



Escuela Técnica Superior
de Ingeniería Agronómica y del Medio Natural



CARACTERIZACIÓN FUNCIONAL DE LOS ELEMENTOS EN CIS INVOLUCRADOS EN LA REGULACIÓN TRADUCCIONAL DE LA MADURACIÓN DEL FRUTO

CARACTERIZACIÓN "IN SILICO" DE LOS 3'UTR DE LOS
ORTÓLOGOS DE *EBF2*



TITULACIÓN: GRADO EN BIOTECNOLOGÍA

CURSO ACADÉMICO: 4º

AUTOR: ANDRÉS LUIS MATRES FERRANDO

TUTOR ACADÉMICO: JAVIER FORMEN MILLET

COTUTOR EXTERNO: JAVIER BRUMÓS FUENTES

FECHA DE ENTREGA: 05-07-2021

LOCALIDAD: VALENCIA

UNIVERSIDAD POLITÉCNICA DE VALENCIA,
Camino de Vera, s/n, 46022 Valencia

Resumen y palabras clave

Título: Caracterización funcional de los elementos en cis involucrados en la regulación traduccional de la maduración del fruto.

Resumen:

El etileno es la hormona vegetal clave que promueve la maduración, senescencia y ablandamiento de los cultivos hortícolas y es uno de los principales determinantes de la calidad y vida útil de las frutas y hortalizas. Una mejor comprensión de cómo el etileno regula la maduración de los frutos podría tener aplicaciones directas en la reducción de su deterioro.

Durante las últimas décadas, los cambios de expresión genética durante la maduración del fruto han sido estudiados a nivel transcripcional, pero los cambios asociados con la maduración en la eficiencia de la traducción se mantienen inexplorados. Nuestros datos preliminares sugieren que un subconjunto de transcritos muestra cambios asociados con la maduración en su eficiencia traduccional. Para identificar transcritos regulados a nivel traduccional durante la maduración del fruto, usamos la tecnología Ribo-seq que permite monitorear la traducción a escala de genoma completo con una resolución de un solo codón.

En presencia de etileno, los 3'UTR de los transcritos de los genes EBFs en Arabidopsis han sido identificados como elementos reguladores en cis suficientes para la inhibición traduccional de la expresión genética. En este estudio, se explorarán los mecanismos moleculares responsables para la regulación de la traducción conferidos por el 3'UTR de los genes EBFs en respuesta a etileno y, por tanto, durante la maduración del fruto. En primer lugar, se llevará a cabo un enfoque filogenético imparcial para estudiar las regiones conservadas presentes en el 3'UTR de los genes EBFs en el Reino Vegetal, en especies de Solanáceas y en accesiones de tomate. En segundo lugar, usando algoritmos de predicción, se modelarán las estructuras secundarias adoptadas por el 3'UTR para buscar estructuras conservadas.

Los resultados de este estudio nos permitirán optimizar el empleo de módulos reguladores de la traducción, los cuales son una herramienta prometedora en las áreas de la biotecnología vegetal y de la biología sintética con potenciales aplicaciones directas en nuestros sistemas agrícolas.

Palabras clave: Etileno; 3'UTR; genes EBFs, elementos reguladores en cis; expresión genética; regulación traduccional

Autor: Andrés Luis Matres Ferrando

Tutor: Javier Brumós Fuentes

Valencia, Julio 2021

Title: Functional characterization of cis-elements involved in translational regulation of fruit ripening.

Abstract:

Ethylene is the key plant hormone that promotes ripening, senescence, and softening of harvested horticultural crops and is a major determinant of fruit and vegetable quality and shelf-life. A better understanding of how ethylene regulates fruit ripening could result in direct applications to reduce spoilage.

Over the last decades, gene expression changes during fruit ripening have been extensively studied at the transcriptional level, but ripening-associated shifts in the efficiency of translation remain uncharted. Our preliminary data suggest that a subset of transcripts displays ripening-associated changes in their translational efficiencies. To identify transcripts regulated at the translational level during fruit ripening, we are using the Ribo-seq technology which can monitor translation at the whole-genome scale with single-codon resolution.

In the presence of ethylene, the Arabidopsis 3'UTR of the EBFs transcripts have been identified as cis-regulatory elements sufficient for the translational inhibition of gene expression. In this study, the molecular mechanisms responsible for the translation regulation conferred by the 3'UTR of the EBFs in response to ethylene and thus, during fruit ripening will be explored. First, an unbiased phylogenetic approach will be taken to study the conserved features present in the 3'UTRs of the EBFs across the plant kingdom, Solanaceae species, and tomato accessions. Second, using prediction algorithms, the secondary structures of the 3'UTRs will be modelled to look for conserved structures.

The results from this study will enable us to optimize the employment of translation regulatory modules which is a promising tool in plant biotechnology and synthetic biology areas with potential direct applications in our agricultural systems.

Keywords: Ethylene; 3'UTR; EBFs, cis-regulatory elements; gene expression; translational regulation

Author: Andrés Luis Matres Ferrando

Tutor: Javier Brumós Fuentes

Valencia, July 2021

AGRADECIMIENTOS

En primer lugar, me gustaría darle las gracias a Javi, por tus consejos, por tu atención y por las charlas de camino a los invernaderos después de las tutorías. Gracias por acogerme cuando lo necesité y por transmitirme la pasión con la que te dedicas a tu trabajo.

Por su puesto, quiero agradecer a mis amigos, a los que he conocido durante estos años, a los de la carrera y a los de fuera, sin vosotros nada de esto habría sido posible. Gracias por los viajes, por las cenas en el Todo Gigante, por las tardes de cervezas, por los días de biblioteca, por las risas en el piso y por ser mi apoyo cuando lo necesité. Desde el primer día hasta el último habéis estado a mi lado, y sé que aunque nuestros caminos se separen siempre estaréis ahí.

ÍNDICE

1. INTRODUCCIÓN	1
1.1 Etileno	1
1.1.1 Ruta de señalización del etileno.....	2
1.1.2 Importancia del EBF2 en la ruta de regulación del etileno	3
1.1.3 Implicación de la región 3'UTR de EBF2 en la respuesta a etileno	3
1.2 Conservación de secuencia	4
1.2.1 Conservación de la secuencia de regiones no traducidas.....	5
1.3 Modo de regulación de la región 3'UTR.....	5
1.4 Contexto del presente trabajo de final de grado	6
2. OBJETIVOS	7
3. MATERIALES Y MÉTODOS	8
3.1 OBTENCIÓN DE SECUENCIAS.....	8
3.1.1 Secuencias de referencia de EBF2.....	8
3.1.2 Secuencias de aminoácidos de EBF2 de especies del Reino Vegetal	8
3.1.3 Secuencias de aminoácidos de EBF2 en Brásicas y Solanáceas	9
3.1.4 Secuencias de aminoácidos de EBF2 en Monocotiledóneas.....	9
3.1.5 Secuencias de nucleótidos del 3'UTR de EBF2 de todas las especies utilizadas	10
3.2 ALINEAMIENTOS MÚLTIPLES.....	11
3.2.1 Alineamientos de proteínas	12
3.2.3 Alineamientos de 3'UTR.....	12
3.2.3 Visualización de los alineamientos.....	13
3.3 ÁRBOLES FILOGENÉTICOS	13
3.4 EVALUACIÓN ORTOLOGÍA DE LOS GENES USADOS	14
3.4.1 Mantenimiento de la sintenia	15
3.4.2 Patrones de expresión.....	15
3.5 PLEGAMIENTO DE LA REGIÓN 3'UTR DE LOS EBF2	15
3.5.1 RNAstructure fold.....	16
3.5.2 ViennaRNA	16
3.5.3 RNAstructure multiling.....	17
4. RESULTADOS	18
4.1 Estudios filogenéticos de EBF2.....	18
4.2 Alineamientos múltiples.....	24
4.3 Mantenimiento de la sintenia	26

4.4 Patrones de expresión.....	28
4.4.1 Expresión del gen elegido para cada especie.....	28
4.5 Estructuras de los 3'UTR	31
4.5.1 Coincidencia de zonas conservadas con apareamiento en las estructuras	35
5. DISCUSIÓN.....	38
6. REFERENCIAS BIBLIOGRÁFICAS	39

ÍNDICE DE FIGURAS

1.1 Ruta de señalización del etileno

1.2 Comparación por alineamiento de la conservación de la secuencia de un mismo gen entre dos especies

1.3 Interacción entre zona plegada del mRNA y un elemento cis-regulador

2.1 Contenido del script para la generación de las secuencias 3'UTR

2.2 Valores de los parámetros empleados en el alineamiento múltiple de proteínas realizado con el programa MEGA-X

2.3 Valores de los parámetros empleados en el alineamiento múltiple de secuencias de nucleótidos realizado con el programa MEGA-X

2.4 Valores de los parámetros empleados para la generación de los árboles filogenéticos de las regiones 3'UTR

4.1 Distribución de las especies en el Reino Vegetal en función de la secuencia de aminoácidos de la proteína EBF2

4.2 Filogenias distribución de las especies en base a la secuencia de aminoácidos de EBF2

4.3 Filogenias distribución de las especies en base a la secuencia de nucleótidos de la región 3'UTR de EBF2

4.4 Alineamientos múltiples de la secuencia de nucleótidos de la región 3'UTR de EBF2

4.5 Sintenia en la región 5 genes upstream y 5 genes downstream del gen EBF2

4.6 Patrones de expresión relativa en los tejidos de varias especies

4.7 Patrones de expresión de EBF2 en *A. thaliana*

4.8 Patrones de expresión *S. lycopersicum*

4.9 Patrones de expresión *S. tuberosum*

4.10 Patrones de expresión *O. sativa*

4.11 Estructuras de la región 3'UTR de el EBF2 para las especies pertenecientes a la familia Solanáceas

4.12 Estructuras de la región 3'UTR de el EBF2 para las especies pertenecientes a la familia Brásicas

4.13 Estructuras de la región 3'UTR de el EBF2 para las especies pertenecientes al grupo de las Monocotiledóneas

4.14 Relación de regiones conservadas en la secuencia del 3'UTR de ortólogos de Solanáceas con el plegamiento de estas regiones

4.15 Relación de regiones conservadas en la secuencia del 3'UTR de ortólogos de Brásicas con el plegamiento de estas regiones

4.16 Relación de regiones conservadas en la secuencia del 3'UTR de ortólogos de Monocotiledóneas con el plegamiento de estas regiones

1. INTRODUCCIÓN

En el último siglo, la población mundial se ha multiplicado de manera exponencial y se espera que alcance los 9.100 millones en la década de 2050, lo que según las previsiones hará que la demanda de alimentos se incremente un 70%. Sin embargo, se estima que en la actualidad hasta un 30% de toda la comida producida para el consumo humano es desperdiciada en algún punto de la cadena de producción y suministro (Eriksson y Spångberg, 2017; Rezaei y Liu, 2017; *The World Population Prospects*, 2015). La pérdida de comida es un problema con graves implicaciones económicas, sociales y medioambientales.

Actualmente, una de las mayores causas de pérdida de alimentos es su degradación durante la cadena de transporte, haciendo que los alimentos lleguen en mal estado a los puntos de distribución o venta y sean por tanto descartados. Una gran parte de estos alimentos, hasta un 54% (Blanke, 2014), son frutas y verduras. La mayoría de estas pérdidas se deben a la excesiva maduración o senescencia, resultando en el ablandamiento, colapso e infecciones oportunistas de los tejidos. El desarrollo de cultivos menos susceptibles a la sobre-maduración y pérdidas poscosecha es una necesidad urgente. Disminuir, a corto plazo, la pérdida de comida puede convertirse en una contribución clave en el aumento de disponibilidad de alimentos demandados por la creciente población.

La maduración del fruto es un mecanismo empleado por gran cantidad de especies vegetales para maximizar la dispersión de las semillas. La maduración consiste en una serie de pasos estrechamente regulados que tienen como objetivo final producir un fruto atractivo para los organismos encargados de dispersar las semillas. Dentro de los frutos, podemos distinguir frutos climatéricos y no climatéricos. En el caso de los primeros, la producción de etileno por el propio fruto es fundamental para su maduración (Giovannoni, 2001). Esta hormona además juega un papel fundamental en una gran variedad de procesos del desarrollo de la planta y en respuesta a estímulos ambientales críticos para la supervivencia de la planta (Merchante et al., 2015).

1.1 Etileno

El etileno es una molécula sencilla, siendo el más simple de los alquenos, pero a su vez se trata de una hormona gaseosa producida por las plantas que regula gran cantidad de procesos del crecimiento y desarrollo de los organismos vegetales (Alonso y Stepanova, 2004). Además, la simplicidad y propiedades químicas de la molécula de etileno le permiten difundir a través de membranas y citoplasma, haciendo innecesaria la presencia de un sistema de transporte al interior celular (Müller y Munné-Bosch, 2015).

El etileno es una hormona implicada en un amplio rango de procesos durante el desarrollo y crecimiento de las plantas. También es responsable de la respuesta a condiciones de estrés. De especial interés es su implicación en la regulación de la maduración en los frutos climatéricos. En los cuales, el característico pico “climatérico” de respiración viene acompañado de un claro

incremento en la producción y respuesta a etileno del fruto (Bleecker y Kende, 2000; Cao et al., 2008).

1.1.1 Ruta de señalización del etileno

La ruta de señalización de etileno muestra un alto nivel de conservación en las especies terrestres del Reino Vegetal. Los receptores de etileno son una serie de proteínas asociadas a la membrana del retículo endoplásmico que, en ausencia de etileno, reprimen de forma activa la ruta de señalización (Müller y Munné-Bosch, 2015).

En *Arabidopsis* el etileno es sentido por hasta 5 receptores asociados a la membrana del retículo endoplasmático (RE) (Chang y Stadler, 2001). Aguas abajo de estos receptores, se encuentra el resto de componentes de la ruta de señalización como Constitutive Triple Response 1 (CTR1), Ethylene Insensitive 2 (EIN2), EIN3/EIN3 Like1 (EIL1) los reguladores clave de los Ethylene Response Factors (ERFs) (Potuschak et al., 2003).

CTR1 es una proteína quinasa que, en ausencia de etileno, fosforila la región C-terminal de EIN2 y de este modo previene su escisión y translocación al núcleo. La permanencia de la región C-terminal de EIN2 en el RE impide cualquier señalización a los subsiguientes componentes de la ruta de señalización, EIN3/EIL1. De todos modos, en ausencia de etileno, los niveles de EIN3/EIL1 son muy bajos, debido a la actividad de las proteínas F-box EIN3 Binding Factor (EBF1 y EBF2) que marcan EIN3/EIL1 para su degradación por el proteasoma 26S (Merchante et al., 2015).

En presencia de etileno, CTR1 se inactiva, provocando la reducción de los niveles de fosforilación de la región C-terminal de EIN2, lo cual desencadena su escisión y translocación desde el RE al núcleo donde participa en la estabilización de EIN3/EIL1. La estabilización de EIN3/EIL1 es en gran parte debida a la actividad de la región C-terminal de EIN2 reprimiendo la traducción de los RNA mensajeros (mRNAs) de EBF1/2 en presencia de etileno. Sin la producción de las proteínas EBF1/2, EIN3/EIL1 no son marcadas para su degradación y pueden regular la respuesta transcripcional a etileno (Potuschak et al., 2003).

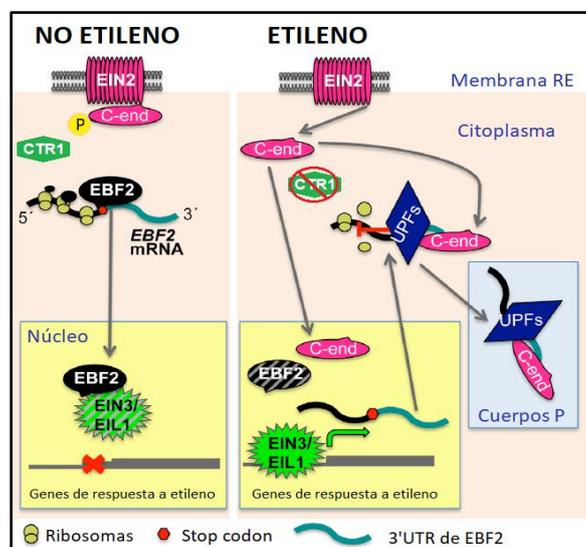


Figura 1.1 Ruta de señalización del etileno. Adaptación de (Merchante et al., 2015): *Implicación de los componentes de la ruta del etileno en condiciones de ausencia y presencia de la hormona.*

1.1.2 Importancia del EBF2 en la ruta de regulación del etileno

EBF1 y EBF2 son dos proteínas F-Box pertenecientes a la familia de proteínas de repetición rica en leucina (LRR), localizadas en el núcleo y expresadas de forma constitutiva en varios órganos de la planta. Ambas proteínas EBF1/2 tienen como diana principal a EIN3 y EIL1, a las cuales, en ausencia de etileno, marcan para su ubiquitinación y posterior degradación. Sin embargo, la proteólisis de dichos componentes se inhibe en presencia de etileno, lo que permite la acumulación en el núcleo de EIN3/EIL1, responsables de la regulación de todas las respuestas transcripcionales mediadas por etileno.

EIN3 y EIL1 son moléculas clave en la ruta de señalización del etileno. Su ubiquitinación está mediada por la ubiquitin ligasa SCF, la cual es un complejo formado por las subunidades SKP1, Cullin 1 y RBX1, y que además está asociado a una de las dos proteínas F-Box EBF1 o EBF2. A pesar de que ambas proteínas tienen funciones similares, se ha visto que solo la expresión transcripcional de EBF2 es activada por etileno. EBF2 además se une a EIN3 usando una región conservada C-terminal de repeticiones ricas en leucina.

EIN3 puede unirse directamente al promotor de EBF2, activando su transcripción, lo que confirma la existencia de un “feedback loop” regulatorio entre ambos componentes de la ruta, necesario para el correcto funcionamiento de la señalización por etileno. En este feedback, EBF2 es un regulador negativo de EIN3/EIL1 pero el etileno y, por tanto, EIN3 activan la transcripción de EBF2, provocando la producción de mRNAs de esta proteína F-Box. Completando este feedback loop regulatorio, existe una región aguas abajo de la región codificante del mRNA de EBF2, la región 3'UTR del mRNA, responsable de la regulación a nivel traduccional de los niveles de expresión de la proteína EBF2. La región C-terminal de EIN2 interacciona con el 3'UTR de EBF2 e inhibe su traducción, provocando un descenso en los niveles de proteína de EBF1/2 en presencia de etileno (Binder et al., 2007; Konishi y Yanagisawa, 2008; Potuschak et al., 2003).

1.1.3 Implicación de la región 3'UTR de EBF2 en la respuesta a etileno

El etileno afecta los niveles de traducción de mRNAs pertenecientes a varios genes, entre los cuales se encuentran EBF1 y EBF2, cuya regulación traduccional depende de sus regiones 3'UTR. En presencia de etileno, los niveles de transcripción del EBF2 aumentan. Sin embargo, sus niveles de traducción se ven reprimidos. Esta regulación de la traducción se da como consecuencia de la unión de la región C-terminal de EIN2 a la 3'UTR de los mRNAs de EBF1 y EBF2, lo cual estimula el reclutamiento de las proteínas Regulator of nonsense transcripts o proteínas UPF que inhiben la traducción de estos mRNAs. De este modo, en presencia de etileno, la eficiencia en la traducción de estos mRNAs se reduce al 40% de la eficiencia mostrada en condiciones control. Las proteínas UPF, junto a la región C-terminal de EIN2, forman un complejo ribonucleoproteico con los mRNAs que se localiza en los Processing-bodies (Cuerpos P), donde quedan almacenados de forma temporal hasta que el etileno deja de estar presente en la célula. Una vez el etileno desaparece, los complejos ribonucleoproteicos se descomponen y los mRNAs son finalmente traducidos, produciendo proteínas EBF1/2 que se translocan al núcleo y marcan para degradación a EIN3/EIL1 causando el cese de la respuesta a etileno (Figura 1.1) (Weber et al., 2008).

