



¿Cómo responden al déficit hídrico las plantas noduladas?

Una aproximación desde el estudio del traductoma

Mauro Martínez Moré

Tesis de Maestría en Ciencias Biológicas
Opción Biología Celular y Molecular

Directora: Dra. María Martha Sainz

Co-Directora: Dra. Carla Valeria Filippi

Laboratorio de Bioquímica

Facultad de Agronomía - UDELAR

Tribunal Evaluador

Dra. Luisa Berná

Dra. Luciana Fleitas

Dra. Joaquina Farías

RESUMEN

La soja obtiene la mayoría del nitrógeno que necesita mediante una interacción simbiótica con bacterias fijadoras de nitrógeno. El déficit hídrico es uno de los factores que más afecta el rendimiento del cultivo. La nodulación puede influenciar positivamente las respuestas al déficit hídrico en leguminosas gracias a modular el metabolismo del nitrógeno y de hormonas, el ajuste osmótico y los sistemas antioxidantes. Sin embargo, existe escasa evidencia sobre los mecanismos moleculares responsables de la alteración de la respuesta a déficit hídrico de una planta nodulada respecto a una no nodulada. A su vez, el control traduccional de la expresión génica es importante en condiciones de estrés y ha sido determinado como relevante en plantas bajo déficit hídrico. En este trabajo, se analizaron las respuestas diferenciales de raíces de soja, noduladas o no, en déficit hídrico a nivel transcripcional y/o traduccional. Por ende, se estudiaron los transcriptomas y traductomas de raíces en cuatro tratamientos combinados. Se encontró que las respuestas de plantas noduladas en déficit hídrico fueron más complejas a nivel de genes diferencialmente expresados (DEGs) totales en comparación con plantas no noduladas. Dicha complejidad se observó también a nivel de DEGs con regulación traduccional y de factores de transcripción DEGs. Además, se identificaron módulos de co-expresión génica que fueron asociados con procesos biológicos relevantes en las respuestas diferenciales de las plantas. La biosíntesis de metabolitos secundarios, particularmente de fenilpropanoides y flavonoides, se vio afectada negativamente por el déficit hídrico. A su vez, se reprogramó la vía de señalización mediada por MAPK y genes del metabolismo de la galactosa se sobreexpresaron en plantas noduladas bajo déficit hídrico. Notablemente, en estas condiciones, la regulación traduccional de genes participantes en los procesos de fosforilación oxidativa e iniciación de la traducción fue clave. Se identificaron genes candidatos robustos sujetos a control traduccional en raíces de plantas de soja noduladas sometidas a déficit hídrico para futuramente analizarlos funcionalmente. Para ello se puso a punto el sistema de transformación transitoria mediado por *Rhizobium rhizogenes*.

Palabras claves: soja, RNA-seq, transcriptoma, traductoma, WGCNA, nodulación, estrés hídrico

AGRADECIMIENTOS

A Martita y Carla por guiarme académica y personalmente durante todo el desarrollo de la tesis, por estar siempre disponibles para responder interrogantes e inquietudes, por tenerme en cuenta como integrante de sus respectivos proyectos de investigación, por permitir e incitarme a realizar pasantías y participar en eventos y por recibirme cálidamente en el laboratorio.

A Martita particularmente por la transferencia de su conocimiento de laboratorio y de biología molecular de plantas, y por su fortaleza de generar preguntas y discutir las causas teóricas de los resultados.

A Carla particularmente por el extenso conocimiento de herramientas bioinformáticas y enseñarme a utilizarlas eficientemente, cortando por lo sano cuando hay que hacerlo, y por ayudarme a razonar qué interpretación biológica tienen los análisis bioinformáticos.

Al Laboratorio de Bioquímica, particularmente a Mariana, Gastón y Selene, por las charlas, la sobremesa, el café y el apoyo constante.

Al tribunal por haber leído y colaborado sustancialmente con la solidez del manuscrito.

A todos los del INTA de Córdoba (Argentina) por haberme recibido de pasantía durante dos meses y haber compartido buenos momentos, en especial a Germán por haberme instruido en la técnica de transformación transitoria, la cual pude transferir satisfactoriamente al laboratorio.

A la Comisión Académica de Posgrado (CAP) por brindarme una beca de apoyo a la realización de la Maestría y a la Comisión Sectorial de Investigación Científica (CSIC) por financiar el proyecto en el cual se basa este trabajo.

Al Departamento de Biometría, Estadística y Computación (DBEC), particularmente a Virginia, por permitirme e insistirme a que me ausente de mis tareas laborales durante la finalización de este manuscrito, lo cual fue extremadamente necesario.

A mis viejos por regalarme el tiempo.

A todas las empresas de ómnibus que, por absurdo, me incitan a desarrollarme académica y profesionalmente a efectos de no recurrir a sus servicios diariamente.

Tabla de contenido

1. INTRODUCCIÓN.....	6
1.1 Leguminosas.....	6
1.2 Cultivo de soja en Uruguay	6
1.3 Simbiosis rizobio–leguminosa.....	6
1.4 Déficit Hídrico: percepción y respuestas de las plantas	10
1.5 Efectos de la Interacción de las Condiciones de Nodulación y Déficit Hídrico	12
1.6 Estructura Genómica de la Soja	13
1.7 Regulación de la Expresión Génica	13
1.7.1 Factores de Transcripción: importancia en las respuestas de la planta	14
1.7.2 Regulación Génica Postranscripcional	15
1.8 Herramientas de Evaluación de la Expresión Génica.....	17
1.8.1 Análisis de Redes de Co-expresión Génica Ponderadas (WGCNA)	20
1.8.2 Análisis de Expresión Diferencial de Genes	21
1.8.3 Análisis de Enriquecimiento Funcional	22
1.8.4 Predicción de Función Génica mediante EGAD	22
1.9 Análisis Funcional de Genes mediante Sistema de Transformación Transitoria	23
1.10 Hipótesis y Objetivos.....	25
2. MATERIALES Y MÉTODOS	26
2.1 Material vegetal, condiciones de crecimiento, nodulación y déficit hídrico	26
2.2 Purificación de fracciones de ARN total y asociado a polisomas.....	28
2.3 Extracción de ARN y secuenciación	29
2.4 Análisis de datos	30
2.4.1 Procesamiento de Datos de Secuenciación	30
2.4.2 Análisis de Redes de Co-expresión Génica Ponderadas (WGCNA)	30
2.4.3 Análisis de Expresión Diferencial de Genes (DEG)	31
2.4.4 Análisis WGCNA + DEGs	32
2.4.5 Análisis de Enriquecimiento GO.....	32
2.4.6 Análisis de Redes de Interacción Proteína-Proteína (PPI)	32
2.4.7 Búsqueda, Identificación y Análisis de Enriquecimiento de Motivos Génicos en Regiones 5'-UTR.....	33
2.4.8 Predicción y Validación de Función Génica	33

2.5 Puesta a Punto de Sistema de Transformación Transitoria mediado por <i>R. rhizogenes</i> para el Análisis Funcional de Genes.....	33
2.5.1 Selección de la Colonia Bacteriana Efectiva.....	35
2.5.2 Transformación de K599 con Plásmidos de Sobreexpresión y Silenciamiento por Electroporación.....	35
2.5.3 Infección de plántulas, confirmación y evaluación de la transformación	36
3. RESULTADOS	37
3.1 Caracterización hídrica de plantas de soja no noduladas y noduladas en diferente contexto hídrico	37
3.2 Análisis de Datos de Secuenciación	40
3.3 Análisis de Componentes Principales	43
3.4 Análisis de Expresión Diferencial de Genes	47
3.5 Identificación de DEGs preferenciales en cada tratamiento y nivel de regulación	50
3.6 Análisis de Red de Co-expresión Génica Ponderada (WGCNA)	52
3.7 Análisis de WGCNA+DEGs	55
3.8 Análisis de Enriquecimiento de Procesos Biológicos, Vías Metabólicas e Interacciones Proteína-Proteína en Módulos de Co-expresión	60
3.8.1 DEGs comunes a nodulación y comunes a estrés.....	60
3.8.2 DEGs preferenciales de cada contraste	63
3.9 Análisis de Factores de Transcripción diferencialmente expresados	75
3.9.1 Factores de Transcripción diferencialmente expresados a nivel traduccional con targets putativos diferencialmente expresados a nivel transcripcional	76
3.10 Predicción y Validación de Función Génica por EGAD.....	78
3.11 Puesta a Punto del Sistema de Transformación Transitoria para Análisis Funcional de Genes	80
4. DISCUSIÓN	85
5. CONCLUSIONES.....	94
6. REFERENCIAS BIBLIOGRÁFICAS	95
7. ANEXO	110

1. INTRODUCCIÓN

1.1 Leguminosas

La familia de plantas leguminosas (Fabaceae) incluye 745 géneros y más de 19.000 especies distribuidas mundialmente, siendo la tercera familia más numerosa de angiospermas (Wink 2013). Esta familia es importante debido a la significancia económica de varias de las especies que la componen, entre ellas diversos granos como la soja (*Glycine max* (L.) Merr.), quien tiene un valor de exportación global estimado en más de U\$S 200 billones (OECD/FAO 2021). Adicionalmente, muchas especies de leguminosas son de valor ecológico debido a la habilidad de fijación de nitrógeno. Esta habilidad, característica de la familia, surge de la formación de estructuras especiales adaptadas para la fijación de nitrógeno en las cuales se da una interacción simbiótica entre bacterias fijadoras de nitrógeno (rizobios) y células de la raíz de la planta. Se estima que un total de más de 60 millones de toneladas nitrógeno atmosférico son fijadas anualmente (Yu and Zhuang 2020), de los cuales aproximadamente 16 millones se deben a los cultivos sojeros (Nakei et al. 2022). Es así que las leguminosas tienen un rol significativo en la sostenibilidad ambiental y agronómica, ya que la capacidad fijadora de nitrógeno de los rizobios implica que éstos pueden ser un método barato y natural de fertilización nitrogenada ayudando a reducir la necesidad de usar fertilizantes sintéticos (Van Heerden et al. 2007; Gresshoff et al. 2015; Foyer et al. 2016; Ferguson et al. 2019).

1.2 Cultivo de soja en Uruguay

La soja (*Glycine max* (L.) Merr.) es de gran importancia agrícola a nivel mundial, siendo uno de los cultivos más importantes ya que su aporte en proteínas y aceites lo tornan valioso tanto para la alimentación humana como animal. A su vez, es considerada una especie valiosa para la obtención de biocombustibles y la conservación de los suelos por la capacidad de fijar nitrógeno atmosférico (Schmutz et al. 2010). A nivel mundial, el cultivo de soja tuvo una gran expansión en los últimos veinte años, fomentado en gran medida por la creciente demanda de alimentos de alto valor proteico. La producción mundial de granos de soja fue de 356 millones de toneladas en 2020 (OECD/FAO 2021). Según datos de la Dirección de Estadísticas Agropecuarias (DIEA-MGAP 2023) la soja es el cultivo de grano con mayor área de siembra en el Uruguay con un total de 1.009.899 hectáreas cultivadas (DIEA-MGAP 2023). La producción de soja en 2023 se estimó en aproximadamente 650 mil toneladas, lo que representó la cuarta parte de la cosecha de la zafra de 2022 cuando se obtuvieron alrededor de 2 millones de toneladas. El cultivo presentó bajo rendimiento en la zafra 2023 como resultado del déficit hídrico experimentado a nivel de todo el territorio uruguayo (DIEA-MGAP 2023). Entre los factores ambientales que más inciden en el desarrollo y rendimiento de la soja se encuentra justamente el déficit hídrico (Abdel-Haleem et al. 2011).

1.3 Simbiosis rizobio-leguminosa

La interacción simbiótica rizobio-leguminosa surgió hace aproximadamente 58 millones de años (Ferguson et al. 2019). El término genérico rizobio incluye los siguientes géneros: *Bradyrhizobium*, *Mesorhizobium*, *Rhizobium*, y *Ensifer* (anteriormente *Sinorhizobium*) (Concha and Doerner 2020; Nakei et al. 2022). Se ha observado que varias especies de estos cuatro géneros participan en la simbiosis con la

soja (Nakei et al. 2022). En Uruguay se recomienda el uso de las cepas U1301 y U1302 de *Bradyrhizobium elkanii* (Beyhaut et al. 2020).

El establecimiento de la simbiosis depende del reconocimiento específico entre los rizobios y la planta a través de una serie de interacciones moleculares. El proceso es iniciado por la planta hospedera mediante la exudación, desde sus raíces, de sustancias fenólicas tipo flavonoides, incluyendo flavonas, flavonoles e isoflavonas, que serán reconocidas por rizobios compatibles capaces de establecer la interacción simbiótica con la especie (Ferguson et al. 2010; Abdel-Lateif et al. 2012). La percepción de los flavonoides induce quimiotaxis en rizobios provocando la quimioatracción de ellos hacia la raíz (Subramanian et al. 2007; Ferguson et al. 2010). Asimismo, se da la activación de genes *nod* en rizobios que se traduce en la liberación de lipoquitinolígosacáridos, más conocidos como factores Nod (D'Haeze and Holsters 2002). Estos factores son determinantes de la especificidad en la interacción rizobio-leguminosa (Antolín-Llovera et al. 2012; Wang et al. 2019).

En soja, los factores Nod son percibidos por receptores NFR1 α /NFR5 α (Nod Factor Receptor), disparando una compleja cascada de señalización en la planta hospedera mediada por oscilaciones de calcio y quinasas, que conlleva la activación de factores de transcripción (por ejemplo, factores NSP1/NSP2 [*Nodulation Signalling Pathway*] o factores NIN [*Nodulation Inception*]) y de nodulinas tempranas como la ENOD40 (*Early Nodulin*) (Figura 1). Esto provoca alteraciones bioquímicas, fisiológicas y anatómicas en la raíz de la planta (Gage 2004; Ferguson et al. 2010).

El modo de infección más común es la formación de un hilo de infección en pelos radiculares en desarrollo, el cual involucra rediferenciación de células del pelo radicular para permitir la formación del hilo de infección. Cuando los pelos radiculares perciben la presencia de factores *Nod* ocurre una interrupción transitoria del crecimiento celular que provoca el retraimiento del extremo del pelo radicular y formación de un bolsillo de infección. Los rizobios quedan atrapados dentro del bolsillo de infección y continúan dividiéndose para luego formar un foco de infección dentro del bolsillo a partir del cual se empieza a desarrollar el hilo de infección (Gage 2004; Ferguson et al. 2010; Oldroyd et al. 2011). Es a través del hilo de infección que los rizobios continúan avanzando hacia el córtex interno de las raíces, donde son rodeados por una membrana peribacteroidal de la célula vegetal conllevando la organogénesis de nuevos órganos denominados nódulos (Figura 1)(Ferguson et al. 2010, 2019).

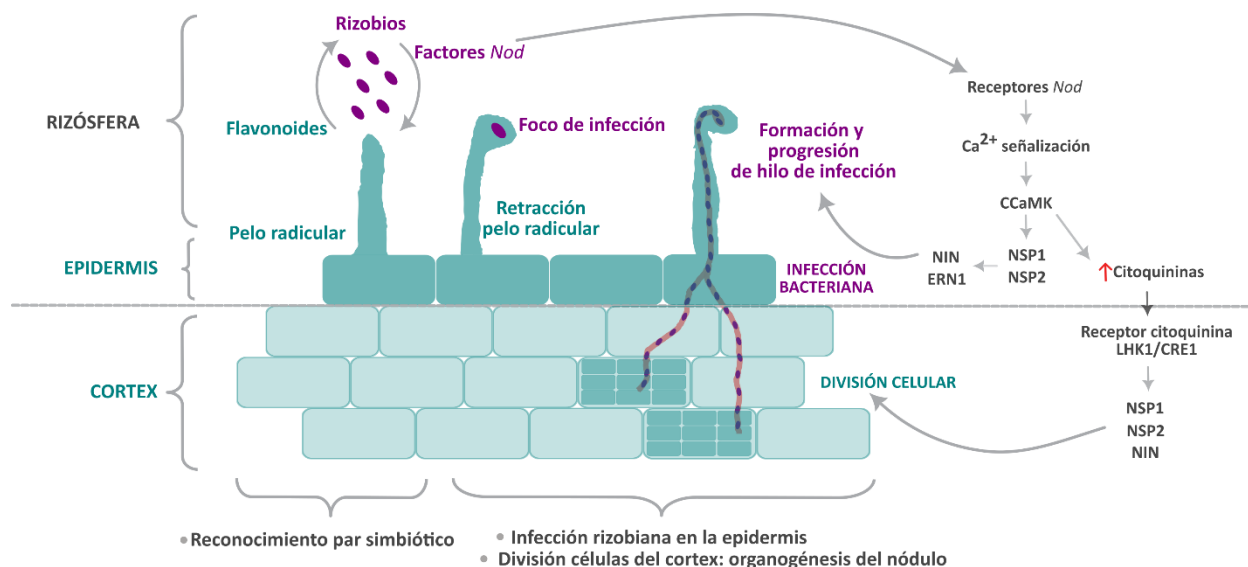


Figura 1 Principales eventos durante las etapas iniciales de la simbiosis rizobio-leguminosa. La interacción comienza en los pelos radiculares a través de un diálogo molecular entre flavonoides de la planta y factores *Nod* de los rizobios. Cuando la comunicación es efectiva, los rizobios se adhieren al pelo quien se retrae atrapando a las bacterias y formando un foco o bolsillo de infección. Se forma un hilo de infección que permite a los rizobios infectar el cortex, lugar donde se origina la división celular y origen del meristemo nodular (organogénesis del nódulo). **CCaMK**: Calcium-activated kinase. Factores de Transcripción **NSP** (*Nodulation Signalling Pathway*) y **NIN** (*Nodulation inception*). **ERN1**: *Early Nodulin 1*. Figura adaptada de (Sainz et al. 2024b).

La organogénesis del nódulo requiere de la integración del reconocimiento de factores *Nod* y de señalización mediada por hormonas, particularmente de auxinas y citoquininas (Figura 1). Con respecto a las citoquininas, se ha observado que la activación de señalización por citoquininas es necesaria y suficiente para la inducción de la morfogénesis del nódulo (Figura 1). Además, durante la nodulación, la señalización por citoquininas está restringida a células del periciclo y células corticales, donde ocurre la división celular durante la formación del nódulo (Oldroyd et al. 2011). Las células del periciclo y de la endodermis participan en el desarrollo del nódulo, pero no son infectadas por el rizobio (Xiao et al. 2017; Concha and Doerner 2020). Un resultado directo de esta vía de señalización por citoquininas es una alteración localizada del transporte polar de auxinas. En *Medicago truncatula* se ha observado una acumulación localizada de auxinas dada por una supresión localizada del transporte de auxinas e inducción de genes responsivos a auxinas como los genes *WOX5* y *YUCCA* (Oldroyd et al. 2011; Concha and Doerner 2020). El transporte de auxinas es restablecido una vez iniciado el proceso de infección (Oldroyd et al. 2011). En otros términos, el balance de auxinas y citoquininas dicta la organogénesis de nódulos (Oldroyd et al. 2011; Concha and Doerner 2020).

Los nódulos presentan una región periférica con células no invadidas y una región central donde se encuentran las células invadidas por rizobios. Los rizobios dentro del nódulo sufren cambios morfofisiológicos y bioquímicos transformándose en células irregulares que no se dividen, denominadas bacteroides (Kereszt et al. 2011). Según la morfología y desarrollo nodular se distinguen dos tipos de nódulos: indeterminados y determinados (Figura 2) (Kohlen et al. 2018; Concha and Doerner 2020). En nódulos indeterminados, los primeros eventos de división celular ocurren en células corticales internas vecinas a la endodermis (Figura 2A). Además, los nódulos indeterminados presentan un meristema persistente que resulta en nódulos con forma cilíndrica, como los observados en *Medicago sativa* o *M.*

truncatula. En la madurez, debido a la continua división celular, dichos nódulos contienen una población heterogénea de bacteroides formando así un gradiente de estadios de desarrollo mientras el nódulo continúa elongándose (Ferguson et al. 2010). Por otro lado, los nódulos determinados usualmente presentan forma esférica, no presentan meristemo persistente y no muestran un gradiente de estadios de desarrollo evidente. Dentro de las leguminosas que forman nódulos determinados se encuentran *Lotus japonicus* y la soja (Ferguson et al. 2010; Kohlen et al. 2018). Los primeros eventos de división celular en nódulos determinados surgen de células subepidérmicas en el córtex medio (*L. japonicus*) o externo (soja) (Figura 2B) (Ferguson et al. 2010; Oldroyd et al. 2011; Kohlen et al. 2018). Los nódulos determinados también forman estructuras denominadas lenticelas, las cuales sirven para potenciar el intercambio gaseoso (Ferguson et al. 2010). Otra importante diferencia entre nódulos indeterminados y determinados es que los procesos de infección del rizobio, fijación de nitrógeno y senescencia de la fijación se encuentran separados espacialmente en nódulos indeterminados (Figura 2C), mientras que dicha separación no ocurre en nódulos determinados (Figura 2D) (Kohlen et al. 2018).

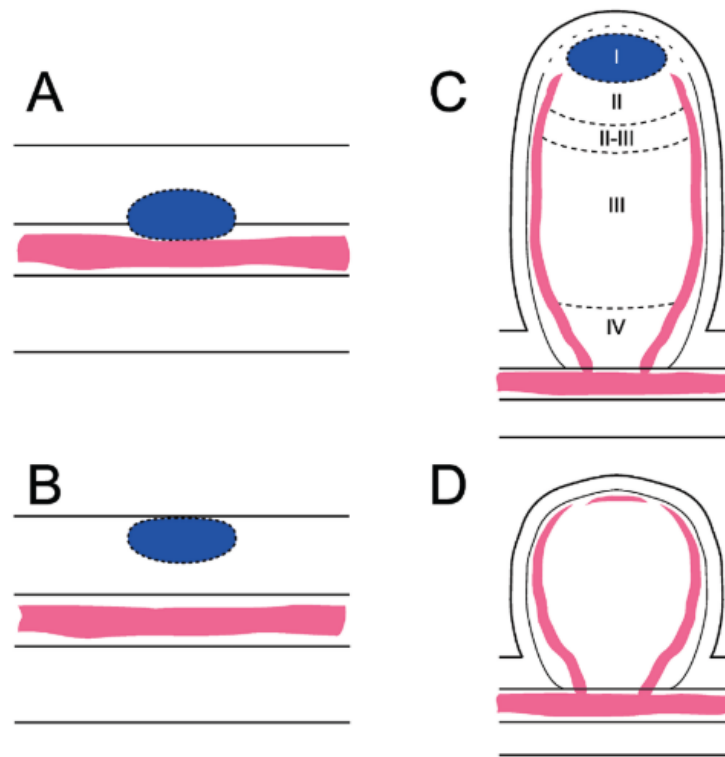


Figura 2 Nódulos indeterminados (A,C) y determinados (B,D). Posición de las primeras divisiones celulares (A,B), lugar que coincide con el aumento localizado de auxinas. Estructura nodular madura (C,D). En **azul**, tejido meristemático (divisiones celulares). En **rosado**, tejido vascular. Nódulos indeterminados (C) mantienen un meristema activo (I) seguido de una zona de infección (II), zona de transición (II-III), zona de fijación (III), y una zona de senescencia (IV) cuando el nódulo es viejo. Nódulos determinados (D) carecen de la distinción zonal. En estos nódulos, el proceso de senescencia comienza a partir del centro del nódulo. Zonas definidas por (Hirsch et al. 1997). Imagen tomada de (Kohlen et al. 2018).

El nódulo permite que exista un ambiente ideal para que suceda el proceso de fijación biológica de nitrógeno (FBN), el cual es llevado a cabo por los bacteroides. El establecimiento de la simbiosis provoca la síntesis de leghemoglobina, una hemoproteína con alta afinidad por el oxígeno, la cual permite el suministro del O_2 necesario como aceptor terminal de electrones de la cadena respiratoria de los bacteroides (Kunert et al. 2016; Concha and Doerner 2020). De esta manera se mantienen concentraciones bajas de este gas en el nódulo protegiendo a la enzima nitrogenasa, la cual está encargada de la conversión de N_2 en NH_3 y que se inactiva irreversiblemente por oxígeno (Kunert et al. 2016; Concha and Doerner 2020). A su vez, el nitrógeno fijado por los bacteroides difunde al espacio peribacteroidal como NH_3 , donde mediante protonación se transforma en amonio (NH_4^+) y pasa al citosol de la célula vegetal (Ferguson et al. 2019; Concha and Doerner 2020). Posteriormente, el nitrógeno es translocado vía xilema hacia la parte aérea de la planta predominantemente bajo formas orgánicas como amidas (asparaginas o glutaminas) en nódulos indeterminados y ureidos (alantoína o ácido alantoico) en nódulos determinados como en el caso de la soja (Van Heerden et al. 2007). Mediante este proceso, los bacteroides proveen de amonio (una forma asimilable de nitrógeno) a la planta hospedera, quien a cambio brinda carbohidratos a los bacteroides como fuente de carbono, siendo la sacarosa el principal fotosintato suministrado (Oldroyd et al. 2011; Ferguson et al. 2019; Concha and Doerner 2020).

La simbiosis y la FBN no solo potencian la disponibilidad de nitrógeno para la planta, sino que también estimulan la fijación de carbono mediante liberación de restricciones de la fotosíntesis. Esto sucede debido a que los bacteroides tienen una alta tasa de consumo de carbohidratos, lo que aumenta la fuerza de fosa de carbono en la raíz, removiendo así limitantes de fosas tanto en raíces como en hojas. De esta manera, se produce un aumento de la fotosíntesis y un desarrollo potenciado de ambos órganos (Concha and Doerner 2020). Por ejemplo, Kaschuk et al. (2012) reportan que en soja la simbiosis aumentó la fotosíntesis hasta en un 31% debido a un aumento de la fuerza de fosa de carbono. Esto significa que la FBN puede estimular tanto el crecimiento como la fotosíntesis de la planta (Kaschuk et al. 2010; Concha and Doerner 2020).

1.4 Déficit Hídrico: percepción y respuestas de las plantas

Las plantas, como organismos sésiles que son, deben lidiar con los constantes cambios ambientales que usualmente se tornan estímulos estresantes para el crecimiento y desarrollo (Zhu 2002, 2016a). El déficit hídrico es un tipo de estrés abiótico y una de las principales causas que limita el rendimiento de los cultivos agrícolas mundialmente (Pareek et al. 2010; Zhu 2016; Moloi and Ngara 2023). Algunos de los principales impactos de la sequía relacionados al rendimiento son alteraciones en estadios fenológicos, arquitectura radicular, producción de semilla, biomasa, índice de cosecha y tasa de germinación (Sánchez-Bermúdez et al. 2022).

Las plantas están sometidas a déficit hídrico cuando las lluvias disminuyen durante la temporada de crecimiento y cuando la tasa de transpiración excede la capacidad de absorción de agua por las raíces debido a condiciones de sequía (Moloi and Ngara 2023). Cómo las plantas sensan los estímulos estresantes y cómo se adaptan a condiciones ambientales son preguntas biológicas fundamentales (Zhu 2002, 2016; Pareek et al. 2010). En el caso de la sequía es conveniente distinguir entre efectos de estrés primarios y secundarios. El efecto primario causado por la sequía es la aparición de estrés hiperosmótico, el cual es referido simplemente como estrés osmótico ya que condiciones hipoosmóticas generalmente no resultan un problema significativo para las células vegetales (Zhu 2016). Los efectos secundarios de estrés por sequía son complejos e involucran daños causados por estrés oxidativo como disrupción de la

estructura y función celular, incluyendo integridad de membrana, fotosíntesis, respiración y procesos de crecimiento (Zhu 2002, 2016; Moloi and Ngara 2023). Aunque algunas respuestas celulares se derivan de efectos de estrés primarios, otras surgen principalmente de efectos secundarios (Zhu 2016). El grado de estos efectos depende de la duración y severidad de la restricción hídrica, especie de planta, genotipo y/o estadio de desarrollo (Moloi and Ngara 2023).

Las plantas han desarrollado diversos mecanismos para mantener el crecimiento, estructura y función celular ante un estímulo ambiental estresante prevaleciente (Osakabe et al. 2014; Moloi and Ngara 2023). Para mitigar los efectos adversos de la sequía sobre las células, las plantas acumulan proteínas y metabolitos y activan una variedad de genes relacionados a estrés mediante cambios en el metabolismo celular y regulación de la expresión génica (Shao et al. 2009; Moloi and Ngara 2023).

Una de las respuestas más tempranas a la sequía es la acumulación de la fitohormona ácido abscísico (ABA) provocada por los efectos hiperosmóticos (Pareek et al. 2010; Zhu 2016; Moloi and Ngara 2023). El ABA es sintetizado principalmente en las raíces e induce la activación de varias respuestas adaptativas en plantas como por ejemplo la disminución de la apertura estomática (Pareek et al. 2010; Zhu 2016). En este sentido, las raíces pueden ser consideradas como sensores primarios del déficit hídrico ya que están en contacto directo con el suelo (Sánchez-Bermúdez et al. 2022). Asimismo, algunos de los cambios moleculares sucedidos en respuesta a estrés pueden estar mediados por vías dependiente de ABA mientras que otros son mediados por vías independientes de ABA (Yoshida et al. 2014; Moloi and Ngara 2023). En ambos casos se ven alteradas las vías de señalización, la regulación génica y funciones de protección de la célula vegetal (Shinozaki and Yamaguchi-Shinozaki 2006; Moloi and Ngara 2023).

Aparte del ABA, otras fitohormonas como el ácido jasmónico, ácido salicílico, etileno, auxinas, giberelinas, citoquininas, brasinosteroides y moléculas peptídicas pequeñas también participan en las respuestas de la planta a sequía mediante complejas interacciones de sus vías de señalización (Pareek et al. 2010; Jogawat et al. 2021; Sánchez-Bermúdez et al. 2022; Moloi and Ngara 2023). Algunas de estas fitohormonas, como por ejemplo ABA, auxinas y etileno se acumulan en respuesta a déficit hídrico para facilitar, por ejemplo, el ajuste osmótico o la captura de especies oxidantes (Moloi and Ngara 2023). A su vez, la interacción entre fitohormonas puede tener efectos negativos o positivos en las hormonas interactuantes y por ende en alivianar los efectos producidos por el déficit hídrico (Jogawat et al. 2021; Moloi and Ngara 2023). Además, la biosíntesis, catabolismo y transporte de fitohormonas, y la conversión entre formas activas e inactivas, y almacenamiento pueden influenciar cómo la planta reprograma el metabolismo celular para maximizar la supervivencia ante condiciones ambientales desfavorables (Pareek et al. 2010; Moloi and Ngara 2023).

La acumulación de metabolitos osmoprotectores, productos del metabolismo secundario, para evitar pérdida de agua celular es otro cambio metabólico común en el déficit hídrico. Dentro de los osmoprotectores más comunes se encuentran la prolina, azúcares, ácido málico y otros compuestos como poliaminas (Pareek et al. 2010; Sánchez-Bermúdez et al. 2022; Ali et al. 2022). Por otro lado, bajo estrés abiótico se acumulan especies oxidantes producidas en el metabolismo celular que pueden actuar como mensajeros secundarios en las respuestas a estrés. Sin embargo, la acumulación de ellas puede resultar perjudicial para la célula causando daños en ADN, carbohidratos y proteínas (Foyer et al. 2016; Sánchez-Bermúdez et al. 2022). Las células vegetales contienen compuestos antioxidantes y enzimas para prevenir excesivas acumulaciones de especies oxidantes y así balancear las respuestas a estrés oxidativo. Algunos

de estos compuestos incluyen carotenoides, ascorbato, glutatión, y enzimas como la catalasa, la superóxido dismutasa y la ascorbato peroxidasa (Desikan 2003; Sánchez-Bermúdez et al. 2022).

1.5 Efectos de la Interacción de las Condiciones de Nodulación y Déficit Hídrico

En el caso de la simbiosis rizobio-leguminosa hay estudios realizados en distintas especies (*Medicago sativa* –alfalfa–, *M. truncatula*, *Phaseolus vulgaris* –poroto–, *Pisum sativum* –arveja–, *Lotus corniculatus*, entre otras) que sugieren que la condición de nodulación de una planta puede afectar sus estrategias de respuesta al déficit (Antolín et al. 1995; Lodeiro et al. 2000; Staudinger et al. 2016; Wang et al. 2016; Turetschek et al. 2020; Liu et al. 2022; Álvarez-Aragón et al. 2023; Lumactud et al. 2023). Por ejemplo, en alfalfa, Antolín et al. (1995) encontraron que la nodulación provocó una acumulación de osmolitos que se tradujo en plantas con mayor tolerancia a la sequía y una tasa fotosintética mayor durante el período de déficit hídrico en comparación con plantas no noduladas. Asimismo, la simbiosis con rizobios mejoró la tolerancia a estrés en alfalfa mediante aumento de la actividad de enzimas antioxidantes (Wang et al. 2016). Por otro lado, la mejora de la tolerancia a sequía en leguminosas nodulada podría estar dada por rizobios osmotolerantes y no por los cambios metabólicos producidos en la planta debido a la interacción simbiótica *per se* (Mhadhbi et al. 2011; Kunert et al. 2016; Rong Li et al. 2020). Por ejemplo, en *P. vulgaris*, la nodulación con una cepa de rizobios altamente tolerante a los efectos osmóticos aumentó la tolerancia al déficit hídrico en comparación con nodulaciones con cepas de menor tolerancia mediante la inducción de nódulos con un metabolismo del nitrógeno más eficiente y una disminución de la sobreproducción de especies oxidantes (Mhadhbi et al. 2011). A su vez, con dicha cepa, la estructura y función del nódulo se mantuvo más estables, principalmente gracias a modular los niveles de leghemoglobinas (Mhadhbi et al. 2011).

Si bien existe bibliografía que reporta cambios moleculares a nivel de raíz en plantas noduladas (Larrainzar et al. 2015; Yuan et al. 2016; Sakamoto et al. 2019), a nivel de nódulos (Lee et al. 2004; Manthey et al. 2004; Kant et al. 2016) y de nódulos bajo condiciones de sequía (Clement et al. 2008; Cilliers et al. 2018; Sańko-Sawczenko et al. 2019; da Silva et al. 2024), existe escasa, pero reciente, bibliografía que ahonde en los mecanismos moleculares responsables de la alteración de la respuesta a sequía de una planta nodulada respecto a una no nodulada. En este sentido, Álvarez-Aragón et al. (2023) encuentran, en arveja y poroto, genes diferencialmente expresados que son exclusivos en las respuestas al déficit hídrico de plantas noduladas. Dentro de ellos se encuentran genes vinculados a: metabolismo de compuestos fenólicos, como metilesterasas (MES1, MES3 y MES10), la glucosiltransferasa UGT74F1 y una oxigenasa tipo DMR6 (DLO1); procesos de desintoxicación, como tiorredoxinas y peroxidases; procesos osmóticos, como el gen STP13; y a la biosíntesis de homoserina, como la aspartoquinasa AK2 y la deshidrogenasa AKHSDH2 (Álvarez-Aragón et al. 2023). Por su parte, Sainz et al. (2024a) en soja, encuentran que genes vinculados al metabolismo del nitrato (por ej., transportadores de nitrato y enzimas que median su reducción), procesos de desintoxicación (por ej., nitrato reductasas como INR2 y NR) y procesos de óxido-reducción (por ej., amino-oxidasas, deshidrogenasas y tiorredoxinas) son regulados preferencialmente en plantas noduladas sometidas a déficit hídrico. Ambas investigaciones destacan la importancia de los factores de transcripción en la regulación de dichas respuestas (Álvarez-Aragón et al. 2023; Sainz et al. 2024a).

1.6 Estructura Genómica de la Soja

El genoma de soja tiene un tamaño de 1,1 gigabases (Gb), siendo el genoma de mayor tamaño secuenciado completamente en plantas, el cual fue obtenido mediante secuenciación del genotipo Williams 82 y que actualmente es utilizado como referencia (Schmutz et al. 2010). El genoma se organiza en 20 cromosomas conteniendo un total de 59148 genes, de los cuales aproximadamente el 80% codificaría para proteínas, un 8% serían pseudogenes, un 7% no codificantes y 5% ARNs pequeños (Schmutz et al. 2010; Sayers et al. 2022). Sin embargo, únicamente el 10% de ellos se encuentran validados experimentalmente, el restante 90% surge de predicciones computacionales sustentadas en homología de secuencias y anotaciones genómicas (Sayers et al. 2022). A su vez, el 25% de los genes se encuentra anotado como “no caracterizado”, es decir que no tiene una descripción del tipo de producto que genera. La cantidad de transcritos y proteínas, contemplando isoformas, es de aproximadamente 88000 y 74000, respectivamente (Sayers et al. 2022).

La soja está enriquecida en familias génicas de genes que contienen dominios de sitios de unión a nucleótidos (NB-ARC) y dominios de repetidos enriquecidos en leucina (LRR), ambos asociados con el sistema inmune de la planta (Schmutz et al. 2010). Asimismo, genes con dominios responsivos a auxinas, *F-box*, citocromo P450, quinasas, nodulina Nod19 y AP2, son dominios comunes en los genes de la especie (Schmutz et al. 2010).

Una característica resaltante de este genoma es que casi el 60% de las secuencias génicas se encuentran en regiones centroméricas (Schmutz et al. 2010). Se han identificado casi 40000 elementos repetidos en el genoma, que en conjunto con elementos truncados y otros fragmentos conforman aproximadamente el 60% del genoma. A su vez, aproximadamente el 42% del genoma de soja está conformado por retrotransposones con repetidos terminales largos. Por otro lado, respecto a las duplicaciones génicas, se sabe que en promedio más del 60% de los genes duplicados se encuentran en bloques que involucran dos cromosomas. Excepciones a ello son los genes duplicados ubicados en el cromosoma 14, que se encuentran distribuidos en 14 cromosomas restantes, y los ubicados en el cromosoma 20, que son altamente homólogos a los del cromosoma 10 (Schmutz et al. 2010). Con respecto a factores de transcripción, 3747 loci que codifican para 6150 factores de transcripción están identificados y clasificados en 57 familias, siendo las familias bHLH, MYB y bZIP las de mayor representación con 548, 430 y 352 factores de transcripción, respectivamente (Jin et al. 2017). La distribución genómica de los factores es relativamente homogénea a lo largo de cada cromosoma, presentando una abundancia relativa de entre 8 y 10% de genes codificantes para factores de transcripción en cada cromosoma (Schmutz et al. 2010).

1.7 Regulación de la Expresión Génica

La expresión génica en eucariotas es influenciada por una amplia variedad de mecanismos y estímulos (Brown 1981; Pareek et al. 2010). La regulación la expresión génica se da mayormente a tres niveles: a nivel de la cromatina, a nivel transcripcional y a nivel postranscripcional. A su vez, la regulación postranscripcional puede suceder a dos niveles: a nivel traduccional y a nivel postraduccional ([Figura 3](#)).

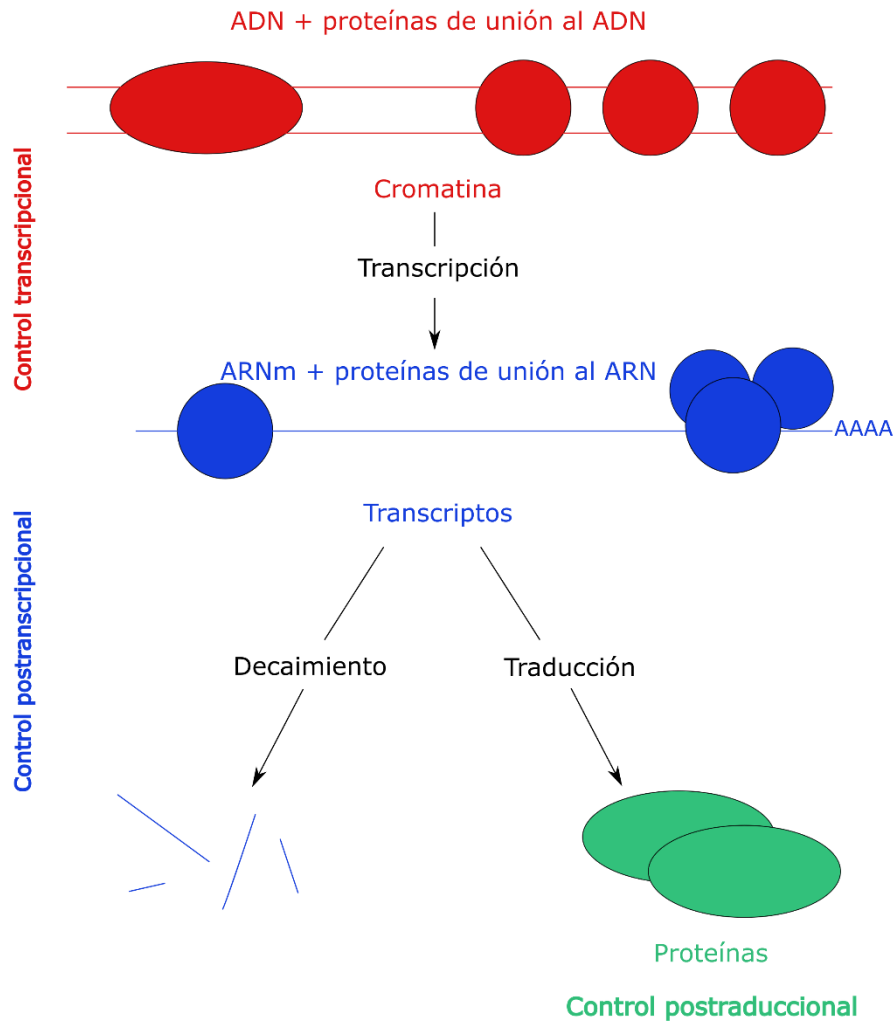


Figura 3 Etapas típicas de la regulación de la expresión génica. Las tasas de transcripción y decaimiento del ARNm determinan los niveles de transcritos estables. Los niveles de transcritos junto con la tasa de traducción determinan la cantidad de proteína producida. La transcripción es controlada mediante proteínas de unión al ADN, por ejemplo, factores de transcripción. La regulación postranscripcional, como el decaimiento de ARNm, y la traducción están reguladas por proteínas de unión al ARN que forman complejos ribonucleoproteicos en conjunto con los transcritos. Existen otras formas de regulación postranscripcional como el procesamiento, exportación y localización del ARNm, así como también existen niveles de regulación postraduccion como modificaciones y degradación de proteína. **AAAA**: cola poli-A. Adaptado de (Mata et al. 2005).

1.7.1 Factores de Transcripción: importancia en las respuestas de la planta

Las estrategias de respuesta diferenciales —dependientes de la condición de nodulación— a la falta de agua pueden ser explicadas por cambios en la regulación de la expresión génica. Los factores de transcripción (FT) juegan un rol relevante en este sentido, ya que las respuestas de tolerancia al déficit hídrico de las plantas son reguladas por numerosos genes que codifican proteínas con roles metabólicos o regulatorios, y entre este segundo grupo se encuentran los FT (Pareek et al. 2010a; Wani et al. 2013).

El déficit hídrico induce una reprogramación global de la expresión génica para ajustar el desarrollo ante la nueva situación ambiental. Dicha reprogramación involucra cambios en los diferentes niveles de la regulación de la expresión, como por ejemplo cambios en la transcripción de genes (Pareek et al. 2010; Sánchez-Bermúdez et al. 2022), en los cuales los FTs tienen especial relevancia (Pareek et al. 2010). La regulación mediada por la interacción de FTs con sus genes blanco activa (aunque raramente inhibe) la transcripción mediante la unión de los FTs en secuencias específicas de ADN denominadas elementos *cis* (Figura 3) (Imagawa 1996). Los elementos *cis* son secuencias cortas de ADN ubicadas en la región reguladora de los genes, que en conjunto con los FT y otras moléculas interactoras conforman el complejo de iniciación de la transcripción (Imagawa 1996; Singh 1998). De esta manera, mediante efectos combinados del empaquetamiento cromatínico y de la acción de los FT, la transcripción es controlada limitando la cantidad de ARNm que es producido por un gen en particular (Singh 1998; Jacquier 2009; Pareek et al. 2010; Joshi et al. 2016).

La percepción del estrés por las plantas involucra en primer lugar la acción de sensores o receptores de membrana. Posteriormente se inician vías de transducción de señales mediadas por mensajeros secundarios que derivan en la activación o represión de FT. Los FT a continuación regulan la expresión de sus genes blanco interactuando con elementos *cis* específicos en sus regiones promotoras (Joshi et al. 2016). Además, los FT interactúan entre sí (o con otras proteínas) previo a la unión a promotores de los genes blancos, resaltando la relevancia de los complejos de transcripción multiméricos en la capacidad de las plantas para integrar y responder a múltiples estímulos ambientales (Ratcliffe and Riechmann 2002).

Miembros de las familias de FT MYB y DREB son inducidos en diversos tipos de estrés abióticos como frío o sequía en varios cultivos (Shen et al. 2003; Vannini et al. 2007; Chen et al. 2009). La familia NAC tiene un rol importante en las respuestas a frío, salinidad, sequía y deshidratación en trigo y arroz (Fujita et al. 2004; Nakashima et al. 2007). La familia ERF participa en respuestas a salinidad, frío y sequía en arroz, tabaco y soja (Guo and Ecker 2004; Cao et al. 2006; Zhang et al. 2009). La familia WRKY está involucrada en respuestas a estrés térmico y sequía en arroz (QIU and YU 2009; Peng et al. 2011). Por su parte, factores *heat shock* (HSF) son FT cruciales que están involucrados en numerosos tipos de estrés abiótico, regulando la expresión de varias proteínas *heat shock* que funcionan como chaperonas moleculares protegiendo y estabilizando proteínas esenciales (Atkinson and Urwin 2012). Particularmente, se ha encontrado que dentro de los FT que predominan en las respuestas de plantas noduladas frente a déficit hídrico se encuentran aquellos pertenecientes a las familias MYB, bHLH, NAC, C2H2, AP2/ERF, WRKY, HSF y GATA (Álvarez-Aragón et al. 2023; Sainz et al. 2024a).

1.7.2 Regulación Génica Postranscripcional

La regulación de la expresión génica a nivel postranscripcional incluye mecanismos como *splicing* de ARN, edición de ARNm o modificaciones químicas del mismo que determinan la identidad y actividad de los productos proteicos (Figura 3). Mientras que otros mecanismos controlan los niveles de proteína mediante la regulación de la exportación de ARNm a citoplasma, y de la estabilidad y traducción de ARNm (Larsson et al. 2013). La regulación a nivel traduccional puede lograrse en sus tres etapas: iniciación, elongación y terminación, siendo el control a nivel de la iniciación el más frecuente.

La iniciación de la traducción en la mayoría de eucariotas depende de la presencia de una secuencia nucleotídica denominada 5'-CAP e involucra el escaneo de la región 5'-UTR en la búsqueda de

un codón de inicio por parte de los ribosomas. Por lo tanto, este mecanismo es denominado CAP-dependiente. Por otra parte, existe un mecanismo CAP-independiente, en el cual algunos ARNm contienen secuencias nucleotídicas complejas, denominadas sitios internos de entrada ribosomal (IRES: *Internal Ribosomal Entry Sites*), en donde el ribosoma se une y da comienzo a la iniciación de la traducción. Los IRES actúan como potenciadores de la traducción en respuesta a estímulos específicos y bajo control de diferentes FT. Es así que, el control traduccional CAP-independiente puede haber evolucionado para regular las respuestas celulares en condiciones de estrés transitorias que de otra manera podrían ser fatales para la célula (Jackson et al. 2010; Solodushko and Fouty 2023).

La vía de decaimiento de ARNm sin sentido (NMD: *nonsense-mediated mRNA decay*) es otro ejemplo de un mecanismo que regula la expresión génica a nivel postranscripcional. Sinópticamente, la vía NMD es un mecanismo de control de calidad co-traduccional que reconoce y elimina ARNm que contienen codones *stop* prematuros, evitando la producción de productos proteicos truncados no funcionales y/o aberrantes (Figura 3) (Luha et al. 2024). La respuesta génica mediada por NMD tiene por objeto mantener la homeostasis celular y disparar respuestas ante situaciones de estrés ambientales o internas (Raxwal and Riha 2023; Luha et al. 2024). Estudios recientes han demostrado el potencial rol de la vía NMD en la adaptación de plantas ante situaciones estresantes (Jeong et al. 2011; Sureshkumar et al. 2016; Shi et al. 2019; Raxwal and Riha 2023; Luha et al. 2024).

Está bien establecido que, una vez percibidas y transmitidas las señales de estrés, la expresión de una multitud de genes es alterada, conllevando una adaptación de las células vegetales a la situación de estrés (Pareek et al. 2010). De esta manera, los genes son transcritos diferencialmente y los ARN producidos son utilizados de diversas formas (Brown 1981; Pareek et al. 2010). Debido a que la transcripción es el primer paso en el inicio de la expresión génica, y a que la mayoría de los genes son regulados a nivel transcripcional, muchos estudios han concentrado sus análisis en la regulación transcripcional (Imagawa 1996). Diversos autores han observado variaciones en las abundancias de ARNm en plantas sometidas a déficit hídrico (Zhu 2002; Clement et al. 2008; Gong et al. 2010; Betti et al. 2012). Sin embargo, se ha demostrado que las diferentes formas de regulación postranscripcional contribuyen a una falta de correlación entre los niveles de transcripto y proteína (Kawaguchi et al. 2004; Vogel and Marcotte 2012; Larsson et al. 2013), donde en la mayoría de los casos el producto funcional de la expresión génica es justamente la proteína, y que además no existe dependencia entre los cambios a nivel del transcriptoma y los cambios que suceden a nivel de los ARNm que están siendo activamente traducidos (traductoma) (Tebaldi et al. 2012).

Por lo anterior, estudiar la regulación traduccional mediante el análisis del traductoma permite una medición más precisa y completa de la expresión génica en comparación con lo que se obtiene cuando únicamente se analizan los niveles de ARNm total, es decir, el transcriptoma (Wang et al. 2013; Ingolia 2014; Lukoszek et al. 2016). Asimismo, hay pocos estudios que analicen traductomas en leguminosas que se encuentran en interacción simbiótica con rizobios o no (Reynoso et al. 2013; Coate et al. 2014; Traubenik et al. 2020). Por ejemplo, un estudio que conllevó análisis transcriptómico y traductómico en soja, reveló la importancia de los sistemas tiorredoxina y glutarredoxina en las respuestas de la planta al déficit hídrico, y a que su vez dichos sistemas se encontraban regulados a nivel traduccional (Sainz et al. 2024a).

