



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE QUÍMICA

LAS PROPIEDADES FÍSICOQUÍMICAS DE LAS PROTEÍNAS LEA Y SU PARTICIPACIÓN EN LA TOLERANCIA A LA DESECACIÓN DE DIVERSOS ORGANISMOS

T E S I S

QUE PARA OPTAR POR EL GRADO DE:

QUÍMICO FARMACÉUTICO BIÓLOGO

PRESENTA:

MICHELLE LOZANO ORTEGA

DIRECTOR DE TESIS:

DR. CÉSAR LUIS CUEVAS VELAZQUEZ

CIUDAD UNIVERSITARIA, CDMX, 2023



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

JURADO ASIGNADO:

Presidente: **DRA. SANCHEZ NIETO SOBEIDA**

Vocal: **DRA. COELLO COUTIÑO MARTHA PATRICIA**

Secretario: **DR. CUEVAS VELAZQUEZ CESAR LUIS**

Suplente 1: **DRA. LARA NUÑEZ AURORA**

Suplente 2: **DRA. GONZALEZ SEGURA LILIAN**

SITIO DONDE SE DESARROLLÓ EL TEMA:

Departamento de Bioquímica, Facultad de Química, Universidad Autónoma de México, Conjunto E,
laboratorio 103

ASESOR DEL TEMA: DR. CÉSAR LUIS CUEVAS VELÁZQUEZ

SUPERVISOR TÉCNICO: DRA. MARÍA FERNANDA CORNEJO GRANADOS

SUSTENTANTE: MICHELLE LOZANO ORTEGA

AGRADECIMIENTOS

Esta investigación fue realizada gracias al Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) de la UNAM IA203422; por lo que agradezco a la DGAPA-UNAM la beca recibida.

De igual manera agradezco al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el donativo de Ciencia de Frontera 2019 CONACYT con clave CF-2019 / 252952 al Dr. César Cuevas.

Adicionalmente doy las gracias a la Dra. Fernanda Cornejo y al Dr. César Cuevas por guiarme y cuestionarme para poder obtener como resultado la tesis que hoy tengo en mis manos. Por último, le agradezco a mi familia por motivarme a dar lo mejor de mí en todo lo que me propongo.

Índice

Resumen	1
I.Introducción	2
I.I Organismos tolerantes a la desecación	2
I.I.I. Plantas de resurrección	2
I.I.II Tardígrados	4
I.II Proteínas Intrínsecamente Desordenadas	6
I.III Proteínas abundantes en la embriogénesis tardía	7
I.IV Propiedades fisicoquímicas de las proteínas	7
II. Antecedentes	10
III. Hipótesis	12
IV. Objetivo general	12
V. Objetivos particulares	12
VI. Metodología	12
VI.I Limpieza de proteomas	12
VI.I.I Análisis de las propiedades fisicoquímicas	12
VI.I.II Determinación de rangos	13
VI.I.III Análisis ontológico	13
VII. Resultados	14
VIII. Discusión	29
IX. Conclusiones	31
X. Perspectivas	31
XI. Bibliografía	32
Anexo	35

Resumen

El déficit hídrico es un tipo de estrés abiótico que las plantas experimentan a lo largo de su vida. Las plantas han desarrollado diferentes mecanismos para contrarrestar este tipo de condiciones adversas. Algunos árboles extienden sus raíces hasta el subsuelo para conseguir nuevas fuentes de agua, mientras que las plantas suculentas (dentro de las cuales se encuentran las cactáceas) almacenan grandes cantidades de agua para futuras emergencias. Bajo ciertas condiciones, el déficit hídrico puede ser muy severo, alcanzando niveles de más del 90 % de pérdida de agua. A esta condición se le conoce como desecación. Algunas plantas experimentan desecación natural durante la última etapa de la maduración de las semillas. En esta etapa se acumulan un tipo de proteínas conocidas como proteínas abundantes en la embriogénesis tardía (LEA, *Late Embryogenesis Abundant*). La acumulación de proteínas LEA es uno de los mecanismos que contribuye a que las semillas sean tolerantes a la desecación.

Además de las semillas, en la naturaleza existen otros organismos tolerantes a la desecación, como algunos tipos de algas unicelulares, y organismos multicelulares como las plantas de resurrección y los tardígrados. Más allá de las plantas, es difícil encontrar ortólogos de las proteínas LEA. Existen muy pocas proteínas con similitud de secuencia con las proteínas LEA (Covarrubias et. al, 2017).

Previamente, se ha sugerido en la literatura que las propiedades fisicoquímicas codificadas en la secuencia de las proteínas LEA, y no la secuencia de aminoácidos por sí misma, podrían ser suficientes para conferir una función protectora durante condiciones de desecación en semillas (Garay-Arroyo et. al 2000). Por ello, el objetivo de este trabajo fue encontrar las propiedades fisicoquímicas particulares de las proteínas LEA y utilizar dichas propiedades para buscar proteínas similares en el proteoma de otro organismo tolerante a la desecación, el tardígrado *Hypsibius dujardini*, el cual no contiene ortólogos que codifican para proteínas LEA.

Los resultados obtenidos a partir del estudio de *A. thaliana*, *Oryza sativa* (arroz), *Triticum aestivum* (trigo) y *Zea mays* (maíz) muestran que las propiedades de las proteínas consideradas como tipo-LEA tienen valores de hidropatía (una medida de la proporción de residuos hidrofóbicos) menores o iguales a 0.46, valores de composición de aminoácidos promotores del desorden estructural mayores o iguales a 0.76, y un valor de carga neta media menor o igual a 0.03. Tras aplicar estos cortes, primeramente de hidropatía y desorden, y posteriormente de carga neta media, se delimitaron a las proteínas tipo-LEA, que corresponden al 1.3 % del proteoma del tardígrado.

Por último, se realizó un análisis de ontología genética en *Arabidopsis thaliana* que mostró que, de acuerdo a su función molecular, la mayoría de las proteínas tipo-LEA están asociadas a la respuesta a diferentes tipos de estrés abiótico. Los datos obtenidos en esta tesis representan la base para estudiar a un mayor número de organismos tolerantes a la desecación no solo mediante estudios *in silico*, si no también en estudios *in vitro* e *in vivo*.

I. Introducción

I.I Organismos tolerantes a la desecación

Existe una diferencia entre la tolerancia a la deshidratación y la tolerancia a la desecación ya que aunque en ambos casos hay poca disponibilidad de agua en el ambiente, en la primera la cantidad de agua intracelular es relativamente elevada y en la segunda es casi nula (Deeba and Pandey, 2017). La desecación puede definirse como una situación en donde el contenido de agua es menor o igual a 0.1 g de agua por gramo de peso seco (Deeba and Pandey, 2017). Así es cómo a diferencia de las suculentas, por ejemplo, los organismos tolerantes a la desecación, en lugar de acumular agua en caso de emergencia, tienen que pasar por diferentes procesos ya sea fisiológicos, morfológicos o genéticos para adaptarse a la falta de la misma.

I.I.I. Plantas de resurrección

Las plantas de resurrección son un tipo de plantas que se adaptan a la humedad relativa del ambiente. Pueden sobrevivir en un estado deshidratado, con un contenido celular de agua menor al 5.0 % (w/w) por varios años y después regresar a un estado fisiológico activo al rehidratarse (Deeba and Pandey, 2017). Se encuentran mayoritariamente donde hay temporadas estacionales bien establecidas en lugar de climas extremos (Deeba and Pandey, 2017).

Existen dos tipos de plantas de resurrección, las estrictamente tolerantes a la desecación y las “modificadas”. Las primeras carecen de los mecanismos necesarios para acumular agua por lo que su respuesta ante esta condición de estrés es rápida. Las segundas se subdividen en dos grupos: PDT (*Poikilochlorophyllous Desiccation-Tolerant*) y HDT (*Homoiochlorophyllous Desiccation-Tolerant*).

Hasta el año de 2017 se conocían cinco familias del tipo PDT: *Cyperaceae*, *Liliaceae*, *Anthericaceae*, *Poaceae* y *Velloziaceae* (Deeba and Pandey, 2017). Estas destruyen sus cloroplastos de manera controlada, dando lugar a “desicoplastos” que permiten la formación de clorofila en el momento de la rehidratación (Tuba et. al, 1993). Contrario a lo que ocurre en las hojas de plantas no tolerantes, en las PDT la destrucción de los cloroplastos es un mecanismo de defensa y no señal de daño (Deeba and Pandey, 2017). Por su parte las HDT no modifican su contenido de clorofila ni muestran pérdidas significativas de proteínas insolubles ni estructurales, como se observa en las PDT (Gaff and Hallman, 1974)¹.

Las plantas de resurrección no sólo experimentan adecuaciones fisiológicas, sino que morfológicamente también toman medidas para tolerar la desecación. La mayoría de las plantas de resurrección enrollan sus hojas hacia sí en periodos de falta de agua y esto desencadena la acumulación de compuestos fenólicos que sirven para proteger a la planta contra la radiación solar, y por ende previenen la formación de especies reactivas de oxígeno (Deeba and Pandey, 2017). A nivel celular también hay un cambio consecuente de la pérdida de agua, las paredes

¹ El artículo no ahonda en el mecanismo de acción de las proteínas HDT.

celulares se tornan un poco más flexibles y se compactan, lo cual permite mantener la viabilidad de la comunicación intercelular mediante plasmodesmos (Jones and McQueen-Manson, 2004). Esto a gran escala provoca que la planta se encoja y que por ello las hojas se retraigan sin embargo, este cambio es reversible.

Uno de los primeros cambios fisiológicos que toda planta lleva a cabo es el cierre de sus estomas, en cuanto detecta una disminución de agua en el ambiente. De este modo, las plantas previenen la pérdida de agua interna por evaporación (Deeba and Pandey, 2017). Las plantas de resurrección además de evitar la pérdida de agua mediante el cierre de sus estomas, tienen una ventaja adicional debido al enrollamiento de sus hojas; ya que esa acción reduce la superficie total de la hoja y por consiguiente reduce la transpiración de la planta. Así regula su temperatura y la protege contra la exposición a la radiación solar, como ya se mencionó anteriormente (Dinakar et. al, 2012). Se ha visto que este proceso está ligado a los niveles de citoquininas (CK) y ácido abscísico (ABA) en plantas. Se sabe que cuando hay altas concentraciones de CK en el xilema, los estomas se vuelven menos sensibles al ABA y se abren. Es importante destacar que el estado de las raíces afecta los niveles de las CK como el ribósido y el ribósido de zeatina (Deeba and Pandey, 2017).

Las raíces de las plantas de resurrección juegan una parte importante en la supervivencia de este tipo de plantas. En ocasiones, redireccionan su crecimiento en busca de fuentes de agua (hidrotropismo positivo) o cesan el desarrollo de raíces secundarias o terciarias (inhibición del meristemo de raíces laterales) (Deak and Malamy, 2005). Cuando partes de la raíz comienzan a secarse, la concentración de CK disminuye; y eso conlleva que en el xilema la concentración de ABA se vea aumentada y por ende la conductancia del xilema disminuye y los poros (estomas) se cierran (Stoll et. al, 2000).

Lamentablemente, el cierre de estomas por un periodo prolongado tiene consecuencias negativas para la planta. Debido a este bloqueo para la obtención de dióxido de carbono (CO₂), se detiene la fijación de carbono, que provoca una sobreexcitación de la clorofila, la cual cede un radical al oxígeno dando la formación de especies reactivas de oxígeno (ROS, *Reactive Oxygen Species*) (Deeba and Pandey, 2017). Ante la presencia de ROS se cree que algunas enzimas se sobreexpresan; pero no se ha demostrado una clara relación entre sus niveles en desecación y la rehidratación (Deeba and Pandey, 2017).

Adicionalmente hay otras proteínas no enzimáticas que actúan contrarrestando el estrés por déficit hídrico en estas plantas. Un ejemplo de estas son las aquaporinas, canales presentes en la membrana celular y responsables de facilitar el paso de agua de un lado de la membrana a otro en las células. Por otro lado, están las proteínas abundantes en la embriogénesis tardía (LEA, *Late Embryogenesis Abundant*), que son aquellas en las que se centra esta tesis. Estas proteínas a grandes rasgos ayudan a conservar la estructura de la membrana celular y protegen a otras proteínas evitando que se desnaturalicen bajo ciertas condiciones de estrés (Deeba and Pandey, 2017). Estas proteínas se analizarán con más detalle en el apartado I.III.

Finalmente, otra macromolécula que se ha visto involucrada con la tolerancia a la desecación en plantas de resurrección son los carbohidratos, en particular dos azúcares: sacarosa y trehalosa. La sacarosa puede actuar por medio de una interacción directa con membranas y/o macromoléculas o indirectamente al afectar la fluidez del citoplasma provocando un estado vítreo del mismo. La trehalosa al ser muy estable y poco reactiva también posee una función protectora de la cual se hablará más a fondo en la sección de tardígrados.

I.I.II Tardígrados

Los tardígrados son animales microscópicos conocidos por su gran habilidad de supervivencia. Al ser organismos extremófilos pueden vivir en condiciones completamente opuestas con la misma facilidad; desde volcanes hasta glaciares, sin que las temperaturas extremas afecten su viabilidad. Los tardígrados inclusive son capaces de sobrevivir décadas sin una fuente de alimento y/o agua al entrar en un estado latente llamado criptobiosis que les permite afrontar varias condiciones de estrés del ambiente, tales como choque osmótico, radiación, o inclusive el vacío en el espacio (Møbjerg and Cardoso, 2021). Para fines de esta tesis nos centraremos particularmente en su habilidad para tolerar la desecación. En términos del mecanismo de acción para tolerar la desecación hay varios factores a los cuales se les ha atribuido parcial o totalmente esta característica.