1.2 Conservación de secuencia

La evolución de una secuencia está mediada por la relación entre la función de esa secuencia y la presión y eficacia de la selección natural. En general, los nucleótidos con una función molecular importante evolucionan a un ritmo menor. En las zonas de DNA codificante, la evolución de las secuencias depende de varios factores, como por ejemplo, el nivel de expresión de la proteína, la categoría funcional, características estructurales y participación en interacciones moleculares (Kim et al., 2006). Además, dentro de las secuencias codificantes, existen regiones específicas, clave para la estructura final de la proteína, como pueden ser los sitios activos de enzimas o ciertos residuos clave para el mantenimiento estructural de la proteína, y cuyos ratios de evolución son mucho menores que en el resto de la secuencia (Cooper y Brown, 2008).

Las zonas de DNA no codificantes también están sometidas a cambios evolutivos mediados por mutaciones y selección de las mismas. Dentro de estas regiones, en las secuencias reguladoras, existen puntos clave como los sitios de unión de factores de transcripción, que contribuyen a la regulación de la expresión de las zonas codificantes de los genes. Estas regiones reguladoras tienden a acumular menos sustituciones, ya que debido a su importancia funcional, la mayoría de cambios en la secuencia tiene efectos deletéreos (Cooper y Brown, 2008). Por tanto, la evolución de ciertos nucleótidos frente a otros en una determinada secuencia, no está solo relacionada con el aminoácido codificado por esa secuencia, sino que en regiones no codificantes también influyen los factores reguladores de los niveles de expresión. Asimismo, la ausencia de conservación en regiones no codificantes no siempre implica ausencia de función (Chamary y Hurst, 2005; Johnsson et al., 2014).

Especie1	AAACCCATAGATTTACGTCACAACCTCAGTTTCTT---TCCCAGAACGCAGCTGCTGT---
Especie2	TGGCTCGTG----TCATAGGCAGTGGGCTTTCTTCTGTACAGGACTAGGTTGTTTTCCA
	* * * *** ** ***** ** * * * * * * * * * *
Especie1	-----TTTTTTC-----AATTTTTAGCAGG--TTCCTTCTAAACAAAGG-AGCCTGTA
Especie2	GGGACCTTTGGCCAGTATTGATTTTTTGCAGGGTCTCCTCTTTTACAGGGGAAGCCTGTT
	*** ** ***** ***** ***** * * * * * * *****
Especie1	TCTACAGCTGTT-----GTGTTTATGGGAATCCTGTGAGTGAGTGCTGGTAA
Especie2	TC-AGAGGTGTTTGGCTGCCACTGATTGGCTACAGCTTTCCTCTAGCCTGGTCTCGGTTC
	** * * * * * * * * * * * * * * * * * * * * * * * *
Especie1	TCTAG-TTAGGGTTTGGTCTTGACCGTCTTTTGGAG-TTGTTGTCTCTATGCTTTCG
Especie2	TTGAGATTACAACCTTGCCTAGCCCTGGTTTATGCTTGGAACTCTGCTTGTTTTTG
	* * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
Especie1	TCCTGTTCCGTCAAGTTCTTGT-GGGAGTCCTTGTCTCTGGT-TTAGGTTTGTGAACGT
Especie2	CCATTTCTAGGTGCCCTTTTATAGAGGGTCTGGTTATGTATCGCTGGGGCAGTG-----
	* * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Figura 1.2 Comparación por alineamiento de la conservación de la secuencia de un mismo gen entre dos especies (alineamiento realizado con la herramienta MUSCLE en la web de EMBL-EBI (Edgar, 2004)): Nucleótidos conservados (*)

1.2.1 Conservación de la secuencia de regiones no traducidas

En organismos eucariotas, el número de transcritos de un gen presentes en una célula, no siempre está directamente relacionado con los niveles de la proteína que codifican. Esto se debe a la presencia de mecanismos de regulación postranscripcionales. En muchas ocasiones, los elementos encargados de regular la expresión postranscripcional se encuentran en forma de secuencias no traducidas en las regiones 3' o 5' (3' y 5' UTRs) del mRNA. El estudio de estos elementos, identificación y caracterización, es clave para obtener una visión global de los distintos mecanismos de regulación de expresión, desde la transcripción del DNA en RNA hasta la traducción de esos RNAs en proteínas (Calvo et al., 2009).

Existen regiones aguas arriba del codón de inicio del mRNA, en el 5' UTR, que poseen upstream open reading frames o uORFs involucradas en la regulación de la traducción. Estas regiones son escaneadas por los ribosomas en la búsqueda del codón de inicio, los ribosomas pueden invertir demasiado tiempo en estas regiones del 5' UTR e incluso traducir estos pequeños uORFs sin llegar a traducir el ORF principal, reduciendo radicalmente la traducción del ORF principal de determinados transcritos (Vaughn et al., 2012).

La estructura del mRNA afecta su estabilidad y su eficiencia de traducción. La localización del mRNA también influye en su capacidad de traducción. Las regiones 3'UTR de los transcritos a menudo, contienen varios elementos reguladores que controlan la expresión espacio-temporal de los mRNAs (Kuersten y Goodwin, 2003). La conservación de secuencias identificada en estas regiones 3'UTR puede estar relacionada con la estructura secundaria que adoptan estas regiones al plegarse en busca de la conformación más estable. Esta estructura podría funcionar como un interruptor, cambiando la conformación al unirse a un metabolito o proteína, alterando así su splicing y/o la estabilidad del transcrito (Burgess y Freeling, 2014; Freeling y Subramaniam, 2009).

1.3 Modo de regulación de la región 3'UTR

La regulación postraduccional juega un papel fundamental en muchos estadios del desarrollo y afecta a un amplio espectro de mRNAs. A pesar de que varias características del mRNA pueden contribuir al control de su traducción, en la mayoría de ocasiones, la traducción de los transcritos está regulada por secuencias UTRs en el 5' y/o 3' fuera de las regiones traducidas. Además, la región situada en el extremo 3' presenta una gran implicación en dicha regulación (Moor et al., 2005; Wilkie et al., 2003). Estas regiones contienen elementos cis-reguladores que controlan la localización y niveles de traducción con los cuales proteínas u otros RNAs pueden interactuar (Figura 1.3). La cola poli-A en el extremo 3' del mRNA es también un importante determinante de la eficiencia de la traducción (Preiss y Hentze, 1999; Wilkie et al., 2003). En muchos organismos, interacciones mediadas por proteínas entre la región terminal del transcrito resultan en la formación de un bucle, el cual afecta la regulación de la traducción (Mazumder et al., 2003).

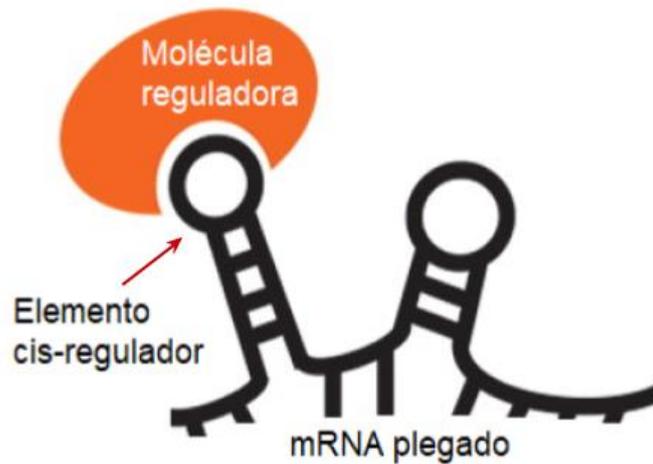


Figura 1.3 Interacción entre zona plegada del mRNA y un elemento cis-regulador. Adaptación de (Hung and Chang, 2010).

1.4 Contexto del presente trabajo de final de grado

Este trabajo de final de grado queda enmarcado dentro de uno de los objetivos de un proyecto centrado en encontrar mecanismos de regulación de la traducción durante la maduración del fruto de tomate. Este proyecto se basa en conocimientos previamente adquiridos en *Arabidopsis* aplicando, entre otras técnicas, el Ribosome footprinting o Ribo-seq (Merchante et al., 2015). La región 3'UTR del transcrito de EBF2 de *Arabidopsis* es necesaria y suficiente para mediar la activación de la ruta de señalización de etileno. El 3'UTR de EBF2 se asocia con el extremo C-terminal de la proteína EIN2, resultando en una inhibición de la traducción de los transcritos de EBF2 y, por tanto, provocando la estabilización de las proteínas EIN3/EIL1 y la consiguiente activación de la respuesta transcripcional a etileno. En *Arabidopsis*, la región 3'UTR del EBF2 está formada por más de 600 nucleótidos, superando en 450 a la media del resto de 3'UTRs de la especie (Kawaguchi y Bailey-Serres, 2005). Esta característica también se repite en otras especies como en tomate, con un 3'UTR de 700 nucleótidos frente a los 257 nucleótidos de media de la especie (Aoki et al., 2010), o en arroz, donde el 3'UTR tiene una longitud de casi 800 nucleótidos en lugar de los 289 de media (Shen et al., 2008). Dicha característica, la atípica enorme longitud de los 3'UTR de distintos EBFs, parece indicar una conservación del mecanismo de regulación descrito en *Arabidopsis*.

En este trabajo de final de grado se ha llevado a cabo la caracterización "in silico" de los distintos 3'UTR de los ortólogos del EBF2 de *Arabidopsis*. Este trabajo contribuye al objetivo global de mejorar nuestro conocimiento acerca de este mecanismo de regulación de la traducción. Para ello, se debe confirmar la presencia del gen EBF2 en las especies del Reino Vegetal e identificar la extensa región 3'UTR de los ortólogos seleccionados. La búsqueda de elementos responsable de la regulación a nivel traduccional en respuesta a etileno se ha llevado a cabo tanto a nivel de secuencia, como a nivel de estructura de esta región 3' del transcrito.

2. OBJETIVOS

1. Identificar de forma inequívoca y simple los ortólogos de EBF2 de *Arabidopsis* y sus 3'UTRs en otras especies del Reino Vegetal a través de distintas aproximaciones.

- Por similitud de secuencia utilizando BLAST.
- Por estudios filogenéticos globales en el Reino Vegetal y específicos en subgrupos que contienen especies más cercanas evolutivamente.
- Por análisis de sintenia confirmando la conservación del orden génico a lo largo de la región cromosómica evaluada.
- Por analogía de los patrones de expresión de los ortólogos en ciertos estadios del desarrollo y bajo determinadas condiciones.

2. Intentar definir elementos reguladores en los 3'UTR responsables de la inhibición de la traducción del transcrito de EBF2 en respuesta a etileno.

- Búsqueda de motivos o regiones de secuencias conservados en el 3'UTR de los ortólogos de EBF2.
- Generación y análisis de los modelos de la estructura secundaria de los 3'UTR de los ortólogos de EBF2.

3. MATERIALES Y MÉTODOS

3.1 OBTENCIÓN DE SECUENCIAS

3.1.1 Secuencias de referencia de EBF2

Como secuencia de referencia para la búsqueda de genes ortólogos de EBF2 en otras especies se empleó la secuencia de aminoácidos de dicho gen de *Arabidopsis thaliana*, la cual se obtuvo de la web "TAIR" (TAIR, s. f.) anotada como AT5G25350.

Para posteriores análisis de BLAST contra la familia de las Solanáceas, se empleó como secuencia de referencia la secuencia de aminoácidos de EBF2 de *Solanum lycopersicum*. En este caso, la secuencia fue obtenida realizando un BLAST de proteínas (BLASTp) con la secuencia del gen de *Arabidopsis thaliana* en la base de datos de "Non-redundant protein sequences" del National Center for Biotechnology Information (NCBI), limitando la búsqueda únicamente a secuencias de *S. lycopersicum*. Tras la búsqueda se seleccionó la secuencia con mayor score, anotada como EIN3-binding F-box protein 2 (EBF2). La misma secuencia es obtenida al realizar un BLAST en la plataforma de Solgenomics (Fernandez-Pozo et al., 2015). A modo de confirmación, se llevó a cabo el BLASTp inverso, con la secuencia de *Solanum lycopersicum* se hizo el BLAST contra las secuencias de *A. thaliana* en NCBI obteniendo AT5G25350 como la secuencia con el score más alto.

3.1.2 Secuencias de aminoácidos de EBF2 de especies del Reino Vegetal

Para la búsqueda de las secuencias de aminoácidos, se emplearon tres plataformas distintas: NCBI, PLAZA (Van Bel et al., 2018; Vandepoele et al., 2013), y Solgenomics. Las tres bases de datos ofrecen la posibilidad de utilizar la herramienta BLAST para realizar búsquedas empleando las secuencias de referencia.

En primer lugar, se realizó una búsqueda de secuencias ortólogas al gen de EBF2 de *Arabidopsis thaliana* utilizando el NCBI, realizando un BLASTp con dicha secuencia frente a la base de datos "Non-redundant protein sequences" y limitando la búsqueda a organismos del taxón Green Plants. Se emplearon los valores predeterminados de los parámetros para realizar la búsqueda con el BLASTp, con un Expected threshold de 0.05, el cual indica que sólo se mostrarán aquellos hits o alineamientos con un E-value (probabilidad de que un determinado hit aparezca por casualidad, para un score determinado) menor de 0.05; un valor de Window-size de 6, el cual indica el tamaño de palabra (número de aminoácidos) empleado en la búsqueda; y un máximo coincidencias en un query range de 0. Los parámetros de puntuación empleados fueron el uso de la matriz de puntuación BLOSUM62 y un coste de "gap" de 11. La búsqueda se amplió también a especies de dicotiledóneas anotadas en la web PLAZA, en la versión Dicots PLAZA 4.5, empleando los valores predeterminados de los parámetros, con un E-value de $1e-5$ y la matriz de puntuación BLOSUM62.

Los requerimientos para seleccionar las secuencias fueron:

- Score alto al hacer BLAST con las secuencias de referencia y un bajo E-value.

- Región 3'UTR anotada y disponible para hacer el posterior análisis.
- Anotación adecuada como EBF2.

Se descartaron secuencias anotadas como “predicted”, “low quality protein”, “unnamed protein product” o “hypothetical protein”. También fueron descartados aquellos hits bien anotados como EBF2, pero cuya región 3'UTR no se encontraba disponible, pues no serían útiles para posteriores análisis. Aquellas secuencias con un alto score y anotadas como EBF1 fueron examinadas para comprobar si realmente se trataba de ortólogos de EBF1 y si dichas especies poseían más parálogos (como EBF2). O si en caso contrario, esas especies solamente poseían una copia de los EBF en su genoma. Para aquellas especies que poseen más de un EBF, las secuencias de EBF1 se eliminaron de nuestros análisis y solo se mantuvieron las secuencias de los confirmados EBF2s para análisis más detallados.

3.1.3 Secuencias de aminoácidos de EBF2 en Brásicas y Solanáceas

Con la intención de acotar el análisis a especies más cercanas evolutivamente y que pudieran mostrar unos niveles de conservación más elevados en sus secuencias 3' UTR no codificantes, se decidió realizar una búsqueda de secuencias similares en las especies relacionadas y cercanas evolutivamente a las especies de mayor interés en nuestro proyecto, *Arabidopsis thaliana* y *Solanum lycopersicum*, identificando secuencias en la familia *Brassicaceae*, a la que pertenece *A. thaliana*; y en la familia *Solanaceae* a la que pertenece *S. lycopersicum*.

Para la búsqueda de secuencias homólogas a EBF2 en Brásicas y Solanáceas, se empleó la herramienta BLASTp del NCBI con las secuencias de referencia de EBF2 de *S. lycopersicum* y *A. thaliana*, de nuevo contra la base de datos “Non-redundant protein sequences” y limitando la búsqueda específicamente a especies de la familia *Brassicaceae* o *Solanaceae*. Se emplearon los valores predeterminados de los parámetros generales y de puntuación comentados anteriormente.

La búsqueda de Solanáceas se amplió con las bases de datos de PLAZA y Solgenomics. Con PLAZA se realizó nuevamente un BLASTp con los parámetros predeterminados, empleando la secuencia de referencia de *S. lycopersicum*. En Solgenomics se realizó también una búsqueda por BLASTp con la secuencia de referencia del EBF2, empleando los valores predeterminados de los parámetros, con un valor de e-value de $1e-10$, un Word-size de 3 y empleando la matriz de sustitución BLOSUM62. La búsqueda en este caso, se realizó de forma específica para cada especie anotada en Solgenomics y se empleó la base de datos más actualizada para cada una de ellas. Los hits con mayor score, anotados correctamente como EBF2 y con la región 3'UTR disponible fueron seleccionados para las especies de cada una de las familias.

3.1.4 Secuencias de aminoácidos de EBF2 en Monocotiledóneas

Las Monocotiledóneas son un grupo de plantas angiospermas o plantas con flor cuyo embrión posee únicamente un cotiledón en lugar de dos, como es el caso de las dicotiledóneas. Existen gran cantidad de cultivos de interés agroeconómico pertenecientes a este grupo vegetal; es el caso de cereales como el maíz (*Zea mays*), el mijo (*Setaria italica*) o el arroz (*Oryza sativa*) y cultivos frutales como la banana (*Musa acuminata*) (Acevedo et al., 2006; Cheadle, 1942; D'Hont et al., 2012; Hochholdinger et al., 2008; Li y Brutnell, 2011). La ruta de señalización de etileno

está altamente conservada en este grupo de plantas de elevado interés agronómico. Se decidió añadir un grupo de especies pertenecientes a esta clase de plantas para completar los análisis y para también tratar de identificar secuencias o características de sus 3' UTR que puedan mediar la regulación de la traducción de estos transcritos en respuesta a etileno.

Al llevar a cabo las búsquedas utilizando los BLASTp, como se ha descrito con anterioridad en las diferentes bases de datos, no se encontraron secuencias similares al EBF2 de *A. thaliana* o *S. lycopersicum* de Monocotiledóneas presentes en las primeras páginas de resultados de secuencias que producen un BLAST con alineamientos significativos, donde preferencialmente se encontraban las secuencias de especies dicotiledóneas, probablemente por encontrarse evolutivamente separadas del resto de especies utilizadas en el análisis. Es por ello que para la obtención de secuencias de las especies de Monocotiledóneas es necesario acotar mejor su búsqueda.

Se emplearon, de nuevo, las bases de datos de NCBI y de PLAZA en su versión "Monocots PLAZA 4.5" para la búsqueda de secuencias. En ambas plataformas, se usó la herramienta BLASTp y se decidió utilizar la secuencia de referencia de *A. thaliana* para realizar la búsqueda. Se utilizaron los valores predeterminados de los parámetros y la matriz de sustitución BLOSUM62 en ambas búsquedas. Se seleccionaron las secuencias con mayor score, anotadas como EBF2 y con la secuencia de la región 3'UTR disponible, entre las que encontramos *Z. mays*, *O. sativa* o *S. italica*.