Debido a que los ARNm que presentan una mayor actividad traduccional están asociados con más ribosomas, las técnicas de análisis de traducción del ARNm están enfocadas en determinar el grado de

asociación de cada ARNm con polisomas (grupo de ribosomas unidos a una sola molécula de ARNm) y estimar el nivel de traducción (Larsson et al. 2013; Ingolia 2014; Sainz et al. 2024b). En este sentido, las técnicas de *polysome profiling sequencing* (Poly-seq) y *ribosome profiling* o *ribosome footprint* (Ribo-seq) son las dos técnicas principales actuales para el análisis de la traducción (Sainz et al. 2024b). Ambas técnicas tienen en común que se miden cambios en los niveles de ARNm para estudiar la relación entre transcripción y traducción bajo diferentes condiciones y para distinguir genes sujetos a control traducción y/o traduccional (Sainz et al. 2024b). La diferencia más relevante entre ambas es que Poly-seq utiliza ARN polisómico para la secuenciación, mientras que Ribo-seq está restringido a la secuenciación de fragmentos de ARN protegidos por los ribosomas (Sainz et al. 2024b).

La técnica de Ribo-seq se basa en que cuando un ribosoma se encuentra leyendo a un ARNm para producir proteína, cubre una región de aproximadamente 30 nucleótidos del mensajero, por lo que dicha región está protegida de la acción de nucleasas, lo que permite retener esa región cuando se extrae el ARNm y se lo somete a una digestión. La metodología descrita por Ingolia (2014) utiliza secuenciación profunda de los fragmentos de ARNm protegidos por ribosomas para determinar la posición de los ribosomas en las secuencias de ARNm a una resolución de codón (Sainz et al. 2024b). Con respecto a la utilización de esta técnica en plantas de soja, se ha logrado obtener huellas ribosomales, pero sin conseguir concentraciones suficientes para continuar con la etapa de secuenciación (Zardo 2021). Por otro lado, si bien Poly-seq es un abordaje que requiere una labor intensiva, se trata de un método simple y robusto que utiliza una ultracentrifugación para separar ARNm que están siendo traducidos activamente y determinar la identidad y abundancia de los transcriptos mediante una secuenciación profunda (Sainz et al. 2024b). Con respecto a esta técnica, se encuentran solo dos referencias que la utilice para analizar la expresión génica en plantas de soja (Sainz et al. 2022a, 2024a).

1.8 Herramientas de Evaluación de la Expresión Génica

La secuenciación de los genomas y los avances tecnológicos han permitido ampliar de manera significativa el conocimiento de los diversos procesos que intervienen en la regulación de la expresión génica en plantas (Pareek et al. 2010; Jores et al. 2023). Particularmente, la secuenciación del ARNm es una técnica relativamente reciente que permite, en conjunto con técnicas bioinformáticas, identificar y determinar la abundancia de transcriptos (Wang et al. 2009; Krishnamurthy et al. 2018; Denyer et al. 2019; Shulze et al. 2019), y por ende, analizar tanto transcriptomas como traductomas (Jacquier 2009).

La secuenciación del ARNm se realiza mediante técnicas de secuenciación de próxima generación (NGS: *Next-Generation Sequencing*) las cuales refieren a técnicas de alto rendimiento y profundidad de secuenciación del material genético (Kulski 2016). Sucintamente, la secuenciación mediante la plataforma *Illumina* sigue el siguiente flujo de trabajo experimental. En primer lugar, se extrae el ARN total a partir de las muestras, al que posteriormente se le realizan controles de cantidad y calidad necesarios para la construcción de librerías, estos son: una cantidad de ARNm mayor o igual a 1 µg y un RIN (*RNA Integrity Number*) mayor o igual a 7 que indica una buena integridad de la molécula. En segundo lugar, se selecciona el ARNm mediante el uso de *kits* de depleción de ARN ribosomal (ARNr) mediada por digestión enzimática y métodos de captura del ARNm utilizando oligonucleótidos que interaccionan con las colas poli-A de los ARNm (Kulski 2016). En tercer lugar, se transforma el ARNm en secuencias de ADN complementario (ADNc), el cual es fragmentado aleatoriamente y dichos fragmentos se ligan con adaptadores en los extremos 5' y 3' (Figura 4). En cuarto lugar, los fragmentos son amplificados mediante PCR y se someten a controles de calidad y cantidad. En este punto, la librería de secuenciación está pronta para ser

secuenciada. Posteriormente, la librería es cargada en el equipo de secuenciación en una *flow cell*, típicamente una placa de vidrio, en donde los fragmentos son capturados en una superficie que contiene oligonucleótidos complementarios a los adaptadores, permitiendo el anclaje de los fragmentos. Cada fragmento es amplificado en distintos *cluster* clonales mediante amplificación por puente. Dicha amplificación sucede gracias a que los oligos de la superficie de la *flow cell* contienen *primers* que están adheridos mediante un *linker* flexible. De esta forma, en cada ciclo de amplificación, el extremo del fragmento que se extiende hacia fuera de la *flow cell* se hibrida con su complementario ubicado en la superficie formando una estructura de puente (Figura 4) (Shendure and Ji 2008).

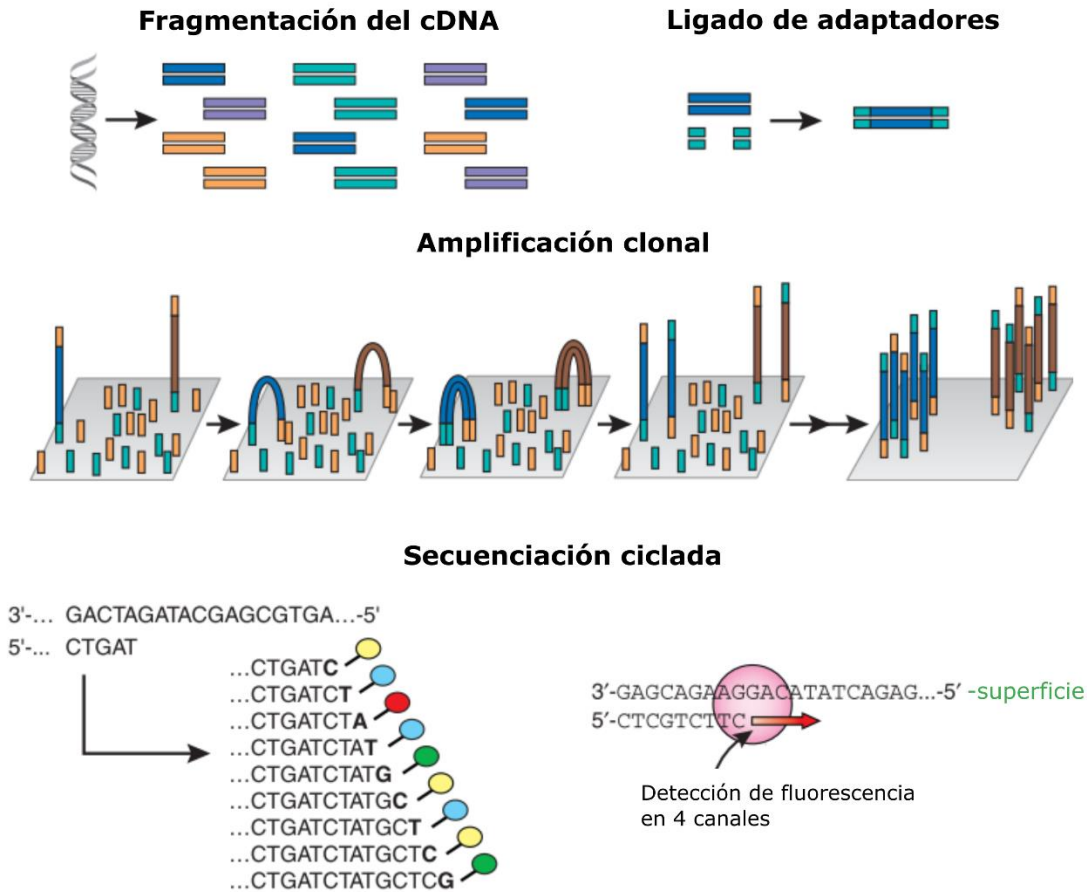


Figura 4 Esquema de la metodología de secuenciación masiva. cDNA: ADN complementario. En secuenciación ciclada: superficie representa la superficie de oligonucleótidos de la *flow cell* que permiten la adhesión de los fragmentos de cDNA mediante adaptadores específicos. Adaptado de (Shendure and Ji 2008).

Finalmente, se procede con la secuenciación por síntesis, en la cual se detecta cada base nucleotídica incorporada en las hebras de ADN molde. Esto se logra gracias a una secuenciación cíclica donde cada ciclo consiste de una extensión de una única base nucleotídica utilizando una polimerasa modificada y una mezcla de los cuatro nucleótidos necesarios. Dichos nucleótidos presentan en el extremo terminal 3' un fluoróforo diferente cada uno, lo que permite determinar qué nucleótido fue adicionado en cada posición en cada *cluster* clonal (Figura 4) (Shendure and Ji 2008). A su vez, el *cluster* tiene el cometido de amplificar la señal para que pueda ser detectada por el equipo mediante fluorescencia en cuatro canales (uno para cada fluoróforo), y ya que cada *cluster* proviene de un único

fragmento, se logra la secuenciación e identificación de los mismos ([Figura 4](#)) (Shendure and Ji 2008; Kulski 2016). El resultado de la secuenciación son archivos *fastq* para cada muestra que contienen la información de las lecturas (*reads*) de los fragmentos nucleotídicos junto con un *score* de calidad para cada base secuenciada (Kulski 2016; Costa-Silva et al. 2017).

Una vez obtenidos los archivos *fastq* es necesario realizar un chequeo de calidad de los mismos que permite analizar la calidad de la secuenciación. En este control se identifican *reads*, o partes de ellos, con baja calidad de secuenciación, presencia de adaptadores, tamaño de *reads* y sesgos de secuenciación (Zhou et al. 2018). Por ejemplo, un valor de *score* menor a 20 se considera de baja calidad, valores entre 20 y 28 son de calidad media y valores superiores a 28 son considerados de buena calidad (Sheng et al. 2016; Zhou et al. 2018). En caso de que los adaptadores hayan sido secuenciados, deben removerse del archivo *fastq*, lo que se logra gracias a que las secuencias de los mismos son únicas y no se existen en la naturaleza génica. Por otro lado, los sesgos de secuenciación pueden incluir anomalías en el contenido de guanina y citosina (GC), alta duplicación de secuencias y/o secuencias sobrerrepresentadas, lo que alteraría los posteriores análisis (Sheng et al. 2016).

A partir de los *reads* procesados se estima la abundancia de transcriptos presentes en cada muestra. Estimar dicha abundancia es una tarea fundamental en genómica, pues aporta información sustancial, por ejemplo, en la clasificación y seguimiento de enfermedades y sus subtipos o en el entendimiento de cambios en la expresión durante el desarrollo y frente a estímulos (Patro et al. 2017). Los algoritmos de cuantificación utilizan alineamientos de los *reads* contra genomas o transcriptomas de referencia (Patro et al. 2017). Cuando se mapea contra genoma las secuencias de referencia son más grandes debido a la presencia de otras regiones génicas más allá de la codificante, como por ejemplo intrones, permitiendo que *reads* que provengan de regiones que no sean codificantes o que no estén anotadas, puedan ir a sus verdaderas locaciones. En cambio, cuando se mapea contra transcriptoma, se fuerza, de alguna manera, a que el tipo de *reads* mencionado anteriormente mapeen contra el transcriptoma. Debido a esto, se utilizan métodos de mapeo contra transcriptoma que consideren secuencias *decoy*, es decir, secuencias genómicas con alta similitud con secuencias transcriptómicas (Srivastava et al. 2020).

La estimación de la expresión se realiza usualmente con el objetivo de realizar análisis de expresión diferencial, y por ejemplo, entender cambios en los patrones de expresión génica durante el desarrollo, o alteraciones del programa genético en respuesta a diferentes estímulos (Patro et al. 2014; Srivastava et al. 2020). Es así que, realizar una cuantificación precisa y eficiente de la abundancia de transcriptos es una tarea fundamental en genómica. Existen diferentes algoritmos de cuantificación, como por ejemplo, *Sailfish* (Patro et al. 2014), *kallisto* (Bray et al. 2016) o *Salmon* (Patro et al. 2017). *Sailfish* realiza la cuantificación en dos fases: indexación y cuantificación. En la indexación, los transcriptos de referencia son fragmentados en *k*-meros (porciones del transcripto con un largo específico definido por el algoritmo), lo que reduce costos computacionales y aumenta la velocidad evitando realizar un mapeo contra transcriptos completos. Este proceso de indexación es posteriormente denominado cuasi-mapeo en la herramienta *Salmon* (Patro et al. 2017). En la fase de cuantificación se utilizan los *reads* y la indexación generada para estimar la abundancia de cada transcripto (Patro et al. 2014). Por su parte, *kallisto* también utiliza el algoritmo de *k*-meros pero mejora la precisión utilizando un proceso denominado pseudoalineamiento enfocándose en identificar a partir de qué transcriptos potenciales se pudo haber originado un *read* y no determinar exactamente cómo los *reads* se alinean con los transcriptos de referencia (Bray et al. 2016). Esto reduce la velocidad dos órdenes de magnitud en comparación con

Sailfish, manteniendo la precisión de la cuantificación (Bray et al. 2016). En el caso de *Salmon*, el algoritmo consiste de tres componentes: un modelo de cuasi-mapeo (o alineamiento si se prefiere), una fase de estimación inicial de los niveles de expresión y una fase de refinamiento de dicha estimación. La ventaja que tiene *Salmon* sobre *Sailfish* y *kallisto* es que corrige sesgos a nivel de muestras vinculados a contenido de GC en fragmentos específicos de secuencias, lo que aumenta la precisión de la estimación (Patro et al. 2017). Esta corrección la hace utilizando la distribución observada de fragmentos secuenciados para cada posible valor de contenido de GC y la incorpora en los modelos estadísticos utilizados para la cuantificación (Patro et al. 2017). El resultado de *Salmon* es una matriz de conteos que contiene la cantidad de transcritos estimados por gen y que puede ser usada como insumo en otros análisis bioinformáticos.

1.8.1 Análisis de Redes de Co-expresión Génica Ponderadas (WGCNA)

El análisis de redes de co-expresión génica ponderadas (o *Weighted Gene Coexpression Network Analysis* [WGCNA]) es un método bioinformático utilizado para describir patrones de correlación entre la expresión de diferentes genes. El WGCNA puede ser usado para encontrar módulos de genes altamente correlacionados, relacionar módulos entre ellos y relacionar módulos con *traits*, donde como *trait* pueden utilizarse tanto características fenotípicas cuantitativas como condiciones o tratamientos experimentales. En este último caso se estarían identificando módulos que asocian en mayor medida con un tratamiento específico (Zhang and Horvath 2005; Langfelder and Horvath 2008). Como insumo de entrada se utiliza una matriz de expresión, la cual contiene la cantidad de transcritos estimados por gen normalizados en función del tamaño de librerías.

La construcción de la red es conceptualmente directa: los nodos representan genes y los nodos están conectados si los correspondientes genes están significativamente co-expresados a lo largo de las muestras en estudio (Zhang and Horvath 2005). De esta manera, cada gen presenta un patrón de expresión a lo largo de las muestras utilizadas, el cual es utilizado en conjunto con una medida de correlación para determinar si un par de genes se encuentran co-expresados. A su vez, se generan módulos de co-expresión, los cuales representan *clusters* de genes altamente interconectados, definiendo un umbral de correlación para establecer si la co-expresión es significativa o no. Como medida de co-expresión se puede utilizar el valor absoluto de la correlación entre patrones de expresión génica elevado a un valor de potencia y se denomina adyacencia. Determinar el valor de potencia resulta crucial en el análisis pues tiene que ser uno que satisfaga el criterio de topología libre de escala, que determina que las propiedades de la red son invariantes con respecto a cambios de escala (Zhang and Horvath 2005; Langfelder and Horvath 2008). El WGCNA implementa un algoritmo interno para elegir el valor más adecuado, definiéndolo como aquel que produzca un R^2 igual o superior a 0,80 en un modelo lineal del logaritmo de la probabilidad de conectividad en función de la conectividad.

Detectar genes significativamente co-expresados mediante WGCNA facilita la identificación de genes candidatos en las respuestas a tratamientos o estímulos específicos pues permite correlacionar módulos con *traits* y reducir la complejidad justamente mediante el agrupamiento de genes (Zhang and Horvath 2005; Langfelder and Horvath 2008). Esta característica se potencia gracias a poder identificar genes *hub* dentro de cada módulo, los que dadas las propiedades de las redes son genes con alta conectividad y por ende participación clave en el módulo (Zhang and Horvath 2005). Diversos autores han explotado estos beneficios para identificar genes claves en las respuestas a déficit hídrico como proteínas

de ubiquitinación en *Boea hygrometrica* (Lin et al. 2019), el factor de transcripción ERF053 en *Medicago falcata* (Li et al. 2022) y factores de transcripción en *Ipomoea batatas* (boniato) (Zong et al. 2023), en *Helianthus annuus* (girasol) (Wu et al. 2022) y en *Saccharum officinarum* (caña de azúcar) (Tang et al. 2023). En el caso de la soja, el WGCNA ha sido utilizado, por ejemplo, para identificar genes involucrados en el contenido de aceites y de flavonas en semillas (Azam et al. 2023; Zhao et al. 2024), en las respuestas de infecciones virales (Niu et al. 2024) y en respuestas a déficit hídrico (Shahriari et al. 2022; Li et al. 2022).

1.8.2 Análisis de Expresión Diferencial de Genes

El abordaje más común dentro del análisis comparativo de datos de RNA-seq es testear si la hipótesis nula de que el *fold change* logarítmico en base 2 (LFC) entre la expresión de un gen en situaciones control versus tratamientos es exactamente cero, es decir evaluar si el gen es afectado de alguna manera por los tratamientos (Love et al. 2014). El LFC es una medida de cambio en la expresión génica entre condiciones, en el cual un valor de 1 indica que la expresión génica aumentó el doble en la condición evaluada frente a la condición control. Con respecto a las herramientas bioinformáticas de análisis de la expresión diferencial, algunas tienen como base modelos no paramétricos, por ejemplo, *NOIseq* y *SAMseq*, y otras adoptan modelos binomiales negativos como base del análisis, tales son *edgeR*, *baySeq* o *DESeq2* (Costa-Silva et al. 2017). La ventaja de usar modelos paramétricos es que se captura toda la información dentro de parámetros (Costa-Silva et al. 2017). Dentro de los que utilizan modelos paramétricos *DESeq2* resulta más eficiente en controlar la tasa de errores tipo I, es decir de detectar falsos positivos (Love et al. 2014).

Para el análisis se debe utilizar información de expresión génica, pudiendo utilizarse la matriz de conteos obtenidas mediante *Salmon*, la cual contiene el número de *reads* mapeados de forma no ambigua para cada gen. En el caso de *DESeq2* se ajusta un modelo lineal generalizado para cada gen, obteniéndose un coeficiente que indica la fuerza de expresión global del gen y un valor de LFC entre condiciones (Love et al. 2014; Costa-Silva et al. 2017). La ventaja que ofrece *DESeq2* sobre otras herramientas es que mejora la estabilidad y reproducibilidad de los análisis mediante el uso de estimadores de contracción en la estimación de los LFC, y el uso de transformaciones para estabilizar la varianza entre muestras que facilita la visualización de diferencias entre muestras y condiciones (Love et al. 2014). Por un lado, se usan estimadores de contracción en la estimación del parámetro de dispersión que cuantifica la variabilidad entre réplicas y en la estimación del LFC. En el caso de la estimación de dispersión permite no subestimar el parámetro debido a tamaños de muestras chicos producto de realizar experimentos en condiciones controlados. En el caso de la estimación del LFC se reduce el ruido inherente de genes con conteos bajo, lo que causa heterocedasticidad en el LFC y complica posteriores análisis (Love et al. 2014). Por otro lado, el uso de transformaciones para estabilizar la varianza, como la transformación logarítmica regularizada, se usa para reducir la heterocedasticidad producida por los altos y bajos conteos facilitando la visualización multivariada de los genes mediante, por ejemplo, análisis de componentes principales y mejorando la interpretabilidad de los datos (Love et al. 2014). De esta manera, se mejora la precisión en la correcta identificación de genes diferencialmente expresados entre condiciones, siendo esto clave para comprender variaciones fenotípicas en los contextos biológicos evaluados (Costa-Silva et al. 2017).

1.8.3 Análisis de Enriquecimiento Funcional

El análisis de enriquecimiento funcional es un método ampliamente utilizado para analizar información de expresión génica y tiene por objeto descubrir anotaciones funcionales que se encuentran sobrerrepresentadas en una lista de genes con respecto a una lista de referencia. Es preciso diferenciar entre anotaciones genómicas y anotaciones funcionales, ya que las anotaciones genómicas refieren a información génica, como ubicación en el genoma e identificación de regiones génicas, y las funcionales informan sobre los roles biológicos de genes o proteínas (Dessimoz and Skunca 2017). Dichas anotaciones son empleadas para interpretar mecanismos moleculares, procesos biológicos y vías metabólicas asociadas con las condiciones experimentales en estudio (Garcia-Moreno et al. 2022). En este sentido, la ontología de genes (GO: *Gene Ontology*) provee un marco conceptual y define un universo de posibles funciones de los productos génicos, por lo que una anotación GO es una sentencia sobre la función biológica de un gen en particular.

La GO considera tres aspectos sobre cómo pueden ser descritas las funciones génicas: función molecular (actividad o proceso a nivel molecular), componente celular (ubicación específica relativa a la célula donde el producto génico lleva a cabo la función) y proceso biológico (objetivo biológico mayor conformado por varias funciones moleculares). Es así que, la ontología de procesos biológicos es la más grande y diversa de las tres, ya que refleja la multiplicidad de niveles de organización biológica en la que programas génicos pueden ser identificados, lo que la convierte en la ontología de preferencia a la hora de analizar datos (Dessimoz and Skunca 2017). Brevemente, la asignación de un término GO a un gen se realiza mediante dos abordajes: evidencia experimental, que involucra un minucioso curado de la información obtenida en experimentos, y predicciones de función, que usualmente se basan en similitud de secuencia o correlaciones en la expresión génica (Dessimoz and Skunca 2017).

A su vez, una herramienta complementaria en el análisis de funciones génicas es el análisis de vías metabólicas KEGG (*Kyoto Encyclopedia of Genes and Genomes*). La anotación KEGG se centra en analizar vías metabólicas completas utilizando información de las interacciones entre moléculas, lo que permite entender redes bioquímicas complejas (Kanehisa 2002; Dessimoz and Skunca 2017). De esta forma, complementa el uso de anotaciones GO, ya que genera el plus de información dado por las interacciones. El conocimiento de las interacciones surge tanto de evidencia experimental como de búsqueda de genes ortólogos y homología de secuencias (Kanehisa 2002; Kanehisa et al. 2012; Dessimoz and Skunca 2017). Sucintamente, cuando la interacción o participación en una vía no está definida, las secuencias de genes o proteínas a analizar son alineadas contra la base de datos KEGG con la consecuente asignación en grupos ortólogos. Los grupos ortólogos son grupos de genes de diferentes especies originadas a partir de un ancestro en común y que usualmente cumplen la misma función biológica. Una vez hecho esto se identifica la vía KEGG mediante información conocida por evidencia experimental de la participación del grupo ortólogo en una vía determinada (Kanehisa et al. 2016).

1.8.4 Predicción de Función Génica mediante EGAD

EGAD (Extending “Guilt-by-Association” by Degree), de alguna manera traducido como asociación por culpabilidad, es un método que permite evaluar redes génicas y predecir funciones génicas dadas las interacciones génicas en las redes (Ballouz et al. 2017). El método requiere de la entrada de dos tipos de datos: una red de interacciones génicas, pudiendo utilizarse la red generada por el WGCNA, y un set de

anotaciones biológicas, por ejemplo, las obtenidas mediante GO o KEGG (Ballouz et al. 2017; Ribone et al. 2023). Concretamente, se determina qué tan bien una red génica agrupa un set de genes de una anotación en particular mediante un algoritmo basado en el principio de “*guilt-by-association*” (GBA), el cual postula que genes con funciones similares están preferentemente conectados (Ballouz et al. 2017). Por lo tanto, se puede inferir que si dos genes están interactuando entonces también cumplen funciones en un mismo proceso biológico (Gillis and Pavlidis 2012).

La predicción de la función génica mediante GBA es llevada a cabo ocultando información de la red de interacción y evaluando la capacidad de la misma de recuperar dicha información. Por ejemplo, en un caso hipotético, se conoce que 10 genes participan en un proceso determinado, se oculta la información de que dos de ellos participan en ese proceso y se evalúa que también se predice el proceso para ellos dos dada las interacciones de la red usada como insumo (Gillis and Pavlidis 2012; Ballouz et al. 2017). Esta evaluación se realiza mediante validación cruzada obteniéndose para cada término un *score* (área bajo la curva –AUC– de características de funcionamiento del receptor –ROC: *Receiver Operating Characteristic*–) que caracteriza el grado con el cual la red puede ser utilizada para predecir y anotar el término en genes que no lo poseen (Ballouz et al. 2017). Este *score* informa sobre qué tan buena es la predicción de la función, es decir la capacidad de distinguir entre un falso y un verdadero positivo. Valores por encima de 0,7 y de 0,9 se consideran muy buenos y excelentes, respectivamente (Ballouz et al. 2017). Esto significa que la predicción de función para un gen es 70 y 90% probable de que sea verdaderamente positiva, respectivamente.

El uso de GBA ha sido probado satisfactoriamente en plantas (Depuydt and Vandepoele 2021). Por ejemplo, (Silva et al. 2016) predijeron funciones para genes no caracterizados en *Arabidopsis thaliana* mediante el uso de datos transcriptómicos sobre la dinámica de la transición de semilla a plántula de la especie. De igual manera, un estudio de transcriptómica comparativa identificó módulos de genes de biosíntesis de lignina conservados en angiospermas, pudiendo así predecir nuevos candidatos en la biosíntesis de pared de celular en diferentes especies (Sibout et al. 2017). Asimismo, Ribone et al. (2023) identificaron genes con funciones vinculadas a respuestas de defensas en *Helianthus annuus* mediante el uso de datos transcriptómicos de raíz y de hoja. Además, el genoma de girasol está parcialmente anotado, contando con al menos una anotación GO en aproximadamente el 60% de los genes, de los cuales la tercera parte tiene al menos un proceso biológico GO anotado (Ribone et al. 2023), por lo que esta herramienta resulta útil tanto para predecir funciones nuevas como para aportar evidencia en términos anotados por predicción (Ballouz et al. 2017; Depuydt and Vandepoele 2021; Ribone et al. 2023). Con respecto al genoma de soja, como fue mencionado anteriormente, presenta un bajo porcentaje de anotaciones validadas experimentalmente (10%) y un alto porcentaje de genes no caracterizados (25%), por lo que usar EGAD resulta sumamente útil para apoyar las actuales predicciones computacionales y para generar nuevas predicciones utilizando datos de calidad.

1.9 Análisis Funcional de Genes mediante Sistema de Transformación Transitoria

Investigar las funciones génicas es crucial para entender el *background* genético del fenotipo en diferentes cultivos y así asistir al desarrollo de cultivares que presenten características de interés, como por ejemplo mayor rendimiento o mayor resistencia ante diferentes tipos de estrés. Una forma de realizar dicho estudio es mediante la transformación estable de plantas (Nath Radhamony et al. 2005; Kereszt et al. 2007; Clemow et al. 2011). Tal es el caso de la transformación mediada por *Agrobacterium tumefaciens*, que es utilizada para generar líneas mutantes que sobreexpresen o presenten pérdida de función de

determinado/s gene/s en especies como *A. thaliana* o *Nicotiana tabacum* (Nath Radhamony et al. 2005; Clemow et al. 2011). Sin embargo, especies como *G. max* o *P. sativum* son consideradas recalcitrantes para la transformación estable (Clemow et al. 2011) debido a diversas dificultades presentadas como largos períodos de desarrollo de las plantas (hasta 9 meses), ploidías alteradas, anormalidades fenotípicas o bajas fertilidades reproductivas, que termina traduciéndose en una baja eficiencia de la transformación (Kereszt et al. 2007; Clemow et al. 2011). No obstante, cuando el gen de interés está involucrado en la biología de la raíz, por ejemplo, interacciones simbióticas, toma de nutrientes, transporte de hormonas o respuestas a estrés, los problemas anteriores pueden subsanarse mediante el uso de *Rhizobium rhizogenes* (Kereszt et al. 2007; Clemow et al. 2011; Mohammadi-Dehcheshmeh et al. 2014).

La bacteria *R. rhizogenes* tiene la capacidad de transferir e integrar en el genoma de su hospedero genes contenidos dentro de su ADN de transferencia (ADN-T) ubicado en un replicón extracromosómico denominado plásmido Ri. La integración del ADN-T en el genoma hospedero provoca la formación de raíces neoplásicas, denominadas *hairy roots* (HR) que pueden contener el ADN-T (Kereszt et al. 2007). Cada evento transgénico (HR), sucedido a partir de la infección y posterior división de una única célula, representa un evento de transformación independiente, por lo que se obtiene un alto número de transformantes que pueden ser analizadas en un corto período de tiempo (Kereszt et al. 2007). A este tipo de transformación se la denomina transformación transitoria, pues el transgén no se pasa a la descendencia en caso de que la planta produzca semillas. Esto es debido a que la parte aérea de la planta se mantiene silvestre, obteniendo una planta quimérica compuesta por una sistema radicular transgénico y parte aérea no transgénica (Kereszt et al. 2007; Clemow et al. 2011).

A partir de esta técnica, entonces, se puede realizar el análisis funcional de genes de interés mediante la transformación de la bacteria con un vector binario que contiene el gen de interés, y la posterior infección y transformación de las plántulas. Para ello, la técnica implica una serie de pasos que pueden ser resumidos como: transformación de *R. rhizogenes* con el vector binario de interés, selección de colonias transformantes, identificación de la colonia más efectiva y por último la transformación de las plántulas (Kereszt et al. 2007; Mohammadi-Dehcheshmeh et al. 2014). A su vez, esta técnica tiene la ventaja de que cuando se transforma la planta se pueden identificar las raíces transgénicas de forma no invasiva, pues la inserción del transgén está acompañada de la inserción de un gen reportero cuyo producto puede ser visualizado bajo luz fluorescente (por ejemplo, GFP o DsRED) (Mohammadi-Dehcheshmeh et al. 2014) o incluso a simple en el caso del sistema RUBY que provoca por la sobreexpresión de enzimas responsables de la formación de betalaína a partir de precursores presentes en la planta (He et al. 2020). Finalmente, se debe realizar un fenotipado adecuado para analizar si el gen tiene una participación relevante en los contextos biológicos de interés estudiados.

1.10 Hipótesis y Objetivos

La respuesta diferencial al déficit hídrico de plantas noduladas en comparación con plantas sin nodular puede ser explicada por cambios en la expresión génica. La hipótesis de este trabajo se basa en que la regulación a nivel traduccional contribuye de manera significativa en la modulación de dicha expresión génica.

Objetivo general:

Contribuir a la mejora de la tolerancia al déficit hídrico en soja mediante la incorporación de nuevas fuentes de tolerancia en las respuestas a estrés. Mediante el análisis de las respuestas estrés-específicas a nivel transcripcional y traduccional se pretende identificar los cambios asociados a la maquinaria traduccional de la raíz, con el fin de definir posibles genes candidatos que estén involucrados en la respuesta de tolerancia de las plantas al estrés.

Objetivos específicos:

1. Análisis del transcriptoma y del traductoma de raíces de plantas de soja noduladas y sometidas a déficit hídrico.
2. Identificación de los genes sujetos a control traduccional en raíces de plantas de soja noduladas y sometidas a déficit hídrico.
3. Puesta a punto del sistema de transformación transitoria mediado por *Rhizobium rhizogenes* para el análisis funcional de genes.

2. MATERIALES Y MÉTODOS

2.1 Material vegetal, condiciones de crecimiento, nodulación y déficit hídrico

El experimento fue llevado a cabo utilizando el genotipo de soja (*Glycine max* (L.) Merr.) Génesis 5601, cedido por el Instituto Nacional de Investigaciones Agropecuarias (INIA). Las plantas fueron crecidas en botellas de 0,5 L conteniendo sustrato compuesto por una mezcla de arena:vermiculita (1:1). Se utilizó una cámara de crecimiento de condiciones controladas estableciendo un ciclo de temperatura día/noche de 28/20°C, respectivamente, y un fotoperíodo de 16/8 h, luz/oscuridad. La humedad relativa ambiente fue de $39,5 \pm 7,7\%$ durante todo el período de crecimiento. El ambiente lumínico estuvo formado por lámparas de halogenuro metálico con una irradiación de $620 \mu\text{moles fotones m}^{-2} \text{s}^{-1}$ (137 W m^{-2}) de un espectro blanco. Se colocaron tres semillas por maceta y una vez expandidos los cotiledones se seleccionó la plántula más sana para continuar con el experimento. A su vez, para la selección de plántulas se tuvo en cuenta la homogeneidad de desarrollo para evitar posibles influencias vinculadas a estadios de desarrollo diferenciales. Para las plantas inoculadas, se utilizó la cepa U1302 del rizobio *Bradyrhizobium elkanii*, la cual fue crecida en medio líquido YEM (Vincent 1971). La inoculación se realizó el día de la siembra.

El diseño experimental comprendió un diseño completamente aleatorizado (DCA) de cuatro tratamientos combinados con cinco réplicas biológicas, haciendo un total de 20 unidades experimentales. Una unidad experimental quedó definida como una maceta con una planta a la cual se le aplicó un tratamiento. Los tratamientos combinaron las condiciones de nodulación (N) y no nodulación (NN), y de plantas a capacidad de campo y con restricción hídrica (CC y RH, respectivamente). Plantas a capacidad de campo no presentan restricciones hídricas en el sustrato, por ende, el tratamiento CC es considerado el tratamiento control de la condición hídrica. Es así que los cuatro tratamientos son: plantas no noduladas a capacidad de campo (NN+CC), plantas no noduladas y con restricción hídrica (NN+RH), plantas noduladas a capacidad de campo (N+CC), y plantas noduladas con restricción hídrica (N+RH) (Figura 5A). De esta manera, los contrastes biológicos analizados fueron: i) N+RH vs N+CC, ii) N+RH vs NN+RH, iii) NN+RH vs NN+CC y iv) N+CC vs NN+CC (Figura 5B).

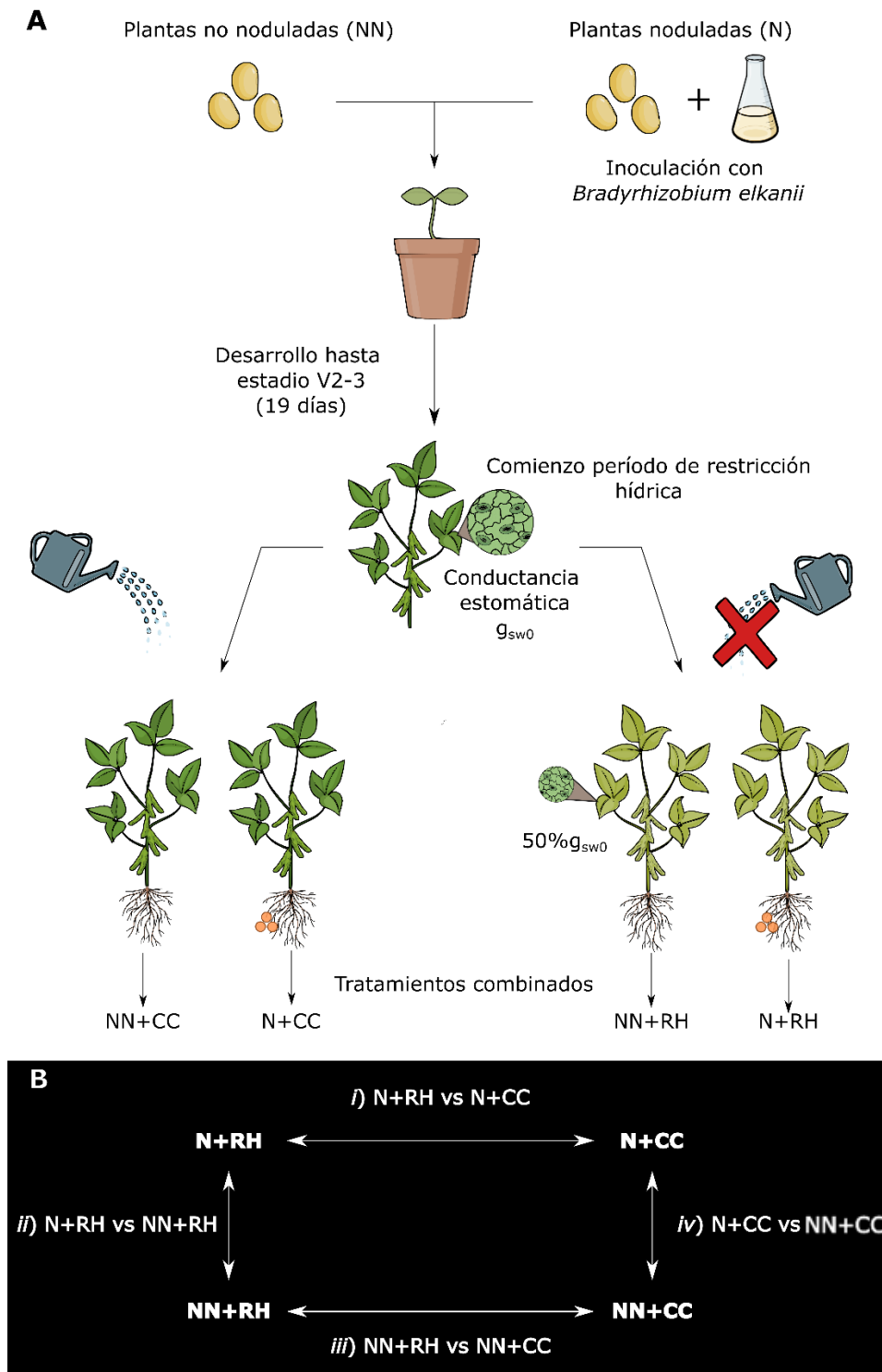


Figura 5 Esquema de la definición de tratamientos combinados y contrastes entre ellos. A Etapas seguidas para el crecimiento de plantas, con inoculación en el caso de plantas noduladas, y establecimiento del período de restricción hídrica. Plantas en restricción hídrica fueron cosechadas cuando la conductancia estomática fue el 50% del valor obtenido al comienzo del período de restricción hídrica. g_{sw0} : conductancia estomática en el día 0 del período de restricción hídrica. Figura hecha con [BioRender](#). **B** Contrastes entre tratamientos combinados. **NN**: no noduladas. **N**: noduladas. **CC**: Capacidad de Campo. **RH**: restricción hídrica. Estadio **V2-3** según (Fehr and Caviness 1977).

Durante los primeros 19 días pos-siembra todas las plantas crecieron sin restricciones hídricas manteniendo el sustrato a capacidad de campo (CC) con medio ByD (Broughton y Dilworth), suplementado con KNO₃ (concentración final 0,5 mM y 5 mM para plantas noduladas y no noduladas, respectivamente). En este tiempo las plantas se encontraron en un estadio de desarrollo vegetativo V2-V3 (Fehr and Caviness 1977) . A partir del día 20, día cero del período de déficit hídrico, se dejaron de regar aquellas plantas correspondientes a los tratamientos de restricción hídrica (RH). Las plantas sin restricción hídrica (CC) se mantuvieron a CC durante todo el ensayo (Figura 5A).

El contenido de agua del sustrato fue medido diariamente mediante gravimetría durante todo el experimento, tanto en período de déficit hídrico como fuera de él. Esto se hizo cuantificando el peso del sistema suelo-planta-maceta (SSPM) mediante una balanza y restando los pesos asociados a la maceta. El peso de la planta se consideró despreciable en el total. Para obtener el peso del sustrato a CC se dejó que el agua de cuatro macetas saturadas percole durante 72 horas. Las macetas se taparon para evitar cualquier pérdida por evaporación del suelo. El peso del sustrato con el agua que queda retenida en el sustrato una vez percola el exceso es considerado el peso del sustrato a CC (Cassel and Nielsen 2018).

La conductancia estomática en la cara abaxial de la hoja correspondiente a V2-V3 fue medida en el día 20 (día 0 del período de déficit hídrico) para cada planta utilizando un porómetro Decagon modelo SC-1 (Decagon Device, United States). De esta manera se determinó el valor de conductancia estomática al inicio del período de déficit hídrico. Se hicieron mediciones de conductancia estomática diarias hasta el final del período de déficit hídrico, el cual fue determinado individualmente para cada planta E cuando el valor de conductancia estomática fue de aproximadamente el 50% del obtenido en el día 0 del período de déficit hídrico (Figura 5A). Se cosecharon las raíces de cada planta E al final de dicho período y se mantuvieron almacenadas a -80°C hasta posteriores determinaciones. Las raíces de plantas C fueron cosechadas en conjunto con las de plantas E y también fueron congeladas a -80°C para ulteriores determinaciones.

2.2 Purificación de fracciones de ARN total y asociado a polisomas

La purificación del conjunto de ARN asociado a polisomas fue realizada mediante centrifugación en colchones de sacarosa (Sainz et al. 2022b). Se homogeneizó un volumen de 2 mL de raíces congeladas previamente pulverizadas mediante nitrógeno líquido en 4 mL de buffer de extracción de polisomas (Sainz et al. 2022b). Los homogeneizados se mantuvieron en hielo hasta que las muestras fueron procesados mediante centrifugación a 16000 g durante 15 minutos. Posteriormente, se filtró la muestra y el paso de centrifugación fue repetido. En este momento se reservaron 500 µL del sobrenadante para el aislamiento del ARN total (TOTAL) (Figura 6). Se cargaron 2 mL del sobrenadante restante en colchones de sacarosa compuesto por dos capas de sacarosa, una al 12% y otra al 33,5%, preparados en tubos de 13,2 mL (Ultra-Clear, Beckman Coulter, Estados Unidos, 344059). Se centrifugó a 35000 rpm a 4°C durante 2 horas en una ultracentrífuga Beckman L-100K de clase S (W40 Ti *swinging bucket* rotor). Luego de la centrifugación, se recuperó la fracción polisómica a partir del pellet y se resuspendió en 200 µL de buffer de resuspensión de polisomas (Sainz et al. 2022b). Los polisomas resuspendidos se mantuvieron a 4°C durante 30 minutos para posteriormente realizar una purificación de ARN regular y obtener la fracción de ARNm asociado a polisomas (PAR: *Polysome-associated RNA*) (Figura 6).

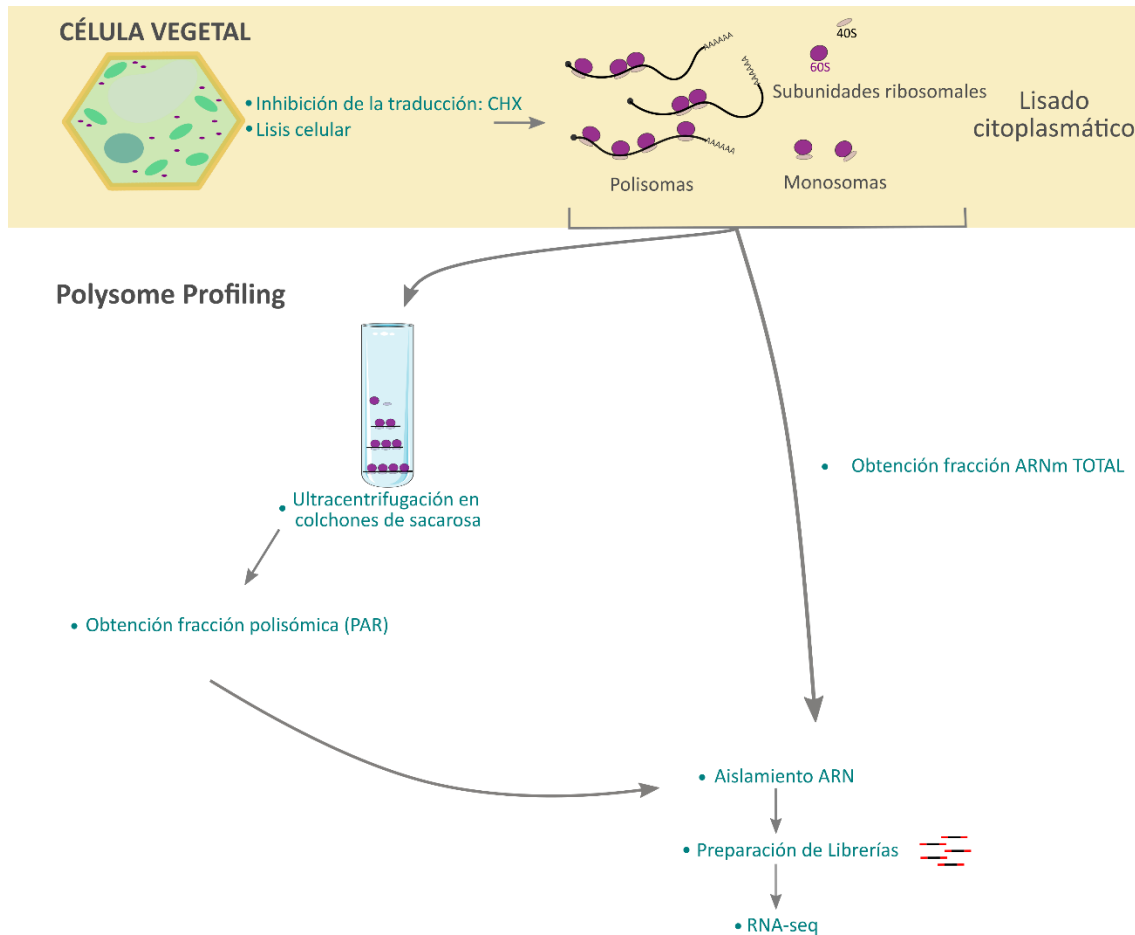


Figura 6 Esquema de la metodología seguida para purificar las fracciones de ARNm TOTAL y PAR (asociada a polisomas). Luego de que el material vegetal es colectado y procesado, la traducción es inhibida con el uso de cicloheximida (CHX) y el lisado celular es obtenido. A partir del lisado se obtiene el ARNm TOTAL mediante un protocolo de extracción directo, y el ARNm PAR mediante *polysome profiling*. Adaptado de (Sainz et al. 2024b)

2.3 Extracción de ARN y secuenciación

La extracción de ARN de las fracciones de ARNm TOTAL y PAR fue hecha utilizando TRizol LS (Invitrogen, Estados Unidos, 10296-028). Tanto el pellet polisómico resuspendido como el extracto reservado para el aislamiento de ARNm total se homogeneizaron en 750 µL de TRizol, siguiendo el procedimiento de (Sainz et al. 2022b). Se cuantificó la concentración e integridad de cada ARN extraído mediante BioAnalyzer Agilent 2100 (Agilent Technologies, Inc., Estados Unidos). Se enviaron para preparación de librerías y secuenciación a Macrogen Inc. (Corea del Sur) aquellas muestras que presentaron un RIN (*RNA integrity number*) mayor a 7,0 y una concentración mayor a 1,0 µg. Las librerías de cDNA fueron hechas mediante el kit *TruSeq Stranded mRNA paired-end* (Illumina). La secuenciación se realizó en plataformas de secuenciación de alto rendimiento de *Illumina Novaseq 6000*. Se enviaron a secuenciar muestras de ARNm TOTAL y PAR de raíces de tres réplicas biológicas de cada tratamiento.

2.4 Análisis de datos

2.4.1 Procesamiento de Datos de Secuenciación

Se inspeccionó visualmente la calidad de cada muestra secuenciada utilizando *FastQC* (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Mediante el uso de *Trimmomatic* (versión 0.39, Bolger et al. 2014) se removieron adaptadores y se *trimmearon* bases con baja calidad de secuenciación. Se retuvieron solamente secuencias *trimmeadas* mayores a 80 pares de bases (pb) y con una calidad general mayor a 30. Se utilizaron los siguientes argumentos con los correspondientes parámetros: HEADCROP:1, TRAILING:3, SLIDINGWINDOW:4:15 y MINLEN:80. Se cuantificó la expresión génica a nivel de transcripto utilizando la herramienta bioinformática *Salmon* en modo *quasi*-mapeo (versión 1.8.0, Patro et al. 2017). Los parámetros de *Salmon* se dejaron por default excepto que se habilitó la corrección en el sesgo de GC. El *index* para el mapeo se construyó a partir de la versión más reciente del transcriptoma de *Glycine max* (GCF_000004515.6, *Glycine max* v4.0, sin ARNr). Debido a que un mismo gen puede tener transcritos alternativos, los conteos a nivel de transcritos fueron agregados a nivel de genes usando el paquete de R *tximport* (versión 1.26.1, Soneson et al. 2016). Se realizó un análisis descriptivo de este proceso, describiendo los conteos de *reads* iniciales y *pos-trimmed*, porcentaje de *reads* retenidos luego del control de calidad y tasa de mapeo.

2.4.2 Análisis de Redes de Co-expresión Génica Ponderadas (WGCNA)

Se realizó un WGCNA utilizando el paquete de R WGCNA (versión 1.72-5, Langfelder and Horvath 2008). Como datos de entrada se utilizaron los conteos a nivel de genes normalizados mediante función *normalized* del paquete R DESeq2 (versión 1.38.3, Love et al. 2015). Se removieron genes con bajo coeficiente de variación y/o bajos conteos (definidos como menos de 50 conteos en más del 50% de las muestras), ya que características con baja expresión tienden a reflejar ruido en la señal. Posteriormente, la función *pickSoftThreshold* del paquete WGCNA fue usada para seleccionar la potencia del *soft threshold* teniendo por objeto satisfacer la topología libre de escala de la red. De esta manera, se construyó la red utilizando la función *blockwiseModules*, que detecta módulos y construye la red en único paso, estableciendo un valor de potencia de 9. El argumento *maxBlockSize* fue establecido en 27000 (número de genes totales pos-filtrado), *mergeCutHeight* en 0,20, *networkType* en *unsigned* (adyacencia no signed), *TOMType* en *signed* (para preservar la continuidad de la correlación), y el resto de argumentos se mantuvieron por default.

Para cada módulo se calculó el *eigengene* (ME: *module eigengene*) (primer componente principal del módulo que oficia como un gen representativo de la expresión del módulo) utilizando la función *moduleEigengenes* del paquete WGCNA, con el objetivo de determinar si existió asociación entre módulos y tratamientos. Se construyó un *heatmap* mediante el paquete de R *pheatmap* (versión 1.0.12, Kolde 2015) para visualizar la relación entre tratamientos y *eigengenes*. Los *eigengenes* fueron agrupados mediante clusterización jerárquica usando como medida de distancia entre ellos (1 – correlación entre MEs) según (Zhang and Horvath 2005). El criterio de clusterización utilizado para la creación del dendrograma fue el de enlace promedio (Dudek 2020) . La selección del número de clusters se realizó

mediante el índice *Silhouette* (Dudek 2020) utilizando la función *NbClust* del paquete de R *NbClust* (Charrad et al. 2014).

La visualización de la red fue hecha mediante el paquete de R *network* (Butts 2008), estableciendo un valor de conexión entre nodos (genes) mayor a 0,15. Los módulos fueron coloreados para mejorar la visualización. A su vez, se estimó la conectividad intramodular de cada gen utilizando la función *intramodularConnectivity* del paquete WGCNA, para así definir genes *hub* como el 10% de los genes con mayor conectividad (Seo et al. 2009).