Anteriormente se creía que quizás la estrategia de los tardígrados era la síntesis o acumulación de trehalosa; pero tras realizar diferentes análisis se ha constatado que los niveles de trehalosa en tardígrados y algunos rotíferos microscópicos como los de la clase Bdelloidea, son muy bajos o nulos (Lapinski and Tunnacliffe, 2003); por lo que los niveles de trehalosa no tiene correlación con su capacidad para tolerar la desecación (Tunnacliffe *et. al* 2005). En el caso de los tardígrados es otra macromolécula la que se vitrifica brindándole protección al contenido intracelular; en lugar de tratarse de un carbohidrato hay tres familias de proteínas solubles a altas temperaturas que pueden llevar a cabo esa función (Boothby *et. al*, 2020). Las tres familias son: las proteínas citoplasmáticas abundantes solubles en calor (CAHS, *Cytoplasmic Abundant Heat Soluble*), las proteínas secretadas abundantes solubles en calor (SAHS, *Secreted Abundant Heat Soluble*), y las proteínas mitocondriales abundantes solubles en calor (MAHS, *Mitochondrial Abundant Heat Soluble*). Todas estas tienen en común que no suelen encontrarse en una conformación tridimensional establecida sino que, al ser proteínas intrínsecamente desordenadas (IDP, *Intrinsically Disordered Proteins*), se encuentran en un constante cambio conformacional. Es así como las tres familias se clasifican en el grupo de las proteínas intrínsecamente desordenadas de tardígrados (TDPs, *Tardigrade Disordered Proteins*) (Boothby *et. al*, 2020). Actualmente no se han realizado experimentos concluyentes para determinar los mecanismos de acción de cualquiera de las tres familias; es por eso que a continuación solo se mencionan diferentes hipótesis a partir de lo que se conoce de su caracterización.

Respecto a las CAHS, se cree que pueden llevar a cabo dos acciones debido a su distribución homogénea de cargas a lo largo de su secuencia. La primera es similar a lo que se menciona para la trehalosa y es la vitrificación de los componentes intracelulares; lo cual lograría evitar la desnaturalización de varias proteínas y ayudaría a mantener la composición de la membrana.

Tras realizar estudios de calorimetría, se llegó a valores distintos de temperaturas de transición vítrea para tardígrados enteros y la proteína recombinante CAHS *in vitro* y en levadura; por lo cual se concluyó que CAHS seguramente dependía de otras moléculas y co-solutos para llevar a cabo esta acción (Boothby et. al, 2020).

Derivado de esto, se analizó que podría pasar con la poca agua que permanece en el tardígrado cuando este se encuentra en un estado vítreo y se plantearon dos vertientes. La primera sugiere que el agua remanente forma una ligera capa protectora alrededor de proteínas sensibles. Las CAHS tienen dos motivos octa-peptídicos cargados y algunas de las proteínas en esa familia adicionalmente tienen una distribución de cargas homogénea; lo cual hace que actúen como polianfolitos (Yamaguchi et. al, 2012). En estos casos, esa característica les permitiría formar enlaces de puentes de hidrógeno con el agua al estar preferentemente en una conformación extendida (Boothby et. al, 2017). Adicionalmente, las CAHS podrían participar en la exclusión preferente del agua. Las CAHS son más grandes que la trehalosa y por lo mismo podrían realizar un mejor trabajo al alejar el agua de sí y acercarla a las proteínas propensas a desnaturalizarse durante la desecación del tardígrado (Boothby et. al, 2020).

Durante el proceso de desecación el tardígrado forma una estructura extracelular especial (SES, *Special Extracellular Structure*) alrededor de sus células secretoras; similar a la pared celular de las células vegetales (Boothby et. al, 2020). Aquí es donde posiblemente las proteínas SAHS juegan un papel importante. Por un lado, se ha visto que las células secretoras abundan durante la desecación y desaparecen conforme el tardígrado es rehidratado; esta es una conducta parecida a la de las proteínas SAHS ya que estas también incrementan su abundancia en el estado de desecación del tardígrado (Richaud et. al, 2020). Ni la función de las SES en la desecación ni la presencia de las SAHS en los SES está completamente clara; sin embargo la correlación de su aparición con el estado de desecación indican un posible rol en la protección ante el daño por esta condición de estrés (Boothby et. al, 2020).

En general, ante cualquier condición de estrés la mitocondria y el peroxisoma producen especies reactivas de oxígeno (ROS, *Reactive Oxygen Species*), pero al mismo tiempo son los organelos responsables de la eliminación de dichas moléculas; por lo que es de suma importancia para el tardígrado proteger ambos organelos (Boothby et. al, 2020). Una manera de mitigar los daños realizados a la mitocondria durante la desecación podría ser un reajuste a nivel de membrana, dado que una disminución del área superficial de la membrana interna impediría que las enzimas productoras de ROS que se encuentran en ella pudieran llevar a cabo su función (Cogliati et. al, 2013). Para poder lograr eso, lo principal es mantener la fluidez de la membrana mitocondrial en la transición de un ambiente hidratado a deshidratado y viceversa. Hasta ahora se ha visto que dentro de otra familia de IDP, las proteínas LEA de plantas, existe al menos un grupo de proteínas capaz de salvaguardar la fluidez de esta membrana; y estas son las proteínas LEA_4 (Boothby et. al, 2020). En condiciones de desecación las proteínas LEA_4 adquieren una conformación de alfa-hélice que acomoda los aminoácidos cargados en su secuencia de tal manera que le permite intercalarse entre los fosfolípidos de la membrana. Esto mantiene un espaciamiento adecuado lo cual evita que la membrana mitocondrial pase a un estado de gel aberrante (Boothby et. al, 2020). Como las proteínas MAHS muestran una distribución similar de aminoácidos cargados y no

cargados, se propone que serían capaces de tomar una conformación helicoidal y actuar del mismo modo que las proteínas LEA_4, pero aún no se sabe cómo es que las condiciones intracelulares en el tardígrado podrían afectar esta acción (Boothby et. al, 2020).

Actualmente falta realizar más estudios para clarificar los mecanismos de acción de las TDPs y queda abierta la pregunta de si las TDPs son proteínas que surgieron como modificaciones de las proteínas LEA simplemente para adaptarse a la biología específica de los tardígrados o si las diferencias a nivel de secuencia les confirieron nuevas funciones (Boothby et. al, 2020). En esta tesis, se ahondará en las características fisicoquímicas de las proteínas LEA y quizás esto ayude al lector a acercarse a una respuesta a las preguntas anteriores.

I.II Proteínas Intrínsecamente Desordenadas

Las proteínas son un tipo de macromolécula esencial para la vida. Hay proteínas que participan en procesos de señalización y de transporte, otras que actúan como enzimas, y también están las proteínas estructurales. En animales, adicionalmente participan en la contracción muscular y actúan como hormonas; por lo que son un tema de estudio muy amplio. Su función radica en dos propiedades principales: su secuencia de aminoácidos, y su estructura tridimensional (terciaria y cuaternaria).

El enlace peptídico por sí mismo tiene un comportamiento planar pero debido a la diferencia de cargas de cada aminoácido es que las proteínas pueden obtener conformaciones secundarias y terciarias. En contraste con las proteínas globulares, las proteínas intrínsecamente desordenadas (IDP, *Intrinsically Disordered Proteins*) carecen de estructuras secundarias o terciarias bien definidas, y se comportan como un ensamble de conformaciones constantemente cambiante. Debido a su falta de estructura, el estudio de las IDP es complicado. Actualmente se han creado parámetros para cuantificar diferentes propiedades de las IDP como la cantidad de cargas en una proteína y la distribución de las mismas. La fracción de residuos cargados (FCR, *Fraction of Charged Residues*) de la mano de la distribución de cargas (cuantificado con un parámetro conocido como *kappa*) han permitido crear un modelo para predecir si una proteína tendrá conformación tipo horquilla, con cargas polarizadas en segmentos, o si será desordenada al tener una distribución homogénea de cargas a lo largo de su secuencia (Das y Pappu 2013).

Las IDP juegan un papel importante en el desarrollo de las plantas. En plantas tienen funciones tipo-chaperona, de factores de transcripción, entre otros; que afectan el correcto funcionamiento de diferentes procesos fisiológicos y la formación de semillas (Covarrubias et. al, 2017). Adicionalmente, se han predicho IDP que están fuertemente relacionadas con la respuesta al estrés, aunque no todas han sido confirmadas como IDP experimentalmente. Actualmente, se conocen al menos tres familias prometedoras: las Proteínas LEA, las proteínas pequeñas de choque térmico (sHSP, *small Heat Shock Proteins*), y las proteínas de unión a RNA ricas en glicina (GR-RBP, *Glycine Rich RNA Binding Proteins*) (Covarrubias et. al, 2017). De las anteriores, en este proyecto ahondaremos en las proteínas LEA.

I.III Proteínas abundantes en la embriogénesis tardía

Existe un subgrupo de las IDP llamadas proteínas LEA, que tienen propiedades particulares. Se acumulan en la última etapa de la embriogénesis de las semillas la cual, corresponde al momento en que la semilla adquiere tolerancia a la desecación (Battaglia et. al, 2008). Uno de los rasgos que las caracterizan es su baja incidencia de aminoácidos como cisteína y triptófano; y alta incidencia de aminoácidos pequeños y/o hidrofílicos como glicina, alanina, glutamina, lisina, arginina y treonina (Battaglia et. al, 2008).

Las proteínas LEA se clasifican en grupos de acuerdo a motivos de aminoácidos altamente conservados evolutivamente, no solo en *Arabidopsis thaliana* sino en las proteínas LEA de distintas especies como algunas algas, angiospermas, gimnospermas, pteridofitos entre otros (Battaglia et. al, 2008). Una de estas clasificaciones es la de *Pfam* (Mistry et. al 2021). En esta clasificación los grupos o familias son: LEA_1, LEA_2, LEA_3, LEA_4, LEA_5, LEA_7, dehidrinas, PvLEA18, AtM y SMP (Battaglia et. al, 2008). A su vez, estas se subclasifican en dos grupos de acuerdo a su hidropatía: las atípicas o hidrofóbicas (LEA_2, LEA_3, SMP), y las típicas o hidrofílicas (LEA_1, LEA_4, LEA_5, LEA_7, PvLEA18, AtM) (Battaglia et. al, 2008). A la fecha no se han encontrado otras proteínas similares en cuanto a su secuencia primaria cuya función ya se conozca; por lo que estudiar a las proteínas LEA resulta un gran reto (Covarrubias et. al, 2017).

Se sabe que las proteínas LEA pueden tener más de una función y que algunas de ellas actúan como chaperonas por su capacidad de protección hacia enzimas en respuesta a condiciones de estrés. (Olvera-Carrillo et. al, 2011). Otra vertiente habla de cómo las proteínas LEA son capaces de proteger a otras proteínas al tomar una conformación de alfa hélice debido al hacinamiento macromolecular, condición prevalente durante condiciones de déficit hídrico (Cuevas-Velazquez et. al, 2016). En diversos estudios usando mutantes de *Arabidopsis* que carecían, de forma independiente, de genes que codifican para proteínas de los grupos LEA_1, LEA_2, LEA_3, y LEA_4, se observó que estas proteínas eran necesarias para que la planta pudiera adaptarse a la desecación, el frío y concentraciones elevadas de sal (Covarrubias et. al, 2017). Aunque hasta ahora su mecanismo de acción no sea claro, se cree que el ser IDP está relacionado con su capacidad de tener múltiples funciones en los organismos que las poseen (Tompa et. al, 2005).

I.IV Propiedades fisicoquímicas de las proteínas

Las propiedades fisicoquímicas son, como su nombre lo dice, aquello que describe a un elemento, un compuesto, una sustancia, una mezcla o una molécula, en términos de su comportamiento físico y químico. Algunas propiedades fisicoquímicas son: punto de fusión, punto de ebullición, solubilidad y pH. Si nos enfocamos en macromoléculas, sus propiedades fisicoquímicas dependen en parte de las propiedades fisicoquímicas de sus monómeros. En el caso de las proteínas, analizar la composición de los aminoácidos que las conforman sirve para describir sus propiedades fisicoquímicas. El conocer las propiedades fisicoquímicas de cualquiera de los ejemplos antes mencionados sirve para entender su comportamiento en diferentes circunstancias; en particular para este proyecto tiene relevancia comprender el comportamiento de las proteínas LEA ante el déficit hídrico. Actualmente existen varios

parámetros fisicoquímicos, cuantificables a partir de la secuencia primaria, que permiten caracterizar cualquier proteína como lo son los mostrados a continuación.

Kappa (K): *Kappa* cuantifica el nivel de distribución de cargas opuestas a lo largo de una proteína. Un valor elevado de *kappa* corresponde a proteínas que tienden a una única conformación y, por el contrario, un valor bajo describe a proteínas que constantemente cambian de conformación (Holehouse et. al, 2015).

Fracción de residuos cargados (FCR, Fraction of Charged Residues): Cuantifica la proporción de residuos con carga positiva o negativa en una proteína. FCR es un valor absoluto, es decir, es indistinto si la carga es positiva o negativa, sino que se enfoca en que tan cargada es una proteína en general (Das y Pappu 2013). Una proteína altamente cargada tendrá un valor de FCR alto y siguiendo la misma lógica una proteína poco cargada tendrá valores bajos de FCR. Este parámetro surge a partir del diagrama de fases de Das-Pappu descrito en el Anexo (Holehouse et. al, 2015).

Cabe resaltar que entre mayor sea el valor de FCR, *kappa* cobrará mayor relevancia; es decir, que entre más cargada esté una proteína tiene mayor relevancia la distribución de las cargas a lo largo de la secuencia (Holehouse et. al, 2015). En 2013 se llevó a cabo un estudio en donde se analizó la relación de *kappa* con FCR empleando 364 IDP (Das y Pappu 2013). A partir de dicho estudio notaron que la mayoría de las proteínas IDP se agrupaban cerca de valores de *kappa* de 0.2 (Das y Pappu 2013).

Hidropatía: La hidropatía es la tendencia de repeler el agua (Kaliszan 2015). Usualmente, en proteínas, se ha asociado la hidropatía con la presencia de residuos no polares en su secuencia (Caron, 2007); lo cual hasta ese punto sería meramente cualitativo. Para darle un valor numérico a la hidropatía, *Kyte* y *Doolittle* le asignaron un valor a cada aminoácido en diferentes escalas o intervalos: la escala original de *Kyte & Doolittle* con valores de -4.5 a 4.5; la escala de hidropatía media o *mean_hydrophathy*, con valores de 0 a 9; y la escala de *Uversky*, con valores de 0 a 1. A partir de los valores para cada aminoácido se obtiene la media para toda la proteína y ese es el valor reportado como hidropatía de la proteína. Valores altos de hidropatía hablan de proteínas hidrofóbicas y por el contrario, valores bajos indican que las proteínas son hidrofílicas (Holehouse et. al, 2015).

