



Facultad de Ciencias Bioquímicas y Farmacéuticas - Universidad Nacional
de Rosario

Tesis de Doctorado

**Estudios sobre la regulación de la expresión génica por
microARNs en plantas mediante estrategias
bioinformáticas**

Presentada por: Uciel Pablo Chorostecki

Rosario, Argentina

2016

**Estudios sobre la regulación de la expresión génica por
microARNs en plantas mediante estrategias
bioinformáticas**

Uciel Pablo Chorostecki

Licenciado en Ciencias de la Computación

Universidad Nacional de Rosario

Esta Tesis es presentada como parte de los requisitos para optar al grado académico de Doctor en Ciencias Biológicas, de la Universidad de Rosario y no ha sido presentada previamente para la obtención de otro título en esta u otra Universidad. La misma contiene los resultados obtenidos en investigaciones llevadas a cabo en el Instituto de Biología Molecular y Celular de Rosario (IBR-CONICET), dependiente de la Facultad de Cs. Bioquímicas y Farmacéuticas, durante el período comprendido entre Abril de 2012 y Marzo de 2016, bajo la dirección del Dr. Javier Palatnik.

Parte de los resultados de esta tesis fueron publicados en los siguientes artículos:

- Mateos JL, Bologna NG, Chorostecki U, Palatnik JF. Curr Biol. 2010 Jan 12;20(1):49-54. doi: 10.1016/j.cub.2009.10.072. Epub 2009 Dec 10.
- Chorostecki U, Crosa VA, Lodeyro AF, Bologna NG, Martin AP, Carrillo N, Schommer C, Palatnik JF. Nucleic Acids Res. 2012 Oct;40(18):8893-904. doi: 10.1093/nar/gks625. Epub 2012 Jul 5.
- Bologna NG, Schapire AL, Zhai J, Chorostecki U, Boisbouvier J, Meyers BC, Palatnik JF. Genome Res. 2013 Oct;23(10):1675-89. doi: 10.1101/gr.153387.112. Epub 2013 Aug 29.
- Chorostecki U, Palatnik JF. Bioinformatics. 2014 Jul 15;30(14):2066-7. doi: 10.1093/bioinformatics/btu147. Epub 2014 Mar 14.

I would like to dedicate this thesis to ...

Índice general

Índice de figuras	xi
Índice de tablas	xiii
Introducción	1
1.1 Silenciamiento génico y ARN pequeños en plantas	1
1.1.1 Variedades de ARN pequeños en plantas	1
1.2 miARNs en plantas	3
1.3 Biogénesis de miARNs	4
1.3.1 Procesamiento de miARNs en animales	4
1.3.2 Procesamiento de miARNs en plantas	5
1.3.3 Precursores en plantas	7
1.4 Regulación de la expresión génica por miARNs	7
1.4.1 Regulación por corte del ARN blanco	7
1.4.2 Regulación de la traducción por miARNs	7
1.4.3 Generación de ta-siRNAs y siRNAs secundarios	9
1.5 Procesos biológicos regulados por miARNs en plantas	9
1.6 Predicción de genes blanco de miARNs	10
1.6.1 Métodos para predicción de genes blanco de miARNs en animales .	10
1.6.2 Métodos para predicción de genes blanco de miARNs en plantas .	11
1.6.3 identificación de genes blanco regulados por miARNs mediante técnicas de secuenciación de alto rendimiento	12
Objetivos	15
2.1 Objetivo general	15
2.2 Objetivos específicos	15
Materiales y Métodos	17
3.1 Obtención de <i>Arabidopsis</i> transformadas	17

3.1.1	Ecotipos utilizados y condiciones de crecimiento	17
3.1.2	Transformación de <i>Arabidopsis</i>	17
3.1.3	Recolección y esterilización de semillas	18
3.1.4	Selección de transformantes	18
3.1.5	Transformación nuclear de tabaco	18
3.2	Cuantificación del nivel de expresión génica	19
3.2.1	Extracción de ARN	19
3.2.2	Cuantificación y chequeo de la integridad del ARN purificado	20
3.2.3	Tratamiento del ARN preparado con ADNasa	20
3.2.4	Retrotranscripción (RT)	20
3.2.5	Reacción en cadena de la polimerasa en tiempo real	21
3.3	Mapeo del sitio de corte	22
3.4	Métodos utilizados para la predicción de genes regulados por miARNs en plantas	24
3.4.1	MiARN consensos	24
3.4.2	Predicción de genes regulados por miARNs	25
3.5	ComTAR: una herramienta para la predicción de genes blanco regulados por miARNs en plantas	26
3.5.1	MiARN y transcriptos	26
3.5.2	Búsqueda de genes blanco	27
3.5.3	Herramienta web y almacenamiento de datos	27
3.6	Protocolo de SPARE multiplex	28
3.7	Protocolo de PARE	33
3.8	Protocolo de ARN pequeños	33
3.9	Estudios genómicos sobre la biogénesis de precursores de miARN en plantas	36
3.9.1	Procesamiento de precursores de miARNs en plantas	36
3.10	Estrategia bioinformática para caracterizar la relación entre la evolución de los precursores de miARNs en plantas y los mecanismos de procesamientos	36
Capítulo 1		41
4.1	Introducción	41
4.2	Resultados y Discusión	42
4.2.1	Predicción de genes regulados por miARNs.	42
4.2.2	comTAR: una herramienta para la predicción de genes blanco regulados por miARNs en plantas.	56

Capítulo 2	61
5.1 Introducción	61
5.2 Resultados y Discusión	63
5.2.1 Procesamiento de precursores de miARNs en plantas	63
5.2.2 Identificación de intermediarios de procesamiento en plantas mutantes en proteínas de procesamiento	69
Capítulo 3	79
6.1 Introducción	79
6.2 Resultados y Discusión	79
6.2.1 Estrategia	80
6.2.2 Búsqueda de ortólogos de precursores de miARNs	80
6.2.3 Alineamiento de los precursores en base a su secuencia primaria . .	81
6.2.4 Mecanismo de base a loop cortos	88
6.2.5 Mecanismo de base a loop secuenciales	88
6.2.6 Mecanismo de loop a base cortos	90
6.2.7 Mecanismo de loop a base secuenciales	93
6.2.8 Procesamiento mixto de miembros de la familia del miR170/miR171	93
6.2.9 Mutaciones puntuales afectan el procesamiento de miARNs en plantas	96
6.2.10 Precursores de plantas conservados en monocotiledóneas	99
6.2.11 Precursores de animales	101
Capítulo 4	107
7.1 Introducción	107
7.2 Resultados y Discusión	108
7.2.1 Biogénesis y actividad de ARN pequeños de plantas en distintas temperaturas	108
7.2.2 Pipeline automatizado para el análisis genómico de ARN pequeños y su actividad <i>in vivo</i> en plantas	111
Conclusiones	115
8.1 Aplicaciones bioinformáticas para el estudio de interacciones miARN-gen blanco	115
8.2 Estudios genómicos sobre la biogénesis de miARNs	116
Bibliografía	117
Anexo	129

Índice de figuras

1.1	Familias de miARNs conservados en plantas	3
1.2	Biogénesis de miARNs en animales	6
1.3	Biogénesis y actividad de los miARNs en plantas	8
3.1	Diagrama de la estrategia utilizada para la identificación de los productos de corte de miARNs	23
3.2	comTAR. Diagrama de flujo	27
3.3	Esquema general para la construcción de bibliotecas de SPARE	28
3.4	Esquema general para la construcción de bibliotecas de PARE	34
3.5	Esquema general para la construcción de bibliotecas de ARN pequeños	35
3.6	Especies almacenadas en Phytozome	37
4.1	Variabilidad de la secuencia del maduro	43
4.2	Estrategia	45
4.3	Conservación de potenciales genes blanco en distintas especies	48
4.4	Selección de genes blanco por conservación evolutiva de la secuencia	49
4.5	Nuevos genes blanco validados en <i>A. thaliana</i>	53
4.6	Identificación de un nuevo gen blanco de miARN	55
4.7	Identificación de genes blanco de miARN, específicos de <i>Solanaceae</i>	57
4.8	Resultados del comTAR para el miR398	59
4.9	Salida del comTAR	59
5.1	Tamaño de precursores	62
5.2	Esquema del procedimiento para analizar los datos de SPARE	64
5.3	Identificación y caracterización de precursores de miARNs procesados de base a loop	65
5.4	Identificación y caracterización de precursores de miARNs procesados de loop a base	66
5.5	Estructura secundaria de precursores de base a loop	67
5.6	Estructura secundaria de precursores de loop a base	67
5.7	Mecanismos de procesamiento	68

5.8	Técnica de SPARE	69
5.9	Análisis informático utilizado para procesar los datos de SPARE	71
5.10	Captura de pantalla de la herramienta de SPARE para el miR165a	73
5.11	Captura de pantalla de la herramienta de SPARE para el miR169d	74
5.12	Captura de pantalla de la herramienta de SPARE para el miR156a	76
5.13	Captura de pantalla de la herramienta de SPARE para el miR159b	77
6.1	Especies detectadas	81
6.2	Alineamiento del precursor del miR172a.	83
6.3	Estructura secundaria del miR172a en distintas especies	84
6.4	Circos del miR172a	86
6.5	MEME del miR172a	87
6.6	Circos del miR172a	89
6.7	Circos de precursores con mecanismos de procesamientos secuenciales de base a loop	91
6.8	Circos de precursores con mecanismos de procesamientos cortos de loop a base	92
6.9	Circos de precursores con mecanismos de procesamientos secuenciales de loop a base	94
6.10	Procesamiento mixto de miembros de la familia del miR170/miR171	95
6.11	Apareamiento de bases en el precursor del miR164a	96
6.12	Estructura secundaria predicha del precursor del miR164a	97
6.13	Circos del miR164a	98
6.14	Apareamiento de bases en el precursor del miR394b	98
6.15	Circos del miR394b	99
6.16	Circos del miR172a	100
6.17	Tamaños de precursores de miARNs	101
6.18	Circos del miARN hsa-let-7a-1	103
6.19	Tamaño del loop terminal en precursores de distintos reinos	104
6.20	Circos de animales vs plantas	106
7.1	Distribución del tamaño de las lecturas totales de ARN pequeños	108
7.2	Distribución del tamaño de las lecturas de ARN pequeños que mapean contra el genoma de <i>A. thaliana</i>	109
7.3	Distribución de ARN pequeños en el genoma de <i>A. thaliana</i>	110
7.4	Diagrama de Venn de ARN pequeños más abundantes a diferente temperaturas	111
7.5	Estrategia para el análisis de bibliotecas de ARN pequeños de <i>A. thaliana</i> .	112
7.6	ARNm blanco con señal de corte en PARE a distintas temperaturas	113

Índice de tablas

3.1	Programa de termociclado	22
3.2	Oligos utilizados para la RT-qPCR	22
3.3	Oligonucleótidos utilizados para 5' RACE	24
4.4	miARNs y sus genes blanco en plantas	46
4.5	Detection of miRNA targets using different filters	50
6.6	Especies de <i>Metazoa</i> utilizadas	102
7.7	Bibliotecas de ARN pequeños	109
9.8	Especies y base de datos utilizadas para la búsquedas de genes blanco de miARNs conservados	132
9.9	Genes blanco validados de familias miARNs conservados en plantas	133
9.10	Abundancia de miARN en bibliotecas de ARN pequeños	134
9.11	Contiunuación: Abundancia de miARN en bibliotecas de ARN pequeños . .	135

Resumen

Los microARNs (o miARNs) son ARN no codificantes que regulan la expresión génica en animales y plantas y están implicados en procesos biológicos muy variables, como el desarrollo, la diferenciación y el metabolismo. Con un largo de aproximadamente 21 nucleótidos, los miARNs reconocen secuencias parcialmente complementarias en los ARNm blanco, provocando su corte o arresto de la traducción. Los miARNs han saltado rápidamente a la primera plana del interés de la comunidad científica como un nuevo nivel en el control de la expresión génica en eucariotas. Estudios recientes han puesto de manifiesto que los miARNs están estrechamente involucrados en distintas enfermedades de importancia. Los cálculos actuales consideran que cerca del 40% de los genes de humanos se encuentran regulados por miARNs.

Está generalmente aceptado que los miARNs en plantas tienen una extensiva complejidad con sus genes blanco y su predicción por lo general se basa en el uso de parámetros empíricos deducidos de interacciones conocidas del par miARN-gen blanco. En este trabajo, primero desarrollamos una estrategia para la identificación de genes blanco regulados por miARNs en plantas, basado en la conservación evolutiva del par miARN-gen blanco. Además, pudimos encontrar genes blanco específicos de Solanaceae y demostrar que la estrategia se puede utilizar para la búsqueda de genes blanco pertenecientes a un grupo determinado de especies.

A partir de estos resultados, desarrollamos una herramienta bioinformática para identificar genes blanco de miARNs basada principalmente en la conservación durante la evolución de la interacción del par miARN-gen blanco en distintas especies. Esta herramienta fue usada para predecir nuevas interacciones y validar experimentalmente genes blanco no conocidos anteriormente en *Arabidopsis thaliana*. Algunos de ellos podrían participar de las mismas vías que genes blanco conocidos anteriormente, sugiriendo que algunos miARNs pueden controlar diferentes aspectos de un proceso biológico.

La biogénesis de los miARNs es un proceso clave porque determina la secuencia exacta de nucleótidos del ARN pequeño funcional. Poco se sabía sobre el reconocimiento de los precursores de plantas por la maquinaria de procesamiento. En la segunda parte de este trabajo presentamos una estrategia para estudiar aspectos mecanísticos de la biogénesis de

los miARNs en plantas. Tratamos de dilucidar la dirección de procesamiento en precursores de miARNs en *A. thaliana* a partir de los patrones de evolución de los precursores.

Introducción

1.1 Silenciamiento génico y ARN pequeños en plantas

El control de la expresión génica es un proceso vital para los organismos unicelulares. Este proceso se utiliza para que las células se ajusten a los cambios en su ambiente de manera de optimizar el desarrollo, la respuesta celular al estrés y al entorno, entre muchos otros procesos. En plantas, la generación de los numerosos tipos celulares diferentes que forman en conjunto un organismo multicelular depende de que los genes se activen en las células adecuadas y en los momentos precisos del desarrollo. Muchos de los mecanismos regulatorios operan a nivel genético generando un control a nivel transcripcional y post-transcripcional.

En los últimos años se ha descubierto que pequeñas moléculas de 20 a 25 nucleótidos de longitud, son reguladores críticos de la expresión génica en eucariotas. En plantas existen distintos tipos de ARN pequeños, que por diferencias en su biogénesis y modos de acción, han sido clasificados en distintas clases. Estos ARN pequeños comparten algunos elementos entre ellos, pero su función y biogénesis son distintas.

1.1.1 Variedades de ARN pequeños en plantas

Los ARN pequeños provienen del procesamiento de regiones helicoidales de precursores de ARN. Los mismos se pueden clasificar en dos grupos principales. Los derivados de precursores de doble hebra (ARNdh) formados por hibridación intermolecular de dos cadenas de ARN complementarias y los derivados de precursores de RNA de una sola hebra (ARNsh) que forman una estructura de tallo-burbuja.

Los ARN pequeños más abundantes, derivados de precursores de ARNdh, son conocidos como siARNs (small interfering RNAs). Pueden ser de origen endógeno o exógeno. A su vez, los siARN se pueden clasificar en dos grupos, aquellos que producen un silenciamiento a nivel transcripcional (TGS, del inglés: Transcriptional Gene Silencing) [28, 143] y aquellos que lo producen a nivel post-transcripcional (PTGS, del inglés: Post Transcriptional Gene Silencing)[57]. El primer grupo se origina a partir de ciertas regiones genómicas como

transposones u otras secuencias repetitivas que son transcriptos por la ARN polimerasa IV. Estos ARNsh se convierten en ARNdh por acción de una RDR. A continuación los ARNdh son procesados por DCL3 para finalmente liberar siARN de 24 nt de longitud. El segundo grupo, los PTGS son originados a partir de ARNs de virus o ARNm proveniente de un transgén. Por acción de una RDR se convierten en ARNdh pero en este caso es procesado pro DCL2 o DCL4 para finalmente librer siARN de 21-22 nt que serán encargados del silenciamiento post-transcripcional [148]. Los roles biológicos del silenciamiento mediado por siARNs incluyen resistencia contra virus [142], protección del genoma contra elementos móviles de ADN y roles regulatorios sobre genes endógenos [17, 28, 143].

Existen otros ARNdhs menos abundantes que han sido descubiertos posteriormente. Este grupo está constituido por un lado por los ta-siARN (del inglés, trans-acting siARNs) y por el otro lado los nat-siARNs y nat-miARNs (del inglés, natural antisense siRNAs y miRNAs) [28, 109, 137, 143].

Los genes que darán lugar a ta-siRNAs son transcriptos que generalmente son denominados TAS. Este transcripto luego es identificado y cortado por acción de un miARN que posee una secuencia complementaria al mismo [5]. Los fragmentos de corte de un TAS se estabilizan y por la acción de ARN polimerasas dependientes de ARN, se convierten en un ARNdhs que es procesado por DCL4, generando ARNs pequeños de 21 nucleótidos. Finalmente algunos ta-siARNs se incorporan a un complejo RISC y guían el corte de los transcriptos de otros genes de manera similar a la acción de un miARN [5, 48, 146, 154].

En contraste con los otros tipos de siARNs, que dependen de un RDR para sintetizar el precursor ARNdhs, se cree que los precursores de ARNdhs de NAT-siARNs surgen de la hibridación de transcriptos separados [23, 67]. Estos ARNs separados pueden ser complementarios debido a que fueron transcriptos por hebras opuestas en el mismo locus, y se denominan cis-NAT-siARNs. Como alternativa, los ARNs hibridizados pueden surgir de genes que no se superopnen, estos se denominan trans-NAT-siARNs. Sólo cis-NAT-siARNs han sido descritos en plantas; los trans-NAT-siARNs permanecen sólo como una posibilidad hipotética. El ARNdhs formado es procesado por DCL1 o DCL2 liberando siARN primarios que inducen el corte sobre el ARNm expresado constitutivamente. Posteriormente, una RDR sintetiza la hebra complementaria del ARNm cortado. Esto da lugar nuevamente a un ARNdhs el cual es procesado por DCL1 para producir una sucesión de siARN ubicados en fase de 21 nt cada uno denominados nat-siARN secundarios [23, 67].

La clase más abundante de ARN pequeños derivados de precursores de ARNsh son los miARNs. Tienen entre 20 y 22 nt de longitud y son originados de largos precursores con extensa estructura secundaria en forma tallo-burbuja. Estos precursores luego son procesados por acción de DCL1, liberando el ARN maduro. Actúan como reguladores negativos de

la expresión génica reprimiendo la expresión de genes endógenos principalmente a nivel post-transcripcional. Este regulación se puede dar a través de la inhibición de la traducción o del corte de los ARN mensajeros. Tanto en animales como en plantas, existen un gran número de miARNs conservados en distintas especies. En animales se han encontrado miARNs conservados en especies muy distantes como gusanos hasta humanos [106]. Y en plantas sucede lo mismo, donde existen miARNs conservados desde musgos hasta dicotiledóneas [12, 13, 35] (Figura 1.1). No se han encontrado miARNs conservados en animales y en plantas, lo que sugiere que han aparecido en forma independientes durante la evolución [35]. Si bien los miARNs en plantas y en animales comparten componentes básicos, su biogénesis y su modo de acción difiere entre ellos.

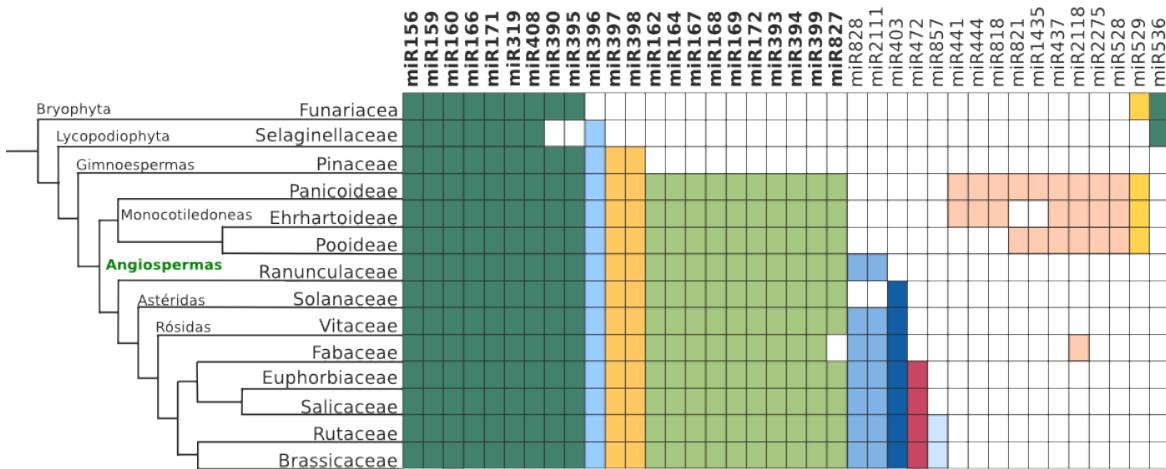


Figura 1.1 Familias de miARNs conservados en plantas. Se muestran las familias de miARNs (columnas) y para cada una de ella se muestra en qué especies están conservadas (filas). En diferentes colores se muestra el rango de especies en el que se encuentran conservadas, para distintos grupos de familias. En negrita se marcan las familias conservadas en angiospermas. MiR319 y miR159 que codifican para miARNs similares, fueron considerados como familias diferentes ya que regulan a genes blanco distintos [104]. Adaptado de [35].

1.2 miARNs en plantas

Los miARNs son generados a partir de loci endógenos, tanto en animales como en plantas. Controlan una gran variedad de procesos biológicos, como el desarrollo, la diferenciación, proliferación y respuesta a estrés [11, 27, 35, 124, 143]

Hasta hoy, en *A. thaliana* se han identificado más de 300 [70] miARNs. Se han utilizado distintos enfoques para identificar los miARNs: el clonado directo de ARN pequeños, mediante secuenciación de alto rendimiento y mediante estudios genéticos y mediante

predicciones bioinformáticas [35], siendo esta última la más común para la mayoría de las especies.