2.4.3 Análisis de Expresión Diferencial de Genes (DEG)

A los efectos de observar los patrones de agrupamiento y distancias entre réplicas y tratamientos se realizó un análisis de componentes principales (PCA) de las muestras y un *heatmap* de distancias entre muestras utilizando los paquetes de R *FactoMineR* (versión 2.11, Lê et al. 2008), *pheatmap* (versión 1.0.12, Kolde 2015) y *ggplot2* (versión 3.5.1, Wickham 2016). Para ello se utilizaron los genes que presentaron al menos un conteo en la suma de todas las muestras y se aplicó la transformación VSD de los datos para estabilizar la varianza siguiendo lo recomendado por (Love et al. 2015). Además, se realizó un análisis de enriquecimiento de procesos biológicos de *Gene Ontology* (GO), utilizando el paquete *topGO* (versión 2.50.0, Alexa and Rahnenfuhrer 2021), sobre los 50 genes que más contribuyeron a las dimensiones 1 y 2 del PCA con el objetivo de ver cuáles procesos son los que influyeron en mayor medida en la variabilidad de los datos. El análisis estadístico de la expresión diferencial se computó mediante la función DESeq del paquete DESeq2 (versión 1.38.3, Love et al. 2015), utilizándose el siguiente modelo con arreglo factorial:

$$Y_{ijk} = \mu + N_i + H_j + (N * H)_{ij} + \varepsilon_{ijk}$$

Donde:

- Y_{ijk} es la cantidad de conteos del tratamiento con el i -ésimo nivel de N , el j -ésimo nivel de H y la k -ésima repetición.
- μ es la media general
- N_i es el i -ésimo nivel del factor nodulación
- H_j es el j -ésimo nivel del factor condición hídrica
- $(N * H)_{ij}$ es el efecto de interacción del i -ésimo nivel del factor nodulación y el j -ésimo nivel del factor condición hídrica
- ε_{ijk} es el error asociado a cada unidad experimental

Se modelaron las fracciones de ARN por separado, utilizando el mismo diseño, para luego retener aquellos genes que presentaron al menos un conteo en todas las muestras. Un gen diferencialmente expresado (DEG) en un tratamiento frente a otro (contraste), y para cada fracción de ARN, fue definido cuando el valor absoluto del \log_2 del *fold change* fue mayor a 1 y el p-valor ajustado con la corrección de Benjamini-Hochberg fue menor a 0,05. Las listas de genes DEG de cada contraste y fracciones de ARN se intersecaron a fin de identificar genes DEG específicos de cada contraste y fracción.

Por un lado, dentro de cada contraste, ej. N+RH vs N+CC, se intersecaron las listas de DEGs en TOTAL y en PAR, obteniendo los genes DEGs que presentaron regulación transcripcional (únicos de

TOTAL), traduccional (PAR), o mixta, es decir transcripcional y traduccional (intersección entre TOTAL y PAR). Por otro lado, se intersecaron pares de contrastes con el objetivo de quitar para un contraste de interés aquellos genes DEG que puedan deberse al efecto de un tratamiento biológico y no al efecto del tratamiento combinado. Siguiendo el ejemplo anterior, si se interseca la lista de DEGs en NN+RH vs NN+CC (genes DEG debido al déficit hídrico sin nodulación previa) se desquita de la lista de N+RH vs N+CC aquellos genes en que su expresión se ve alterada debido a la falta de agua y se obtienen los genes que se alteran específicamente cuando la planta experimenta déficit hídrico, pero a su vez se encuentra nodulada.

De igual forma, en el segundo contraste de interés, el cual estudia las diferencias en la expresión génica de plantas sometidas a déficit hídrico en diferente contexto de nodulación (N+RH vs NN+RH), la lista de DEGs a intersecar resultó de la obtenida en el contraste N+CC vs NN+CC para poder remover genes DEGs que se deban un efecto de la nodulación y así obtener la lista de genes que altera su expresión en déficit hídrico dado que la planta se encuentra nodulada frente a una planta que no se encuentra nodulada.

2.4.4 Análisis WGCNA + DEGs

Los genes DEGs en cada fracción y en cada contraste entre tratamientos combinados [i) N+RH vs N+CC, ii) N+RH vs NN+RH, iii) NN+RH vs NN+CC, iv) N+CC vs NN+CC] fueron co-localizados en los módulos de la red de co-expresión. Genes con expresión diferencial fueron coloreados en la red según su estado de regulación, es decir, *up/down*-regulados o estables. De esta manera se identificaron módulos donde la presencia de DEGs se vio enriquecida.

2.4.5 Análisis de Enriquecimiento GO

La anotación de términos GO del genoma de soja se realizó con *PANNZER* (Törönen and Holm 2022). Se realizaron análisis de enriquecimiento GO en los módulos que tuvieron una alta localización de DEGs. El análisis de enriquecimiento fue realizado para la ontología correspondiente a procesos biológicos, utilizando el test exacto de Fisher y el método *weight01* para construir la estructura del grafo GO. Un término GO fue considerado significativamente enriquecido cuando presentó un *false discovery rate* (FDR) menor a 0,05. El análisis de enriquecimiento fue realizado utilizando el paquete de R/BioConductor *topGO* (versión 2.50.0, Alexa and Rahnenfuhrer 2021). Los gráficos de enriquecimiento GO fueron contruidos utilizando el paquete de R *ggplot2* (versión 3.5.1, Wickham 2016).

2.4.6 Análisis de Redes de Interacción Proteína-Proteína (PPI)

Se utilizaron las herramientas *CytoScape* (Smoot et al. 2011) y *STRING* (Szklarczyk et al. 2021) para construir redes de interacción proteína-proteína y buscar potenciales interacciones entre proteínas codificadas por DEGs. Las redes fueron construidas utilizando los DEGs co-localizados en los módulos seleccionados para análisis. Se realizaron análisis de enriquecimiento de procesos biológicos GO y de vías KEGG en las redes PPI. En ambos casos un término fue considerado significativamente enriquecido cuando presentó un FDR menor a 0,05.

2.4.7 Búsqueda, Identificación y Análisis de Enriquecimiento de Motivos Génicos en Regiones 5'-UTR

Las regiones 5'-UTR de los transcritos se obtuvieron a partir del archivo GTF de anotación genómica (GCF_000004515.6, *Glycine max* v4.0) utilizando un *script* público (https://github.com/saketkc/gencode_regions). Posteriormente, se utilizó la herramienta bioinformática *seqtk* (<https://github.com/lh3/seqtk>) para recuperar las secuencias *fasta* de las 5'-UTR, usando como insumo el archivo *bed* de posiciones de las 5'-UTR obtenido en el paso anterior. Para la búsqueda e identificación de motivos génicos en las regiones 5'-UTR se utilizó la herramienta bioinformática MEME (versión 5.5.7, Bailey and Elkan 1994). Se buscaron hasta 10 motivos con un largo de entre 6 y 50 nucleótidos, utilizando el modelo de búsqueda clásico y cero o una ocurrencia de motivos por secuencia. El análisis de enriquecimiento de motivos se realizó con la herramienta SEA (*Simple Enrichment Analysis*), perteneciente al mismo conjunto de herramientas bioinformáticas que MEME (Bailey et al. 2015). Se dejaron los parámetros de búsqueda por default: E-value menor a 10, modelo de orden Markoviano cero para ajustar por sesgos de único nucleótido y alineamiento entre secuencias en el centro (Bailey et al. 2015).

2.4.8 Predicción y Validación de Función Génica

Se utilizó el paquete de R EGAD (Ballouz et al. 2017) para evaluar la capacidad de la red de co-expresión génica de conectar genes con términos GO compartidos y así predecir o validar procesos biológicos GO asociados. Basándose en el principio de *Guilt-by-Association* (GBA) una red puede ser evaluada ocultando el término GO anotado para un subgrupo de genes dado y posteriormente testeando si el término puede ser predicho a partir de la co-expresión (correlación entre expresiones génicas) del resto de genes anotados. Se utilizaron términos GO que tengan más de 20 y 1000 genes anotados, haciendo un total de 146 términos. Se utilizó el área bajo la curva ROC (AUC) validada mediante 5-fold *cross validation* como métrica de precisión de las predicciones. Se seleccionaron aquellos términos donde el AUC fue mayor o igual a 0,7 (Ballouz et al. 2017). Como red de co-expresión se utilizó la misma matriz de adyacencia obtenida y utilizada en el WGCNA.

2.5 Puesta a Punto de Sistema de Transformación Transitoria mediado por *R. rhizogenes* para el Análisis Funcional de Genes

La transformación mediada por *R. rhizogenes* requiere de tres etapas: transformación de la bacteria, selección de la colonia bacteriana efectiva e infección en el material (Figura 7). Si bien la transformación puede realizarse con la bacteria sin contenido de plásmido o gen de interés, en la mayoría de los casos el interés es analizar la funcionalidad de un gen mediante sobreexpresión y/o silenciamiento del mismo. Por lo tanto, se debe disponer de los plásmidos adecuados, incorporarlos en la bacteria y probar la efectividad de la infección bacteriana. En este trabajo se puso a punto la técnica sin realizar ningún clonado de gen dentro de los plásmidos, pero sí utilizando los plásmidos de sobreexpresión y silenciamiento vacíos, es decir sin la incorporación de genes de interés.

Se utilizó la cepa hipervirulenta K599 de *R. rhizogenes*, la cual contiene el plásmido RI 2659 de tipo cucumopina (Xiang 2016) y resistencia a espectinomicina 50 $\mu\text{g} \cdot \mu\text{L}^{-1}$ (SmR). La cepa fue proporcionada por el Dr. Germán Robert del Instituto de Fisiología y Recursos Genéticos Vegetales (IFRGV) del Instituto

Nacional de Tecnología Agropecuaria (INTA), Córdoba, Argentina. Como material vegetal se utilizaron semillas de soja del genotipo Génesis 5601.

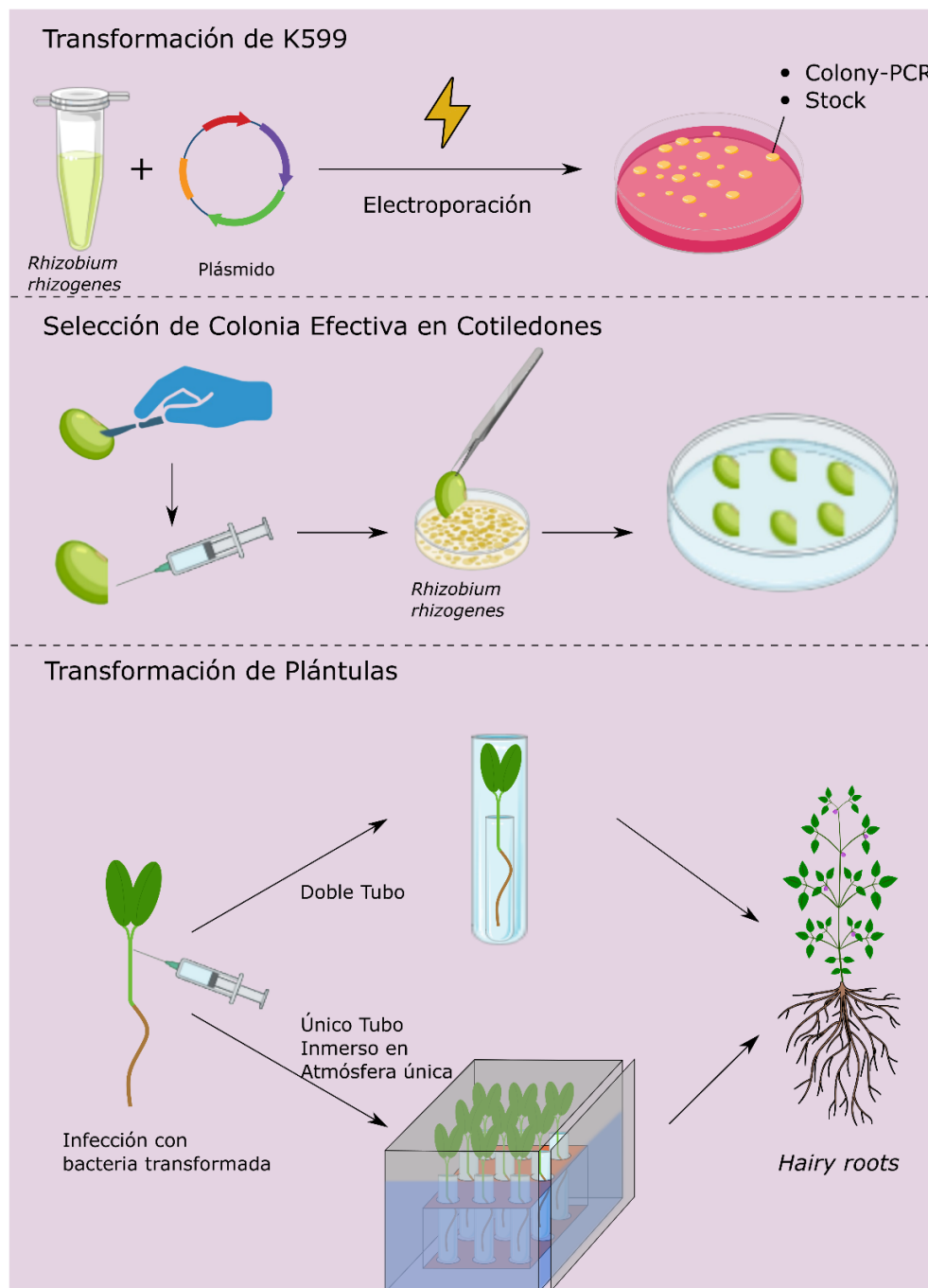


Figura 7 Esquema de las etapas seguidas en la transformación transitoria de plantas de soja mediada por *Rhizobium rhizogenes*. Para el stock se realiza cultivo líquido en medio LB y se almacena a -80°C en glicerol al 15%. En transformación de plántulas, en sistema de transformación en único tubo inmerso en atmósfera (TUTIA), se representa en azul el agua dentro de la caja, utilizada para generar la atmósfera saturada necesaria para el crecimiento y desarrollo de la infección. **K599:** cepa hipervirulenta de *R. rhizogenes*.

2.5.1 Selección de la Colonia Bacteriana Efectiva

A partir de una alícuota de K599 se realizaron estriados en dos placas LB-agar con $50 \mu\text{g} \cdot \mu\text{L}^{-1}$ de Sm, las cuales se incubaron durante 48 h a 28°C . Se aislaron cuatro colonias de cada una y se pusieron a crecer en 3 mL de medio LB líquido con $50 \mu\text{g} \cdot \mu\text{L}^{-1}$ de Sm y acetosiringona (AS) 150 mM durante 48 h a 28°C en agitación a 180 rpm. La acetosiringona es un flavonoide que favorece la infección del rizobio en la planta ya que es reconocido para la proteína virulenta VirA del rizobio, activando subsecuentemente otros genes *vir* (Kereszt et al. 2007; De Saeger et al. 2021). Posteriormente, para cada colonia, se impregnaron 50 μL del cultivo líquido en una placa LB-agar con $50 \mu\text{g} \cdot \mu\text{L}^{-1}$ de Sm y AS 150 mM durante 48 h a 28°C de forma de obtener un césped de células para realizar la infección en material vegetal. Además, cada colonia fue alicuotada en glicerol 15% para así disponer de la colonia efectiva una vez elegida.

La evaluación de la efectividad de colonia se realizó en cotiledones de soja. Se utilizaron seis cotiledones por colonia para seleccionar la colonia efectiva. Las semillas fueron desinfectadas con hipoclorito de sodio al 15% durante 30 minutos realizando un posterior triple lavado con agua destilada. Luego se colocaron sobre papel y algodón humedecidos dentro de un *tupper*, y se llevaron a estufa a 28°C durante aproximadamente 48-72 h. En este tiempo se comienza a ver el inicio del crecimiento de la radícula, momento suficiente para realizar la infección.

Se obtuvieron los cotiledones mediante remoción de la testa de la semilla. En cada cotiledón se realizó una remoción de una porción basal, cercana al nudo cotiledonar, asegurándose de que el primordio foliar sea removido. A su vez, este corte serviría para realizar la infección bacteriana, pues se precisa justamente de una herida para que suceda el proceso de infección. En dicha herida se realizaron además pequeñas incisiones con el extremo de una aguja de jeringa, a los efectos de favorecer en mayor medida el ingreso de la bacteria. Finalmente, se embetunó la herida con el césped bacteriano y se colocaron en una cámara de crecimiento con un fotoperíodo de 16/8 h (luz/oscuridad), a 25°C y una radiación de $200 \mu\text{mol de fotones} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ (Figura 7). Se monitoreó la aparición de callos de infección y *hairy roots* (HR) cada 5 días durante 15-20 días, tiempo suficiente para la aparición y crecimiento adecuado de HR para su evaluación (Kereszt et al. 2007).

2.5.2 Transformación de K599 con Plásmidos de Sobreexpresión y Silenciamiento por Electroporación

Para realizar la transformación de la bacteria por electroporación es necesario previamente generar bacterias electrocompetentes. Para ello se utilizó la colonia efectiva seleccionada previamente, la cual fue plaqueada en LB-agar con iguales condiciones mencionadas en el párrafo anterior, sin AS. Se preparó un inóculo de 5 mL a 28°C *overnight* en medio LB con Sm. Se pipetearon 10 μL del inóculo en 100 mL de medio LB con Sm a 28°C durante 16-20 h en agitación a 180 rpm. Estas condiciones son suficientes para alcanzar una densidad óptica (DO) del cultivo a 550 nm de aproximadamente 1, condición necesaria para proseguir. Se realizaron tres centrifugaciones de 10 minutos a $3000g$ y a 4°C , descartando sobrenadante y manteniendo el pellet en hielo. Luego de cada centrifugación se agregaron 50 mL de glicerol 10% estéril y frío, excepto luego de la tercera que se agregaron 25 mL. Finalmente, se realizó una centrifugación de 5 minutos en iguales condiciones, se descartó el sobrenadante y se resuspendió el pellet en 0,5 mL de glicerol. Se hicieron alícuotas de 50 μL y se almacenaron a -80°C . El protocolo de electrocompetencia fue cedido por el Dr. Germán Robert del IFRGV, INTA, Córdoba, Argentina.

La transformación se hizo con un electroporador (Gene Pulse Xcell Total System, Biorad), usando el programa "Agrobacterium". Se utilizó una alícuota de 50 μL de bacterias competentes obtenidas anteriormente a la cual se le adicionó 1 μL de ADN plasmídico (200 $\text{ng}/\mu\text{L}$) y se la traspasó a una cubeta de electroporador. Se incubó la cubeta durante 30 minutos en hielo, luego 90 segundos a 42°C y finalmente 2 minutos en hielo. Se colocó la cubeta en el electroporador y se aplicó un impulso eléctrico a 2,4 kV, 200 ohm y 25 μF . Inmediatamente después se agregaron 200 μL de medio SOC y se incubaron las células durante 4 horas a 28°C . Finalmente, se plaquearon las células en medio LB-agar conteniendo Sm y los antibióticos adecuados según el plásmido, kanamicina 50 $\mu\text{g}\cdot\mu\text{L}^{-1}$, en los vectores de sobreexpresión, y espectinomicina 100 $\mu\text{g}\cdot\mu\text{L}^{-1}$, en el vector de edición génica. Si bien la presencia del vector podría ser evidenciada por haber crecido en selección, puede suceder la aparición de resistencia espontánea en colonias bacterianas, por lo que se chequeó y confirmó la inserción del vector mediante *colony*-PCR empleando *primers* para amplificar el gen DsRED. Las secuencias de dichos *primers* son las siguientes: GGATCAGATTGTCGTTTCC (DsRED Fw) y CACTGATAGTTTAATTCCCGATC (DsRED Rv), esperándose amplicones de aproximadamente 1800 pares de bases.

Se utilizaron los plásmidos de sobreexpresión 35S:RUBY y pEAQHT-RED (Sainsbury et al. 2009), y el plásmido de *knockout* (KO) por edición génica piLuCAS-RED, este último generado y cedido por la Dra. Luciana Fleitas del Laboratorio de Biología Molecular Vegetal (FCien-Udelar). El 35S:RUBY presenta como marcador el RUBY que puede ser visualizado a simple vista, pues las HR positivas son de color violeta. Los plásmidos pEAQHT-RED y piLuCAS-RED contienen como marcador el DsRED, que puede ser visualizado bajo fluorescencia, para la selección de las *hairy roots* transformadas.

2.5.3 Infección de plántulas, confirmación y evaluación de la transformación

Se cultivó un césped bacteriano de cada bacteria transformada siguiendo los pasos establecidos en el apartado de selección de colonia efectiva. Paralelamente, se pusieron a germinar semillas de soja siguiendo los pasos mencionados anteriormente. Por un lado, se evaluó la eficiencia de cada construcción infectando cotiledones. Por otro lado, se infectaron plántulas con cada construcción para observar el funcionamiento del sistema de transformación en el material de interés. En este caso, las semillas se dejaron germinar durante 4-5 días, a efectos de que la radícula logre un crecimiento óptimo para el uso del sistema. En los casos que fue necesario se removió la testa de la semilla previo a la infección cuidando de herir los cotiledones, en algunas semillas la testa ya se encontraba desprendida totalmente.

Para realizar la infección se hizo una incisión en forma de cuña a la altura del comienzo del hipocótilo, procurando que la profundidad del corte alcance los tejidos vasculares, esto es una profundidad apenas menor al 50% del diámetro del hipocótilo. En el corte se realizaron además pequeñas incisiones con una aguja justamente para favorecer la infección bacteriana. Se embetunó el corte con el césped bacteriano y se colocó en un tubo conteniendo medio de crecimiento ByD estéril (Broughton and Dilworth 1971). Dicho tubo se tapó con papel de aluminio al cual se le realizó un orificio para que ingrese la radícula. Asimismo, el aluminio ofició como soporte de los cotiledones. La herida infectada quedó por encima del aluminio. A su vez, dado que se quiere alta humedad (más del 80%) para la emergencia de las *hairy roots*, se colocó el tubo dentro un tubo más grande, el cual contenía agua. El tubo externo fue sellado con papel film para evitar pérdidas de humedad y mantener una atmósfera saturada. Cuando la plántula creció y alcanzó el nivel del film, se quitó el film y se colocó por fuera una bolsa transparente humedecida para evitar cambios bruscos de humedad ambiente y que estos perjudiquen a la planta en algún sentido.

Asimismo, se experimentó en un sistema de transformación de único tubo inmerso en atmósfera (TUTIA), el cual establece una atmósfera común para todos los tubos mediante el uso de una caja de plástico, con el objetivo de reducir tiempos y esfuerzos en la implementación del sistema, dado que en cada instancia de infección se deben realizar infecciones de entre 80 y 100 plántulas para aumentar la obtención de resultados positivos.

Luego de 15-20 días las HR eran suficientemente grandes para observar bajo la lupa la fluorescencia de raíces transformadas positivas. Las raíces que no fueron positivas se removieron, así como la raíz principal. Inmediatamente, se probó si la planta podía sobrevivir y desarrollarse normalmente en maceta con sustrato en las condiciones de crecimiento utilizadas por el laboratorio y mencionadas anteriormente. Asimismo, se inocularon las plantas con *B. elkanii* U1302 para probar si la nodulación efectivamente ocurre en HR.

3. RESULTADOS

Los tratamientos analizados en este trabajo combinaron las condiciones de nodulación (N) y no nodulación (NN), y de plantas a capacidad de campo y con restricción hídrica (CC y RH, respectivamente). Plantas a capacidad de campo no presentaron restricciones hídricas en el sustrato, por ende, el tratamiento CC fue considerado el tratamiento control de la condición hídrica. De esta manera, los cuatro tratamientos estudiados fueron: plantas no noduladas a capacidad de campo (NN+CC), plantas no noduladas y con restricción hídrica (NN+RH), plantas noduladas a capacidad de campo (N+CC), y plantas noduladas con restricción hídrica (N+RH). Habiendo definido dichos tratamientos, los contrastes biológicos analizados fueron: *i*) N+RH vs N+CC, que evidencia el efecto de la restricción hídrica en plantas noduladas; *ii*) N+RH vs NN+RH, donde se estudia el efecto de la restricción hídrica en contextos de nodulación diferentes; *iii*) NN+RH vs NN+CC, en el cual se analizó el efecto de la restricción hídrica en plantas sin nodular; y *iv*) N+CC vs NN+CC, donde el efecto de la nodulación es analizado en plantas que siempre experimentaron condiciones hídricas óptimas.

3.1 Caracterización hídrica de plantas de soja no noduladas y noduladas en diferente contexto hídrico

Durante el período de restricción hídrica, todas las plantas -tanto plantas noduladas (N) como no noduladas (NN)- correspondientes a tratamientos involucrando sustrato con condiciones hídricas de capacidad de campo (CC) se mantuvieron a un porcentaje gravimétrico de agua de 40% (Figura 8A), porcentaje equivalente a CC. Las raíces de plantas a CC fueron cosechadas aleatoriamente durante el período de restricción hídrica (RH). Es así que se cosecharon plantas N+CC y NN +CC hasta el quinto y séptimo día del período de RH, respectivamente (Figura 8A). La conductancia estomática media de plantas a CC monitoreada durante el período de RH estuvo entre 176 y 249 mmol H₂O m⁻² s⁻¹ en plantas N, y entre 123 y 204 mmol H₂O m⁻² s⁻¹ en plantas NN (Figura 8B).

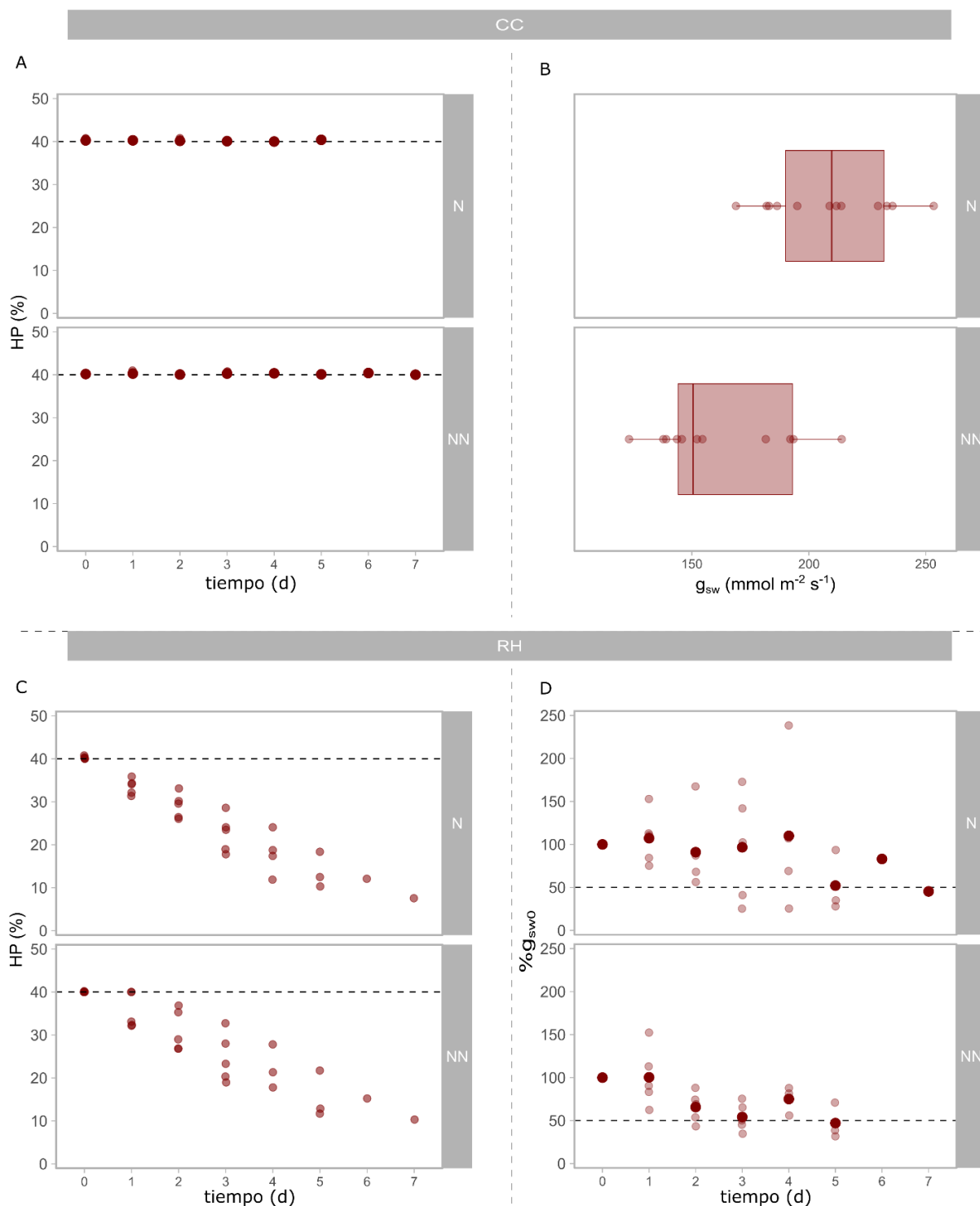


Figura 8 Caracterización hídrica de todas las unidades experimentales utilizadas en el ensayo durante el período de restricción hídrica. A Contenido de agua en peso (HP) del sustrato en plantas mantenidas a Capacidad de Campo (CC), noduladas y no noduladas, arriba y abajo, respectivamente. **B** Variabilidad de la conductancia estomática (g_{sw}) del segundo trifolio de plantas noduladas y no noduladas mantenidas a CC. **C** Contenido de agua en peso (HP) del sustrato en plantas con restricción hídrica (RH), noduladas y no noduladas, arriba y abajo, respectivamente. **D** Conductancia estomática en porcentaje con respecto a la conductancia estomática en el día inicial de la restricción hídrica (g_{sw0}) del segundo trifolio de plantas noduladas y no noduladas, arriba y abajo, respectivamente. En todos los casos el tiempo en días luego de iniciado el período de restricción hídrica.

En el caso de plantas en RH, el porcentaje gravimétrico de agua del sustrato fue de 40% en el día 0 de la restricción, disminuyendo hasta un valor promedio de 22% en el tercer día, y entre 7-13% entre el quinto y séptimo día (Figura 8C). La conductancia estomática de plantas en RH en el inicio de la restricción hídrica tuvo valores promedios entre 145 y 254 mmol H₂O.m⁻².s⁻¹ en plantas N y entre 114 y 156 mmol H₂O.m⁻².s⁻¹ en plantas NN. Para cada planta, el valor de conductancia estomática en el día inicial fue la referencia (100%) para posteriormente cosecharla cuando el valor promedio de conductancia alcanzara un valor igual o menor al 50% del valor inicial (Figura 8D). La conductancia promedio de plantas N+RH se mantuvo en torno al 100% hasta el cuarto día de la restricción hídrica (Figura 8D). En el caso de plantas NN+RH la conductancia estomática promedio fue alrededor del 100% hasta el primer día de la RH (Figura 8D). En el segundo día de la RH, la primera planta NN+RH presentó valores de conductancia menores al 50% del valor inicial (Figura 8D), umbral definido para su cosecha (Figura 9). En plantas N+RH esto sucedió en el tercer día de la RH (Figura 8D y Figura 9). El cuarto día de RH fue la fecha de cosecha promedio para todas las plantas en RH. Para plantas NN+RH el día promedio de cosecha fue el tercero de la RH. Mientras que para N+RH el día de cosecha promedio fue mayor, siendo éste el quinto día (Figura 9).

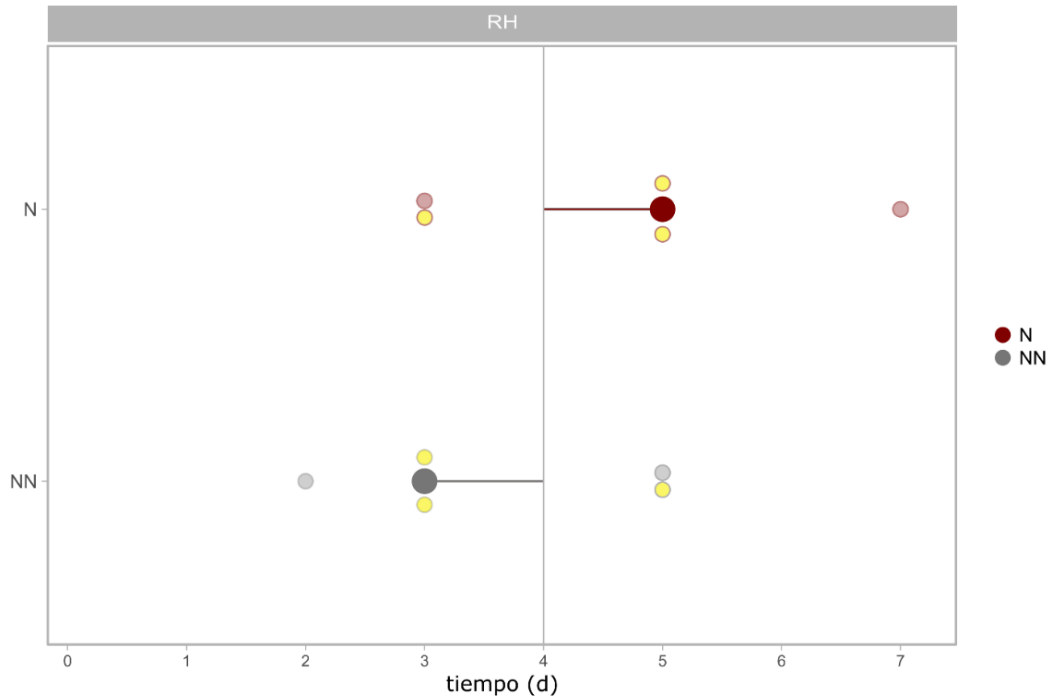


Figura 9 Día de cosecha de plantas noduladas (N) y no noduladas (NN) en restricción hídrica (RH). El tiempo en días luego de iniciado el período de restricción hídrica. La línea central al cuarto día corresponde al día medio de cosecha de todas las plantas en RH. Puntos de mayor tamaño corresponden al día medio de cosecha según condición de nodulación. En amarillo plantas que fueron utilizadas para análisis RNA-seq.

3.2 Análisis de Datos de Secuenciación

Una vez secuenciadas las dos fracciones de ARNm de cada réplica biológica se obtuvieron archivos *fastq* conteniendo entre 20 y 31 millones de *paired-end reads*. Luego de realizar la eliminación de adaptadores y control de calidad mediante *Trimmomatic* se retuvieron entre 19 y 30 millones de *reads*, haciendo ello un porcentaje de *reads* descartados de entre 2,8 y 6,8% (Tabla 1). Por otro lado, la tasa de mapeo de *reads* retenidos contra transcriptoma de referencia (GCF_000004515.6, *Glycine max* v4.0, sin ARNr) mediante el programa *Salmon* fue de entre 28 y 94%. Se observaron dos franjas de tasa de mapeo, una comprendida entre 28 y 54% y otra entre 88 y 94% (Tabla 1). Los mapeos entre 28 y 54% sucedieron todos en fracciones de ARNm PAR de una réplica biológica del tratamiento NN+CC, dos del NN+RH, una del tratamiento N+CC y tres del N+RH. Esta disminución del porcentaje de mapeo fue producto de que se mapeó contra transcriptoma excluyendo transcritos codificantes para ARNr y, por ende, para las muestras mencionadas la depleción del ARNr no fue total. Las restantes cuatro fracciones PAR (dos NN+CC, una N+RH y una N+CC) presentaron tasas de mapeo entre 88 y 94% (Tabla 1).

Tabla 1 Estadísticas descriptivas de procesamiento y mapeo de *reads* para cada muestra biológica y sus fracciones de ARN.

Muestra	Fracción de ARN	# PE <i>reads</i>	# PE <i>reads</i> retenidos	% <i>Reads</i> descartados	Tasa de mapeo (%)
NN+CC.1	TOT	28724521	27835677	3.09	92.44
NN+CC.1	PAR	21921840	21120617	3.65	32.56
NN+CC.2	TOT	21753412	20264985	6.84	88.16
NN+CC.2	PAR	22394571	21051722	6.00	92.08
NN+CC.3	TOT	21810738	21114180	3.19	93.19
NN+CC.3	PAR	22617433	21809961	3.57	39.34
NN+RH.1	TOT	31556167	30146963	4.47	93.44
NN+RH.1	PAR	31552221	30112902	4.56	88.41
NN+RH.2	TOT	26705209	25876982	3.1	91.38
NN+RH.2	PAR	20975297	20114247	4.11	34.58
NN+RH.3	TOT	23536085	22833397	2.99	93.26
NN+RH.3	PAR	20207273	19569084	3.16	94.05
N+CC.1	TOT	23861620	23190883	2.81	92.02
N+CC.1	PAR	24698171	23831711	3.51	40.99
N+CC.2	TOT	24346013	23323586	4.2	91.52
N+CC.2	PAR	21389089	20551342	3.92	41.46
N+CC.3	TOT	22633413	21748412	3.91	88.31
N+CC.3	PAR	23903868	23002868	3.77	28.62
N+RH.1	TOT	24718265	23911783	3.26	92.74
N+RH.1	PAR	24653064	23705724	3.84	46.76
N+RH.2	TOT	27897774	27032943	3.1	90.79
N+RH.2	PAR	29472901	28450341	3.47	54.09
N+RH.3	TOT	31553305	29413819	6.78	90.13
N+RH.3	PAR	21731838	20478608	5.77	93.01

TOT: fracción de ARNm total. **PAR:** fracción de ARNm asociado a polisomas. **PE:** *paired-end*. *Reads* **retenidos** luego de procesamiento con *Trimmomatic* (versión 0.39, Bolger et al. 2014). Tasa de **mapeo** de *reads* retenidos contra transcriptoma de referencia sin ARNr (GCF_000004515.6, *Glycine max* v4.0, sin ARNr).

El rango de conteos “crudos” de genes, es decir directamente obtenidos de la cuantificación vía *Salmon* y sin realizar ningún procesamiento posterior, en todas las muestras fue entre 0 y 1 millón, con una mediana que varió entre 14 y 99 conteos ([Figura 10A](#)). En estos datos están incluidos los 59148 genes del genoma de soja obtenidos de la base de datos de NCBI. Luego de realizar un paso de filtrado de genes con al menos un conteo entre todas las muestras se retuvieron 43070 genes. En la [Figura 10B](#), se muestra la dispersión de los conteos normalizados de dichos genes a lo largo de todas las muestras. Debido a que la normalización es un paso interno dentro del procesamiento mediante la función *DESeq2* del paquete homónimo de R cuando se realiza el análisis de expresión diferencial se muestra cómo se visualizarían los conteos en dicho paso. El rango de conteos normalizados estuvo entre 0 y 3,2 millones, con una mediana entre 84 y 102 conteos ([Figura 10B](#)). Gracias a la normalización, las diferencias en cantidad de transcritos por muestras quedan subsanadas.

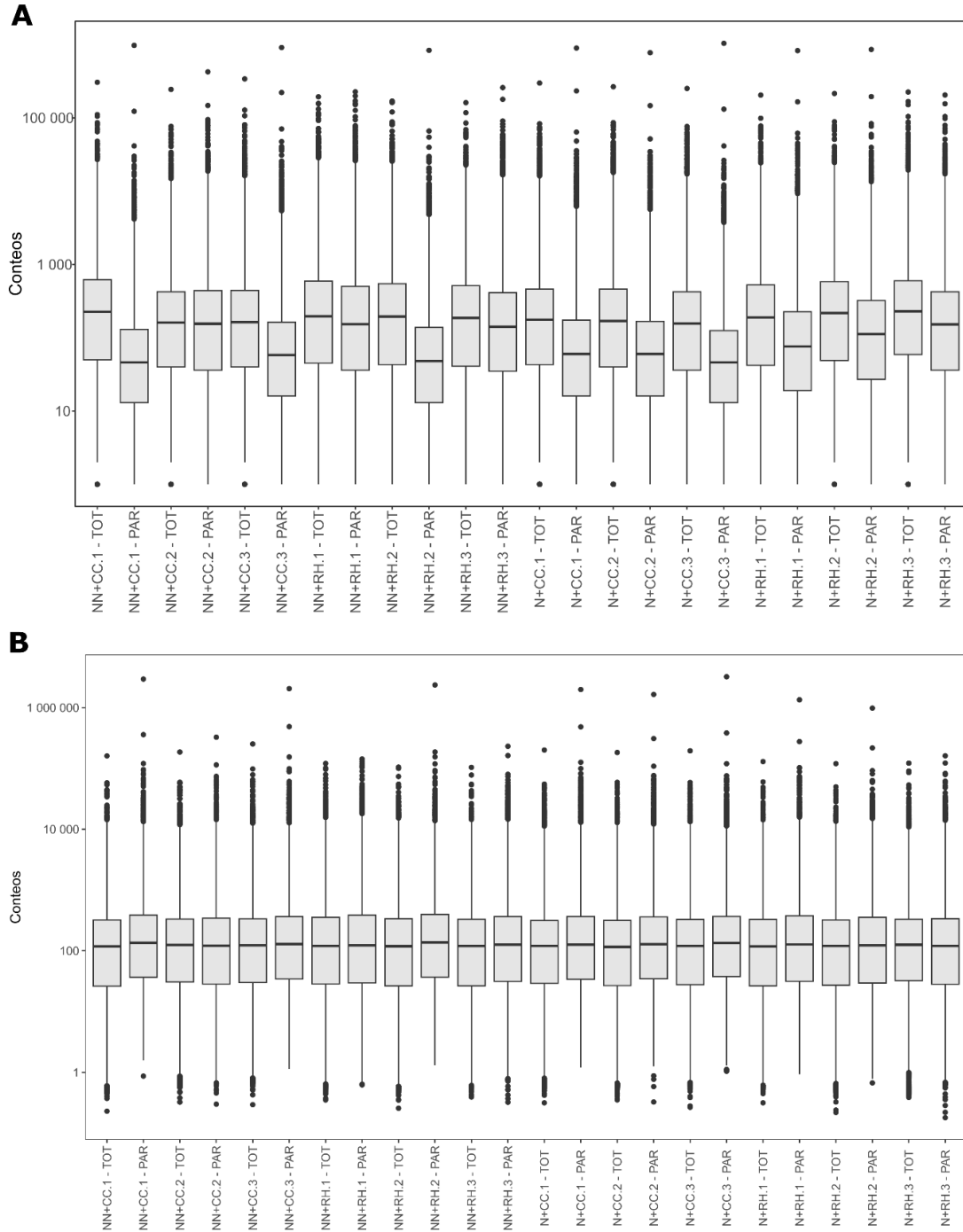


Figura 10 Boxplots de conteos de genes obtenidos con *Salmon* y el paquete de R *tximport* para las fracciones de ARNm total (TOT) y ARNm asociado a polisomas (PAR) de cada tratamiento combinado (NN+CC, NN+RH, N+CC, N+RH). **A** Conteos “crudos”. **B** Conteos normalizados de genes con al menos un conteo entre todas las muestras.

3.3 Análisis de Componentes Principales

El análisis exploratorio de datos de RNA-seq mediante análisis de matriz de distancias y análisis de componentes principales (PCA) (Figura 11A y B) mostró que las réplicas biológicas *clusterizaron* juntas. En la matriz de distancias (Figura 11A) se observó la formación de dos grandes *clusters*, agrupando por un lado a todas las muestras correspondientes al tratamiento de plantas noduladas en restricción hídrica (N+RH) y al resto de las muestras en otro grupo. Mediante el PCA no se observó efecto *batch* sobre los datos de secuenciación, es decir que no hubo efecto de la fecha o lote de secuenciación (Figura S1). A su vez, se observó que los cuatro tratamientos combinados se diferenciaron claramente en cuanto a la expresión génica (Figura 11B) y que las fracciones de ARNm total y asociado a polisomas (PAR) se ubicaron conjuntamente para cada una de las muestras biológicas (Figura 11B, círculos y triángulos, respectivamente). Esto devela la buena calidad de los datos y permite continuar con los análisis sin presentar ningún inconveniente extra.

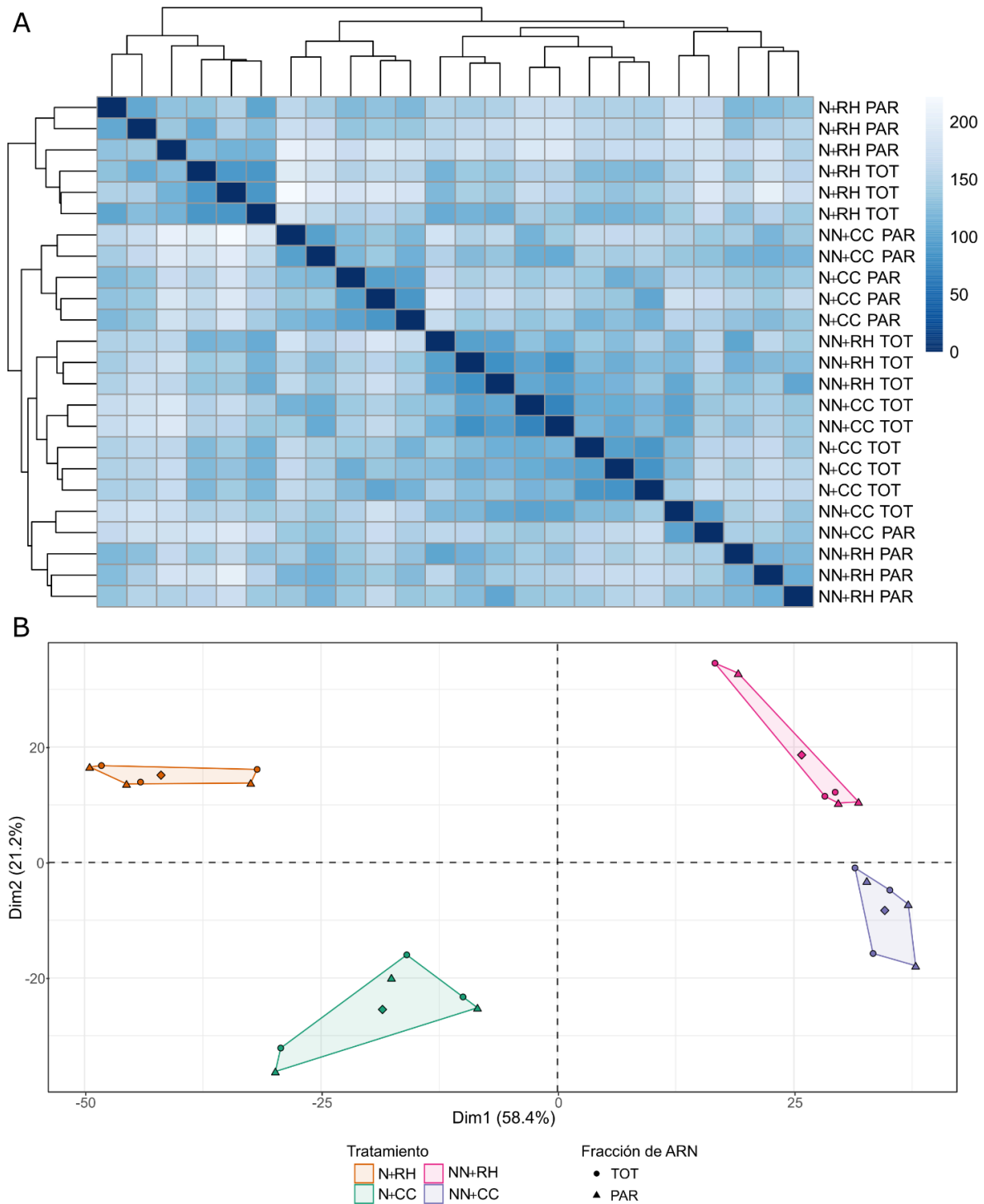


Figura 11 Análisis descriptivo de los datos de RNA-seq. A *Heatmap* de distancias euclidianas entre muestras. Se incluyen las tres réplicas de cada tratamiento combinado (NN+CC, NN+RH, N+CC, N+RH) contemplando las fracciones de ARNm TOTAL y PAR. **B** Análisis de Componentes Principales. Dim1 y Dim2 refieren a las dimensiones 1 y 2, respectivamente. TOT: fracción de ARNm total. PAR: fracción de ARNm asociado a polisomas.

La dimensión 1 (Dim1) del PCA explicó un 58,4% del total de la varianza separando las muestras por condición de nodulación, ubicándose en valores negativos los tratamientos que contemplaron una condición de nodulación (N+CC y N+RH), y en valores positivos los tratamientos de plantas que no fueron noduladas (NN+CC y NN+RH) (Figura 11A). Por otro lado, la dimensión 2 (Dim2) explicó un 21,2% de la varianza separando las muestras según condición hídrica, tomando valores negativos los tratamientos a capacidad de campo, es decir sin restricción hídrica (NN+CC y N+CC), y valores positivos los tratamientos que experimentaron restricción hídrica (NN+RH y N+RH) (Figura 11B).

Lo anterior quedó sustentado cuando se realizó un análisis de enriquecimiento de procesos biológicos en los 50 genes que más contribuyeron a la generación de la Dim1 y de la Dim2 (Figura 12). Para la Dim1, se encontró que la mayoría de ellos correspondían a genes vinculados al proceso de nodulación como leghemoglobinas A, C1, C2 y C3, nodulinas tempranas y diversas nodulinas (Tabla 2). Estos genes presentaron valores negativos de Dim1, lo que significó que muestras ubicadas hacia la izquierda del gráfico (Figura 11B) presentaron una expresión mayor de dichos genes. Las muestras de los tratamientos de nodulación (N+CC y N+RH) quedaron ubicadas en dicho espacio. Asimismo, el proceso biológico de nodulación fue el mayor contribuyente en la generación de la Dim1 aportando aproximadamente un 30% de la variabilidad (Figura 12B). Los procesos de proteólisis, transporte de oxígeno, transporte de fosfato y regulación de la transcripción le siguieron en importancia jerárquica contribuyendo entre 3 y 8% en la variabilidad de la Dim1 (Figura 12B). Esto indicaría que dichos procesos son importantes en el metabolismo de plantas noduladas.