3.1.5 Secuencias de nucleótidos del 3'UTR de EBF2 de todas las especies utilizadas

Una vez obtenidas las secuencias de aminoácidos de los genes ortólogos al EBF2 de *Arabidopsis* en varias especies del Reino Vegetal, Brásicas, Solanáceas y Monocotiledóneas, se procedió a obtener la secuencia de la región 3'UTR de dichas especies para poder llevar a cabo su posterior análisis. Como se ha comentado anteriormente, sólo fueron seleccionadas las secuencias de aminoácidos de EBF2 de las que la región 3'UTR se encontraba también disponible en la base de datos.

Para la obtención de estas secuencias en la web del NCBI, se empleó la herramienta tBLASTn (BLAST de proteína a nucleótido traducido), contra la base de datos de "Nucleotide collection", utilizando la secuencia de referencia correspondiente: *A. thaliana* para Reino Vegetal, Brásicas y Monocotiledóneas; y *S. lycopersicum* para Solanáceas. Se utilizaron los valores predeterminados de los parámetros, y se limitó en cada caso la búsqueda, realizando el BLAST contra los taxones Green Plants, *Solanaceae*, *Brassicaceae* y *Monocotyledoneae*. Para cada búsqueda, se escogieron las secuencias de nucleótidos correspondientes al gen EBF2 de aquellas especies cuya secuencia de aminoácidos había sido previamente seleccionadas, y se generó un archivo en formato GenBank con la información de todas ellas.

Los archivos GenBank, son archivos con la información almacenada de cada registro del NCBI, y contienen características anotadas de las secuencias como el locus o el gen al que pertenecen (Sayers et al., 2019). La información de interés que se desea obtener de estos archivos es la secuencia de nucleótidos que forma el mRNA del gen EBF2 de cada especie seleccionada y las posiciones inicial y final de la CDS, a partir de las cuales se pueden encontrar las regiones no traducidas o UTR. Para trabajar con estos datos y obtener la región 3'UTR, se generó un script

en lenguaje python, en el que se empleó la herramienta SeqIO de la librería BioPython (Cock et al., 2009), la cual permite leer y parsear diferentes formatos de secuencia, entre los que se encuentra el formato GenBank. Con este script, se generaron varios archivos “.fasta” en los que se almacenó la secuencia 3’UTR junto con el nombre de la especie a la que pertenece dicha secuencia (Figura 2.1).

```
1 # 3'UTR generator
2 from Bio import SeqIO
3
4 # CHANGE THE NAME OF THE INPUT FILE HERE:
5 input_file = 'input_file.gb'
6
7 outfile = open('utr_file.fasta', 'w')
8 for record in SeqIO.parse(input_file, 'gb'):
9     for feature in record.features:
10        if feature.type == 'CDS':
11            cds = feature.location
12            start_utr = int(cds.end)
13
14            organism = record.annotations['organism']
15            _3utr = str(record.seq[start_utr:])
16
17        outfile.write('>' + organism + ' 3UTR\n' + _3utr + '\n')
18 outfile.close()
```

Figura 2.1 Contenido del script para la generación de las secuencias 3’UTR Se importa el módulo SeqIO de la librería BioPython (2). Se parsea el archivo GenBank (8). Se obtienen las características de la secuencia (9). Se obtiene la posición de inicio del 3’UTR (12). Se obtiene el nombre del organismo (14). Se obtiene la secuencia del 3’UTR (15).

Para aquellas secuencias obtenidas en la base de datos de la web PLAZA, se realizó también una búsqueda empleando la herramienta tBLASTn de dicha página web, contra la base de datos “PLAZA transcript sequences”. Se utilizó para la búsqueda los valores predeterminados de los parámetros, con un Expected threshold de $1e-5$ y la matriz BLOSUM62, y la secuencia de referencia correspondiente. Para las especies cuya secuencia de aminoácidos de EBF2 había sido seleccionada en esta página web, se obtuvo también la secuencia de nucleótidos del 3’UTR.

3.2 ALINEAMIENTOS MÚLTIPLES

Una vez obtenidas las secuencias, analizamos la similitud entre ellas. Una forma de comprobar la conservación en las diferentes especies es mediante un alineamiento múltiple de sus secuencias. Para llevar a cabo dicho análisis, se empleó el programa bioinformático MEGA-X: Molecular Evolutionary Genetics Analysis a través de plataformas informáticas, el cual además incluye múltiples métodos analíticos y herramientas para llevar a cabo análisis filogenéticos (Kumar et al., 2018).

Se realizaron los alineamientos múltiples tanto de las secuencias de aminoácidos de EBF2 como de la región 3’UTR de los transcritos del gen, para las especies agrupadas en: Reino Vegetal, Brásicas, Solanáceas y Monocotiledóneas. Para llevar a cabo dichos alineamientos, se agruparon las secuencias en archivos “.fasta”, los cuales se utilizaron como input del programa; siguiendo los siguientes pasos: Align, Edit/Build Alignment/Retrieve a sequence from a file.

3.2.1 Alineamientos de proteínas

Una vez seleccionado el conjunto de secuencias de aminoácidos para llevar a cabo el alineamiento, éste se realizó usando el algoritmo de alineamiento MUSCLE (Edgar, 2004b), el cual usa dos medidas de distancia para una pareja de secuencias:

- Distancia por k-mers, la cual se mide en parejas de secuencias no alineadas. Los k-mers son subsecuencias de longitud k. Al compararse los k-mers de dos secuencias relacionadas, estos tienden a alinearse un mayor número de veces que las que se espera que se den por casualidad en dos secuencias no relacionadas. (Edgar, 2004a)
- Distancia Kimura, la cual se mide entre parejas de secuencias ya alineadas, calculando la identidad de ambas secuencias y convirtiéndola en una estimación de la distancia. (Kimura, 1989)

Para los alineamientos se emplearon los valores predeterminados de los parámetros de penalización por gap (Figura 2.2), los cuales penalizan la necesidad de abrir huecos en la secuencia y/o extender su longitud a consecuencia del alineamiento y multiplican dicha penalización si estos huecos aparecen en regiones hidrofóbicas, donde las inserciones y deleciones son menos frecuentes (Edgar, 2004c).

Como método de clustering, se escogió el método Neighbor Joining (Figura 2.2), ya que se pensó que era el más adecuado para posteriormente construir árboles filogenéticos con los alineamientos, debido a su menor tendencia a generar árboles con ataduras que el método UPGMA (Backeljau et al., 1996; (Saitou y Nei, 1987; Takezaki, 1998).

Option	Setting
GAP PENALTIES	
Gap Open	-2.90
Gap Extend	0.00
Hydrophobicity Multiplier	1.20
MEMORY/ITERATIONS	
Max Memory in MB	2048
Max Iterations	16
ADVANCED OPTIONS	
Cluster Method (Iterations 1,2)	<i>Neighbor Joining</i>
Cluster Method (Other Iterations)	<i>Neighbor Joining</i>
Min Diag Length (Lambda)	24

Figura 2.2 Valores de los parámetros empleados en el alineamiento múltiple de proteínas realizado con el programa MEGA-X Gap Open Penalty, penalización por abrir un gap en el alineamiento; Gap Extend Penalty, penalización por extender un espacio en un residuo; Hydrophobicity Multiplier, multiplicador para brecha abierta o cerrada en regiones hidrofóbicas; Min Diag Length, mínima longitud de la diagonal

3.2.3 Alineamientos de 3'UTR

El alineamiento de las secuencias de nucleótidos de las regiones 3'UTR se llevó a cabo de manera similar al alineamiento de proteínas, empleándose también el algoritmo de alineamiento MUSCLE. Se usaron los valores predeterminados de los parámetros. De nuevo, el método de clustering escogido fue el Neighbor Joining (Figura 2.3).

Option	Setting
GAP PENALTIES	
Gap Open	-400.00
Gap Extend	0.00
MEMORY/ITERATIONS	
Max Memory in MB	2048
Max Iterations	16
ADVANCED OPTIONS	
Cluster Method (Iterations 1,2)	<i>Neighbor Joining</i>
Cluster Method (Other Iterations)	<i>Neighbor Joining</i>
Min Diag Length (Lambda)	24

Figura 2.3 Valores de los parámetros empleados en el alineamiento múltiple de secuencias de nucleótidos realizado con el programa MEGA-X. Gap Open Penalty, penalización por abrir un gap en el alineamiento; Gap Extend Penalty, penalización por extender un espacio en un residuo; Min Diag Length, mínima longitud de la diagonal

3.2.3 Visualización de los alineamientos

Todos los alineamientos fueron almacenados en archivos “.MAS” los cuales posteriormente serán utilizados para su representación de forma gráfica y para la construcción de árboles filogenéticos.

Para la visualización gráfica de los alineamientos, se empleó la página web NCBI Multiple Sequence Alignment Viewer (NCBI, s. f.). Fue necesario pasar a formato fasta los archivos de alineamiento para su correcta visualización. Con esta herramienta se pudo diferenciar tanto aminoácidos como nucleótidos por colores, permitiendo una mejor visualización de las zonas conservadas.

3.3 ÁRBOLES FILOGENÉTICOS

Una vez generados los archivos de alineamiento, éstos fueron empleados para generar los árboles filogenéticos con los que poder observar la evolución tanto de la secuencia de la proteína de EBF2 como de la región 3’UTR de los mRNAs. Se generaron árboles filogenéticos tanto para las secuencias de aminoácidos como para las del 3’ UTR, en una figura general donde se incluyen las especies del Reino Vegetal, y otras figuras más específicas centradas en los grupos de Brásicas, Solanáceas y Monocotiledóneas. Se empleó de nuevo el programa bioinformático MEGA-X, el cual tomó como input los archivos de alineamiento generados anteriormente.

Para la generación de los árboles, se decidió emplear el método estadístico de Maximum likelihood, el cual permite encontrar el árbol que mejor refleja la evolución más probable de las especies (Felsenstein, 1981; Price et al., 2010). Como test de filogenia se empleó el método bootstrap, el cual permite evaluar la confianza de un árbol en base a la cantidad de árboles que muestran el mismo clado (Efron et al., 1996), y se decidió emplear un número de réplicas bootstrap de 1000 para dar una mayor robustez a los resultados, a pesar de que a partir de las 500 réplicas, la precisión del árbol final es bastante elevada (Pattengale et al., 2009). Para el resto de parámetros se emplearon los valores predefinidos, tanto para árboles de aminoácidos

como de nucleótidos. Para la estimación del número de sustituciones de nucleótidos se empleó el modelo de sustitución Tamura-Nei (Tamura, 1994), mientras que en los árboles filogenéticos de aminoácidos se empleó el modelo de sustitución Jones-Taylor-Thornton. Se consideró una ratio uniforme de evolución y se emplearon todos los sitios, tanto nucleótidos como gaps, para construir los árboles.

Por último, el árbol inicial se realizó con la opción predeterminada de Neighbor-Joining y como método heurístico para refinar la topología del árbol, se empleó el método de los intercambios de vecinos más cercanos (NNI) (Figura 2.4).

Option	Setting
ANALYSIS	
Statistical Method	→ <i>Maximum Likelihood</i>
PHYLOGENY TEST	
Test of Phylogeny	→ <i>Bootstrap method</i>
No. of Bootstrap Replications	→ <i>1000</i>
SUBSTITUTION MODEL	
Substitutions Type	→ <i>Nucleotide</i>
Model/Method	→ <i>Tamura-Nei model</i>
RATES AND PATTERNS	
Rates among Sites	→ <i>Uniform Rates</i>
No of Discrete Gamma Categories	→ <i>Not Applicable</i>
DATA SUBSET TO USE	
Gaps/Missing Data Treatment	→ <i>Use all sites</i>
Site Coverage Cutoff (%)	→ <i>Not Applicable</i>
TREE INFERENCE OPTIONS	
ML Heuristic Method	→ <i>Nearest-Neighbor-Interchange (NNI)</i>
Initial Tree for ML	→ <i>Make initial tree automatically (Default - NJ/BioNJ)</i>
Initial Tree File	→ <i>Not Applicable</i>
Branch Swap Filter	→ <i>None</i>

Figura 2.4 Valores de los parámetros empleados para la generación de los árboles filogenéticos de las regiones 3'UTR. Método estadístico, Maximum likelihood; Test de filogenia Bootstrap con 1000 réplicas; Modelo de sustitución, modelo Tamura-Nei; Ratios de variación uniformes; Uso de todas las posiciones en la construcción del árbol; Método heurístico NNI; Generación del árbol inicial por el método de Neighbor-Joining

3.4 EVALUACIÓN ORTOLOGÍA DE LOS GENES USADOS

Con el fin de confirmar que los genes utilizados en nuestros análisis eran realmente los ortólogos de EBF2 y con ello dar validez a los estudios filogenéticos, se emplearon herramientas que permiten comprobar el mantenimiento de la sintenia u orden de los genes dentro del cromosoma , junto con otras que permiten el estudio de los patrones de expresión del gen de interés en cada una de las especies estudiadas y de este modo comprobar que en todas ellas la función de los genes de interés se mantiene.

Este tipo de comprobación proporciona la confirmación del correcto uso de los genes seleccionados como ortólogos y empleados en los análisis filogenéticos. Dando validez tanto al uso de estos genes en nuestros análisis como robustez a los resultados obtenidos de ellos.

3.4.1 Mantenimiento de la sintenia

Para la comprobación del mantenimiento de la sintenia en los genes empleados en los análisis, se utilizó de nuevo la página PLAZA genomics (Van Bel et al., 2018; Vandepoele et al., 2013), en sus versiones “Dicots PLAZA 4.5” para dicotiledóneas y “Monocots PLAZA 4.5” para Monocotiledóneas. Para los análisis de sintenia, se emplearon de nuevo los genes de las secuencias de referencia anteriormente utilizadas, en este caso el nombre con el que están anotados en la base de datos. El gen de *A. thaliana* “AT5G25350”, se empleó para la comprobación con las especies de Monocotiledóneas y Brásicas, mientras que el gen de *S. lycopersicum* “Solyc08g060810.2” se empleó para la comprobación de las especies de Solanáceas. Se utilizó un tamaño de ventana de 5 genes a cada lado del gen de EBF2 en cada especie. Para la comprobación de las especies de cada familia, se seleccionaron las especies presentes en la base de datos, utilizadas anteriormente en el resto de análisis. Por último, para una mejor visualización, se proyectó el gen central, el supuesto gen ortólogo en todas las especies, en la dirección forward, de modo que se pudiera detectar con mayor facilidad si se estaba produciendo el mantenimiento o no de la sintenia.

3.4.2 Patrones de expresión

Otra aproximación complementaria para confirmar que los genes utilizados en los análisis son los ortólogos de nuestras secuencias de referencia, es analizar los patrones de expresión de cada ortólogo en su especie correspondiente. Si estos patrones de expresión están conservados, el gen ortólogo se expresa durante específicos procesos del desarrollo, en ciertos tejidos, o en respuesta a determinados estímulos, se puede afirmar con mayor confianza que se trata de genes ortólogos.

Para esta confirmación, se analizaron los datos de expresión en varias especies en distintos experimentos recogidos en el buscador eFP en la base de datos de la página web BAR Toronto (Winter et al., 2007). Se seleccionaron 5 especies que formaban parte de los anteriores análisis, de las cuales se disponía de datos de expresión: *Arabidopsis thaliana*, *Medicago truncatula*, *Solanum lycopersicum*, *Solanum tuberosum* y *Oryza sativa*. Los nombres de los genes de estas especies fueron obtenidos de la web PLAZA, tras comprobar que estaban anotados como genes codificantes de EBF2 o proteína F-box. Para cada especie se obtuvo la información de la expresión del gen en 3 tejidos: Raíz, hoja y flor/inflorescencia, junto con una imagen que representa gráficamente la expresión absoluta en cada tejido.

3.5 PLEGAMIENTO DE LA REGIÓN 3’UTR DE LOS EBF2

Una de las hipótesis planteadas sobre el posible modo en que el 3’UTR de los transcritos de EBF2 modula su traducción a proteína, está asociada a la estructura tridimensional que adquiere el mRNA a la hora de plegarse. Es posible que dicha estructura sea capaz de favorecer la interacción con la región C-terminal de EIN2 y así inhibir la traducción de estos transcritos. Para generar los modelos de las estructuras secundarias de los 3’UTRs de los distintos ortólogos de EBF2, los cuales están sometidos al mismo tipo de regulación traduccional en presencia de etileno, se utilizaron diversas herramientas bioinformáticas capaces de predecir y generar la estructura

secundaria de las secuencias de RNA aplicando algoritmos distintos. De esta manera, las estructuras producidas por una de las herramientas deben mantener cierto grado de similitud o ciertas características conformacionales con las estructuras producidas por otra herramienta, dando a entender que las características de las estructuras de los transcritos plegados se deben realmente a que son las óptimas y no son artefactos dependientes del uso de un algoritmo u otro. Para la generación de estas estructuras, los softwares buscan la conformación más estable de las secuencias, o lo que es lo mismo la que menor energía libre presenta (Taliaferro et al., 2016).

Para ello se escogieron las secuencias del 3'UTR de EBF2 de seis especies de Solanáceas, Brásicas y Monocotiledóneas y se almacenaron en archivos ".fasta" para ser usadas como input en las herramientas de simulación de estructuras.

3.5.1 RNAstructure fold

El primer programa utilizado fue RNA structure version 6.3 (Reuter y Mathews, 2010), con el que se predice el plegamiento del RNA en función de la estructura con la mínima energía libre (MFE), obteniendo la estructuras con la mayor precisión. Se usó la opción "Fold RNA single strand", la cual toma como input las secuencias de RNA individuales en formato FASTA. Se emplearon los valores predeterminados para los parámetros:

- Máximo porcentaje de diferencia de energía de 10%, la cual indica que cualquier estructura generada, cuya energía libre sea mayor a la de la estructura generada con menor energía, en más de un 10% será rechazada.
- Número máximo de estructuras de 5, indicando que el límite de estructuras generadas que cumplan el primer requisito es de 5.
- Tamaño de ventana de 0, parámetro que controla la cómo de diferentes pueden ser las estructuras subóptimas. Un tamaño de ventana mayor permitiría la generación de estructuras con más diferencias entre ellas.

Las estructuras son generadas en formato ".ct" y visualizadas con el programa StructureEditor.

3.5.2 ViennaRNA

El segundo programa utilizado para obtener las estructuras de los transcritos fue ViennaRNA (Lorenz et al., 2011), el cual también emplea la minimización de la energía para la generación de los plegamientos de RNA. Se empleó la función RNAfold, con la que el programa calcula las estructuras de RNA con la mínima energía libre. Esta función, requiere el uso del siguiente comando:

```
RNAfold -p -d2 --noLP < especie.fasta > especie.out
```

El cual toma como input las secuencias en formato fasta y genera como output un archivo PostScript con la estructura en 2D del plegamiento de RNA.