Composición de aminoácidos promotores del desorden estructural: Esta predicción de desorden se basa en el contenido de ciertos aminoácidos, que ya se ha demostrado que promueven el desorden. Los aminoácidos considerados como promotores del desorden estructural son: treonina, alanina, glicina, arginina, ácido aspártico, histidina, glutamina, lisina, serina, ácido glutámico, y prolina. Los aminoácidos que promueven el orden estructural son: triptófano, fenilalanina, tirosina, isoleucina, metionina, leucina, valina, asparagina, y cisteína. (Holehouse et. al, 2015).

Carga neta media (MNC, Mean Net Charge): Cuantifica el valor absoluto del promedio de cargas presentes en una secuencia de aminoácidos sin importar su distribución. (Holehouse et. al, 2015).

II. Antecedentes

Ante condiciones extremas de estrés, algunos organismos han desarrollado diversos mecanismos para sobrevivir. Las plantas por ejemplo se enfrentan a cambios drásticos en su entorno durante periodos de sequías por falta de lluvias. A diferencia de los animales e insectos que tienen la capacidad de migrar en busca de un ambiente adecuado para evitar el estrés, las plantas no tienen esta opción y por ello su supervivencia es más sensible a los cambios ambientales. Se espera que el cambio climático incremente el número y la duración de condiciones adversas como la sequía. Por tal motivo, resulta de interés entender de qué manera las plantas responden y se aclimatan a condiciones de falta de agua (deshidratación y/o desecación), debido a que los cultivos son la fuente principal de alimentos en el mundo.

Previamente se han realizado estudios en organismos tolerantes a la desecación y dentro de los múltiples factores que participan de manera simultánea para conferir tolerancia a esos organismos se han señalado a dos macromoléculas principalmente: carbohidratos y proteínas. La trehalosa es un disacárido no reductor que tiene la habilidad de mitigar o casi detener las afecciones dentro de las células por medio de una vitrificación (Crowe, 1998); sin embargo, recientemente se ha visto que algunos organismos que no poseen las herramientas necesarias para sintetizar trehalosa aún son capaces de tolerar la desecación (Boothby *et. al* 2020); y lo que tienen en común dichos organismos son las proteínas intrínsecamente desordenadas (IDP, *Intrinsically Disordered Proteins*).

En los tardígrados existe un grupo de proteínas conocidas como proteínas desordenadas de tardígrado (TDP, *Tardigrade Disordered Proteins*), las cuales se ha propuesto que le confieren la capacidad de tolerar condiciones de desecación (Boothby *et. al* 2017). Boothby y colaboradores realizaron un estudio en el que demostraron que a pesar de los niveles bajos o ausencia de trehalosa, los tardígrados aún eran capaces de tolerar la desecación y que, en contraste, los tardígrados presentaron sensibilidad a la desecación cuando se silenciaron algunos genes que codifican para TDP. Como conclusión, se encontró que los tardígrados requieren de las proteínas TDP para sobrevivir condiciones de desecación (Boothby *et. al* 2017).

Otra familia de proteínas intrínsecamente desordenadas son las proteínas abundantes en la embriogénesis tardía (LEA, *Late Embryogenesis Abundant*), las cuales son ubicuas en el reino vegetal, incluyendo las plantas de resurrección (Battaglia *et. al*, 2008). La mayoría de las proteínas LEA tienden a ser altamente hidrofílicas, contienen una alta proporción del aminoácido glicina (Battaglia *et. al*, 2008) y, en su mayoría, se predice que son estructuralmente desordenadas, lo cual hace que pertenezcan al grupo de las IDP (Hundertmark y Hinch, 2008). En diversos estudios se ha observado que los genes que codifican para las proteínas LEA están asociados con distintas respuestas a condiciones de estrés como respuesta a frío y desecación (Hundertmark y Hinch, 2008); por lo que surge la pregunta de si las proteínas LEA son las responsables de conferir la capacidad de tolerar la desecación a los organismos que las poseen.

Resulta interesante que dos tipos de organismos tolerantes a la desecación que son evolutivamente distantes, como las plantas de resurrección y los tardígrados, acumulen IDP para contrarrestar los efectos causados por la falta de agua. Las TDP y las LEA no comparten motivos

conservados a nivel de secuencia primaria sin embargo, su composición de aminoácidos es similar. Inclusive, las proteínas LEA de las siete diferentes familias no comparten homología de secuencia entre sí, sin embargo, su composición de aminoácidos es muy similar (Battaglia et. al, 2008). Dichas observaciones sugieren que las propiedades fisicoquímicas de las proteínas efectoras que participan en la tolerancia a la desecación son fundamentales para ejercer su función protectora, pero a la fecha no se ha realizado una caracterización y cuantificación sistemática de las propiedades fisicoquímicas de este tipo de proteínas.

III. Hipótesis

Los diferentes grupos de proteínas LEA de plantas poseen propiedades fisicoquímicas que las distinguen del resto del proteoma, y dichas propiedades están conservadas en proteínas del tardígrado *Hypsibius dujardini*, un animal tolerante a la desecación.

IV. Objetivo general

Caracterizar las propiedades fisicoquímicas de las proteínas LEA de las plantas *Arabidopsis thaliana*, *Oryza sativa*, *Triticum aestivum* y *Zea mays*, y buscar proteínas con propiedades similares en el proteoma del tardígrado *Hypsibius dujardini*.

V. Objetivos particulares

- Analizar las propiedades fisicoquímicas de las proteínas LEA de *A. thaliana*, *O. sativa*, *T. aestivum* y *Z. mays*; codificadas en su secuencia primaria.
- Contrastar las propiedades fisicoquímicas de las proteínas LEA con las del resto del proteoma de las plantas seleccionadas.
- Establecer los parámetros que distinguen a las proteínas LEA de plantas.
- Buscar proteínas con propiedades fisicoquímicas similares a las proteínas LEA en el proteoma de *Hypsibius dujardini*.
- Analizar la ontología genética de aquellas proteínas que cumplen con las propiedades fisicoquímicas y rangos seleccionados en *Arabidopsis thaliana*.

VI. Metodología

VI.I Obtención y limpieza de proteomas

Se descargaron los proteomas de referencia (Tabla 1) de *Arabidopsis thaliana*, *Oryza sativa* (arroz), *Triticum aestivum* (trigo), *Zea mays* (maíz) e *Hypsibius dujardini* (tardígrado) de la base de datos UniProt (*Universal Protein Resource*, <https://www.uniprot.org>). Primero se limpió el archivo de proteínas con *protfasta* en la plataforma de Spyder (versión 5.0.1) para eliminar las proteínas con aminoácidos o caracteres ambiguos (por ejemplo: B/U/X/*/-).

Para obtener las proteínas LEA de cada especie, se buscaron artículos en donde se caracterizó experimentalmente su presencia en las plantas de interés (Tabla 1). Al hacer esto, existía la posibilidad de encontrar proteínas enlistadas dos veces en la base de datos de los proteomas que se generó, dado que las proteínas LEA provenían de diferentes artículos y no de Uniprot. Para depurar la lista se realizó una segunda limpieza de los datos mediante BLASTp para eliminar las proteínas duplicadas con el programa *Bioedit* (versión 7.0.9.0). Los criterios de BLASTp que se consideraron fueron: *bits* mayores a 50, porcentaje de identidad mayor a 30 y un *E-value* menor a 1×10^{-20} . De este modo, se eliminaron los primeros *hits* de cada proteína LEA que indicaran duplicaciones.

VI.I.I Análisis de las propiedades fisicoquímicas

Los parámetros fisicoquímicos de las proteínas LEA y del resto de los proteomas se calcularon a partir de su secuencia primaria. Primero se utilizó *splitfasta* para dividir los archivos multifasta en archivos FASTA individuales, es decir, un archivo independiente para cada secuencia. Después, los archivos individuales se analizaron con *localCIDER* v0.1.18 (Holehouse et. al, 2015) para obtener los valores de hidropatía (en la escala de uversky), la composición de aminoácidos promotores del desorden, *kappa*, y carga neta media para cada proteína. Con estos datos se construyeron gráficas de dispersión utilizando *RStudio* (versión 1.3.1093) con la librería *ggplot2*.

VI.I.II Determinación de rangos estadísticos

La determinación de los rangos de las propiedades fisicoquímicas que distinguen a las proteínas LEA se realizó mediante un análisis estadístico. En primer lugar, se juntaron los datos de los parámetros de hidropatía y composición de aminoácidos promotores del desorden de todos los proteomas de las plantas analizadas (*A. thaliana*, *O.sativa*, *T. aestivum* y *Z. mays*) y se crearon unas gráficas de caja con la paquetería *plotly* en R. Con estas gráficas se estableció el intervalo de cada parámetro bajo el siguiente criterio: que contuviera al 75 % de las proteínas LEA. Los intervalos seleccionados fueron entre 0.000000 y 0.459596 para hidropatía, y entre 0.7553957 y 1.0000 para la composición de aminoácidos promotores del desorden. Considerando estos intervalos, se filtraron las proteínas de los cuatro organismos; esto se denominó como el primer corte.

A continuación se compararon los parámetros de *kappa* y carga neta media para determinar un segundo corte. Para ello se crearon dos gráficas de caja empleando únicamente las proteínas de plantas que pasaron el primer corte. A partir de esto se concluyó que el 75 % de las proteínas LEA, que pasaron el primer corte, se encontraban entre 0.0000000 y 0.03479037. A todas las proteínas, de los proteomas de plantas, que pasaron ambos cortes se les denominó como proteínas tipo-LEA. Finalmente, se aplicaron estos dos rangos estadísticos (cortes) al proteoma de *H. dujardini*.

VI.I.III Análisis ontológico

Se tomaron los identificadores de UniProt de las proteínas que no pasaron ningún corte, las que sí pasaron el primer corte, y las que pasaron el segundo corte de *A. thaliana* para llevar a cabo un análisis de enriquecimiento con *Gene Ontology* (Ashburner et. al 2000) a tres niveles en el sitio <http://geneontology.org>: función molecular, proceso biológico y componente celular. Se aplicó una prueba exacta de Fisher a los datos con una tasa de corrección de falsos descubrimientos (FDR) y considerando un valor de $p < 0.05$.

VII. Resultados

El objetivo de este proyecto fue caracterizar las propiedades fisicoquímicas que distinguen a las proteínas LEA y determinar si dichas propiedades están presentes en proteínas del tardígrado *Hypsibius dujardini*, un animal tolerante a la desecación. Para ello, primero se analizaron las propiedades fisicoquímicas con el algoritmo CIDER para los proteomas completos de *A. thaliana*, arroz, maíz, y trigo. Después se crearon gráficos con el fin de encontrar aquellos parámetros que mejor distinguieran a las proteínas LEA del resto del proteoma. Posteriormente, se analizó el proteoma de *H. dujardini* bajo los mismos criterios y finalmente se realizó un estudio de ontología genética a *A. thaliana*.

Debido a que la planta modelo *A. thaliana* es la planta más estudiada, se comenzó trabajando con ella para comparar las propiedades fisicoquímicas de sus 51 proteínas LEA (Hundertmark y Hinch, 2008) y el resto de proteínas de su proteoma (Tabla 1), y así tener una idea general de las propiedades particulares de las proteínas LEA. En primer lugar, se analizó la fracción de residuos cargados (FCR, *Fraction of Charged Residues*) y la distribución de las cargas positivas y negativas a lo largo de cada secuencia. La distribución de cargas se cuantificó utilizando el parámetro de *kappa*, el cual indica el grado de segregación de las cargas positivas y negativas (Holehouse *et. al*, 2015). Una secuencia con una distribución homogénea de cargas tiende a un valor de 0, mientras que una secuencia con cargas totalmente segregadas tiende a un valor de 1. Cerca del 97 % de las proteínas del proteoma de *A. thaliana* tiene valores de FCR entre 0.1 y 0.4, así como valores de *kappa* entre 0.1 y 0.3 (Figura 1). Al analizar a las proteínas LEA, se encontró que en general presentan valores de *kappa* y FCR similares al resto del proteoma de *A. thaliana* (Figura 1).

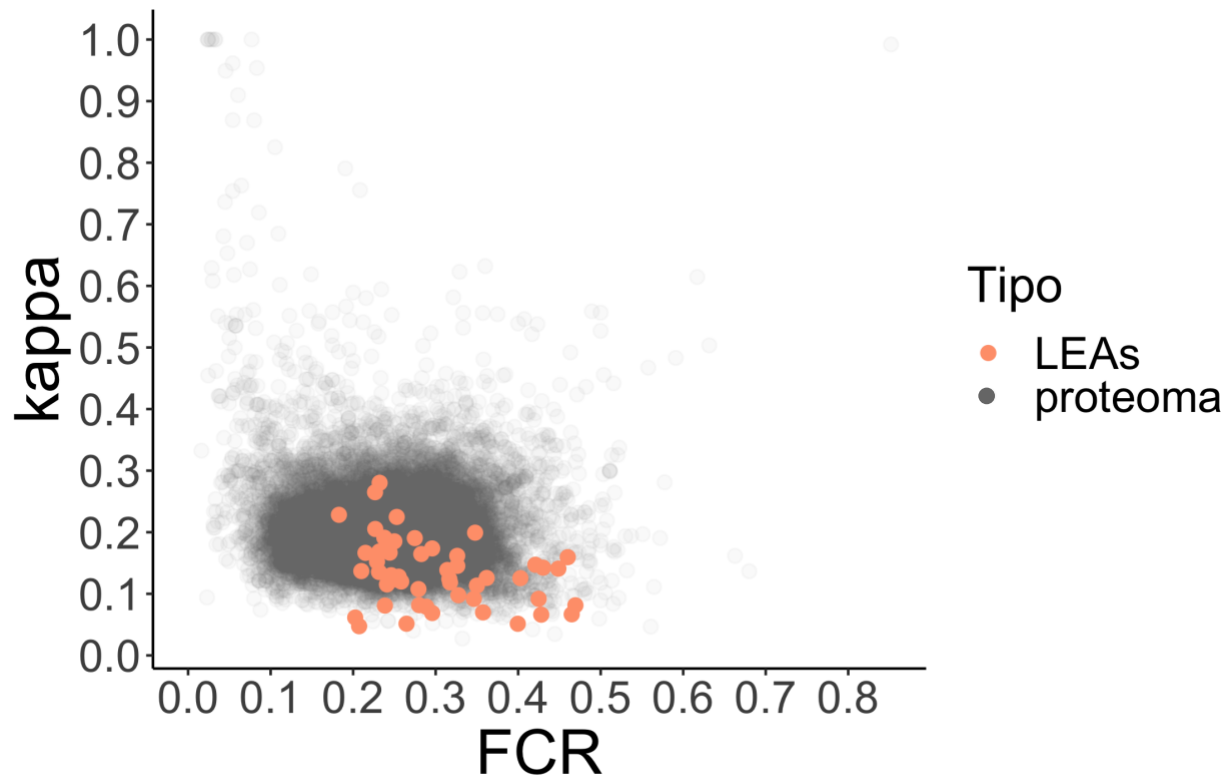


Figura 1. Distribución de cargas opuestas (*kappa*) frente a proporción de cargas (FCR) en *A. thaliana*. En color salmón se muestran las proteínas LEA mientras que los puntos en color gris ($\alpha = 1/26$) son el resto de las proteínas del proteoma de *A. thaliana*.