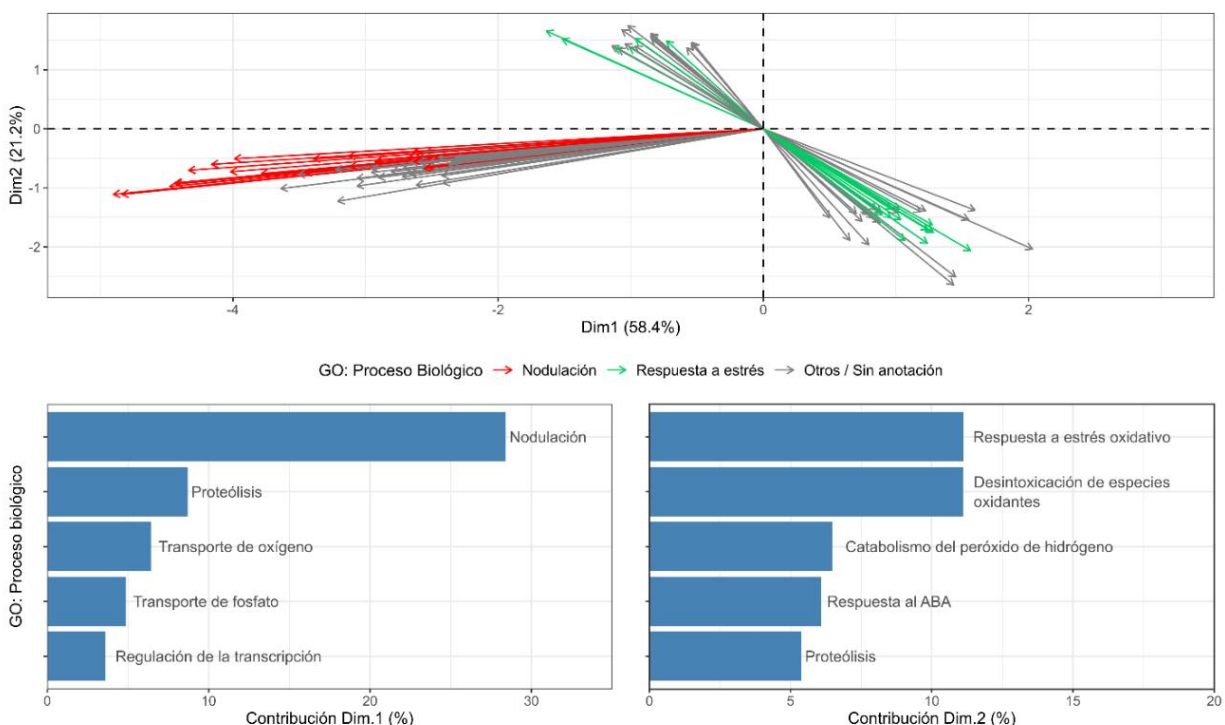


Figura 12 Análisis de enriquecimiento funcional de *Gene Ontology* (GO) sobre los principales genes que contribuyen en la generación de las dos primeras dimensiones del Análisis de Componentes Principales de datos de RNA-seq. (A) Análisis de Componentes Principales mostrando 50 genes que más contribuyen en las dimensiones 1 y 2. Se muestran coloreados según término GO anotado. (B) y (C) Procesos biológicos más contribuyentes en la generación de las dimensiones 1 y 2, respectivamente. ABA: ácido abscísico.

Tabla 2 Lista de genes anotados con proceso biológico de nodulación que más contribuyeron a la generación de la dimensión 1 del Análisis de Componentes Principales de datos de RNA-seq.

Descripción	GeneID	GlymaID
leghemoglobina A	100527427	Glyma_10G199100
leghemoglobina C1	100785236	Glyma_10G199000
leghemoglobina C2	100527379	Glyma_20G191200
leghemoglobina C3	100527391	Glyma_10G198800
nodulina temprana ENOD55-2	547770	Glyma_17G073400
nodulina temprana-55-2-like	100784860	Glyma_02G204500
nodulina temprana-70	100817841	Glyma_11G238500
nodulina temprana-75	100775551	Glyma_08G131925
nodulina temprana-93	547773	Glyma_06G216500
nodulina-16	100527599	Glyma_02G265200
nodulina-16-like	547914	Glyma_02G265400
nodulina-16-like	106796780	Glyma_02G265300
nodulina-20	100527279	Glyma_13G328800
nodulina-21	547974	Glyma_05G121600
nodulina-22	548101	Glyma_15G045000
nodulina-23	547902	Glyma_20G024200
nodulina-24	100527648	Glyma_14G052400
nodulina-26	547904	Glyma_08G120100
nodulina-26b (AA 1-213)	547905	Glyma_19G074000
LOC100527289 no caracterizado	100527289	Glyma_01G122500

GeneID: identificador de base de datos NCBI. GlymaID: identificador "Glyma" versión 4 utilizado en varias bases de datos genómicos.

Por otro lado, los procesos de respuesta a estrés oxidativo y de desintoxicación de especies oxidantes fueron los dos más contribuyentes en la generación de la Dim2 explicando aproximadamente un 12% de la varianza de dicha dimensión ([Figura 12C](#)). Le siguieron en importancia jerárquica los procesos de catabolismo del peróxido de hidrógeno, respuesta al ácido abscísico y proteólisis, aportando entre 5 y 7% de la variabilidad de la Dim2 ([Figura 12C](#)), indicando que, en las respuestas a déficit hídrico, independientemente de la condición de nodulación, dichos procesos biológicos serían relevantes. Asimismo, se encontró que dentro de los 50 genes que contribuyeron en mayor medida a la generación de la Dim2, 14 genes estuvieron vinculados a procesos de respuesta a estrés ([Figura 12A](#)). Dentro de ellos se encontraron genes que codifican para dos dioxigenasas 9-cis-epoxicarotenoides NCED1 (enzima de la vía de biosíntesis del ABA), una proteína LEA5 y dos galactinol sintasas ([Tabla 3](#)) que presentaron valores positivos de la Dim2. Esto significó que la expresión de dichos genes estuvo asociada con los tratamientos de restricción hídrica, ya que dichos tratamientos (NN+RH y N+RH) quedaron ubicados en zonas positivas de la Dim2 ([Figura 11B](#)). Por otro lado, genes que codifican para cuatro receptores de ácido abscísico tipo PYL4, una glucosidasa y dos peroxidasas ([Tabla 3](#)) presentaron valores negativos de Dim2, significando que la expresión de dichos genes en los tratamientos de RH fue menor en comparación con los tratamientos que involucraron una condición hídrica de capacidad de campo. Por lo que dichos genes estarían teniendo relevancia en la modulación de las respuestas a estrés independientemente de la condición de nodulación. Es de importancia notar que en el listado de estos 14 genes asociados a

respuestas a estrés hay varios genes que codifican para una misma proteína, por ejemplo, los cuatro receptores de ABA tipo PYL4 (no se tratan de isoformas pues los loci son diferentes).

Tabla 3 Lista de genes anotados con el proceso biológico de respuesta a estrés que más contribuyeron a la generación de la dimensión 2 del Análisis de Componentes Principales de datos de RNA-seq.

Descripción	GeneID	GlymaID
dioxigenasa 9-cis-epoxicarotenoide NCED1 (+)	100777166	Glyma_08G176300
dioxigenasa 9-cis-epoxicarotenoide NCED1 (+)	100814399	Glyma_15G250100
proteína contra desecación LEA 5 (+)	547885	Glyma_17G027400
galactinol sintasa 2 (+)	100792129	Glyma_03G229800
galactinol sintasa 2 (+)	100800249	Glyma_19G227800
proteína 65 kDa inducida por bajas temperaturas (+)	100791989	Glyma_20G158500
proteína proteasa aspártica (-)	100805846	Glyma_08G321200
receptor ácido abscísico PYL4-like (-)	100798803	Glyma_02G261900
receptor ácido abscísico PYL4-like (-)	100808416	Glyma_11G233300
receptor ácido abscísico PYL4-like (-)	100804400	Glyma_14G056300
receptor ácido abscísico PYL4-like (-)	121174020	Glyma_18G023966
glucano endo-1,3-beta-glucosidasa (-)	547822	Glyma_03G132700
peroxidasa 5 (-)	100815671	Glyma_09G277800
peroxidasa 5 (-)	100796311	Glyma_09G284700

GeneID: identificador de base de datos NCBI. (+) y (-) refieren a si los genes tomaron valores positivos y negativos en la dimensión 2, respectivamente. GlymaID: identificador "Glyma" versión 4 utilizado en varias bases de datos genómicos.

3.4 Análisis de Expresión Diferencial de Genes

Debido a que en el análisis exploratorio de los datos de secuenciación los tratamientos combinados quedaron diferenciados en cuanto a la expresión génica global (réplicas de un mismo tratamiento y fracciones de ARN de una misma réplica biológica se parecieron entre si), se procedió analizando la expresión diferencial de genes entre los diferentes tratamientos combinados. En primer lugar, se estudiaron los contrastes generales, es decir cómo los diferentes contextos de nodulación (N vs NN) independiente de la condición hídrica, y cómo las diferentes condiciones hídricas (RH vs CC) independientemente de los contextos de nodulación afectaron la expresión génica a nivel de transcriptoma y de traductoma (Figura 13). La restricción hídrica provocó una mayor alteración de la expresión génica frente a la nodulación ya que aproximadamente 2000 genes fueron diferencialmente expresados (DEGs), contabilizando tanto genes *down*-regulados como *up*-regulados, sea a nivel transcriptómico (Figura 13C) como traductómico (Figura 13D), en el contraste RH vs CC. Mientras que en el contraste N vs NN, aproximadamente 1700 y 1100 genes fueron DEGs a nivel transcriptómico y traductómico, respectivamente (Figura 13A y B).

Por otra parte, la condición de nodulación provocó una notoria mayor inducción de genes en comparación contra la represión tanto a nivel de transcripción como de traducción, donde el número de genes *up* fue más del doble que el número de genes *down* en ambas fracciones (Figura 13A y B). En cambio, la condición de restricción hídrica provocó relativamente una mayor represión génica, aunque no

tan exacerbada, en donde existieron entre un 10 y 24% más de genes *down*-regulados que *up*-regulados en ambas fracciones de ARNm (Figura 13C y D). En otros términos, significa que la nodulación induce, en mayor medida, la activación de programas génicos y que la falta de agua reprime la expresión en mayor medida.

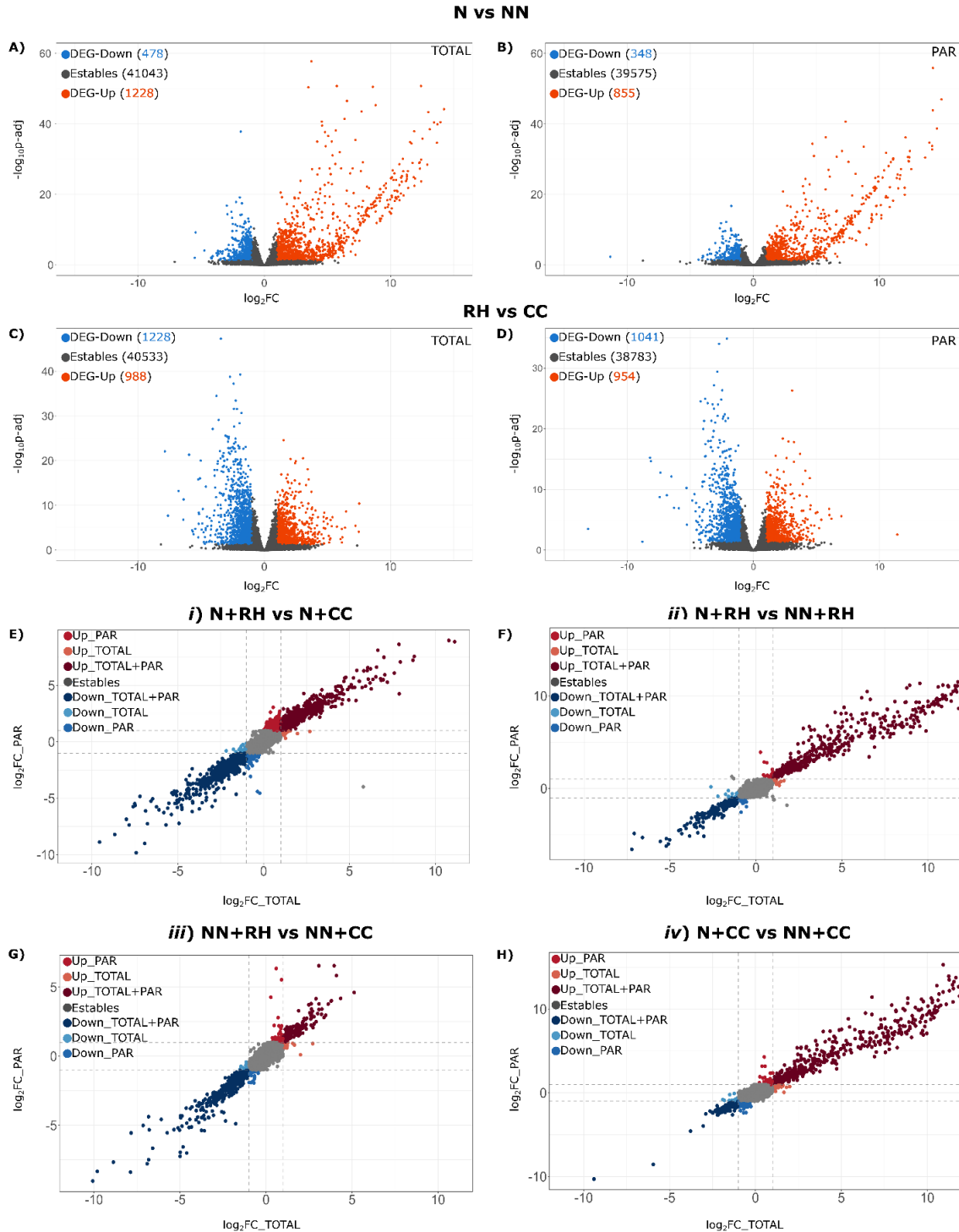


Figura 13 Respuestas transcripcionales y traduccionales. Volcano plots mostrando genes estables y genes diferencialmente expresados *down*- y *up*-regulados en el contraste global de plantas noduladas comparadas con plantas no noduladas a nivel transcripcional -TOTAL- (A) y traduccional -PAR- (B). C y D muestran mismas condiciones de genes para el contraste global entre plantas en déficit hídrico y plantas sin restricción hídrica, para las fracciones de ARNm TOTAL y PAR, respectivamente. E-H Gráficos de dispersión del *fold change* de genes en fracción PAR en función del *fold change* en fracción TOTAL para las comparaciones entre tratamientos combinados. E Contraste i) N+RH vs N+CC. F Contraste ii) N+RH vs NN+RH. G Contraste iii) NN+RH vs NN+CC. H Contraste iv) N+CC vs NN+CC.

Entrando más en detalle, se compararon los niveles de expresión génica en la fracción PAR (conjunto de ARNm asociados a polisomas) en función de los niveles en la fracción TOTAL (ARNm total) para cada uno de los cuatro tratamientos combinados: *i*) N+RH vs N+CC, *ii*) N+RH vs NN+RH, *iii*) NN+RH vs NN+CC y *iv*) N+CC vs NN+CC (Figura 13E, F, G y H). Cuando se analizó el set de DEGs con mayores valores de *fold change* (FC) a nivel traduccional en el contraste *i*, se encontraron genes vinculados al metabolismo de las hormonas ácido abscísico (ABA), etileno, giberelina y auxinas tales como genes que codifican para proteína tipo SAUR responsiva a auxina, ácido indol-3-acético (IAA) sintetasa GH 3.6, proteína regulada por auxinas, proteínas reguladas por giberelinas, 1-aminociclopropano-1-carboxilato (ACC) sintasa, así como factores de transcripción vinculados al metabolismo hormonal como ABI5 y ERF062 que responden al ABA y etileno, respectivamente. Esto indicaría que modular el metabolismo de hormonas en una planta nodulada y sometida a déficit hídrico es de importancia meridiana y que la regulación predomina a nivel traduccional. También se encontraron factores de transcripción de las familias bHLH y WRKY, receptores quinasas, transportadores de aminoácidos, iones y azúcares dentro de los genes con regulación traduccional preferencial en el contraste *i* (N+RH vs N+CC). Por otro lado, algunos DEGs que presentaron regulación mayoritariamente traduccional en el contraste *ii* (N+RH vs NN+RH) fueron genes que participan en el metabolismo de hormonas como proteínas con motivo VQ, proteínas tipo *scarecrow* y el factor responsivo a etileno ERN1, y genes vinculados a la biosíntesis de isoflavonoides y de pared celular, como por ejemplo la chalcona sintasa y proteínas relacionadas a la arabinogalactona, respectivamente.

3.5 Identificación de DEGs preferenciales en cada tratamiento y nivel de regulación

Posteriormente, para poder tener las identidades génicas que se expresaron diferencialmente exclusivamente en cada contraste y fracción de ARNm estudiadas (o conjuntamente entre contrastes y fracciones) se intersecaron las listas de DEGs obtenidas para cada situación. Es decir, se intersecaron las listas de DEGs de las fracciones TOTAL y PAR de un contraste, obteniendo así los genes que presentaron regulación transcripcional o traduccional exclusiva y genes con regulación a ambos niveles para un contraste. Dicho proceso se hizo para DEGs *down*- y *up*-regulados por separado. A su vez, dichas listas DEGS se intersecaron entre contrastes de forma de remover de cada lista genes que fueron diferencialmente expresados debido al efecto biológico de un tratamiento (por ejemplo, el efecto de la restricción hídrica) y no del tratamiento combinado.

Por un lado, y poniendo en ejemplo, las listas del contraste *i* (N+RH vs N+CC), que evidencia el efecto de la restricción hídrica en un contexto de nodulación, se intersecaron con las listas del contraste *iii*, que muestra el efecto de la restricción hídrica en un contexto de no nodulación, a efectos de obtener los DEGs que se debieron a un efecto de la restricción hídrica independientemente del contexto de nodulación y de los DEGs preferenciales de cada contraste (Figura 14). Esto resultó en la obtención, por ejemplo, de 463 y 56 DEGs *up*-regulados a nivel traduccional (PAR) en el contraste *i* y *ii*, respectivamente, y 5 DEGs con regulación PAR compartidos entre ambos contrastes (Figura 14A). A su vez, cuando se observó la cantidad de genes con regulación traduccional preferencial con respecto a la cantidad total exclusiva de cada contraste, independientemente del estado, se vio que la cantidad fue mayor en el contraste *i* (N+RH vs N+CC) que en el contraste *iii* (NN+RH vs NN+CC), representando un 27% y un 15%, respectivamente (Figura 14A y C). Esto indica que bajo condiciones de déficit hídrico una planta nodulada, además de presentar mayor número de DEGs, estaría presentando mayor cantidad de DEGs relativos con regulación traduccional en comparación con una planta no nodulada.

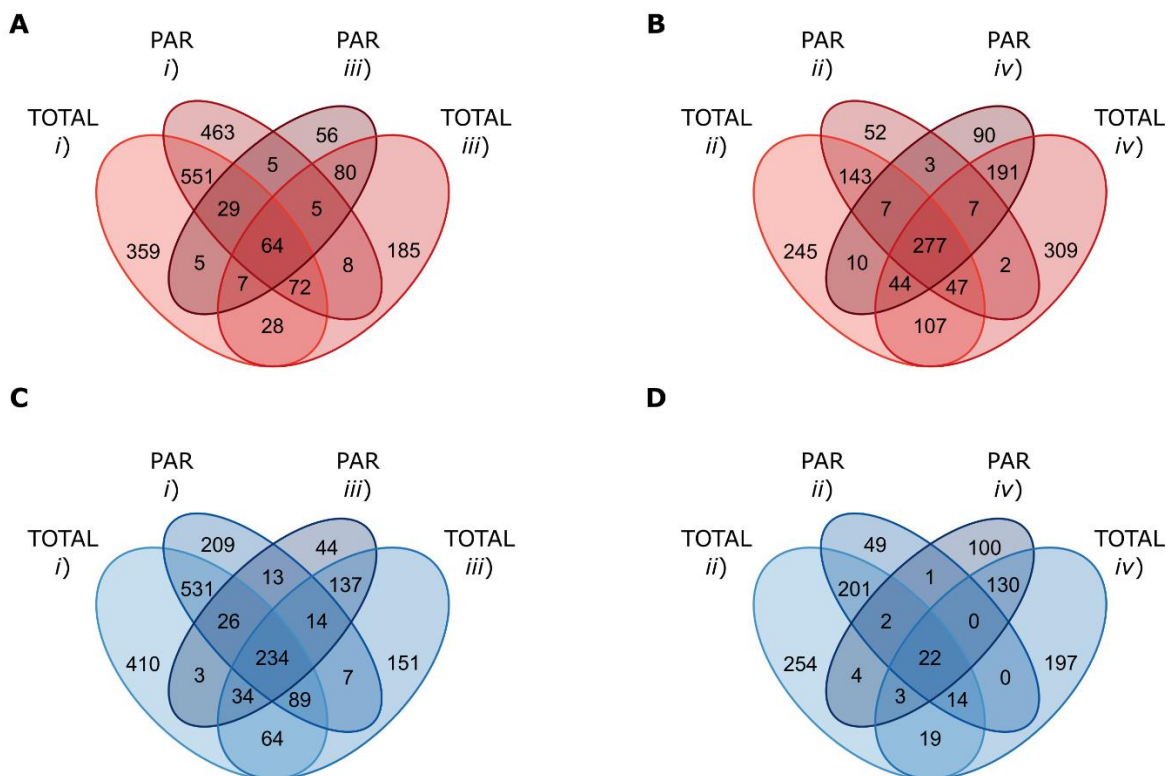


Figura 14 Diagramas de Venn intersecando listas de DEGs en los diferentes contrastes de tratamientos combinados. **A y C** Intersección de listas de DEGs a nivel transcripcional (TOTAL) y traduccional (PAR) entre los contrastes *i*) N+RH vs N+CC y *iii*) NN+RH vs NN+CC, para DEGs *up*- y *down*-regulados respectivamente. **B y D** Intersección de listas de DEGs a nivel transcripcional (TOTAL) y traduccional (PAR) entre los contrastes *ii*) N+RH vs NN+RH y *iv*) N+CC vs NN+CC, para DEGs *up*- y *down*-regulados respectivamente.

Por otro lado, se repitió el proceso pero entre los contrastes *ii* (N+RH vs NN+RH) y *iv* (N+CC vs NN+CC). El contraste *ii* evidencia la respuesta al déficit hídrico de plantas noduladas respecto a plantas no noduladas. Mientras que el contraste *iv* evidencia la respuesta de las plantas al proceso de nodulación sin involucrar restricción hídrica. De esta manera se logró discriminar entre genes que fueron DEGs comunes debido a la nodulación (N+RH y N+CC), y genes que fueron DEGs debido a cada tratamiento combinado (N+RH, por un lado, y N+CC por otro) (Figura 14B y D). Por ejemplo, se obtuvieron 52 y 90 DEGs *up*-regulados a nivel traduccional en el contraste *ii* y *iv*, respectivamente, y 3 DEGs compartidos entre ambos contrastes, también con regulación traduccional (Figura 14B y D). El contraste *ii* (N+RH vs NN+RH) y el contraste *iv* (N+CC vs NN+CC) presentaron un 11% y un 19% de genes con regulación traduccional preferencial, independientemente del estado, relativos a la cantidad total exclusiva de cada contraste, respectivamente (Figura 14B y D), siendo dichos porcentajes menores que en el contraste *i* y mayores que en el *ii*.

3.6 Análisis de Red de Co-expresión Génica Ponderada (WGCNA)

A los efectos de detectar módulos de genes y genes candidatos en las respuestas a los diferentes tratamientos, particularmente en las respuestas de plantas noduladas y en déficit hídrico, se construyó una red de co-expresión génica. Para realizar el WGCNA se utilizaron las 24 muestras correspondientes a las réplicas de los cuatro tratamientos combinados y las dos fracciones de ARNm de cada uno. Se utilizaron los conteos de un total de 26356 genes en el análisis, los cuales resultaron de retener aquellos genes que presentaron al menos 50 conteos totales en un mínimo de 12 muestras. El valor de potencia utilizado en la función de adyacencia para la construcción de la red de co-expresión génica ponderada fue de nueve, ya que dicho valor satisface el criterio de topología libre de escala, lo que significa que el valor de R^2 del modelo de topología ajustado con dicha potencia superó el umbral de 0,80 (Figura S2). Esto resultó en la construcción de 33 módulos de co-expresión génica más un módulo restante (módulo 0) que agrupa los genes que no pudieron clasificarse dentro de ninguno de los módulos restantes (Figura 15). El número de genes contenidos en cada módulo estuvo entre 30 y 5494, siendo el módulo 1 y el módulo 33 los módulos con mayor y menor cantidad de genes, respectivamente (Figura 15). Los módulos fueron posteriormente agrupados en *clusters* de acuerdo al perfil de expresión de sus *eigengenes* (ME), resultando en un total de nueve *clusters* definidos en función del índice *Silhouette* (Figura S3).

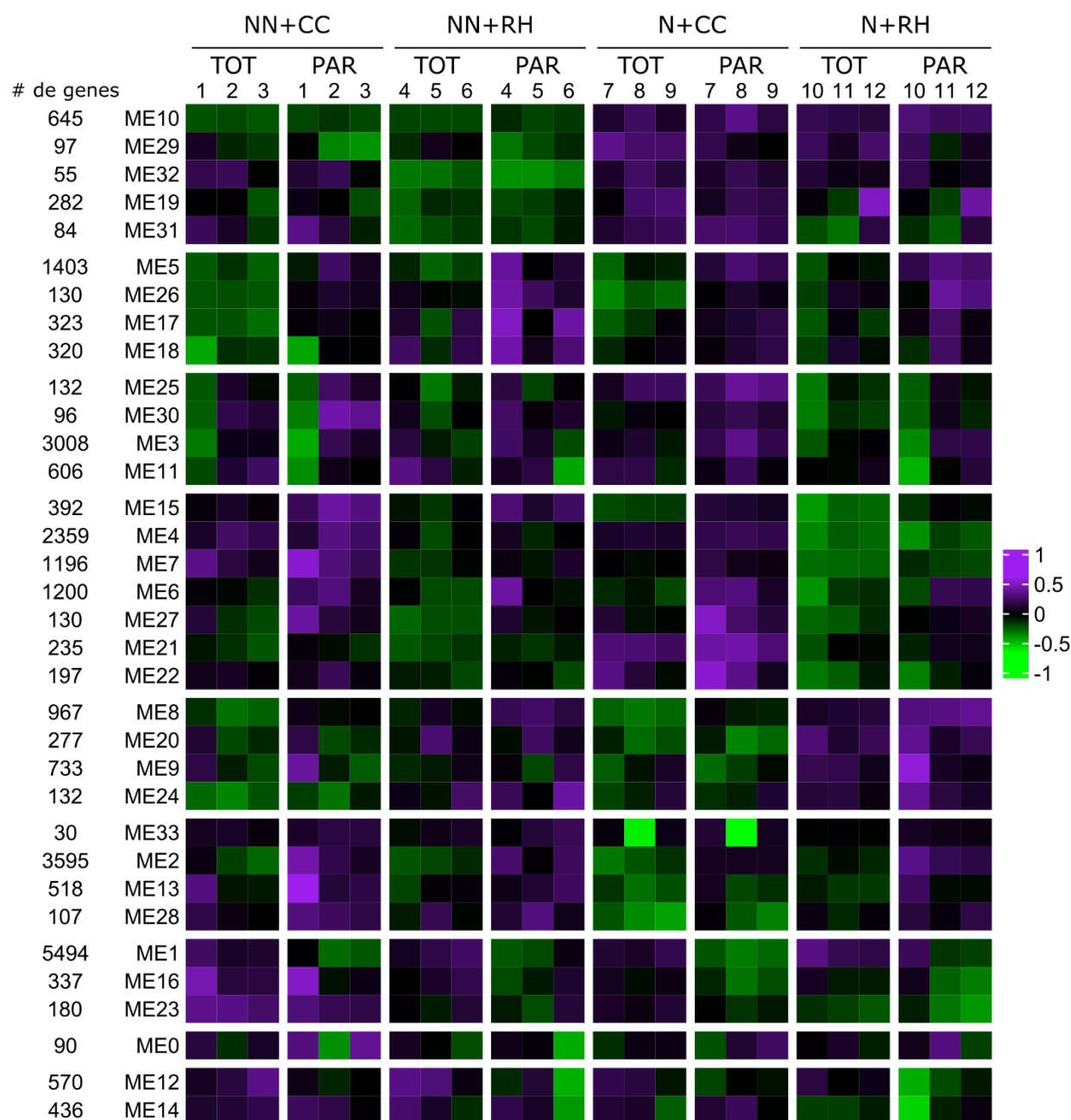


Figura 15 Red de co-expresión génica, a nivel de ARNm total y asociado a polisomas, de plantas de soja sujetas a cuatro tratamientos combinando condiciones de nodulación e hídricas. Heatmap clusterizado del perfil de expresión del *eigengene* de cada módulo (ME) en los tratamientos mencionados (NN+C, NN+E, N+C, N+E) y fracciones de ARNm total (TOT) y asociado a polisomas (PAR). La escala representa la expresión génica promedio normalizada de cada ME. Se muestra el número de cada réplica biológica y el número total de genes en cada módulo.

El módulo 0 (M0) conformó un único *cluster* ya que el perfil de expresión de su *eigengene* (ME0) no agrupó con ningún otro módulo. El *cluster* compuesto por los ME10, ME29, ME32, ME19 y ME31 fue conformado de acuerdo a la expresión de sus genes en función del contexto de nodulación. Esto sucedió ya que la expresión de dichos ME fue mayor en los tratamientos combinados N+CC y N+RH en comparación con los tratamientos de no nodulación, y puede ser visualizado por un color más violeta en

la (Figura 15). A su vez, la expresión promedio de los genes ubicados en dichos genes fue menor en el tratamiento NN+RH (expresión más verde) (Figura 15). Dentro de dicho *cluster* se encontró el ME32, el cual fue también el ME que tuvo la mayor correlación negativa con el tratamiento NN+RH (Figura 16). Es decir que genes ubicados en el módulo 32 presentaron una expresión más baja en el tratamiento de plantas no noduladas sometidas a déficit hídrico.

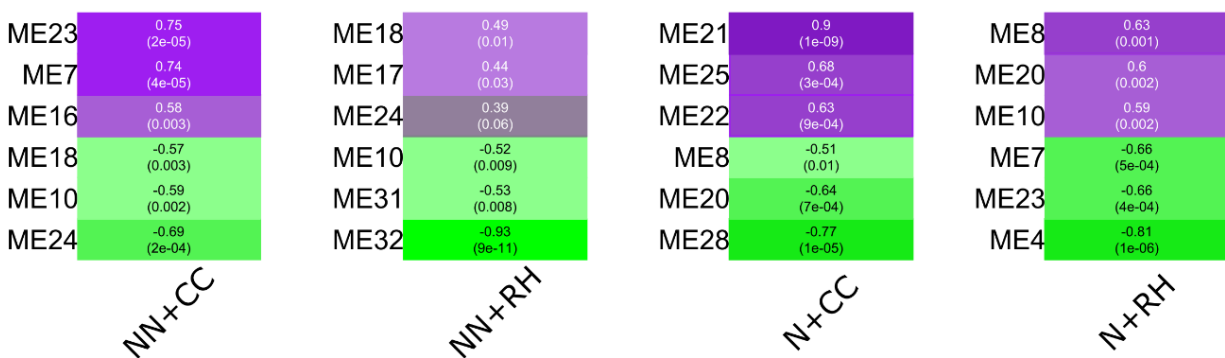


Figura 16 Correlación de Pearson entre *eigengenes* (ME) y los diferentes tratamientos combinados. Se muestran las seis correlaciones principales para cada tratamiento combinado. En tonos violáceos las correlaciones positivas y en verdes las negativas. Entre paréntesis se muestra el p-valor de cada correlación.

Un tercer *cluster* quedó conformado por los ME5, ME26, ME17 y ME18, en el cual se observaron mayores valores de expresión promedio en las muestras de fracciones de ARNm asociado a polisomas en todos los tratamientos, viéndose la mayor expresión en plantas no noduladas en restricción hídrica (NN+RH) (Figura 15). Por su parte, un cuarto *cluster* compuesto por los ME15, ME4, ME7, ME6, ME27, ME21 y ME22 puede ser visto como un *cluster* con genes de menor expresión génica promedio en la condición de restricción hídrica, ya que para los tratamientos NN+RH y N+RH se observaron tonos más verdes en el *heatmap* (Figura 15). A su vez, de dicho *cluster* se pudo interpretar que los ME15, ME4 y ME7 presentaron una mayor expresión en plantas no noduladas a capacidad de campo (NN+CC). Asimismo, el ME7 correlacionó positivamente con el tratamiento combinado NN+CC (Figura 16). La expresión de los ME15, ME4 y ME7 disminuyó en el resto de los tres tratamientos combinados, siendo mayor dicha disminución en la restricción hídrica de plantas noduladas (N+RH) (Figura 15). Esto quedó confirmado en la Figura 16, donde se observó que los ME4 y ME7 correlacionaron negativamente con N+RH.

De forma opuesta al anterior *cluster*, un quinto *cluster* compuesto por los ME8, ME20, ME9 y ME24 incluyó genes preferencialmente expresados en condiciones de restricción hídrica (NN+RH y N+RH), siendo mayor la expresión promedio en N+RH (Figura 15). Los ME20 y ME8 presentaron correlaciones positivas con N+RH (Figura 16). Finalmente, el *cluster* comprendido por los ME33, ME2, ME13 y ME28 presentó expresiones más bajas en el tratamiento de plantas noduladas y a capacidad de campo (N+CC) (Figura 15), en donde el ME28 correlacionó negativamente con el tratamiento N+CC (Figura 16).

Además de la representación de los ME (perfil de expresión promedio) es útil visualizar la interconexión de la co-expresión mediante una visualización de redes. Para ello se realizó la representación topológica de la red construida mediante WGCNA, la cual está proyectada en la Figura 17.

A excepción de genes del módulo 28, genes de todos los módulos presentaron un valor de solapamiento topológico mayor a 0,15, umbral utilizado para poder graficar la red, por lo que dichos módulos se visualizan en la red. De esta manera, la red se conformó por 12848 nodos, que representan genes, y 3951052 aristas, que representan la conexión entre genes. Las aristas no fueron graficadas debido a limitaciones computacionales.

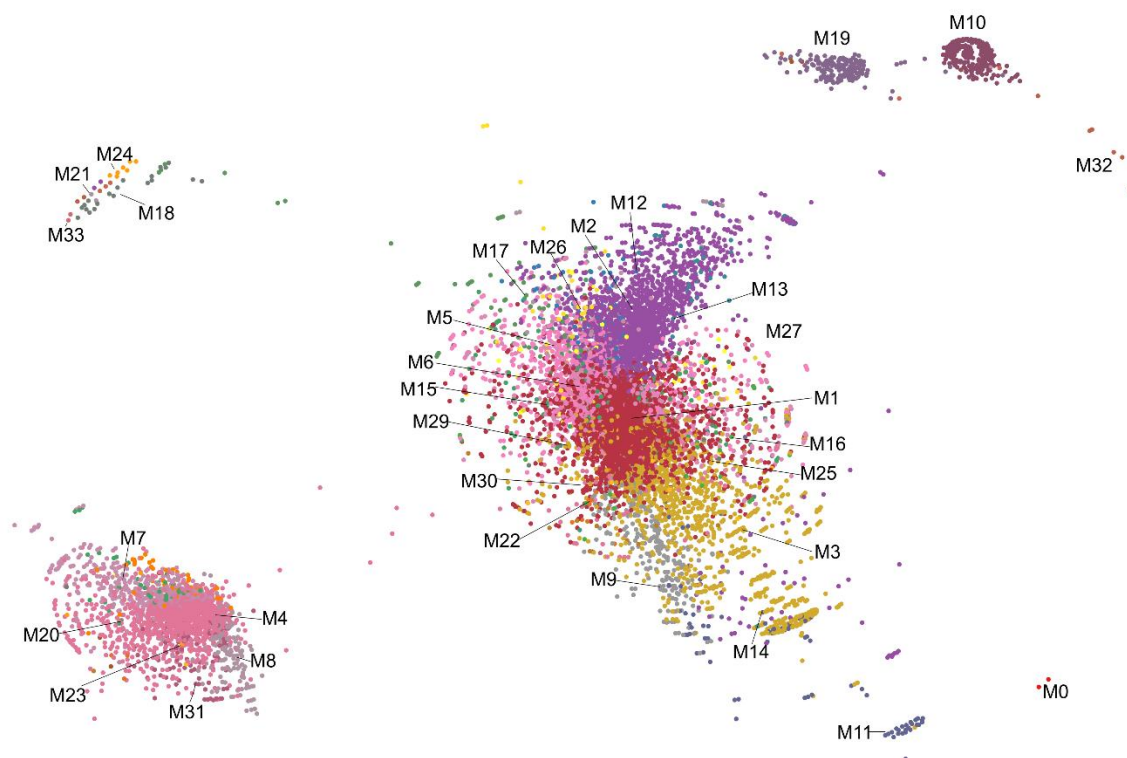


Figura 17 Red de co-expresión génica ponderada (WGCNA). Se proyecta la red de solapamiento topológico obtenida mediante WGCNA coloreando genes en función del módulo de expresión al que pertenecen (M). De los 34 módulos obtenidos, el M28 no se visualiza debido a un bajo valor de solapamiento topológico (<0,15) de los genes que lo integran. Cada nodo representa un gen, existiendo un total de 12848 nodos.

3.7 Análisis de WGCNA+DEGs

A los efectos de analizar las respuestas diferenciales de plantas de soja en los diferentes tratamientos se cruzó la información obtenida en el WGCNA con los DEGs de cada contraste y estado de regulación (*up*- o *down*- regulado). En primer lugar, se localizaron los DEGs comunes a restricción hídrica (contrastes *i* y *iii*, N+RH vs N+CC y NN+RH vs NN+CC) y los DEGs comunes a nodulación (contrastes *ii* y *iv*, N+RH vs NN+RH y N+CC vs NN+CC) en los módulos de la red de co-expresión (Tabla S1). En el caso de DEGs comunes a RH, se ubicaron mayoritariamente en los módulos M4, M7, M8 y M23, por lo que estos módulos podrían estar asociados a respuestas a estrés, y en el caso de DEGs comunes a nodulación se ubicaron prácticamente exclusivamente en el módulo M10 (Tabla 4).

Tabla 4 Cantidad de genes expresados diferencialmente a nivel transcriptómico (TOTAL) y traductómico (PAR) comunes a déficit hídrico -contrastes *i)* y *iii)*- y a nodulación -contrastes *ii)* y *iv)*- de acuerdo a su localización en módulos de co-expresión génica (M) de la red de WGCNA.

Contrastes	Estado	Nivel de regulación	M4	M7	M8	M10	M23
<i>i)</i> N+RH vs N+CC y <i>iii)</i> NN+RH vs NN+CC	Down	TOTAL	17	7	2	0	8 (5)
		PAR	5 (1)	3	1	0	1
		TOTAL+PAR	144 (84)	50 (21)	4 (2)	0	0
	Up	TOTAL	6	0	0	0	0
		PAR	1	0	1	0	0
		TOTAL+PAR	34 (14)	1	16 (7)	0	0
<i>ii)</i> N+RH vs NN+RH y <i>iv)</i> N+CC vs NN+CC	Down	TOTAL	2 (1)	2	0	2	1 (1)
		PAR	0	0	0	1	0
		TOTAL+PAR	4 (4)	9 (4)	0	5	0
	Up	TOTAL	2	1	1	6 (1)	3
		PAR	0	0	0	1	0
		TOTAL+PAR	1	0	0	113 (63)	0

Se muestra si los genes fueron *up*-regulados (Up) o *down*-regulados (Down). Entre paréntesis se muestra el número de genes diferencialmente expresados que además fueron *hub* del módulo correspondiente. Genes *hub* fueron definidos como el 10% de los genes con mayor conectividad intramodular.

Por otro lado, se localizaron los DEGs exclusivos de cada contraste en los módulos de co-expresión y fueron proyectados en la red de co-expresión (Figura 18). La mayor cantidad de DEGs del contraste *i* (N+RH vs N+CC), que involucró la respuesta a la restricción hídrica de plantas noduladas, se localizaron en los módulos M1, M4, M7, M8 y M20 (Figura 18; Tabla 5; Tabla S2). En el caso del contraste *ii* (N+RH vs NN+RH), aquel que involucró las respuestas de plantas frente a restricción hídrica en diferentes contextos de nodulación, los DEGs se ubicaron en su mayoría en los módulos M4, M15, M19 y M32 (Figura 18; Tabla 5; Tabla S2). Los DEGs obtenidos en el contraste entre plantas no noduladas en diferente condición hídrica (contraste *iii*, NN+RH vs NN+CC) se localizaron preferencialmente en los módulos M7, M17, M18 y M24 (Figura 18; Tabla 5; Tabla S2). Por su parte, los DEGs del contraste que permitió evidenciar respuestas al proceso de nodulación en plantas sin restricción hídrica (contraste *iv*, N+CC vs NN+CC) se localizaron en los módulos M7, M10, M21 y M28 (Figura 18; Tabla 5; Tabla S2).

Tabla 5 Módulos (M) de la red de WGCNA con mayor cantidad de genes diferencialmente expresados en los contrastes biológicos analizados.

Contraste	Estado	Nivel de regulación	M1	M4	M7	M8	M19	M20
i) N+RH vs N+CC	Down	TOTAL	8	60 (2)	21 (1)	19	0	8 (1)
		PAR	2	70 (4)	19 (6)	11	0	8
		TOTAL+PAR	3	240 (57)	43 (19)	20 (3)	0	21 (1)
	Up	TOTAL	4	49	2	33 (5)	0	6 (1)
		PAR	228 (99)	63 (1)	0	22 (5)	0	18
		TOTAL+PAR	31 (1)	158 (20)	3	120 (42)	0	51 (23)
ii) N+RH vs NN+RH	Down	TOTAL	5	82 (31)	16 (2)	6 (1)	0	1
		PAR	2	14 (5)	6 (1)	0	0	0
		TOTAL+PAR	1	126 (61)	35 (8)	2 (1)	0	0
	Up	TOTAL	14 (1)	28 (4)	2 (1)	6	27 (11)	1 (1)
		PAR	2	14 (3)	1	3 (1)	7 (1)	0
		TOTAL+PAR	4	31 (9)	4	7 (1)	11 (4)	1
iii) NN+RH vs NN+CC	Down	TOTAL	4	9	25 (2)	4	0	2
		PAR	0	4	15 (3)	2	1	0
		TOTAL+PAR	0	8	51 (10)	0	0	0
	Up	TOTAL	3	6	7	3	0	0
		PAR	6	2	1	7	0	2
		TOTAL+PAR	1	2	3	2	0	0
iv) N+CC vs NN+CC	Down	TOTAL	5	7	12 (4)	6 (1)	0	3 (1)
		PAR	4	0	17 (1)	1	0	2
		TOTAL+PAR	3	4	28 (10)	4 (1)	0	2
	Up	TOTAL	6	6	5	2	8	1
		PAR	1	1	2	0	5	0
		TOTAL+PAR	1	0	7	0	2	0

Se muestra si los genes fueron *up*-regulados (Up) o *down*-regulados (Down). Entre paréntesis se muestra el número de genes diferencialmente expresados que además fueron *hub* del módulo correspondiente. Genes *hub* fueron definidos como el 10% de los genes con mayor conectividad intramodular.

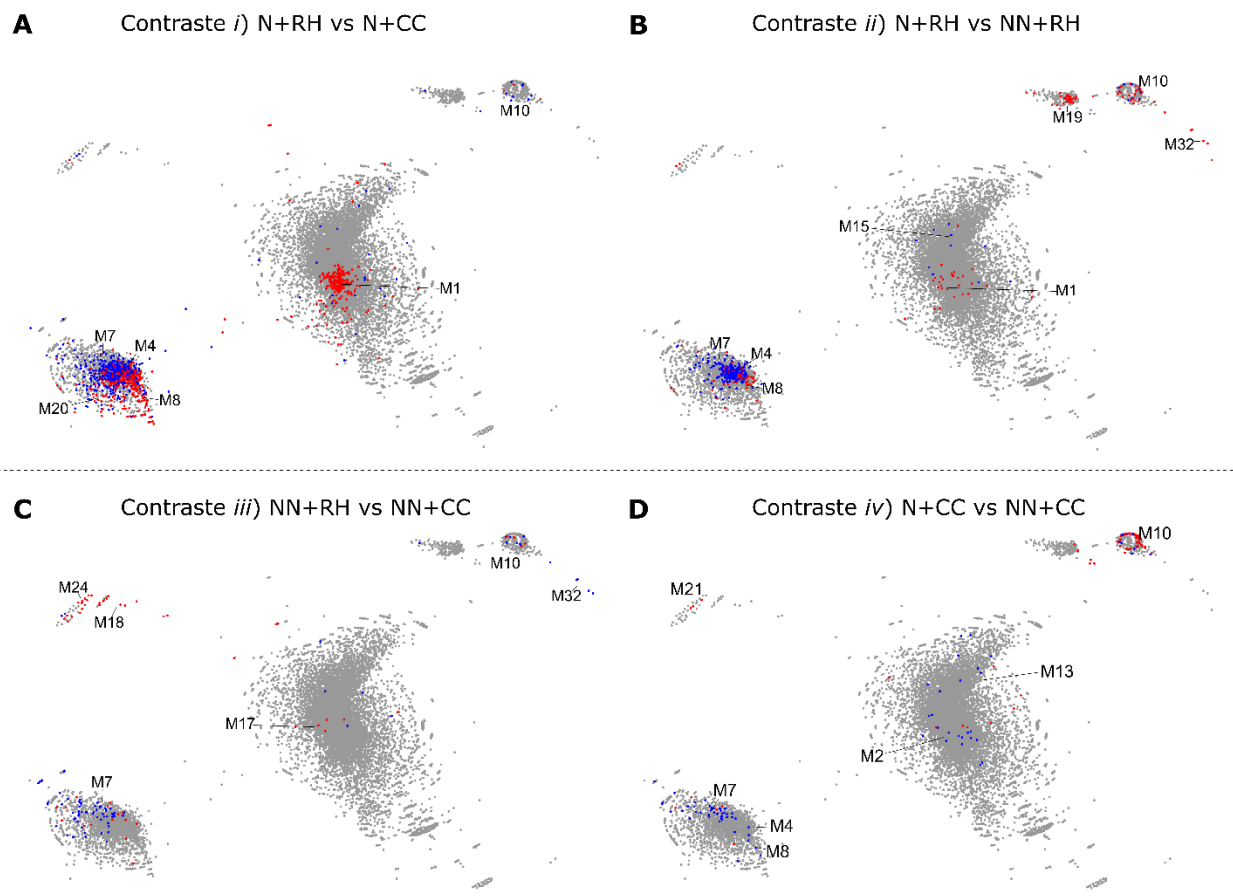


Figura 18 Localización de DEGs, a nivel de los cuatro contrastes estudiados y en ambas fracciones de ARNm (TOTAL y/o PAR), dentro de la red de co-expresión génica ponderada. En rojo, genes *up*-regulados. En azul, genes *down*-regulados. **A** Contraste i) N+RH vs N+CC. **B** Contraste ii) N+RH vs NN+RH. **C** Contraste iii) NN+RH vs NN+CC. **D** Contraste iv) N+CC vs NN+CC.

Interesantemente, se encontraron dos módulos que presentaron genes que en su mayoría revirtieron su estado de regulación en diferentes contrastes. El M21 es uno de ellos y presentó genes que se *up*-regulaban en una planta nodulada frente una no nodulada cuando no experimentaron restricciones hídricas (contraste iv, N+CC vs NN+CC), pero que se reprimieron cuando la planta nodulada experimentó restricción hídrica (contraste ii, N+RH vs N+CC) independientemente del nivel de regulación (Tabla 6). Asimismo, el M32 presentó genes que se reprimieron cuando plantas no noduladas se sometieron a restricción hídrica (contraste iii, NN+RH vs NN+CC), pero que se indujeron cuando la restricción hídrica fue analizada en diferentes contextos de nodulación (contraste ii, N+RH vs NN+RH). En este caso, la reversión de estado de estos DEGs sucedió a nivel de regulación mixta (TOTAL+PAR) (Tabla 6).

Tabla 6 Módulos (M) de la red de WGCNA que presentaron genes diferencialmente expresados que revierten su estado (*up*- o *down*-regulado) entre alguno de los contrastes biológicos analizados.

Contraste	Estado	Nivel de regulación	M21	M32
i) N+RH vs N+CC	Down	TOTAL	15 (1)	0
		PAR	13 (2)	0
		TOTAL+PAR	39 (18)	0
	Up	TOTAL	0	0
		PAR	3	1
		TOTAL+PAR	1	0
ii) N+RH vs NN+RH	Down	TOTAL	0	0
		PAR	0	0
		TOTAL+PAR	0	1
	Up	TOTAL	0	1
		PAR	0	1
		TOTAL+PAR	0	20 (6)
iii) NN+RH vs NN+CC	Down	TOTAL	0	2
		PAR	0	0
		TOTAL+PAR	0	17 (6)
	Up	TOTAL	0	1
		PAR	0	0
		TOTAL+PAR	0	1
iv) N+CC vs NN+CC	Down	TOTAL	4	0
		PAR	3	0
		TOTAL+PAR	1	0
	Up	TOTAL	22 (1)	0
		PAR	8 (2)	0
		TOTAL+PAR	53 (19)	0

Se muestra si los genes fueron *up*-regulados (Up) o *down*-regulados (Down). Entre paréntesis se muestra el número de genes diferencialmente expresados que además fueron *hub* del módulo correspondiente. Genes *hub* fueron definidos como el 10% de los genes con mayor conectividad intramodular.

En función de los módulos mencionados que presentaron mayor cantidad de DEGs en cada contraste, se seleccionaron los módulos M4, M7, M8, M10 y M23 para análisis de DEGs comunes a restricción hídrica y comunes a nodulación (Tabla 4) y los módulos M1, M4, M7, M8, M20, M21 y M32 para posteriores análisis de DEGs provenientes de cada contraste (Tabla 5; Tabla 6). En ambas situaciones se discriminó el nivel de regulación (transcripcional, traduccional o mixta) de los DEGs. De esta manera, los DEGs comunes en restricción hídrica presentaron regulación mixta preferencial, en donde los genes *up*-regulados se localizaron en los módulos M4 y M8, con 34 y 16 DEGs *up*-regulados, respectivamente (Tabla 4). Mientras que los genes *down*-regulados comunes en restricción hídrica con regulación mixta se localizaron en los módulos M4 y M7, con 144 y 50 DEGs *down*-regulados, respectivamente (Tabla 4). Respecto a DEGs comunes a nodulación, también presentaron regulación mixta preferencial, casi en su totalidad *up*-regulados y fueron co-localizados en el M10, con un total de 133 genes con dichas características (Tabla 4).

