UNIVERSIDAD AUTÓNOMA DE SAN LUIS POTOSÍ

FACULTAD DE CIENCIAS QUÍMICAS

POSGRADO EN CIENCIAS EN BIOPROCESOS

El RNA no codificante *EIN2* modula la respuesta de dos fitohormonas en plantas

TESIS QUE PARA OBTENER EL GRADO DE

DOCTOR EN CIENCIAS EN BIOPROCESOS

PRESENTA:

Nieto Hernández Jesús

DIRECTORA DE TESIS:

Dra. Catalina Arenas Huertero

SAN LUIS POTOSÍ, S. L. P. JULIO 2026

Este proyecto se realizó en el Laboratorio del Metabolismo del RNA de la Facultad de Ciencias de la Universidad Autónoma de San Luis Potosí en el periodo comprendido entre agosto del 2020 y junio del 2026, bajo la dirección de la Dra. Catalina Arenas Huertero y fue apoyado con recursos personales de la Dra. Arenas para la realización del trabajo.

El programa de Doctorado en Ciencias en Bioprocesos de la Universidad Autónoma de San Luis Potosí pertenece al Sistema Nacional de Posgrados de Calidad (SNP) del SECIHTI, registro 000590, en el Nivel Consolidado. Número de la beca otorgada por SECIHTI: 778040. Número de CVU: 862449.

Los datos del trabajo titulado “El RNA no codificante *EIN2* modula la respuesta de dos fitohormonas en plantas” se encuentran bajo el resguardo de la Facultad de Ciencias y pertenecen a la Universidad Autónoma de San Luis Potosí.

UNIVERSIDAD AUTÓNOMA DE SAN LUIS POTOSÍ
Facultad de Ciencias Químicas
Centro de Investigación y Estudios de Posgrado
Posgrado en Ciencias en Bioprocesos
Programa de Doctorado

Solicitud de Registro de Tesis Doctorado

San Luis Potosí SLP a junio 30 del 2026

Comité Académico

En atención a: Dra. Alicia Grajales Lagunes

Por este conducto solicito a Usted se lleve a cabo el registro de título de tesis de Doctorado, el cual quedó definido de la siguiente manera:

El RNA no codificante *EIN2* modula la respuesta de dos fitohormonas en plantas.

que desarrollará el estudiante: **Jesús Nieto Hernández.**

bajo la dirección de la: **Dra. Catalina Arenas Huertero.**

Asimismo, le comunico que el proyecto en el cual trabajará el alumno involucrará el manejo de y/o generación de plantas de *Arabidopsis thaliana* genéticamente modificadas que no tienen valor comercial agroalimentario y, por lo tanto, requiere de aval de Comité de Ética e investigación de la FCQ.

(Complete la opción que aplique en su caso):

( ) Sí debido a que:

( ) No

( ) No Aplica

Sin otro particular, quedo de Usted.

A T E N T A M E N T E

Jesús Nieto Hernández	Dra. Catalina Arenas Huertero
Nombre y firma del estudiante	Nombre y firma de la Directora de Tesis

UNIVERSIDAD AUTÓNOMA DE SAN LUIS POTOSÍ

FACULTAD DE CIENCIAS QUÍMICAS

Posgrado en Ciencias en Bioprocesos

El RNA no codificante *EIN2* modula la respuesta de dos fitohormonas en plantas.

Tesis que para obtener el grado de:

Doctorado en Ciencias en Bioprocesos

Presenta:

Nieto Hernández Jesús

SINODALES:

Presidente:	Dra. Claudia Escudero Lourdes
Secretaria:	Dra. Margarita Rodríguez y Domínguez Kessler
Vocal:	Dr. J. Sergio Casas Flores
Vocal:	Dra. Luz María Teresita Paz Maldonado
Vocal:	Dra. Catalina Arenas Huertero

SAN LUIS POTOSÍ, S. L. P. JULIO, 2026

INTEGRANTES DEL COMITÉ TUTORIAL ACADÉMICO

Dra. Claudia Escudero Lourdes. Adscrita al Posgrado en Ciencias en Bioprocesos de la Universidad Autónoma de San Luis Potosí, San Luis Potosí, S.L.P.

Dra. Margarita Rodríguez y Domínguez Kessler. Secretaria. Adscrita al Posgrado en Ciencias en Bioprocesos de la Universidad Autónoma de San Luis Potosí, San Luis Potosí, S.L.P.

Dr. J. Sergio Casas Flores. Sinodal Vocal. Profesor invitado al posgrado en Ciencias en Bioprocesos de la Universidad Autónoma de San Luis Potosí, San Luis Potosí, S.L.P. Adscrito al Instituto Potosino de Investigación Científica y Tecnológica, A.C., San Luis Potosí, San Luis Potosí, S.L.P.

Dra. Luz María Teresita Paz Maldonado. Vocal. Adscrita al posgrado en Ciencias en Bioprocesos de la Universidad Autónoma de San Luis Potosí, San Luis Potosí, S.L.P.

Dra. Catalina Arenas Huertero. Vocal. Adscrita al posgrado en Ciencias en Bioprocesos de la Universidad Autónoma de San Luis Potosí, San Luis Potosí, S.L.P.

Carta Cesión de Derechos

San Luis Potosí SLP a junio 30 del 2026

En la ciudad de San Luis Potosí el día 30 del mes de Junio del año 2026 El que suscribe, Jesús Nieto Hernández, alumno del programa de posgrado Ciencias en Bioprocesos adscrito a la Facultad de Ciencias Químicas, manifiesta que es autor intelectual del presente trabajo terminal, realizado bajo la dirección de: Dra. Catalina Arenas Huertero y cede los derechos del trabajo titulado “El RNA no codificante *EIN2* modula la respuesta de dos fitohormonas en plantas” a la **Universidad Autónoma de San Luis Potosí**, para su difusión con fines académicos y de investigación.

Los usuarios de la información no deben reproducir de forma total o parcial texto, gráficas, imágenes o cualquier contenido del trabajo sin el permiso expreso del o los autores. Éste, puede ser obtenido directamente con el autor o autores escribiendo a la siguiente dirección catalina.arenas@uaslp.mx. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

_____

Jesús Nieto Hernández

UNIVERSIDAD AUTÓNOMA DE SAN LUIS POTOSÍ
Facultad de Ciencias Químicas
Centro de Investigación y Estudios de Posgrado
Posgrado en Ciencias en Bioprocesos
Programa de Doctorado

Carta de Análisis de Similitud

San Luis Potosí SLP a junio 30 del 2026

L.B. Reyna Nayeli Ortíz Quintero

Biblioteca de Posgrado FCQ

Asunto: Reporte de porcentaje de similitud de tesis de grado

Por este medio me permito informarle el porcentaje de similitud obtenido mediante Ithenticate para la tesis titulada "*El RNA no codificante EIN2 modula la respuesta de dos fitohormonas en plantas*" presentada por el autor Jesús Nieto Hernández. La tesis es requisito para obtener el grado de Doctorado en el Posgrado en Ciencias en Bioprocesos. El análisis reveló un porcentaje de similitud de Porcentaje de Similitud del 21 % excluyendo referencias y metodología.

Agradezco sinceramente su valioso tiempo y dedicación para llevar a cabo una exhaustiva revisión de la tesis. Quedo a su disposición para cualquier consulta o inquietud que pueda surgir en el proceso.

Sin más por el momento, le envío un cordial saludo.

A T E N T A M E N T E

_____

Dra. Alicia Grajales Lagunes

Coordinadora Académico del Posgrado en Ciencias en Bioprocesos

Agradecimientos

RESUMEN

Las plantas poseen complejas redes regulatorias que integran diversas señales hormonales, ambientales y del desarrollo para coordinar su crecimiento y su adaptación al ambiente. Entre los componentes de estas redes regulatorias se encuentran los RNAs largos no codificantes (lncRNAs, de sus siglas en inglés), moléculas capaces de modular la expresión génica mediante su interacción con proteínas, otros RNAs o regiones del DNA. A pesar de su creciente caracterización, su origen evolutivo y sus características estructurales en plantas aún han sido poco estudiados. En este trabajo se caracterizaron dos lncRNAs asociados con genes reguladores en *Arabidopsis thaliana*: el lncRNA localizado río abajo del gen *ETHYLENE INSENSITIVE 2* (*EIN2*) (*AT5G03285*) y el lncRNA intergénico asociado a *TBP2,* (*TALIR*). La caracterización funcional de *AT5G03285* mostró que las líneas mutantes *knockdown* presentan alteraciones en la respuesta al ET, incluido un fenotipo intermedio de triple respuesta y respuestas alteradas a la fitohormona ABA. Asimismo, líneas reporteras del promotor de *AT5G03285*, revelaron patrones de expresión específico de tejido y de respuesta a hormonas. Los análisis comparativos de sintenia indicaron que el *locus* de *AT5G03285* se encuentra conservado en múltiples especies de plantas, especialmente en angiospermas, manteniendo su proximidad genómica con *EIN2*. Aunque la conservación nucleotídica es limitada, se identificaron regiones conservadas con pequeños marcos de lectura (sORF), y los análisis de estructura secundaria revelaron estructuras tallo-asa potencialmente implicadas en la interacción RNA-EIN2. Finalmente, *TALIR* mostró conservación de sintenia dentro del orden Brassicales, así como evidencias de conservación estructural y regulatoria en su región promotora. En conjunto, estos resultados sugieren que los lncRNAs asociados con genes reguladores centrales pueden modular las respuestas hormonales, del desarrollo y de adaptación en plantas.

Palabras clave: ABA, *Arabidopsis thaliana*, EIN2, etileno, lncRNA, TBP2.

ABSTRACT

Plants possess complex regulatory networks that integrate hormonal, environmental, and developmental signals to coordinate growth and adaptation to their environment. Among the components of these regulatory networks are long non-coding RNAs (lncRNAs), molecules capable of modulating gene expression through interactions with proteins, other RNAs or DNA regions. Despite their increasing characterization, the evolutionary origin and structural features of many plant lncRNAs remain poorly understood. In this work, two lncRNAs associated with regulatory genes in *Arabidopsis thaliana* were characterized: the lncRNA located downstream of the *ETHYLENE INSENSITIVE 2* (*EIN2*) gene (*AT5G03285*) and the intergenic lncRNA associated with *TBP2* (*TALIR*). Functional characterization of *AT5G03285* revealed that *knockdown* mutant lines exhibit altered ethylene responses, including an intermediate triple-response phenotype and modified responses to the phytohormone ABA. In addition, reporter lines carrying the *AT5G03285* promoter revealed tissue-specific expression patterns and regulation in response to hormonal signals. Comparative synteny analyses indicated that the *AT5G03285 locus* is conserved across multiple plant species, particularly among angiosperms, maintaining its genomic proximity to *EIN2*. Although overall nucleotide conservation is limited, conserved regions containing small open reading frames (sORFs) were identified, and secondary structure analyses revealed stem–loop structures potentially involved in RNA–EIN2 interactions. Finally, *TALIR* showed synteny conservation within the order Brassicales, as well as evidence of structural and regulatory conservation in its promoter region. Together, these results suggest that lncRNAs associated with key regulatory genes may play important roles in modulating hormonal signaling, development, and environmental adaptation in plants.

Key words: ABA, *Arabidopsis thaliana*, EIN2, Ethylene, lncRNA, TBP2.

INDICE GENERAL

INDICE DE FIGURAS

INTRODUCCIÓN

Respuesta de las plantas al ambiente

Las plantas son organismos sésiles que, a lo largo de su ciclo de vida, deben percibir e integrar una amplia diversidad de señales ambientales para ajustar su crecimiento, desarrollo y supervivencia. A diferencia de los organismos móviles, las plantas no pueden evadir condiciones adversas, por ello, dependen de mecanismos fisiológicos y moleculares altamente regulados que les permiten responder de manera eficiente a estímulos externos cambiantes (Zhu, 2016).

Entre los principales factores que afectan el crecimiento y desarrollo de las plantas se encuentra el estrés de tipo biótico, ocasionado por la interacción con otros organismos vivos y el estrés abiótico, asociado a condiciones como sequía, deshidratación, salinidad, temperaturas extremas y limitaciones nutricionales. Estos factores abióticos pueden actuar de manera independiente o en combinación con el estrés de tipo biótico, afectando las distintas etapas del crecimiento y desarrollo de las plantas (Bita y Gerats, 2013; Zhu, 2016, Zandalinas *et al*., 2021).

En las últimas décadas, la frecuencia y la intensidad de los factores de estrés abiótico se han incrementado como consecuencia de las variaciones ambientales globales, generando un escenario particularmente desafiante para los sistemas agrícolas y los ecosistemas naturales. Este incremento ha puesto de manifiesto la necesidad de comprender con mayor profundidad los mecanismos que subyacen a la capacidad adaptativa de las plantas frente a este tipo de cambio (Bailey-Serres *et al*., 2019, Zandalinas *et al*., 2021).

La respuesta de las plantas a los cambios ambientales implica la activación y regulación de redes complejas de señalización hormonal que permiten la percepción de los estímulos y la generación de respuestas celulares y moleculares específicas. Estas redes integran señales provenientes del entorno con el estado fisiológico interno de la planta, dando lugar a ajustes coordinados en la expresión génica (Gilroy *et al*., 2016; Suzuki *et al*., 2012).

Estos procesos constituyen sistemas regulatorios altamente sofisticados que opera a múltiples niveles. Comprender cómo se coordinan estas respuestas requiere profundizar en los mecanismos moleculares que controlan la expresión génica y la integración de señales ambientales. En este contexto, diversas clases de ácidos ribonucleicos (RNAs), como los microRNAs (miRNAs), pequeños RNAs interferentes (siRNAs) y RNAs largos no codificantes (lncRNAs) desempeñan funciones clave en la regulación de estos procesos (Chen *et al*., 2020; Lämke y Bäurle, 2017).

Los RNAs no codificantes como reguladores de la expresión génica

Los ajustes fisiológicos y de desarrollo que permiten a las plantas responder eficientemente a condiciones ambientales variables dependen, en gran medida, de una regulación precisa de la expresión génica. En este contexto, el RNA ha dejado de considerarse únicamente como un intermediario entre el DNA y las proteínas, para reconocerse como un elemento activo en los mecanismos regulatorios celulares. Una fracción considerable del transcriptoma vegetal está constituida por RNAs no codificantes **(ncRNAs)**, es decir, transcritos que presentan un potencial bajo para codificar péptidos o proteínas, pero que desempeñan funciones regulatorias esenciales a distintos niveles de la expresión génica (Chekanova, 2015; Hajjari y Salavaty, 2015, Yang *et al*., 2023).

De manera general, los ncRNAs pueden agruparse en RNAs constitutivos (RNAs de transferencia y RNAs ribosomales) y RNAs regulatorios. Los primeros incluyen secuencias altamente conservadas y esenciales para la función celular básica, como los RNAs de transferencia **(tRNA)** y los RNAs ribosomales **(rRNA)**. En contraste, los RNAs regulatorios participan en la modulación de la expresión génica y suelen clasificarse con base en su longitud en RNAs pequeños no codificantes, con longitudes menores de 200 nucleótidos, y RNAs largos no codificantes **(lncRNA)**, con longitudes superiores a 200 nucleótidos (Hajjari y Salavaty, 2015).

Dentro del grupo de los RNAs pequeños no codificantes se incluyen los microRNAs **(miRNAs)** y los RNAs pequeños interferentes **(siRNAs)**, los cuales difieren entre sí en sus vías de biogénesis y mecanismos de acción. A su vez, los siRNAs

comprenden diversos subtipos, entre ellos los siRNAs de acción *trans* (ta-siRNAs, trans-acting small interfering RNAs), los siRNAs de acción *cis* (ca-siRNAs, cis-acting small interfering RNAs), los siRNAs derivados de transcritos antisentido naturales (nat-siRNAs, natural antisense transcript-derived small interfering RNAs) y los siRNAs en fase (phasiRNAs, phased small interfering RNAs), participan en procesos regulatorios específicos mediante mecanismos que pueden actuar a nivel transcripcional, postranscripcional o epigenético. Esta diversidad de RNAs pequeños refleja la complejidad de las redes regulatorias mediadas por RNA y su papel central en la integración de señales internas y externas en las plantas (Chekanova, 2012).

Por otra parte, los lncRNAs han emergido como reguladores clave de múltiples procesos biológicos debido a su diversidad estructural y funcional.

lncRNAs

Los lncRNAs se definen como transcritos de RNA con una longitud mayor de 200 nucleótidos y con un potencial mínimo o nulo para la codificación de péptidos o proteínas. De acuerdo con su localización genómica respecto a genes codificantes, los lncRNAs pueden clasificarse en lncRNAs intergénicos (lincRNAs), intrónicos, antisentido, bidireccionales y derivados de regiones promotoras. Sin embargo, en los últimos años se han descrito una gran diversidad de lncRNAs con funciones duales, capaces tanto de actuar como RNA no codificantes como de codificar pequeños péptidos funcionales. En comparación con los RNAs mensajeros, los lncRNAs suelen presentar niveles de expresión más bajos y patrones de expresión más restringidos a tejidos, etapas del desarrollo o condiciones fisiológicas específicas. Los lncRNAs participan en la regulación de la expresión génica a distintos niveles, incluyendo cambios epigenéticos, control transcripcional y regulación postranscripcional. Estos transcritos pueden interactuar con diversas moléculas como DNA, proteínas y otros RNAs actuando como guías, andamios o plataformas de ensamblaje para complejos regulatorios, lo que amplía considerablemente el repertorio de mecanismos de control génico en las células vegetales (Zhang *et al*., 2019) (Figura 1).

Así mismo, el desarrollo y la implementación de tecnologías de secuenciación de nueva generación han permitido la identificación de miles de lncRNAs en diversos organismos eucariotas, incluyendo vertebrados y plantas. Sin embargo, a pesar de su abundancia y relevancia funcional, los lncRNAs se caracterizan por presentar altas tasas de recambio evolutivo y una conservación limitada a nivel de secuencia, lo que ha dificultado el análisis comparativo y la inferencia de sus funciones desde una perspectiva evolutiva (Kapusta y Feschotte, 2014; Ruiz-Orera y Alba, 2019, Yang *et al*., 2023).

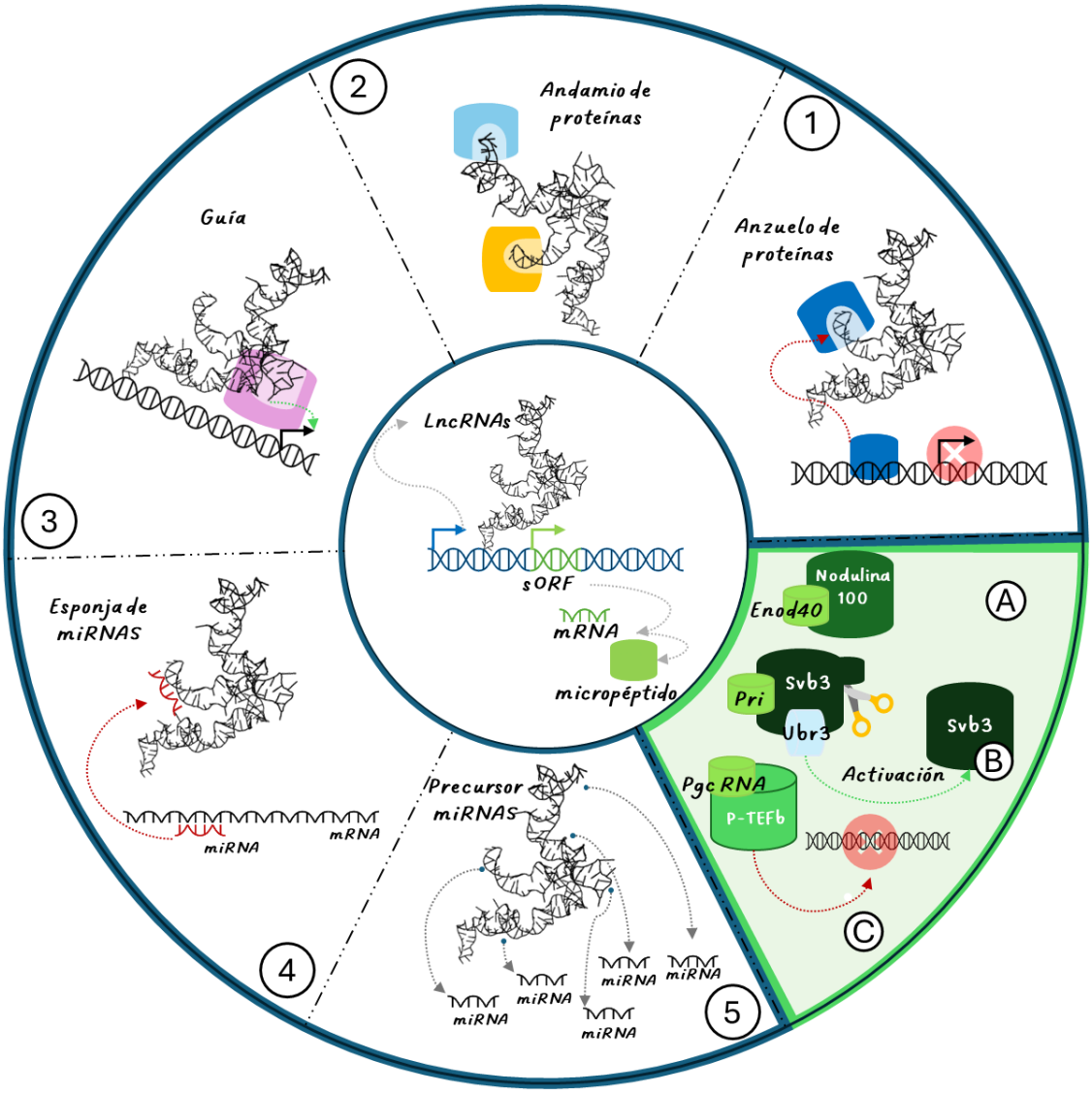


Figura 1. Mecanismos de acción de los lncRNAs en la regulación de la expresión génica y su función codificante.

Los principales mecanismos mediante los cuales los lncRNAs regulan la expresión génica, incluyendo funciones como anzuelo de proteínas (1), andamio de proteínas (2), guía de complejos regulatorios hacia regiones específicas del DNA (3), esponja de miRNAs (4) y precursores de miRNAs (5). En el mecanismo de anzuelo (1), los lncRNAs secuestran proteínas regulatorias e impiden su unión al DNA; como andamio (2), facilitan la formación de complejos proteicos; como guía (3), dirigen dichos complejos a *loci* específicos; como esponja (4), secuestran miRNAs evitando la represión de sus ARNm blanco; y como precursores (5), generan miRNAs funcionales. Adicionalmente, algunos lncRNAs contienen ORFs cortos (sORFs) capaces de codificar micropéptidos funcionales. En los paneles (A–C) se ilustran ejemplos representativos: (A) ENOD40 produce micropéptidos que interactúan con nodulina 100; (B) Pri regula la activación del factor de transcripción Shavenbaby (Svb) mediante Ubr3; y (C) el péptido derivado de pgc interactúa con P-TEFb, inhibiendo la elongación transcripcional. Estos ejemplos evidencian que los lncRNAs constituyen una fuente funcional de micropéptidos reguladores (Chugunova *et al*., 2018). ENOD40 (EARLY NODULIN 40); Pri (POLISHED RICE); Svb (SHAVENBABY); Ubr3 (UBIQUITIN PROTEIN LIGASE E3 COMPONENT N-RECOGNIN 3); P-TEFb (POSITIVE TRANSCRIPTION ELONGATION FACTOR b); sORF (SHORT OPEN READING FRAME).

Por otra parte, la estructura de los lncRNAs desempeña un papel fundamental en su actividad regulatoria, ya que determina su capacidad para establecer interacciones específicas con otras moléculas. Estas estructuras pueden contener motivos de reconocimiento altamente conservados que facilitan la unión a proteínas, RNAs u otras regiones del genoma, lo que confiere a los lncRNAs una notable versatilidad funcional. En este sentido, se ha propuesto que la conservación estructural, más que la conservación de la secuencia nucleotídica constituye una característica clave para comprender la funcionalidad y la relevancia evolutiva de los lncRNAs en distintos linajes (Di *et al*., 2014; Uroda *et al*., 2019; Gultyaev *et al*., 2023; Zhao *et al*., 2024).

En conjunto, estas características posicionan a los lncRNAs como componentes fundamentales de las redes regulatorias celulares, capaces de integrar señales internas y externas mediante mecanismos flexibles y altamente especializados.

ncRNAs en plantas y su participación en la señalización hormonal

En plantas, los lncRNAs han sido ampliamente estudiados por su papel en regulación de procesos del desarrollo y en la respuesta a condiciones ambientales adversas a través de la modulación de la expresión génica asociada a múltiples rutas de señalización, lo que resalta su importancia como componentes de regulación fina en procesos fisiológicos complejos (Bardou *et al*., 2014; Nejat *et al*., 2018; Li *et al*; Wang *et al*., 2023).

Los lncRNAs pueden actuar como reguladores en *cis* o en *trans*, modulando la expresión de genes involucrados en la percepción del estrés, la transducción de señales y la activación de respuestas adaptativas. Ejemplos de esto incluyen el lncRNA *APOLO*, implicado en la regulación del desarrollo radicular en respuesta a señales hormonales, así como *COLDAIR* y *COOLAIR*, lncRNAs asociados a la respuesta al frío y a la regulación epigenética del *locus FLOWERING LOCUS C* (*FLC*) durante la vernalización. Asimismo, *HID1* ha sido descrito como un lncRNA involucrado en la regulación del desarrollo fotomorfogénico, lo que destaca la diversidad funcional de estos transcritos en distintos contextos ambientales y de desarrollo (Ariel *et al*., 2014; Nejat *et al*., 2018).

En este sentido, se ha demostrado que los lncRNAs pueden influir en la biosíntesis, la señalización y la respuesta a distintas fitohormonas, incluyendo auxinas, ácido abscísico y ácido jasmónico, mediante mecanismos transcripcionales y postranscripcionales que afectan genes clave de estas vías (Ariel *et al*., 2014; Wang *et al*., 2023).

La participación de los lncRNAs en la señalización hormonal sugiere que estos transcritos forman parte de niveles regulatorios finos que permiten ajustar la intensidad, duración y especificidad de las respuestas hormonales. De esta manera, los lncRNAs pueden actuar como mediadores que conectan señales ambientales con rutas hormonales, facilitando la integración coordinada de múltiples componentes regulatorios y favoreciendo la adaptación de las plantas a contextos fisiológicos variables (Wang *et al*., 2017; Zhao *et al*., 2020).

En conjunto, estas evidencias indican que los lncRNAs constituyen un nivel regulatorio adicional en las plantas, con un papel relevante en la integración de señales hormonales y ambientales, lo que subraya la necesidad de profundizar en el estudio de sus mecanismos de acción y su interacción con componentes clave de las vías de señalización vegetal.

lncRNAs en la vía del etileno

El etileno (ET) es una fitohormona gaseosa que regula una amplia variedad de procesos fisiológicos en las plantas, incluidos la germinación, el crecimiento, la senescencia, la respuesta al estrés, así como la maduración de frutos. Debido a su participación en respuestas tanto de desarrollo como adaptativas, la señalización por ET constituye un sistema altamente regulado que integra señales ambientales y endógenas para ajustar la expresión génica de manera precisa (Wang *et al*.,2025).

En los últimos años, diversos estudios han comenzado a revelar la participación de los lncRNAs en la regulación de la vía de señalización del ET (Yu *et al*., 2022; Dey *et al*., 2022; Cao *et al*., 2022; Lee *et al*., 2024; Cheng *et al*., 2024; Ma *et al*., 2024).

En especies de interés agronómico se han identificado lncRNAs específicos asociados a procesos dependientes del ET. Por ejemplo, en jitomate (*Solanum lycopersicum*), se ha demostrado que lncRNAs como lncRNA1459 y lncRNA1840 regulan la expresión de genes implicados en la maduración del fruto, un proceso estrechamente vinculado a la acción del ET. De manera similar, en pepino (*Cucumis sativus*) se han descrito lncRNAs, como *CslncRNA1* y *CslncRNA2*, asociados a la regulación de componentes de la señalización del ET y las respuestas al estrés. En manzano (*Malus domestica*), lncRNAs como *MdLNC610* y *MdLNC499* han sido implicados en la modulación de la maduración del fruto y de otros procesos fisiológicos regulados por esta fitohormona (Yu *et al*., 2022; Dey *et al*., 2022; Cao *et al*., 2022; Lee *et al*., 2024; Cheng *et al*., 2024; Ma *et al*., 2024).

Estas evidencias sugieren que los lncRNAs podrían participar en la regulación de la señalización del ET, contribuyendo a la regulación espacial y temporal de la

expresión génica asociada a esta vía. Aunque los mecanismos específicos de acción aún no han sido completamente caracterizados, los patrones de expresión observados sugieren que los lncRNAs podrían influir en la intensidad y especificidad de la respuesta al ET, facilitando la integración de señales hormonales y ambientales en contextos fisiológicos específicos (Maric *et al*., 2025).

En conjunto, la evidencia disponible indica que la regulación mediada por lncRNAs constituye una capa adicional de control en la vía de señalización del ET. Sin embargo, a pesar de los avances recientes, los mecanismos moleculares mediante los cuales estos transcritos se integran con los componentes centrales de la vía, en particular aquellos responsables de la transducción de la señal al núcleo permanecen en gran medida poco caracterizados. Esta falta de integración mecanística resalta la necesidad de profundizar en el estudio de la función de los lncRNAs en el sistema de señalización del ET.

EIN2 como nodo central de la señalización por etileno

A nivel molecular, la vía de señalización del ET presenta características distintivas, ya que opera mediante un mecanismo de regulación negativa poco común en plantas, en el que los receptores de ET permanecen activos en ausencia de la hormona y se inactivan tras su percepción (Gallie, 2015).

La elucidación de la vía de percepción y respuesta al ET ha sido posible principalmente gracias al uso de líneas mutantes insensibles al ET o con fenotipos constitutivos de respuesta, particularmente en la planta modelo *Arabidopsis thaliana*. El análisis genético y molecular de estas líneas mutantes permitió establecer un modelo canónico de la vía de señalización del ET, organizado de manera lineal, en el que la percepción hormonal se traduce en cambios transcripcionales específicos que conducen a respuestas fisiológicas bien caracterizadas, como la denominada triple respuesta, que consiste en el acortamiento y engrosamiento del hipocótilo, así como la formación de un gancho apical exagerado en plántulas germinadas en oscuridad y expuestas a ET o a su precursor, el ácido 1-aminociclopropano-1-carboxilico (ACC) (Alonso *et al*., 1999; Binder *et al*., 2020).

