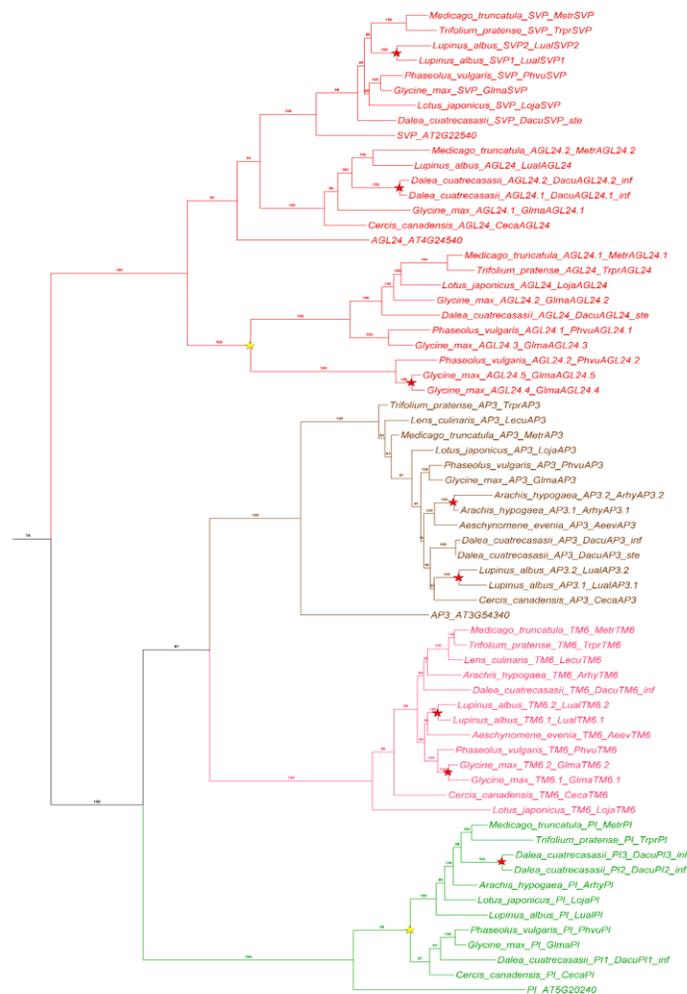


Nota. Los números en cada rama indican los valores de Ultrafast Bootstrap (UFBS).

Los clados del linaje de genes *API/FUL*, *euAPI*, *AGL79* y *FUL* se recuperan con altos soportes. En el clado *API* (UFBS = 100) se evidencia una duplicación específica de las

Papilionoideae (UFBS = 98), y tres duplicaciones locales, dos en *Glycine max* y una en *Cercis canadensis*. En el clado *AGL79* (UFBS =100), se identifica una duplicación idéntica en las Papilionoideae (UFBS = 100), así como una duplicación local en *Glycine max*. Finalmente, en el clado *FUL* (UFBS = 94), se observa una duplicación a gran escala (UFBS = 59), posiblemente afectando todas las leguminosas, con la presencia adicional de seis duplicaciones locales, dos en *Glycine max*, una en *Arachis hypogaea* y tres en *Dalea cuatrecasasii* (Figura 4).

Figura 5. Análisis ML que muestra en detalle los genes SVP, AGL24, AP3, TM6 y PI en las leguminosas. Las estrellas amarillas indican eventos de duplicación a gran escala, las estrellas rojas representan duplicaciones especie-específicas.

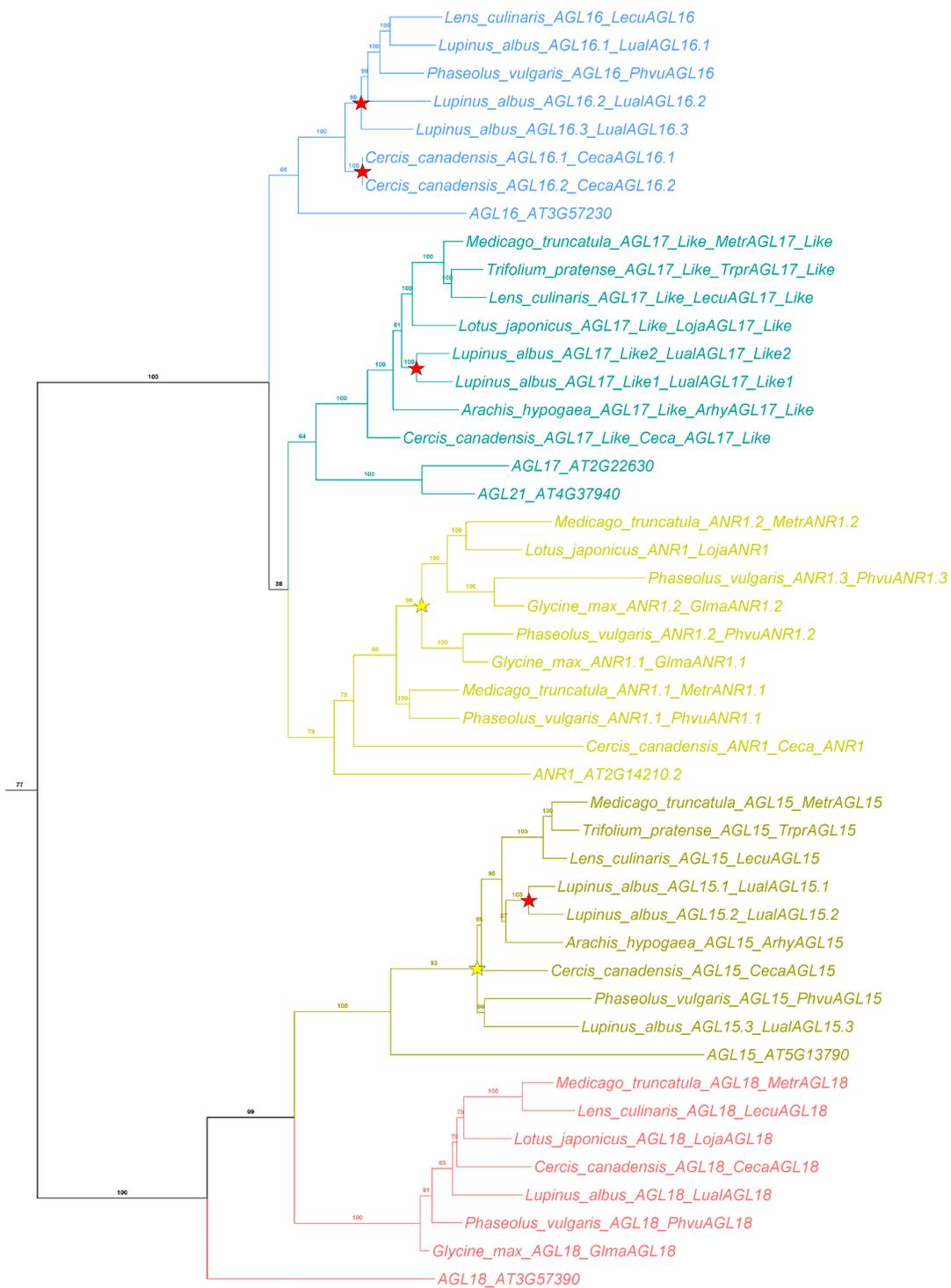


Nota. Los números en cada rama indican los valores de Ultrafast Bootstrap (UFBS).