Posteriormente, se compararon los parámetros de carga neta media, composición de aminoácidos promotores del desorden e hidropatía. Para facilitar las comparaciones, cada uno de los parámetros mencionados se graficó contra los valores de FCR. Al analizar la carga neta media, es decir, el valor absoluto de la carga de la secuencia normalizada por la longitud de la misma, se encontró que el 99.7 % (39,214) de las proteínas del proteoma de *A. thaliana* muestran valores entre 0 y 0.2 (Figura 2). La distribución de carga neta media en las proteínas LEA mostró una tendencia a valores bajos, sin embargo, fue similar a la observada en las proteínas del proteoma (Figura 2).

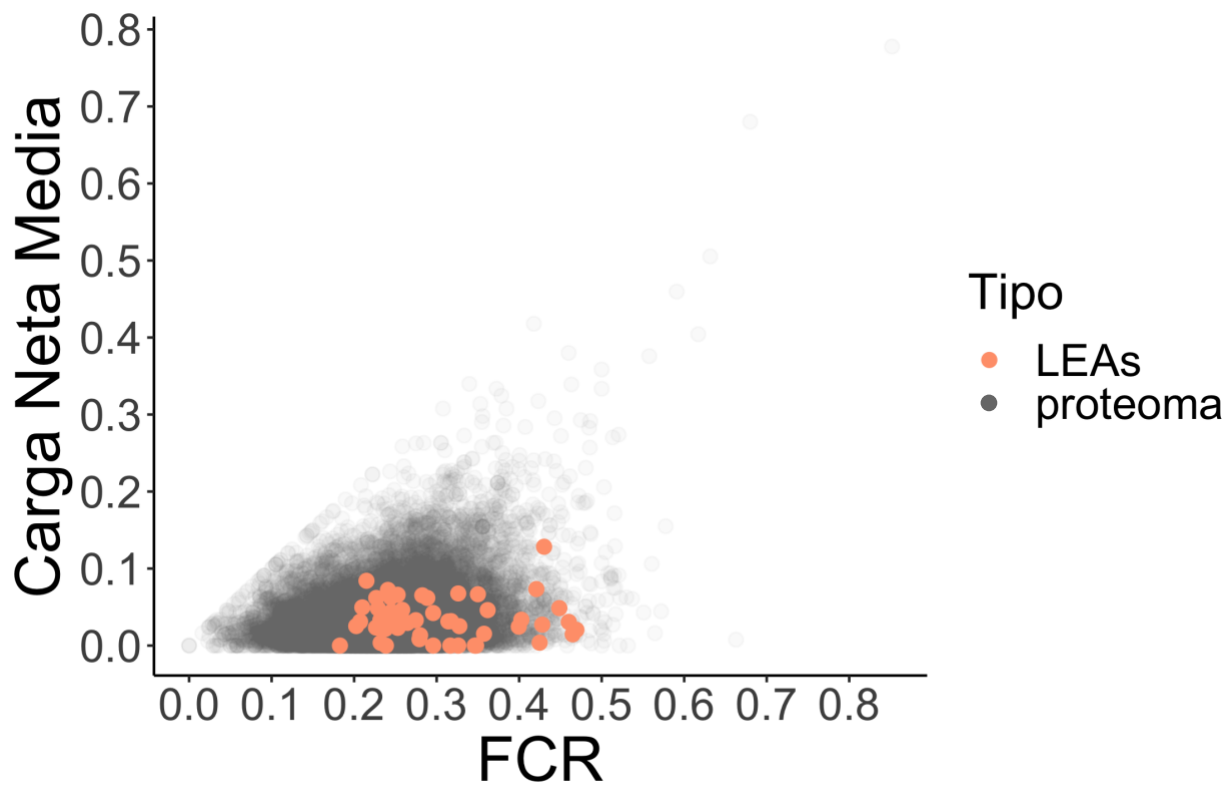


Figura 2. Carga neta media frente a proporción de cargas (FCR) en *A. thaliana*. En color salmón se muestran las proteínas LEA mientras que los puntos en color gris ($\alpha=1/26$) son el resto de las proteínas del proteoma de *A. thaliana*.

En el caso de la hidropatía, que es un parámetro que cuantifica la proporción de aminoácidos hidrofóbicos o hidrofílicos en una secuencia (las proteínas hidrofílicas presentan niveles de hidropatía que tienden a 0, mientras que las proteínas hidrofóbicas presentan niveles de hidropatía que tienden a 1), se encontró que 38,873 proteínas del proteoma de *A. thaliana*, es decir el 98.8 % del proteoma, muestran niveles de hidropatía entre 0.35 y 0.6 (Figura 3). Por otro lado, las proteínas LEA mostraron un patrón interesante. Algunas mostraron niveles de hidropatía comparables con el resto del proteoma sin embargo, se encontró un grupo de proteínas LEA con valores de hidropatía menores a 0.4 (Figura 3), lo que sugiere que los bajos niveles de hidropatía podrían ser un parámetro característico de algunas de las proteínas LEA.

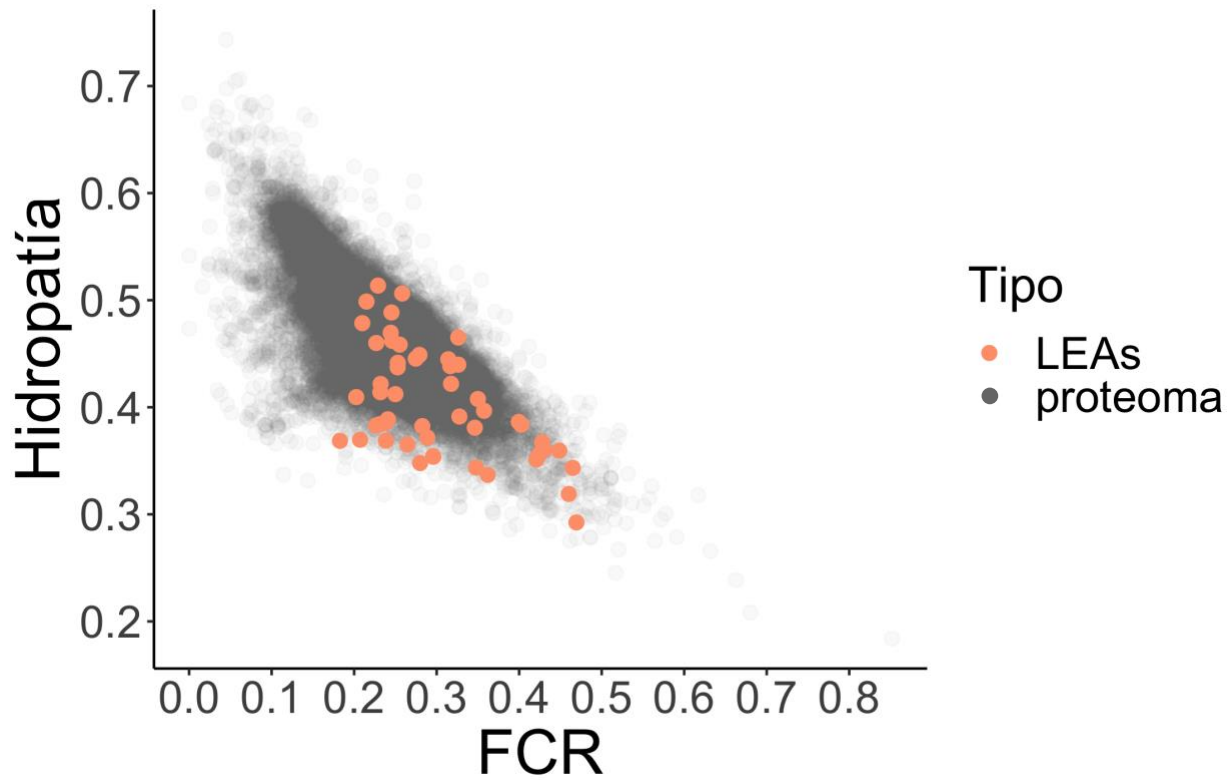


Figura 3. Hidropatía frente a proporción de cargas (FCR) en *A. thaliana*. En color salmón se muestran las proteínas LEA mientras que los puntos en color gris ($\alpha = 1/26$) son el resto de las proteínas del proteoma de *A. thaliana*.

Finalmente, se analizó la composición de aminoácidos promotores del desorden. En este caso se observó que el 92.8 % del proteoma de *A. thaliana* (lo que corresponde a 36,488 proteínas) tienen una composición de aminoácidos que promueven el desorden entre 0.4 y 0.7 (Figura 4). De acuerdo a lo reportado previamente, la mayoría de las proteínas LEA (44 de 51, es decir 86.3 %) presentan niveles mayores a 0.7, siendo este parámetro el que mejor distingue a las proteínas LEA (Figura 4).

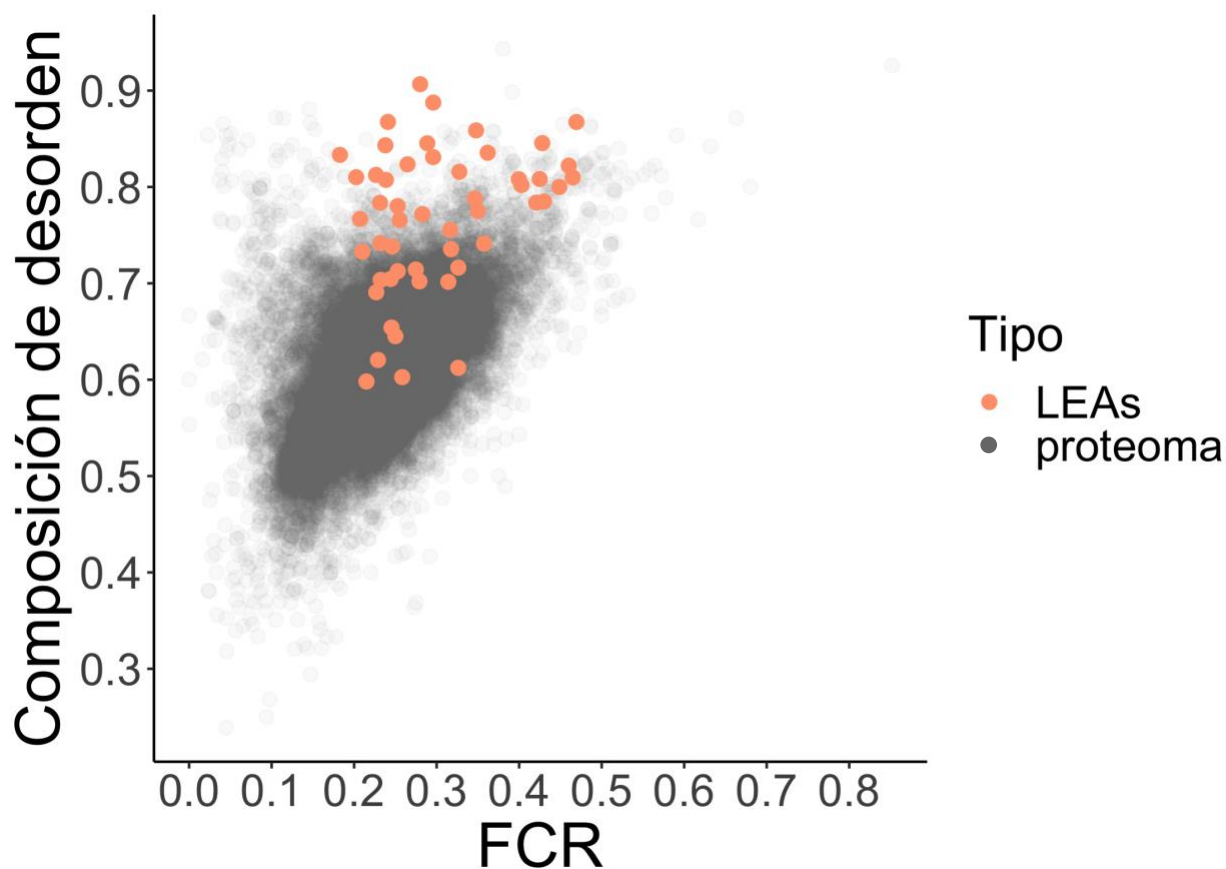


Figura 4. Composición de aminoácidos promotores del desorden estructural frente a proporción de cargas (FCR) en *A. thaliana*. En color salmón se muestran las proteínas LEA mientras que los puntos en color gris ($\alpha = 1/26$) son el resto de las proteínas del proteoma de *A. thaliana*.

Debido a que la composición de aminoácidos promotores del desorden y la hidropatía fueron los parámetros que mejor distinguen a las proteínas LEA del resto del proteoma de *A. thaliana*, se realizó una representación gráfica utilizando estos parámetros. Como se observa en la Figura 5, la mayoría de las proteínas LEA tienen una alta proporción de aminoácidos hidrofílicos y que promueven el desorden.

