



Centro de Investigación Científica de Yucatán, A.C.
Posgrado en Ciencias Biológicas

**EL DESORDENOMA DE PLANTAS Y SU RELACIÓN
CON LA RESPUESTA A LA FALTA DE AGUA**

Tesis que presenta
JESÚS ALEJANDRO ZAMORA BRISEÑO

En opción al título de
DOCTOR EN CIENCIAS
(Ciencias Biológicas: Opción Biotecnología)

Mérida, Yucatán, México

2020

CENTRO DE INVESTIGACIÓN CIENTÍFICA DE YUCATÁN, A. C.
POSGRADO EN CIENCIAS BIOLÓGICAS



RECONOCIMIENTO

Por medio de la presente hago constar que el trabajo de tesis de **Jesús Alejandro Zamora Briseño**, titulado “**El desordenoma de plantas y su relación con la respuesta a la falta de agua**”, fue realizado bajo la dirección del Dr. Luis Carlos Rodríguez Zapata en el Laboratorio de Fisiología Vegetal y Transformación Genética de Plantas de la Unidad de Biotecnología del Centro de Investigación Científica de Yucatán, A.C., dentro de la opción de Biotecnología, perteneciente al Programa de Doctorado del Posgrado en Ciencias Biológicas de este Centro.

Atentamente:



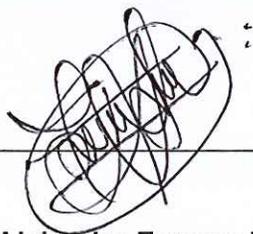
Dra. Cecilia Hernández Zepeda
Directora de Docencia

Mérida, Yucatán, México, a 9 de septiembre de 2020.

DECLARACIÓN DE PROPIEDAD

Declaro que la información contenida en la sección de Materiales y Métodos Experimentales, los Resultados y Discusión de este documento proviene de las actividades de experimentación realizadas durante el período que se me asignó para desarrollar mi trabajo de tesis, en las Unidades y Laboratorios del Centro de Investigación Científica de Yucatán, A.C., y que a razón de lo anterior y en contraprestación de los servicios educativos o de apoyo que me fueron brindados, dicha información, en términos de la Ley Federal del Derecho de Autor y la Ley de la Propiedad Industrial, le pertenece patrimonialmente a dicho Centro de Investigación. Por otra parte, en virtud de lo ya manifestado, reconozco que de igual manera los productos intelectuales o desarrollos tecnológicos que deriven o pudieran derivar de lo correspondiente a dicha información, le pertenecen patrimonialmente al Centro de Investigación Científica de Yucatán, A.C., y en el mismo tenor, reconozco que si derivaren de este trabajo productos intelectuales o desarrollos tecnológicos, en lo especial, estos se registrarán en todo caso por lo dispuesto por la Ley Federal del Derecho de Autor y la Ley de la Propiedad Industrial, en el tenor de lo expuesto en la presente Declaración.

Firma: _____



Jesús Alejandro Zamora Briseño

Este trabajo se llevó a cabo en la Unidad de Biotecnología del Centro de Investigación Científica de Yucatán, bajo la dirección del Dr. Luis Carlos Rodríguez Zapata.

AGRADECIMIENTOS

Al CONACYT, por los financiamientos de los proyectos que sirvieron para pagar las publicaciones, específicamente a los proyectos con número FC2016-1572, 221208 y S52089-Z.

Al CICY, por la infraestructura que me brindó durante mis estudios de doctorado.

Al Dr. Luis Carlos Rodríguez Zapata, por su amistad y por haberme dado la confianza de entrar a su laboratorio para hacer un trabajo en el que pocos creían.

A mi comité tutorial, el cual participó de una forma activa y expedita en mi formación, lo cual ayudó a enriquecer la calidad de este trabajo.

Un especial reconocimiento se merecen el Dr. Eliel Ruiz May y el Biól. José Miguel Elizalde Contreras por todo el apoyo que me prestaron para llevar a cabo los experimentos de proteómica, así como por su amistad. En este sentido el Dr. Alejandro Pereira Santana merece mi más merecido reconocimiento por todo su apoyo, sus ideas y también por su amistad.

Al QFB. Miguel Keb Llanes, por el apoyo técnico brindado durante el desarrollo de mi tesis en el Laboratorio de Fisiología Molecular y Transformación Genética de Plantas.

Quiero agradecer a todos los integrantes, amigos y compañeros del Laboratorio de Fisiología Molecular y Transformación Genética de Plantas, por todo el aprendizaje que recibí de ustedes, por su amistad y por el increíble equipo que pudimos conformar. Les deseo lo mejor en sus vidas.

DEDICATORIAS

Este trabajo se lo dedico a mis seres más amados: mis padres, hermanos y mis tres angelitos, Iori, Alejandrito y Rafaelito. Los amo.

ÍNDICE

CAPÍTULO I	1
ANTECEDENTES	1
1.1 Desorden estructural de proteínas como paradigma emergente.....	1
1.2 Características principales de las IDRs/IDPs.....	2
1.3 Respuestas celulares a condiciones que propician deshidratación en plantas.....	4
1.4 IDPs y estrés hídrico en plantas.....	6
JUSTIFICACIÓN	9
HIPÓTESIS	9
OBJETIVO GENERAL	9
OBJETIVOS PARTICULARES	9
ESTRATEGIA EXPERIMENTAL	9
CAPÍTULO II HACIA LA COMPRESIÓN DEL PAPEL DEL DESORDEN INTRÍNSECO EN LA ADAPTACIÓN DE LAS PLANTAS A DESAFÍOS AMBIENTALES	11
2.1 INTRODUCCIÓN.....	11
2.2 MATERIALES Y MÉTODOS.....	13
2.3 RESULTADOS.....	14
2.4 DISCUSIÓN.....	22
2.5 CONCLUSIÓN.....	26
CAPÍTULO III DINÁMICA GLOBAL DE LAS IDPs Y SU RELACIÓN CON LA DESECACIÓN EN SEMILLAS. LA CARIOPSIS DE MAÍZ COMO MODELO DE ESTUDIO	27
3.1 INTRODUCCIÓN.....	27
3.2 MATERIALES Y MÉTODOS.....	31
3.2.1 Adquisición de datos transcriptómicos.....	31
3.2.2 Análisis de desorden proteico y agrupamiento de datos.....	31
3.2.3 Minería de datos transcriptómicos y análisis de expresión diferencial.....	32
3.2.4 Análisis de redes de co-expresión.....	33
3.3 RESULTADOS.....	34

3.4 DISCUSIÓN.....	43
3.5 CONCLUSIÓN.....	52
CAPÍTULO IV CARACTERIZACIÓN DEL PROTEOMA TERMO-ÁCIDO-SOLUBLE DE HOJAS DE CAÑA DE AZÚCAR <i>Saccharum officinarum</i> EN RESPUESTA AL ESTRÉS POR FALTA DE AGUA.....	53
4.1 INTRODUCCIÓN.....	53
4.2 MATERIALES Y MÉTODOS.....	55
4.2.1 Condiciones de crecimiento de las plantas.....	55
4.2.2 Tratamiento de estrés, determinación de parámetros fotosintéticos y obtención de proteínas termo-ácido-solubles.....	56
4.2.3 Digestión enzimática de proteínas.....	57
4.2.4 Análisis por nano-LC-MS / MS.....	58
4.2.5 Análisis bioinformático.....	58
4.3 RESULTADOS.....	60
4.4 DISCUSIÓN.....	69
4.4.1 Características generales de las proteínas identificadas.....	69
4.4.2 Proteínas diferenciales identificadas y su implicación en la respuesta a la falta de agua.....	72
4.5 CONCLUSIÓN.....	74
CAPÍTULO V CONCLUSIÓN GENERAL Y PERSPECTIVAS.....	75
5.1 CONCLUSIÓN GENERAL.....	75
5.2 PERSPECTIVAS.....	79
BIBLIOGRAFÍA.....	80

ÍNDICE DE FIGURAS

Capítulo II

Figura 2.1 Distribución de proteínas desordenadas entre los proteomas de monocotiledóneas, eudicotiledóneas y algas.....	15
Figura 2.2 Comparación entre la abundancia relativa de proteínas de acuerdo con su contenido de desorden relativo de desorden.....	16
Figura 2.3 Relación entre el nivel de desorden con el tamaño de las proteínas.....	17
Figura 2.4 Mapa de calor de la composición porcentual de los aminoácidos de acuerdo con el nivel de desorden intrínseco de las proteínas a lo largo de los diferentes clados.....	18
Figura 2.5 Enriquecimiento ontológico de las proteínas entre las diferentes categorías de desorden.....	19
Figura 2.6 Enriquecimiento de procesos biológicos para las proteínas de <i>A. thaliana</i> de acuerdo con su nivel de desorden.....	20
Figura 2.7 Análisis de enriquecimiento de componentes celulares para las proteínas de <i>A. thaliana</i> de acuerdo con su nivel de desorden.....	21

Capítulo III

Figura 3.1 Relación entre el nivel de desorden proteico y la longitud de las proteínas.....	34
Figura 3.2 Características generales de las proteínas según el contenido del desorden proteico.....	35
Figura 3.3 Diagramas de Venn de los genes expresados (TPM > 0) y compartidos de cada etapa, de acuerdo con el contenido de desorden de las proteínas codificadas	36
Figura 3.4 Patrones de expresión de transcritos de acuerdo con el contenido de desorden	37

Figura 3.5 Dinámica de la composición relativa de los transcritos expresados a lo largo del desarrollo de semillas en función del porcentaje de desorden estructural de las proteínas para las que codifican.....	39
Figura 3.6 Perfil de expresión génica de transcritos que codifican para IDPs con más del 75% de desorden en su secuencia.....	40
Figura 3.7 Análisis de redes de co-expresión utilizando como reguladores los TFs regulados de manera positiva.....	43
Figura 3.8 Modelo del ajuste del desordenoma durante el desarrollo de semillas ortodoxas.....	52

Capítulo IV

Figura 4.1 Evaluación fisiológica de las plantas de caña empleadas en este estudio y distribución de las proteínas termo-ácido-solubles identificadas.....	61
Figura 4.2 Distribución de las proteínas identificadas de acuerdo con su contenido porcentual de desorden estructural en su secuencia.....	63
Figura 4.3 Comparación de aminoácidos de las proteínas termo-ácido-solubles identificadas y las proteínas del PDB.....	64
Figura 4.4 Gráfica de los dominios pfam identificados más abundantes de entre las proteínas identificadas.	65
Figura 4.5 Distribución de los términos ontológicos de las proteínas termo-ácido-solubles identificadas.....	66
Figura 4.6 Anotación general de las proteínas termo-ácido-solubles identificadas.....	67

ÍNDICE DE CUADROS

Capítulo IV

Cuadro 4.1 Estadísticas generales de desorden de las proteínas termo-ácido-solubles identificadas.....	62
---	----

RESUMEN

Las proteínas intrínsecamente desordenadas (IDPs) son proteínas funcionales que no tienen una estructura tridimensional definida. En este trabajo se plantean varias razones del porqué en las plantas el conjunto de IDPs en el proteoma (desordenoma) podría estar sujeto a una regulación activa durante condiciones de escasez de agua. Primero, se estimó el contenido relativo de IDPs en proteomas de algas, monocotiledóneas y eudicotiledóneas. Se encontró una variación del contenido de IDPs entre los clados principales, estando las monocotiledóneas particularmente enriquecidas en este tipo de proteínas. También se halló una relación inversa entre el nivel de desorden de las IDPs y su longitud, siendo las IDPs más pequeñas que las proteínas ordenadas. Además, la abundancia relativa de aminoácidos depende del desorden intrínseco, pero también varía entre clados. Las IDPs son más abundantes en el núcleo que las proteínas ordenadas. También están especializadas en funciones reguladoras, de unión de DNA y RNA y en respuesta a estímulos ambientales. Luego, para correlacionar la modulación del contenido de las IDPs en un escenario de tolerancia a la desecación, se analizó el perfil del transcriptoma durante el desarrollo de la semilla de maíz de acuerdo con el contenido de desorden de sus proteínas codificadas. Se encontró que el perfil de expresión total de los transcritos que codifican proteínas estructuradas aumenta durante la fase media de desarrollo de la semilla, mientras que el contenido relativo de desorden proteico aumenta conforme avanza su desarrollo. Al final de la fase tardía, se identificó un grupo de genes sobre-regulados que codifica IDPs y que coinciden con el comienzo de la adquisición de la tolerancia a la desecación. Finalmente, se utilizó un enfoque proteómico para estudiar las proteínas termoestables inducidas por la sequía de las hojas de caña de azúcar. Se observó una inducción generalizada de IDPs involucradas en la respuesta a la deshidratación, mientras que su número fue casi tres veces mayor que en condiciones de control. En conclusión, se considera que hay evidencia suficiente como para proponer que el desorden estructural juega un papel importante en la adaptación de las plantas a las condiciones de falta de agua.

ABSTRACT

Intrinsically disordered proteins (IDPs) are proteins without a rigid three-dimensional structure; they are emerging multifunctional cellular components of proteomes. Here, we disclose several reasons why the collection of intrinsically disordered proteins in a proteome (disordome) could be subjected to an active regulation during conditions of water scarcity in plants. First, it was estimated the relative content of intrinsic protein disorder in plant proteomes (*sensu lato*) including algae, monocots, and eudicots. A variation in the content of intrinsic protein disorder among these major clades was found. In turn, there is an inverse relationship between intrinsic protein disorder level and protein length, being higher disordered smaller proteins. Relative abundance of amino acids depends on intrinsic disorder, but also varies among clades. Moreover, IDPs have a greater abundance of proteins located in nucleus than ordered proteins. They are also specialized in regulatory functions, nucleic acid binding, RNA processing, and in response to environmental stimuli. Then, to correlate the modulation of the content of IDPs with the desiccation tolerance acquisition, the transcriptome profile during maize seed development was analyzed according to the disorder content of their encoding proteins. There is a sharply increase in the total expression profile of transcripts encoding for structured proteins during the middle phase, while the relative content of protein disorder increases as development progresses. A gene cluster encoding for IDPs at the end of the late phase was identified, which coincides with the beginning of the acquisition of desiccation tolerance stage. Finally, a proteomic approach was used to study the drought-induced boiling-stable proteins of sugarcane leaves. We observed an induction of IDPs, while its number was almost three times higher in drought stress than in control conditions. In conclusion, there is sufficient evidence to propose that the disordome play a major role in the plant adaptation to water limiting conditions.

CAPÍTULO I

ANTECEDENTES

1.1 Desorden estructural de proteínas como paradigma emergente

El paradigma de estructura-función de proteínas es para la bioquímica lo que el dogma central es para la biología molecular. Este paradigma sugiere que la función de una proteína está dada por su estructura. Sin embargo, esto no siempre se cumple. De hecho, con base en una gran cantidad de información generada a partir de análisis bioinformáticos y experimentales, a la fecha se conoce un sinnúmero de regiones proteicas e incluso proteínas enteras que carecen de una estructura tridimensional definida en su forma nativa cuando están en solución (Uversky y Dunker, 2010). Estas regiones/proteínas han sido denominadas proteínas intrínsecamente desordenadas (IDPs, por sus siglas en inglés) cuando de lo que se trata es de llamar a la cadena polipeptídica entera y regiones intrínsecamente desordenadas (IDRs, por sus siglas en inglés) cuando son parte de una proteína estructurada. Las IDPs/IDRs no pertenecen a una familia proteica en particular, sino que más bien es una característica ampliamente distribuida entre las diferentes proteínas del proteoma. Por ello, para un organismo dado, a la colección IDPs/IDRs se le ha denominado desordenoma.

En años recientes, se han resaltado algunas características emergentes que presentan las IDPs/IDRs, tales como su flexibilidad estructural y su hidrofiliidad. Ambas características son consecuencia de su composición, pues poseen una gran proporción de residuos cargados y polares y de prolina (Williams et al., 2001). Al ser esta una característica fisicoquímica compartida por una gran cantidad de proteínas, el desordenoma resulta ser un conglomerado de proteínas que juegan diferentes roles en la célula. Siguiendo esta idea, es posible que existan funciones particulares que se encuentren enriquecidas en IDPs y tales proteínas podrían estar estrechamente reguladas. Si esto es correcto y existe, por ejemplo, una porción del desordenoma implicado en la respuesta a estrés, probablemente exista una presión selectiva en tal porción del desordenoma que actúe a su vez sobre cada una de las proteínas que lo conforman, independientemente de si cada proteína está o no relacionada filogenéticamente entre sí. De existir, a esta porción del desordenoma se le podría llamar el desordenoma del estrés. Para darle soporte a esta consideración, sólo hace

falta observar algunos ejemplos de bacterias y arqueas que típicamente poseen una proporción relativa mayor de IDPs que sus contrapartes mesofílicas y/o sensibles a sal y pH (Xue et al., 2010). Sin embargo, en plantas hace falta información sobre la regulación que sufre el desordenoma en respuesta a diferentes tipos de estreses, en el que además se involucren variables como el tipo de tejido o el estado de desarrollo. Esta es una consideración pertinente en plantas, debido a que su naturaleza sésil les impide escapar de condiciones ambientales estresantes que pueden comprometer la disponibilidad de agua (Tompa y Kovacs, 2010). Es por esto mismo que las plantas han tenido que desarrollar mecanismos que les permitan adaptarse a las exigencias de su entorno.

1.2 Características principales de las IDRs/IDPs

En términos estructurales, las IDRs/IDPs se caracterizan por su alta heterogeneidad de conformaciones espacio-temporales, pues son proteínas que en vez de estar plegadas en un número discreto de estructuras tridimensionales, con unas coordenadas atómicas relativamente fijas y una baja dinámica conformacional como ocurre en las proteínas globulares, estas son más bien ensamblajes estructurales rápidamente interconvertibles de conformaciones alternativas (Kovacs *et al.*, 2008; Uversky y Dunker, 2010; Mittal *et al.*, 2013). Las IDPs/IDRs fueron descubiertas por primera vez como regiones no resueltas en estructuras de proteínas determinadas por difracción de rayos X o como proteínas o regiones sin restricciones suficientes como para definir una estructura fija en los espectros de NMR. Ahora se sabe que dichas observaciones son indicativas de regiones altamente móviles (Pazos *et al.*, 2013).

Las IDPs/IDRs presentan baja complejidad de secuencia y muestran cargas netas altas a pH neutro como consecuencia de la presencia de muchos residuos cargados no interactuantes y pocos residuos hidrofóbicos (Uversky, 2002). Esto resulta en una alta flexibilidad intramolecular y una baja o inexistente estructura terciaria. A pesar de esto, las IDPs/IDRs desempeñan roles biológicamente importantes (Stender *et al.*, 2015). Además, estas características les proveen de una inusual estabilidad a altas temperaturas y solubilidad en condiciones ácidas (Ismail, Hall y Close, 1999; Cortese *et al.*, 2005; Galea et al., 2006; Zamora-Briseño y de Jiménez, 2016).

Estructuralmente, las IDPs pueden ser definidas con base a dos diferentes criterios: si contienen al menos un segmento desordenado mayor de 30 aminoácidos o si al menos el 30% de sus residuos son desordenados (Liu *et al.*, 2006; Deiana, 2012). Sin embargo, la última definición es más universal, debido a que la primera, dependiendo de la longitud de la proteína, pueden hacer referencia sólo a proteínas globulares que contienen IDRs. Por otra parte, el valor funcional que poseen las IDPs/IDRs en un contexto celular emerge en parte de la predicción que indica que entre un 30 a un 40% de los proteomas eucariontes son desordenados (Xue *et al.*, 2012). De hecho, el desorden estructural es un fenómeno universal entre los dominios biológicos y está relacionado con la complejidad, tanto a nivel orgánico como a nivel de procesos celulares, existiendo una correlación directa entre el número de IDPs y el número de proteínas en el genoma involucrado, incrementando su contenido desde bacterias a hongos a organismos eucariontes. Esto a su vez, refleja la importancia de considerar el desorden desde una perspectiva evolutiva (Deiana y Giansanti, 2008; Xue *et al.*, 2012; Lobanov y Galzitskaya, 2015; Liu *et al.*, 2017).

Funcionalmente, las IDPs están implicadas en la regulación de procesos celulares, como por ejemplo la transcripción, la traducción, la transducción de señales o el ciclo celular, (Romero *et al.*, 2001; Tompa *et al.*, 2005; Liu *et al.*, 2006; Uversky, 2011). Por otra parte, cabe destacar que los sitios en los que ocurren varios tipos de modificaciones post-traduccionales están desproporcionadamente localizados entre IDRs (Iakoucheva *et al.*, 2004; Kurotani *et al.*, 2014). Esto hace sentido, dado el requerimiento de contactos cercanos entre la proteína siendo modificada y la enzima encargada de hacer la modificación postraduccionales. Es decir, cuando el lado de la cadena bajo modificación se encuentra entre una IDR, el desorden estructural facilita la unión al sustrato al permitir que la región desordenada se pliegue de una forma tal que permita el trabajo de la enzima modificadora (Uversky, 2013). Esto explica por qué aproximadamente el 70% de las proteínas de señalización en eucariotas se predicen que albergan IDRs o son IDPs (Iakoucheva *et al.*, 2002). A su vez, el hecho de que sean blancos de modificaciones postraduccionales tiene implicaciones sobre diferentes mecanismos regulatorios, tales como la localización subcelular y las interacciones con ligandos, la regulación de la estabilidad proteica y su tasa de recambio (Seet *et al.*, 2006).

Ciertas IDPs son capaces de adoptar diferentes estructuras bajo determinadas condiciones o al momento de entrar en contacto con sus interactores, permitiendo interacciones versátiles con varios blancos mediante interacciones rápidas, transitorias y débiles (Gallagher *et al.*, 2006; Kovacs *et al.*, 2008). Esto permite el re-uso de la misma proteína en múltiples rutas, facilitando la ganancia de información de manera dinámica sin la necesidad de la inversión adicional de energía, al permitir que una proteína realice múltiples funciones. Aunque estas son sólo algunas consideraciones generales de las funciones que desempeña las IDPs/IDRs, muchas otras han sido extensamente discutidas en detalle por otros autores (van der Lee *et al.*, 2014). Sin embargo, las funciones enumeradas aquí son suficientes para exponer el amplio abanico de funciones que están relacionadas con el desorden estructural. También hay que destacar que la evidencia de las funciones de la mayoría de las IDRs/IDPs han sido obtenidas principalmente de modelos no vegetales, por lo que, en comparación a los animales, la información disponible para las IDPs de plantas es mucho menor (Sun *et al.*, 2011; Marin *et al.*, 2013; Liu *et al.*, 2017; Álvarez-Ponce *et al.*, 2018).

1.3 Respuestas celulares a condiciones que propician deshidratación en plantas

La naturaleza sésil de las plantas ha propiciado la selección de una variedad de mecanismos que les permiten tolerar diferentes desafíos ambientales, tanto de naturaleza biótica como abiótica, ante los que están expuestas a lo largo de su vida. El estrés afecta adversamente el desempeño fisiológico de las plantas y los procesos involucrados en su desarrollo y reproducción, siendo el estrés por falta de agua es uno de los que afecta de manera más marcada las funciones de la planta (Andjelkovic y Thompson, 2006).

Dentro de las estrategias seleccionadas en las plantas para contender con el estrés por falta de agua, está la maduración temprana antes del inicio de periodos de sequía, la reproducción sólo en condiciones favorables, el desarrollo de un sistema radicular para mejorar la absorción de agua, la reducción de la transpiración, el almacenamiento en tejidos carnosos y/o simplemente enfrentando activamente la deshidratación del tejido mediante la activación de mecanismos de resistencia al estrés (van Bavel, 1996). Esta última estrategia puede ser considerada como la auténtica estrategia de tolerancia al estrés, en la cual los

tejidos vegetativos son capaces de sobrevivir varios niveles de falta de agua sin sufrir daños irreparables (Simpson y Niklas, 2006). Sin embargo, esto usualmente ocurre a expensas del desarrollo vegetal, con un impacto negativo en la ganancia de biomasa, debido a que la energía tiene que ser redirigida para el mantenimiento de las funciones celulares mínimas.

Aunque las plantas pueden presentar más de un mecanismo de respuesta a la falta de agua y poseen diferentes niveles de tolerancia, es posible distinguir dos fenómenos de acuerdo con el nivel de deshidratación que la planta es capaz de resistir: la tolerancia a la sequía y la tolerancia a la desecación. La tolerancia a la sequía se puede definir como la tolerancia a la deshidratación moderada, esto es, a una disminución del contenido de agua disponible por debajo del cual no hay agua citoplásmica almacenada, esto es equivalente a un ~23% de agua con base en peso fresco o bien, $\sim 0.3 \text{ g de H}_2\text{O} \cdot \text{g de peso seco}^{-1}$ (Hoekstra *et al.*, 2001). A su vez, se pueden presentar diferentes niveles de tolerancia a la sequía, dependiendo del tipo de planta, órgano y tejido. Por ejemplo, la mayoría de las angiospermas son sensibles a la deshidratación y contienen porcentajes relativos de agua de entre el 85 y el 100% (Mitra *et al.*, 2013). De esta forma, si el contenido relativo de agua cae por debajo de 59 a 30%, la mayoría de estas plantas son incapaces de sobrevivir (Dinakar y Bartels, 2013). Por otra parte, la tolerancia a la desecación se define como la habilidad para sobrevivir a niveles de sequía por debajo de un contenido de agua absoluto de $\sim 0.1 \text{ g de H}_2\text{O} \cdot \text{g de peso seco}^{-1}$, equivalente a una humedad relativa del aire del 50% a 20°C y $\leq -100 \text{ MPa}$ de potencial de agua (Moore *et al.*, 2008). Diversas plantas y órganos especializados, tales como helechos, musgos, polen y semillas ortodoxas, poseen esta capacidad. Quizá el ejemplo más sobresaliente que existe sean las plantas de la resurrección, las cuales presentan niveles de tolerancia a la desecación en su tejido vegetativo de hasta el 10% de contenido relativo de agua o menos (Mitra *et al.*, 2013).

Aunado a la falta de agua *per se*, las plantas también deben defenderse de otros procesos secundarios concomitantes, incluido el estrés oxidante, la desestabilización de membranas y la pérdida de integridad macromolecular (Illing *et al.*, 2005). Esto último es un aspecto muy importante para la función de las proteínas ya que puede comprometer sus funciones. Esto se debe a que la presencia de agua es esencial para el mantenimiento de su estructura, de forma que la pérdida de agua induce el desplegamiento y la agregación

proteica, como consecuencia del colapso hidrofóbico (Chakrabortee *et al.*, 2007). En este sentido, la falta de agua implica que un gran número de proteínas y otras macromoléculas estén presentes en un espacio celular compacto (efecto conocido en inglés como macromolecular crowding). Dicho efecto compromete la función de las proteínas por medio del decremento en el efecto hidrofóbico cuando hay menos agua presente. En algunos casos, la densidad macromolecular debe ser ajustada dramáticamente, por lo que diversos mecanismos deben encenderse para poder mantener la viabilidad celular. Una vía por la cual las plantas realizan esto es mediante la acumulación de solutos compatibles que promueven la estabilización de componentes críticos de la célula, entre los que se incluyen osmolitos compatibles, tales como prolina, sacarosa, polioles, trehalosa, glicina betaína, alanina betaína, prolina betaína o pipercolato betaína (Rhodes y Hanson, 2003), pero también enzimas protectoras y chaperonas, así como enzimas detoxificadoras. Esta situación es aún más dramática en condiciones de desecación, como es el caso de las semillas ortodoxas quiescentes o en el tejido vegetativo de plantas de la resurrección en estado de desecación. De hecho, las semillas ortodoxas son un ejemplo típico de modelos que ejemplifican el encendido de los mecanismos de resistencia a la desecación en plantas. Un ejemplo de proteínas que juegan un rol en la protección en plantas durante la falta de agua son las proteínas abundantes de la embriogénesis tardía (LEA, por sus siglas en inglés). Estas proteínas se acumulan en altos niveles durante la última etapa de formación de la semilla, cuando se da lugar la desecación de los tejidos. En estas condiciones, las proteínas LEA pueden llegar a representar hasta un 4% de las proteínas celulares en estos tejidos. Además, muchas de ellas se inducen de manera abundante durante los periodos de déficit de agua en tejidos vegetativos y representan el mayor grupo de IDPs implicadas en múltiples estreses, incluido el estrés por falta de agua (Garay-Arroyo *et al.*, 2000; Roberts *et al.*, 2007; Amara *et al.*, 2012).

1.4 IDPs y estrés hídrico en plantas

Hasta la fecha, hay relativamente pocos los ejemplos de IDPs caracterizadas que estén involucradas en estrés abiótico y por mucho, las proteínas LEA son las más estudiadas. Recientemente un subgrupo de proteínas LEA, el de las proteínas LEA4, ha sido propuesto como modelo para el estudio de IDPs en plantas (Covarrubias *et al.*, 2017). Sin embargo, incluso para las proteínas LEA, sólo una fracción de estas se han investigado en detalle, lo que hace que sea complicado hacer comparaciones funcionales (Bremer *et al.*, 2017). Para

estas proteínas, se han propuesto diversos mecanismos de acción. Por ejemplo, se ha encontrado que ciertos miembros de las proteínas LEA pueden secuestrar agua y azúcares y formar una red intrincada de biomoléculas vía interacciones por puentes de hidrógeno, lo que a su vez permite la formación de geles estables. Otras funciones asociadas a las proteínas LEA incluyen la de remover especies reactivas de oxígeno, proteger membranas, unir metales o poseer actividades antioxidantes (Hundertmark *et al.*, 2011; Liu *et al.*, 2013; Cuevas-Velázquez *et al.*, 2016; Bremer *et al.*, 2017).

Otro ejemplo de proteínas que contienen IDRs, son los factores de transcripción. Se estima que entre el 82 y el 94% de los factores de transcripción poseen extensas IDRs (Liu *et al.*, 2006). La evidencia sugiere que las IDRs de las familias de factores de transcripción NAC, bZIP y GRAS, participan en la interacción con ligandos (Thieulin-Pardo *et al.*, 2015). Este es el caso del Radical Induced Cell Death1 (RCD1), una proteína clave como organizador de la señalización en respuesta a diversos tipos de estrés, la cual actúa uniéndose a las IDRs de los factores de transcripción, incluido el de miembros de la familia NAC involucrados en las respuestas a estrés abiótico (Sun *et al.*, 2011; Kragelund *et al.*, 2012).

La falta de una estructura tridimensional definida predispone a las IDPs/IDRs a poseer capacidades que las proteínas globulares no tienen, entre las que destaca la actividad antiagregación. Esta resulta de la interferencia física por parte de ciertas IDPs y que reduce la frecuencia de encuentros indeseables entre proteínas (Goya *et al.*, 2005; Li *et al.*, 2006; Chakrabortee *et al.*, 2010; Toxopeus *et al.*, 2014; Lobanov y Galzitskaya, 2015; Cuevas-Velázquez *et al.*, 2016; Liu *et al.*, 2017). Consistente con esta observación, es el hecho de que las IDRs constituyen alrededor del 50% de las cadenas de las chaperonas de RNA y alrededor de 30% de las chaperonas de proteínas (Tompa y Csermely, 2004). Por esto es probable que además de servir como mecanismo para modificar rápidamente las redes de interacción y responder a desafíos ambientales, las IDPs/IDRs podrían ayudar a que el proteoma de las plantas sea más resiliente a la falta de agua y sus consecuencias a nivel molecular, pues muchas IDRs son multiespecíficas (Tompa y Kovacs, 2010; Tunnacliffe *et al.*, 2010; Kragelund *et al.*, 2012). Esta capacidad multifuncional puede permitir que muchas de ellas puedan actuar como organizadores en procesos de señalización. La regulación del desordenoma puede ser una forma dinámica para contrastar los efectos negativos del

estrés y ser un elemento adicional de los mecanismos de protección típicos como lo son la inducción de metabolitos compatibles. La idea de tener proteomas ricos en IDPs que sean más estables a la falta de agua que los proteomas en condiciones normales parece factible y varias especies cuyos proteomas abundantes en IDPs (e.g. bacteria esporuladoras, bacterias y archaea halófilas) son tolerantes a la desecación (Kriško *et al.*, 2010). En las semillas quiescentes, ocurre una acumulación de proteínas LEA y otras proteínas hidrofílicas termoestables (Amara *et al.*, 2012). En este sentido, una idea es que los ajustes a nivel del desordenoma pueden ocurrir en células estresadas para adaptar sus funciones a ese nuevo escenario celular, esto no sólo con base a un ajuste energético o funcional, sino además tomando en consideración la capacidad de las proteínas para trabajar bajo condiciones que favorecen la agregación molecular. Esto podría ser importante en un contexto de estrés por falta de agua, pues podría representar un ahorro energético en un momento en que el gasto energético debe mantenerse tan bajo como sea posible.

En 2013, Pietrosemoli *et al.* llevaron a cabo el primer análisis a escala global del repertorio de IDPs/IDRs en *Arabidopsis thaliana* y encontraron varios procesos enriquecidos en IDPs, incluidas las respuestas a estímulos abióticos. Hay que considerar que las inferencias basadas en datos genómicos no toman en cuenta la plasticidad dinámica de la regulación de la maquinaria celular, tanto la que se da a nivel de tejido como la que ocurre en respuestas a cambios ambientales. Por ejemplo, bajo esta aproximación, una carióspside quiescente de maíz tendría el mismo contenido de IDPs/IDRs que cada una de las estructuras vegetativas de la planta, esto por compartir la misma información genética. En la realidad, esto claramente no funciona así. Por ello, las inferencias bioinformáticas obtenidas a partir de este enfoque deben ser asumidas con cautela. Por otra parte, hay evidencias que sugieren que el desordenoma de plantas responde estímulos ambientales. Por ejemplo, una estimación experimental sobre la interacción entre *Colletotrichum acutatum* y fresa mostró que el proteoma activado durante la propagación del hongo en las fresas susceptibles presenta una menor cantidad de IDRs/IDPs comparado con sus contrapartes resistentes a la infección. Esto sugiere que las IDPs están involucradas en la respuesta a patógenos (Baraldi *et al.*, 2015), lo que concuerda con el hecho de que las IDRs de proteínas efectoras son clave en las interacciones planta-patógeno (Marin *et al.*, 2013). Aunque no en plantas, recientemente se ha corroborado experimentalmente que las IDPs

participan en la resistencia de los tardígrados a la desecación extrema (Boothby y Pielak, 2017). En plantas, hay datos que sugieren que varias IDPs/IDRs se asocian inversamente con el contenido de agua en plantas de maíz y de soya (Amara *et al.*, 2012; Liu *et al.*, 2017).