Los clados *SVP*, *AGL24A* y *AGL24B* (Ramírez-Ramírez et al., 2021) se recuperan aquí parcialmente. El clado *SVP* (UFBS = 100), se recupera predominantemente de copia única con solo una duplicación local en *Lupinus albus*. El clado *AGL24A* (UFBS = 100), se recupera sin el homólogo de *Arabidopsis* (*AGL24*), también se mantiene de copia única, y solo se detecta una duplicación local en *Dalea cuatrecasasii*. Finalmente, el clado *AGL24B* (UFBS = 100) se rescata sin homólogos de *Arabidopsis* y con una duplicación a gran escala (UFBS = 100), que parece solo retenerse en la tribu Phaseoleae (que incluye a *Phaseolus* y *Glycine*). Además, se detectan dos duplicaciones locales en *Glycine max* y *Dalea cuatrecasasii* (inflorescencia) (Figura 5).

En cuanto a los clados genes clase B, se recuperan los tres clados de genes *AP3*, *TM6* y *PI* (todos con UFBS = 100). Los genes *AP3* y *TM6* no revelan duplicaciones a gran escala, solo duplicaciones locales de *TM6* en *Lupinus albus* y *Glycine max*, así como duplicaciones locales de *AP3* en *Arachis hypogaea* y *Glycine max*. Finalmente, en el clado *PI*, se identificó una duplicación posiblemente afectando a todas las leguminosas (UFBS = 95), o que al menos predata la diversificación de las Dalbergioides (*Dalea cuatrecasasii*) y una única duplicación local en *Dalea cuatrecasasii* (Figura 5).

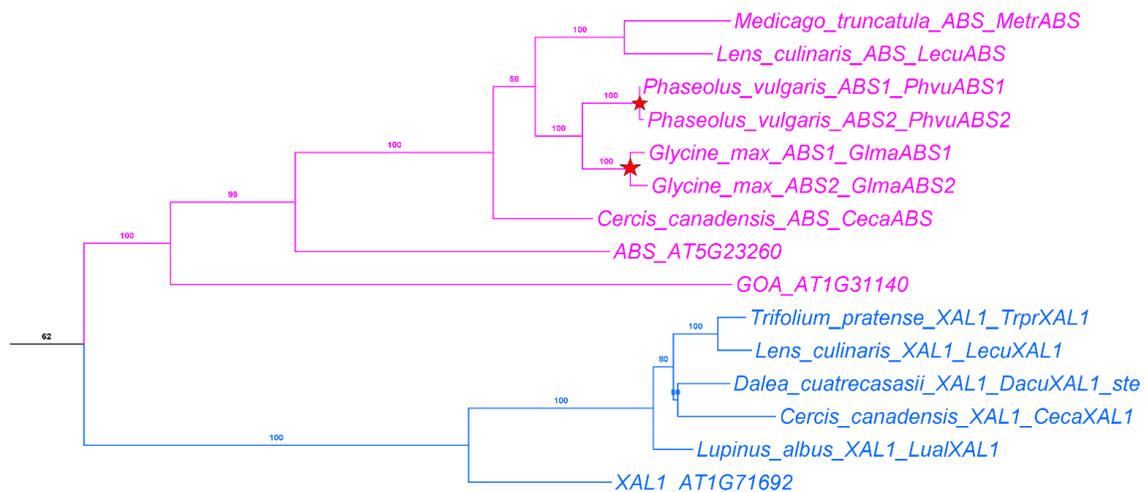
Figura 6. Análisis ML que muestra en detalle los genes *AGL15*, *AGL16*, *AGL17*, *AGL18* y *ANR1* en las leguminosas. Las estrellas amarillas indican eventos de duplicación a gran escala, las estrellas rojas representan duplicaciones especie-específicas



Nota. Los números en cada rama indican los valores de Ultrafast Bootstrap (UFBS).

En el clado *AGL16* (UFBS = 100) se identifican duplicaciones locales en *Cercis canadensis* y *Lupinus albus*. En el clado *AGL17 Like* (UFBS = 100), también hay una duplicación local en *Lupinus albus*. En el clado *ANR1* (UFBS = 38), se identifica una duplicación a gran escala (UFBS = 99), que parece ocurrir antes de la diversificación de las tribus Phaseolae, Loteae, Fabeae y Trifolieae, pero que solo se retiene en las Phaseolae (*Phaseolus* y *Glycine*). En cuanto clado *AGL15* (UFBS = 99), se encuentra una duplicación a gran escala (UFBS = 93), difícil de temporizar, junto con una duplicación local en *Lupinus albus*. Finalmente, el clado *AGL18* (UFBS = 99) se mantiene de copia única en todas las leguminosas muestreadas (Figura 6).