La percepción del ET ocurre a nivel celular en la membrana del retículo endoplásmico, donde un conjunto de receptores, entre los que se incluyen ETR1, ETR2, ERS1, ERS2 y EIN4, requiere del cotransportador RESPONSE TO ANTAGONIST 1 (RAN1) para facilitar la incorporación de cobre, elemento esencial para la unión del ET a estos receptores (Hirayama *et al*., 1999). En ausencia de ET, los receptores permanecen activos y mantienen la actividad de la proteína CTR1, una quinasa que fosforila y reprime al componente central río abajo, ETHYLENE INSENSITIVE 2 (EIN2). La presencia de ET conduce a la inactivación de los receptores y de CTR1, lo que permite la desfosforilación de EIN2 y la activación de la señalización hormonal (Gallie, 2015).

EIN2 constituye un nodo central en la vía de señalización del ET, ya que actúa como el principal integrador de la señal percibida en el retículo endoplásmico y su transducción hacia el núcleo y otros compartimentos celulares. En presencia de ET, el extremo C-terminal de EIN2 (CEND-EIN2) es escindido y translocado tanto al núcleo como a los cuerpos de procesamiento de RNA (P-bodies), donde regula de manera directa respuestas transcripcionales y postranscripcionales asociadas a la señalización por ET (Qiao *et al*., 2012; Li *et al*., 2015).

En los P-bodies, el fragmento CEND-EIN2 se une a motivos ricos en uracilo localizados en la región 3′UTR de los transcritos que codifican para las ubiquitinas ligasas EIN3 BINDING F-BOX PROTEIN 1 (EBF1) y EIN3 BINDING F-BOX PROTEIN 2 (EBF2). A través de su interacción con la exorribonucleasa EIN5 y otros componentes asociados a los P-bodies, EIN2 dirige estos transcritos hacia dichos complejos citoplasmáticos, donde se reprime su traducción. En ausencia de ET, EBF1 y EBF2 promueven la degradación de los factores de transcripción EIN3 y EIN3-LIKE 1 (EIL1) mediante la vía del proteosoma. Por lo tanto, la acción de EIN2 sobre estos transcritos resulta

esencial para la estabilización de EIN3 y EIL1 y la activación de la expresión de genes de respuesta al ET (Li *et al*., 2015; Reichel *et al*., 2016) (Figura 2).

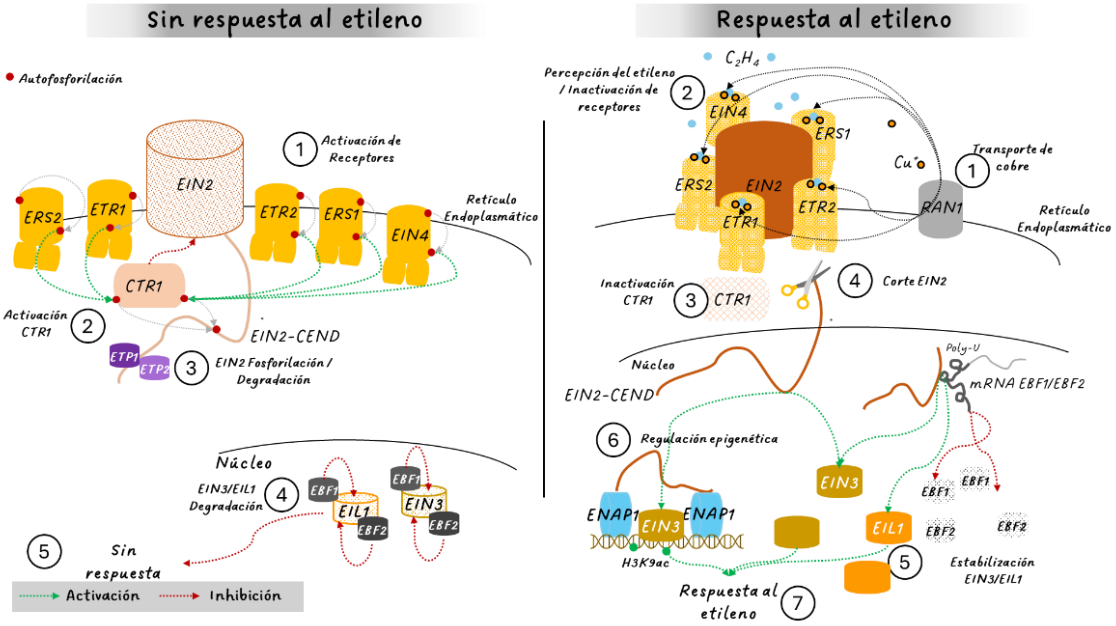


Figura 2. Vía de señalización del etileno en *Arabidopsis thaliana.*

La ruta de percepción del ET en ausencia (panel izquierdo) y en presencia del ET (panel derecho). En ausencia de ET, los receptores localizados en la membrana del retículo endoplásmico (ETR1, ETR2, ERS1, ERS2 y EIN4) se encuentran activos y son autofosforilados (1), lo que promueve la activación de la proteína quinasa CTR1 (2). CTR1 fosforila el extremo C-terminal de EIN2, promoviendo su degradación (3). Como consecuencia, los factores de transcripción EIN3 y EIL1 son ubiquitinados y degradados por las proteínas EBF1 y EBF2 (4), impidiendo la activación de genes de respuesta al ET, lo que resulta en la ausencia de respuesta a esta hormona (5). En presencia de ET, el cobre es transportado hacia los receptores por la proteína RAN1 (1), permitiendo la percepción de la hormona (C_2H_4) (2). La unión de ET provoca la inactivación de los receptores ETR1, ETR2, ERS1, ERS2, EIN4 y la de la proteína quinasa CTR1 (3). Esto conduce a la desfosforilación y escisión del extremo C-terminal de EIN2 (EIN2-CEND), el cual posteriormente es translocado al núcleo (4), donde inhibe la traducción de las proteínas EBF1 y EBF2. La disminución en los niveles de EBF1/EBF2 permite la estabilización de EIN3 y EIL1 (5), que activan la respuesta al ET. Adicionalmente, EIN2-CEND participa en la regulación epigenética de genes de respuesta al ET mediante modificaciones de cromatina, como la acetilación de histonas (H3K9ac) asociada a ENAP1 (6). En conjunto, estos procesos conducen a la activación de la respuesta al ET (7). Las flechas verdes indican activación, las líneas rojas punteadas representan represión y las líneas grises señalan el proceso de autofosforilación de los receptores ETR1, ETR2, EIN4, ERS1 y ERS2. CTR1 (CONSTITUTIVE TRIPLE RESPONSE 1); EIN3 (ETHYLENE

INSENSITIVE); EIL1 (EIN3-LIKE 1); EBF1/2, (EIN3-BINDING F-BOX PROTEIN 1/2); RAN1 (RESPONSIVE TO ANTAGONIST 1); ENAP1 (EIN2 NUCLEAR ASSOCIATED PROTEIN 1).

En el núcleo, EIN3 y EIL1 inducen la transcripción de múltiples genes sensibles al ET, incluidos factores de transcripción de la familia ERF y genes clave como *HOOKLESS1* (*HLS1*), involucrado en la formación del gancho apical durante la triple respuesta. Además de su función postranscripcional, se ha demostrado que CEND-EIN2 participa en la regulación epigenética de la respuesta al ET mediante su interacción con la proteína EIN2 NUCLEAR-ASSOCIATED PROTEIN 1 (ENAP1), promoviendo la acetilación de histonas y facilitando una activación transcripcional rápida y eficiente de genes blanco de EIN3 (Zhang *et al*., 2017) (Figura 2).

Evidencias recientes han revelado que la vía de señalización del ET no opera de manera estrictamente lineal, sino que presenta múltiples niveles de regulación y ramificaciones funcionales. En este contexto, EIN2 ha sido implicada en procesos adicionales como la regulación de la elongación y proliferación celular a través del eje glucosa–TOR–EIN2, el cual se encuentra desacoplado de la señalización canónica dependiente de CTR1 y está mediado por sitios específicos de fosforilación en el extremo C-terminal de EIN2 (Fu *et al*., 2021). Asimismo, análisis de Western blot y de espectrometría de masas sugieren que CEND-EIN2 puede procesarse proteolíticamente en múltiples fragmentos, lo que añade un nivel adicional de complejidad a su regulación funcional (Qiao *et al*., 2012).

En conjunto, estas evidencias posicionan a EIN2 como un regulador central y multifuncional dentro de la vía de señalización del ET, sujeto a una regulación altamente compleja e integrada con otras rutas de señalización hormonal y metabólica. La diversidad de mecanismos regulatorios asociados a EIN2 resalta su importancia como punto de convergencia de múltiples señales, lo que permite a las plantas ajustar sus respuestas fisiológicas ante condiciones ambientales variables.

La complejidad de la señalización mediada por ET, caracterizada por múltiples niveles de control transcripcional, postranscripcional y epigenético, ofrece un contexto interesante para la acción reguladora de los lncRNAs. Dada su capacidad para

interactuar con DNA, RNA y proteínas, los lncRNAs podrían influir en distintos puntos de la vía, modulando la estabilidad, localización o actividad de componentes clave de la señalización, así como la expresión de genes de respuesta al ET. Sin embargo, los mecanismos moleculares específicos mediante los cuales estos transcritos ejercen su función reguladora aún no han sido completamente dilucidados (Binder *et al*., 2020).

En este contexto, la posición central de EIN2 en la vía de señalización del ET resulta particularmente relevante. EIN2 integra la señal percibida a nivel del retículo endoplásmico y coordina respuestas tanto transcripcionales como postranscripcionales, lo que lo convierte en un punto potencial de interacción con mecanismos regulatorios mediados por RNA. La existencia de una regulación postranscripcional dependiente de motivos ricos en uracilo y la participación de EIN2 en procesos epigenéticos sugieren escenarios en los que los lncRNAs podrían contribuir a la regulación de la respuesta al ET (Yu *et al*., 2022; Dey *et al*., 2022; Cao *et al*., 2022; Lee *et al*., 2024; Cheng *et al*., 2024; Ma *et al*., 2024).

A pesar de estos indicios, la mayoría de los estudios disponibles se han centrado en describir asociaciones funcionales entre lncRNAs y procesos dependientes de ET, sin abordar directamente cómo estos transcritos se integran con los componentes centrales de la vía de señalización. Esta falta de integración mecanística limita la comprensión de cómo la señalización por ET se ajusta con precisión en distintos contextos fisiológicos y ambientales.

En conjunto, la regulación mediada por lncRNAs emerge como un componente potencialmente crucial en la señalización del ET, capaz de aportar un nivel adicional de control y especialización a esta vía hormonal. Profundizar en la caracterización de estos mecanismos resulta esencial para comprender cómo las plantas integran señales hormonales y ambientales a través de redes regulatorias complejas, y cómo esta integración contribuye a la adaptación y plasticidad de las respuestas vegetales.

ANTECEDENTES

El lncRNA *AT5G03285* en *Arabidopsis thaliana*

En estudios previos, se identificó y se caracterizó inicialmente el lncRNA *AT5G03285* en *A. thaliana*, una región codificante para un transcrito localizado río abajo del gen *EIN2,* componente central de la vía de señalización del ET (Nieto-Hernández,2020). La localización genómica de *AT5G03285* en una región cercana a EIN2 sugiere un posible papel regulador de este lncRNA en procesos controlados por fitohormonas asociadas al gen *EIN2*.

El análisis de expresión de *AT5G03285* reveló que este lncRNA presenta un patrón regulado por condiciones ambientales y hormonales específicas como el calor, el frío y la respuesta a ABA (Jin *et al*., 2022). En particular, se observó que la acumulación del transcrito se incrementa significativamente bajo condiciones de estrés por calor y tras tratamientos con ácido abscísico (ABA), mientras que su expresión disminuye bajo otras condiciones de estrés, como el frío. Estos resultados indicaron que *AT5G03285* responde de manera diferencial a estímulos ambientales y hormonales, lo cual es consistente con lo descrito para otros lncRNAs funcionales involucrados en la respuesta al estrés en plantas (Ariel *et al*., 2014; Nejat *et al*., 2018) (Figura 3).

Adicionalmente, diversos estudios mencionan que la región promotora de *AT5G03285* puede asociarse a diversos factores de transcripción asociados a la respuesta al estrés y a señales hormonales, particularmente relacionados con ABA. Esto es congruente con estudios previos que han demostrado que lncRNAs asociados a regiones reguladoras enriquecidas en elementos de respuesta hormonal pueden participar en la modulación fina de la expresión génica dentro de redes de señalización complejas (Tan *et al*., 2020; Wang *et al*., 2023).

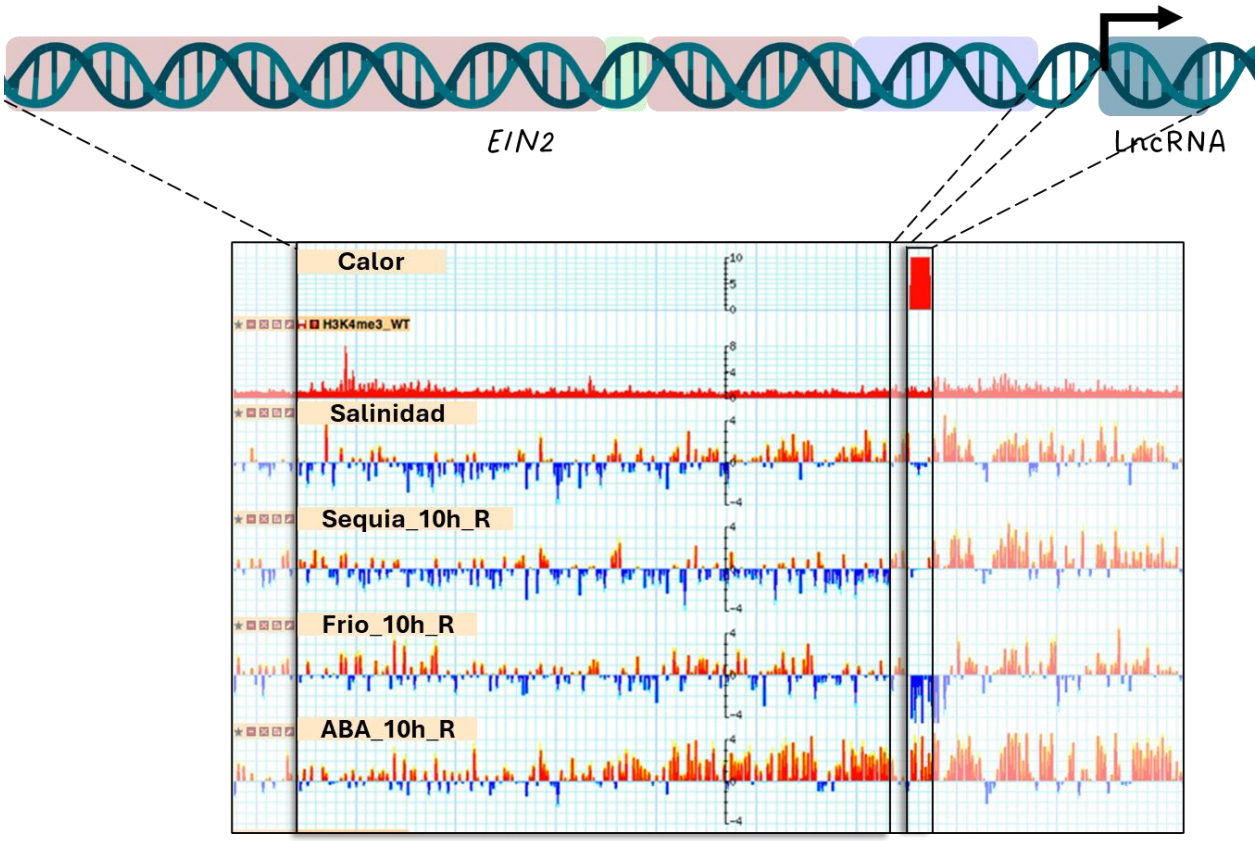


Figura 3. Contexto genómico y perfil de expresión de *AT5G03285* bajo condiciones de estrés abiótico.
Representación del *locus* genómico de *AT5G03285*, que muestra la región promotora, la dirección de transcripción y el cuerpo génico. (B) Visualización tipo genome browser que muestra los patrones de expresión de *AT5G03285* bajo diferentes condiciones de estrés, incluyendo calor, salinidad, sequía (10 h), frío (10 h) y tratamiento con ácido abscísico (ABA, 10 h). Las señales en rojo y azul representan incrementos y disminuciones relativas de la expresión, respectivamente. La región resaltada indica la ubicación del *locus* de *AT5G03285*. Los datos de expresión se presentan sin modificación de escala ni de proporción.

De manera complementaria, análisis y revisiones bibliográficas basados en información experimental y en bases de datos curadas permitieron situar a *AT5G03285* dentro de un entorno regulatorio dominado por factores de transcripción asociados a la señalización por ABA, a las respuestas al estrés y a procesos de senescencia. En este contexto, se identificaron factores como ABI5, miembro de la familia ABF, así como reguladores asociados a la senescencia y al estrés térmico.

Estos antecedentes sugieren que *AT5G03285* se encuentra inmerso en un entorno regulatorio estrechamente vinculado a la respuesta al estrés y a la integración de señales hormonales, en particular las mediadas por ABA y ET. Dado que la vía de señalización del ET se caracteriza por la participación de mecanismos postranscripcionales y epigenéticos altamente especializados, la proximidad genómica y el patrón de expresión de *AT5G03285* sugieren que este lncRNA podría participar en la modulación fina de dicha vía, potencialmente a través de mecanismos reguladores aún no descritos (Nieto-Hernández, 2020).

En conjunto, la evidencia disponible posiciona a *AT5G03285* como un candidato relevante para el estudio de la regulación mediada por RNA en la señalización hormonal. Además, el papel funcional específico de este lncRNA, así como los mecanismos moleculares mediante los cuales podría influir en la señalización por ET y su interacción con ABA, permanece sin caracterizar. Estas interrogantes constituyen el punto de partida del presente trabajo, orientado a profundizar en la función de *AT5G03285* dentro de las redes regulatorias que controlan la respuesta hormonal y al estrés en las plantas.

JUSTIFICACIÓN

A pesar del creciente interés en los lncRNAs en plantas, la información disponible sobre la función específica de muchos de estos transcritos sigue siendo limitada. En la mayoría de los casos, el conocimiento existente proviene de análisis transcriptómicos a gran escala, que permiten identificar patrones de expresión, sin embargo, estos resultan insuficientes para esclarecer su papel funcional y los mecanismos moleculares mediante los cuales participan en la regulación de procesos celulares específicos.

El lncRNA *AT5G03285* representa un ejemplo de este vacío de conocimiento. La evidencia disponible sugiere que su expresión está modulada por distintos estímulos ambientales y hormonales, incluyendo su sobreexpresión en condiciones de estrés por calor y bajo tratamientos exógenos con ABA, así como su regulación negativa en

condiciones de frío. Adicionalmente, su localización genómica en proximidad al gen *EIN2*, un componente central de la vía de señalización del ET sugiere un posible papel regulador en rutas hormonales clave de la respuesta al estrés.

Considerando la complejidad de la señalización hormonal en plantas y el papel integrador de EIN2 en la respuesta al ET, así como su interacción con otras fitohormonas, se plantea la necesidad de investigar si *AT5G03285* participa en la regulación de la respuesta al ABA a través de mecanismos asociados a la señalización del ET. Para abordar esta pregunta, en el presente trabajo se realizará la caracterización funcional de AT5G03285 mediante el análisis de líneas mutantes, evaluando su respuesta frente a tratamientos con ABA y ET. Esta estrategia permitirá generar evidencia experimental sobre la función biológica de este lncRNA y su posible integración en redes regulatorias hormonales.

Desde una perspectiva biológica y fisiológica, resulta especialmente relevante determinar si la función de *AT5G03285* se encuentra asociada con la respuesta a estrés abiótico. La sequía representa una de las principales causas de pérdida de cultivos en la agricultura (Zandalinas *et al*., 2021), por lo que la identificación de nuevos reguladores moleculares involucrados en la tolerancia a este tipo de estrés constituye un área de interés estratégico. En este sentido, el estudio de lncRNAs como *AT5G03285* podría abrir la posibilidad de desarrollar, en trabajos futuros, estrategias orientadas a modular su función con fines de mejorar la adaptación de las plantas a condiciones adversas.

Por otra parte, en caso de que la actividad de *AT5G03285* esté relacionada con la señalización del ET, su caracterización podría aportar información valiosa para el diseño de estrategias orientadas a optimizar la respuesta de las plantas frente a estrés biótico, así como para la manipulación de procesos de desarrollo como la senescencia y la maduración de frutos. Este conocimiento tendría implicaciones potenciales en el ámbito agrícola y agroindustrial, particularmente en la prolongación de la vida de anaquel.

En conjunto, la presente investigación se justifica por la necesidad de profundizar en la comprensión funcional de los lncRNAs en plantas, abordando el estudio de *AT5G03285* como un regulador potencialmente relevante en la integración de señales hormonales y en las respuestas al estrés, y contribuyendo así al avance del conocimiento básico y aplicado en biología vegetal.

HIPÓTESIS

El RNA largo no codificante *AT5G03285* participa en la regulación de la respuesta hormonal y al estrés en *A. thaliana*, actuando como un modulador de la señalización por ET y de su interacción con otras fitohormonas, particularmente con ABA, a través de mecanismos regulatorios asociados al componente central EIN2.

OBJETIVOS

Objetivo general

Caracterizar la función del lncRNA *AT5G03285* en *A. thaliana* y su participación en la regulación de la señalización por ET y su interacción con ABA, mediante el análisis de su papel en respuestas hormonales y al estrés, así como su posible relación con el componente central EIN2.

Objetivos específicos

1. Generar y caracterizar fenotípica y genotípicamente líneas mutantes del lncRNA *AT5G03285* en *A. thaliana* bajo tratamientos exógenos de ácido abscísico (ABA) y ácido 1-aminociclopropano-1-carboxilico (ACC), así como en condiciones de estrés abiótico relacionadas con la señalización por ET.

2. Caracterizar la expresión del lncRNA *AT5G03285* en líneas mutantes de genes involucrados en la percepción y respuesta a ABA y ET, bajo tratamientos exógenos de ABA y ACC.

3. Caracterizar la línea reportera del promotor del lncRNA *AT5G03285* bajo tratamientos exógenos de ABA y ACC.

4. Analizar bioinformáticamente la conservación evolutiva del lncRNA *AT5G03285* mediante estudios de sintenia y la predicción de su estructura secundaria.

5. Identificar la localización celular del lncRNA *AT5G03285* en respuesta a tratamientos de ABA y ACC exógenos.

6. Analizar la posible interacción molecular entre el lncRNA *AT5G03285* y la proteína EIN2.

Capitulo 1. Artículo científico derivado de la tesis doctoral

El presente capítulo corresponde al manuscrito titulado "LncRNA-encoded peptides: the case of the lncRNA gene located downstream of EIN2", resultado de la investigación desarrollada en esta tesis doctoral.

Este trabajo fue publicado en la revista Functional & Integrative Genomics bajo la siguiente referencia:

Nieto-Hernández, J., Arenas-Huertero, C., & Ibarra-Laclette, E. (2023). LncRNA-encoded peptides: the case of the lncRNA gene located downstream of EIN2. Functional & Integrative Genomics, 23(2), 108. https://doi.org/10.1007/s10142-023-01038-8

Title: LncRNAs-encoded peptides; the case of the lncRNA-gene located downstream of *EIN2*.

Jesús Nieto-Hernández[1,2], Catalina Arenas-Huertero[2]*, Enrique Ibarra-Laclette[3]*

[1]Facultad de Ciencias Químicas, Universidad Autónoma de San Luis Potosí, Manuel Nava 6, Zona Universitaria, 78290, San Luis Potosí, México.
[2]Laboratorio de Metabolismo del RNA, Facultad de Ciencias, Universidad Autónoma de San Luis Potosí. Av. Chapultepec, 1570, 78295, San Luis Potosí, México.
[3]Red de Estudios Moleculares Avanzados (REMAV), Instituto de Ecología, A.C. (INECOL), Xalapa 91073, Veracruz, México.

*Corresponding author: Enrique Ibarra-Laclette, and Catalina Arenas-Huertero
Email address: enrique.ibarra@inecol.mx, and catalina.arenas@uaslp.mx

Jesús Nieto-Hernández, Orcid ID: 0000-0002-3069-2461
Catalina Arenas-Huertero, Orcid ID: 0000-0003-0830-4810

Enrique Ibarra-Laclette, Orcid ID: 0000-0003-1441-1938

ABSTRACT

Here, by *in silico* analysis, we identified a putative lncRNA-encoded peptide which appears to have an ancestral origin and has remained through plant evolution in a position downstream of the *EIN2* gene.

Keywords: Plant lncRNAs, small ORFs, micro-peptides, translatable lncRNAs

INTRODUCTION

The long non-coding RNAs (lncRNAs) are a group of heterogeneous regulatory RNA greater than 200 nucleotides (nt) in length with a low evident protein-coding. These molecules are important regulators because they can function both in *cis* or in *trans* in the genome or interact with other RNAs and proteins (Nagano and Fraser 2011, Bhogireddy *et al*. 2021), also with the genome position and the relation with other genes, they are classified as an antisense orientation, intronic, promoter and intergenic (Ariel *et al*. 2015). The lncRNAs are transcribed by RNA polymerase II, IV and V; also, their polyadenylation could be missing and located in the nucleus, cytoplasm, or both (Lucero *et al*. 2020).

Although it has been argued that lncRNA and microRNAs transcripts preferentially act as functional RNAs (Ben Amor *et al*. 2009), recently, not only in mammals but also in plants, it has been proved that surely many of them can be also translated into small peptides with specific physiological functions (Pan *et al*. 2018, Xing *et al*. 2020 and Kaur *et al*. 2022). *ENOD40* gene, retained in legumes and other species, is might the first-characterized plant lncRNA with dual function. It is a highly structured lncRNA containing two small open reading frames (sORF) involved in root nodule *Rhizobium* interaction (Gultyaev and Roussis 2007). Considering the impact that could have at functional and regulation levels, here, by *in silico* analysis, we analyze the possible existence of a lncRNA-encoded peptide located downstream of the central component

of ethylene signaling (Ethylene Insensitive 2; *EIN2*), the same is retained on distinct angiosperm plant species and, at least in *Arabidopisis thaliana* and according with different databases, it has some tissues-, and/or developmental stage-specific expression patterns (Jin *et al*. 2013, Waese *et al*. 2017).

RESULTS AND DISCUSSION

The clustering of orthologs (and paralogs) genes shared between some flowering plants was performed using OrthoMCL software v2.0.9 (Li *et al*. 2003). The predicted protein sequences from 64 plant species belonging to 27 distinct orders (Supplementary Table S1), were used for this purpose. Once ortho-groups were formed, we identified the one in which the EIN2 protein from *A. thaliana* (AT5G03280) was grouped. Multiple-sequence alignment performed by MUSCLE (Edgar 2004) was manually reviewed and some erroneous-predicted gene models were corrected. To do so, first, the complete genomic region (or loci) of each *EIN2* gene was extracted (extractseq program from EMBOSS suite v6.6.0; Rice *et al*. 2000), and then, using the software GeneWise (https://www.ebi.ac.uk/Tools/psa/genewise/; Birney *et al*. 2004) and the orthologs protein from the closest species, the correct coding sequences (CDS) were identified. Phylogenetic relationships for 79 EIN2 proteins were resolved using Bayesian inference model (Fig. 1a) (Rambaut *et al*. 2018, Ronquist *et al*. 2012). Besides, for all analyzed species, we surveyed the genomic context of the *EIN2* gene using a reciprocal best BLAST hits and a synteny analysis (SynMap within CoGe; https://genomevolution.org/coge/SynMap.pl). In all analyzed species, a strongly syntenic and colinear block was observed (Supplementary Fig. S1). This block includes a region flanked by the EIN2 protein and either an ortholog of the isocitrate dehydrogenase V (IDHV) or phenylalanine ammonia-lyase (PAL) (Supplementary Fig. S2). With only a few exceptions (*Solanum tuberosum* (n=3), *Capsicum annum* (n=4)), in most of the analyzed species, one or, sometimes two long non-coding RNA (lncRNA) genes, have been identified downstream of *EIN2* gene (Supplementary Fig. S2) within a variable-size intergenic region (ranged from 571 bp in *A. thaliana* to 163,376 bp in *Capsicum annum*) (Supplementary Fig. S2). The number of exons in these lncRNAs-

coding genes ranges from 0 to 4 depending on the species. Their transcripts have an average 2,393 nt length, the shortest of 204 nt in *A. thaliana*, and the longest of 31,157 nt from *Helianthus annuus* species (Supplementary Fig. S2). Interestingly, species with the longest intergenic regions show portions with high similarity to some transposable elements (Supplementary Fig. S2). This was expected mainly in those species which, like *C. annum* a single type of LTR retrotransposon explains most of its genome expansion (Park *et al*. 2012). It is worth mentioning that all these lncRNAs-coding genes were predicted/identified based on transcriptional evidence; the above is stated because we consider in our analysis only plant species whose gene models have been predicted by the Eukaryotic Genome Annotation Pipeline of NCBI (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/process/).