Los miARNs en plantas están codificados por familias de genes de 1 a 32 miembros que dan lugar a miARNs maduros idénticos o muy similares. Cada *locus* perteneciente a una familia codifica un miARN maduro idéntico o casi idéntico. Hasta el momento han sido definidas unas 42 familias de miARNs en plantas, las que regulan una amplia variedad de procesos biológicos. Doce de dichas familias tienen como blanco ARN mensajeros que codifican factores de transcripción involucrados en el desarrollo, mientras que otras están relacionadas con rutas de respuesta a señales ambientales y hormonales, entre otros, estando la mayoría de ellas conservadas entre mono y dicotiledóneas [64]. Muchos de estos pequeños ARNs han aparecido recientemente en la evolución y por lo tanto aparecen en un número pequeño de especies [11, 13]. Además está claro si tienen algún rol biológico [13, 35].

Sin embargo, existen 22 familias de miARNs que están altamente conservadas en las plantas, estando presentes en angiospermas, gimnospermas y algunas de ellas aún en plantas basales como los musgos [9, 11, 157] (Figura 1.1). Estos últimos miARNs cumplen funciones esenciales para la biología de las plantas [64].

1.3 Biogénesis de miARNs

1.3.1 Procesamiento de miARNs en animales

En animales, el procesamiento de los precursores es realizado en el núcleo y el citoplasma. Luego de la transcripción, el complejo formado por la ribonucleasa de tipo III, DROSHA, y la proteína DiGeorge syndrome critical region gene 8 (DGCR8) produce un primer corte sobre el pri-miARN para liberar una estructura de tallo y burbuja denominado pre-miARN (Figura 1.2) [40, 53, 58, 155].

A continuación, el pre-miARN es trasladado desde el núcleo hacia el citoplasma por la proteína de transporte nuclear EXPORTIN-5 (EXP5) [84, 153]. En el citoplasma DICER conjuntamente con TRPB, otra proteína de unión a ARNdh, produce un segundo corte sobre el precursor para que liberar el dúplex miARN/miARN*. Este segundo corte ocurre a ~21 nt de distancia del sitio donde se había producido el primer corte por DROSHA. Finalmente el miARN maduro es incorporado al complejo RISC [68] (Figura 1.2).

Además de esta ruta canónica de procesamiento, recientemente se han descripto vías alternativas. Para algunos precursores, el requerimiento de DROSHA puede ser suplantado por la maquinaria de corte y empalme (splicing) [101, 118]. En el caso de otros, es necesaria la interacción de proteínas específicas con el loop terminal [55, 136]. Por último, existen

casos en donde, el procesamiento de ciertos precursores puede estar específicamente inhibido en ciertos tejidos o determinadas condiciones [61, 100, 141].

1.3.2 Procesamiento de miARNs en plantas

Los miARNs se diferencian de otros ARNs pequeños por su particular biogénesis que implica su escisión de un precursor con extensa estructura secundaria localizado en un largo transcripto primario (Figura 1.3). En general, la biogénesis de estos ARN pequeños comienza con la transcripción por la ARN polimerasa II [146] a partir de unidades transcripcionales propias distribuidas en el genoma [112]. Los transcriptos primarios, llamados pri-miARNs, pueden tener varias kilobases de longitud y sufrir modificaciones post-transcripcionales como ser splicing, capping y poliadenilación. Estos transcriptos contienen precursores para miARNs con extensa estructura secundaria en forma de tallo-burbuja (stem-loop) [64] (Figura 1.3).

En animales, el procesamiento comienza en el núcleo por DROSHA y finaliza en el citoplasma por la acción de DICER. En plantas no existe un homólogo de DROSHA. Los precursores son procesados completamente en el núcleo a través de la acción de una ribonucleasa llamada DCL1 [112, 121] (del inglés DICER LIKE 1) en asociación con el cofactor proteico de unión a ARN de doble hebra HYL1 [60] (del inglés HYPONASTIC LEAVES 1) y la proteína SERRATE [82] (Figura 1.3).

Al parecer es la estructura secundaria por sobre la secuencia primaria del precursor la más importante en la determinación del correcto procesamiento del mismo [20]. El producto generado a partir de los cortes llevados a cabo por DCL1, es un dúplex miARN-miARN* que luego continúa siendo procesado por otros componentes enzimáticos hasta dar lugar al miARN maduro de 21 nt.

El paso final de la biogénesis de los miARN es la incorporación asimétrica, a partir del dúplex miARN-miARN*, del miARN maduro dentro de un complejo de silenciamiento. Este complejo se denomina RISC (del inglés RNAi Silencing Complex). El componente central de todos los complejos de silenciamiento es un miembro de la familia de proteínas ARGONAUTA (AGO). En *Arabidopsis* existen distintas proteínas AGO que participan en diferentes procesos biológicos [138] y la incorporación de los ARN pequeños en los distintos complejos depende de la identidad del nucleótido del extremo 5' y de la vía de biogénesis [92, 95, 133] (Figura 1.3).

En la mayoría de los miARNs el nucléotido extremo 5' es una U y en general la principal efectora de la actividad es AGO1 [92, 139, 143]. Complejos RISC similares se encuentran presentes en células animales. Más recientemente han sido identificadas proteínas adicionales que regularían la actividad de la maquinaria de procesamiento [20].

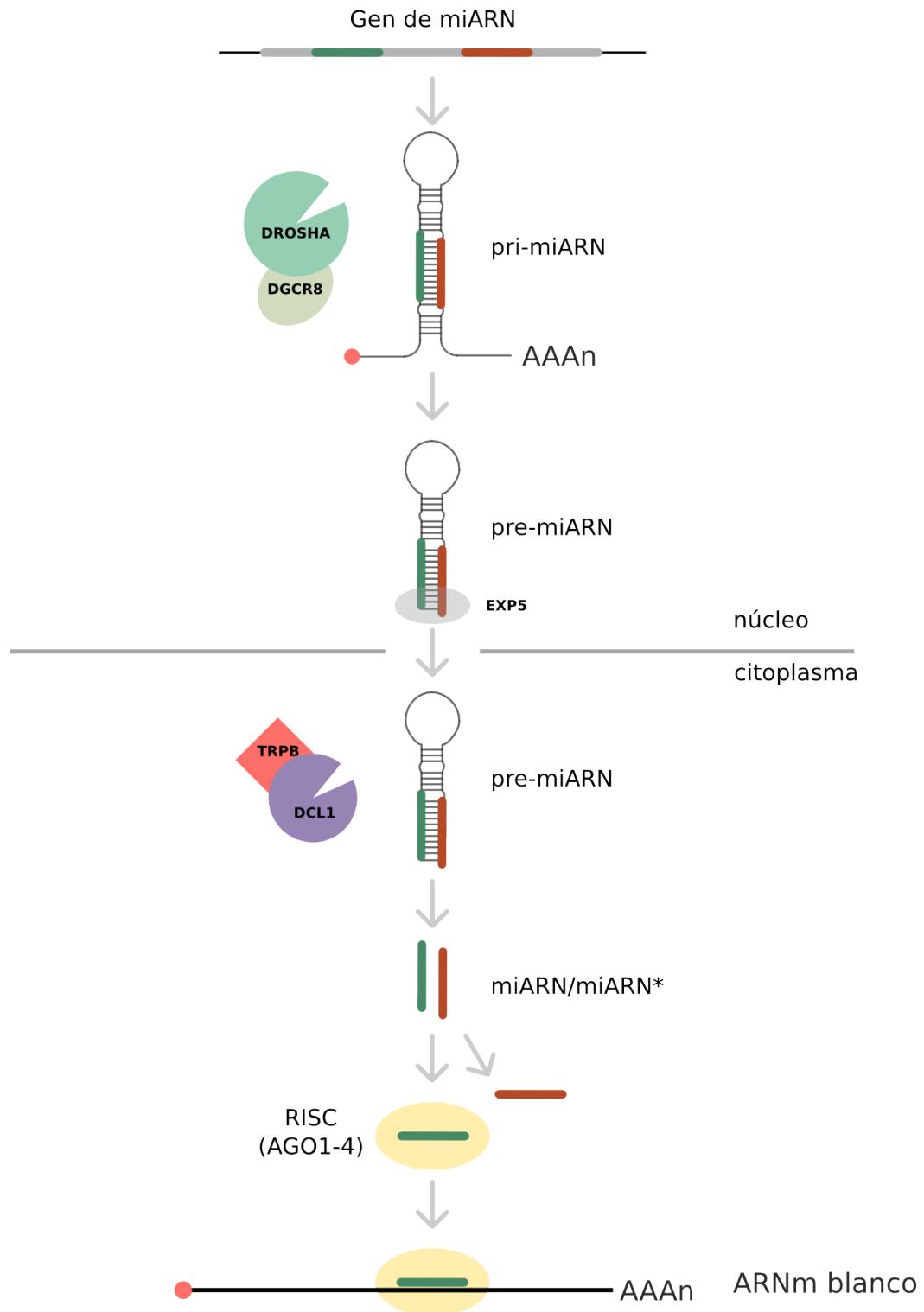


Figura 1.2 Mecanismo de procesamiento de miARNs en animales.

En animales, los miARNs reconocen principalmente a la región 3' no codificante de ARN mensajeros blanco inhibiendo su traducción. En plantas es más común que los miARNs se unan a secuencias complementarias en los ARNm blanco en la región codificante señalándolos para su degradación [64]. En cualquier caso, es el miARN el que proporciona la especificidad contra las moléculas de ARN blanco [16].

1.3.3 Precursores en plantas

1.4 Regulación de la expresión génica por miARNs

1.4.1 Regulación por corte del ARN blanco

En animales existe un gran número de genes blanco mediado por miARNs y un ARNm puede estar regulado por varios miARNs. En cambio los miARNs en plantas regulan un número limitado de genes blanco que además tienen un único sitio que es altamente complementario al miARN [143].

El ARN pequeño guía al complejo RISC hacia una molécula de ARNm complementario. Luego del reconocimiento de ARNm blanco por complementariedad de bases, la proteína AGO1 del complejo RISC introduce un corte en un enlace fosfodiester del ARNm. Este corte ocurre entre las posiciones 10 y 11 desde el extremo 5' del miARN, independientemente de la longitud del miARN [46, 81, 85, 103, 147].

Luego del corte mediado por el miARN, los fragmentos 3' son degradados mediante la actividad de la enzima citoplasmática 5'-3' EXORIBONUCLEASA4 (XRN4) en *A. thaliana* [127] Los fragmentos 5' también pueden ser degradados por un complejo denominado Exosoma, el cual está encargado de diferentes funciones de degradación y procesamiento de ARNs [29]. En *Arabidopsis* la degradación del fragmento 5', puede ser acelerada por uridilación en el extremo 3' por la enzima una enzima denomina HESO1 [113].

1.4.2 Regulación de la traducción por miARNs

Los miARNs en animales, en general, son parcialmente complementarios a uno o más sitios presentes en la región 3' no traducida de los ARNm blancos [27, 43, 77] y raramente sufren el tipo de corte antes mencionado. Además la limitada complementariedad de secuencia, permite que los miARNs de animales regulen generalmente la expresión de muchos genes blanco. Un mecanismo que involucra la inhibición de la traducción del ARNm blanco por el miARN explica la represión de la expresión de los blancos de miARNs en animales [43]. En otras ocasiones, los miARNs de animales disminuyen la vida media de los transcriptos

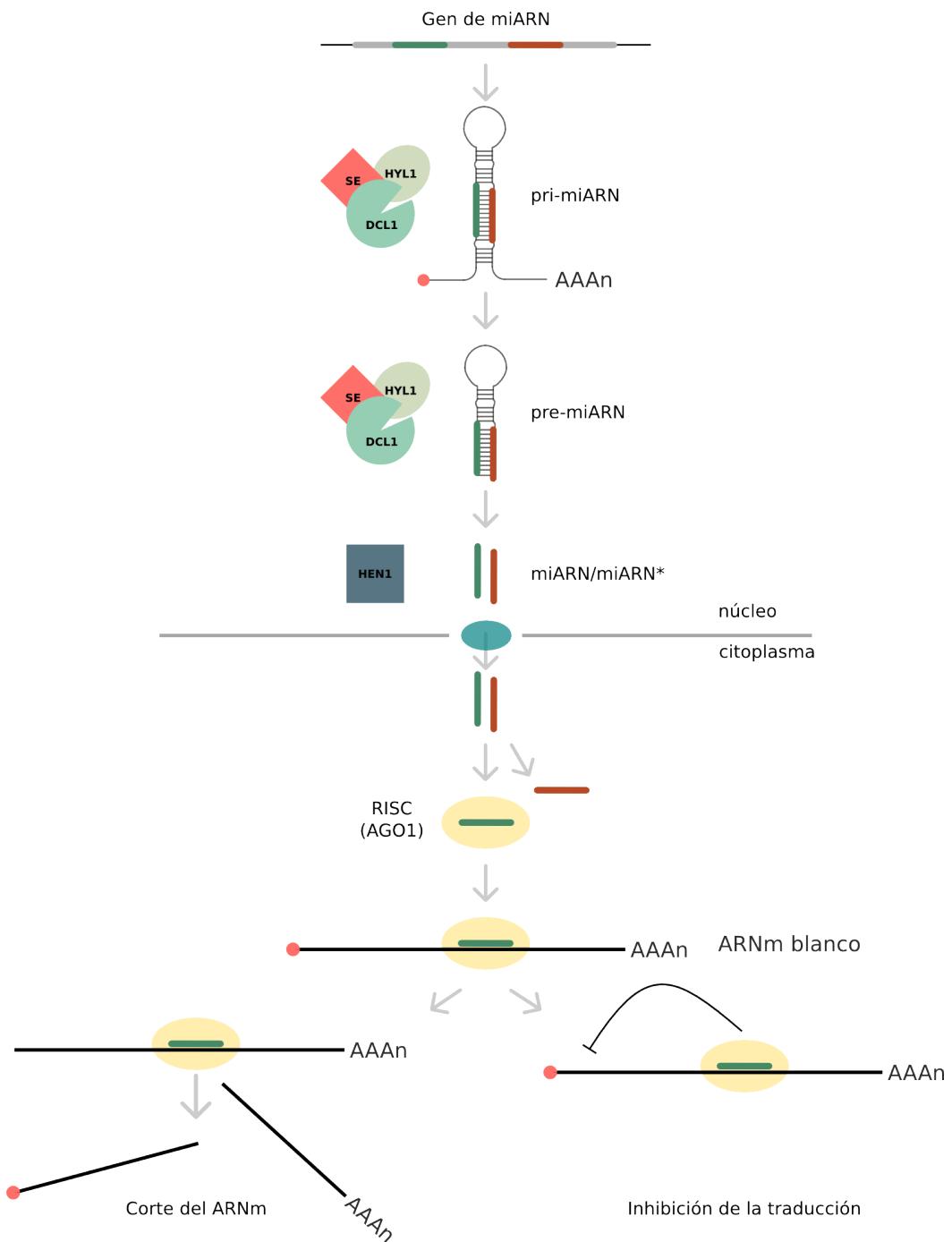


Figura 1.3 Los pri-miARNs son transcriptos por la ARN polimerasa II y tienen extensa estructura secundaria en forma de tallo-burbuja (stem-loop). Estas estructuras son procesados en el núcleo a través de la acción de una ribonucleasa llamada DICER LIKE 1 (DCL1) en asociación con el cofactor proteico de unión a ARN de doble hebra HYPONASTIC LEAVES 1 (HYL1) y la proteína SERRATE (SE). El producto generado a partir de los cortes llevados a cabo por DCL1, es un dúplex miARN-miARN* que luego continúa siendo procesado por otros componentes enzimáticos hasta dar lugar al miARN maduro de 21 nt. Luego el miARN maduro se asocia con una proteína Argonauta (AGO), mayormente AGO1 y se incorpora dentro de un complejo de silenciamiento RISC. En el complejo RISC, el miARN principalmente induce el corte de los ARNm blanco, aunque algunos miARNs también regulan la expresión de sus genes blancos por represión traduccional.

a los que se unen [56]. En algunos pocos miARN de plantas también se ha demostrado la existencia de un mecanismo de represión traduccional, además del corte del transcripto [10, 26, 31, 42, 75, 123].

1.4.3 Generación de ta-siRNAs y siRNAs secundarios

El corte de un transcripto por un miARN puede inducir la generación de ARN pequeños secundarios [5, 30, 34, 86, 96, 96]. Existe un mecanismo por el cual se genera un ARN pequeño específico en plantas que se denomina ta-siARN (trans-acting short-interfering RNA). Los genes que darán lugar a ta-siRNAs son transcriptos que generalmente son denominados TAS. El transcripto es identificado y cortado por acción de un miARN que posee una secuencia complementaria al transcripto [5]. Los fragmentos de corte de un TAS se estabilizan y por la acción de ARN polimerasas dependientes de ARN, se convierten en un ARN de doble hebra. Luego el ARN doble hebra es procesado por una Ribonucleasa de tipo III, llamada DICER- LIKE4 (DCL4), generando ARNs pequeños de 21 nucleótidos, los ta-siARNs. Finalmente estos ta-siARNs se incorporan a un complejo RISC y guían el corte de los transcriptos de otros genes de manera similar a la acción de un miARN [5, 48, 146, 154].

1.5 Procesos biológicos regulados por miARNs en plantas

Los procesos biológicos regulados por miARNs en plantas son muchos, y algunos miARNs desempeñan papeles importantes en el desarrollo, otros en la trasducción de señales hormonales, respuesta a estrés y respuesta a señales del ambiente [116, 117, 143].

De las 22 familias de miARNs conservados en plantas, 11 regulan factores de transcripción y la mayoría de ellos están involucrados en procesos de desarrollo o diferenciación celular [64] (Tabla 4.4). Por ejemplo los genes blanco del miR172 pertenecen a la familia de factores de transcripción AP2 (APETALA2) que está involucrado en el tiempo de floración y en el desarrollo de la hoja [10, 31]. La familia del miR319 regula factores de transcripción TCP (TEOSINTE BRANCHED1), la familia del miR165/166 factores de transcripción HD-ZIPIII (HOMEODOMAIN-LEUCINE ZIPPER de clase III) y la familia del miR396 regulan factores de transcripción GRF (GROWTH-REGULATING FACTOR). Todas estas familias intervienen en el desarrollo de la hoja regulando distintos factores de transcripción [25, 103, 115]. El miR164, participa en el establecimiento del meristema apical del tallo, en el desarrollo de la raíz y definición de los bordes de los órganos mediante la regulación de miembros de la familia de factores de transcripción NAC (NAM, ATAF1/2 and CUC2) [76].

Por otra parte, entre los genes blanco que no corresponden a factores de transcripción, hay genes involucrados en diversos aspectos de la biología vegetal. Algunos codifican proteínas

pertenecientes a la familia F-Box o enzimas E2 conjugantes de ubiquitina, las cuales están implicadas en la marcación de proteínas para la degradación por el proteosoma [116]. Otros genes con función conocida codifican para proteínas involucradas en la transducción de señales hormonales, o proteínas involucradas en el metabolismo energético, la respuesta a estrés o déficit de nutrientes.

Por otro lado, en otras especies que se han realizado estudios de ARN pequeños por técnicas de alto rendimiento, se han encontrado miARNs no conservados [35, 108, 115]. Los genes blanco regulados por estos miARNs tienen funciones más variables que los blanco de los miARNs conservados, de todos modos se ha demostrado el rol biológico de un pequeño número de estas interacciones [35].

En esta tesis nos enfocaremos principalmente en las familias de miARNs conservadas entre distintas especies de plantas (Tabla 4.4). Estas familias han sido estudiadas profundamente, por lo que su mayoría se encuentran ampliamente caracterizadas.

1.6 Predicción de genes blanco de miARNs

1.6.1 Métodos para predicción de genes blanco de miARNs en animales

Existen muchos métodos para la predicción de genes blanco de miARNs que fueron desarrollados por distintos grupos de investigación. Gran parte de ellos continúan manteniendo y actualizando sus algoritmos. El grupo de Cohen en el EMBL propuso el primer predictor de genes blanco de miARNs en 2003 [130] que luego fue actualizado en el 2005 [129]. TargetScan y TargetScanS fueron desarrollados por Bartel en el MIT y Burge en Cambridge [47, 54, 79, 80]. DIANA-microT, otra herramienta popular que fue creada por el grupo de Hatzigeorgiou, fue actualizada recientemente a la versión 5.0 [87, 88, 105]. El grupo de Rajewsky publicó su predictor de genes blanco en el 2005 y fue actualizado en el 2011 [8, 111].

Hay dos categorías de modelos predictivos: heurísticos y empíricos. Los modelos heurísticos utilizan algoritmos de detección que buscan posiciones a lo largo de la secuencia del ARNm y funciones de puntajes que filtran los blanco mediante la combinación de varios valores de entrada. Los primeros predictores aplican enfoques heurísticos, debido a la falta de una cantidad suficiente de datos para construir los modelos basados en el conocimiento empírico.

Los modelos predictivos utilizan entradas que se derivan del conocimiento de los detalles mecanísticos de las interacciones miARN-ARNm. La entrada más utilizada de los modelos

predictivos, es la complementariedad de base entre el miARN y ARNm. En contraste con el apareamiento de bases casi perfecto en las plantas [114], los miARNs de animales por lo general se unen al ARNm con sólo algunas posiciones que están apareadas [74]. La complementariedad del apareamiento de bases en la región semilla (seed), que comprende los ocho primeros nucleótidos en el extremo 5' del miARN, es particularmente importante. Los parámetros de entradas de accesibilidad del sitio y de conservación evolutiva se utilizan para aumentar la especificidad. La mayoría de los predictores usan softwares existentes, como el Vienna RNA package [83], mFold [162], DINAMelt [89] y mFold [41], para el cálculo de la energía libre. El uso de la conservación evolutiva de los genes blanco de miARNs está motivada por la premisa de que las especies 'similares' deben compartir miARNs comunes y genes blanco comunes. Sin embargo, esto conduce a la omisión de los genes blanco no conservados [78, 134].

1.6.2 Métodos para predicción de genes blanco de miARNs en plantas

Los algoritmos mencionados anteriormente son utilizados generalmente para predecir genes blanco de miARNs en animales. Algunos de ellos, como MiRanda, se utilizan tanto en animales como en plantas. Existen herramientas diseñadas específicamente para la predicción de genes blanco de miARNs en plantas. Algunas de ellas fueron implementadas como herramientas independientes, otras como servidores web y algunas como ambos.