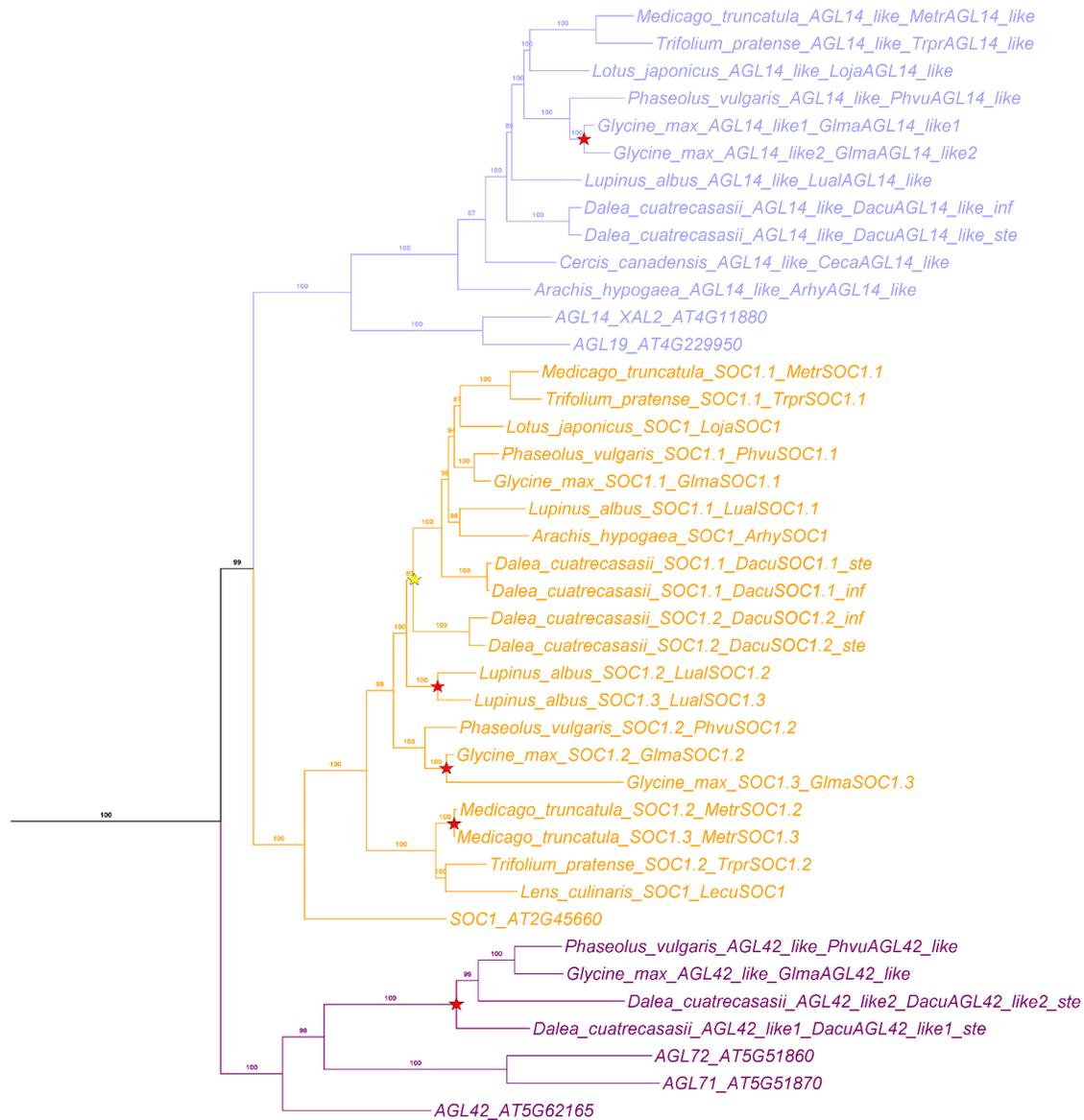
Figura 7. Análisis ML que muestra en detalle los genes *ABS* y *XAL1* en las leguminosas. Las estrellas amarillas indican eventos de duplicación a gran escala, las estrellas rojas representan duplicaciones especie-específicas



Nota. Los números en cada rama indican los valores de Ultrafast Bootstrap (UFBS).

En el caso del clado *ABS* (UFBS = 62), se detectaron dos duplicaciones locales en *Glycine max* y *Phaseolus vulgaris*. Los genes *XAL1*, solo fueron aislados de unas pocas especies, y la reconstrucción filogenética no rescata ninguna duplicación (Figura 7).

Figura 8. Análisis ML que muestra en detalle los genes *AGL42_Like*, *AGL14_Like* y *SOC1* en las leguminosas. Las estrellas amarillas indican eventos de duplicación a gran escala, las estrellas rojas representan duplicaciones especie-específicas de especies

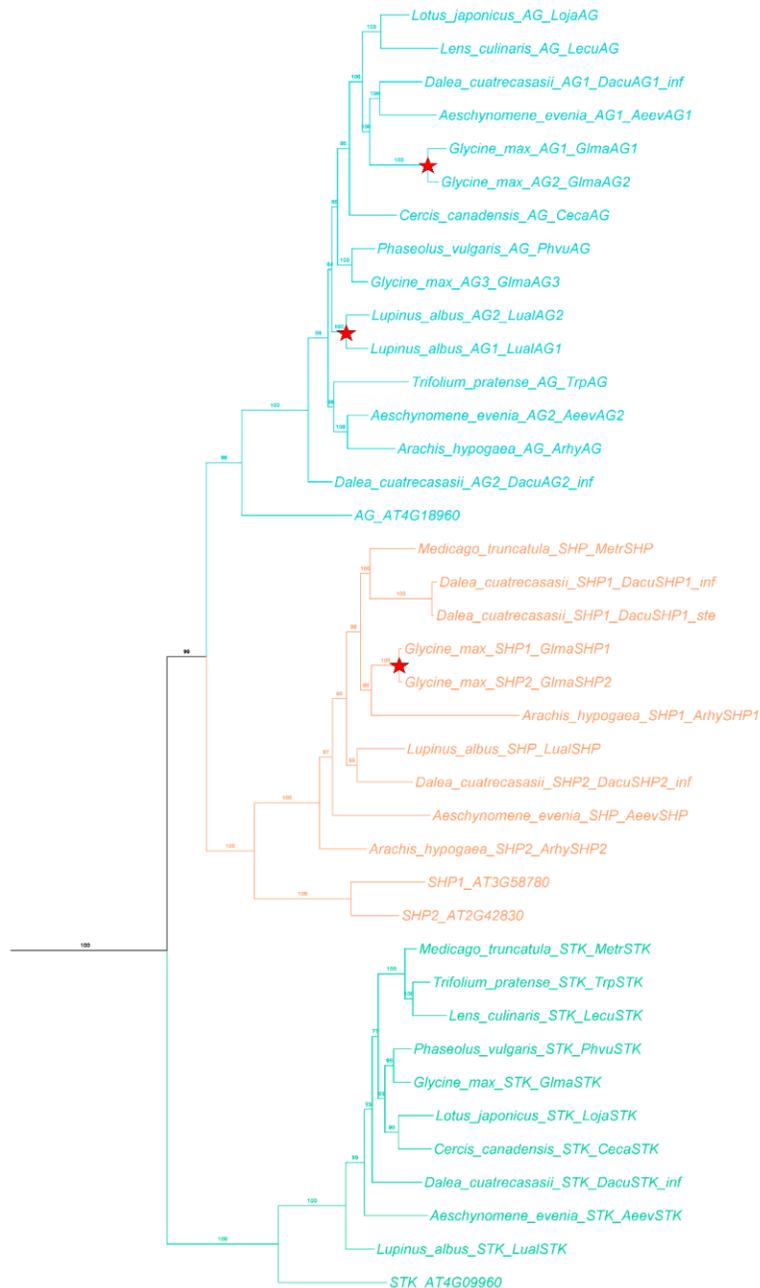


Nota. Los números en cada rama indican los valores de Ultrafast Bootstrap (UFBS).

En cuanto al clado *AGL14 Like* (UFBS = 99), solo se observa una duplicación local en *Glycine max*. En el clado *SOC1*, se detecta al menos una duplicación a gran escala (UFBS = 83) y

tres duplicaciones locales en *Glycine max*, *Lupinus albus* y *Medicago truncatula*. Finalmente, en el clado *AGL42 Like*, se observa una única duplicación local en *Dalea cuatrecasasii* (Figura 8).

Figura 9. Análisis ML que muestra en detalle los genes AG, SHP y STK en las leguminosas. Las estrellas amarillas indican eventos de duplicación a gran escala, las estrellas rojas representan duplicaciones especie-específicas



Nota. Los números en cada rama indican los valores de Ultrafast Bootstrap (UFBS).

Por último, en los clados *AG* y *SHP* (UFBS = 99), no hay duplicaciones a gran escala. Estos genes tienden a mantenerse de copia única. En el clado *AG* se observan dos duplicaciones locales en *Glycine max* y *Lupinus albus*. En el clado *SHP*, solo se identifica una duplicación local en *Glycine max*. Finalmente, en el clado *STK* (UFBS = 100) no se detecta ninguna duplicación (Figura 9).