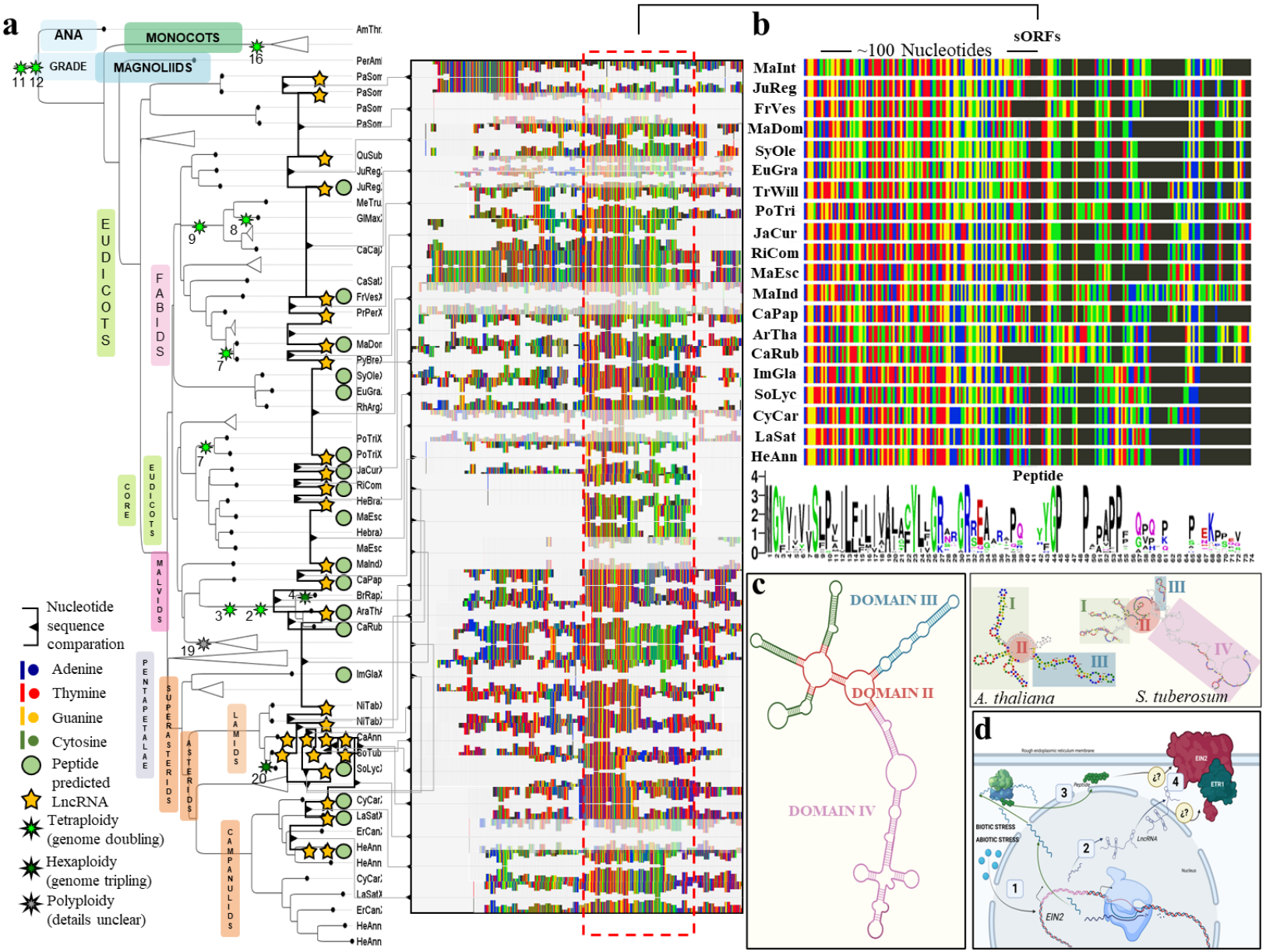

a
ANA
MONOCOTS
GRADE
MAGNOLIIDS
EUDICOTS
FABIDS
CORE EUDICOTS
MALVIDS
PENTAPETALAE
SUPERASTERIDS
ASTERIDS
LAMIIDS
CAMPANULIDS
Nucleotide sequence comparation
Adenine
Thymine
Guanine
Cytosine
Peptide predicted
LncRNA
Tetraploidy (genome doubling)
Hexaploidy (genome tripling)
Polyploidy (details unclear)
b
~100 Nucleotides
sORFs
MaInt
JuReg
FrVes
MaDom
SyOle
EuGra
TrWill
PoTri
JaCur
RiCom
MaEsc
MaInd
CaPap
ArTha
CaRub
ImGla
SoLyc
CyCar
LaSat
HeAnn
Peptide
c
DOMAIN III
DOMAIN II
DOMAIN IV
A. thaliana
S. tuberosum
d
EIN2
ETR1
BIOTIC STRESS
ABIOTIC STRESS
LncRNA

Fig. 1. Solved angiosperm plant phylogenetic tree for *EIN2* and comparative analysis of the located downstream lncRNA-gene. **a** Phylogenetic tree inferred by Bayesian method for a total of 67 angiosperm plant species belonging to 27 distinct orders. The tree was visualized with the FigTree version 1.4.4 software (http://tree.bio.ed.ac.uk/software/figtree/). He abbreviations for each species' name are listed in Supplementary Table S1. Whole-genome duplication (WGD) events (*paleo*polyploidy level) are represented by 8-pointed stars (light-, dark-green and gray), most of them ordered and shown according to information available on CoGepedia web page (https://genomevolution.org/). Besides, those lncRNAs located downstream of *EIN2* were represented by a yellow 5-point star and the ORFs contained within are shown whit a green circle. At the right side of the phylogeny, pairwise alignment for the lncRNA nucleotide (nt) sequences are shown. Each plant species was aligned against its closest evolutionary relative. To find conserved regions in all analyzed species, the product of the pairwise alignments was split into small fragments (10 nt) using the slide window method, and then, they were compared to each other. To generate a visual representation of these alignments ordered based on the highly conserved region (dotted line red box), a colored vector per species was generated. Pair bases are shown as Adenine-blue, Thymine-red, Guanine-yellow, and Cytosine-green. Color schemes to plot lncRNA alignments were graphed using Rstudio (v1.1453) and the ggplot2 package (Wickham 2016). **b** Global multiple alignments of the small ORFs (sORF) encoded within lncRNA-genes is also shown. At the top, the ~100 highly conserved nt are delimited, and at the bottom, the sequence LOGO (Crooks *et al*. 2004) generated from the aligned translated peptides is shown. **c** Secondary structures obtained for lncRNAs located downstream of *EIN2*. On left side, and as reference, a consensus structure is shown; on it, the domains I-IV are highlighted (green, red, blue, and purple, respectively). Highly, poorly, and moderately conserved regions, i.e., domains I-III, belongs to the small peptide encoded into the lncRNA. On the right side, and as examples, the secondary structures of the lncRNAs from *A. thaliana* and *S. tuberosum* are shown. Notice that structural similarities consist of three- (sometimes four) linear stems (domain I) joined by a loop (domain II) with an additional linear stem. Besides,

this conserved structure is complemented by a highly variable portion (domain IV), which is absent in *A. thaliana*, and it is the pieces less conserved across all compared species. **d** Diagram for the hypothesized mechanism of action from lncRNA located downstream of *EIN2*. In response to changing biotic or abiotic conditions in which ethylene signaling is involved (1), the lncRNA maybe is transcribed as such and could act once the secondary structure is folded (2) or maybe together with the peptide encoded in the sORF contained on it (3); in both cases, (4) either by direct interaction with EIN2, or additional elements part of the EIN2-ETR1 receptor complex (Bisson and Groth 2011). **c** and **d** diagrams were Created with BioRender.com.

Based on the high percent (~60%) of plant species in which lncRNAs coding genes were present within analyzed intergenic region (Supplementary Fig. S2), we suggest that at least one lncRNA-coding gene downstream of *EIN2* is "conserved/retained" across all (or the majority) of angiosperm plant species. We do not entirely discard that its absence could be related to some evolutionary process (e.g., fractionation or gene loss, among others; see Ibarra-Laclette *et al*. 2013); however, it is highly probable that its absence is due to that any lncRNAs annotation process has not yet been performed (e.g., *Persea americana* genome) or its prediction still is deficient due the transcriptional evidence used during its annotation processes and which corresponds to collections of low-depth sequenced transcriptomes. Notice that it has been reported that lncRNAs show mostly tissue-, cell-, or developmental stage-specific expression, and they are differentially expressed in response to biotic or abiotic stress conditions (Wang and Chekanova 2017). It also was interesting to notice that in a considerable number of species (29), *EIN2* is present as a single copy gene but in some species with multiple copies (paralogs), the presence of a lncRNA-gene downstream of *EIN2* was observed preferentially in only one of the copies (Fig. 1a: e.g., Campanulids taxonomic clade).

Considering the not-uncommon rate of whole-genome duplications (WGDs) events (or *paleo*polyploidy events) that have taken place throughout the evolution in the different lineages of plants (Supplementary Fig. S3), we suggest that it is highly probable that the lncRNAs-coding genes located downstream of *EIN2* arose from an ancient protein-coding gene that, after a WGD and due to faster rates of evolution that

take place in duplicated genes, evolved to a functional lncRNA (neofunctionalization process) (Barrera-Redondo *et al*. 2018). Interestingly, this lncRNA syntenic-retained in several plant species shows a lack of conservation at the sequence level. Despite this lack of conservation (Fig. 1a), we found, in all these lncRNAs, a small region ranged from 165 to 216 nt and corresponded to an sORF (or peptide) (Fig. 1a-b). This region, when is compared across distinct plant species, shows a high similarity at the nucleotide and amino acids sequences level (≥ 55 identity %; Fig. 1b) as well as at the secondary structure level (Fig. 1c). Once lncRNAs secondary structures are folded, they typically show four domains (I-IV). Domain I is composed of ~100 nt, is highly conserved and is part of the start of the peptide synthetized by the sORF (C-terminus). At the structural level, domain I form three or four stem structures. Meanwhile, domains II and III correspond to the second half of the sORF (N-terminus), and they are poorly and/or moderately conserved at the nucleotide level. Domain II, at the structural level, is a segment of unpaired nucleotides; that is, it forms a loop that joins the stems formed by domain I with an additional stem formed by domain III. Finally, in the most of species (but not in *A. thaliana*), domain IV is a no conserved region and without evident similarity, neither in sequence level nor at a structural level (Fig. 1c).

The ≈65 amino acids peptides encoded into the lncRNAs located downstream of *EIN2* are enriched with hydrophobic amino acids, mainly in the highly conserved region (Domain I). Interestingly, this region in the sORFs seems to be a transmembrane domain which locates the peptide in the endoplasmic reticule (Supplementary Fig. S4). This feature in the peptides encoded by sORF firmly caught our attention because even when it has been argued that lncRNAs show poor evolutionary conservation and that those sORFs contained in them could be subject to substantial purifying selection (Fensenko *et al*. 2021), it is also true that it has been proven that the few lncRNAs-encoded peptides which are conserved among distinct species, on average, show higher transcriptional level than non-conserved sORFs; they are usually enriched with AT-rich codons encoding hydrophobic amino acids and contain low-complexity regions or transmembrane domains (Fensenko *et al*. 2021).

Since a few species lacks the sORF on the transcribed lncRNA but still show a high identity percent at nucleotide sequence level within the small-conserved region part of them, we suggest that peptide, presumably produced by the sORF it has been retained through evolution and is a reminiscence of the protein-coding region of the ancestral gene which gave rise to the lncRNA-coding gene. Despite mentioned above, it is possible that the retention of the lncRNA-coding gene downstream of *EIN2*, possesses functional activities at the structural level but also like a peptide. This hypothesis can be reinforced if we consider those species in which the lncRNA-coding gene (and sORF is contained) located downstream of *EIN2*, is duplicated, which suggests that new copies birth later in the evolution by gene duplication processes (Supplementary Fig. S2: e.g., *C. annuum* (4), *S. tuberosum* (3), *Mangifera indica* (2).

CONCLUSIONS

The performed analyses show that a putative lncRNA is conserved downstream of the *EIN2* gene, containing an sORF (Fig. 1a-b). The first 100 nucleotides in the sORF are more conserved than the rest at (nt) and peptide levels, suggesting being an important domain whether it functions either as a peptide or a lncRNA. A transmembrane domain was identified in the peptide, which presumably locates it in the endoplasmic reticulum (Supplementary Fig. S4). At the structural level, the two conserved domains (I and III) form almost linear stem-loop structures (Fig. 1c), hence could be necessary for the function meanwhile, another domain (domain IV) tends to be variable between species (Fig, 1c: e.g., *S. tuberosum*). Considering some well-characterized lncRNAs in animals and plants that act on its neighboring *locus* (e.g., Ariel *et al*. 2020, Seo *et al*. 2017 and Engreitz *et al* 2016), the conservation and closeness of the analyzed lncRNA-gene suggests that its function could be related to the activity of EIN2 (data no show). This lncRNA-gene could be transcribed and function both as mRNA translatable into a peptide or, once folded, the transcript itself can take a functional secondary structure and have a relevant biological role (Fig. 1d). The presence of transmembrane domain in the peptide encoded in the lncRNAs is a relevant finding because EIN2, the central regulator of ethylene signaling, is localized

at the ER membrane where it interacts with the ethylene receptor ETR1 (Bisson *et al*. 2009). Likewise, the highly conserved portion in the lncRNA, which is retained in many plant species (and which represents an sORF), could be the product of an "evolutionary strategy" which selects functional portions of emerging genes which acquire new functions as lncRNA (and may preserve reminiscences of protein-coding regions) after a WGD event. Finally, considering that in *A. thaliana* the lncRNA-gene located downstream of *EIN2* is the smallest gene from analyzed plant species that possess an equivalent gene. We suggest that, as such, this lncRNA which also encodes a small peptide, might represent the most compressed version of this gene which evolved in a species with a clearly contracted genome (Oyama *et al*. 2008).

DECLARATIONS SECTION

Ethical Approval and Consent to participate.

Not applicable.

Human and Animal Ethics

Not applicable.

Consent for publication

Not applicable.

Availability of supporting data

All data generated or analyzed during this study were included in this published article (and its supplementary information files).

Competing interests

The authors declare that they have no conflict of interest.

Funding

This work was supported by Fondo de Apoyo al Investigador, C20-FAI-10-05.05, Universidad Autónoma de San Luis Potosí, México (awarded to Catalina Arenas Huertero).

Authors' contributions

Conceptualization: C.A.-H., E.I.-L. Data curation: J.N.-H. Formal Analysis: J.N.-H and E.I.-L. Funding acquisition: C.A.-H. Writing – original draft: J.N.-H, C.A.-H., E.I.-L. Writing – review & editing: all authors.

Acknowledgements

The authors wish to thank the Institute of Ecology A.C. (INECOL) for allow accessing to the high-performance computing system (HUITZILIN) required to carry out this research. We thank Emanuel Villafán for his assistance as the administrator of this resource and Mizraim Olivares for his valuable discussion on several of the performed analyses. Finally, we also thank to CONACyT for the scholarship provided to Jesus Nieto-Hernández to perform his postgraduate studies.

Authors' information

First author: Jesús Nieto-Hernández[1,2]

Corresponding author: Enrique Ibarra-Laclette[3]; Orcid ID: 0000-0003-1441-1938, Catalina Arenas-Huertero[2]; Orcid ID: 0000-0003-0830-4810

[1]Facultad de Ciencias Químicas, Universidad Autónoma de San Luis Potosí, Manuel Nava 6, Zona Universitaria, 78290, San Luis Potosí, México.

[2]Laboratorio de Metabolismo del RNA, Facultad de Ciencias, Universidad Autónoma de San Luis Potosí. Av. Chapultepec, 1570, 78295, San Luis Potosí, México.

[3]Instituto de Ecología, A.C. (INECOL), Red de Estudios Moleculares Avanzados (REMAV), Xalapa, Veracruz, 91073, México.

REFERENCES

Ariel F, Lucero L, Christ A, Mammarella MF, Jegu T, Veluchamy A, Mariappan K, Latrasse D, Blein T, Liu C, Benhamed M, Crespi M (2020) R-loop mediated *trans* action of the APOLO long noncoding RNA. Molecular cell 77:1055-1065.e1054. https://doi.org/ 10.1016/j.molcel.2019.12.015

Ariel F, Romero-Barrios N, Jégu T, Benhamed M, Crespi M (2015) Battles and hijacks: noncoding transcription in plants. Trends in plant science 20:362-371. https://doi.org/ 10.1016/j.tplants.2015.03.003

Barrera-Redondo J, Ibarra-Laclette E, Vázquez-Lobo A, Gutiérrez-Guerrero YT, Sánchez de la Vega G, Piñero D, Montes-Hernández S, Lira-Saade R, Eguiarte LE (2019) The genome of *Cucurbita argyrosperma* (silver-seed gourd) reveals faster rates of protein-coding gene and long noncoding RNA turnover and neofunctionalization within *Cucurbita*. Molecular plant 12:506-520. https://doi.org/ 10.1016/j.molp.2018.12.023

Ben Amor B, Wirth S, Merchan F, Laporte P, d'Aubenton-Carafa Y, Hirsch J, Maizel A, Mallory A, Lucas A, Deragon JM, Vaucheret H, Thermes C, Crespi M (2009) Novel long non-protein coding RNAs involved in *Arabidopsis* differentiation and stress responses. Genome Res 19:57-69. https://doi.org/ 10.1101/gr.080275.108

Bhogireddy, S., Mangrauthia, S. K., Kumar, R., Pandey, A. K., Singh, S., Jain, A., Budak, H., Varshney, R. K., & Kudapa, H. (2021). Regulatory non-coding RNAs: a new frontier in regulation of plant biology. Functional & integrative genomics, 21(3-4), 313–330. https://doi.org/ 10.1007/s10142-021-00787-8Birney E, Clamp M, Durbin R (2004) GeneWise and Genomewise. Genome Res 14:988-995. https://doi.org/ 10.1101/gr.1865504

Bisson MM, Bleckmann A, Allekotte S, Groth G (2009) EIN2, the central regulator of ethylene signalling, is localized at the ER membrane where it interacts with the ethylene receptor ETR1. The Biochemical journal 424 (1):1-6. https://doi.org/10.1042/bj20091102

Bisson MM, Groth G (2011) New paradigm in ethylene signaling: EIN2, the central regulator of the signaling pathway, interacts directly with the upstream receptors. Plant signaling & behavior 6:164-166. https://doi.org/ 10.4161/psb.6.1.14034

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14:1188-1190. https://doi.org/ 10.1101/gr.849004

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792-1797. https://doi.org/ 10.1093/nar/gkh340

Engreitz, J. M., Haines, J. E., Perez, E. M., Munson, G., Chen, J., Kane, M., McDonel, P. E., Guttman, M., & Lander, E. S. (2016). Local regulation of gene expression by lncRNA promoters, transcription and splicing. Nature, 539(7629), 452–455. https://doi.org/ 10.1038/nature20149

Fesenko I, Shabalina SA, Mamaeva A, Knyazev A, Glushkevich A, Lyapina I, Ziganshin R, Kovalchuk S, Kharlampieva D, Lazarev V, Taliansky M, Koonin EV (2021) A vast pool of lineage-specific microproteins encoded by long non-coding RNAs in plants. Nucleic Acids Res 49 (18):10328-10346. https://doi.org/ 10.1093/nar/gkab816

Gultyaev AP, Roussis A (2007) Identification of conserved secondary structures and expansion segments in enod40 RNAs reveals new enod40 homologues in plants. Nucleic Acids Res 35:3144-3152. https://doi.org/ 10.1093/nar/gkm173

Ibarra-Laclette E, Lyons E, Hernandez-Guzman G, Perez-Torres CA, Carretero-Paulet L, Chang TH, Lan T, Welch AJ, Juarez MJ, Simpson J, Fernandez-Cortes A, Arteaga-Vazquez M, Gongora-Castillo E, Acevedo-Hernandez G, Schuster SC, Himmelbauer H, Minoche AE, Xu S, Lynch M, Oropeza-Aburto A, Cervantes-Perez SA, de Jesus Ortega-Estrada M, Cervantes-Luevano JI, Michael TP, Mockler T, Bryant D, Herrera-Estrella A,

Albert VA, Herrera-Estrella L (2013) Architecture and evolution of a minute plant genome. Nature 498:94-98. https://doi.org/ 10.1038/nature12132

Jin J, Liu J, Wang H, Wong L, Chua NH (2013) PLncDB: plant long non-coding RNA database. Bioinformatics (Oxford, England) 29:1068-1071. https://doi.org/ 10.1093/bioinformatics/btt107

Kaur, S., Kumar, S., & Mohapatra, T. (2022). MicroRNA: noncoding but still coding, another example of self-catalysis. Functional & integrative genomics, 23(1), 4. https://doi.org/ 10.1007/s10142-022-00926-9

Li L, Stoeckert CJ, Jr., Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13:2178-2189. https://doi.org/ 10.1101/gr.1224503

Lucero L, Fonouni-Farde C, Crespi M, Ariel F (2020) Long noncoding RNAs shape transcription in plants. Transcription 11:160-171. https://doi.org/ 10.1080/21541264.2020.1764312

Nagano T, Fraser P (2011) No-nonsense functions for long noncoding RNAs. Cell 145:178-181. https://doi.org/ 10.1016/j.cell.2011.03.014

Oyama RK, Clauss MJ, Formanová N, Kroymann J, Schmid KJ, Vogel H, Weniger K, Windsor AJ, Mitchell-Olds T (2008) The shrunken genome of *Arabidopsis thaliana*. Plant Systematics and Evolution 273:257-271. https://doi.org/ 10.1007/s00606-008-0017-z

Pan J, Meng X, Jiang N, Jin X, Zhou C, Xu D, Gong Z (2018) Insights into the noncoding RNA-encoded peptides. Protein and peptide letters 25:720-727. https://doi.org/ 10.2174/0929866525666180809142326

Park M, Park J, Kim S, Kwon J-K, Park HM, Bae IH, Yang T-J, Lee Y-H, Kang B-C, Choi D (2012) Evolution of the large genome in *Capsicum annuum* occurred through accumulation of single-type long terminal repeat retrotransposons and their derivatives. The Plant Journal 69:1018-1029. https://doi.org/ 10.1111/j.1365-313X.2011.04851.x

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systematic biology 67:901-904. https://doi.org/ 10.1093/sysbio/syy032

Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends in genetics : TIG 16:276-277

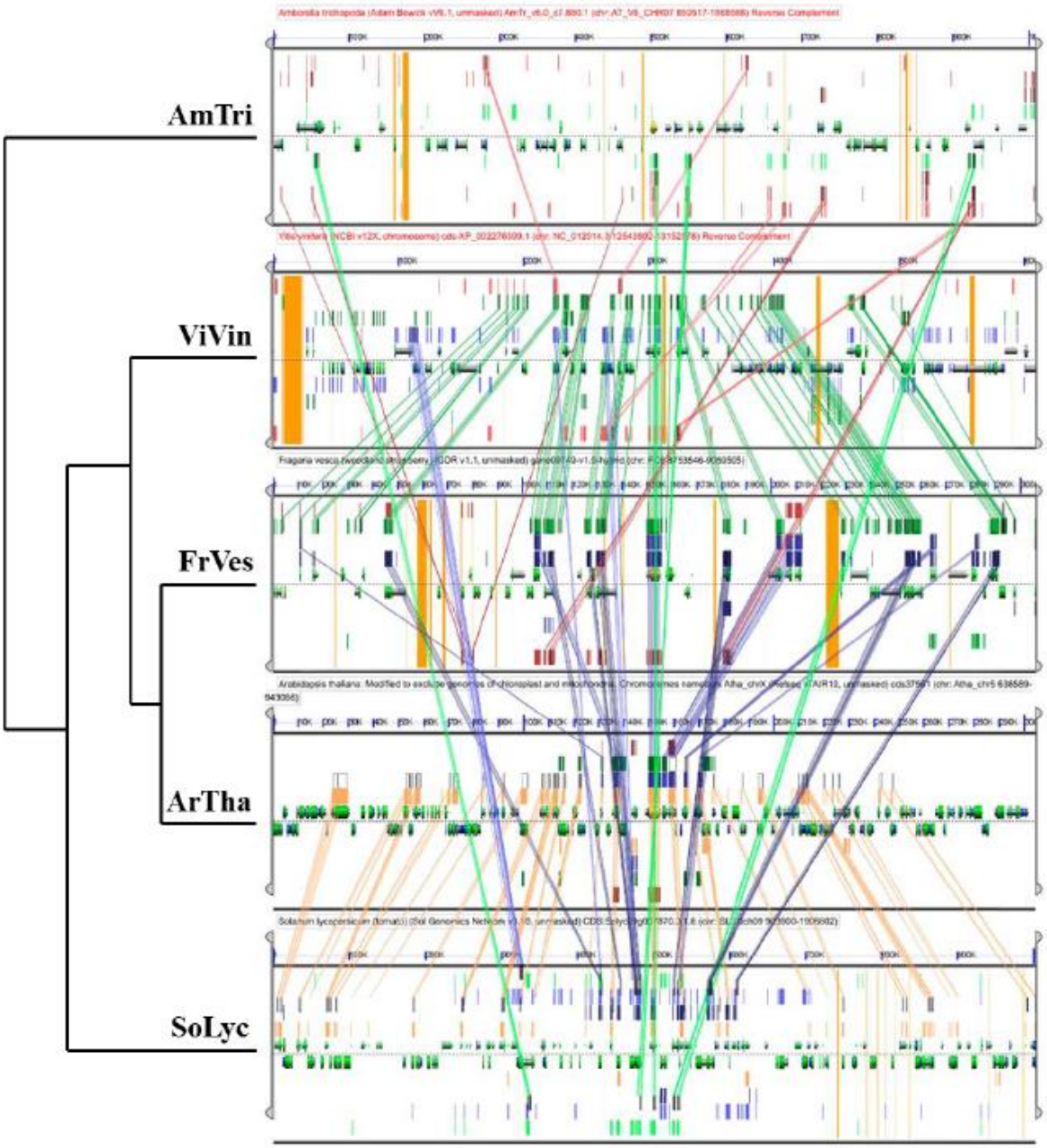
Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic biology 61:539-542. https://doi.org/ 10.1093/sysbio/sys029

Seo JS, Sun HX, Park BS, Huang CH, Yeh SD, Jung C, Chua NH (2017) ELF18-INDUCED LONG-NONCODING RNA associates with mediator to enhance expression of innate immune response genes in *Arabidopsis*. The Plant cell 29:1024-1038. https://doi.org/ 10.1105/tpc.16.00886

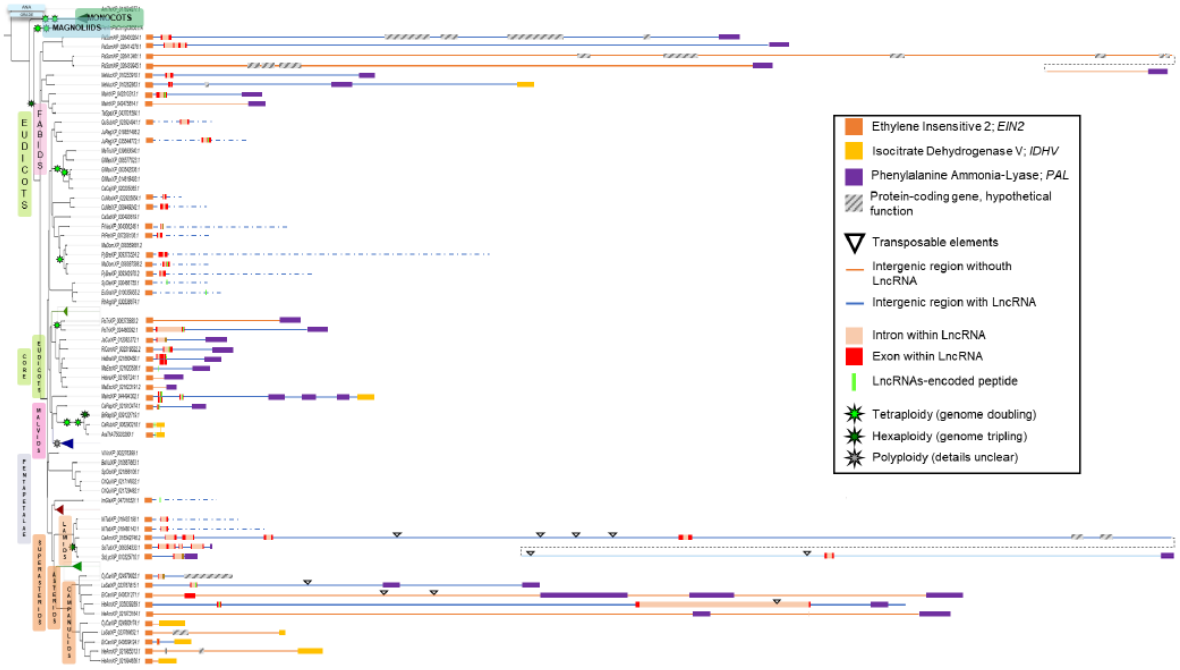
Waese J, Fan J, Pasha A, Yu H, Fucile G, Shi R, Cumming M, Kelley LA, Sternberg MJ, Krishnakumar V, Ferlanti E, Miller J, Town C, Stuerzlinger W, Provart NJ (2017) ePlant: visualizing and exploring multiple levels of data for hypothesis generation in plant biology. The Plant cell 29:1806-1821. https://doi.org/ 10.1105/tpc.17.00073

Wang HV, Chekanova JA (2017) Long noncoding RNAs in plants. Advances in experimental medicine and biology 1008:133-154. https://doi.org/ 10.1007/978-981-10-5203-3_5

Wickham H (2016) ggplot2: elegant graphics for data analysis. Springer International Publishing Xing J, Liu H, Jiang W, Wang L (2020) LncRNA-encoded peptide: functions and predicting methods. Frontiers in oncology 10:622294. https://doi.org/ 10.3389/fonc.2020.622294


AmTri
ViVin
FrVes
ArTha
SoLyc

Supplementary Fig. S1. Syntenic blocks from different angiosperm plant species with a no high close evolutionary relationship [*Amborella trichocarpa* (Amborellales; Amborellaceae), *Vitis vinifera* (Vitales; Vitaceae), *Fragaria vesca* (Rosales; Rosaceae), *Arabidopsis thaliana* (Brassicales; Brassicaceae), *Solanum lycopersicum* (Solanales; Solanaceae)] but which containing an ortholog single-copy EIN2 gene (used as anchor gene and shown in yellow). These genomic regions are strongly syntenic with extensive collinear arrangement which also contain a considerable number of orthologous gene pairs. Notice that gene islands are more compact and much higher in gene density in those species with smaller genomes [e.g., *A. thaliana* or *F. vesica* (135and 240 Mpb, respectively) versus *V. vinifera* (500 Mpb) or even versus *A. trichocarpa* (870 Mpb) or *S. lycopersicum* (950 Mpb)]. Genomic windows shown are ranged from 300 Kbp to 1 Mbp. Lines have been drawn connecting syntenic genomic regions or even gene pairs. Abbreviations: AmTri; *A. trichocarpa*, ViVin; *V. vinífera*, FrVes; *F. vesca*, ArTha; *A. thaliana*, and SoLyc *S. lycipersicum*.