Por otro lado, con respecto a DEGs obtenidos para cada contraste biológico, es de destacar varios aspectos. En primer lugar, que en el M7 se localizaron en su mayoría genes que fueron *down*-regulados en todos los contrastes (Tabla 5). El número de DEGs *down*-regulados, independientemente del nivel de regulación, del M7 fue de 83, 57, 91 y 57 en los contrastes *i*, *ii*, *iii* y *iv*, respectivamente (Tabla 5). En segundo lugar, en el M4 se localizaron preferentemente DEGs, tanto *down*- como *up*-regulados e independientemente del nivel de regulación, provenientes de los contrastes *i* y *ii*, es decir, del efecto de la restricción hídrica en una planta nodulada, y del efecto de la restricción hídrica en diferentes contextos de nodulación (Tabla 5). El número de DEGs con dichas características fue de 640 y 295 en los contrastes *i* y *ii*, respectivamente. En tercer lugar, los DEGs del M20 provinieron casi exclusivamente del contraste *i* (N+RH vs N+CC). Dichos DEGs fueron *down*- o *up*-regulados, siendo un total de 112 genes independientemente del nivel de regulación (Tabla 5). En cuarto lugar, la mayoría de los DEGs co-localizados en el M8 correspondieron al contraste *i* y fueron *up*-regulados, siendo ellos un total de 175 genes (Tabla 5). En quinto y último lugar, el 70% de los DEGs del M1 (228 de 338) presentaron regulación traduccional exclusiva (PAR), los cuales se obtuvieron de genes *up*-regulados del contraste *i* (Tabla 5).

Por otro lado, el M21 presentó 83 DEGs *up*-regulados en plantas noduladas frente a no noduladas sin restricción hídrica (contraste *iv*, N+CC vs NN+CC) y 67 DEGs *down*-regulados cuando plantas noduladas fueron sometidas a restricción hídrica (N+RH vs N+CC) (Tabla 6). De ellos, 55 genes revirtieron su estado de regulación (*down*- o *up*-), independientemente del nivel de regulación, en los contrastes mencionados (*i* y *iv*). Con respecto al M32, se localizaron 17 DEGs que fueron *down*-regulados en el contraste *iii* (NN+RH vs NN+CC), y 20 DEGs *up*-regulados en el contraste *ii* (N+RH vs NN+RH) (Tabla 6), en ambos con una regulación mixta, de los cuales 17 revirtieron su estado de regulación.

3.8 Análisis de Enriquecimiento de Procesos Biológicos, Vías Metabólicas e Interacciones Proteína-Proteína en Módulos de Co-expresión

Se realizaron análisis de enriquecimientos de procesos biológicos mediante *Gene Ontology* (GO) en módulos de co-expresión seleccionados previamente (M10, M4, M7, M20, M8 y M1) con el objetivo de englobar las respuestas de co-expresión y comprender qué procesos están teniendo respuestas similares entre tratamientos. El M10 fue seleccionado debido a la exclusividad de DEGs en tratamientos de nodulación. El M7 lo fue debido presentar preferencia de genes que reprimen su expresión en todos los contrastes. El M4 debido a presentar DEGs preferencialmente en los contrastes *i* y *ii*. Los M20 y M8 fueron seleccionados debido a la preferencia de la localización de DEGs en el contraste *i*, y el M1 lo fue debido a la alta exclusividad de DEGs con regulación traduccional exclusiva que localizaron en él. Con los DEGs co-localizados en cada uno de los módulos se realizaron redes de interacción proteína-proteína (PPI), y posterior análisis de enriquecimiento GO y de vías KEGG cuando fue posible, es decir cuando la red presentó interacciones significativas mayores a las esperadas en un set de proteínas tomadas al azar del genoma con igual tamaño que el entregado a *String* para realizar la PPI.

3.8.1 DEGs comunes a nodulación y comunes a estrés

Debido a que los tratamientos aplicados en el estudio fueron los de nodulación y déficit hídrico, se analizó la localización de DEGs comunes a uno u otro tratamiento en módulos de co-expresión. Es así que, en el M10, módulo común a tratamientos de nodulación, los procesos biológicos de respuesta a estímulos, transporte transmembrana, regulación de compuestos con bases nucleotídicas y nodulación

fueron los procesos que presentaron mayor cantidad de conteos significativos (Figura 19A). Cuando se construyó una red de interacciones proteína-proteína (PPI) para los 113 DEGs *up*-regulados comunes a nodulación se encontró que justamente la nodulación fue el único término GO sobrerrepresentado (Figura 19B). Dicha red mostró interacción entre proteínas, las cuales fueron las leghemoglobinas A, C1, C2 y C3 y nodulinas 21, 22 y 44, todas proteínas claves en el proceso de nodulación (Figura 19B).

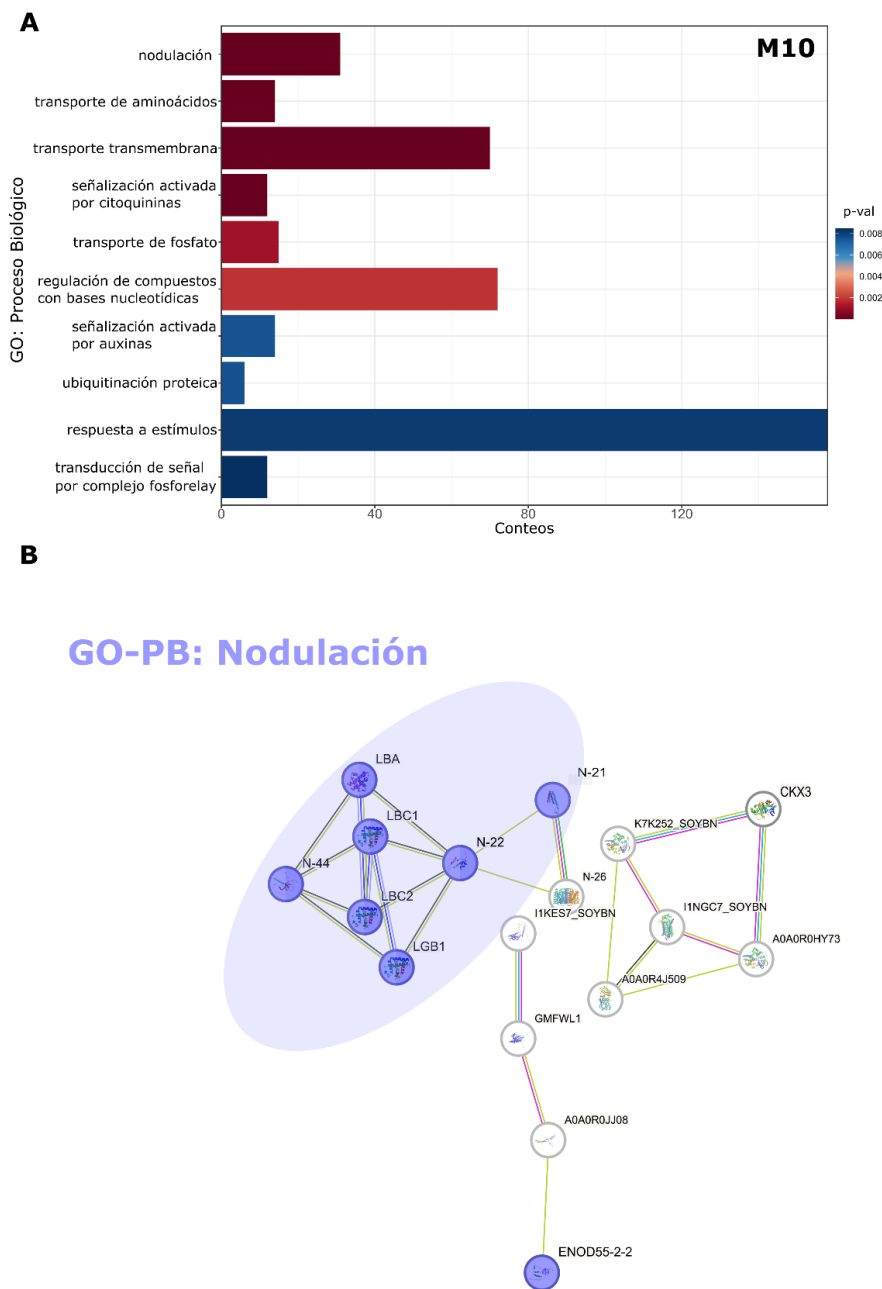


Figura 19 Análisis de enriquecimiento funcional en módulo de co-expresión número 10 (M10). **A** Enriquecimiento de Procesos Biológicos (PB) de Ontología de Genes (GO) en la totalidad de genes del M10. Se muestran los diez términos con mayor cantidad de conteos significativos (p -valor < 0.05). **B** Red de interacción proteína-proteína obtenida a partir de los genes *up*-regulados comunes a nodulación co-localizados en M10 (Tabla 4, contrastes *ii* y *iv*). La red se construyó utilizando *STRING* (Szklarczyk et al. 2021).

Por otro lado, para los DEGs comunes a estrés se realizó una red PPI utilizando los DEGs de los módulos 4, 7, 8 y 23 (Tabla 4; Figura 20). El transporte transmembrana fue un proceso sobrerrepresentado en esta red, en el cual existió interacción entre transportadores de amonio y de nitrato, todos *down*-regulados, y un transportador de cloruro *up*-regulado. A su vez, dentro de los términos sobrerrepresentados en esta red se encontró el proceso de modificación de pared celular, y vías de biosíntesis de fenilpropanoides, señalización vía MAPK y transducción de señales mediada por hormonas (Figura 20). Con respecto a las modificaciones de pared, cuatro genes que codifican para pectinesterasas y uno para poligalacturonasa fueron reprimidos en condiciones de déficit hídrico. Para la biosíntesis de fenilpropanoides, se encontró una glicosiltransferasa como nodo central que interacciona con 13 peroxidasas, todas *down*-reguladas. Asimismo, la transducción de señales mediada por ABA también presentó genes *down*-regulados, encontrándose cuatro receptores de ABA tipo PYL4, uno tipo PYL6 y una alfa-dioxigenasa. Esto sugiere que independientemente de la nodulación o no, el transporte transmembrana, la señalización mediada por ABA y las modificaciones de pared celular son procesos que experimentan represión de genes en condiciones de restricción hídrica.

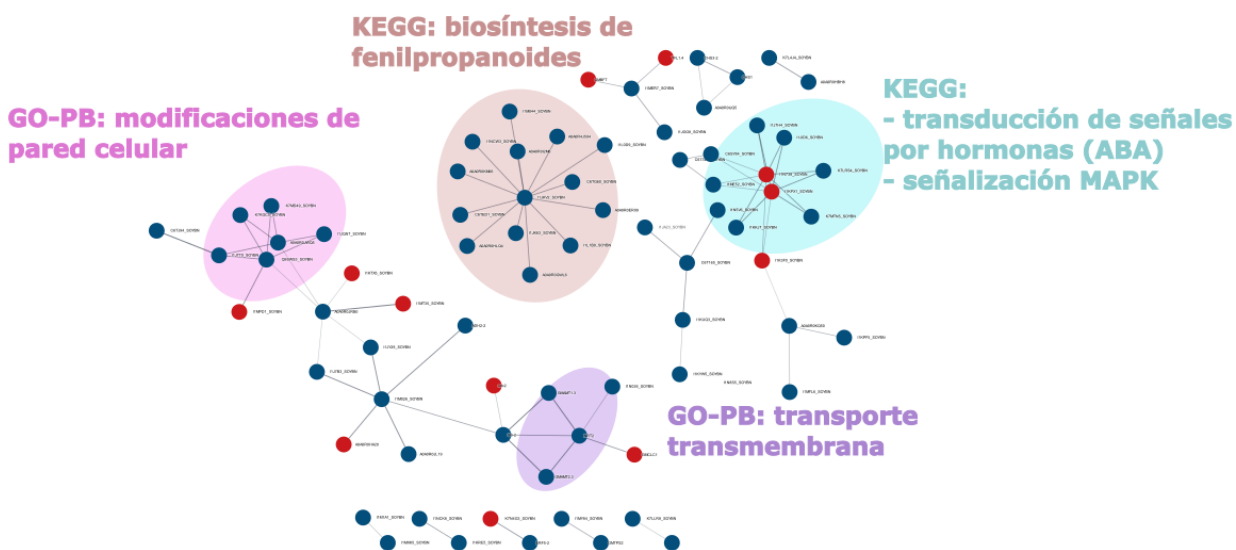


Figura 20 Análisis de enriquecimiento funcional en red de interacción proteína-proteína para el set de genes diferencialmente expresados comunes a tratamientos de restricción hídrica (N+RH vs N+CC y NN+RH vs NN+CC, contrastes *i* y *iii*, respectivamente). Se utilizaron genes que localizaron en módulos de co-expresión M4, M7, M8, M10 y M23 (Tabla 4; Tabla S2). Nodos en azul y rojo representan genes *down*- y *up*-regulados, respectivamente, independientemente de la fracción de ARNm. **GO-PB:** Proceso Biológico de Ontología de Genes. La red se construyó utilizando *Cytoscape* (Smoot et al. 2011) y *STRING* (Szklarczyk et al. 2021).

3.8.2 DEGs preferenciales de cada contraste

Gracias al análisis combinado de WGCNA y DEGs se identificó un módulo de co-expresión génica que presentó DEGs *down*-regulados en todos los contrastes analizados, tanto a nivel transcripcional y/o traduccional. Dicho módulo, el M7, puede entenderse como un módulo preferentemente de genes en los que su expresión y asociación a polisomas fue reprimida cuando existió la presencia de un estímulo sobre la planta. Esto es así ya que en cada contraste se estudia el efecto de un estímulo sobre una condición que no lo presenta. En el contraste *iv* (N+CC vs NN+CC) el estímulo es el efecto biótico de la nodulación, en los contrastes *i* (N+RH vs N+CC) y *iii* (NN+RH vs NN+CC) el estímulo es el efecto abiótico de la restricción hídrica (en contextos de nodulación diferente). Mientras que en el contraste *ii* (N+RH vs NN+RH) el estímulo diferencial es la nodulación, pues en ambos tratamientos las plantas estuvieron sometidas a déficit hídrico. Cuando se analizó el enriquecimiento funcional sobre la totalidad del módulo se encontró que estuvo enriquecido mayoritariamente en los procesos biológicos de respuesta a estrés, regulación de la transcripción, fosforilación de proteínas, ubiquitinación proteica y biosíntesis de flavonoides y de fenilpropanoides ([Figura 21A](#)).

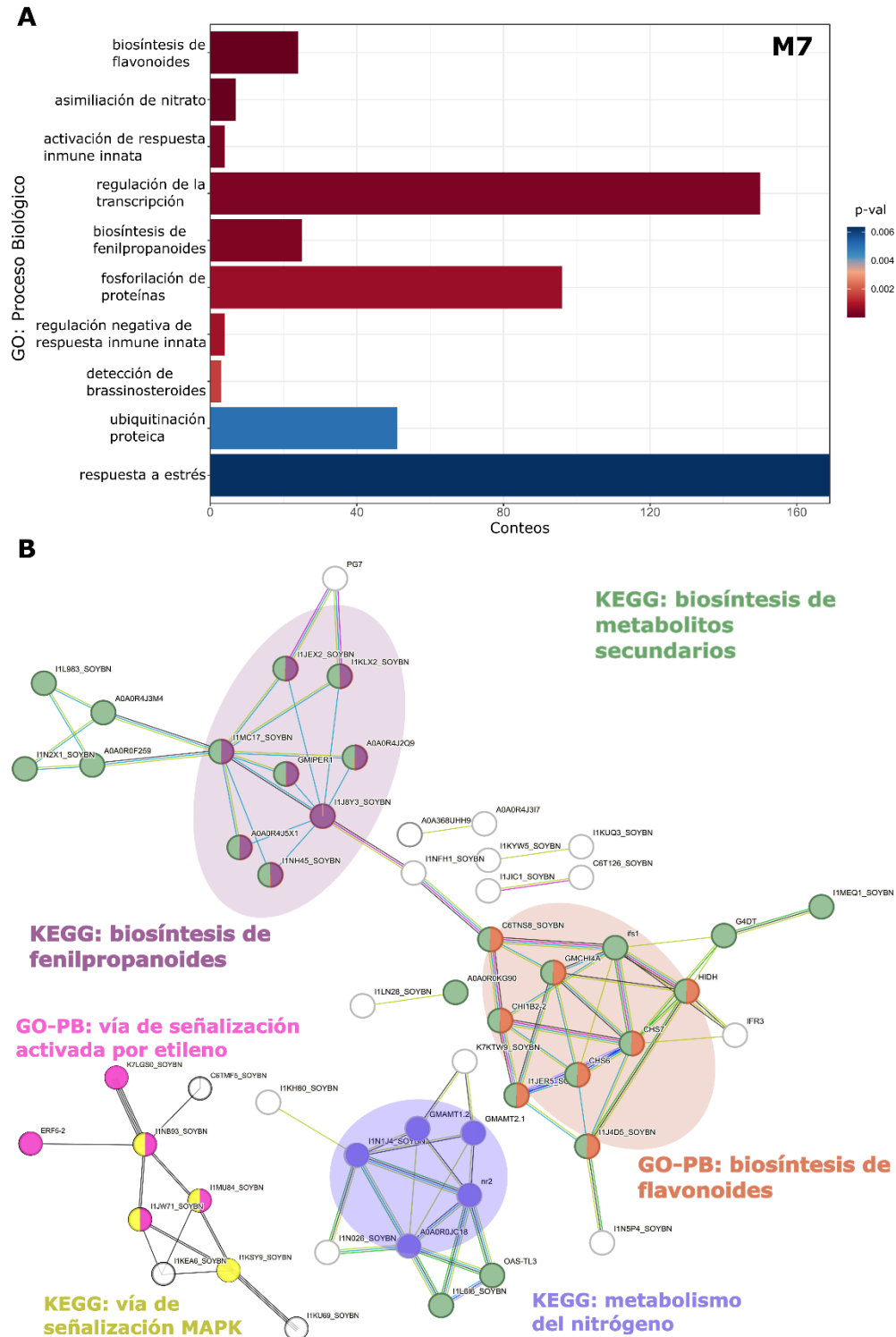


Figura 21 Análisis de enriquecimiento funcional en módulo de co-expresión número 7 (M7). A Enriquecimiento de Procesos Biológicos (PB) de Ontología de Genes (GO) en la totalidad de genes del M7. Se muestran los diez términos con mayor cantidad de conteos significativos (p -valor < 0.05). B Red de interacción proteína-proteína obtenida a partir de los genes diferencialmente expresados *down*-regulados en todos los contrastes co-localizados en el M7 (Tabla 5). La red se construyó utilizando STRING (Szklarczyk et al. 2021).

En la red PPI de todos los genes del M7 que estuvieron *down*-regulados, independientemente del nivel de regulación, en los cuatro contrastes se observó la sobrerrepresentación principalmente de la vía KEGG biosíntesis de metabolitos secundarios, dentro de la cual el proceso de biosíntesis de flavonoides y la vía de biosíntesis de fenilpropanoides se encontraron sobrerrepresentadas (Figura 21B). Con respecto a genes con rol en la biosíntesis de metabolitos secundarios se encontró una trehalosa 6-P fosfatasa, una fructosa-bisfosfato aldolasa, una glicinol 4-dimetilaliltransferasa, una Mg-protoporfirina quelatasa (cataliza la inserción de Mg en la protoporfirina) y dos cisteína sintasas. Por su parte, genes que codifican para una glicosiltransferasas, una proteína con dominio PKS-ER (funciones policétido eniolreductasa sintasa) y seis peroxidasas fueron genes *down*-regulados de la vía de biosíntesis de fenilpropanoides. Asimismo, genes de la vía de biosíntesis de flavonoides que reprimen su expresión debido a estímulo fueron tres chalcona sintasas, dos chalcona-flavonona isomerasas, una isoflavona sintasa, una epimerasa y una 2-hidroxiisoflavonona deshidratasa. Es así que disminuir la expresión de determinados genes asociados al metabolismo secundario pareciera ser una respuesta conservada en los cuatro contrastes biológicos estudiados.

Otras vías que presentaron represión de la expresión génica debida a la existencia de un estímulo y que se encontraron en la red PPI del M7 fueron la vía de señalización activada por etileno, la señalización mediada por MAPK y el metabolismo del nitrógeno (Figura 21B). Respecto al metabolismo del nitrógeno se encontraron genes que codifican para una nitrato reductasa y dos transportadores de amonio. Mientras que para las vías de señalización se encontraron *down*-regulados dos factores de transcripción, uno de respuesta a etileno (ERF5) y otro de respuesta a auxinas, una proteína con dominio *F-box*, una quinasa y una histidina quinasa. Esto sugiere una modulación conjunta de la transducción de señales por quinasas y hormonas en respuesta a estímulo.

El M4, que fue seleccionado debido a concentración de DEGs en los contrastes de interés, contrastes *i* y *ii* (N+RH vs N+CC y N+RH vs NN+RH), es decir genes que son modulados en una planta nodulada bajo déficit hídrico, estuvo altamente relacionado al proceso GO de respuesta a estímulos (más de 600 conteos significativos), seguido de los términos GO de fosforilación de proteínas y transporte transmembrana, así como los de respuesta a ABA y de señalización activada por auxinas (Figura 22A). La red PPI del set de DEGs provenientes de los contrastes *i* y *ii* presentó enriquecimiento de las vías KEGG: señalización por proteínas quinasas (MAPK), metabolismo de la galactosa, metabolismo de la arginina y prolina, biosíntesis de carotenoides y biosíntesis de fenilpropanoides (Figura 22B).

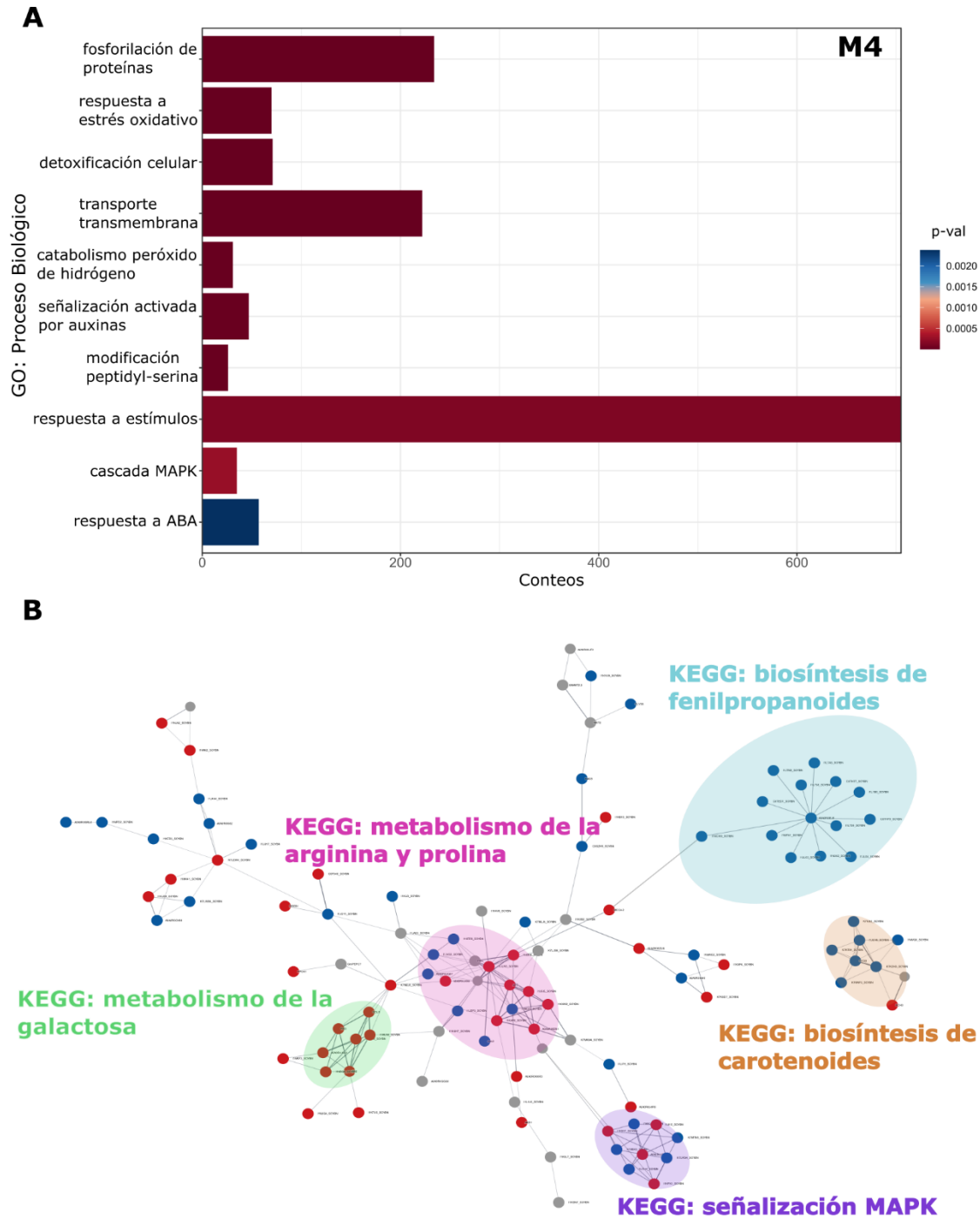


Figura 22 Análisis de enriquecimiento funcional en módulo de co-expresión número 4 (M4). **A** Enriquecimiento de Procesos Biológicos (PB) de Ontología de Genes (GO) en la totalidad de genes del M4. Se muestran los diez términos con mayor cantidad de conteos significativos (p -valor < 0.05). **B** Red de interacción proteína-proteína obtenida a partir de los genes diferencialmente expresados del contraste *i* y *ii* co-localizados en el M4 (*Tabla 5*). Nodos azules y rojos representan genes *down*- y *up*-regulados, respectivamente. La red se construyó utilizando *CytoScape* (Smoot et al. 2011) y *STRING* (Szklarczyk et al. 2021).

En los genes codificantes para proteínas vinculadas a la vía de señalización mediada por MAPK se encontraron, tanto en el contraste *i* como el *ii*, dos fosfatasa tipo PPM (fosfatasa Serina/Treonina ligadas a un manganeso o magnesio en el sitio activo) *up*-reguladas (Figura 22B). Además, se encontró que un gen codificante para una peroxidasa que participa en la señalización por MAPK fue *down*-regulado (Figura 22B). Por lo tanto, la modulación de la vía de señalización por quinasas activadas por tirosinas o citosinas (MAPK) sería de importancia en las respuestas de plantas noduladas en un contexto de déficit hídrico.

Por otro lado, los genes de esta red PPI del M4 (Figura 22B) que participan en el metabolismo de la galactosa fueron *up*-regulados en ambos contrastes (N+RH vs N+CC y vs NN+RH), encontrándose varias glicosiltransferasas y una rafinosa sintasa. Sumado a esto, el metabolismo de la arginina y prolina aparece como una vía central en esta red PPI que caracteriza la modulación de las respuestas plantas N+RH (Figura 22B). En esta vía se encontraron, por ejemplo, genes codificantes para una P5CS sintasa (enzima involucrada en la síntesis de prolina), dos amino oxidasas, dos arginina decarboxilasas y una proteína PAB5 (involucrada en la síntesis de espermidina). Dichos genes también fueron *up*-regulados en los dos contrastes que involucran plantas noduladas en déficit hídrico (contrastos *i* y *ii*). En este sentido, tanto el metabolismo de la galactosa, como el de la arginina y prolina, serían vías que se activan cuando el agua escasea en plantas noduladas.

Adicionalmente, la biosíntesis de metabolitos secundarios cobró importancia en la red de DEGs del M4, particularmente las vías de biosíntesis de fenilpropanoides y de carotenoides (Figura 22B). Los genes participantes en la biosíntesis de fenilpropanoides estuvieron todos *down*-regulados, independientemente del contraste. Dentro de ellos se encontraron 13 genes que codifican para peroxidasa y uno para una metiltransferasa (Figura 22B). Con respecto a la biosíntesis de carotenoides, una alfa/beta hidrolasa y tres enzimas citocromo P450 estuvieron *down*-reguladas y una hidroxilasa de ácidos grasos estuvo *up*-regulada.

El M20, relacionado con respuestas de plantas noduladas en restricción hídrica (contraste *i*, N+RH vs N+CC), estuvo fuertemente asociado al proceso biológico de regulación de la transcripción (Figura 23) con 43 conteos significativos. No se pudieron construir redes PPI para los DEGs provenientes del contraste mencionado. Sin embargo, análisis de enriquecimiento funcional sobre dichos DEGs revelaron que la fosforilación de proteínas y la respuesta al ABA fueron los procesos más representados en los genes *down*-y *up*-regulados, respectivamente (5 y 4 conteos, respectivamente, no se muestra figura). Los genes que participan en el proceso de fosforilación de proteínas son cuatro quinasas con dominios LRR y una quinasa dependiente de calcio/calmodulina. Con respecto a los *up*-regulados que participan en las respuestas al ABA se encontraron dos factores de transcripción de la familia NAC, uno de la familia MYB y una proteína LEA de protección contra la deshidratación.

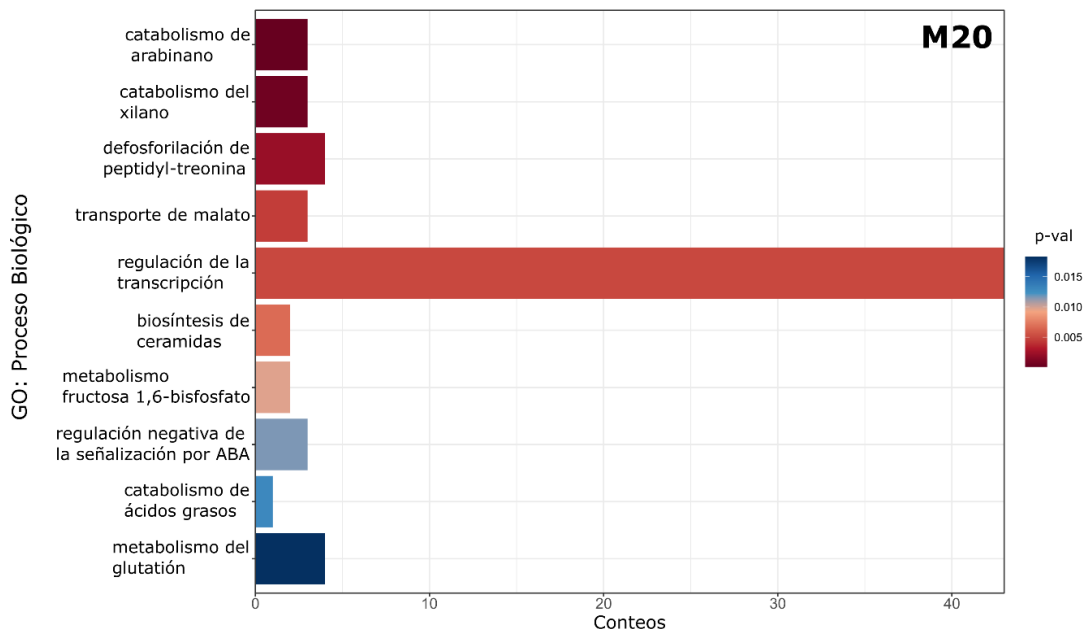


Figura 23 Análisis de enriquecimiento funcional en módulo de co-expresión número 20 (M20). Enriquecimiento de procesos biológicos de la *Gene Ontology* (GO) en la totalidad del M20, seleccionado por presentar genes diferencialmente expresados en el contraste *i* (N+RH vs N+CC) que evidencia las respuestas de plantas noduladas al déficit hídrico. Se muestran los diez términos con mayor cantidad de genes significativos ($p\text{-valor} < 0,05$).

El M8, altamente relacionado con la regulación de genes en el contraste *i* (N+RH vs N+CC) a todos los niveles de regulación, presentó enriquecimiento de los términos GO biosíntesis de sustancias orgánicas, respuesta a estrés, respuesta a estímulos endógenos, biosíntesis de compuestos antociánicos, entre otros (Figura 24A). Cuando se analizó el enriquecimiento funcional de la red PPI construida con los DEGs en el contraste *i*, se vio una sobrerrepresentación de las vías KEGG metabolismo de la galactosa y biosíntesis de metabolitos secundarios, particularmente de estilbenos y flavonoides (Figura 24B). Con respecto al metabolismo de la galactosa, se encontraron genes que codifican para dos UDP-glucosa 4-epimerasas y una glicosil hidrolasa. Mientras que para la biosíntesis de metabolitos secundarios se encontraron genes codificantes para: una chalcona/estilbeno sintetasa, tres aldo/ceto reductasas, una aldehído deshidrogenasa y una amino oxidasa (Figura 24B). Tanto los genes que participan en el metabolismo de la galactosa como en la biosíntesis de metabolitos secundarios se encontraron *up-regulados*, lo que significaría que estas vías estarían siendo favorecidas en el contraste *i* (N+RH vs N+CC).

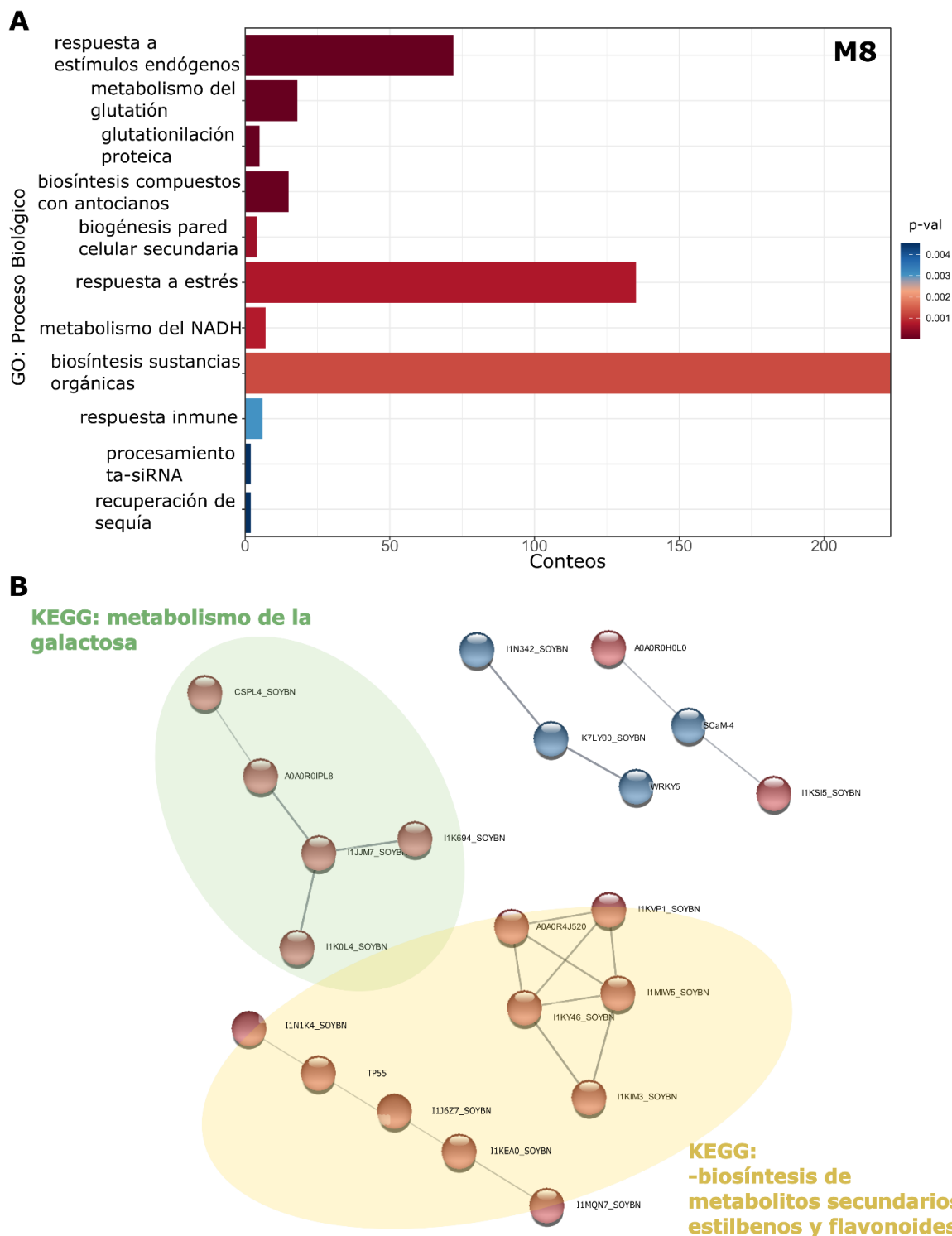


Figura 24 Análisis de enriquecimiento funcional en módulo de co-expresión número 8 (M8). **A** Enriquecimiento de Procesos Biológicos (PB) de Ontología de Genes (GO) en la totalidad de genes del M8. Se muestran los diez términos con mayor cantidad de conteos significativos (p -valor < 0.05). **B** Red de interacción proteína-proteína obtenida a partir de los genes *up*-regulados en el contraste *i* (N+RH vs N+CC) (Tabla 5). La red se construyó utilizando *CytoScape* (Smoot et al. 2011) y *STRING* (Szklarczyk et al. 2021).

Por su parte, el M1, fuertemente relacionado con la regulación traduccional exclusiva de genes en el contraste N+RH vs N+CC, se vio enriquecido en los términos GO de fosforilación y traducción, presentando más de 400 conteos en cada uno ([Figura 25A](#)). Interesantemente, la red PPI construida con los 228 genes *up*-regulados a nivel traduccional en dicho contraste *i*, presentó enriquecimiento en proteínas vinculadas a los procesos de fosforilación oxidativa, traducción y desintoxicación de especies oxidantes ([Figura 25B](#)). Por un lado, respecto a la fosforilación oxidativa, se encontraron genes que codifican para: las proteínas ribosomales mitocondriales (RP) L51 y S25, una proteína de tipo dedo de zinc CCCH, dos proteínas ACP (que participan en la síntesis de ácidos grasos) y una subunidad 6 del complejo citocromo *b*, que estaría sugiriendo una fuerte demanda de ATP en plantas N+RH mediada por un aumento en la traducción de algunos de sus componentes. A su vez, se encontraron genes que codifican para las subunidades G y F de ATPasas vacuolares (V-ATPasas) con procesos anotados de fosforilación oxidativa y de señalización mTORC1 regulada por aminoácidos. Esto significa que la vía de señalización mTORC1 junto con el metabolismo energético están siendo regulados a nivel traduccional y estarían participando en las respuestas de plantas noduladas frente condiciones de restricción hídrica.

Por otro lado, respecto a la traducción, se encontraron genes que codifican para proteínas participantes en dos vías definidas por la base de datos *Reactome*: la formación de un pool de subunidades ribosomales 40S y la iniciación de la traducción ([Figura 25B](#)). Dentro de ellas se encontraron proteínas ribosomales asociadas a subunidades mayores (RPL) y menores (RPS) como las RPL5, RPL6, RPL7, RPL28, RPL29, RPL36, RPL37, RPL38, RPS13, RPS15, RPS19, RPS21, RPS24, RPS25, entre otras, indicando que en condiciones de déficit hídrico las plantas noduladas aumentarían la síntesis (a nivel traduccional) de proteínas de la maquinaria traduccional, y por ende, favoreciendo el proceso de traducción en dichas condiciones.

A su vez, en la red se encontraron dos genes codificantes para tiorredoxinas, uno para una glutatión peroxidasa y uno para una enzima tipo superóxido dismutasa, los cuales participan en la desintoxicación de especies oxidantes, indicando que la expresión de dichas proteínas fue inducida a nivel traduccional en el contraste *i* ([Figura 25B](#)).

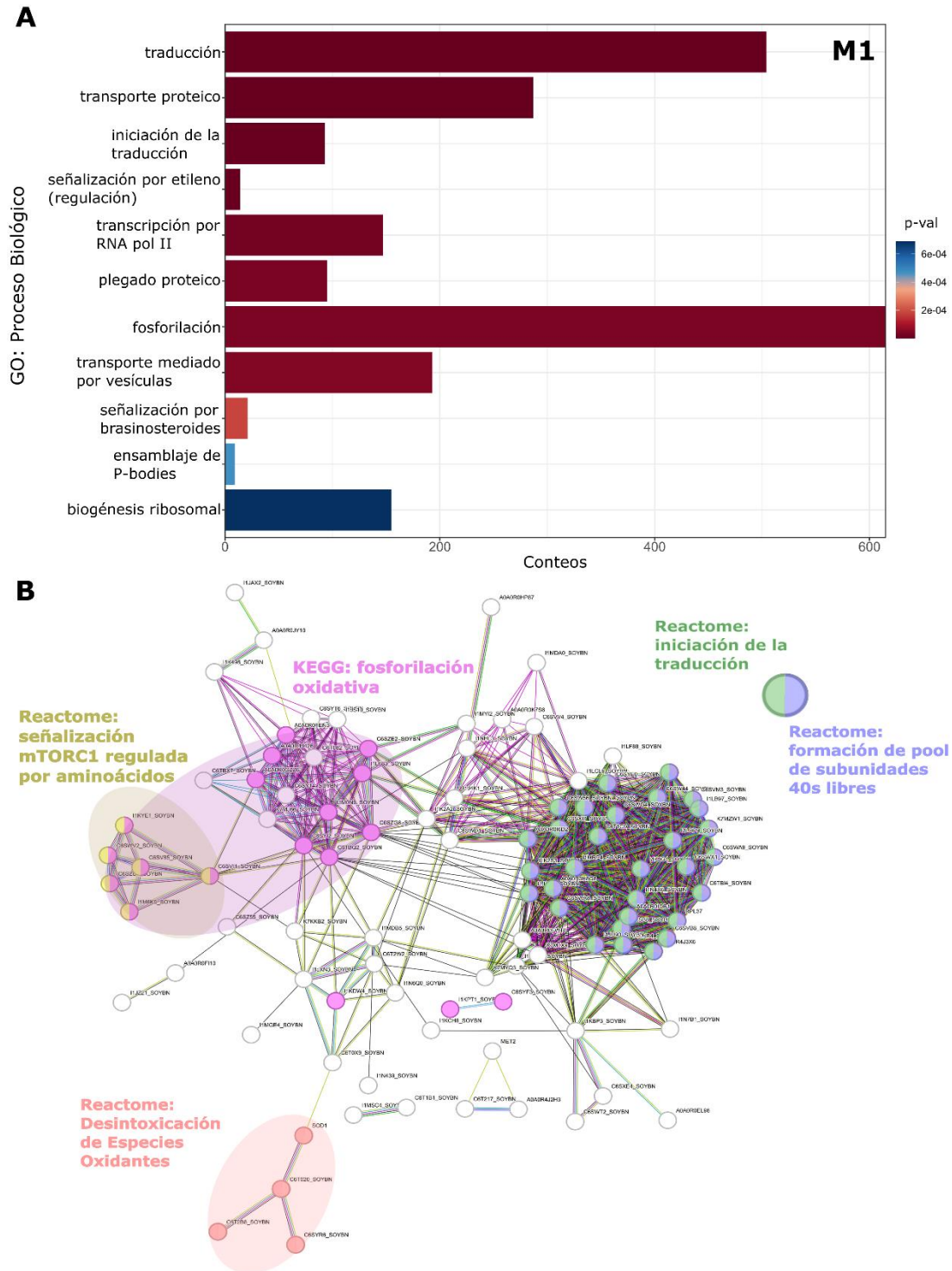


Figura 25 Análisis de enriquecimiento funcional en módulo de co-expresión número 1 (M1). A Enriquecimiento de Procesos Biológicos (PB) de Ontología de Genes (GO) en la totalidad de genes del M1. Se muestran los diez términos con mayor cantidad de conteos significativos (p -valor < 0.05). B Red de interacción proteína-proteína obtenida a partir de los genes *up*-regulados a nivel traduccional (PAR) en el contraste *i* (N+RH vs N+CC) (Tabla 5). La red se construyó utilizando *STRING* (Szklarczyk et al. 2021).

Motivos en región 5'-UTR de transcriptos de genes con regulación traduccional en contraste *i*

Adicionalmente, mediante la herramienta MEME, se analizaron los motivos que estuvieron sobrerrepresentados en la región 5'-UTR de los 228 transcriptos del M1 *up*-regulados a nivel traduccional en el contraste *i* (N+RH vs N+CC). Esto se realizó debido al potencial regulatorio de ciertos motivos frente a situaciones de estrés. Interesantemente se encontró que el 46% de los 228 transcriptos presentaron un motivo polipirimidínico (enriquecido en citosinas –C– y/o timinas –T–) THHYYYYTYCTCTYTYTYTCTYYNTY (donde N puede ser cualquiera de las cuatro bases nucleotídicas, H puede ser adenina, citosina o timina e Y pueden ser C o T) (Figura 26). El motivo 5' terminal oligopirimidínico (TOP) tiene un rol conocido en potenciar la *up*-regulación a nivel traduccional de proteínas ribosomales, factores de iniciación y elongación de la traducción, entre otras proteínas. A su vez, 21 genes de los que presentaron este motivo codifican para proteínas ribosomales. Esto estaría sugiriendo que el motivo encontrado permitiría a las células modular rápidamente la expresión de proteínas involucradas en la biogénesis ribosomal para sostener y potenciar la maquinaria traduccional de plantas noduladas sometidas a déficit hídrico.

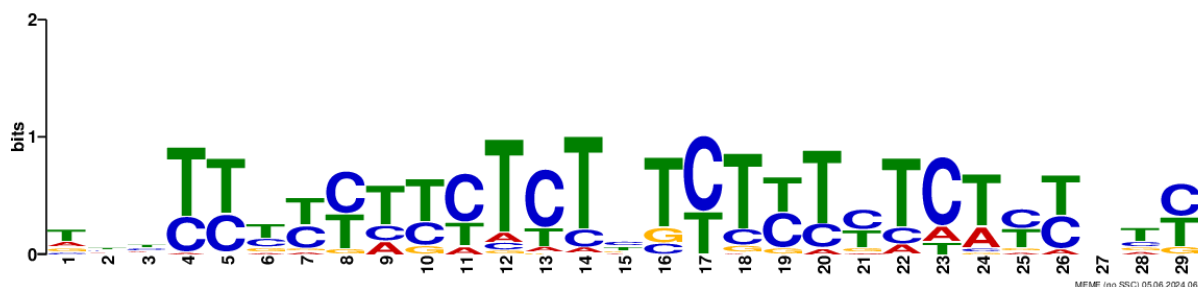


Figura 26 Motivo observado en la región 5'-UTR de transcriptos *up*-regulados a nivel traduccional en el contraste *i* (N+RH vs N+CC) y co-localizados en el módulo 1 de la red WGCNA. E-value de $2,0 \times 10^{-229}$, con 156 sitios de bases nucleotídicas contribuyendo a la generación del motivo.

Asimismo, se realizó un análisis de enriquecimiento para observar si el motivo estuvo sobrerrepresentando en genes con *up*-regulación traduccional en el contraste *i* (N+RH vs N+CC) en comparación con genes con diferente estado y nivel de regulación en dicho contraste (Tabla 7).

Se observó que, para el contraste *i*, el enriquecimiento del motivo TOP fue significativo cuando se comparó el set de genes que presentó regulación traduccional frente a los que no presentaron dicha regulación en el mismo contraste (comparación A) (Tabla 7), y cuando se compararon los genes *up*-regulados a nivel traduccional del M1 (set donde se encontró el motivo) contra los que no presentaron regulación traduccional en el contraste (comparación C) (Tabla 7). Esto sugeriría que dicho motivo efectivamente estaría cumpliendo una función en potenciar la regulación traduccional de los genes que presentan el motivo TOP en las regiones 5'-UTR de sus transcriptos. Cuando se analizó el enriquecimiento del motivo TOP en la totalidad de genes *up*-regulados a nivel traduccional en el contraste *i* frente a los *up*-regulados a nivel transcripcional en dicho contraste, no se observaron resultados significativos (comparación B) (Tabla 7). No obstante, el motivo estuvo significativamente enriquecido en el conjunto de genes del M1 *up*-regulados a nivel traduccional del contraste *i* frente a aquellos genes con regulación transcripcional (comparación D). Esto significaría que, para los 228 genes del M1 mencionados anteriormente, la regulación traduccional podría estar siendo potenciada debido a la presencia del motivo

TOP y, por ende, reforzaría la hipótesis de dicho motivo es el que estaría permitiendo modular rápidamente la biogénesis ribosomal.

Sumado a lo anterior, se analizó, para los genes *up*-regulados traduccionalmente del contraste *i*, el *fold change* promedio de los genes del M1 en comparación a los que no co-localizaron en el M1, observándose que el *fold change* de los genes del M1 fue significativamente mayor (1,40 vs 1,24, *p-value* < 0.01). Esto indicaría que, la presencia del motivo TOP en los genes del M1 estaría generando un diferencial, observado como un aumento de la expresión dado por una regulación traduccional mayor, frente a genes que presentaron un aumento de expresión dado por regulación transcripcional. Si bien dicho motivo no estaría teniendo un rol diferencial en la inducción de la expresión por regulación traduccional en plantas noduladas y sometidas a déficit hídrico en comparación a los regulados transcripcionalmente a nivel global, si estaría siendo influyente en la mayor expresión de genes codificantes para proteínas de la maquinaria traduccional (genes del M1 *up*-regulados a nivel PAR en contraste *i*).

Tabla 7 Análisis de enriquecimiento del motivo génico 5'-TOP THHHYYYTYCTCTYTYTYTCTYYNTY en el contraste *i*.

Contraste	Comparación	Enrichment Ratio	E-value
<i>i</i>) N+RH vs N+CC	A: <i>Up</i> -PAR vs <i>Not Up</i> -PAR	35,95	5,34 x 10⁻¹¹
	B: <i>Up</i> -PAR vs <i>Up</i> -TOTAL	1,04	0,28
	C: M1 <i>Up</i> -PAR vs <i>Not Up</i> -PAR	1,73	1,31 x 10⁻⁷
	D: M1 <i>Up</i> -PAR vs <i>Up</i> -TOTAL	1,44	4,19 x 10⁻³

M1: módulo de co-expresión 1. ***Not Up-PAR:*** genes que no presentaron *up*-regulación traduccional, es decir, genes estables, *down*-regulados PAR, o *down*- o *up*-regulados transcripcionalmente. Se resaltan en negritas las comparaciones significativas.

Genes que revierten el estado de regulación entre contrastes

En el M21 se dio la particularidad de que se vio enriquecido en genes que estuvieron *up*-regulados cuando se estudió el efecto de la nodulación en condiciones hídricas de capacidad de campo (contraste *iv*, N+CC vs NN+CC), y que estuvieron *down*-regulados cuando una planta nodulada se sometió a restricción hídrica (contraste *i*, N+RH vs N+CC). Dentro de estos genes se encontraron genes que codifican para proteínas vinculadas al metabolismo de hormonas como la 1-aminocyclopropano-1-carboxilato sintasa (ACC) y dos 8'-hidroxilasas del ABA y genes codificantes para factores de transcripción como el MYB124 y el MYB178. También se encontraron genes que participan en el metabolismo de pared celular como hidrolasas de xiloglucanos, endo-1,3-beta-glucosidasas de glicanos y xilosiltransferasas. Esto significaría que estos genes participantes en el metabolismo de la pared celular y de hormonas son requeridos en mayor medida, y por ende se induce la expresión, cuando en condiciones de capacidad de campo la planta se encuentra nodulada, pero que de alguna manera no estaría siendo requeridos en gran medida cuando dicha planta nodulada es sometida a déficit hídrico.