3. Discusión

Tendencias de duplicación de genes en leguminosas y específicamente en Dalea cuatrecasasii

Diferentes tendencias de duplicación fueron detectadas para cada uno de los genes estudiados en leguminosas. Mientras los genes *FUL*, *PI*, *SEP3*, y *SEP4* se duplicaron previo a la diversificación de todas las leguminosas, la mayoría de las duplicaciones registradas tuvo lugar durante la diversificación de la subfamilia Papilionoideae, y están presentes de copia única en *Cercis*. Estas duplicaciones han ocurrido en los genes *AGL6*, *AGL24*, *AGL79*, *AP1*, *SEP1*, y *SOC1*. Por otra parte, solo algunos genes se han duplicado en grupos más específicos de Papilionoideae. Este es el caso de los genes *ANRI*, cuya duplicación predata la diversificación de Phaseolae, Loteae, Fabeae y Trifolieae. Así mismo, la duplicación de los genes *AGL6* *sister* solo afecta a las Fabeae y las Trifolieae. Es interesante que a pesar de los esfuerzos de identificación de los genes MADS-box en leguminosas este sea el primer estudio en identificar tendencias de duplicaciones en la familia, ya que estudios previos habían comparado genes de una sola especie de las Fabaceae con homólogos de *Arabidopsis*, y nunca habían realizado un muestreo comparativo entre leguminosas (Dong et al., 2021; Kumar et al., 2021; Mou et al., 2022; Okay et al., 2024).

En general los datos sugieren que cada linaje de genes MADS-box pudo haber tenido un tiempo evolutivo específico de duplicación en las leguminosas. Es también interesante que linajes como *CAL/FUL/AP1*, *AP3/PI*, *SEP* y *AG/SHP* exhiben un alto grado de redundancia mientras que los genes, *AGL6*, *AGL12*, *AGL15*, y *ANRI* mantienen menos copias y quizás mayor autonomía funcional, lo que ya había sido resaltado en un estudio a gran escala de los genes MADS-box en soya (Shu et al 2013). Uno de los casos ya reportados de retención de duplicados corresponde a la presencia de tres parálogos de genes *SOC1* en *Cajanus cajan*, aunque en este reporte no se usan los homólogos de *Arabidopsis* para la reconstrucción del árbol de genes MADS-box, por lo que es difícil entender la homología de las copias (Kumar et al., 2021). Sin embargo, si sugiere que algunos grupos de leguminosas han sufrido eventos de duplicación de genes que pudiesen estar asociados a nuevas funciones, en el caso particular de los genes *SOC1* asociadas a cambios en las señales que controlan de floración. Otros casos similares de duplicaciones locales han sido reportados en genes que no son MADS-box, como es el caso de por ejemplo la familia de genes *LEED*, *PEED* (LP) en *Medicago truncatula* (A17) está compuesta por 13 genes que codifican

pequeños péptidos, presuntamente secretados. Sin embargo, esta familia no está presente en los genomas de *Glycine max*, *Lotus japonicus* o en la especie IRLC *Cicer arietinum*, lo que sugiere que esta familia de genes LP surgió en los últimos 25 millones de años (Trujillo et al., 2014).

Con respecto a la especie *Dalea cuatrecasasii* se identificaron varias instancias de genes duplicados, específicamente el caso de los genes *AGL6*, *AGL14* y *SOC1*, que además presentan variaciones sutiles en longitud y mutaciones. Es claro que las duplicaciones son la fuente más importante de cambios de función que puedan ocurrir a nivel evolutivo, de hecho, las duplicaciones pueden conducir a la neofuncionalización o a la pseudogenización, o a la subfuncionalización, o a la redundancia, dependiendo de si las copias mantienen su función o la pierden debido a mutaciones (Jiang et al., 2013).

En el análisis de las secuencias de *FUL* en *Dalea cuatrecasasii*, se observa que la secuencia correspondiente a *FUL2 – DacuFUL2* que se expresa activamente en el tallo presenta diferencias significativas en comparación con las demás secuencias de este grupo. Específicamente, esta secuencia carece de 30 nucleótidos en el extremo 3', en una región clave que incluye el dominio (PWMLRP__E*), el cual es importante para la función de unión de proteínas. Esta ausencia sugiere que podrían existir cambios en la función de la proteína codificada por este gen, potencialmente afectando su capacidad de interacción con otras proteínas o de regulación en procesos celulares. Esta variación podría, en consecuencia, tener implicaciones funcionales relevantes en el desarrollo o adaptación de *Dalea cuatrecasasii*.

3.1 Funciones de genes MADS-box asociados con cambios fenotípicos conocidos en leguminosas

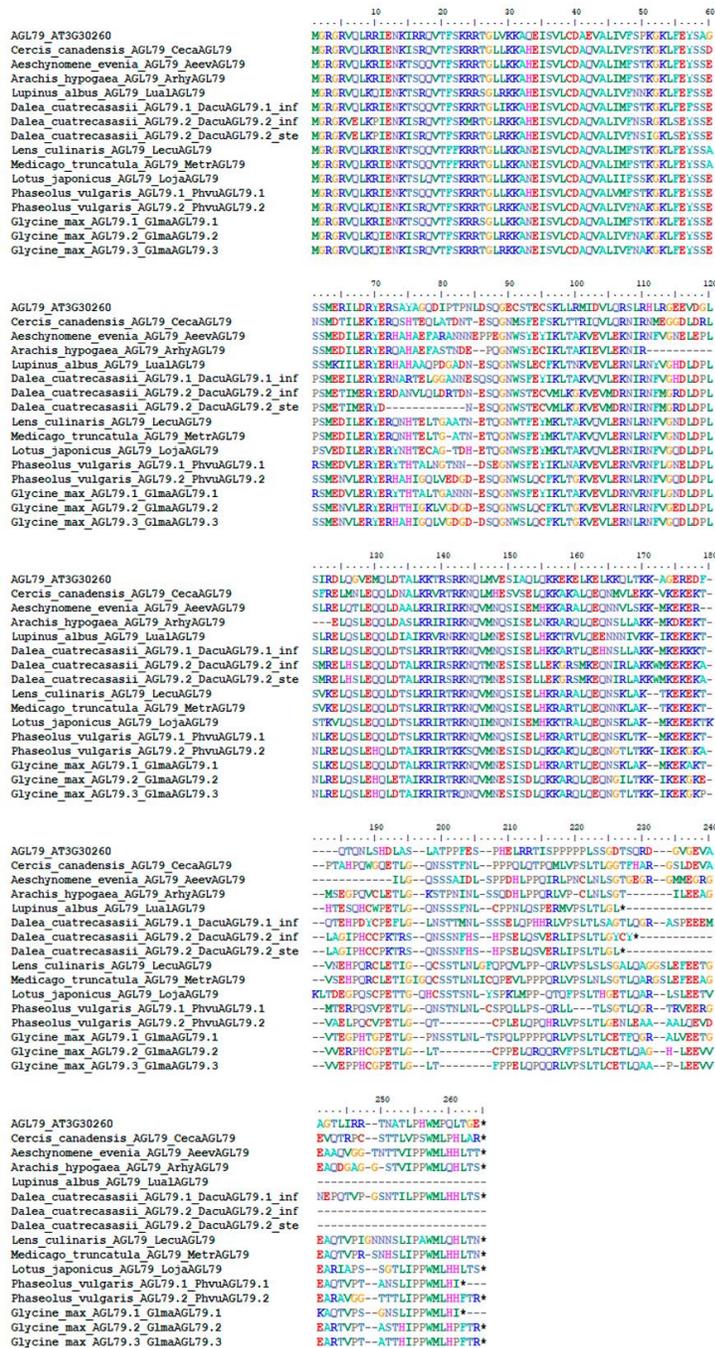
Figura 10. Alineamiento de la región C-terminal de las proteínas SHP identificadas en este análisis. La caja punteada señala la región SSQA identificada en *Medicago truncatula* como una de las más importantes para la lignificación y la torsión de los frutos



