

CARACTERIZACIÓN *IN SILICO* Y ANÁLISIS DE LA EXPRESIÓN GÉNICA DE PROTEÍNAS ABUNDANTES EN LA EMBRIOGÉNESIS TARDÍA DE *AGAVE TEQUILANA* WEBER VAR. AZUL

IN SILICO CHARACTERIZATION AND GENE EXPRESSION ANALYSIS OF LATE-EMBRYOGENESIS ABUNDANT PROTEINS OF *AGAVE TEQUILANA* WEBER VAR. AZUL

 JORGE VILLEGAS-CAMAS¹,  KARINA VERDEL-ARANDA²,  JOEL LARA-REYNA¹ AND  AÍDA MARTÍNEZ-HERNÁNDEZ^{1*}

¹ Colegio de Postgraduados campus Campeche, Champotón, Campeche, México.

² Tecnológico Nacional de México, Instituto Tecnológico de Chiná, Campeche, México.

*Author for correspondence: aida.martinez@colpos.mx

Resumen:

Conocimiento previo/especie: *Agave tequilana* Weber var. *azul* es un importante cultivo en México, utilizado para la producción de tequila. Muchas especies de *Agave* son tolerantes a condiciones áridas. Sin embargo, las bases moleculares de los mecanismos seleccionados en los agaves para confrontar el estrés abiótico, no han sido descritas.

Hipótesis: Las proteínas abundantes en la embriogénesis tardía (LEAPs), una superfamilia asociada a las respuestas ante el estrés abiótico en plantas, son un elemento clave en las respuestas de los agaves ante ambientes áridos.

Métodos: Datos transcriptómicos de *A. tequilana* fueron utilizados para realizar análisis *in silico* e identificar genes que codifican *Agave* LEAPs. Comparamos sus características estructurales y su similitud/divergencia con LEAPs de otras plantas, utilizando bioinformática. La abundancia de los transcritos de AteqLEAP en órganos vegetativos y en respuesta a altas temperaturas fue determinada mediante qRT-PCR.

Resultados: Identificamos tres AteqLEAPs estructuralmente diferentes. Las AteqLEA_5Bs muestran similitud (relativamente baja) con LEAPs conocidas como “atípicas” (LEA_3) y exhiben, inesperadamente, altos niveles de expresión constitutiva en hojas. Los transcritos de AteqLEA_5C (LEA_2) mostraron baja expresión en todos los órganos analizados. Dos isoformas de AteqDHN tipo SK3 muestran el típico desorden estructural e hidrofilicidad de las dehidrinas y son altamente expresadas en hojas no desarrolladas, meristemo vegetativo y tallo (piña).

Conclusiones: Las AteqLEAP_5B parecen tener un papel protector preventivo en las hojas fotosintéticas plenamente funcionales; mientras que las AteqDHNs parecen proteger tejidos en proceso de diferenciación como meristemos y hojas en desarrollo; así como tejidos de almacenamiento, como el tallo del agave.

Palabras clave: *Agave*LEAP, análisis estructural *in silico*, dehidrinas, patrón de expresión génica.

Abstract

Background/ Studied species: *Agave tequilana* Weber var. *azul* is an important crop in Mexico, used for tequila production. Many species of *Agave* grow in arid lands and are highly tolerant to those environmental conditions. However, the molecular bases of mechanisms selected in agaves to face abiotic stress, have not been described.

Hypotheses: Late-embryogenesis abundant proteins (LEAPs), a superfamily closely associated to abiotic stress responses in plants, are a key element in agave responses to arid environments.

Methods: Transcriptomic data from *A. tequilana* were used to perform *in silico* analyses to identify genes encoding *Agave* LEAPs. We compare their structural characteristics and their similarity/divergence with described LEAPs from other plant species using bioinformatics. The abundance of the AteqLEAP transcripts in vegetative organs and in response to high temperatures was determined by qRT-PCR.

Results: We identified three structurally different AteqLEAPs. AteqLEA_5Bs show low similarity to LEAPs known as “atypical” (containing LEA_3 domain); and they exhibit an unexpected high and constitutive expression pattern in leaves. But AteqLEA_5C (LEA_2) transcripts are low-expressed in all the organs analyzed. Two isoforms of AteqDHN type SK3 show the typical structural disorder and hydrophilicity of dehydrins and are highly expressed in undeveloped leaves, vegetative meristem and stem.

Conclusions: AteqLEAP_5Bs seem to play a major and preventive protective role in fully photosynthetic leaves as in a “primed state”; whereas AteqDHNs presumably play a role protecting tissues in differentiation stage as meristem and developing leaves, and storage tissues such as the agave stem.

Keywords: *Agave*LEAP, dehydrins, gene expression pattern, *in silico* structural analysis.

Este artículo se encuentra bajo los términos de la licencia Creative Commons Attribution License CC BY-NC (4.0) internacional.

<https://creativecommons.org/licenses/by-nc/4.0/>



Las proteínas abundantes en la embriogénesis tardía o LEAPs (del inglés *Late-embryogenesis abundant proteins*), son un amplio grupo de proteínas ubicuas descritas como altamente hidrofílicas, ricas en aminoácidos (aa) pequeños (Gly, Ala y Ser) y bajo contenido de Cys, Trp y Phe; por lo que su estructura posee alta flexibilidad y se considera como intrínsecamente desordenada (Battaglia *et al.* 2008). Las LEAPs se identificaron inicialmente por su elevada abundancia en la última etapa de la embriogénesis, durante el período de deshidratación de las semillas (Dure *et al.* 1981, Galau & Dure 1981). Actualmente se conoce que las LEAPs pertenecen a una superfamilia multifuncional presente en plantas, invertebrados y bacterias, y han sido ampliamente descritas como proteínas efectoras que también se expresan en tejidos vegetativos de plantas y en respuesta a diversas condiciones de estrés ambiental como frío, sequía, salinidad; así como a ácido abcísico (ej: Galau *et al.* 1986, Yu *et al.* 2016, Xiang & Man 2018, Chen *et al.* 2019, Guo *et al.* 2019, Dirk *et al.* 2020). Se han identificado grupos de LEAPs también inducidas por condiciones de estrés, conocidas como *atípicas* por presentar relativamente alta hidrofobicidad (Galau *et al.* 1993, Maitra & Cushman 1994).

Diferentes LEAPs se localizan en diferentes compartimentos subcelulares. Candat *et al.* (2014) mostraron experimentalmente que 36 de las 51 LEAPs de *A. thaliana* son citosólicas, algunas con difusión al núcleo, mientras que las restantes se localizan exclusivamente en plástidos, mitocondria, retículo endoplásmico, vacuola, pexofagosomas o son secretadas al espacio extracelular. Las LEA_4 se distribuyen ubicuamente dentro de las células.

El mecanismo bioquímico y molecular mediante el cual las LEAPs contribuyen a la protección de las plantas cuando confrontan condiciones ambientales adversas, no está bien definido. Por su naturaleza flexible, ha sido difícil determinar experimentalmente su estructura y definir su función. De acuerdo a ensayos bioquímicos *in vitro* y a fenotipos evidenciados por mutaciones en genes codificantes de LEAPs, se ha propuesto que estas proteínas podrían realizar más de una función: como secuestradores de iones, como antioxidantes, como chaperonas y escudos contra el calor, como estabilizadoras de membranas y antiagregantes de proteínas (Amara *et al.* 2014, Cuevas-Velázquez *et al.* 2014, Farrant *et al.* 2015). Posibles mecanismos de cómo las LEAPs protegen al proteoma susceptible ante condiciones de estrés, modificando el ambiente intracelular para mantener la homeostasis, se han revisado recientemente (Dirk *et al.* 2020).

Las LEAPs son codificadas en el genoma de las plantas por varias familias génicas que se diferencian entre sí por su secuencia proteínica, dominios conservados y motivos distintivos. Se han propuesto diferentes criterios de nomenclatura y clasificación de las LEAPs. En este trabajo nos basamos en las ocho familias reconocidas por los dominios descritos en Pfam (El-Gebali *et al.* 2019) y los grupos propuestos por Battaglia *et al.* (2008).

El número de genes para LEAPs en cada especie vegetal es variable. En *Arabidopsis thaliana*, se reportan 51 genes codificando nueve distintos grupos de LEAPs (Hundertmark & Hincha 2008). Varios de estos genes se han caracterizado funcionalmente (Guo *et al.* 1992, Artus *et al.* 1996, Manfre *et al.* 2006, Dang *et al.* 2014). En *Oryza* se identifican 26 *loci* en el ‘Rice Genome Annotation Project’ (rice.plantbiology.msu.edu), funcionalmente anotados por homología en EnsemblPlants como genes de LEAPs (plants.ensembl.org/Oryza_sativa/Info/Index). En plantas xerófitas adaptadas a sobrevivir bajo condiciones ambientales adversas y que muestran tolerancia a la desecación temporal (DT) de sus tejidos vegetativos, como *Xerophyta viscosa*, se identificaron 126 genes de LEAPs mostrando una expansión genómica respecto a otras plantas. Noventa genes mostraron una mayor acumulación de sus transcritos ante tratamientos de deshidratación y rehidratación de plantas adultas (Costa *et al.* 2017), sugiriendo una función de estos genes en la adaptación fisiológica a dichas condiciones. Análisis transcriptómicos en *Craterostigma planlagioides* indican que la acumulación de transcritos de LEAPs es uno de los componentes principales asociados a la DT (Giarola & Bartels 2015).

Una revisión comparativa de las ocho familias multigénicas de LEAPs (Pfam) presentes en 60 genomas de plantas se publicó recientemente por Artur *et al.* (2019), revelando la expansión de algunas familias en ciertas especies y sugiriendo que la diversificación estructural y funcional de las LEAPs ha contribuido a la adaptación de algunas plantas a ambientes con condiciones hídricas limitadas. Describimos abajo características de tres grupos de LEAPs estructuralmente distintas, las cuales son de interés para este trabajo:

Dehidrinas (Grupo 2). Contienen el dominio PF00257 de Pfam, se caracterizan por su alta hidrofilicidad y se distinguen por la presencia de los siguientes motivos altamente conservados (Battaglia *et al.* 2008): el segmento-K rico en Lys se encuentra en todas las dehidrinas de una a 11 veces; el segmento-Y se distribuye de una a 35 copias en tándem hacia la región N-terminal de la proteína; un tercer motivo presente en algunas dehidrinas es el segmento-S, rico en Ser. La presencia y ubicación de estos motivos divide estructuralmente a las dehidrinas en cinco subgrupos: K, SK, YSK, YK y KS. Algunas dehidrinas poseen también un segmento-F (Strimbeck 2017, Riley *et al.* 2019). A la fecha se han caracterizado funcionalmente dehidrinas de diversas especies, describiendo los patrones de abundancia de sus transcritos ante condiciones como sequía, alta temperatura, salinidad y estrés osmótico, y comparando entre variedades susceptibles y tolerantes a sequía, así como en plantas xerófitas. También se ha analizado el efecto de silenciar sus genes o sobre-expresarlos heterólogamente en bacterias o en otras especies vegetales (Shi *et al.* 2016, Lv *et al.* 2017, Guo *et al.* 2019, , Cui *et al.* 2020). Las dehidrinas de *A. thaliana* como ERD14, COR47, ERD10, LTI45 y RAB18 participan en la tolerancia al frío, aparentemente protegiendo membranas plasmáticas y de cloroplastos (Artus *et al.* 1996, Puhakainen *et al.* 2004). Todas ellas parecen localizarse en el citosol (Candat *et al.* 2014). Los posibles mecanismos mediante los cuales las dehidrinas protegen membranas y enzimas, eliminan especies reactivas de oxígeno, y/o previenen la desnaturización de ciertas proteínas, evitando la formación de agregados proteicos, han sido revisados por varios autores (Cuevas-Velázquez *et al.* 2014, Graether & Boddington 2014).

LEAPs atípicas (Grupo 5). Originalmente identificadas en *Craterostigma*, arroz y soya (Piatkowski *et al.* 1990, Galau *et al.* 1993, Maitra & Cushman 1994), las LEAPs *atípicas* tienen un contenido de aminoácidos hidrofóbicos relativamente alto, pocos aminoácidos con carga iónica y tienen estructura secundaria más definida que otras LEAPs (Singh *et al.* 2005). Se clasifican en tres subgrupos denominados 5A, 5B, 5C; los cuales contienen los dominios Pfam SMP (PF04927), LEA_3 (PF03242) y LEA_2 (PF03168), respectivamente (Battaglia *et al.* 2008). Se ha propuesto que, en solución, estas LEAPs adquieren conformación globular, la cual es inestable en altas temperaturas (Singh *et al.* 2005). Este grupo ha sido menos caracterizado que otras LEAPs.

Grupo 5B (LEA_3).- Pocas LEAPs de este grupo han sido estudiadas. En *Arabidopsis*, dos de los cuatro *loci* que codifican LEAP_5B han sido descritas funcionalmente. SAG21/AtLEA5, previene la senescencia y participa en la tolerancia a estrés oxidativo (Mowlia *et al.* 2006, Salleh *et al.* 2012). AtDI21 se acumula en el proceso de sequía y se regula por ácido abscísico (Gosti *et al.* 1995). Tres LEAP_5B de *Arabidopsis* son mitocondriales (Candat *et al.* 2014). El gen *CpLEA5* de *Chimonanthus praecox* (planta que se desarrolla en ambientes gélidos), sobreexpresado en bacterias, levaduras o plantas, contribuye a la protección contra el frío, salinidad o sequía y previene del daño a membranas artificiales (Liu *et al.* 2015). La proteína *LjIDP1* de *Lotus japonicus* es intrínsecamente desordenada en solución, pero cambia en condiciones de deshidratación o desnaturizantes, adoptando una estructura de α -hélice y previniendo la inactivación de enzimas al evitar que formen agregados o pierdan su conformación (Haaning *et al.* 2008). Recientemente se ha identificado que este grupo, además de contener el dominio Pfam PF03242, se definen por la presencia de un motivo W (Trp) (Singh & Graether 2020) y otros ubicados en las regiones N- y C-terminal.