Por otro lado, en el M32 localizaron 17 genes *down*-regulados en una planta no nodulada sometida a restricción hídrica (contraste *iii*, NN+RH vs NN+CC) y que fueron *up*-regulados cuando la restricción hídrica fue analizada en diferentes contextos de nodulación (contraste *ii*, N+RH vs NN+RH). Dentro de estos genes se encontraron una peroxidasa, cuatro cisteínas oxidasas y factores de transcripción pertenecientes a las familias LOB (Glyma_13G046300 y Glyma_19G046300), bHLH (Glyma_06G092000) y MYB (Glyma_03G189600). Esto podría significar que, con respecto a los factores de transcripción, la regulación génica mediada por dichos factores no estaría siendo requerida cuando una planta sin nodular es sometida a déficit hídrico (contraste *iii*), pero que cuando el déficit hídrico se da en una planta nodulada sí se estarían utilizando en mayor medida dichos factores (contraste *ii*).

Asimismo, se encontró que casi la totalidad de los genes DEGs que se localizaron en el módulo 19 (M19) fueron obtenidos del contraste *ii*, por lo que adicionalmente se seleccionó el módulo para aportar información referida a las respuestas al déficit hídrico en contextos de nodulación diferentes, complementando así lo obtenido en el párrafo anterior a partir del M32. De esta manera, el M19 se vio enriquecido en los procesos biológicos de respuesta de defensa y regulación de la transcripción con 43 y 36 conteos, respectivamente (Figura 27), seguido en tercer lugar por el proceso de ubiquitinación proteica con 15 conteos (Figura 27). Cuando se analizaron los 55 genes *up*-regulados, independiente del nivel de regulación, del contraste *ii* que localizaron en el M19, no se encontraron evidencias de interacción proteica mediante redes PPI.

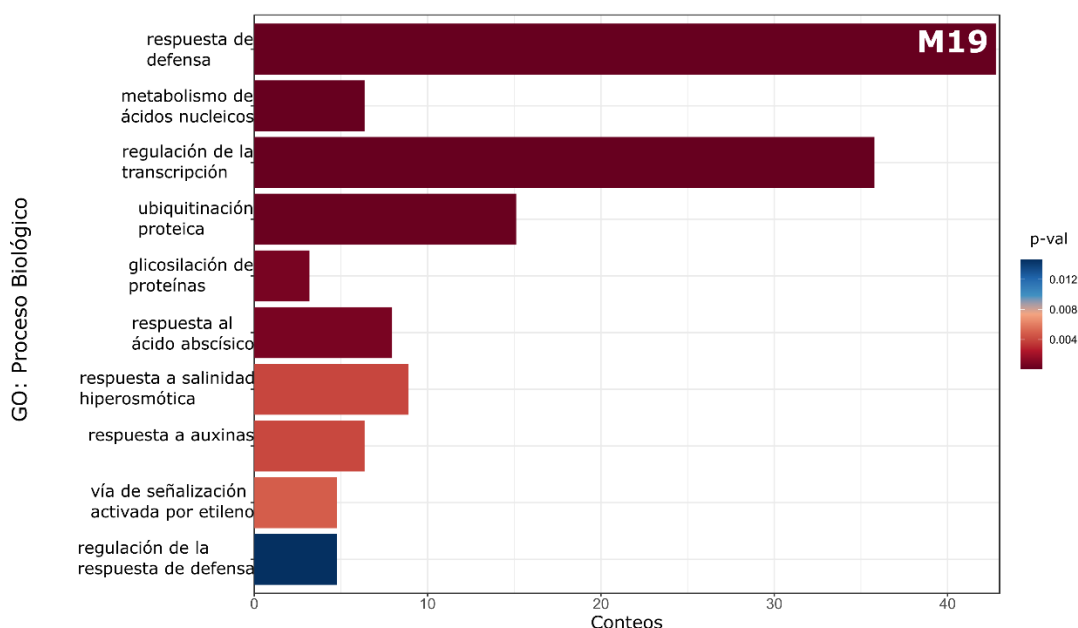


Figura 27 Análisis de enriquecimiento funcional en módulo de co-expresión número 19 (M19). Enriquecimiento de procesos biológicos de la *Gene Ontology* (GO) en la totalidad del M19, seleccionado por presentar genes diferencialmente expresados y *up*-regulados casi exclusivamente en el contraste *ii* (N+RH vs NN+RH), el cual evidencia las respuestas de plantas al déficit hídrico en diferente contexto de nodulación. Se muestran los diez términos con mayor cantidad de genes significativos (p -valor $< 0,05$).

Con respecto al proceso de respuesta de defensa se encontró que, dentro de estos 55 genes, una proteína HS1pro1 estuvo regulado a nivel combinado y que una proteína tipo BON1 lo estuvo a nivel traduccional. Asimismo, dos factores de transcripción responsivos a etileno, que participan en la respuesta

de defensa, estuvieron regulados, uno a nivel combinado (Glyma_06G290000) y otro a nivel traduccional (Glyma_01G231200). Dichos factores también participan en el proceso de regulación de transcripción, en el cual a su vez se encontró un factor de dedo de zinc tipo C2H2 y uno responsivo a giberelina e inducible por quitina, regulados a nivel combinado y traduccional, respectivamente. Los factores de transcripción mencionados son de interés para sumarse en la lista final de genes candidatos obtenidos en este trabajo debido a que estarían participando en las respuestas al déficit hídrico diferencialmente según si la planta está nodulada o no.

3.9 Análisis de Factores de Transcripción diferencialmente expresados

Los FTs están clasificados en familias, y ya que FTs de la misma familia tienden a tener roles similares en las respuestas de la planta debido a que los *targets* generalmente son genes que codifican para proteínas similares e incluso a veces tienen los mismos *targets*, se realizó un análisis exploratorio para observar qué familias de FTs predominan en las respuestas de la planta a un nivel de regulación traduccional (PAR) o combinado (TOT+PAR) en cada contraste. El contraste *i* (N+RH vs N+CC) presentó 148 FTs DEGs en dichas condiciones, siendo el contraste con mayor cantidad de FTs diferencialmente expresados, seguido de los contrastes *ii* (N+RH vs NN+RH) y *iv* (N+CC vs NN+RH) con 56 y 53 FTs DEGs, respectivamente. El contraste *iii* (NN+RH vs NN+CC) tuvo un total de 37 FTs diferencialmente expresados (Figura 28).

Se destacan en el contraste *i* las familias WRKY, NAC, MYB, ERF y bHLH con un alto número de FTs DEGs. Con respecto a los WRKY, 15 FTs fueron *down*-regulados a nivel traduccional o combinado, superando ampliamente los números en los restantes contrastes: tres y un FT WRKY fueron *down*-regulados en los contrastes *ii* (N+RH vs NN+RH) y *iii* (NN+RH vs NN+CC), respectivamente, ninguno lo fue en el contraste *iv* (N+CC vs NN+CC) (Figura 28). Con respecto a la familia NAC, predominó la *up*-regulación de los mismos, donde 9 FTs estuvieron *up*-regulados en el contraste *i* a nivel combinado o traduccional. En el resto de contrastes existieron entre 1 y 4 FTs DEGs de dicha familia a cualquier estado de regulación (Figura 28). Por otro lado, FTs de la familia MYB presentaron regulación diferencial tanto por inducción como por represión de la expresión a nivel combinado o traduccional en el contraste *i* (N+RH vs N+CC), en donde 13 y 7 FTs fueron *up*- y *down*-regulados, respectivamente (Figura 28). Asimismo, los FTs de las familias ERF y bHLH también presentaron un mayor número de DEGs tanto por *up*- como por *down*-regulación en el contraste *i* en comparación con el resto de contrastes. En el caso de los ERF, 8 FTs fueron *down*-regulados y 8 fueron *up*-regulados, y en el caso de los bHLH 10 y 8 fueron *down*- y *up*-regulados, respectivamente. Para ambas familias dichos números corresponden al contraste *i*. FTs DEGs de la familia bHLH también fueron predominantes comparando las respuestas de plantas noduladas frente a no noduladas cuando fueron sometidas a restricción hídrica (contraste *ii*, N+RH vs NN+RH), donde 8 FTs fueron *down*-regulados a nivel traduccional o combinado. En el resto de contrastes, el número de FTs DEGs (*down* o *up*-regulados) de la familia bHLH estuvo entre 1 y 3 (Figura 28).

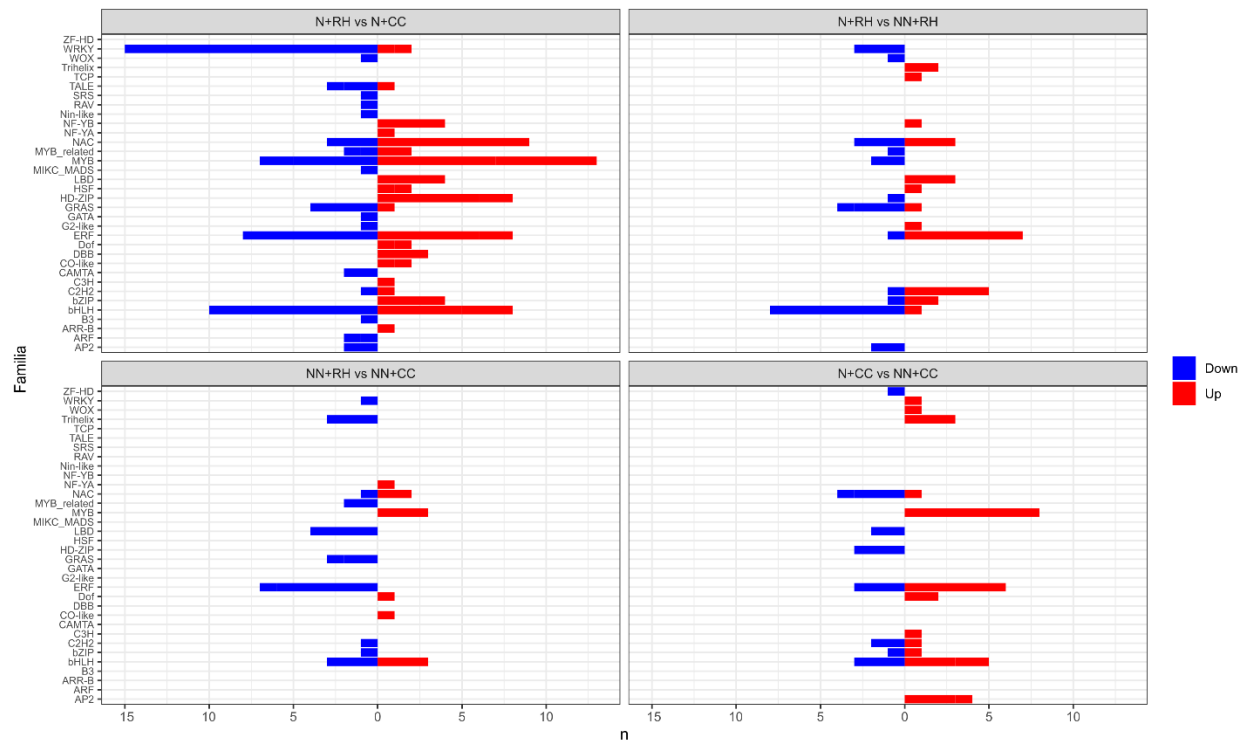


Figura 28 Cantidad de factores transcripción, agrupados por familia, diferencialmente expresados con regulación traduccional o combinada en los cuatro contrastes analizados. N+RH vs N+CC (contraste *i*), N+RH vs NN+RH (contraste *ii*), NN+RH vs NN+CC (contraste *iii*), N+CC vs NN+CC (contraste *iv*). **Down**: down-regulados. **Up**: up-regulados.

3.9.1 Factores de Transcripción diferencialmente expresados a nivel traduccional con targets putativos diferencialmente expresados a nivel transcripcional

Debido a la importancia de los FTs en la modulación de programas génicos en respuestas a diferentes estímulos, a la importancia de la regulación génica a nivel traduccional y al interés de las respuestas de plantas noduladas sometidas a déficit hídrico, se buscaron FTs que hayan presentado regulación a nivel traduccional (mayoritariamente o combinada con transcripcional) tanto por inducción como por represión en el contraste *i* (N+RH vs N+CC). A su vez, se analizó si los *targets* putativos de dichos FTs también presentaron el mismo estado de regulación (*up* o *down*), filtrando la lista de FTs candidatos en aquellos donde el estado de regulación coincidía. Asimismo, dado que los FTs regulan la abundancia de sus *targets*, se filtró la lista reteniendo *targets* que hayan presentado un nivel de regulación transcripcional (exclusivo o combinado con traduccional).

El 70% de los FTs DEGs regulados a nivel PAR o TOT+PAR del contraste *i* (N+RH vs N+CC) caen en los módulos de interés M1, M4, M7, M8 y M20. Aplicando los filtros anteriores, y seleccionando FTs ubicados en los módulos analizados previamente, se encontraron cinco FTs co-localizados en el M4, uno en el M20, uno en el M4 y uno en el M1 *up*-regulados en el contraste *i* con 156 *targets* putativos *up*-regulados y distribuidos en los mismos cuatro módulos. Los FTs seleccionados pertenecen a las familias bZIP, HD-ZIP, NAC y MYB y resultan fuertes candidatos para ser evaluados en estudios futuros (Tabla 8). No se encontraron evidencias de interacción, analizadas por red PPI, entre ninguno de los factores con los

156 *targets* putativos. Análogamente, se repitió el proceso pero para FTs y *targets* que fueron *down*-regulados en el contraste *i*, en los módulos M4, M7, M8 y M20. Se encontraron dos FTs ubicados en el M4 con regulación combinada (transcripcional y traduccional) para los cuales 27 *targets* putativos también estuvieron *down*-regulados en el contraste *i*. Dichos *targets* se distribuyeron en los cuatro módulos. Los FTs encontrados son uno vinculado a la expresión de terpenoides (ET1) y otro vinculado a la señalización por etileno (PPLZ02) (Tabla 88). No se encontraron evidencias de interacción analizadas por red PPI.

Finalmente, a partir de todo el análisis realizado y debido a la importancia de los FTs en la regulación de expresión génica y al interés particular del grupo de investigación se culminó con una lista de genes candidatos, todos FTs, a tener un rol relevante en las respuestas al déficit hídrico en plantas noduladas debido a la regulación traduccional de los mismos. Sumado a los FTs que han sido mencionados anteriormente, se le agrega a la lista un set de FTs seleccionados debido a, además de presentar regulación traduccional o combinada, ser genes *hub* en el módulo que fueron co-localizados. Los genes *hub* fueron definidos como el 10% de los genes con mayor conectividad intramodular, lo que significa que son genes que tienen una alta conexión con el resto de los genes de su módulo. De aquí que son interesantes ya que interactuar con gran cantidad de otros genes los estaría haciendo relevantes en las respuestas biológicas de las plantas en los tratamientos analizados (Tabla 8).

Tabla 8 Lista de genes candidatos a tener un rol relevante en las respuestas de plantas noduladas frente al déficit hídrico. Se muestran los criterios de selección utilizados.

GlymaID	Factor de Transcripción	Módulo de co-expresión	Contraste	Nivel de regulación	Estado	Criterio de selección
Glyma_16G218900	MYB52	M1	<i>i</i>	PAR	Up	<i>Targets</i> putativos con igual estado de regulación a nivel transcripcional o combinado
Glyma_03G003400	ABI5	M4	<i>i</i>	PAR		
Glyma_04G039300	bZIP TRAB1	M4	<i>i</i>	TOTAL+PAR		
Glyma_13G279900	NAC27	M4	<i>i</i>	TOTAL +PAR		
Glyma_16G021000	ATHB-12	M4	<i>i</i>	TOTAL +PAR		
Glyma_16G043200	NAC	M4	<i>i</i>	TOTAL +PAR		
Glyma_14G152700	NAC	M8	<i>i</i>	TOTAL +PAR		
Glyma_04G101900	MYB93	M20	<i>i</i>	PAR		
Glyma_04G027000	Dedo de Zinc B-box 21	M8	<i>i</i>	PAR	Up	Gen hub
Glyma_13G030900	NAC	M20	<i>i</i>	TOTAL +PAR		
Glyma_18G040700	MYB20	M20	<i>i</i>	TOTAL +PAR		
Glyma_17G240100	ERF RAP2-1	M20	<i>i</i>	TOTAL +PAR		
Glyma_10G010300	MYB78	M20	<i>i</i>	TOTAL +PAR		
Glyma_11G127100	Dedo de Zinc B-box 24	M20	<i>i</i>	TOTAL +PAR		
Glyma_05G109500	HAT22	M20	<i>i</i>	TOTAL +PAR		
Glyma_07G052100	HOMEBOX	M4	<i>i</i>	TOTAL +PAR		
Glyma_19G180300	NAC	M4	<i>ii</i>	TOTAL +PAR	Down	
Glyma_05G029000	WRKY 72A	M4	<i>ii</i>	TOTAL +PAR		
Glyma_06G290000	ERF9	M19	<i>ii</i>	TOTAL +PAR	Up	Exclusivo de contraste <i>ii</i> (N+RH vs NN+RH)
Glyma_01G036600	Dedo de Zinc tipo C2H2	M19	<i>ii</i>	TOTAL +PAR		
Glyma_01G231200	ERF021	M19	<i>ii</i>	PAR		
Glyma_13G285400	GRAS	M19	<i>ii</i>	PAR		

M: módulo. **Up:** *up*-regulado. **Down:** *down*-regulado. Genes **hub** fueron definidos como el 10% de los genes con mayor conectividad intramodular. **GlymaID:** identificador “Glyma” versión 4 utilizado en varias bases de datos genómicos.

3.10 Predicción y Validación de Función Génica por EGAD

En este trabajo se buscó predecir funciones génicas mediante el uso de la información de expresión génica obtenida. Para ello es necesario tener información de características a predecir, pudiendo ser por ejemplo información de términos GO. Es así que se obtuvo la anotación de términos GO del genoma de soja mediante *PANNZER*, la cual resultó en un 71% de genes anotados con al menos un término y en un 58% de genes con al menos un proceso biológico anotado. Sumado a este bajo porcentaje, únicamente el 10% de la totalidad de genes tiene una anotación funcional validada experimentalmente. De esta manera el uso de EGAD resulta interesante tanto para aportar sustento en los genes que actualmente presentan funciones predichas por otros métodos como en la predicción de nuevas funciones para genes que no poseen ningún término anotado.

Como red de interacción de entrada se utilizó la matriz de co-expresión obtenida mediante WGCNA, en la cual se encontraron 3864 procesos biológicos GO con representación de al menos un gen.

Sin embargo, únicamente 146 términos presentaron entre 20 y 1000 genes anotados, criterio utilizado para filtrar términos con pocos conteos y que desencadenan una subestimación de la predicción mediante EGAD. La medida de precisión (AUC) promedio de las predicciones para términos fue de 0,55 dentro de un rango entre 0,42 y 0,92. Se seleccionaron procesos biológicos que hayan presentado un AUC mayor o igual a 0,7 ya que como reporta la bibliografía, valores por encima de este umbral son consideradas buenas predicciones. De esta forma, se muestran los procesos biológicos que se predicen con un buen *score* en la [Figura 29](#), dentro de los cuales destacan los de localización de proteína a membrana, ensamblaje de subunidades ribosomales menores, fosforilación oxidativa, nodulación y traducción. Dichos términos presentaron un AUC mayor a 0,8 y además, interesantemente, fueron términos sobrerrepresentados en los módulos y redes analizadas anteriormente.

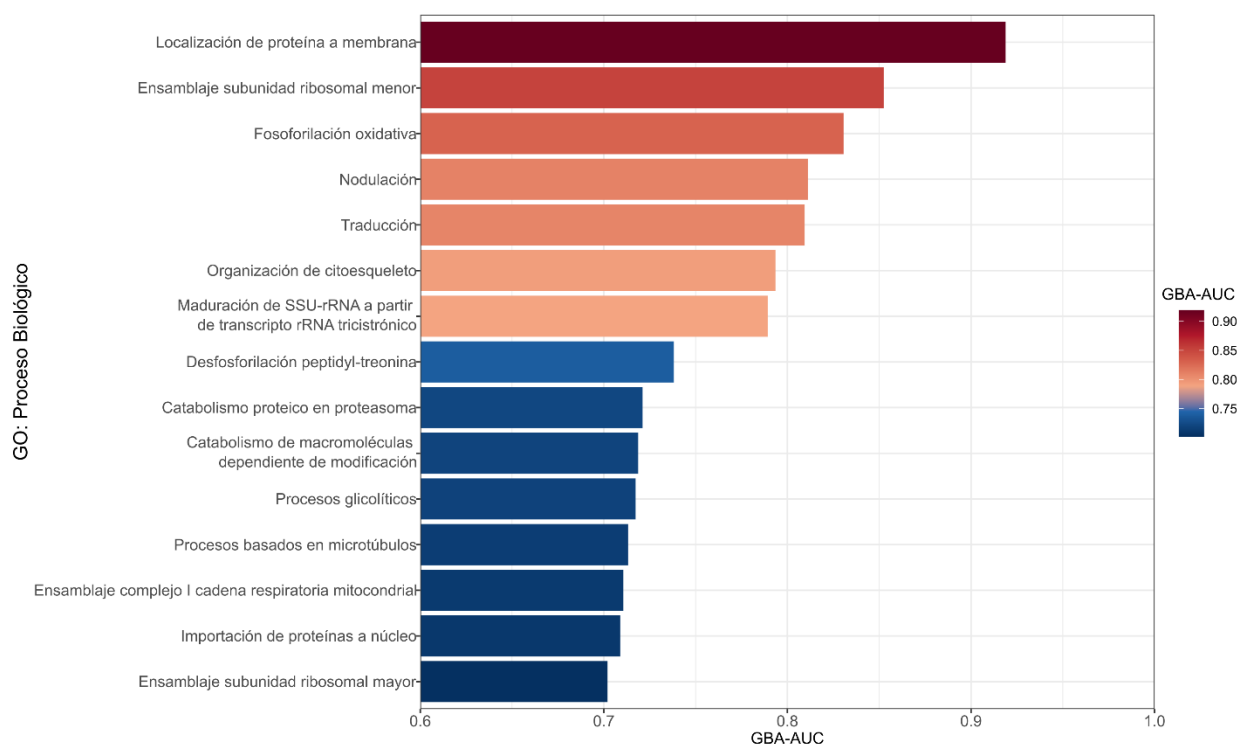


Figura 29 Score de predicción de procesos biológicos GO utilizando la co-expresión de genes de la red WGCNA. GBA: *Guilt-by-Association*. AUC: área bajo la curva ROC (*Area Under Curve*). GO: *Gene Ontology*.

Debido al interés particular de los procesos de nodulación y de traducción, se muestran los genes a los que se les imputó dicho proceso mediante EGAD ([Tabla S3](#)). Por un lado, un gen que codifica para transportador de metal Nramp3a (anotado con los procesos de transporte de metales y elongación de pelo radicular) y uno que codifica para una expansina-A4 (anotado con el proceso de organización de pared celular) fueron imputados con el proceso de nodulación ([Tabla S3](#)). Por otro lado, 44 genes fueron imputados con el proceso de traducción, dentro de los cuales se encuentran diversas proteínas ribosomales de las subunidades mayores y menores del ribosoma ([Tabla S3](#)). Interesantemente, se encontraron cinco locus no caracterizados (anotados con los procesos de organización de membrana tilacoidal, inserción de proteína en membrana mitocondrial, plegado proteico, hidrólisis de enlace fosfodiéster de ARNs y procesos metabólicos) a los cuales se les imputó el proceso de traducción mediante EGAD ([Tabla S3](#)). Se proyectaron los genes imputados con el proceso de traducción en la red WGCNA a

efectos de mostrar el principio GBA que establece que genes conectados tienen potencialidad de participar en los mismos procesos (Figura 30). En este caso se observa que los genes imputados se ubicaron en la vecindad de genes que poseían el proceso de traducción anotado.

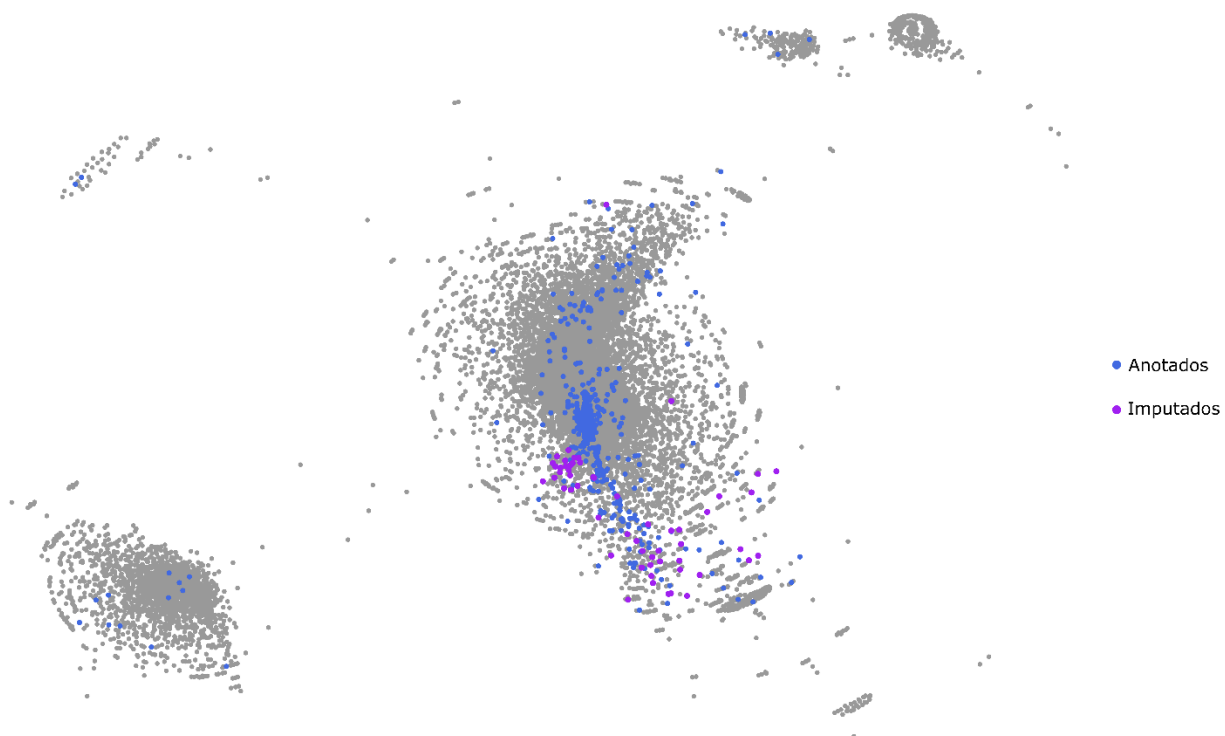


Figura 30 Visualización de predicción de función génica mediante EGAD para el proceso de traducción en la red WGCNA. Anotados: genes anotados con el proceso de traducción utilizados como datos de entrada en la predicción de función. Imputados: genes a los que se le imputa el proceso de traducción mediante EGAD.

3.11 Puesta a Punto del Sistema de Transformación Transitoria para Análisis Funcional de Genes

Una vez obtenidos los genes candidatos a participar en las respuestas de la planta al déficit hídrico cuando se encuentran noduladas, resulta de interés realizar un análisis funcional de cada uno de ellos a los efectos de estudiar el rol que cumplen y validar que su participación en las respuestas sea relevante. Esto puede ser logrado mediante sobreexpresión o silenciamiento génico, para lo cual utilizar el sistema de transformación transitoria mediado por *R. rhizogenes* es altamente útil en plantas como soja, que presenta dificultades al momento de transformarla estable –mediante *Agrobacterium tumefaciens*–, y cuando se está trabajando en biología de la raíz como lo es en este trabajo.

En primer lugar, se evaluó la efectividad de cuatro colonias aisladas de la cepa K599 de *R. rhizogenes*. Las colonias 1 y 2 fueron las más efectivas, generando *hairy roots* en el 58% de los cotiledones infectados cuando no se adicionó acetosiringona en el medio de cultivo bacteriano (Tabla 9). En todos los casos, el uso de acetosiringona en el medio de cultivo bacteriano aumentó la infectividad de la colonia

(Tabla 9). En el caso de la colonia 2 el aumento fue mayor, alcanzando un 83% de cotiledones con presencia de *hairy roots*. Se seleccionó y alicuotó dicha colonia para proseguir con la puesta a punto del sistema.

Tabla 9 Selección de la colonia efectiva de la cepa K599 de *Rhizobium rhizogenes*.

Colonia	Acetosiringona 150 mM	% de cotiledones con HR	Número de cotiledones con HR
1	Sí	67	16
	No	58	14
2	Sí	83	20
	No	58	14
3	Sí	67	16
	No	50	12
4	Sí	17	4
	No	8	2

Se evaluó el uso de acetosiringona en el medio cultivo sobre la efectividad de la colonia. Por colonia se utilizaron cuatro placas conteniendo seis cotiledones cada uno, lo que hace un total de 24 cotiledones por tratamiento. **HR:** *Hairy Roots*.

La transformación de la colonia seleccionada con los plásmidos 35S:RUBY, pEAQHT y piLuCAS resultó efectiva, comprobándose la amplificación por *colony*-PCR (se chequeó en pEAQHT y piLuCAS) (Figura 31). Las cinco colonias evaluadas para cada construcción resultaron positivas, ya que se visualizaron bandas de aproximadamente 1800 pares de bases, cayendo en lo esperado (Figura 31).

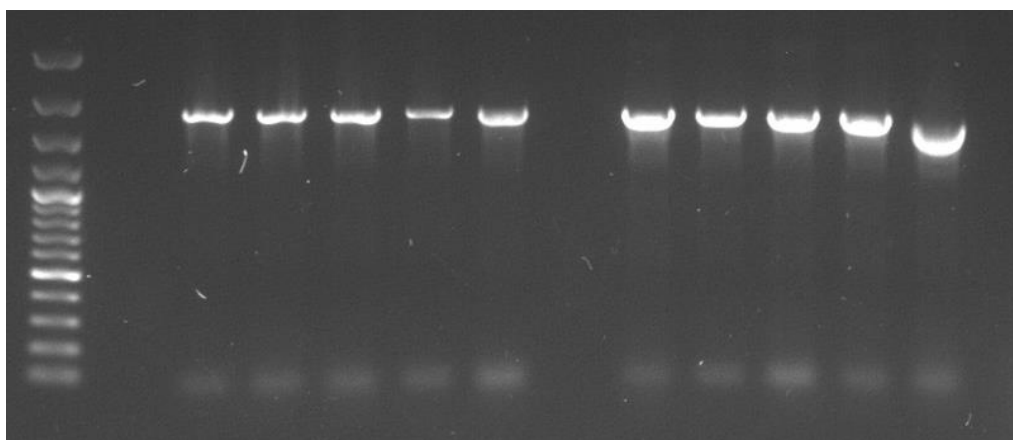


Figura 31 Revelado de colony-PCR de bacterias de la cepa K599 de *Rhizobium rhizogenes* transformadas con los plásmidos de sobreexpresión pEAQHT y de silenciamiento mediante edición génica piLuCAS. En primer carril: marcado de peso molecular 100 pb plus (Thermo). Primer grupo de cinco carriles: colony-PCR de bacterias transformadas con pEAQHT. Segundo grupo de cinco carriles: colony-PCR de bacterias transformadas con piLuCAS.

En la Figura 32A y B se observan placas de Petri con cotiledones produciendo *hairy roots* abundantemente. Con respecto a la confirmación del 35S:RUBY, se realizó mediante inspección visual de las *hairy roots* una vez que se infectaron cotiledones. Estas *hairy roots* positivas muestran una coloración violeta provocada por la sobreexpresión de enzimas responsables de la formación de betalaína a partir de

precursores presentes en la planta (Figura 32C). Si bien la transformación, tanto de la bacteria como de los cotiledones fue efectiva, se descartó continuar usando este plásmido ya que el desarrollo de las HR era más lento en comparación a los otros dos plásmidos.

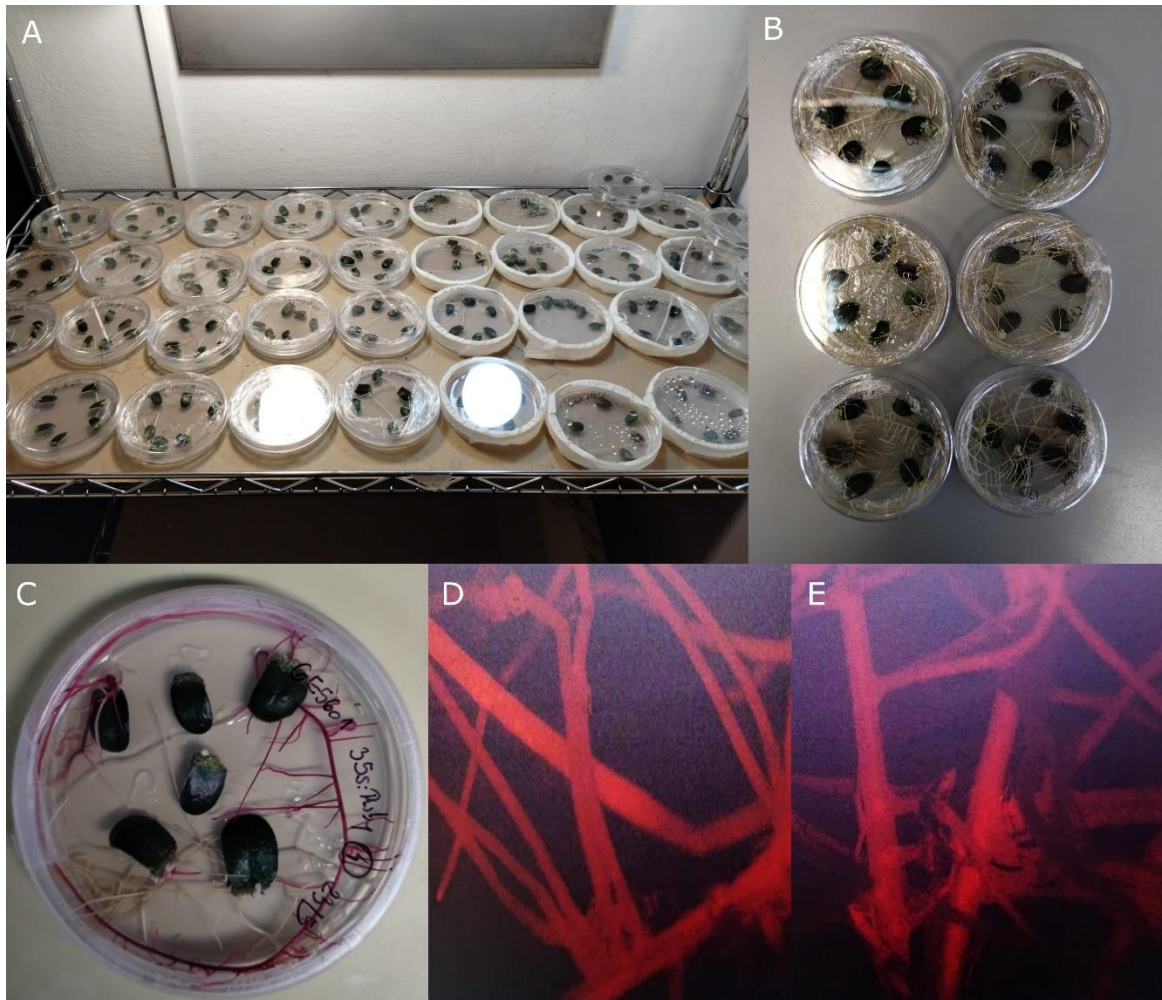


Figura 32 Hairy roots en cotiledones de soja. A,B cotiledones en placas de Petri con agar-agua al 1%. C Hairy roots positivas obtenidas infectando con bacterias conteniendo el plásmido 35:RUBY. D,E Hairy roots positivas obtenidas infectando con *Rhizobium rhizogenes* conteniendo los plásmidos pEAQHT y pILuCAS, respectivamente. Fluorescencia bajo lupa con emisión en luz verde.

Se comprobó la presencia, en cotiledones, de hairy roots positivas con construcciones pEAQHT y pILuCAS mediante observación de fluorescencia bajo lupa con emisión de luz verde (Figura 32D y E). Se utilizaron 24 cotiledones por construcción. El número de hairy roots por cotiledones fue de entre 1 y 11, la cantidad de hairy roots positivas fue entre 0 y 6. Esto resultó en una mediana de 50% de raíces positivas para ambas construcciones (Figura 33).

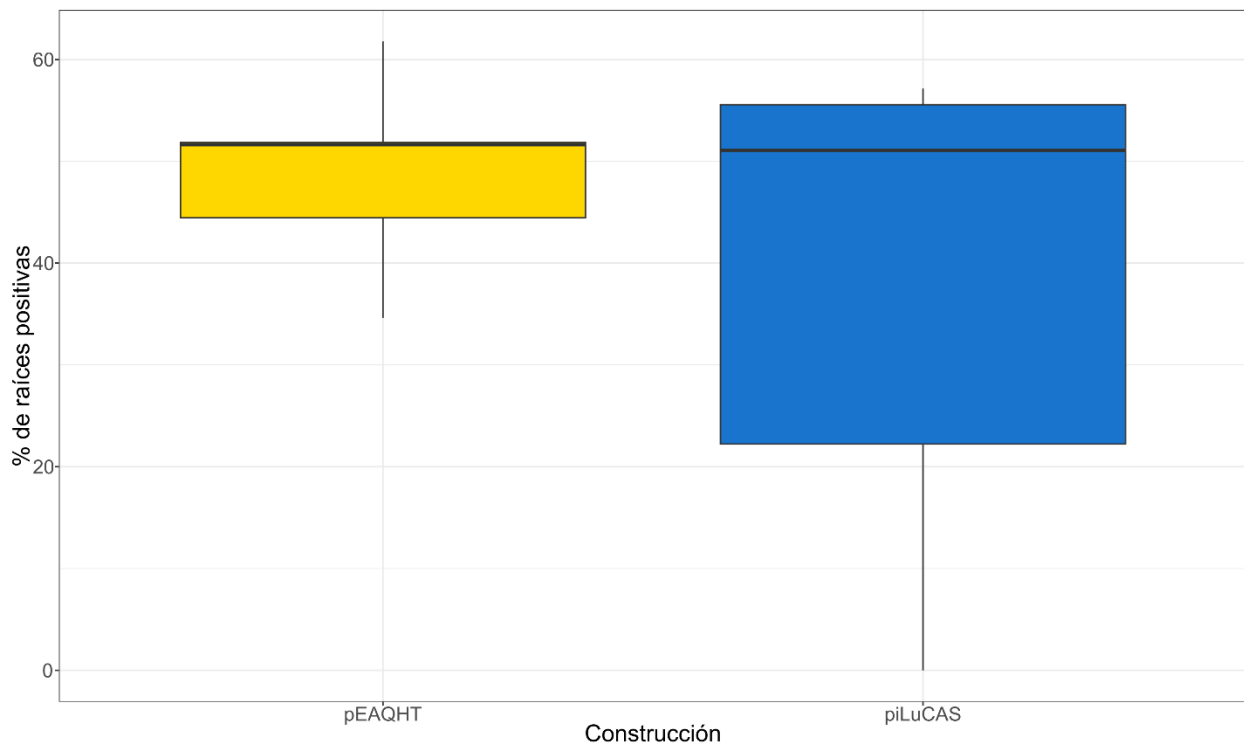


Figura 33 Porcentaje de raíces positivas (transformadas) para cada construcción. Se utilizaron 24 cotiledones por construcción.

La técnica resultó efectiva cuando se probó el sistema de infección y transformación en plántulas (Figura 34). La aparición de callos y *hairy roots* fue notoria, y el desarrollo del sistema radicular conformado por HR también fue notorio (Figura 34A, B y C). Asimismo, el proceso de nodulación ocurrió normalmente en HR (Figura 34B). Se confirmó la presencia de *hairy roots* positivas bajo lupa (Figura 34C). No se cuantificó la eficiencia de la transformación, pero la cantidad de raíces remanentes luego de remover raíz principal y raíces negativas (no transformadas) fue suficiente para que la planta desarrolle un sistema radicular como el observado en las imágenes (Figura 34B y C) y que la planta prospere adecuadamente. Las plántulas transformadas se pusieron a crecer tanto en sistema de doble tubo (Figura 34D) como en sistema de Transformación en Único Tubo Inmerso en Atmósfera (TUTIA) (Figura 34E, F y G), desarrollándose normalmente en cualquiera de los dos y sin diferencias entre ambos. De esta manera, el sistema TUTIA queda establecido como sistema óptimo para futuras implementaciones pues requiere de menor tiempo y esfuerzo en la preparación y realización de la técnica.



Figura 34 Hairy roots en plántulas. **A** Plántula con 7 días posinfección mostrando callos y principio de *hairy root*. **B** Sistema de raíces *hairy roots* noduladas obtenidas de plantas transformadas y traspasadas a sustrato. **C** Sistema de raíces *hairy roots* obtenidas de plantas transformadas y traspasadas a sustrato. Recuadro arriba a la derecha muestra visualización de raíces positivas en lupa con emisión de luz verde previo al traspaso de la plántula al sustrato. **D** Sistema de doble tubo: tubo interior contiene solución nutriente ByD, tubo externo contiene agua. **E, F y G** Sistema de Transformación en Único Tubo Inmerso en Atmósfera (TUTIA), en el cual el agua dentro de la caja plástica genera una atmósfera saturada necesaria para el desarrollo de la infección.

4. DISCUSIÓN

La respuesta de las plantas de soja al estrés por déficit hídrico ha sido ampliamente estudiada en plantas no noduladas (Van Heerden et al. 2007; Song et al. 2016; Wang et al. 2022). En este estudio, esto se analizó en el contraste *iii* (NN+RH vs NN+CC) (Figura 4B). Sin embargo, la mayoría de países productores de soja utilizan plantas inoculadas con rizobios para hacer uso de la capacidad de fijación de nitrógeno debido a la simbiosis que las plantas leguminosas son capaces de establecer (Maluk et al. 2023). Por ende, si se quieren comprender las respuestas de los cultivos en el campo, evaluar las respuestas de plantas sin nodular frente a estímulos abióticos no es el abordaje más pertinente. Sumado a ello, diversos trabajos sugieren que existen respuestas diferenciales al déficit hídrico entre plantas noduladas y no noduladas (Lodeiro et al. 2000; Staudinger et al. 2016; Liu et al. 2022). Plantas de *M. truncatula* noduladas presentaron una senescencia foliar inducida por déficit hídrico retrasada, que se debió principalmente a aumento de osmolitos, interacciones entre fitohormonas y una mayor regulación traduccional (Staudinger et al. 2016). A su vez, en *Astragalus sinicus* la nodulación mejoró la tolerancia al déficit hídrico mediante aumentos en la biosíntesis de poliaminas (por ejemplo, arginina y prolina) y aminoácidos libres gracias a presentar una mejor asimilación de amonio dada por la fijación de nitrógeno (Liu et al. 2022). En el presente trabajo, haciendo uso de la conductancia estomática como rasgo definitorio para la cosecha de plantas de *G. max* en restricción hídrica, se observó que plantas noduladas alcanzaron el 50% de la conductancia inicial (día 0 del período de déficit hídrico) dos días más tarde que plantas sin nodular (Figura 9). Aunque no se encontraron reportes específicos sobre la respuesta de la conductancia estomática al déficit hídrico en diferentes contextos de nodulación en leguminosas, existen investigaciones sobre respuestas entre genotipos contrastantes en un mismo contexto de nodulación. Por ejemplo, Chowdhury et al. (2016) estudian los efectos del déficit hídrico sobre la fotosíntesis y conductancia estomática en tres genotipos de soja tolerantes (Shohag, BARI Soybean-6 y BD2331) y uno susceptible (BGM2026) en estadio vegetativo y sin nodular, encontrando que el susceptible presentó una mayor reducción tanto de la tasa fotosintética como de la conductancia estomática. A su vez, Fenta et al. (2014) caracterizaron las respuestas de dos genotipos de soja (Prima y A-5409RG) al déficit hídrico, estando noduladas, mediante el fenotipado de parámetros fotosintéticos, entre otros. En este estudio, el genotipo menos tolerante a la restricción hídrica (A-5409RG) presentó una rápida disminución de la apertura estomática, alcanzando el 50% de la conductancia en el tercer día (Fenta et al. 2014). Mientras que el genotipo caracterizado como más tolerante alcanzó dicho umbral en el quinto día (Fenta et al. 2014). Dichos números son coincidentes con las respuestas observadas en este estudio (Figura 9), donde, considerando los valores de conductancia estomática, la condición de nodulación estaría aportando a una mayor tolerancia al déficit hídrico en comparación con plantas no noduladas. Sin embargo, se desconocen cuáles son los mecanismos moleculares que gobiernan la mayor tolerancia al déficit hídrico en plantas noduladas respecto a no noduladas.

Sumado a lo anterior, en el caso de plantas noduladas, la respuesta a la restricción hídrica es más compleja en términos de número de DEGs encontrados, pues el contraste *i* (N+RH vs N+CC) presentó 2523 DEGs, el mayor número de DEGs entre los cuatro contrastes analizados (944, 653 y 1017 DEGs en los contrastes *ii*, *iii* y *iv*, respectivamente; Figura 17; Tabla S2) y el contraste *iii* (NN+RH vs NN+CC) fue el que presentó menor número de DEGs (653). Bhaskarla et al. (2020) compararon las respuestas al déficit hídrico de dos genotipos de *Cicer arietinum* con respuestas contrastantes, encontrando que el genotipo tolerante presentó mayor número de DEGs tanto en fase vegetativa como reproductiva. Esto podría indicar que, por analogía, la nodulación contribuiría en la adaptación al déficit hídrico a través de un mayor número

de genes diferencialmente expresados. Asimismo, en el contraste *i* el 26 % de DEGs exclusivos presentó regulación traduccional preferencial, siendo el mayor porcentaje entre los cuatro contrastes analizados (Figura 14). Por su parte, dicho porcentaje tuvo el valor más bajo (10%) en el contraste *iii* (NN+RH vs NN+CC). Esto indica que, en términos de número de DEGs, dentro de las complejas respuestas al déficit hídrico, la regulación traduccional tiene un peso relativo mayor en plantas noduladas. En este sentido, las estrategias de respuesta diferencial, en función del contexto de nodulación frente al déficit hídrico en plantas de soja podrían ser explicadas por cambios en la regulación de la expresión génica. Particularmente, la regulación traduccional de la expresión génica permite a las células responder rápidamente ante la presencia de estímulos, lo que aumenta la adaptabilidad del organismo (Hummel et al. 2009; Merchante et al. 2017; Sablok et al. 2017). Debido a ello, en este trabajo se estudiaron las respuestas transcriptómicas y traductómicas en raíces a los efectos de identificar genes candidatos a tener un rol relevante en las respuestas a déficit hídrico que estén sujetos a control traduccional.

Mediante el análisis de datos transcriptómicos y traductómicos, utilizando una operativa que involucró WGCNA+DEGs (Sánchez-Baizán et al. 2022), se pudieron identificar módulos de co-expresión génica (Figura 15; Figura 17) enriquecidos con los DEGs obtenidos en los diferentes contrastes. A su vez, dado que los DEGs co-localizaron específicamente en algunos módulos en vez de distribuirse uniformemente en toda la red de co-expresión (Figura 18; Tabla 4; Tabla 5), se pudieron asociar módulos de co-expresión con procesos biológicos subyacentes a las respuestas de las plantas en los diferentes escenarios. Esto repercutió además en una reducción de la complejidad de la información facilitando la interpretación de los datos (Li et al. 2020; Sferra et al. 2023), en un contexto en donde la soja presenta 59148 genes. Se pasa de ese número a un total de 11510 genes si se tienen en cuenta la totalidad de los módulos seleccionados (M1, M4, M7, M8, M10, M19, M20, M21 y M32) (Figura 15) y a 2020 genes si se tienen en cuenta únicamente los DEGs co-localizados en dichos módulos (Tabla 4; Tabla 5).

Una alternativa común a la estrategia de WGCNA+DEGs utilizada en este trabajo, que se encuentra en la literatura respecto a estudios transcriptómicos de plantas, consiste en filtrar los datos de transcriptoma primariamente por DEGs y luego realizar el WGCNA en el banco de datos filtrados, es decir DEGs+WGCNA. Aunque esta estrategia es menos demandante computacionalmente, no es recomendada por los desarrolladores de WGCNA ya que los módulos de co-expresión se identifican usando *clusterización* no supervisada, por lo que el filtrado previo de DEGs puede afectar negativamente la arquitectura de la red y por ende conducir a interpretaciones sesgadas (Zhang and Horvath 2005; Langfelder and Horvath 2008; Sánchez-Baizán et al. 2022).

Como primera capa de análisis, la estrategia WGCNA+DEGs permitió analizar DEGs comunes a las respuestas de las plantas a nodulación (contrastos *ii* y *iv*) y a déficit hídrico (contrastos *i* y *iii*) (Tabla 4). Por un lado, en el M10 co-localizaron 113 DEGs, casi el 70% de los comunes a nodulación, por lo que dicho módulo puede ser referido como el principal módulo asociado a las respuestas de la planta frente a la nodulación independientemente de la condición hídrica (Tabla 4; Tabla S1). Concordantemente, el proceso biológico de nodulación estuvo sobrerrepresentado tanto en el módulo como en los DEGs del módulo (Figura 19). Dentro de estos genes se encontraron diversos genes codificantes para nodulinas como las leghemoglobinas, proteínas características del proceso de nodulación. A su vez, la nodulación provocó mayormente la inducción de la expresión génica observada por un aumento en la cantidad de genes *up*-regulados (Figura 13A y B), estando los 113 DEGs en esta categoría. Esto concuerda con lo encontrado por Hayashi et al. (2012) en etapas tempranas de nodulación en soja, y por Hakim et al. (2023) en *Vigna radiata* en etapas tempranas y ya establecidas de la nodulación, quienes encontraron una mayor

inducción de la expresión génica en raíces de plantas noduladas. Otro proceso biológico sobrerrepresentado en el M10 es el de transporte transmembrana (Figura 19). O'Rourke et al. (2014) determinaron que en el proceso de la nodulación, la modulación de la expresión de transportadores de nitrógeno inorgánico tiene un rol importante. Esto es así ya que el proceso de fijación de nitrógeno conlleva la transformación y transporte del mismo bajo otras formas biológicas (como ureidos) y por ende deben modularse los transportadores a efectos de atender cambios en el metabolismo del nitrógeno (Collier and Tegeder 2012; O'Rourke et al. 2014; Pellizzaro et al. 2017).