Nota. Este análisis revela que esta secuencia no está conservada en las demás leguminosas estudiadas, que además no poseen torsión del fruto.

Todos los genes MADS-box son reguladores esenciales del desarrollo reproductivo en las plantas y definen cambios en características como la floración, la estructura floral y el crecimiento de semillas, que han sido importantes en procesos de domesticación (Schilling et al., 2018). Un caso específico es el de los genes *SHATTERPROOF* (*SHP*) que regulan en gran medida la formación de la zona de dehiscencia en *Arabidopsis* (Liljegren et al., 2000) y el control de la lignificación en las márgenes de las valvas de las leguminosas (Fourquin et al., 2013). Estudios detallados comparados en el género *Medicago* muestran que existe una secuencia distintiva en el dominio C terminal de la proteína *SHP* correlacionada con la lignificación del fruto y la torsión de las valvas que resulta en frutos enrollados. La secuencia polimórfica es el motivo SSQA en frutos enrollados y el cambio a _SQT/S en frutos no enrollados (Fourquin et al., 2013). Nuestros datos muestran que la mayoría de las leguminosas muestreadas tienen el motivo clásico de los frutos no enrollados; sin embargo, las secuencias de *Dalea* son muy diferentes en este motivo. En primer lugar, *Dalea* posee dos genes *SHP*, a diferencia de las demás leguminosas muestreadas. En segundo lugar, mientras que la proteína *DacuSHP1* posee solamente una serina (_S_) en esta región del alineamiento, la proteína *DacuSHP2* posee un motivo muy divergente (PAQT). Estos cambios podrían estar vinculados a cambios en la morfo-anatomía de los frutos extra cortos de *Dalea* (Figura 10).

Figura 11. Alineamiento de secuencias de proteínas del clado AGL79 en leguminosas. La copia *DacuAGL79.2* expresada en la inflorescencia termina en el aminoácido 229 y no presenta el motivo (LPPWML)



Nota. La variante en el tallo tiene una delección en los aminoácidos 72–83 en el dominio I, posterior al dominio MADS.

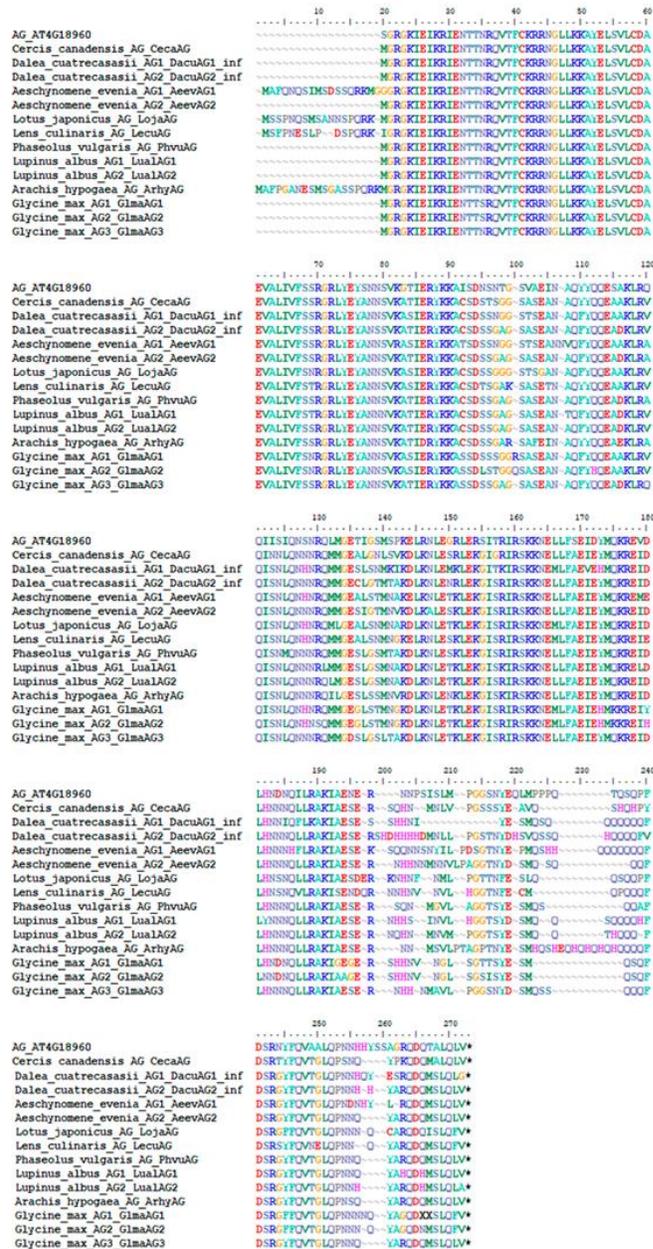
Otro caso muy interesante de rasgos fenotípicos asociados a cambios en genes MADS-box es el de los genes *AGL79*, que están asociados con la identidad reproductiva de meristemas de inflorescencia secundarios. Cuando los homólogos de los genes *AGL79* en arveja, (llamados *VEGETATIVE 1 - VEG1*) se ven afectados, las inflorescencias secundarias se transforman en ramas vegetativas (Berbel et al., 2012). En *Dalea cuatrecasasii* hay dos variantes de *AGL79*, que hemos llamado *DacuAGL79.1* y *AGL79.2*, ambas se expresan en la inflorescencia y solo una (*AGL79.2*) se expresa en el tallo. De hecho, la copia *AGL79.2* termina en la posición 229 en aminoácidos, y carece del motivo LPPWML típico de genes *FUL* (Litt & Irish, 2003). Además, la variante *DacuAGL79.2* que se expresa en el tallo, posee además una delección del aminoácido 72-83 en el dominio I, después del dominio MADS (Figura 11), lo que teóricamente podría afectar sus interacciones proteicas. Será de gran importancia estudiar la expresión y la función de estos homólogos de *VEG1* en *Dalea*, ya que, así como otras Fabaceae (ej. *Arachis*, *Lupinus*), en *Dalea* hay inflorescencias apicales (de primer orden) y axilares (de segundo orden).