3.5.3 RNAstructure multiling

Una vez realizada la búsqueda de las estructuras individuales de cada especie modeladas con la mínima energía libre, se empleó el mismo programa RNA structure version 6.3 con la opción “RNA Multiling”. Esta herramienta toma como input varias secuencias individuales de RNA, generando como resultado un número equivalente de estructuras con la mínima energía libre y con una conformación común entre ellas. Esta opción, se empleó para buscar una estructura o estructuras similares entre todas las secuencias ya que, al compartir todas ellas la misma función, ésta muy posiblemente depende de la estructura secundaria del RNA. La formación de la estructura secundaria del RNA no es una simple cuestión de maximizar el número de puentes estables y minimizar la energía libre, sino que ésta se encuentra restringida por muchos más factores como los procesos de plegamiento guiados por proteínas y ribozimas o la interacción con iones (Vandivier et al., 2016).

Se emplearon los valores predeterminados de los parámetros, exceptuando el número máximo de estructuras generadas, el cual fue reducido para mejorar el rendimiento:

- Máximo porcentaje de diferencia de energía de 20%.
- Número máximo de estructuras de 2.
- Tamaño de ventana de 2.
- Tamaño de ventana del alineamiento de 1.
- Penalización por gap de 0.4

Las estructuras son generadas en formato “.ct” y visualizadas con el programa StructureEditor.

4. RESULTADOS

4.1 Estudios filogenéticos de EBF2

Para evaluar la relación evolutiva de las proteínas de EBF2 y así identificar de forma concisa los ortólogos del gen de *Arabidopsis* se construyeron los árboles filogenéticos con la secuencia de aminoácidos de varias especies. Empleando las herramientas de análisis filogenético, se pudo observar como el gen de EBF2 se conserva en todas las especies y se distribuye de forma esperada en todo el Reino Vegetal, pudiendo agruparse en las familias de Brásicas y Solanáceas y el grupo de las Monocotiledóneas, las cuales conforman clados separados del resto de especies (Figuras 4.1 y 4.2). Al realizar los árboles filogenéticos centrándose sólo en las especies de estos grupos, también se observó la conservación de la secuencia entre especies más cercanas, apareciendo agrupadas las especies del mismo género (Figura 4.1). Así mismo, se pudieron descartar secuencias pertenecientes a los ortólogos de EBF1 u otros miembros cercanos con una alta similitud de secuencia pertenecientes a la familia de las proteínas F-Box.

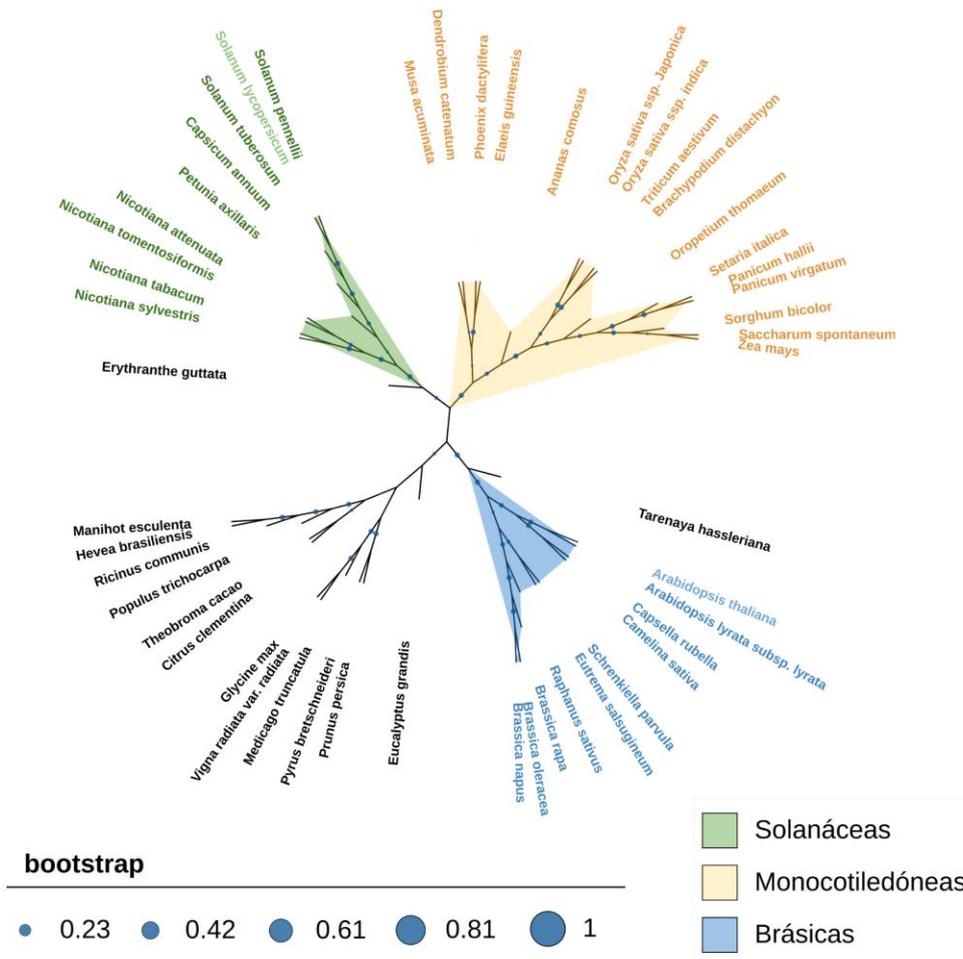
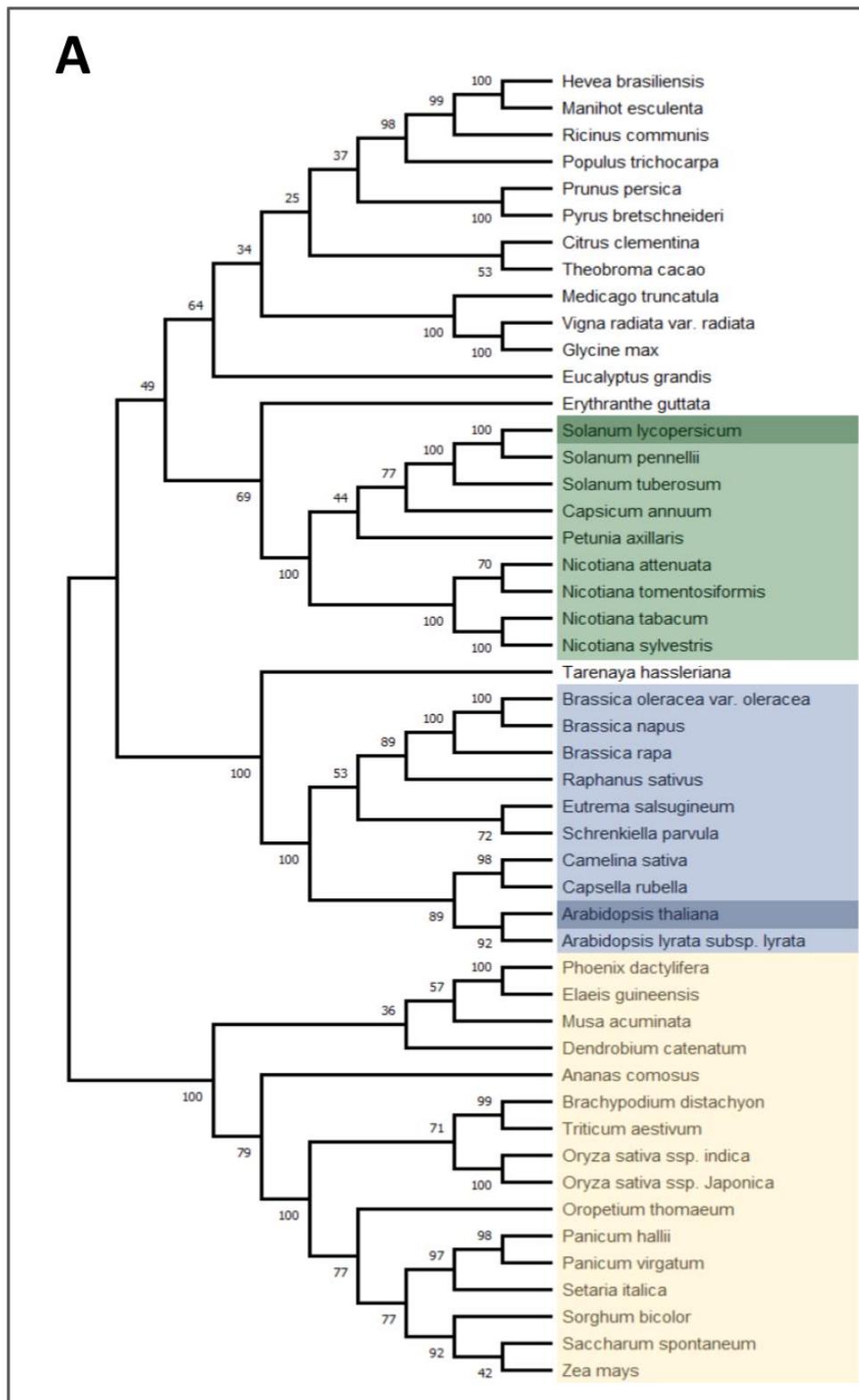


Figura 4.1 Distribución de las especies en el Reino Vegetal en función de la secuencia de aminoácidos de la proteína EBF2: Destacados los clados de Monocotiledóneas (amarillo) y las familias de Solanáceas (verde) y Brásicas (azul). Especies destacadas *S. lycopersicum* (verde claro) y *A. thaliana* (azul claro). Valores de bootstrap en cada nodo.



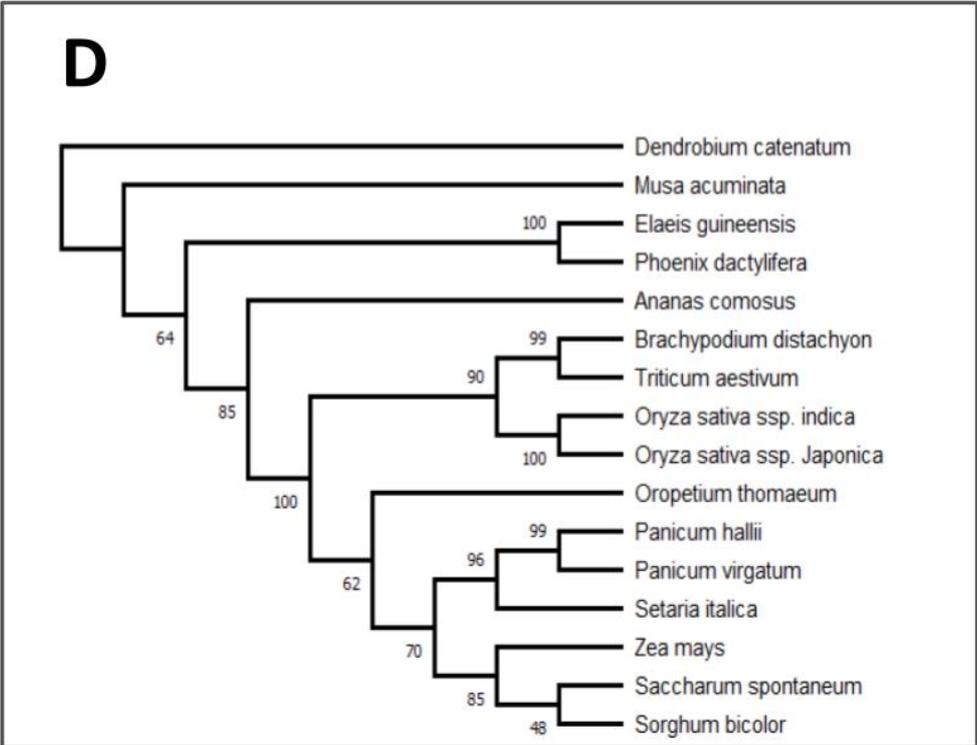
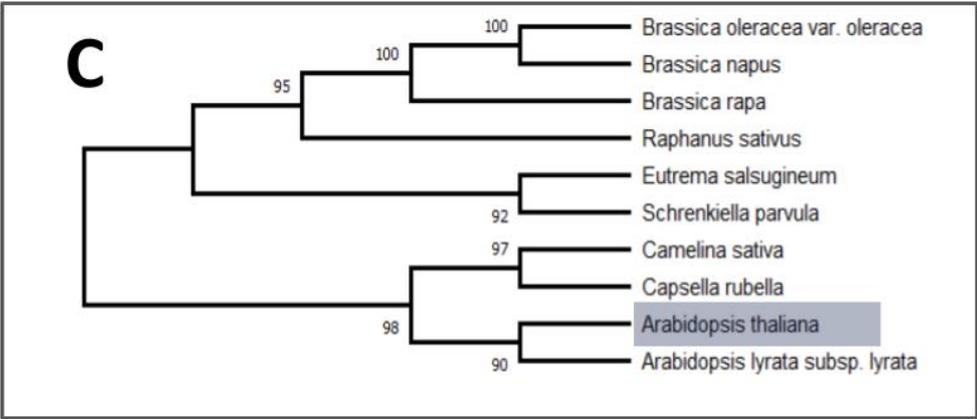
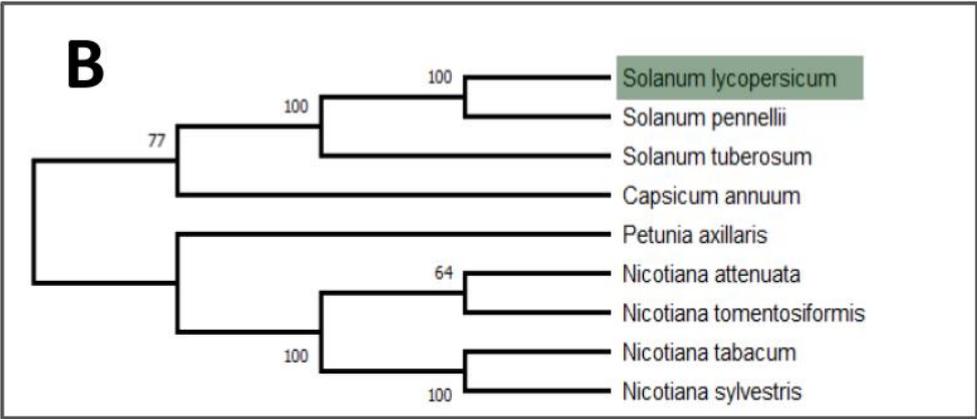
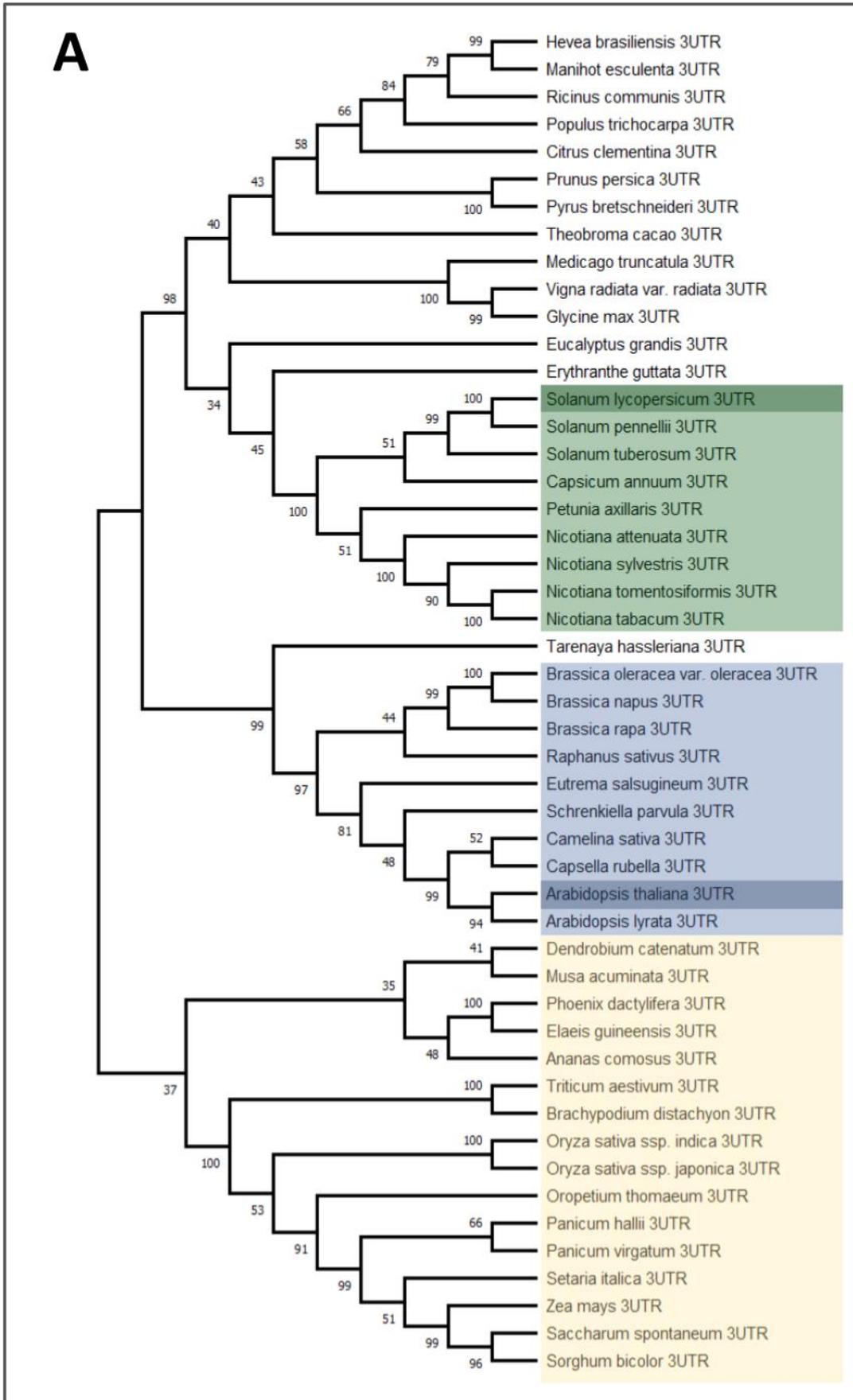


Figura 4.2 Filogenias distribución de las especies en base a la secuencia de aminoácidos de EBF2: (A) Filogenia Reino Vegetal: Monocotiledóneas (amarillo), Solanáceas (verde), Brásicas (azul), *Solanum lycopersicum* (verde oscuro), *Arabidopsis thaliana* (azul oscuro); **(B)** Filogenia familia Solanáceas: *S. lycopersicum* (verde oscuro); **(C)** Filogenia familia Brásicas: *A. thaliana* (azul oscuro); **(D)** Filogenia grupo Monocotiledóneas. Árbol realizado por análisis de Maximum Likelihood y con test de filogenia Bootstrap (nº Bootstrap Replicates = 1000); valores bootstrap en cada nodo.

Igualmente, se estudió la relación evolutiva de la región 3'UTR de los ortólogos de EBF2 en las especies analizadas, y se pudo observar de nuevo una distribución esperada, con las especies de las familias de Brásicas y Solanáceas y el grupo de Monocotiledóneas agrupadas en distintos clados evolutivos (Figura 4.3).

Estos resultados dan validez a posteriores análisis de la región 3'UTR y de su potencial función a la hora de regular la traducción de EBF2, pues su presencia en todas las especies estudiadas y la conservación de su secuencia, confirman la hipótesis de su importancia en esta función en el Reino Vegetal.