JUSTIFICACIÓN

Debido a la falta de información relacionada con la forma en cómo el desordenoma se regula en las plantas ante la falta de agua y dado que el conjunto de proteínas que lo componen está comparativamente menos caracterizado que las proteínas estructuradas, es posible que su análisis permita identificar proteínas asociadas con el estrés que no han sido previamente caracterizadas.

HIPÓTESIS

En las plantas, el aumento en el contenido relativo de proteínas intrínsecamente desordenadas es un mecanismo adaptativo de respuesta ante la falta de agua.

OBJETIVO GENERAL

Identificar la relación que hay entre el contenido de proteínas intrínsecamente desordenadas y la respuesta de las plantas a la falta de agua.

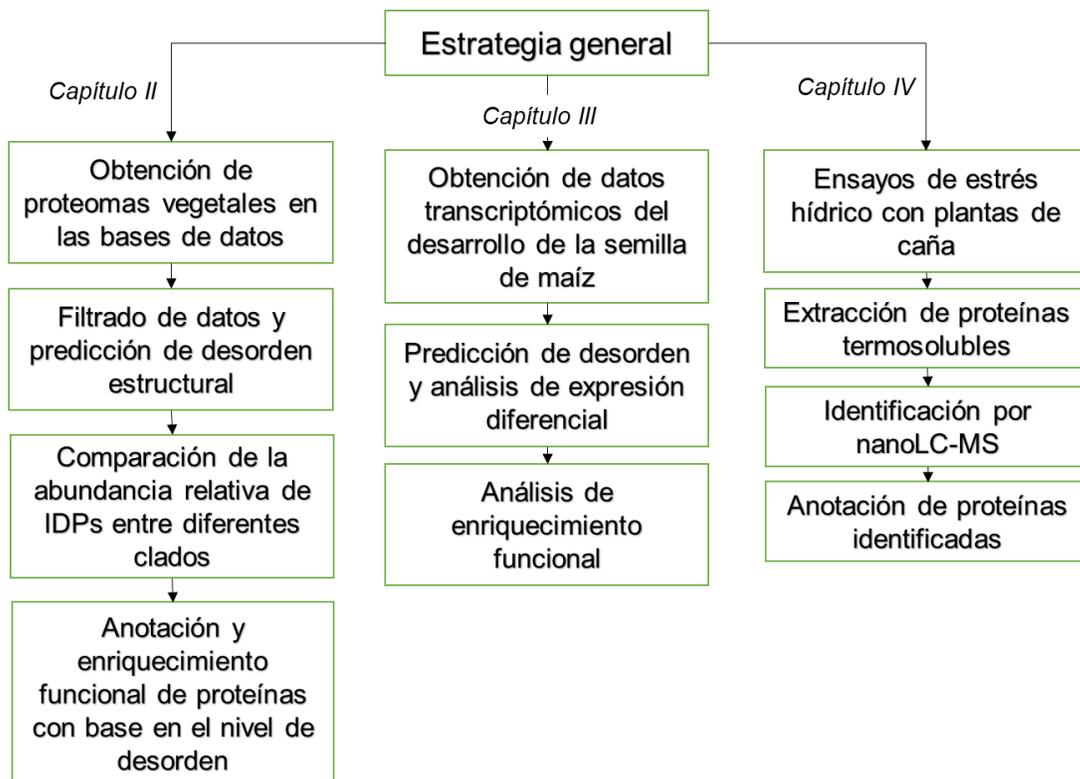
OBJETIVOS ESPECÍFICOS

1. Analizar la distribución del desorden en los proteomas vegetales y predecir las funciones moleculares en las que las IDPs están enriquecidas.
2. Caracterizar el patrón de expresión de las IDPs en un contexto de desecación.
3. Identificar las alteraciones proteómicas que ocurren en las proteínas termoestables en respuesta a estrés por falta de agua.

ESTRATEGIA EXPERIMENTAL

En este trabajo se decidió emplear tres enfoques diferentes para resolver la hipótesis planteada. Primero, se realizó un análisis global de la distribución del desorden en diferentes modelos vegetales cuya información genómica está disponible. Esto permitió

saber si existen grupos taxonómicos de plantas con una mayor proporción de desorden estructural, para luego poder anotar y catalogar las proteínas en función de su nivel de desorden estructural e inferir si en las IDPs existe enriquecimiento de funciones ontológicas asociadas a las respuestas a estrés. Posteriormente, se empleó un enfoque bioinformático para analizar el comportamiento de los perfiles de expresión de las IDPs en un contexto de desecación, empleando como modelo los datos transcriptómicos disponibles que hay sobre el desarrollo de la cariopsis de maíz, con lo que se infirió el papel que juegan en la adquisición de la tolerancia a la desecación. Por último, se empleó un enfoque experimental para analizar la participación de las IDPs en la respuesta a sequía en plantas de caña. Para ello, se estableció un experimento de sequía en plantas de caña y se evaluaron las alteraciones proteómicas que ocurren en su desordenoma con el fin de determinar las funciones en las que este fenómeno puede estar involucrado. Todo lo anterior se resume a continuación:



CAPÍTULO II

HACIA LA COMPRESIÓN DEL PAPEL DEL DESORDEN INTRÍNSECO EN LA ADAPTACIÓN DE LAS PLANTAS A DESAFÍOS AMBIENTALES

2.1 INTRODUCCIÓN

Las IDPs son proteínas que están presentes de manera ubicua en todos los dominios de la vida (Xue *et al.*, 2012; Xue *et al.*, 2010). Se estima que entre el 23 y el 28% de las proteínas eucariotas, son altamente desordenadas y que más del 50% de las proteínas eucarióticas y el 70% de proteínas involucradas en señalización contienen IDRs largas, mayores a 30 aa (Iakoucheva *et al.*, 2002; Xue *et al.*, 2012). De hecho, se reconoce que algunas familias génicas están particularmente enriquecidas en IDPs (Dai *et al.*, 2016). En general, se considera que el desorden estructural es significativamente más alto en eucariotas que en procariontes y esto ha sido asociado con la complejidad orgánica (Liu *et al.*, 2006; Peng *et al.*, 2014).

A nivel de estructura primaria, las IDPs/IDRs se caracterizan por su sesgo en la composición de aa, la baja complejidad en sus secuencias y por su bajo contenido de aminoácidos hidrofóbicos voluminosos (Romero *et al.*, 2001; Wright y Dyson, 2015). Además, poseen varios residuos subrepresentados conocidos como residuos promotores de orden (W, C, F, I, Y, V, L, H, T y N). En contraste, poseen un abundante contenido de prolina y de residuos polares y cargados, conocidos todos ellos como residuos promotores de desorden (K, E, P, S, Q, R, D y M). Finalmente, el contenido de A y G se considera que es semejante entre IDPs y sus contrapartes ordenadas (Radivojac, 2003). Estas características estructurales, les confiere una elevada carga neta y una baja hidrofobicidad media (Uversky *et al.*, 2000). Además, el desorden intrínseco favorece la flexibilidad dinámica de las proteínas y esta flexibilidad permite la transición entre diferentes estados estructurales, lo que ayuda a que estas proteínas generalmente posean funciones multiespecíficas (Romero *et al.*, 2001; Radivojac, 2003; Uversky, 2011; Sun *et al.*, 2013; Covarrubias *et al.*, 2017; Zamora-Briseño *et al.*, 2018), particularmente las asociadas a la regulación de la transcripción, la señalización y las respuestas a estrés (Sun *et al.*, 2013; Pietrosemoli *et al.*, 2013). Esto ha incentivado el desarrollo de programas predictores de desorden intrínseco. A la fecha, se

han desarrollado diversos predictores con base a los atributos fisicoquímicos de estas proteínas. Algunos de estos predictores han demostrado ser altamente confiables (Romero *et al.*, 2001; Peng *et al.*, 2006; Mészáros *et al.*, 2009; Xue *et al.*, 2010; Walsh *et al.*, 2012; Dosztányi, 2018). Por medio de ellos, hoy en día es posible estimar de manera bastante confiable el contenido global de proteínas desordenadas en un proteoma (Walsh *et al.*, 2012; Yruela y Contreras-Moreira, 2012; Kurotani *et al.*, 2014). Esto ha hecho posible una importante cantidad de estudios orientados a contestar diversas preguntas sobre el desorden estructural a escala genómica en una gran cantidad de modelos (Xue *et al.*, 2012; Peng *et al.*, 2014; Pietrosemoli *et al.*, 2013; Schad *et al.*, 2011). No obstante, frecuentemente resulta difícil comparar los resultados obtenidos a partir de diferentes estudios y producir generalizaciones a partir de ellos. Esto debido a que cada estudio emplea diferentes predictores (cada uno con diferente nivel de confianza) y diferentes criterios para medir y clasificar el desorden estructural (Pancsa y Tompa, 2012).

A pesar de lo anterior, en plantas los análisis de IDPs realizados a escala global siguen siendo escasos y se limitan a *Arabidopsis thaliana* y otros pocos modelos vegetales (Yruela y Contreras-Moreira, 2012; Pancsa y Tompa, 2012; Pietrosemoli *et al.*, 2013; Kurotani *et al.*, 2014; Vincent y Schnell, 2016; Liu *et al.*, 2017; Álvarez-Ponce *et al.*, 2018). La escasa información de este tópico en plantas limita la identificación de los roles biológicos de IDPs que carecen de funciones homólogas en otros modelos. Además, debido al hecho de que las IDPs participan en cascadas de señalización y en los procesos de repuesta a estrés, las IDPs pueden resultar relevantes en el desarrollo y la adaptación de las plantas (Kovacs *et al.*, 2008; Pietrosemoli *et al.*, 2013; Álvarez-Ponce *et al.*, 2018; Zamora-Briseño *et al.*, 2018). Todo esto mediante mecanismos específicos de plantas, dado que estas han tenido que evolucionar para desarrollar sistemas propios que le permitan adaptarse al ambiente del cual no pueden escapar (Moore *et al.*, 2008; Schad *et al.*, 2011; Xue *et al.*, 2012; Pietrosemoli *et al.*, 2013; Peng *et al.*, 2014). Por tanto, aunque las conclusiones derivadas a partir de otros modelos pueden ser aplicables a plantas, esto no siempre es así. Por ejemplo, en un estudio en el que se evaluó la correlación entre la ocurrencia de modificaciones postraduccionales (PTM) en las IDPs/IDRs de plantas, se encontró que mientras las fosforilaciones, acetilaciones y las O-glicosilaciones presentan preferencia por las IDPs/IDRs de la misma forma en como ocurre para animales, no ocurre lo mismo con

las metilaciones, las cuales tienden a ocurrir preferentemente en regiones ordenadas (Kurotani *et al.*, 2014).

Por todo lo anterior, en este capítulo, se realizó la predicción del desorden intrínseco de proteínas de 107 proteomas de plantas y algas. A diferencia de otros trabajos, aquí se decidió clasificar los resultados obtenidos en cuatro diferentes categorías basadas en el contenido relativo del desorden estructural de las secuencias (0-25, 25-50, 50-75 y 75-100% de desorden estructural). Usando este criterio, se identificó que existe un sesgo en el contenido relativo de desorden intrínseco entre los diferentes taxones analizados, existiendo un patrón diferencial claro y significativo entre monocotiledóneas, eudicotiledóneas y algas, el cual puede tener relación con la forma en cómo se distribuyen las funciones entre los proteomas de los taxones analizados. Esto puede tener implicaciones adaptativas, dado que las funciones de las proteínas están relacionadas con su contenido relativo de desorden, lo cual también influye en el tamaño de las proteínas, su distribución en la célula, su función e incluso su tasa de evolución. Finalmente, se discute cómo estas diferencias pueden tener implicaciones biológicas en las respuestas adaptativas de las plantas ante desafíos de su entorno.

2.2 MATERIALES Y MÉTODOS

Para analizar la distribución del desorden en los proteomas vegetales, se emplearon los proteomas de especies disponibles en los navegadores genómicos Ensembl Plants (<http://plants.ensembl.org/species.html>) y Phytozome (<https://phytozome.jgi.doe.gov/>). Se eliminaron todas las secuencias por debajo de 30 aa de longitud y se eliminaron todas las posiciones inespecíficas. A cada proteoma, se le hizo la predicción de desorden con el programa ESpritz (Walsh *et al.*, 2012) y los resultados de estas predicciones se agruparon de acuerdo con el contenido relativo de desorden de cada secuencia en cuatro diferentes categorías de desorden estructural definidas *a priori*: 0-25, 25-50, 50-75 y 75-100%. Se cuantificó el número de secuencias de cada categoría para cada especie, las cuales se agruparon con phyloT (<http://phylot.biobyte.de/>) y los resultados se exportaron a iTOL (Letunic y Bork, 2016).

Para visualizar las diferencias en la composición de aa de las proteínas de cada categoría, se cuantificó la composición relativa de aa de las proteínas por categoría para algas, monocotiledóneas y eudicotiledóneas y los datos se graficaron de acuerdo con el tipo de aa. Para encontrar las funciones ontológicas enriquecidas entre las proteínas de cada categoría independientemente de su clado, primero se hizo la anotación de las secuencias proteicas con el programa InterproScan5 (Jones *et al.*, 2014). Esto permitió tener proteínas anotadas con un criterio homogéneo. De los resultados obtenidos, se tomó una submuestra aleatoria de las anotaciones de 25 000 proteínas de cada categoría, las cuales se compararon entre sí con WEGO (Ye *et al.*, 2006) y se seleccionaron los términos ontológicos parentales que fueron significativamente enriquecidos para una $p < 0.001$, los cuales se exportaron y graficaron como mapa de calor. Por otra parte, se comparó la longitud de las secuencias proteicas entre cada categoría de desorden para algas, monocotiledóneas y eudicotiledóneas, así como su contenido porcentual de desorden. Las diferencias estadísticas entre la longitud de las proteínas y el contenido del desorden intrínseco de cada categoría se compararon utilizando una prueba de Kruskal-Wallis y una prueba t, respectivamente, las cuales se llevaron a cabo en R (R Development Core Team, 20016). Todos los datos se graficaron en ggplot2 (Ginestet, 2011).

Finalmente, para observar cómo el desorden intrínseco influye tanto en los procesos biológicos de las proteínas como en su localización celular, se tomó como ejemplo los datos del proteoma de *A. thaliana*. Para ello, se seleccionaron las proteínas de acuerdo con su categoría de desorden y a cada categoría se le realizó un análisis de enriquecimiento ontológico empleando ShinyGO v0.61 (Ge y Jung, 2018; Ge *et al.*, 2019). El valor de corte de la significancia estadística del enriquecimiento ontológico se fijó a un valor de $p < 0.001$. Con esta misma herramienta, se estimaron los motivos de unión a factores de transcripción que se encontraron sobre-representados en los promotores (300 pb río arriba) de los genes que codifican para cada proteína de cada categoría de desorden.

2.3 RESULTADOS

El análisis de distribución del desorden mostró que existen clados en los que la proporción porcentual de desorden intrínseco es mayor (Figura 2.1).

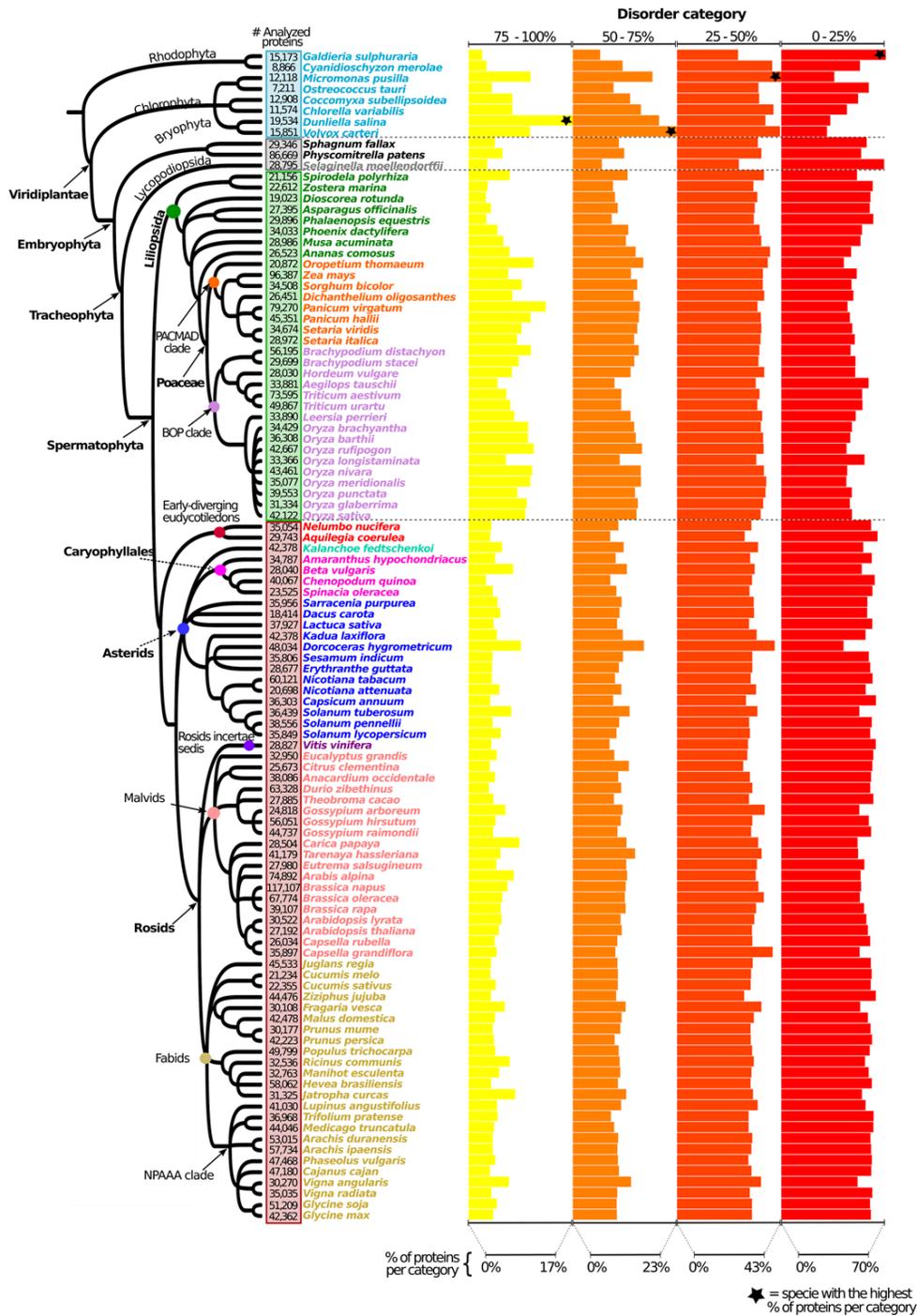


Figura 2.1 Distribución de proteínas desordenadas entre los proteomas de monocotiledóneas, eudicotiledóneas y algas. Se puede observar que la abundancia relativa de desorden no es constante entre los clados analizados. Con estrellas, se marca la especie con la abundancia más alta de proteínas en cada categoría.

De manera general, los proteomas de las monocotiledóneas y las algas poseen una mayor proporción de proteínas desordenadas. Además, el contenido relativo de proteínas con un desorden de al menos el 25%, es estadísticamente mayor para las algas y para las monocotiledóneas con respecto a las eudicotiledóneas. En contraste, las eudicotiledóneas poseen un mayor contenido relativo de proteínas ordenadas (<25% de desorden) con respecto a algas y monocotiledóneas (Figura 2.2).

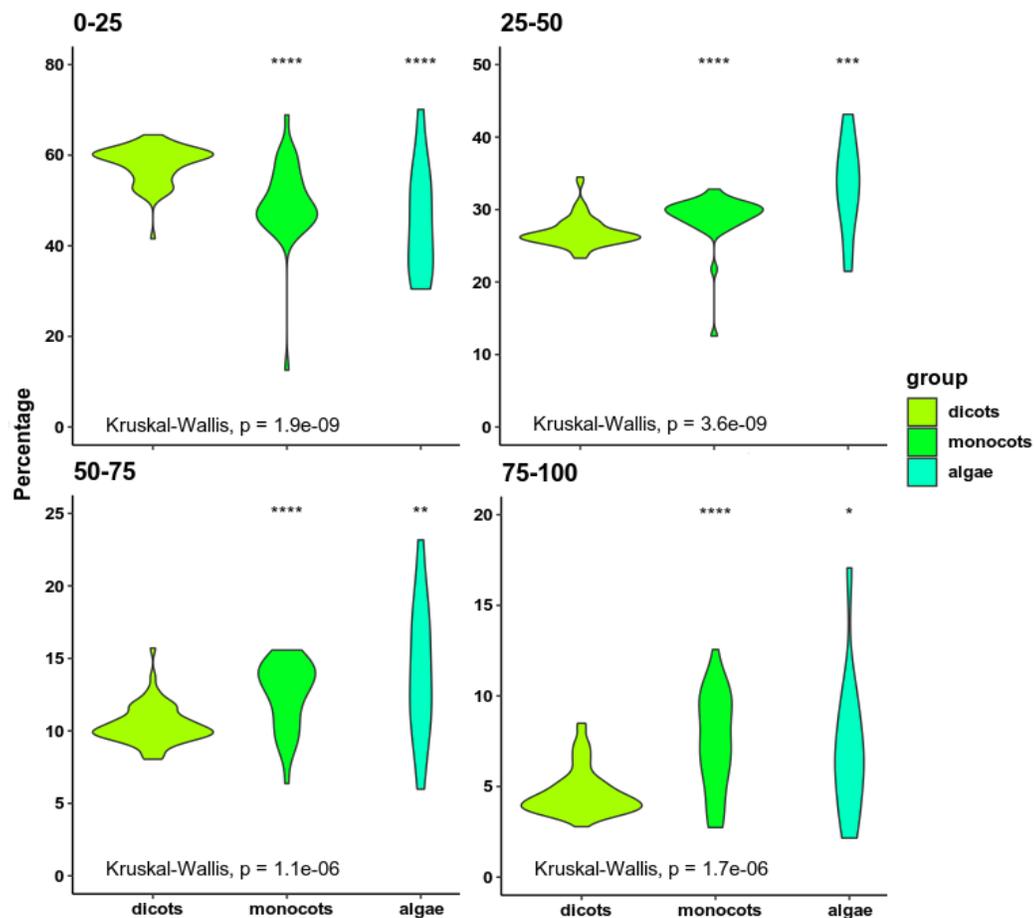


Figura 2.2 Comparación entre la abundancia relativa de cada categoría de acuerdo con su contenido de desorden. Se puede observar que la abundancia relativa de desorden no es constante entre los clados analizados.

También se encontró que, a mayor nivel de desorden, las proteínas tienden a ser más pequeñas independientemente del clado al que pertenezcan (Figura 2.3).

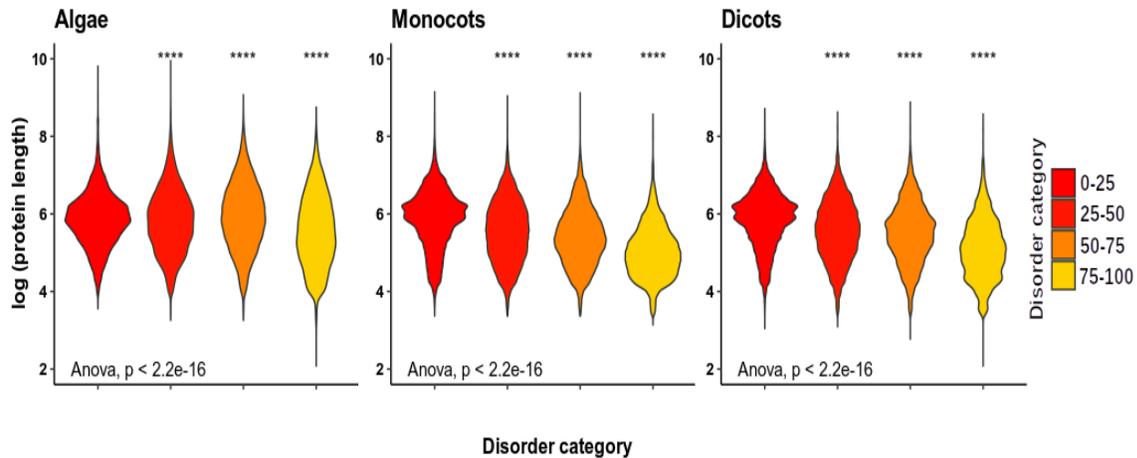


Figura 2.3 Relación entre el nivel de desorden con el tamaño de las proteínas. Es claro que, a mayor nivel de desorden, la longitud de las proteínas disminuye. Este patrón se observa de manera independiente del clado del que se trate. Las comparaciones pareadas se hicieron entre parejas comparando cada categoría con respecto a la categoría de 0-25. Se empleó una prueba t-student. **** = $p < 0.0001$.

De acuerdo con lo esperado, en el análisis de la composición de aminoácidos en las diferentes categorías proteicas para cada clado, la proporción de aminoácidos hidrofóbicos es mayor para proteínas más estructuradas. Además, si se compara su perfil de abundancias entre los diferentes clados, se puede observar que estos mantienen una proporción bastante uniforme en comparación con los aminoácidos pequeños, hidrofílicos y cargados, los cuales están enriquecidos en las IDPs (Figura 2.4).

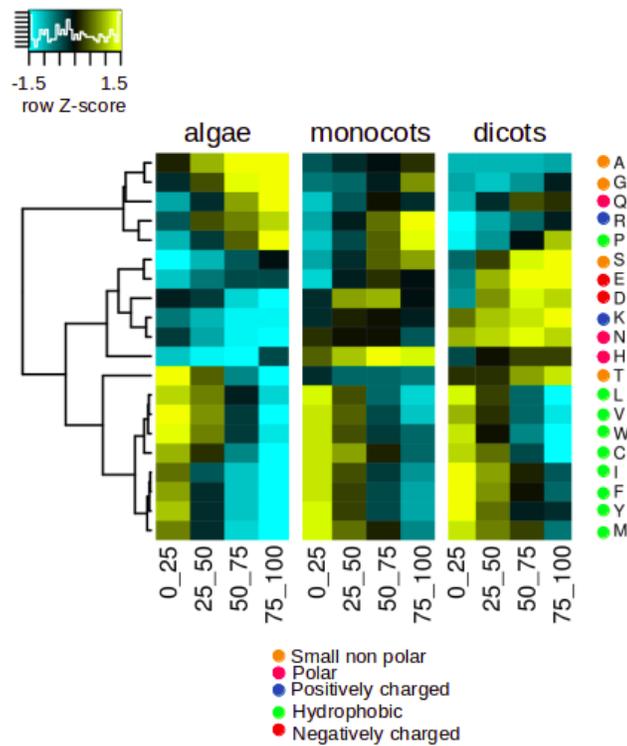


Figura 2.4 Mapa de calor de la composición porcentual de los aminoácidos que conforman a las proteínas de acuerdo con su nivel de desorden. Se puede observar que los aminoácidos hidrofóbicos se separan perfectamente del resto a excepción de la P. Estos aminoácidos se encuentran enriquecidos en las proteínas más estructuradas y presentan un patrón de abundancias más uniforme comparado con el resto de aminoácidos. Por otro lado, los aminoácidos promotores de desorden se encuentran enriquecidos (K, E, P, S, Q, R, D) en las proteínas con mayor nivel de desorden, a excepción de M, que parece no seguir este comportamiento típico. Además, el enriquecimiento en la abundancia relativa de estos aminoácidos no es constante a lo largo de los diferentes clados, lo que parece estar influenciado por factores taxonómicos.

Con la intención de comprender la división funcional entre cada una de las cuatro categorías de desorden intrínseco, se realizó un análisis de anotación funcional para cada una. De este análisis se desprende que existen categorías ontológicas enriquecidas para cada una de las categorías de desorden (Figura 2.5).

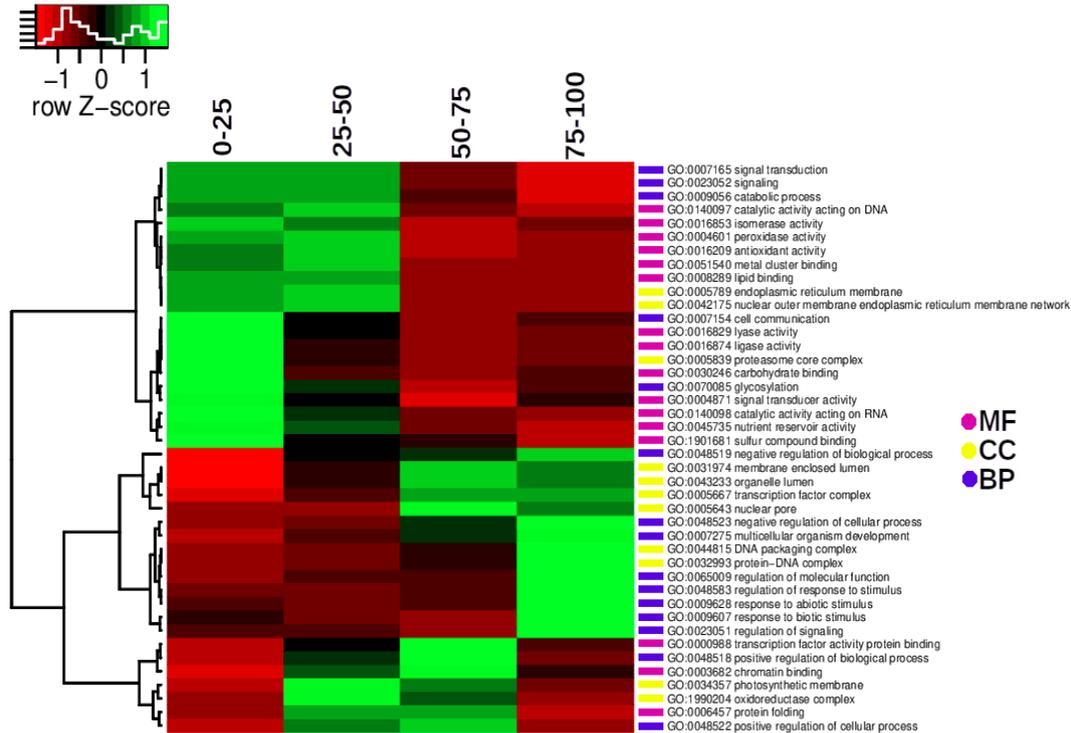


Figura 2.5 Enriquecimiento ontológico de las proteínas pertenecientes a las diferentes categorías de nivel de desorden. Entre otros aspectos, es interesante destacar que mientras las proteínas menos desordenadas (< 25% de desorden intrínseco), están enriquecidas en funciones catalíticas, las proteínas más desordenadas (>75% de desorden), están enriquecidas en términos ontológicos asociados con regulación de las actividades del núcleo, la regulación de procesos metabólicos, las respuestas a estímulos ambientales tanto del tipo biótico como abiótico, así como en procesos regulatorios.

Lo anterior coincide con lo encontrado a nivel específico para las proteínas de *A. thaliana*, sólo que, en este caso, el nivel de detalle es mayor. En este sentido, se puede observar que hay una clara separación entre los procesos biológicos en los que están enriquecidos cada categoría. La categoría de 0-25% está enriquecida en procesos sintéticos, tales como el metabolismo de lípidos, transporte de membranas, transporte de iones a través de membranas, procesos catabólicos o procesos asociados al metabolismo de fósforo. A medida que el desorden relativo se incrementa, se enriquecen los procesos biológicos regulatorios, tales como la regulación de la expresión génica, la regulación de la transcripción o la regulación de procesos metabólicos (categorías 25-50 y 50-75%). Sin embargo, lo que más llama la atención es que las proteínas más desordenadas (>75% de desorden intrínseco), están especializadas en procesos biológicos asociados con la

regulación y el procesamiento de RNA, como el splicing alternativo, el transporte de RNA y la regulación negativa de procesos metabólicos y de la actividad quinasa (Figura 2.6).