Otro linaje de genes que vale la pena explorar en detalle en las leguminosas es el linaje *AGL6* y su gen hermano aquí identificado por primera vez *AGL6 sister*, que parece ser el resultado de una duplicación específica en leguminosas. El gen *AGL6* está implicado en la regulación del meristemo floral, facilitando la transición hacia la formación de órganos florales, participa en el desarrollo de estructuras específicas como la palea en gramíneas, los integumentos del óvulo, y las semillas, y además está implicado en el desarrollo de gametofitos masculinos y femeninos (Dreni & Zhang, 2016). Otros estudios sugieren que homólogos de *AGL6*, pueden desempeñar un papel importante en la morfología de los estambres en los Zingiberales (Yockteng et al., 2013). En las leguminosas, especialmente en la soja, los genes *AGL6* son cruciales para regular la identidad de los órganos florales (Wong et al., 2013). Debido a que es el primer caso conocido en eudicotiledóneas centrales donde los genes *AGL6* están duplicados, será de gran importancia caracterizar su expresión en leguminosas, ya que es posible que nuevas funciones vinculadas a la transición floral o a la especialización de órganos florales estén acompañando la retención de los dos linajes de genes, *AGL6* y *AGL6 sister* en leguminosas.

3.2 Genes exclusivos de las inflorescencias, del tallo y compartidos entre los dos puntos en Dalea cuatrecasasii

Este estudio permitió identificar los genes MADS-box MIKCC de *Dalea cuatrecasasii* expresados exclusivamente en el tallo, en la inflorescencia y aquellos que se expresan de forma más amplia en las dos regiones. Los factores de transcripción activos en tallo y en inflorescencia, incluyen a *AGL14*, *SOC1.1*, *SOC1.2*, *AP3*, *AGL6.1*, *AGL6.2*, *AGL79.2*, *FUL2*, *SEP4.1*, y *SHP1*. Aquellos exclusivamente encontrados en el tallo incluyen *AGL24*, *AGL42 Like1*, *AGL42 Like2*, *FUL1*, *FUL3*, y *SVP*. Finalmente, los genes que son exclusivamente expresados en la inflorescencia incluyen *AG1*, *AG2*, *AGL6 sister*, *AGL24.1*, *AGL24.2*, *AGL79.1*, *DacuAPI Like1*, *DacuAPI Like2*, *FUL1*, *PI1*, *PI2*, *PI3*, *SEP1 Like1*, *SEP1 Like2*, *SEP4.2*, *SEP4.3*, *SEP3*, *SHP2*, *STK* y *TM6*. De estos datos es notable resaltar que varios de los integradores de la floración, como son los homólogos de *AGL24*, *SVP* y *SOC1* están activos en el tallo y en la inflorescencia de *Dalea*. La especie *Dalea cuatrecasasi* es la única hospedera conocida de la especie de Apodanthaceae *Pilostyles boyacensis*, y en este contexto es de particular importancia la presencia de genes de la planta hospedera que estén presentes en el tallo y cuyo mRNA pueda ser potencialmente movilizado hacia el endófito de la planta parásita. La movilidad de ARN en plantas, especialmente en especies representativas como *Arabidopsis thaliana*, ha sido un campo de investigación esclarecedor sobre cómo determinados ARN pueden moverse a través del floema para llegar a diferentes tejidos y desempeñar funciones reguladoras específicas. Por ejemplo, Thieme et al. (2015) demostraron que el ARN mensajero no autónomo es capaz de moverse a través del floema en *Arabidopsis*, llegando incluso a zonas alejadas de la planta. Este descubrimiento ofrece una base firme para sugerir que algunos ARN en *Dalea cuatrecasasii*, especialmente aquellos asociados a la familia de genes MADS-box, también podrían mostrar movilidad en los tejidos, lo que les facilitaría el contacto directo con el endófito *Pilostyles boyacensis* en el tallo. Estudios en otras especies han demostrado que la movilidad de ARN puede estar asociada a procesos de regulación no solo del desarrollo sino de respuesta a patógenos o simbiontes (Palauqui et al., 1997), lo cual refuerza la plausibilidad de esta hipótesis. Debido a que varios de estos genes presentes en el tallo de dalea pueden jugar roles importantes en la floración es interesante especular que la planta hospedera regula la floración de la planta parásita, más aún cuando en plantas endoholoparasitas no hay hojas que son las principales fuentes de señales de floración asociadas a las variables ambientales, como la luz y la temperatura.

Figura 12. Alineación múltiple de secuencias de proteínas de la familia AG en varias especies de leguminosas



Nota. La duplicación de los genes AG en *D. cuatrecasasii* y otras leguminosas, que generalmente presentan una sola copia en otras plantas, sugiere que estas copias adicionales, con cambios significativos en el dominio C terminal y distintas interacciones tetraméricas, podrían estar experimentando subfuncionalización o neofuncionalización.

Por otra parte, los genes *AG1*, *AG2*, *AGL6 sister*, *AGL24.1*, *AGL24.2*, *AGL79.1*, *DacuAP1 Like1*, *DacuAP1 Like2*, *FUL1*, *PI1*, *PI2*, *PI3*, *SEP1 Like1*, *SEP1 Like2*, *SEP4.2*, *SEP4.3*, *SEP3*, *SHP2*, *STK* y *TM6*, cuya expresión ocurre de forma exclusiva en la inflorescencia deben ser predominantemente aquellos que controlan exclusivamente la identidad de los órganos florales de *Dalea cuatrecasasii*. En efecto, varios están vinculados a la identidad de todos los órganos florales en conjunto (*SEP*) o a funciones más restringidas a la identidad de sépalos y pétalos (*API*), pétalos y estambres (*AP3*, *PI* y *TM6*), y la identidad de estambres y carpelos (*AG*), así como a la formación de óvulos (*STK*). De gran interés en este listado es la ocurrencia de un gen específico de leguminosas, que es el *AGL6 sister*, cuya función es aún desconocida y la duplicación de los genes *AG*, que suelen ser de copia única en otras plantas. La presencia de dos copias de *AG* con grandes cambios en el dominio C terminal y la capacidad de establecer diferentes interacciones tetraméricas es un aspecto que debe estudiarse con más atención en el futuro, ya que estos genes pueden estar sufriendo procesos de subfuncionalización o neofuncionalización en *D. cuatrecasasii* y en las otras leguminosas con duplicaciones locales (Figura 12).