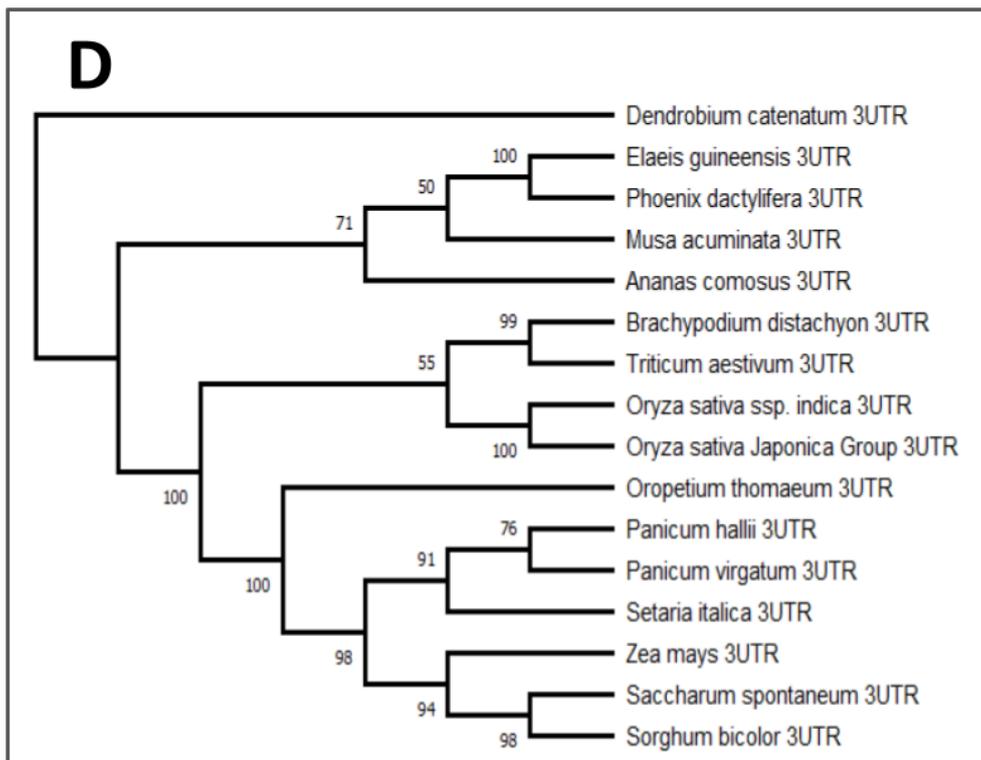
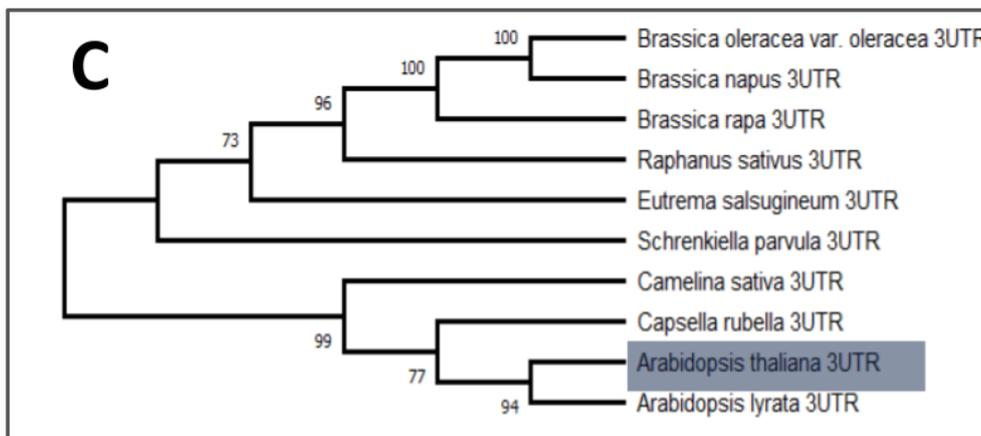
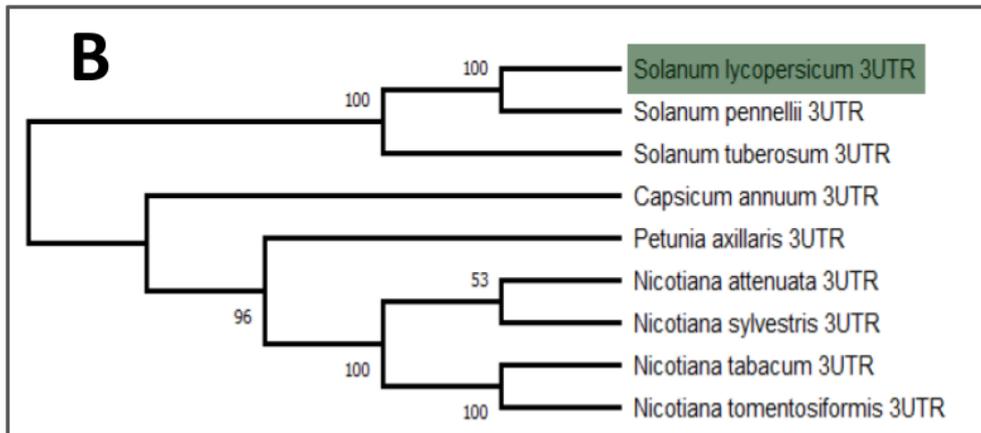
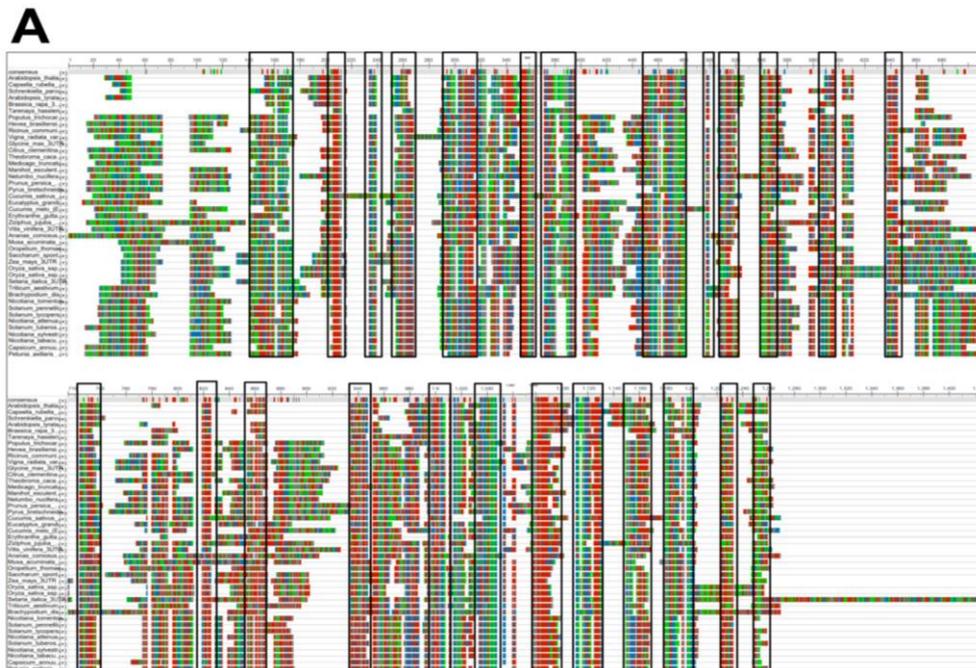


Figura 4.3 Filogenias distribución de las especies en base a la secuencia de nucleótidos de la región 3'UTR de EBF2: (A) Filogenia Reino Vegetal: Monocotiledóneas (amarillo), Solanáceas (verde), Brásicas (azul), *Solanum lycopersicum* (verde oscuro), *Arabidopsis thaliana* (azul oscuro); (B) Filogenia familia Solanáceas: *S. lycopersicum* (verde oscuro); (C) Filogenia familia Brásicas: *A. thaliana* (azul oscuro); (D) Filogenia grupo Monocotiledóneas. Árbol realizado por análisis de Maximum Likelihood y con test de filogenia Bootstrap (nº Bootstrap Replicates = 1000); valores bootstrap en cada nodo.

4.2 Alineamientos múltiples

A partir de los resultados obtenidos de los alineamientos múltiples, se puede observar cierta conservación en las secuencias del 3'UTR de los EBF2 del Reino Vegetal. Cuanto más relacionadas están las secuencias, grupo Monocotiledóneas y familias de Brásicas y Solanáceas, mayor es el número de regiones conservadas identificadas. No obstante, la conservación se diluye a medida que el análisis es realizado con especies más alejadas evolutivamente, manteniendo sólo algunas zonas conservadas repetidas periódicamente. Son estas regiones alineadas y su presencia en todas las especies analizadas del Reino Vegetal, las que podrían tener una mayor importancia e implicación a la hora de regular la traducción del gen EBF2 (Figura 4.4).

La conservación, tanto de la secuencia como del gran tamaño de las regiones 3'UTR en todas las especies analizadas, comprobada tanto con las filogenias como con el análisis de los alineamientos múltiples, permite hipotetizar con la importancia de esta región y su función reguladora de la traducción, muy posiblemente conservada a lo largo del Reino Vegetal.



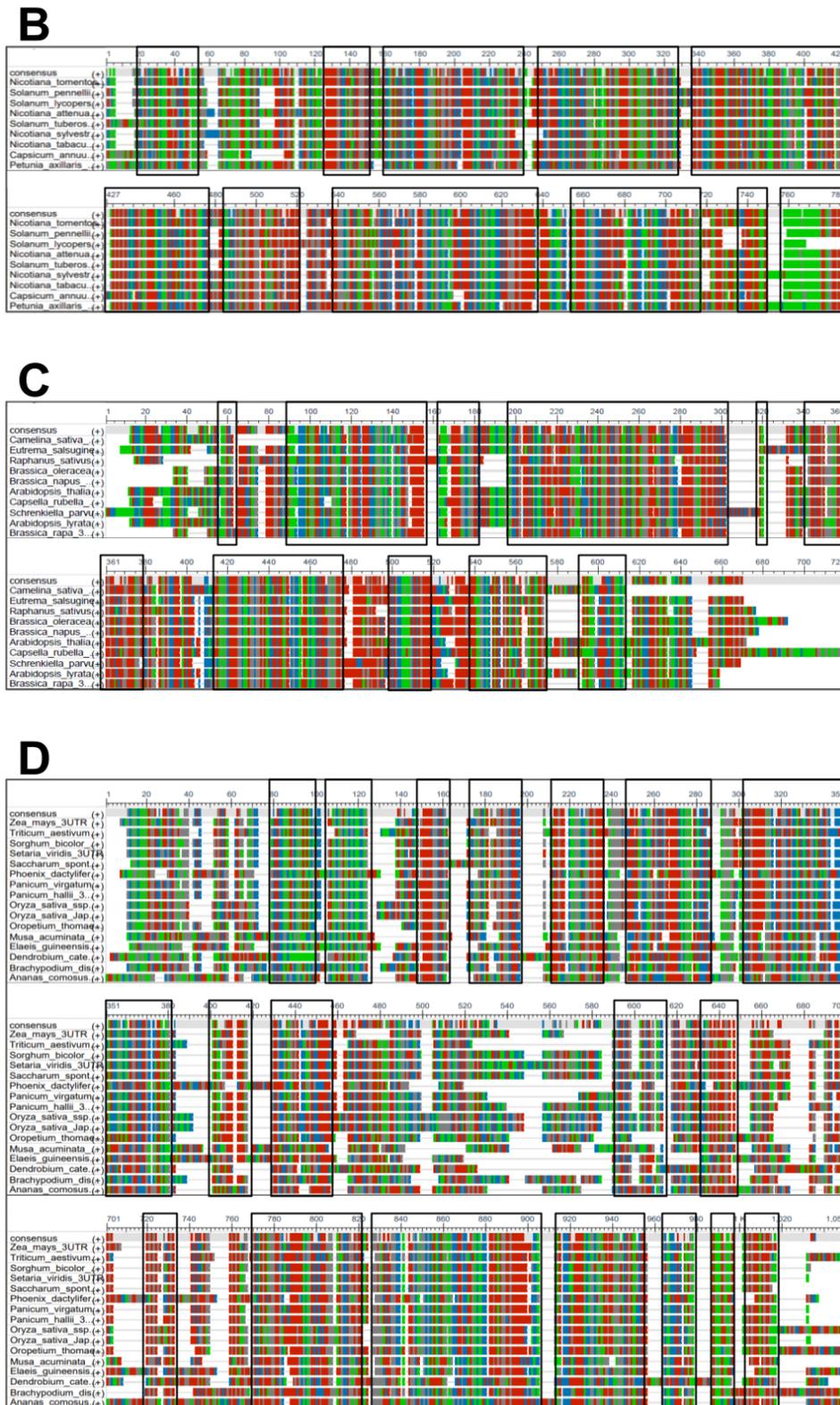


Figura 4.4 Alineamientos múltiples de la secuencia de nucleótidos de la región 3'UTR de EBF2: (A) Especies del Reino Vegetal, (B) Especies familia Solanáceas, (C) Especies familia Brásicas, (D) Especies grupo Monocotiledóneas. Alineamiento realizado con el programa MEGA-X, mediante Muscle Alignment con método de clusters Neighbor Joining. Color dependiente del nucleótido: A (verde), G (gris), T (rojo), C (azul). Zonas más conservadas señaladas.

4.3 Mantenimiento de la sintenia

La sintenia es la conservación de regiones de un cromosoma en las que se pueden observar los mismos conjuntos de genes (Ghiurcuta y Moret, 2014). Dentro de la definición de sintenia, encontramos el concepto de microsintenia, el cual mantiene la definición anterior, pero reduce la conservación a regiones pequeñas dentro de un cromosoma, de tan solo unos pocos genes (Engström et al., 2007). La sintenia permite evaluar las fuerzas evolutivas que actúan sobre la estructura de los cromosomas y además es una herramienta de gran utilidad a la hora de determinar la ortología de genes entre varias especies. Aquellos genes cuyos genes vecinos ocupan posiciones similares, es decir un mantenimiento de la microsintenia, son muy probablemente genes ortólogos, ya que, durante la evolución de las especies, los genes tienden a moverse en bloques (Ghedini et al., 2004).

La sintenia de los ortólogos de EBF2 se analizó en tres grupos distintos de especies vegetales. En los resultados de estos análisis se puede observar un claro mantenimiento de los genes en la región del cromosoma cercana a la del ortólogo con el que se trabajó para cada especie. Como se puede ver en la Figura 4.5, en el caso de la familia de Solanáceas, parece mostrar una mayor variabilidad en la región evaluada, en los casos de Brásicas y Monocotiledóneas, la sintenia parece mantenerse de forma más robusta. Estos resultados contribuyen a la confirmación de que los genes identificados como ortólogos de EBF2 de *Arabidopsis* realmente lo son. De este modo se validan las secuencias seleccionadas en los análisis filogenéticos y se confirma el uso de estas secuencias para estudios adicionales. Como última etapa en la confirmación de la validez del uso de estas secuencias para intentar definir elementos reguladores en los 3'UTR responsables de la inhibición de la traducción del transcrito de EBF2 en respuesta a etileno se compararon los patrones de expresión mostrados por varios de los ortólogos en sus especies correspondientes, en distintos estadios de desarrollo y varios tejidos.

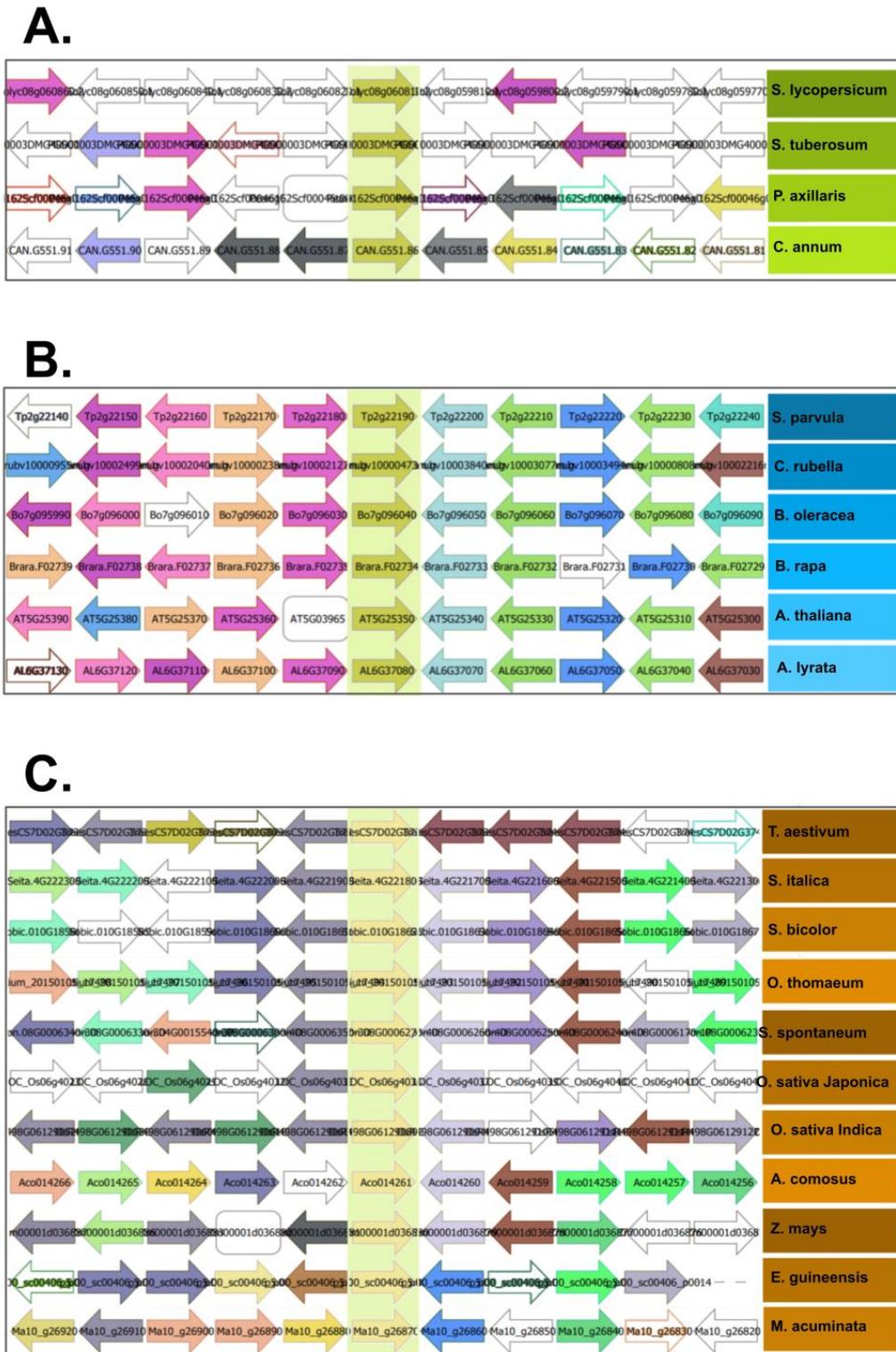


Figura 4.5 Sintenia en la región 5 genes upstream y 5 genes downstream del gen EBF2 (fondo verde claro): (A) Sintenia en especies de la familia de Solanáceas; (B) Sintenia en especies de la familia de Brásicas; (C) Sintenia en especies del grupo de Monocotiledóneas. Imágenes adaptadas de datos obtenidos en web Plaza, flechas del mismo color indican genes ortólogos.