Orange blocks indicate gaps (unsequenced NTs). This analysis may be regenerated by CoGe at https://genomevolution.org/r/1nvwn

Supplementary Fig. S2. Intergenic region flanked by the Ethylene Insensitive 2 (EIN2) protein and either an ortholog of the isocitrate dehydrogenaseV (IDHV) or phenylalanine ammonia-lyase (PAL). The length of lines representing intergenic regions represent, at scale, the distance (in bp) contained between flanked genes. Theses intergenic regions contain a lncRNA-gene downstream of *EIN2*sometimes containing introns and/or sometimes containing and small ORF (sORF) codifying for a putative peptide of around 65 amino acids. Notice that in some species, this lncRNA have multiple copies, e.g., *C. annuum* (4), *S. tuberosum* (3), *Mangifera indica* (2), among others. These copies seem have arisen independently to whole genome duplications events history, which suggest that might these orthologs (and paralogs) are involved in a key functional role still unknown which provide functional redundancy to strengthen the fitness. Names of analyzed species and they abbreviations in the phylogeny are listedin Supplementary Table S1.

Capítulo 2. Artículo científico derivado de la tesis doctoral

El presente capítulo corresponde al manuscrito titulado "A synteny conserved TATA binding protein 2-associated long intergenic non-coding RNA influences leaf development and ABA response in Arabidopsis thaliana", resultado de la investigación desarrollada en esta tesis doctoral.

Este trabajo fue publicado en la revista Plant and Cell Physiology bajo la siguiente referencia:

Espinoza-López, B. S., Nieto-Hernández, J., Quezada-Eguía, Á. E., Vargas-Camacho, S. I., Aportela-Cortez, J., Reyes, J. L., & Arenas-Huertero, C. (2026). A synteny conserved TATA binding protein 2-associated long intergenic non-coding RNA influences leaf development and ABA response in Arabidopsis thaliana. Plant and Cell Physiology. Advance online publication. https://doi.org/10.1093/pcp/pcag040

Title: A Synteny Conserved TATA Binding Protein 2-associated Long Intergenic Non-coding RNA Influences Leaf Development and ABA Response in Arabidopsis thaliana.

Byanka Sthefany Espinoza-López [1], Jesús Nieto-Hernández [1], Ángel Eduardo Quezada Eguía [2], Susana Isabel Vargas Camacho [1], Jaime Aportela Cortez [2], José Luis Reyes [3], Catalina Arenas-Huertero [2] *

[1]Posgrado en Ciencias en Bioprocesos, Facultad de Ciencias Químicas, Universidad Autónoma de San Luis Potosí, Av. Dr. Salvador Nava Mtz. s/n, Zona Universitaria, C.P. 78290, San Luis Potosí, S.L.P., Mexico.

[2]Facultad de Ciencias, Universidad Autónoma de San Luis Potosí, Av. Chapultepec #1570, Priv. del Pedregal., San Luis Potosí, S.L.P., C.P. 78295, Mexico.

[3]Departamento de Biología Molecular de Plantas, Instituto de Biotecnología, UNAM, Av. Universidad 2001, Cuernavaca, CP, 62210, Mexico.

*Corresponding author, catalina.arenas@uaslp.mx

Byanka Sthefany Espinoza-López, ORCID ID: 0009-0008-6413-6720

Jesús Nieto-Hernández, ORCID ID: 0000-0002-3069-2461

Ángel Eduardo Quezada Eguía, ORCID ID: 0009-0003-7818-4823

Susana Isabel Vargas Camacho, ORCID ID: 0009-0001-6735-5055

Jaime Aportela Cortez, ORCID ID: 0009-0004-2846-4744

José Luis Reyes, ORCID ID: 0000-0001-5129-9741

Catalina Arenas-Huertero, ORCID ID: 0000-0003-0830-481

ABSTRACT

Long non-coding RNAs are molecules with regulatory potential involved in many biological processes of plants. Despite their significant role, only a limited number have been studied in detail. Here, we describe at the molecular and bioinformatic levels a *TBP2-associated long intergenic non-coding RNA* (*TALIR*) encoded upstream of *TATA-binding protein 2* (*TBP2*) of *Arabidopsis thaliana.* Downregulation of *TALIR* leads to plants with reduced leaf and rosette size, as well as decreased lateral root density. The expression of *TBP2* in *talir* mutant lines is increased, which affects the timing of expression of *MIR396* and *GRF2*, as well as *MIR156* and *SPL9*. Therefore, these dysregulations may, in part, contribute to the altered leaf size observed. Transcriptional evidence from *TBP2 loci* suggests the presence of a putative *TALIR* ortholog gene conserved across the Brassiceae, Eutremeae, and Camelineae tribes at the DNA sequence and synteny level. Furthermore, *TALIR* is upregulated upon ABA and NaCl treatments in WT plants, whereas *talir* mutants exhibit insensitivity to ABA during germination. Together, these findings expand current knowledge of lncRNAs and provide new insights into the regulation of *MIR396* and *MIR156* in the control of leaf growth. Further, the evolutionary conservation of *TALIR* among tribes within the Brassicaceae family suggests a conserved functional role in this group of plants.

Keywords

Abiotic stress, *Arabidopsis thaliana*, Brassicales, growth control, long non-coding RNAs, microRNAs

INTRODUCTION

Non-coding RNAs (ncRNAs) are biomolecules with little or no capacity to translate into proteins that regulate many aspects of plant life. The long non-coding RNAs (lncRNAs) and microRNAs (miRNAs) are two classes of natural ncRNAs that regulate gene expression at a transcriptional and post-transcriptional level, respectively, controlling various biological processes in plants (Wu *et al*.,2020). This classification is based on ncRNA size. MiRNAs are typically 20 – 24 nucleotides inlength and have been identified in 88 plant species. The genes encoding miRNAs contain in their promoter the TATA-box core promoter element and transcription factor binding motifs which modulate their transcription. They execute their function by base-pairing to their target transcripts and directing translation repression or inducing mRNA cleavage (Meng *et al*., 2021). In contrast, lncRNAs are molecules with a length longer than 200 nt. They generally exhibit lower expression levels than protein-coding genes, but their expression can be cell- and tissue-specific as well as stage-dependent. Additionally, lncRNAs show low nucleotide sequence conservation but may be conserved in secondary structure, function and their transcription from syntenic *loci* (Diederichs, 2014; Wu *et al*., 2020). LncRNAs are grouped based on their location and orientation relative to the adjacent or overlapping coding genes, to include sense lncRNA, natural antisense lncRNA, intronic lncRNA, and intergenic lncRNA (Chen *et al*., 2020). LncRNAs contribute to genome organization and the control of gene activity in *cis* and *trans* through multiple mechanisms which depend on interactions with other macromolecules such as chromatin remodelers, regulatory proteins, or lncRNA-DNA and lncRNA-RNA interactions (Wu *et al*.,2020). Both miRNAs and lncRNAs play crucial roles in nearly every aspect of plant development and responses to biotic and abiotic stress (Chen *et al*., 2020; Meng *et al*., 2021). Plant development is a finely tuned process, regulated at the molecular level and influenced by environmental factors. Leaf development involves a series of steps: leaf initiation, growth and expansion and maturation culminating in the final size and shape of the leaf. miR396 and miR156 are master modulators of these processes. Specifically, miR396 regulates members of the

GROWTH- REGULATING FACTOR (*GRF*) gene family at the post-transcriptional level controlling cell proliferation (Rodriguez *et al*., 2010). In addition, miR156 modulates the expression of the *SQUAMOSA BINDING PROTEIN* (*SPL*) gene members resulting in the vegetative phase change control (Wu *et al*., 2009). Additionally, lncRNAs can serve as a precursor of microRNAs, one example is bra-*miR156aHG*, a precursor lncRNA of *miR156a*, which modulates leaf morphology in *Brassica campestris* and *Arabidopsis thaliana* (Zhou *et al*., 2023).

Moreover, leaf development is also regulated by other lncRNAs, such as *TWISTED LEAF* (*TL*), a lncRNA transcribed from the opposite strand of its target gene, *OsMYB60,* and acts in maintaining leaf blade flattening in rice (*Oryza sativa*) (Liu *et al*., 2018). Other examples are the natural antisense lncRNAs *NAT1*UGT73C6 and *NAT2*UGT73C6, transcribed from the UDP-glycosyltransferase gene, *UGT73C6*, which act in *trans* by modulating the expression of transcription factors involved in leaf growth (Meena *et al*., 2023). Additionally, lncRNAs are involved in root development, for instance, the *AUXIN REGULATOR ELEMENT DOWNSTREAM SOLITARYROOT* (*ARES*) which acts as a modulator of the auxin response by regulating, in *trans*, genes involved in lateral root formation (Roulé T., *et al*., 2023). Another lncRNA implicated in this process is *APOLO*, which modulates the expression of its neighboring gene, *PID,* through the formation of a chromatin loop between the *APOLO* locus and the *PID* promoter. *APOLO* also controls gene expression in *trans* through sequence complementarity and R-loop formation (Lucero *et al*., 2021). In contrast, only a few lncRNAs are known to be involved in plant stress responses. For instance, *DRIR* is induced by abscisic acid (ABA) treatment and enhances plant drought tolerance by prompting the expression of stress-response genes (Qin *et al*., 2017). The lncRNA *ARTA* is also induced by ABA, and it is implicated in plant responses by modulating *ABI5* expression resulting in a reduced plant drought tolerance. (Yang *et al*., 2023). In *A. thaliana*, more than 6 000 intergenic lncRNAs and 37 238 sense-antisense lncRNAs have been identified (Wang *et al*, 2014). However, only a limited number of these have been biologically characterized. Further researc is needed to characterize their biological significance, as well as to determine

their conservation in other plant species and integrate them into gene regulatory networks involving other molecules such as microRNAs.

In this study, we have characterized a lncRNA, *TBP2-associated long intergenic non-coding RNA* (*TALIR*), required for proper leaf size, morphology and ABA response in *A. thaliana* by modulating the abundance of *TATA-binding protein* 2 (*TBP2*), a protein involved in the basal transcription of genes that contain the TATA-box motif. The absence of *TALIR* alters the timing of *TBP2* expression. This alteration leads to the dysregulation of *MIR396* and *MIR156*, resulting in a reduced rosette size. Additionally, *TALIR* is evolutionarily conserved in DNA sequence and synteny across Brassicales suggesting a shared ancestral origin and common function. Our findings provide new insights into the role of lncRNAs in modulating gene expression and their involvement in leaf and root development and ABA response, expanding the repertoire of lncRNA functions in plants.

RESULTS

Temporal Regulation of Arabidopsis Development by TALIR

We characterized a new long intergenic non-coding RNA (lincRNA) from *Arabidopsis thaliana* (*A. thaliana*). We identified this gene through a plant lncRNA database (Jin *et al*., 2021). It is located downstream of the *E3 RINGU* gene (At1g55530), which encodes an uncharacterized E3 ubiquitin ligase (Takahashi *et al*., 2009), and upstream of the *TATA-box binding protein 2* (TBP2, At1g55520) gene (Heard *et al*., 1993). Based on its genomic position, we named it *TBP2*-associated long intergenic non-coding RNA (*TALIR*, At1g55525) encoding a 417 nt transcript (Supplementary Fig. 1).

We carried out a phenotyping analysis of different mutant lines, one based on RNA interference (*talir-Ri*), a second one using an artificial microRNA line (*talir-amiR*) and another one is a T-DNA insertion mutant line (*talir-1*) (Fig. 1a) to obtain information about plant development. To validate that these mutant lines resulted in altered levels of *TALIR*, we evaluated its relative transcript accumulation by quantitative RT-PCR

(RT-qPCR), confirming that its abundance levels were down- regulated (Fig. 1b). Detailed observations of *talir-1*, *talir-Ri*, and *talir-amiR* plants at 21 days after sowing (DAS) showed a smaller rosette size compared to WT plants. These observations were consistent with the quantitative data of the rosette area in mutant and WT plants (Fig. 1c, e). Furthermore, we observed a reduced number of leaves in *talir-Ri* and *talir-amiR* plants compared to the WT. In contrast, *talir-1* plants had a similar number of leaves as WT, but their leaves were smaller (Fig. 1d, f). Leaf morphology is a characteristic trait of the *A. thaliana* shoot; thus, we observed morphological variations in the fifth leaf in mutant plants. They exhibited reduced elongation and lacked the typical adult leaf serrations seen in the WT phenotype (Poethig, 2013). These features were more evident in plants at 21 DAS (Fig. 1d). However, the reduction in size and total area of the rosettes persisted in the mutant plants over time, as observed at 24 and 27 DAS (Supplementary Fig. 2 and 3). These findings suggest that *TALIR* plays a role in modulating leaf growth.

To determine if *TALIR* has a role in root development, we evaluated the primary root growth during seven days post germination (DPG). This assay revealed no difference in primary root length between mutants and WT plants. Nevertheless, microscopic examination of cleared roots revealed that mutant plants developed fewer lateral roots than WT plants. Quantification of lateral root primordia and emerged lateral roots indicated that *TALIR* primarily affects lateral root formation rather than the emergence process (Supplementary Fig. 4). Consistent with a proposed role in leaf and root development, we found that transcripts of both *TALIR* and one of its potentially regulated targets TBP2, are present in these two organs (Supplementary Fig. 5).

To further study the involvement of *TALIR* in the development of *A. thaliana*, we focused on leaf size and morphology. There are several factors participating in this process such as *ANGUSTIFOLIA3* (*AN3*), *AINTEGUMENTA* (*ANT*), *CUP SHAPED COTYLEDON2* (*CUC2*), and *PIN FORMED1* (*PIN1*) (Ali *et al*., 2020). Furthermore, it has also been described that microRNAs contribute to leaf development. Additionally, several lncRNAs are involved in plant growth by regulating the transcription of

neighboring genes (Lucero *et al*., 2021; Roulé *et al*., 2023). Thus, we investigated whether *TALIR* expression is associated with that of its neighboring *loci* (Fig. 2a). To this end, we measured the relative transcript levels of *TBP2*, *E3 RINGU*, and *TBP1* (At3g13445) at 21, 24 and 27 DAS. To evaluate the relative expression of these genes, we selected two mutant lines, which interfere with *TALIR* activity at the transcriptional (T-DNA) and post-transcriptional (amiR) levels. Due to the evident changes observed in the fifth leaf during the juvenile-to-adult transition (Fig. 1d), we chose this leaf for expression analysis at different developmental stages.

We found that at 21 DAS, *E3 RINGU* and *TBP2* mRNA accumulation was diminished in both lines. In contrast, *TBP1*, a gene located in a different chromosome but a homolog of *TBP2*, showed no changes in relative abundance in the mutant lines tested (Fig. 2b). At 24 DAS, *E3 RINGU* did not show changes in its accumulation, while the relative accumulation of *TBP2* and *TBP1* transcripts was upregulated in both genotypes. (Fig. 2c). At 27 DAS, E3 *RINGU* showed reduced accumulation in both genotypes. On the other hand, *TBP2* did not show changes in its abundance while *TBP1* displayed increased abundance compared to WT (Fig. 2d). These results suggest that *TALIR* affects the transcription of *TBP2* transiently during leaf development. In addition, it is possible that *TBP2* levels affect positively *TBP1* accumulation, suggesting that the role of *TALIR* on *TBP1* expression is indirect.

Moreover, some MIRNA genes contain TATA- box motifs in their promoter regions and it has been described that *TBP2* interacts with TOUGH (TGH), a protein involved in miRNA biogenesis (Xie *et al*., 2005; Calderon-Villalobos *et al*., 2005), thus we explored their potential interconnection. The *miR396* has been described as a key modulator that regulates leaf size by controlling cell proliferation through post-transcriptional regulation of the *GROWTH REGULATING FACTOR* (*GRF*) gene family (Rodriguez *et al*., 2010). Meanwhile, miR156 is involved in regulating the juvenile-to-adult vegetative phase transition by modulating genes of the *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE* (*SPL*) family (Wu *et al*., 2009). Therefore, we

evaluated the expression of these miRNAs and their target genes at 21, 24, and 27 DAS.

To investigate whether *TALIR* exerts a role in *MIR396* expression, we performed a reporter assay in which transgenic plants carrying the MIR396b promoter fused to the reporter gene *GUS* were transformed with the amiR-*TALIR* plasmid. We then selected three independent transgenic lines carrying pMIR396b:GUS in the talir mutant background (Fig. 2e). Consistent with our previous observations at 21 DAS, in 2-week-old seedlings, the size of the fifth leaf is smaller than that of the WT, and the expression pattern of miR396b in the transgenic leaves is lower than in the parental line.

We then analyzed the accumulation levels of miR396b in the WT, *talir-1*, and *talir-amiR* lines. The expression level of miR396 in the fifth leaf of WT (Fig. 2f), showed changes in their expression consistent with previous results reported by Rodriguez *et al*. (2010). We observed that in the WT line the relative *miR396b* transcript accumulation reaches its maximum point at 21 DAS and decreases at 27 DAS. In contrast, the transcript levels of miR396b in the *talir-1* mutant line were lower at 21 and 24 DAS, however at 27 DAS, there was a higher accumulation of miR396b than in WT plants. On the other hand, the *talir-amiR* line showed low accumulation levels of miR396b throughout the span of the experiment when compared to WT (Fig. 2f). To evaluate the effects of altered miR396b levels on the regulation of its target gene *GRF2*, we performed a gene expression analysis. (Fig. 2h). As expected, GRF2 transcript levels were reduced in samples with increased miR396b abundance with this trend being particularly evident in the *talir-1* line. Interestingly, at 27 DAS, the expression of GRF2 in the *talir-1* line was higher than in the *talir-amiR* and WT plants, suggesting that *TALIR* regulation during leaf development might be time-dependent.

WT plants showed an accumulation of miR156 in the fifth leaf after 21 DAS (Fig. 2g), reaching its highest level at 24 DAS and then declining to a minimum at 27 DAS. On the other hand, *talir-1* and *talir-amiR* mutant plants reached lower levels of miR156 accumulation than the WT at 21 and 24 DAS. However, at 27 DAS, the *talir-1* mutant showed higher levels of miR156 in contrast with t*alir- amiR* line, which had a relative

accumulation similar to WT. To corroborate the relative abundance of the miR156 target genes, we quantified the relative transcript level of the *SPL9* gene. While the WT displayed the characteristic behavior of a gene subjected to post-transcriptional regulation by a microRNA, with a maximum accumulation at 27 DAS (Fig. 2i), the *talir-1* and *talir-amiR* lines showed a differential expression pattern, with both lines exhibiting lower *SPL9* accumulation at 27 DAS. These results suggest that the lack of adult phase traits in the mutant leaves is partly due to the absence of *SPL9*, which along with other *SPLs* such as *SPL13* and *SPL15*, contributes to vegetative traits related to the adult phase of leaves (Wu *et al*., 2009).

The expression of TALIR is regulated under ABA and NaCl treatments

Bioinformatic analysis of the *TALIR* promoter revealed the presence of cis-acting binding elements associated with MYB/MYC transcription factors (Supplementary Fig. 6), which are involved in the ABA-dependent pathway for the upregulation of abiotic stress-related genes (Abe *et al*., 2003). These *cis*-acting elements are also present in the promoters of the putative *TALIR* orthologs in other species within the Brassicaceae family (Supplementary Fig. 6), and thus we performed stress experiments to evaluate this possibility. We grew WT plants for 2 weeks under control conditions that were then treated with 5 µM ABA or 175 mM NaCl for 24 hours. RT- qPCR data revealed a clear increase in accumulation of *TALIR* upon addition of ABA and in response to NaCl compared to untreated plants (Fig 3a). We used the *GEA6* (At2g40170) gene which encodes a protein belonging to the LATE EMBRYOGENESIS ABUNDANT (*LEA*) gene family and activated by ABI5, as a marker for stress (Fig. 3b).

Furthermore, we evaluated the germination percentage of WT and *talir* mutant lines under control conditions and with increasing concentrations of ABA for 5 days (Fig. 3c). We found that there was no difference in germination between WT and mutant plants under control conditions. In contrast, while treatment with 5 µM ABA caused the expected delay in germination for the WT line, surprisingly the mutant lines displayed a reduced sensitivity to the hormone. Additionally, we investigated the transcriptional activity of the *TALIR*-associated *loci* under ABA treatment conditions. WT and *talir*

mutant lines were treated with 5 μM ABA for 24 hours. We confirmed the increased accumulation of *TALIR* in WT plants treated with ABA, and a reduced *TALIR* accumulation in the treated mutant lines (Fig. 3d). Under these conditions, *TBP2* transcript levels were significantly dysregulated in mutant plants, whereas *TBP1* expression increased in both genotype lines treated with ABA (Fig. 3e, f). In contrast, we did not observe statistically significant changes in the expression levels of *E3 RINGU* under these conditions (Fig. 3g). Finally, we corroborated the increased accumulation of mRNA for the ABA inducible *GEA6* gene in the treated plants (Fig. 3h). These results suggest that the response of *TBP1* to ABA is independent of the presence of *TALIR*, while *TALIR* negatively regulates TBP2 in this condition.

Subcellular localization of TALIR

Various lncRNAs are localized in the nucleus, where they participate in transcriptional regulation through mechanisms involving interactions with DNA, other RNAs or proteins (Wu *et al*., 2020). To determine the localization of *TALIR*, we employed a method developed by Kinoshita *et al* (2018). In this method, two constructs were used: a) an MS2 stem-loop motifs RNA was fused to *TALIR* and expressed under the control of the 35S viral promoter. This construct also included an independent transcriptional unit encoding an mCherry reporter gene fused to a nuclear localization signal (NLS) to mark transformed nuclei. b) a protein component formed by bacteriophage MS2 coat protein (MSCP) fused to GFP and an NLS. These constructs were expressed in *Nicotiana benthamiana* leaves via *R. radiobacter* co-infiltration and the leaves were observed using confocal microscopy. The MS2-*TALIR*-mCherry-NLS signal appeared as punctate dots, indicating transformed nuclei. MSCP-GFP-NLS localized to both the cytoplasm and the nucleus. When we co-expressed both constructs, we observed a discrete yellow signal that colocalized at the same nuclear position (Fig. 4), similar to *NPC60* RNA, which was previously shown to have a nuclear localization (Kinoshita *et al*., 2018), serving here as a positive control (Supplementary Fig. 7). To confirm this localization, we performed cellular fractionation using 3-week-old plants treated with 3 μM ABA for 24 h. We isolated nuclei, and total RNA was

extracted to evaluate transcript accumulation by end-point RT- PCR. We confirmed that *TALIR* was present in the nuclear fraction and verified the nuclear enrichment amplifying transcripts corresponding to pri-MIR396b and the lncRNA MAS (Supplementary Fig. 8). These results indicate that *TALIR* is principally nuclear, which is consistent with a potential role in the modulation of *TBP2* expression.

The Origin of TALIR is in Brassicales

Conservation analyses can provide indirect evidence of functional relevance for lncRNAs, as lncRNA loci that are maintained across evolutionary time are more likely to be biologically meaningful. Here, a phylogenetic framework was used to guide comparisons among closely related Brassicaceae species, increasing sensitivity for detecting nucleotide-level conservation among *TALIR*-associated intergenic regions that may be lost in broader comparisons. Therefore, we investigated whether *TALIR* is conserved within the Brassicaceae family (order Brassicales) by examining (i) conservation of genomic context (microsynteny) and (ii) nucleotide-level conservation across species.

Because TBP proteins are highly conserved across eukaryotes, we used *TBP2* as an anchor to define the genomic region in which *TALIR* is located. Using *A. thaliana* TBP2 as query, we retrieved 27 TBP2-related protein sequences distributed across 11 species: *Arabidopsis thaliana*, A. lyrata, *Brassica oleracea*, *B. rapa*, *B. napus*, *Capsella rubella*, *Camelina sativa*, *Eutrema salsugineum*, *Raphanus sativus*, *Tarenaya hassleriana*, and *Mangifera indica* (used as an outgroup) (Table S1). Phylogenetic reconstruction grouped Brassicaceae TBP2 homologs into three tribes —Brassiceae, Eutremeae, and Camelineae—consistent with established phylogenetic relationships (Fig. 5a). Within Brassiceae (*R. sativus*, *B. oleracea*, *B. rapa*, and *B. napus*), we identified 11 sequences distributed across multiple *loci*, forming two main clades, one of which included proteins annotated as *TBP*-like. In Eutremeae (*E. salsugineum*), we identified a single *TBP2* homolog. In Camelineae (*A. lyrata*, *A. thaliana*, *C. rubella*, and *C. sativa*), *TBP2* was present as a single copy in the first three species, whereas *C. sativa* contained seven TBP2-related copies, consistent with its polyploid history.

Outside Brassicaceae, four TBP2-related sequences were identified in *T. hassleriana*, and a single *TBP2* homolog was found in the outgroup species *M. indica* (Table S1). Together, these data support the broad conservation of TBP2 within and beyond Brassicaceae.

To test whether the *TALIR locus* is conserved in genomic position, we performed microsynteny analyses of the chromosomal regions flanking *TBP2* and the adjacent *E3 RINGU* gene. We examined 40-kb regions upstream and downstream of *TBP2* across the selected Brassicaceae species. Overall, gene order and content were strongly conserved across the three Brassicaceae tribes, with the notable exception of several *C. sativa loci*, in which *TBP2* was retained, but the surrounding region showed reduced conservation (Fig. 5b). As expected for intergenic regions, the DNA segment between *TBP2* and *E3 RINGU* exhibited greater divergence than the flanking protein-coding genes. In particular, the intergenic interval was expanded and more variable in *C. sativa* and in some *loci* of *B. napus* and *B. rapa* (Fig. 5b), suggesting lineage- and *locus*-specific structural diversification.

We next searched for evidence of transcriptional activity within the intergenic region between *TBP2* and *E3 RINGU* as a proxy for the presence of *TALIR*-like loci. Using publicly available NCBI resources, we identified transcriptional evidence supporting the existence of 19 putative *TALIR*- related transcripts in addition to *A. thaliana TALIR* itself (Supplementary Fig. 6). Notably, sequence comparisons revealed unexpected nucleotide-level conservation across multiple Brassicaceae species within this same intergenic region (Fig. 6a, b). This conservation was strongest among phylogenetically close species, consistent with selective constraints acting on *TALIR*-related loci. Moreover, local structural comparisons highlighted the presence of conserved stem–loop–like motifs among closely related Brassicaceae species (A. thaliana, *A. lyrata* and *C. sativa*). Although global RNA structures were divergent, these locally conserved elements suggest selection constraints acting on specific *TALIR*-associated RNA features (Supplementary Fig. 9). While *TALIR* is annotated as a non-coding RNA in *A. thaliana* and in one *C. sativa locus* only, the combined presence of transcriptional

evidence and conserved sequence strongly suggests that the unannotated intergenic *loci* represent *TALIR* homologs present in other species (Fig. 6; Supplementary Fig. 6).

Copy-number patterns varied across lineages. In most Camelineae species, we detected a single *TALIR*-like candidate adjacent to *TBP2*; in contrast, *C. sativa* exhibited multiple candidates, including one *locus* containing four *TALIR*-like segments and two *loci* with two candidates each. Within Brassicaceae, most species showed one or two *TALIR*-like candidates that grouped into a single clade, except for one *B. rapa* candidate that clustered with *TBP2* sequences annotated as *TBP*-like proteins (Fig. 5a, b). Together, these observations are consistent with the existence of *TALIR*-related *loci* that are positionally linked to *TBP2* and conserved at both synteny and sequence levels across Brassicaceae tribes.

In contrast, although *T. hassleriana* contained multiple TBP2-related copies, we did not detect sequences with significant similarity to *A. thaliana TALIR*, nor did we find comparable transcriptional evidence in the corresponding intergenic regions. Similarly, *M. indica* contained *TBP2* and *E3RINGU* homologs but displayed a larger and more divergent intergenic region, limited microsyntenic conservation relative to Brassicaceae, and no detectable *TALIR*-like conservation or transcriptional support (Supplementary Fig. 10). These observations suggest that *TALIR* likely emerged after the divergence of Brassicaceae from more distantly related Brassicales lineages.

Because *A. thaliana* also contains a *TBP2* paralog, *TBP1*, we asked whether a similar intergenic lncRNA might be present near the *TBP1 locus*. Microsynteny analysis anchored on *TBP1* did not reveal a *TALIR*-like candidate in the intergenic region between *TBP1* and the adjacent *E3 RINGU* gene (Supplementary Fig. 11), supporting the idea that *TALIR* is specifically associated with the *TBP2* genomic context.

Finally, comparative analyses of *TBP*-associated regions in *A. thaliana*, *M. indica*, and *Carica papaya* revealed strong conservation within species but marked differences among species outside Brassicaceae, including substantial changes in intergenic length and, notably, an apparent change in *TBP2* orientation in *M. indica* and *C. papaya*

relative to *A. thaliana* (Supplementary Fig. 12).Collectively, these data support a model in which *TALIR* originated within Brassicaceae (order Brassicales), potentially facilitated by genome reorganization events associated with ancient whole-genome duplications and subsequent fractionation. However, additional comparative genomic data from intermediate lineages will be required to resolve the precise timing and mechanism underlying *TALIR* emergence.

DISCUSSION

In this study, we characterized an intergenic lncRNA *TALIR*, which is implicated in leaf and root development, ABA response and it is conserved across plant species in the Brassicales. In the context of development, previous studies have identified several lncRNAs, *TL*, *bra-miR156aHG*, *NAT1*$_{UGT73C6}$ and *NAT2*$_{UGT73C6}$ identified in *Oryza sativa*, *Brassica campestris* and *A. thaliana*, respectively, that regulate growth and leaf morphology, while *ARES* and *APOLO* both participate in auxin signaling implicated in lateral root formation (Lucero *et al*., 2021; Liu *et al*., 2018; Zhou *et al*., 2023; Meena *et al*., 2023; Roulé *et al*., 2023). Using different strategies to generate loss-of-function mutants, we showed that talir knockdown mutants exhibited reduced leaf size and a diminished rosette area as well as a lesser LR number and LR density which suggests that *TALIR* modulates both processes.