Biological process

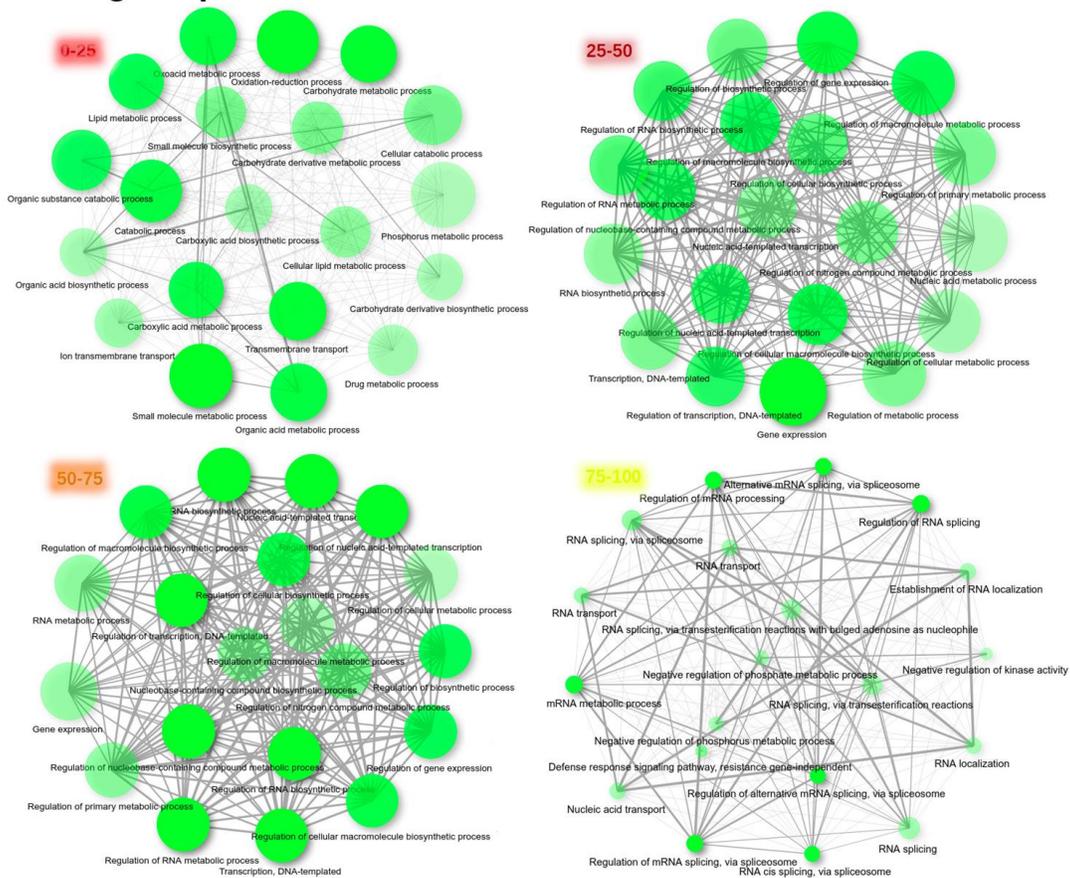


Figura 2.6 Enriquecimiento de procesos biológicos para las proteínas de *A. thaliana* de acuerdo con las diferentes categorías de desorden estructural. Se puede observar que los procesos que desarrolla cada categoría varían entre ellos.

Lo anterior coincide con el enriquecimiento de componentes celulares (Figura 2.7).

Cellular component

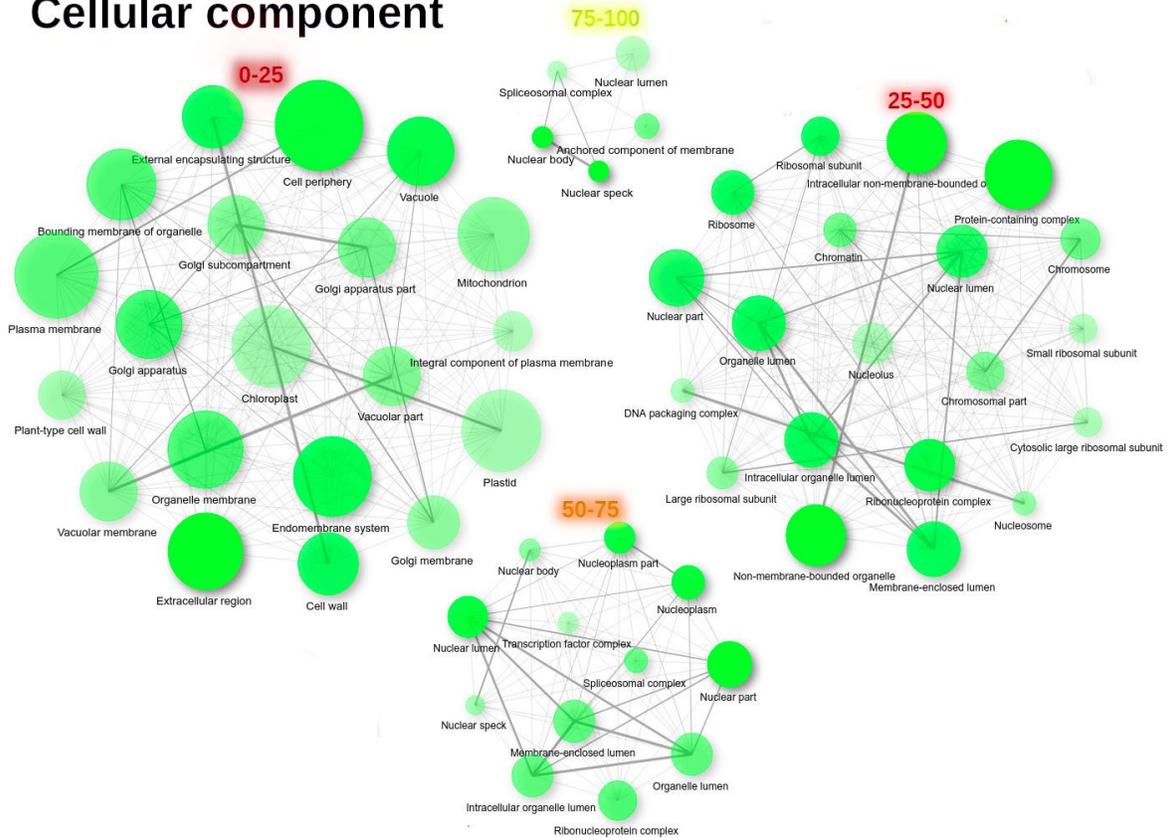


Figura 2.7 Análisis de enriquecimiento de componentes celulares para las proteínas de *A. thaliana* de acuerdo con las diferentes categorías de desorden estructural. Se puede observar que la distribución de las proteínas de acuerdo con su nivel de desorden en la célula no es azarosa. Así, mientras que el núcleo está enriquecido en proteínas desordenadas, las proteínas con menos de 25% de desorden tienen una distribución más amplia en la célula, pero no están enriquecidas en el núcleo.

En este caso, es interesante destacar la marcada separación de las proteínas en función de su nivel porcentual de desorden. Así, las proteínas más ordenadas (<25%), están ampliamente enriquecidas en términos ontológicos asociados a diversos espacios sub-celulares, como la mitocondria, el aparato de Golgi, el cloroplasto o la pared celular. Es destacable que, para este tipo de proteínas, el componente celular del núcleo no se

enriqueció. Para el caso de las proteínas con 25-50% de desorden, los términos ontológicos enriquecidos incluyen al ribosoma, el nucleosoma, la cromatina, el lumen nuclear, el nucleólo, organelos no unido a membrana, entre otros. Sin embargo, a medida que aumenta el nivel de desorden, disminuye la diversidad de espacios sub-celulares enriquecidos y es más evidente que a mayor desorden, las proteínas se localizan de manera preferente en el núcleo. De tal forma, las proteínas con 50% de desorden o más, se enriquecen exclusivamente en términos ontológicos asociados al núcleo, particularmente el nucleolo, el complejo de corte y empalme, el poro nuclear o el complejo de transcripción. Algo parecido ocurre con la categoría de 75-100% de desorden, en cuyo caso se observó el enriquecimiento de términos tales como el lumen nuclear, el complejo de corte y empalme o el cuerpo nuclear. De esta manera, tal parece que la distribución de las proteínas en las células en parte depende de su nivel de desorden y que el núcleo está enriquecido en IDPs.

2.4 DISCUSIÓN

En comparación con el total de trabajos enfocados a caracterizar el papel funcional del desorden intrínseco en los organismos, los trabajos dedicados a comprender este fenómeno en plantas siguen siendo escasos (Yizhi *et al.*, 2018; Zamora-Briseño *et al.*, 2019). A su vez, en gran parte de estos últimos, el estudio del rol del desorden intrínseco en plantas o no ha sido el objetivo principal (Xue *et al.*, 2012; Peng *et al.*, 2014), o bien sólo se han centrado en estudiar una o muy pocas especies, a partir de lo cual se extraen conclusiones generales que, si bien son valiosas, merecen ser validadas (Pietrosemoli *et al.*, 2013; Kurotani *et al.*, 2014; Choura *et al.*, 2019). Por ello, en este trabajo se realizó un análisis extensivo de la distribución de los proteomas de 107 modelos de plantas. De este análisis se desprenden varios datos que resultan interesantes de analizar.

Algunas estimaciones realizadas anteriormente entorno a las variaciones en el contenido de desorden entre clados de plantas han arrojado resultados contradictorios. Primero se encontró que el contenido de desorden intrínseco no varía entre plantas de diferentes clados (Yruela y Contreras-Moreira, 2012), aunque posteriormente se encontró que el contenido de desorden es mayor entre monocotiledóneas y eudicotiledóneas (Kurotani *et al.*, 2014; Choura *et al.*, 2019). Es probable que las diferencias en las conclusiones de estos estudios se deban al pequeño tamaño de la muestra empleado, así como en las diferencias

en los criterios para estimar el desorden. Debido a estas conclusiones contrastantes, se decidió comparar el contenido relativo de desorden entre los diferentes clados. A partir de esto, se encontró que para todas las categorías de desorden mayor al 25%, el contenido de desorden intrínseco es mayor en algas, seguido de monocotiledóneas y eudicotiledóneas. En cambio, sólo la categoría de 0-25% de desorden presenta un comportamiento opuesto. Dado que, para fines comparativos, se consideró que ésta última categoría está conformada por proteínas estructuradas, se puede decir que las eudicotiledóneas tienen un contenido relativo de proteínas ordenadas más elevado que las monocotiledóneas y las algas.

La categorización del desorden intrínseco en estos cuatro grupos distintos es un aspecto que se debe resaltar. Aunque esta división es artificial y se realizó sin seguir ningún tipo de criterio biológico *a priori*, lo cierto es que permitió observar que las proteínas desempeñan roles diferentes dependiendo de su nivel porcentual de desorden. Por ello, resulta importante el hecho de que catalogar las proteínas en proteínas ordenadas e intrínsecamente desordenadas como se hace en la mayoría de los trabajos, resulta demasiado simplista y puede provocar que se pierda información valiosa. En otras palabras, no sólo importa saber si una proteína es o no intrínsecamente desordenada, si no también qué tanto, pues de esto también depende su función. De alguna manera, de esta forma se intenta capturar parte de los diferentes niveles que hay descritos para el desorden intrínseco y que, en definitiva, afectan la función de las proteínas (Dunker et al., 2008; Walsh et al., 2012; Forcelloni y Giansanti, 2020).

Dicho lo anterior, resulta interesante observar que a medida que el desorden intrínseco aumenta, existe una disminución en el tamaño de las proteínas. Esta relación inversa entre desorden intrínseco y longitud de la cadena polipeptídica, ya se había reportado anteriormente y es un hecho generalmente aceptado (Peng et al., 2014; Afanasyeva et al., 2018). En cierta medida esto es algo esperado dada las diferencias en la composición aminoacídica entre las IDPs y sus contrapartes ordenadas, pues de alguna forma la ocurrencia de la mayoría de los aminoácidos guarda relación con la longitud de las proteínas (Carugo, 2008). Sabiendo el hecho de que las proteínas más largas en general se conservan más que las proteínas pequeñas (Lipman et al., 2002), es posible suponer que

las proteínas más desordenadas son menos conservadas que las más ordenadas. Dicha aseveración también se apoya en el hecho de que los cambios de aa son más rápidos para las proteínas que poseen mayor proporción de aa expuestos al solvente tal como ocurre con las IDPs (Lin *et al.*, 2007), lo que a su vez provoca que en términos generales las IDPs poseen una mayor tasa mutacional que las proteínas globulares, pues además presentan una alta tolerancia a las mutaciones antes de perder su función (Brown *et al.*, 2002; Forcelloni y Giansanti, 2020). Lo anterior sugiere que los aminoácidos que están enriquecidos en las IDPs (aa promotores de desorden), están sometidos a una menor presión selectiva y por tanto poseen mayor tasa de mutaciones que los aa que favorecen el orden estructural. Eso explicaría el por qué este último grupo de aa se separa claramente en el mapa de calor de la distribución de aa y que a su vez explica que presenten un patrón de distribución más conservado entre los clados en comparación con los aa promotores de desorden. Por otra parte, la abundancia relativa de los aa no es la misma para los diferentes clados. En general se considera que, en comparación con las proteínas estructuradas, las IDPs muestran una reducción en el contenido de C, W, Y, F, I, V y L y están enriquecidas significativamente en M, K, R, S, Q, P y E (Dunker *et al.*, 2008). Esta regla general parece no seguir el mismo patrón en plantas, pues si observamos el mapa de calor, se puede ver que al menos la M no está enriquecida en ninguno de los clados si no todo lo contrario. Además, A y G son aa que normalmente no se consideran enriquecidos en las IDPs (Radivojac, 2003) y que sí parecen estarlo en algas y monocotiledóneas. Es más, el enriquecimiento de los aa promotores de desorden parece estar en función de los clados. De hecho, da la impresión de que las monocotiledóneas están en una especie de transición entre las algas y las eudicotiledóneas en cuanto a la distribución de aa observados entre las diferentes categorías de desorden, aunque esta observación requiere ser corroborada.

En plantas existe una correlación positiva entre la tasa de recombinación genética y la frecuencia de desorden proteico y se ha propuesto que la recombinación genética podría considerarse una fuerza evolutiva que contribuye al desorden estructural en las proteínas (Yrueala y Contreras-Moreira, 2013). El hecho de que las IDPs tengan una mayor tasa de recombinación y una mayor mutabilidad resulta de particular interés en las respuestas de las plantas a condiciones desafiantes y podría estar relacionado con el hecho de que una

de las categorías ontológicas enriquecidas en las proteínas más desordenadas sea precisamente la respuesta a estímulos bióticos y abióticos. Es interesante que las proteínas de más de 75% de desorden estructural estén particularmente enriquecidas en procesos biológicos relacionados con la regulación y transporte del RNA, así como con el corte y empalme de RNA. Esto es consistente con datos previos que indican que un gran número de proteínas que se unen a RNA presentan amplias IDRs. Por ejemplo, se estima que hasta más del 50% de residuos de aa de chaperonas de RNA caen en IDRs (Tompá y Csermely, 2004). Además, el hecho de que haya enriquecimiento en el corte y empalme alternativo tiene amplias consecuencias. Por ejemplo, se sabe que este es un proceso muy importante para las respuestas inducidas por el estrés en las plantas, que puede modular los rasgos fenotípicos de las plantas y pueden contribuir a sus adaptaciones a diferentes estresores ambientales (Ling *et al.*, 2019; Mastrangelo *et al.*, 2012).

Del análisis de enriquecimiento ontológico de componentes celulares, se desprende que el nivel de desorden influye en la localización subcelular de las proteínas y es evidente que el núcleo está enriquecido en IDPs tal como lo sugieren otros trabajos (Frege y Uversky, 2015; Skupien-Rabian *et al.*, 2015). Además, considerando que en monocotiledóneas y algas la proporción de proteínas con más del 25% de desorden es mayor que en eudicotiledóneas, es factible pensar que haya variaciones en la proporción de las proteínas nucleares entre clados. Así, es de esperarse que las monocotiledóneas guarden una proporción de proteínas nucleares mayor que las eudicotiledóneas.

Finalmente, dado que gran parte del proteoma con funciones desconocidas está constituido por IDPs (Bhowmick *et al.*, 2016), se puede inferir que gran parte de este proteoma representa un reservorio de funciones potenciales no descubiertas y que podrían ser aprovechables con fines biotecnológicos. De esta forma, la caracterización de genes que codifican para IDPs y que responden a estrés puede llevar al descubrimiento de nuevos mecanismos de respuesta a estrés en plantas.

2.5 CONCLUSIÓN

Este trabajo representa el análisis del desordenoma de proteomas de plantas más extenso hasta ahora. El nivel de desorden intrínseco influye en diversas características funcionales de las proteínas más allá de su falta de plegamiento.

CAPÍTULO III

DINÁMICA GLOBAL DE LAS IDPs Y SU RELACIÓN CON LA DESECACIÓN EN SEMILLAS. LA CARIOPSIS DE MAÍZ COMO MODELO DE ESTUDIO

3.1 INTRODUCCIÓN

Durante los últimos años, se ha generado nueva información sobre el papel de las proteínas intrínsecamente desordenadas (IDPs), lo que ha cambiado nuestra comprensión de la bioquímica de proteínas. Estos hallazgos llevaron a la categorización de las IDPs como un grupo interesante de proteínas que difieren de las proteínas globulares en términos de sus modos de acción y sus características fisicoquímicas. En términos generales, las IDPs poseen baja complejidad de secuencia y una composición sesgada de aminoácidos (aa), con escasos residuos hidrofóbicos y aromáticos, y enriquecidas en aa polares y cargados (Wright y Dyson, 2015). Estas características en la estructura primaria les confieren una alta carga neta y baja hidrofobicidad media (Uversky *et al.*, 2000). Por estas razones, tales secuencias de proteínas no pueden plegarse en estructuras tridimensionales estables, rígidas y globulares. En contraste, se comportan como conjuntos dinámicamente interconvertibles de conformaciones espaciales que van desde coils extendidos hasta glóbulos colapsados (Tompa y Kovacs, 2010; Wright y Dyson, 2015). Ciertas proteínas están completamente desordenadas (IDPs en *sensu stricto*), mientras que otras no están intrínsecamente desordenadas, sino que tienen segmentos desordenados (regiones intrínsecamente desordenadas, IDRs; Romero *et al.*, 2001). A su vez, a la colección de IDPs/ IDRs de un organismo dado en una condición específica se le puede denominar colectivamente desordenoma (Zamora-Briseño *et al.*, 2018).

Las IDPs están enriquecidas en ontologías génicas asociadas con señalización y regulación de procesos celulares clave, tales como el control de la división celular, la apoptosis, la modificación postraduccional o la transcripción, por mencionar algunos (Iakoucheva *et al.*, 2002, 2004). Por ejemplo, para el 2014, el análisis bioinformático de la base de datos Swiss-Prot reveló que existen al menos 238 términos GO en la categoría de procesos celulares distribuidos exclusivamente entre las IDP/IDRs (Liu y Huang, 2014). La conformación flexible de las IDPs les confiere ventajas estructurales frente a las proteínas globulares, las

cuales incluyen economizar recursos genómicos y proteicos, superar restricciones estéricas durante las interacciones con otras moléculas, poseer interacciones con alta especificidad y baja afinidad, facilitar modificaciones postraduccionales, permitir enlaces flexibles, prevenir la agregación intermolecular, proporcionar resistencia a condiciones no nativas, favorecer una mayor velocidad de interacción, la capacidad de combinar una alta especificidad mediante uniones débiles y reversibles, la capacidad de llevar a cabo más de una función y finalmente, permitir la interacción con una mayor cantidad de proteínas (Tompa *et al.*, 2014).

En comparación con otros organismos eucariotas, la información a nivel genómico disponible respecto a las IDPs en plantas se encuentra todavía muy limitada y en general los datos exploratorios están circunscritos a *Arabidopsis thaliana* y otros escasos modelos vegetales (Pietrosemoli *et al.*, 2013; Álvarez-Ponce *et al.*, 2018). El hecho de que muchas IDPs puedan servir como integradores en cascadas de señalización y procesos de respuesta al estrés, permite suponer que su papel puede ser especialmente importante en el desarrollo y la adaptación de las plantas a las alteraciones en las condiciones ambientales, probablemente al proporcionarles un mecanismo rápido para formar complejos y redes moleculares altamente interconectadas (Pazos *et al.*, 2013). De hecho, se ha señalado que existen diversas IDPs en plantas que juegan roles importantes en las respuestas a diferentes tipos de estreses (Pietrosemoli *et al.*, 2013; Álvarez-Ponce *et al.*, 2018) y en particular al estrés por falta de agua (Kovacs *et al.*, 2008; Kovacs *et al.*, 2008; Mouillon *et al.*, 2008). Se ha sugerido que las plantas pueden usar la regulación de IDPs para adaptarse rápida y eficientemente a los cambios ambientales de los que no pueden escapar (Pietrosemoli *et al.*, 2013). Por ello, aunque se ha logrado un progreso importante en nuestra comprensión de las funciones de las IDPs, las respuestas a muchas preguntas interesantes siguen siendo difíciles de resolver. Una característica importante que evaluar es su papel en la tolerancia a la desecación (DT). La idea detrás de esta relación subyace en que uno de los efectos observados cuando se reduce el contenido de agua celular es el aumento en la densidad de las macromoléculas, un efecto conocido como hacinamiento macromolecular (Mouillon *et al.*, 2008). En ciertos casos, tal densidad llega a ajustarse de manera muy importante, por ejemplo, en las semillas ortodoxas en estado quiescente.

En el caso particular de las semillas ortodoxas ocurre que al final de su desarrollo, estas semillas presentan un período programado de deshidratación, el cual conduce a la pérdida masiva de agua en toda la estructura (Manfre *et al.*, 2009). Tal deshidratación provoca una reducción en el volumen celular. El hacinamiento macromolecular resultante de la compactación citoplasmática proporciona un entorno propicio para numerosas interacciones indeseables que pueden conducir a la fusión de membranas celulares y a la desnaturalización y agregación de ciertas proteínas especialmente susceptibles a tales condiciones (Chakrabortee *et al.*, 2007). Por otra parte, es importante hacer notar que la adquisición de la tolerancia a la desecación o la capacidad de resistir estos potenciales hídricos muy bajos y el consiguiente hacinamiento macromolecular se ha correlacionado con la acumulación de diversos compuestos protectores, incluidas proteínas protectoras y azúcares (Manfre *et al.*, 2009). Además, se ha propuesto que un posible mecanismo para ayudar parcialmente a hacer frente a este efecto es la inducción de proteínas menos propensas a la pérdida de función inducida por la desnaturalización (Zamora-Briseño *et al.*, 2018). En este sentido, es probable que la inducción generalizada de IDPs represente un mecanismo compensatorio para mitigar el efecto del hacinamiento macromolecular en la célula, ya que estas proteínas no pueden perder su función mediante un mecanismo mediado por la desnaturalización. Por su parte, también es factible una reducción en el contenido de proteínas estructuradas con propensión a la desnaturalización mediada por la falta de agua.

Por otro lado, está bien establecido que el patrón de expresión espaciotemporal de los genes representa una fuente invaluable de información acerca de si la función de un gen contribuye o no en un contexto celular determinado (Das *et al.*, 2015; Pereira-Santana *et al.*, 2017). Por lo tanto, el perfil de expresión global de genes en un contexto de desecación podría ayudar a establecer el papel que desempeñan las IDPs bajo tales condiciones. En este sentido, las semillas ortodoxas pueden ser modelos particularmente útiles para evaluar la asociación entre la regulación de las proteínas desordenadas a nivel global en un contexto anhidrobiótico, ya que estas son estructuras especializadas que han desarrollado adaptaciones evolutivas altamente conservadas para resistir la pérdida extrema de agua (Moore *et al.*, 2009). Particularmente un modelo de semilla ortodoxa muy estudiada es la cariopsis del maíz. Este es un modelo vegetal que cuenta con bastante información

genómica que puede ser útil para entender la relación entre las IDPs y la adquisición a la tolerancia a la desecación en semillas ortodoxas, por contar, entre otras cosas, con un genoma completamente secuenciado y bien anotado, con una importante cantidad de datos transcriptómicos públicamente disponibles sobre el desarrollo de semillas, así como una descripción detallada de los cambios anatómicos, fisiológicos y bioquímicos que ocurren durante su desarrollo (Prioul *et al.*, 2008; Chen *et al.*, 2014; Silva-Sánchez *et al.*, 2014; Nannas y Dawe, 2015; Yi *et al.*, 2019).

Por estas razones, en este capítulo se probó la hipótesis de que en un contexto de desecación se favorece la acumulación de proteínas menos propensas a la desnaturalización mediada por la pérdida de agua. En este sentido, se considera que el aumento del contenido de IDPs podría representar una estrategia para aumentar la resistencia general del proteoma como respuesta a condiciones extremas de pérdida de agua. Para evaluar esta idea, se analizó la expresión de genes de acuerdo con el contenido del desorden estructural de las proteínas codificadas por los transcritos expresados durante el desarrollo de la semilla de maíz incluido el inicio del programa de adquisición de la tolerancia a la desecación. En general, se encontró que el contenido global de las IDPs difiere a lo largo del proceso de desarrollo de la semilla y este cambio está estrechamente asociado con las diferentes etapas de desarrollo en las que se divide la semilla de maíz. Curiosamente, se puede observar una reducción general en la expresión de transcritos que codifican proteínas estructuradas. Además, se identificaron diversos grupos de proteínas altamente desordenadas cuya expresión está correlacionada con el inicio del programa que conduce a la adquisición de tolerancia a la deshidratación en la semilla. Por todo lo anterior, se considera que la semilla de maíz es un buen modelo para estudiar los roles de las IDPs y comprender sus implicaciones durante el desarrollo, así como en la respuesta al estrés por sequía en general. Finalmente, se propuso un modelo general que implica una reducción global de las proteínas estructuradas y un aumento en el contenido de IDPs como mecanismo para reducir el impacto negativo de la pérdida extrema de agua en la célula.

3.2 MATERIALES Y MÉTODOS

3.2.1 Adquisición de datos transcriptómicos

Para correlacionar el contenido de IDPs con la progresión de la pérdida de agua durante el desarrollo de la semilla de maíz, se emplearon los datos transcriptómicos disponibles de la semilla de maíz durante su desarrollo. Los datos del transcriptoma se descargaron del EBI Expression Atlas (<https://www.ebi.ac.uk/gxa/home>). Estos datos fueron generados por Chen *et al.*, 2014 (Chen *et al.*, 2014) y corresponden al experimento identificado como "Perfil de transcripción por secuenciación de alto rendimiento de embrión y endospermo durante el desarrollo de la semilla de maíz". Estos datos incluyen las cuentas de transcritos estandarizados en transcritos por millón de lecturas (TPM) para las semillas de maíz desde el día 0 hasta el día 38 de la post-polinización (DAP). Los identificadores de cada transcrito se recuperaron de la tabla de expresión para obtener las secuencias de proteínas codificadas por cada uno de ellos. Los identificadores correspondientes se descargaron del genoma de maíz B73 versión 4, disponibles en el Ensembl Plants (plants.ensembl.org/index.html).

3.2.2 Análisis de desorden proteico y agrupamiento de datos

La predicción del desorden intrínseco en las proteínas se llevó a cabo mediante Espritz (Walsh *et al.*, 2012) usando X-ray y Best Sw como parámetros. Las isoformas más largas de cada transcrito se usaron y catalogaron de acuerdo con el porcentaje de desorden estructural de su proteína codificada. Los transcritos se catalogaron en cuatro categorías diferentes de menor a mayor porcentaje de desorden estructural de las proteínas codificadas: (I) 0-25, (II) 25-50, (III) 50-75 y (IV) 75-100 de porcentaje de desorden. Las secuencias proteicas de cada categoría se analizaron para cuantificar la composición relativa de aa. El porcentaje de la composición de aa de cada categoría se graficó como un mapa de calor. Estos datos también se utilizaron para realizar un análisis de componentes principales (PCA) para identificar si las categorías se podían distinguir entre sí. Ambos análisis se realizaron con ClustVis (Metsalu y Vilo, 2015). Por otro lado, las proteínas fueron anotadas con Blast2GO (Conesa *et al.*, 2005) y el enriquecimiento ontológico (GO) se realizó con WEGO (Ye *et al.*, 2006). Los términos GO enriquecidos se exportaron y graficaron en forma de mapas de calor mediante ClustVis (Metsalu y Vilo, 2015). Para correlacionar la longitud de la proteína y el desorden estructural, se clasificaron las

proteínas en seis categorías diferentes de acuerdo con su longitud: (I) 30-250, (II) 250-500, (III) 500-750, (IV) 750-1000, (V) 1000-1500, y (VI) > 1500 aa. A su vez, estas categorías se agruparon en cuatro categorías de proteínas desordenadas según su nivel de desorden (0-25, 25-50, 50-75 y 75-100% de desorden).

3.2.3 Minería de datos transcriptómicos y análisis de expresión diferencial

Para obtener la proporción de transcritos que codifican proteínas desordenadas con respecto los que codifican para proteínas estructuradas, se estimó el contenido relativo de las transcritos de cada categoría. Se obtuvo la suma total de TPM de cada categoría y se dividió entre el número total de TPM por día y se multiplicó por 100 y los resultados se representaron con respecto al tiempo (0-38 DAP). Estas cinéticas se dividieron en cinco etapas de desarrollo diferentes (R1, R2, R3, R4 y R5), según Hanway (Hanway, 1966). Además, se cuantificó la cantidad porcentual del número de genes expresados de cada categoría (considerando como genes expresados aquellos con un TPM > 0). Para confirmar la correlación entre la etapa de desarrollo y el perfil de expresión global de las IDPs, se realizó un análisis de correlación de Pearson de transcritos que codifican proteínas con un contenido de desorden > 25%.

Para identificar transcritos expresados de manera específica entre las etapas de desarrollo y los compartidos entre cada una de ellas, se construyeron diagramas de Venn de los transcritos expresados (TPM > 0). Para simplificar la agrupación de los datos, los transcritos se dividieron en transcritos que codifican proteínas estructuradas (0-25% del desorden) y transcritos que codifican IDPs (> 25% del desorden). Para obtener transcritos que codifican proteínas altamente desordenadas (>75% del desorden) que también se expresan diferencialmente, se utilizó el paquete EdgeR (McCarthy *et al.*, 2012), utilizando un valor de corte de FDR < 0.001. A partir de estos datos, se definieron dos grupos de transcritos que codifican IDPs, los transcritos que están regulados positivamente y los que están regulados negativamente a lo largo del desarrollo de la semilla. Además, se realizó un análisis de enriquecimiento funcional, tanto para los genes sobregulados como para los desregulados. Este análisis se realizó con Blast2GO (Conesa *et al.*, 2005).

3.2.4 Análisis de redes de co-expresión

Para averiguar si existe una especialización funcional de los factores de transcripción (TFs) en función de su nivel de desorden estructural, se infirieron las redes de regulación formadas de acuerdo con el nivel de desorden estructural de los TFs identificados como regulados al alza en las etapas temprana (R1), media (R2 y R3) y tardía (R4 y R5). Los TFs regulados al alza se obtuvieron del análisis diferencial realizado con EdgeR, empleando la lista de TFs expresados en semilla que fueron identificados previamente (Chen *et al.*, 2014).

Los valores de TPM de los genes correspondientes al conjunto de datos se filtraron para reducir la complejidad y el ruido de los datos. Para el entrenamiento de la matriz de co-expresión, sólo se emplearon aquellos genes con al menos un conteo por tratamiento de punto de muestreo y con una suma total de conteos por tratamiento mayor de 21 (hubo 21 puntos de muestreo). Para construir la matriz de co-expresión por pares, se utilizó el paquete de bioconductor GENIE3 (Huynh-Thu *et al.*, 2010; Aibar *et al.*, 2017), que se basa en el modelo de aprendizaje automático Random Forest (Breiman, 2002) empleando los parámetros pre-establecidos por defecto. Se construyeron cuatro redes de coexpresión diferentes usando como "nodos reguladores" los 73 factores de transcripción expresados diferenciales (DETF) encontrados en el análisis de expresión diferencial y como matriz de valores de expresión, la matriz de conteos TPM y filtrada para un valor de TPM > 21. Ésta última matriz se empleó para estimar los nodos objetivo de la red de co-expresión. Los 73 DETF se agruparon según su nivel de desorden (0-25, 25-50, 50-75 y 75-100 %). Para la primera red, se utilizaron como nodos reguladores los cinco DETF con 0-25% de desorden; para la segunda red se utilizaron 34 DETF dentro de un 25-50% de desorden intrínseco; la tercera red se basó en 22 DETF con 50-75% de desorden, mientras que la cuarta red se basó en 12 DETF dentro de un 75-100% de desorden intrínseco. Los archivos resultantes del software GENIE3 se cargaron como tablas en Cytoscape v.3.6.0 (Shannon *et al.*, 2003). Todas las redes de co-expresión se tomaron como no dirigidas pero ponderadas. Las redes se agruparon mediante el uso del algoritmo de agrupación GLay en Cytoscape (Su *et al.*, 2010). Finalmente, los grupos de transcritos se separaron de acuerdo con el nivel de desorden TF y la etapa de desarrollo en la que se indujo su expresión.

3.3 RESULTADOS

Se encontró una clara relación entre la longitud de la proteína y el contenido de desorden intrínseco (Figura 3.1). La proporción de proteínas con menor longitud es mayor a medida que el desorden aumenta. Por lo tanto, las IDPs tienden a ser más cortas .

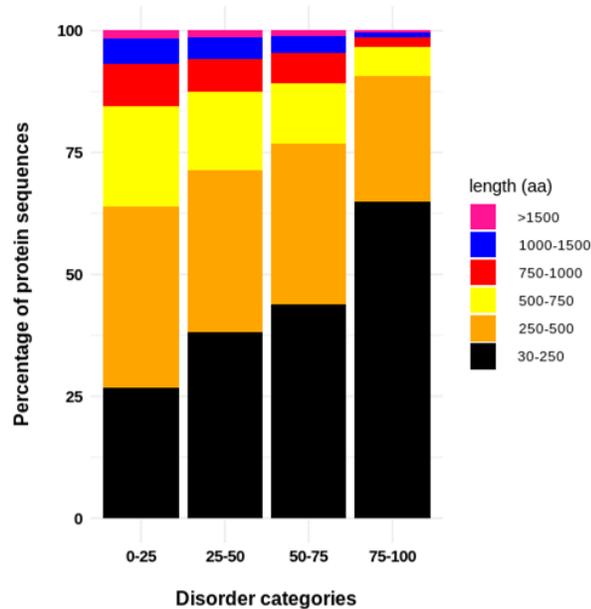


Figura 3.1 Relación entre el nivel de desorden proteico y la longitud de la proteína. Se puede observar que mientras más desordenada es una proteína, mayor es la proporción relativa de proteínas pequeñas de cada categoría.