Grupo 5C (LEA_2).- Esta familia, a pesar de haber sido originalmente clasificadas como “atípicas” por sus características estructurales, se ha reconocido recientemente como la más extendida en los genomas de las plantas (Artur *et al.* 2019). Contienen el dominio Pfam PF03168 y un motivo distintivo conservado (Hundertmark & Hincha 2008). Las LEAPs_5C se han identificado en diversas especies y sus transcritos se acumulan ante múltiples tipos de estrés (Piatkowski *et al.* 1990, Galau *et al.* 1993, Maitra & Cushman 1994, Zegzouti *et al.* 1997, Kimura *et al.* 2003, Kim *et al.* 2005, Wang *et al.* 2014, Zheng *et al.* 2019). Estas LEAPs presentan actividad protectora *in vitro* asociada a salinidad, congelamiento, calor, radiación UV, estrés oxidativo y osmótico (Singh *et al.* 2005, He *et al.* 2012, Hu *et al.* 2019) y confieren tolerancia a estrés ambiental cuando se sobre-expresan de forma heteróloga (Kim *et al.* 2005, Wang *et al.* 2014, Zheng *et al.* 2019). Las LEAPs_5C, ampliamente distribuidas en la naturaleza, se encuentran en algunas especies bacterianas conteniendo el dominio WHy (de *Water stress and Hypersensitive response*) similar a LEA_2,

presumiblemente participando en su tolerancia a la desecación (Ciccarelli & Bork 2005). Las LEAPs_5C de *Arabidopsis* son citosólicas (Candat *et al.* 2014) y se ha reportado que en solución son más hidrofóbicas, estructuradas y estables que otras LEAPs del grupo 5, adoptando una conformación $\alpha\beta$ con una α -hélice y siete cadenas β que forman dos láminas β antiparalelas (Singh *et al.* 2005, He *et al.* 2012). LEAPs_5C recombinantes purificadas protegen *in vitro* la actividad enzimática de la lactato deshidrogenasa (LDH) durante el proceso de desecación y subsecuente rehidratación (Dang *et al.* 2014, Hu *et al.* 2019).

Bases moleculares de la tolerancia ante el estrés abiótico en Agave spp. Muchas especies del género *Agave* crecen en ambientes áridos (Nobel 1988, Gentry 2003) y se consideran como tolerantes al calor y a la sequía (Nobel & Smith 1983, Ramírez-Tobías *et al.* 2014, Stewart 2015). Se han descrito las adaptaciones morfofisiológicas que poseen como plantas suculentas para evitar la pérdida excesiva de agua, desde su gruesa cutícula hasta su metabolismo CAM (Nobel & Smith 1983, Nobel & Valenzuela 1987). Sin embargo, a pesar de que durante la pasada década transcriptomas de varias especies de agaves se han secuenciado (Gross *et al.* 2013, Cervantes-Pérez *et al.* 2018, Huang *et al.* 2019, Sarwar *et al.* 2019), existen pocos estudios enfocados a conocer las bases moleculares de su tolerancia al estrés abiótico (Luján *et al.* 2009, Deng *et al.* 2019, Shakeel *et al.* 2013). Conocer dichas bases permitirá una mejor comprensión de su fisiología a nivel molecular, así como de las estrategias evolutivas de este género; herramientas necesarias para el mejoramiento genético. Es posible hipotetizar diferencias en los patrones de expresión de los genes “efectores” involucrados en conferir tolerancia al estrés abiótico, o bien en las características bioquímicas de proteínas efectoras, como las LEAPs.

De acuerdo al conocimiento existente que identifica a las LEAPs como efectores ubicuos con participación en la tolerancia al estrés abiótico, proponemos que estas proteínas realizan un papel similar en los agaves. Para elucidar si sus propiedades bioquímicas son similares o se han diferenciado respecto a las de otras especies vegetales, en este trabajo identificamos secuencias codificantes para LEAPs a partir de datos transcriptómicos de *A. tequilana* y se clasificaron con base en sus dominios conservados, verificando la presencia de motivos presentes en proteínas LEA ya caracterizadas. También analizamos con herramientas bioinformáticas sus características moleculares y fisicoquímicas tales como la hidrofilicidad y su propensión al desorden estructural. Adicionalmente evaluamos su similitud y diferenciación evolutiva respecto a LEAPs de otras especies vegetales y describimos el patrón de acumulación de sus transcritos en diferentes órganos de *A. tequilana* y en respuesta a calor.

Materiales y métodos

Identificación de posibles genes de A. tequilana codificantes de LEAPs. La AgaveDB contiene secuencias parciales de 35,000 transcritos (ESTs del inglés *Expressed sequence tags*) de *A. tequilana* (Simpson *et al.* 2011). Mediante blastx (Altschul *et al.* 1990) de los ESTs contra la base de datos nr de NCBI, se identificaron aquellos con *hits* (secuencias similares de otras plantas) presumiblemente codificantes de *late embryogenesis abundant proteins* (LEAPs); así como los *contigs* (secuencias consenso) construidos a partir de los *clusters* (grupos) de ESTs similares entre sí. Los *contigs* del transcriptoma de seqRNA (Gross *et al.* 2013) depositado en NCBI (NCBI Resource Coordinators 2016 www.ncbi.nlm.nih.gov) se consultó mediante blastn y tblastn utilizando como *query* o *semilla* las secuencias codificantes de LEAPs previamente identificadas en la AgaveDB. Los *contigs* identificados como posibles genes de LEAPs se compararon mediante blastn contra la base de datos nr/nt de NCBI.

Ánalisis in silico de las LEAPs identificadas en A. tequilana. Los *contigs* de *A. tequilana* con *hits* a transcritos o *loci* codificantes de LEAPs, se tradujeron con ExPaSy translate (Gasteiger *et al.* 2003); los péptidos traducidos se compararon contra la base de datos nt/nr de NCBI mediante blastp, para verificar que corresponden a LEAPs por su similitud con las de otras especies vegetales. Las proteínas similares a las AteqLEAPs con e-value < 10⁻¹⁰ se alinearon usando el software Clustal Omega disponible en el EMBL-EBI (www.ebi.ac.uk/Tools/msa/clustalo). Los alineamientos se editaron usando la herramienta Jalview2 (Waterhouse *et al.* 2009) y ESPript (escript.ibcp.fr/EScript/EScript/index.php). Para

confirmar que las secuencias proteínicas de *A. tequilana* codifican LEAPs mediante la identificación de dominios, motivos y firmas conservados, se compararon mediante blastp contra las bases de datos de Pfam (El-Gebali *et al.* 2019), PROSITE (Sigrist *et al.* 2013, PANTHER (Thomas *et al.* 2003) e IntePRO (Mitchell *et al.* 2019), disponibles en el EMBL-EBI (www.ebi.ac.uk/services/proteins). Las dehidrinas se analizaron adicionalmente con el ScanProsite disponible en prosite.expasy.org. Los péptidos de tránsito se predijeron mediante el Phobious del EMBL-EBI.

Las reconstrucciones filogenéticas se realizaron comparando las AteqLEAP contra secuencias de *A. thaliana* codificantes de LEAPs ortólogas (Hundertmark & Hincha 2008) y con LEAPs homólogas de otras especies vegetales (identificadas mediante blastp contra nr del NCBI), aplicando el método de inferencia Neighbor-Joining, y calculando las distancias evolutivas con el método de corrección de Poisson. Se realizó un test de *bootstrap* de 1000 réplicas usando el paquete MegaX (www.megasoftware.net). El cladograma estructural se construyó con secuencias homólogas del UniProt/SwissProt identificadas mediante blastp, alineadas en Clustal Omega, y aplicando la matriz de distancias generada de los puntajes por pares en MEGA X.

Las características fisicoquímicas de las LEAPs de *A. tequilana* (peso molecular, punto isoeléctrico (pI), índice de inestabilidad, índice alifático e hidrofilicidad (GRAVY) se predijeron utilizando el programa ProtParam de Expasy (web.expasy.org/protparam). El análisis de carga hidropática se calculó utilizando el programa PONDR (www.ponder.com), distinguiendo entre proteínas ordenadas por la carga media absoluta.

Detección y cuantificación de transcritos de AteqLEAPs. Para comparar los niveles de transcritos de AteqLEAPs en diferentes órganos, tres plantas de *A. tequilana* Weber var. *azul* provenientes de la región de Tequila, Jalisco, establecidas en campo ($19^{\circ} 29' 50.2''$ N; $90^{\circ} 32' 38.1''$ W), a partir de hijuelos, por dos años, se colectaron en diciembre del 2018 para diseccionar órganos vegetativos. De cada planta se diseccionó la zona media de tres hojas perimetrales, de tres hojas de la zona intermedia de la roseta y de las hojas no desarrolladas enrolladas dentro del cono central o cogollo, así como el meristemo vegetativo y la zona central del tallo o piña. Las muestras se congelaron con nitrógeno líquido y se mantuvieron en ultracongelación hasta su procesamiento. Las muestras congeladas se molieron, se pesaron y se mezclaron en cantidades iguales para obtener una sola muestra de cada órgano, representativa de las tres plantas.

Para analizar el efecto del calor en los niveles de transcritos de LEAPs, en enero del 2019 se colectaron hojas a partir de tres plantas de *A. tequilana*, del mismo predio que el experimento anterior. De cada planta se colectaron nueve hojas: tres ubicadas en la zona cercana al cogollo, tres de la zona media de la roseta y tres de la zona perimetral. Tres hojas por planta, representando cada zona de la roseta, fueron tratadas durante dos horas a 25, 45 o 55 °C. Cada tratamiento se aplicó a las hojas de las tres plantas, al mismo tiempo. Se muestreó la zona media de las hojas y se congelaron con nitrógeno líquido. Las muestras de las tres plantas por tratamiento fueron trituradas y posteriormente mezcladas en cantidades iguales, para generar una sola muestra por tratamiento.

El ARN de cada muestra se extrajo con Trizol (Invitrogen, Carlsbad, California USA, cat. No. 15596026), eliminando carbohidratos con el PureLink™ RNA Mini Kit (Ambion ®, Carlsbad, California cat. 12183018A). La concentración y pureza del ARN se evaluaron con el Nanodrop 1000 (Thermo Scientific™, Waltham, Massachusetts). La integridad del ARN se visualizó mediante electroforesis en geles de agarosa (Invitrogen™ cat. 16500500) al 1.2 %. La síntesis de cDNA se realizó con SuperScript III (Invitrogen™ cat. No. 18080093) a partir de 1.6 mg de ARN total por reacción, y las amplificaciones mediante qPCR con IQ SYBRGreen Supermix (BioRad, Hercules, California, cat. 170-8880); conforme lo reportado en Monroy-González & Martínez Hernández (2019), utilizando por reacción 20 pmoles de los oligonucleótidos iniciadores descritos en la [Tabla S1](#). Los iniciadores se diseñaron con base en la secuencia de los contigs aC2782_14/590_7 (AteqLEA5B), aC380_6 (AteqLEA5C), aC891/13_585/4 (AteqDHN1-2) y aC2884_3 (AteqDHN3). Para los controles de expresión constitutiva se diseñaron iniciadores contra genes de poliubiquitina *AteqpUBQ* y del factor de elongación *AteqEF1-α*.

Las reacciones de PCR para amplificar los ADNc consistieron de un ciclo inicial de desnaturización a 95 °C por 5 min, seguido de 25-30 ciclos de amplificación de: 30 s a 95 °C, 30 s a la temperatura de alineamiento de cada iniciador reportada en la [Tabla S1](#), 30 s a 72 °C, finalizando con un ciclo a 72 °C por 7 min. Para el PCR punto final

se utilizaron 25 ciclos de amplificación para *AteqLEA5B*, *AteqDHN1-2* y 30 ciclos para *AteqLEA5C*, *AteqDHN3*, *AteqpUBQ* y *AteqEF1- α* . Los productos de amplificación se visualizaron por electroforesis en geles de agarosa al 1.2 % usando el marcador de peso molecular de 1 Kb plus (Invitrogen™ cat. 10787018). Para el qPCR los ciclos umbral (Ct) de cada reacción y las temperaturas de disociación (Tm) de los productos amplificados a partir de *AteqLEAPs* y de los genes control de expresión constitutiva (*AteqpUBQ* y *AteqEF1- α*) se determinaron con el iQ5 Optical System Software 2.0. Se incluyeron en paralelo controles negativos sin DNA molde (NTC). La acumulación relativa de los transcritos amplificados por qRT-PCR se analizó a través del método $2^{-\Delta\Delta CT}$ (Livak & Schmittgen 2001) de la media de los Cts por triplicado de cada muestra normalizada contra la media de los Cts del gen control, se calculó el error propagado del ΔCT y se calibró contra el ΔCT del tejido, gen o condición comparativa. La selección del gen control para normalizar en cada experimento se realizó calculando el $2^{-\Delta CT}$. El análisis estadístico de la acumulación relativa de los transcritos se realizó con el método de análisis de varianza para comparar múltiples medias ANOVA y TUKEY aplicando una $\alpha = 0.05 \%$.

Resultados

De una búsqueda realizada entre los hits del blastX contra los ESTs de la base de datos AgaveDB, identificamos 10 grupos de ESTs similares entre sí (*clusters*) cuyas secuencias consenso (*contigs*) presentan similitud con una LEAP. Para confirmar si estas secuencias tienen similitud con genes o transcritos reportados en otras especies, las comparamos mediante blastn contra la base de datos de nucleótidos del NCBI. En la [Tabla 1](#) se resumen los datos de similitud encontrados del mejor *hit* identificado para cada *contig* de *A. tequilana*. Encontramos que cinco de los 10 *contigs* de *A. tequilana* fueron similares a secuencias parciales de LEAPs reportadas en *A. angustifolia*. Entre ellos, identificamos que el *contig* de *aC2782_14* incluye a un EST (GW667495.1) que reportamos previamente (Martínez-Hernández *et al.* 2010). Adicionalmente, identificamos dos *contigs* que fueron altamente similares (E value = 0) a una secuencia genómica de *Asparagus officinalis*, cuyo mRNA no se ha caracterizado pero contiene un dominio LEA_2. Otros tres *contigs* fueron similares a secuencias parciales de dehidrinas (LEAPs Grupo 2) reportadas en otra asparagal (*Hyacinthus orientalis*).