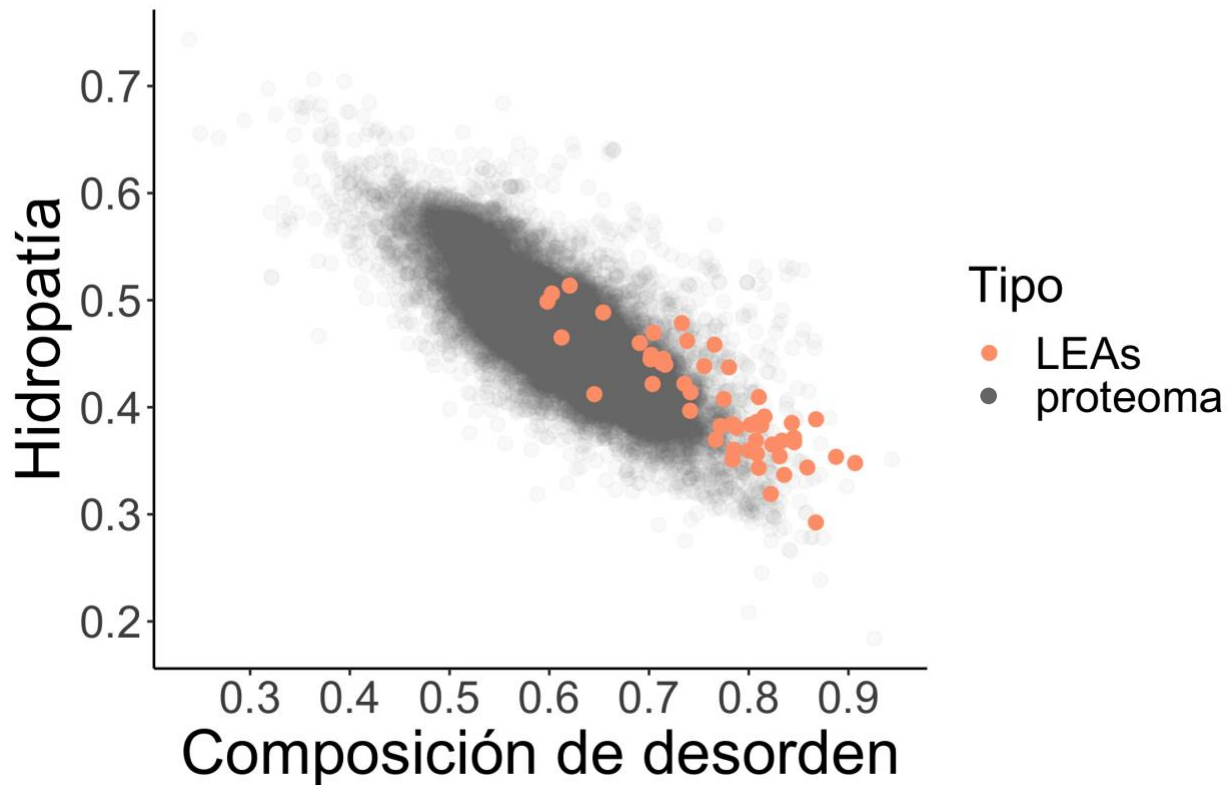


Figura 5. Hidropatía frente a la composición de aminoácidos promotores del desorden estructural. En color salmón se muestran las proteínas LEA mientras que los puntos en color gris ($\alpha = 1/26$) son el resto de las proteínas del proteoma de *A. thaliana*.

El análisis anterior se extendió utilizando otras plantas altamente estudiadas que cuentan con una caracterización detallada en la literatura de los genes LEA. Las plantas seleccionadas fueron: *Oryza sativa* (arroz), *Triticum aestivum* (trigo), y *Zea mays* (maíz). El tamaño de los proteomas y el número de proteínas LEA se describen en la Tabla 1.

La comparación de la composición de aminoácidos promotores del desorden y la hidropatía de las proteínas LEA de *A. thaliana*, arroz, trigo y maíz y el resto de proteínas del proteoma se muestra en la Figura 6. Este análisis nos permitió confirmar las observaciones obtenidas sólo con *A. thaliana*.

Tabla 1. Proteomas analizados en este trabajo.

Organismo	ID Uniprot	# de proteínas	# de genes LEAs	Referencia
<i>Arabidopsis thaliana</i>	UP000006548	39,337	51	(Hundertmark y Hinch, 2008)
<i>Oryza sativa</i>	UP000007015	37,383	37	(Wang et. al, 2007)
<i>Triticum aestivum</i>	UP000019116	130,673	114	(Liu et. al, 2019)
<i>Zea mays</i>	UP000007305	99,253	32	(Li y Cao, 2015)

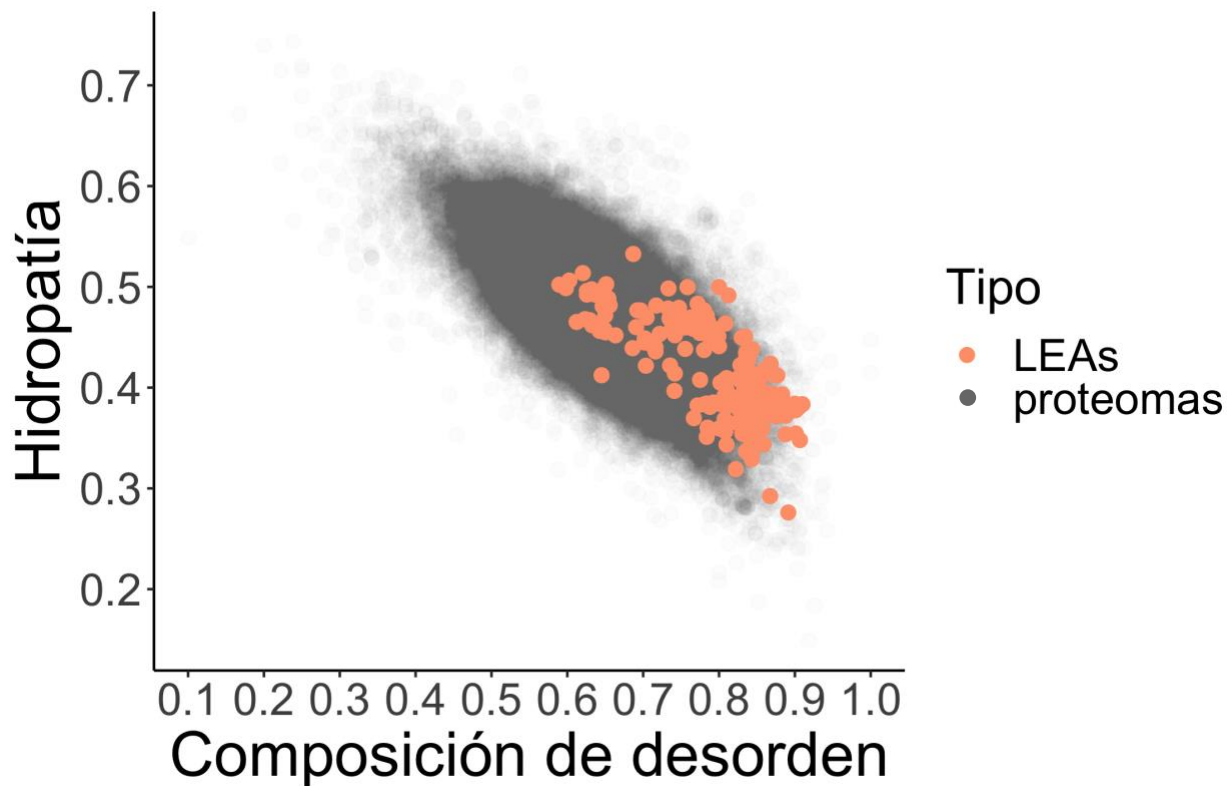


Figura 6. Hidropatía frente a la composición de aminoácidos promotores del desorden estructural. En color salmón se muestran las proteínas LEA de *A. thaliana*, *Oryza sativa* (arroz), *Triticum aestivum* (trigo), y *Zea mays* (maíz) mientras que en color gris ($\alpha = 1/50$) se muestra el resto de las proteínas de los proteomas de plantas.

Con el objetivo de establecer los valores específicos de hidropatía y composición de aminoácidos que promueven el desorden para definir a las proteínas LEA de plantas, se procedió a realizar un análisis estadístico. Para considerar a la mayoría de las proteínas LEA, se realizó un análisis de “box plot” para determinar el valor del cuartil Q1 en el caso de la composición de aminoácidos que promueven el desorden (Figura 7) y el valor del cuartil Q3 en el caso de la hidropatía (Figura 8). Esto permitió realizar cortes que consideran al menos al 75 % de las proteínas LEA en cada caso (arriba del cuartil Q1 para la composición de aminoácidos que promueven el desorden y debajo del cuartil Q3 para la hidropatía). De esta manera, se observó que el 66.2 % de las proteínas LEA pasan el primer corte, es decir, tienen una composición de aminoácidos que promueven el desorden mayor a 0.76 y un valor de hidropatía menor a 0.46 (Figura 7 y Figura 8). En contraste, solo el 3.7 % de las proteínas de los proteomas cumplen con los valores mencionados anteriormente. Es decir, al realizar el primer corte se logró separar a las proteínas LEA del 96.3 % del resto de proteínas de los proteomas de plantas.

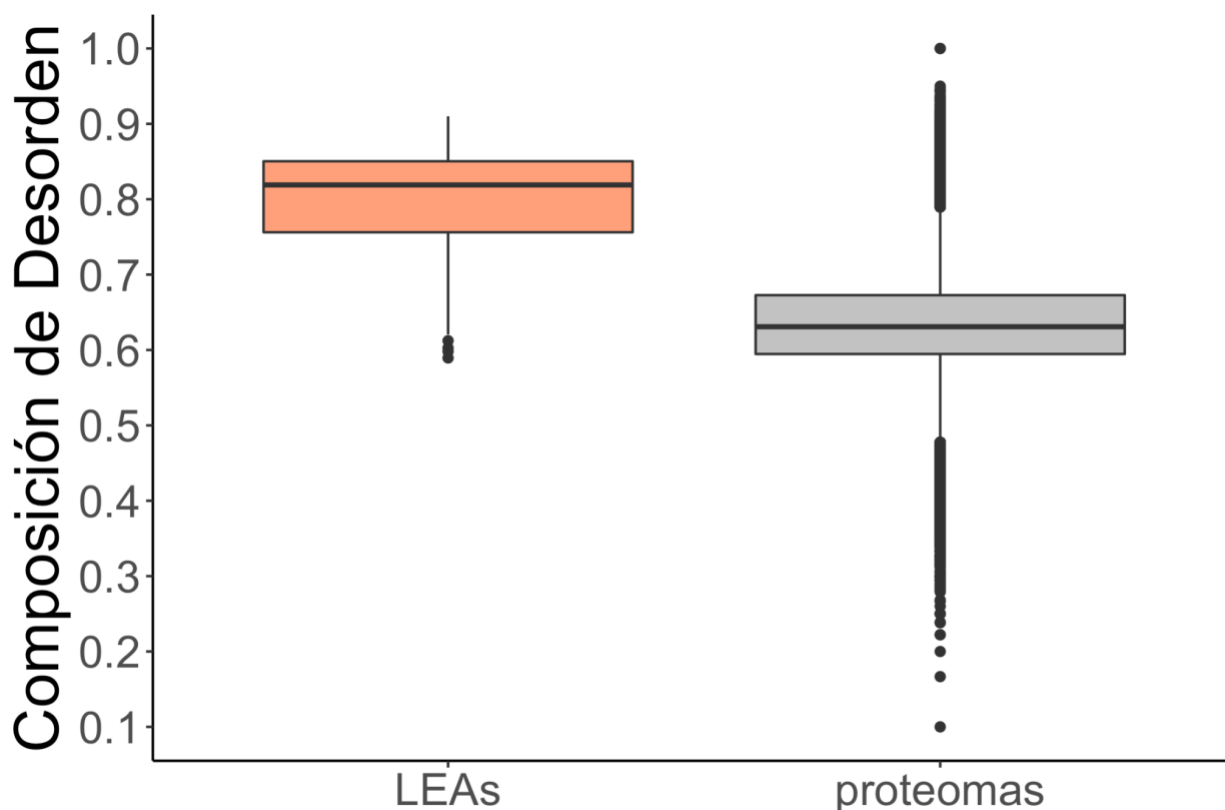


Figura 7. Análisis estadístico de composición de aminoácidos promotores del desorden estructural. En color salmón se muestran las proteínas LEA de *A. thaliana*, *Oryza sativa* (arroz), *Triticum aestivum* (trigo), y *Zea mays* (maíz) mientras que en color gris se muestra el resto de las proteínas de los proteomas de plantas. Las cajas delimitadas por los cuartiles Q1 y Q3 contienen el 50 % de los datos, la mediana representa el cuartil Q2 y en las prolongaciones superior e inferior se encuentra representado el otro 50 % de los datos.

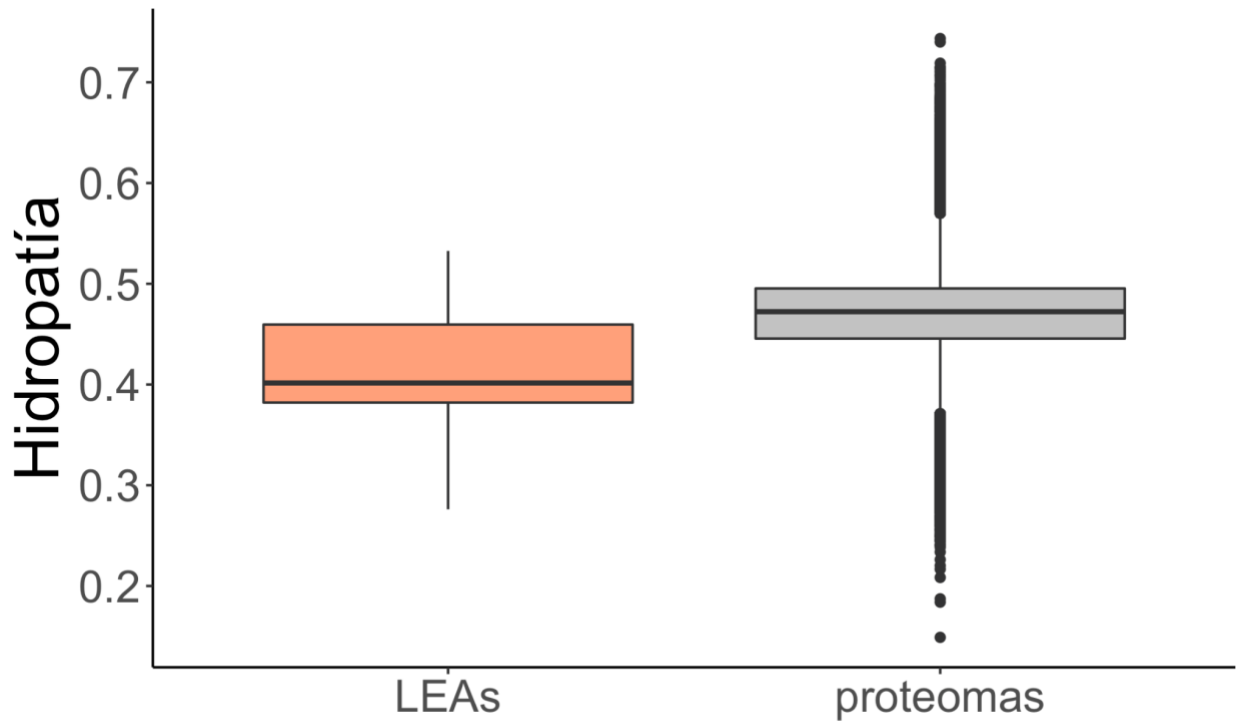


Figura 8. Análisis estadístico de hidropatía. En color salmón se muestran las proteínas LEA de *A. thaliana*, *Oryza sativa* (arroz), *Triticum aestivum* (trigo), y *Zea mays* (maíz) mientras que en color gris se muestra el resto de las proteínas de los proteomas de plantas.