La herramienta psRNATarget [37] utiliza un enfoque de programación dinámica, alineando las secuencias utilizando el algoritmo de Smith-Waterman modificado y aplicando el algoritmo 'RNAup'. Targetfinder [18] implementa un programa "FASTA" junto con un sistema de puntuación para "mismatches", "bulges" o "gaps" para alinear las secuencias. TAPIR [22] está integrado con dos opciones de búsqueda, el motor de búsqueda "FASTA" (para búsquedas rápidas), y el motor de búsqueda "RNA hybrid" (para obtener resultados precisos). La herramienta Target-align [145] también emplea el método de puntuación basado en el algoritmo de Smith-Waterman para predecir las complementariedades entre los miARNs y los RNAs mensajeros. Por último, la herramienta p-TAREF [62] implementa asupport vector regression (RVS) y utiliza la información de la "variación de la densidad de dinucleótido" alrededor del sitio target de conjunt de datos de *A. thaliana*, *Oryza sativa*, *Medicago truncatula* y *Solanum lycopersicum*.

Además existen otras herramientas que son utilizadas para predecir genes blanco de miARNs pero que no fueron específicamente diseñadas para esto. Por ejemplo, la herramienta Web miRNA designer, WMD3 [102] es utilizada para el diseño del miARNs artificiales para silenciar la expresión de genes blanco específicos. Pero esta herramienta también es empleada para la predicción de genes blanco de miARNs en plantas. RNAHybrid [71] es una

herramienta utilizada para buscar la energía mínima libre de un ARN largo y uno corto. La secuencia corta se hibrida con la parte más adecuada de la secuencia larga. La herramienta está destinada principalmente para la predicción de genes blanco de miARNs, pero no es su único uso.

Patmatch [150] que está integrado en "The Arabidopsis Information Resource" (TAIR)¹ permite realizar búsqueda de secuencias de nucleótidos o de péptidos corto, como dominios pequeños o motivos. Esta herramienta puede ser adaptada, para utilizarla como herramienta de búsqueda de genes blanco de miARNs en plantas.

La mayoría de las herramientas específicas de plantas fueron desarrolladas para predecir genes blanco con una alta especificidad en el organismo modelo *A. thaliana*, si se utilizan los parámetros de predicciones óptimos. Y los sistemas de puntajes optimizados de Arabidopsis no se pueden utilizar como un umbral en el análisis de organismos no modelos [128]. Además estas herramientas en general detectan una gran cantidad de potenciales genes blanco (que incluyen los validados experimentalmente), pero su debilidad radica en que predicen una gran número de falsos positivos.

1.6.3 identificación de genes blanco regulados por miARNs mediante técnicas de secuenciación de alto rendimiento

Con los avances de los métodos de secuenciación de alto rendimiento o "next-generation sequencing", la identificación sistemática de genes blanco de miARNs específicos en un tiempo corto ya es un hecho. Un método denominado PARE (Construction of Parallel Analysis of RNA Ends) [49]. La misma consiste en una combinación de la estrategia modificada 5'RACE, para la identificación de genes blanco de miARNs y permite realizar estudios del degradoma de ARN, mediante técnicas de alto rendimiento.

A partir de este método, se han desarrollado diferentes herramientas como SeqTar [160], que es un método para la identificación de corte guiado por miARNs de degradoma de transcriptos de plantas. Por otro lado, CleaveLand [3] es un pipeline computacional para la detección de cortes de miARNs en datos de degradoma. Esta herramienta toma como entrada secuencias de degradoma, ARNs pequeños y una base de datos de ARNm y devuelve los potenciales genes blanco de esos ARNs pequeños. sPARTA (small RNA-PARE target analyzer) [65] es un pipeline paralelizado, para el análisis integrado de miARNs en plantas y un conjunto de ARNm. Este pipeline incluye un software de identificación de nuevos genes blanco regulados por miARNs y utiliza datos de PARE.

¹<https://arabidopsis.org>

Por otro lado, la herramienta PAREsnip permite realizar búsqueda de potenciales genes blanco de todos los ARN pequeños obtenidos de técnicas de secuenciación de alto rendimiento. Mediante la búsqueda de genes blanco de todo el "sRNAome" se pude facilitar identificación de genes blanco de ARN pequeños a grande escala.

Objetivos

2.1 Objetivo general

Uno de los objetivos general de este trabajo de Tesis consiste en identificar a los genes regulados por miARNs y descubrir sus roles en plantas. Además, como objetivo queremos contribuir al conocimiento de la regulación del procesamiento de los miARNs en plantas. Se espera que los resultados de esta Tesis sirvan no solo para alcanzar los objetivos de investigación planteados sino también para promover el desarrollo de la Bioinformática como una disciplina que brinda una oportunidad única para que, a partir de investigaciones en las ciencias básicas, pueda hallarse el camino hacia el desarrollo de aplicaciones de interés estratégico para el país.

2.2 Objetivos específicos

1. Identificar genes regulados por miARNs en plantas.
 - Diseñar una estrategia para la identificación de genes blanco regulados por miARNs en plantas, basado en la conservación evolutiva del par miARN-gen blanco.
 - Desarrollar una herramienta web para la predicción de genes blanco de miARNs en diferentes especies de plantas.
2. Estudiar la biogénesis de los miARNs en plantas.
 - Identificación y caracterización de precursores de miARNs con mecanismos de procesamiento distintos.
 - Identificación de intermediarios de procesamiento en plantas mutantes en proteínas de procesamiento.
 - Caracterizar la relación entre la evolución de los precursores de miARNs en plantas y los mecanismos de procesamiento determinados previamente.

Materiales y Métodos

3.1 Obtención de *Arabidopsis* transformadas

3.1.1 Ecotipos utilizados y condiciones de crecimiento

Las plantas de *A. thaliana* utilizadas para los experimentos en esta parte del trabajo corresponden a el ecotipo Columbia-0 Col-0. Las plantas fueron cultivadas en una cámara de crecimiento con un régimen de 16 h de luz ($100 \mu\text{E.m.}^{-2}\text{s}^{-1}$) y 8 h de oscuridad (condición día largo). La temperatura de crecimiento fue de 23°C durante el ciclo luz/oscuridad, mientras que la humedad fue mantenida en 65% de humedad relativa. Las plantas fueron regadas 2 veces por semana con agua. Para el crecimiento directo en tierra, las semillas fueron estratificadas a 4°C por 2 días en tubos de microcentrífuga con 1ml de 0,1% (p/v) agar, y luego sembradas en tierra. Las plantas de *Nicotiana tabacum* (cv Petit Havana) fueron crecidas en condición día largo durante 8 semanas y la segunda hoja fue utilizada para el análisis de ARN.

3.1.2 Transformación de *Arabidopsis*

La transformación de plantas de *Arabidopsis* se realizó mediante la técnica de "floral-dip". Se cultivó las cepas de *A. tumefaciens* ASE transformadas con los distintos plásmidos binarios en 100 ml de LB conteniendo 50 $\mu\text{g}/\text{ml}$ de espectinomicina, 50 $\mu\text{g}/\text{ml}$ de kanamicina y 25 $\mu\text{g}/\text{mg}$ de cloranfenicol a 28 °C hasta llegar a una DO600 cercana a 2. Se cosechó las células por centrifugación a 5.000 g durante 10 min a temperatura ambiente y se resuspendió en un medio con 5% (p/v) sacarosa y 200 $\mu\text{l}/\text{l}$ de Silwet L-77. Las inflorescencias de plantas de *Arabidopsis* regadas el día anterior se sumergieron en la suspensión de células durante 30 segundos. Luego de escurrir el exceso de medio de cultivo, se colocaron las plantas en posición horizontal dentro de bolsas plásticas para conservar la humedad. Luego de 24 hs se sacaron de las bolsas y se prosiguió con el cultivo de las plantas en la cámara de crecimiento hasta la maduración de las semillas.

3.1.3 Recolección y esterilización de semillas

La recolección de semillas se realizó a partir de plantas que finalizaron su ciclo de vida y que presentaban semillas maduras. Se retiró las plantas del pote de tierra y frotando las mismas con las manos se dejó caer las semillas sobre una hoja de papel. Se eliminaron los restos de silicua y hojas por tamizado. Las semillas se almacenaron en bolsas de papel en un recipiente seco hasta su posterior utilización.

Para la esterilización de las semillas, las mismas se incubaron con una solución de 70% (v/v) etanol, 0,1% (v/v) tritón-X-100, durante 8 min. Luego se realizó un enjuague con 96% (v/v) etanol y finalmente se las dejó secar en el flujo laminar.

La manipulación del material vegetal en condiciones de esterilidad se realizó en una cabina de flujo laminar.

3.1.4 Selección de transformantes

Para la selección de plantas transformadas, las semillas esterilizadas se sembraron en placas de Petri con medio MS-agar suplementado con 25 µg/ml de kanamicina. Las placas se mantuvieron en oscuridad a 4 °C por 48 hs. Transcurrido este período se transfirieron las placas a la cámara de cultivo y en estas condiciones las plantas se dejaron crecer 5-10 días. Luego se analizaron las placas y aquellas plantas que presentaban un fenotipo de resistencia al agente de selección, evidenciado por el desarrollo de la raíz primaria y del primer par de hojas verdaderas, se transfirieron a potes con tierra. Los potes se ubicaron en bandejas con agua las cuales se cubrieron con papel de nylon transparente, de modo de evitar la pérdida de humedad. En estas condiciones se cultivaron en la cámara durante 2 días, momento en el que se retiró el nylon y se continuó con su cultivo.

3.1.5 Transformación nuclear de tabaco

Las plantas transgénicas de tabaco fueron preparadas por el método de transformación de discos de hojas por co-cultivo con *A. tumefaciens*. Se cultivó la cepa de *A. tumefaciens* GV3101 pMP90 transformada con los distintos plásmidos binarios en 10 ml de 2YT conteniendo los antibióticos necesarios (100 µg ml⁻¹ rifampicina, 50 µg ml⁻¹ gentamicina, 60 µg ml⁻¹ cloranfenicol) a 28 °C hasta llegar a una DO600 entre 0,5 y 1. Se utilizaron plantas de tabaco (*N. tabacum* cv *Petit Havana*) crecidas en medio de cultivo sólido de Murashige-Skoog con 3% (p/v) sacarosa (MS-0) en condiciones estériles. Las hojas fueron cortadas en piezas de aproximadamente 1 cm² usando un bisturí estéril. Las porciones de hoja fueron colocadas en una caja de Petri conteniendo 100 ml de medio MS líquido, al que se agregaron 10 ml

del cultivo de *A. tumefaciens*, y se incubó entre 15 y 30 min. El exceso de líquido de las porciones de hoja fue eliminado con papel de filtro estéril y luego se las colocó en placas de Petri conteniendo medio MS-10 (MS-0 suplementado con 1 mg l-1 kinetina y 0,1 mg l-1 ácido naftalenacético). Tras 4 días de incubación a 28 °C, los discos de hoja fueron cambiados a cajas de Petri con medio MS-10 suplementado con 100 µg l-1 kanamicina, para seleccionar células vegetales transformadas, y 100 µg ml-1 cefotaxima, para eliminar el Agrobacterium. Luego de 3 a 4 semanas de crecimiento en cámara de cultivo, se generaron vástagos que fueron transplantados a medio MS-0 suplementado con los mismos antibióticos, para que desarrolle raíces y continúen con su crecimiento. Después de 6 a 8 semanas de crecimiento, las plántulas que desarrollaron raíces fueron transplantadas a tierra y aproximadamente 12 semanas después se cosecharon las semillas.

3.2 Cuantificación del nivel de expresión génica

3.2.1 Extracción de ARN

La extracción de ARN de tejido vegetal de *Arabidopsis* se realizó utilizando el reactivo TRIzol (Invitrogen). La recolección de las muestras de tejido se realizó en tubos de microcentrífuga que inmediatamente fueron sumergidos en N2 líquido. El material fue reducido a un polvo fino utilizando un pilón. A continuación, sobre este polvo se adicionó TRIzol (1 ml de reactivo para un máximo de 100 mg de tejido) y se agitó de modo de resuspender el tejido mortereado. El homogenizado se centrifugó a 12000 g a 4 °C durante 10 min y finalmente se transfirió el sobrenadante a un tubo de microcentrífuga nuevo.

A continuación se agregó 0,2 volúmenes de cloroformo por cada volumen de TRIzol original. Esta mezcla se invirtió vigorosamente por 15 segundos y se centrifugó a 12000 g a 4 °C durante 15 min. Se transfirió la fase superior a un nuevo tubo de microcentrífuga y se precipitó el ARN mediante el agregado de un volumen de isopropanol. Esta mezcla se incubó por 2 hs a -20 °C. Luego se centrifugó a 12000 g a 4 °C durante 10 min. Finalizada la centrifugación se descartó el sobrenadante.

Se lavó el precipitado de ARN mediante el agregado de 1 ml de 70% (v/v) etanol frío y agitación con vortex. Luego se centrifugó a 7500 g a 4 °C durante 5 min, descartando el sobrenadante una vez finalizada la centrifugación. Este paso de lavado se repitió una vez más.

El precipitado de ARN obtenido se secó en estufa 37 °C por 10 min y luego se resuspendió en 50 µl de agua Milli-Q esteril.

3.2.2 Cuantificación y chequeo de la integridad del ARN purificado

Se determinó la absorbancia a 230, 260 y 280 nm. Se estimó la pureza de la preparación a partir de la relación de las medidas de absorbancia a Abs260/ Abs230 y Abs260/ Abs280. La integridad del ARN purificado se determinó mediante la electroforesis en geles de 1,5% (p/v) agarosa de 5 μ l del ARN preparado. Las bandas de ARN ribosomal se visualizaron por tinción de los geles con bromuro de etidio.

3.2.3 Tratamiento del ARN preparado con ADNasa

Se preparó la siguiente mezcla de reacción: En un volumen final de reacción de 20 μ l se adicionó: 0,5 a 1 μ g de ARN total, 2 μ l de buffer "RQ1 RNase-Free DNase" (Promega), 1 U "RQ1 RNase-Free DNase" (Promega) y agua Milli-Q para completar el volumen de reacción.

La mezcla de reacción se incubó 30 min a 37 °C. Luego se inactivó la ADNasa mediante la adición al tubo de reacción de 1 μ l de "DNase Stop Solution" (Promega) e incubación por 10 min a 65 °C. De los 21 μ l de reacción finales, 12 μ l se utilizaron para sintetizar ADN complementario (ADNc). Los 9 μ l restantes se utilizaron como control negativo en la PCR en tiempo real.

3.2.4 Retrotranscripción (RT)

La síntesis del ADNc a partir del ARN preparado se llevó cabo según el siguiente protocolo. En una primera etapa se preparó la siguiente mezcla: en un volumen final de 13,5 μ l se adicionó 0,25 μ g de oligo dT, 12 μ l de ARN tratado con ADNasa, dNTPs a una concentración final de 0,4 mM cada uno. El volumen de la mezcla se completó con agua Milli-Q.

La mezcla de reacción se incubó durante 5 min a 65 °C y a continuación en hielo durante al menos un minuto. Luego se realizó una centrifugación rápida y se adicionó a cada tubo de reacción 4 μ l de "5X First Strand Buffer", 1 μ l de 0,1 M DTT, 1 μ l "RNase OUT Recombinant Ribonuclease Inhibitor" (Invitrogen) y 100 U de "SuperScript III Reverse Transcriptase" (Invitrogen).

Se mezcló por inversión y se incubó 60 min a 50 °C. Luego se realizó una centrifugación rápida y se inactivó la reacción mediante la incubación a 70 °C durante 15 min. Finalmente el ADNc a usar como molde en la reacciones de amplificación se diluyó al menos 40 veces en agua Milli-Q.

3.2.5 Reacción en cadena de la polimerasa en tiempo real

La cuantificación relativa de los niveles de expresión génica se llevó a cabo mediante la técnica de PCR en tiempo real (Real-Time PCR) según el método $2^{-\Delta\Delta C_t}$ [107].

Como calibrador se utilizó un gen de expresión constitutiva que codifica por la Ser/Thr proteína fosfatasa 2 (PP2A; At1g13320) [36]. Un aspecto importante que se tuvo en cuenta para la aplicación del método de cuantificación $2^{-\Delta\Delta C_t}$ es que las eficiencias de amplificación de los fragmentos del gen de interés y del gen de referencia sean similares y próximas a 2 [107].

Para el diseño de los oligonucleótidos para la cuantificación de blancos de miARNs se tuvieron en cuenta las siguientes consideraciones. Los miARNs en plantas regulan la expresión de sus blancos principalmente mediante el corte del ARNm. De esta manera para cuantificar el nivel de expresión de un gen que es blanco de un miARN es fundamental distinguir el ARNm completo del cortado. Dado que en la RT se sintetizan las hebras de ADNc a partir del oligo dT que hibrida en el extremo 3' de los ARN mensajeros, la secuencia que se ubica 5' del sitio complementario al miARN solamente será retrotranscrita en aquellos ARNm que no hayan sido cortados por el miARN. De esta manera si los oligonucleótidos utilizados en la PCR en tiempo real hibridan en esta región se amplificará específicamente un fragmento presente en el ADNc generado de ARNm intactos.

Las reacciones se realizaron en un equipo de qPCR "Mastercycler® ep realplex" (Eppendorf) en microtubos apropiados para esta técnica. Para determinar el perfil de amplificación durante la PCR se utilizó el colorante fluorescente "SYBR Green I Nucleic Acid Gel Stain" (Roche).

Las reacciones de amplificación se llevaron a cabo bajo las siguientes condiciones. Para un volumen final de reacción de $20\mu\text{l}$ se adicionó a cada tubo: $2\mu\text{l}$ de "10X PCR Buffer" (Invitrogen), 3mM MgCl_2 , $0,2\text{mM}$ de cada dNTP, $5\mu\text{l}$ de ADNc, $0,5\text{ U}$ de "Platinum Taq DNA Polymerase" (Invitrogen), 20 pmol de cada cebador, y $0,8\mu\text{l}$ de una dilución $1000\times$ en agua del "SYBR Green I Nucleic Acid Gel Stain (Roche)" original. El volumen se completó con agua Milli-Q.

Como control negativo se preparó un tubo de reacción para cada muestra en el cual se agregaron los mismos componentes en las mismas condiciones, pero en lugar de utilizar ADNc como molde se colocó $5\mu\text{l}$ del ARN sobrante luego del tratamiento con ADNasa, diluido en la misma proporción.

El termociclado de las mezclas de reacción se llevó a cabo según el programa que se muestra en la tabla 3.1.

Se incluyó una etapa final de determinación de la temperatura de fusión de los productos amplificados, lo que permite conocer la especificidad de la reacción. Además, la primera vez

Table 3.1 Programa de termociclado

Etapa	Temperatura (°C)	Tiempo	Nº de ciclos
Desnaturalización	95	1 min	1
Ciclos	95	15 seg	40
	55	30 seg	
	72	40 seg	
	Lectura de fluorescencia		

que se utilizaron los oligonucleótidos el producto de la reacción de amplificación se analizó en un gel de agarosa para confirmar la especificidad de la reacción.

Los oligonucleótidos utilizados en este estudio son que se detallan en la Tabla 3.2.

Table 3.2 Oligos utilizados para la RT-qPCR

Gene	Locus ID	Forward primer	Reverse Primer
PAA2	At5g21930	GTCCTCTTATCAGGGGACAGG	CATAGTTGCTTGCAAGACTCAG
MYB33	At5g06100	CTATGAAACCGACATTACCTG	CTTGGCTTCCAGAAGCAACATATCG
NZZ	At4g27330	TCGGGTCAAGTTATGATCGA	AGGGTTTCCCTCATGTAGCTCC
PP2A	At1g13320	CCTGCCGTAATAACTGCATCT	CTTCACCTAGCTCCACCAAGCA
tMT2A	tobacco	TACCCAGATTGAGCTACAACGAG	GCAGGAGATTCACCCATTCCATT
tMT2B	tobacco	TACCCAGATTGAGCTACAACGAA	AGGGGATTCACCCCATTCCATT

3.3 Mapeo del sitio de corte

En la Figura 3.1 se muestra un esquema para la identificación de los productos de corte de miARNs. Inicialmente, se realizó una extracción 50mg de ARN total de plántulas de Col-0 y se realizó una purificación de ARN utilizando el kit comercial "PolyATract®" de (Promega). El ARNm contenido en un volumen de 100 μ l de agua se precipitó mediante el agregado de 2 μ l 10 mg/ μ l glucógeno, 10 μ l de 3M Acetato Sodio pH 5.2 y 220 μ l de etanol absoluto, e incubación toda la noche a -20 °C. Luego se centrifugó 20 min a 4 oC a 20000 g. Se lavó con etanol 80%, y se centrifugó 5 min a 4 °C a 20000 g. Se removió el sobrenadante y se secó 10 min a 37 °C. Finalmente se resuspendió en 10 μ l agua.

A continuación se realizó la ligación del Oligo Adaptador ARN (5' CAGCUGGAGCAC-GAGGACACUGACAUGGACUGAAGGAGUAGAAA 3') mediante el siguiente protocolo. Se mezcló 250 ng de ARNm con 0,25 μ l de oligo ARN (100 μ M), y se incubó 5 min a 65 oC. Luego se incubó en hielo 2 min. Se agregó la mezcla de reacción: 1 μ l 10X Buffer Ligasa, 1 μ l 10mM ATP, 1 μ l BSA, 40 U RNaseOut (Invitrogen), 5 U T4 RNA ligasa (Invitrogen), agua para completar 10 μ l. Se incubó a 37 °C durante 1 hora, luego se centrifugó y se dejó en hielo. Se precipitó el ARNm mediante el protocolo antes descripto. Posteriormente se realizó una retrotranscripción y finalmente se eliminó el exceso de oligo ARN GeneRacer incubando 30 min a 37 °C con 0,04 g/ μ l de RNAsa A.