Con el fin de corroborar si los 10 distintos *contigs* de *A. tequilana* corresponden a transcritos completos codificantes de LEAPs, identificamos sus marcos de lectura abiertos (ORFs) y obtuvimos la secuencia de la proteína codificada mediante traducción *in silico*. La búsqueda de dominios y motivos conservados presentes en los productos proteínicos codificados por cada *contig*, mediante comparación contra las bases de datos Pfam, PROSITE, PANTHER e InterPRO del EBI, condujo a la identificación de dominios conservados en tres tipos de LEAPs ([Tabla S2](#)). Los primeros cinco *contigs* de *A. tequilana* codifican proteínas que contienen el dominio LEA_3, por lo que se clasificaron como LEAPs del Grupo 5B, conforme la clasificación propuesta por Battaglia *et al.* (2008). Los dos *contigs* que contienen al dominio LEA_2, se clasificaron como codificantes de proteínas LEA del Grupo 5C. En estos *contigs* InterPRO identificó también la presencia del dominio WHy. Los últimos tres *contigs*, ya identificados como similares a genes para dehidrinas (DHN), codifican proteínas que contienen los dominios típicos de dehidrinas reportados en Pfam e InterPRO, así como dos motivos de dehidrinas con aminoácidos altamente conservados reportados en PROSITE. Así, se concluyó que todos los *contigs* de *A. tequilana* identificados son transcritos completos y codifican una LEAP que pertenece a los grupos 5B, 5C ó 2 (dehidrina). Los denominamos entonces *AteqLEA5(B o C) #cluster* o *AteqDHN #cluster*, en donde el número indicado al final del nombre de cada *contig* ([Tabla 1](#)) corresponde a el número de ESTs que conforma cada *cluster* y es un reflejo de la abundancia de cada tipo de transcríto en el transcriptoma analizado. Por lo que los cinco *contigs* codificantes de LEAPs 5B conforman el grupo más abundante en la base de datos AgaveDB sumando 72 transcritos secuenciados, seguido de los tres *clusters* similares a dehidrinas que suman veinte ESTs, en tanto que, los dos *clusters* similares a LEAPs 5C son los menos abundantes (10 ESTs).

Tabla 1. Mejores hits obtenidos con blastn comparando los *contigs* de *A. tequilana* contra la base de datos de NCBI.

Contig	GenBank accession	Best Hit	% Cover	E value	% Identity
<i>aC3235_35</i>	MW496331	<i>Agave angustifolia Marginata</i> LEA mRNA, partial cds KU315011.1	50	7.00E-56	80.82
<i>aC2782_14*</i>	MW496332	<i>Agave angustifolia Marginata</i> LEA mRNA, partial cds KU315011.1	54	2.00E-50	79.34
<i>aC708_10</i>	MW496333	<i>Agave angustifolia Marginata</i> LEA mRNA, partial cds KU315011.1	51	3.00E-47	79.1
<i>aC590_7</i>	MW496334	<i>Agave angustifolia Marginata</i> LEA mRNA, partial cds KU315011.1	56	7.00E-49	78.93
<i>aC734_6</i>	MW496335	<i>Agave angustifolia Marginata</i> LEA mRNA, partial cds KU315011.1	55	9.00E-48	78.51
<i>aC380_6</i>	MW496336	PREDICTED: <i>Asparagus officinalis</i> uncharacterized (LOC109838014), mRNA XM_020406424.1	100	0	87.72
<i>aC746_4</i>	MW496337	PREDICTED: <i>Asparagus officinalis</i> uncharacterized (LOC109838014), mRNA XM_020406424.1	99	0	84.89
<i>aC891_7_13</i>	MW496338	<i>Hyacinthus orientalis</i> dehydrin mRNA, partial cds AY389583.1	79	5.00E-99	72.02
<i>aC585_10_4</i>	MW496340	<i>Hyacinthus orientalis</i> dehydrin mRNA, partial cds AY389583.1	66	7.00E-98	72.02
<i>aC2884_3</i>	MW496339	<i>Hyacinthus orientalis</i> dehydrin mRNA, partial cds AY389583.1	61	2.00E-85	71.61

El último número del nombre del *contig* indica el total de ESTs incluidos en ese *cluster*.

*Este *cluster* incluye al EST previamente en el GenBank (GW667495.1) por Martínez-Hernández *et al.*(2010).

Una búsqueda en los 195,967 *contigs* de seqRNA de *A. tequilana* (Gross *et al.* 2013) mediante blastn y tblastn, utilizando como query los *contigs* de *AteqLEA/DHN* y sus secuencias proteínicas correspondientes, condujo a identificar un número relativamente limitado de *hits* (Tabla S3). La proporción relativa de genes de LEAPs 5B : 5C : DHN observada resultó similar entre ambos transcriptomas. No fue posible en el marco de este trabajo identificar variantes proteínicas de LEAPs 5B, 5C o dehidrinas en ese transcriptoma, ya que los *contigs* reportados presentan numerosos codones de paro y quimeras. Secuencias parciales idénticas a los *contigs* encontrados en la AgaveDB también se identificaron en el transcriptoma de seqRNA.

En la Figura 1 se muestra una comparación entre los distintos tipos de LEAPs identificadas en *A. tequilana* y

LEAPs de otras especies reportadas en SwissProt, incluyendo *A. thaliana*. Las 10 AteqLEAPs conforman tres grupos acorde a sus dominios y clasificación, formando clados con los genes de *A. thaliana* codificantes de LEAPs que contienen los dominios de Pfam LEA_3, LEA_2 y Dehidrinas (Hundertmark & Hincha 2008). El cladograma muestra que los cinco genes *AteqLEA5B* se agrupan con dos de los cuatro genes de *A. thaliana* codificantes de LEAPs 5B (locus At1g02820 y At4g02380), los cuales son sus posibles ortólogos; pero forman una rama independiente y diversificada. La diversificación de estos genes potencialmente se debe a duplicaciones génicas o genómicas; o bien, corresponden a alelos, ya que los transcriptos de *A. tequilana* provienen de un conjunto de plantas. Los dos genes *AteqLEA5C* se agrupan con el locus At2g440160 (posible ortólogo), a diferencia de la LEA_5C de *C. plantagineum* que se agrupa junto con otras dicotiledóneas con los otros dos *loci* de *A. thaliana* codificantes de proteínas LEA_5C. La proteína membranal MJ1138 de la arqueobacteria *Methanocaldococcus jannaschii* contiene un dominio WHy similar a LEA_2 y se separa de las LEAPs de plantas. Los tres genes de dehidrinas de *A. tequilana* forman un clado con el gen codificador de la DHN1 de *O. sativa* y cuatro *loci* de *A. thaliana*, separándose de los otros seis *loci* codificantes de dehidrinas en *A. thaliana*.

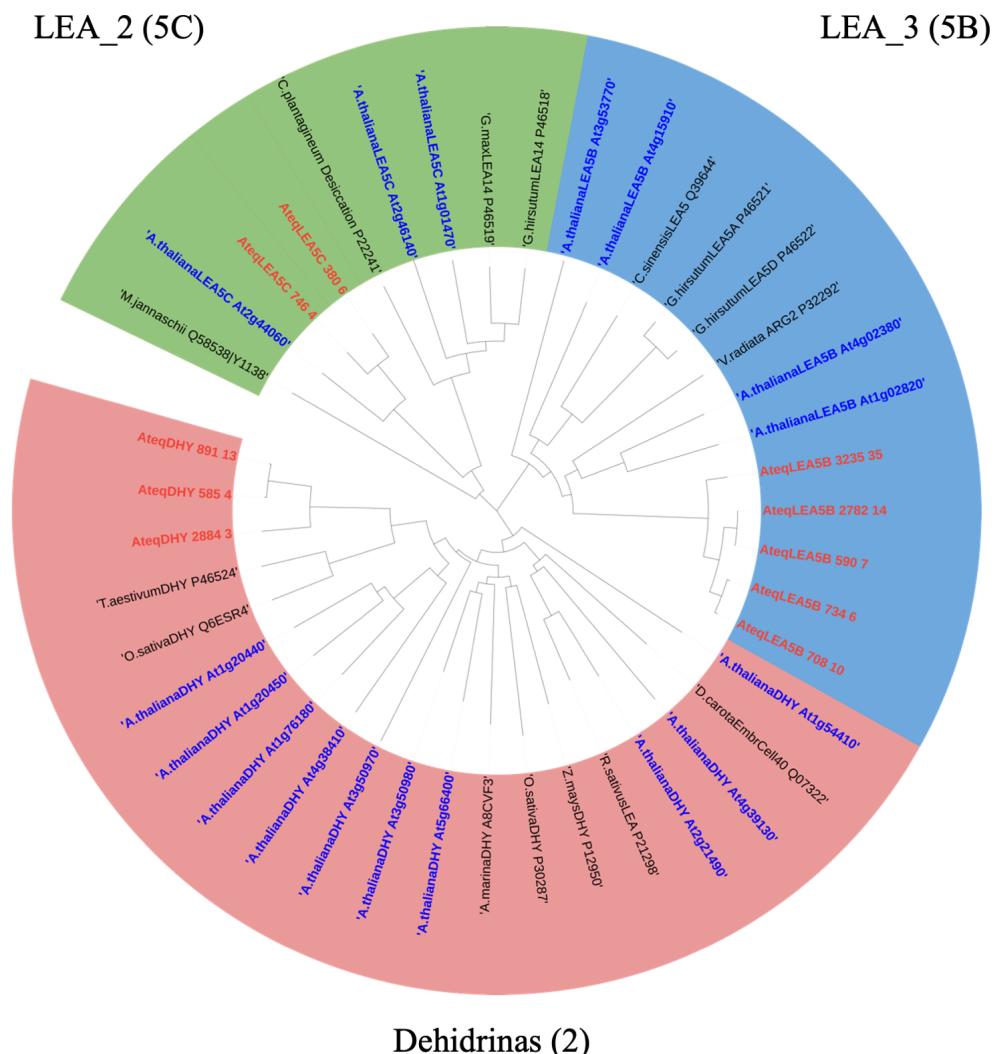


Figura 1. Cladograma comparativo de familias (Grupos) de LEAPs con diferente secuencia y estructura, halladas en el transcriptoma de *A. tequilana*. Las tres familias identificadas, conforme los dominios de Pfam, están indicadas por colores: azul (LEA5B, dominio LEA_3), verde (LEA5C, dominio LEA_2), salmón (Dehidrinas). En rojo las LEAPs de *A. tequilana* identificadas en este trabajo y en azul locus de *A. thaliana* codificantes de LEAPs. Los nombres en las secuencias de *A. thaliana* identifican LEAPs funcionalmente caracterizadas cercanas a las AteqLEAP.

Para corroborar si las distintas LEAPs identificadas de *A. tequilana* contienen los motivos conservados descritos en otras especies vegetales, comparamos su secuencia aminoacídica con proteínas con alta similitud de otras especies, las cuales se identificaron mediante blastp; así como contra secuencias de *A. thaliana* pertenecientes a estos grupos (Hundertmark & Hincha 2008).

El alineamiento de LEAPs 5B de la [Figura 2](#) muestra que las secuencias de agaves son de menos de 100 aa y contienen zonas conservadas respecto a las LEAPs 5B de otras especies, aunque su porcentaje de identidad es relativamente bajo porque también presentan zonas variables respecto a otras especies. Sin embargo sí contienen el dominio LEA_3 de Pfam, conservan sin ambigüedades el motivo W (Hundertmark & Hincha 2008) y presentan los motivos MAaRS, RRGYA4 y DAELR identificados por Singh & Graether (2020). Aunque en el motivo DAELR las AteqLEA5B presentan un par de variaciones respecto al consenso reportado, y en el dominio MAaRS cuatro AteqLEA5B carecen de la Ser consenso de la cuarta posición, la cual está sustituida por una Phe. Así mismo, la Lys está sustituida por una Asn y los últimos seis aminoácidos del consenso están ausentes (*gap*) entre la posición 16-17 de las AteqLEA5B respecto a las de otras especies. Las AteqLEA5B, como otras LEAPs de este grupo, presentan un péptido señal transmembranal y mediante *phobious* se predice que no son citoplasmáticas. Las dos argininas de la posición 17-18 corresponden al sitio de corte para escindir el péptido señal.

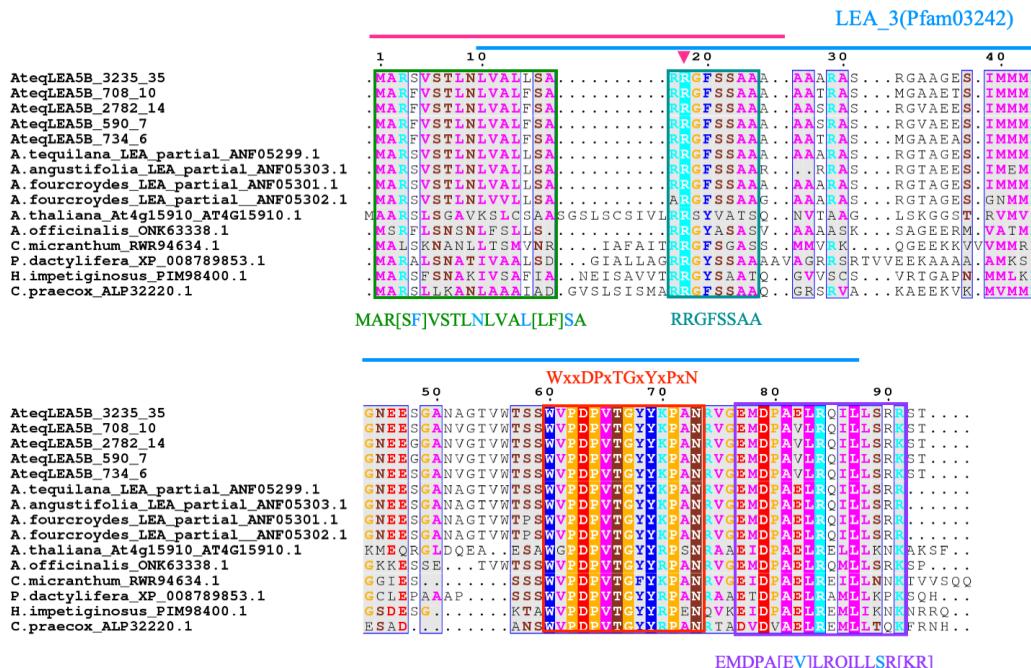


Figura 2. Dominio y motivos conservados en AteqLEAPs del Grupo 5B. La línea azul muestra el dominio Pfam (03242) de las LEA_3. La línea y triángulo rosa indican el péptido señal y el sitio de corte entre el par de argininas (R). En recuadros de color motivos conservados en proteínas LEA_3 identificados en AteqLEA5B: en rojo el motivo W: (W-x(2)-D-P-x-T-G-x-[WY]-x-P-x-[DGNST]) (Hundertmark & Hincha 2008), las x indican posiciones que pueden estar ocupadas por cualquier residuo; en verde el motivo MARS, en azul/verde el RRGYA₄, y en morado el DAELR (Singh & Graether 2020). Las secuencias de los motivos (en el color del recuadro) indican los aminoácidos encontrados en *A. tequilana* que corresponden a los consensos. Las letras azules muestran aminoácidos en *A. tequilana* distintos a los consensos.