4.4 Patrones de expresión

Los patrones de expresión se analizaron empleando la base de datos de BAR (Winter et al., 2007), donde se obtuvo información de especies pertenecientes a las familias de Solanáceas (*Solanum lycopersicum* y *Solanum tuberosum*) y Brásicas (*Arabidopsis thaliana*), al grupo de las Monocotiledóneas (*Oryza sativa*) y una especie dicotiledónea extra (*Medicago truncatula*) para completar los resultados. Los patrones de expresión de los ortólogos analizados muestran una alta similitud. Los distintos ortólogos de EBF2 tienden a expresarse con mayor intensidad en las estructuras de flor e inflorescencia y generalmente en la raíz, con alguna excepción como en el caso de *A. thaliana*; mientras que su expresión en órganos vegetativos como las hojas es mucho más baja. Esta pauta se repite en las especies analizadas (Figura 4.6), lo que hace aumentar la confianza de que los genes identificados son ortólogos.

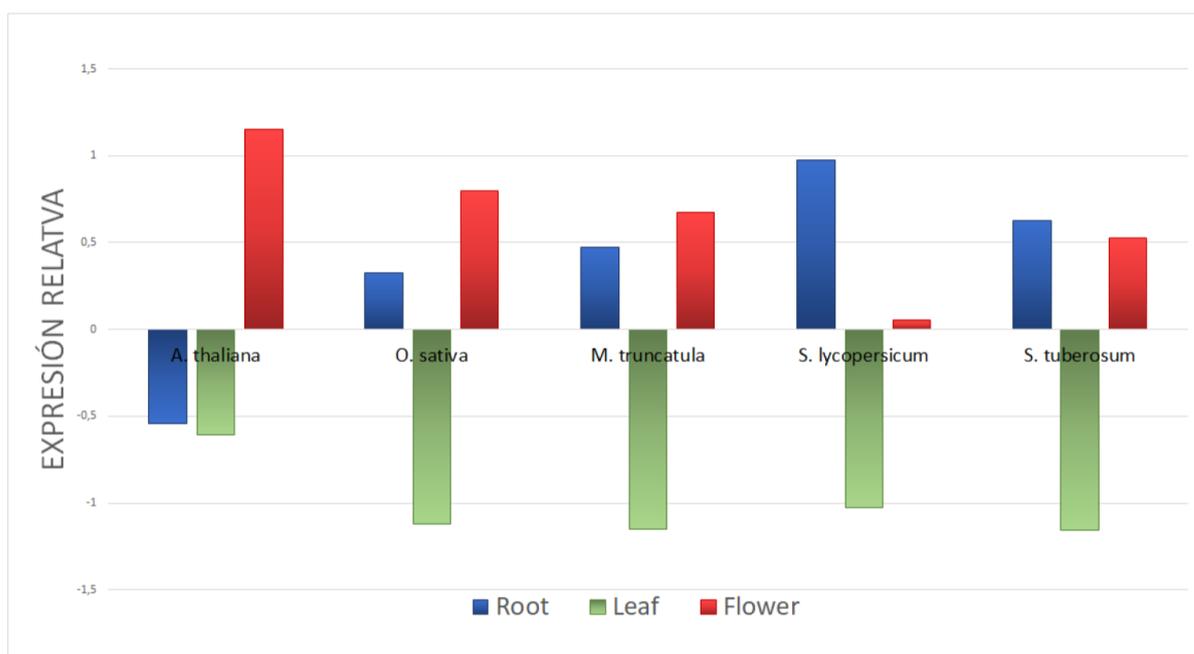


Figura 4.6 Patrones de expresión relativa en los tejidos de varias especies: Expresión relativa de EBF2 en los tejidos de *A. thaliana*, *O. sativa*, *M. truncatula*, *S. lycopersicum*, *S. tuberosum*; raíz (azul), hoja (verde), flor/inflorescencia (rojo).

4.4.1 Expresión del gen elegido para cada especie

Tras obtener los datos de expresión relativa en tres tejidos en las especies previamente mencionadas, se obtuvieron varias figuras con información visual sobre la expresión absoluta del gen en cada una de las especies. En estas figuras, las estructuras y órganos con mayor expresión tienen tonalidades rojas o naranjas, frente a aquellas regiones en las que el gen muestra valores de expresión muy bajos que están coloreadas con una tonalidad amarilla o amarilla anaranjada.

En *Arabidopsis*, *A. thaliana* (Figura 4.7), se puede observar como la estructura vegetativa de la planta, junto con los estados tempranos de la semilla presentan una baja expresión del gen, mientras que otras estructuras como la flor o la semilla en estados avanzados de maduración

presentan altos niveles de expresión. También destaca la mayor importancia del gen en la hoja senescente, pues la actividad en este tejido, junto con los mencionados anteriormente, está muy ligada a la presencia de la hormona etileno (Grbić y Bleecker, 1995; Hoerberichts et al., 2002; Sun et al., 2020), relacionada con la función de EBF2.

Al igual que en *A. thaliana*, en *Solanum lycopersicum* y *Solanum pimpinellifolium* se puede observar una expresión baja de EBF2 en los tejidos vegetativos como la hoja o el tallo, y por el contrario, EBF2 muestra unos niveles de expresión mucho más elevados en la raíz y en los órganos reproductivos como flores y frutos (Figura 4.8), donde la importancia del etileno es también fundamental (Catchpole y Hillman, 1969). En *S. tuberosum*, el fruto y la flor parecen ser de nuevo los tejidos donde el gen muestra valores de expresión más elevados, junto con la raíz y los tubérculos (Figura 4.9), donde también el etileno juega un papel importante en el desarrollo y crecimiento de estos órganos (Ghiurcuta y Moret, 2014).

Por último, los datos de *O. sativa* muestran de nuevo una menor incidencia del gen en tejidos vegetativos como las hojas, o en los estados tempranos de la semilla y un alto nivel de expresión en tejidos como la inflorescencia, la semilla en estados avanzados y la raíz (Figura 4.10).

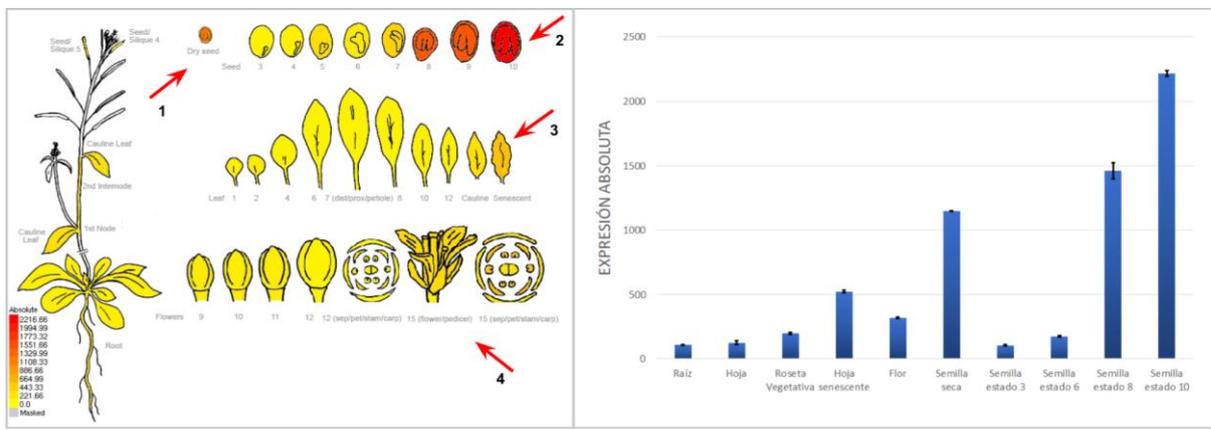


Figura 4.7 Patrones de expresión de EBF2 en *A. thaliana*: Expresión absoluta de EBF2 en varios tejidos de *A. thaliana* (izquierda); expresión de EBF2 en los tejidos de semilla, hoja y flor durante los estados del desarrollo (derecha, de arriba a abajo). Flechas rojas indicando tejidos donde se da mayor expresión de EBF2: **(1)** Semilla seca; **(2)** Semilla en último estado de desarrollo; **(3)** Hoja senescente; **(4)** Flor. A la derecha gráfica con valores absolutos de la expresión de EBF2 en los tejidos, barras de error mostradas corresponden con los valores de desviación estándar. Datos disponibles en (Matas et al., 2011).

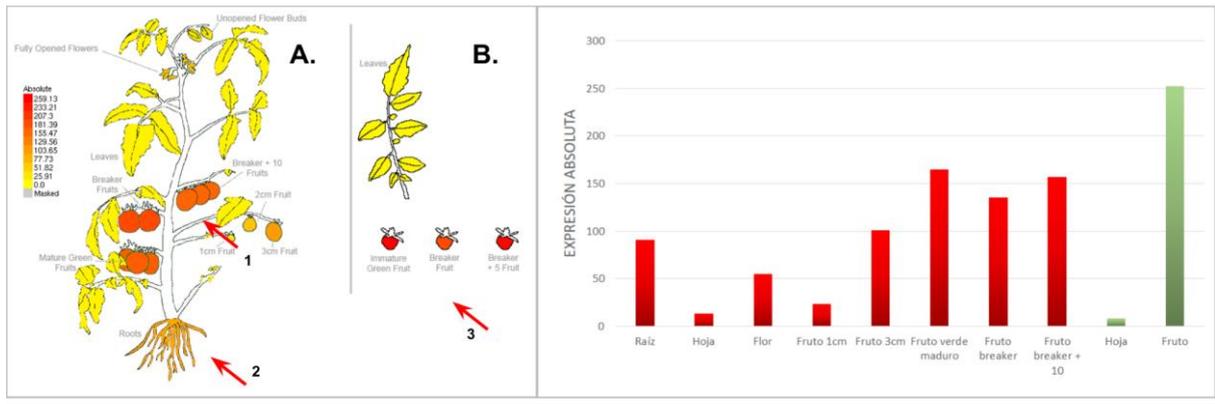


Figura 4.8 Patrones de expresión *S. lycopersicum*: Expresión absoluta de EBF2 en varios tejidos de *S. lycopersicum* (A) y *S. pimpinellifolium* (B). Flechas rojas indicando tejidos donde se da mayor expresión de EBF2: (1) Frutos *S. lycopersicum*, (2) Raíz, (3) Frutos *S. pimpinellifolium*. A la derecha gráfica con valores absolutos de la expresión de EBF2 en los tejidos de *S. lycopersicum* (rojo) y *S. pimpinellifolium* (verde), datos correspondientes a la desviación estándar no disponibles. Datos disponibles en (Massa et al., 2011).

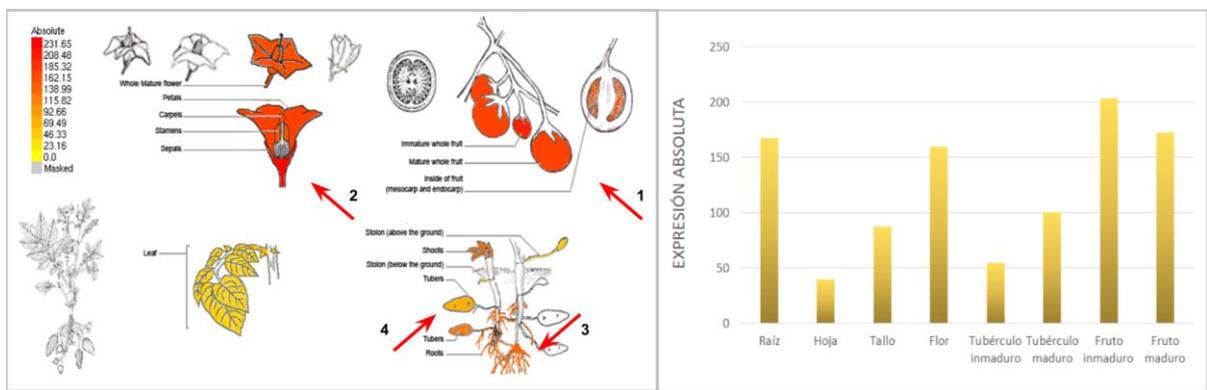


Figura 4.9 Patrones de expresión *S. tuberosum*: Expresión absoluta de EBF2 en varios tejidos de *S. tuberosum*. Flechas rojas indicando tejidos donde se da mayor expresión de EBF2: (1) Fruto, (2) Flor, (3) Raíz, (4) Tubérculo. A la derecha gráfica con valores absolutos de la expresión de EBF2 en los tejidos, datos correspondientes a la desviación estándar no disponibles. Datos disponibles en (Jain et al., 2007).

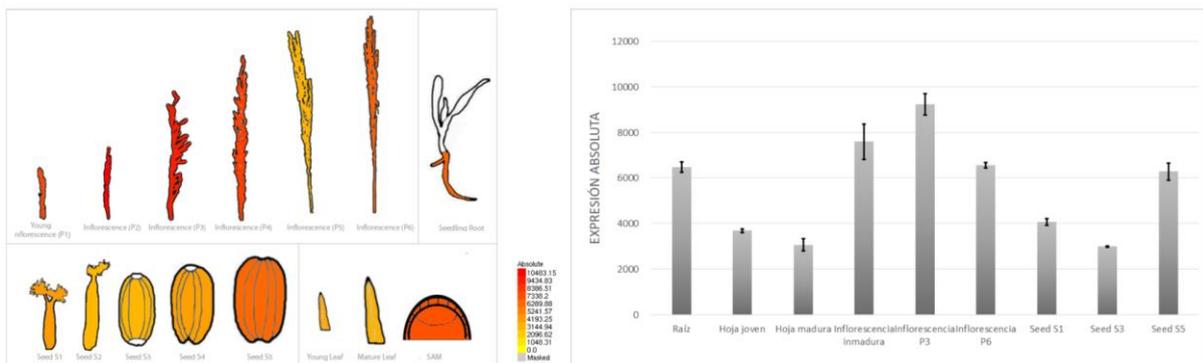


Figura 4.10 Patrones de expresión O. sativa: Expresión absoluta de EBF2 en varios tejidos de *O. sativa* durante diferentes estados del desarrollo. **(A)** Inflorescencia, **(B)** Raíz, **(C)** Semilla, **(D)** Hoja. A la derecha gráfica con valores absolutos de la expresión de EBF2 en los tejidos, barras de error mostradas corresponden con los valores de desviación estándar. Datos disponibles en (NASCArrays at The BAR, s. f.).

A partir de los resultados de los patrones de expresión, siendo similares en las distintas especies y afectando principalmente a órganos en los que el etileno juega un papel importante, junto con los resultados obtenidos en los análisis de sintenia, se puede afirmar con más confianza que los genes empleados en los análisis son genes ortólogos.

4.5 Estructuras de los 3'UTR

Los resultados obtenidos a partir del modelado de las estructuras secundarias de las regiones 3'UTR de los transcritos, mostraron para todos los algoritmos utilizados, estructuras con regiones apareadas de gran longitud y pocos bucles o regiones sin aparear intermedias. La presencia de estas zonas de apareamiento son una característica de complejidad estructural, las cuales potencialmente podrían tener un efecto en la eficiencia de la traducción de estas estructuras. Además, estos patrones de apareamiento estructural son susceptibles de ser modificados en respuesta a la unión de proteínas u otras moléculas al transcrito (Schmid et al., 2005).

Al analizar las estructuras obtenidas con la misma función "RNAfold" en ambos softwares, RNAstructure 6.3 y ViennaRNA, se puede observar su gran semejanza (Figuras 4.11, 4.12 y 4.13). Esto unido a la similitud en los valores de las energías libres de estas estructuras para ambos programas, permite hipotetizar que el 3'UTR de esas especies debe adquirir una configuración *in vivo* que podría asemejarse a los modelos generados. También se puede comprobar, que las estructuras obtenidas con la función "RNA Multiling" presentan una energía libre mayor que las obtenidas con la función "RNAfold" (Figuras 4.11 y 4.12). No obstante, estas mantienen una conformación similar con las del resto de especies, lo cual podría tener un alto sentido biológico, al compartir todas ellas la función de regular la traducción de los transcritos de EBF2 en presencia de etileno.

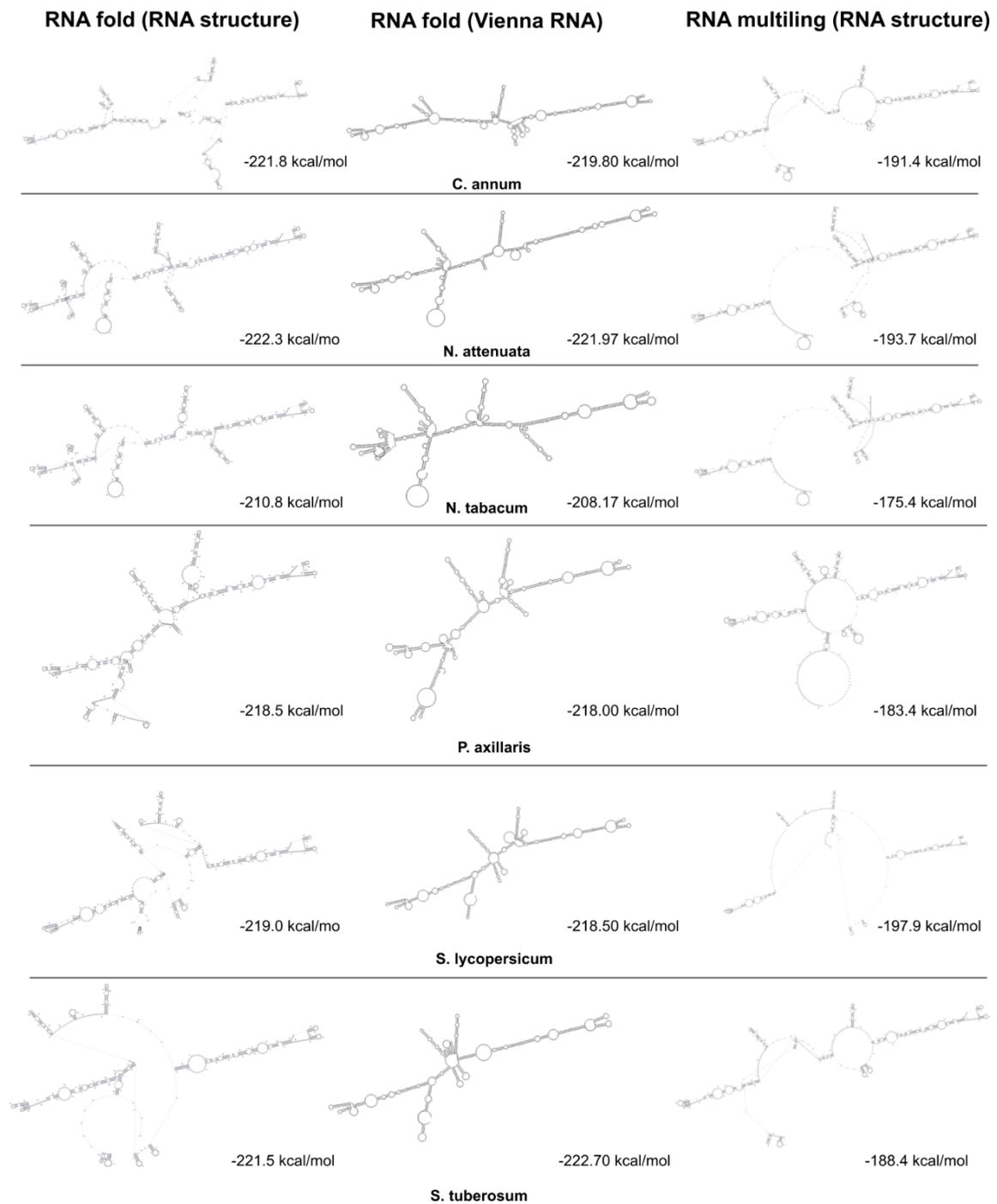


Figura 4.11 Estructuras de la región 3'UTR de el EBF2 para las especies pertenecientes a la familia Solanáceas: (Izquierda) Obtenidas con la herramienta "RNAfold" del programa RNAstructure 6.3; (Centro) Obtenidas con la herramienta "RNAfold" del programa ViennaRNA; (Derecha) Obtenidas con la herramienta "RNA Multiling" del programa RNAstructure 6.3. Valores de la energía libre de Gibbs asociados a cada estructura expresados en Kcal/mol.