Desde un punto de vista funcional, existe una clara separación entre las funciones enriquecidas de proteínas estructuradas ($\leq 25\%$ de desorden) y las proteínas desordenadas. Las proteínas desordenadas se enriquecen en funciones reguladoras, en contraste con las proteínas estructuradas que se enriquecen en las funciones catalíticas (Figura 3.2 A). Además, es posible separar cuatro categorías basadas en la composición de aa, ya que se pueden diferenciar las proteínas estructuradas ($< 25\%$ del desorden proteico) de las que presentan un desorden moderado (25-50% y 50-75%) y proteínas altamente desordenadas ($> 75\%$ de desorden proteico; Figura 3.2 B). Esta separación ocurre debido a que las proteínas estructuradas tienden a enriquecerse en residuos hidrofóbicos, mientras que los aa polares e hidrofílicos están sobre-representados en proteínas altamente desordenadas (Figura 3.2 C).

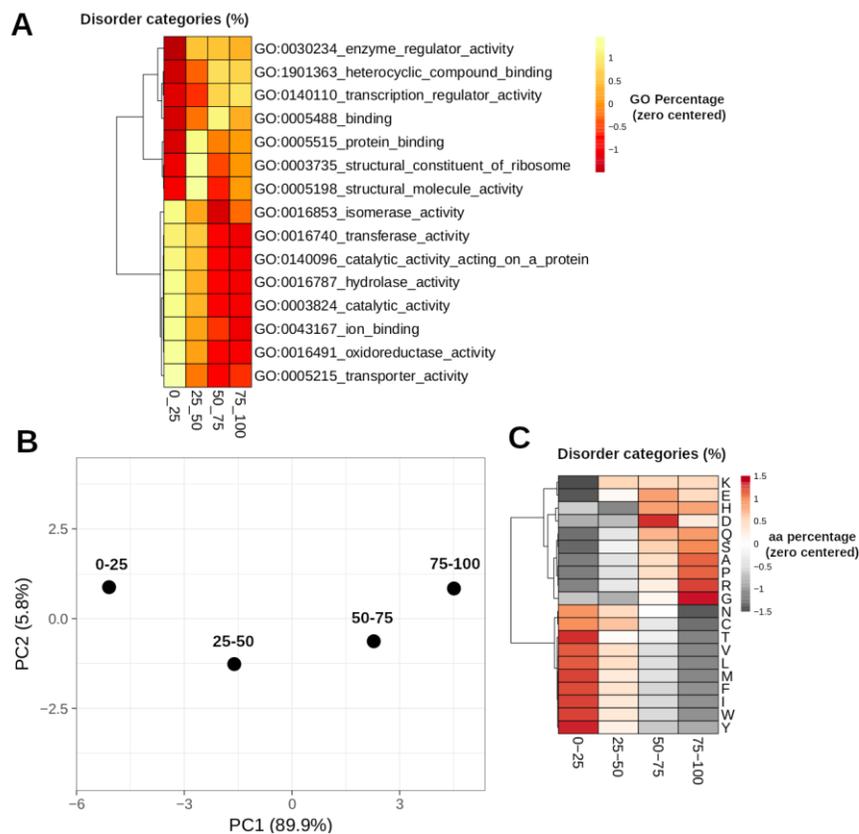


Figura 3.2 Características generales de las proteínas según el contenido del desorden. A) Mapa de calor de los términos GO enriquecidos entre las cuatro categorías para un valor $p < 0.01$. Los colores representan el porcentaje de cada categoría GO con escala aplicada a las filas y los valores centrados a cero. B) Análisis de componentes principales (PCA) de la composición aa de las proteínas que constituyen cada categoría. Los ejes X e Y muestran el componente principal 1 y el componente principal 2, que explican el 89,9% y el 5,8% de la varianza total, respectivamente. C) Mapa de calor de la composición aa de las diferentes categorías de desorden.

Además, durante la progresión del desarrollo de la semilla hay una reducción en el número de genes expresados que codifican proteínas estructuradas (Figura 3.3). Aparentemente, los genes que codifican para estas proteínas se activan en las primeras etapas del desarrollo de la semilla y se inhiben más adelante (Figura 3.3 A). En contraste, esto no ocurre para los transcritos que codifican proteínas con $>25\%$ de desorden estructural. En este caso, hay un subconjunto de genes que se activan específicamente en cada una de las etapas de desarrollo (Figura 3.3 B).

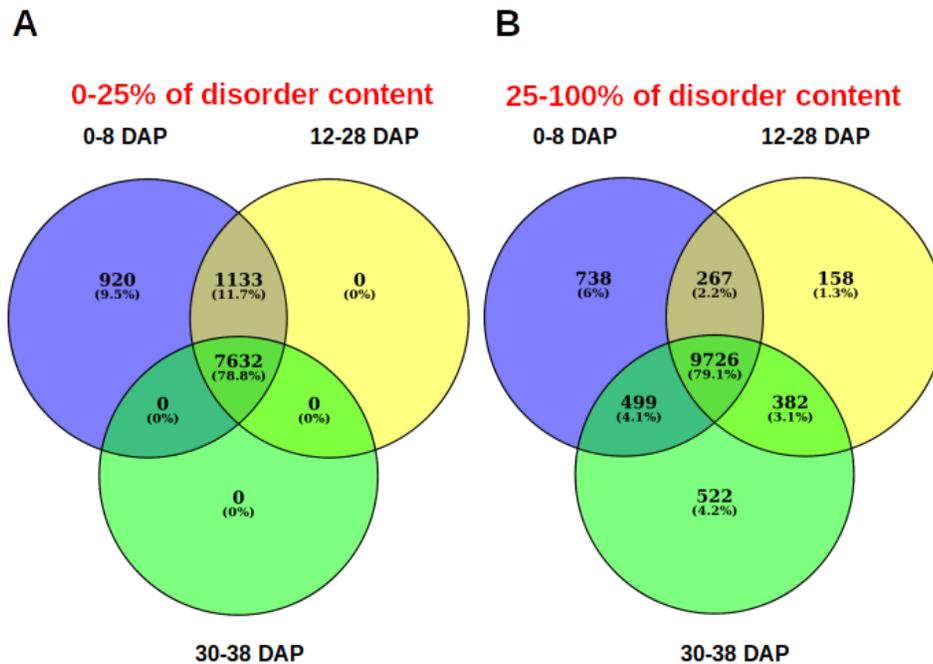


Figura 3.3 Diagramas de Venn de los genes expresados (TPM > 0) y compartidos de cada etapa, de acuerdo con el contenido de proteínas desordenadas codificadas por cada transcrito. A) Los transcritos que codifican proteínas estructuradas (0-25% del contenido de desorden) se expresan temprano (0-8 DAP) y luego se desactivan en las siguientes etapas. B) Hay transcritos que codifican IDPs que se expresan específicamente en cada etapa.

Por otra parte, el contenido relativo de los transcritos que codifican proteínas estructuradas no es constante durante el desarrollo de la semilla. Si se divide el conjunto de datos en las cinco fases diferentes del desarrollo de la semilla de maíz (R1, R2, R3, R4 y R5), que incluyen el período de tiempo analizado en este trabajo (0-38 DAP), es interesante notar que en las fases R2 y R3, hay un aumento en la proporción de transcritos que codifican proteínas estructuradas (Figura 3.4 A).

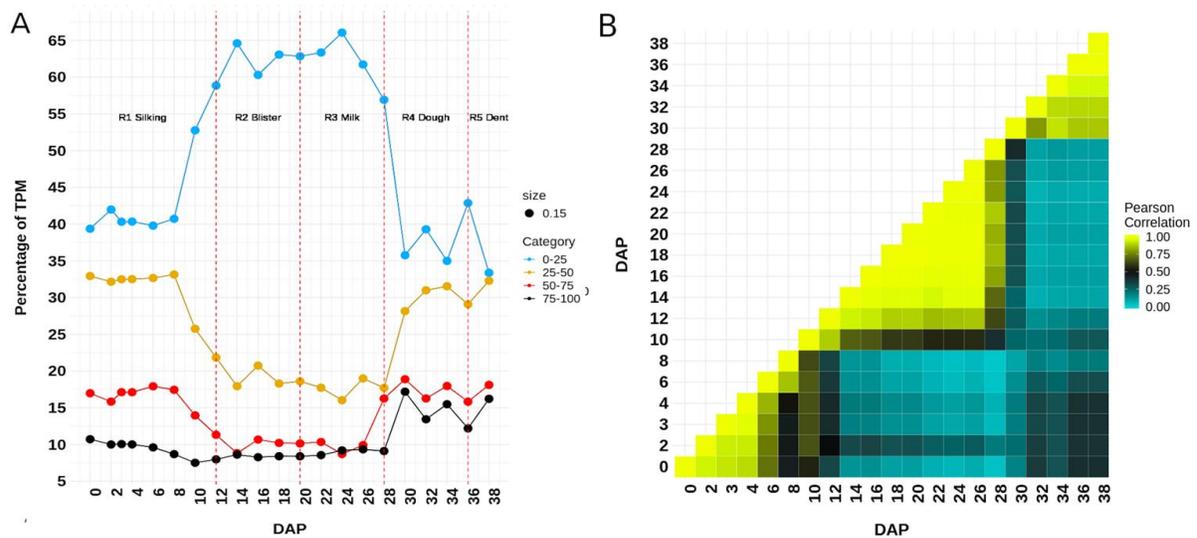


Figura 3.4 Patrones de expresión de transcritos de acuerdo con el contenido del desorden. A) Porcentaje de TPM de transcritos que codifican proteínas con diferentes grados de desorden. Hay una inducción en la expresión general de transcritos que codifican proteínas estructuradas durante la fase media del desarrollo de la semilla. B) Análisis de correlación de Pearson de la expresión de transcritos que codifican proteínas con > 25% de desorden. Hay una clara separación de las diferentes fases de desarrollo, lo que indica la expresión específica de los transcritos durante cada una de las etapas de desarrollo.

Con base en lo anterior, se identificaron los transcritos más abundantes de estas fases y se seleccionaron los 15 más abundantes. De este modo, se identificaron los siguientes: alfa zeína A20 (Zm00001d019155), alfa zeína 19D1 (Zm00001d030855), zeína 15kD (Zm00001d035760), zeína 10 kD (Zm00001d045937), alfa zeína 3 22kD (Zm00001d048809, Z0004D048809), alfa zeína (Zm00001d048813), zeína 3 (Zm00001d048817), alfa zeína 19B1 (Zm00001d048848), alfa zeína (Zm00001d048849), alfa zeína PMS1 (Zm00001d048850), alfa zeína z1C2 y Zm2-z3 (Zm) Zm00001d049476).

Los recuentos totales de estos transcritos representan hasta 39.3 ± 7.94 (promedio \pm DE) por ciento del total de TPMs en la fase media (R2 y R3), mientras que en las fases temprana y tardía sus abundancias normalizadas representan 0.62 ± 0.84 y 2.70 ± 3.43 , respectivamente. Curiosamente, la expresión global de IDPs (> 25% de desorden proteico) depende en gran medida de la fase de desarrollo y esto se ve reflejado en la expresión total de este tipo de transcritos con respecto al tiempo (Figura 3.4 B).

Es evidente que los transcriptomas de 0 a 10 DAP (R1), 12 a 28 DAP (R2 y R3) y 30 a 38 DAP (R4 y R5) formaron grupos que coinciden con las fases de desarrollo temprana, media y tardía, respectivamente (Chen *et al.*, 2014). R1 es un período activo de división y desarrollo celular, mientras que R2 y R3 corresponden a las fases de desarrollo de la morfogénesis y la maduración. Durante las etapas R2 y R3, en el embrión ocurre una síntesis activa de DNA, división celular y diferenciación, seguido de síntesis de moléculas de reservas y osmoprotectores (Vernoud *et al.*, 2005). Un grupo distinto después de las etapas R4 y R5 (30 a 38 DAP) abarca el final de la acumulación de compuestos de almacenamiento en el endospermo y la activación de procesos biológicos involucrados en la latencia y la deshidratación (Chen *et al.*, 2014).

De hecho, al final de la etapa R5 la proporción de transcritos expresados que codifican para IDPs aumenta con respecto a los transcritos que codifican para proteínas estructuradas (Figura 3.5) con un pico atípico de transcritos que codifican proteínas estructuradas a los 22 DAP. Por lo tanto, es de esperarse que la longitud promedio de las proteínas en la fase tardía del desarrollo de la semilla tienda a reducirse en comparación con el inicio de la embriogénesis.

Estos resultados sugieren que los cambios en la proporción de desorden estructural es consecuencia de los cambios funcionales que ocurren en las diferentes fases del desarrollo. Cuando comienza el desarrollo de semillas, se requieren una gran cantidad de recursos celulares para la división y diferenciación celular. Este proceso requiere una importante regulación genética. Sin embargo, cuando la semilla entra en la fase media de desarrollo, el metabolismo se vuelve principalmente biosintético, dado que se necesitan varias enzimas para la síntesis de almidón y otros compuestos de almacenamiento. Al final de la fase de llenado y al comienzo de la etapa tardía de desarrollo (R5), se inicia el programa de deshidratación.

Por lo anterior, se decidió identificar los transcritos sobre-regulados que codifican para las IDPs. Así, se identificaron 1015 transcritos que codifican para IDPs (> 25% de desorden) que se expresan positivamente al final de la etapa R5.

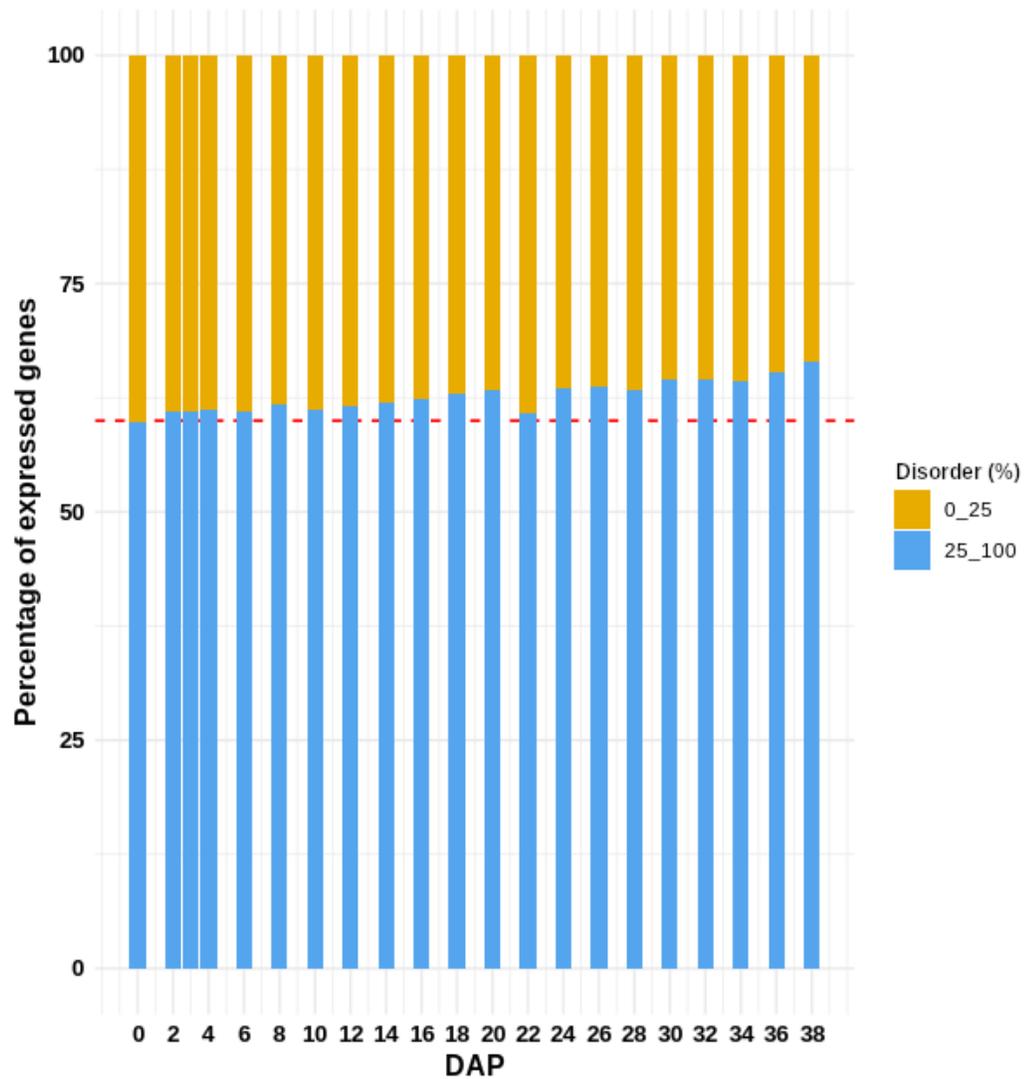


Figura 3.5. Dinámica en la composición relativa de los transcritos expresados a lo largo del desarrollo de semillas en función del porcentaje de desorden. Existe una clara reducción en la composición relativa de proteínas estructuradas. Los datos atípicos observados en el día 22 DAP coinciden con el pico de llenado de grano de las semillas.

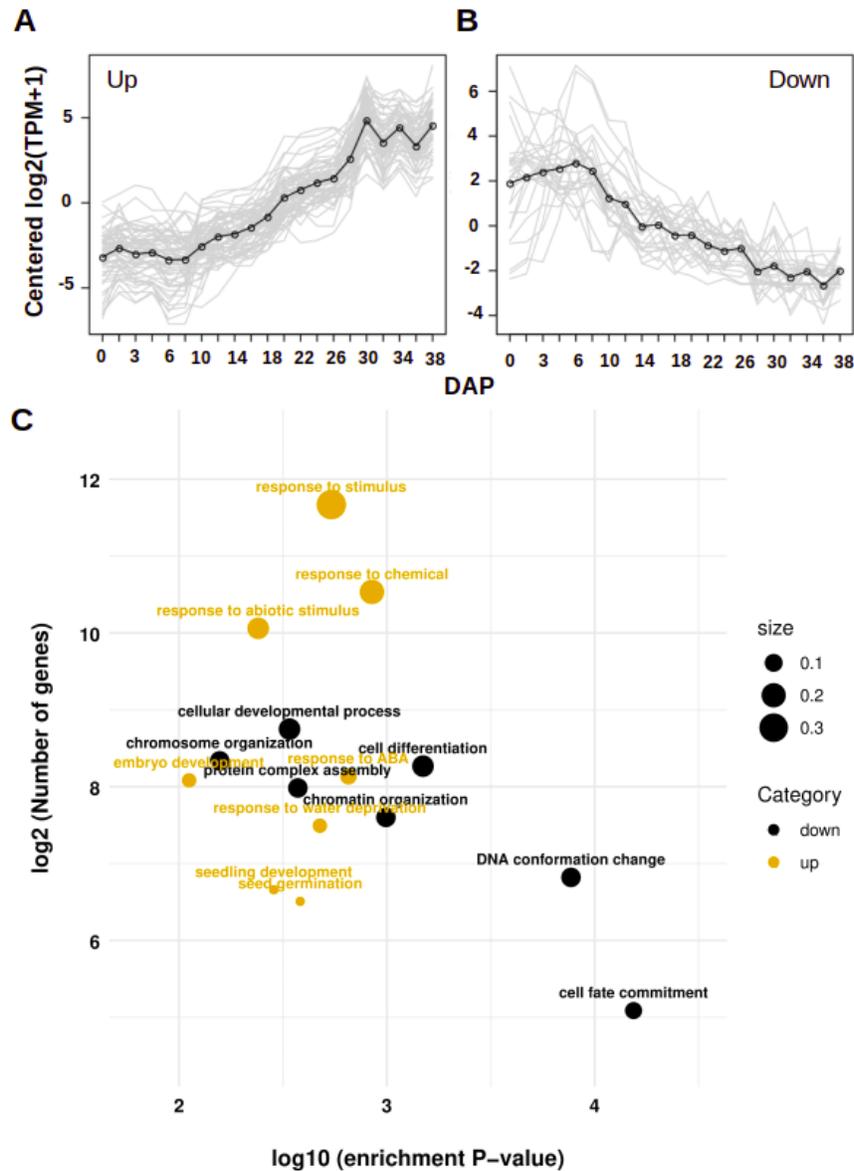


Figura 3.6 Perfil de expresión génica de transcritos que codifican para IDPs con más del 75% de desorden en su secuencia. A) y B) Hay un grupo de transcritos que es inducido y otro grupo que es reprimido durante todo el período evaluado. C) Análisis de enriquecimiento funcional para cada grupo de transcritos regulados hacia arriba y hacia abajo y que codifican para IDPs altamente desordenadas. El tamaño es proporcional al número de genes enriquecidos entre el número total de sus genes asociados a cada término GO.

Entre los transcritos que codifican IDPs, se identificaron 193 transcritos con regulación positiva que codifican proteínas altamente desordenadas (>75% de desorden proteico). En

contraste, se identificaron 157 transcritos con regulación negativa que codifican para este tipo de proteínas (Figura 3.6 A y B). Por su parte, el análisis de enriquecimiento GO de ambos conjuntos de proteínas refleja el cambio en el metabolismo que ocurre durante el desarrollo de la semilla de maíz (Figura 3.6 C). Las ontologías de genes asociadas con los genes regulados hacia arriba son claramente diferentes de las asociadas con los genes regulados hacia abajo. En el grupo regulado a la baja, se enriquecen varios términos GO asociados con el desarrollo y la diferenciación celulares. Entre estos términos GO, algunos son representativos, como por ejemplo "Cellular developmental process", "Chromosome organization", "Cell differentiation" y "DNA conformation change". El conjunto de transcritos sobre-regulados muestra el enriquecimiento de los términos GO asociados con la adaptación de la semilla a las condiciones de desecación, los cuales se hacen evidentes. Algunos procesos sobre-representados son "response to abiotic stimulus", "Response to water deprivation", "Response to ABA", "Seed germination" y "Embryo development".

En el grupo de genes regulados hacia abajo, se identificaron genes que codifican proteínas desordenadas involucradas en la regulación del DNA y la regulación del RNA como las histonas H1 (Zm00001d013066 y Zm00001d013067), H1a (Zm00001d018981 y Zm00001d034479), H2A (Zm00001d044246), la proteína de alta movilidad (Zm00001d044246), la proteína de alta movilidad del grupo proteína 3 (Zm00001d051427), la proteína de alta movilidad de la familia A (Zm00001d032239), el factor de ensamblaje de nucleosomas / cromatina D (Zm00001d052749), la subunidad 3 del complejo H (Zm00001d051116), o la subunidad 1 de la ribonucleoproteína del complejo ACA (Zm00001d052952). También se identificaron varios transcritos que participan en la señalización celular y los TFs involucrados en el desarrollo, tales como la proteína de unión a Calmodulina (Zm00001d002630 y Zm00001d028841), Calreticulina-2 (Zm00001d005460), la proteína de unión a Calmodulina 2 (Zm00001d038838), el factor protodérmico 1 (Z400001d040435), la proteína de nodulina 75 (Zm00001d031878), el factor de transcripción HMG 13 (Zm00001d021433), el factor de transcripción Homeobox 41 (Zm00001d017422) y el factor de transcripción WRKY 48 (Zm00001d015515).

Mientras tanto, en la lista de genes sobre-regulados que codifican para proteínas altamente desordenadas, se identificaron varios TF implicados en la respuesta al estrés y la latencia,

incluidos bHLH-167 (Zm00001d003677), MYB-139 (Zm00001d005300), bZIP-91 (Zm00001d007042), MYB- relacionado-111 (Zm00001d026017), proteína de unión a DRE 1 (Zm00001d032295), bZIP-29 (Zm00001d034571), bHLH87 (Zm00001d038863), ZF-HD-9 (Zm00001d044662) y bZIP (Zm00001d025), así como varias proteínas involucradas en la maduración de la semilla, la protección de los componentes celulares y la respuesta al estrés hídrico, como el homólogo de la proteína asociado a la latencia 3 (Zm00001d047503), LEA 3 (Zm00001d043709 y Zm00001d038870), Xero 1 (Zm00001d043730), las proteínas de maduración de la semilla (Zm00001d04, Zm00001d024414, Zm00001d033782 y Zm00001d035000) y la proteína estructural de la pared celular rica en glicina (Zm00001d017033).

Utilizando el análisis de redes de co-expresión, se identificó un número importante de factores de transcripción clave formando grupos reguladores de transcritos. Como se esperaba, la mayoría de los TFs identificados que forman grupos tienen un cierto nivel de desorden estructural (Figura 3.7). Aunque durante la fase intermedia hay un aumento en el contenido total de transcritos que codifican proteínas estructuradas, en este período no se identificó ningún grupo de TF estructurado que formara grupos de regulación. Del mismo modo, el desorden estructural parece jugar un papel preponderante en algunas familias de TFs, incluidas las familias AP2, NAC, bZIP, HSF, Opaque 2 y Myb.

Finalmente, en este análisis se identificaron varios TFs de las familias NAC, Ethylene insensitive, HSF y bZIP como reguladores clave de la fase tardía. En general, los datos generados en este capítulo sugieren que hay un ajuste activo en el contenido del desorden durante el desarrollo de la semilla de maíz.

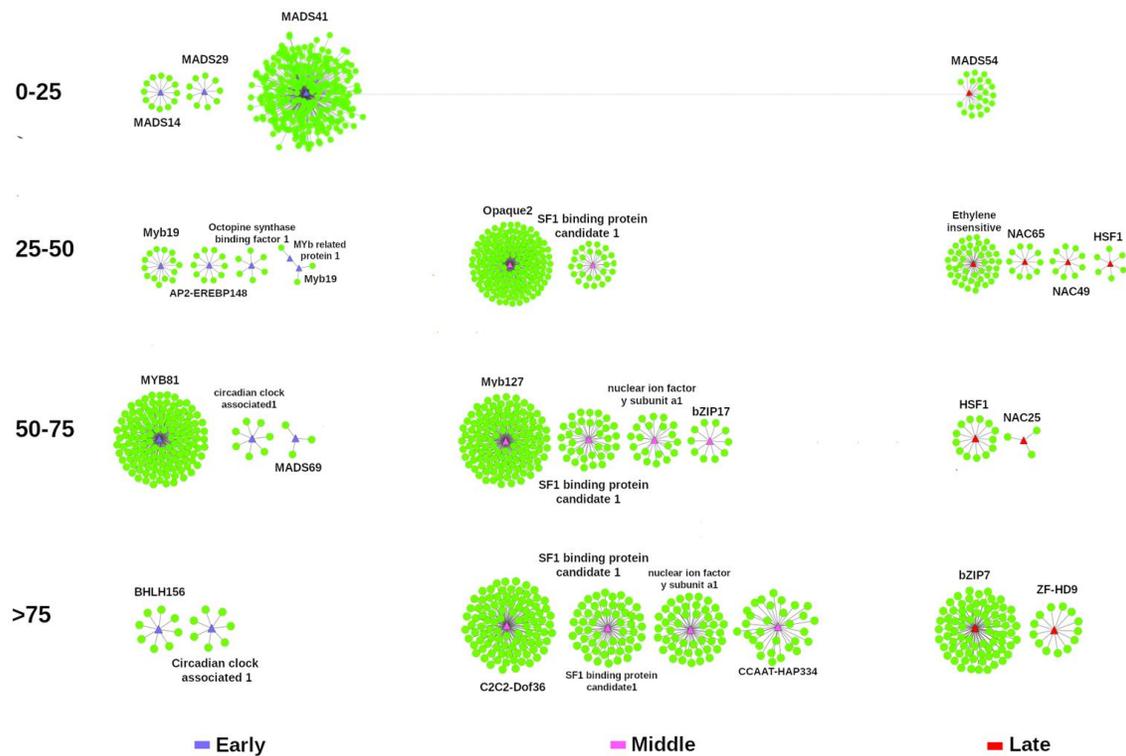


Figura 3.7 Análisis de redes de co-expresión utilizando como reguladores los TFs regulados de manera positiva. Los TFs expresados diferencialmente de cada fase se usaron para inferir los TFs clave de cada etapa, de acuerdo con su contenido de desorden. En la etapa inicial, las familias de TFs están asociadas con el proceso de desarrollo. Durante la fase intermedia, se inducen los TFs clave para controlar los procesos sintéticos, como la acumulación de proteínas de almacenamiento. Al comienzo de la fase tardía, las familias de TFs inducidas están especializadas de manera preponderante en el control de los mecanismos de adquisición de desecación. Sólo los TFs con más de dos genes diana se incluyeron en la figura.

3.4 DISCUSIÓN

Una característica potente de los estudios transcriptómicos es la capacidad de sacar conclusiones de los mecanismos que están siendo regulados bajo una condición biológica definida. Un enfoque complementario sería analizar la expresión génica en relación con las propiedades conformacionales predichas de las proteínas que codifican. Sin embargo, el uso de este enfoque no está tan explotado (Arvidsson y Wright, 2018). Este tipo de información no es obvia y debe cuantificarse. La consideración de las características

estructurales globales de las proteínas que se sintetizan bajo una condición dada podría dar información adicional sobre los mecanismos de adaptación utilizados por los organismos para responder a su entorno. Por ejemplo, recientemente se corroboró que las proteínas cuyos niveles de expresión aumentan por el calor en *A. thaliana*, se enriquecen en aa cargados y tienen una baja proporción de aa polares e hidrofóbicos, en comparación con las proteínas des-reguladas (Álvarez-Ponce *et al.*, 2018). La regulación de los tipos de proteínas en función de su composición aa podría representar un tipo de respuesta adaptativa a los cambios ambientales que no se ha tenido en cuenta previamente.

Hay suficientes ejemplos de la acumulación de IDPs en respuesta a la pérdida de agua, como para sugerir que el aumento del desorden estructural es parte de las respuestas al estrés hídrico (Chakrabortee *et al.*, 2007, 2010; Sun *et al.*, 2013; Boothby *et al.*, 2017; Boothby y Pielak, 2017; Wetzler *et al.*, 2018). Por ejemplo, en las semillas ortodoxas, las proteínas LEA son muy abundantes durante las últimas etapas del desarrollo de las semillas de las plantas cuando el embrión se vuelve tolerante a la desecación y su inducción coincide con el inicio de la tolerancia a la desecación (Sun *et al.*, 2013; Zamora-Briseño y de Jiménez, 2016). Anteriormente, se señaló que existe una fuerte correlación entre la desecación y el desorden intrínseco en las proteínas, ya que las IDPs están involucrados en la vitrificación, el reemplazo de agua, el blindaje molecular, la estabilización membranal, la preservación de la organización y estructura celular, la retención de agua y la eliminación de especies reactivas de oxígeno (Boothby y Pielak, 2017). Otra medida factible que las células pueden implementar para reducir los efectos de la falta de agua es la reducción de proteínas que son más sensibles a la pérdida de función mediada por la falta de agua. Con base a estas ideas, en este capítulo se analizó la regulación de la expresión de transcritos de acuerdo con el contenido de desorden de las proteínas para las que codifican a lo largo del desarrollo de la semilla de maíz, incluido el comienzo de la fase tardía.

Durante el desarrollo de la semilla de maíz hay una progresión de un programa morfogénico que incluye la adquisición de tolerancia a la desecación al final del programa de desarrollo. Esto es acompañado de una fuerte reducción en el contenido de agua. Curiosamente, en este trabajo se encontró que esto también se acompaña de una

reducción en el número de genes expresados que codifican proteínas estructuradas. Esto representa la primera evaluación global de la expresión de los transcritos expresados de acuerdo con el desorden estructural de sus proteínas codificadas en un contexto de desarrollo de semillas y reducción de agua.

Primero se describieron las características generales de las diferentes categorías en las que las proteínas se dividieron de acuerdo con su contenido de desorden. Es interesante observar que existe una relación clara entre el nivel de desorden intrínseco y la longitud de la proteína. Previamente se había descrito que el porcentaje de ocurrencia de la mayoría de los tipos de residuos depende significativamente de la dimensión de la proteína, pero la razón de esta dependencia no fue explicada (Carugo, 2008). La especialización funcional de las IDPs con respecto a las proteínas estructuradas puede ayudar a explicar tales observaciones.

El análisis de enriquecimiento ontológico mostró que la especialización de proteínas depende del nivel de desorden de las proteínas. Las IDPs están especializadas en actividades reguladoras, mientras que las proteínas estructuradas se enriquecen en funciones catalíticas. Por ejemplo, el desorden es menor en enzimas, ya que muchas de estas proteínas son estructuradas (Yizhi *et al.*, 2018). Este sesgo funcional puede explicarse si se considera que el desorden intrínseco es una consecuencia directa de la particular composición de aa. Este sesgo composicional es un atributo común de las IDPs (Uversky *et al.*, 2000; Williams *et al.*, 2001; Iakoucheva *et al.*, 2002; Uversky, 2013). Las IDPs generalmente están enriquecidas en Arg, Gln, Glu, Lys, Pro, y Ser, mientras que son deficientes en Cys, Ile, Leu, Phe, Trp, Tyr, y Val (Romero *et al.*, 2001). Estos datos están de acuerdo con las predicciones, las cuales muestran una clara diferencia para cada categoría de desorden.

Es interesante observar que los transcritos que codifican proteínas estructuradas se inducen activamente durante la fase de llenado de la semilla (etapa R2 y R3) y coinciden con la síntesis y la acumulación de moléculas de reserva (Chen *et al.*, 2014). Durante esta fase, la conversión sacarosa y el uso de aa para la síntesis de almidón y proteínas de almacenamiento, respectivamente, puede representar alrededor del 90% de la materia

seca total del grano, siendo el llenado de grano más rápido entre los 21 y los 25 DAP, independientemente del fondo genético de la semilla (Méchin *et al.*, 2004; Prioul *et al.*, 2008; Liu *et al.*, 2011; Jin *et al.*, 2013).