Posteriormente, se aplicó un nuevo parámetro para separar aún más a las proteínas LEA del resto del proteoma: la carga neta media. Como se observa en la Figura 2, las proteínas LEA mostraron una tendencia a valores bajos (menores a 0.1) de carga neta media en *A. thaliana*, antes de aplicar el primer corte. Cuando se tomaron solo a las proteínas que pasaron el primer corte (proteínas LEA y proteínas del resto de los proteomas) y se graficaron en función de FCR y carga neta media, se encontró que la tendencia que tienen las proteínas LEA a presentar valores bajos de carga neta media se acentuó aún más (Figura 9).

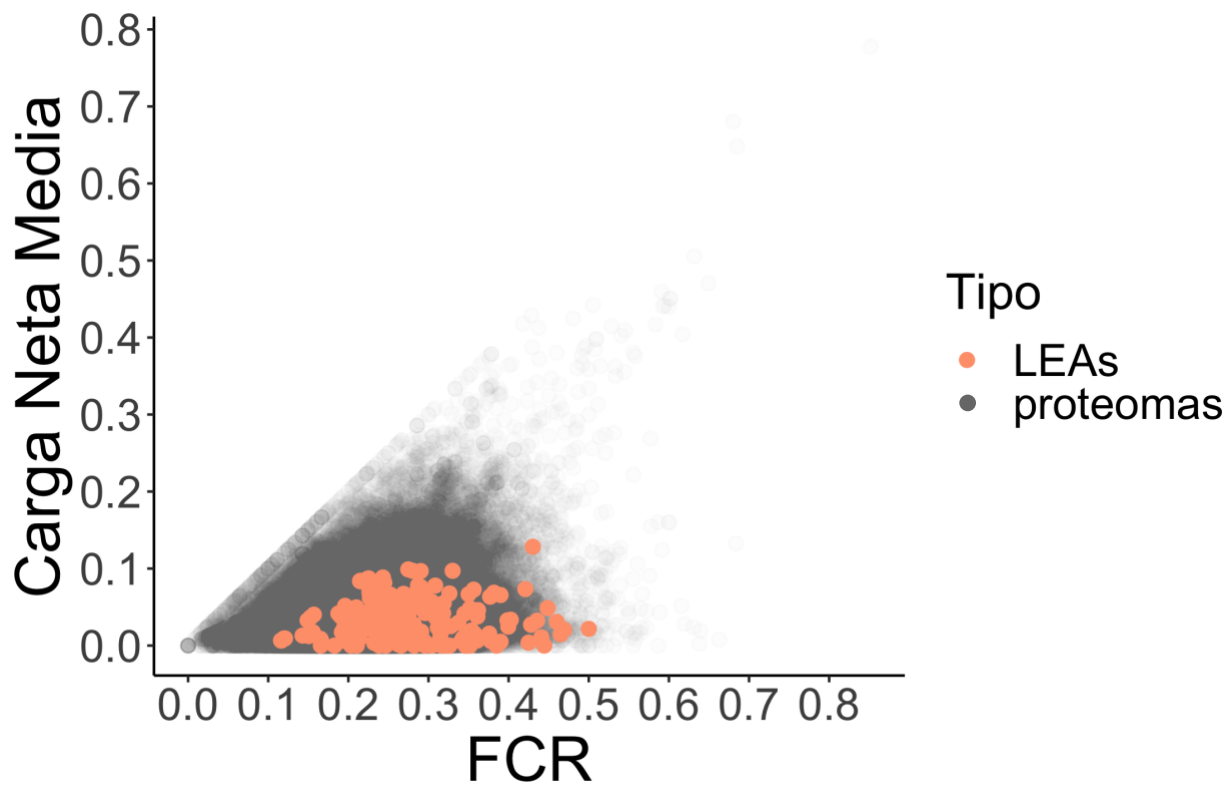


Figura 9. Carga Neta Media frente a FCR en plantas. En color salmón se muestran las proteínas LEA de *A. thaliana*, *Oryza sativa* (arroz), *Triticum aestivum* (trigo), y *Zea mays* (maíz) mientras que en color gris ($\alpha = 1/26$) se muestra el resto de las proteínas de los proteomas de plantas. Las cajas delimitadas por los cuartiles Q1 y Q3 contienen el 50 % de los datos, la mediana representa el cuartil Q2 y en las prolongaciones superior e inferior se encuentra representado el otro 50 % de los datos.

Al tomar en cuenta la carga neta media como un nuevo parámetro, se estableció un segundo corte tomando en cuenta el mismo criterio que el primero, es decir, manteniendo al 75 % de las proteínas LEA. De esta forma, se cuantificó el cuartil Q3 para el grupo de las proteínas LEA, encontrando que el 74.8 % de las proteínas LEA que pasaron el primer corte tienen valores de carga neta media menores a 0.035 (Figura 10), lo cual corresponde al 49.6 % de todas las proteínas LEA de las plantas que se analizaron. De esta forma, al tomar en cuenta los dos cortes se logró separar a las proteínas LEA del 98.6 % de las proteínas del proteoma.

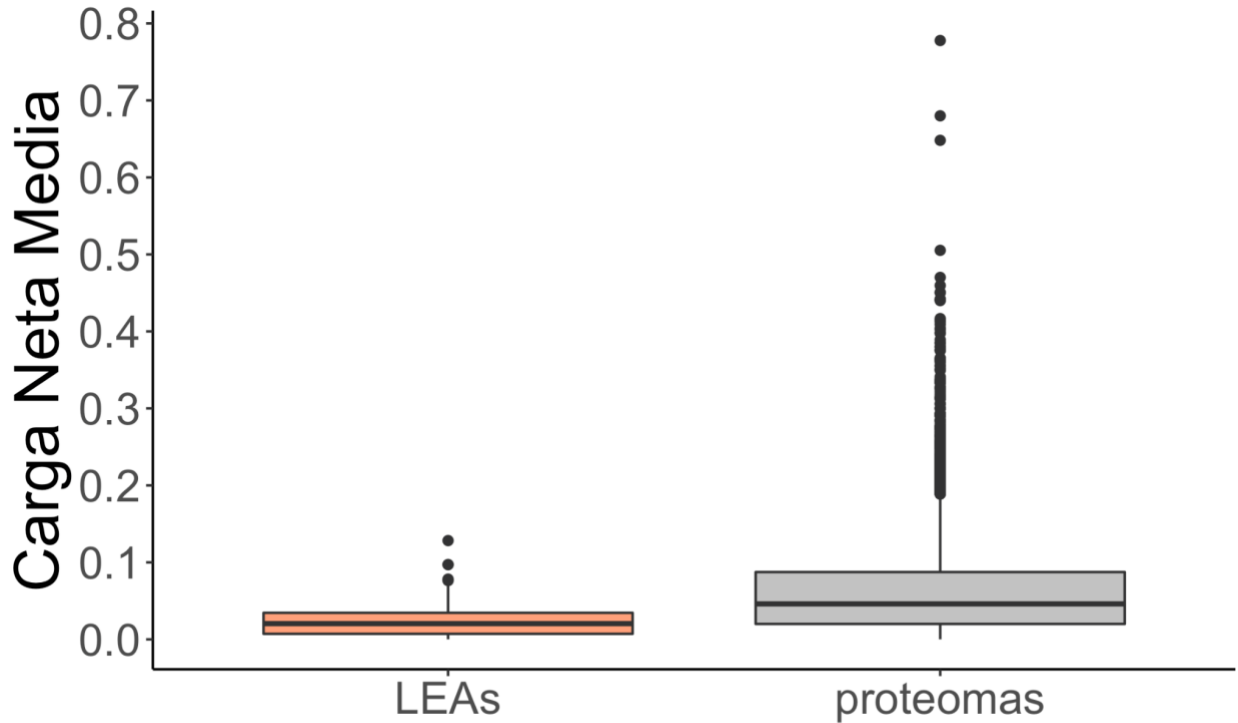


Figura 10. Análisis estadístico de carga neta media. En color salmón se muestran las proteínas LEA de *A. thaliana*, *Oryza sativa* (arroz), *Triticum aestivum* (trigo), y *Zea mays* (maíz) mientras que en color gris se muestra el resto de las proteínas de los proteomas de plantas.

Una vez que se seleccionaron los parámetros que definen a las proteínas LEA, nos preguntamos si otros organismos tolerantes a la desecación tienen proteínas con dichas propiedades fisicoquímicas. De esta forma, se analizó el proteoma del tardígrado *H. dujardini*, que como se mencionó anteriormente, es un organismo tolerante a la desecación que no posee proteínas LEA homólogas. El proteoma de *H. dujardini* contiene 14,867 proteínas, las cuales se graficaron en función de su hidropatía y la composición de aminoácidos que promueven el desorden (Figura 11). Al aplicar los criterios del primer corte, se observó que 230 proteínas pasan dichos criterios, lo que corresponde al 2.8 % del proteoma (Figura 11).

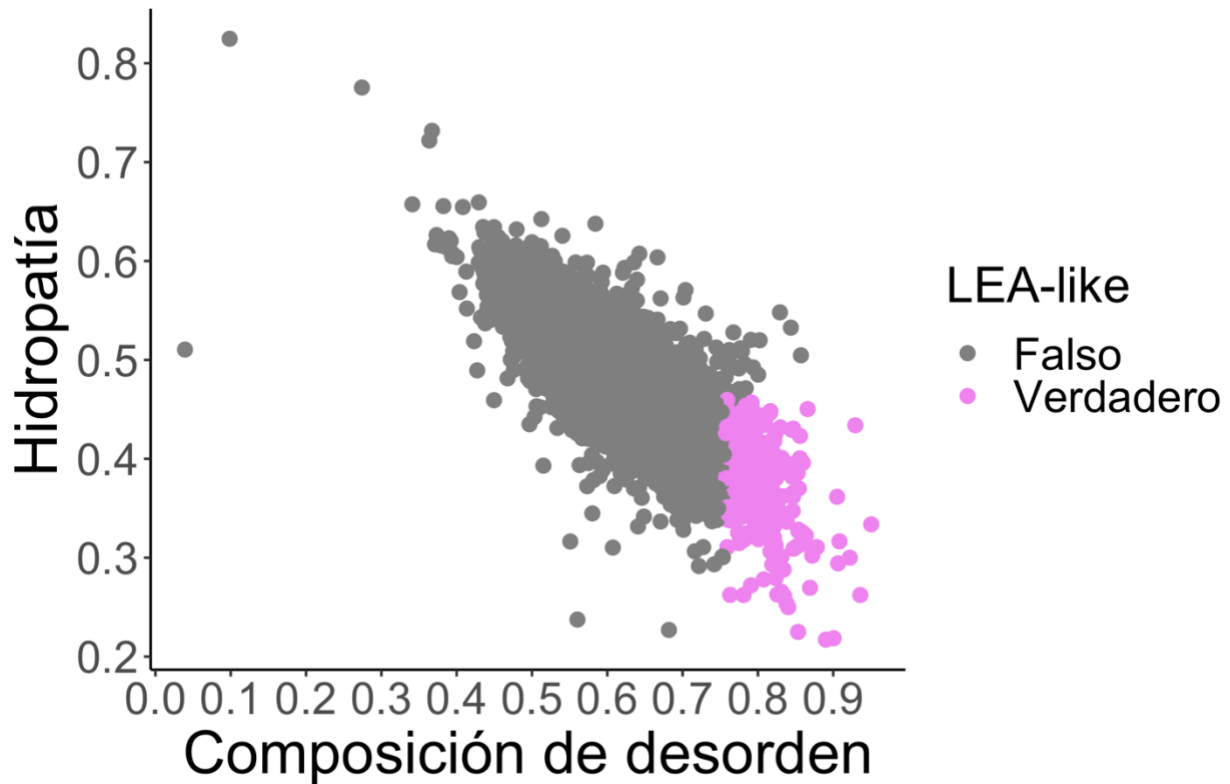


Figura 11. Hidropatía frente a composición de aminoácidos promotores del desorden estructural en *H. dujardini*. En color violeta se muestran las proteínas de tardígrado que caen dentro del intervalo que corresponde al primer corte mientras que en color gris se muestra el resto de las proteínas del proteoma de *H. dujardini*, las cuales no cumplen con dicho criterio.

Posteriormente, se procedió a realizar el segundo corte, es decir, carga neta media menor a 0.035. Así, se encontró que el tardígrado *H. dujardini* posee 190 proteínas que pasan los dos cortes, lo que corresponde al 1.3% de su proteoma (Figura 12). Debido a que las proteínas encontradas no se consideran proteínas LEA *sensu stricto* (porque no poseen motivos conservados de secuencia primaria) pero sí comparten sus propiedades fisicoquímicas, a este grupo se le llamó proteínas tipo-LEA.