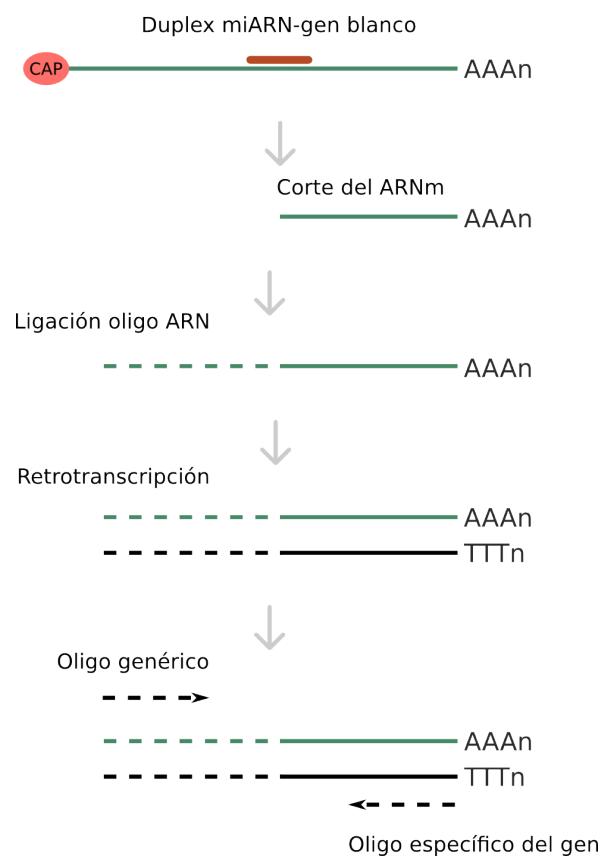


Figura 3.1 Diagrama de la estrategia utilizada para la identificación de los productos de corte de miARNs.

Para identificar los productos de corte específicos del gen en estudio se realizó una PCR con un oligo que hibrida sobre la secuencia del Oligo Adaptador ARN (oligogenérico) y otro oligo que hibrida sobre la secuencia del gen en una región 3' del sitio putativo de corte del miARN. Los productos de la PCR se clonaron y se secuenció al menos 15 clones independientes para determinar la secuencia del extremo 5'.

Los oligonucleótidos utilizados en este estudio son que se detallan en la Tabla 3.3.

Table 3.3 Oligonucleótidos utilizados para 5' RACE

Gen	Locus ID	5' RACE	5' RACE nested
General		CGACTGGAGCAGCAGGACACTGA	GGACACTGACATGGACTGAAGGGAGTA
PAA2	At5g21930	GACTTATGGAGCTGCAGAAAGTAATG	CATAGTTGCTTGCAAGACTCAG
IAR3	At1g51760	ATCTTCGTATCCCATTAAATGGITGCATCTCG	CATATTCAACGCTCGCTTGCTTGTGATAACC
NZZ	At4g27330	CATTAAAGCTTCAAGGACAATCAATGGTATTAGG	AGGGTTCTTCCATGTAGCTCC
MMG4.7	At5g43060	ATGGTAACACCTTAGCATTTTCC	CTTCGGTATCAACCCWCCATT
UDP	At2g47650	AATGGGGGCCACATGTTCTCC	CCTCGGTATGATGTCATGGT
SVP	At2g22540	GCAACTTTCTTCATTCAATC	TTTCATCTGCCTCAGCTCAC
loricrin-related	At5g64550	ACCATAGAGCTTGAGTAGT	CCTCAGCACTTCTGTACAG
	At3g14110	CGGAAGGATCAGTCAGTCTC	CCAGCCTGGTATAACAGTC
	At3g22110	GTTTCATCGCCAAGGTAAC	CCAGGCGAATAAACTAGAG
AVA-P2	At1g19910	CTCTAGACTGACCAAGCTGA	GGATGATACCAACAATGAGA

3.4 Métodos utilizados para la predicción de genes regulados por miARNs en plantas

En la primer parte de esta tesis diseñamos una estrategia para la identificación de genes blanco regulados por miARNs basado en la conservación evolutiva del par miARN-gen blanco. La metodología aplicada es la siguiente.

3.4.1 MiARN consensos

Las 22 familias de miARNs conservadas en angiospermas fueron consideradas para esta parte del trabajo [13, 45]. MiR319 y miR159 que codifican para miARNs similares, fueron considerados como familias diferentes ya que regulan a genes blanco distintos [104]. Consideramos todos los miembros de estas familia, obtenidos de miRBASE², pertenecientes a *A. thaliana*, *Populus trichocarpa* y *Oryza Sativa*. Variaciones en las posiciones 1, 20 y 21 son muy comunes en las familias de miARNs [39]. Por esto, definimos como secuencia consenso, a las secuencias más comunes (posiciones 2-19) de distintos miembros de cada familia (tabla 4.4).

²<http://mirbase.org>

3.4.2 Predicción de genes regulados por miARNs

Conjunto de datos de plantas

Los datos de las secuencias pertenecen a librerías extraídas de "Gene Index Project"³, que consiste en una base de datos de ESTs ensamblados. Seleccionamos un conjunto de datos pertenecientes a Angiospermas. Además utilizamos secuencias de ARNm completos de *A. thaliana*⁴ y *Oryza Sativa*⁵ (ver tabla 9.8). La búsqueda la realizamos utilizando PatMatch [150], que es un programa de búsqueda de patrones de nucleótidos cortos o péptidos. El programa puede ser usado para encontrar coincidencias con un patrón de secuencia específico y permite el uso de códigos de secuencias ambiguas y expresiones regulares y por esto se puede utilizar la búsqueda con "mismatches", inserciones y delecciones. Realizamos la búsqueda de potenciales genes blanco permitiendo tres "mismatches" con las secuencias consenso, mientras que las interacciones G:U y los bulges fueron considerados mismatches. Para realizar el alineamiento del par miARN-gen blanco, desarrollamos una versión modificada del algoritmo de programación dinámica Needleman-Wunsch [97], utilizando el lenguaje Perl⁶. Además, desarrollamos scripts para integrar los módulos de Blastx [7] utilizando el proteoma de Arabidopsis y el módulo RNAhybrid [51] que es una herramienta que permite encontrar la menor energía libre de hibridación (MFE) de dos secuencias de ARN.

Filtros

Las secuencias candidatas fueron etiquetadas con el identificador del locus (locus ID) con mejor puntuación (best hit) en *A. thaliana*, utilizando el módulo de Blastx (Corte del evalúe de 10e⁻⁵). De este modo, genes blanco de distintas especies que tenían la misma etiqueta fueron agrupados juntos, ya que tendrían el mismo homólogo en *A. thaliana*. El filtro de conservación evolutiva hace referencia al número mínimo de especies donde la misma etiqueta estaba presente para un miARN particular. El filtro empírico está basado en trabajos previos [123] y hace referencia a la energía de interacción MFE (mínima energía libre de hibridación de al menos 72% del apareamiento perfecto). El otro filtro empírico requiere que entre el par miARN-gen blanco, solamente está permitido un "mismatch" entre la posición 2 y la 12 del miARN (1-11 de nuestra búsqueda modificada con las secuencias consenso).

³<http://compbio.dfci.harvard.edu/tgi/>

⁴<http://arabidopsis.org>

⁵<http://rice.plantbiology.msu.edu>

⁶<http://perl.org>

Controles

Como control, realizamos las búsquedas del mismo modo que lo hicimos para los miARNs conservados, pero utilizando secuencia al azar. Para cada miARN conservado, generamos 20 secuencias al azar (scramble) dividiendo las secuencias originales de a di-nucleótidos y luego generando nuevas secuencias al azar conservando esa composición de los di-nucleótidos como fue descrito previamente [63]. De estas 20 secuencias al azar, elegimos las 10 que tenían el número más similar del total de genes blanco para el miARN real correspondiente. La relación señal/ruido fue calculada como el cociente entre el número de genes blanco para los miARNs y el número de genes blanco del promedio obtenido para las secuencias al azar. Como un control adicional, seleccionamos dos miARNs que no están conservados durante la evolución, que son el miR158 y el miR173.

3.5 ComTAR: una herramienta para la predicción de genes blanco regulados por miARNs en plantas

A partir de los resultados positivos obtenidos de la estrategia descrita anteriormente, decidimos desarrollar una herramienta web y dejarla disponible para la comunidad científica denominada comTAR que está disponible en un sub-dominio de la página web institucional del IBR: <http://rnabiology.ibr-conicet.gov.ar/comtar>.

3.5.1 MiARN y transcriptos

Como las secuencias del maduro del miARN puede variar en distintas especies, especialmente en la posición 1, 20 y 21 ([32], utilizamos secuencias del 2-19 (18nt) para realizar las búsquedas. Como además existen variaciones en las secuencias en los distintos miARNs de las mismas familias, utilizamos la más representativa teniendo en cuenta los genomas de *Arabidopsis*, álamo y arroz. De este modo comTAR contiene datos pre-calculados, de potenciales genes blanco para 22 miARNs conservados en plantas (ver tabla 4.4) donde el usuario puede navegar los resultados y cambiar los parámetros de entrada. Además, el usuario puede realizar la búsqueda de nuevos ARNs pequeños teniendo en cuenta esta consideración. El cálculo se hace en el cluster del CCT-Rosario y los datos se obtienen luego de unas horas. Como la herramienta web la realizamos tiempo después de haber hecho la estrategia para predicción de genes blanco, utilizamos una nueva base de datos más actualizada y completa denominada Phytozome⁷ [52]. La misma corresponde a secuencias de transcriptos de plantas

⁷<http://phytozome.jgi.doe.gov>

formado por archivos de nucleótidos en formato FASTA de transcriptos de ARNm (UTR, exones) con variantes de splicing.

3.5.2 Búsqueda de genes blanco

La búsqueda de genes blanco la realizamos de la misma manera que la descrita anteriormente con algunos cambios. Además de actualizar la base de datos y utilizar la de Phytozome, actualizamos la base de datos de *A. thaliana* por la del TAIR10. Las secuencias candidatas fueron etiquetadas con el mejor hit del locus ID del Arabidopsis TAIR10, utilizando los archivos de anotación de Phytozome, y lo utilizamos como "TAG" (etiqueta). Por último, cada TAG de Arabidopsis fue indexado con una breve descripción funcional y computacional obtenida del TAIR10 y los genes blanco candidatos fueron agrupados por familias teniendo en cuenta la clasificación de familias del TAIR10.

3.5.3 Herramienta web y almacenamiento de datos

ComTAR fue diseñado como una aplicación web con un framework open-source en PHP denominado Codeigniter para la interfaz gráfica, pero el análisis está basado en un back-end escrito en Perl. Los datos que surgen de ese análisis fueron almacenados en una base de datos en MySQL⁸. El back-end es el encargado de realizar la búsqueda de secuencias y además ahí es donde se integraron las herramientas y scripts para aumentar la especificidad y sensibilidad de comTAR. También el back-end es el encargado de generar los resultados finales. Mientras el front-end es el responsable de mostrar los resultados (Figura 3.2). El TAG del mejor hit en Arabidopsis es el que determina el número de especies donde un gen blanco está presente, y el número mínimo de especie es un parámetro que es definido por el usuario.

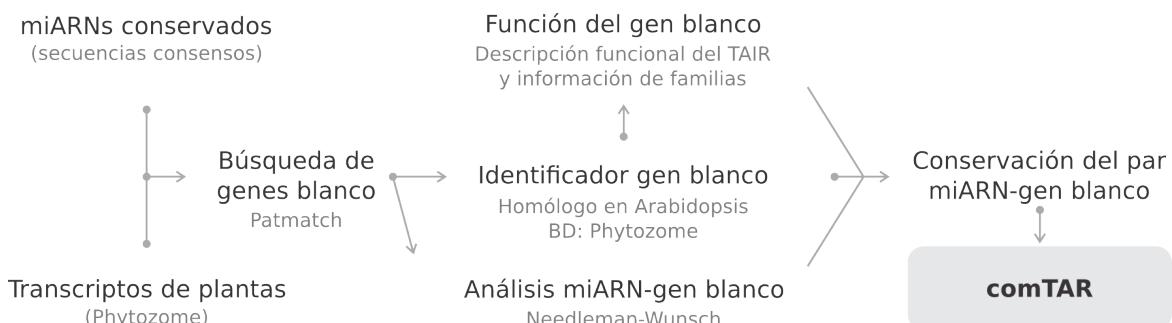


Figura 3.2 comTAR. Diagrama de flujo que describe la herramienta

⁸<http://mysql.com>

3.6 Protocolo de SPARE multiplex

Construcción de bibliotecas de SPARE

La técnica de SPARE (del inglés Specific Parallel Analisys of 5' RNA Ends) fue desarrollada con el objetivo de caracterizar el procesamiento de los precursores de miARNs de *A. thaliana*. El esquema de la técnica se muestra en la Figura 3.3 y se detalla a continuación.

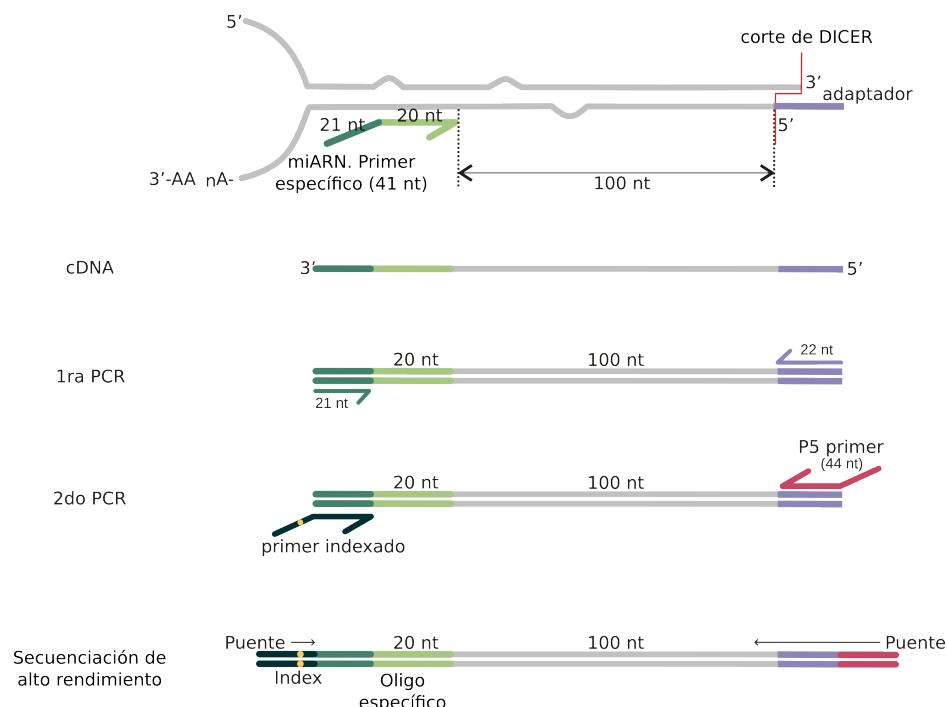


Figura 3.3 Esquema general para la construcción de bibliotecas de SPARE.

Extracción de ARN

La extracción de ARN fue realizada como se detalla en sección 3.2.1. Las plántulas de *A. thaliana* fueron crecidas durante 10 días con luz continua (independizándonos de este modo la variación asociada al ritmo circadiano a la hora de la colecta de muestra).

Tratamiento con ADNasa

1. Preparar la siguiente reacción:

- ARN (40 µg de ARN en 89 µl finales)
- Buffer DNasa 10 µl

- DNasa (Promega) 1 μ l
2. Incubar 30 minutos a 37 °C
 3. Agregar 1 μ l de DNasa Stop Solution
 4. Incubar 10 minutos a 65 °C

Precipitación de ARN

1. Agregar:
 - 1 μ l de Glucógeno (20mgr/ μ l)
 - 10 μ l de Acetato Sodio 3M pH 5.2
 - 220 μ l de Etanol Abs
2. ON -20 °C
3. Centrifugar 30 minutos a 4 °C a máxima velocidad
4. Descartar sobrenadante
5. Lavar con Etanol 80
6. Centrifugar 5 minutos a 4 °C a máxima velocidad
7. Descartar sobrenadante. Repetir desde el paso 5 una vez.
8. Spin y remover excedente.
9. Secar 5-10 minutos a 37 °C
10. Resuspender en 9 μ l de agua

Ligación con Oligo ARN

1. 1. Mezclar 9 μ l de ARN con 1 μ l 5-ARN adapter(200 uM) u oligo target.
2. Incubar 5 minutos a 65 °C
3. Poner en hielo aproximadamente 2 minutos
4. Agregar a la mezcla de reacción (Vf: 20 μ l)

- 2 μ de 10X Buffer Ligasa
 - 2 μ de 10mM ATP
 - 2 μ de BSA
 - 2 μ de RNaseOut (40U/ μ l)
 - 2 μ de T4 RNA ligasa (5U/ μ l)
5. Incubar a 37°C durante 1 hora
 6. Centrifugar y poner en hielo.
 7. Inactivación de la ligasa: 72 °C durante 15'

Purificación de ARNm con dynabeads

1. Seguir las instrucciones del kit
2. Resuspender en 12 μ l de agua.

Transcripción Reversa

1. Preparar la siguiente reacción:
 - ARNm-ligado purificado: 12 μ l (dividir en 8 tubos del PCR del strip pipetando 4 μ l en cada uno)
 - dNTP mix 10 mM:1 μ l
 - Mix Oligo específico: 0,5 μ l (ver indicaciones de cómo prepararla más abajo)
2. Calentar a 65 °C durante 5 minutos
3. Incubar en hielo al menos 1 minuto
4. Spin down.
5. Agregar:
 - 5X First Strand Buffer 4 μ l
 - DTT 0,1 M 1 μ l
 - Inhibidor de RNasa (Invitrogen) 1 μ l
 - SuperScript III 0,5 μ l

6. Incubar 60 minutos a 50 °C
7. Spin down
8. Inactivar la reacción 15 minutos a 70 °C

Preparación de la mix de oligos Los stocks de oligos están 100 μ M, como en cada reacción de síntesis de ADNc se levantan aproximadamente 15 precursores, conviene trabajar de la siguiente manera: en un eppendorf se coloca 1 μ l de cada oligo (las mix 1-7 tienen 15 oligos y la mix 8 tiene 16 oligos) y se completa con cantidad suficiente para 20 μ l. De esta dilución 1:20 se toman 10 μ l y se lleva a 100 μ l, para obtener una dilución 1:200. Esto último se repite para obtener una dilución final 1:2000 de la que se toman 0,5ul para realizar la síntesis de ADNc.

PCRs

PCR1

La función de esta PCR es enriquecer los intermediarios de procesamiento, sin la interferencia de los oligos indexados, que son de mayor tamaño y por ende dan mayor número de dímero de primer. Se toman 2,5uL de cada ADNc para un total de 20 μ L que se emplean como molde de la PCR1.

- Oligo Fw: 1,5 μ l
- Oligo Rv: 1,5 μ l
- dNTPs 10mM: 1 μ l
- buffer GC: 10 μ L
- Phusion pol: 0,5 μ l
- cDNA: 20 μ l
- DMSO: 1,5 μ L
- H₂O csp 50 μ l.

Programa

- 1' 98 °C x 1 ciclo
- 30" 98 °C

- 30" 63 °C
- 50" 72 °C X 20 ciclos
- 10' 72 °C 1 ciclo

Nota: la PCR1 se hace en condiciones de no-saturación para poder trabajar en condiciones semi-cuantitativas.

La PCR1 se corre en un gel de agarosa al 1,5% conjunto a un marcador de peso molecular 100 (PM100). En estas condiciones los productos de PCR no son visibles en el gel. Se corta y purifica la región comprendida entre 100-700pb con un kit Wizard SV gel and PCR Clean-up system, eluyendola en un volumen de 35 μ l.

PCR2

Se realiza con el objetivo de incorporar el oligo indexado al producto de la PCR1. De la PCR1 se toman 10 μ l para realizar la segunda PCR. Oligo indexado: 0,75 μ l (no se coloca en la master mix, se pone un oligo indexado diferente por biblioteca).

- 5X GC buffer: 10 μ l
- dNTPs: 1 μ l
- Oligo P5: 0,75 μ l
- Oligo Indexado: 0,75 μ l
- Molde: 10 μ l
- DMSO: 1,5 μ l
- Phusion: 0,5 μ l
- H₂O: 22,5 μ l

Volumen final: 50 μ l. Programa de PCR

- 98 °C 1'
- 98 °C 30"
- 63 °C 30" x 10 ciclos
- 72 °C 50"
- 72 °C 10'

Se corre un gel al 2% conjunto a un marcado de peso de molecular de 100pb y otro de 50pb. Se corta la región desde 150-800pb y se purifica por kit Wizard SV gel and PCR Clean-up system eluyendo en un volumen final de 22 μ L.

Análisis bioinformático de los datos obtenidos de la secuenciación

Una vez obtenidos los datos crudos de la secuencias se procedió a hacer el procesamiento de los datos siguiendo los siguientes pasos:

- Comprobar la calidad de las secuencias.
- Remover secuencias adaptadoras (trimming)
- Agrupar secuencias iguales y determinar abundancia (conversión de archivos fastq file a archivos tag-count).
- Mapear las secuencias únicas con las secuencias de precursores (sub-genoma definido)
- Combinar los resultados del mapeo con la abundancia de cada secuencia
- Analizar y comparar los intermediarios de procesamiento para plantas silvestre y mutantes de procesamiento

3.7 Protocolo de PARE

El protocolo de bibliotecas de ARN pequeños se preparó como se Methods [156] y muestra en la Figura 3.4.

3.8 Protocolo de ARN pequeños

El protocolo de bibliotecas de ARN pequeños se preparó como se detalla en catálogo de ILLUMINA (# RS9301012) y muestra en la Figura 3.5.