En la [Figura 2](#) también se observa que los cinco genes *AteqLEA5B* codifican proteínas muy similares entre sí, en tanto que los *contigs* aC2782_14 y aC590_7 codifican exactamente la misma secuencia proteínica, aunque difieren entre sí en la secuencia nucleotídica. AteqLEA5B_3235_35 es una proteína casi idéntica a una secuencia parcial de LEAP de *A. tequilana* previamente reportada (ANF05299.1), pero presenta una sustitución A/T en el aminoácido 34. AteqLEA5B_3235_35 muestra algunas variaciones de aminoácidos respecto a las otras cuatro AteqLEA5B aquí

identificadas. Esas variaciones se encuentran también presentes en secuencias reportadas en otros agaves, por lo que parecen corresponder a esta isoforma, como también indica el filograma de la [Figura S1](#). Aunque la identidad proteínica de las AteqLEA5B respecto a las LEA_5B de otras plantas es relativamente baja, el árbol de la [Figura S1](#) muestra que, acorde a la relación filogenética de las especies incluidas, las AteqLEA5B son homólogas a la de *A. officinalis* y se agrupan con las LEA_5B de otras monocotiledóneas incluyendo asparagales como las orquídeas y come-línidas (palmas, gramíneas y bromelias), separándose de las LEAPs del Grupo 5B de dicotiledóneas, entre las cuáles se encuentran las secuencias de *A. thaliana*. Aunque magnólicas como *Cinnamomum micranthum* y *Chimonanthus praecox* quedaron incluidas en la rama de las monocotiledóneas y las orquídeas no se agruparon con las Asparagales.

Conforme el análisis de la composición de aminoácidos y la predicción de las características fisicoquímicas de las AteqLEA5B ([Tabla 2](#)), estas proteínas cumplen con las características descritas para su grupo: una masa molecular menor de 10 KDa, baja abundancia de aminoácidos cargados, relativamente hidrofóbicas (GRAVY Index) y ordenadas ([Figura S2A-B](#)); un pI básico. Conforme el índice de inestabilidad calculado a partir de la composición de aminoácidos ([Tabla 2](#)), se predice que cuatro AteqLEA5B tienen una estructura estable, pero AteqLEA5B_35 forma una estructura inestable, con o sin péptido de tránsito.

Tabla 2. Características fisicoquímicas de la LEAPs de *A. tequilana* calculadas en PONDR.

	Number of AA	Molecular weight	Theoretical pI	AA+	AA-	Instability index	Classification	GRAVY
AteqLEA5B_3235_35	93	9788.11	10.37	7	11	48.31	unstable	-0.167
AteqLEA5B_2782_14	93	9938.37	10.14	7	11	38.34	stable	-0.055
AteqLEA5B_708_10	93	9901.36	10.14	6	10	29.37	stable	0.015
AteqLEA5B_590_7	93	9938.37	10.14	7	11	38.34	stable	-0.055
AteqLEA5B_734_6	94	9972.44	10.14	6	10	29.16	stable	0.034
AteqLEA5C_380_6	316	35168.15	4.89	56	41	15.98	stable	-0.338
AteqLEA5C_746_4	316	35203.21	4.74	59	40	16.91	stable	-0.302
AteqDHN_891_7_13	240	26638.69	4.94	64	44	52.77	unstable	-1.111
AteqDHN_585_10_4	240	26606.63	4.93	64	44	52.21	unstable	-1.1
AteqDHN_2884_3	240	26310.23	5.03	59	41	52.18	unstable	-1.07

En el caso de las LEAPs del Grupo 5C de *A. tequilana*, el alineamiento de la [Figura 3](#) mostró que son proteínas de peso molecular relativamente alto (más de 300 aa), muy conservadas, que contienen solo un dominio Pfam de LEA_2 y se identificó el motivo reportado por Hundertmark & Hincha (2008) con variaciones en dos aminoácidos. Este análisis mostró que la proteína de *A. thaliana* difiere en la región N-terminal, por la inserción de nueve aminoácidos. Aunque son muy similares, las LEAPs del Grupo 5C de monocotiledóneas y dicotiledóneas se separan en dos distintas ramas dentro del filograma de la [Figura S3](#). En cuanto a sus características fisicoquímicas, las AteqLEA5C son distintas a las otras AteqLEA atípicas (5B) ya que tienen pI ácido, un número de aminoácidos cargados relativamente alto y son ligeramente hidrofílicas, menos que las dehidrinas pero más que las AteqLEA5B. Su índice de inestabilidad es el más bajo, por lo que son de estructura estable y ordenadas ([Tabla 2](#), [Figura S2C-D](#)).

Las AteqDHN presentaron alta variabilidad respecto a las dehidrinas similares de otras especies, identificadas por blastp ([Figura 4](#)). Sin embargo, se identificó el dominio Pfam de dehidrinas (00257), así como los segmentos F,

S y tres K. Estos segmentos presentan algunas ambigüedades respecto a los consensos reportados (Battaglia *et al.* 2008, Riley *et al.* 2019), por lo que en PROSITE se detectó solo un motivo K. El análisis filogenético ([Figura S4](#)) mostró que dos AteqDHN son muy similares entre sí, por lo que pueden corresponder a dos alelos, mientras que la AteqDHN3 corresponde a un segundo gen, todos ellos homólogos a genes de dehidrinas reportados en Asparagales. Este grupo de dehidrinas se separa de los de otras monocotiledóneas, incluyendo comelínidas y gramíneas. Las dehidrinas de dicotiledóneas forman una rama independiente. Conforme lo esperado, las AteqDHN son las AteqLEAPs con el mayor número de aminoácidos cargados, su pI es ácido, son muy hidrofílicas y su estructura se predice como inestable y desordenada ([Tabla 2](#), [Figura S2E-F](#)).

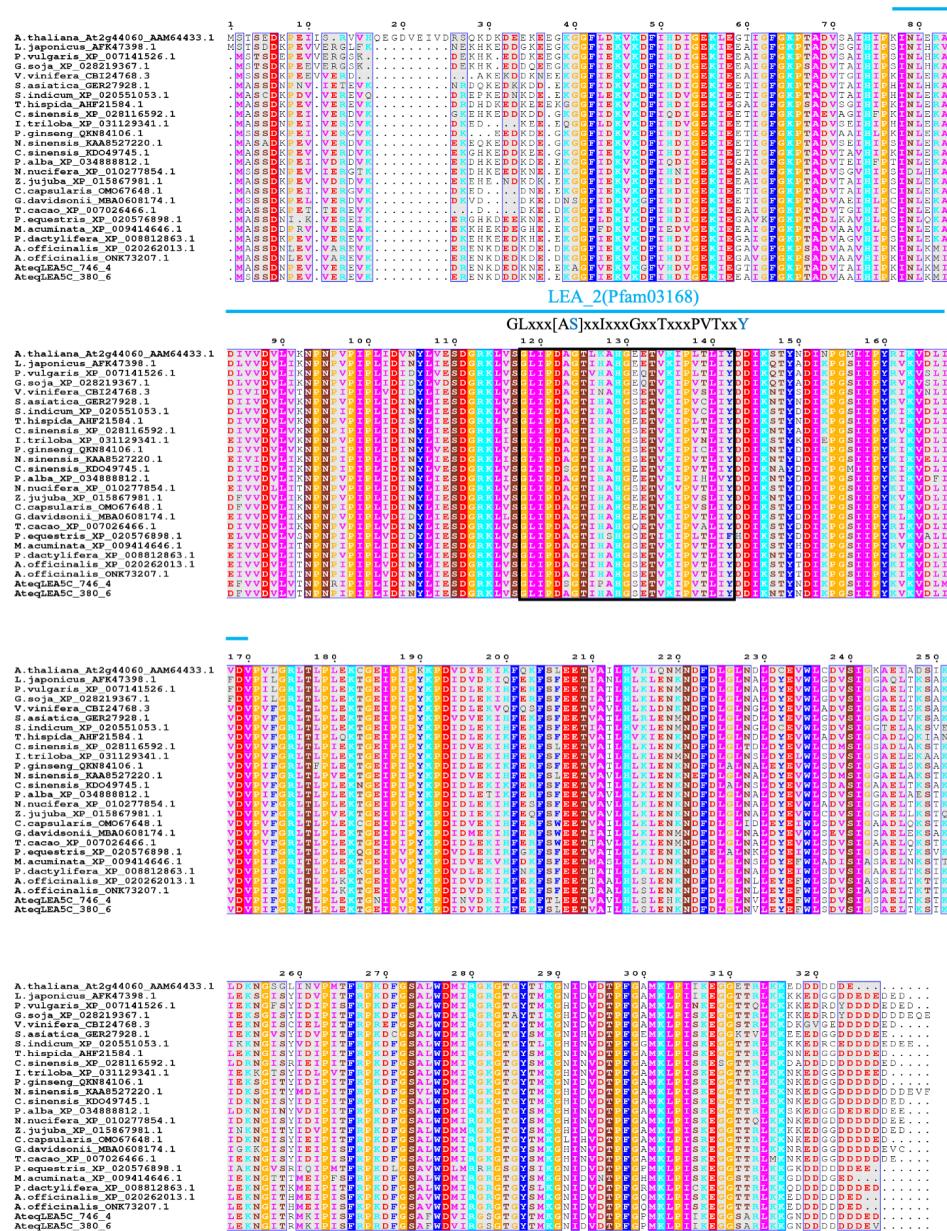


Figura 3. Dominio y motivo conservados en AteqLEAPs del Grupo 5C. La línea azul indica el dominio Pfam (03168) de las LEA_2. El recuadro negro indica el motivo G-L-x(2)-[AIL]-x-[IV]-x-[GV]-x(2)-[PT]-x-[PS]-[ILV]-[NPST]-x(2)-[GI] identificado en *A. thaliana* (Hundertmark & Hincha 2008). La secuencia en letras negras muestra los aminoácidos encontrados en *A. tequilana* que corresponden al consenso de *A. thaliana*. Las letras azules muestran aminoácidos en *A. tequilana* distintos al consenso. La x indica una posición que puede estar ocupada por cualquier residuo.