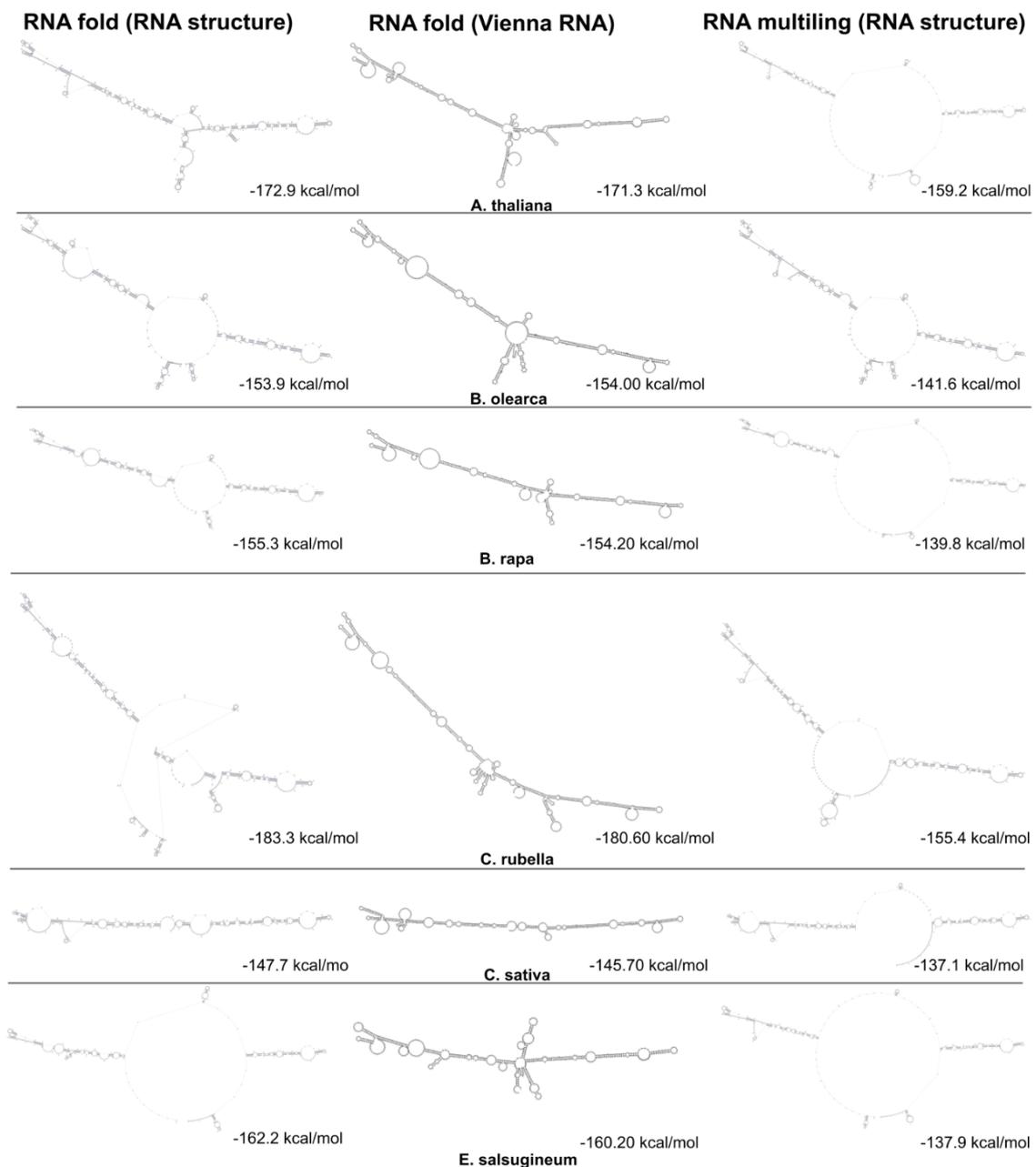


Figura 4.12 Estructuras de la región 3'UTR de el EBF2 para las especies pertenecientes a la familia Brásicas: (Izquierda) Obtenidas con la herramienta "RNAfold" del programa RNAstructure 6.3; (Centro) Obtenidas con la herramienta "RNAfold" del programa ViennaRNA; (Derecha) Obtenidas con la herramienta "RNA Multiling" del programa RNAstructure 6.3. Valores de la energía libre de Gibbs asociados a cada estructura expresados en Kcal/mol.

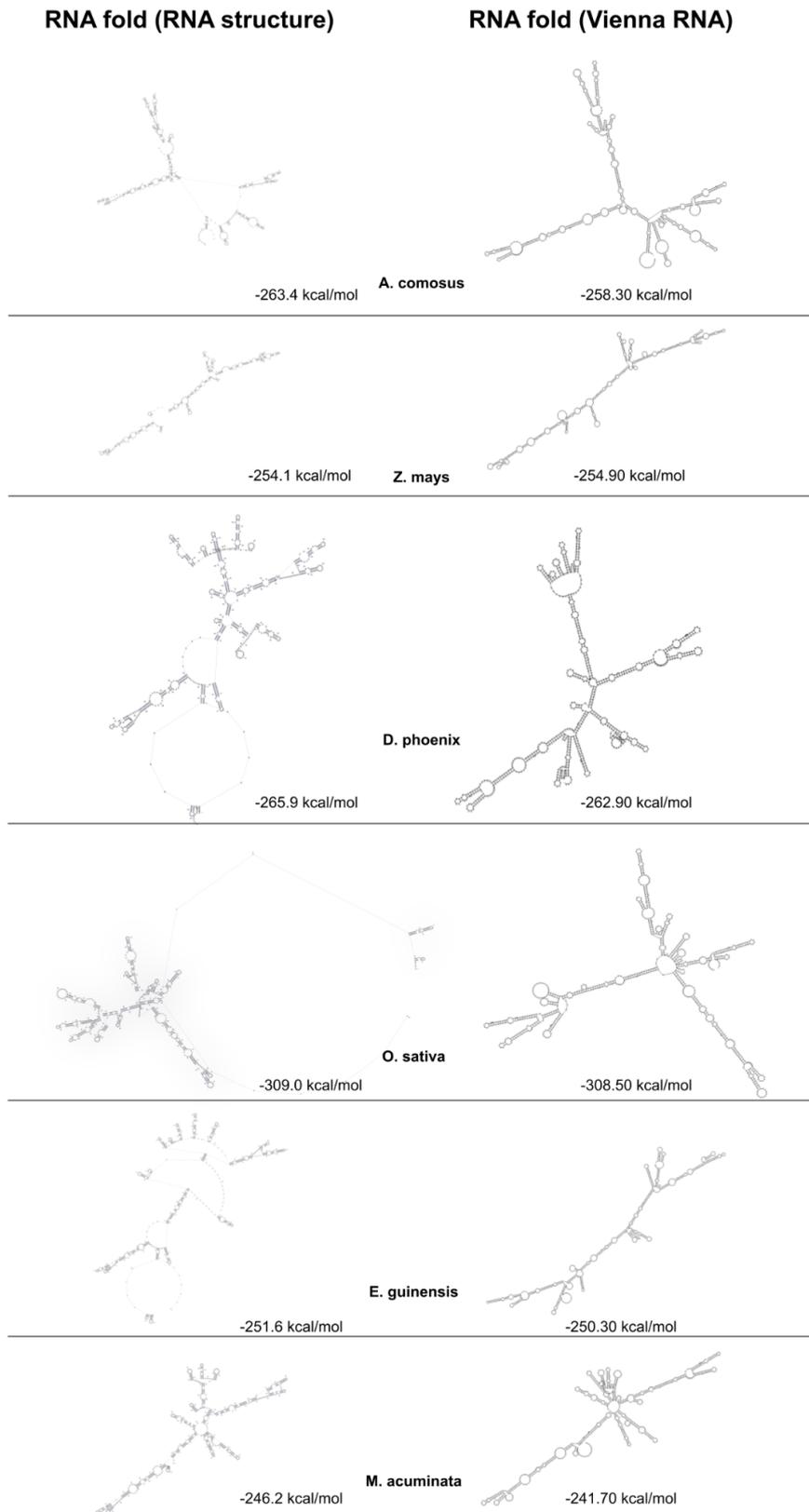


Figura 4.13 Estructuras de la región 3'UTR de el EBF2 para las especies pertenecientes al grupo de las **Monocotiledóneas**: (Izquierda) Obtenidas con la herramienta "RNAfold" del programa RNAstructure 6.3; (Derecha) Obtenidas con la herramienta "RNAfold" del programa ViennaRNA.

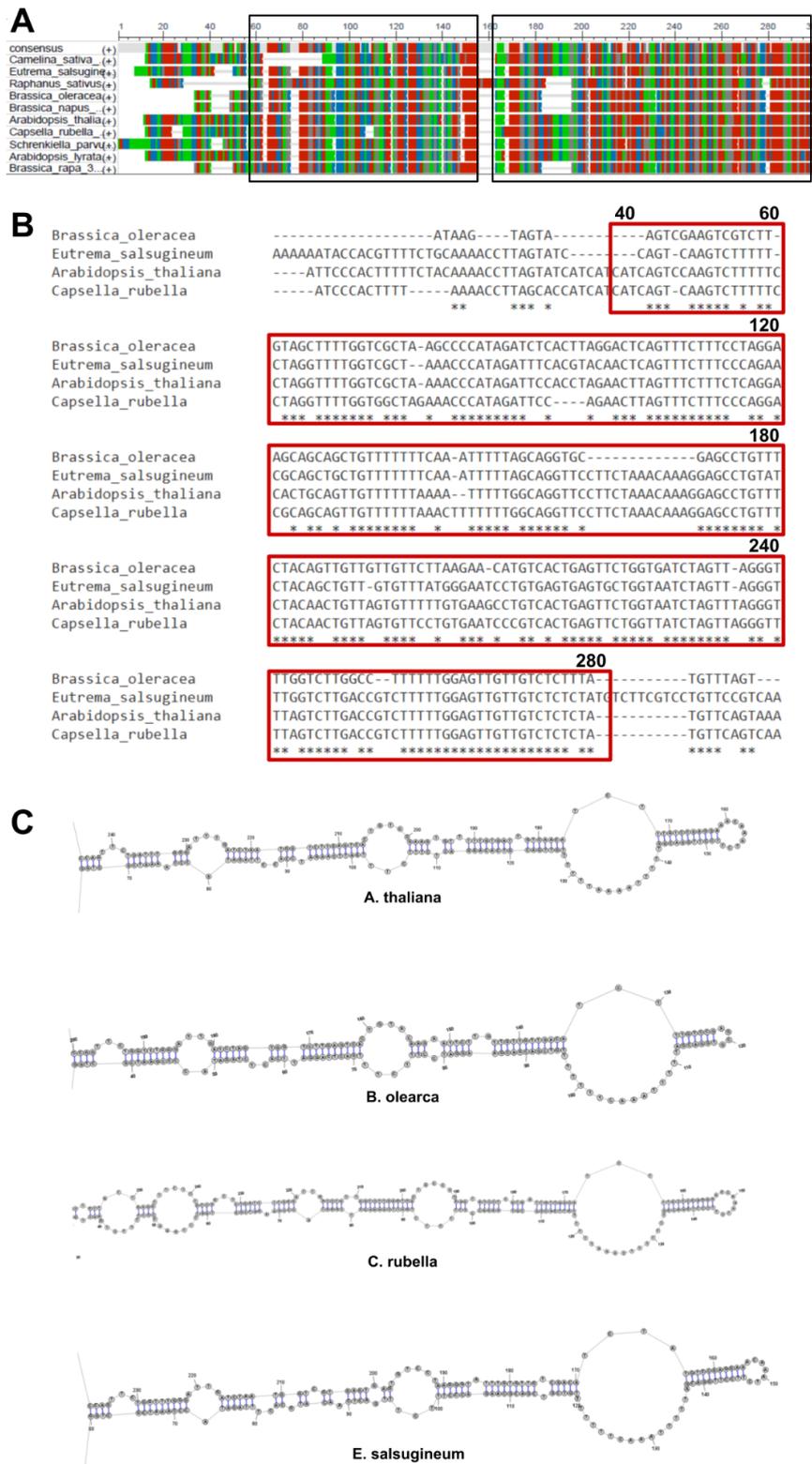


Figura 4.15 Relación de regiones conservadas en la secuencia del 3'UTR de ortólogos de Brásicas con el plegamiento de estas regiones: (A) Regiones señaladas con mayor conservación, (B) Secuencia de las regiones conservadas, alineamiento realizado en la web de EMBL-EBI, (C) Zonas de las estructuras con apareamiento similar, correspondientes a las secuencias mostradas.

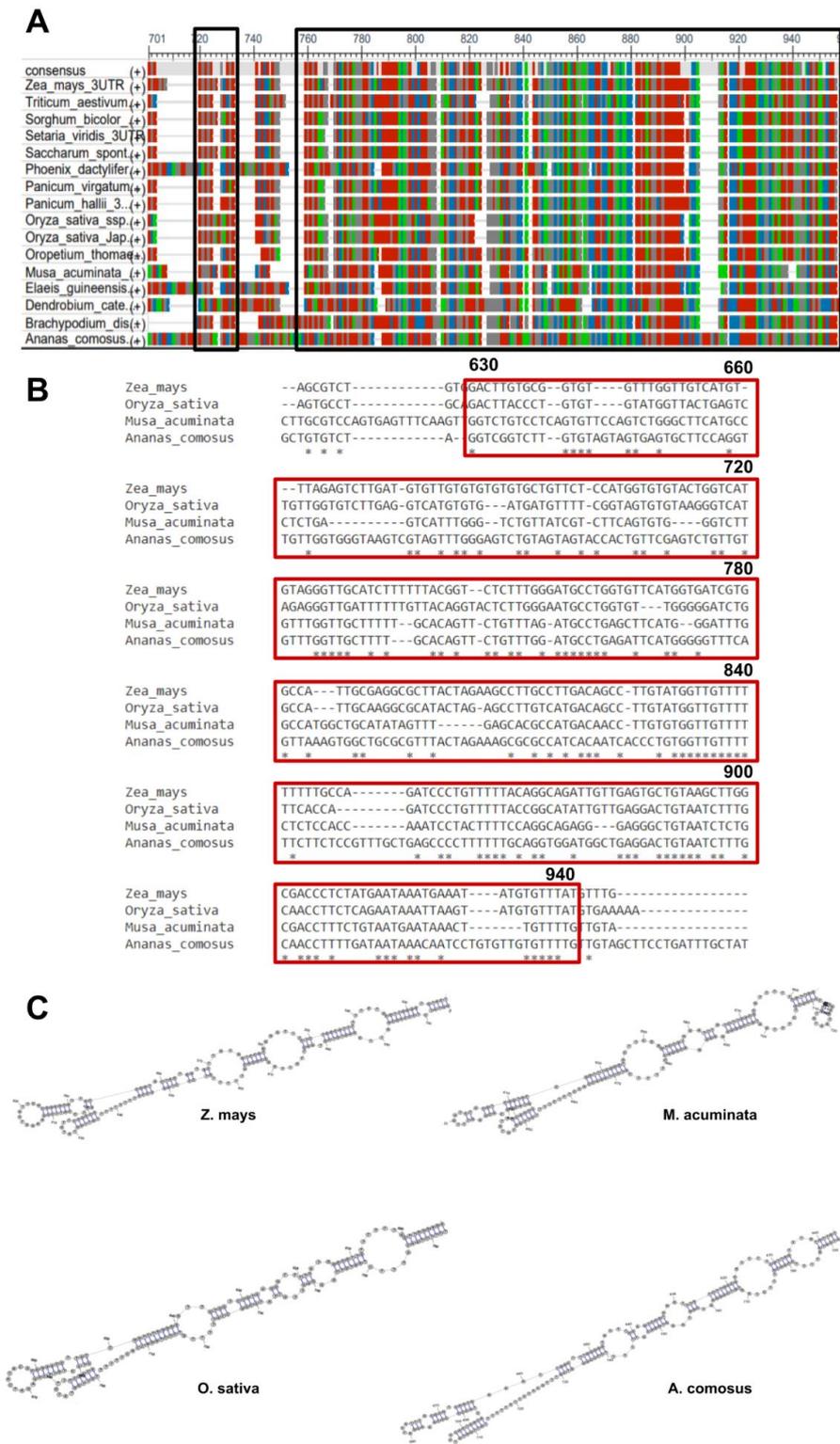


Figura 4.16 Relación de regiones conservadas en la secuencia del 3'UTR de ortólogos de Monocotiledóneas con el plegamiento de estas regiones: (A) Regiones señaladas con mayor conservación, (B) Secuencia de las regiones conservadas, alineamiento realizado en la web de EMBL-EBI, (C) Zonas de las estructuras con apareamiento similar, correspondientes a las secuencias mostradas.

5. DISCUSIÓN

La respuesta de las plantas a la hormona etileno se ha estudiado ampliamente y se ha identificado una vía de señalización lineal responsable de desencadenar la multitud de respuestas a esta hormona. Es importante destacar que todos los cambios de expresión génica a nivel transcripcional desencadenados por esta hormona requieren de la totalidad de la vía, incluidos EIN2 y los reguladores transcripcionales EIN3 y EIL1 (Alonso y Stepanova, 2004).

Recientemente, se han descrito también respuestas postranscripcionales a etileno, en las cuales, el etileno es capaz de afectar la eficiencia de traducción de un grupo específico de genes. La actividad de EIN2 es necesaria para la regulación de la traducción del mRNA de *EBF2*. La región C-terminal de EIN2 se asocia al 3'UTR del transcrito de *EBF2* inhibiendo su traducción.

Este trabajo final de grado trata de iniciar una nueva línea de investigación encaminada a la identificación de elementos reguladores en *cis* involucrados en la regulación traduccional de *EBF2*. Para ello se han analizado las secuencias del 3'UTR de los verdaderos ortólogos del *EBF2* de *Arabidopsis* en busca de secuencias conservadas que potencialmente puedan ser responsables de la regulación de la traducción de este mRNA a través del reino vegetal. Esta búsqueda no solo se ha llevado a cabo a nivel de secuencia, sino también a nivel de estructura de esta región 3' del transcrito. Se han logrado identificar regiones en los 3'UTRs de estos transcritos con cierto grado de conservación a nivel de secuencia que claramente corresponde con regiones de los modelos estructurales muy estables energéticamente que podrían poseer una función biológica implicada en la interacción con moléculas reguladoras. Esta asociación puede ser clave para la regulación de la traducción de estos mRNAs en presencia de etileno.