Por otro lado, con respecto a los DEGs comunes a déficit hídrico (contrastes *i* y *iii*), independientemente del contexto de nodulación, se observó que los mismos co-localizaron predominantemente en los M4, M7 y M8, presentando mayoritariamente una *down*-regulación (Tabla 4). Esto quedó evidenciado también en la Figura 13 donde se observa una mayor represión de la expresión génica en plantas sometidas a déficit hídrico. Los procesos biológicos asociados a esta respuesta común a déficit hídrico fueron los de señalización MAPK, transducción de señales por hormonas, transporte transmembrana, modificaciones de pared celular y biosíntesis de fenilpropanoides, todos *down*-regulados (Figura 20). Wan et al. (2022) encontraron en alfalfa que los procesos de transducción de señales por hormonas, señalización MAPK y biosíntesis de fenilpropanoides se vieron afectados por el déficit hídrico. Interesantemente, encontraron que prácticamente todos los DEGs involucrados en las vías de síntesis de metabolitos secundarios, como monoterpenoides, isoflavonoides y alcaloides estuvieron *down*-regulados en déficit hídrico, indicando que el metabolismo secundario es afectado negativamente por la falta de agua (Wan et al. 2022). Con respecto a las vías de señalización se ha observado tanto *up*- como *down*-regulación de diferentes genes en un mismo experimento (Van Heerden et al. 2007; Bhaskarla et al. 2020; Wan et al. 2022), indicando que la modulación de la señalización es más compleja y se requiere tanto la activación como la represión de genes participantes en ellas (Pareek et al. 2010). Interesantemente, la expresión de los receptores de ABA tipo PYL se reprime en alfalfa bajo déficit hídrico (Wan et al. 2022), hecho que ocurrió en este trabajo, donde la expresión de receptores tipo PYL4 fue menor en plantas con restricción hídrica (Figura 12; Tabla 3). Incluso dichos receptores estuvieron *down*-regulados, a nivel transcripcional y traduccional, cuando se analizaron las respuestas al déficit hídrico en diferentes contextos de nodulación (contraste *ii*, N+RH vs NN+RH), por lo que modular la vía de señalización mediada por ABA sería una respuesta diferencial entre plantas noduladas y no noduladas.

Para poder investigar qué mecanismos moleculares estuvieron involucrados en las respuestas de las plantas a la combinación de las distintas condiciones hídricas y contextos de nodulación se realizaron análisis de enriquecimiento funcional y análisis de redes PPI en los módulos más interesantes de acuerdo a la co-localización de DEGs (Tabla 5). En este sentido, el M7 fue identificado como un módulo relacionado a la represión transcripcional y traduccional cuando una planta experimentó un estímulo. Es decir, en cada contraste se estudió el efecto de un estímulo sobre una condición que no lo presentó y además, en cada contraste, co-localizaron genes *down*-regulados (Tabla 5). Procesos como respuesta a estrés, regulación de la transcripción, fosforilación de proteínas y metabolismo secundario estuvieron sobrerrepresentados en el M7 (Figura 21A). Particularmente, la biosíntesis de metabolitos secundarios, como de fenilpropanoides y de flavonoides, fue el proceso que cobró mayor relevancia en la red PPI analizada para los genes *down*-regulados en todos los contrastes. Esto significa que, como fue mencionado anteriormente, el metabolismo secundario es afectado negativamente por el déficit hídrico. Dentro de los genes que reprimen su expresión debido a un estímulo (ya sea nodulación o déficit hídrico) se encuentran chalcona sintasas, chalcona-flavonona isomerasas, isoflavona sintasas, epimerasas y 2-

hidroxiisoflavanona dehidratasa para la vía de flavonoides y varias peroxidasas en la de fenilpropanoides. Por lo tanto, contemplando los contrastes de mayor interés (contrastos *i* y *ii*), una planta nodulada disminuye la expresión de dichos genes cuando es sometida a restricción hídrica (*i*: N+RH vs N+CC), pero dicha disminución es mayor comparada con la de plantas no noduladas bajo restricción hídrica (*ii*: N+RH vs NN+RH). Nuevamente, sumado a lo que fue mencionado anteriormente, donde Wan et al. (2022) encuentran que la biosíntesis de fenilpropanoides se vio afectada negativamente bajo restricciones hídricas en alfalfa, existen evidencias de represión de genes involucrados en la biosíntesis de flavonoides en nodulación (Hakim et al. 2023) y en déficit hídrico (Marinho et al. 2019).

Con respecto al metabolismo de pigmentos se observó que, en otro genotipo de soja estudiado por el grupo de investigación, la expresión de genes fue preferencialmente inducida en raíces de plantas noduladas y sometidas a déficit hídrico respecto a su control (contraste *i*: N+RH vs N+CC) (Sainz et al. 2024a). Este genotipo presenta una respuesta diferencial al déficit hídrico con respecto al genotipo de este trabajo (datos no publicados, caracterizados por el laboratorio), por lo que la diferencia en el estado de regulación, es decir, *up*- o *down*-regulación de genes involucrados en el metabolismo de pigmentos podría, de alguna manera, estar contribuyendo en la respuesta contrastante al déficit hídrico de ambos genotipos.

El M4 presentó un alto enriquecimiento de genes vinculados al proceso de respuesta a estímulos (Figura 22A). Sin embargo, a diferencia del M7, el M4 tuvo una alta concentración de DEGs, tanto *down*- como *up*-regulados, en los contrastes *i* y *ii* (N+RH vs N+CC y N+RH vs NN+RH) (Tabla 5), es decir que capitaliza genes que son modulados en una planta nodulada bajo déficit hídrico. Reforzando lo observado en módulos anteriores, en la red PPI construida con DEGS, se encontró que la vía de señalización MAPK tuvo tanto inducción como represión de genes (Figura 22B), demostrando que dicha vía sería de importancia en las respuestas de plantas noduladas en un contexto de déficit hídrico sin presentar predominancia de inducción o represión, sino más bien una reprogramación génica de la misma. Asimismo, otro set de genes de la biosíntesis de fenilpropanoides estuvieron *down*-regulados en plantas N+RH, encontrándose 13 peroxidasas y una metiltransferasa, los cuales se suman al listado de lo mencionado anteriormente. Interesantemente, el metabolismo de galactosa estuvo sobrerrepresentado en la red PPI del M4, donde los genes encontrados fueron inducidos en plantas N+RH (Figura 22B). Dentro de estos genes se encuentran varias glicosiltransferasas y una rafinosa sintasa. La rafinosa tiene diversas funciones en la planta, como por ejemplo proteger la semilla contra desecación (Pareek et al. 2010). A su vez, se ha encontrado que en *Zea mays* (maíz) y en *Cucumis sativus* (pepino) que la rafinosa se acumula en condiciones de déficit hídrico (Liu et al. 2023; Dong et al. 2024). Asimismo, la sobreexpresión de una rafinosa sintasa en maíz mejoró la tolerancia a déficit hídrico tanto en *Arabidopsis* como en maíz (Li et al. 2020). En el caso de la soja, se encontró también que la expresión de genes vinculados al metabolismo de la galactosa fue inducida en raíces de plantas de soja noduladas y sometidas a déficit hídrico en un genotipo con respuesta diferente frente al déficit hídrico en comparación con el de este trabajo (Sainz et al. 2024a).

Otro de los módulos seleccionado para análisis de enriquecimiento funcional fue el M20; en él la mayoría de los DEGs co-localizados fueron del contraste *i* (N+RH vs N+CC), es decir de plantas noduladas en restricción hídrica (Tabla 5). A su vez, dicho módulo estuvo fuertemente asociado al proceso biológico de regulación de la transcripción (Figura 23), por lo que se trata de un módulo que comprende predominantemente genes encargados de regular la transcripción génica. Si bien no se pudieron construir redes PPI utilizando los DEGs del módulo debido a no encontrarse evidencia de interacción entre las

proteínas codificadas por los genes en la base de datos de *String*, se encontraron dos factores de transcripción de la familia NAC y uno de la familia MYB *up*-regulados. Dichos factores participan justamente en la regulación de la transcripción y además están anotados como participantes en las vías de señalización mediadas por ABA, lo que estaría indicando que modular dicha vía ABA a nivel de transcripción sería una respuesta inducida en plantas N+RH. Diversos autores han encontrado que la modulación de la interacción entre factores NAC y MYB junto con el ABA aumenta la tolerancia al déficit hídrico en *A. thaliana* y arroz (Gao et al. 2010; Chen et al. 2012; Yin et al. 2017). Por ejemplo, la sobreexpresión de un factor de transcripción NAC aumentó la sensibilidad al ABA en *A. thaliana* mejorando la tolerancia al déficit hídrico sin detener el crecimiento de la planta durante condiciones de estrés (Gao et al. 2010). En el caso del arroz, plantas sobreexpresando un factor de transcripción MYB mostraron mayor tolerancia al déficit, modificando la sensibilidad al ABA y realizando un ajuste osmótico mediante el aumento de los niveles de azúcares solubles (Yin et al. 2017).

Un quinto módulo, el M8, que fue seleccionado también por co-localizar la mayoría de DEGs del contraste *i*, aunque en este caso con preferencia de genes *up*-regulados (Tabla 5), presentó enriquecimiento de los procesos de respuesta a estrés y de biosíntesis de sustancias orgánicas (Figura 24A). A su vez, el metabolismo de galactosa fue un término sobrerrepresentado en la red PPI presentando una glicosil hidrolasa y dos UDP-glucosa 4-epimerasas (UGE) *up*-reguladas (Figura 24B). Las UGEs son enzimas encargadas de convertir glucosa en galactosa y tienen funciones conocidas en la modulación de la biosíntesis de compuestos de pared celular (Rösti et al. 2007; Pareek et al. 2010). A su vez, estudios han demostrado que algunas UGEs aumentan la tolerancia a estrés salino en *Brassica rapa* (Jung et al. 2015) y que estarían involucradas en la resistencia a otros tipos de estrés abiótico mediante la acumulación de metabolitos secundarios (Rösti et al. 2007; Hou et al. 2021). Sumado a ello, inducir la transformación de glucosa en galactosa en condiciones de estrés serviría para permitir la biosíntesis de galactinol, que en conjunto con la rafinosa (*up*-regulada en M4) funcionarían como metabolitos osmoprotectores en condiciones de déficit hídrico (Rösti et al. 2007; Jung et al. 2015). Por otro lado, en el M8 co-localizaron DEGs *up*-regulados que codifican para intermediarios de la biosíntesis de flavonoides (Figura 24B), como una chalcona/estilbeno sintetasa, tres aldo/ceto reductasas, una aldehído deshidrogenasa y una amino oxidasa. Esto contradice lo encontrado en los M4 y M7 donde se observó que genes de la biosíntesis de compuestos fenólicos como fenilpropanoides y flavonoides fue reprimida. Sin embargo, la modulación de la biosíntesis de fenólicos se ha observado que es compleja, dependiente de especies y genotipos, sucediendo tanto activación como represión de genes de la vía (Pareek et al. 2010; Kumar et al. 2019; Marinho et al. 2019; Tyagi et al. 2020; Wan et al. 2022). En este sentido, este trabajo sugiere que la biosíntesis de metabolitos secundarios es reprogramada en plantas noduladas sometidas a restricción hídrica, predominando fuertemente la represión de genes involucrados en las vías de biosíntesis de fenilpropanoides y flavonoides, por lo que la regulación de dichas vías tendría un rol diferencial en plantas noduladas comparadas con no noduladas.

Interesantemente, en el M1 el 70% de los DEGs que co-localizaron en él presentaron una *up*-regulación traduccional preferencial en el contraste *i* (N+RH vs N+CC) (Tabla 5). Asimismo, los procesos de transporte proteico, fosforilación, biogénesis ribosomal y traducción estuvieron sobrerrepresentados en dicho módulo (Figura 25A). En la red PPI construida con los DEGs regulados a nivel traduccional se encontró un enriquecimiento de las vías de desintoxicación de especies oxidantes, señalización mTORC1, fosforilación oxidativa, iniciación de la traducción y formación de pool de subunidades ribosomales 40S (Figura 25B). Las plantas presentan varios sistemas antioxidantes involucrados en la homeostasis de

especies oxidantes, como las enzimas superóxido dismutasas o las oxidorreductasas tiorredoxinas y glutaredoxinas (Meyer et al. 2006; Mittler 2017). Sainz et al. (2022a) encontraron en soja que algunos miembros del sistema tiorredoxina/glutarredoxina presentaron regulación traduccional en plantas noduladas sometidas a déficit hídrico. En este trabajo se encontraron dos genes codificantes para tiorredoxinas, uno para una glutatión peroxidasa y uno para una enzima tipo superóxido dismutasa que presentaron regulación traduccional (Figura 25B). Con respecto a la superóxido dismutasa, encargada de transformar el radical superóxido en peróxido de hidrógeno, se han observado incrementos de su producción bajo déficit hídrico en genotipos tolerantes de *Amaranthus tricolor* (Sarker and Oba 2018) y *Hordeum vulgare* –cebada– (Acar et al. 2001). De esta manera, este trabajo sugiere que en las respuestas al déficit hídrico de plantas noduladas los sistemas antioxidantes enzimáticos están siendo prevalentes en la desintoxicación de especies oxidantes en comparación con la biosíntesis de metabolitos secundarios, la cual se vio principalmente reprimida (módulos 4 y 7).

Sumado a lo anterior, la fosforilación oxidativa es otra vía relevante que está siendo regulada a nivel traduccional en plantas N+RH (Figura 25B). Dentro de los genes *up*-regulados asociados a esta vía se encuentran proteínas ribosomales en fosforilación oxidativa, proteínas de síntesis de ácidos grasos y subunidades del citocromo *b*. La fosforilación oxidativa es una vía que forma parte de la respiración celular catalizando el paso de transformación de ADP en ATP. Por lo tanto, estos resultados estarían sugiriendo un rol relevante del metabolismo energético, particularmente de una fuerte demanda de ATP, en plantas N+RH mediada por un aumento en la traducción de algunos de sus componentes. Coutinho et al. (2023) observaron un aumento en la expresión de proteínas involucradas en la respiración celular bajo condiciones de déficit hídrico.

Por otro lado, respecto al proceso de iniciación de la traducción (Figura 25B), se encontraron diversos genes que codifican para proteínas ribosomales que se asocian a las subunidades mayores del ribosoma indicando que en condiciones de déficit hídrico las plantas noduladas aumentarían la traducción de proteínas de la propia maquinaria traduccional. El control traduccional de la expresión génica en plantas es ampliamente utilizado en diferentes contextos biológicos, ya que permite tener una rápida respuesta frente a estímulos debido a que no se requiere síntesis de más transcritos, sino que se modula la traducción de ARNm presentes en las células (Gebauer and Hentze 2004; Hummel et al. 2009; Merchante et al. 2017; Sablok et al. 2017). Una característica de los ARNm que favorece una rápida regulación traduccional en condiciones adversas son los motivos oligopirimidínicos en la región 5'-UTR (Sheshukova et al. 2017; Kabardaeva et al. 2019; Cockman et al. 2020; Tungalag et al. 2023). Vinculado a esto, en este trabajo se encontró que un motivo de tipo oligopirimidínico (Figura 26) estuvo sobrerrepresentado en los transcritos de las proteínas ribosomales regulados a nivel traduccional en el contraste *i*, lo que sugiere que la regulación traduccional mediada por motivo oligopirimidínico podría coordinar la síntesis de componentes proteicos requeridos para la biogénesis ribosomal. En este sentido, bajo déficit hídrico las plantas noduladas sobreexpresan componentes de la maquinaria traduccional mediante la regulación traduccional de los mismos. Si bien estudios recientes muestran evidencias de la modulación de la biogénesis ribosomal en respuestas a estrés (Moore et al. 2016; Moin et al. 2017; Benevenuto et al. 2021; Dias-Fields and Adamala 2022; Yadav et al. 2024), no se encuentra bibliografía que reporte dicha modulación a nivel traduccional en soja. Las proteínas ribosomales son constituyentes estructurales de los ribosomas, quienes son necesarios para la síntesis proteica (Shamimuzzaman and Vodkin 2018). De aquí surge la interrogante de ¿para sintetizar qué proteínas es que se están produciendo

más componentes ribosomales? Se hipotetiza que la mayor y rápida producción de componentes ribosomales en plantas noduladas bajo situaciones de restricción hídrica pueda ser requerida para la traducción de transcritos relacionados a los procesos que se vieron *up*-regulados como el metabolismo de la galactosa, biosíntesis de algunos flavonoides, fosforilación oxidativa, desintoxicación de especies oxidantes y regulación de la transcripción. Se requieren estudios adicionales para su comprobación.

Es de interés notar que, la regulación traduccional preferencial de genes vinculados al metabolismo energético y a la iniciación de la traducción en raíces de soja noduladas y sometidas a déficit hídrico, no fue encontrada, en las mismas condiciones experimentales (plantas noduladas y sometidas a déficit hídrico respecto a capacidad de campo, contraste *i*: N+RH vs N+CC), en el genotipo de respuesta diferencial estudiado por el laboratorio (Sainz et al. 2024a). De esta manera, esta diferencia podría estar siendo un componente que, en conjunto con las diferencias observadas respecto al metabolismo de pigmentos, explique las respuestas diferenciales de ambos genotipos cuando son sometidos a déficit hídrico.

Resumiendo hasta este punto lo observado en este trabajo respecto a las respuestas al déficit hídrico de plantas noduladas (N+RH vs N+CC), se tiene que: se reprograma la expresión génica de la vía de señalización mediada por MAPK; se modula la biosíntesis de fenilpropanoides y flavonoides, predominantemente por represión; se induce la expresión de genes del metabolismo de la galactosa, y se activan algunas vías enzimáticas de desintoxicación de especies oxidantes, el metabolismo energético y la iniciación de la traducción, éstas últimas tres con una regulación traduccional preferencial. Por ende, se hipotetiza que estas vías podrían ser relevantes a la hora de mejorar la tolerancia al déficit hídrico en soja, y que estas vías (o algunos de sus componentes) tendrían roles claves en la mejor adaptabilidad al déficit cuando la planta se encuentra nodulada como ha sido observado en varias leguminosas (Antolín et al. 1995; Lodeiro et al. 2000; Staudinger et al. 2016; Liu et al. 2022; Álvarez-Aragón et al. 2023; Lumactud et al. 2023).

Adicionalmente, el M19 fue seleccionado debido a co-localizar principalmente DEGs *up*-regulados en las respuestas contrastantes de plantas sometidas a déficit hídrico en contextos de nodulación diferente (contraste *ii*, N+RH vs NN+RH) (Tabla 5). El M19 presentó enriquecimiento significativo de procesos biológicos como los de respuesta de defensa y de regulación de la transcripción (Figura 27). Se encontraron dos factores de transcripción responsivos a etileno inducidos a nivel traduccional y que participan tanto en la regulación de la transcripción como en la respuesta de defensa, por lo que la vía de señalización mediada por etileno podría ser un plus que se suma a las respuestas de plantas N+RH cuando son comparadas frente a plantas NN+RH.

Debido a la importancia de los factores de transcripción en la reprogramación génica y coordinación de procesos biológicos y vías metabólicas, y a la importancia del control traduccional en las respuestas a estímulos (Pareek et al. 2010), se analizó la participación de factores de transcripción agrupados por familias y con regulación principalmente traduccional o combinada (transcripcional + traduccional) en los cuatro contrastes. Interesantemente las familias bHLH, ERF, MYB, NAC y WRKY se destacan con un alto número de DEGs en las respuestas asociadas al contraste *i* (N+RH vs N+CC) (Figura 28). Esto es coincidente con lo reportado en la bibliografía de que la mayoría de factores de transcripción involucrados en la respuesta a déficit hídrico pertenecen a las familias bHLH, MYB, NAC y WRKY (Pareek et al. 2010; Wan et al. 2022). Dichas familias tienen roles caracterizados frente a déficit hídrico, activando y reprimiendo procesos biológicos como el metabolismo de hormonas, de moléculas antioxidantes y de

proteínas vinculadas a estrés (Pareek et al. 2010; Manna et al. 2021). La coordinación de factores de transcripción en las respuestas a diferentes tipos de estrés es compleja, ya que como se observa en la [Figura 28](#) existe tanto inducción como represión de los factores de las mencionadas familias. Sin embargo, la gran mayoría de trabajos apuntan a estudiar factores de transcripción que se inducen bajo déficit hídrico (Gao et al. 2010; Rushton et al. 2010; Chen et al. 2012; Dias et al. 2016; Jiang et al. 2017; Yin et al. 2017; Manna et al. 2021). Por ende, se procedió a obtener una lista de genes candidatos a tener un rol relevante en plantas sometidas a restricciones hídricas, la cual quedó conformada exclusivamente por factores de transcripción que presentaron regulación traduccional o combinada en los contrastes *i* y *ii* ([Tabla 8](#)). A su vez, se aplicaron otros criterios de selección como fueron ser genes *hub* del módulo en el que co-localizaron o que sus *targets* putativos, obtenidos de la base de datos *PlantTFDB* (Jin et al. 2017), hayan presentado el mismo estado de regulación, a nivel transcripcional o combinado. Dicha lista quedó conformada por 22 factores de transcripción ([Tabla 8](#)), pudiendo ser útil para realizar futuros estudios y que aporte evidencia para programas de mejoramiento.

Respecto a las funciones génicas predichas mediante EGAD, se imputó el proceso biológico de nodulación para los genes Glyma_04G044000 (transportador de metal tipo Nramp3a) y Glyma_18G153500 (expansina A4). Dichos genes tienen anotados los procesos biológicos de elongación de pelo radicular y organización de pared celular, ambos procesos del desarrollo radicular ([Tabla S3](#)). En este sentido, Dong et al. (2021) encontraron que los genes *short-root* y *scarecrow*, ambos genes reguladores del desarrollo radicular (Helariutta et al. 2000), fueron redistribuidos evolutivamente para mediar la organogénesis de nódulos en *M. truncatula*. Debido a esto, podría ser interesante analizar si los dos genes de soja mencionados anteriormente, participantes en el desarrollo de radicular, efectivamente cumplen funciones en el proceso de nodulación predicho por EGAD.

Sistema de Transformación Transitoria para el Análisis Funcional de Genes

La sobreexpresión o silenciamiento de genes son métodos ampliamente utilizados para estudiar las funciones biológicas de los genes (Nath Radhamony et al. 2005; Kereszt et al. 2007; Clemow et al. 2011). En este sentido, la generación de líneas mutantes generadas por transformación estable permite estudiar fenotipos en función de la pérdida de función o sobreexpresión de genes de interés a nivel planta entera. Especies como *A. thaliana* o *N. tabacum* son permisivas para realizar dicha transformación (Clemow et al. 2011). Sin embargo, la soja es considerada recalcitrante para la transformación estable, por lo que se debe optar por otras técnicas (Clemow et al. 2011). En este trabajo, debido a que se estudiaron las respuestas génicas a nivel de raíz, utilizar el sistema de transformación transitoria resulta adecuado para analizar las funciones de los genes candidatos. Dicho sistema permite la sobreexpresión o silenciamiento de genes a nivel de sistema radicular, formado por las *hairy roots*, obteniendo una planta compuesta por la parte aérea silvestre y la raíz transgénica, por lo que se pueden analizar las funciones génicas en la biología de la raíz (Kereszt et al. 2007; Clemow et al. 2011; Mohammadi-Dehcheshmeh et al. 2014). Debido a esto, se puso a punto el sistema de transformación transitoria mediado por *R. rhizogenes* para poder realizar el análisis funcional de los genes candidatos.

En primer lugar, como era de esperar, el uso de acetosiringona en el medio de cultivo bacteriano promovió la aparición de *hairy roots* ([Tabla 9](#)). Esto va acorde con lo reportado en bibliografía ya que la acetosiringona es un flavonoide que favorece la infección de la bacteria en la planta mediante una interacción de la molécula con la proteína virulenta VirA de la bacteria (De Saeger et al. 2021). Esto se debe a las características naturales de una infección bacteriana, en la cual cuando una planta presenta

una herida secreta compuestos fenólicos que son detectados por VirA, lo que desencadena una cascada de señalización que activa otros genes virales (De Saeger et al. 2021). De esta forma, la incorporación de acetosiringona favorece dicho proceso.

Por otro lado, se probó el uso de dos genes reporteros, el RUBY que puede visualizarse a ojo desnudo, y el DsRED que debe ser visualizado bajo lupa o microscopio. Ambos reporteros funcionaron bien en la transformación, es decir que se obtuvieron *hairy roots* que los expresaban. Sin embargo, la tasa aparición de *hairy roots* con RUBY se vio enlentecida, obteniendo *hairy roots* con un tamaño considerable a los aproximadamente 30 días luego de la infección, mientras que para *hairy roots* con DsRED dicho tiempo fue de entre 14 y 20 días. Debido a esto se continuó trabajando con el DsRED como gen reportero. Lo anterior podría deberse a que en el caso del RUBY se provoca la sobreexpresión de enzimas responsables de la formación de betaláina a partir de precursores presentes en la planta que podría provocar un enlentecimiento del metabolismo (He et al. 2020). Sin embargo, Parthasarathy (2023) obtiene *hairy roots* en soja mediante uso del reportero RUBY en 14 días utilizando la cepa R1000 de *R. rhizogenes*. Además del uso de otra cepa diferente a este trabajo, el uso de otro genotipo podría explicar la diferencia observada, ya que las tasas de transformación y aparición de *hairy roots* son genotipo dependientes (Kereszt et al. 2007; Clemow et al. 2011; He et al. 2020; De Saeger et al. 2021; Parthasarathy 2023).

El porcentaje promedio de raíces positivas, expresando DsRED, en cotiledones fue de 50% (Figura 33). Este porcentaje es bueno, ubicándose en el centro del rango reportado en la bibliografía (20-80%) (Kereszt et al. 2007; Harvey et al. 2008; Clemow et al. 2011; De Saeger et al. 2021; Cheng et al. 2021). Los porcentajes más elevados han sido encontrados cuando los cotiledones infectados fueron depositados, en vez de sobre agar-agua, sobre un medio enriquecido en sales, vitaminas y la selección correspondiente al vector y/o acetosiringona (Harvey et al. 2008; Rana et al. 2016; Kong et al. 2019; Rong Li et al. 2020; Cheng et al. 2021). Podría probarse cambiar el protocolo actual establecido para el laboratorio a efectos de obtener mayores eficiencias. No obstante, dicha mejora sería implementada para la etapa de infección cotiledones cuando en realidad el objetivo final es obtener mayores eficiencias en infecciones en plántulas. Por lo tanto, el uso de un medio enriquecido en este caso no sería directamente viable pues en el método de infección en plántulas no se requiere una etapa de incubación del material vegetal en un medio sólido como es el caso de en cotiledones, sino que se infecta directamente la herida de la plántula con el césped bacteriano. Sin embargo, podría probarse aplicar sobre la herida, previo al embetunamiento con bacteria, una capa de una solución conteniendo acetosiringona y/o antibiótico correspondiente a la selección del vector a efectos de intentar levantar el porcentaje de eficiencia de transformación. Dado lo comentado anteriormente, ambos compuestos podrían ser los causantes de aumentar la eficiencia. Por un lado, el uso de antibiótico favorece la selección de bacterias conteniendo el vector que, si bien en todo el protocolo se manejan condiciones asépticas y se van seleccionando las bacterias en diferentes etapas, no resultaría menor incorporar otro paso de selección. Por otro lado, el uso de acetosiringona, como fue mencionado anteriormente, favorece la virulencia de la bacteria.

Finalmente, se probó en este trabajo el uso de una única atmósfera controlada y generalizada para proveer de la humedad suficiente y necesaria para el desarrollo de las *hairy roots* en plántulas en sustitución del uso de una atmósfera individual, es decir para cada planta (Figura 34). Tanto las plántulas como las *hairy roots* se desarrollaron normalmente en este nuevo sistema, por lo que el uso de él es sumamente apropiado ya que disminuye los tiempos y esfuerzos de la técnica.

5. CONCLUSIONES

- El análisis combinado de WGCNA+DEGs resultó útil para explorar el transcriptoma y traductoma de soja bajo diferentes condiciones hídricas y contextos de nodulación.
- Se identificaron módulos de co-expresión génica enriquecidos en DEGs que pudieron ser asociados a procesos biológicos.
- Se asociaron respuestas específicas de las plantas con módulos de co-expresión génica.
- La regulación traduccional de la expresión génica fue relevante en las respuestas de plantas noduladas a déficit hídrico, en particular para los procesos de iniciación de la traducción, fosforilación oxidativa y formación de pool de subunidades ribosomales 40S libres.
- Se identificaron genes candidatos robustos sujetos a control traduccional en raíces de plantas de soja noduladas sometidas a déficit hídrico.
- Se puso a punto el sistema de transformación transitoria mediado por *Rhizobium rhizogenes* para el análisis funcional de genes candidatos.

6. REFERENCIAS BIBLIOGRÁFICAS

- Abdel-Haleem H, Lee GJ, Boerma RH (2011) Identification of QTL for increased fibrous roots in soybean. *Theoretical and Applied Genetics* 122:935–946. <https://doi.org/10.1007/s00122-010-1500-9>
- Abdel-Lateif K, Bogusz D, Hoher V (2012) The role of flavonoids in the establishment of plant roots endosymbioses with arbuscular mycorrhiza fungi, rhizobia and Frankia bacteria. *Plant Signal Behav* 7:636–641. <https://doi.org/10.4161/psb.20039>
- Acar O, Türkan I, Özdemir F (2001) Superoxide dismutase and peroxidase activities in drought sensitive and resistant barley (*Hordeum vulgare* L.) varieties. *Acta Physiol Plant* 23:351–356. <https://doi.org/10.1007/s11738-001-0043-8>
- Alexa A, Rahnenfuhrer J (2021) topGO: Enrichment Analysis for Gene Ontology
- Ali S, Ullah S, Khan MN, et al (2022) The Effects of Osmosis and Thermo-Priming on Salinity Stress Tolerance in *Vigna radiata* L. *Sustainability* 14:12924. <https://doi.org/10.3390/su141912924>
- Álvarez-Aragón R, Palacios JM, Ramírez-Parra E (2023) Rhizobial symbiosis promotes drought tolerance in *Vicia sativa* and *Pisum sativum*. *Environ Exp Bot* 208:105268. <https://doi.org/10.1016/j.envexpbot.2023.105268>
- Antolín MC, Yoller J, Sánchez-Díaz M (1995) Effects of temporary drought on nitrate-fed and nitrogen-fixing alfalfa plants. *Plant Science* 107:159–165. [https://doi.org/10.1016/0168-9452\(95\)04108-7](https://doi.org/10.1016/0168-9452(95)04108-7)
- Antolín-Llovera M, Ried MK, Binder A, Parniske M (2012) Receptor Kinase Signaling Pathways in Plant-Microbe Interactions. *Annu Rev Phytopathol* 50:451–473. <https://doi.org/10.1146/annurev-phyto-081211-173002>
- Atkinson NJ, Urwin PE (2012) The interaction of plant biotic and abiotic stresses: from genes to the field. *J Exp Bot* 63:3523–3543. <https://doi.org/10.1093/jxb/ers100>
- Azam M, Zhang S, Li J, et al (2023) Identification of hub genes regulating isoflavone accumulation in soybean seeds via GWAS and WGCNA approaches. *Front Plant Sci* 14:. <https://doi.org/10.3389/fpls.2023.1120498>
- Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *AAAI Press* 28–36
- Bailey TL, Johnson J, Grant CE, Noble WS (2015) The MEME Suite. *Nucleic Acids Res* 43:W39–W49. <https://doi.org/10.1093/nar/gkv416>
- Ballouz S, Weber M, Pavlidis P, Gillis J (2017) EGAD: Ultra-fast functional analysis of gene networks. *Bioinformatics* 33:612–614. <https://doi.org/10.1093/bioinformatics/btw695>
- Benevenuto RF, Zanatta CB, Guerra MP, et al (2021) Proteomic profile of glyphosate-resistant soybean under combined herbicide and drought stress conditions. *Plants* 10:. <https://doi.org/10.3390/plants10112381>
- Betti M, Pérez-Delgado C, García-Calderón M, et al (2012) Cellular stress following water deprivation in the model legume *Lotus japonicus*. *Cells* 1:1089–1106. <https://doi.org/10.3390/cells1041089>

- Beyhaut E, Barlocco C, Mortalena M, et al (2020) CONVENIO INIA-MGAP: Sistema de Control de Calidad de Inoculantes. *Revista INIA* 91–93. https://doi.org/10.1007/978-3-642-37241-4_7
- Bhaskarla V, Zinta G, Ford R, et al (2020) Comparative Root Transcriptomics Provide Insights into Drought Adaptation Strategies in Chickpea (*Cicer arietinum* L.). *Int J Mol Sci* 21:1781. <https://doi.org/10.3390/ijms21051781>
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/BIOINFORMATICS/BTU170>
- Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 34:525–527. <https://doi.org/10.1038/nbt.3519>
- Broughton W, Dilworth M (1971) Control of leghaemoglobin synthesis in snake beans. *Biochemical Journal* 125:. <https://doi.org/10.1042/bj1251075>
- Brown DD (1981) Gene Expression in Eukaryotes. *Science* (1979) 211:667–674. <https://doi.org/10.1126/science.6256857>
- Butts CT (2008) network: A Package for Managing Relational Data in R. *J Stat Softw* 24:. <https://doi.org/10.18637/jss.v024.i02>
- Cao Y, Song F, Goodman RM, Zheng Z (2006) Molecular characterization of four rice genes encoding ethylene-responsive transcriptional factors and their expressions in response to biotic and abiotic stress. *J Plant Physiol* 163:1167–1178. <https://doi.org/10.1016/j.jplph.2005.11.004>
- Cassel DK, Nielsen DR (2018) Field Capacity and Available Water Capacity. pp 901–926
- Charrad M, Ghazzali N, Boiteau V, Niknafs A (2014) NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *J Stat Softw* 61:. <https://doi.org/10.18637/jss.v061.i06>
- Chen L, Song Y, Li S, et al (2012) The role of WRKY transcription factors in plant abiotic stresses. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1819:120–128. <https://doi.org/10.1016/j.bbagrm.2011.09.002>
- Chen M, Xu Z, Xia L, et al (2009) Cold-induced modulation and functional analyses of the DRE-binding transcription factor gene, GmDREB3, in soybean (*Glycine max* L.). *J Exp Bot* 60:121–135. <https://doi.org/10.1093/jxb/ern269>
- Cheng Y, Wang X, Cao L, et al (2021) Highly efficient *Agrobacterium* rhizogenes-mediated hairy root transformation for gene functional and gene editing analysis in soybean. *Plant Methods* 17:73. <https://doi.org/10.1186/s13007-021-00778-7>
- Chowdhury J, Karim M, Khaliq Q, et al (2016) Effect of drought stress on gas exchange characteristics of four soybean genotypes. *Bangladesh Journal of Agricultural Research* 41:195–205
- Cilliers M, van Wyk SG, van Heerden PDR, et al (2018) Identification and changes of the drought-induced cysteine protease transcriptome in soybean (*Glycine max*) root nodules. *Environ Exp Bot* 148:59–69. <https://doi.org/10.1016/j.envexpbot.2017.12.005>

- Clement M, Lambert A, Herouart D, Boncompagni E (2008) Identification of new up-regulated genes under drought stress in soybean nodules. *Gene* 426:15–22. <https://doi.org/10.1016/j.gene.2008.08.016>
- Clemow SR, Clairmont L, Madsen LH, Guinel FC (2011) Reproducible hairy root transformation and spot-inoculation methods to study root symbioses of pea. *Plant Methods* 7:46. <https://doi.org/10.1186/1746-4811-7-46>
- Coate JE, Bar H, Doyle JJ (2014) Extensive Translational Regulation of Gene Expression in an Allopolyploid (*Glycine dolichocarpa*). *Plant Cell* 26:136–150. <https://doi.org/10.1105/TPC.113.119966>
- Cockman E, Anderson P, Ivanov P (2020) Top mrnps: Molecular mechanisms and principles of regulation. *Biomolecules* 10:1–18
- Collier R, Tegeder M (2012) Soybean ureide transporters play a critical role in nodule development, function and nitrogen export. *The Plant Journal* 72:355–367. <https://doi.org/10.1111/j.1365-313X.2012.05086.x>
- Concha C, Doerner P (2020) The impact of the rhizobia–legume symbiosis on host root system architecture. *J Exp Bot* 71:3902–3921. <https://doi.org/10.1093/jxb/eraa198>
- Costa-Silva J, Domingues D, Lopes FM (2017) RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS One* 12:e0190152. <https://doi.org/10.1371/journal.pone.0190152>
- Coutinho FS, Mesquita RO, Rodrigues JM, et al (2023) Alterations in the root phenylpropanoid pathway and root–shoot vessel system as main determinants of the drought tolerance of a soybean genotype. *Physiology and Molecular Biology of Plants* 29:559–577. <https://doi.org/10.1007/s12298-023-01307-7>
- da Silva HAP, Caetano VS, Pessôa DDV, et al (2024) Unraveling the drought-responsive transcriptomes in nodules of two common bean genotypes during biological nitrogen fixation. *Front Plant Sci* 15:. <https://doi.org/10.3389/fpls.2024.1345379>
- De Saeger J, Park J, Chung HS, et al (2021) Agrobacterium strains and strain improvement: Present and outlook. *Biotechnol Adv* 53:107677. <https://doi.org/10.1016/j.biotechadv.2020.107677>
- Denyer T, Ma X, Klesen S, et al (2019) Spatiotemporal Developmental Trajectories in the Arabidopsis Root Revealed Using High-Throughput Single-Cell RNA Sequencing. *Dev Cell* 48:840–852.e5. <https://doi.org/10.1016/J.DEVCEL.2019.02.022>
- Depuydt T, Vandepoele K (2021) Multi-omics network-based functional annotation of unknown Arabidopsis genes. *The Plant Journal* 108:1193–1212. <https://doi.org/10.1111/tpj.15507>
- Desikan R (2003) ABA, hydrogen peroxide and nitric oxide signalling in stomatal guard cells. *J Exp Bot* 55:205–212. <https://doi.org/10.1093/jxb/erh033>
- Dessimoz C, Skunca N (2017) *The Gene Ontology Handbook*. Springer New York, New York, NY
- D’Haeze W, Holsters M (2002) Nod factor structures, responses, and perception during initiation of nodule development. *Glycobiology* 12:79R–105R. <https://doi.org/10.1093/glycob/12.6.79R>

- Dias LP, de Oliveira-Busatto LA, Bodanese-Zanettini MH (2016) The differential expression of soybean [Glycine max (L.) Merrill] WRKY genes in response to water deficit. *Plant Physiology and Biochemistry* 107:288–300. <https://doi.org/10.1016/j.plaphy.2016.06.018>
- Dias-Fields L, Adamala KP (2022) Engineering Ribosomes to Alleviate Abiotic Stress in Plants: A Perspective. *Plants* 11:. <https://doi.org/10.3390/plants11162097>
- DIEA-MGAP (2023) Estadísticas Agropecuarias (DIEA) Coordinador Tareas preparatorias Relevamiento de información Muestreo y procesamiento de información Redacción del informe
- Dong D, Qi C, Zhang J, et al (2024) CsHSFA1d Promotes Drought Stress Tolerance by Increasing the Content of Raffinose Family Oligosaccharides and Scavenging Accumulated Reactive Oxygen Species in Cucumber. *Plant Cell Physiol* 65:809–822. <https://doi.org/10.1093/pcp/pcae023>
- Dong W, Zhu Y, Chang H, et al (2021) An SHR–SCR module specifies legume cortical cell fate to enable nodulation. *Nature* 589:586–590. <https://doi.org/10.1038/s41586-020-3016-z>
- Dudek A (2020) Silhouette Index as Clustering Evaluation Tool. pp 19–33
- Fehr WR, Caviness CE (1977) Stages of soybean development
- Fenta B, Beebe S, Kunert K, et al (2014) Field Phenotyping of Soybean Roots for Drought Stress Tolerance. *Agronomy* 4:418–435. <https://doi.org/10.3390/agronomy4030418>
- Ferguson BJ, Indrasumunar A, Hayashi S, et al (2010) Molecular Analysis of Legume Nodule Development and Autoregulation. *J Integr Plant Biol* 52:61–76. <https://doi.org/10.1111/j.1744-7909.2010.00899.x>
- Ferguson BJ, Mens C, Hastwell AH, et al (2019) Legume nodulation: The host controls the party. *Plant Cell Environ* 42:41–51. <https://doi.org/10.1111/pce.13348>
- Foyer CH, Lam H-M, Nguyen HT, et al (2016) Neglecting legumes has compromised human health and sustainable food production. *Nat Plants* 2:16112. <https://doi.org/10.1038/nplants.2016.112>
- Fujita M, Fujita Y, Maruyama K, et al (2004) A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway. *The Plant Journal* 39:863–876. <https://doi.org/10.1111/j.1365-313X.2004.02171.x>
- Gage DJ (2004) Infection and Invasion of Roots by Symbiotic, Nitrogen-Fixing Rhizobia during Nodulation of Temperate Legumes. *Microbiology and Molecular Biology Reviews* 68:280–300. <https://doi.org/10.1128/MMBR.68.2.280-300.2004>
- Gao F, Xiong A, Peng R, et al (2010) OsNAC52, a rice NAC transcription factor, potentially responds to ABA and confers drought tolerance in transgenic plants. *Plant Cell, Tissue and Organ Culture (PCTOC)* 100:255–262. <https://doi.org/10.1007/s11240-009-9640-9>
- Garcia-Moreno A, López-Domínguez R, Villatoro-García JA, et al (2022) Functional Enrichment Analysis of Regulatory Elements. *Biomedicines* 10:590. <https://doi.org/10.3390/biomedicines10030590>
- Gebauer F, Hentze MW (2004) Molecular mechanisms of translational control. *Nat Rev Mol Cell Biol* 5:827–835. <https://doi.org/10.1038/nrm1488>

- Gillis J, Pavlidis P (2012) “Guilt by Association” Is the Exception Rather Than the Rule in Gene Networks. *PLoS Comput Biol* 8:e1002444. <https://doi.org/10.1371/journal.pcbi.1002444>
- Gong P, Zhang J, Li H, et al (2010) Transcriptional profiles of drought-responsive genes in modulating transcription signal transduction, and biochemical pathways in tomato. *J Exp Bot* 61:3563–3575. <https://doi.org/10.1093/JXB/ERQ167>
- Gresshoff PM, Hayashi S, Biswas B, et al (2015) The value of biodiversity in legume symbiotic nitrogen fixation and nodulation for biofuel and food production. *J Plant Physiol* 172:128–136. <https://doi.org/10.1016/j.jplph.2014.05.013>
- Guo H, Ecker JR (2004) The ethylene signaling pathway: new insights. *Curr Opin Plant Biol* 7:40–49. <https://doi.org/10.1016/j.pbi.2003.11.011>
- Hakim S, Imran A, Hussain MS, Mirza MS (2023) RNA-Seq analysis of mung bean (*Vigna radiata* L.) roots shows differential gene expression and predicts regulatory pathways responding to taxonomically different rhizobia. *Microbiol Res* 275:127451. <https://doi.org/10.1016/j.micres.2023.127451>
- Harvey JJW, Lincoln JE, Gilchrist DG (2008) Programmed cell death suppression in transformed plant tissue by tomato cDNAs identified from an *Agrobacterium* rhizogenes-based functional screen. *Molecular Genetics and Genomics* 279:509–521. <https://doi.org/10.1007/s00438-008-0329-2>
- Hayashi S, Reid DE, Lorenc MT, et al (2012) Transient Nod factor-dependent gene expression in the nodulation-competent zone of soybean (*Glycine max* [L.] Merr.) roots. *Plant Biotechnol J* 10:995–1010. <https://doi.org/10.1111/j.1467-7652.2012.00729.x>
- He Y, Zhang T, Sun H, et al (2020) A reporter for noninvasively monitoring gene expression and plant transformation. *Hortic Res* 7:152. <https://doi.org/10.1038/s41438-020-00390-1>
- Helariutta Y, Fukaki H, Wysocka-Diller J, et al (2000) The SHORT-ROOT Gene Controls Radial Patterning of the Arabidopsis Root through Radial Signaling. *Cell* 101:555–567. [https://doi.org/10.1016/S0092-8674\(00\)80865-X](https://doi.org/10.1016/S0092-8674(00)80865-X)
- Hirsch AM, Larue TA, Doyle J (1997) Is the Legume Nodule a Modified Root or Stem or an Organ *sui generis* ? *CRC Crit Rev Plant Sci* 16:361–392. <https://doi.org/10.1080/07352689709701954>
- Hou J, Tian S, Yang L, et al (2021) A systematic review of the Uridine diphosphate-Galactose/Glucose-4-epimerase (UGE) in plants. *Plant Growth Regul* 93:267–278. <https://doi.org/10.1007/s10725-020-00686-1>
- Hummel M, Rahmani F, Smeekens S, Hanson J (2009) Sucrose-mediated translational control. *Ann Bot* 104:1–7. <https://doi.org/10.1093/aob/mcp086>
- Imagawa M (1996) Negative regulation of gene expression in eukaryotes
- Ingolia NT (2014) Ribosome profiling: New views of translation, from single codons to genome scale. *Nat Rev Genet* 15:205–213
- Jackson RJ, Hellen CUT, Pestova T V. (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* 11:113–127. <https://doi.org/10.1038/nrm2838>

- Jacquier A (2009) The complex eukaryotic transcriptome: Unexpected pervasive transcription and novel small RNAs. *Nat Rev Genet* 10:833–844
- Jeong H-J, Kim YJ, Kim SH, et al (2011) Nonsense-Mediated mRNA Decay Factors, UPF1 and UPF3, Contribute to Plant Defense. *Plant Cell Physiol* 52:2147–2156. <https://doi.org/10.1093/pcp/pcr144>
- Jiang J, Ma S, Ye N, et al (2017) WRKY transcription factors in plant responses to stresses. *J Integr Plant Biol* 59:86–101
- Jin J, Tian F, Yang D-C, et al (2017) PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res* 45:D1040–D1045. <https://doi.org/10.1093/nar/gkw982>
- Jogawat A, Yadav B, Chhaya, et al (2021) Crosstalk between phytohormones and secondary metabolites in the drought stress tolerance of crop plants: A review. *Physiol Plant* 172:1106–1132. <https://doi.org/10.1111/ppl.13328>
- Jores T, Hamm M, Cuperus JT, Queitsch C (2023) Frontiers and techniques in plant gene regulation. *Curr Opin Plant Biol* 75:102403. <https://doi.org/10.1016/j.pbi.2023.102403>
- Joshi R, Wani SH, Singh B, et al (2016) Transcription factors and plants response to drought stress: Current understanding and future directions. *Front Plant Sci* 7:. <https://doi.org/10.3389/fpls.2016.01029>
- Jung YJ, Kyoung JH, Nou IS, et al (2015) Molecular characterization of the UDP-glucose 4-epimerase (BrUGE) gene family in response to biotic and abiotic stress in Chinese cabbage (*Brassica rapa*). *Plant Biotechnol Rep* 9:339–350. <https://doi.org/10.1007/s11816-015-0370-7>
- Kabardaeva K V., Tyurin AA, Pavlenko OS, et al (2019) Fine Tuning of Translation: A Complex Web of Mechanisms and Its Relevance to Plant Functional Genomics and Biotechnology. *Russian Journal of Plant Physiology* 66:835–849
- Kanehisa M (2002) The KEGG Database. pp 91–103
- Kanehisa M, Goto S, Sato Y, et al (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 40:D109–D114. <https://doi.org/10.1093/nar/gkr988>
- Kanehisa M, Sato Y, Kawashima M, et al (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 44:D457–D462. <https://doi.org/10.1093/nar/gkv1070>
- Kant C, Pradhan S, Bhatia S (2016) Dissecting the Root Nodule Transcriptome of Chickpea (*Cicer arietinum* L.). *PLoS One* 11:e0157908. <https://doi.org/10.1371/journal.pone.0157908>
- Kaschuk G, Hungria M, Leffelaar PA, et al (2010) Differences in photosynthetic behaviour and leaf senescence of soybean (*Glycine max* [L.] Merrill) dependent on N₂ fixation or nitrate supply. *Plant Biol* 12:60–69. <https://doi.org/10.1111/j.1438-8677.2009.00211.x>
- Kaschuk G, Yin X, Hungria M, et al (2012) Photosynthetic adaptation of soybean due to varying effectiveness of N₂ fixation by two distinct *Bradyrhizobium japonicum* strains. *Environ Exp Bot* 76:1–6. <https://doi.org/10.1016/j.envexpbot.2011.10.002>

- Kawaguchi R, Girke T, Bray EA, Bailey-Serres J (2004) Differential mRNA translation contributes to gene regulation under non-stress and dehydration stress conditions in *Arabidopsis thaliana*. *The Plant Journal* 38:823–839. <https://doi.org/10.1111/J.1365-313X.2004.02090.X>
- Kereszt A, Li D, Indrasumunar A, et al (2007) *Agrobacterium rhizogenes*-mediated transformation of soybean to study root biology. *Nat Protoc* 2:948–952. <https://doi.org/10.1038/nprot.2007.141>
- Kereszt A, Mergaert P, Kondorosi E (2011) Bacteroid Development in Legume Nodules: Evolution of Mutual Benefit or of Sacrificial Victims? *Molecular Plant-Microbe Interactions*® 24:1300–1309. <https://doi.org/10.1094/MPMI-06-11-0152>
- Kohlen W, Ng JLP, Deinum EE, Mathesius U (2018) Auxin transport, metabolism, and signalling during nodule initiation: indeterminate and determinate nodules. *J Exp Bot* 69:229–244. <https://doi.org/10.1093/jxb/erx308>
- Kolde R (2015) Package ‘pheatmap.’ R package 1:790
- Kong Y, Schiestel K, Zheng Y (2019) Pure blue light effects on growth and morphology are slightly changed by adding low-level UVA or far-red light: A comparison with red light in four microgreen species. *Environ Exp Bot* 157:58–68. <https://doi.org/10.1016/j.envexpbot.2018.09.024>
- Krishnamurthy A, Ferl RJ, Paul AL (2018) Comparing RNA-Seq and microarray gene expression data in two zones of the *Arabidopsis* root apex relevant to spaceflight. *Appl Plant Sci* 6:e01197. <https://doi.org/10.1002/APS3.1197>
- Kulski JK (2016) Next-Generation Sequencing — An Overview of the History, Tools, and “Omic” Applications. In: *Next Generation Sequencing - Advances, Applications and Challenges*. InTech
- Kumar M, Chauhan AS, Kumar M, et al (2019) Transcriptome Sequencing of Chickpea (*Cicer arietinum* L.) Genotypes for Identification of Drought-Responsive Genes Under Drought Stress Condition. *Plant Mol Biol Report* 37:186–203. <https://doi.org/10.1007/s11105-019-01147-4>
- Kunert KJ, Vorster BJ, Fenta BA, et al (2016) Drought Stress Responses in Soybean Roots and Nodules. *Front Plant Sci* 7:. <https://doi.org/10.3389/fpls.2016.01015>
- Langfelder P, Horvath S (2008) WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* 9:. <https://doi.org/10.1186/1471-2105-9-559>
- Larrainzar E, Riely BK, Kim SC, et al (2015) Deep Sequencing of the *Medicago truncatula* Root Transcriptome Reveals a Massive and Early Interaction between Nodulation Factor and Ethylene Signals. *Plant Physiol* 169:233–265. <https://doi.org/10.1104/pp.15.00350>
- Larsson O, Tian B, Sonenberg N (2013) Toward a Genome-Wide Landscape of Translational Control. *Cold Spring Harb Perspect Biol* 5:a012302. <https://doi.org/10.1101/CSHPERSPECT.A012302>
- Lê S, Josse J, Rennes A, Husson F (2008) FactoMineR: An R Package for Multivariate Analysis
- Lee H, Hur C-G, Oh CJ, et al (2004) Analysis of the Root Nodule-enhanced Transcriptome in Soybean. *Mol Cells* 18:53–62. [https://doi.org/10.1016/S1016-8478\(23\)13081-0](https://doi.org/10.1016/S1016-8478(23)13081-0)