Un fenómeno que caracteriza este período es la acumulación de zeínas en el endospermo. Las zeínas son una clase de proteínas de almacenamiento que se han clasificado como zeínas α , β , γ y δ en función de su solubilidad y de sus propiedades estructurales (Esen 1987; Woo *et al.*, 2001; Coleman *et al.*, 2007; Wilson, 2008). El grupo de las proteínas α -zeínas es el más grande entre ellos y conforma el ~80% del total de proteínas tipo zeínas (Li y Vasal, 2015). Las α -zeínas son proteínas estructuradas altamente hidrosolubles. inusualmente ricas en aminoácidos hidrofóbicos, incluidos leucina, alanina y prolina, junto con la glutamina la cual tiene carácter hidrofílico (Coleman *et al.*, 2007; Wilson, 2008; Meng y Cloutier, 2014). Por ejemplo, las zeínas de 22kDa y 19kDa están conformadas por nueve hélices contiguas, topológicamente antiparalelas agrupadas dentro de un cilindro distorsionado (Song *et al.*, 2001). En este trabajo se encontró que una gran parte del aumento en el contenido de los transcritos que codifican proteínas estructuradas puede explicarse por el increíble aumento en la expresión de los genes de la familia zeína.

Por otro lado, el control transcripcional de los genes que codifican las proteínas zeína de 22 kD está mediado por opaque-2 (o2), un TF bien conocido que es un regulador transcripcional clave durante la fase intermedia (Schmidt *et al.*, 2007). Curiosamente, Opaque-2 (Zm00001d018971) fue identificado como un regulador de varias zeínas en los análisis de redes, lo que ayuda a respaldar las predicciones.

Por otro lado, el número de genes expresados que codifican proteínas estructuradas muestra una reducción continua después de la fase media. Al comienzo de la fase de maduración (etapa R5), hay un subconjunto de proteínas altamente desordenadas que se induce activamente. Sin embargo, la regulación de estos genes comienza en la etapa R4 que le precede. La regulación del programa de desecación comienza inmediatamente después del final de la fase de síntesis más activa, antes de que la adquisición de tolerancia a la desecación se convierta en una característica visible. Esta es una fase que puede considerarse como preparatoria. Antes de la etapa R5, los granos a lo largo de la mazorca

comienzan a secarse en la parte superior. Ya en la etapa R5, los granos comienzan a desecarse y el contenido de humedad se reduce severamente, llegando a menos del 50% (Hanway, 1966; Zamora-Briseño *et al.*, 2018). En esta etapa, la tasa de llenado de grano decrece y culmina en la madurez fisiológica y la formación de capas negras, lo que a su vez está relacionado con la latencia de las semillas (Prioul *et al.*, 2008).

Además, estos datos sugieren que hay una reducción continua en el número de transcritos que codifican para proteínas estructuradas. Curiosamente, hay un pico atípico en el contenido relativo de este tipo de transcritos a los 22 DAP y coincide con la tasa máxima de acumulación de almidón, la cual ocurre a los 21 DAP (Prioul *et al.*, 2008; Jin *et al.*, 2013).

La etapa inicial involucra la división celular, después de lo cual las células del endospermo se agrandan y como resultado de varios procesos metabólicos adquieren proteínas de almacenamiento y almidón (Gibbon y Larkins, 2005). En la fase final de la embriogénesis, se adquiere tolerancia a la desecación y se establece la latencia (Devic y Roscoe, 2016). Por lo tanto, el análisis de expresión diferencial es muy útil para definir las IDPs asociadas con el primer proceso de desarrollo y con el comienzo de la fase de madurez. Dentro de estos dos tipos de transcritos, la especialización de las IDPs puede diferenciarse fácilmente mediante las ontologías génicas asociadas a procesos claramente diferentes. Si bien los transcritos que codifican para IDPs y que son regulados negativamente están dedicados a procesos asociados con el desarrollo, los inducidos están involucrados en funciones relacionadas con el estrés abiótico y las respuestas a estímulos ambientales.

Dentro de los transcritos que codifican para IDPs identificados en la primera categoría, hay varias proteínas asociadas con el empaquetamiento de ácidos nucleicos, como las histonas y las proteínas del grupo de alta movilidad (HMG). Es un hecho reconocido que muchas clases funcionales de proteínas dependen en gran medida del desorden. Entre estas funciones se encuentran la interacción con ácidos nucleicos y el ensamblaje de complejos de proteínas. Anteriormente, se demostró que todos los miembros de la familia de las histonas poseen IDRs. De hecho, el desorden intrínseco es absolutamente necesario para diversas funciones de las histonas, desde la heterodimerización hasta la formación de oligómeros de orden superior, las interacciones con el DNA y otras proteínas y las

modificaciones postraduccionales (Peng *et al.*, 2012). Las proteínas HMG son proteínas nucleares que se unen de forma transitoria a los nucleosomas, cambian la arquitectura local de la fibra de cromatina y afectan varias actividades relacionadas con el DNA, como la transcripción, la replicación y la reparación del DNA (Reeves, 2009). Las HMG son proteínas altamente desordenadas y se enrollan principalmente en forma aleatoria en solución, pero se someten a plegamiento al unirse con DNA u otros interactores (Love *et al.*, 2004; Slama-Schwok *et al.*, 2000).

Además, se identificó un número importante de transcritos regulados negativamente y que codifican proteínas involucradas en el desarrollo de plantas y la señalización mediada por calcio, como la proteína reprimida por auxina (ARP) y la calreticulina. Han surgido pruebas acerca de que los miembros de la familia ARP son IDPs. Los genes ARP responden a las hormonas involucradas en las respuestas al estrés biótico, como el ácido salicílico (SA) y el metil jasmonato (MeJA), así como a las hormonas que regulan el crecimiento y desarrollo de las plantas, incluidas las auxinas (Shi *et al.*, 2013; Zhao *et al.*, 2014; Wu *et al.*, 2017). La calreticulina (CALR) es reconocida como una chaperona molecular de la proteína de unión al Ca^{2+} que ayuda al plegamiento de las glucoproteínas recién sintetizadas y modula la homeostasis del Ca^{2+} en el lumen del retículo endoplásmico (RE) (Borisjuk *et al.*, 1998). En el RE, CALR se une no sólo a Ca^{2+} , sino que también interactúa con muchas proteínas del RE y con algunos mRNAs, lo que a su vez determina su destino. Las funciones potenciales de CALR son tan numerosas que la identificación de todas ellas es una tarea compleja. Curiosamente, su versatilidad funcional es proporcionada por sus propiedades de desorden intrínseco (High *et al.*, 2000; Michalak *et al.*, 2009; Qiu *et al.*, 2012; Varricchio *et al.*, 2017).

En el caso de los genes regulados positivamente, se identificó un número importante de familias de IDPs, como las proteínas LEA, las deshidratinas, las proteínas ricas en glicina, las proteínas ricas en prolina (PRP) y las proteínas de maduración de semillas (SMP), todas las cuales están implicadas en la DT y la capacidad de almacenamiento de semillas (Delahaie *et al.*, 2013; Dure y Galau, 1981; Hundertmark y Hincha, 2008). Por ejemplo, las proteínas LEA, especialmente las del grupo 3, son IDPs altamente hidrofílicas, cuya expresión está asociada con la adquisición de tolerancia a la desecación en semillas

durante la maduración (Chakrabortee *et al.*, 2007; Hundertmark e Hinch, 2008; Olvera-Carrillo *et al.*, 2011). En maíz, se descubrió que el gen LEA3 responde tanto a ABA, como a la hiperosmolaridad y luego se demostró que puede reducir los efectos de la pérdida de volumen celular bajo condiciones de deshidratación (Thomann *et al.*, 2008). Por otro lado, las PRPs son IDPs que se identificaron primero como proteínas que se acumulan en la pared celular en respuesta al daño físico y se expresan durante el desarrollo de la hoja, el tallo, la raíz y la cubierta de la semilla, el crecimiento de las plántulas y en los tipos de células asociados con la lignificación (Kishor, 2015). Funcionalmente, las PRPs se insolubilizan y reticulan en las paredes celulares con la participación de ROS y juegan un papel en la integridad estructural de las células vegetales, participan en actividades relacionadas con la defensa y median interacciones de la superficie celular de las plantas (Kishor, 2015).

Por su parte, las proteínas ricas en glicina regulan diversos procesos celulares en el desarrollo de las plantas y la respuesta al estrés. En tabaco, las *NtGRP* están altamente reguladas bajo estrés osmótico, mientras que *AtGRP* y *OsGRP* se han identificado actuando como chaperonas de RNA que regulan la exportación de mRNAs desde el núcleo al citoplasma en condiciones de estrés. Además, las proteínas GRP son importantes para la regulación de algunos pasos en el procesamiento postranscripcional de RNA como el empalme y la poliadenilación (Park *et al.*, 2009; Kim *et al.*, 2010; Lu *et al.*, 2019). La detección de estas reconocidas IDPs, cuya participación en DT ha sido ampliamente estudiada, ayuda a respaldar las predicciones de desorden de este trabajo. Sin embargo, hay otras IDPs que se inducen en esta fase, como las remorinas y las metalotioneínas que se han asociado con la tolerancia a la sequía (Benešová *et al.*, 2012; Checker y Khurana, 2013; Jaiswal *et al.*, 2018) y cuya información sobre su relación con la DT sigue siendo limitada, por lo que se debe abordar un análisis más detallado de sus implicaciones en la adquisición de tolerancia a la desecación.

En el análisis de co-expresión, se detectaron varios TFs clave para cada etapa de desarrollo, cada uno con diferentes niveles de desorden. Los TFs identificados como reguladores clave en la etapa inicial de desarrollo de la semilla se especializan principalmente en la coordinación de procesos relacionados con el desarrollo y la

diferenciación celulares, como por ejemplo los MADS y MYB. Los MADS están involucrados en el control de muchos procesos de desarrollo en plantas con flores, que van desde el desarrollo del polen y el saco embrionario hasta el desarrollo de raíces, flores y frutos (Theißen y Gramzow, 2016). Las proteínas MYB son clave en las redes reguladoras que controlan el desarrollo, pero también el metabolismo y las respuestas al estrés biótico y abiótico (Dubos *et al.*, 2010).

En la fase intermedia, se identificaron varios TFs intrínsecamente desordenados que al parecer son clave en la acumulación de reservas de almacenamiento. Dentro de este grupo, opaque2 es particularmente importante, debido a su rol en la regulación de la acumulación de proteínas de reserva y su increíble capacidad para integrar el metabolismo de N y C (Hartings *et al.*, 2011). Opaque 2 es un TF tipo bZIP y se expresa específicamente en el endospermo, regulando la activación de los genes de α -zeína de 22 kDa y la β -zeína de 15 kDa. Opaque2 también regula directa o indirectamente una importante cantidad de genes que codifican proteínas con funciones adicionales a las de almacenamiento, como la lisina-cetoglutarato reductasa y la proteína de choque térmico 70 (Motto *et al.*, 1997; Brochetto-Braga *et al.*, 2008). En el análisis de redes de co-expresión, estas y otras proteínas relevantes, tales como el receptor de ABA, PYL5, el TF NAC87, la proteína de la subunidad ribosomal 60S L7-1, la proteína de la subunidad ribosomal 40S S24, la giberelina-2 oxidase 1, la aconitasa 1, el TFDII subunidad 9, entre otras, fueron predichas como reguladas por el TF opaque2.

Durante la fase tardía, la mayoría de los reguladores clave identificados pertenecen a la familia NAC. Los NACs comprenden una de la familia más grande de TFs en plantas (Nuruzzaman *et al.*, 2013; Pereira-Santana *et al.*, 2015). Estos TFs juegan un papel vital en las complejas redes de señalización durante las respuestas al estrés en plantas. Parece que una proporción importante de los genes NAC funcionan en la respuesta al estrés de acuerdo con los datos de expresión de los análisis de expresión global en muchas plantas (Fang *et al.*, 2008; Le *et al.*, 2011; Hu *et al.*, 2015). Curiosamente, el factor de choque térmico 1, -HSF1- (Zm00001d005888, 50-75% de desorden), parece estar involucrado en la regulación del estrés oxidante, ya que muchas proteínas relacionadas con el estado redox fueron identificadas como reguladas por este TF. Por ejemplo, en este grupo se

encuentran las proteínas inducidas por heridas y por estallidos oxidantes, como la proteína del subgrupo III de glutaredoxina, la glutaredoxina y la HSP de clase I de 17.4 kDa. En tomates se descubrió que el HsfA1 es el regulador maestro en respuesta al estrés por choque térmico (Mishra *et al.*, 2002) y se ha propuesto que algunos miembros de la familia HSF pueden actuar como sensores redox (Miller y Mittler, 2006).

Por otro lado, el TFs bZIP7 parece jugar un papel esencial en esta etapa de desarrollo. Además, hasta donde sabemos, este gen no se ha caracterizado funcionalmente. En general, pocos genes bZIP de maíz se han caracterizado funcionalmente (Ying *et al.*, 2012). En las plantas, los bZIP regulan diversas funciones, incluidos procesos como el desarrollo y la respuesta al estrés. En el análisis de redes, el bZIP7 aparece como regulador de genes implicados en la tolerancia a la desecación, como las proteínas ricas en glicina, proteínas ricas en hidroxiprolina, GRAS46, metalotioneína 2, glutatión transferasa 19, glutatión transferasa 37, proteína rica en prolina específica de anteras, cupinas, SMP, etc. Entonces, un tema interesante podría ser caracterizar el papel de este TF o sus homólogos. Desafortunadamente, este trabajo está limitado por la falta de datos sobre el perfil transcriptómico de la etapa R6, donde la maduración de las semillas termina y el proceso de desecación de la cariósida concluye. Se deben realizar análisis adicionales durante el final de la fase de embriogénesis tardía para descubrir el patrón global de desorden en la etapa R6. Sin embargo, con los datos obtenidos es posible proponer un modelo general sobre las alteraciones que ocurren durante el ajuste del nivel de desorden del proteoma como mecanismo de tolerancia a la desecación en semillas ortodoxas. En este modelo, el contenido de proteínas estructuradas es mayor durante las etapas metabólicas más activas del desarrollo de la semilla, justo cuando el contenido de agua es mayor. Posteriormente, hay una reducción gradual en el número de proteínas estructuradas, así como en su contenido total. Al mismo tiempo, el contenido relativo de IDPs aumenta y con esto, la proporción de proteínas menos propensas a perder su función debido a la pérdida de estructura. Este mecanismo puede facilitar el mantenimiento de la integridad del proteoma, debido al hecho de que habría una menor cantidad de proteínas susceptibles de perder su estructura espacial y su función molecular. La visión general de tales ideas se resume en la Figura 3.8.

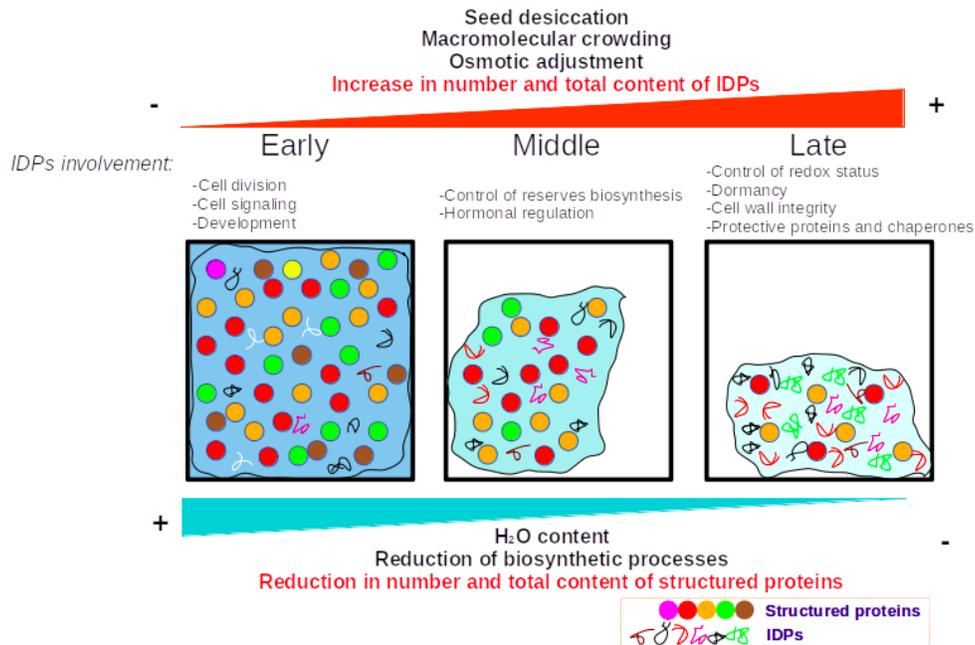


Figura 3.8 Modelo del ajuste del desordenoma durante el desarrollo de semillas ortodoxas. Al comienzo del desarrollo de la semilla, la síntesis de compuestos de reserva es muy activa, el agua disponible es abundante y el contenido de proteínas estructuradas es alto (A). Con el progreso del programa de desarrollo, se produce una reducción importante en el contenido de agua, la disminución del volumen celular y el hacinamiento macromolecular aumentan. Este efecto provoca interacciones no deseadas entre proteínas, pérdida de estructura terciaria y agregación de proteínas. Para reducir este efecto, las células activan mecanismos protectores (como la síntesis de solutos compatibles y la inducción de chaperonas). Sin embargo, se propone que también hay una reducción en el contenido de las proteínas más lábiles, sin reemplazo de proteínas similares. En contraste, el contenido de IDPs aumenta y todas las proteínas sintetizadas *de novo* son también IDPs (B).

3.5 CONCLUSIÓN

Los datos obtenidos permiten establecer una relación entre la modulación global del desordenoma y la progresión del desarrollo de la semilla, incluido el inicio del programa de tolerancia a la desecación. Hay evidencia para proponer que el inicio del programa de desecación en las semillas provoca una reducción de las proteínas que son más propensas a perder su función en un entorno que limita el agua, acompañado de un aumento en el contenido de IDPs.

CAPÍTULO IV

CARACTERIZACIÓN DEL PROTEOMA TERMO-ÁCIDO-SOLUBLE DE HOJAS DE CAÑA DE AZÚCAR *Saccharum officinarum* L. EN RESPUESTA AL ESTRÉS POR FALTA DE AGUA

4.1 INTRODUCCIÓN

Hoy en día se ha descrito un importante conjunto de proteínas que carecen de una estructura tridimensional determinada, a pesar de ser funcionales. Estas proteínas se conocen como proteínas intrínsecamente desordenadas (IDPs; van der Lee *et al.*, 2014). Las IDPs a menudo están asociadas a regiones de baja complejidad, se caracterizan por poseer largos segmentos conformados primordialmente por residuos cargados e hidrofílicos y con bajo contenido de aa hidrofóbicos y que promueven el orden, tales como cisteína y triptófano, los cuales normalmente están sub-representados o ausentes en estas proteínas (Dunker *et al.*, 2008; Sun *et al.*, 2013). Esta particular composición de aminoácidos impide la formación de núcleos hidrofóbicos que faciliten su posterior plegado.

Actualmente, se estima que hasta un tercio de las proteínas eucariotas son desordenadas y hasta un 50% contienen al menos un segmento desordenado mayor a 30 residuos en su secuencia (Uversky y Dunker, 2010). Dentro de las IDPs se encuentran proteínas que desempeñan funciones clave en la célula, tales como la regulación transcripcional, la señalización y el control de las respuestas al estrés (Pazos *et al.*, 2013; Pietrosemoli *et al.*, 2013). En las plantas, estas funciones cobran especial relevancia puesto que su naturaleza sésil les impide moverse ante situaciones ambientales desfavorables, lo que a su vez ha obligado a que las plantas hayan desarrollado una amplia batería de mecanismos de respuestas adaptativas ante tales condiciones. En este sentido, muchas de las proteínas que forman parte de los mecanismos de percepción, regulación y respuesta al estrés forman parte del universo de las IDPs. En plantas, quizá el ejemplo más conocido de este tipo de proteínas son las proteínas LEA y las dehidrinas (Battaglia y Covarrubias, 2013; Covarrubias *et al.*, 2017).

Aunque experimentalmente existen diversas aproximaciones para determinar la falta de estructura proteica (Tompa, 2002) , hoy en día sigue siendo un reto poder hacer una caracterización a escala proteómica de las IDPs. De hecho, la mayor parte de la evidencia disponible sobre IDPs se basa en predicciones bioinformáticas, que, si bien han permitido generar un conjunto importante de datos acerca de las funciones en las que estas proteínas están involucradas, la información experimental que soporta esta evidencia sigue siendo necesaria (Tantos y Tompa, 2012). Cabe indicar que en el caso particular de plantas, incluso los trabajos en los que se han empleado aproximaciones bioinformáticas para predecir las IDPs a una escala global siguen siendo escasos (Pietrosemoli *et al.*, 2013; Vincent y Schnell, 2016; Zamora-Briseño *et al.*, 2019).

Dentro de las aproximaciones experimentales más empleadas para poder diferenciar entre proteínas globulares e IDPs, está el uso de condiciones y/o agentes desnaturizantes, tales como la temperatura y el pH extremos, ya que las propiedades fisicoquímicas de las IDPs contribuyen a mantenerlas solubles bajo condiciones que normalmente hacen precipitar a las proteínas globulares (Irar *et al.*, 2006; Campos *et al.*, 2011; Amara *et al.*, 2012; Tantos y Tompa, 2012; Yizhi *et al.*, 2018). De hecho, se ha sugerido que una de las propiedades distintivas de las IDPs es su habilidad para permanecer solubles bajo tales condiciones (Campos *et al.*, 2011) . Este tipo de ensayos ha sido ampliamente empleado en diferentes modelos, aunque en plantas su uso se restringe al análisis de proteínas de la semilla (Boudet *et al.*, 2006; Galea *et al.*, 2009; Oliveira y Fávoro, 2010; Campos *et al.*, 2011; Amara *et al.*, 2012; Zamora-Briseño y de Jiménez, 2016; Liu *et al.*, 2017; Zhang *et al.*, 2018). A pesar de lo escaso de este tipo de trabajos en plantas, se ha observado que hay un número importante de IDPs que están involucradas en la adquisición de la tolerancia a la desecación, lo cual coincide con lo reportado para otros modelos no vegetales (Boudet *et al.*, 2006; Amara *et al.*, 2012; Liu *et al.*, 2017).

Por otro lado, los agregados proteicos que se generan como consecuencia de ciertos tipos de estrés, están conformados por proteínas intrínsecamente susceptibles a la agregación, lo que en última instancia, termina por afectar su función (Weids *et al.*, 2016). Además, se ha propuesto que la respuesta a la falta de agua incluye la inducción de proteínas con características hidrofílicas y resistentes a la desnaturización (Garay-Arroyo *et al.*, 2000;

Chakrabortee *et al.*, 2007; Zamora-Briseño *et al.*, 2018), lo que puede implicar que las proteínas que forman parte de la respuesta a estrés sean proteínas menos susceptibles para formar agregados insolubles, con lo que de esa forma evitan comprometer su función.

Por todo lo anterior, resulta lógico pensar que existe un subconjunto de proteínas que se inducen ante el estrés por falta de agua que tienen como características el ser hidrofílicas, desordenadas y/o resistentes a la formación de agregados proteicos. La evaluación proteómica de este sub-conjunto de proteínas puede resultar en el hallazgo de nuevas proteínas que participan en la respuesta a estrés, puesto que al enriquecer los extractos proteicos con proteínas que son solubles en condiciones de choque térmico y pH ácidos (proteínas termo-ácido-solubles) resulta posible la identificación de proteínas que normalmente no son identificadas porque son enmascaradas por proteínas altamente abundantes como es el caso de la ribulosa-1,5-bisfosfato carboxilasa oxigenasa (Rubisco), las cuales implican un importante desafío cuando se emplean enfoques proteómicos (Zhang *et al.*, 2013).

En este capítulo se decidió evaluar los cambios en las proteínas termo-ácido-solubles en respuesta a la falta de agua empleando un enfoque de proteómica libre de gel. Los resultados obtenidos permiten proponer que la respuesta a estrés por falta de agua incluye la activación de un subconjunto de proteínas que son resistentes a la precipitación en condiciones desnaturalizantes. Además, las proteínas diferenciales identificadas que responden a estrés pueden resultar útiles para mejorar la comprensión de las respuestas a la falta de agua y ser blancos moleculares putativos para el mejoramiento genético en esta especie. Cabe remarcar que, hasta donde sabemos, este estudio representa el primer análisis de proteínas termo-ácido-solubles en respuesta a la falta de agua en tejidos vegetativos de plantas.

4.2 MATERIALES Y MÉTODOS

4.2.1 Condiciones de crecimiento de las plantas

En el presente estudio se emplearon plántulas cultivadas *in vitro* de *Saccharum officinarum* cv 69-290, las cuales fueron multiplicadas empleando el medio Murashige y Skoog

suplementado con sacarosa (2 %) y BAP (0.2 mg·l⁻¹). Las plántulas fueron mantenidas con un fotoperiodo de 16 h luz/8 h oscuridad a una temperatura de 28°C, durante 28 días. Posteriormente, fueron dispuestas en fase de maduración, la cual consistió en tres ciclos de subcultivo de 45 días cada uno, hasta alcanzar la homogeneidad en términos de tamaño y vigorosidad. En esta fase, las plántulas fueron mantenidas en las mismas condiciones que las empleadas durante la multiplicación, omitiendo a BAP como suplemento. Una vez terminada la fase de maduración, se pasaron a condiciones *ex vitro*. Se eligieron plántulas de 5 cm de longitud (de pseudotallo) y con al menos 4 hojas. A las raíces se les dio un lavado con agua corriente para quitar los residuos del medio de cultivo, posterior a esto se dispusieron en una solución fungicida de Mancozeb y Amistar, ambos al 1%. Las plántulas se acomodaron en charolas de 30 cm de largo x 20 de ancho x 10 cm de alto con una profundidad de sustrato de 4 cm a capacidad de campo, por cada charola se dispusieron 10 plántulas. El sustrato empleado fue Berger M2. Las plántulas se acondicionaron en un cuarto de cultivo durante 15 días con un fotoperiodo 16/8 horas luz/oscuridad, a una temperatura de 28 °C ± 2 y con riego cada tercer día. Pasado este tiempo, las plantas se individualizaron en maceteros con el mismo sustrato, manteniendo las mismas condiciones. Se mantuvieron en crecimiento durante 45 días.

4.2.2 Tratamiento de estrés, determinación de parámetros fotosintéticos y obtención de proteínas termo-acido-solubles

Se manejaron 12 plantas por condición (tiempo 0 y 10 días de estrés hídrico). El tratamiento por estrés hídrico consistió en suspensión del riego a las plantas durante 10 días. Al cabo de este tiempo, se estimaron determinados parámetros fisiológicos incluidos la tasa de fotosíntesis (A) y de transpiración (T), las cuales se midieron en la hoja bandera de cada planta. Para esto se empleó un sistema portátil Li-6400 (Li-Cor, Lincoln, NE, USA). Las determinaciones se realizaron en cuatro plantas diferentes.

Para el análisis proteómico se realizó la extracción de proteínas hidrosolubles a partir de la porción aérea de las plantas, incluyendo las hojas y el pseudotallo. Los análisis proteómicos de realizaron por triplicado. Para esto, los tejidos de cuatro plantas diferentes fueron triturados con N₂ líquido. Para la extracción de las proteínas hidrosolubles se usó un buffer de extracción (Tris pH 8.0 100 mM, NaCl 100 mM y EDTA 100 mM) adicionado con

inhibidores de proteasas Cocktail cOplete™ de Roche (1x). Se tomó 1.8 g de tejido macerado y se adicionó a un volumen de 25 mL de buffer. La mezcla se homogeneizó mediante vortex y se hirvió a una temperatura de 100°C durante 15 min y luego la mezcla se mantuvo en hielo durante 30 min. Pasado este tiempo, el homogeneizado se centrifugó a 20 000 x g por 10 min y 4 °C para retirar el debris. El sobrenadante se transfirió a un tubo nuevo y se repitió la centrifugación en las mismas condiciones. Se recuperó el sobrenadante y a este se le agregó lentamente ácido tricloroacético (TCA) hasta una concentración final del 3%. Durante la aplicación del TCA, la mezcla se mantuvo en agitación constante, en frío, tras lo cual la mezcla se incubó 10 min en hielo. Posteriormente, el homogeneizado se centrifugó a 20 000 x g por 10 min y 4 °C para retirar el precipitado. El sobrenadante se trasvasó y se le adicionó un volumen de acetona fría, se mezcló vigorosamente con vórtex y se incubó toda la noche a -20°C. Posteriormente, la mezcla se centrifugó a 20 000 x g por 10 min y 4°C. El sobrenadante se desechó, la pastilla se lavó con 5 mL acetona fría al 100% y luego se centrifugó a 15 000 x g por 10 min y 4 °C. Se desechó el sobrenadante. Este paso se repitió nuevamente. Posteriormente, las muestras se dejaron evaporar a temperatura ambiente por 20 min para remover la acetona. Finalmente, las proteínas se liofilizaron y se almacenaron a -70°C hasta su análisis proteómico.

4.2.3 Digestión enzimática de proteínas

Para el análisis proteómico, las proteínas se solubilizaron en buffer de resuspensión (1% SDS en buffer de fosfatos 100 mM, pH 7.5) y se cuantificaron con el paquete comercial micro BCA assay (Thermo Scientific ®). A partir de esto, se redujo un total de 100 µg de proteína con 10 mM de Tris (2-car-boxietil) fosfina (TCEP) durante 45 min a 60°C, se alquiló con 30 mM yodoacetamida (IA) durante 1 h a temperatura ambiente y en oscuridad. La IA restante fue inactivada con 30 mM DTT durante 10 min. Las proteínas fueron precipitadas con 4 volúmenes de acetona fría durante la noche a -20°C. La mezcla se centrifugó a 10,000 x g durante 15 minutos a 4°C y la pastilla se secó a temperatura ambiente durante 15 minutos. Luego, las proteínas se suspendieron con 100 µl de buffer de digestión (dicarbonato de trietilamonio 50 mM (TEAM) y SDS al 0,1%) y se digirieron con tripsina (Trypsin Gold, Mass Spectrometry Grade, Promega, Madison WI) en una proporción de tripsina/proteína de 1:30 p/p durante toda la noche a 37°C. Posteriormente, se adicionó

tripsina recién preparada en una proporción de 1:60 p/p durante cuatro horas más a 37°C para completar la digestión. La reacción se detuvo desnaturalizando la tripsina a 80°C durante 10 min.

4.2.4 Análisis por nano-LC-MS / MS

Las muestras se analizaron mediante nano-LC-MS/MS utilizando un espectrómetro de masas Orbitrap Fusion Tribrid TM (Thermo-Fisher Scientific®) equipado con una fuente de nano iones “EASY Spray” (Thermo-Fisher Científico) acoplado a un sistema UltiMate 3000 RSLC (Dionex ®). Cada muestra reconstituida (5 µL) se inyectó en una columna nanoviper de C18 (3 µm, 75 µm × 2 cm, Dionex ®) a una velocidad de flujo de 3 µL·min⁻¹ y luego se separó en una columna de EASY spray RSLC C18 (2 µm, 75 µm × 25 cm), utilizando un gradiente de 100 min con un caudal de 300 nL·min⁻¹ de ácido fórmico al 0.1% en agua grado LC-MS (disolvente A) y ácido fórmico al 0,1% en acetonitrilo al 90% (disolvente B). El gradiente fue el siguiente: 10 min de solvente A, 7–20% de solvente B dentro de 25 min, 20% de solvente B por 15 min, 20–25% de solvente B por 15 min, 25–95% solvente B durante 20 min y 8 min para solvente A. El espectrómetro de masas fue operado en modo de iones positivos con voltaje de nanoespray establecido en 3.5 kV y temperatura de la fuente a 280°C. Como calibrante externo, se incluyó cafeína, Met-Arg-Phe-Ala (MRFA) y Ultramark 1621.

4.2.5 Análisis bioinformático

Los datos de los espectros crudos se analizaron con el paquete Proteome Discoverer 2.1 (PD, Thermo Fisher Scientific Inc.). Las búsquedas posteriores fueron llevadas a cabo utilizando los motores de búsqueda Mascot (versión 2.4.1, Matrix Science, Boston, MA), SEQUEST HT (Eng *et al.*, 1994), AMANDA (Dorfer *et al.*, 2014) y Byonic (Bern *et al.*, 2012). Se seleccionaron sólo las proteínas con un score de 15 para al menos 1 motor de búsqueda. La búsqueda con los motores se realizó contra una base de datos personalizada, debido a que no se dispone todavía de un genoma de referencia para esta caña de azúcar. Para construir esta base de datos, se descargaron todas las secuencias proteicas disponibles en el NCBI para los géneros *Officinarum*, *Zea*, *Sorghum*, *Panicum*, *Oryza*, *Setaria*, *Aegilops*, *Brachypodium*, *Triticum*, *Dichantelium* y *Arundo*. Además, se descargaron las secuencias

de RNA disponibles para el género *Officinarum* y las secuencias ensambladas *de novo* en un trabajo previo de nuestro laboratorio usando este modelo (Pereira-Santana *et al.*, 2017). Estos transcritos se tradujeron con TransDecoder (Haas *et al.*, 2013) y las proteínas predichas se juntaron con las secuencias de las proteínas descargadas. Para reducir la redundancia de las secuencias, el archivo con formato fasta se depuró con CD-HIT (Li y Godzik, 2006), con valores de 0.95, 0 y 5, para los parámetros c, t y n, respectivamente. También se filtraron las secuencias proteicas de longitud menor a 31 aminoácidos y se eliminaron los residuos inespecíficos. El archivo resultante se empleó para preparar la base de datos personalizada con la cual alimentar a los motores de búsqueda empleados para resolver los espectros de masas.