En un futuro próximo, se llevarán a cabo estudios adicionales más exhaustivos de estas regiones del 3'UTR de *EBF2* en colaboración con el Dr. Marcos de la Peña en el IBMCP. También se tratará de confirmar "in vivo" la función de estas regiones identificadas en los 3'UTR de los ortólogos de *EBF2*. Paralelamente, se pueden llevar a cabo ensayos de expresión transitoria y generación de líneas transgénicas expresando un "reporter gene" compuesto por la proteína fluorescente GFP bajo el control transcripcional de un promotor constitutivo como 35S o Ubiquitina10 y fusionada a las regiones del 3'UTR a testar. Este "reporter gene" en condiciones control mostrará elevados valores de expresión y la proteína fluorescente será fácilmente detectable. En cambio, en presencia de etileno, la traducción de este "reporter gene" se verá severamente inhibida.

Estos resultados pueden no solo mejorar nuestro conocimiento de la regulación molecular de la respuesta a etileno, hasta ahora principalmente estudiada solo a nivel transcripcional. Sino que pueden contribuir al establecimiento de las primeras herramientas de biología sintética específicas para el control y regulación de la traducción de cualquier proceso molecular en respuesta a etileno.

6. REFERENCIAS BIBLIOGRÁFICAS

- ACEVEDO, M. A.; CASTRILLO, W. A. Y BELMONTE, U. C. (2006). Origen, evolución y diversidad del arroz. *Agronomía Tropical*, 56(2), 151-170.
- ALONSO, J. M., Y STEPANOVA, A. N. (2004). The Ethylene Signaling Pathway. *Science*, 306(5701), 1513-1515.
- Aoki, K.; Yano, K.; Suzuki, A.; Kawamura, S.; Sakurai, N.; Suda, K.; Kurabayashi, A.; Suzuki, T.; Tsugane, T.; Watanabe, M.; Ooga, K.; Torii, M.; Narita, T.; Shin-i, T.; Kohara, Y.; Yamamoto, N.; Takahashi, H.; Watanabe, Y.; Egusa, M.; ... Shibata, D. (2010). Large-scale analysis of full-length cDNAs from the tomato (*Solanum lycopersicum*) cultivar Micro-Tom, a reference system for the Solanaceae genomics. *BMC Genomics*, 11(1), 210.
- BACKELJAU, T.; DE BRUYN, L.; DE WOLF, H.; JORDAENS, K.; VAN DONGEN, S. Y WINNEPENINCKS, B. (1996). Multiple UPGMA and Neighbor-joining Trees and the Performance of Some Computer Packages. *Molecular Biology and Evolution*, 13(2), 309-313.
- BINDER, B. M.; WALKER, J. M.; GAGNE, J. M.; EMBORG, T. J.; HEMMANN, G.; BLEECKER, A. B. Y VIERSTRA, R. D. (2007). The Arabidopsis EIN3 Binding F-Box Proteins EBF1 and EBF2 Have Distinct but Overlapping Roles in Ethylene Signaling. *The Plant Cell*, 19(2), 509-523.
- BLANKE, M. M. (2014). Reducing ethylene levels along the food supply chain: A key to reducing food waste? *Journal of the Science of Food and Agriculture*, 94(12), 2357-2361.
- BLEECKER, A. B. Y KENDE, H. (2000). Ethylene: A Gaseous Signal Molecule in Plants. *Annual Review of Cell and Developmental Biology*, 16(1), 1-18.
- BURGESS, D. Y FREELING, M. (2014). The Most Deeply Conserved Noncoding Sequences in Plants Serve Similar Functions to Those in Vertebrates Despite Large Differences in Evolutionary Rates. *The Plant Cell*, 26(3), 946-961.
- CALVO, S. E.; PAGLIARINI, D. J. Y MOOTHA, V. K. (2009). Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proceedings of the National Academy of Sciences*, 106(18), 7507-7512.
- CAO, Y.-R.; CHEN, S.-Y. Y ZHANG, J.-S. (2008). Ethylene signaling regulates salt stress response. *Plant Signaling y Behavior*, 3(10), 761-763.
- CATCHPOLE, A. H. Y HILLMAN, J. (1969). Effect of Ethylene on Tuber Initiation in *Solanum tuberosum* L. *Nature*, 223(5213), 1387-1387.
- CHAMARY, J. Y HURST, L. D. (2005). Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. *Genome Biology*, 6(9), R75.
- CHANG, C. Y STADLER, R. (2001). Ethylene hormone receptor action in Arabidopsis. *BioEssays*, 23(7), 619-627.
- CHEADLE, V. I. (1942). The Occurrence and Types of Vessels in the Various Organs of the Plant in the Monocotyledoneae. *American Journal of Botany*, 29(6), 441-450.
- COCK, P. J. A.; ANTAO, T.; CHANG, J. T.; CHAPMAN, B. A.; COX, C. J.; DALKE, A.; FRIEDBERG, I.; HAMELRYCK, T.; KAUFF, F.; WILCZYNSKI, B. Y DE HOON, M. J. L. (2009). Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11), 1422-1423.

- COOPER, G. M. Y BROWN, C. D. (2008). Qualifying the relationship between sequence conservation and molecular function. *Genome Research*, 18(2), 201-205.
- D'HONT, A.; DENOEUDE, F.; AURY, J.-M.; BAURENS, F.-C.; CARREEL, F.; GARSMEUR, O.; NOEL, B.; BOCS, S.; DROC, G.; ROUARD, M.; DA SILVA, C.; JABBARI, K.; CARDI, C.; POULAIN, J.; SOUQUET, M.; LABADIE, K.; JOURDA, C.; LENGELLÉ, J.; RODIER-GOUD, M.; ... WINCKER, P. (2012). The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature*, 488(7410), 213-217.
- EDGAR, R. C. (2004a). Local homology recognition and distance measures in linear time using compressed amino acid alphabets. *Nucleic Acids Research*, 32(1), 380-385.
- EDGAR, R. C. (2004b). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792-1797.
- EDGAR, R. C. (2004c). MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, 5(1), 113.
- EFRON, B.; HALLORAN, E. Y HOLMES, S. (1996). Bootstrap confidence levels for phylogenetic trees. *Proceedings of the National Academy of Sciences*, 93(23), 13429.
- ENGSTRÖM, P. G.; SUI, S. J. H.; DRIVENES, Ø.; BECKER, T. S. Y LENHARD, B. (2007). Genomic regulatory blocks underlie extensive microsynteny conservation in insects. *Genome Research*, 17(12), 1898-1908.
- ERIKSSON, M. Y SPÅNGBERG, J. (2017). Carbon footprint and energy use of food waste management options for fresh fruit and vegetables from supermarkets. *Waste Management*, 60, 786-799.
- FELSENSTEIN, J. (1981). Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution*, 17(6), 368-376.
- FERNANDEZ-POZO, N.; MENDA, N.; EDWARDS, J. D.; SAHA, S.; TECLE, I. Y.; STRICKLER, S. R.; BOMBARELY, A.; FISHER-YORK, T.; PUJAR, A.; FOERSTER, H.; YAN, A. Y MUELLER, L. A. (2015). The Sol Genomics Network (SGN)—From genotype to phenotype to breeding. *Nucleic Acids Research*, 43(D1), D1036-D1041.
- FREELING, M. Y SUBRAMANIAM, S. (2009). Conserved noncoding sequences (CNSs) in higher plants. *Current Opinion in Plant Biology*, 12(2), 126-132.
- GHEDIN, E.; BRINGAUD, F.; PETERSON, J.; MYLER, P.; BERRIMAN, M.; IVENS, A.; ANDERSSON, B.; BONTEMPI, E.; EISEN, J.; ANGIUOLI, S.; WANLESS, D.; VON ARX, A.; MURPHY, L.; LENNARD, N.; SALZBERG, S.; ADAMS, M. D.; WHITE, O.; HALL, N.; STUART, K.; ... EL-SAYED, N. M. A. (2004). Gene synteny and evolution of genome architecture in trypanosomatids. *Molecular and Biochemical Parasitology*, 134(2), 183-191.
- GHIURCUTA, C. G. Y MORET, B. M. E. (2014). Evaluating synteny for improved comparative studies. *Bioinformatics*, 30(12), i9-i18.
- GIOVANNONI, J. (2001). Molecular Biology of Fruit Maturation and Ripening. *Annual Review of Plant Physiology and Plant Molecular Biology*, 52(1), 725-749.
- GRBIĆ, V. Y BLEECKER, A. B. (1995). Ethylene regulates the timing of leaf senescence in Arabidopsis. *The Plant Journal*, 8(4), 595-602.
- HOCHHOLDINGER, F.; WEN, T.-J.; ZIMMERMANN, R.; CHIMOT-MAROLLE, P.; SILVA, O. D. C. E, BRUCE, W.; LAMKEY, K. R.; WIENAND, U. Y SCHNABLE, P. S. (2008). The maize (*Zea mays* L.) roothairless3 gene encodes a putative GPI-anchored, monocot-specific, COBRA-like protein that significantly affects grain yield. *The Plant Journal*, 54(5), 888-898.

- HOEBERICHTS, F. A.; VAN DER PLAS, L. H. W. Y WOLTERING, E. J. (2002). Ethylene perception is required for the expression of tomato ripening-related genes and associated physiological changes even at advanced stages of ripening. *Postharvest Biology and Technology*, 26(2), 125-133.
- HUNG, T. Y CHANG, H. Y. (2010). Long noncoding RNA in genome regulation. *RNA Biology*, 7(5), 582-585.
- JAIN, M.; NIJHAWAN, A.; ARORA, R.; AGARWAL, P.; RAY, S.; SHARMA, P.; KAPOOR, S.; TYAGI, A. K. Y KHURANA, J. P. (2007). F-Box Proteins in Rice. Genome-Wide Analysis, Classification, Temporal and Spatial Gene Expression during Panicle and Seed Development, and Regulation by Light and Abiotic Stress. *Plant Physiology*, 143(4), 1467-1483.
- JOHANSSON, P.; LIPOVICH, L.; GRANDÉR, D. Y MORRIS, K. V. (2014). Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1840(3), 1063-1071.
- KAWAGUCHI, R. Y BAILEY-SERRES, J. (2005). mRNA sequence features that contribute to translational regulation in Arabidopsis. *Nucleic Acids Research*, 33(3), 955-965.
- KIM, P. M.; LU, L. J.; XIA, Y. Y GERSTEIN, M. B. (2006). Relating Three-Dimensional Structures to Protein Networks Provides Evolutionary Insights. *Science*, 314(5807), 1938-1941.
- KIMURA, M. (1989). The neutral theory of molecular evolution and the world view of the neutralists. *Genome*, 31(1), 24-31.
- KONISHI, M. Y YANAGISAWA, S. (2008). Ethylene signaling in Arabidopsis involves feedback regulation via the elaborate control of EBF2 expression by EIN3. *The Plant Journal*, 55(5), 821-831.
- KUERSTEN, S. Y GOODWIN, E. B. (2003). The power of the 3' UTR: Translational control and development. *Nature Reviews Genetics*, 4(8), 626-637.
- KUMAR, S.; STECHER, G.; LI, M.; KNYAZ, C. Y TAMURA, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, 35(6), 1547-1549.
- LEPPEK, K.; DAS, R. Y BARNA, M. (2018). Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nature Reviews Molecular Cell Biology*, 19(3), 158-174.
- LI, P. Y BRUTNELL, T. P. (2011). *Setaria viridis* and *Setaria italica*, model genetic systems for the Panicoid grasses. *Journal of Experimental Botany*, 62(9), 3031-3037.
- LORENZ, R.; BERNHART, S. H.; HÖNER ZU SIEDERDISSEN, C.; TAFER, H.; FLAMM, C.; STADLER, P. F. Y HOFACKER, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, 6(1), 26.
- MASSA, A. N.; CHILDS, K. L.; LIN, H.; BRYAN, G. J.; GIULIANO, G. Y BUELL, C. R. (2011). The Transcriptome of the Reference Potato Genome *Solanum tuberosum* Group Phureja Clone DM1-3 516R44. *Plos one*, 6(10), e26801.
- MATAS, A. J.; YEATS, T. H.; BUDA, G. J.; ZHENG, Y.; CHATTERJEE, S.; TOHGE, T.; PONNALA, L.; ADATO, A.; AHARONI, A.; STARK, R.; FERNIE, A. R.; FEI, Z.; GIOVANNONI, J. J. Y ROSE, J. K. C. (2011). Tissue- and Cell-Type Specific Transcriptome Profiling of Expanding Tomato Fruit Provides Insights into Metabolic and Regulatory Specialization and Cuticle Formation. *The Plant Cell*, 23(11), 3893-3910.
- MAZUMDER, B.; SESHADRI, V. Y FOX, P. L. (2003). Translational control by the 3'-UTR: The ends specify the means. *Trends in Biochemical Sciences*, 28(2), 91-98.

- MERCHANTE, C.; BRUMOS, J.; YUN, J.; HU, Q.; SPENCER, K. R.; ENRÍQUEZ, P.; BINDER, B. M.; HEBER, S.; STEPANOVA, A. N. Y ALONSO, J. M. (2015). Gene-Specific Translation Regulation Mediated by the Hormone-Signaling Molecule EIN2. *Cell*, 163(3), 684-697.
- MOOR, C. H. DE, MEIJER, H. Y LISSENDEN, S. (2005). Mechanisms of translational control by the 3' UTR in development and differentiation. *Seminars in Cell y Developmental Biology*, 16(1), 49-58.
- MÜLLER, M. Y MUNNÉ-BOSCH, S. (2015). Ethylene Response Factors: A Key Regulatory Hub in Hormone and Stress Signaling. *Plant Physiology*, 169(1), 32-41.
- NASCARRAYS AT THE BAR. (s. f.), visto el 4 de julio de 2021 <http://www.bar.utoronto.ca/NASCArrays/index.php?ExpID=172>
- NCBI MULTIPLE SEQUENCE ALIGNMENT VIEWER 1.20.0. (s. f.). visto el 4 de julio de 2021 https://www.ncbi.nlm.nih.gov/projects/msaviewer/?appname=ncbi_msav&openuploadialog
- PATTENGALE, N. D.; ALIPOUR, M.; BININDA-EMONDS, O. R. P.; MORET, B. M. E. Y STAMATAKIS, A. (2009). How Many Bootstrap Replicates Are Necessary? *Journal of computational biology*, 17(3), 337-354.
- POTUSCHAK, T.; LECHNER, E.; PARMENTIER, Y.; YANAGISAWA, S.; GRAVA, S.; KONCZ, C. Y GENSCHIK, P. (2003). EIN3-Dependent Regulation of Plant Ethylene Hormone Signaling by Two Arabidopsis F Box Proteins: EBF1 and EBF2. *Cell*, 115(6), 679-689.
- PREISS, T. Y HENTZE, M. W. (1999). From factors to mechanisms: Translation and translational control in eukaryotes. *Current Opinion in Genetics y Development*, 9(5), 515-521.
- PRICE, M. N.; DEHAL, P. S. Y ARKIN, A. P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *Plos one*, 5(3), e9490.
- REUTER, J. S. Y MATHEWS, D. H. (2010). RNAstructure: Software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, 11(1), 129.
- REZAEI, M. Y LIU, B. (2017). Food loss and waste in the food supply chain. *Nutfruit*, 71, 26-27.
- SAITOU, N. Y NEI, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4), 406-425.
- SAYERS, E. W.; CAVANAUGH, M.; CLARK, K.; OSTELL, J.; PRUITT, K. D. Y KARSCH-MIZRACHI, I. (2019). GenBank. *Nucleic Acids Research*, 47(D1), D94-D99.
- SCHMID, M.; DAVISON, T. S.; HENZ, S. R.; PAPE, U. J.; DEMAR, M.; VINGRON, M.; SCHÖLKOPF, B.; WEIGEL, D. Y LOHMANN, J. U. (2005). A gene expression map of Arabidopsis thaliana development. *Nature Genetics*, 37(5), 501-506.
- SHEN, Y.; JI, G.; HAAS, B. J.; WU, X.; ZHENG, J.; REESE, G. J. Y LI, Q. Q. (2008). Genome level analysis of rice mRNA 3'-end processing signals and alternative polyadenylation. *Nucleic Acids Research*, 36(9), 3150-3161.
- SUN, Y.; LI, J. Q.; YAN, J. Y.; YUAN, J. J.; LI, G. X.; WU, Y. R.; XU, J. M.; HUANG, R. F.; HARBERD, N. P.; DING, Z. J. Y ZHENG, S. J. (2020). Ethylene promotes seed iron storage during Arabidopsis seed maturation via ERF95 transcription factor. *Journal of Integrative Plant Biology*, 62(8), 1193-1212.
- TAIR (s. f.). visto el 4 de julio de 2021 <https://www.arabidopsis.org/>
- TAKEZAKI, N. (1998). Tie trees generated by distance methods of phylogenetic reconstruction. *Molecular Biology and Evolution*, 15(6), 727-737.

- TALIAFERRO, J. M.; LAMBERT, N. J.; SUDMANT, P. H.; DOMINGUEZ, D.; MERKIN, J. J.; ALEXIS, M. S.; BAZILE, C. A. Y BURGE, C. B. (2016). RNA Sequence Context Effects Measured In Vitro Predict In Vivo Protein Binding and Regulation. *Molecular Cell*, 64(2), 294-306.
- TAMURA, K. (1994). Model selection in the estimation of the number of nucleotide substitutions. *Molecular biology and evolution*, 11(1), 154-157.
- THE WORLD POPULATION PROSPECTS: 2015 Revision. visto el 4 de julio de 2021. [Report]. <https://www.un.org/en/development/desa/publications/world-population-prospects-2015-revision.html>
- VAN BEL, M.; DIELS, T.; VANCAESTER, E.; KREFT, L.; BOTZKI, A.; VAN DE PEER, Y.; COPPENS, F. Y VANDEPOELE, K. (2018). PLAZA 4.0: An integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Research*, 46(D1), D1190-D1196.
- VANDEPOELE, K.; BEL, M. V.; RICHARD, G.; LANDEGHEM, S. V.; VERHELST, B.; MOREAU, H.; PEER, Y. V. DE, GRIMSLEY, N. Y PIGANEAU, G. (2013). Pico-PLAZA, a genome database of microbial photosynthetic eukaryotes. *Environmental Microbiology*, 15(8), 2147-2153.
- VANDIVIER, L. E.; ANDERSON, S. J.; FOLEY, S. W. Y GREGORY, B. D. (2016). The Conservation and Function of RNA Secondary Structure in Plants. *Annual Review of Plant Biology*, 67(1), 463-488.
- VAUGHN, J. N.; ELLINGSON, S. R.; MIGNONE, F. Y ARNIM, A. VON. (2012). Known and novel post-transcriptional regulatory sequences are conserved across plant families. *RNA*, 18(3), 368-384.
- WEBER, C.; NOVER, L. Y FAUTH, M. (2008). Plant stress granules and mRNA processing bodies are distinct from heat stress granules. *The Plant Journal*, 56(4), 517-530.
- WILKIE, G. S.; DICKSON, K. S. Y GRAY, N. K. (2003). Regulation of mRNA translation by 5'- and 3'-UTR-binding factors. *Trends in Biochemical Sciences*, 28(4), 182-188.
- WINTER, D.; VINEGAR, B.; NAHAL, H.; AMMAR, R.; WILSON, G. V. Y PROVART, N. J. (2007). An "Electronic Fluorescent Pictograph" Browser for Exploring and Analyzing Large-Scale Biological Data Sets. *Plos one*, 2(8), e718.