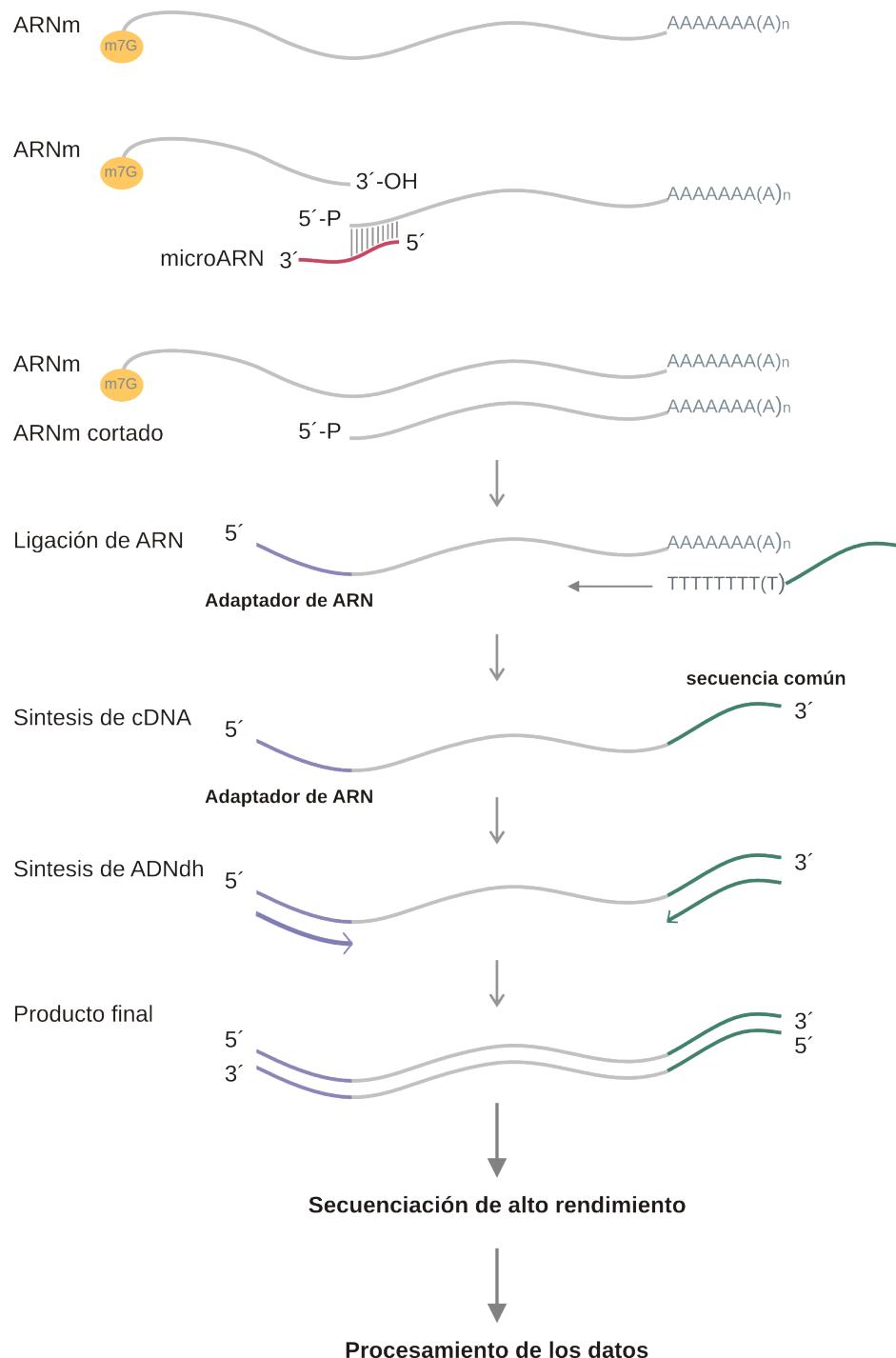


Figura 3.4 Esquema general para la construcción de bibliotecas de PARE.

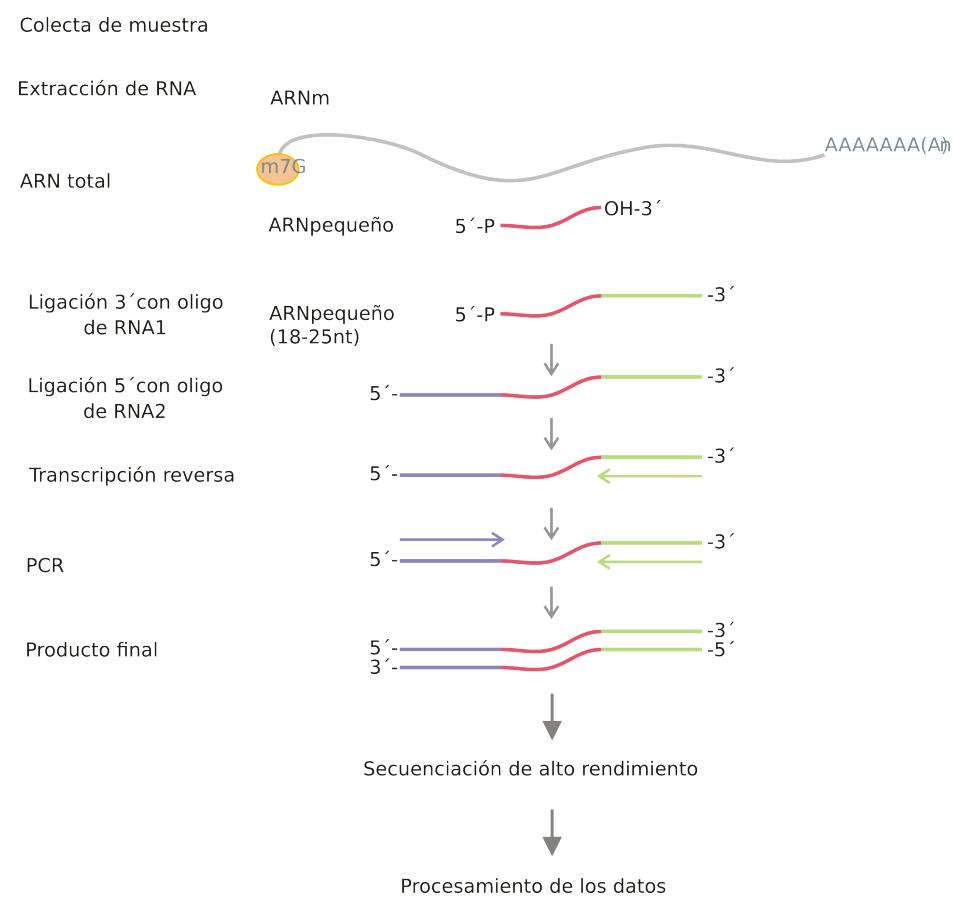


Figura 3.5 Esquema general para la construcción de bibliotecas de ARN pequeños.

3.9 Estudios genómicos sobre la biogénesis de precursores de miARN en plantas

3.9.1 Procesamiento de precursores de miARNs en plantas

Análisis bioinformático

Obtuvimos las estructuras secundarias para cada precursor calculada a partir de la herramienta mfold [162] con los parámetros por default a 37 °C de temperatura. El lado proximal del duplex miARN/miARN* fue definido como la posición +1. Analizamos la estructura secundaria y consideremos las posiciones que había un match como un 0, mientras que bulges y "mismatches" los consideramos como 1. Además hicimos un promedio para todos los precursores siguiendo la misma estrategia. Implementamos un pipeline bioinformático utilizando "in-house" scripts y datos públicos de miRBASE, para asistir con el análisis de las bibliotecas de secuenciación masiva. Las secuencias de los ARN pequeños fueron obtenidas de la base de datos de nueva generación de *Arabidopsis*⁹ [65] y de la base de datos de miRBASE [70].

Acceso a los datos

Los datos de secuenciación masiva con los resultados del SPARE de la primer parte del trabajo, están accesibles mediante el NCBI Gene Expression Omnibus (GEO¹⁰ con el código de acceso GSE46429.

3.10 Estrategia bioinformática para caracterizar la relación entre la evolución de los precursores de miARNs en plantas y los mecanismos de procesamientos

Búsqueda de ortólogos y extensión de la secuencia

De la base de datos biológica miRBASE (release 19), obtuvimos las secuencias de 92 precursores que incluye cada miembro de las familias de *A. thaliana* conservados en dicotiledóneas. Además obtuvimos las secuencias de 117 precursores que incluye cada miembro de las familias de *O. Sativa* conservados en monocotiledóneas. Con estas secuencias de *A. thaliana* y de *O. Sativa*, buscamos los ortólogos en dicotiledóneas y monocotiledóneas respectivamente.

⁹<https://http://mpss.udel.edu/>

¹⁰<http://ncbi.nlm.nih.gov/geo>

Esto lo hicimos utilizando la técnica de blast recíproco, mediante un script desarrollado por nosotros que utiliza el blast del paquete de BLAST+ del NCBI (versión 2.2.25). Los datos genómicos para la búsqueda de ortólogos, fueron extraídos de la base de datos de Phytozome y utilizamos 30 especies de dicotiledóneas y 6 especies de monocotiledóneas (Figura 3.6). Además extendimos las secuencias, obteniendo más bases de las que originalmente están definidas en miRBASE y empezamos nuestro análisis con una definición arbitraria de los precursores de plantas incluyendo 150 nt fuera del par miARN/miARN*. Esta definición se basa en que todos los determinantes estructurales validados experimentalmente necesarios para el procesamiento de miARNs están dentro de este rango.

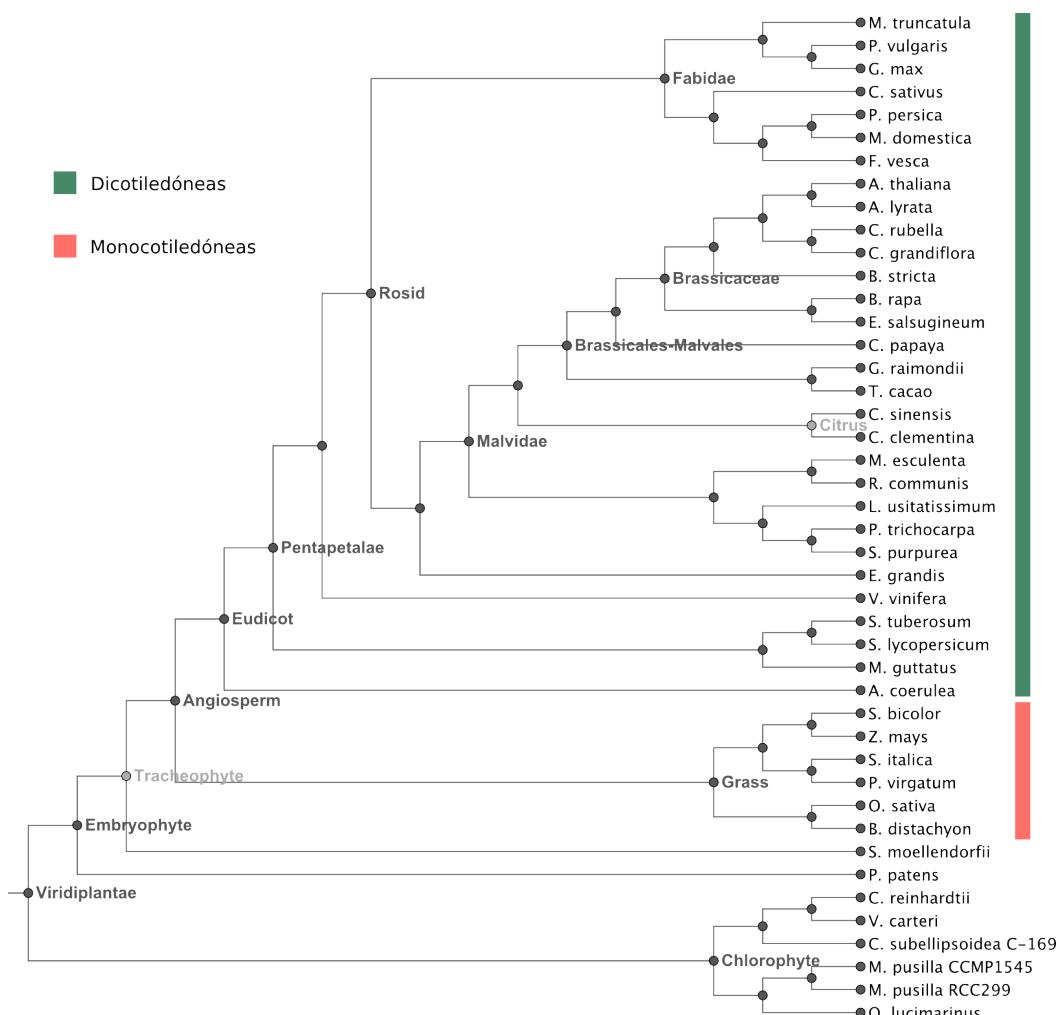


Figura 3.6 Especies almacenadas en Phytozome.

Alineamientos de precursores y determinación de estructura secundaria

Los alineamientos de precursores en base a su secuencia primaria se realizaron utilizando la herramienta T-coffee (versión 11.00.8cbe486) [99] como se muestra a continuación :

```
t_coffee -in=miR.fasta -mode=regular -method=slow_pair -output=fasta_aln -quiet=stdout -out=miR.aln
```

Por defecto, T-coffee colorea los alineamientos en base a un puntaje de las bibliotecas construida a partir de distintos algoritmos de alineamiento. Como a nosotros nos interesa poder visualizar las diferencias en base a la conservación de los precursores, utilizamos T-Coffee para colorear el alineamiento en base a el nivel de conservación

```
t_coffee -other_pg seq_reformat -in miR.aln -in3 miR.aln -action +3evaluate idmat -out=miR_T_Coffee.pdf
```

Los alineamientos de precursores, considerando su estructura secundaria, fueron realizados con la herramienta R-coffee [141] como se muestra a coninuación.

```
t_coffee -in=miR.aln -mode=rcoffee -method=slow_pair -output=score_pdf score_ascii -run_name=miR_R_Coffee
```

Y para colorear los alineamientos en base a su conservación, utilizamos el siguiente comando:

```
t_coffee -other_pg seq_reformat -in miR.aln -in3 miR.aln -action +3evaluate idmat -out= miR.aln_R_Coffee.pdf
```

El plegamiento de los precursores para determinar la estructura secundaria lo realizamos utilizando el paquete RNAfold (versión 2.1.1) [83].

Búsqueda de motivos

La búsqueda de motivos conservados la realizamos utilizando la herramienta MEME (versión 4.10.1) [15]. Hicimos dos búsquedas de motivos, la primera para encontrar como máximo 2 motivos conservados en los precursores, de alrededor de 20nt. Utilizamos el comando y los parámetros como se muestra a continuación:

```
mem miR.fasta -dna -oc miR -nostatus -time 18000 -mod zoops -nmotifs 2 -minw 19 -maxw 23 -revcomp
```

Y luego hicimos otra búsqueda de motivos, pero esta vez con tamaño más variable (de 5 a 50 nt) y permitimos encontrar hasta 10 motivos con un umbral de corte por e-value. Utilizamos el comando y los parámetros como se muestra a continuación:

```
mem miR.fasta -dna -oc miR -nostatus -time 18000 -mod anr -nmotifs 10 -minw 5 -maxw 50 -revcomp -evt 1e-3
```

Representación gráfica de los precursores

En base a la información de los alineamientos y estructuras secundarias, representamos los datos utilizando el paquete Circos [72]. Por simplicidad en la representación de los

precursores, eliminamos los gaps y sólo están representadas las bases dentro del precursor. El archivo de configuración principal se muestra en el anexo en 9.1 y los archivos de configuración del histograma y las relaciones se muestran en el anexo en 9.2 y 9.3 respectivamente. Los archivos karyotype.txt, highlight.txt y link.txt son distintos para cada precursor en particular y se realizaron mediante un script, con el formato que se especifica en <http://circos.ca/>.

Código fuente

Para todo el trabajo de Tesis se utilizaron la mayoría de programas de Software libre y código abierto en un entorno Linux. Los programas, scripts y algoritmos desarrollados en esta Tesis están disponibles para su uso en el siguiente repositorio....

Capítulo 1

Aplicaciones bioinformáticas para el estudio de interacciones miARN-gen blanco

4.1 Introducción

Los miARNs son ARN pequeños de alrededor de 21 nucleótidos, son importantes reguladores de la expresión de genes en eucariotas y controlan procesos claves como el desarrollo y la respuesta a estrés. Los miARNs controlan la expresión génica ajustando los niveles finales de las proteínas o eliminando los transcriptos de ARN en la célula. La identidad de los genes blancos está especificada por la molécula de miARN, la cual reconoce una secuencia por complementariedad de bases. En plantas, los miARNs tienen una buena complementariedad con sus blancos y en muchos casos regulan más de un gen de una misma familia [64].

Los miARNs conservados en plantas se encuentran generalmente como pequeñas familias de genes que codifican ARN pequeños de secuencias similares o idénticas (Figura 4.1). Una de las ventajas de tener familias génicas con múltiples miembros es que proporciona flexibilidad en la manera en la que cada uno de ellos es regulado [116, 117]. Estas diferencias pueden ser por variaciones en elementos regulatorios en los promotores o en las estructuras de los precursores de miARNs que llevan a eficiencias de procesamiento diferenciales [116].

Por otro lado, las diferencias en las secuencias de los miARNs maduros podrían causar que miARNs de una misma familia regulen diferentes conjuntos de genes. Esto se ha demostrado previamente para los miARNs miR319 y miR159. Estos dos miARNs son muy similares en secuencia, al punto que varios autores los consideran miembros de una misma familia [64]. Sin embargo ha sido demostrado que regulan genes diferentes [103, 104, 123]. Así, mientras el miR319 puede reprimir la expresión de factores de transcripción de la familia TCP y MYB, diferencias de secuencia en 4 nucleótidos previenen la actividad del miR159 sobre los TCP [104].

La identificación de genes blanco regulados por miARNs en plantas es muy importante para poder conocer el rol de los miARNs. En general está identificación de genes blanco,

se obtiene de diferentes estrategias computacionales donde tienen en cuenta la complementariedad con sus mensajeros blanco. Un de los mayores desafío ha sido poder encontrar la mayoría de los genes regulados por estos ARN pequeños con una baja frecuencia de predicciones falsas.

A partir de predicciones de genes blanco y luego validación experimental, se han determinado algunos parámetros empíricos para la identificación de genes blanco. Por un lado, el hecho de que determinados genes con un número de "mismatches" fuesen blancos de regulación por miARNs mientras que otros con igual número no, sugería que debían existir otros factores que afectaran la interacción de los miARN con los ARNm blanco. Es por esto que se han utilizado otros enfoques para abordar este problema, por ejemplo el requerimiento de que el sitio complementario al miARN en los ARNm estuviera conservados entre genes homólogos de Arabidopsis y arroz.

Más recientemente se desarrolló una tecnología, en este caso experimental, conocida como "degradoma", que permite obtener el perfil global de los productos de degradación de los ARN mensajeros [1, 50]. Este enfoque permitió detectar *in vivo* y a nivel genómico la actividad de los miARNs al identificarse los productos de corte de los ARNm mediado por estos ARNs pequeños. Dicha información es muy útil para la identificación de nuevos posibles blancos.

En esta parte de la Tesis se profundizó esta cuestión y se desarrolló un método bioinformático para la predicción de genes regulados por miARNs conservados en plantas.

4.2 Resultados y Discusión

4.2.1 Predicción de genes regulados por miARNs.

Diseño de una estrategia para la identificación de genes blanco regulados por miARNs basado en la conservación evolutiva del par miARN-gen blanco.

Enfocamos nuestro análisis en 22 miARNs que están conservados en Angiospermas [35, 39]. En general estos miARNs están codificados por pequeñas familias hasta 32 miembros. En los genomas completos de Arabidopsis, poplar y arroz es común encontrar variaciones en la secuencia de los miARNs pertenecientes a una misma familia, especialmente en el primer nucleótido y los nucleótidos 20 y 21 [39].

Sin embargo, observamos que la región entre la posición 2 y 19 está bastante conservada y pudimos encontrar una secuencia consenso presente en la mayoría de los miembros de cada

4.2 Resultados y Discusión

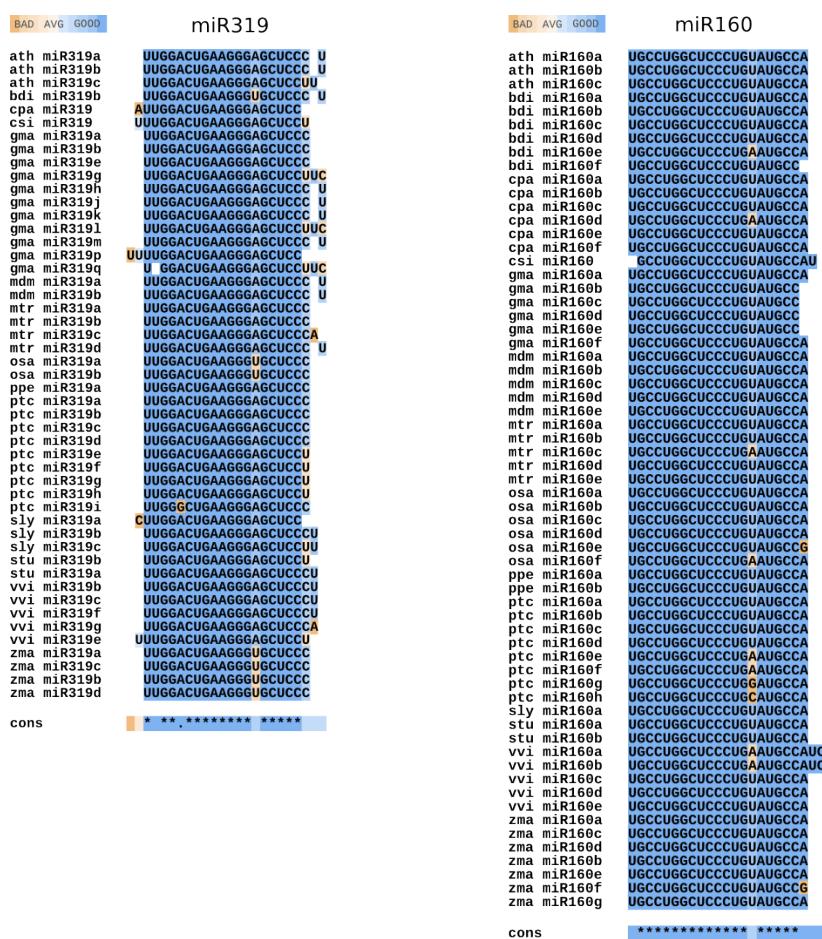


Figura 4.1 Variabilidad de la secuencia del maduro.

familia de miARNs en esas tres especies (tabla 4.4). Curiosamente, las bases variables fuera de esta región conservada son propensas a tener "mismatches" con genes blanco conocidos, lo que indica que podría existir una correlación entre la interacción miARN-gen blanco y la conservación de la secuencia del miARN.