We focused on changes in leaf size and morphology analyzing them at 21, 24 and 27 DAS. We observed the most evident phenotype at 21 DAS, when the leaves of *talir* mutants were smaller than those of WT plants. The changes observed in leaf size and shape in *talir* mutants can be explained, at least in part, by the transient regulatory effect *TALIR* exerts on its neighboring genes. TBP2 is a subunit of the general transcription factor TFIID, which, along with other general transcription factors, assembles the preinitiation complex which is required for basal transcription (Heard *et al*., 1993). Heterozygous *A. thaliana* plants that constitutively express *TBP2* exhibit an increased number of rosette's leaves and a reduced leaf size, along with other developmental defects. Plants overexpressing *TBP2* display an even more severe phenotype, with smaller rosette leaves and an increased number of apical shoots being

more exacerbated in homozygous plants compared to heterozygous plants (Li *et al*., 2001). Our observations showed a reduced accumulation of *TBP2* transcripts at 21 DAS and an increased accumulation at 24 DAS in the *TALIR* loss-of-function, which suggests that *TALIR* modulates *TBP2* accumulation both positively and negatively.

TBP2 interacts with TGH, an evolutionary conserved protein implicated in miRNA biogenesis by associating with the DCL1 complex, which participates in the processing of primary miRNA transcripts (Calderon-Villalobos *et al*., 2005). Having established changes in TBP2 abundance, and considering its role in miRNA biogenesis, we examined the accumulation of specific miRNAs involved in leaf development. miR396 and miR156 are master modulators of *A. thaliana* leaf growth.

In WT plants our observations about miR396 were consistent with a previous report (Rodriguez *et al*., 2010) about both histochemical assay and RT-qPCR analysis. However, in *talir* mutant plants the transient dysregulation of *miR396b* abundance leads to a transient increase in the *GRF2* accumulation at 24 DAS. Nevertheless, unlike plants that overexpress *GRF2* and exhibit larger leaves, our plants did not show an increase in leaf size, indicating a contrast with the reported evidence (Rodriguez *et al*., 2010). This suggests that leaf size may be negatively modulated independently of the *GRF* gene family, suggesting that *TALIR* acts downstream of *GRF2* or participates in an alternative pathway that regulates genes involved in cell growth.

talir mutant plants showed alterations in *MIR156* expression as well as *SPL9*, one of its target genes. *SPL9* mutants develop a larger number of leaves and rounder leaves under short-day conditions (Wu *et al*., 2009). However, our results showed an opposite phenotype to that reported in the literature, partly because our experiments were performed under long-day conditions. Besides, single mutants of *SPL9* and *SLP15* genes grown in long days display, on average, one additional leaf than WT (Schwarz *et al*., 2008).

Taken together, these findings show *TALIR*-dependent regulation of TBP2 may contribute to the dysregulation of *MIR396b* and *MIR156* and their targets in talir mutant lines, resulting in an altered leaf development phenotype.

This general view shows that *TALIR* loss-of-function mutants have a differential temporal regulation of TBP2 resulting in a pleiotropic phenotype. Additionally, plants lacking *TGH* exhibit decreasing *miRNA* accumulation and severe defects in development (Ren *et al*., 2012). According to our results, the decrease in *miRNAs* accumulation agrees with the reduced expression of *TBP2* at 21 DAS in talir mutants. However, the transient overaccumulation of *TBP2* at 24 DAS does not correlate with miRNAs overexpression. Therefore, we suggest that *TALIR*-dependent regulation of *TBP2* indirectly influences MIRNAs expression. The phenotypes observed in *TALIR* loss-of-function lines may be synergistic with the regulation that TBP2 exerts over other genes that participate in these developmental programs such as *STM* and *KNAT1* genes (Li *et al*., 2001).

Furthermore, we observed that the abundance levels of the miRNAs and their target genes varied among the different *talir* mutant lines at 24 DAS. This variability is likely due to the methods used to generate *TALIR* loss-of-function. In the T-DNA line, *TALIR* was repressed at the transcriptional level, whereas in the amiR line, *TALIR* was inactivated at the posttranscriptional level. It is possible that *TALIR* forms a secondary structure capable of mediating gene regulation, which could be inaccessible for amiRNA-directed cleavage. It has been reported for the lncRNA *HID1*, that its secondary structure is essential for its function during photomorphogenesis (Wang *et al*., 2014).

Only a few lncRNAs are known to be involved in ABA response such as *DRIR* and *ARTA*, which are induced by ABA treatment and enhance plant drought tolerance by prompting the expression of stress-response genes and participating in the inhibition of *ABI5* (Qin *et al*., 2017; Yang *et al*., 2023). Here, we found that *TALIR* accumulation was induced by ABA and NaCl treatments. *TALIR* increased accumulation is necessary to maintain transcriptional control of *TBP2*, *TALIR* loss-of-function resulted in

dysregulated *TBP2* transcription, suggesting it can act as a negative regulator of *TBP2* under ABA treatment. We also observed an ABA-insensitive germination phenotype, potentially related to *TBP2* regulation of ABA-dependent genes. In *Vigna radiata* L., a TBP closely related to TBP2 has been identified as a positive regulator of salt stress (Wu *et al*., 2024), whereas in *Oryza sativa*, *TBP2* accumulates under drought conditions in agreement with our results. Moreover, *knockdown* mutants of *TBP2* displayed reduced tolerance to drought stress (Zhang *et al*., 2020). These examples indicate that *TBP2* expression is regulated under stress conditions, potentially involving both ABA-dependent and ABA-independent pathways, suggesting a role in plant responses that could involve *TALIR*. Additionally, under control conditions, we did not observe differences in germination between WT and *talir* mutants, indicating that *TALIR* functions as a regulator of ABA response during germination but is not required for normal development at this stage.

A common mechanism for lncRNAs which suppress the transcription of its neighboring genes is recruitment of chromatin remodelers, such is the case of *COOLAIR* which recruits *PRC2* to silence *FLC*, or *AG-inRNA4*, which binds *CLF* to repress *AGAMOUS* transcription (Lucero *et al*., 2021). Another mechanism involves the *lncRNA-TBP*, which interacts with TBP protein to suppress genes involved in myoblast proliferation in Xinghua chicken (Ma *et al*., 2023). Analyzing the subcellular localization of lncRNAs is a key step toward understanding their function and mechanism of action. We elucidated the subcellular localization of *TALIR* using a two-component system: a) an MS2 stem-loop motifs RNA fused to *TALIR* and tagged with mCherry-NLS, and b) a protein MSCP fused to GFP-NLS. Both constructs were co-infiltrated in *Nicotiana benthamiana* leaves using *R. radiobacter*. Confocal microscopy revealed nuclear localization of both signal components, supporting the hypothesis that *TALIR* functions in the nucleus by modulating *TBP2* transcription. Nevertheless, the precise mechanism by which *TALIR* functions remains unclear. Notably, most nuclear-localized lncRNAs are associated with chromatin interactions, transcriptional regulation, and genome organization (Wu *et al*., 2020). A key open question is

identifying the molecular partners with which *TALIR* interacts to exert its function. Approaches such as RNA immunoprecipitation (RIP) or RNA pull-down assay may provide insights into the mode of action of *TALIR*.

In plants a few lncRNAs are conserved at the sequence or synteny level. For instance, *lncCOBRA1*, is conserved at the sequence level in five species of the Brassiceae tribe and contributes to regulate germination and development (Kramer *et al*., 2022). Another example is a lncRNA downstream of *EIN2*, which exhibits conserved synteny in many angiosperm plant species (Nieto-Hernández *et al*., 2023).

Interestingly, we found that *TALIR* is conserved at the sequence and synteny level across Brassiceae, Eutremeae, and Camelineae tribes belonging to Brassicales family, which together with its strict positional association with *TBP2 locus*, suggests that *TALIR* emerged as part of a regulatory module rather than as an isolated non-coding transcript. Such positional conservation is consistent with a role in *cis*-mediated regulation of transcription. *TALIR* is located almost exclusively upstream of *TBP2* gene but not *TBP1*, suggesting these two genes have different regulatory mechanisms for their expression (Heard *et al*., 1993) The conservation of *TALIR* at the level of stem–loop motifs, despite sequence divergence in specific lineages such as *Camelina sativa*, supports the idea that structural features rather than primary sequence are under selective constraint. When considered alongside the developmental defects, altered miRNA accumulation, and ABA-related phenotypes observed in *TALIR* loss-of-function lines, these evolutionary data support a model in which *TALIR* contributes to the integration of developmental and responsive transcriptional programs in plants.

A lncRNA can originate through pseudogenization processes. A well characterized example is *Xist*, which is conserved in eutherians. *Xist* evolved from the loss-of-function of protein-coding *Lex3*, along with transposable element insertion and later amplification, giving *Xist* a new function in the transcriptional silencing of X chromosome genes (Elisaphenko *et al*., 2008). Since *TALIR* is only present in Brassicales family, we hypothesized that it emerged within this group. Comparative analysis of the genomic regions adjacent to *TBP1* and *TBP2* in M. indica, *C. papaya*,

and *A. thaliana* revealed intra-species synteny but no synteny among these plant species and no evidence of any *TALIR* ortholog. Differences were noted in the intergenic region in *A. thaliana*, where the intergenic region of *TBP2* is smaller than that of *TBP1*. These changes in size of the genomic regions adjacent to *TBP1* and *TBP2* may be associated with polyploidy events. Polyploidy is a common trait in plants, characterized by the presence of three or more chromosomal sets, and can be derived from a whole-genome duplication (WGD) event. It has played a crucial role in plant speciation and diversification altering the organism complexity (del Pozo and Ramirez-Parra 2015). The WGD triggers genomic reorganization, the α-WGD event which is exclusive to the Brassicaceae represents a pivotal genomic event that may have provided structural context for the emergence for new genes (Barker et. al. 2009). *A. thaliana* contains only a single *TBP2* copy, in contrast to copy number variation among other species, which appears to correlate with WGD events. Although this paleopolyploid species experienced and ancient duplication (At-α), it has since undergone diploidization, leading to gene loss and fractionation (del Pozo-Parra and Ramírez Parra, 2015). Our data suggests that *TALIR* emerged within the Brassicales family; however, we cannot yet determine the precise timing or mechanism that triggered its origin. Several WGD events following the β-WGD generated numerous orthologous genes located on homologous chromosomes or chromosomal regions. The high sequence similarity between homologous regions likely facilitated illegitimate recombination (Wang *et al*., 2009), which may have contributed to the reorientation of the *E3 RINGU* gene and its surrounding regions. This process likely leads to pseudogenization and gene conversion, ultimately contributing to the emergence of *TALIR*.

Our findings reveal a novel lncRNA involved in leaf development and ABA response. Moreover, its evolutionary conservation is noteworthy, suggesting that the functions of *TALIR* may be conserved in the Brassicaceae species where it is found upstream of *TBP2*.

MATERIALS AND METHODS

Plant material and growth conditions

All *Arabidopsis thaliana* plants used in this study, including wild-type (WT) and transgenic lines were in the Columbia-0 ecotype background. The T-DNA insertion mutant (*talir-1*) (Salk_089677) was obtained from the Salk T-DNA collection (Nottingham Arabidopsis Stock Centre). RNA interference mutant (*talir-Ri*) and artificial microRNA mutant (*talir-amiR*) lines were generated in this work. To generate the RNAi construct, a 227 bp fragment was cloned into a binary vector pKGWIWGII(2) (Karimi *et al*., 2002). Three independent homozygous lines were obtained but only one was used for subsequent analysis. To obtain amiR lines, the Web MicroRNA Designer tool was used to obtain specific oligonucleotides (Ossowski *et al*., 2008), followed by a series of PCR amplifications based on the protocol developed by Schwab *et al*. (2006). The fragment obtained was cloned into a binary vector pBA-DC (Zhang *et al*., 2005). Two independent homozygous lines were obtained but only one was used for subsequent analysis. The pmiR396: *GUS* line was generously donated by Dr. Javier Palatnik from the Instituto de Biología Molecular y Celular de Rosario, Santa Fe, Argentina.

Seeds were sterilized in 96% ethanol for 2 min, followed by incubation in 30% commercial bleach (sodium hypochlorite) for 7 min, and washed seven times with sterile distilled water. Seedlings were grown in Petri dishes in 0.2 X and 0.5 X Murashige and Skoog (MS) medium, pH 5.7, 1% sucrose (J.T. Baker, USA) and 0.8% agar (w/v, BD Bioxon, Mexico). For root growth characterization, plants were grown in a vertical orientation. For stress tolerance assays, seeds from WT, T-DNA, RNAi, and amiR genotypes were grown on 0.5 X MS medium for two weeks. Next, the seedlings were transferred to a fresh medium supplemented with 5 µM ABA or left without supplementation and incubated for 24 h. Seedlings were collected and frozen in liquid nitrogen for later processing. For seed germination assays, seeds of all genotypes were sown in 0.5 X MS medium supplemented with 0 µM, 1 µM, 3 µM, or 5 µM ABA. Germination was evaluated starting 24 h after exposure to light until the fifth day. Seeds in which the radicle emerged were considered as germinated. Two replicates were

performed for each experiment. All plants were grown at 21°C ± 2°C, under a 16 h light / 8 h dark long-day photoperiod.

Leaf analysis

Plants used for this analysis were grown in a long-day photoperiod for two weeks. Following this initial period, they were transferred to soil and collected at 21, 24, or 27 days after sowing. The leaves were examined, classified, and scanned using an HP Scanjet G4050 to measure their surface area with the Fiji software. All experiments were conducted with three independent biological replicates. For each *talir* mutant, 24 leaves were analyzed in each replicate experiment.

Root analysis

The quantification of root growth was performed on vertically grown seedlings cultured in 0.2X MS medium. The position of the root tip was marked every 24 h throughout 7 days post-germination (DPG). Following this period, the plates were photographed using a Canon Powershot SX50 HS camera, and these images were used to measure the increment in root length using Fiji software. Lateral root density was determined using cleared roots. Briefly, roots were fixed with 4% paraformaldehyde in prepared 0.025 M phosphate buffer and stored at 4°C for 24 h. Roots were then treated with an acidified methanol solution (0.24 N HCl in 20% methanol) and incubated at 62°C for 4 min. This solution was then replaced with a basic solution (7% NaOH in 60% ethanol) and incubated at room temperature for 20 min. Roots were rehydrated through a sequential washing process with ethanol solutions at concentrations of 40%, 20%, and 10%, at 24 h intervals. Finally, an equal volume of 50% glycerol was added for a final concentration of 5% ethanol and 25% glycerol solution, in which the roots were maintained for 24 h. The roots were placed on slides containing a 50% glycerol solution and sealed with coverslips to observe them under an optical microscope (Leica DM500, Germany) (Dubrovsky *et al*. 2006). The number of lateral roots and the length of the branching zone, spanning from the last emerged

lateral root to the hypocotyl, were determined by observing cleared roots. All parameters were evaluated for each individual root.

RNA Extraction

Total RNA was isolated from 2-week-old plants or fifth leaves at 21- 24- and 27-days post-sowing using TRIzol LS Reagent (Thermo Fisher), according to the manufacturer's protocol. The RNA was resuspended in sterile Mili-Q water and incubated at 60°C for 10 min. The concentration of total RNA was determined using a NanoDrop 2000c (Thermo Scientific).

Expression analysis

One µg total RNA extracted from plants was treated with DNAse I (Thermo Scientific). The cDNA for each line was synthesized using the RevertAid First Strand cDNA Synthesis Kit (Thermo Fischer Scientific) following the manufacturer's instructions. cDNA was diluted 1:2 and 1 µL was used for each reaction except for negative control. The relative quantification of the transcripts was conducted using a RT-qPCR detection System (Bio-Rad, CA, USA). The reactions were carried out using the Power SYBR Green PCR Master Mix for CFX (AppliedBiosystems, Thermo Fischer Scientific) in a volume of 10 µL following the manufacturer's protocol. The sequences of the oligonucleotides used in this analysis are listed in Table S2. Transcript abundance levels were measured using the comparative ΔΔCq method and normalized against the endogenous reference gene ACT2. Three technical replicates were performed for each sample. Two biological replicates were made for each experiment.

Analysis of microRNAs and their targets

To analyze the level of abundance of microRNAs in the fifth leaf, first-strand cDNA synthesis and the quantitative RT-qPCR for each line were performed using the Clontech Kit (Takara Bio), following the manufacturer's instructions. The method is based on polyadenylation and reverse transcription of RNA in a single reaction. Quantitative RT-qPCR detection was carried out using a RT-qPCR detection System (Bio-Rad, CA, USA). The sequences of the oligonucleotides used in this assay are

listed in Table S2. Transcript abundance levels were measured using the comparative ΔΔCq method, normalized against the endogenous reference genes U6 snRNA or ACT2. Three technical replicates were performed for each sample. Two biological replicates were conducted for each experiment.

GUS Histochemical Assay

To evaluate the expression pattern of p*MIR396* in the *talir* mutant background, plants carrying the pMIR396b:GUS construct were transformed with amiR *TALIR* plasmid by the floral dip method described by Zhang *et al*. (2006). Three independent lines of pMIR396b:GUS – *amiR TALIR* were obtained. Twenty seeds were sowed in 0.5X MS and cultivated in vitro for two weeks under a long-day photoperiod. The *GUS* staining protocol was carried out as previously described by Jefferson *et al*. (1987) with minor modifications. Tissue samples were stained by incubating the roots in a *GUS* staining buffer at 37°C overnight, followed by four cycles of decolouring in a 9:1 ethanol-glycerol solution for 15 min each. Samples were maintained in the same solution for 3 days and photographed and observed using an optic microscope (Leica DM500, Germany).

Subcellular localization of lncRNA

A transient expression assay of *TALIR* was performed in *Nicotiana benthamiana* leaves following the protocol described by Kinoshita *et al*. (2018). Briefly, two constructs were generated: one containing the lncRNA fused to a vector with 35S-6xMS2-mCherry-NLS, and the other with 35S-MSCP-GFP-NLS (pSCJ216 and pSCJ380, Kinoshita *et al*., 2018). Both plasmids were transformed into *Rhizobium radiobacter* C58 (*R. radiobacter*) cells independently and cultured at 28°C for 2 days. The transformed *R. radiobacter* strains carrying the constructs were mixed, and co-infiltrated into the epidermal cells of the leaves of 4-week-old plants. Cell visualization was performed two days post infiltration using confocal microscopy (CONFOCAL OLYMPUS FV1000 upright) at the Laboratorio Nacional de Microscopía Avanzada,

Instituto de Biotecnología, UNAM. The excitation– emission wavelengths for GFP and mCherry were at 488–559/30 nm, and 561–617/30 nm, respectively.

Phylogenetic analyses

To elucidate the phylogenetic relationship between *Arabidopsis thaliana TBP2* and related genes in other Brassicales species, the protein amino acid sequence was used in BLASTP as query (E-value <10-10), resulting in the selection and identification of twenty-seven homologues, which were downloaded from the National Center for Biotechnology Information (NCBI) database. Species names and accession numbers are listed in Table S1. Sixty sequences were aligned using MAFFT software (Katoh *et al*., 2002), followed by a manual alignment in PhyDE® software v. 0.9971. Phylogenetic analysis was conducted in CIPRES Science Gateway by a Bayesian Inference method (Yang and Rannala, 1997), employing the program MrBayes on XSEDE v. 3.2.7a, the most appropriate evolutionary models were selected, resulting in the HKY+G model for TBP2 and the GTR+G model for TBP1. Subsequently, phylogenetic trees were generated using MrBayes (Rambaut *et al*., 2018), with parameters set to ten million generations, four chains, a sampling frequency of 1000, and a runtime of 2.5 hours. The DNA evolution model was estimated with jModeltest2 v. 2.1.6 (Darriba *et al*., 2012). Phylogenetic tree was visualized using FigTree software v. 1.4.4 (accessible at http://tree.bio.ed.ac.uk/software/fgtree/).

Synteny analysis of TALIR

Synteny analyses were performed by considering a genomic region of 40 kb upstream and downstream of the TBP2 gene. This region was evaluated for each species within the Brassicaceae family using the CoGE, Gevo, and Multi-Genome Synteny Viewer tools (accessible at https://genomevolution.org/coge/SynMap.pl). Data visualization and Microsynteny analysis were accomplished by MCscan (Python version) (Tang *et al*. 2008) and ‘ggenes’ extension of the ggplot library within the RStudio environment (v1.1453).

Statistical analysis

Student's *t* test and ANOVA were utilized to compare the phenotypic and genetic features between WT and *talir* mutants. A *p*-value of less than 0.05 was considered as statistically significant. To analysis the relative abundance, the statistical analysis was determined by One way ANOVA followed by post hoc Dunnett´s test. All statistical analyses in this study were performed using SigmaPlot 12.0 (Palo Alto, CA), and the data were visualized with RStudio software (v1.1453).

ACKNOWLEDGEMENTS

We would like to thank the School of Science and the School of Chemistry Science of the Autonomous University of San Luis Potosí, for providing the facilities to perform all analyses. To SECIHTI, formerly CONAHCYT for the scholarship (718765 to Byanka Espinoza-Lopez). M.C. Coral Martínez Martínez and Lic. Xochitl Alvarado Affantranger for technical support for RNA localization experiments and the National Laboratory of Advanced Microscopy, UNAM. J.L.R., B.E.E.P, M.M.R., M.I.M.V., C.A.H., for useful contribution to discussion.

AUTHOR CONTRIBUTIONS

B.S.E.L. wrote the manuscript and performed phenotypic and genotypic analyses. J.N.H. and A.E.Q.E. conducted the bioinformatic analyses and contributed to the bioinformatic results section. S.I.V.C. and J.A.C. performed the subcellular localization assays and contributed to the characterization of the T-DNA, RNAi, and amiR lines. C.A.H. conceived and designed the study. C.A.H. and J.L.R. critically reviewed and edited the manuscript.

DISCLOSURES

Conflict of Interest: The authors declare that they have no competing interests.

Data availability

The datasets analyzed in this study were obtained from the National Center for Biotechnology Information (NCBI) databases (https://www.ncbi.nlm.nih.gov/guide/genes-expression/;

https://www.ncbi.nlm.nih.gov/protein/). Accession numbers are provided in Supplementary table 1.

REFERENCES

Abe, H., Urao, T., Ito, T., Seki, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2003). *Arabidopsis AtMYC2* (bHLH) and k(MYB) function as transcriptional activators in abscisic acid signaling. Plant Cell 15:63–78.

Ali, S., Khan, N., and Xie, L. (2020). Molecular and Hormonal Regulation of Leaf Morphogenesis in *Arabidopsis*. Int. J Mol. Sci. 21: 5132.

Barker, M. S., Vogel, H., and Schranz, M. E. (2009). Paleopolyploidy in the Brassicales: analyses of the *Cleome* transcriptome elucidate the history of genome duplications in *Arabidopsis* and other Brassicales. Genome Biol. Evol. 1:391–399.

Calderon-Villalobos, L. I., Kuhnle, C., Dohmann, E. M., Li, H., Bevan, M., and Schwechheimer, C. (2005). The evolutionarily conserved TOUGH protein is required for proper development of *Arabidopsis thaliana*. Plant Cell 17:2473–2485.

Chen, L., Zhu, Q. H., and Kaufmann, K. (2020). Long non-coding RNAs in plants: emerging modulators of gene activity in development and stress responses. *Planta* 252:92.

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9:772.

del Pozo, J. C., and Ramirez-Parra, E. (2015). Whole genome duplications in plants: an overview from *Arabidopsis*. J. Exp. Bot. 66:6991–7003.

Diederichs S. (2014). The four dimensions of noncoding RNA conservation. *Trends Genet.* 30:121–123.

Dubrovsky, J. G., Gambetta, G. A., Hernández-Barrera, A., Shishkova, S., and González, I. (2006). Lateral root initiation in *Arabidopsis*: developmental window, spatial patterning, density and predictability. *Ann. Bot.*, 97:903–915.

Elisaphenko, E. A., Kolesnikov, N. N., Shevchenko, A. I., Rogozin, I. B., Nesterova, T. B., Brockdorff, N. *et al*. (2008). A dual origin of the *Xist* gene from a protein-coding gene and a set of transposable elements. *PloS One* 3:e2521.

Guo, X., Liu, J., Hao, G., Zhang, L., Mao, K., Wang, X. *et al*. (2017). Plastome phylogeny and early diversification of *Brassicaceae*. *BMC Genomics*, 18:176.

Heard, D. J., Kiss, T., and Filipowicz, W. (1993). Both *Arabidopsis* TATA binding protein (TBP) isoforms are functionally identical in RNA polymerase II and III transcription in plant cells: evidence for gene-specific changes in DNA binding specificity of TBP. *EMBO J.* 12:3519–3528.

Jefferson, R. A., Kavanagh, T. A., and Bevan, M. W. (1987). GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J*., 6:3901–3907.

Jin, J., Lu, P., Xu, Y., Li, Z., Yu, S., Liu, J. *et al*. (2021). PLncDB V2.0: a comprehensive encyclopedia of plant long noncoding RNAs. *Nucleic Acids Res.* 49:D1489–D1495.

Karimi, M., Inzé, D., and Depicker, A. (2002). GATEWAY vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* 7:193–195.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30:3059–3066.

Kinoshita, N., Arenas-Huertero, C., and Chua, N. H. (2018). Visualizing nuclear-localized RNA using transient expression system in plants. *Genes Cells* 23:05–111.

Kramer, M. C., Kim, H. J., Palos, K. R., Garcia, B. A., Lyons, E., Beilstein, M. *et al*. (2022). A Conserved Long Intergenic Non-coding RNA Containing snoRNA Sequences, *lncCOBRA1*, Affects *Arabidopsis* Germination and Development. *Front. Plant Sci.* 13:906603.

Li, Y. F., Dubois, F., and Zhou, D. X. (2001). Ectopic expression of TATA box-binding protein induces shoot proliferation in *Arabidopsis*. *FEBS Lett.* 489:187–191.

Liu, X., Li, D., Zhang, D., Yin, D., Zhao, Y., Ji, C. *et al*. (2018). A novel antisense long noncoding RNA, *TWISTED LEAF*, maintains leaf blade flattening by regulating its associated sense R2R3-*MYB* gene in rice. *New Phytol.* 218:774–788.

Lucero, L., Ferrero, L., Fonouni-Farde, C., & Ariel, F. (2021). Functional classification of plant long noncoding RNAs: a transcript is known by the company it keeps. *New Phytol.* 229:1251–1260.

Ma, M., Cai, B., Zhou, Z., Kong, S., Zhang, J., Xu, H. *et al*. (2023). LncRNA-TBP mediates TATA-binding protein recruitment to regulate myogenesis and induce slow-twitch myofibers. *Cell Commun. Signal.* 21:7.

Meena, S. K., Heidecker, M., Engelmann, S., Jaber, A., de Vries, T., Triller, S. *et al*. (2023). Altered expression levels of long non-coding natural antisense transcripts overlapping the *UGT73C6* gene affect rosette size in *Arabidopsis thaliana*. *Plant J.* 113: 460–477.

Meng, X., Li, A., Yu, B., and Li, S. (2021). Interplay between miRNAs and lncRNAs: Mode of action and biological roles in plant development and stress adaptation. *Comput. Struct. Biotechnol. J.* 19:2567–2574.

Nieto-Hernández, J., Arenas-Huertero, C., and Ibarra-Laclette, E. (2023). LncRNA-encoded peptides: the case of the lncRNA gene located downstream of *EIN2*. *Funct. Integr. Genomics* 23:108.

Ossowski, S., Schwab, R., and Weigel, D. (2008). Gene silencing in plants using artificial microRNAs and other small RNAs. *Plant J.* 53:674–690.

Poethig, R. S. (2013). Vegetative phase change and shoot maturation in plants. *Curr. Top. Dev. Biol.* 105:125–152.

Qin, T., Zhao, H., Cui, P., Albesher, N., and Xiong, L. (2017). A Nucleus-Localized Long Non-Coding RNA Enhances Drought and Salt Stress Tolerance. *Plant Physiol.* 175:1321–1336.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* 67:901–904.

Ren, G., Xie, M., Dou, Y., Zhang, S., Zhang, C., and Yu, B. (2012). Regulation of miRNA abundance by RNA binding protein TOUGH in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA.* 109:12817–12821.

Rodriguez, R. E., Mecchia, M. A., Debernardi, J. M., Schommer, C., Weigel, D., and Palatnik, J. F. (2010). Control of cell proliferation in *Arabidopsis thaliana* by microRNA miR396. *Development.* 137:103–112.

Roulé, T., Legascue, M. F., Barrios, A., Gaggion, N., Crespi, M., Ariel, F., and Blein, T. (2023). The long intergenic noncoding RNA ARES modulates root architecture in *Arabidopsis*. *IUBMB life.* 75:880–892.

Schwab, R., Ossowski, S., Riester, M., Warthmann, N., and Weigel, D. (2006). Highly specific gene silencing by artificial microRNAs in *Arabidopsis*. *Plant cell* 18:1121–1133.

Schwarz, S., Grande, A. V., Bujdoso, N., Saedler, H., and Huijser, P. (2008). The microRNA regulated SBP-box genes *SPL9* and *SPL15* control shoot maturation in *Arabidopsis*. *Plant Mol. Biol.* 67:183–195.

Takahashi, H., Nozawa, A., Seki, M., Shinozaki, K., Endo, Y., and Sawasaki, T. (2009). A simple and high-sensitivity method for analysis of ubiquitination and polyubiquitination based on wheat cell-free protein synthesis. *BMC Plant Biol.* 9:39.

Tang, H., Bowers, J. E., Wang, X., Ming, R., Alam, M., and Paterson, A. H. (2008). Synteny and collinearity in plant genomes. *Science*, 320:486–488.

Wang, H., Chung, P. J., Liu, J., Jang, I. C., Kean, M. J., Xu, J. *et al*. (2014). Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in *Arabidopsis*. *Genome Res.* 24:444–453.

Wang, X., Tang, H., Bowers, J. E., and Paterson, A. H. (2009). Comparative inference of illegitimate recombination between rice and sorghum duplicated genes produced by polyploidization. *Genome Res.* 19:1026–1032.

Wang, Y., Fan, X., Lin, F., He, G., Terzaghi, W., Zhu, D. *et al*. (2014). *Arabidopsis* noncoding RNA mediates control of photomorphogenesis by red light. *Proc. Nat. Acad. Sci. USA* 111:10359–10364.

Wu, G., Park, M. Y., Conway, S. R., Wang, J. W., Weigel, D., and Poethig, R. S. (2009). The sequential action of miR156 and miR172 regulates developmental timing in *Arabidopsis*. *Cell.* 138:750–759.