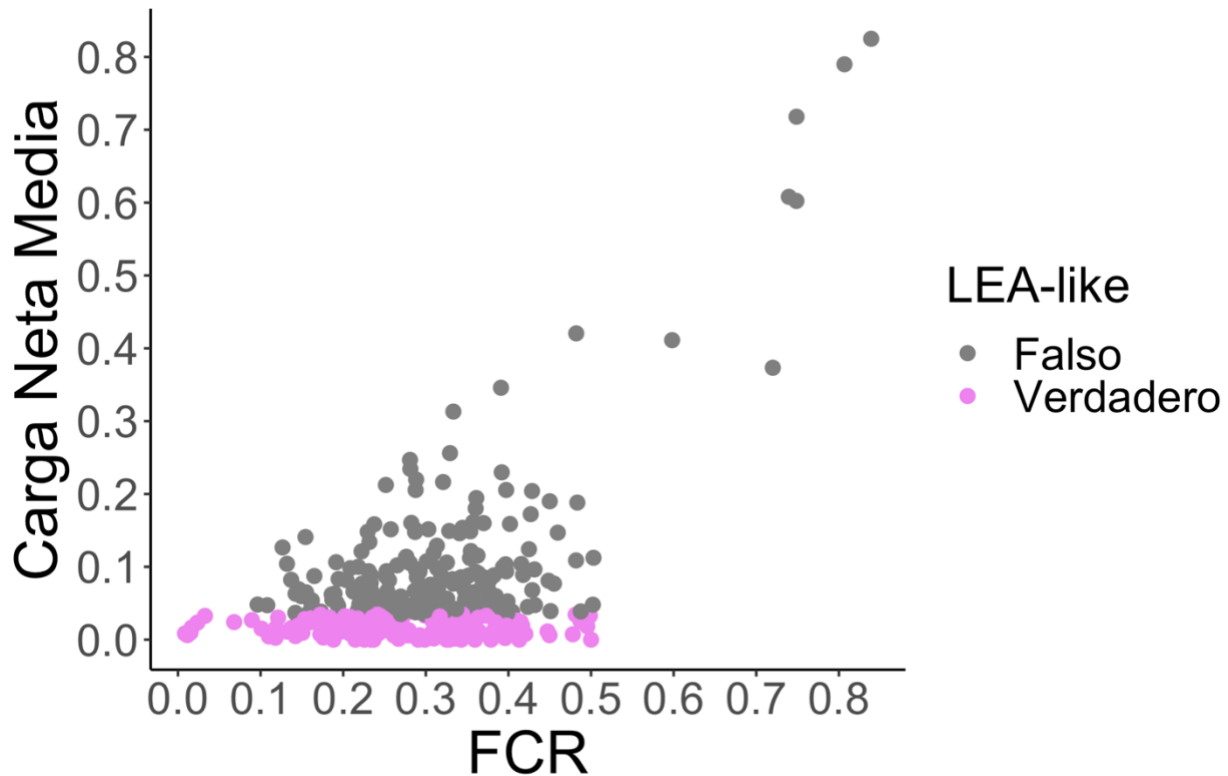


Figura 12. Carga Neta Media frente a FCR en *H. dujardini*. Los puntos mostrados representan aquellas proteínas que pasaron el primer corte; haciendo una distinción entre las proteínas que caen en el intervalo del segundo corte (color violeta) y aquellas que no (color gris).

Finalmente, con el objetivo de establecer una asociación funcional a las proteínas tipo-LEA, se realizó un análisis de ontología genética. En el caso de las plantas, se utilizó a *A. thaliana* debido a que es el organismo con la anotación de ontología genética más completa. Se realizó un análisis de enriquecimiento con *Gene Ontology* a tres niveles: componente celular, función molecular y proceso biológico. Debido a la esperada representación de funciones involucradas en respuesta al estrés, se presentan los resultados del análisis de proceso biológico. En este análisis se compararon las proteínas que no pasan ningún corte, las que solo pasan el primer corte y las que pasan tanto el primero como el segundo corte. Como se observa en la Figura 13, las proteínas que pasaron algún corte están asociadas a procesos biológicos de respuesta a estrés abiótico. En particular, se observó que las proteínas que pasaron los dos cortes (proteínas tipo-LEA) están enriquecidas en procesos de respuesta o aclimatación al frío y respuesta a la privación de agua. El p-valor para cada grupo de proteínas muestran que es probable que las proteínas que no pasaron ningún corte solo estén asociadas a esos procesos biológicos por azar (Tabla 2).

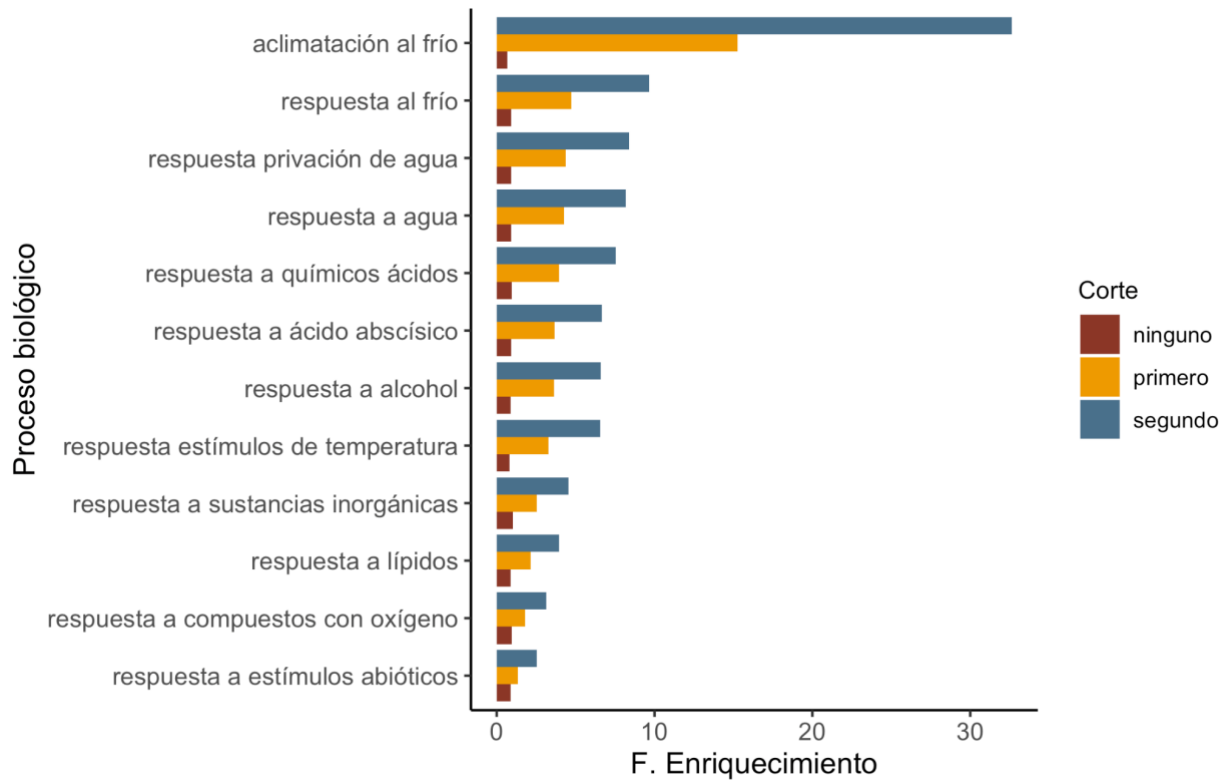


Figura 13. Análisis de enriquecimiento de los procesos biológicos en *A. thaliana*. La gráfica de barras muestra a que procesos biológicos se encuentran asociadas las proteínas del proteoma de *A. thaliana* donde el color vino representa a todas aquellas que no pasaron ningún corte, el amarillo a las que pasaron únicamente el primero y en azul a las que pasaron el primer y segundo corte.

Ya que en el servidor de Gene Ontology no cuenta con las anotaciones del proteoma de *H. dujardini*, sólo fue posible realizar un análisis parcial considerando las anotaciones de UniProt del 31.1 % de las proteínas LEA-like de tardígrado. Se observó que en general las proteínas tipo-LEA de *H. dujardini* están asociadas a las siguientes categorías: motivos de reconocimiento, ribonucleoproteínas, factores de *splicing*, factores ribosomales y proteínas de unión y transporte. Lo que indica que estas proteínas están enriquecidas en funciones relacionadas con procesos de transcripción y traducción.

Tabla 2. P-valores asociados a cada corte en *A.thaliana*

Proceso biológico	Código Gene Ontology	p-valor Ninguno	p-valor Primero	p-valor Segundo
respuesta al frío	GO:0009409	3.69E-01	2.31E-09	1.78E-11
aclimatación al frío	GO:0009631	1.65E-01	7.78E-10	3.60E-10
respuesta estímulos de temperatura	GO:0009266	9.83E-03	4.20E-07	1.30E-09
respuesta privación de agua	GO:0009414	4.27E-01	8.23E-08	8.05E-09
respuesta a agua	GO:0009415	3.59E-01	1.19E-07	1.06E-08
respuesta a químicos ácidos	GO:0001101	6.47E-01	4.11E-07	2.70E-08
respuesta a ácido abscísico	GO:0009737	2.20E-01	6.59E-07	1.07E-07
respuesta a alcohol	GO:0097305	1.93E-01	7.78E-07	1.21E-07
respuesta a sustancias inorgánicas	GO:0010035	7.33E-01	6.09E-05	2.92E-06
respuesta a compuestos con oxígeno	GO:1901700	3.30E-01	2.00E-03	9.78E-06
respuesta a lípidos	GO:0033993	4.26E-02	1.49E-03	3.05E-05
respuesta a estímulos abióticos	GO:0009628	2.57E-03	9.75E-02	5.11E-05

VIII. Discusión

Dado que la literatura sugería que las propiedades fisicoquímicas dadas por la secuencia de aminoácidos podían proporcionarles a las proteínas LEA de *A. thaliana* algunas funciones tolerar la desecación en semillas, se empleó *localCIDER* para obtener los valores de diversos parámetros procedentes de la secuencia primaria de las proteínas LEA y las proteínas del resto del proteoma. Los primeros resultados con *A. thaliana* (Figura 5) mostraron como al graficar hidropatía y desorden la mayor parte de las proteínas LEA formaban un grupo separado de la mayoría del proteoma, sugiriendo que esos valores eran característicos de las proteínas LEA. Lo anterior coincide con lo observado por Uversky, donde valores bajos de hidropatía aunados a valores altos de carga neta iban de la mano con la falta de una estructura conformacional compacta. (Uversky, 2002). Hundertmark y Hinch a corroboraron lo visto por Uversky ya que en su estudio emplearon los mismos parámetros y obtuvieron la misma tendencia; obtuvieron valores bajos de hidropatía media para la mayoría de las proteínas LEA de *A. thaliana* (Hundertmark y Hinch a, 2008). En este trabajo se confirmó lo anterior tras realizar el mismo análisis incorporando más organismos tolerantes a la desecación con proteínas LEA: *Oryza sativa* (arroz), *Triticum aestivum* (trigo), y *Zea mays* (maíz) (Figura 6).

Este comportamiento no era tan evidente con otros parámetros así que hidropatía y composición de aminoácidos promotores del desorden fueron los primeros en considerarse. Para ello creamos las gráficas de caja (Figuras 7 y 8) que, por medio de los cuartiles, permitían fácilmente determinar en qué intervalo se encontraban la mayoría de las proteínas LEA, en específico el 75 % de ellas. A la par, también se buscaba que al seleccionar ese porcentaje de proteínas LEA se eliminara la mayoría de las proteínas de los proteomas al no caer dentro del intervalo seleccionado. De este modo el grupo que quedara seleccionado no solo sería representativo del comportamiento de las proteínas LEA sino que englobaría a las proteínas del proteoma que tuvieran una mayor similitud fisicoquímica con las proteínas LEA. Así es como se estableció el primer corte de hidropatía y la composición de aminoácidos promotores del desorden. Tras aplicar ese filtro, se eliminó cerca del 96 % de las proteínas del conjunto de proteomas y se conservó alrededor del 66 % de las proteínas LEA de arroz, maíz, trigo y *A. thaliana*. Al aplicar un segundo corte, estableciendo un intervalo respecto a la carga neta media se separaron a las proteínas LEA del 99 % de las proteínas de los proteomas de plantas, obteniendo un grupo de proteínas denominado proteínas tipo-LEA.

El siguiente paso debía guiar hacia la determinación de una relación entre la función de dichas proteínas con la tolerancia a la desecación en aquellos organismos que las poseen. Por ello se realizó un análisis ontológico en *A. thaliana* (Figura 5). Únicamente seleccionamos a *Arabidopsis thaliana* para este análisis dado que ha sido ampliamente caracterizada y nos permitiría realizar el análisis de enriquecimiento de procesos biológicos. Los resultados muestran que tanto las proteínas que únicamente pasan el primer corte como aquellas que pasan ambos cortes (proteínas tipo-LEA) están enriquecidas en los procesos de respuesta a estrés abiótico en contraste con las proteínas que no pasan ningún corte. El obtener esta concordancia entre ambos grupos nos indica que los parámetros seleccionados para ambos cortes fueron adecuados ya que soportan la hipótesis planteada hasta ese punto. Las proteínas LEA y las proteínas con

propiedades fisicoquímicas similares a las proteínas LEA están relacionadas con la tolerancia a la desecación.

Para completar la hipótesis faltaba analizar al menos un organismo tolerante a la desecación que no poseyera proteínas LEA y por ende se aplicaron los cortes de hidropatía y composición de aminoácidos promotores del desorden (Figura 4.A), y carga neta media (Figura 4.B) al proteoma de *H. dujardini*. A partir de ellos se logró delimitar un grupo pequeño de proteínas con propiedades fisicoquímicas similares a las proteínas LEA: 2.8 % del proteoma para el primer corte y un 1.3 % del proteoma para ambos cortes. De las proteínas tipo-LEA encontradas en el tardígrado hay varias proteínas asociadas a RNA como: motivos de reconocimiento, ribonucleoproteínas, factores de *splicing*, factores ribosomales y proteínas de unión y transporte. Estas proteínas juegan un papel importante en la transcripción y traducción los cuales son dos procesos claves para que *H. dujardini* tolere la desecación (Boothby et. al, 2017). Como último paso podría realizarse a futuro un análisis ontológico a *H. dujardini* como se menciona en las perspectivas (Boothby et. al, 2017).

IX. Conclusiones

A través de análisis bioinformáticos se logró caracterizar diferentes propiedades fisicoquímicas codificadas en la secuencia primaria de las proteínas LEA de plantas. Se encontró una diferencia significativa entre su comportamiento y el comportamiento del resto del proteoma en términos de tres parámetros fisicoquímicos: hidropatía, composición de aminoácidos promotores del desorden y carga neta media. Al analizar el proteoma de *H. dujardini* bajo esos mismos tres parámetros fisicoquímicos se encontró que dentro del grupo de proteínas tipo-LEA se encuentran algunas proteínas asociadas a RNA.

Aunado a esto, el análisis ontológico que se realizó para *A. thaliana* mostró cómo las proteínas tipo-LEA están asociadas a la respuesta a condiciones de estrés abiótico, incluyendo respuesta a la privación de agua. Los resultados de esta tesis contribuyen a entender cómo es que las características fisicoquímicas mencionadas anteriormente están relacionadas con funciones moleculares asociadas a la respuesta a condiciones de estrés.