Diseñamos una estrategia para identificar nuevos pares miARN-gen blanco principalmente basada en la conservación evolutiva de la secuencia del gen blanco (Figura 4.2). Las secuencias consenso de 18 nt de cada familia de miARN fueron usadas inicialmente para realizar la búsqueda de genes blanco en contigs de ESTs, de 41 especies de plantas, obtenidos de "Gene Index Project" un proyecto mantenido y administrado por la universidad de Harvard que contiene un catálogo completo de genes en una amplia gama de organismos incluyendo plantas. Además se utilizaron ARNm completos para *A. thaliana* y *Oryza Sativa* (para ver la lista completa de especies, ver tabla 9.8). Utilizando las secuencias consenso de 18nt y permitiendo 3 "mismatches", la búsqueda de genes blanco arrojó como resultado 38.597 genes distribuídos en las 43 especies (Figura 4.2, bin 1). Las interacciones G-U y los bulges fueron considerados como "mismatches" en esta primera búsqueda. Todos los genes blanco de *A. thaliana* conocidos hasta ese momento fueron identificados usando esta estrategia con la excepción de CSD2, un gen blanco del miR398 que contiene 4 mismatches (tabla 9.8).

Teniendo en cuenta que la mayoría de los genes blanco arrojados presentan una escasa descripción del tipo genómica funcional, realizamos un BLASTx contra el proteoma de *A. thaliana*. El "locus ID" obtenido como "best hit" se utilizó como tag (etiqueta) para identificar al candidato en distintas especies (Figura 4.2). A pesar que esta estrategia no necesariamente identifica el gen ortólogo de Arabidopsis, sirve como propósito de clasificación de cada potencial gen blanco de miARN. Aunque la mayoría de los potenciales genes blanco pudieron ser fácilmente asignados con una etiqueta, algunos pocos casos, que incluye a los genes que representan ARNs no codificantes fueron perdidos en este paso.

Este enfoque permite la selección de los mejores candidatos basándose en la presencia de los genes blanco en un número distinto de especies. Utilizando 4 especies como el mínimo de especies requeridas (ya que tiene una buena especificidad), dio como resultado 3.781 genes que corresponden a 533 tags diferentes (Figura 4.2, bin 2).

La búsqueda también se puede hacer en combinación con filtros empíricos de interacción par miARN-gen blanco que tienen en cuenta la energía de interacción y la posición de los "mismatches" (ver Materiales y métodos). De los 38.597 candidatos iniciales, 9.375 pasan estos filtros (Figura 4.2, bin 4). Combinando filtros de energía y filtro de conservación evolutiva, la búsqueda arrojó como resultado 563 candidatos correspondientes a 146 tags (Figura 4.2, bin 5).

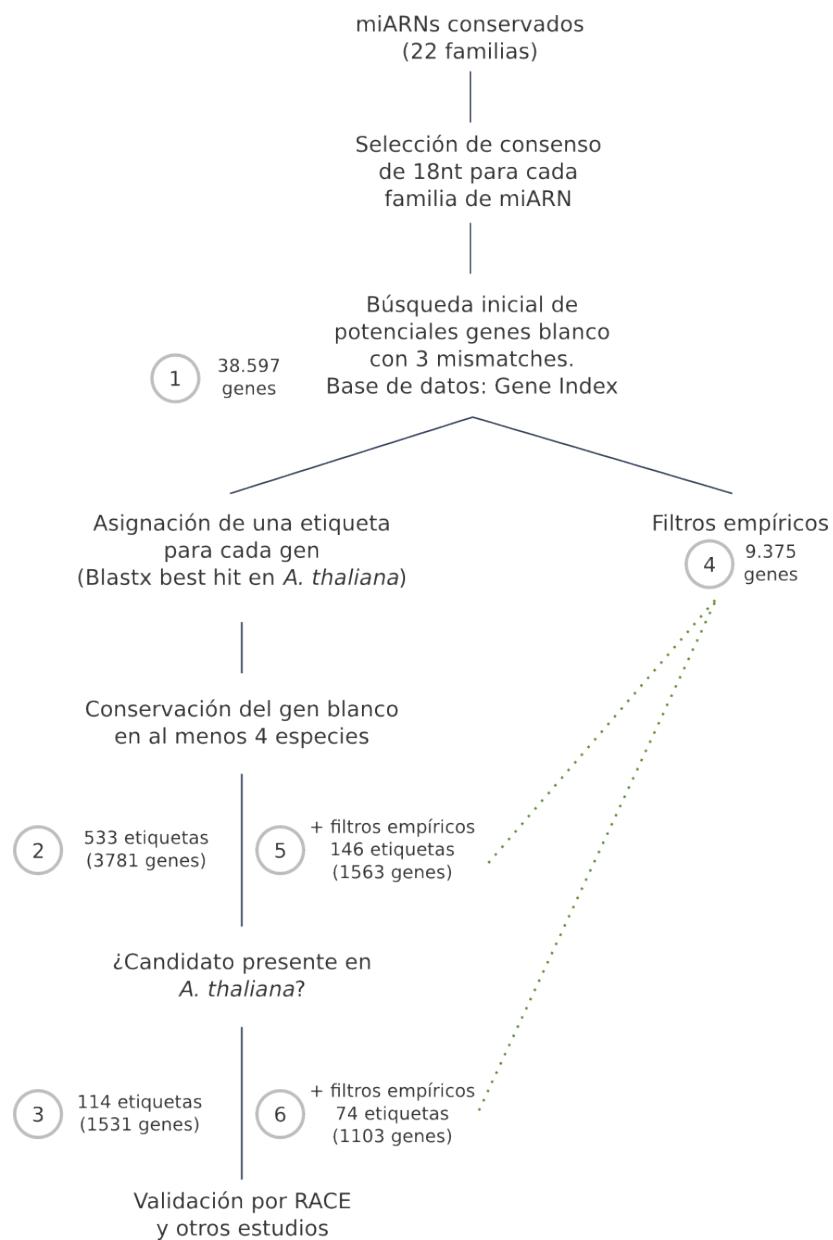


Figura 4.2 Esquema de la estrategia para la identificación de nuevos genes blanco. El número de genes blanco está identificado en cada paso. Luego de aplicar el análisis de conservación, todos los genes que tienen el mismo hit en *Arabidopsis*, fueron considerados como un solo gen blanco. El lado derecho muestra la búsqueda hecha con filtros empíricos: bin 5 y 6 incluyen genes blanco seleccionados con ambos filtros, empíricos y de conservación. Mientras que el bin 2 y 3 muestra los potenciales genes blanco seleccionados sólo con el filtro de conservación.

Table 4.4 miARNs y sus genes blanco en plantas

miARN	Consenso (18 nt)	Targets conocidos ^(a,b)
miR156	GACAGAACAGAGATGAGCA	factores de transcripción SPL
miR159	TTGGATTGAAGGGAGCTC	factores de transcripción MYB, NOZZLE (NZL)
miR160	GCCTGGCTCCCTGTATGC	factores de transcripción ARF
miR162	CGATAAACCTCTGCATCC	DCL1
miR164	GGAGAACAGGGCACGTG	factores de transcripción NAC
miR166	CGGACCAGGCTTCATTCC	factores de transcripción HDZip
miR167	GAAGCTGCCAGCATGATC	factores de transcripción ARF, IAA-ALANINE RESISTANT 3 (IAR3)
miR168	CGCTTGTTGAGGTGGGG	AGO1
mir169	AGCCAAGGGATGACTTGCC	factores de transcripción CCAT-HAP2
mir171	TTGAGCCGTGCCAATATC	factores de transcripción GRAS
miR172	GAATCTGATGATGCTGC	factores de transcripción AP2
miR319	TGGACTGAAGGGAGCTC	factores de transcripción TCP
miR390	AGCTCAGGAGGGATAGCG	TAS RNA
miR393	CCAAAGGGATGCCATTGA	TIR1 proteína, F-BOX proteína
miR394	TGGCATTCTGTCCACCTC	proteínas F-BOX
miR395	TGAAGTGTGTTGGGGAAAC	ATP-sulfurilasas, transportadores de sulfato
miR396	TCCACAGCTTCTTGAAC	factores de transcripción GRF, MMG4.7, FLUORESCENT IN BLUE LIGHT (FLU)
miR397	CATTGAGTGCAGCGTTGA	Laccases
miR398	GTGTTCTCAGGTACCCCC	Cu/Zn SODs, CytC oxidase protein subunit, Chaperona de cobre (CCS)
miR399	GCCAAAGGAGATTGCCC	Enzima E2 de conjugación de ubiquitina
miR408	TGCACTGGCCTTCCCTG	Blue copper proteins, Laccases, P-TYPE ATPase (PAA2), PAC1 (Proteasome component)
miR827	TAGATGACCATCAGCAA	SPX proteína

a Los genes blanco fueron agrupados según sus funciones.

b Nuevos genes blanco validados experimentalmente en este estudio están indicados en negrita.

Parámetros empíricos y de conservación evolutiva pueden actuar de manera sinérgica para identificar genes blanco regulados por miARNs.

Potenciales genes blanco de miARNs fueron clasificados de acuerdo al mínimo número de especie en donde fueron detectados (Figura 2A-E). Como control para cada miARN generamos 10 secuencias “scramble” (al azar), dividiendo las secuencias originales de a di-nucleótidos y luego generando nuevas secuencias al azar conservando la composición de los di-nucleótidos. Estas secuencias al azar fueron utilizadas para realizar búsqueda de genes blanco del mismo modo que lo hicimos para las secuencias originales. La relación señal/ruido fue calculada como el cociente entre el número de genes blanco para los miARNs y el número promedio obtenido de las secuencias al azar. El radio fue de 1,2 para todos los miARNs juntos sin requerir conservación y esa relación incrementa con el número de especie en donde los genes blanco fueron detectados (Figura 4.3 A, recuadro). Los datos para todos los miARNs y sus potenciales genes blanco conservados en al menos 4 especies están incluidos en la tabla 4.5.

Luego estudiamos la selección de candidatos teniendo en cuenta los filtros empíricos. Para esto aplicamos una versión modificada de los filtros descritos anteriormente y requiriendo (i) una energía mínima de hibridación (MFE) de al menos 72% del apareamiento perfecto de cada secuencia consenso y (ii) que sólo un "mismatch" pudiera estar presente entre la posición 1 y la 11 de la secuencia consenso (2-12 del miARN). De la búsqueda inicial 9.375 genes pasaron estos filtros conteniendo el 97% de los genes validados anteriormente de *Arabidopsis*. (Figura 4.2, bin 4).

Al aplicar solamente este filtro empírico, dio como resultado una relación señal/ruido de 1,7, al agrupar todos los miARNs juntos (Figura 4.3 A). Observamos que aplicar si-

multáneamente los filtros empíricos y de conservación aumentaron significativamente la relación señal/ruido para todos los miARNs juntos (Figura 4.3 A recuadro) y también de cada miARN individualmente (Figura 4.3 B-E, recuadros y tabla 4.5). En varios casos, esta relación llega hasta 10 cuando se requiere de que el gen blanco este presente en más de 5 especies y que pase los filtros empíricos (Figure 4.3 A-D). Este efecto sinérgico indica que el filtro de conservación evolutiva y los parámetros empíricos pueden estar seleccionando aspectos diferentes de la interacción miARN-gen blanco.

Observamos que el número de genes blanco candidato y la relación señal/ruido es variable entre los distintos miARNs. El miR396 tiene la mayor cantidad de potenciales genes blanco, 92 de ellos presentes en al menos 4 especies y 26 de ellos pasan además los filtros empíricos (Tabla 4.5 y Figura 4.3 B). El miR408 y el miR398 también tienen un número alto de potenciales genes blanco y buenas relaciones de señal/ruido (Figura 4.3 C-D).

En contraste, ciertos miARNs como el miR162, miR168 y miR399 tienen un solo potencial gen blanco conservado en al menos 4 especies de acuerdo con nuestra búsqueda (Tabla 4.5 y Figura 4.3 E). Al menos en el caso del miR162 y del miR168 este resultado podría estar reflejando su rol específico en la regulación por retroalimentación de la biogénesis del miARN, ya que controlan los niveles de expresión DCL1 y AGO1 respectivamente [140, 147].

Como control adicional para nuestra estrategia hicimos la búsqueda de genes blanco del miR158 y miR173, que son miARNs presentes solamente en *A. thaliana* y especies bien cercanas (17). Como era esperado estos miARNs no generaron más candidatos que sus versiones al azar (Tabla 4.5 y Figura 4.3 F).

Luego chequeamos si los pares miARN-gen blanco altamente conservados tenían una interacción más fuerte que los que están presentes en pocas especies. Para esto calculamos la energía mínima de hibridación para cada interacción detectada en nuestro trabajo. Observamos que los pares miARN-gen blanco presentes en muchas especies tienden a tener energía de interacción mayores que los que están presentes en menos especies (Figure 4.4 A). De todos modos, la correlación no fue notoria y algunas interacciones miARN-gen blanco tuvieron una baja energía de hibridación (Figure 4.4 A). Estos resultados muestran que una alta conservación podría no ser necesariamente equivalente a una fuerte interacción, la misma podría proporcionar una explicación para los efectos sinérgicos causados por los filtros de evolución y empíricos sobre la relación señal/ruido.

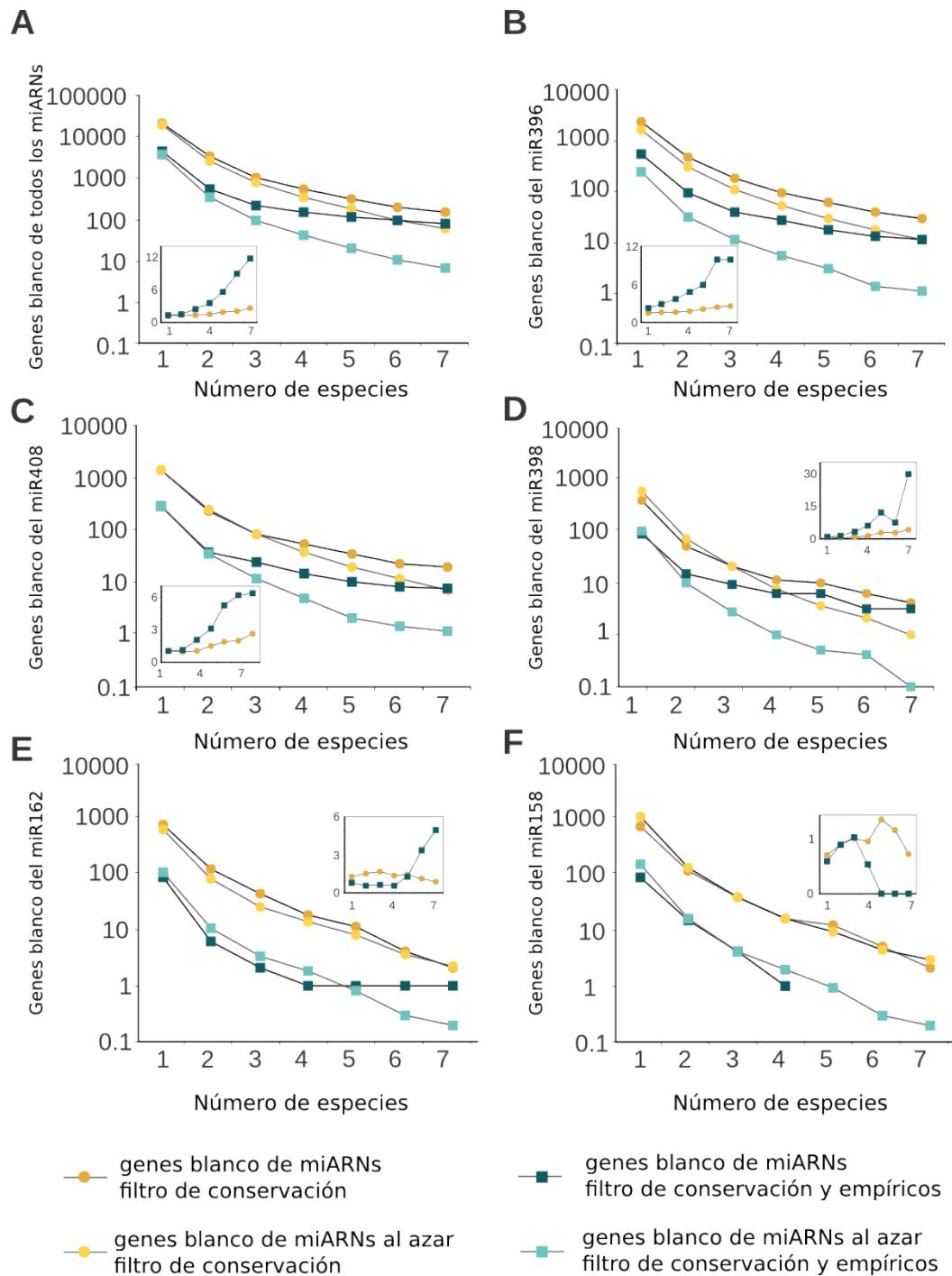


Figura 4.3 Conservación de potenciales genes blanco en distintas especies. Todos los miARNs (A), miR396 (B), miR408 (C), miR398 (D), miR162 (E), miR158 (F). Puntos naranja representan los genes blanco de miARNs usando filtro evolutivo. Puntos amarillos representan los genes blanco de las secuencias al azar usando filtro evolutivo. El cuadrado azul muestra los genes blanco de miARNs luego de aplicar filtros empíricos y evolutivos, mientras que el cuadrado celeste representa los genes blanco de las secuencias al azar en las mismas condiciones. Los recuadros muestran la relación señal/ruido.

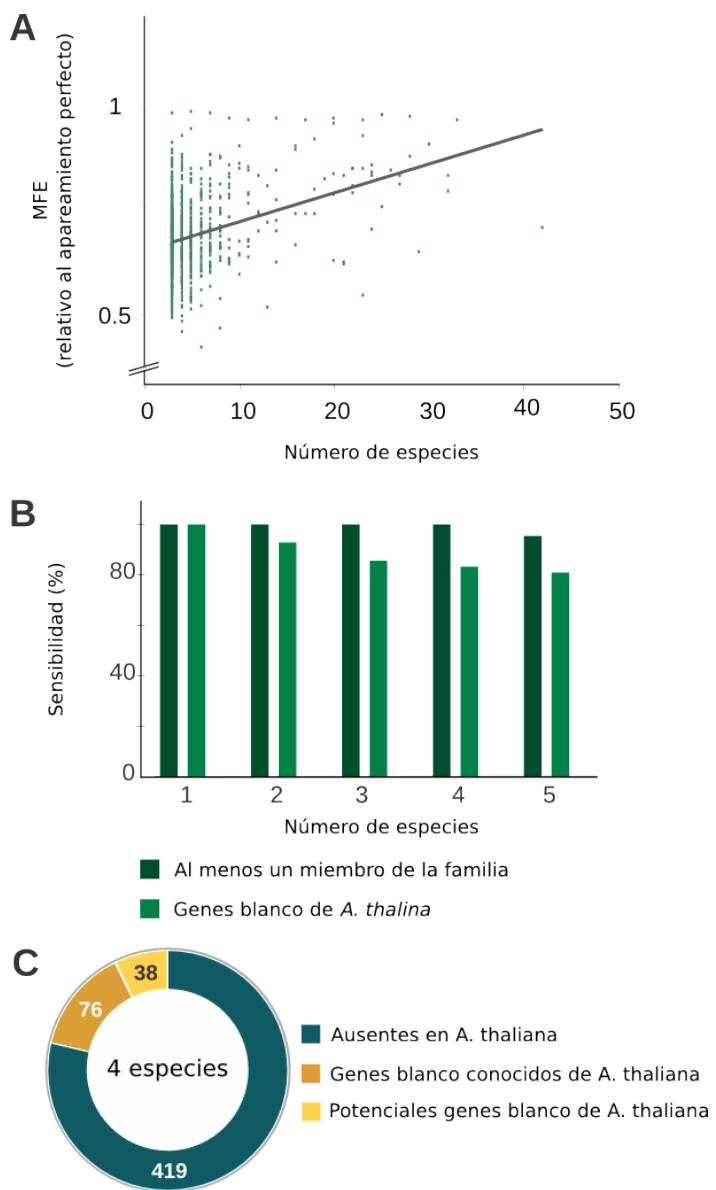


Figura 4.4 Selección de genes blanco por conservación evolutiva de la secuencia. (A) Relación entre MFE y el número de especies en donde cada gen blanco fue detectado. (B) Sensibilidad de la estrategia, analizado de dos modos distintos. Verde clarito: evaluando la presencia de genes validados en *Arabidopsis* y en verde oscuro teniendo en cuenta la presencia de por lo menos un gen blanco de cada familia regulada por miARNs. (C) Clasificación de los potenciales genes blanco presentes en al menos 4 especies.