Referencias

- Arias-Agudelo, L. M., González, F., Isaza, J. P., Alzate, J. F., & Pabón-Mora, N. (2019). Plastome reduction and gene content in New World Pilostyles (Apodanthaceae) unveils high similarities to African and Australian congeners. *Molecular Phylogenetics and Evolution*, *135*, 193–202. <https://doi.org/10.1016/j.ympev.2019.03.014>
- Ayra, L., Reyero-Saavedra, M. del R., Isidra-Arellano, M. C., Lozano, L., Ramírez, M., Leija, A., Fuentes, S.-I., Girard, L., Valdés-López, O., & Hernández, G. (2021). Control of the rhizobia nitrogen-fixing symbiosis by common bean MADS-domain/AGL transcription factors. *Frontiers in plant science*, *12*. <https://doi.org/10.3389/fpls.2021.679463>
- Azani, N., Babineau, M., Bailey, C. D., Banks, H., Barbosa, A. R., Pinto, R. B., Boatwright, J. S., Borges, L. M., Brown, G. K., Bruneau, A., Candido, E., Cardoso, D., Chung, K.-F., Clark, R. P., Conceição, A. de S., Crisp, M., Cubas, P., Delgado-Salinas, A., Dexter, K. G., ... Zimmerman, E. (2017). A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny: The Legume Phylogeny Working Group (LPWG). *Taxon*, *66*(1), 44–77. <https://doi.org/10.12705/661.3>
- Barneby, R. C. (1977). Daleae imagines. *Memoirs of the New York Botanical Garden* 27: 1–892, 1980. Three new species of *Dalea sect. Parosela* (Leguminosae: Amorpheae) from western and southern Mexico. *Brittonia* 32: 392–396, 1990. Two new taxa in *Dalea* (Fabaceae: Amorpheae) from southern Mexico and northern Chile. *Brittonia* 42: 89–91.
- Berbel, A., Ferrándiz, C., Hecht, V., Dalmais, M., Lund, O. S., Sussmilch, F. C., Taylor, S. A., Bendahmane, A., Ellis, T. H. N., Beltrán, J. P., Weller, J. L., & Madueño, F. (2012). VEGETATIVE1 is essential for development of the compound inflorescence in pea. *Nature Communications*, *3*(1), 797. <https://doi.org/10.1038/ncomms1801>
- Dong, X., Deng, H., Ma, W., Zhou, Q., & Liu, Z. (2021). Genome-wide identification of the MADS-box transcription factor family in autotetraploid cultivated alfalfa (*Medicago sativa* L.) and expression analysis under abiotic stress. *BMC Genomics*, *22*(1). <https://doi.org/10.1186/s12864-021-07911-9>

-
- Dong, Z.-C., Zhao, Z., Liu, C.-W., Luo, J.-H., Yang, J., Huang, W.-H., Hu, X.-H., Wang, T. L., & Luo, D. (2005). Floral patterning in *Lotus japonicus*. *Plant Physiology*, 137(4), 1272–1282. <https://doi.org/10.1104/pp.104.054288>
- Dreni, L., & Zhang, D. (2016). Flower development: the evolutionary history and functions of the AGL6 subfamily MADS-box genes. *Journal of Experimental Botany*, 67(6), 1625–1638. <https://doi.org/10.1093/jxb/erw046>
- Dreni, L., & Zhang, D. (2016). Flower development: the evolutionary history and functions of the AGL6 subfamily MADS-box genes. *Journal of Experimental Botany*, 67(6), 1625–1638. <https://doi.org/10.1093/jxb/erw046>
- Estrada-C., A. E., Villarreal-Q., J. A., & López-E., L. (2011). A new species of *Dalea* ser. *Versicolores* (Leguminosae: Amorphaeae) from Durango, Mexico. *Brittonia*, 63(4), 465–468. <https://doi.org/10.1007/s12228-011-9198-9>
- Feng, L., Tao, J., Yang, X., Jiao, Q., Wang, C., Cheng, Y., & Yin, Y. (2019). Characterization and phylogenetic analysis of the complete chloroplast genome of *Cercis canadensis* “Forest Pansy”. *Mitochondrial DNA. Part B, Resources*, 5(1), 154–155. <https://doi.org/10.1080/23802359.2019.1698359>
- Fernandez, D. E., Heck, G. R., Perry, S. E., Patterson, S. E., Bleecker, A. B. and Fang, S. C. (2000). The embryo MADS domain factor AGL15 acts postembryonically. Inhibition of perianth senescence and abscission via constitutive expression. *Plant Cell* 12, 183-198.
- Fourquin, C., del Cerro, C., Victoria, F. C., Vialette-Guiraud, A., de Oliveira, A. C., & Ferrándiz, C. (2013). A change in SHATTERPROOF protein lies at the origin of a fruit morphological novelty and a new strategy for seed dispersal in medicago genus. *Plant Physiology*, 162(2), 907–917. <https://doi.org/10.1104/pp.113.217570>
- Gonzalez, F. & Mora, N. P. (2014b). *Pilostyles boyacensis*, a new species of Apodanthaceae (Cucurbitales) from Colombia. *Phytotaxa* 178: 138-145.

- González, F. & Pabón-Mora, N. (2014a). First reports and generic descriptions of the achlorophyllous holoparasites Apodanthaceae (Cucurbitales) of Colombia. *Actualidades Biológicas* 36: 123-135.
- Harding, E. W., Tang, W., Nichols, K. W., Fernandez, D. E. and Perry, S. E. (2003). Expression and maintenance of embryogenic potential is enhanced through constitutive expression of AGAMOUS-Like 15. *Plant Physiol.* 133, 653- 663
- Heck, G. R., Perry, S. E., Nichols, K. W. and Fernandez, D. E. (1995). AGL15, a MADS domain protein expressed in developing embryos. *Plant Cell* 7, 1271- 1282
- Hoang, D.T., Chernomor, O., Von Haeseler, A., Minh, B.Q., Vinh, L.S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35 (2), 518–522.
- Jiang, A., Gan, L., Tu, Y., Ma, H., Zhang, J., Song, Z., He, Y., Cai, D., & Xue, X. (2013). The effect of genome duplication on seed germination and seedling growth of rice under salt stress. *Australian Journal of Crop Science*, 7(12), 1814–1821.
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6), 587–589. <https://doi.org/10.1038/nmeth.4285>
- Kumar, K., Srivastava, H., Das, A., Tribhuvan, K. U., Durgesh, K., Joshi, R., Sevanthi, A. M., Jain, P. K., Singh, N. K., & Gaikwad, K. (2021). Identification and characterization of MADS box gene family in pigeonpea for their role during floral transition. *3 Biotech*, 11(2). <https://doi.org/10.1007/s13205-020-02605-7>
- Kutter, C., Schöb, H., Stadler, M., Meins, F., Jr and Si-Ammour, A. (2007). MicroRNA-mediated regulation of stomatal development in *Arabidopsis*. *Plant Cell* 19, 2417-2429.
- Lewis, G. P., Schrire, B. D., Mackinder, B. A., Rico, L., & Clark, R. (2013). A 2013 linear sequence of legume genera set in a phylogenetic context — A tool for collections management and taxon sampling. *Suid-Afrikaanse Tydskrif Vir Plantkunde [South African Journal of Botany]*, 89, 76–84. <https://doi.org/10.1016/j.sajb.2013.06.005>