A. tequilana LEA proteins

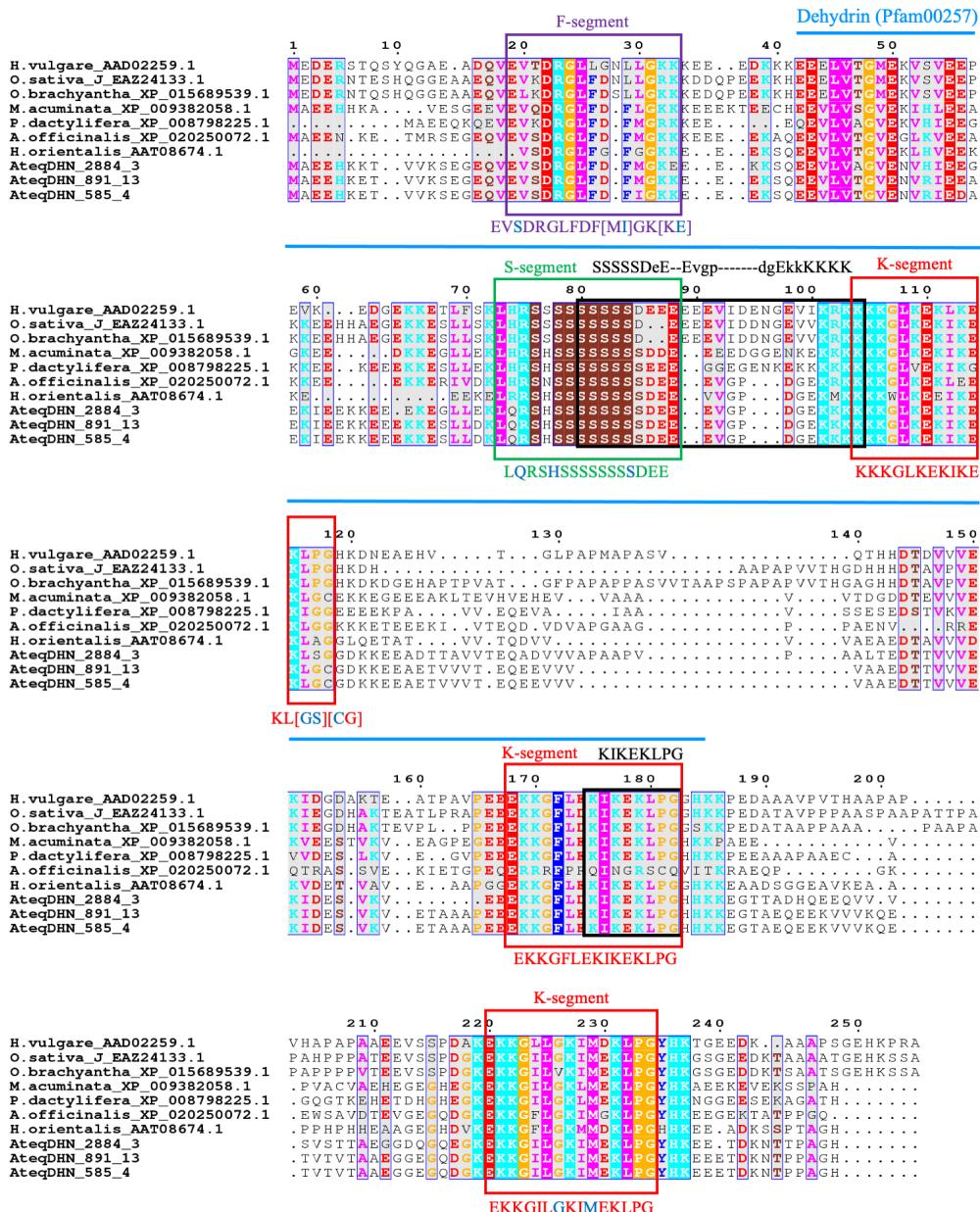


Figura 4. Dominio y motivos conservados en AteqLEAPs del Grupo 2 (Dehidrinas). La línea azul indica el dominio Pfam (00257). Los recuadros y letras negras indican las firmas encontradas en PROSITE. En recuadros y letras de colores los motivos conservados que definen la estructura de las dehidrinas: en rojo los segmentos-K: [EKR]-KKG-[ILMF]-[MLK]-[DE]-KIK-[ED]-K-[LI]-PG y en verde el segmento-S: [LH]-HRSG-[SH]-[WS]-SSSSS-[DE]-[DE]-[DE] (Battaglia *et al.* 2008); en morado el segmento-F: E-[VT]-[KTQ]-DRG-[LM]-F-[DG]-F-[LFM]-[GK]-KK (Strimbeck 2017, Riley *et al.* 2019). Las secuencias de los motivos (en el color del recuadro) indican los aminoácidos encontrados en *A. tequilana* que corresponden a los consensos. Las letras azules aminoácidos de *A. tequilana* que difieren de los consensos reportados.

Con el fin de determinar si las LEAPs de *A. tequilana* se expresan en órganos vegetativos funcionalmente diferenciados, evaluamos el nivel de expresión de sus transcriptos, mediante qRT-PCR y punto final. Este análisis se realizó con ARN extraído a partir de hojas expandidas de la zona intermedia y externa de la roseta, hojas en desarrollo no fotosintéticas enrolladas dentro del cono central (cogollo), de tallo (piña) y de meristemo colectados a partir de plantas establecidas en campo por dos años ([Figura 5](#)). Se establecieron las condiciones de reacción ([Tabla S1](#)) y se verificó que para cada reacción se genera un solo producto, midiendo su Tm ([Figura S5](#)) y corroborando su tamaño

mediante electroforesis. Se identificó al gen *AteqpUBQ* como control endógeno de expresión constitutiva en este experimento ya que la abundancia media de ese transcripto es similar en todos los órganos evaluados, sin mostrar diferencias estadísticamente significativas (Figuras 6A, S6-A); a diferencia de *AteqEF1- α* que se acumula en niveles significativamente diferentes en distintos órganos (Figura S6-B).

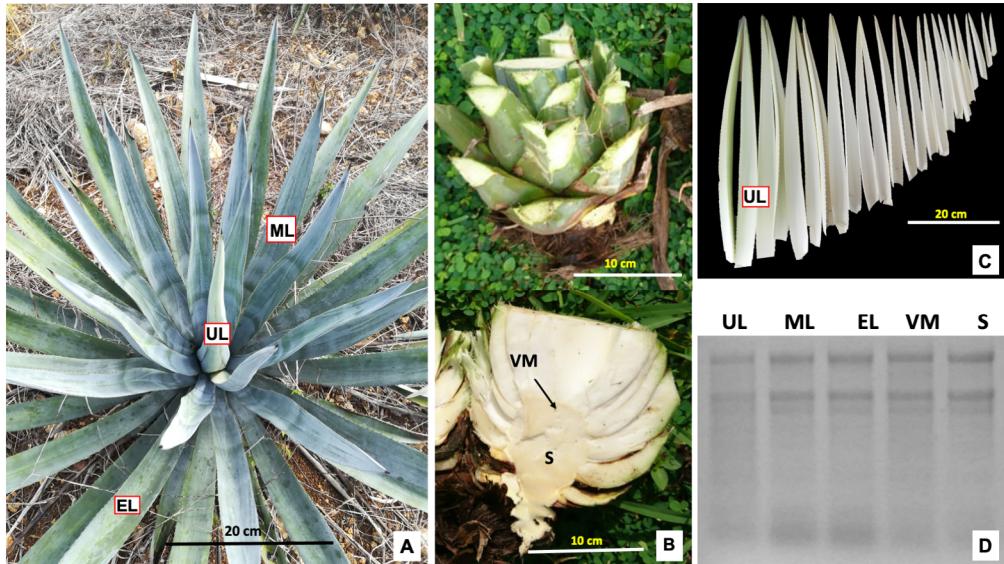


Figura 5. Tejidos de *A. tequilana* Weber var. *azul* colectados y ARN extraído para los análisis de expresión de genes codificantes de proteínas LEAPs. A: Ejemplo de planta colectada, señalando el cono central (cogollo) contenido hojas sin desplegar en proceso de desarrollo (UL), hojas desplegadas de la región intermedia de la roseta (ML), hojas desplegadas de la región externa de la roseta (EL). B: Ejemplo de tallo (piña en desarrollo) colectado (arriba), mostrando en el corte longitudinal (abajo) el tejido de almacenamiento (S) y meristemático (VM) colectados. C: hojas en desarrollo contenidas dentro del cono central (UL). D: ARNs obtenidos a partir de los tejidos señalados.

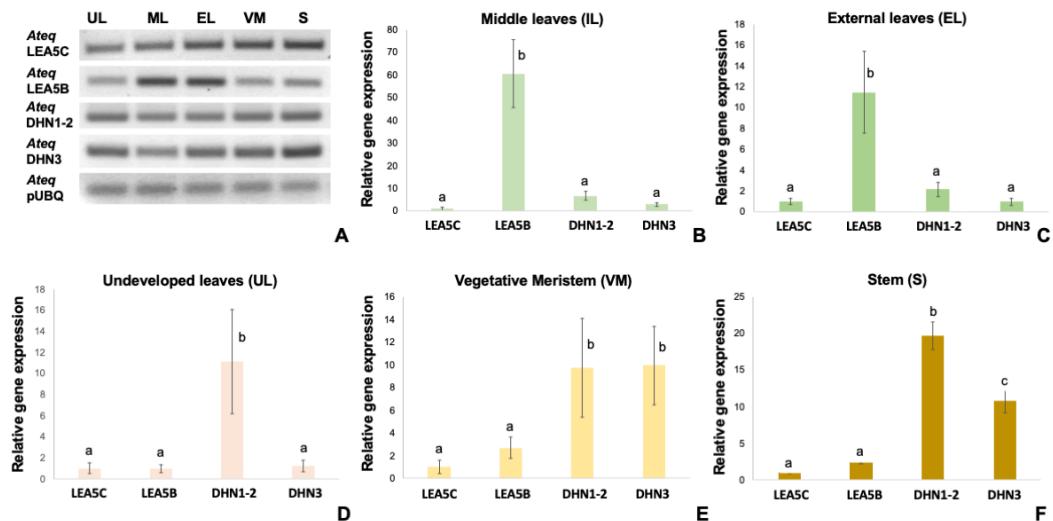


Figura 6. Perfil de expresión génica de LEAPs de *A. tequilana* en diferentes órganos vegetativos. A. Análisis con RT-PCR punto final de la acumulación de transcriptos *AteqLEA5B* (2782_14/590_7), *AteqLEA5C* (380_6), dehidrinas (*AteqDHN1-2*: *AteqDHN_891_7_13* y *AteqDHN_585_10_4*; *AteqDHN3*: *AteqDHN_2884_3*) y poliubiquitina (*AteqpUBQ*) en hojas no desplegadas o cogollo (UL), hojas de posición intermedia en la roseta (ML), hojas externas en la roseta (EL), meristemo vegetativo (VM) y tallo o piña (S). B-F: Acumulación relativa de los transcriptos de AteqLEAPs en los diferentes órganos de *A. tequilana*, determinada por qRT-PCR. Las gráficas muestran el $2^{-\Delta\Delta Ct}$ normalizando contra el gen control de expresión constitutiva *AteqpUBQ*, comparando contra el nivel de acumulación de *AteqLEA5C* como calibrador en cada tejido y calculando el error propagado (barras). Columnas con distintas letras, tienen medias significativamente diferentes, de acuerdo con la prueba de Tukey a una $P \leq 0.05$.

Al comparar la expresión relativa de cada gen en distintos órganos por RT-PCR punto final ([Figura 6A](#)) y por qRT-PCR ([Figura S7](#)); así como al determinar la abundancia relativa de los distintos *Ateq*LEAPs en un mismo órgano ([Figura 6B-F](#)); encontramos que todos los *Ateq*LEAP se expresaron en todos los órganos analizados, pero a diferente nivel y con un patrón diferencial. En hojas intermedias y externas los *Ateq*LEA5B presentaron un mayor nivel de acumulación de hasta 60 veces más que los transcriptos de otras LEAPs ([Figuras 6A-C, S7B](#)); pero se expresaron a bajo nivel en otros órganos ([Figura 6D-F](#)). Por el contrario, en piña, meristemo y en hojas en desarrollo (cogollo) los transcriptos para las dehidrinas, especialmente *Ateq*DHN1-2, fueron los más abundantes, acumulándose de 10 a 20 veces más que los transcriptos de otras LEAPs ([Figura 6D-F](#) y [S7C](#)). *Ateq*DHN3 tuvo un patrón de expresión similar a *Ateq*DHN1-2 pero a más bajo nivel ([Figuras 6E-F, S7-D](#)). Los transcriptos *Ateq*LEA5C mostraron baja expresión en todos los órganos ([Figura 6B-F](#)).

Para evaluar si la expresión de *Ateq*LEAPs se afecta por tratamientos con calor, su acumulación relativa se cuantificó en hojas disectadas sometidas por dos horas a 45 o 55 °C, comparando contra sus niveles de expresión en hojas mantenidas a 25 °C. En este experimento se normalizó contra *Ateq*EF1- α , el cual mostró niveles de acumulación menos variables que *Ateq*UBQ ante estos tratamientos ([Figura S6C-D](#)). En la [Figura 7A](#) se observa que los transcriptos de *Ateq*LEA5B mantuvieron su alto nivel de acumulación, aún en hojas tratadas a 45 o 55 °C. Mientras que los de *Ateq*DHN1-2, que en esta mezcla de hojas mostraron un nivel similar al de *Ateq*LEA5B a 25 °C, decayeron por efecto del tratamiento con calor, de forma estadísticamente significativa ([Figura 7C](#)). *Ateq*LEA5C y *Ateq*DHN3, que son de bajo nivel de expresión en hojas, no mostraron un cambio significativo por efecto del tratamiento con calor ([Figura 7B, D](#)).

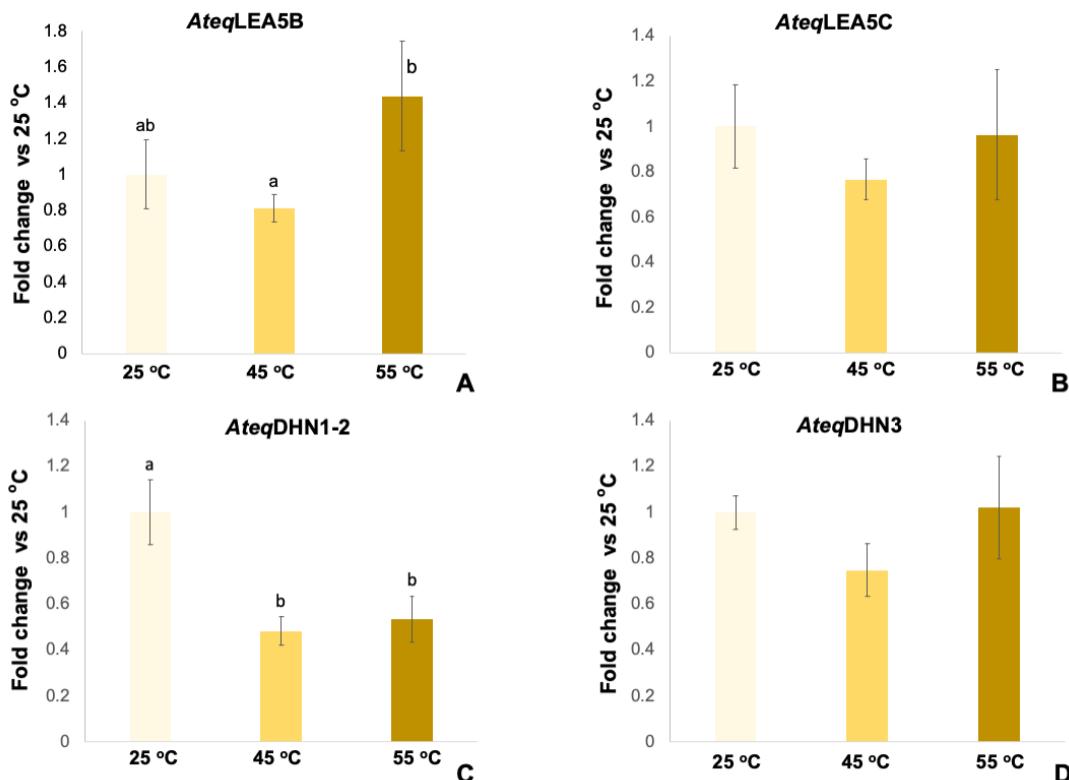


Figura 7. Efecto del tratamiento con calor sobre el nivel de transcriptos de LEAPs en hojas de *A. tequilana*. A-D: Acumulación relativa de cada *Ateq*LEAP en cada tratamiento, determinada por qRT-PCR. Las gráficas muestran el $2^{-\Delta\Delta Ct}$ normalizando contra el gen control de expresión constitutiva *Ateq*EF1- α , comparando contra el nivel de transcriptos a 25 °C y calculando el error propagado (barras). Columnas con distintas letras tienen medias significativamente diferentes, de acuerdo con la prueba de Tukey a una $P \leq 0.05$. La comparación de medias (ANOVA) en los niveles relativos de acumulación de *Ateq*LEA_5C (B) y DNH_3 (D) no mostró cambios estadísticamente significativos bajo ningún tratamiento.