Table 4.5 Detection of miRNA targets using different filters

	Sin filtros				Filtros empíricos				Conservación 4 especies				Todos los filtros			
	miARN	scramble	ratio	miARN	scramble	ratio	miARN	scramble	ratio	miARN	scramble	ratio	miARN	scramble	ratio	
miR156	3915	3994.4	± 149.9	1.0	890	704.7	± 45.2	1.3	34	39.7	± 3.1	0.9	10	5.4	± 1.1	1.9
miR159	1663	1283.7	± 47.8	1.3	472	254.9	± 21.9	1.9	20	10.1	± 1.1	2.0	6	1.5	± 0.5	4.0
miR160	793	695.6	± 30.5	1.1	277	157.5	± 28.8	1.8	5	4.4	± 0.9	1.1	4	0.5	± 0.3	8.0
miR162	1191	930.2	± 139.5	1.3	108	164.7	± 24.1	0.7	18	13.5	± 3.5	1.3	1	1.8	± 0.5	0.6
miR164	2486	1480.2	± 60.4	1.7	678	333.2	± 32.2	2.0	39	12.4	± 1.9	3.1	12	1.5	± 0.5	8.0
miR166	879	815.5	± 45.0	1.1	231	129	± 14.5	1.8	16	10.6	± 1.4	1.5	6	0.9	± 0.4	6.7
miR167	1777	1364.2	± 146.6	1.3	478	214.8	± 27.5	2.2	22	20.2	± 3.6	1.1	4	1.8	± 0.5	2.2
miR168	962	797.5	± 48.5	1.2	209	185	± 14.2	1.1	6	4.4	± 0.8	1.4	1	1.1	± 0.5	0.9
miR169	1540	1047.2	± 69.7	1.5	464	181.4	± 15.6	2.6	26	11.1	± 2.1	2.3	10	1.2	± 0.2	8.3
miR171	884	723.4	± 32.1	1.2	202	113.8	± 13.4	1.8	7	6.6	± 1.4	1.1	2	0.7	± 0.3	2.9
miR172	3007	1693.7	± 124.7	1.8	540	288.1	± 40.3	1.9	34	17.7	± 1.7	1.9	5	2.2	± 0.6	2.3
miR319	1363	1274.2	± 113.6	1.1	324	249.2	± 22.3	1.3	18	15	± 2.8	1.2	7	1.8	± 0.5	3.9
miR390	873	814.4	± 64.3	1.1	335	173	± 22.5	1.9	8	4.7	± 1.2	1.7	3	0.7	± 0.5	4.3
miR393	986	844.6	± 58.7	1.2	276	124.6	± 11.1	2.2	14	7.1	± 1.2	2.0	5	0.5	± 0.2	10.0
miR394	1569	1531.4	± 57.5	1.0	188	237.1	± 25.0	0.8	26	21.4	± 2.2	1.2	3	2.9	± 0.5	1.0
miR395	1472	1226.7	± 66.7	1.2	426	217.6	± 16.5	2.0	11	8.8	± 1.3	1.3	6	1.3	± 0.3	4.6
miR396	4641	2979.3	± 246.6	1.6	1246	390.5	± 38.8	3.2	92	51.4	± 5.9	1.8	26	5.4	± 1.0	4.8
miR397	1426	1050.9	± 27.9	1.4	368	236.5	± 23.5	1.6	26	9.7	± 0.8	2.7	10	1.6	± 0.3	6.3
miR398	935	834	± 34.5	1.1	376	144	± 18.1	2.6	11	7.5	± 1.6	1.5	6	1	± 0.3	6.0
miR399	1192	1137.6	± 72.0	1.0	275	207.8	± 24.9	1.3	5	13.6	± 1.7	0.4	1	1.5	± 0.7	0.7
miR408	2782	2502.9	± 103.6	1.1	695	468.7	± 50.8	1.5	51	35.1	± 3.0	1.5	14	4.6	± 0.8	3.0
miR827	2261	2000.1	± 119.8	1.1	317	297.1	± 45.0	1.1	44	23.4	± 3.9	1.9	4	2.3	± 0.8	1.7
Total	38597	31021.7	± 1859.8	1.2	9375	5473.2	± 576.3	1.7	533	348.4	± 47.0	1.5	146	42.2	± 11.3	3.5
Control						±										
miR158	1364	1462.8	± 69.1	0.9	170	208.7	± 15.8	0.8	15	16	± 1.7	0.9	1	1.9	± 0.4	0.5
miR173	1386	1232.1	± 101.7	1.1	243	215.6	± 23.4	1.1	11	12	± 2.4	0.9	1	1.5	± 0.4	0.7

a Sin filtros, búsqueda inicial utilizando los miARN consenso de 18nt y 3 mismatches.

b Filtros empíricos, energía de al menos 72% del apareamiento perfecto y 1 mismatch en la posición 2-12 del par miARN-gen blanco.

c Conservación del ID tag en al menos cuatro especies.

d Todos los filtros, combinación de los filtros empíricos y de conservación en al menos cuatro especies.

e miARN, genes blanco para cada miARN específico.

f scramble, promedio de los genes blanco de 10 versiones al azar de cada miARN ± error estándar.

Identificación de nuevos genes blanco en *A. thaliana* por conservación de la secuencia del gen blanco.

Para encontrar nuevos genes blanco nos enfocamos en los genes potenciales que fueron seleccionados de nuestra estrategia utilizando solamente conservación evolutiva, debido a que los parámetros empíricos ya fueron utilizados extensamente en trabajos anteriores. [5, 63, 123]. En primer lugar, analizamos la detección de genes blanco validados previamente en *A. thaliana* [basado en [45]] usando nuestra estrategia y encontramos que el 84% de ellos estaban presentes en al menos 4 especies (Figura 4.4 B). Consideramos esto como un buen resultado ya puede ser que no todos los genes blanco de *Arabidopsis* estén conservados evolutivamente.

Generalmente los miARNs en plantas regulan genes que codifican para proteínas de la misma familia, es por esto que evaluamos si por lo menos un miembro de cada familia era detectado en nuestro enfoque. Encontramos genes blanco pertenecientes a casi todas las familias de genes codificantes para proteínas presentes en cuatro especies (Figura 4.4 B), con la excepción de TAS3, que es regulado por el miR390, al ser un ARN no codificante no es detectado por Blastx.

Para encontrar nuevos genes blanco regulados por miARNs, nos enfocamos únicamente en los potenciales genes blanco conservados en 4 especies, donde una de ellas es *A. thaliana* (Figura 1, bin 1). Genes blanco de miARNs que no están presentes en *A. thaliana* podrían incluir genes que perdieron su regulación durante la evolución o genes que hayan adquirido control por un miARN conservado más reciente en otras especies. La conservación en cuatro especies fue elegida como un filtro evolutivo porque provee buena sensibilidad para genes blanco conocidos.

Identificamos 114 potenciales genes que satisfacen este criterio. Donde 76 de ellos son genes validados anteriormente o genes muy relacionados (Figura 4.4 C). Curiosamente encontramos 38 genes que no tienen relación con genes blanco conocidos de miARNs y decidimos estudiar este grupo con mayor detalle. Nos enfocamos primero en los genes que estaban presentes en un gran número de especies para tener mejor especificidad (Figura 4.3) e intentamos validarlos utilizando 5' RACE PCR modificada [66, 81].

Un potencial gen blanco del MiR408 era At5g21930 que codifica para P-TYPE ATPase OF ARABIDOPSIS 2 (PAA2) y estaba presente en 22 especies distintas incluido monocotiledóneas y dicotiledóneas. MiR408 es inusual debido a que tiene un 5'-A, sin embargo >30% de las secuencias maduro del miR408 corresponden a una variante corrida 1 nt que empieza con 5'-U [91] (Figura 4.5 A). La validación experimental reveló fragmentos de ARNm compatible con este último sitio de corte (Figura 4.5 A). PAA2 es necesaria para

el transporte de iones de cobre a plastocianina [98], y su regulación por el miR408 está relacionada con el rol de este miARN en la homeostasis de cobre [149].

Otro potencial candidato del miR408 era At3g22110 que codifica para PROTEASOME ALPHA SUBUNIT C1 (PAC1) y estaba presente en 20 especies. Por medio de 5' RACE PCR demostramos que este gen es gen blanco del miR408 (Figura 4.5 A). Curiosamente la interacción del par miARN-gen blanco tiene 3 mismatches en la región 5', y se hubiera perdido como potencial gen blanco si se aplicaban solamente los filtros empíricos.

Luego estudiamos los genes blanco del miR396, donde los genes SVP y SUI1 estaban presentes en 29 y 19 especies respectivamente. Pero en ambos casos fallamos al obtener producto de la PCR utilizando 5' RACE PCR modificada. La falta de regulación de este gen por el miR396 podría estar relacionado a la débil energía de hibridación del par miARN-gen blanco, aunque no podemos descartar que el miR396 esté controlando su traducción.

Otros dos potenciales genes blanco del miR396 eran At5g43060 y At3g14110 que codifican para la proteasa MMG4.7 y FLUORESCENT IN BLUE LIGTH (FLU), respectivamente. Y en ambos casos pudimos detectar el corte (Figura 4.5 C y D).

En contraste con el miR408 y miR396, donde tienen varios potenciales genes blanco, obtuvimos un solo potencial gen blanco para el miR159, un factor de transcripción MYB que regula desarrollo del estambre y polen [94]. El otro potencial gen blanco era At4g27330, conocido como NOZZLE/SPOROCYTLESS. Este factor de transcripción, que participa en desarrollo del estambre y óvulo [122, 152], fue también validado por 5' RACE PCR (Figura 4.5 E). Es interesante notar que al menos las funciones de NOZZLE y PAA2 pueden estar directamente relacionadas con el rol de genes blanco, ya descritos anteriormente, del miR159 y miR408 respectivamente.

PAA2, FLU y NOZZLE fueron detectados en mono y dicotiledóneas mientras que PAC1 y MMG4.7 fueron detectadas solamente en dicotiledóneas (Figura 4.5 A-E). Las posiciones del sitio de unión del miARN-gen blanco están altamente conservadas y muchas de las posiciones variables corresponden a mismatches con el miARN o variaciones del tipo G-C/G-U. Además este método no requiere que el sitio del gen blanco esté conservada, sino más bien que haya una interacción predicha con el miARN en distintas especies. De esta manera el sitio de NOZZLE, donde la secuencia cambia en diferentes especies (Figura 4.5 E), pudo ser detectado por este enfoque.

Identificación de nuevos genes blanco permitiendo interacciones G-U.

Los genes blanco identificados utilizando la estrategia descrita anteriormente, tienen varios "mismatches" y "bulges" con sus miARNs, lo que puede ayudar a explicar por que se perdieron en trabajos anteriores. También notamos que muchas de estas nuevas interacciones

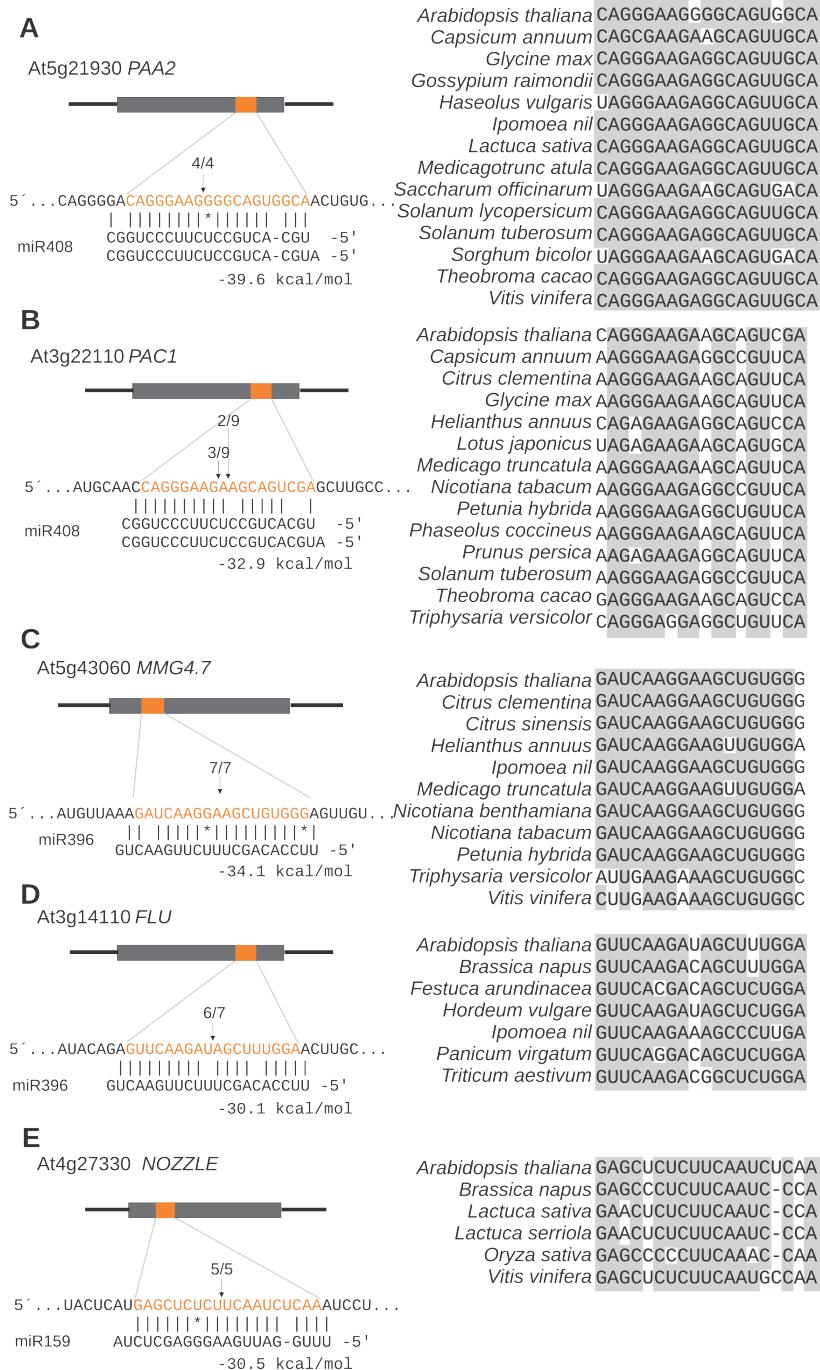


Figura 4.5 Nuevos genes blanco validados en *A. thaliana*. El alineamiento entre el miARN y los nuevos genes blanco identificados se muestran en la izquierda. La conservación evolutiva de la secuencia del sitio reconocido por el miARN en las especies seleccionadas se muestra a la derecha. La figura muestra las interacciones del miR408 con PAA2 (**A**), miR408 con PAC1 (**B**), miR396 con MMG4.7 (**C**), miR396 con FLU (**D**), miR159 con NOZZLE (**E**). Las flechas marcan el sitio de corte determinado por 5'RACE-PCR y los números indican la frecuencia de clonada de cada fragmento.

miARN-gen blanco contenían posiciones que variaban alternadamente entre G-C y G-U en distintas especies. Como consideramos G-U como "mismatch" en nuestra búsqueda inicial, decidimos realizar nuevamente la búsqueda con los miARNs consenso de 18nt pero permitiendo ahora 4 "mismatches", donde al menos uno de ellos tiene que ser del tipo G-U. Esta búsqueda permitiría interacciones miARN-gen blanco con sólo 14 bases apareadas perfectamente.

Para compensar el uso de estos parámetros relajados en términos de "mismatches", requerimos que el gen blanco aparezca en al menos 10 especies distintas para aumentar la especificidad (Figura 4.6 A). Encontramos 125 potenciales genes blanco en *A. thaliana* teniendo en cuenta este criterio (Figura 4.6 A) y 34 de ellos no aparecían en las búsquedas anteriores. El gen blanco CSD2 regulado por el miR398, que no apareció anteriormente, fue detectado con estos parámetros.

Luego examinamos el último grupo de potenciales genes regulados por miARNs que estaban realizando funciones auxiliares a los genes blanco ya descritos para cada miARN. Y encontramos que el miR167 que regula factores de respuesta a auxina (ARFs), también regulaba potencialmente a un gen denominado IAA-ALANINE RESISTANT 3 (IAR3) (Figura 4.6 B y C), que está involucrado en el control de niveles libre de auxina [38, 110].

IAR3 en *Arabidopsis* tiene 3 "mismatches" con respecto al miR167, pero en la posición 12 de la interacción miARN-gen blanco, hay una interacción G-U en varias especies (Figura 4.6 B y C). La técnica de 5' RACE PCR confirmó que el gen realmente era gen blanco del miR167 (Figura 4.6 C).

Identificación de genes blanco específicos de Solanaceae.

Pensamos que la estrategia mostrada también se puede utilizar para encontrar genes blanco presentes específicamente en un grupo de especies relacionadas. Por lo tanto intentamos demostrar esto, encontrando potenciales genes blanco específicos de la familia de *Solanaceae*.

Elegimos esta familia en particular, ya que 6 especies estaban bien representadas en la biblioteca utilizada. La relación señal/ruido entre los genes blanco y las secuencias al azar era más de 2 cuando el filtro empírico o de conservación (en al menos 3 de las 6 especies *Solanaceae*) fueron aplicados (Figura 4.7 A). Curiosamente, al aplicar ambos filtros dio como resultado una relación señal/ruido por encima de 6 (Figura 4.7 A), confirmando nuestros previos hallazgos de que ambos filtros mejoran la detección de genes blanco de miARNs.

Encontramos 132 potenciales genes blanco presentes en al menos 3 especies *Solanaceae*. De este grupo, 41 genes no fueron detectados en otras especies (Figura 6B). El gen blanco más común fue la metalotioneína MT2A, presente en las 6 *Solanaceae*, como potencial gen

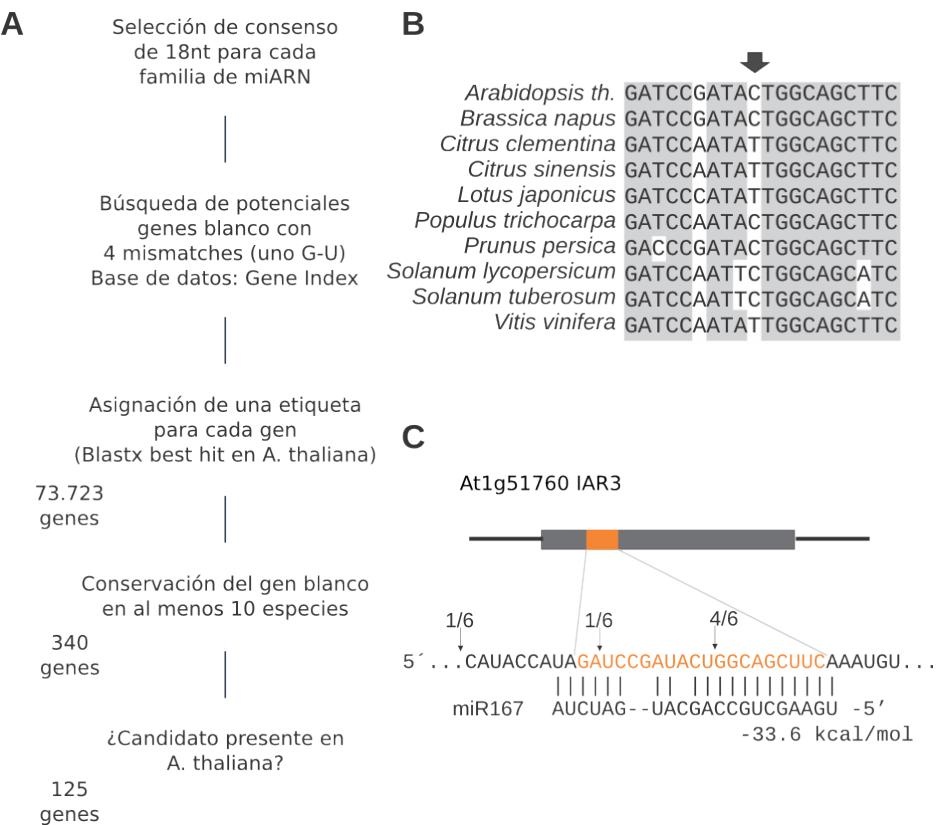


Figura 4.6 Identificación de un nuevo gen blanco de miARN, relajando los parámetros de interacción pero incrementando el parámetro de conservación evolutiva. (A) Esquema de la estrategia modificada para identificar genes blanco de miARNs. (B) Conservación del sitio blanco reconocido por el miARN en distintas especies. La flecha indica una variación de G-C o G-U con el miARN dependiendo de la especie. (C) Alineamiento en *A. thaliana* del gen blanco IAR3 con el miR167. La flecha indica la posición del corte indicada por 5'RACE-PCR y el número indica la frecuencia de clonado de cada fragmento.

blanco del miR398, mientras que MT2B, homólogo de este gen, fue detectado en 5 especies (Figura 4.7 B-D).

Luego, aprovechamos las plantas transgénicas de tabaco que contienen un transgén 35S.mir398 (A.F. Lodeyro, N. Carrillo y J.F. Palatnik resultados no publicados) y chequeamos la expresión de estos genes. Encontramos que CSD2, un gen blanco conservado del miR398, disminuía su expresión > 10 veces en las plantas transgénicas 35S:miR398 comparadas con la planta salvaje (Figura 4.7 E). Curiosamente, observamos que tanto MT2A como MT2B disminuyeron sus niveles de transcripción > 5 veces en estas plantas (Figura 4.7 E). Estos resultados concuerdan con la regulación de MT2A y MT2B por el miR398, aunque no necesariamente demuestra una interacción directa. Además, estos resultados demuestran que los genes blanco presentes en un grupo específico de especies pueden ser encontrados utilizando esta estrategia.

4.2.2 comTAR: una herramienta para la predicción de genes blanco regulados por miARNs en plantas.

A partir de la estrategia descrita en el capítulo anterior, que fue utilizada para encontrar y validar experimentalmente genes blanco regulados por miARNs en *A. thaliana*, desarrollamos una herramienta web denominada comTAR¹¹ (Conserved plant miRNA target prediction tool) [33]. La misma se puede utilizar para predecir potenciales genes blanco regulados por miARNs en plantas y está basada en la conservación evolutiva del par miARN-gen blanco con un número relajado de "mismatches". ComTAR permite distintas opciones/parámetros de búsqueda que pueden ser modificados por el usuario:

- Filtro de mismatch: Solamente un "mismatch" está permitido entre la posición 1 y la 11 de la secuencia del miARN consenso. (Sí/No).
- Corte por energía de hibridación: Se define que un gen blanco es predicho si la mínima energía de hibridación está por debajo del corte elegido.
- El número mínimo de especies donde un mismo TAG está presente para un miARN particular.