- Li M, Li H, Sun A, et al (2022) Transcriptome analysis reveals key drought-stress-responsive genes in soybean. *Front Genet* 13:. <https://doi.org/10.3389/fgene.2022.1060529>
- Li T, Zhang Y, Liu Y, et al (2020a) Raffinose synthase enhances drought tolerance through raffinose synthesis or galactinol hydrolysis in maize and *Arabidopsis* plants. *Journal of Biological Chemistry* 295:8064–8077. <https://doi.org/10.1074/jbc.RA120.013948>
- Li Y, Cai X, Gu W, Wang G (2020b) Transcriptome analysis of carotenoid biosynthesis in *Dunaliella salina* under red and blue light. *J Oceanol Limnol* 38:177–185. <https://doi.org/10.1007/s00343-019-9064-2>
- Lin C-T, Xu T, Xing S-L, et al (2019) Weighted Gene Co-expression Network Analysis (WGCNA) Reveals the Hub Role of Protein Ubiquitination in the Acquisition of Desiccation Tolerance in *Boea hygrometrica*. *Plant Cell Physiol* 60:2707–2719. <https://doi.org/10.1093/pcp/pcz160>
- Liu Y, Guo Z, Shi H (2022) Rhizobium Symbiosis Leads to Increased Drought Tolerance in Chinese Milk Vetch (*Astragalus sinicus* L.). *Agronomy* 12:725. <https://doi.org/10.3390/agronomy12030725>
- Liu Y, Li T, Zhang C, et al (2023) Raffinose positively regulates maize drought tolerance by reducing leaf transpiration. *The Plant Journal* 114:55–67. <https://doi.org/10.1111/tpj.16116>
- Lodeiro AR, González P, Hernández A, et al (2000) Comparison of drought tolerance in nitrogen-fixing and inorganic nitrogen-grown common beans. *Plant Science* 154:31–41. [https://doi.org/10.1016/S0168-9452\(99\)00246-0](https://doi.org/10.1016/S0168-9452(99)00246-0)
- Love MI, Anders S, Kim V, Huber W (2015) RNA-Seq workflow: gene-level exploratory analysis and differential expression. *F1000Res* 4:1070. <https://doi.org/10.12688/f1000research.7035.1>
- Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:. <https://doi.org/10.1186/s13059-014-0550-8>
- Luha R, Rana V, Vainstein A, Kumar V (2024) Nonsense-mediated mRNA decay pathway in plants under stress: general gene regulatory mechanism and advances. *Planta* 259:51. <https://doi.org/10.1007/s00425-023-04317-7>
- Lukoszek R, Feist P, Ignatova Z (2016) Insights into the adaptive response of *Arabidopsis thaliana* to prolonged thermal stress by ribosomal profiling and RNA-Seq. *BMC Plant Biol* 16:. <https://doi.org/10.1186/s12870-016-0915-0>
- Lumactud RA, Dollete D, Liyanage DK, et al (2023) The effect of drought stress on nodulation, plant growth, and nitrogen fixation in soybean during early plant growth. *J Agron Crop Sci* 209:345–354. <https://doi.org/10.1111/jac.12627>
- Maluk M, Giles M, Wardell GE, et al (2023) Biological nitrogen fixation by soybean (*Glycine max* [L.] Merr.), a novel, high protein crop in Scotland, requires inoculation with non-native bradyrhizobia. *Frontiers in Agronomy* 5:. <https://doi.org/10.3389/fagro.2023.1196873>
- Manna M, Thakur T, Chirom O, et al (2021) Transcription factors as key molecular target to strengthen the drought stress tolerance in plants. *Physiol Plant* 172:847–868. <https://doi.org/10.1111/ppl.13268>

- Manthey K, Krajinski F, Hohnjec N, et al (2004) Transcriptome Profiling in Root Nodules and Arbuscular Mycorrhiza Identifies a Collection of Novel Genes Induced During *Medicago truncatula* Root Endosymbioses. *Molecular Plant-Microbe Interactions*® 17:1063–1077. <https://doi.org/10.1094/MPMI.2004.17.10.1063>
- Marinho JP, Coutinho ID, da Fonseca Lameiro R, et al (2019) Metabolic alterations in conventional and genetically modified soybean plants with GmDREB2A;2 FL and GmDREB2A;2 CA transcription factors during water deficit. *Plant Physiology and Biochemistry* 140:122–135. <https://doi.org/10.1016/J.PLAPHY.2019.04.040>
- Mata J, Marguerat S, Bähler J (2005) Post-transcriptional control of gene expression: a genome-wide perspective. *Trends Biochem Sci* 30:506–514. <https://doi.org/10.1016/j.tibs.2005.07.005>
- Merchante C, Stepanova AN, Alonso JM (2017) Translation regulation in plants: an interesting past, an exciting present and a promising future. *The Plant Journal* 90:628–653. <https://doi.org/10.1111/tpj.13520>
- Meyer Y, Riondet C, Constans L, et al (2006) Evolution of redoxin genes in the green lineage. *Photosynth Res* 89:179–192. <https://doi.org/10.1007/S11120-006-9095-3/METRICS>
- Mhadhbi H, Bacem M, Chihaoui S-A, Moez J (2011) A highly osmotolerant rhizobial strain confers a better tolerance of nitrogen fixation and enhances protective activities to nodules of *Phaseolus vulgaris* under drought stress. *Afr J Biotechnol* 10:4555–4563
- Mittler R (2017) ROS Are Good. *Trends Plant Sci* 22:11–19. <https://doi.org/10.1016/j.tplants.2016.08.002>
- Mohammadi-Dehcheshmeh M, Ebrahimie E, Tyerman SD, Kaiser BN (2014) A novel method based on combination of semi-in vitro and in vivo conditions in *Agrobacterium rhizogenes*-mediated hairy root transformation of Glycine species. *In Vitro Cellular and Developmental Biology - Plant* 50:282–291. <https://doi.org/10.1007/s11627-013-9575-z>
- Moin M, Bakshi A, Madhav MS, Kirti PB (2017) Expression profiling of ribosomal protein gene family in dehydration stress responses and characterization of transgenic rice plants overexpressing RPL23A for water-use efficiency and tolerance to drought and salt stresses. *Front Chem* 5:. <https://doi.org/10.3389/fchem.2017.00097>
- Moloi SJ, Ngara R (2023) The roles of plant proteases and protease inhibitors in drought response: a review. *Front Plant Sci* 14:.. <https://doi.org/10.3389/fpls.2023.1165845>
- Moore M, Gossmann N, Dietz KJ (2016) Redox Regulation of Cytosolic Translation in Plants. *Trends Plant Sci* 21:388–397
- Nakashima K, Tran LP, Van Nguyen D, et al (2007) Functional analysis of a NAC-type transcription factor OsNAC6 involved in abiotic and biotic stress-responsive gene expression in rice. *The Plant Journal* 51:617–630. <https://doi.org/10.1111/j.1365-313X.2007.03168.x>
- Nakei MD, Venkataramana PB, Ndakidemi PA (2022) Soybean-Nodulating Rhizobia: Ecology, Characterization, Diversity, and Growth Promoting Functions. *Front Sustain Food Syst* 6:.. <https://doi.org/10.3389/fsufs.2022.824444>

- Nath Radhamony R, Mohan Prasad A, Srinivasan R (2005) T-DNA insertional mutagenesis in Arabidopsis: a tool for functional genomics
- Niu J, Zhao J, Guo Q, et al (2024) WGCNA Reveals Hub Genes and Key Gene Regulatory Pathways of the Response of Soybean to Infection by Soybean mosaic virus. *Genes (Basel)* 15:566. <https://doi.org/10.3390/genes15050566>
- OECD/FAO (2021) OECD-FAO Agricultural Outlook 2021-2030. OECD
- Oldroyd GED, Murray JD, Poole PS, Downie JA (2011) The Rules of Engagement in the Legume-Rhizobial Symbiosis. *Annu Rev Genet* 45:119–144. <https://doi.org/10.1146/annurev-genet-110410-132549>
- O'Rourke JA, Iniguez LP, Fu F, et al (2014) An RNA-Seq based gene expression atlas of the common bean. *BMC Genomics* 15:866. <https://doi.org/10.1186/1471-2164-15-866>
- Osakabe Y, Osakabe K, Shinozaki K, Tran L-SP (2014) Response of plants to water stress. *Front Plant Sci* 5:. <https://doi.org/10.3389/fpls.2014.00086>
- Pareek A, Sudhir ·, Sopory K, et al (2010) Abiotic Stress Adaptation in Plants Physiological, Molecular and Genomic Foundation
- Parthasarathy S (2023) Optimization of the hairy-root transformation in soybean to facilitate bioassays with the root pathogen *Phytophthora sojae*. Québec, Canada
- Patro R, Duggal G, Love MI, et al (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* 2017 14:4 14:417–419. <https://doi.org/10.1038/nmeth.4197>
- Patro R, Mount SM, Kingsford C (2014) Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nat Biotechnol* 32:462–464. <https://doi.org/10.1038/nbt.2862>
- Pellizzaro A, Alibert B, Planchet E, et al (2017) Nitrate transporters: an overview in legumes. *Planta* 246:585–595. <https://doi.org/10.1007/s00425-017-2724-6>
- Peng X, Tang X, Zhou P, et al (2011) Isolation and Expression Patterns of Rice WRKY82 Transcription Factor Gene Responsive to Both Biotic and Abiotic Stresses. *Agric Sci China* 10:893–901. [https://doi.org/10.1016/S1671-2927\(11\)60074-6](https://doi.org/10.1016/S1671-2927(11)60074-6)
- QIU Y, YU D (2009) Over-expression of the stress-induced OsWRKY45 enhances disease resistance and drought tolerance in Arabidopsis. *Environ Exp Bot* 65:35–47. <https://doi.org/10.1016/j.envexpbot.2008.07.002>
- Rana M, Han Z-X, Song D-P, et al (2016) Effect of Medium Supplements on Agrobacterium rhizogenes Mediated Hairy Root Induction from the Callus Tissues of *Camellia sinensis* var. *sinensis*. *Int J Mol Sci* 17:1132. <https://doi.org/10.3390/ijms17071132>
- Ratcliffe OJ, Riechmann JL (2002) Arabidopsis Transcription Factors and the Regulation of Flowering Time: A Genomic Perspective. *Curr Issues Mol Biol*. <https://doi.org/10.21775/cimb.004.077>
- Raxwal VK, Riha K (2023) The biological functions of nonsense-mediated mRNA decay in plants: RNA quality control and beyond. *Biochem Soc Trans* 51:31–39. <https://doi.org/10.1042/BST20211231>

- Reynoso MA, Blanco FA, Bailey-Serres J, et al (2013) Selective recruitment of mRNAs and miRNAs to polyribosomes in response to rhizobia infection in *Medicago truncatula*. *The Plant Journal* 73:289–301. <https://doi.org/10.1111/TPJ.12033>
- Ribone AI, Fass M, Gonzalez S, et al (2023) Co-Expression Networks in Sunflower: Harnessing the Power of Multi-Study Transcriptomic Public Data to Identify and Categorize Candidate Genes for Fungal Resistance. *Plants* 12:. <https://doi.org/10.3390/plants12152767>
- Rong Li, Chen H, Yang Z, et al (2020) Research status of soybean symbiosis nitrogen fixation. *Oil Crop Science* 5:6–10. <https://doi.org/10.1016/j.ocsci.2020.03.005>
- Rösti J, Barton CJ, Albrecht S, et al (2007) UDP-Glucose 4-Epimerase Isoforms UGE2 and UGE4 Cooperate in Providing UDP-Galactose for Cell Wall Biosynthesis and Growth of *Arabidopsis thaliana*. *Plant Cell* 19:1565–1579. <https://doi.org/10.1105/tpc.106.049619>
- Rushton PJ, Somssich IE, Ringler P, Shen QJ (2010) WRKY transcription factors. *Trends Plant Sci* 15:247–258. <https://doi.org/10.1016/j.tplants.2010.02.006>
- Sablok G, Powell JJ, Kazan K (2017) Emerging Roles and Landscape of Translating mRNAs in Plants. *Front Plant Sci* 8:. <https://doi.org/10.3389/fpls.2017.01443>
- Sainsbury F, Thuenemann EC, Lomonosoff GP (2009) pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol J* 7:682–693. <https://doi.org/10.1111/j.1467-7652.2009.00434.x>
- Sainz MM, Filippi CV, Eastman G, et al (2022a) Analysis of Thioredoxins and Glutaredoxins in Soybean: Evidence of Translational Regulation under Water Restriction. *Antioxidants* 11:1622. <https://doi.org/10.3390/antiox11081622>
- Sainz MM, Filippi CV, Eastman G, et al (2022b) Polysome Purification from Soybean Symbiotic Nodules. *Journal of Visualized Experiments*. <https://doi.org/10.3791/64269>
- Sainz MM, Filippi C V, Eastman G, et al (2024a) Water deficit response in nodulated soybean roots: a comprehensive transcriptome and translome network analysis. *BMC Plant Biol*. <https://doi.org/10.1186/s12870-024-05280-5>
- Sainz MM, Sotelo-Silveria M, Filippi C V, Zardo S (2024b) Legume-rhizobia symbiosis: Translatome analysis. *Genet Mol Biol*. <https://doi.org/10.1590/1678-4685-GMB-2023-0284>
- Sakamoto K, Ogiwara N, Kaji T, et al (2019) Transcriptome analysis of soybean (*Glycine max*) root genes differentially expressed in rhizobial, arbuscular mycorrhizal, and dual symbiosis. *J Plant Res* 132:541–568. <https://doi.org/10.1007/s10265-019-01117-7>
- Sánchez-Baizán N, Ribas L, Piferrer F (2022) Improved biomarker discovery through a plot twist in transcriptomic data analysis. *BMC Biol* 20:208. <https://doi.org/10.1186/s12915-022-01398-w>
- Sánchez-Bermúdez M, del Pozo JC, Pernas M (2022) Effects of Combined Abiotic Stresses Related to Climate Change on Root Growth in Crops. *Front Plant Sci* 13:. <https://doi.org/10.3389/fpls.2022.918537>

- Sańko-Sawczenko I, Łotocka B, Mielecki J, et al (2019) Transcriptomic Changes in *Medicago truncatula* and *Lotus japonicus* Root Nodules during Drought Stress. *Int J Mol Sci* 20:1204. <https://doi.org/10.3390/ijms20051204>
- Sarker U, Oba S (2018) Catalase, superoxide dismutase and ascorbate-glutathione cycle enzymes confer drought tolerance of *Amaranthus tricolor*. *Sci Rep* 8:16496. <https://doi.org/10.1038/s41598-018-34944-0>
- Sayers EW, Bolton EE, Brister JR, et al (2022) Database resources of the national center for biotechnology information. *Nucleic Acids Res* 50:D20–D26. <https://doi.org/10.1093/nar/gkab1112>
- Schmutz J, Cannon SB, Schlueter J, et al (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463:178–183. <https://doi.org/10.1038/nature08670>
- Seo CH, Kim J-R, Kim M-S, Cho K-H (2009) Hub genes with positive feedbacks function as master switches in developmental gene regulatory networks. *Bioinformatics* 25:1898–1904. <https://doi.org/10.1093/bioinformatics/btp316>
- Sferra G, Fantozzi D, Scippa GS, Trupiano D (2023) Key Pathways and Genes of *Arabidopsis thaliana* and *Arabidopsis halleri* Roots under Cadmium Stress Responses: Differences and Similarities. *Plants* 12:1793. <https://doi.org/10.3390/plants12091793>
- Shahriari AG, Soltani Z, Tahmasebi A, Pocza P (2022) Integrative System Biology Analysis of Transcriptomic Responses to Drought Stress in Soybean (*Glycine max* L.). *Genes (Basel)* 13:1732. <https://doi.org/10.3390/genes13101732>
- Shamimuzzaman Md, Vodkin L (2018) Ribosome profiling reveals changes in translational status of soybean transcripts during immature cotyledon development. *PLoS One* 13:e0194596. <https://doi.org/10.1371/journal.pone.0194596>
- Shao H-B, Chu L-Y, Jaleel CA, et al (2009) Understanding water deficit stress-induced changes in the basic metabolism of higher plants – biotechnologically and sustainably improving agriculture and the ecoenvironment in arid regions of the globe. *Crit Rev Biotechnol* 29:131–151. <https://doi.org/10.1080/07388550902869792>
- Shen Y-G, Zhang W-K, Yan D-Q, et al (2003) Characterization of a DRE-binding transcription factor from a halophyte *Atriplex hortensis*. *Theoretical and Applied Genetics* 107:155–161. <https://doi.org/10.1007/s00122-003-1226-z>
- Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nat Biotechnol* 26:1135–1145. <https://doi.org/10.1038/nbt1486>
- Sheng Q, Vickers K, Zhao S, et al (2016) Multi-perspective quality control of Illumina RNA sequencing data analysis. *Brief Funct Genomics* elw035. <https://doi.org/10.1093/bfpg/elw035>
- Sheshukova E V., Komarova T V., Ershova NM, et al (2017) An alternative nested reading frame may participate in the stress-dependent expression of a plant gene. *Front Plant Sci* 8:. <https://doi.org/10.3389/fpls.2017.02137>

- Shi Y, Su Z, Yang H, et al (2019) Alternative splicing coupled to nonsense-mediated mRNA decay contributes to the high-altitude adaptation of maca (*Lepidium meyenii*). *Gene* 694:7–18. <https://doi.org/10.1016/j.gene.2018.12.082>
- Shinozaki K, Yamaguchi-Shinozaki K (2006) Gene networks involved in drought stress response and tolerance. *J Exp Bot* 58:221–227. <https://doi.org/10.1093/jxb/erl164>
- Shulze CN, Cole BJ, Ciobanu D, et al (2019) High-Throughput Single-Cell Transcriptome Profiling of Plant Cell Types. *Cell Rep* 27:2241–2247.e4. <https://doi.org/10.1016/J.CELREP.2019.04.054>
- Sibout R, Proost S, Hansen BO, et al (2017) Expression atlas and comparative coexpression network analyses reveal important genes involved in the formation of lignified cell wall in *Brachypodium distachyon*. *New Phytologist* 215:1009–1025. <https://doi.org/10.1111/nph.14635>
- Silva AT, Ribone PA, Chan RL, et al (2016) A Predictive Coexpression Network Identifies Novel Genes Controlling the Seed-to-Seedling Phase Transition in *Arabidopsis thaliana*. *Plant Physiol* 170:2218–2231. <https://doi.org/10.1104/pp.15.01704>
- Singh KB (1998) Transcriptional Regulation in Plants: The Importance of Combinatorial Control
- Smoot ME, Ono K, Ruscheinski J, et al (2011) Cytoscape: new features for data integration and network visualization. *Bioinformatics* 27:431–432. <https://doi.org/10.1093/bioinformatics/btq675>
- Solodushko V, Fouty B (2023) Terminal hairpins improve protein expression in IRES-initiated mRNA in the absence of a cap and polyadenylated tail. *Gene Ther* 30:620–627. <https://doi.org/10.1038/s41434-023-00391-4>
- Song L, Prince S, Valliyodan B, et al (2016) Genome-wide transcriptome analysis of soybean primary root under varying water-deficit conditions. *BMC Genomics* 17:57. <https://doi.org/10.1186/s12864-016-2378-y>
- Srivastava A, Malik L, Sarkar H, et al (2020) Alignment and mapping methodology influence transcript abundance estimation. *Genome Biol* 21:239. <https://doi.org/10.1186/s13059-020-02151-8>
- Staudinger C, Mehmeti-Tershani V, Gil-Quintana E, et al (2016) Evidence for a rhizobia-induced drought stress response strategy in *Medicago truncatula*. *J Proteomics* 136:202–213. <https://doi.org/10.1016/j.jprot.2016.01.006>
- Subramanian S, Stacey G, Yu O (2007) Distinct, crucial roles of flavonoids during legume nodulation. *Trends Plant Sci* 12:282–285. <https://doi.org/10.1016/j.tplants.2007.06.006>
- Sureshkumar S, Dent C, Seleznev A, et al (2016) Nonsense-mediated mRNA decay modulates FLM-dependent thermosensory flowering response in *Arabidopsis*. *Nat Plants* 2:16055. <https://doi.org/10.1038/nplants.2016.55>
- Szklarczyk D, Gable AL, Nastou KC, et al (2021) The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* 49:D605–D612. <https://doi.org/10.1093/nar/gkaa1074>

- Tang Y, Li J, Song Q, et al (2023) Transcriptome and WGCNA reveal hub genes in sugarcane tiller seedlings in response to drought stress. *Sci Rep* 13:12823. <https://doi.org/10.1038/s41598-023-40006-x>
- Tebaldi T, Re A, Viero G, et al (2012) Widespread uncoupling between transcriptome and translome variations after a stimulus in mammalian cells. *BMC Genomics* 13:1–15. <https://doi.org/10.1186/1471-2164-13-220/FIGURES/4>
- Törönen P, Holm L (2022) PANNZER—A practical tool for protein function prediction. *Protein Science* 31:118–128. <https://doi.org/10.1002/pro.4193>
- Traubenik S, Reynoso MA, Hobecker K, et al (2020) Reprogramming of Root Cells during Nitrogen-Fixing Symbiosis Involves Dynamic Polysome Association of Coding and Noncoding RNAs. *Plant Cell* 32:352–373. <https://doi.org/10.1105/TPC.19.00647>
- Tungalag S, Shinriki S, Hirayama M, et al (2023) Ribosome profiling analysis reveals the roles of DDX41 in translational regulation. *Int J Hematol* 117:876–888. <https://doi.org/10.1007/s12185-023-03558-2>
- Turetschek R, Staudinger C, Wienkoop S (2020) Rhizobial symbiosis influences response to early salt and drought stress of the *Medicago truncatula* root proteome. In: *The Model Legume Medicago truncatula*. Wiley, pp 253–260
- Tyagi K, Shukla P, Rohela GK, et al (2020) Plant Phenolics: Their Biosynthesis, Regulation, Evolutionary Significance, and Role in Senescence. In: *Plant Phenolics in Sustainable Agriculture*. Springer Singapore, Singapore, pp 431–449
- Van Heerden PDR, De Beer M, Mellet DJ, et al (2007) Growth media effects on shoot physiology, nodule numbers and symbiotic nitrogen fixation in soybean. *South African Journal of Botany* 73:600–605. <https://doi.org/10.1016/j.sajb.2007.05.009>
- Vannini C, Campa M, Iriti M, et al (2007) Evaluation of transgenic tomato plants ectopically expressing the rice *Osmyb4* gene. *Plant Science* 173:231–239. <https://doi.org/10.1016/j.plantsci.2007.05.007>
- Vincent J (1971) A manual for the practical study of the root-nodule bacteria. *BIBP Handbk 15* Oxford and Edinburgh: Blackwell Scientific Publications
- Vogel C, Marcotte EM (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* 13:227–232. <https://doi.org/10.1038/nrg3185>
- Wan L, Li Y, Li S, Li X (2022) Transcriptomic Profiling Revealed Genes Involved in Response to Drought Stress in Alfalfa. *J Plant Growth Regul* 41:92–112. <https://doi.org/10.1007/s00344-020-10287-x>
- Wang ET, Tian CF, Chen WF, et al (2019) *Ecology and Evolution of Rhizobia*. Springer Singapore, Singapore
- Wang T, Cui Y, Jin J, et al (2013) Translating mRNAs strongly correlate to proteins in a multivariate manner and their translation ratios are phenotype specific. *Nucleic Acids Res* 41:4743–4754. <https://doi.org/10.1093/NAR/GKT178>
- Wang X, Wu Z, Zhou Q, et al (2022) Physiological Response of Soybean Plants to Water Deficit. *Front Plant Sci* 12:. <https://doi.org/10.3389/fpls.2021.809692>

- Wang Y, Zhang Z, Zhang P, et al (2016) Rhizobium symbiosis contribution to short-term salt stress tolerance in alfalfa (*Medicago sativa* L.). *Plant Soil* 402:247–261. <https://doi.org/10.1007/s11104-016-2792-6>
- Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10:57–63
- Wani SH, Singh NB, Devi TR, et al (2013) Engineering Abiotic Stress Tolerance in Plants: Extricating Regulatory Gene Complex
- Wickham H (2016) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York
- Wink M (2013) Evolution of secondary metabolites in legumes (Fabaceae). *South African Journal of Botany* 89:164–175. <https://doi.org/10.1016/j.sajb.2013.06.006>
- Wu Y, Wang Y, Shi H, et al (2022) Time-course transcriptome and WGCNA analysis revealed the drought response mechanism of two sunflower inbred lines. *PLoS One* 17:e0265447. <https://doi.org/10.1371/journal.pone.0265447>
- Xiang T (2016) Cucumopine type *Agrobacterium* rhizogenes K599 (NCPB2659) T-DNA mediated plant transformation and its application. *Bangladesh J Bot*
- Xiao X, Fan M, Wang E, et al (2017) Interactions of plant growth-promoting rhizobacteria and soil factors in two leguminous plants. *Appl Microbiol Biotechnol* 101:8485–8497. <https://doi.org/10.1007/s00253-017-8550-8>
- Yadav S, Yadava YK, Meena S, et al (2024) Novel insights into drought-induced regulation of ribosomal genes through DNA methylation in chickpea. *Int J Biol Macromol* 266:. <https://doi.org/10.1016/j.ijbiomac.2024.131380>
- Yin X, Cui Y, Wang M, Xia X (2017) Overexpression of a novel MYB-related transcription factor, OsMYBR1, confers improved drought tolerance and decreased ABA sensitivity in rice. *Biochem Biophys Res Commun* 490:1355–1361. <https://doi.org/10.1016/j.bbrc.2017.07.029>
- Yoshida T, Mogami J, Yamaguchi-Shinozaki K (2014) ABA-dependent and ABA-independent signaling in response to osmotic stress in plants. *Curr Opin Plant Biol* 21:133–139. <https://doi.org/10.1016/j.pbi.2014.07.009>
- Yu T, Zhuang Q (2020) Modeling biological nitrogen fixation in global natural terrestrial ecosystems. *Biogeosciences* 17:3643–3657. <https://doi.org/10.5194/bg-17-3643-2020>
- Yuan S, Li R, Chen S, et al (2016) RNA-Seq Analysis of Differential Gene Expression Responding to Different Rhizobium Strains in Soybean (*Glycine max*) Roots. *Front Plant Sci* 7:. <https://doi.org/10.3389/fpls.2016.00721>
- Zardo S (2021) Aproximación metodológica para el estudio del control traduccional en plantas de soja noduladas en respuesta a sequía. *Facultad de Ciencias*
- Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 4:. <https://doi.org/10.2202/1544-6115.1128>

- Zhang Z, Zhang H, Quan R, et al (2009) Transcriptional Regulation of the Ethylene Response Factor LeERF2 in the Expression of Ethylene Biosynthesis Genes Controls Ethylene Production in Tomato and Tobacco . *Plant Physiol* 150:365–377. <https://doi.org/10.1104/pp.109.135830>
- Zhao X, Zhang Y, Wang J, et al (2024) GWAS and WGCNA Analysis Uncover Candidate Genes Associated with Oil Content in Soybean. *Plants* 13:1351. <https://doi.org/10.3390/plants13101351>
- Zhou Q, Su X, Jing G, et al (2018) RNA-QC-chain: comprehensive and fast quality control for RNA-Seq data. *BMC Genomics* 19:144. <https://doi.org/10.1186/s12864-018-4503-6>
- Zhu JK (2016a) Abiotic Stress Signaling and Responses in Plants. *Cell* 167:313–324
- Zhu JK (2002a) Salt and drought stress signal transduction in plants. *Annu Rev Plant Biol* 53:247–273
- Zhu JK (2002b) Salt and drought stress signal transduction in plants. *Annu Rev Plant Biol* 53:247–273
- Zhu J-K (2016b) Abiotic Stress Signaling and Responses in Plants. *Cell* 167:313–324. <https://doi.org/10.1016/j.cell.2016.08.029>
- Zong J, Chen P, Luo Q, et al (2023) Transcriptome-Based WGCNA Analysis Reveals the Mechanism of Drought Resistance Differences in Sweetpotato (*Ipomoea batatas* (L.) Lam.). *Int J Mol Sci* 24:14398. <https://doi.org/10.3390/ijms241814398>

7. ANEXO

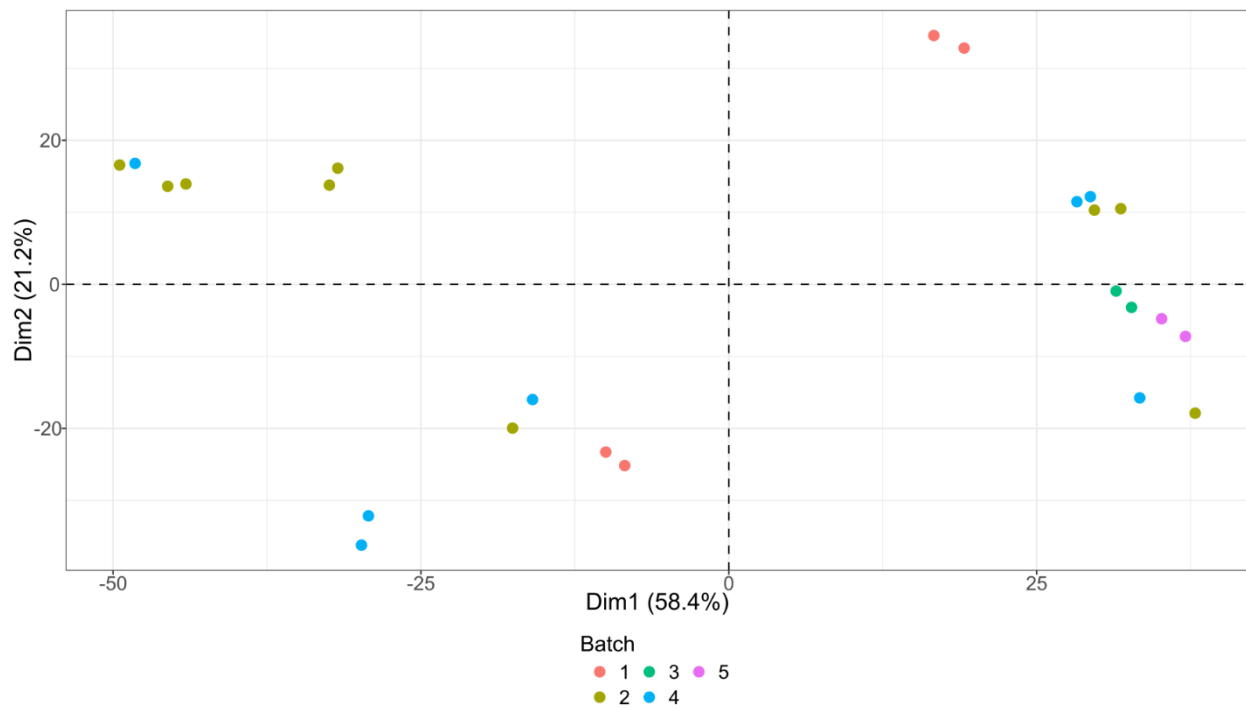


Figura S1 Análisis de Componentes Principales. Dim1 y Dim2 refieren a las dimensiones 1 y 2, respectivamente. Batch corresponde a los diferentes lotes de secuenciación.

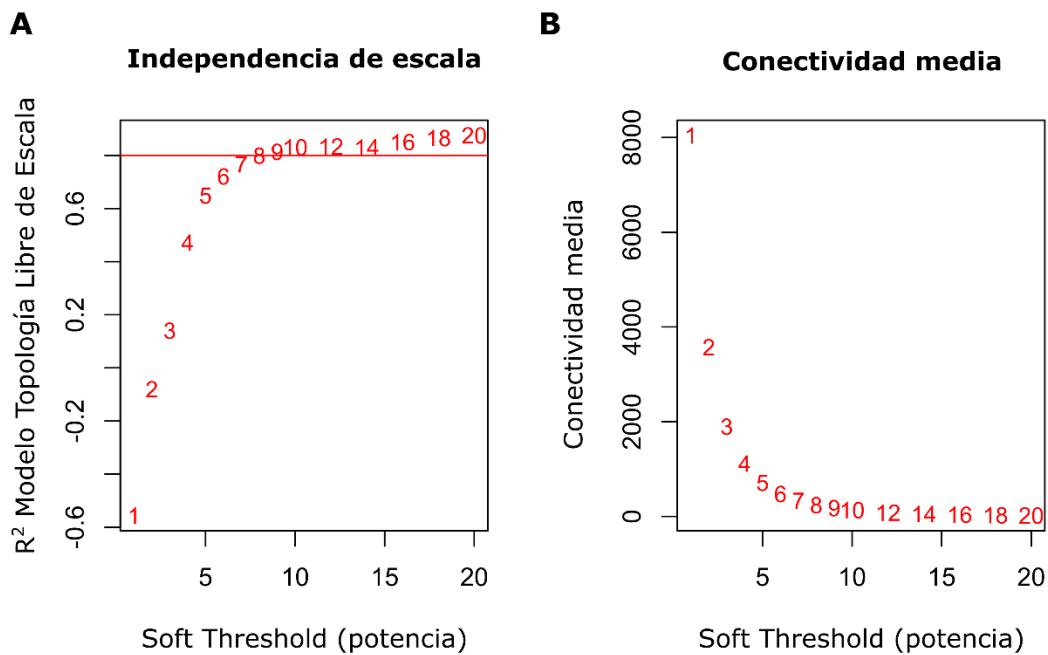


Figura S2 Selección del valor de potencia para la función de adyacencia en la construcción de módulos de co-expresión génica. **A** Valor del R² en función del valor de potencia utilizado para calcular la adyacencia entre genes. Valores por encima de 0,8 satisfacen el criterio de topología libre de escala. **B** Número de genes promedio por módulo en función del valor de potencia.

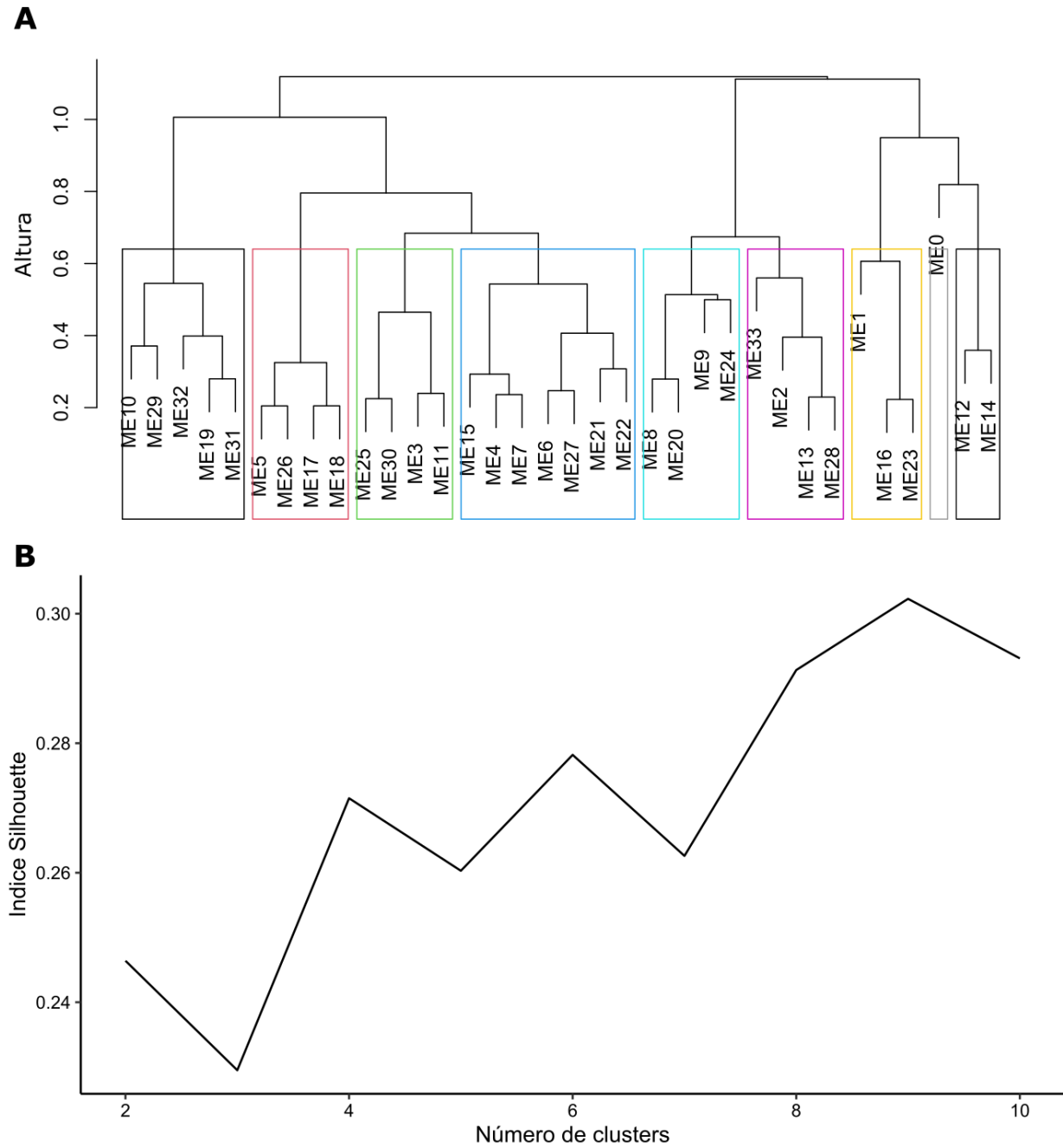


Figura S3 Clusterización jerárquica de *eigengenes* de módulos (ME) de red WGCNA. A Distancia entre ME calculada siguiendo el manual del paquete de R WGCNA (1 – correlación entre MEs). Criterio de clusterización utilizado para la creación del dendrograma: enlace promedio. Selección de número de clusters mediante índice *Silhouette* (Dudek 2020). B *Silhouette* plot. Índice *silhouette* calculado usando el paquete de R *NbClust* (Charrad et al. 2014).

Tabla S1 Cantidad de genes expresados diferencialmente a nivel transcriptómico (TOTAL) y/o traductómico (PAR) comunes a déficit hídrico -contrastes i) y iii)- y a nodulación -contrastes ii) y iv)- de acuerdo a su localización en módulos de co-expresión génica (M) de la red de WGCNA.

Contrastes	Estado	Nivel de regulación	M0	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18	M19	M20	M21	M22	M23	M24	M25	M26	M27	M28	M29	M30	M31	M32	M33	
<div><div><div>ij)</div><div>N+RH vs N+CC y</div></div><div><div>iii)</div><div>NN+RH vs NN+CC</div></div></div>	Down	TOTAL	0	0	0	0	17	0	0	7	2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	8 (5)	0	0	1	0	0	0	0	0	2	0	0
		PAR	0	0	0	0	5 (1)	0	0	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
		TOTAL+PAR	0	0	0	0	144 (84)	0	0	50 (21)	4 (2)	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	0	0	0	0	1	0	0	0	2 (1)	1	0
	Up	TOTAL	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		PAR	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		TOTAL+PAR	0	0	0	0	34 (14)	0	0	1	16 (7)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
<div><div><div>ij)</div><div>N+RH vs NN+RH y</div></div><div><div>iv)</div><div>N+CC vs NN+CC</div></div></div>	Downregulated	TOTAL	0	0	0	0	2 (1)	0	1	2	0	0	2	0	0	0	0	2	0	0	0	0	0	0	1 (1)	0	0	0	0	0	0	0	0	0	0	0	
		PAR	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		TOTAL+PAR	0	0	0	0	4 (4)	0	0	9 (4)	0	0	5	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	Upregulated	TOTAL	0	0	0	0	2	0	0	1	1	0	6 (1)	0	0	0	0	0	0	0	0	8 (1)	0	0	0	3	0	0	0	0	0	1	0	0	0	0	
		PAR	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
		TOTAL+PAR	0	0	0	0	1	0	0	0	0	0	113 (63)	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Tabla S2 Cantidad de genes expresados diferencialmente a nivel transcriptómico (TOTAL) y/o traductómico (PAR) en cada contraste de acuerdo a su localización en módulos de co-expresión génica (M) de la red de WGCNA

Contra ste	Estado	Nivel de regulación	M 0	M1	M2	M 3	M4	M 5	M 6	M7	M8	M9	M1 0	M1 1	M1 2	M1 3	M1 4	M1 5	M1 6	M17	M1 8	M19	M20	M21	M2 2	M2 3	M2 4	M2 5	M2 6	M2 7	M2 8	M2 9	M3 0	M3 1	M3 2	M 33	
i) N+RH vs N+CC	Down	TOTAL	0	8	13	3	60 (2)	3	2	21 (1)	19	3	6	0	1	2	0	2	2	2	1	0	8 (1)	15 (1)	5	11 (2)	0	9 (1)	2	1	2 (1)	3 (1)	0	1	0	1	
		PAR	0	2	8 (1)	4	70 (4)	0	0	19 (6)	11	6	2	0	1	0	2	5 (3)	0	0	0	0	8	13 (2)	11 (2)	1	2	3	0	0	0	0	0	5	0	0	
		TOTAL+PAR	0	3	10	3	240 (57)	0	1	43 (19)	20 (3)	5 (1)	1	0	0	1	0	5	0	0	0	0	21 (1)	39 (18)	29 (8)	1	0	6 (3)	0	6	0	1	0	2	0	0	
	Up	TOTAL	0	4	6	3	49	0	0	2	33 (5)	2	1	0	1	0	0	2	0	0	0	0	6 (1)	0	0	2	0	6	3 (1)	0	9 (4)	0	1	0	0	0	
		PAR	0	228 (99)	11	9	63 (1)	0	0	0	22 (5)	14 (1)	2	0	0	3	3	6 (1)	0	0	0	0	18	3	3	0	2	6 (1)	0	0	5	1	1	0	1	2	
		TOTAL+PAR	0	31 (1)	12	4	158 (20)	0	1	3	120 (42)	0	0	0	0	0	0	3 (1)	0	0	0	0	51 (23)	1	2 (1)	1	0	6 (3)	4 (1)	0	3 (2)	0	0	0	0	0	
ii) N+RH vs NN+R H	Down	TOTAL	0	5	1	1	82 (31)	1	1	16 (2)	6 (1)	2	2	0	5 (1)	0	1	6	1	0	0	0	1	0	0	9 (1)	0	0	0	0	0	0	0	0	0		
		PAR	0	2	0	0	14 (5)	0	1	6 (1)	0	1	7	0	0	0	0	4 (1)	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0		
		TOTAL+PAR	0	1	0	1	126 (61)	0	1	35 (8)	2 (1)	0	8	0	0	0	0	12 (4)	0	0	0	0	0	0	0	4 (2)	0	0	0	0	0	0	0	0	1	0	
	Up	TOTAL	0	14 (1)	2	3	28 (4)	0	1	2 (1)	6	1	11	0	0	0	1	1	0	0	2	27 (11)	1 (1)	0	0	1	0	0	0	0	0	4	0	0	1	0	
		PAR	0	2	0	0	14 (3)	0	0	1	3 (1)	0	5	0	1	0	0	2 (2)	0	0	0	7 (1)	0	0	0	0	1	0	0	0	0	0	0	0	1	0	
		TOTAL+PAR	0	4	0	0	31 (9)	1	0	4	7 (1)	0	16	0	0	0	0	7	0	1	1	11 (4)	1	0	0	1 (1)	0	0	0	0	0	3	0	0	20 (6)	0	
iii) NN+R H vs NN+C C	Down	TOTAL	0	4	2	1	9	0	1	25 (2)	4	0	3	0	0	1	0	3	3 (1)	0	2	0	2	0	0	8 (2)	0	0	3	0	1	0	0	5	2	0	
		PAR	0	0	2	0	4	0	0	15 (3)	2	0	1	0	0	0	0	0	0	0	3	1	0	0	0	1	2	1	0	1	0	0	0	1	0	0	
		TOTAL+PAR	0	0	2	0	8	1	0	51 (10)	0	0	4	0	0	1	0	0	1	1	8	0	0	0	0	4 (1)	2	0	0	0	0	0	0	1	17 (6)	0	
	Up	TOTAL	0	3	1	1	6	0	0	7	3	1	6	0	0	0	0	3	3	12	23 (8)	0	0	0	1	3	16 (1)	0	0	0	0	0	0	0	0	1	0
		PAR	0	6	1	0	2	1	0	1	7	0	3	0	0	0	0	0	0	2 (1)	5 (2)	0	2	0	0	1	3	0	0	0	0	0	0	0	0	0	
		TOTAL+PAR	0	1	0	0	2	0	0	3	2	0	3	0	0	0	0	0	0	8 (1)	16 (8)	0	0	0	0	3 (2)	17 (9)	0	0	0	0	0	0	0	0	1	0
iv) N+CC vs NN+CC	Down	TOTAL	0	5	4	4	7	1	3	12 (4)	6 (1)	1	9	0	0	3	0	1	2	0	2	0	3 (1)	4	0	4 (1)	2	2	1	0	13 (4)	2	0	0	0	0	
		PAR	0	4	9 (1)	5	0	2	0	17 (1)	1	0	5	0	0	10 (1)	0	0	3	0	0	0	2	3	1	7 (2)	0	0	1	0	6 (1)	0	0	0	0	0	
		TOTAL+PAR	0	3	11	8	4	0	1	28 (10)	4 (1)	0	4	0	0	16 (1)	0	3	3	0	3	0	2	1	0	2 (1)	0	0	0	0	12 (3)	4 (2)	0	0	0	0	
	Up	TOTAL	0	6	7	0	6	7	2	5	2	0	19	0	1	4	0	7	2	0	6	8	1	22 (1)	7 (1)	1	4	3	0	0	3	10 (4)	0	0	0	0	
		PAR	0	1	3	3	1	1	0	2	0	0	19	1	0	2	0	0	4	2	3	5	0	8 (2)	2	1	2	0	0	0	0	0	0	0	0	0	
		TOTAL+PAR	0	1	9	4	0	2	1	7	0	1	52	0	0	2	0	0	4	0	0	2	0	53 (19)	2	3 (1)	1	1	0	0	6 (2)	1	0	0	0	0	

Tabla S3 Lista de genes imputados para los procesos biológicos (PB) de la *Gene Ontology* (GO) de traducción y nodulación mediante EGAD.

GlymaID	Descripción	GO-PB Anotado	GO-PB Imputado
Glyma_07G085900	LOC100790976 no caracterizado	organización membrana tilacoidal	traducción
Glyma_19G072000	LOC100499753 no caracterizado	inserción de proteína en membrana mitocondrial	
Glyma_17G196800	factor de splicing tipo 7	splicing ARNm via spliceosoma	
Glyma_18G078700	proteína ribosomal L19-1 50S	biosíntesis de ATP	
Glyma_13G051500	alfa/beta hidrolasa	proteólisis	
Glyma_18G007100	correpresor dr1	regulación de la transcripción	
Glyma_19G151400	histona H2AX	metabolismo dolichol; manosiación de proteína	
Glyma_08G151900	proteína ribosomal L6 60S	ensamblaje subunidad ribosomal mayor	
Glyma_09G185900	aspartil-proteasa APCB1	proteólisis; defensa contra hongos	
Glyma_16G159900	LOC100527783 no caracterizado	plegado proteico	
Glyma_01G167300	proteína ribosomal L8-1 60S	respuesta a auxina	
Glyma_02G196900	proteína NSA2 biogénesis ribosomal	procesamiento ARNr	
Glyma_08G222100	proteína ribosomal L7-4 60S	maduración de ARNr a partir de transcripto tricistrónico	
Glyma_12G028500	proteína ribosomal L7a-2 60S	biogénesis ribosomal	
Glyma_07G214900	proteína ribosomal S16 40S	maduración de ARNr a partir de transcripto tricistrónico	
Glyma_17G095400	proteína NIP7 biogénesis ribosomal	ensamblaje ribosomal	
Glyma_08G014100	quinasa BSK7	señalización mediada por brassinosteroides	
Glyma_12G053500	ligasa E3 ARI2	ubiquitinación proteica	
Glyma_19G150400	regulador de complejo MON1-CCZ1	regulación de la autofagia	
Glyma_01G114000	proteína transportadora de auxina	señalización activada por auxinas	
Glyma_15G167400	proteína ribosomal L7-4 60S	maduración de ARNr a partir de transcripto tricistrónico	
Glyma_19G155700	Rho GTPasa mitocondrial	transducción de señal mediada por GTP; respuesta a ABA	
Glyma_07G028400	2-metoxi-6-poliprenil-1,4-benzoquinol metilasa	metilación	
Glyma_05G241200	proteína ribosomal S16 40S	maduración de ARNr a partir de transcripto tricistrónico	
Glyma_09G222300	proteína tipo 10A5 inducida por auxina	respuesta a auxina	
Glyma_19G208900	factor GAI-associated dedo de zing	regulación de la transcripción	
Glyma_07G020000	proteína ribosomal L30/L7	maduración de ARNr a partir de transcripto tricistrónico	
Glyma_13G214600	proteína carrier acil putativa	biosíntesis de ácidos grasos	

Glyma_05G045200	quinasa nucleósido difosfato	biosíntesis UTP, GTP y CTP	
Glyma_13G347400	subunidad TIM23-2 de translocasa mitocondrial	transporte transmembrana	
Glyma_08G141700	proteína ribosomal L6-1 60S	ensamblaje subunidad ribosomal mayor	
Glyma_13G094600	proteína FGYY	metabolismo de carbohidratos	
Glyma_04G169200	proteína <i>EMBRYO DEFECTIVE</i> 514	terminación desarrollo embrionario en dormancia	
Glyma_04G224200	ligasa E3 RNFT1	regulación positiva de vía ERAD	
Glyma_01G193800	proteína ribosomal L26-1 60S	biogénesis subunidad ribosomal mayor	
Glyma_08G042300	proteína ribosomal S15 40S	biogénesis subunidad ribosomal menor	
Glyma_03G215300	LOC100499789 no caracterizado	hidrólisis enlace fosfodiéster ARN	
Glyma_11G103700	proteína ribosomal L7a-2 60S	biogénesis ribosomal	
Glyma_14G190900	probable transportador de aniones probable	transporte transmembrana	
Glyma_08G109300	chalcona sintasa	biosíntesis flavonoides	
Glyma_07G117000	subunidad delta ATP sintasa	biosíntesis de ATP	
Glyma_20G192300	factor de transcripción NAC 53	regulación de la transcripción	
Glyma_08G190300	proteína caseína quinasa	fosforilación proteica	
Glyma_20G130500	LOC100809521 no caracterizado	procesos metabólicos	
Glyma_04G044000	transportador de metal Nramp3a	transporte de metales; elongación pelo radicular	nodulación
Glyma_18G153500	expansina-A4	organización pared celular	