Una vez identificadas las proteínas de cada condición, primero se analizó cuáles de las proteínas encontradas se compartieron entre tratamientos y cuáles no. Para ello se construyó un diagrama de Venn con el programa Venny 2.0 (Oliveros, 2007). Para poder obtener una idea general del grado de enriquecimiento de desorden intrínseco, las secuencias de las proteínas identificadas se analizaron con Espritz (Walsh *et al.*, 2012). Con los valores arrojados por esta herramienta, se construyó la distribución de desorden intrínseco predicho para el proteoma global que incluye al total de proteínas no redundantes identificadas en cada condición y se generaron las estadísticas generales de desorden intrínseco predicho. En este sentido, también se exploró el sesgo en la composición de aminoácidos de las proteínas identificadas. Para ello, se empleó el programa Composition Profiler (Vacic *et al.*, 2007). El programa se corrió con 10,000 bootstraps y con un alfa de 0.05. La base de referencia empleada contra la que se realizó la comparación fue el PDB select 25.

Por otro lado, se procedió a realizar la anotación funcional de las proteínas. Para ello, las proteínas se compararon con las proteínas de la familia Poaceae mediante BLASTp (Altschul *et al.*, 1990), esto debido a que muchas de las proteínas identificadas fueron secuencias predichas por TransDecoder a partir de secuencias del transcriptoma y por tanto eran secuencias sin anotar. Por otra parte, también se obtuvieron las ontologías, así como las anotaciones de los dominios pFam, todo estos mediante InterproScan 5 (Jones *et al.*, 2014).

Las ontologías generadas por InterproScan 5, fueron analizadas con WEGO (Ye *et al.*, 2006). En este caso, se seleccionaron sólo los términos GO de primer nivel. Todo lo anterior se complementó con las anotaciones GO generadas con PANNZER2 (Törönen *et al.*, 2018), con el que se anotaron por separado las proteínas identificadas exclusivamente en la condición control, las identificadas exclusivamente en estrés y las proteínas identificadas en ambas condiciones. En este caso, se seleccionaron todos los términos ontológicos identificados que tuviesen una puntuación mayor a 0.70 en la distribución hipergeométrica. En ambos casos, los resultados generados se exportaron y se graficaron con ggplot2 (Ginestet, 2011).

4.3 RESULTADOS

Las plantas de caña sometidas a falta de riego presentaron una clara reducción en su talla, una evidente pérdida de vigor, así como una importante reducción en sus parámetros fisiológicos, específicamente en la tasa de fotosíntesis (A), medida como asimilación de dióxido de carbono y en la tasa de transpiración (T) (Figura 4.1).

El hecho de que la tasa de asimilación resultara negativa es evidencia de que las plantas estresadas se encontraban realizando fotorespiración, lo que implica una disminución en la fijación de carbono. Este es un proceso dependiente de la luz, en el que se toma oxígeno molecular y se libera concomitantemente dióxido de carbono a partir de compuestos orgánicos (Lafta y Lorenzen, 1995), lo que permite comprender el hecho de que las plantas estuvieran generando dióxido de carbono en vez de estarlo fijando. Por otro lado, la reducción observada en la tasa transpiratoria es una prueba indirecta de que las plantas bajo estas condiciones presentaban cierre estomático, lo cual es una respuesta a la falta de agua (Brodribb y Holbrook 2003). Todo esto implica que las plantas, bajo las condiciones probadas, aunque estuvieron bajo estrés, todavía estaban fisiológicamente activas.

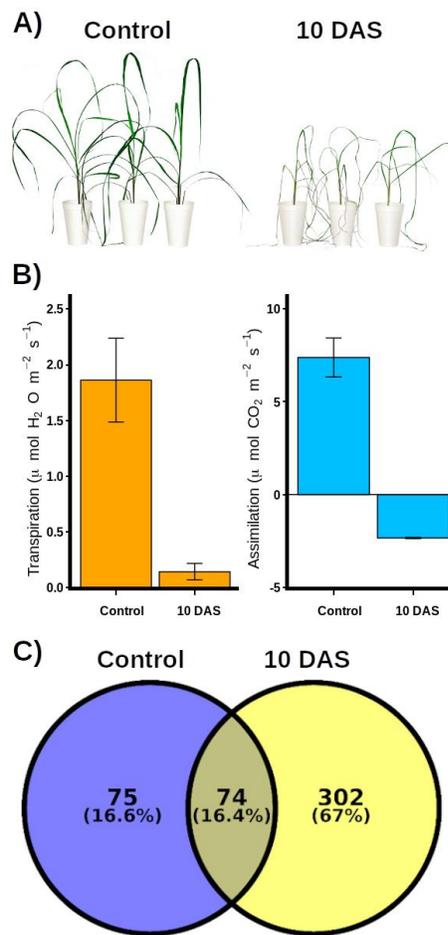


Figura 4.1 Evaluación fisiológica de las plantas de caña empleadas en este estudio y distribución de las proteínas termo-ácido-solubles identificadas. A) Las plantas de caña sometidas a estrés por falta de agua mostraron una importante disminución en su tamaño y en su porte. B) Tanto la tasa de transpiración (A) como la tasa de fotosíntesis (T) presentaron alteraciones en las plantas sometidas a falta de agua, registrando ambos parámetros una reducción en las plantas estresadas. C) Diagrama de Venn que muestra la manera en cómo se comparten las proteínas identificadas en las condiciones evaluadas. Se puede observar que existe una importante cantidad de proteínas que se identificaron exclusivamente en plantas sometidas a estrés, por lo que se puede suponer que muchas de estas proteínas forman parte de los mecanismos de respuesta a estrés. Por otra parte, cabe señalar que el número de proteínas termo-ácido-solubles identificadas exclusivas bajo condiciones de estrés es casi tres veces mayor que las identificadas exclusivamente en condiciones normales de crecimiento. 10 DAS: 10 Days After Stress.

Por otra parte, es interesante indicar que bajo las condiciones de estrés analizadas se identificó un mayor número de proteínas termo-ácido-solubles comparado con las condiciones de crecimiento normales, lo que sugiere que algunas de estas proteínas

participan en la respuesta a la falta de agua en caña. Es interesante destacar que el número de estas proteínas es casi tres veces mayor en condiciones de estrés que las proteínas identificadas en las condiciones control.

Una de las características que permiten que las proteínas permanezcan solubles bajo condiciones de desnaturalización tan drásticas como las empleadas es la falta intrínseca de estructura tridimensional fija. Por ello, una de las características fisicoquímicas que se predijo es el nivel de desorden intrínseco de las proteínas identificadas. En este sentido, cabe resaltar que la mayoría de las proteínas identificadas presentaron regiones intrínsecamente desordenadas (IDRs) mayores a 30 residuos. En general, los resultados de desorden encontrados tanto para las proteínas identificadas en el control como en el tratamiento fueron bastante semejantes (Cuadro 4.1).

Cuadro 4.1 Estadísticas generales de desorden de las proteínas termo-ácido-solubles identificadas. Se puede observar que la mayoría de las proteínas analizadas poseen en su secuencia al menos una IDR mayor a 30 residuos.

Característica	Control	Estrés
aa totales evaluados	55319	148679
Número total de proteínas	149	376
% de proteínas con al menos 1 región desordenada mayor a 30 residuos	71.14	73.14
% de proteínas con al menos una región desordenada mayor a 50 residuos	54.36	54.26
Número total de IDRs mayores a 30 residuos	172	430
Número total de IDRs mayores a 50 residuos	109	276
Número de segmentos desordenados	599	1537

La distribución de desorden estructural, indica que la categoría de 25-50% de desorden fue la más abundante, aunque también hay una importante cantidad de proteínas con un porcentaje de desorden mayor a este rango (Figura 4.2). En total, el porcentaje de proteínas con 25 % de desorden estructural o más suma 67.49 % del total de proteínas identificadas en ambas condiciones, lo cual es un porcentaje bastante aproximado al total de las proteínas que poseen al menos una IDR con >30 aa en su secuencia (72.14%).

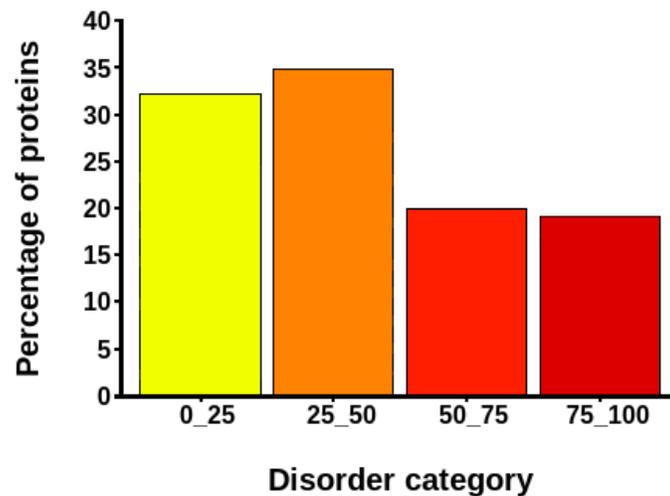


Figura 4.2 Distribución de las proteínas identificadas de acuerdo con su contenido de desorden. Se puede observar que la mayor parte de las proteínas poseen niveles de desorden mayor al 25% en su secuencia.

Para conocer si en las proteínas identificadas existen aminoácidos que están sobre- o sub-representados, se compararon respecto a un conjunto de proteínas estructuradas con estructura cristalográfica y que están depositadas en el PDB. Los resultados obtenidos permiten ver que los aa que componen a la fracción analizada presentan un sesgo respecto a los aa de las proteínas depositadas en el PDB (Figura 4.3).

Es importante destacar que los aminoácidos promotores de orden están sub-representados, mientras que ciertos aminoácidos promotores de desorden están sobre-representados. En cambio, todos los aminoácidos cargados, están sub-representados, independientemente de si son o no promotores de desorden. Es evidente que los aminoácidos con mayor potencial de solvatación están sobre-representados, en contraste con los que poseen un bajo potencial de solvatación.

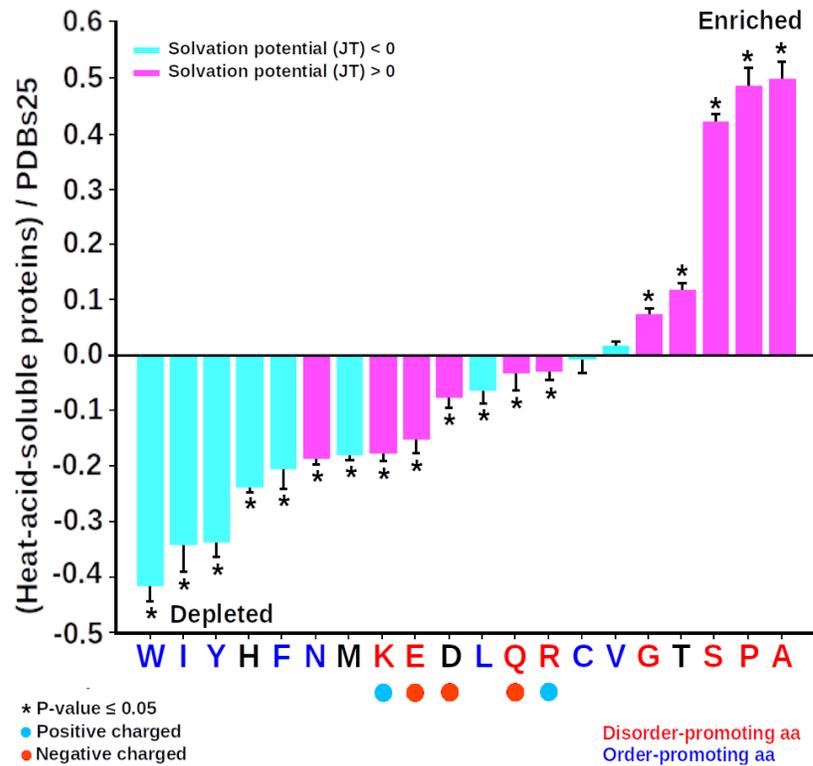


Figura 4.3 Comparación de aa de las proteínas termo-ácido-solubles identificadas y las proteínas del PDB. Se puede apreciar que las proteínas identificadas presentan una mayor proporción de aminoácidos con alto potencial de solvatación (barras rosas), en contraste con los aminoácidos con bajo potencial de solvatación, los cuales se encuentran en una menor proporción. En rojo, destacan los aminoácidos que son promotores de desorden. C y V no presentan una abundancia significativamente diferente ($\alpha < 0.05$).

En contraste, existe un porcentaje importante de proteínas que aparentemente no son desordenadas. De hecho, existe buena parte de estas proteínas participando en actividades

catalíticas. Esto se puede observar graficando los dominios pfam más abundantes en el conjunto de proteínas identificado (Figura 4.4).

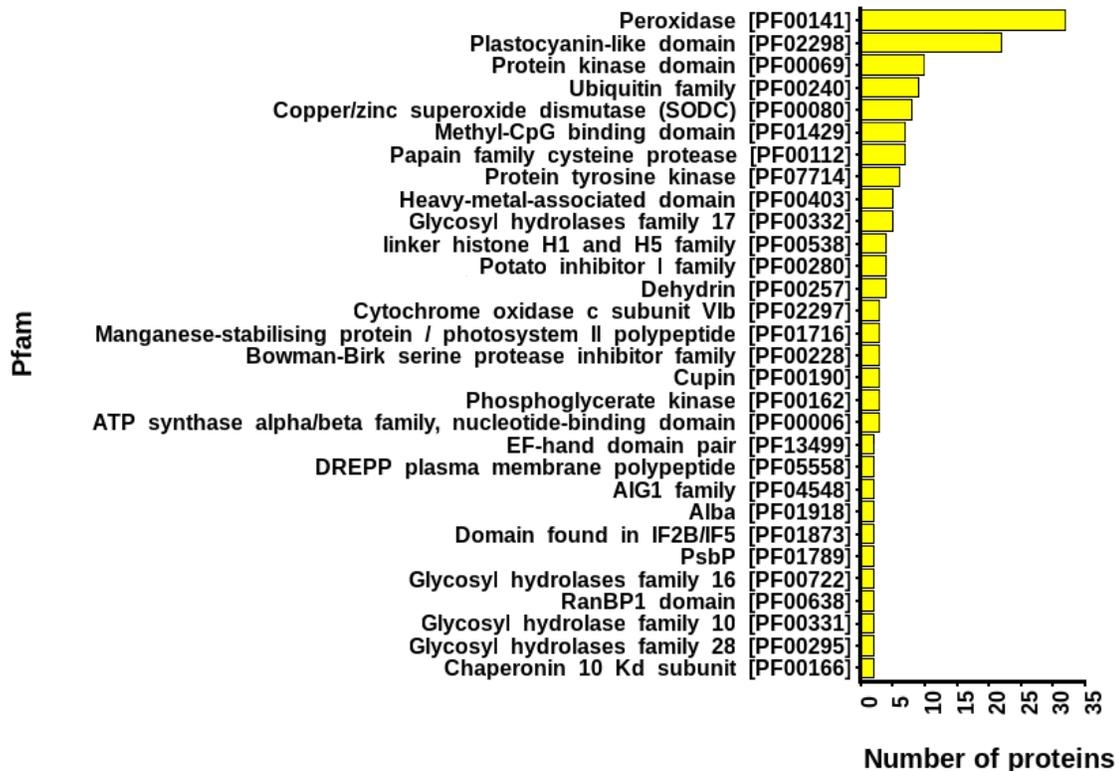


Figura 4.4 Gráfica de los dominios pfam más abundantes de entre las proteínas identificadas. Un importante número de proteínas con función enzimática fueron identificadas, siendo la familia de las peroxidases la más abundante.

Cabe indicar que clasificar las proteínas con base en los dominios pfam presenta sus inconvenientes, pues resulta que una fracción significativa de las proteínas y regiones sin anotar y sin función conocida (lo que se conoce como proteoma oscuro), son IDPs. Es más, se ha llegado a considerar que estos dos aspectos del proteoma son casi sinónimos (Bhowmick et al., 2016; Hatos *et al.*, 2019). Para poner en contexto lo anterior, se estima que sólo el 45% de los residuos de todas las proteínas humanas están representadas en las anotaciones pfam (Mistry *et al.*, 2013). En este trabajo, sólo a 203 proteínas se les pudo asignar un dominio pfam, lo que representa justamente el 45.01% del total de proteínas termo-ácido-solubles identificadas. No obstante, a pesar de sus limitaciones, es interesante encontrar que contrario a lo esperado, en la fracción proteica analizada se identificó un número importante de enzimas que son mayormente proteínas estructuradas.

En concordancia con lo anterior, se puede observar que la distribución de los términos ontológicos refleja que un importante número de proteínas con actividades enzimáticas fueron anotadas (Figura 4.5).

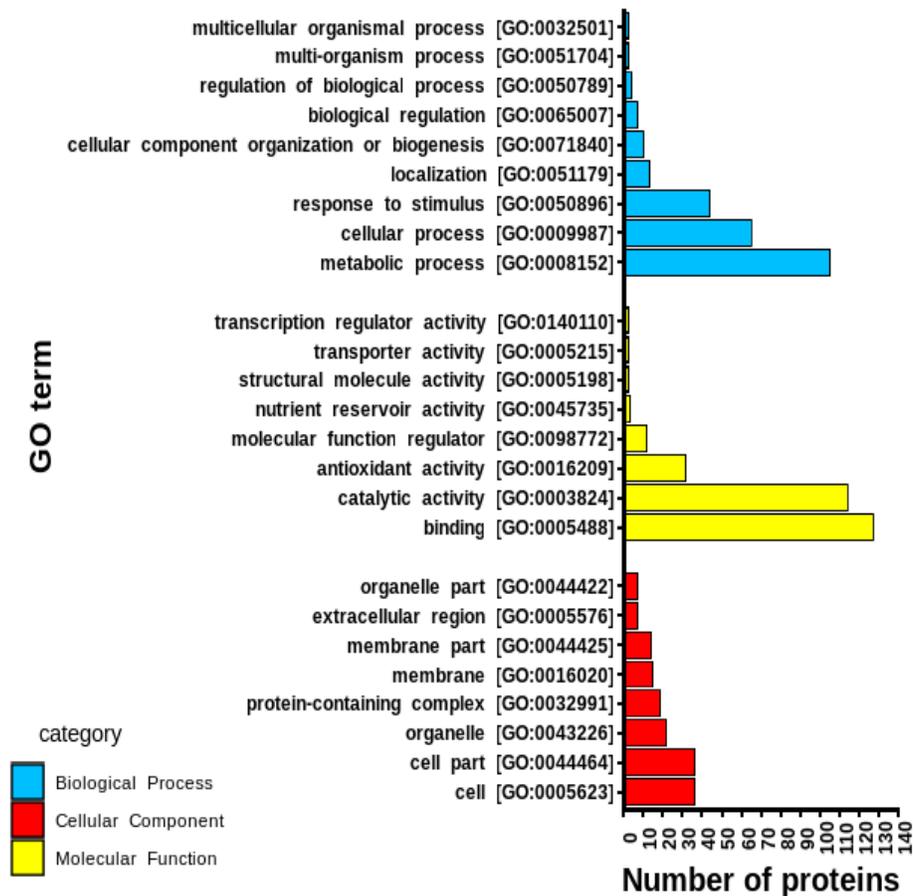


Figura 4.5 Distribución de los términos ontológicos de las proteínas termo-ácido-solubles identificadas. Se puede observar que una buena parte de las 218 proteínas anotadas poseen actividad catalítica y están involucradas en procesos metabólicos y están ligadas a la respuesta a estímulos.

La capacidad de anotar las proteínas identificadas mediante esta estrategia fue ligeramente mayor, pues se pudieron anotar un total de 218 proteínas con sus ontologías asociadas, lo que representa el 48.34% del total. A pesar de esto, los términos GO obtenidos, siguen siendo muy generales, por lo que se decidió realizar una anotación ontológica más detallada. Lo que se consiguió con PANNZER2 (Figura 6).

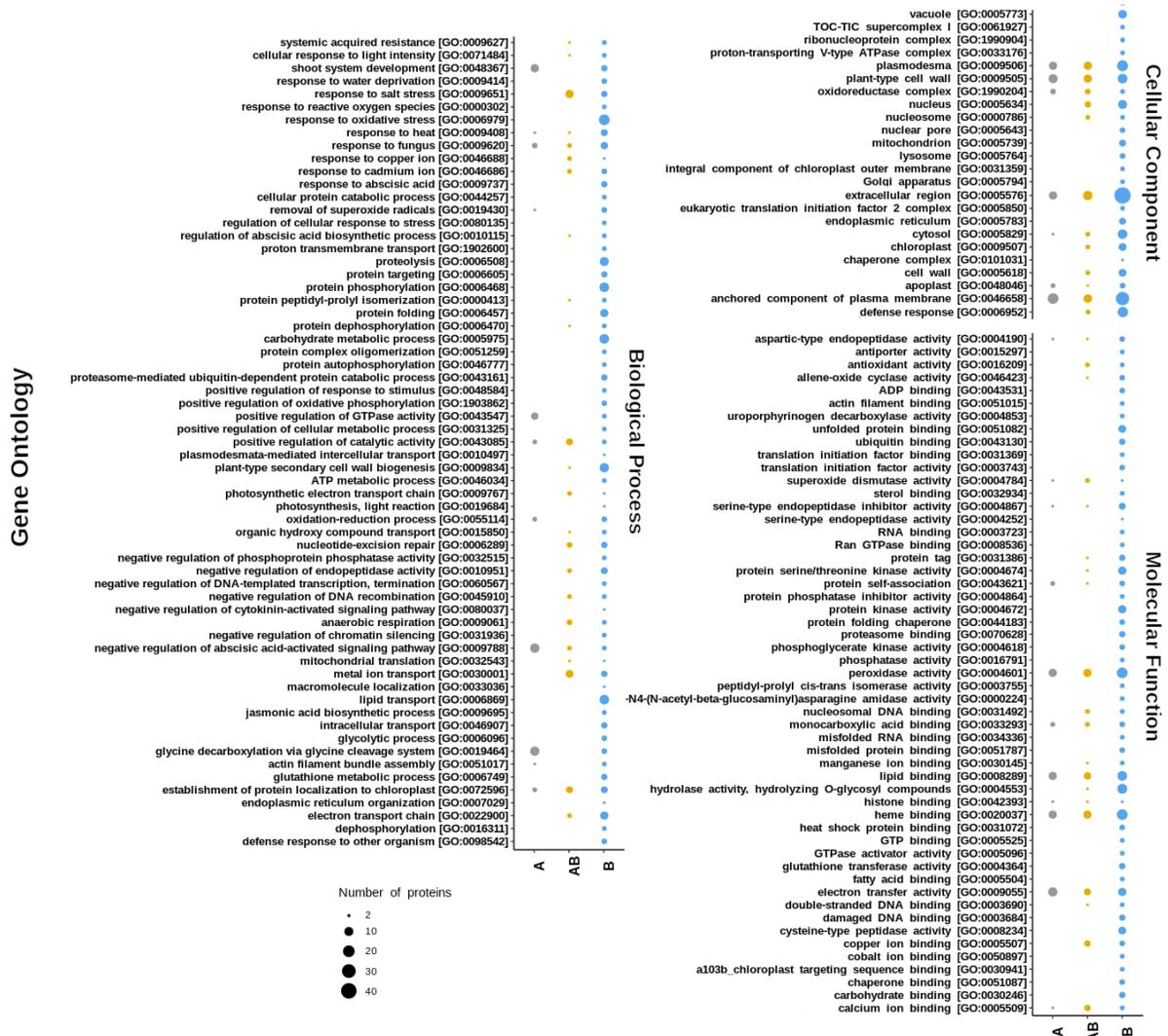


Figura 4.6 Anotación general de las proteínas termo-ácido-solubles identificadas. Las anotaciones presentadas se generaron con PANNZER2. Para un mejor desglose de los términos ontológicos, las proteínas identificadas se subdividieron en tres diferentes subgrupos: A: proteínas identificadas exclusivamente en condiciones control; AB: proteínas identificadas tanto en condiciones control como en condiciones de estrés; B: proteínas identificadas exclusivamente en condiciones de estrés. Sólo se visualizaron los términos ontológicos con al menos dos proteínas anotada en al menos uno de los diferentes sub-grupos definidos.

Mediante esta herramienta fue posible realizar una anotación más detallada de las proteínas identificadas. Es notable destacar que buena parte de las proteínas anotadas al parecer forman parte del medio extracelular. Además, es evidente que muchas de las proteínas identificadas exclusivamente en la condición de estrés estén especializadas en la respuesta a estrés abiótico y estrés oxidante.

También es interesante notar que las funciones asociadas a fosforilación de proteínas, proteólisis, biogénesis de pared secundaria, plegamiento de proteínas, defensa, metabolismo de carbohidratos, transporte de lípidos, biosíntesis de ácido jasmónico o respuesta a ácido abscísico fueron algunas de las funciones en las que se clasificaron las proteínas exclusivamente identificadas en estrés. Por su parte, es relevante destacar que en el caso de las proteínas exclusivamente encontradas en las condiciones control se identificaron varias proteínas asociadas a desarrollo, como las implicadas en el desarrollo del sistema de brotes, a la regulación negativa de la vía de señalización mediada por ABA o la regulación positiva de la actividad GTPase. Dentro de las proteínas compartidas tanto en el tratamiento como en el control, destacan las asociadas a la respuesta a estrés salino y transporte de iones metálicos.

Dentro de las proteínas identificadas como diferenciales y que normalmente son consideradas como proteínas marcadoras de la respuesta a la sequía, se encuentran tres proteínas LEA del grupo 3, cuatro deshidratinas, una rab28, una osmotina, dos quinasas de respuesta a frío, una proteína pequeña de choque térmico, una proteína de 65 kDa inducida por bajas temperaturas y una proteína rica en glicina 2, entre otras. En el caso de las proteínas implicadas en las respuestas a estrés oxidante, además de las peroxidasas y de las superóxido dismutasas, también se identificaron dos glutatión-S-transferasas y una tioredoxina. Por otro lado, también se identificó un número importante de proteínas de unión a metales, una proteína transportadora de cobre ATX1, una ATPasa transportadora de cobre PAA2 cloroplástica, una chaperona de cobre, una proteína transportadora de cobre ATOX1 y dos proteínas estabilizadoras de manganeso (encontrada en ambas condiciones). Cabe destacar que ocho de las 25 proteínas tipo plastocianinas identificadas se encontraron exclusivamente bajo condiciones de estrés. De hecho, este fue el segundo grupo de proteínas más abundante de los que se pudo identificar.

Como parte de las proteínas de respuesta a estrés, también se encontraron diversas proteínas implicadas con el empaquetamiento de DNA, tales como cuatro proteínas con dominio PF00538 de la familia de histonas de unión H1 y H5 (dos de las cuales se encontraron exclusivamente bajo condiciones de estrés) y siete proteínas de unión a metil-

CpG, (tres de las cuales se encontraron exclusivamente en la condición a estrés). Cabe indicar que ambos tipos de proteínas estuvieron dentro de las proteínas con mayor nivel de desorden con un porcentaje de desorden estimado mayor al 93%. Este nivel de desorden es comparable con el encontrado en ciertas dehidrasas y proteínas LEA 3 identificadas.

Es interesante notar que se identificaron 13 fasciclinas, de las cuales 8 fueron encontradas exclusivamente bajo condiciones de estrés. Cabe indicar que estas proteínas no fueron identificadas mediante la anotación de su dominio pfam, por lo que no aparecen en las anotaciones hechas con esta herramienta. Además, se encontraron 3 cupinas expresadas en condiciones de estrés, así como 7 proteasas con dominio tipo papaína y 2 chaperoninas, entre otras diversas proteínas, muchas de las cuales no se pudieron anotar.

4.4 DISCUSIÓN

4.4.1 Características generales de las proteínas identificadas

En este trabajo, se obtuvo la fracción proteica termo-ácido-soluble de la porción aérea de plantas de cañas de azúcar sometidas a estrés por falta de agua. En términos generales, esta fracción está enriquecida en IDPs, pues más del 60% de las proteínas analizadas poseen >25% de desorden estructural en su secuencia. La razón de que esta fracción proteica se enriquezca en IDPs, se debe a que bajo las condiciones desnaturizantes en las que estas proteínas fueron aisladas, ocurre una precipitación diferencial de proteínas. El que esta estrategia permita un enriquecimiento de IDPs, radica en que las proteínas globulares se pliegan alrededor de núcleos hidrofóbicos y su plegamiento está, al menos, dirigido por el efecto hidrofóbico (Scheeff y Fink, 2005). Esto provoca que las proteínas globulares posean regiones hidrofóbicas fuera del alcance de las interacciones polares del solvente. Cuando este tipo de proteínas se someten a cambios de temperatura, estas regiones quedan expuestas en la amplia mayoría de las proteínas globulares. Al exponerse las regiones hidrofóbicas de moléculas diferentes, también se favorecen las colisiones hidrofóbicas intermoleculares, con las que se forman agregados moleculares de mayor tamaño. Cuando estos agregados alcanzan una masa crítica, precipitan. Las IDPs en cambio, no precipitan por el mismo hecho de que carecen de regiones hidrofóbicas (Tompa, 2002), de tal forma que cuando ocurre cambios en el medio que obligan a modificar la

disposición de sus residuos en el espacio, al no haber regiones hidrofóbicas que puedan dar lugar a esos agregados, simplemente no precipitan. Por su parte, a pesar de su amplio uso, el mecanismo por el cual el TCA induce la precipitación de las proteínas no se comprende del todo. Un mecanismo plausible es que a bajas concentraciones, los iones de tricloroacetato promueven la pérdida de estructura, mediante la ruptura de interacciones electrostáticas que desestabilizan la conformación nativa de las proteínas. El desplegamiento parcial de las proteínas da lugar a la exposición de la cara no polar al solvente, lo que resulta en la coalescencia intermolecular de moléculas de proteínas proximales, lo que permite su precipitación (Rajalingam *et al.*, 2009).

Por otra parte, la composición de aa de las IDPs explica por qué resisten más a la precipitación por TCA, apoyados en que su alta proporción de residuos hidrofílicos favorece amplias zonas de interacción con el solvente y probablemente, pocas zonas de interacción intra e intermoleculares que contribuyan al colapso intermolecular, la coalescencia y finalmente, la precipitación de las proteínas. El modelo anterior, sin embargo, no es suficiente para explicar el hecho de que algunas proteínas con pI ácido y que son capaces de permanecer solubles después del choque térmico, se insolubilizan en presencia de TCA (Campos *et al.*, 2011). Sin embargo, se completa, si se toma en cuenta que el TCA, por tratarse de un ácido, también induce la precipitación isoeléctrica de proteínas, es decir, cuando al cambiar el pH, la carga neta de las proteínas tienden a cero y las fuerzas de repulsión electrostáticas se anulan, haciendo que las proteínas se agreguen y precipiten (Rajalingam *et al.*, 2009).

Dentro de las proteínas identificadas en el proteoma se identificaron diversas proteínas intrínsecamente desordenadas cuya presencia en el extracto se puede explicar por su falta de estructura, la cual las hace resilientes a precipitar bajo condiciones desnaturalizantes. Dentro de estas, se encuentran proteínas como las dehidrinas o las histonas H1. Cabe indicar que un número importante de proteínas predichas como intrínsecamente desordenadas no se pudieron anotar con ninguna de las bases de datos empleada. En este sentido, la caracterización funcional de muchas de estas IDPs que además responden a estrés cobra mayor importancia y puede propiciar el descubrimiento de nuevas funciones en los que estas proteínas pueden estar involucradas, por lo que se considera necesario

poner más énfasis en el estudio de IDPs con función desconocida, tales como las identificadas en este trabajo y que responden al estrés.

No obstante, es interesante notar que aunque lo anterior puede explicar la razón por la que se identificó una cantidad importante de IDPs en los extractos proteicos, que de acuerdo con las estimaciones bioinformáticas fueron cerca del 70% del total de las proteínas identificadas, lo cierto es que esta explicación no es suficiente para explicar por qué también se identificaron proteínas globulares. Por ejemplo, los datos de anotación funcional indican que dentro de las proteínas identificadas se encuentran varias enzimas, es decir, proteínas tradicionalmente consideradas como proteínas estructuradas. Esta idea subyace a la idea generalizada de que se requiere de una estructura única para la catálisis enzimática, esto a pesar de que estudios recientes indican la presencia de desorden intrínseco en diversas enzimas funcionales en estado nativo (Palombo *et al.*, 2017). Las enzimas identificadas más numerosas fueron las de la familia de peroxidasas. Es bien conocido el hecho de que una importante cantidad de peroxidasas son termoestables. De hecho, las peroxidasas se encuentran entre las enzimas vegetales más estables al calor (Lee y Pennesi, 1984; Saraiva *et al.*, 2010). Por otra parte, si se analiza con detenimiento el caso de las 32 peroxidasas identificadas, se puede observar que el contenido predicho de desorden para estas proteínas fluctúa entre un 12.8 a 33.5%, lo que implica que de acuerdo con la clasificación de que las proteínas con un nivel de desorden porcentual mayor al 30% son desordenadas (Deiana *et al.*, 2019), algunas de ellas podrían ser catalogadas como proteínas desordenadas, lo cual resulta muy interesante para ser enzimas, puesto que los ejemplos disponibles en la literatura de enzimas desordenadas son todavía bastante escasos (Palombo *et al.*, 2017). Otro ejemplo de enzimas identificadas en este trabajo es las superóxido dismutasas de cobre-zinc. En este caso, existe una amplia evidencia de proteínas que pertenecen a esta familia que son altamente estables a condiciones de calor (Hallewell *et al.*, 1991; Shin *et al.*, 2009). Por ejemplo, la enzima *BtCu,ZnSOD* es una de las enzimas que posee una temperatura de fusión de entre 89 y 104°C dependiendo de las condiciones (Lepock *et al.*, 1985). En el caso de las superóxido dismutasas y de proteínas estructuradas que son termoestables son varios los factores que se han considerado para explicar estas características fisicoquímicas, incluido la pequeña área de superficie expuesta, una mayor densidad de empaquetamiento que reduce las cavidades en el núcleo

hidrofóbico, un incremento en la hidrofobicidad del núcleo hidrofóbico, uniones de hidrógeno en los residuos polares, un incremento en la ocurrencia de residuos de prolina en los loops, un incremento en el área de superficie polar, un incremento en el contenido de helicoidal, una extensa red de pares iónicos y una mejor formación de puentes salinos, entre otros factores (Vetriani *et al.*, 1998; Kumar y Nussinov 2001; Yogavel *et al.*, 2008; Kumar *et al.*, 2012). A pesar de lo anterior, los mecanismos que explican tal fenómeno no son los mismos en todas las proteínas que presentan termoestabilidad, por lo que es contexto dependiente.