Discusión

El número de genes y familias de LEAPs identificados en A. tequilana es reducido. La búsqueda dirigida de proteínas LEA en datos transcriptómicos de *A. tequilana* condujo a la identificación de relativamente pocos genes codificantes de LEAPs, pertenecientes a solo tres de los ocho familias reportadas en Pfam: Grupo 5B (LEA_3), Grupo 5C (LEA_2) y Grupo 2 (Dehidrinas); tanto en la base de datos de ESTs, como en los de seqRNA. En ambos transcriptomas es similar la abundancia relativa de transcriptos de los distintos tipos de LEAPs, siendo más abundantes los que codifican proteínas LEA5B, seguidos de los transcriptos para dehidrinas y finalmente los de LEA5C, los cuales mostraron el más bajo nivel. Esto resulta contrario a la hipótesis de que este tipo de genes podría ser abundante y diverso en el transcriptoma de los agaves considerando que las LEAPs podrían significar una estrategia evolutiva de protección particularmente diseminada en algunas especies. Para *X. viscosa* se ha planteado que la sobre-representación de algunos tipos específicos de LEAPs, cuyos miembros, estructuras y funciones se han expandido y diversificado, está evolutivamente asociada con la contribución de estos genes a su adaptación al estrés hídrico (Costa *et al.* 2017, Artur *et al.* 2019).

En los genomas de dicotiledóneas como *Malus domestica*, *Glycine max*, *Brassica* spp. los genes codificantes de LEAPs se han expandido (Artur *et al.* 2019), alcanzando hasta 242 genes en *Gossypium hirsutum* (Magwanga *et al.* 2018). Entre las monocotiledóneas secuenciadas a nivel genómico, las gramíneas presentan una mayor abundancia relativa de genes LEAPs, conteniendo algunas especies como *Zea mays* y *Hordeum vulgare* más de 100 genes, mientras que el genoma de *Phalaenopsis equestris* (asparagal) contiene 69 genes de LEAPs. El limitado número de genes codificantes de LEAPs identificados en *A. tequilana* y el bajo número de familias génicas observadas sugiere que están subrepresentadas en el transcriptoma analizado, posiblemente porque fue secuenciado a partir de plantas cultivadas, sin signos de estrés. Otra posibilidad es que su participación en la tolerancia de esta especie, adaptada a desarrollarse en condiciones de estrés ambiental, ocurre principalmente mediante tipos específicos de LEAPs, como los aquí identificados. Los análisis genómicos y transcriptómicos han revelado que, si bien el número de genes de LEAPs es extenso en plantas, no todas las LEAPs se expresan en respuesta a estrés abiótico y su patrón de expresión titular es variable (ej: Kimura *et al.* 2003, Hundertmark & Hincha 2008, Nagaraju *et al.* 2019). Más aún, muchas LEAPs se expresan solo durante el proceso de desecación en la embriogénesis y no en tejidos vegetativos. Los transcriptomas de agaves actualmente disponibles no incluyen esos tejidos. Un futuro análisis de los genomas de *Agave* spp. en proceso de ensamble, permitirán corroborar el tamaño y diversidad de las familias génicas de LEAPs en este género.

Si bien una abundancia y participación limitada de las LEAPs en los agaves es inesperada, en otras especies xerófitas como la planta de la resurrección *C. plantagineum* se ha propuesto que; además de los mecanismos comunes que participan en la tolerancia a la desecación; podrían existir mecanismos moleculares distintos seleccionados evolutivamente por las condiciones limitantes, como los denominados genes restringidos taxonómicamente (TRGs) (Giarola & Bartels 2015). La identificación de otros posibles efectores taxonómicamente seleccionados en *Agave* spp. ante estrés ambiental, así como de mecanismos reguladores alternos, debe ser considerada.

Subrepresentación de las LEAPs del grupo 5C (LEA_2) en A. tequilana. Aunque las LEAPs del Grupo 5 fueron denominadas como *atípicas* debido a sus características bioquímicas diferentes (hidrofóbicas y más estructuradas) (Galau *et al.* 1993, Battaglia *et al.* 2008), hoy conocemos que esta familia es la más extendida en los genomas vegetales y se ha propuesto que, a través de duplicaciones recurrentes en tandem se ha diversificado funcionalmente contribuyendo a la adaptación de las plantas a ambientes con limitación de agua (Artur *et al.* 2019). Además, LEAPs que contienen al dominio WHY (*Water stress and Hypersensitive response*), similar al LEA_2, se encuentran en bacterias y arqueobacterias en las cuales puede haberse adquirido recientemente por transferencia horizontal (Mertens *et al.* 2018).

Sin embargo, no muchas LEAPs 5C de plantas se han caracterizado funcionalmente. Por ejemplo, transcriptos con dominio LEA_2 de *X. viscosa* mostraron respuesta a la sequía, incrementando sus niveles rápidamente ante este

tratamiento. Después de la rehidratación, los transcritos disminuyeron. Se sugiere que una función de éstas y otras LEAPs, con comportamiento similar, está relacionada con mantener la homeostasis celular durante la deshidratación (Costa *et al.* 2017). En *Ipomoea pes-caprae* L., planta halófita marina, el nivel de transcritos de una LEA_2 se incrementa por tratamientos con sal, choque osmótico, oxidantes, congelamiento y ácido abscísico. La sobre-expresión de este gen incrementa la tolerancia a estrés salino y sequía en plántulas de *A. thaliana*. Se sugiere que este gen está involucrado en la respuesta de la planta mediando la homeostasis hídrica o como un *scavenger* de las especies reactivas de oxígeno (Zheng *et al.* 2019). Genes codificantes de las LEAPs 5C de arroz también han mostrado participar en la tolerancia a estrés abiótico (He *et al.* 2012, Hu *et al.* 2019).

Aunque las LEAPs del grupo 5C están altamente conservadas en la naturaleza, se han multiplicado ampliamente en plantas y al menos algunas están asociadas a una función protectora contra estrés ambiental; en *A. tequilana* solo detectamos transcritos de dos isoformas de AteqLEA5C, con bajo nivel de acumulación en los tejidos evaluados y sin respuesta al tratamiento por calor. La baja diversidad y expresión observada sugiere, inesperadamente, poca relevancia de esta subfamilia en *A. tequilana* o bien, una función específica en órganos distintos a los estudiados. Sin embargo, en otras monocotiledóneas como sorgo y arroz las LEA5C sí se expresan en tejidos vegetativos (He *et al.* 2012, Nagaraju *et al.* 2019). La falta de inducción de las AteqLEA5C con el tratamiento con calor contrasta con la inducción observada en miembros de esta familia en otras plantas, por lo que la participación de las AteqLEAP5C bajo otras condiciones de estrés ambiental como sequía o frío será analizada.

Potencial papel de las dehidrinas SK3 de A. tequilana protegiendo tejidos no fotosintéticos en desarrollo, de reserva y meristemáticos. Las dehidrinas, LEAPs altamente hidrofílicas e intrínsecamente desordenadas, conforman el grupo de LEAPs más estudiado en diversas especies (ej: Puhakainen *et al.* 2004, Hughes *et al.* 2013, Cuevas-Velázquez *et al.* 2014, Graether & Boddington 2014, Riley *et al.* 2019, Cui *et al.* 2020). DHN1 y WCOR410 son dehidrinas SK3 identificadas en *Oryza sativa* y *Triticum aestivum*, inducibles por frío (Guo *et al.* 1992, Lee *et al.* 2005), homólogas a la *cold-regulated* COR47 de *A. thaliana* (At1g20440). Conforme lo reportado en UniProtKB, DHN1 se expresa en panícula y también se induce por sequía, salinidad y ácido abscísico; mientras que WCOR410 se expresa en raíz, espigas y hojas durante la aclimatación a frío y se induce en variedades tolerantes a congelamiento. Las dehidrinas de *A. tequilana* son homólogas a las proteínas WCOR410, por lo que su respuesta a frío deberá ser evaluada.

Las proteínas AteqDHN identificadas contienen un segmento-S y tres K. Sin embargo, dos de sus segmentos-K presentan variaciones respecto a la firma identificada en PROSITE y al consenso reportado por Battaglia *et al.* (2008). Considerando que las sustituciones aminoacídicas observadas en el segmento-K de las AteqDHN, respecto al consenso, se conservan en otras monocotiledóneas, proponemos que éstas pertenecen al tipo SK3. Sin embargo, es importante considerar que las sustituciones observadas en los segmentos K y F podrían afectar sus funciones bioquímicas, ya que estos segmentos poseen propensión a formar α -hélices amfipáticas las cuáles se asocian a la función de las dehidrinas para la crioprotección de membranas artificiales, a las cuáles se unen de forma fosforilación-dependiente (Hughes *et al.* 2013, Candat *et al.* 2014, Strimbeck 2017). También se identificó en las AteqDHN una secuencia similar al segmento-F (Strimbeck 2017, Riley *et al.* 2019), el cual se reporta presente en dehidrinas específicamente inducidas por estrés abiótico como sequía, calor y salinidad (Artur *et al.* 2019).

Los transcritos de las dos isoformas de AteqDHN aquí identificadas mostraron mayor nivel de expresión en tallo y meristemo vegetativo que las otras AteqLEAPs. Esta expresión constitutiva observada en órganos vegetativos de *A. tequilana* es atípica, respecto a los patrones de expresión observados en otras plantas. AteqDHN1-2 son además los más expresados en las hojas en desarrollo, enrolladas dentro del cono central (cogollo). En los agaves, el meristemo y las hojas del cogollo permanecen viables aún después de estar varios meses sin riego, en tanto que las hojas expandidas externas pierden viabilidad. Ante el riego, las hojas del cogollo se despliegan formando hojas desarrolladas, con capacidad fotosintética. La expresión de las AteqDHN en estos órganos sugiere que pueden participar en la protección de la viabilidad de células meristemáticas y en proceso de desarrollo, para asegurar la supervivencia de la planta mediante la reactivación de estos tejidos, aún después de etapas prolongadas de estrés hídrico. Este hipótesis es coincidente con lo previamente reportado por Luján *et al.* (2009), quien mostró que las *heat-shock proteins* (HSP)

pequeñas se expresan en altos niveles en las hojas del cono central, tanto a temperatura ambiente como bajo tratamientos con calor, sugiriendo que dicha expresión constitutiva es parte de las estrategias moleculares que protegen al meristemo apical vegetativo (SAM) de daño por altas temperaturas. Aunque *AteqDHN1-2* disminuyó su nivel de acumulación en hojas tratadas con calor y *AteqDHN3* no modificó sus bajos niveles de expresión en dicho tejido, se requiere evaluar la respuesta de estas dehidrinas ante calor en cogollo y meristemo apical.

Otra explicación para el alto nivel de expresión de dehidrinas detectado en estos tejidos es que en las células meristemáticas y en desarrollo, al encontrarse en proceso de división acelerado y de construcción de estructuras celulares y membranas, requieran la presencia de dehidrinas para ensamblar o proteger las proteínas estructurales y metabólicas implicadas en los procesos de división y diferenciación celular, independientemente de las condiciones ambientales. La alta expresión de los transcritos de *AteqDHNs* en la piña (tallo); el cual en los agaves es un tejido especializado en la acumulación de carbohidratos tipo fructanos; puede estar relacionada con el alto potencial osmótico que pueden tener esas células y con la protección de enzimas metabólicas en este órgano de reserva.

*Las LEAP_5B (LEA_3) podrían jugar un papel relevante en la protección de tejidos fotosintéticos en *A. tequilana*.* Las LEAPs de este grupo, por su relativa escasez en los genomas y por su temprana categorización como *atípicas*, han sido poco estudiadas y descritas en la literatura. Las *AteqLEA5B* aquí analizadas conservan, junto con las LEAPs_5B de otras monocotiledóneas, el motivo W identificado en *A. thaliana* por Hundertmark & Hincha (2008). También se identificaron otros motivos estructurales reportados por Sing & Graether (2020), lo que conduce a clasificarlas dentro del subgrupo MAaRS-1W, una de las estructuras más comunes en LEA5B de otras especies. Las *AteqLEA5B* son hidrofóbicas, ordenadas y con estructura estable, por lo que cumplen con todos los parámetros fisicoquímicos reportados para el subgrupo MAaRS (Singh & Graether, 2020). Interesantemente para la *AteqLEA5B_35*, que es el miembro de esta familia génica con mayor nivel de expresión, se predice una estructura inestable. Tal diferencia radica en pocos cambios de aminoácidos entre sus secuencias; en particular notamos que en las cuatro *AteqLEA5B* estables una Phe sustituye a la Ser consenso del dominio MAaRS y una Val sustituye al Glu en el motivo del C-terminal. El péptido señal que se predice en el N-terminal de las *AteqLEA5B* es congruente con los reportes experimentales que muestran que éstas LEAPs son mitocondriales (Candat *et al.* 2014). El motivo MAaRS de las *AteqLEA5B*, forma parte del péptido señal, por lo que sus variaciones respecto al consenso podrían afectar la localización subcelular de las *AteqLEA5B*.

Los cinco genes de *AteqLEA5B* aquí identificados codifican solo tres distintas proteínas, lo cual sugiere que dos pares de estos genes son alelos de un mismo gen, o bien que son duplicaciones génicas que están divergiendo. Aunque las *AteqLEA5B* contienen todos los motivos conservados, su secuencia proteínica es de homología relativamente baja contra LEAPs 5B de otras plantas, incluida la de *A. officinalis*; mostrando que estas LEAPs han divergido en *A. tequilana*.