Wu, L., Liu, S., Qi, H., Cai, H., and Xu, M. (2020). Research Progress on Plant Long Non-Coding RNA. *Plants (Basel, Switzerland)* 9:408.

Wu, R., Jia, Q., Guo, Y., Lin, Y., Liu, J., Chen, J. *et al*. (2024). Characterization of *TBP* and *TAFs* in Mungbean (*Vigna radiata* L.) and Their Potential Involvement in Abiotic Stress Response. *I. J. Mol. Sci.* 25:9558.

Xie, Z., Allen, E., Fahlgren, N., Calamar, A., Givan, S. A., and Carrington, J. C. (2005). Expression of *Arabidopsis MIRNA* genes. *Plant Physiol.* 138:2145–2154.

Yang, J., He, R., Qu, Z., Gu, J., Jiang, L., Zhan, X. *et al*. (2023). Long noncoding RNA *ARTA* controls ABA response through *MYB7* nuclear trafficking in *Arabidopsis*. *Dev. Cell.* 58:1206–1217.e4.

Yang, Z., and Rannala, B. (1997). Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo Method. *Molecular Biol. Evol.* 14:717–724.

Zhang, X., Garreton, V., and Chua, N. H. (2005). The AIP2 E3 ligase acts as a novel negative regulator of ABA signaling by promoting *ABI3* degradation. *Genes Dev.* 19:1532–1543.

Zhang, X., Henriques, R., Lin, S. S., Niu, Q. W., and Chua, N. H. (2006). *Agrobacterium*-mediated transformation of *Arabidopsis thaliana* using the floral dip method. *Nat. Protoc.* 1:641–646.

Zhang, Y., Zhao, L., Xiao, H., Chew, J., Xiang, J., Qian, K. *et al*. (2020). Knockdown of a Novel Gene *OsTBP2.2* Increases Sensitivity to Drought Stress in Rice. *Genes*, 11:629.

Zhou, D., Zhao, S., Zhou, H., Chen, J., and Huang, L. (2023). A lncRNA *bra-miR156HG* regulates flowering time and leaf morphology as a precursor of miR156 in *Brassica campestris* and *Arabidopsis thaliana*. *Plant Sci.* 337:111889.

FIGURE LEGENDS

Figure 1. *TALIR* is required for leaf development in *Arabidopsis thaliana*. (a) Schematic representation of the *TALIR locus*, indicating the sites where T-DNA insertion, RNAi (F and R), and amiR constructs were designed. (b) RT-qPCR analysis of *TALIR* transcript abundance in both WT and mutant plant lines in two-week-old seedlings. Relative abundance levels were normalized against the endogenous reference gene *ACT2*. Error bars represent the mean ± standard deviation (SD); $n = 3$. (c) Phenotypes of 21-day-old from WT and mutant plant lines after sowing. Scale bar = 2 cm. (d) Sequential stages of leaf development including cotyledons. The fifth leaf is enclosed in a black box. Scale bar = 1 cm. (e) Quantification of the total rosette and (f) individual leaf areas in both WT and mutant lines. Error bars represent the mean ± SD; $n = 4$. Asterisks (*) denote statistical significance at $p \leq 0.05$ (determined by the *t*-test). All experimental assays were performed with three independent biological replicates.

Figure 2. Dysregulation of transcript abundance of *miR396b* and *miR156* and the neighboring genes of *TALIR* during leaf development. (a) Schematic representation of

the *TALIR locus*, indicating the adjacent genes. RT-qPCR analysis of *TALIR*, *E3 RINGU*, *TBP2* and *TBP1* in the fifth leaf of both wild-type (WT) and mutant plant lines (*talir-1*, and *talir-amiR*) at (b) 21-days post-sowing, (c) 24-days post-sowing and (d) 27-days post-sowing. Relative abundance levels were normalized against the endogenous reference gene ACT2. Error bars represent the mean ± SD; *n* = 3. Asterisks (*) denote statistical significance at P ≤ 0.01 (determined by t-test). Representative figure from two experiments performed independently. (e) Histochemical GUS staining; expression pattern of the MIR396b promoter in WT plants and in the *talir* mutant background in two-week-old seedlings grown in vitro under control conditions. Representative figure from two experiments performed independently. (f) RT-qPCR analysis of miR396b and (h) its target gene *GRF2* in the fifth leaf of both WT and mutant plant lines (*talir-1* and *talir-amiR*) at 21-, 24-, and 27-days after sowing (DAS). (g) RT-qPCR analysis of *miR156* and (i) its target gene *SPL9* in the fifth leaf of both WT and mutant plant lines (*talir-1* and *talir-amiR*) at 21-, 24-, and 27-DAS. Error bars represent the mean ± SD; n = 3 leaves. Asterisks (*) denote statistical significance at P < 0.001 (determined by One way ANOVA followed by post hoc Dunnett´s test). Representative figure from two experiments performed independently.

Figure 3. Accumulation of TALIR under stress conditions and ABA response during germination. (a) RT-qPCR analyses were carried out to evaluate the transcripts levels of *TALIR* and (b) *GEA6* in two-week-old WT plants. Seedlings were subjected to either control conditions or treated with 5 µM ABA and 175 mM NaCl for 24 h. Relative abundance levels were normalized using the endogenous reference gene ACT2. Error bars represent the mean ± SD; *n* = 3. Asterisks (*) denote statistical significance at P < 0.01 (determined by t-test). (c) Seed germination assays were conducted under control conditions or treated with 1 µM, 3 µM, and 5 µM ABA for 5 days. Seed germination percentage of indicated genotypes shown at 5 days. Error bars represent the mean ± SD, n=50. Asterisks (*) denote statistical significance at P ≤ 0.01 (determined by t - test). All experimental assays were performed with two independent biological replicates. (d) – (h) *TALIR* modulates the accumulation of *TBP2* during ABA treatment.

Two-week-old seedlings of WT and *talir* mutant lines were treated with 5 μM ABA for 24 h. RT-qPCR analyses were carried out to evaluate the transcripts levels of *TBP2*, *TBP1*, *E3 RINGU*, *TALIR*, and *GEA6*. Error bars represent the mean ± SD; *n* = 3. Relative abundance levels were normalized using the endogenous reference gene ACT2. Asterisks (*) denote statistical significance at $P \leq 0.001$ (determined by *t* - test). Representative figure from two experiments performed independently.

Figure 4. *TALIR* shows a nuclear localization in *Nicotiana benthamiana*. Subcellular localization of *TALIR* was analyzed using an MS2-based transient expression system. Leaves were co-infiltrated with constructs expressing *TALIR* fused to six MS2 stem-loop motif sequences under the control of the CaMV 35S promoter and an independent mCherry reporter carrying a nuclear localization signal (MS2-TALIR/mCherry-NLS), together with the MS2 coat protein fused to GFP and an NLS (MSCP-GFP-NLS).

Figure 5. Resolved Brassicales plant phylogenetic tree for *TBP2* and microsynteny conservation. (a) Phylogenetic tree inferred by Bayesian method for eleven species. The whole-genome duplication (WGD) events are represented by 3-pointed stars (pink, green, and yellow). Hexaploidy and alopolyploidy are represented by 2- pointed stars (blue and white). The green, blue, and orange boxes indicate Brassicaceae, Eutremeae and Camelineae tribes, respectively. Black circles represent the orthologs of *TALIR* while blue triangles represent annotated *TALIR*. (b) Conservation of *TALIR* microsynteny at the *loci*, where blue lines represent *TBP2*, green lines represent *E3 RINGU*, and red lines represent localization of *TALIR*. The orthologs of *TALIR* are numbered 1 to 20 in the phylogenetic tree (a) and microsynteny diagram (b).

Figure 6. Zoomed-in view of the phylogenetic tree of *TBP2* and nucleotide conservation of *TALIR*. (a) Phylogenetic tree of *TBP2*. Hexaploidy and alopoliploidy are represented by 2- pointed stars (blue and white). The green, blue, and orange boxes indicate Brassicaceae, Eutremeae and Camelineae tribes, respectively. Black circles represent the orthologs *TALIR* while blue triangles represent annotated TALIR. (b) Alignment of the nucleotide sequence of *TALIR* in species of the order Brassicales.

Nucleotides are represented by colors (adenine, yellow; thymine, blue; guanine, green; cytokine, red; gaps, black).

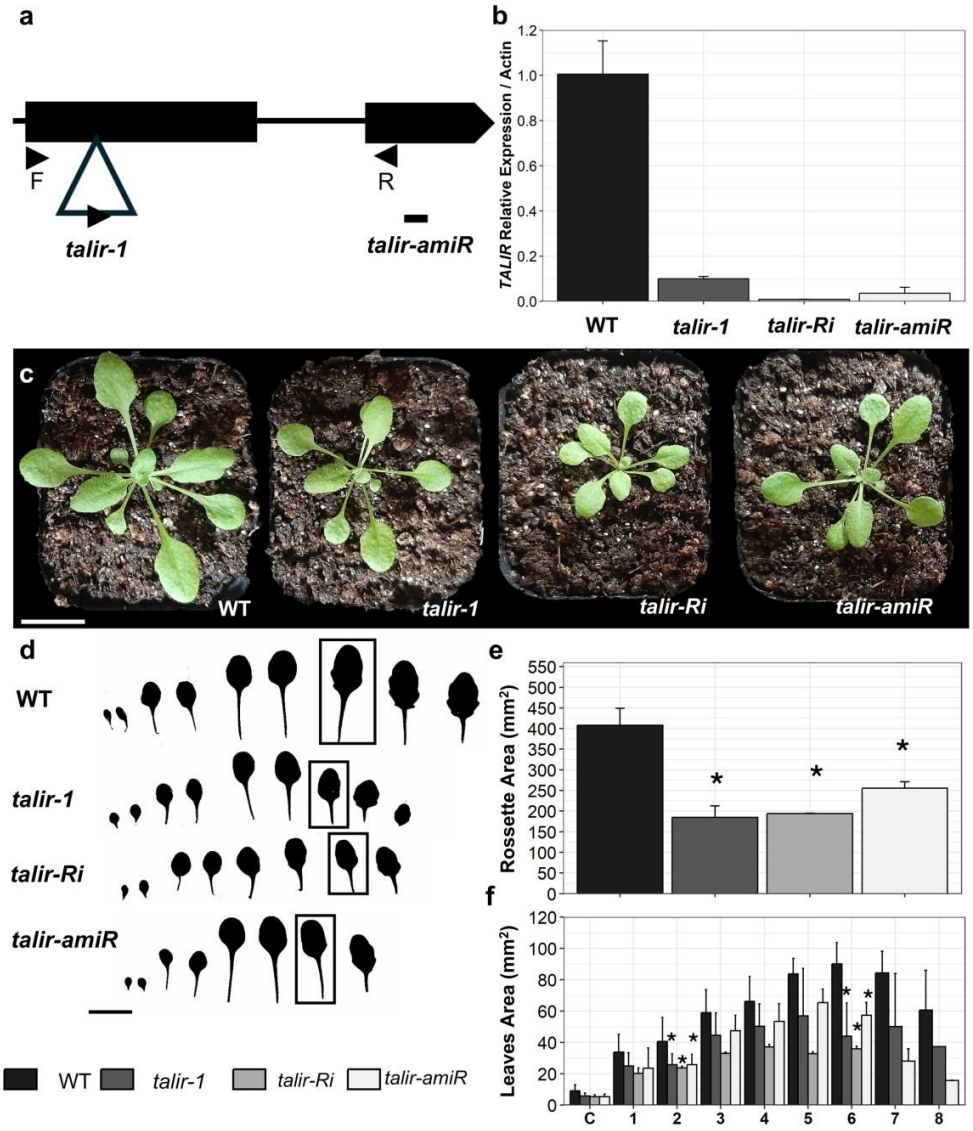


Figure 1. *TALIR* is required for leaf development in *Arabidopsis thaliana*.

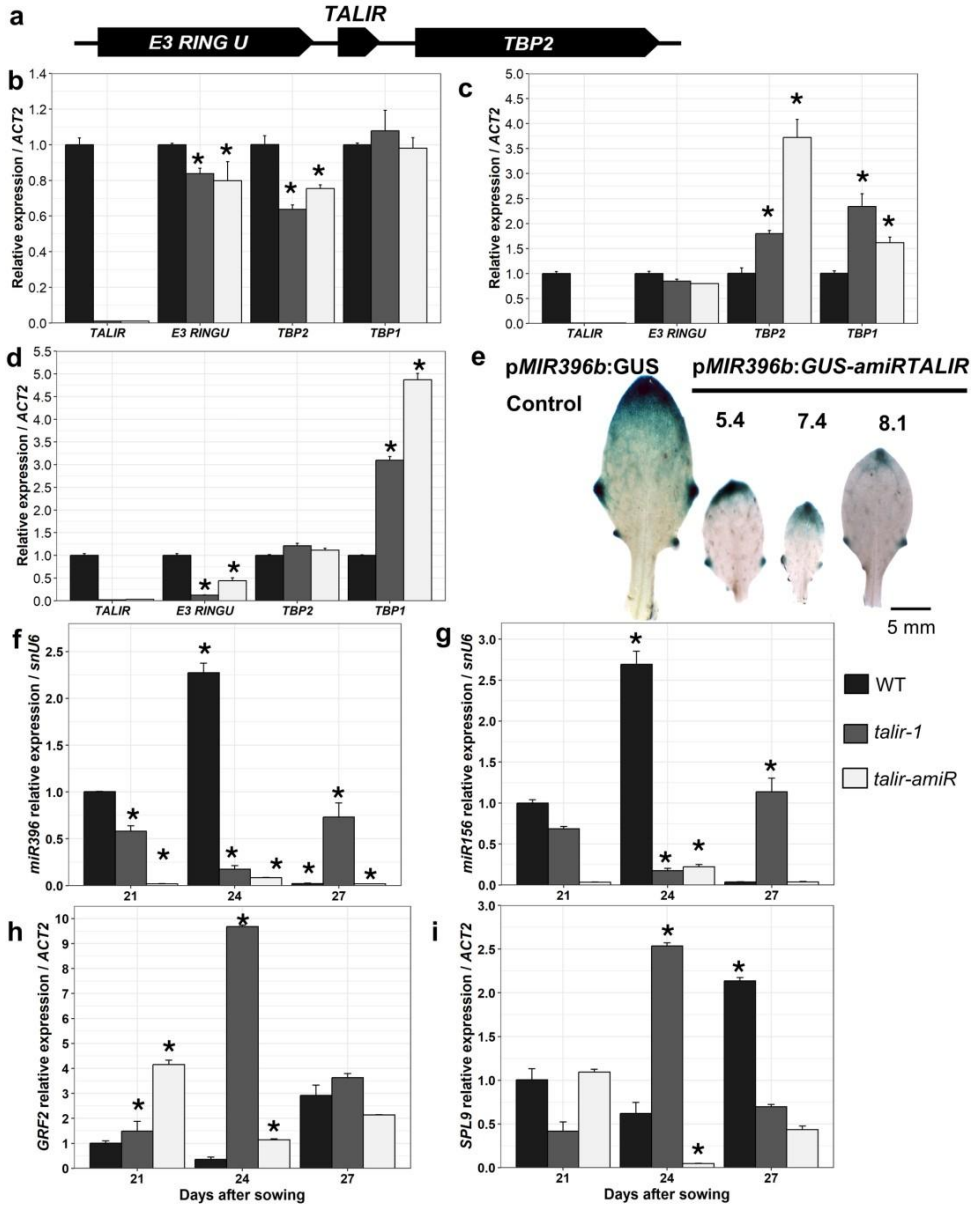


Figure 2. Dysregulation of transcript abundance of *miR396b* and *miR156* and the neighboring genes of *TALIR* during leaf development.

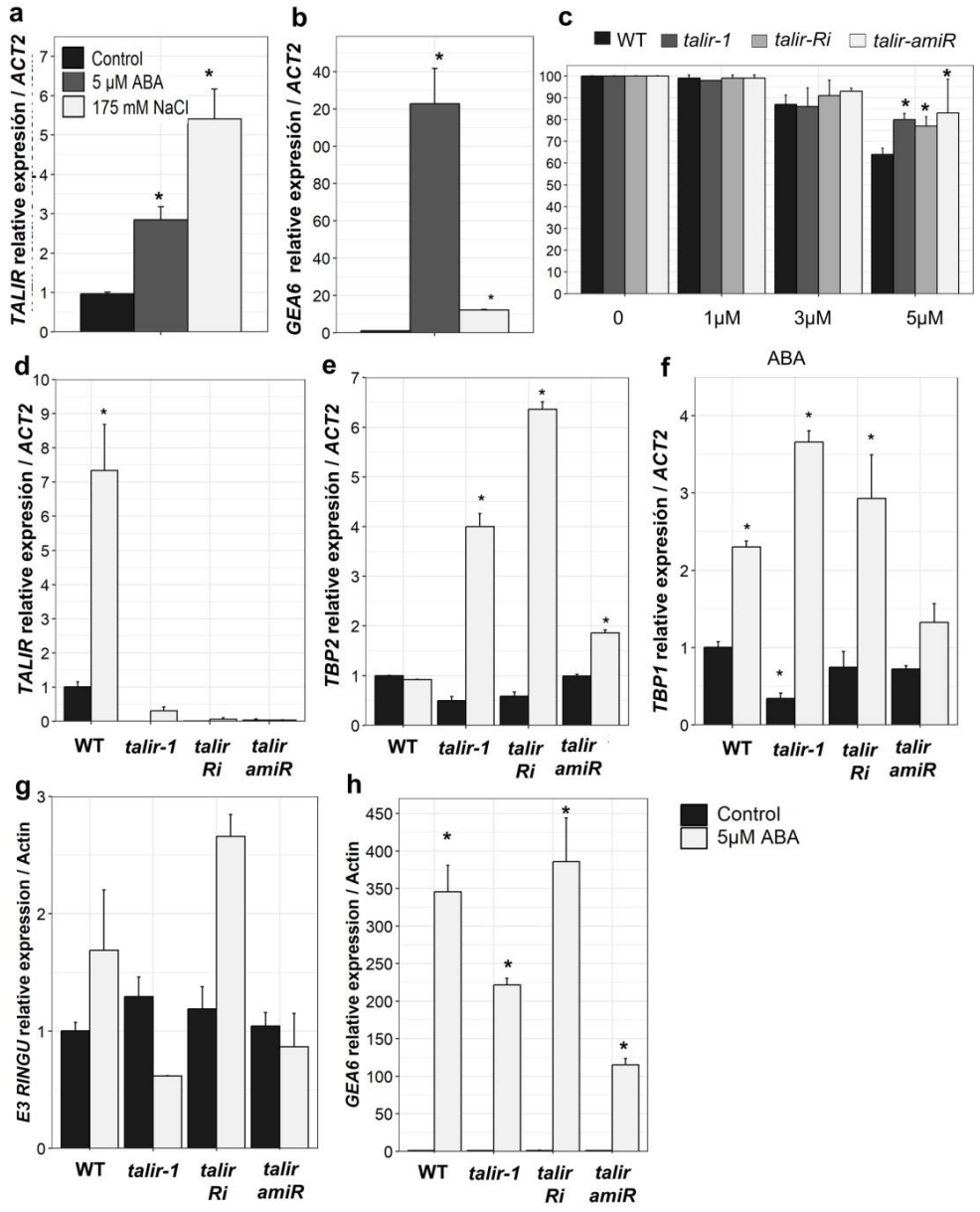


Figure 3. Accumulation of TALIR under stress conditions and ABA response during germination.

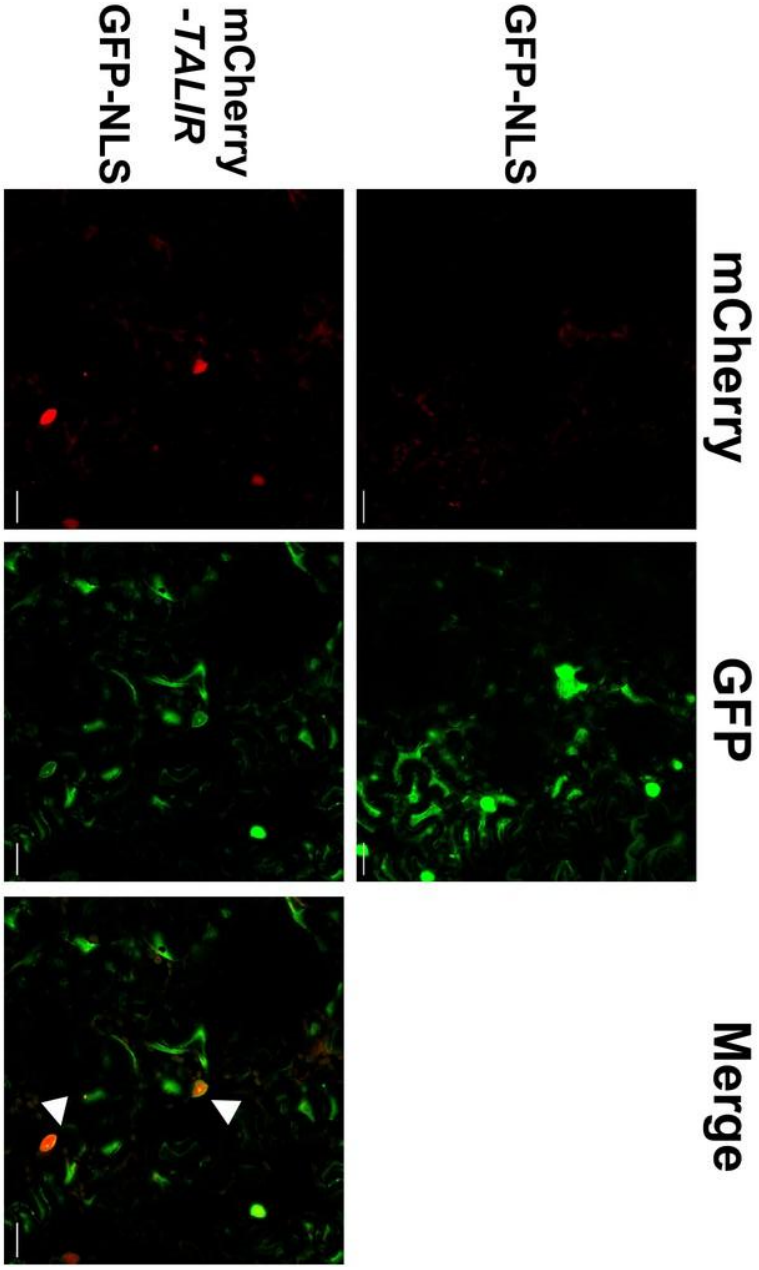


Figure 4. *TALIR* shows a nuclear localization in *Nicotiana benthamiana*

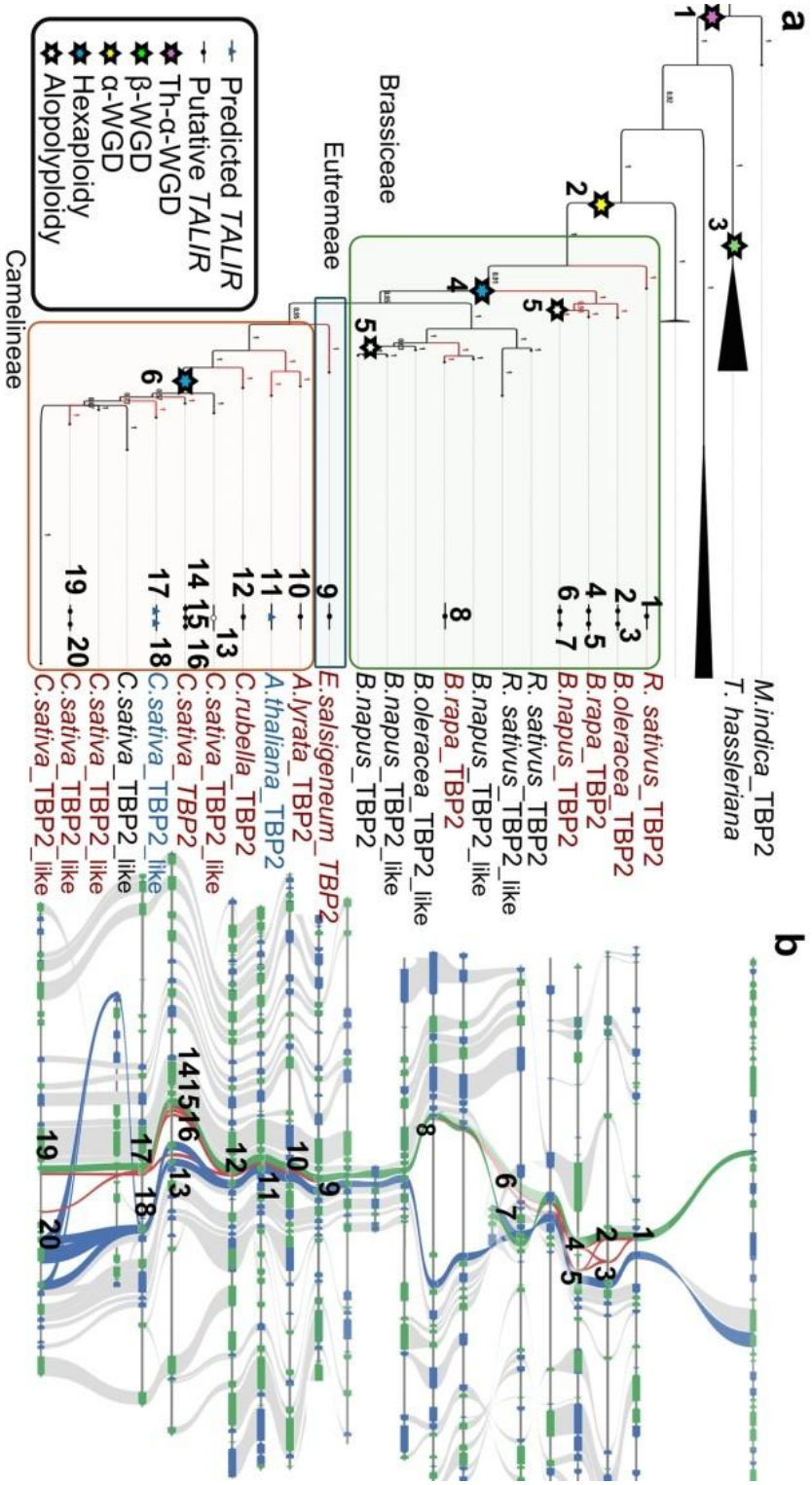


Figure 5. Resolved Brassicales plant phylogenetic tree for *TBP2* and microsynteny conservation.

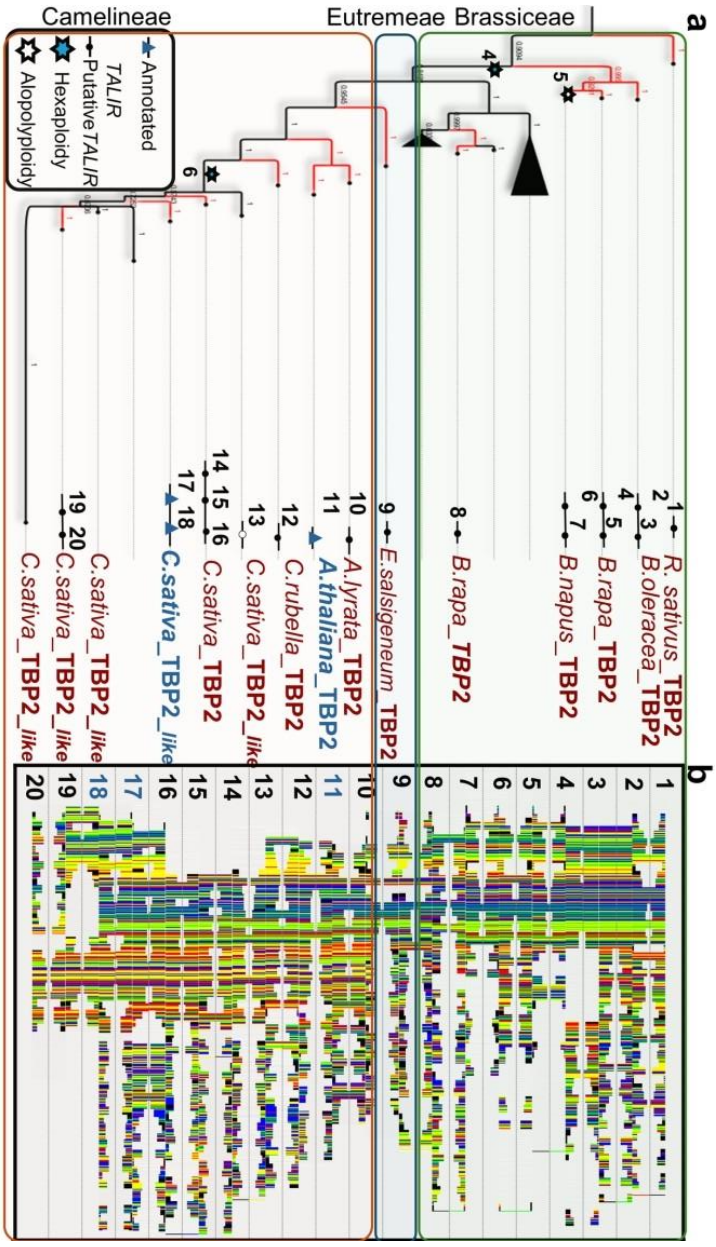


Figure 6. Zoomed-in view of the phylogenetic tree of *TBP2* and nucleotide conservation of *TALIR*

Supplementary Fig 1. *TALIR* genomic information and nucleotide sequence. TALIR (AT1G5525.1) is a non-coding RNA located on chromosome 1 (Chr1:20728329-20728745). TAIR Accession 6533805296. NCBI Reference

Sequence: NC_003070.9.

The full TALIR nucleotide sequence is shown.