4.4.2 Proteínas diferenciales identificadas y su implicación en la respuesta a la falta de agua

A pesar de que en este trabajo sólo se analizó una fracción del proteoma de la caña de azúcar, es de destacar que se identificó una cantidad importante de proteínas diferenciales en respuesta a la falta de agua, muchas de las cuales son reconocidas ampliamente como parte de la respuesta a la falta de agua en plantas. Tal es el caso de las peroxidasas, las cuales juegan un papel en las respuestas a estrés en plantas a tal punto que son consideradas como indicadores de estrés (Siegel, 1993). La inducción de las peroxidasas y de las superóxido dismutasas (SOD), hace pensar que las plantas estuvieron sometidas a un intenso estrés oxidante, puesto que ambos tipos de enzimas constituyen la principal línea de defensa contra las especies reactivas de oxígeno (ROS), las cuales se sobreproducen bajo condiciones de sequía (Alscher *et al.*, 2002; Molina-Rueda *et al.*, 2013). Esto adicional a la identificación de otras proteínas relacionadas con la contención del estrés oxidante, tales como la glutatión-S-transferasa o la metalotioneína y su papel en las respuesta a estrés por sequía está bien establecido (Dinakar y Bartels 2013; Jaiswal *et al.*, 2018; Kumar y Trivedi 2018).

Otras proteínas que forman típicamente parte de las respuestas a estrés por sequía son las proteínas LEA. En este trabajo se identificaron algunos miembros de las proteínas LEA 3 y de las dehidrinas. De hecho las dehidrinas son un sub-grupo de las proteínas LEA, específicamente conforman el grupo de las LEA 2 (Battaglia y Covarrubias, 2013), cuya expresión está estrechamente relacionada con el estrés por deshidratación y aunque su rol bioquímico no está del todo comprendido, se piensa que su función es la de la protección de otras proteínas, la estabilización de membranas, así como la protección contra especies

reactivas de oxígeno (Hanin *et al.*, 2011; Hundertmark *et al.*, 2011; Ismail *et al.*, 1999; Riley *et al.*, 2019).

Por otra parte, es interesante indicar que diversas dehidrinas, así como proteínas LEA 3 tienen capacidad para unir iones metálicos (Hanin *et al.*, 2011; Jin *et al.*, 2013; Liu *et al.*, 2019), lo cual resulta curioso, puesto que muchas de las proteínas identificadas tienen funciones relacionadas con el transporte de iones metálicos o unen iones metálicos, incluidas las superóxido dismutasas, las proteínas estabilizadoras de manganeso, las metalotioneínas, tioredoxinas, las peroxidasas, entre otras proteínas identificadas (Ruchko *et al.*, 1998; Lemaire *et al.*, 1999; Popelkova y Yocum, 2011; Yogavel *et al.*, 2008; Plieth y Vollbehr, 2012). Es probable que ésta sea una característica extendida entre las proteínas termo-ácido-solubles y que merece la pena ser evaluada con mayor detenimiento. En este sentido, es importante resaltar que muchas de estas proteínas poseen regiones con baja complejidad estructural, las cuales recientemente han sido asociadas a funciones relacionadas con la unión a ligandos incluidos RNA, DNA, iones metálicos y polisacáridos o con funciones como chaperonas (Ntountoumi *et al.*, 2019). Funciones que también comparten otras proteínas identificadas incluidas las proteínas de choque térmico, las proteínas ricas en glicina, las chaperoninas, las histonas, las proteínas de unión a metil-CpG o las fasciclinas.

El caso de las proteasas tipo papaína resulta interesante, puesto que todas ellas se encontraron exclusivamente en bajo condiciones de estrés. Los ejemplos de proteínas tipo papaínas (PLCP) que responden a la falta de agua son muy abundantes. Por ejemplo, en trigo se encontró una PLCP que responde positivamente a estreses osmóticos, salino y por frío y genera resistencia en *A. thaliana* (Zang *et al.*, 2010). Aunque cabe indicar que diversos estudios de funcionalización de estos genes han llegado a conclusiones diferentes, ya que aparentemente los mecanismos que median son contextos dependientes y a que las funciones de estas proteínas en el contexto de la sequía son diversas. Así, mientras que el PLCP de trigo generó resistencia a estrés por sequía en *A. thaliana* de manera semejante a como ocurre con la sobre-expresión del gen SPCP2 de camote (Chen *et al.*, 2010), la sobre-expresión del SPCP3 generó el efecto contrario en *Arabidopsis* (Chen *et al.*, 2013). Lo anterior hace suponer que la proteólisis es un mecanismo importante para la respuesta

de las plantas a la falta de agua (Liu *et al.*, 2018), por lo que la actividad de estas proteínas debe ser tomada en consideración en futuros estudios de funcionalización.

4.5 CONCLUSIÓN

De acuerdo con los datos encontrados en este trabajo es posible afirmar que las condiciones de estrés favorecen la acumulación de proteínas con baja tendencia a la formación de agregados. En este sentido, lo que se propone es que las proteínas que forman parte de la respuesta a la falta de agua son proteínas con tendencia a no formar agregados. Por ello, se considera que es necesario tomar en consideración las características fisicoquímicas generales de las proteínas de respuesta a estrés, más allá de las funciones que poseen de manera particular. Así, es posible que bajo condiciones de estrés definidas, el proteoma se ajuste de una forma tal que sea más resiliente a los efectos directos del estrés, tales como la agregación molecular. Esto es similar a lo observado para *A. thaliana* en respuesta a estrés por calor (Álvarez-Ponce *et al.*, 2018).

CAPÍTULO V

CONCLUSIÓN GENERAL Y PERSPECTIVAS

5.1 CONCLUSIÓN GENERAL

En este trabajo se abordó el estudio de las proteínas intrínsecamente desordenadas y su relación con las respuestas a la falta de agua en plantas, esto con el propósito de evaluar si en la respuesta a la falta de agua en las plantas ocurre un aumento en el contenido de IDPs. La idea detrás de todo esto parte de la observación de que una buena parte de los efectos negativos de la falta de agua en las células ocurre como consecuencia de la pérdida de función de proteínas a consecuencia de la pérdida de su estructura nativa, así como del incremento de las interacciones proteicas no específicas que ocurren a consecuencia de la falta de agua. De ahí que se propusiera que una vía para aminorar de los efectos negativos de la pérdida de agua en las plantas pasara por disminuir el contenido de proteínas intrínsecamente susceptibles a la agregación molecular (en general, proteínas globulares con alto grado de estructuración) y en cambio se indujera la cantidad de proteínas resistentes a la pérdida de función como consecuencia de la pérdida de estructura. La idea es simple: las plantas emplean proteínas de respuesta a la falta de agua que en son resistentes a la pérdida de estructura y a la agregación molecular y bajo condiciones de estrés se minimiza la síntesis de proteínas que son susceptibles a la agregación molecular. Bajo esta idea, se estableció el concepto del desordenoma del estrés.

Para evaluar esta propuesta, se emplearon diferentes enfoques. Primero, se realizó una búsqueda exhaustiva para evaluar la distribución del contenido relativo de desorden estimado en los proteomas vegetales de especies cuyos genomas estuvieran disponibles al momento de iniciar el trabajo. Esto permitió evaluar qué grupo taxonómico está enriquecido en desorden estructural. Se pudo observar que las monocotiledóneas presentan una mayor proporción de desorden estructural que las eudicotiledóneas. Sorpresivamente, el contenido relativo de desorden en los proteomas de algas es mayor que en plantas superiores, lo que contrasta con la idea de que el desorden está asociado con el incremento en la complejidad orgánica, pero que puede explicarse si se toma en consideración la multifuncionalidad de las IDPs, de tal forma que una mayor proporción de IDPs, implica ahorros en términos de recursos genómicos al permitir compactar el conjunto

de funciones necesarios para el mantenimiento del organismo con un subconjunto reducido de proteínas.

Por otra parte, se realizó el análisis de enriquecimiento ontológico de las anotaciones de los proteomas analizados, esto para evaluar si existen términos enriquecidos en las proteínas de acuerdo con su nivel de desorden. De acuerdo con estos análisis, las proteínas estructuradas se encuentran enriquecidas en funciones estructurales y catalíticas, mientras que las proteínas que poseen mayor desorden están enriquecidas en funciones relacionadas con la señalización y las respuestas a estrés. Finalmente, se encontró que los aa que están enriquecidos en las IDPs, presentan mayores cambios entre los diferentes taxas analizados (algas, monocotiledóneas y eudicotiledóneas) que aquellos aa enriquecidos en las proteínas estructuradas. Lo anterior sugiere que la tasa de cambios en los residuos promotores de desorden es mayor que los residuos promotores de orden, lo que a su vez implica mayor variación en las proteínas desordenadas. Esto puede tener fuertes implicaciones adaptativas en las plantas ya que el desordenoma puede resultar ser una fuente de variación mayor que el resto del proteoma.

Posteriormente, se consideró una estrategia que permitiera evaluar la forma en cómo se da la regulación transcripcional de las proteínas desordenadas en un contexto de desecación. Para ello, se evaluó la regulación del transcriptoma a lo largo del desarrollo de la semilla de maíz considerando la forma en cómo se comportan los transcritos cuando son clasificados de acuerdo con el nivel de desorden de las proteínas para las que codifican.

Se seleccionó este modelo debido a que es una monocotiledónea bien caracterizada de la que se tiene un profundo conocimiento de los procesos que ocurren durante la embriogénesis de la semilla y porque al final del desarrollo de la semilla ocurre la activación del programa de quiescencia, lo que implica la activación de los mecanismos de adaptación a la desecación. Teniendo en consideración ambos factores, se decidió estudiar el comportamiento de los transcritos que son expresados a lo largo del desarrollo de la semilla hasta el inicio de la embriogénesis tardía, que es cuando ocurre la adquisición de la tolerancia a la desecación. Se observó que tanto el número de genes expresados que codifican para IDPs como su abundancia relativa, son dependientes de la etapa de

desarrollo de la semilla y que es durante la etapa de llenado de grano cuando la abundancia relativa de los transcritos que codifican para las proteínas estructuradas aumenta de manera significativa. Sin embargo, en términos generales, la expresión de transcritos que codifican para IDPs aumenta de forma constante conforme avanza el desarrollo de la semilla y aumenta justo al final de dicho periodo, cuando se establecen las condiciones para la quiescencia.

Por otra parte, se identificó un número importante de genes que codifican para proteínas altamente desordenadas cuya expresión está directamente relacionada con la progresión de la pérdida de agua en la semilla. Dichos transcritos están asociados a funciones relacionadas con la respuesta a estrés por falta de agua. Además, se identificaron factores de transcripción que funcionan como reguladores maestros de diferentes etapas del proceso. Dichos factores de transcripción se clasificaron de acuerdo con su nivel de desorden y se encontró que muchos de ellos son intrínsecamente desordenados. A su vez, se identificaron factores de transcripción que pueden ser claves para la adquisición de la tolerancia a la desecación en la semilla de maíz y que hasta la fecha no han sido caracterizados funcionalmente, lo que merece especial atención. Con toda la información generada se estableció un modelo en el que se propone que bajo un contexto de desecación existe un ajuste activo que tiene como objetivo aumentar el contenido relativo de proteínas intrínsecamente desordenadas, todo esto mediante dos mecanismos que son complementarios. Primero, aumentando el número de genes expresados que codifican para estas proteínas y segundo, reprimiendo la expresión de genes que codifican para proteínas con alto nivel de estructuración. Los datos presentados concuerdan con lo propuesto, por lo que el modelo, a pesar de su simplicidad, se mantiene.

Finalmente, se decidió emplear un enfoque experimental para tratar de corroborar las observaciones anteriores. Para ello, se empleó caña de azúcar como modelo, esto por ser una monocotiledónea, cuya respuesta a falta de agua no ha sido muy estudiada. En este caso, se empleó un enfoque basado en proteómica libre de gel para identificar al conjunto de proteínas que permanecen solubles después de la aplicación de un choque térmico y pH ácido. A dicha fracción se le denominó proteínas termo-ácido-solubles. El que estas

proteínas permanezcan solubles después de aplicar dichas condiciones es una muestra de que son proteínas resistentes a la formación de agregados.

Para tratar de explicar las características que favorecen este comportamiento, se realizó una serie de análisis bioinformáticos para predecir el desorden intrínseco. Se encontró que la mayor parte de estas proteínas (más del 70% de las proteínas identificadas) son desordenadas. Por otro lado, se identificó un importante número de proteínas que responden a la falta de agua y que muchas de ellas son desordenadas y que no poseen anotación funcional, lo que indica que no están caracterizadas. Este hecho implica que dichas proteínas pueden poseer funciones que todavía no han sido descritas y que pueden tener un rol relevante en la respuesta a la falta de agua. Además, se identificaron diversas enzimas que se predicen que son parcialmente desordenadas y que responden a la deshidratación, lo cual es un aspecto novedoso si se considera que en la literatura existen pocos ejemplos de enzimas desordenadas que hayan sido experimentalmente validadas. Un aspecto destacable de esta parte del trabajo, radica en que se identificó que el número de proteínas termo-acido-solubles y por ende, el número de IDPs es mayor en condiciones de falta de agua (casi tres veces mayor que en condiciones normales) lo que sugiere que la síntesis de IDPs es un mecanismo generalizado para activar la respuesta a la falta de agua.

De esta forma, en este trabajo se encontró evidencia que sugiere que las respuestas al estrés y a la falta de agua en particular, pasa por incluir la selección de proteínas que son más resilientes a la pérdida de función mediada por desnaturalización. Tal parece que las proteínas que forman parte de los mecanismos de respuesta al estrés son proteínas que poseen una robusta estabilidad y poseen resiliencia a la agregación mediada por la falta de agua. Si uno se detiene a pensar un poco, resulta que este es un mecanismo sencillo que implica que las proteínas que forman parte de la respuesta a la falta de agua requieren menor asistencia para poder desempeñar sus funciones en el sentido de que son proteínas menos susceptibles a los efectos de la agregación.

5.2 PERSPECTIVAS

Este trabajo a pesar de lo sencillo de sus aproximaciones resulta en una buena base para plantear nuevos experimentos para evaluar el fenómeno aquí estudiado bajo otros contextos. Por ejemplo, partiendo de esta idea sería interesante evaluar este fenómeno en otros modelos, estados de desarrollo u otros tipos de estrés.

Por otra parte, muchos de los datos generados pueden servir para plantear proyectos con fines biotecnológicos, para lo cual se sugiere realzar una cuidadosa selección de proteínas no caracterizadas que responden a estrés, para ser sometidas a evaluaciones funcionales. Por otro lado, es probable que alguna de las enzimas termoestables identificadas posea funciones de interés comercial, en cuyo caso estaría justificada su caracterización funcional.

Es importante hacer hincapié en el potencial biotecnológico de los factores de transcripción que se encontraron como reguladores clave en el proceso de desarrollo de la semilla, especialmente, en las etapas tardías de dicho proceso. En el caso de las proteínas de caña, la identificación experimental de un número importante de proteínas con función desconocida que responden al estrés y que son IDPs, da paso a toda una serie de proyectos interesantes que merecen la pena ser llevados a cabo.

Finalmente, es importante seguir sometiendo a prueba el modelo planteado, para conocer sus alcances y sus limitaciones y así poder completarlo, mejorarlo o encontrar sus excepciones. El empleo de herramientas más exhaustivas para la caracterización del desorden estructural de manera sistemática y masiva permitirá el abordaje de este fascinante fenómeno, sólo hay que esperar a que dicho enfoque pronto esté disponible.

BIBLIOGRAFÍA

- Afanasyeva, A., Bockwoldt, M., Cooney, C. R., Heiland, I. y Gossmann, T. I. (2018). Human long intrinsically disordered protein regions are frequent targets of positive selection. *Genome Research*, 28, 975-982.
- Aibar, S., González-Blas, C. B., Moerman, T., Huynh-Thu, V. A., Imrichova, H., Hulselmans, G. y Aerts, S. (2017). SCENIC: Single-cell regulatory network inference and clustering. *Nature Methods*, 14, 1083-1086.
- Alscher, R. G., Erturk, N. y Heath, L. S. (2002). Role of superoxide dismutases (SODs) in controlling oxidative stress in plants. *Journal of Experimental Botany*, 53, 1331-1341.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. y Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215, 403-410.
- Álvarez-Ponce, D., Ruiz-González, M. X., Vera-Sirera, F., Feyertag, F., Perez-Amador, M. A. y Fares, M. A. (2018). Arabidopsis heat stress-induced proteins are enriched in electrostatically charged amino acids and intrinsically disordered regions. *International Journal of Molecular Sciences*. 19, E2276. doi: <https://doi.org/10.3390/ijms19082276>
- Amara, I., Odena, A., Oliveira, E., Moreno, A., Masmoudi, K., Pags, M. y Goday, A. (2012). Insights into maize LEA proteins: from proteomics to functional approaches. *Plant and Cell Physiology*, 53, 312-29
- Andjelkovic, V. y Thompson, R. (2006). Changes in gene expression in maize kernel in response to water and salt stress. *Plant Cell Reports*, 25, 71-9.
- Arvidsson, G. y Wright, A. P. H. (2018). A protein intrinsic disorder approach for characterising differentially expressed genes in transcriptome data: Analysis of cell-adhesion regulated gene expression in lymphoma cells. *International Journal of Molecular Sciences* 19, 3101. doi: <https://doi.org/10.3390/ijms19103101>
- Baraldi, E., Coller, E., Zoli, L., Cestaro, A., Tosatto, S. C. E. y Zambelli, B. (2015). Unfoldome variation upon plant-pathogen interactions: strawberry infection by *Colletotrichum acutatum*. *Plant Molecular Biology*, 89, 49-65.
- Battaglia, M. y Covarrubias, A. A. (2013). Late Embryogenesis Abundant (LEA) proteins in legumes. *Frontiers in Plant Science*, 4, 190.

doi: <http://doi.org/10.3389/fpls.2013.00190>

- Benešová, M., Holá, D., Fischer, L., Jedelský, P. L., Hnilička, F., Wilhelmová, N., *et al.* (2012). The physiology and proteomics of drought tolerance in maize: early stomatal closure as a cause of lower tolerance to short-term dehydration? *PLoS ONE*, 7(6): e38017.
- Bern, M., Kil, Y. J. y Becker, C. (2012). Byonic: advanced peptide and protein identification software, *Curr Protoc Bioinformatics*. *Curr Protoc Bioinformatics Chapter 13*, Unit13. <https://doi.org/10.1002/0471250953.bi1320s40.Byonic>
- Bhowmick, A., Brookes, D. H., Yost, S. R., Dyson, H. J., Forman-Kay, J. D., Gunter, D., *et al.* (2016). Finding our way in the dark proteome. *Journal of the American Chemical Society* 138, 9730-9742.
- Boothby, T. C. y Pielak, G. J. (2017). Intrinsically disordered proteins and desiccation tolerance: elucidating functional and mechanistic underpinnings of anhydrobiosis. *BioEssays*, 39, 11. doi: 10.1002/bies.201700119.
- Boothby, T. C., Tapia, H., Brozena, A. H., Piszkiwicz, S., Smith, A. E., Giovannini, I., Goldstein, B. (2017). Tardigrades use intrinsically disordered proteins to survive desiccation. *Molecular Cell*, 65, 975-984.
- Borisjuk, N., Sitailo, L., Adler, K., Malysheva, L., Tewes, A., Borisjuk, L. y Manteuffel, R. (1998). Calreticulin expression in plant cells: Developmental regulation, tissue specificity and intracellular distribution. *Planta*, 206, 504-14.
- Boudet, J., Buitink, J., Hoekstra, F. A., Rogniaux, H., Larré, C., Satour, P. y Leprince, O. (2006). Comparative analysis of the heat stable proteome of radicles of *Medicago truncatula* seeds during germination identifies late embryogenesis abundant proteins associated with desiccation tolerance. *Plant Physiology*, 140, 1418-1436.
- Breiman, L. (2002). Statistical Modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, 16, 199-231.
- Bremer, A., Kent, B., Hauß, T., Thalhammer, A., Yepuri, N. R., Darwish, T. A., Hinch, D. K. (2017). Intrinsically disordered stress protein cor15a resides at the membrane surface during dehydration. *Biophysical Journal*, 113, 572-579.

- Brochetto-Braga, M. R., Leite, A. y Arruda, P. (2008). Partial purification and characterization of lysine-ketoglutarate reductase in normal and opaque-2 maize endosperms. *Plant Physiology*, 98, 1139-1147.
- Brodribb, T. J. y Holbrook, N. M. (2003). Stomatal closure during leaf dehydration, correlation with other leaf physiological traits. *Plant Physiology*, 132, 2166-2173.
- Brown, C. J., Takayama, S., Campen, A. M., Vise, P., Marshall, T. W., Oldfield, *et al.* (2002). Evolutionary rate heterogeneity in proteins with long disordered regions. *Journal of Molecular Evolution*, 55, 104-110.
- Campos, F., Guillén, G., Reyes, J. L. y Covarrubias, A. A. (2011). A general method of protein purification for recombinant unstructured non-acidic proteins. *Protein Expression and Purification*, 80, 47-51.
- Carugo, O. (2008). Amino acid composition and protein dimension. *Protein Science*, 17, 2187-2191.
- Chakrabortee, S., Boschetti, C., Walton, L. J., Sarkar, S., Rubinsztein, D. C. y Tunnacliffe, A. (2007). Hydrophilic protein associated with desiccation tolerance exhibits broad protein stabilization function. *Proceedings of the National Academy of Sciences*, 104, 18073-18078.
- Chakrabortee, S., Meersman, F., Kaminski Schierle, G. S., Bertoncini, C. W., McGee, B., Kaminski, C. F. y Tunnacliffe, A. (2010). Catalytic and chaperone-like functions in an intrinsically disordered protein associated with desiccation tolerance. *Proceedings of the National Academy of Sciences*, 107, 16084-16089.
- Chakrabortee, S., Tripathi, R., Watson, M., Kaminski Schierle, G. S., Kurniawan, D. P., Kaminski, *et al.* (2012). Intrinsically disordered proteins as molecular shields. *Molecular BioSystems*, 8, 210-219.
- Checker, V. G. y Khurana, P. (2013). Molecular and functional characterization of mulberry EST encoding remorin (MiREM) involved in abiotic stress. *Plant Cell Reports*, 32, 1729-1741.
- Chen, H. J., Su, C. T., Lin, C. H., Huang, G. J. y Lin, Y. H. (2010). Expression of sweet potato cysteine protease SPCP2 altered developmental characteristics and stress responses in transgenic *Arabidopsis* plants. *Journal of Plant Physiology*, 167, 838-847.

- Chen, H. J., Tsai, Y. J., Shen, C. Y., Tsai, T. N., Huang, G. J. y Lin, Y. H. (2013). Ectopic expression of sweet potato cysteine protease SPCP3 alters phenotypic traits and enhances drought stress sensitivity in transgenic *Arabidopsis* plants. *Journal of Plant Growth Regulation*, 32, 108-121.
- Chen, J., Zeng, B., Zhang, M., Xie, S., Wang, G., Hauck, A. y Lai, J. (2014). Dynamic transcriptome landscape of maize embryo and endosperm development. *Plant Physiology*, 166, 252-264.
- Choura, M., Ebel, C. y Hanin, M. (2019). Genomic analysis of intrinsically disordered proteins in cereals: From mining to meaning. *Gene*, 714, 143984. doi: 10.1016/j.gene.2019.143984.
- Coleman, C. E., Herman, E. M., Takasaki, K. y Larkins, B. A. (2007). The maize g-zein sequesters a-zein and stabilizes its accumulation in protein bodies of transgenic tobacco endosperm. *Plant Cell*, 8, 2335-2345.
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M. y Robles, M. (2005). Blast2GO: A universal tool for annotation, visualization, and analysis in functional genomics research. *Bioinformatics*, 21, 3674-6.
- Cortese, M. S., Baird, J. P., Uversky, V. N. y Dunker, A. K. (2005). Uncovering the unfoldome: Enriching cell extracts for unstructured proteins by acid treatment. *Journal of Proteome Research*, 4, 1610-1618.
- Covarrubias, A. A., Cuevas-Velázquez, C. L., Romero-Pérez, P. S., Rendón-Luna, D. F. y Chater, C. C. C. (2017). Structural disorder in plant proteins: where plasticity meets sessility. *Cellular and Molecular Life Sciences*, 74, 3119-3147.
- Cuevas-Velázquez, C. L., Saab-Rincón, G., Reyes, J. L. y Covarrubias, A. A. (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. *Journal of Biological Chemistry*, 291, 10893–10903.
- Dai, J., Liu, H., Zhou, J. y Huang, K. (2016). Selenoprotein R protects human lens epithelial cells against D-galactose-induced apoptosis by regulating oxidative stress and endoplasmic reticulum stress. *International Journal of Molecular Sciences*, 17, 231. <https://doi.org/10.3390/ijms17020231>

- Das, R. K., Ruff, K. M. y Pappu, R. V. (2015). Relating sequence encoded information to form and function of intrinsically disordered proteins. *Current Opinion in Structural Biology*, 32, 102-12.
- Deiana, A., Forcelloni, S., Porrello, A. y Giansanti, A. (2019). Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell. *PLoS ONE*, 14, e0217889.
- Deiana, A. Y Giansanti, A. (2012). Tuning the precision of predictors to reduce overestimation of protein disorder over large datasets. *Journal of Bioinformatics and Computational Biology*, 11, 1250023.
- Delahaie, J., Hundertmark, M., Bove, J., Leprince, O., Rogniaux, H. y Buitink, J. (2013). LEA polypeptide profiling of recalcitrant and orthodox legume seeds reveals ABI3-regulated LEA protein abundance linked to desiccation tolerance. *Journal of Experimental Botany*, 64, 4559-4573.
- Devic, M. y Roscoe, T. (2016). Seed maturation: Simplification of control networks in plants. *Plant Science*, 252, 335-346.
- Dinakar, C. y Bartels, D. (2013). Desiccation tolerance in resurrection plants: new insights from transcriptome, proteome, and metabolome analysis. *Frontiers in Plant Science*. 4, 482. <https://doi.org/10.3389/fpls.2013.00482>
- Dorfer, V., Pichler, P., Stranzl, T., Stadlmann, J., Taus, T., Winkler, S. y Mechtler, K. (2014). MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. *Journal of Proteome Research*, 13, 3679-3684.
- Dosztányi, Z. (2018). Prediction of protein disorder based on IUPred. *Protein Science*, 27, 331-340.
- Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C. y Lepiniec, L. (2010). MYB transcription factors in *Arabidopsis*. *Trends in Plant Science*, 15, 573-581.
- Dunker, A. K., Silman, I., Uversky, V. N. y Sussman, J. L. (2008). Function and structure of inherently disordered proteins. *Current Opinion in Structural Biology*, 18(6), 756-64. <https://doi.org/10.1016/j.sbi.2008.10.002>
- Dure, L. y Galau, G. a. (1981). Developmental biochemistry of cottonseed embryogenesis and germination 1. *Plant Physiology*, 20, 4162-168.

- Eng, Jimmy K., Ashley L. McCormack, and J. R. Y. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry*, 5, 976-989.
- Esen, A. (1987). A proposed nomenclature for the alcohol-soluble proteins (zeins) of maize (*Zea mays* L.). *Journal of Cereal Science*, 5, 117-128.
- Fang, Y., You, J., Xie, K., Xie, W. y Xiong, L. (2008). Systematic sequence analysis and identification of tissue-specific or stress-responsive genes of NAC transcription factor family in rice. *Molecular Genetics and Genomics*, 280, 547-63.
- Forcelloni, S. y Giansanti, A. (2020). Evolutionary forces and codon bias in different flavors of intrinsic disorder in the human proteome. *Journal of Molecular Evolution*, 88, 164-178.
- Frege, T. y Uversky, V. N. (2015). Intrinsically disordered proteins in the nucleus of human cells. *Biochemistry and Biophysics Reports*, 1:33-51.
- Galea, C. A., High, A. A., Obenauer, J. C., Mishra, A., Park, C. G., Punta, M., *et al.* (2009). Large-scale analysis of thermostable, mammalian proteins provides insights into the intrinsically disordered proteome. *Journal of Proteome Research*, 8, 211-26.
- Galea, C. A., Pagala, V. R., Obenauer, J. C., Park, C. G., Slaughter, C. A. y Kriwacki, R. W. (2006). Proteomic studies of the intrinsically unstructured mammalian proteome. *Journal of Proteome Research*, 5, 2839-2848.
- Gallagher, E., Gao, M., Liu, Y.-C. y Karin, M. (2006). Activation of the E3 ubiquitin ligase Itch through a phosphorylation-induced conformational change. *Proceedings of the National Academy of Sciences*, 103, 1717-1722.
- Garay-Arroyo, A., Colmenero-Flores, J. M., Garcíarrubio, A. y Covarrubias, A. A. (2000). Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. *Journal of Biological Chemistry*, 275, 5668-5674.
- Ge, S. X., Jung, D. y Yao, R. (2019). ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics*, 36, 2628-2629.
- Gibbon, B. C. y Larkins, B. A. (2005). Molecular genetic approaches to developing quality protein maize. *Trends in Genetics*, 21, 227-33.

- Ginestet, C. (2011). ggplot2: Elegant graphics for data analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. https://doi.org/10.1111/j.1467-985x.2010.00676_9.x
- Goyal, K., Walton, L. J. Y Tunnacliffe, A. (2005). LEA proteins prevent protein aggregation due to water stress. *Biochemical Journal*, 388(Pt 1):151-157.
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., *et al.* (2013). *De novo* transcript sequence reconstruction from RNA-seq: reference generation and analysis with Trinity. *Nature Protocols*, 8, 1494-1512.
- Hallewell, R. A., Imlay, K. C., Lee, P., Fong, N. M., Gallegos, C., Getzoff, E. D., *et al.* (1991). Thermostabilization of recombinant human and bovine CuZn superoxide dismutases by replacement of free cysteines. *Biochemical and Biophysical Research Communications*, 181, 474-80.
- Hanin, M., Brini, F., Ebel, C., Toda, Y., Takeda, S. y Masmoudi, K. (2011). Plant dehydrins and stress tolerance: Versatile proteins for complex mechanisms. *Plant Signaling and Behavior*, 6, 1503-1509.
- Hartings, H., Lauria, M., Lazzaroni, N., Pirona, R. y Motto, M. (2011). The *Zea mays* mutants opaque-2 and opaque-7 disclose extensive changes in endosperm metabolism as revealed by protein, amino acid, and transcriptome-wide analyses. *BMC Genomics*, 12, 41. doi: <https://doi.org/10.1186/1471-2164-12-41>
- Hatos, A., Hajdu-Soltész, B., Monzon, A. M., Palopoli, N., Álvarez, L., Aykac-Fas, B., *et al.* (2019). DisProt: intrinsic protein disorder annotation in 2020. *Nucleic Acids Research*, 48(D1), D269-D276.
- High, S., Lecomte, F. J. L., Russell, S. J., Abell, B. M. y Oliver, J. D. (2000). Glycoprotein folding in the endoplasmic reticulum: A tale of three chaperones? *FEBS Letters*, 476, 38-41.
- Hoekstra, F. A., Golovina, E. A. y Buitink, J. (2001). Mechanism of plant desiccation tolerance. *Trends in Plant Science*, 6, 431-438.
- Hu, W., Wei, Y., Xia, Z., Yan, Y., Hou, X., Zou, M., *et al.* (2015). Genome-wide identification and expression analysis of the NAC transcription factor family in cassava. *PLoS ONE*, 10, e0136993.