En cuanto a la función de las LEAPs del Grupo 5B, SAG21 (*senescence-associated gene*), proteína con dominio LEA_3 de *A. thaliana* también denominada AtLEA5, ha sido funcionalmente caracterizada y se expresa en tejidos vegetativos antes de la senescencia y en polen, se induce por estrés oxidativo, ácido absícico, deshidratación, frío, salinidad y por estrés biótico (Mowla *et al.* 2006, Salleh *et al.* 2012). Una LEAP_5B de arroz (OsLEA3) también se induce por ácido absícico y ante condiciones de salinidad en variedades tolerantes (Moons *et al.* 1997). En varias especies de *Agave* spp., Tamayo-Ordóñez *et al.* (2016) reportó el incremento en la expresión relativa de LEAPs ante períodos de 15 o 30 días bajo condiciones de alta salinidad o temperatura. Aquí mostramos que dichas LEAPs son ortólogas a *AteqLEA5B_35*.

Nuestros resultados confirman que las *AteqLEA_5B*, que son las LEAPs más abundantes en los transcriptomas analizados, tienen mayor nivel de expresión que otras LEAPs, particularmente en hojas expandidas, plenamente funcionales y sin signos de estrés. Dicho nivel de expresión se mantiene al ser tratadas con calor, sugiriendo producción o estabilidad del transcripto aún bajo condiciones en las que los transcritos de *AteqUBQ* y *AteqDHN1-2* se ven disminuidos. Esta expresión alta y constitutiva sugiere que las *AteqLEAPs_5B* tienen un papel relevante en la protección preventiva de hojas fotosintéticamente funcionales de *A. tequilana*.

Conclusiones

Tomando en conjunto los resultados aquí obtenidos, y considerando que los patrones de expresión observados a nivel transcripto se mantienen a nivel proteína, sugerimos que las AteqLEA5B atípicas (LEA_3) y las dehidrinas, en particular AteqLEA5B_35 y AteqDHN1-2, podrían tener una función evolutivamente seleccionada en *A. tequilana*, expresándose de forma constitutiva en tejidos vegetativos diferenciales como los fotosintéticos o meristemáticos, en desarrollo y de almacenamiento; respectivamente. La expresión constitutiva de estas AteqLEAPs, forma parte de las estrategias moleculares que confieren al agave una protección preventiva tipo “*primed state*”; como se ha sugerido para las plantas de resurrección, las cuales parecen estar preparadas expresando constantemente los genes codificantes de las enzimas que participan en la biosíntesis de azúcares protectores y antioxidantes, así como genes de proteínas de defensa como las LEAPs (Bechtold 2018). El aislamiento de las AteqLEA5B nativas y la evaluación de sus propiedades bioquímicas *in vitro*, el análisis del efecto de su expresión heteróloga en otros organismos, y la localización subcelular de las LEAPs *in vivo*, contribuirán a definir su función en los agaves.

Agradecimientos

Al CONACYT por la beca a JVC; a O-Martínez y J-Simpson (CINVESTAV) por la construcción de la *AgaveDB* y ESTs de tejidos reproductivos, respectivamente; J-Ramírez (CICY) por la búsqueda de dehidrinas en seqRNA; a AA-Covarrubias (IBT-UNAM) por sus aportaciones asesorando la tesis de JVC; a E-Alfaro por la edición de la bibliografía.

Material suplementario

El material suplementario de este artículo puede ser descargado de esta liga: <https://doi.org/10.17129/botsci.2861>

Literatura citada

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **3**: 403-410. DOI: [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)

Amara I, Zaidi I, Masmoudi K, Ludevid M, Pagès M, Goday A, Brini F. 2014. Insights into Late Embryogenesis Abundant (LEA) Proteins in Plants: From Structure to the Functions. *American Journal of Plant Sciences* **5**: 3440-3455. DOI: <https://doi.org/10.4236/ajps.2014.522360>

Artur MAS, Zhao T, Ligterink W, Schranz E, Hilhorst H. 2019. Dissecting the Genomic Diversification of Late Embryogenesis Abundant (LEA) Protein Gene Families in Plants. *Genome Biology and Evolution* **11**: 459-471. DOI: <https://doi.org/10.1093/gbe/evy248>

Artus NN, Uemura M, Steponkus PL, Gilmour JS, Lin C, Thomashow MF. 1996. Constitutive expression of the cold-regulated *Arabidopsis thaliana* COR15a gene affects both chloroplast and protoplast freezing tolerance. *Proceedings of the National Academy of Sciences* **93**: 13404-13409. DOI: <https://doi.org/10.1073/pnas.93.23.13404>

Battaglia M, Olvera-Carrillo Y, Garciarrubio A, Campos F, Covarrubias AA. 2008. The enigmatic LEA proteins and other hydrophilins. *Plant Physiology* **148**: 6-24. DOI: <https://doi.org/10.1104/pp.108.120725>

Bechtold U. 2018. Plant life in extreme environments: how do you improve drought tolerance? *Frontiers in Plant Science* **9**: 543. DOI: <https://doi.org/10.3389/fpls.2018.00543>

Candat A, Paszkiewicz G, Neveu M, Gautier R, Logan DC, Avelange-Macherel MH, Macherel D. 2014. The ubiquitous distribution of late embryogenesis abundant proteins across cell compartments in *Arabidopsis* offers tailored protection against abiotic stress. *Plant Cell* **26**: 3148-66. DOI: <https://doi.org/10.1105/tpc.114.127316>

Cervantes-Pérez SA, Espinal-Centeno A, Oropeza-Aburto A, Caballero-Pérez J, Falcon F, Aragón-Raygoza A, Sánchez-Segura L, Herrera-Estrella L, Cruz-Hernández A, Cruz-Ramírez A. 2018. Transcriptional profiling of the

CAM plant *Agave salmiana* reveals conservation of a genetic program for regeneration. *Developmental Biology* **442**: 28-39. DOI: <https://doi.org/10.1016/j.ydbio.2018.04.018>

Chen Y, Li C, Zhang B, Yi J, Yang Y, Kong C, Lei C, Gong M. 2019. The role of the Late Embryogenesis-Abundant (LEA) protein family in development and the abiotic stress response: A comprehensive expression analysis of potato (*Solanum tuberosum*). *Genes* **10**: 148. DOI: <https://doi.org/10.3390/genes10020148>

Ciccarelli DF, Bork P. 2005. The WHy domain mediates the response to desiccation in plants and bacteria. *Bioinformatics* **21**: 1304-1307. DOI: <https://doi.org/10.1093/bioinformatics/bti221>

Costa MC, Artur AM, Maia J, Jonkheer E, Derkx FLM, Nijveen H, Williams B, Mundree GS, Jiménez-Gómez JM, Hesselink T, Schijlen E, Ligterink W, Oliver MJ, Farrant JM, Hilhorst HW. 2017. A footprint of desiccation tolerance in the genome of *Xerophyta viscosa*. *Nature Plants* **3**: 17038. DOI: <https://doi.org/10.1038/nplants.2017.38>

Cuevas-Velázquez CL, Rendón-Luna DF, Covarrubias AA. 2014. Dissecting the cryoprotection mechanisms for dehydrins. *Frontiers in Plant Science* **5**: 583. DOI: <https://doi.org/10.3389/fpls.2014.00583>

Cui H, Wang Y, Yu T, Chen S, Chen Y, Lu C. 2020. Heterologous expression of three ammopiptanthus mongolicus dehydrin genes confers abiotic stress tolerance in *Arabidopsis thaliana*. *Plants* **9**: 193. DOI: <https://doi.org/10.3390/plants9020193>

Dang NX, Popova AV, Hundertmark M, Hincha DK. 2014. Functional characterization of selected LEA proteins from *Arabidopsis thaliana* in yeast and in vitro. *Planta* **240**: 325-336. DOI: <https://doi.org/10.1007/s00425-014-2089-z>

Deng G, Huang X, Xie L, Tan S, Gbokie T Jr, Bao Y, Xie Z, Yi K. 2019. Identification and Expression of SAUR Genes in the CAM Plant *Agave*. *Genes* **10**: 555. DOI: <https://doi.org/10.3390/genes10070555>

Dirk LMA, Abdel CG, Ahmad I, Neta ICS, Pereira CC, Pereira FECB, Unêda-Trevisoli SH, Pinheiro DG, Downie AB. 2020. late embryogenesis abundant protein-client protein interactions. *Plants* **9**: 814. DOI: <https://doi.org/10.3390/plants9070814>

Dure L, Greenway SC, Galau GA. 1981. Developmental biochemistry of cottonseed embryogenesis and germination: changing messenger ribonucleic acid populations as shown by in vitro and in vivo protein synthesis. *Biochemistry* **20**: 4162-4168. DOI: <https://doi.org/10.1021/bi00517a033>

El-Gebali S, Nistri J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson JL, Salazar GA, Smart A, Sonnhammer LLE, Hirsh L, Paladin L, Piovesan D, Tosatto ESC, Finn RD. 2019. The Pfam protein families database in 2019. *Nucleic Acids Research* **47**: D427-D432. DOI: <https://doi.org/10.1093/nar/gky995>

Farrant JM, Cooper K, Hilgart A, Abdalla KO, Bentley J, Thomson JA, Dace HJ, Peton N, Mundree SG, Rafudeen MS. 2015. A molecular physiological review of vegetative desiccation tolerance in the resurrection plant *Xerophyta viscosa* (Baker). *Planta* **242**: 407-426. DOI: <https://doi.org/10.1007/s00425-015-2320-6>

Galau GA, Dure L. 1981. Developmental biochemistry of cottonseed embryogenesis and germination: changing messenger ribonucleic acid populations as shown by reciprocal heterologous complementary deoxyribonucleic acid-messenger ribonucleic acid hybridization. *Biochemistry* **20**: 4169-4178. DOI: <https://doi.org/10.1021/bi00517a034>

Galau GA, Hughes DW, Dure III L. 1986. Abscisic acid induction of cloned cotton late embryogenesis-abundant (Lea) mRNAs. *Plant Molecular Biology* **7**: 155-170. DOI: <https://doi.org/10.1007/BF00021327>

Galau GA, Wang HY, Hughes DW. 1993. Cotton Lea5 and Lea14 encode atypical late embryogenesis-abundant proteins. *Plant Physiology* **101**: 695-696. DOI: <https://doi.org/10.1104/pp.101.2.695>

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research* **31**: 3784-3788. DOI: <https://doi.org/10.1093/nar/gkg563>

Gentry HS. 2003. Agaves of Continental North America. *University of Arizona Press, Tucson, AZ, USA*. ISBN: 9780816507757

Giarola V, Bartels D. 2015. What can we learn from the transcriptome of the resurrection plant *Craterostigma planlagioides*? *Planta* **242**: 427-34. DOI: <https://doi.org/10.1007/s00425-015-2327-z>

Gosti F, Bertauche N, Vartanian N, Giraudat J. 1995. Abscisic acid-dependent and -independent regulation of gene

expression by progressive drought in *Arabidopsis thaliana*. *Molecular and General Genetics* **246**: 10-18. DOI: <https://doi.org/10.1007/bf00290128>

Graether SP, Boddington KF. 2014. Disorder and function: a review of the dehydrin protein family. *Frontiers in Plant Science* **5**: 576. DOI: <https://doi.org/10.3389/fpls.2014.00576>

Gross SM, Martin JA, Simpson J, Abraham-Juarez MJ, Wang Z, Visel A. 2013. *De novo* transcriptome assembly of drought tolerant CAM plants, *Agave deserti* and *Agave tequilana*. *BMC Genomics* **14**: 563. DOI: <https://doi.org/10.1186/1471-2164-14-563>

Guo W, Ward RW, Thomashow MF. 1992. Characterization of a cold-regulated wheat gene related to *Arabidopsis cor47*. *Plant Physiology* **100**: 915-922. DOI: <https://doi.org/10.1104/pp.100.2.915>

Guo X, Zhang L, Wang X, Zhang M, Xi Y, Wang A, Zhu J. 2019. Overexpression of *Saussurea involucrata* dehydrin gene SiDHN promotes cold and drought tolerance in transgenic tomato plants. *PLoS One* **14**: 11. DOI: <https://doi.org/10.1371/journal.pone.0225090>

Haaning S, Radutoiu S, Hoffmann SV, Dittmer J, Giehm L, Otzen DE, Stougaard J. 2008. An unusual intrinsically disordered protein from the model legume *Lotus japonicus* stabilizes proteins in vitro. *The Journal of Biological Chemistry* **283**: 31142-52. DOI: <https://doi.org/10.1074/jbc.M805024200>

He S, Tan L, Hu Z, Chen G, Wang G, Hu T. 2012. Molecular characterization and functional analysis by heterologous expression in *E. coli* under diverse abiotic stresses for OsLEA5, the atypical hydrophobic LEA protein from *Oryza sativa* L. *Molecular Genetics and Genomics* **287**: 39-54. DOI: <https://doi.org/10.1007/s00438-011-0660-x>

Hu T, Liu Y, Zhu S, Qin J, Li W, Zhou N. 2019. Overexpression of OsLea14-A improves the tolerance of rice and increases Hg accumulation under diverse stresses. *Environmental Science and Pollution Research* **26**: 10537-10551. DOI: <https://doi.org/10.1007/s11356-019-04464-z>

Huang X, Xiao M, Xi J, He C, Zheng J, Chen H, Gao J, Zhang S, Wu W, Liang Y, Xie L, Yi K. 2019. *De novo* transcriptome assembly of *Agave* H11648 by Illumina sequencing and identification of cellulose synthase genes in *Agave* species. *Genes* **10**: 103. DOI: <https://doi.org/10.3390/genes10020103>

Hughes SL, Schart V, Malcolmson J, Hogarth KA, Martynowicz DM, Tralman-Baker E, Patel SN, Graether SP. 2013. The importance of size and disorder in the cryoprotective effects of dehydrins. *Plant Physiology* **163**: 1376-1386. DOI: <https://doi.org/10.1104/pp.113.226803>

Hundertmark M, Hincha DK. 2008. LEA (Late Embryogenesis Abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics* **9**: 118. DOI: <https://doi.org/10.1186/1471-2164-9-118>

Kim HS, Lee JH, Kim JJ, Kim CH, Jun SS, Hong YN. 2005. Molecular and functional characterization of Ca-LEA6, the gene for a hydrophobic LEA protein from *Capsicum annuum*. *Gene* **344**: 115-23. DOI: <https://doi.org/10.1016/j.gene.2004.09.012>

Kimura M, Yamamoto YY, Seki M, Sakurai T, Abe T, Yoshida S, Manabe K, Shinozaki K, Matsui M. 2003. Identification of *Arabidopsis* genes regulated by high light-stress using cDNA microarray. *Photochemistry and Photobiology* **77**: 226-233. DOI: [https://doi.org/10.1562/0031-8655\(2003\)077<0226:ioagrb>2.0.co;2](https://doi.org/10.1562/0031-8655(2003)077<0226:ioagrb>2.0.co;2)

Lee SC, Lee MY, Kim SJ, Jun SH, An G, Kim SR. 2005. Characterization of an abiotic stress-inducible dehydrin gene, OsDhn1, in rice (*Oryza sativa* L.). *Molecules and Cells* **19**: 212-218.