- Lewis, G.P. & Forest, F. (2005). Cercideae. Pp. 57–67 in: Lewis, G., Schrire, B., Mackinder, B. & Lock, M. (eds.), *Legumes of the World*. Richmond, U.K.: Royal Botanic Gardens, Kew.
- Lewis, G.P. & Schrire, B.D. 2003. Leguminosae or Fabaceae? Pp. 1–3 in: Klitgaard, B. & Bruneau, A. (eds.), *Advances in legume systematics, part 10, Higher level systematics*. Richmond, U.K.: Royal Botanic Gardens, Kew.
- Lewis, G., Schrire B., Mackinder, B. & Lock, M. (eds.) 2005. *Legumes of the World*. Richmond, U.K.: Royal Botanic Gardens, Kew
- Liljegren, S. J., Ditta, G. S., Eshed, Y., Savidge, B., Bowman, J. L., & Yanofsky, M. F. (2000). SHATTERPROOF MADS-box genes control seed dispersal in *Arabidopsis*. *Nature*, 404(6779), 766–770. <https://doi.org/10.1038/35008089>
- Litt, A., & Irish, V. F. (2003). Duplication and diversification in the APETALA1/FRUITFULL floral homeotic gene lineage: implications for the evolution of floral development. *Genetics*, 165(2), 821–833. <https://doi.org/10.1093/genetics/165.2.821>
- Lock, M. (2005). Amorpheae. Pp. 299-305 in Lewis, G. P., Schrire, B., & Lock, M. (eds.). *Legumes of the World*. Kew: Royal Botanic Gardens.
- LPWG, Legume Phylogeny Working Group (2013a). Legume phylogeny and classification in the 21st century: Progress, prospects and lessons for other species-rich clades. *Taxon* 62: 217–248. <https://doi.org/10.12705/622.8>
- Mou, Y., Yuan, C., Sun, Q., Yan, C., Zhao, X., Wang, J., Wang, Q., Shan, S., & Li, C. (2022). MIKC-type MADS-box transcription factor gene family in peanut: Genome-wide characterization and expression analysis under abiotic stress. *Frontiers in plant science*, 13. <https://doi.org/10.3389/fpls.2022.980933>
- Ng, M., & Yanofsky, M. F. (2001). Function and evolution of the plant MADS-box gene family. *Nature Reviews. Genetics*, 2(3), 186–195. <https://doi.org/10.1038/35056041>
- Okay, A., Kırılıoğlu, T., Durdu, Y. Ş., Akdeniz, S. Ş., Büyük, İ., & Aras, E. S. (2024). Omics approaches to understand the MADS-box gene family in common bean (*Phaseolus vulgaris*

- L.) against drought stress. *Protoplasma*, 261(4), 709–724. <https://doi.org/10.1007/s00709-024-01928-z>
- Palauqui, J. C., Elmayan, T., Pollien, J. M., & Vaucheret, H. (1997). Systemic acquired silencing: transgene-specific post-transcriptional silencing is transmitted by grafting from silenced stocks to non-silenced scions. *The EMBO Journal*, 16(15), 4738–4745. <https://doi.org/10.1093/emboj/16.15.4738>
- Piñeros Urrego, P., Suárez-Barón, H., Pabón-Mora, N., & González, F. (2022). Reinstatement of the genus *Psorodendron* and related systematic novelties as revealed from phylogenetic analyses of the tribe Amorpheae (Leguminosae, Papilionoideae). *Caldasia*, 45(1), 49–65. <https://doi.org/10.15446/caldasia.v45n1.96498>
- Piñeros-U, L. P., & González, F. (2020). Revisión taxonómica de *Dalea* (Leguminosae: Papilionoideae) en Colombia. *Caldasia*, 42(2), 220–240. <https://doi.org/10.15446/caldasia.v42n2.83513>
- Piñeros-U., L. P., & González, F. (2019). Nueva especie de *Dalea* (Leguminosae: Papilionoideae) de matorrales interandinos secos del sur de Colombia y clave para identificación de las especies del género en el país. *Caldasia*, 41(1), 215–223. <https://doi.org/10.15446/caldasia.v41n1.75852>
- Ren, Q., Jiang, H., Xiang, W., Nie, Y., Xue, S., Zhi, H., Li, K., & Gai, J. (2022). A MADS-box gene is involved in soybean resistance to multiple Soybean mosaic virus strains. *The Crop Journal*, 10(3), 802–808. <https://doi.org/10.1016/j.cj.2021.10.003>
- Riechmann, J. L., Krizek, B. A., & Meyerowitz, E. M. (1996). Dimerization specificity of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. *Proceedings of the National Academy of Sciences of the United States of America*, 93(10), 4793–4798. <https://doi.org/10.1073/pnas.93.10.4793>
- Rijkema, A. S., Zethof, J., Gerats, T., & Vandenbussche, M. (2009). The petunia AGL6 gene has a SEPALLATA-like function in floral patterning. *The Plant Journal: For Cell and Molecular Biology*, 60(1), 1–9. <https://doi.org/10.1111/j.1365-313X.2009.03917.x>

- Schilling, S., Pan, S., Kennedy, A., & Melzer, R. (2018). MADS-box genes and crop domestication: the jack of all traits. *Journal of Experimental Botany*, 69(7), 1447–1469. <https://doi.org/10.1093/jxb/erx479>
- Shu, Y., Yu, D., Wang, D., Guo, D., & Guo, C. (2013). Genome-wide survey and expression analysis of the MADS-box gene family in soybean. *Molecular Biology Reports*, 40(6), 3901–3911. <https://doi.org/10.1007/s11033-012-2438-6>
- Smaczniak, C., Immink, R. G. H., Angenent, G. C., & Kaufmann, K. (2012). Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. *Development (Cambridge, England)*, 139(17), 3081–3098. <https://doi.org/10.1242/dev.074674>
- Thieme, C. J., Rojas-Triana, M., Stecyk, E., Schudoma, C., Zhang, W., Yang, L., Miñambres, M., Walther, D., Schulze, W. X., Paz-Ares, J., Scheible, W.-R., & Kragler, F. (2015). Endogenous Arabidopsis messenger RNAs transported to distant tissues. *Nature Plants*, 1(4), 15025. <https://doi.org/10.1038/nplants.2015.25>
- Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A., & Minh, B. Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research*, 44(W1), W232-5. <https://doi.org/10.1093/nar/gkw256>
- Trujillo, D. I., Silverstein, K. A. T., & Young, N. D. (2014). Genomic characterization of the LEED, PEEDs, a gene family unique to the medicago lineage. *G3 (Bethesda, Md.)*, 4(10), 2003–2012. <https://doi.org/10.1534/g3.114.011874>
- Wong, C. E., Singh, M. B., & Bhalla, P. L. (2013). Novel members of the AGAMOUS LIKE 6 subfamily of MIKCC-type MADS-box genes in soybean. *BMC Plant Biology*, 13, 105. <https://doi.org/10.1186/1471-2229-13-105>
- Yockteng, R., Almeida, A. M. R., Morioka, K., Alvarez-Buylla, E. R., & Specht, C. D. (2013). Molecular evolution and patterns of duplication in the SEP/AGL6-like lineage of the Zingiberales: a proposed mechanism for floral diversification. *Molecular Biology and Evolution*, 30(11), 2401–2422. <https://doi.org/10.1093/molbev/mst137>