- Hundertmark, M., Dimova, R., Lengefeld, J., Seckler, R. y Hinch, D. K. (2011). The intrinsically disordered late embryogenesis abundant protein LEA18 from *Arabidopsis thaliana* modulates membrane stability through binding and folding. *Biochimica et Biophysica Acta - Biomembranes*, 1808, 446-53.
- Hundertmark, M. y Hinch, D. K. (2008). LEA (Late Embryogenesis Abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics*, 9, 118.
- Huynh-Thu, V. A., Irrthum, A., Wehenkel, L. y Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE*, 5 e12776.
- Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradović, Z. y Dunker, A. K. (2002). Intrinsic disorder in cell-signaling and cancer-associated proteins. *Journal of Molecular Biology*, 323, 573-584.
- Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z. y Dunker, A. K. (2004). The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Research*, 32, 1037-1049.
- Illing, N., Denby, K. J., Collett, H., Shen, A. y Farrant, J. M. (2005). The signature of seeds in resurrection plants: A molecular and physiological comparison of desiccation tolerance in seeds and vegetative tissues. *Integrative and Comparative Biology*, 45, 771-787.
- Irar, S., Oliveira, E., Pagès, M. y Goday, A. (2006). Towards the identification of late-embryogenic-abundant phosphoproteome in *Arabidopsis* by 2-DE and MS. *Proteomics*, 6, S175-185.
- Ismail, A. M., Hall, A. E. y Close, T. J. (1999). Purification and partial characterization of a dehydrin involved in chilling tolerance during seedling emergence of cowpea. *Plant Physiology*, 120, 237-244.
- Jaiswal, P. S., Mittal, N. y Randhawa, G. S. (2018). *Cyamopsis tetragonoloba* type 1 metallothionein (CtMT1) gene is upregulated under drought stress and its protein product has an additional C-X-C motif and unique metal binding pattern. *International Journal of Biological Macromolecules*, 119, 1324-1334.
- Jin, X., Fu, Z., Ding, D., Li, W., Liu, Z. y Tang, J. (2013). Proteomic identification of genes associated with maize grain-filling rate. *PLoS ONE*, 8, e59353.

- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., *et al.* (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, 30, 1236-1240.
- Kishor, P. B. (2015). Role of proline in cell wall synthesis and plant development and its implications in plant ontogeny. *Frontiers in Plant Science*, 6: 544.
- Kim, J. Y., Kim, W. Y., Kwak, K. J., Oh, S. H., Han, Y. S. y Kang, H. (2010). Glycine-rich RNA-binding proteins are functionally conserved in *Arabidopsis thaliana* and *Oryza sativa* during cold adaptation process. *Journal of Experimental Botany*, 61, 2317-2325.
- Kovacs, D., Kalmar, E., Torok, Z. y Tompa, P. (2008). Chaperone activity of erd10 and erd14, two disordered stress-related plant proteins. *Plant Physiology*, 147, 381-390.
- Kovacs, D., Agoston, B. y Tompa, P. (2008). Disordered plant LEA proteins as molecular chaperones. *Plant Signaling and Behavior*, 3,710-713.
- Kragelund, B. B., Jensen, M. K. y Skriver, K. (2012). Order by disorder in plant signaling. *Trends in Plant Science*, 17, 625-632.
- Kriško, A., Smole, Z., Debret, G., Nikolić, N. y Radman, M. (2010). Unstructured hydrophilic sequences in prokaryotic proteomes correlate with dehydration tolerance and host association. *Journal of Molecular Biology*, 402, 775-782.
- Kumar, A., Dutt, S., Bagler, G., Ahuja, P. S. y Kumar, S. (2012). Engineering a thermo-stable superoxide dismutase functional at sub-zero to >50°C, which also tolerates autoclaving. *Scientific Reports*, 2, 387. doi: <https://doi.org/10.1038/srep00387>
- Kumar, S. y Nussinov, R. (2001). How do thermophilic proteins deal with heat? *Cellular and Molecular Life Sciences*. 58, 1216-1233
- Kurotani, A., Tokmakov, A. A., Kuroda, Y., Fukami, Y., Shinozaki, K. y Sakurai, T. (2014). Correlations between predicted protein disorder and post-translational modifications. *Bioinformatics*, 30, 1095-1103.
- Lafta, A. M. y Lorenzen, J. H. (1995). Effect of high temperature on plant growth and carbohydrate metabolism in potato. *Plant Physiology*, 109, 637-643.
- Le, D. T., Nishiyama, R., Watanabe, Y., Mochida, K., Yamaguchi-Shinozaki, K., Shinozaki, K. y Tran, L. S. P. (2011). Genome-wide survey and expression analysis of the plant-specific NAC transcription factor family in soybean during development and dehydration stress. *DNA Research*, 18, 263-276.

- Lee, C. Y. y Pennesi, A. P. (1984). Isolation and further characterization of a heat resistant peroxidase isoenzyme from cauliflower. *Journal of Food Science*, 49, 1616 -1617.
- Lemaire, S., Keryer, E., Stein, M., Schepens, I., Issakidis-Bourguet, E., Gérard-Hirne, C., *et al.* (1999). Heavy-metal regulation of thioredoxin gene expression in *Chlamydomonas reinhardtii*. *Plant Physiology*, 120, 773-778.
- Letunic I, Bork P (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research*, 44, W242-245.
- Lepock, J. R., Arnold, L. D., Torrie, B. H., Andrews, B. y Kruuv, J. (1985). Structural analyses of various Cu²⁺, Zn²⁺-superoxide dismutases by differential scanning calorimetry and Raman spectroscopy. *Archives of Biochemistry and Biophysics*, 241, 243-51.
- Li, J. S. y Vasal, S. K. (2015). Maize: quality protein maize. En *Encyclopedia of Food Grains: Second Edition*. <https://doi.org/10.1016/B978-0-12-394437-5.00223-0>
- Li, W. y Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22, 1658-9.
- Lin, Y. S., Hsu, W. L., Hwang, J. K. y Li, W. H. (2007). Proportion of solvent-exposed amino acids in a protein and rate of protein evolution. *Molecular Biology and Evolution*, 24, 1005-1011.
- Ling, Z., Brockmüller, T., Baldwin, I. T. y Xu, S. (2019). Evolution of alternative splicing in eudicots. *Frontiers in Plant Science*, 10, 707. doi: <https://doi.org/10.3389/fpls.2019.00707>
- Lipman, D. J., Souvorov, A., Koonin, E. V., Panchenko, A. R. y Tatusova, T. A. (2002). The relationship of protein conservation and sequence length. *BMC Evolutionary Biology*, 2, 20. doi: <https://doi.org/10.1186/1471-2148-2-20>
- Liu, H., Hu, M., Wang, Q., Cheng, L. y Zhang, Z. (2018). Role of papain-like cysteine proteases in plant development. *Frontiers in Plant Science*, 9, 1717. doi: <https://doi.org/10.3389/fpls.2018.01717>
- Liu, J., Perumal, N. B., Oldfield, C. J., Su, E. W., Uversky, V. N. y Dunker, A. K. (2006). Intrinsic disorder in transcription factors. *Biochemistry*, 45, 6873-6888.
- Liu, Yang, Li, D., Song, Q., Zhang, T., Li, D. y Yang, X. (2019). The maize late embryogenesis abundant protein ZmDHN13 positively regulates copper tolerance in transgenic yeast and tobacco. *Crop Journal*, 7, 403-410.

- Liu, Y., Wang, L., Xing, X., Sun, L., Pan, J., Kong, X., *et al.* (2013). ZmLEA3, a multifunctional group 3 LEA protein from maize (*Zea mays* L.), is involved in biotic and abiotic stresses. *Plant and Cell Physiology*, 54, 944-959.
- Liu, Y., Wu, J., Sun, N., Tu, C., Shi, X., Cheng, H., *et al.* (2017). Intrinsically disordered proteins as important players during desiccation stress of soybean radicles. *Journal of Proteome Research*, 16, 2393-2409.
- Liu, Z. H., Ji, H. Q., Cui, Z. T., Wu, X., Duan, L. J., Feng, X. X. y Tang, J. H. (2011). QTL detected for grain-filling rate in maize using a RIL population. *Molecular Breeding*. 27, 25-36.
- Liu, Z. y Huang, Y. (2014). Advantages of proteins being disordered. *Protein Science*, 23, 539-50.
- Lobanov, M. Y. y Galzitskaya, O. V. (2015). How common is disorder? Occurrence of disordered residues in four domains of life. *International Journal of Molecular Sciences*, 16, 19490-19507.
- Love, J. J., Li, X., Chung, J., Dyson, H. J. y Wright, P. E. (2004). The LEF-1 high-mobility group domain undergoes a disorder-to-order transition upon formation of a complex with cognate DNA. *Biochemistry*, 43, 8725-8734.
- Lu, Y., Sun, J., Yang, Z., Zhao, C., Zhu, M., Ma, D., *et al.* (2019). Genome-wide identification and expression analysis of glycine-rich RNA-binding protein family in sweet potato wild relative *Ipomoea trifida*. *Gene*, 686, 177-186.
- Manfre, A. J., LaHatte, G. A., Climer, C. R. y Marcotte, W. R. (2009). Seed dehydration and the establishment of desiccation tolerance during seed maturation is altered in the *Arabidopsis thaliana* mutant *atem6-1*. *Plant and Cell Physiology*, 50, 243-53.
- Marin, M., Uversky, V. N. y Ott, T. (2013). Intrinsic disorder in pathogen effectors: protein flexibility as an evolutionary hallmark in a molecular arms race. *Plant Cell*, 25, 3153-3157.
- Mastrangelo, A. M., Marone, D., Laidò, G., De Leonardis, A. M. y De Vita, P. (2012). Alternative splicing: Enhancing ability to cope with stress via transcriptome plasticity, *Plant Science*, 185-186:40-9.

- McCarthy, D. J., Chen, Y. y Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, 40, 4288-4297.
- Méchin, V., Balliau, T., Château-Joubert, S., Davanture, M., Langella, O., Négroni, L., *et al.* (2004). A two-dimensional proteome map of maize endosperm. *Phytochemistry*, 65, 1609-1618.
- Mészáros, B., Simon, I. y Dosztányi, Z. (2009). Prediction of protein binding regions in disordered proteins. *PLoS Computational Biology*, 5, e1000376.
- Metsalu, T. y Vilo, J. (2015). ClustVis: A web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. *Nucleic Acids Research*, 43, W566-570.
- Michalak, M., Groenendyk, J., Szabo, E., Gold, L. I. y Opas, M. (2009). Calreticulin, a multi-process calcium-buffering chaperone of the endoplasmic reticulum. *Biochemical Journal*, 417, 651-666.
- Miller, G. y Mittler, R. (2006). Could heat shock transcription factors function as hydrogen peroxide sensors in plants? *Annals of Botany*, 98, 279-288.
- Mishra, S. K., Tripp, J., Winkelhaus, S., Tschiersch, B., Theres, K., Nover, L. y Scharf, K. D. (2002). In the complex family of heat stress transcription factors, HsfA1 has a unique role as master regulator of thermotolerance in tomato. *Genes and Development*, 16, 1555-1567.
- Mistry, J., Coggill, P., Eberhardt, R. Y., Deiana, A., Giansanti, A., Finn, R. D., *et al.* (2013). The challenge of increasing Pfam coverage of the human proteome. *Database*, 2013, bat023, <https://doi.org/10.1093/database/bat023>
- Mitra, J., Xu, G., Wang, B., Li, M. y Deng, X. (2013). Understanding desiccation tolerance using the resurrection plant *Boea hygrometrica* as a model system. *Frontiers in Plant Science*, 4, 446. <https://doi.org/10.3389/fpls.2013.00446>
- Mittal, J., Yoo, T. H., Georgiou, G. y Truskett, T. M. (2013). Structural ensemble of an intrinsically disordered polypeptide. *Journal of Physical Chemistry B*, 117, 118-124.
- Molina-Rueda, J. J., Tsai, C. J. y Kirby, E. G. (2013). The populus superoxide dismutase gene family and its responses to drought stress in transgenic poplar overexpressing a pine cytosolic glutamine synthetase (GS1a). *PLoS ONE*, 8, e56421.

- Moore, J. P., Le, N. T., Brandt, W. F., Driouich, A. y Farrant, J. M. (2009). Towards a systems-based understanding of plant desiccation tolerance. *Trends in Plant Science*, 14, 110-117.
- Moore, J. P., Vicié-Gibouin, M., Farrant, J. M. y Driouich, A. (2008). Adaptations of higher plant cell walls to water loss: Drought vs desiccation. *Physiologia Plantarum*, 134, 237-245.
- Motto, M., Thompson, R. y Salamini, F. (1997). Genetic regulation of carbohydrate and protein accumulation in seeds. En: Larkins B.A., Vasil I.K. (eds) Cellular and molecular biology of plant seed development. Advances in cellular and molecular biology of plants, vol 4. Springer, Dordrecht. pp 479-522.
- Mouillon, J.-M., Eriksson, S. K. y Harryson, P. (2008). Mimicking the plant cell Interior under water stress by macromolecular crowding: disordered dehydrin proteins are highly resistant to structural collapse. *Plant Physiology*, 148, 1925-1937.
- Ntountoumi, C., Vlastaridis, P., Mossialos, D., Stathopoulos, C., Iliopoulos, I., Promponas, V., *et al.* (2019). Low complexity regions in the proteins of prokaryotes perform important functional roles and are highly conserved. *Nucleic Acids Research*, 47, 9998-10009
- Nuruzzaman, M., Sharoni, A. M. y Kikuchi, S. (2013). Roles of NAC transcription factors in the regulation of biotic and abiotic stress responses in plants. *Frontiers in Microbiology*, 4, 248. doi: <https://doi.org/10.3389/fmicb.2013.00248>
- Oliveros, J. C. 2007. VENNY. An interactive tool for comparing lists with Venn Diagrams. [Online] Disponible en: <http://bioinfogp.cnb.csic.es/tools/venny/index.html>. Acceso: 22 de nov 2019.
- Olvera-Carrillo, Y., Luis Reyes, J. y Covarrubias, A. A. (2011). Late embryogenesis abundant proteins. *Plant Signaling & Behavior*, 6, 586-589.
- Palombo, M., Bonucci, A., Etienne, E., Ciurli, S., Uversky, V. N., Guigliarelli, B., *et al.* (2017). The relationship between folding and activity in UreG, an intrinsically disordered enzyme. *Scientific Reports*. 7, 5977. <https://doi.org/10.1038/s41598-017-06330-9>
- Panca, R. y Tompa, P. (2012). Structural disorder in eukaryotes. *PLoS ONE*, 7, e34687. doi: <https://doi.org/10.1371/journal.pone.0034687>

- Park, S. J., Kwak, K. J., Oh, T. R., Kim, Y. O. y Kang, H. (2009). Cold shock domain proteins affect seed germination and growth of *Arabidopsis thaliana* under abiotic stress conditions. *Plant and Cell Physiology*, 50, 869-878.
- Pazos, F., Pietrosevoli, N., García-Martín, J. A. y Solano, R. (2013). Protein intrinsic disorder in plants. *Frontiers in Plant Science*, 4, 363-368.
- Peng, K., Radivojac, P., Vucetic, S., Dunker, A. K. y Obradovic, Z. (2006). Length-dependent prediction of protein in intrinsic disorder. *BMC Bioinformatics*, 7, 2006. doi: <https://doi.org/10.1186/1471-2105-7-208>
- Peng, Z., Mizianty, M. J., Xue, B., Kurgan, L. y Uversky, V. N. (2012). More than just tails: Intrinsic disorder in histone proteins. *Molecular BioSystems*, 8, 1886-1901.
- Peng, Z., Yan, J., Fan, X., Mizianty, M. J., Xue, B., Wang, K., Gang H., Uversky, V., Kurgan, L. (2014). Exceptionally abundant exceptions: Comprehensive characterization of intrinsic disorder in all domains of life. *Cellular and Molecular Life Sciences*, 72, 137-151.
- Pereira-Santana, A., Alcaráz, L. D., Castaño, E., Sánchez-Calderon, L., Sánchez-Teyer, F. y Rodríguez-Zapata, L. (2015). Comparative genomics of NAC transcriptional factors in angiosperms: Implications for the adaptation and diversification of flowering plants. *PLoS ONE*, 10, e0141866. <https://doi.org/10.1371/journal.pone.0141866>
- Pereira-Santana, A., Alvarado-Robledo, E. J., Zamora-Briseño, J. A., Ayala-Sumuano, J. T., Gonzalez-Mendoza, V. M., Espadas-Gil, F., et al. (2017). Transcriptional profiling of sugarcane leaves and roots under progressive osmotic stress reveals a regulated coordination of gene expression in a spatiotemporal manner. *PLoS ONE*, 12, e0189271. doi: <https://doi.org/10.1371/journal.pone.0189271>
- Pietrosevoli, N., García-Martín, J. A., Solano, R. y Pazos, F. (2013). Genome-wide analysis of protein disorder in *Arabidopsis thaliana*: implications for plant environmental adaptation. *PLoS ONE*, 8, e55524. <https://doi.org/10.1371/journal.pone.0055524>
- Plieth, C. y Vollbehr, S. (2012). Calcium promotes activity and confers heat stability on plant peroxidases. *Plant Signaling and Behavior*, 7, 650-660.
- Popelkova, H. y Yocum, C. F. (2011). PsbO, the manganese-stabilizing protein: Analysis of the structure-function relations that provide insights into its role in photosystem II. *Journal of Photochemistry and Photobiology B: Biology*, 104, 179-90.

- Prioul, J. L., Méchin, V., Lessard, P., Thévenot, C., Grimmer, M., Chateau-Joubert, S., ... Edwards, K. J. (2008). A joint transcriptomic, proteomic, and metabolic analysis of maize endosperm development and starch filling. *Plant Biotechnology Journal*, 6, 855-869.
- Qiu, Y., Xi, J., Du, L. y Poovaiah, B. W. (2012). The function of calreticulin in plant immunity: New discoveries for an old protein. *Plant Signaling and Behavior*, 7, 907-910.
- R Development Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria [Online]. Disponible en: URL <http://www.R-project.org/>.
- Radivojac, P. (2003). Protein flexibility and intrinsic disorder. *Protein Science*. 13, 71-80.
- Rajalingam, D., Loftis, C., Xu, J. J. y Kumar, T. K. S. (2009). Trichloroacetic acid-induced protein precipitation involves the reversible association of a stable partially structured intermediate. *Protein Science*, 18, 980-993.
- Reeves, R. (2009). HMG nuclear proteins: Linking chromatin structure to cellular phenotype. *Biochimica et Biophysica Acta*, 1799, 3-14.
- Rhodes, D. y Hanson, A. D. (2003). Quaternary ammonium and tertiary sulfonium compounds in higher plants. *Annual Review of Plant Physiology and Plant Molecular Biology*, 44, 357-384.
- Riley, A. C., Ashlock, D. A. y Graether, S. P. (2019). Evolution of the modular, disordered stress proteins known as dehydrins. *PLoS ONE* 14, e0211813. <https://doi.org/10.1371/journal.pone.0211813>
- Roberts, J. K., DeSimone, N. A., Lingle, W. L. y Dure, L. (2007). Cellular concentrations and uniformity of cell-type accumulation of two lea proteins in cotton embryos. *Plant Cell*, 5, 769-780.
- Robinson, M. D., McCarthy, D. J. y Smyth, G. K. (2009). edgeR: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26, 139-40.
- Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J. y Dunker, A. K. (2001). Sequence complexity of disordered protein. *Proteins: Structure, Function and Genetics*, 42, 38-48.

- Ruchko, M. V., Gorodnyaya, H. M. y Sorochinsky, B. V. (1998). Plant metallothioneins. *Ukrain'skyi Biokhimichni Zhurnal*, 70, 33-34.
- Saraiva, J. A., Cláudia, C. S., Nunes, S. y Coimbra, M. A. (2007). Purification and characterization of olive (*Olea europaea* L.) peroxidase – Evidence for the occurrence of a pectin binding peroxidase. *Food Chemistry*, 101, 1571-1579.
- Schad, E., Tompa, P. y Hegyi, H. (2011). The relationship between proteome size, structural disorder and organism complexity. *Genome Biology*, 12, R120. <https://doi.org/10.1186/gb-2011-12-12-r120>
- Scheeff, E. D. y Fink, J. L. (2003). Fundamentals of protein structure. *Methods of Biochemical Analysis*, 44, 15-39.
- Schmidt, R. J., Ketudat, M., Aukerman, M. J. y Hoschek, G. (2007). Opaque-2 is a transcriptional activator that recognizes a specific target site in 22-kd zein genes. *Plant Cell*, 6, 237-249.
- Seet, B. T., Dikic, I., Zhou, M.-M. y Pawson, T. (2006). Reading protein modifications with interaction domains. *Nature reviews. Molecular Cell Biology*, 7, 473-483.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., *et al.* (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13, 2498-2504. <https://doi.org/10.1101/gr.1239303>
- Shi, H. Y., Zhang, Y. X. y Chen, L. (2013). Two pear auxin-repressed protein genes, PpARP1 and PpARP2, are predominantly expressed in fruit and involved in response to salicylic acid signaling. *PLoS ONE*, 9, e89926. doi: <https://doi.org/10.1007/s11240-013-0321-3>
- Shin, D. S., DiDonato, M., Barondeau, D. P., Hura, G. L., Hitomi, C., Berglund, J. A., *et al.* (2009). Superoxide dismutase from the eukaryotic thermophile *Alvinella pompejana*: structures, stability, mechanism, and insights into amyotrophic lateral sclerosis. *Journal of Molecular Biology*, 385, 1534-1555.
- Siegel, B. Z. (1993). Plant peroxidases-an organismic perspective. *Plant Growth Regulation*, 12, 303-312.
- Silva-Sanchez, C., Chen, S., Li, J. y Chourey, P. S. (2014). A comparative glycoproteome study of developing endosperm in the hexose-deficient miniature1 (mn1) seed mutant and its wild type Mn1 in maize. *Frontiers in Plant Science*, 5, 63. <https://doi.org/10.3389/fpls.2014.00063>

- Simpson, B. B. y Niklas, K. J. (2006). The evolutionary biology of plants. *Systematic Botany*, 22, 727. doi: <https://doi.org/10.2307/2419438>
- Skupien-Rabian, B., Jankowska, U., Swiderska, B., Lukasiewicz, S., Ryszawy, D., Dziedzicka-Wasylewska, M. y Kedracka-Krok, S. (2015). Proteomic and bioinformatic analysis of a nuclear intrinsically disordered proteome. *Journal of Proteomics*, 130, 76-84.
- Slama-Schwok, A., Zakrzewska, K., Léger, G., Leroux, Y., Takahashi, M., Käs, E. y Debey, P. (2000). Structural changes induced by binding of the high-mobility group I protein to a mouse satellite DNA sequence. *Biophysical Journal*, 78, 2543-2559.
- Song, R., Llaca, V., Linton, E. y Messing, J. (2001). Sequence, regulation, and evolution of the maize 22-kD alpha-zein gene family. *Genome Research*, 11, 1817-1825.
- Stender, E. G., O'Shea, C. y Skriver, K. (2015). Subgroup-specific intrinsic disorder profiles of arabidopsis NAC transcription factors: Identification of functional hotspots. *Plant Signaling and Behavior*, 10, e1010967.
- Su, G., Kuchinsky, A., Morris, J. H., States, D. J. y Meng, F. (2010). GLay: Community structure analysis of biological networks. *Bioinformatics*, 26, 3135-3137.
- Sun, X., Rikkerink, E. H. A., Jones, W. T. y Uversky, V. N. (2013). Multifarious roles of intrinsic disorder in proteins illustrate its broad impact on plant biology. *Plant Cell*, 25, 38-55.
- Sun, X., Xue, B., Jones, W. T., Rikkerink, E., Dunker, A. K. y Uversky, V. N. (2011). A functionally required unfoldome from the plant kingdom: Intrinsically disordered N-terminal domains of GRAS proteins are involved in molecular recognition during plant development. *Plant Molecular Biology*, 77, 205-23.
- Tantos, A. y Tompa, P. (2012). Proteomic methods for the identification of intrinsically disordered proteins. *Methods in Molecular Biology*, 896, 429-37.
- Theißen, G. y Gramzow, L. (2016). Chapter 8 - Structure and Evolution of Plant MADS Domain Transcription Factors A2 - Gonzalez, Daniel H. BT - Plant Transcription Factors. En *Plant Transcription Factors*. pp 127-138.
- Thieulin-Pardo, G., Avilan, L., Kojadinovic, M. y Gontero, B. (2015). Fairy tails: flexibility and function of intrinsically disordered extensions in the photosynthetic world. *Frontiers in Molecular Biosciences*, 2, 23. doi: <https://doi.org/10.3389/fmolb.2015.00023>

- Thomann, E. B., Sollinger, J., White, C. y Rivin, C. J. (2008). Accumulation of group 3 late embryogenesis abundant proteins in *Zea mays* embryos: roles of abscisic acid and the viviparous-1 gene product. *Plant Physiology*, 99, 607-614.
- Tompa, P. (2002). Intrinsically unstructured proteins. *Trends in Biochemical Sciences*, 27, 527-533.
- Tompa, P. y Csermely, P. (2004). The role of structural disorder in the function of RNA and protein chaperones. *FASEB Journal*, 18, 1169-1175.
- Tompa, P. y Kovacs, D. (2010). Intrinsically disordered chaperones in plants and animals. *Biochemistry & Cell Biology*, 88, 167-174.
- Tompa, P., Szász, C. y Buday, L. (2005). Structural disorder throws new light on moonlighting. *Trends in Biochemical Sciences*, 30, 484-489
- Törönen, P., Medlar, A. y Holm, L. (2018). PANNZER2: A rapid functional annotation web server. *Nucleic Acids Research*, 46, W84-W88.
- Toxopeus, J., Warner, A. H. y MacRae, T. H. (2014). Group 1 LEA proteins contribute to the desiccation and freeze tolerance of *Artemia franciscana* embryos during diapause. *Cell Stress and Chaperones*, 19, 939-948.
- Tunnacliffe, A., Hinch, D. K., LePrince, O. y MacHerel, D. (2010). LEA proteins: Versatility of form and function. *Topics in Current Genetics*, 21, 91-108.
- Uversky, V. N. (2002). Natively unfolded proteins: A point where biology waits for physics. *Protein Science*, 11, 739-56.
- Uversky, V. N. (2011). Intrinsically disordered proteins from A to Z. *International Journal of Biochemistry and Cell Biology*, 43, 1090-1103.
- Uversky, V.N. (2013). Unusual biophysics of intrinsically disordered proteins. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1834, 932-951.
- Uversky, V. N. y Dunker, A. K. (2010). Understanding protein non-folding. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1804, 1231-1264.
- Uversky, V. N., Gillespie, J. R. y Fink, A. L. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins: Structure, Function and Genetics*, 41, 415-427.

- Vacic, V., Uversky, V. N., Dunker, A. K. y Lonardi, S. (2007). Composition Profiler: A tool for discovery and visualization of amino acid composition differences. *BMC Bioinformatics*, 8, 1-7. <https://doi.org/10.1186/1471-2105-8-211>
- van Bavel, C. H. M. (1996). Water relations of plants and soils. *Soil Science*, 161, 257-260
- van der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., *et al.* (2014). Classification of intrinsically disordered regions and proteins. *Chemical Reviews*, 114, 6589-5631.
- Varricchio, L., Falchi, M., Dall'Ora, M., De Benedittis, C., Ruggeri, A., Uversky, V. N. y Migliaccio, A. R. (2017). Calreticulin: challenges posed by the intrinsically disordered nature of calreticulin to the study of its function. *Frontiers in Cell and Developmental Biology*, 5, 96.
- Vernoud, V., Hajduch, M., Khaled, A. S., Depège, N. y Rogowsky, P. M. (2005). Maize embryogenesis. *Maydica*, 50, 469-483.
- Vetriani, C., Maeder, D. L., Tolliday, N., Yip, K. S. P., Stillman, T. J., Britton, K. L., *et al.* (1998). Protein thermostability above 100°C: A key role for ionic interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 12300-12305.
- Vincent, M. y Schnell, S. (2016). A collection of intrinsic disorder characterizations from eukaryotic proteomes. *Scientific Data*, 3, 160045 <https://doi.org/10.1038/sdata.2016.45>
- Walsh, I., Martin, A. J. M., Di domenico, T. y Tosatto, S. C. E. (2012). Espritz: Accurate and fast prediction of protein disorder. *Bioinformatics*, 28, 503-509. <https://doi.org/10.1093/bioinformatics/btr682>
- Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. y Jones, D. T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *Journal of Molecular Biology*, 337, 635-645. doi: <https://doi.org/10.1016/j.jmb.2004.02.002>
- Weids, A. J., Ibstedt, S., Tamás, M. J. y Grant, C. M. (2016). Distinct stress conditions result in aggregation of proteins with similar properties. *Scientific Reports*, 6, 24554. <https://doi.org/10.1038/srep24554>
- Wetzler, D. E., Fuchs Wightman, F., Bucci, H. A., Rinaldi, J., Caramelo, J. J., Iusem, N. D. y Ricardi, M. M. (2018). Conformational plasticity of the intrinsically disordered protein *asr1* modulates its function as a drought stress-responsive gene. *PLoS ONE*, 13, e0202808. doi: <https://doi.org/10.1371/journal.pone.0202808>

- Williams, R. M., Obradovi, Z., Mathura, V., Braun, W., Garner, E. C., Young, J., *et al.* (2001). The protein non-folding problem: amino acid determinants of intrinsic order and disorder. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 2001, 89-100.
- Wilson, C. M. (2008). Serial analysis of zein by isoelectric focusing and sodium dodecyl sulfate gel electrophoresis. *Plant Physiology*, 82, 196-202.
- Woo, Y. M., Hu, D. W., Larkins, B. A. y Jung, R. (2001). Genomics analysis of genes expressed in maize endosperm identifies novel seed proteins and clarifies patterns of zein gene expression. *Plant Cell*, 13, 2297-2318.
- Wright, P. E. y Dyson, H. J. (2015). Intrinsically disordered proteins in cellular signalling and regulation. *Nature reviews. Molecular Cell Biology*, 16, 18-29.
- Wu, L., Yu, M., Holowachuk, J., Sharpe, A., Lydiate, D., Hegedus, D. y Gruber, M. (2017). Evaluation of a *Brassica napus* auxin-repressed gene induced by flea beetle damage and *Sclerotinia sclerotiorum* infection. *American Journal of Plant Sciences*, 8, 1921-1957.
- Xue, B., Dunker, A. K. y Uversky, V. N. (2012). Orderly order in protein intrinsic disorder distribution: Disorder in 3500 proteomes from viruses and the three domains of life. *Journal of Biomolecular Structure and Dynamics*, 30, 137-149.
- Xue, B., Williams, R. W., Oldfield, C. J., Dunker, A. K. y Uversky, V. N. (2010). Archaic chaos: intrinsically disordered proteins in Archaea. *BMC Systems Biology*, 4, S1-S21.
- Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., *et al.* (2006). WEGO: A web tool for plotting GO annotations. *Nucleic Acids Research*, 46, W71-W75.
- Yi, F., Gu, W., Chen, J., Song, N., Gao, X., Zhang, X., Lai, J. (2019). High temporal-resolution transcriptome landscape of early maize seed development. *Plant Cell*, 31, 974-992.
- Ying, S., Zhang, D. F., Fu, J., Shi, Y. S., Song, Y. C., Wang, T. Y. y Li, Y. (2012). Cloning and characterization of a maize bZIP transcription factor, ZmbZIP72, confers drought and salt tolerance in transgenic *Arabidopsis*. *Planta*, 235, 253-266.
- Yizhi, Z., Hélène, L., Antoine, S., Régine, L. y Brigitte, G. (2018). Exploring intrinsically disordered proteins in *Chlamydomonas reinhardtii*. *Scientific Reports*, 8, 6805.

- Yogavel, M., Mishra, P. C., Gill, J., Bhardwaj, P. K., Dutt, S., Kumar, S., *et al.* (2008). Structure of a superoxide dismutase and implications for copper-ion chelation. *Acta Crystallographica Section D: Biological Crystallography*, D64, 892-901.
- Yruela, I. y Contreras-Moreira, B. (2012). Protein disorder in plants: A view from the chloroplast. *BMC Plant Biology*, 12, 1. <https://doi.org/10.1186/1471-2229-12-165>
- Yruela, I. y Contreras-Moreira, B. (2013). Genetic recombination is associated with intrinsic disorder in plant proteomes. *BMC Genomics*, 14,1. <https://doi.org/10.1186/1471-2164-14-772>
- Yruela, I., Oldfield, C. J., Niklas, K. J. y Dunker, A. K. (2017). Evidence for a strong correlation between transcription factor protein disorder and organismic complexity. *Genome Biology and Evolution*, 9, 1248-1265.
- Zamora-Briseño, J. A. y de Jiménez, E. S. (2016). A LEA 4 protein up-regulated by ABA is involved in drought response in maize roots. *Molecular Biology Reports*, 43, 221-228.
- Zamora-Briseño, J. A., Pereira-Santana, A., Reyes-Hernández, S. J., Castaño, E. y Rodríguez-Zapata, L. C. (2019). Global dynamics in protein disorder during maize seed development. *Genes*, 10, 502. <https://doi.org/10.3390/genes10070502>
- Zamora-Briseño, J. A., Reyes-Hernández, S. J. y Zapata, L. C. R. (2018). Does water stress promote the proteome-wide adjustment of intrinsically disordered proteins in plants? *Cell Stress and Chaperones*, 23, 807-812.
- Zang, Q. W., Wang, C. X., Li, X. Y., Guo, Z. A., Jing, R. L., Zhao, J. y Chang, X. P. (2010). Isolation and characterization of a gene encoding a polyethylene glycol-induced cysteine protease in common wheat. *Journal of Biosciences*, 35, 379-388.
- Zhang, Yixiang, Gao, P., Xing, Z., Jin, S., Chen, Z., Liu, L., Dai, S. Y. (2013). Application of an improved proteomics method for abundant protein cleanup: Molecular and genomic mechanisms study in plant defense. *Molecular and Cellular Proteomics*, 12, 3431-3442.
- Zhang, Yizhi, Launay, H., Schramm, A., Lebrun, R. y Gontero, B. (2018). Exploring intrinsically disordered proteins in *Chlamydomonas reinhardtii*. *Scientific Reports*, 8, 6805. doi: <https://doi.org/10.1038/s41598-018-24772-7>

Zhao, Y., Li, C., Ge, J., Xu, M., Zhu, Q., Wu, T., *et al.* (2014). Recessive mutation identifies auxin-repressed protein ARP1, which regulates growth and disease resistance in tobacco. *Molecular Plant-Microbe Interactions*, 27, 638-54.