Liu Y, Xie L, Liang X, Zhang S. 2015. *CpLEA5*, the late embryogenesis abundant protein gene from *Chimonanthus praecox*, possesses low temperature and osmotic resistances in prokaryote and eukaryotes. *International Journal of Molecular Sciences* **16**: 26978-26990. DOI: <https://doi.org/10.3390/ijms161126006>

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT method. *Methods* **25**: 402-408. DOI: <https://doi.org/10.1006/meth.2001.1262>

Luján R, Lledías F, Martínez LM, Barreto R, Cassab GI, Nieto-Sotelo J. Small heat-shock proteins and leaf cooling capacity account for the unusual heat tolerance of the central spike leaves in *Agave tequilana* var. Weber. 2009. *Plant, Cell & Environment* **32**: 1791-803. DOI: <https://doi.org/10.1111/j.1365-3040.2009.02035.x>

Lv A, Fan N, Xie J, Yuan S, An Y, Zhou P. 2017. Expression of CdDHN4, a novel YSK2 -type dehydrin gene from

bermudagrass, responses to drought stress through the ABA-dependent signal pathway. *Frontiers in Plant Science* **8**: 748. DOI: <https://doi.org/10.3389/fpls.2017.00748>

Magwanga RO, Lu P, Kirungu JN, Lu H, Wang X, Cai X, Zhou Z, Zhang Z, Salih H, Wang K, Liu F. 2018. Characterization of the late embryogenesis abundant (LEA) proteins family and their role in drought stress tolerance in upland cotton. *BMC Genetics* **19**: 6. DOI: <https://doi.org/10.1186/s12863-017-0596-1>

Maitra N, Cushman JC. 1994. Isolation and characterization of a drought-induced soybean cDNA encoding a D95 family late-embryogenesis-abundant protein. *Plant Physiology* **106**: 805-806. DOI: <https://doi.org/10.1104/pp.106.2.805>

Manfre AJ, Lanni LM, Marcotte WR. 2006. The *Arabidopsis* group 1 LATE EMBRYOGENESIS ABUNDANT protein ATEM6 is required for normal seed development. *Plant Physiology* **140**: 140-149. DOI: <https://doi.org/10.1104/pp.105.072967>

Martínez-Hernández A, Mena-Espino ME, Herrera-Estrella AH, Martínez-Hernández P. 2010. Construcción de bibliotecas de ADNc y análisis de expresión génica por RT-PCR en agaves. *Revista Latinoamericana de Química* **38**: 21-44.

Mertens J, Aliyu H, Cowan DA. 2018. LEA Proteins and the Evolution of the WHy Domain. *Applied and Environmental Microbiology* **84**: e00539-18. DOI: <https://doi.org/10.1128/AEM.00539-18>

Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown D, Chang H-Y, El-Gebali S, Fraser MI, Gough J, Haft DR, Huang H, Letunic I, Lopez R, Luciani A, Madeira F, Marchler-Bauer A, Mi Huaiyu, NDA, Necci M, Nuka G, Orengo C, Pandurangan AP, Paysan-Lafosse T, Pesseat S, Potter SC, Qureshi MA, Rawlings ND, Redaschi N, Richardson LJ, Rivoire C, Salazar GA, Sangrador-Vegas A, Sigrist CJA, Sillitoe I, Sutton GG, Thanki N, Thomas PD, Tosatto SCE, Yit YS, Finn RD. 2019. InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Research* **47**: 351-360. DOI: <https://doi.org/10.1093/nar/gky1100>

Monroy-González Z, Martínez-Hernández A. 2019. Identificación de cDNAs de *Agave tequilana* Weber Var. azul similares a furostanol glicósido 26-o-β-glucosidasas. *Revista Fitotécnica Mexicana* **42**: 21-44. DOI: <https://doi.org/10.35196/rfm.2019.4.439-447>

Moons A, De Keyser A, Van Montagu M. 1997. A group 3 LEA cDNA of rice, responsive to abscisic acid, but not to jasmonic acid, shows variety-specific differences in salt stress response. *Gene* **191**: 197-204. DOI: [https://doi.org/10.1016/s0378-1119\(97\)00059-0](https://doi.org/10.1016/s0378-1119(97)00059-0)

Mowla SB, Cuypers A, Driscoll SP, Kiddle G, Thomson J, Foyer CH, Theodoulou FL. 2006. Yeast complementation reveals a role for an *Arabidopsis thaliana* late embryogenesis abundant (LEA)-like protein in oxidative stress tolerance. *The Plant Journal* **48**: 743-756. DOI: <https://doi.org/10.1111/j.1365-313X.2006.02911.x>

Nagaraju M, Kumar SA, Reddy PS, Kumar A, Rao DM, Kavi Kishor PB. 2019. Genome-scale identification, classification, and tissue specific expression analysis of late embryogenesis abundant (LEA) genes under abiotic stress conditions in *Sorghum bicolor* L. *PLoS One* **14**: e0209980. DOI: <https://doi.org/10.1371/journal.pone.0209980>

NCBI Resource Coordinators. 2016. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* **44**: 7-19. DOI: <https://doi.org/10.1093/nar/gkv1290>

Nobel PS, Smith SD. 1983. High and low temperature tolerances and their relationships to distribution of agaves. *Plant, Cell & Environment* **6**: 711-719. DOI: <https://doi.org/10.1111/1365-3040.ep11589339>

Nobel PS. 1988. *Environmental Biology of Agaves and Cacti*. Cambridge, UK: Cambridge University Press. ISBN: 978-052-1543-34-7

Nobel PS, Valenzuela AG. 1987. Environmental responses and productivity of the CAM plant, *Agave tequilana*. *Agricultural and Forest Meteorology* **39**: 319-334. DOI: [https://doi.org/10.1016/0168-1923\(87\)90024-4](https://doi.org/10.1016/0168-1923(87)90024-4)

Piatkowski D, Schneider K, Salamini F, Bartels D. 1990. Characterization of five abscisic acid-responsive cDNA clones isolated from the desiccation-tolerant plant *Craterostigma plantagineum* and their relationship to other water-stress genes. *Plant Physiology* **94**: 1682-1688. DOI: <https://doi.org/10.1104/pp.94.4.1682>

Puhakainen T, Hess MW, Mäkelä P, Svensson J, Heino P, Palva ET. 2004. Overexpression of multiple dehydrin

genes enhances tolerance to freezing stress in *Arabidopsis*. *Plant Molecular Biology* **54**: 743-753. DOI: <https://doi.org/10.1023/B:PLAN.0000040903.66496.a4>

Ramírez-Tobías HM, Peña-Valdivia CB, Aguirre JR. 2014. Respuestas bioquímico-fisiológicas de especies de *Agave* a la restricción de humedad. *Botanical Sciences* **92**: 131-139. DOI: <https://doi.org/10.17129/botsci.156>

Riley AC, Ashlock DA, Graether SP. 2019. Evolution of the modular, disordered stress proteins known as dehydrins. *PLoS One* **14**: e0211813. DOI: <https://doi.org/10.1371/journal.pone.0211813>

Salleh FM, Evans K, Goodall B, Machin H, Mowla SB, Mur LA, Runions J, Theodoulou FL, Foyer CH, Rogers HJ. 2012. A novel function for a redox-related LEA protein (SAG21/AtLEA5) in root development and biotic stress responses. *Plant Cell & Environment* **35**: 418-429. DOI: <https://doi.org/10.1111/j.1365-3040.2011.02394.x>

Sarwar MB, Ahmad Z, Rashid B, Hassan S, Gregersen PL, Leyva MO, Nagy I, Asp T, Husnain T. 2019. De novo assembly of *Agave sisalana* transcriptome in response to drought stress provides insight into the tolerance mechanisms. *Scientific Reports* **9**: 396. DOI: <https://doi.org/10.1038/s41598-018-35891-6>

Shi J, Liu M, Chen Y, Wang J, Lu C. 2016. Heterologous expression of the dehydrin-like protein gene AmCIP from *Ammopiptanthus mongolicus* enhances viability of *Escherichia coli* and tobacco under cold stress. *Plant Growth Regulation* **79**: 71-80. DOI: <https://doi.org/10.1007/s10725-015-0112-4>

Shakeel SN, Aman S, Haq NU, Heckathorn SA, Luthe D. 2013. Proteomic and transcriptomic analyses of *Agave americana* in response to heat stress. *Plant Molecular Biology Reporter* **31**: 840-851. DOI: <https://doi.org/10.1007/s11105-013-0555-6>

Sigrist CJA, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueret L, Xenarios I. 2013. New and continuing developments at PROSITE. *Nucleic Acids Research* **41**(Database issue): D344-7. DOI: <https://doi.org/10.1093/nar/gks1067>

Simpson J, Martínez-Hernández A, Abraham-Juárez MJ, Delgado-Sandoval S, Sánchez-Villarreal A, Cortés-Romero C. 2011. Genomic resources and transcriptome mining in *Agave tequilana*. *GCB Bioenergy* **3**: 25-36. DOI: <https://doi.org/10.1111/j.1757-1707.2010.01079.x>

Singh S, Cornilescu CC, Tyler RC, Cornilescu G, Tonelli M, Lee MS, Markley JL. 2005. Solution structure of a late embryogenesis abundant protein (LEA14) from *Arabidopsis thaliana*, a cellular stress-related protein. *Protein Science* **14**: 2601-2609. DOI: <https://doi.org/10.1110/ps.051579205>

Singh KK, Graether SP. 2020. Conserved sequence motifs in the abiotic stress response protein late embryogenesis abundant 3. *PLoS One* **15**: e0237177. DOI: <https://doi.org/10.1371/journal.pone.0237177>

Stewart JR. 2015. *Agave* as a model CAM crop system for a warming and drying world. *Frontiers in Plant Science* **24**: 684. DOI: <https://doi.org/10.3389/fpls.2015.00684>

Strimbeck R. 2017. Hiding in plain sight: the F segment and other conserved features of seed plant SKn dehydrins. *Planta* **245**: 1061-1066. DOI: <https://doi.org/10.1007/s00425-017-2679-7>

Tamayo-Ordóñez MC, Rodriguez-Zapata LC, Narváez-Zapata JA, Tamayo-Ordóñez YJ, Ayil-Gutiérrez BA, Barredo-Pool F, Sánchez-Teyer LF. 2016. Morphological features of different polyploids for adaptation and molecular characterization of CC-NBS-LRR and LEA gene families in *Agave* L. *Journal of Plant Physiology* **195**: 80-94. DOI: <https://doi.org/10.1016/j.jplph.2016.03.009>

Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A. 2003. **PANTHER: a library of protein families and subfamilies indexed by function**. *Genome Research* **13**: 2129-41. DOI: <https://doi.org/10.1101/gr.772403>

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189-1191. DOI: <https://doi.org/10.1093/bioinformatics/btp033>

Wang M, Li P, Li C, Pan Y, Jiang X, Zhu D, Zhao Q, Yu J. 2014. SiLEA14, a novel atypical LEA protein, confers abiotic stress resistance in foxtail millet. *BMC Plant Biology* **14**: 290. DOI: <https://doi.org/10.1186/s12870-014-0290-7>

Xiang D, Man L. 2018. EhEm1, a novel Em-like protein from *Eutrema halophilum*, confers tolerance to salt and drought stresses in rice. *Molecular Breeding* **38**: 17. DOI: <https://doi.org/10.1007/s11032-017-0750-5>

Yu J, Lai Y, Wu X, Wu G, Guo C. 2016. Overexpression of OsEm1 encoding a group I LEA protein confers enhanced drought tolerance in rice. *Biochemical and Biophysical Research Communications* **478**: 703-709. DOI: <https://doi.org/10.1016/j.bbrc.2016.08.010>

Zegzouti H, Jones B, Marty C, Lelièvre JM, Latché A, Pech JC, Bouzayen M. 1997. Er5, a tomato cDNA encoding an ethylene-responsive LEA-like protein: characterization and expression in response to drought, ABA and wounding. *Plant Molecular Biology* **35**: 847-854. DOI: <https://doi.org/10.1023/a:1005860302313>

Zheng J, Su H, Lin R, Zhang H, Xia K, Jian S, Zhang M. 2019. Isolation and characterization of an atypical LEA gene (IpLEA) from *Ipomoea pes-caprae* conferring salt/drought and oxidative stress tolerance. *Scientific Report* **9**: 14838. DOI: <https://doi.org/10.1038/s41598-019-50813-w>

Editor de sección: Martín Mata Rosas

Contribución de autores: JVC y KVA, ambos son primer autor. JVC: caracterización de DHNs, análisis fisicoquímico y expresión génica. KVA: minería de seqRNA, identificación de dominios y análisis filogenético (LEAPs). JLR: figuras, análisis estadístico, discusión. AMH: bibliotecas de cDNAs, expresión génica, diseño, supervisión y redacción. Todos los autores revisaron el manuscrito.