>NC_003070.9:c20728745-20728329 Arabidopsis thaliana chromosome 1 sequence

AGATTAGTCATAATCCAAAAGAAGAAGTAATGAATCGTCAGGTTTTATCCCTAA

TCACTTTGATTATTCTCTTGGTCTCCGCGTACAACACGACTATGGGCCGAACT

CTGCCGAGTAAGTGACCGTCTTTGATCATACAAATACAATCACGCATAAACAA

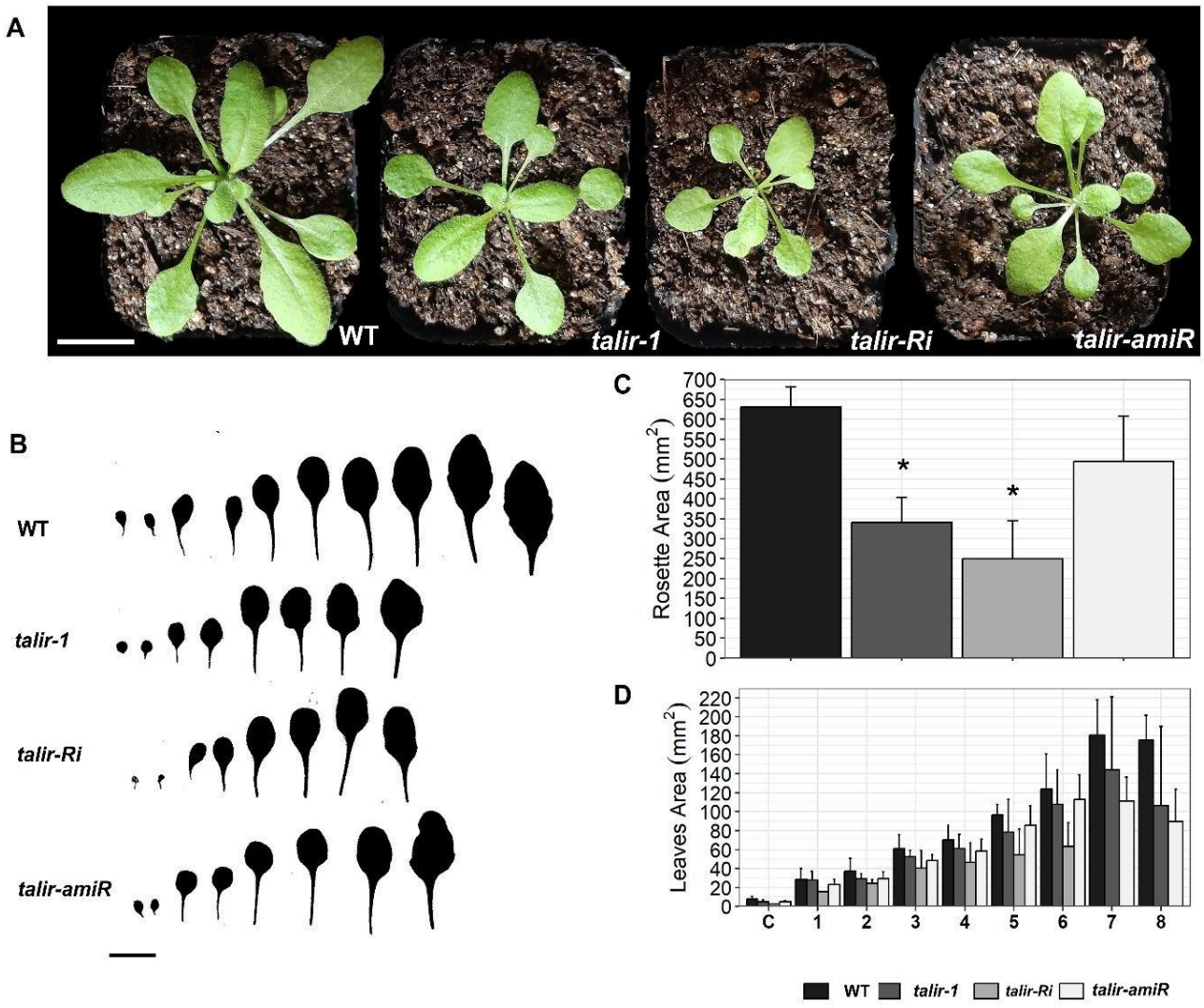
AACTCATATTCAGTTCTTAATCTCTTTTTTTATTATTATTATAAATTAAGGGGAA

GGAAATGGAGCACAACAATTTGATAATCTAAAAGCTGGTGTCAGACGAAGTGA

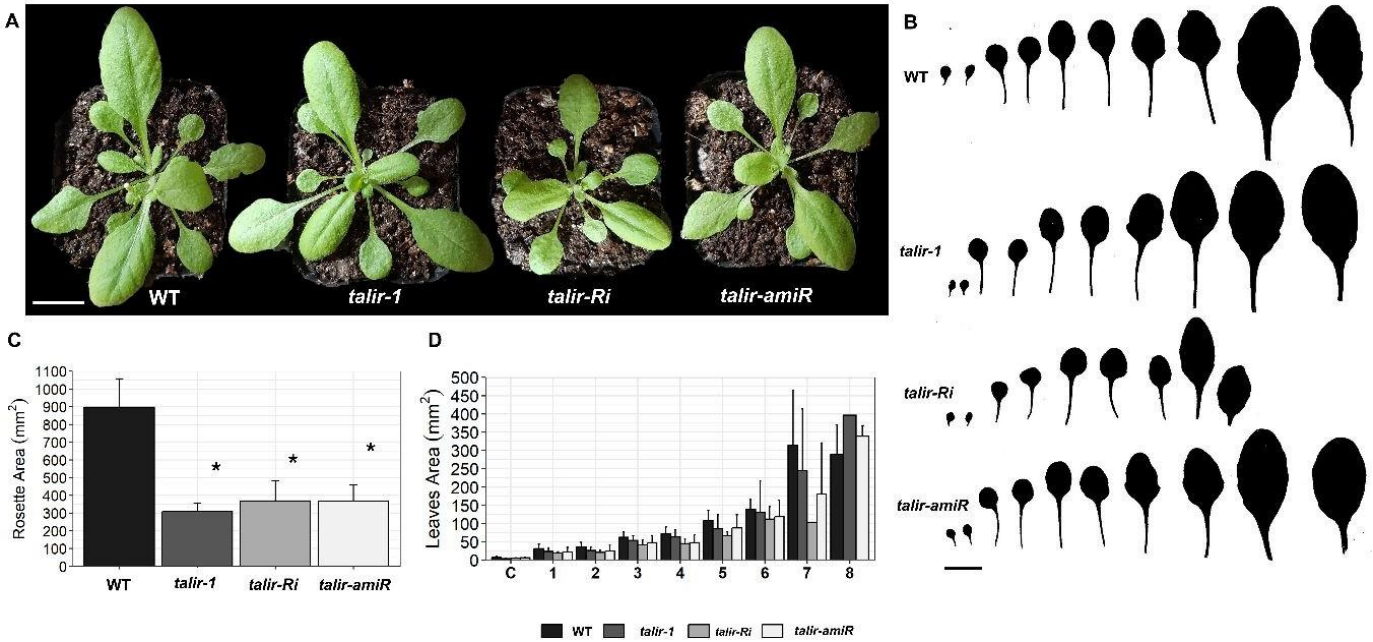
ATCGGCACGTACTCGTCAGAAAGGACCCGGACACTGATGAGTGATGATCTCT

AACAAGTTTTAGTTGGCACCTTCATGCAATAATTGATAATCTACTGCTTTGTAA

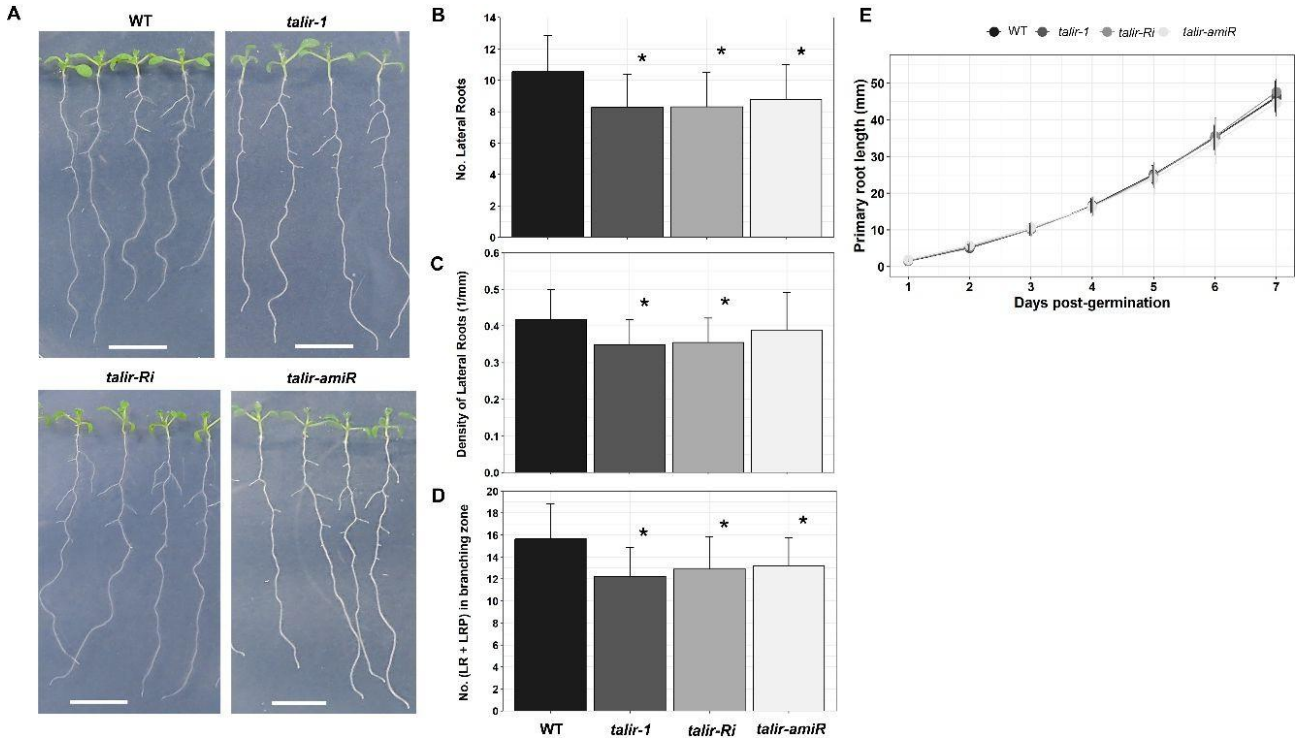
TAAGTCCGTCGTTGTGGCGATTCATAGTCTCTACTCGATTCAT



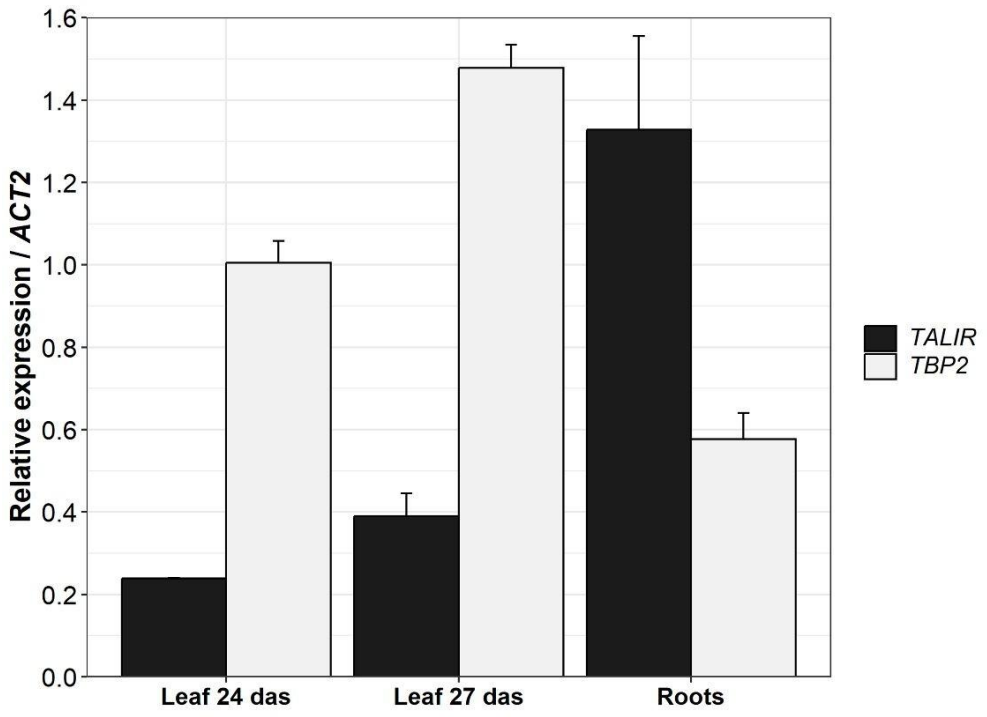
Supplementary Fig. 2. *TALIR* is required for leaf development in *Arabidopsis thaliana*. (A) Phenotypes of 24-day-old WT and mutant plants after sowing. Scale bar = 2 cm. (B) Sequential stages of leaf development including cotyledons. Scale bar = 1 cm. (C) Quantification of the total rosette and (d) individual leaf areas in both WT and mutant lines. Error bars represent the mean ± SD; n = 4. Asterisks (*) denote statistical significance at $P \leq 0.05$ (determined by the t-test). All experimental assays were performed with three independent biological replicates.



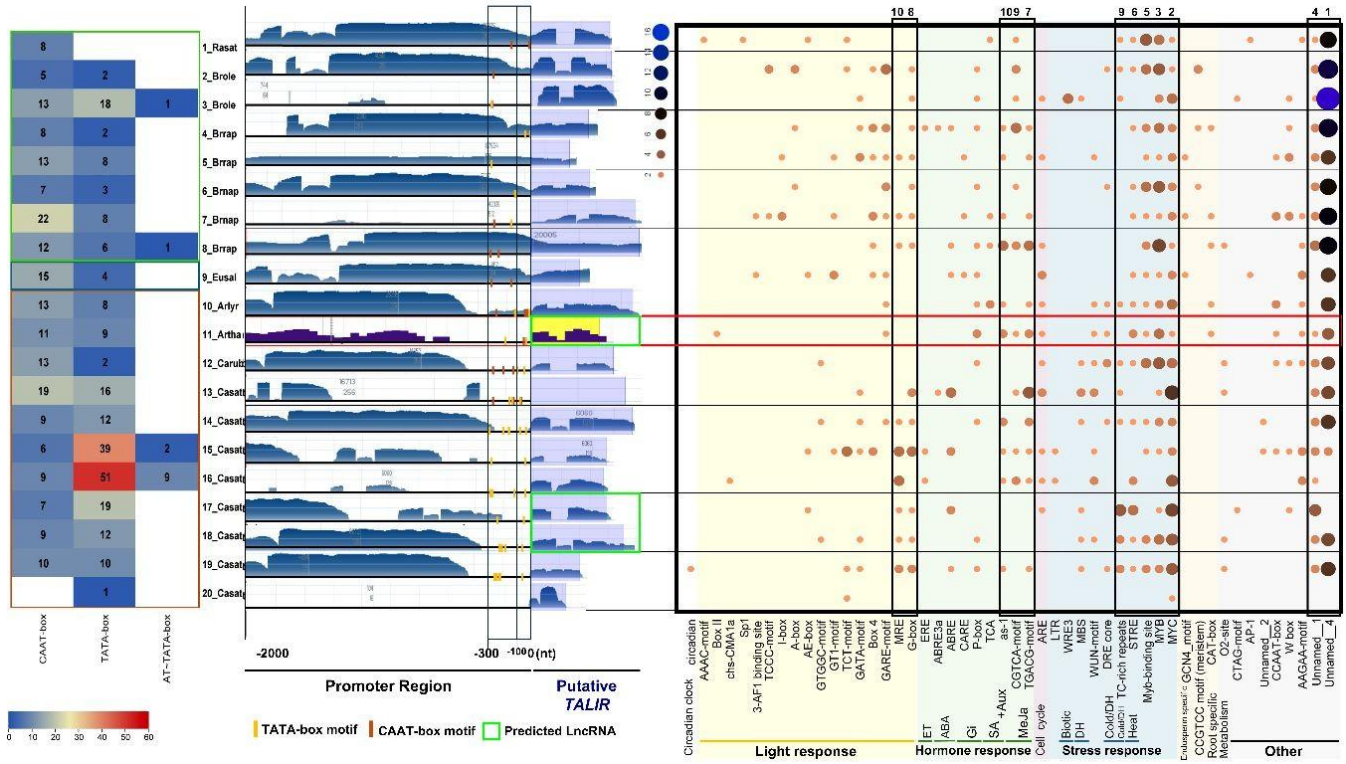
Supplementary Fig. 3. *TALIR* is required for leaf development in *Arabidopsis thaliana*. (A) Phenotypes of 27-day-old WT and mutant plants lines after sowing. Scale bar = 2 cm. (B) Sequential stages of leaf development including cotyledons. Scale bar = 1 cm. (C) Quantification of the total rosette and (D) individual leaf areas in both WT and mutant lines. Error bars represent the mean ± SD; n =4. Asterisks (*) denote statistical significance at $P \leq 0.05$ (determined by the t- test). All experimental assays were performed with three independent biological replicates.



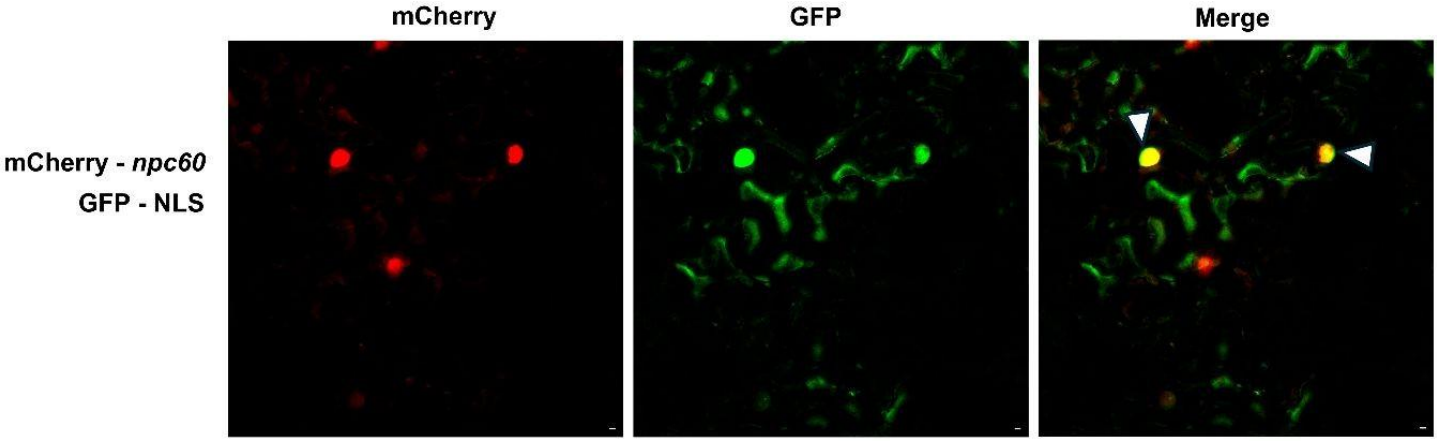
Supplementary Fig. 4. *TALIR* influences the development of roots in *Arabidopsis thaliana*. (A) Phenotypes of WT and *talir* mutant plants at 7 DPG. Scale bar = 1 cm. (B) Total number of lateral roots in both WT and *talir* mutant plants at 7 DPG. (C) Lateral root density in both WT and *talir* mutant plants at 7 DPG. (D) Quantification of the cumulative count of lateral roots and lateral root primordia within the branching zone in both WT and *talir* mutant plants at 7 DPG. Error bars represent the mean ± SD; n = 24 to 32. Asterisks (*) denote statistical significance at $P \leq 0.05$ (determined by the t-test). (E) Temporal changes in primary root elongation over the time during 7 days post-germination (DPG). All experimental assays were performed with two independent biological replicates.



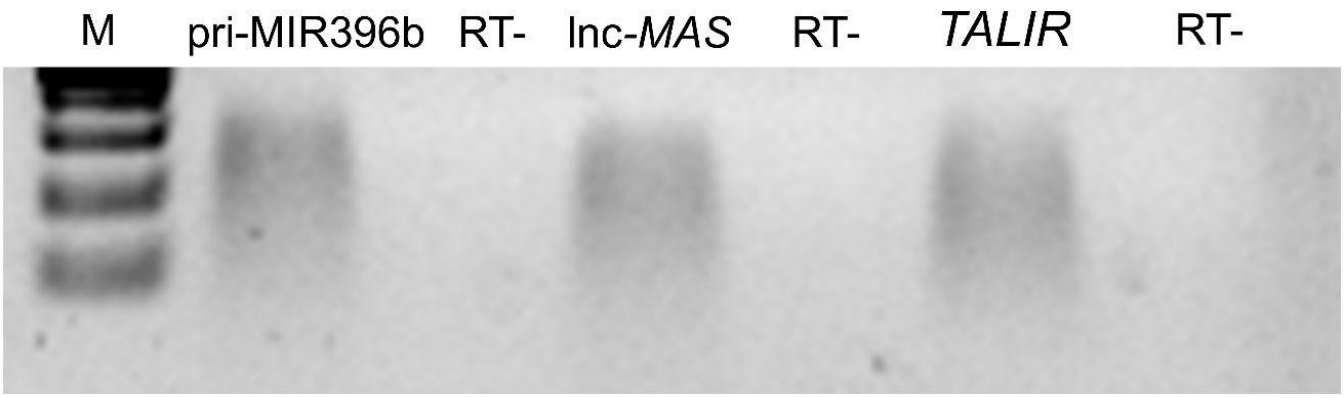
Supplementary Fig. 5. Accumulation of *TALIR* and *TBP2* transcripts in leaves and roots. RT-qPCR analyses were carried out to evaluate the transcript levels of *TALIR* and *TBP2* in leaves from 24 and 27 DAS WT plants. Roots from 7 day- after-germination plants. Relative abundance levels were normalized using the endogenous reference gene ACT2. Error bars represent the mean ± SD; n =3.



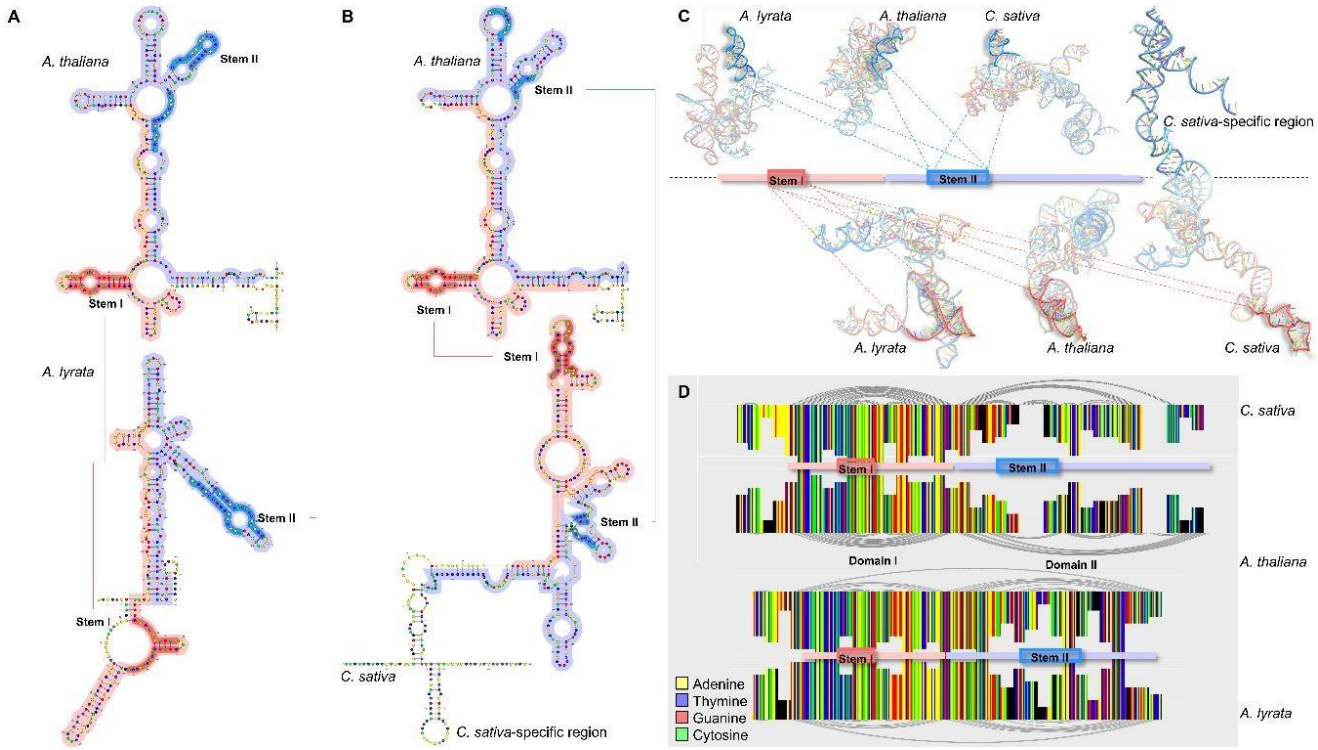
Supplementary Fig. 6. Promoter analysis and transcriptional activity of *TALIR* orthologs. On the left side of the image, the promoter analysis of *TALIR* and its orthologs is shown. The color scale indicates the frequency which the motifs CAAT-box, TATA-box and AT-TATA-box were observed. In the center of the image, the species in which *TALIR* was identified are indicated. The analyzed promoter region is represented in blue (0 to -2000 nt). Within the region -300 to 0 nt region, TATA-box and CAAT-box are marked. Adjacent to the promoter region, the transcriptional activity of each lncRNA orthologs are highlighted within a blue box, while predicted lncRNA are closed in a green box. On the right side of the image, the motifs identified in the promoter of the 20 *TALIR* orthologs are shown. Colored circles, ranging from orange to indigo, represent the frequency with which each motif was observed in the promoter region. Numbers at the top of the image indicate the ranking of motifs prevalent within the promoter region. *TALIR* is highlighted within a red box. Rasat, *R. sativus*; Brole, *B. oleracea*; Brrap, *B. rapa*; Brnap, *B. napus*; Eusal, *E. salsigeneum*; Arlyr, *A. lyrata*; Artha, *A. thaliana*; Carub, *C. rubella*; Casat, *C. sativa*.



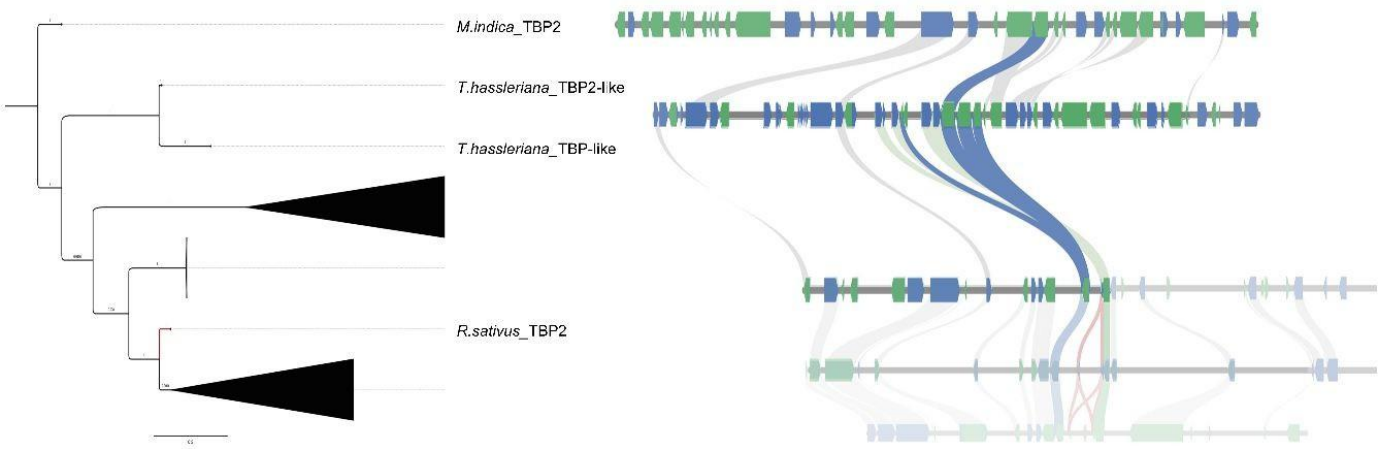
Supplementary Fig. 7. Transient nuclear localization of *NPC60 RNA* in *Nicotiana benthamiana*. Co-localization of MSCP protein fused with a GFP and an NLS along with RNA fused with 6X-MS2-mCherry. *NPC60* was used as a positive control. Arrowheads denote nuclei that show co-localization of both components. Scale bar = 20 μm, 60X objective. All experiments were independently performed twice.



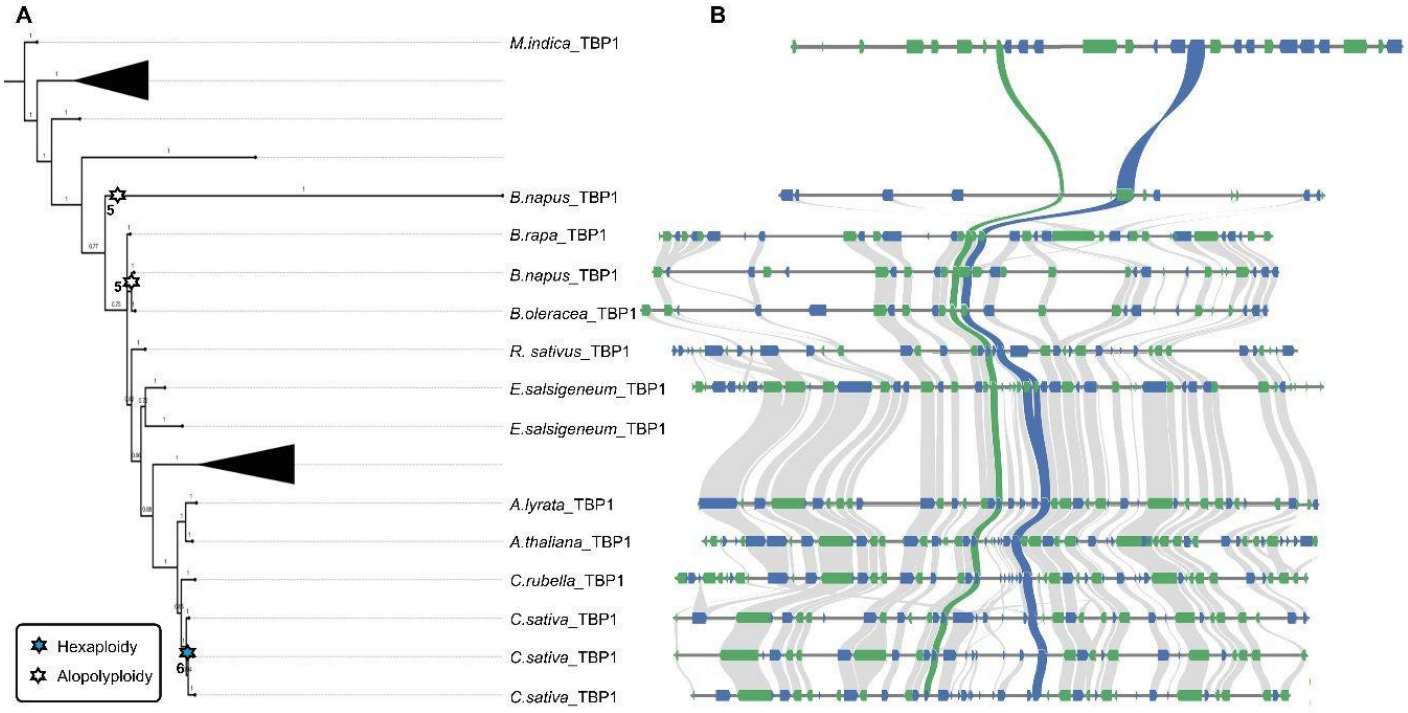
Supplementary Fig. 8. Agarose gel electrophoresis of end-point PCR products obtained from nuclear fraction. Amplification of *TALIR*, pri-*miR396b* and *MAS* transcripts is shown. Reactions lacking reverse transcriptase (RT-) were included as negative controls. M, DNA size marker.