X. Perspectivas

Tras haber observado que algunas de las proteínas que tienen propiedades fisicoquímicas similares (bajo los criterios antes mencionados de hidropatía, composición de aminoácidos promotores del desorden, y carga neta media) están asociadas a la respuesta a la desecación; podría realizarse un análisis de los niveles de expresión de los transcritos que codifican para estas proteínas. Esto para confirmar su presencia en condiciones de desecación. De igual manera, resultaría valioso llevar a cabo un análisis ontológico de *H. dujardini* para evaluar si del mismo modo que las proteínas tipo-LEA de *A. thaliana* el grupo de proteínas del tardígrado que pasaron ambos cortes están asociadas a respuestas ante condiciones de estrés. A la par, podrían analizarse más organismos tolerantes a la desecación para determinar si esta relación entre propiedades fisicoquímicas y proceso biológico se mantiene. Finalmente, es necesario realizar ensayos *in vitro* y/o *in vivo* para determinar si las proteínas LEA y las proteínas tipo-LEA comparten funciones en respuesta a la desecación.

XI. Bibliografía

- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics*, 25(1), 25–29. <https://doi.org/10.1038/75556>
- Battaglia, M., Olvera-Carrillo, Y., Garcarrubio, A., Campos, F., & Covarrubias, A. A. (2008). The enigmatic LEA proteins and other hydrophilins. *Plant physiology*, 148(1), 6–24. <https://doi.org/10.1104/pp.108.120725>
- Boothby, T. C., Tapia, H., Brozena, A. H., Piszkiwicz, S., Smith, A. E., Giovannini, I., Rebecchi, L., Pielak, G. J., Koshland, D., & Goldstein, B. (2017). Tardigrades Use Intrinsically Disordered Proteins to Survive Desiccation. *Molecular cell*, 65(6), 975–984.e5. <https://doi.org/10.1016/j.molcel.2017.02.018>
- Caron, G. (2007). 5.18 - Lipophilicity, Polarity, and Hydrophobicity. In G. Ermondi & R. A. Scherrer (Eds.), *Comprehensive Medical Chemistry II* (Vol. 5, pp. 425–452). Elsevier. <https://doi.org/10.1016/B0-08-045044-X/00135-8>
- Cogliati, S., Frezza, C., Soriano, M. E., Varanita, T., Quintana-Cabrera, R., Corrado, M., Cipolat, S., Costa, V., Casarin, A., Gomes, L. C., Perales-Clemente, E., Salviati, L., Fernandez-Silva, P., Enriquez, J. A., & Scorrano, L. (2013). Mitochondrial cristae shape determines respiratory chain supercomplexes assembly and respiratory efficiency. *Cell*, 155(1), 160–171. <https://doi.org/10.1016/j.cell.2013.08.032>
- Covarrubias, A. A., Cuevas-Velazquez, C. L., Romero-Pérez, P. S., Rendón-Luna, D. F., & Chater, C. (2017). Structural disorder in plant proteins: where plasticity meets sessility. *Cellular and molecular life sciences : CMLS*, 74(17), 3119–3147. <https://doi.org/10.1007/s00018-017-2557-2>
- Crowe, J. H., Carpenter, J. F., & Crowe, L. M. (1998). The role of vitrification in anhydrobiosis. *Annual review of physiology*, 60, 73–103. <https://doi.org/10.1146/annurev.physiol.60.1.73>
- Cuevas-Velazquez CL, Saab-Rincon G, Reyes JL, Covarrubias AA (2016) The unstructured N-terminal region of Arabidopsis group 4 late embryogenesis abundant (LEA) proteins is required for folding and for chaperone-like activity under water deficit. *J Biol Chem* 291(20):10893–10903. doi:10.1074/jbc.M116. 720318
- Das, R. K., & Pappu, R. V. (2013). Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proceedings of the National Academy of Sciences of the United States of America*, 110(33), 13392–13397. <https://doi.org/10.1073/pnas.1304749110>
- Deak KI, Malamy J (2005) Osmotic regulation of root system architecture. *Plant J* 43(1):17–28
- Deeba, F., Pandey, V. (2017). Adaptive Mechanisms of Desiccation Tolerance in Resurrection Plants. In: Shukla, V., Kumar, S., Kumar, N. (eds) *Plant Adaptation Strategies in Changing Environment*. Springer, Singapore. https://doi.org/10.1007/978-981-10-6744-0_2
- Dinakar C, Djilianov D, Bartels D (2012) Photosynthesis in desiccation-tolerant plants: energy metabolism and antioxidative stress defense. *Plant Sci* 182:29–41. <https://doi.org/10.1016/j.Plantsci.2011.01.018>
- Gaff DF, Hallam ND (1974) Resurrecting desiccated plants. *Roy Soc NZ Bull* 12:389–393 Gagneul D, Ainouche A, Duhazé C, Lugan R, Larher FR, Bouchereau A (2007) A reassessment of the function of the so-called compatible solutes in the halophytic Plumbaginaceae *Limonium latifolium*. *Plant Physiol* 144(3):1598–1611

- Garay-Arroyo, A., Colmenero-Flores, J. M., Garcarrubio, A., & Covarrubias, A. A. (2000). Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. *The Journal of biological chemistry*, 275(8), 5668–5674.
- Hesgrove, C., & Boothby, T. C. (2020, November 4). The biology of tardigrade disordered proteins in extreme stress tolerance. *BioMed Central*. <https://biosignaling.biomedcentral.com/articles/10.1186/s12964-020-00670-2>
- Holehouse, A.S., Ahad, J., Das, R.K., and Pappu, R.V. (2015). CIDER: Classification of Intrinsically Disordered Ensemble Regions. *Biophys. J.* 108, 228a.
- Hundertmark, M., & Hinch, D. K. (2008). LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC genomics*, 9, 118. <https://doi.org/10.1186/1471-2164-9-118>
- Jones L, McQueen-Mason S (2004) A role for expansins in dehydration and rehydration of the resurrection plant *Craterostigma plantagineum*. *FEBS Lett* 559:61–65
- Kaliszan, R. (2015). Quantitative Structure–Retention Relationships (QSRRs) in Chromatography. In *Encyclopedia of Separation Science* (pp. 4063–4075). Elsevier. <https://doi.org/10.1016/B978-0-12-409547-2.11535-5>
- Lapinski, J., & Tunnacliffe, A. (2003). Anhydrobiosis without trehalose in bdelloid rotifers. *FEBS letters*, 553(3), 387–390. [https://doi.org/10.1016/s0014-5793\(03\)01062-7](https://doi.org/10.1016/s0014-5793(03)01062-7)
- Li, X., & Cao, J. (2015). *Late Embryogenesis Abundant (LEA) Gene Family in Maize: Identification, Evolution, and Expression Profiles*. *Plant Molecular Biology Reporter*, 34(1), 15–28. doi:10.1007/s11105-015-0901-y
- Liu, D., Sun, J., Zhu, D., Lyu, G., Zhang, C., Liu, J., ... Gao, D. (2019). *Genome-Wide Identification and Expression Profiles of Late Embryogenesis-Abundant (LEA) Genes during Grain Maturation in Wheat (Triticum aestivum L.)*. *Genes*, 10(9), 696. doi:10.3390/genes10090696
- Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E., Tosatto, S., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2021). Pfam: The protein families database in 2021. *Nucleic acids research*, 49(D1), D412–D419. <https://doi.org/10.1093/nar/gkaa913>
- Møbjerg, N., & Cardoso, R. (2021, January 11). New insights into survival strategies of tardigrades. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 254. <https://doi.org/10.1016/j.cbpa.2020.110890>
- Olvera-Carrillo Y, Luis Reyes J, Covarrubias AA (2011) Late embryogenesis abundant proteins: versatile players in the plant adaptation to water limiting environments. *Plant Signal Behav* 6(4):586–589. doi:10.4161/psb.6.4.15042
- Richaud, Myriam & Goff, Emilie & Cazevielle, Chantal & Ono, Fumihisa & Mori, Yoshihisa & Saini, Naurang & Cuq, Pierre & Baghdiguian, Stephen & Godefroy, Nelly & Galas, Simon. (2020). Ultrastructural analysis of the dehydrated tardigrade *Hypsibius exemplaris* unveils an anhydrobiotic-specific architecture. *Scientific Reports*. 10. 10.1038/s41598-020-61165-1.
- RStudio Team (2020). RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>
- Stoll M, Loveys B, Dry P (2000) Hormonal changes induced by partial rootzone drying of irrigated grapevine. *J Exp Bot* 51:1627–1634
- Tompa P, Szasz C, Buday L (2005) Structural disorder throws new light on moonlighting. *Trends Biochem Sci* 30(9):484–489. doi:10.1016/j.tibs.2005.07.008
- Tuba Z, Lichtenthaler HK, Csintalan ZS, Pócs T (1993b) Regreening of desiccated leaves of the poikilochlorophyllous *Xerophyta scabrida* upon rehydration. *J Plant Physiol* 142:103–108

- Tunnacliffe, A., Lapinski, J. & McGee, B. (2005) A Putative LEA Protein, but no Trehalose, is Present in Anhydrobiotic Bdelloid Rotifers. *Hydrobiologia* 546: 315–321 . <https://doi.org/10.1007/s10750-005-4239-6>
- UniProt Consortium (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic acids research*, 49(D1), D480–D489. <https://doi.org/10.1093/nar/gkaa1100>
- Uversky V. N. (2002). Natively unfolded proteins: a point where biology waits for physics. *Protein science: a publication of the Protein Society*, 11(4), 739–756. <https://doi.org/10.1110/ps.4210102>
- Wang, X.-S., Zhu, H.-B., Jin, G.-L., Liu, H.-L., Wu, W.-R., & Zhu, J. (2007). *Genome-scale identification and analysis of LEA genes in rice (Oryza sativa L.)*. *Plant Science*, 172(2), 414–420. doi:10.1016/j.plantsci.2006.10.
- Yamaguchi, A., Tanaka, S., Yamaguchi, S., Kuwahara, H., Takamura, C., Imajoh-Ohmi, S., Horikawa, D. D., Toyoda, A., Katayama, T., Arakawa, K., Fujiyama, A., Kubo, T., & Kunieda, T. (2012). Two novel heat-soluble protein families abundantly expressed in an anhydrobiotic tardigrade. *PloS one*, 7(8), e44209. <https://doi.org/10.1371/journal.pone.0044209>

Anexo

Das-Pappu Phase diagram position: El diagrama de fases considera a la fracción de residuos cargados negativamente (f_{men}) para el eje y, y a la fracción de residuos cargados positivamente (f_{mas}) para el eje x; la intersección de los valores de f_{men} y f_{mas} , de la secuencia de una proteína en el diagrama de fases corresponde al valor de FCR.

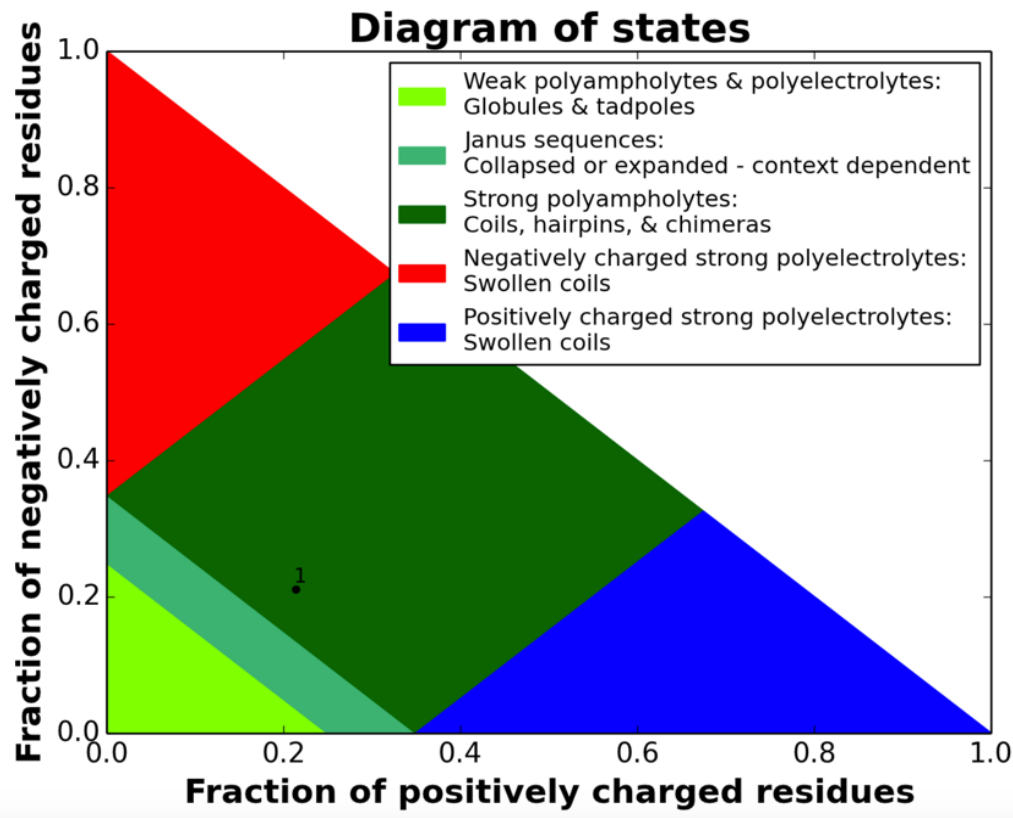


Figura 14. Diagrama de fases de *Das-Pappu*. En esta figura el área que se encuentra entre los valores máximos que puede tomar cada parámetro se encuentra clasificada de la siguiente manera: polianfolitos/polielectrolitos débiles (verde fosforescente) con conformación globular o de *tadpole*; secuencias Janus (verde menta) cuya conformación, colapsada o expandida, depende de las condiciones del ambiente a las que son expuestas; polianfolitos fuertes (verde oscuro) que pueden encontrarse como *hairpin*, *random coil* o *chimera* dependiendo de su valor de κ ; polielectrolitos fuertes cargados negativamente (rojo) con conformación de *swollen coil*; y finalmente polielectrolitos fuertes cargados positivamente (azul) con conformación de *swollen coil* (Das y Pappu 2013). ** Dado que la prolina suele formar hélices de poli-prolina, si su contenido es alto en una proteína, la predicción de fases pierde confiabilidad (Das y Pappu 2013).