Supplementary Fig. 9. Conservation of *TALIR* RNA secondary structure. (A) Predicted secondary structures of TALIR from *Arabidopsis thaliana* and *Arabidopsis lyrata* (upper and lower schematics, respectively). Only conserved stem–loop motifs identified by RNAfold and LocARNA are shown and highlighted in dark red (stem I) and dark blue (stem II). (B) Predicted secondary structures of *TALIR* from *A. thaliana* and *Camelina sativa* (upper and lower schematics, respectively), highlighting conserved stem–loop motifs as in panel A. (C) Three- dimensional models derived from the secondary structures shown in panels A and B. Conserved stem–loop motifs are highlighted in dark red and dark blue. (D) Structure-based sequence alignment generated with LocARNA. Arcs indicate predicted base-pairing interactions. Roman numerals denote conserved stem–loop motifs corresponding to those highlighted in panels A–C.

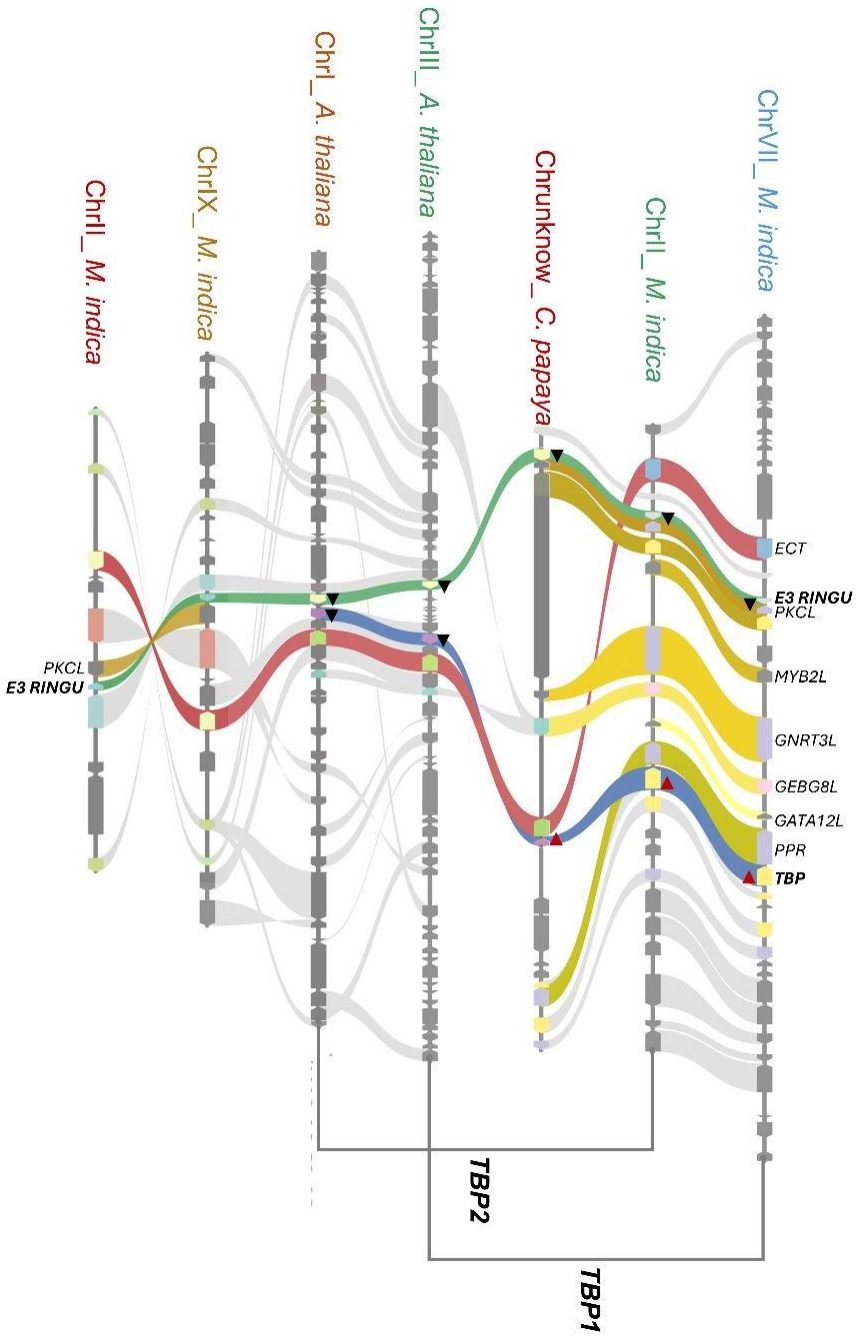


Supplementary Fig. 10. Plant phylogenetic tree for TBP2 and microsynteny conservation. Phylogenetic tree inferred by Bayesian method for four species. At the right side, conservation of microsynteny at the *loci*, where blue lines represent *TBP2*. The translucent image shows the position of *E3 RINGU* and *TALIR* in Brassicales



species.

Supplementary Fig. 11. Brassicales plant phylogenetic tree for TBP1 and microsynteny conservation. (A) Phylogenetic tree inferred by Bayesian method for ten species. Hexaploidy and allopolyploidy are represented by 2- pointed stars (blue and white). (B) Conservation of microsynteny at the *loci*, where blue lines represent *TBP1*, green lines represent E3 *RINGU*.



Supplementary Fig. 12. Putative origin of *TALIR* in the Brassicales. Comparison of *TBP1* and *TBP2 loci* microsynteny in *A. thaliana, M. indica,* and *C. papaya* where blue lines represent *TBP* genes and green lines represent *E3 RINGU* genes, and the red and black arrowheads represent the gene orientation. The orientation of the *TBP* and *E3 RINGU* genes is opposite in *M. indica loci*, unlike in *A. thaliana*, where both genes are coded in the same direction. Regions conserved between the different *loci* studied are represented in similarly color lines.

CONCLUSIÓN

Los resultados presentados en esta tesis evidencian que los lncRNAs son elementos clave que pueden modular diversos procesos biológicos en plantas, en particular la respuesta a las fitohormonas ET y ABA. A diferencia de los genes que codifican a proteínas, los lncRNAs muestran patrones de conservación que no se limitan a la secuencia nucleotídica, sino que también incluye una conservación a nivel estructural y de sintenia, lo que sugiere que la función de los lncRNAs podría estar vinculada tanto a su posición genómica como a su capacidad de interactuar con otras moléculas como proteínas y DNA.

En este contexto, el análisis *in silico* de la secuencia nucleotídica, la identificación de sORFs conservados, así como los análisis de estructurales y de sintenia revelaron que, aunque los lncRNAs presentan una menor conservación de secuencia en comparación con los genes codificantes, mantienen elementos conservados en términos de estructura y microsintenia. Esto sugiere que los lncRNAs pueden desempeñar funciones regulatorias dependientes de su localización genómica, actuando como moduladores *cis* de la expresión génica o como andamios estructurales para la interacción con proteínas y complejos regulatorios. Además, la presencia de sORFs conservados sugiere un rol potencial en la codificación de pequeños péptidos con roles regulatorios.

De manera particular, los análisis realizados sobre el lncRNA localizado río abajo de EIN2 (AT5G03285) mostraron que este locus se conserva en múltiples especies de plantas manteniendo su proximidad genómica con EIN2. Asimismo, se identificaron regiones conservadas que contienen sORFs y dominios estructurales organizados en estructuras tipo tallo-asa, algunas de las cuales presentan patrones de conservación entre especies. Estos resultados sugieren que determinados elementos estructurales y posicionales podrían haber sido conservados durante la evolución y representar componentes funcionalmente relevantes. Por otra parte, el análisis de TALIR reveló conservación de sintenia dentro del orden Brassicales, así como evidencias de conservación estructural y regulatoria en su región promotora. En conjunto, estos

hallazgos muestran que ambos lncRNAs conservan características estructurales, regulatorias y genómicas que podrían estar asociadas con su función biológica, aun cuando presentan una baja conservación de secuencia nucleotídica.

Adicionalmente, los resultados de expresión génica bajo distintas condiciones de estrés abiótico muestran que los lncRNAs *TALIR* y *AT5G03285* presentan patrones de respuesta específicos a estímulos ambientales, lo que refuerza su papel como moduladores de la integración de señales ambientales. Sus cambios en la expresión, junto con su proximidad a *loci* de genes clave en la señalización hormonal, sugieren que estos transcritos participan en la modulación fina de redes regulatorias complejas, contribuyendo a la plasticidad fenotípica y fisiológica de las plantas.

Desde una perspectiva de conservación estructural, los resultados sugieren a que la funcionalidad de los lncRNAs está asociada a la formación de estructuras secundarias que permiten interacciones específicas con otras moléculas, más allá de la simple secuencia lineal. Sin embargo, es necesario realizar análisis adicionales para determinar si esta conservación ocurre a nivel global o se limita a regiones específicas. Esto con el fin de respaldar la idea de que la evolución de los lncRNAs favorece la conservación de dominios estructurales y contextos genómicos, en lugar de secuencias altamente conservadas, lo que representa una trayectoria evolutiva distinta a la observada en genes codificantes.

En conjunto, estos resultados permiten proponer que los lncRNAs *TALIR* y *AT5G03285* forman parte de una capa regulatoria esencial en la señalización hormonal, particularmente en las vías del ET y de ABA, donde actúan como integradores de señales y moduladores de la expresión génica. Esto amplía la comprensión de una regulación lineal en las rutas de señalización en plantas, incorporando a los lncRNAs como elementos funcionales con un papel integrador y adaptable en la modulación de diversas respuestas hormonales.

Finalmente, el estudio de la conservación, la estructura y la expresión de los lncRNAs abre nuevas oportunidades para explorar su potencial en aplicaciones

biotecnológicas, especialmente en el desarrollo de estrategias orientadas a mejorar la respuesta al estrés en cultivos. La identificación de lncRNAs funcionalmente relevantes y conservados a nivel estructural y posicional representa un paso fundamental hacia la comprensión integral de la regulación génica en plantas y su aplicación biotecnológica.

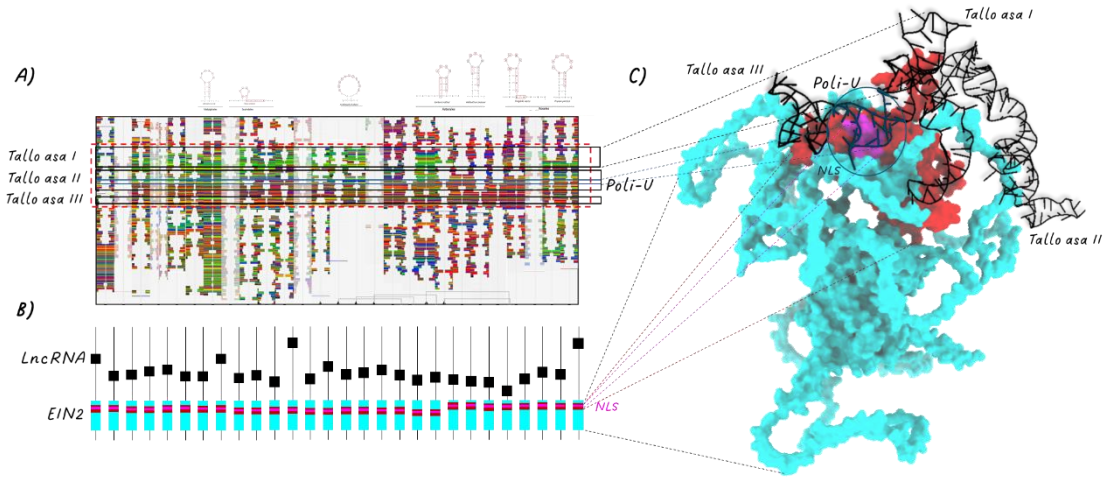


Figura 4. *Modelo de la conserva*ción del lncRNA *AT5G03285* y su modelo de interacción con EIN2.

(A) Alineamiento de las secuencias del lncRNA *AT5G03285* en diferentes especies vegetales, donde se destaca un bloque de secuencia altamente conservado (recuadro rojo); en la parte superior se muestran las predicciones de la estructura secundaria en forma de tallo-asa de regiones ricas en uracilos. (B) Esquema representativo de microsintenia que ilustra la conservación posicional del *locus* del lncRNA y *EIN2* entre especies, resaltando la ubicación de la señal de localización nuclear (NLS) dentro del gen *EIN2*. (C) Modelo propuesto del complejo, donde la proteína EIN2 se representa en superficie (cian) interactuando físicamente con el dominio funcional conservado del lncRNA.

BIBLIOGRAFÍA

Alonso, J. M., Hirayama, T., Roman, G., Nourizadeh, S., & Ecker, J. R. (1999). EIN2, a bifunctional transducer of ethylene and stress responses in Arabidopsis. *Science*, *284*(5423), 2148–2152. https://doi.org/10.1126/science.284.5423.2148

Ariel, F., Lucero, L., Christ, A., Mammarella, M. F., Jegu, T., Veluchamy, A., Mariappan, K., Latrasse, D., Blein, T., Liu, C., Benhamed, M., & Crespi, M. (2020). R-Loop Mediated trans Action of the APOLO Long Noncoding RNA. *Molecular Cell*, *77*(5), 1055–1065.e4. https://doi.org/10.1016/j.molcel.2019.12.015

Bailey-Serres, J., Parker, J. E., Ainsworth, E. A., Oldroyd, G. E. D., & Schroeder, J. I. (2019). Genetic strategies for improving crop yields. Nature, 575(7781), 109–118. https://doi.org/10.1038/s41586-019-1679-0

Bardou, F., Ariel, F., Simpson, C. G., Romero-Barrios, N., Laporte, P., Balzergue, S., Brown, J. W., & Crespi, M. (2014). Long noncoding RNA modulates alternative splicing regulators in Arabidopsis. *Developmental Cell*, *30*(2), 166–176. https://doi.org/10.1016/j.devcel.2014.06.017

Binder, B. M. (2020). Ethylene signaling in plants. The Journal of Biological Chemistry, 295(22), 7710–7725. https://doi.org/10.1074/jbc.REV120.010854

Bita, C. E., & Gerats, T. (2013). Plant tolerance to high temperature in a changing environment: scientific fundamentals and production of heat stress-tolerant crops. *Frontiers in Plant Science, 4*, 273. https://doi.org/10.3389/fpls.2013.00273

Cao, P., Zhan, C., Yin, J., Gong, S., Ma, D., & Li, Y. (2022). Genome-wide identification of long intergenic non-coding RNAs for *Ralstonia solanacearum* resistance in tomato (*Solanum lycopersicum*). *Frontiers in Plant Science*, *13*, 981281. https://doi.org/10.3389/fpls.2022.981281

Chang, C. (2016). Q&A: How do plants respond to ethylene and what is its importance?. *BMC Biology*, *14*, 7. https://doi.org/10.1186/s12915-016-0230-0

Chekanova, J. A. (2015). Long non-coding RNAs and their functions in plants. *Current Opinion in Plant Biology*, *27*, 207–216. https://doi.org/10.1016/j.pbi.2015.08.003

Chen, Y. J., Chen, C. Y., Mai, T. L., Chuang, C. F., Chen, Y. C., Gupta, S. K., Yen, L., Wang, Y. D., & Chuang, T. J. (2020). Genome-wide, integrative analysis of circular RNA dysregulation and the corresponding circular RNA-microRNA-mRNA regulatory axes in autism. Genome Research, 30(3), 375–391. https://doi.org/10.1101/gr.255463.119

Cheng, B., Pei, W., Wan, K., Pan, R., & Zhang, W. (2024). LncRNA cis- and trans-regulation provides new insight into drought stress responses in wild barley. *Physiologia Plantarum, 176*(4), e14424. https://doi.org/10.1111/ppl.14424

Chugunova, A., Navalayeu, T., Dontsova, O., & Sergiev, P. (2018). Mining for Small Translated ORFs. *Journal of Proteome Research*, *17*(1), 1–11. https://doi.org/10.1021/acs.jproteome.7b00707

Dey, S. S., Sharma, P. K., Munshi, A. D., Jaiswal, S., Behera, T. K., Kumari, K., G, B., Iquebal, M. A., Bhattacharya, R. C., Rai, A., & Kumar, D. (2022). Genome wide identification of lncRNAs and circRNAs having regulatory role in fruit shelf life

in health crop cucumber (*Cucumis sativus L.*). *Frontiers in Plant Science*, *13*, 884476. https://doi.org/10.3389/fpls.2022.884476

Di, C., Yuan, J., Wu, Y., Li, J., Lin, H., Hu, L., Zhang, T., Qi, Y., Gerstein, M. B., Guo, Y., & Lu, Z. J. (2014). Characterization of stress-responsive lncRNAs in Arabidopsis thaliana by integrating expression, epigenetic and structural features. *The Plant Journal*, *80*(5), 848–861. https://doi.org/10.1111/tpj.12679

Fu, L., Liu, Y., Qin, G., Wu, P., Zi, H., Xu, Z., Zhao, X., Wang, Y., Li, Y., Yang, S., Peng, C., Wong, C. C. L., Yoo, S. D., Zuo, Z., Liu, R., Cho, Y. H., & Xiong, Y. (2021). The TOR-EIN2 axis mediates nuclear signalling to modulate plant growth. *Nature*, *591*(7849), 288–292. https://doi.org/10.1038/s41586-021-03310-y

Gallie, D. R. (2015). Ethylene receptors in plants - why so much complexity?. *F1000Prime Reports*, *7*, 39. https://doi.org/10.12703/P7-39

Gilroy, S., Białasek, M., Suzuki, N., Górecka, M., Devireddy, A. R., Karpiński, S., & Mittler, R. (2016). ROS, Calcium, and Electric Signals: Key Mediators of Rapid Systemic Signaling in Plants. *Plant Physiology, 171*(3), 1606–1615. https://doi.org/10.1104/pp.16.00434

Gultyaev, A. P., Koster, C., van Batenburg, D. C., Sistermans, T., van Belle, N., Vijfvinkel, D., & Roussis, A. (2023). Conserved structured domains in plant non-coding RNA enod40, their evolution and recruitment of sequences from transposable elements. *NAR Genomics and Bioinformatics, 5*(4), lqad091. https://doi.org/10.1093/nargab/lqad091

Hajjari, M., & Salavaty, A. (2015). HOTAIR: an oncogenic long non-coding RNA in different cancers. *Cancer Biology & Medicine, 12*(1), 1–9. https://doi.org/10.7497/j.issn.2095-3941.2015.0006

Hirayama, T., Kieber, J. J., Hirayama, N., Kogan, M., Guzman, P., Nourizadeh, S., Alonso, J. M., Dailey, W. P., Dancis, A., & Ecker, J. R. (1999). RESPONSIVE-TO-ANTAGONIST1, a Menkes/Wilson disease-related copper transporter, is required for ethylene signaling in Arabidopsis. *Cell*, *97*(3), 383–393. https://doi.org/10.1016/s0092-8674(00)80747-3

Jin, J., Ohama, N., He, X., Wu, H. W., & Chua, N. H. (2022). Tissue-specific transcriptomic analysis uncovers potential roles of natural antisense transcripts in *Arabidopsis* heat stress response. *Frontiers in Plant Science*, *13*, 997967. https://doi.org/10.3389/fpls.2022.997967

Kapusta, A., & Feschotte, C. (2014). Volatile evolution of long noncoding RNA repertoires: mechanisms and biological implications. *Trends in Genetics, 30*(10), 439–452. https://doi.org/10.1016/j.tig.2014.08.004

Lämke, J., & Bäurle, I. (2017). Epigenetic and chromatin-based mechanisms in environmental stress adaptation and stress memory in plants. *Genome Biology*, *18*(1), 124. https://doi.org/10.1186/s13059-017-1263-6

Lee, G., DiBiase, C. N., Liu, B., Li, T., McCoy, A. G., Chilvers, M. I., Sun, L., Wang, D., Lin, F., & Zhao, M. (2024). Transcriptomic and epigenetic responses shed light on soybean resistance to Phytophthora sansomeana. *The Plant Genome*, e20487. Advance online publication. https://doi.org/10.1002/tpg2.20487

Li, W., Ma, M., Feng, Y., Li, H., Wang, Y., Ma, Y., Li, M., An, F., & Guo, H. (2015). EIN2-directed translational regulation of ethylene signaling in Arabidopsis. Cell, 163(3), 670–683. https://doi.org/10.1016/j.cell.2015.09.037

Ma, L., Zheng, Y., Zhou, Z., Deng, Z., Tan, J., Bai, C., Fu, A., Wang, Q., & Zuo, J. (2024). Dissection of mRNA ac4C acetylation modifications in AC and Nr fruits: insights into the regulation of fruit ripening by ethylene. *Molecular Horticulture, 4*(1), 5. https://doi.org/10.1186/s43897-024-00082-7

Maric, A., Agashe, A., Söntgerath, J., & Hartman, S. (2025). Ethylene signal integration through epigenetic mechanisms in plants. *Current Opinion in Plant Biology*, *87*, 102780. https://doi.org/10.1016/j.pbi.2025.102780

Nejat, N., & Mantri, N. (2018). Emerging roles of long non-coding RNAs in plant response to biotic and abiotic stresses. *Critical Reviews in Biotechnology*, *38*(1), 93–105. https://doi.org/10.1080/07388551.2017.1312270

Nieto-Hernández, J., Arenas-Huertero, C., & Ibarra-Laclette, E. (2023). LncRNA-encoded peptides: the case of the lncRNA gene located downstream of EIN2. *Functional & Integrative Genomics*, *23*(2), 108. https://doi.org/10.1007/s10142-023-01038-8

Nieto-Hernández, J. (2020). Caracterización de la función del RNA intermedio no codificante *lmncEIN2* en la vía de percepción del etileno [Tesis de maestría, Universidad Autónoma de San Luis Potosí].

Qiao, H., Shen, Z., Huang, S. S., Schmitz, R. J., Urich, M. A., Briggs, S. P., & Ecker, J. R. (2013). Response to perspective: "separation anxiety: an analysis of

ethylene-induced cleavage of EIN2". *Plant Signaling & Behavior*, *8*(8), e25037. https://doi.org/10.4161/psb.25037

Qiao, H., Shen, Z., Huang, S. S., Schmitz, R. J., Urich, M. A., Briggs, S. P., & Ecker, J. R. (2012). Processing and subcellular trafficking of ER-tethered EIN2 control response to ethylene gas. *Science*, *338*(6105), 390–393. https://doi.org/10.1126/science.1225974

Reichel, M., Liao, Y., Rettel, M., Ragan, C., Evers, M., Alleaume, A. M., Horos, R., Hentze, M. W., Preiss, T., & Millar, A. A. (2016). In Planta Determination of the mRNA-Binding Proteome of Arabidopsis Etiolated Seedlings. *The Plant Cell, 28*(10), 2435–2452. https://doi.org/10.1105/tpc.16.00562

Ruiz-Orera, J., & Albà, M. M. (2019). Translation of Small Open Reading Frames: Roles in Regulation and Evolutionary Innovation. *Trends in Genetics*, *35*(3), 186–198. https://doi.org/10.1016/j.tig.2018.12.003

Suzuki, N., Koussevitzky, S., Mittler, R., & Miller, G. (2012). ROS and redox signalling in the response of plants to abiotic stress. *Plant, Cell & Environment*, *35*(2), 259–270. https://doi.org/10.1111/j.1365-3040.2011.02336.x

Tan, X., Li, S., Hu, L., & Zhang, C. (2020). Genome-wide analysis of long non-coding RNAs (lncRNAs) in two contrasting rapeseed (Brassica napus L.) genotypes subjected to drought stress and re-watering. *BMC Plant Biology, 20*(1), 81. https://doi.org/10.1186/s12870-020-2286-9

Uroda, T., Anastasakou, E., Rossi, A., Teulon, J. M., Pellequer, J. L., Annibale, P., Pessey, O., Inga, A., Chillón, I., & Marcia, M. (2019). Conserved Pseudoknots

in lncRNA MEG3 Are Essential for Stimulation of the p53 Pathway. *Molecular Cell*, *75*(5), 982–995.e9. https://doi.org/10.1016/j.molcel.2019.07.025

Wang, J., Meng, X., Dobrovolskaya, O. B., Orlov, Y. L., & Chen, M. (2017). Non-coding RNAs and Their Roles in Stress Response in Plants. *Genomics, Proteomics & Bioinformatics, 15*(5), 301–312. https://doi.org/10.1016/j.gpb.2017.01.007

Wang, X., Fan, H., Wang, B., & Yuan, F. (2023). Research progress on the roles of lncRNAs in plant development and stress responses. *Frontiers in Plant Science*, *14*, 1138901. https://doi.org/10.3389/fpls.2023.1138901

Wang, X., Wen, H., Suprun, A., & Zhu, H. (2025). Ethylene Signaling in Regulating Plant Growth, Development, and Stress Responses. *Plants*, 14(3), 309. https://doi.org/10.3390/plants14030309

Yang, H., Cui, Y., Feng, Y., Hu, Y., Liu, L., & Duan, L. (2023). Long Non-Coding RNAs of Plants in Response to Abiotic Stresses and Their Regulating Roles in Promoting Environmental Adaption. *Cells, 12*(5), 729. https://doi.org/10.3390/cells12050729

Yu, J., Qiu, K., Sun, W., Yang, T., Wu, T., Song, T., Zhang, J., Yao, Y., & Tian, J. (2022). A long noncoding RNA functions in high-light-induced anthocyanin accumulation in apple by activating ethylene synthesis. *Plant Physiology*, *189*(1), 66–83. https://doi.org/10.1093/plphys/kiac049

Zandalinas, S. I., Sengupta, S., Fritschi, F. B., Azad, R. K., Nechushtai, R., & Mittler, R. (2021). The impact of multifactorial stress combination on plant growth and

survival. *The New Phytologist*, *230*(3), 1034–1048. https://doi.org/10.1111/nph.17232

Zhang, F., Wang, L., Qi, B., Zhao, B., Ko, E. E., Riggan, N. D., Chin, K., & Qiao, H. (2017). EIN2 mediates direct regulation of histone acetylation in the ethylene response. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(38), 10274–10279. https://doi.org/10.1073/pnas.1707937114

Zhang, G., Lan, Y., Xie, A., Shi, J., Zhao, H., Xu, L., Zhu, S., Luo, T., Zhao, T., Xiao, Y., & Li, X. (2019). Comprehensive analysis of long noncoding RNA (lncRNA)-chromatin interactions reveals lncRNA functions dependent on binding diverse regulatory elements. *The Journal of Biological Chemistry, 294*(43), 15613–15622. https://doi.org/10.1074/jbc.RA119.008732

Zhao, M., Wang, T., Sun, T., Yu, X., Tian, R., & Zhang, W. H. (2020). Identification of tissue-specific and cold-responsive lncRNAs in Medicago truncatula by high-throughput RNA sequencing. *BMC Plant Biology*, *20*(1), 99. https://doi.org/10.1186/s12870-020-2301-1

Zhao, Z., Yang, Y., Iqbal, A., Wu, Q., & Zhou, L. (2024). Biological Insights and Recent Advances in Plant Long Non-Coding RNA. *International Journal of Molecular Sciences*, *25*(22), 11964. https://doi.org/10.3390/ijms252211964

Zhu, J. K. (2016). Abiotic Stress Signaling and Responses in Plants. *Cell*, *167*(2), 313–324. https://doi.org/10.1016/j.cell.2016.08.029

Home > Functional & Integrative Genomics > Article

LncRNA-encoded peptides: the case of the lncRNA gene located downstream of *EIN2*

Brief Report | Published: 29 March 2023

Volume 23, article number 108, (2023) Cite this article

Functional & Integrative Genomics

Aims and scope →

Submit manuscript →

Save article

PCP
PLANT & CELL PHYSIOLOGY

A synteny conserved TATA binding protein 2-associated long intergenic non-coding RNA influences leaf development and ABA response in *Arabidopsis thaliana*

Byanka Sthefany Espinoza-López[1], Jesús Nieto-Hernández[1], Ángel Eduardo Quezada-Eguía[2], Susana Isabel Vargas-Camacho[1], Jaime Aportela-Cortez[2], José Luis Reyes[3] and Catalina Arenas-Huertero[2,*]

[1]Posgrado en Ciencias en Bioprocesos, Facultad de Ciencias Químicas, Universidad Autónoma de San Luis Potosí, Av. Dr. Salvador Nava Mtz. s/n, 78290, San Luis Potosí, México, [2]Facultad de Ciencias, Universidad Autónoma de San Luis Potosí, Av. Chapultepec #1570, 78295, San Luis Potosí, México, [3]Departamento de Biología Molecular de Plantas, Instituto de Biotecnología, UNAM, Av. Universidad 2001, 62210, Cuernavaca, México

*Corresponding author: E-mail, catalina.arenas@uaslp.mx

Received 22 September 2025; Accepted 2 March 2026

Regular Paper