Realizar la búsqueda de potenciales genes blanco de miARN

Esta es la búsqueda por defecto. El usuario puede realizar la búsqueda de genes blanco de miARNs conservados. En la primer pantalla se muestra los potenciales genes blanco para un

¹¹<http://rnabiology.ibr-conicet.gov.ar/comtar>

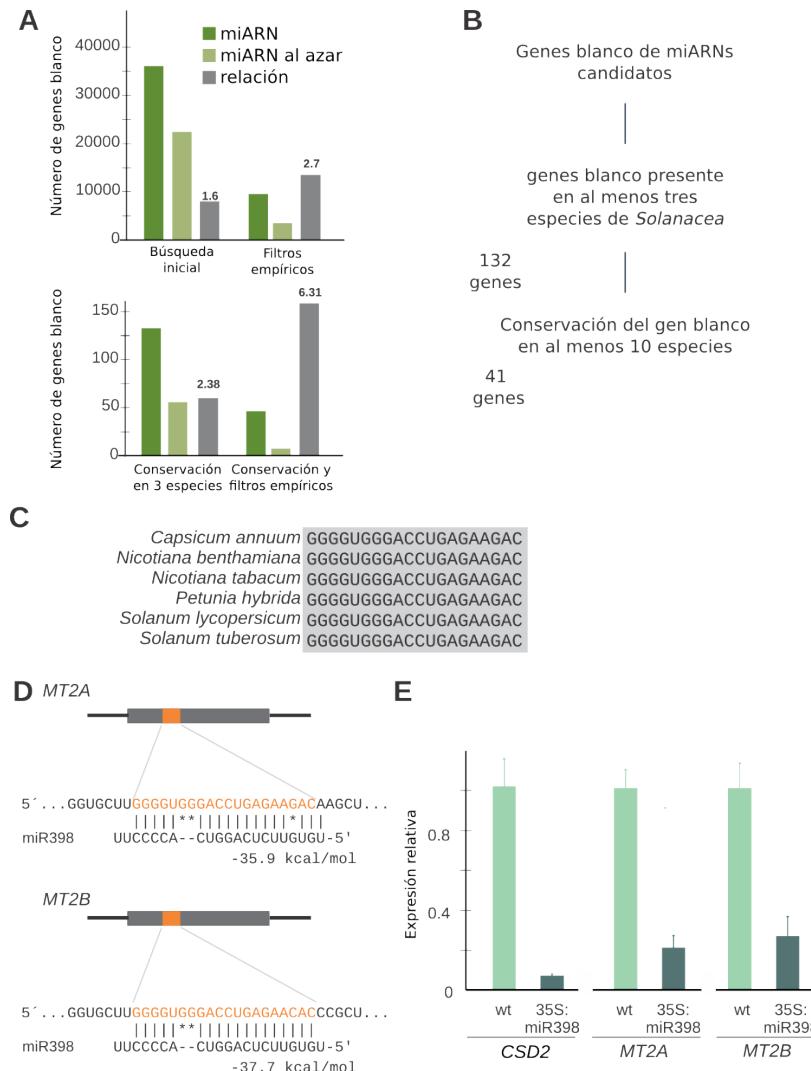


Figura 4.7 Identificación de genes blanco de miARN, específicos de *Solanaceae*. (A) Predicción de genes blanco de miARN en cinco especies de *Solanaceae*. El número de genes blanco de todos los miARNs conservados juntos se muestra luego de aplicar distintos filtros. También se muestran los genes blanco obtenidos a partir de las secuencias al azar. (B) Esquema que muestra la estrategia para identificar genes blanco específicos de *Solanaceae*. (C) Conservación del sitio reconocido por el mir398 con MT2A específico de *Solanaceae*. (D) Esquema que muestra el sitio de unión entre el miR398 y MT2A y MT2B. (E) Niveles de transcriptos de CSD2, MT2A y MT2B en plantas salvajes y plantas transgénicas de tabaco (cv Petit havana) que sobreexpresan el miR398.

miARN dado (Figura 4.8), con una breve descripción del gen, la familia a la que pertenece y además en cuantas y cuáles especies está presente. También, para cada especie que está presente, se tiene acceso por pantalla al alineamiento del miARN-gen blanco, la energía de hibridación y los filtros empíricos de interacciones conocidas del par miARN-gen blanco (Figura 4.9).

Realizar la búsqueda de familias de potenciales genes blanco de miARN

Debido a que los miARNs en plantas en general regulan genes que codifican a proteínas de las misma familias, la herramienta tiene otra funcionalidad donde permite la búsqueda de genes agrupados por familias en vez de agruparlos por TAG. De este modo genes en distintas especies con diferentes TAG, pero que pertenecen a la misma familia pueden ser detectados como familias de potenciales genes blanco.

Realizar la búsqueda de un gen de interés para ver si es potencial gen blanco de algun miARN conservado

El usuario puede introducir un locus TAG en particular (tanto de Arabidopsis como el 'gene ID' del Phytozome) y se identifica si este gen en particular puede ser un potencial gen blanco de algun miARN y en cuantas especies aparece. En Arabidopsis se utiliza el LocusID como identificador, mientrás que en Phytozome este identificador varía según la especie y se puede ver la precedencia de cada especie en el sitio de Phytozome.

Realizar la búsqueda de un miARN nuevo

En esta parte del programa el usuario puede realizar la búsqueda de nuevos ARNs pequeños teniendo en cuenta que la secuencia introducida tiene que ser de 18nt de largo (posiciones 2-19). Luego de la búsqueda, se da un link al usuario y después de unas horas, cuando haya sido procesado el cálculo, el usuario puede entrar a ese link y navegar los resultados por pantalla.

comTAR
conserved plant miRNA target prediction tool

[Find potential miRNA targets](#)
[Find potential miRNA target families](#)
[Is this gene a potential miRNA target?](#)

Targets

miRNA: miR398_GTTCTCAGGTACCC
MFE cutoff: -29.232 kcal/mol
MM Filter: Yes
Species: All

[Go Back](#)

Arabidopsis Tag [?]	Count [?]	Species [?]	Target description [?]	Gene family [?]	Alignments [?]
AT1G12520	16	Show/Hide	copper chaperone for SOD1		View
AT1G08830	9	Show/Hide	copper/zinc superoxide dismutase 1		View
AT3G15640	4	Show/Hide	Rubredoxin-like superfamily protein		View
AT1G80230	3	Show/Hide	Rubredoxin-like superfamily protein		View
AT2G26975	3	Show/Hide	Ctr copper transporter family		View
AT2G40340	3	Show/Hide	Integrase-type DNA-binding superfamily protein	AP2-EREBP Transcription Factor Family	View
AT3G27200	3	Show/Hide	Cupredoxin superfamily protein	Miscellaneous Membrane Protein Families	View
AT3G43860	3	Show/Hide	glycosyl hydrolase 9A4	Glycoside Hydrolase Gene Families	View
AT4G00050	3	Show/Hide	basic helix-loop-helix (bHLH) DNA-binding superfamily protein	basic Helix-Loop-Helix (bHLH) Transcription Factor	View
AT5G12220	3	Show/Hide	las1-like family protein		View

Figura 4.8 Resultados de la búsqueda de comTAR, con parámetros por defecto para el miR398

Sequence ID	Species	5'-target-3' Alignment 3'-miRNA-5' [?]	MFE [?]
Aquca_013_00504.1	Aquilegia coerulea	TGGGCACCTGGAAACAT * * * * CCCCACTGGACTCTTG	-31.7
471402	Arabidopsis lyrata	TGGGAGACCTGGAAACAC * * * CCCCACTGGACTCTTG	-32.1
AT1G12520.1	Arabidopsis thaliana	TGGGAGACCTGGAAACAC * * * CCCCACTGGACTCTTG	-32.1
Bradi5g18900.3	Brachypodium distachyon	TTGGTACCTGGGAACGC ** * * CCCCACTGGACTCTTG	-33.5
Bra026968	Brassica rapa	TGGGCACCTGGAAACAC * * * CCCCACTGGACTCTTG	-32.5
Carubv10011816m	Capsella rubella	TGGGAGACCTGGAAACAC * * * CCCCACTGGACTCTTG	-32.1
evm.model.supercontig_29.47	Carica papaya	TAGGTACCTGAGAACAT ** * * CCCCACTGGACTCTTG	-34.2
Ciclev10021134m	Citrus clementina	TTGGTACCTGGGAACAC * * * CCCCACTGGACTCTTG	-33.9
orange1.1g020436m	Citrus sinensis	TTGGTACCTGGGAACAg ** * * CCCCACTGGACTCTTG	-32.7

Figura 4.9 Parte de la salida de comTAR mostrando el par miR398/SOD1 (At1g12520) en diferentes especies

Capítulo 2

Estudios genómicos sobre la biogénesis de precursores de miARN en plantas

5.1 Introducción

Los miARNs son una clase de ARNs de 20-22nt de largo que son originados de genes endógenos y regulan otros ARNs por complementariedad de bases [143]. Se distinguen de otros ARNs pequeños por su biogénesis única que involucra el corte preciso del precursor, para liberar el miARN maduro. Aunque la evidencia actual indica que los miARNs han surgido y especializada de forma independiente en animales y las plantas, su biogénesis depende del reconocimiento de claves estructurales ubicadas en los precursores de miARN [14, 20, 35].

En nuestro grupo actualmente se está estudiando la biogenesia de miARNs, específicamente como los precursores son reconocidos y procesados en plantas [21]. Estos precursores tienen una estructura de tallo-burbuja característica [64], que se cree que proporciona las claves para el procesamiento del mismo y la liberación de los ARN pequeños de 21 nt.

Mientras que los precursores de miARN en animales tienen estructuras homogéneas, los precursores de miARNs en plantas constituyen una amplia gama de estructuras, y sus longitud pueden variar entre 50 y 900 nucleótidos [21, 35]. Esa variabilidad se da entre distintas familias de precursores, pero a veces también entre una misma familia de precursores en distintas especies (Figura 5.1).

Además, en las plantas, las mutaciones que impiden la biogénesis o actividad de los miARNs, tales como *hyll1*, *henl1* y *ago1*, conducen al aumento en los niveles de los transcriptos de muchos de los genes blanco de miARNs (aproximadamente un 45% de todos los blancos) [5, 24, 60, 151]. Esto sugiere que el mecanismo que involucra el corte y degradación de los ARNm es un componente importante de la represión inducida por los miARNs en plantas [64, 143].

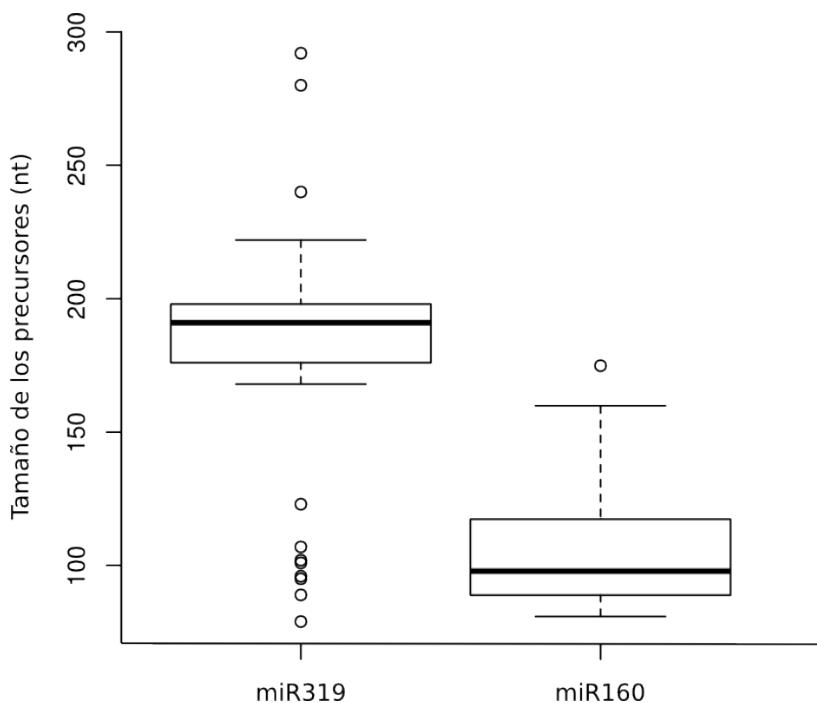


Figura 5.1 Tamaño de precursores. Se muestra el tamaño (nt) de dos familias de precursores en distintas especies.

En particular, la biogénesis de los miARNs es un proceso clave porque determina la secuencia exacta de nucleótidos del ARN pequeño funcional. Si bien en el caso de animales está claro cuáles elementos estructurales son reconocidos en los precursores durante su procesamiento, poco se sabía sobre el reconocimiento de los precursores de plantas por la maquinaria de procesamiento.

Muchos precursores en plantas tienen un tallo de \sim 15 nt debajo del duplex miARN/miARN* seguido por un loop interno, que sirve como una señal estructural de reconocimiento por la maquinaria de procesamiento [59, 90, 126, 144]. Sin embargo, este determinante de procesamiento no se encuentra en todos los precursores [90]. Además, la biogénesis de los miARNs conservados evolutivamente como ser el miR319 y miR159 comienzan con un corte al lado del loop interno y continúa con 3 cortes adicionales en una dirección de burbuja a base hasta que finalmente el miARN es liberado [4, 21]. Se ha demostrado que otros precursores de plantas liberan otros ARNs pequeños además del miARN [73, 159], aunque los mecanismos de procesamiento subyacentes eran desconocidos.

5.2 Resultados y Discusión

5.2.1 Procesamiento de precursores de miARNs en plantas

En esta parte del proyecto de tesis y en el marco de una colaboración con el grupo del Dr. Blake Meyers (Delaware, USA), el cual se especializa en secuenciación y análisis de ARN pequeños, nos propusimos entender cómo se procesan los precursores de miARNs plantas. Colegas del laboratorio realizaron una estrategia para analizar sistemáticamente intermediarios de procesamiento de miARNs y caracterizar la biogénesis de la mayoría de los miARNs conservados presentes en *A. thaliana* mediante técnicas de secuenciación de alto rendimiento, utilizando los equipos de última generación disponibles en Delaware (USA). Esta técnica desarrollada en el laboratorio se conoce como SPARE [120] (del inglés Specific Parallel Amplification of RNA Ends). Utilizando esta técnica encontramos que los miARNs son procesados por cuatro mecanismos, dependientes de la dirección secuencial de la maquinaria de procesamiento y del número de cortes requeridos para liberar el miARN. La clasificación de los precursores, teniendo en cuenta los mecanismos de procesamiento, reveló determinantes estructurales específicos para cada grupo. Se encontró que la complejidad de las vías de procesamiento de miARN se produce tanto en precursores jóvenes como en conservados y que los miembros de la misma familia pueden ser procesados de diferentes maneras. Además hemos observado que diferentes determinantes estructurales compiten por la maquinaria de procesamiento y que miARNs alternativos pueden ser generados a partir de un único precursor. Los resultados ofrecen una explicación para la diversidad estructural de los genes de precursores de miARN en plantas y nuevas perspectivas hacia la comprensión de la biogénesis de los ARNs pequeños [21].

Análisis de datos y precursores detectados

Mediante la cantidad de cortes detectados la técnica de SPARE permite definir si el mecanismo es base a loop o loop a base. Esta técnica arroja una gran cantidad de datos producto de la secuenciación de alto rendimiento, por lo que se necesita de un enfoque bioinformático para la interpretación de los resultados. Por la gran cantidad de precursores a estudiar y el número de bibliotecas se necesitó un análisis previo de los datos y una forma de presentarlos. Para esto construimos e implementamos un pipeline bioinformático utilizando "in-house" scripts y datos disponibles de miRBASE para poder analizar los datos de las bibliotecas de deep-sequencing obtenidos a partir de la técnica de SPARE.

Un precursor fue considerado como detectado si más de tres lecturas corresponden a la secuencia de ese precursor. De esta manera encontramos fragmentos de ARN que corresponden

a 129 precursores, 71 de ellos de miARNs conservados y 58 de miARNs jóvenes (Figura 5.2). Mediante el análisis de los datos arrojados de la estrategia bioinformática pudimos definir la dirección de procesamiento de la mayoría de los precursores en *A. thaliana*. De los cuales 32 de ellos fueron definidos como procesados por un mecanismo de base a loop, ya que se encontraron los cortes en la parte proximal del duplex miARN/miARN* sin detectar cortes en la parte de arriba del duplex, como en el caso del miR168a, miR172b y el miR395b (Figura 5.3). Además encontramos 16 precursores de miARNs conservados con cortes detectados ($>5\%$) en el lado distal del miARN/miARN* los cuales fueron definidos como loop a base (Figura 5.4).

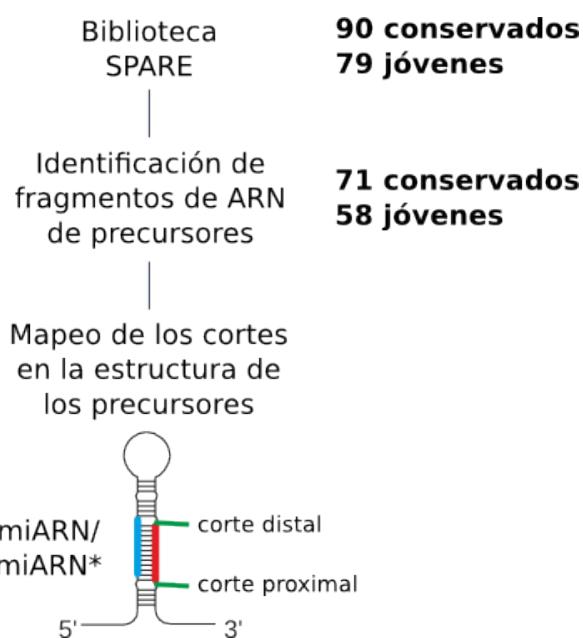


Figura 5.2 Esquema del procedimiento para analizar los datos de SPARE.

Estructura secundaria de los precursores

Para ver si había diferencias estructurales para los precursores con diferentes mecanismos de procesamiento, determinamos la estructura secundaria de precursores detectados que se procesan en dirección base a loop (Figura 5.5) y los que se procesan loop a base (Figura 5.6). Obtuimos las estructuras secundarias para cada precursor. Definimos a una coincidencia en cada posición con un 0, mientras que "bulges" y "mismatches" los consideramos como 1. El lado proximal del duplex miARN/miARN* fue definido como la posición +1 y analizamos desde la posición -25 a la posición +40 (Figura 5.5 y 5.6).

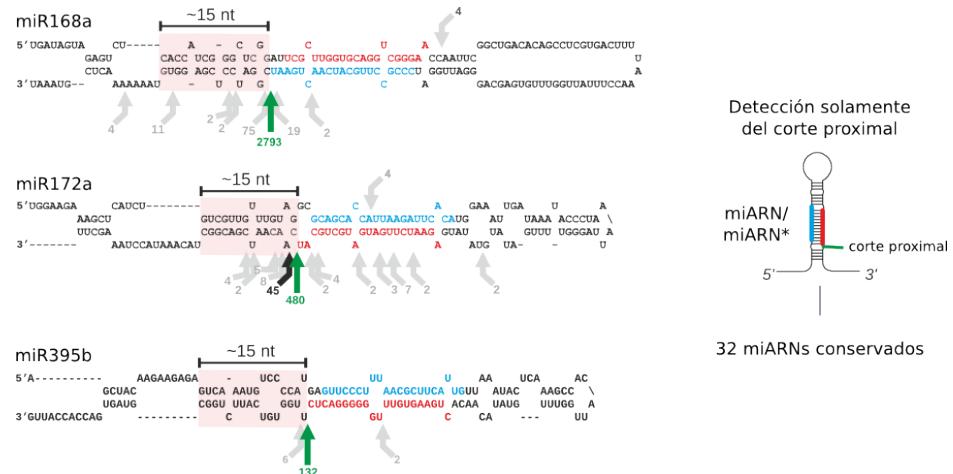


Figura 5.3 Identificación y caracterización de precursores de miARNs procesados de base a loop. Esquema donde se muestra la estructura secundaria del miR168a, miR172b y miR395b. Las flechas indican la posición y número de lecturas de los cortes del precursor identificado. Flechas en verde muestran el corte más abundante, que también coincide con al corte proximal del miARN/miARN*. Flechas en negro muestran otros cortes con al menos 5% de abundancia del número total de cortes, mientras que otros cortes minoritarios se muestran con una flecha gris. Con rosa se resalta el stem de 15nt debajo del corte proximal. El miARN se indica en color rojo y el miARN* en color azul. La imagen de la derecha muestra el patrón de corte típico detectado en la biblioteca de SPARE para estos precursores.

Procesamiento de base a loop

Consideramos la estructura secundaria de 32 precursores analizados en esta parte del trabajo que se procesan de base a loop y todos ellos tienen un claro tallo inferior de 15 nt de largo (Figura 5.5). Además este tallo se pudo ver tanto para los precursores validados experimentalmente que se procesan de base a loop como para todos los precursores conservados (Figura 5.5 en violeta). Pero pudimos observar que las bases inmediatamente debajo del duplex miARN/miARN* (posición -2 y -1) tienden a estar desapareadas (Figura 5.5). Además las posiciones -3 y -4 y las 3 últimas posiciones del stem inferior (-13,-14 y -15) están apareadas casi siempre (Figura 5.5). En general, nuestros resultados muestran que los precursores procesados en una dirección base a loop son más uniformes de lo que se pensaba previamente y que al menos algunos de los precursores no detectados como base a loop probablemente tengan otros determinantes específicos de ARN.

Procesamiento de loop a base

Estos precursores, que tienen un procesamiento loop a base, tienen un corte mayoritario que se puede detectar en nuestras bibliotecas. Este corte es el esperado en la dirección de

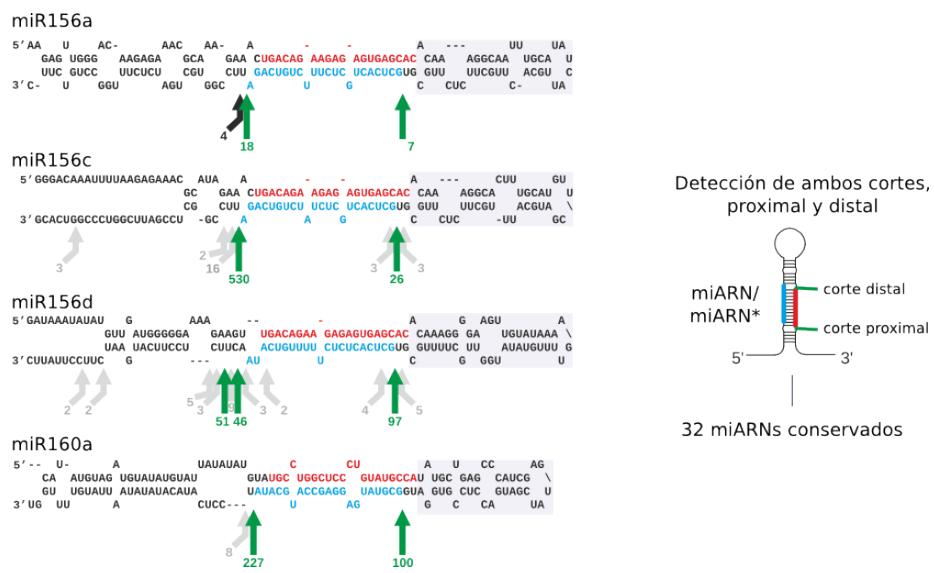


Figura 5.4 Identificación y caracterización de precursores de miARNs procesados de loop a base. Esquema donde se muestra la estructura secundaria del miR156a, miR156c, miR156d y miR160a. Las flechas indican la posición y número de lecturas de los cortes del precursor identificado. Flechas en verde muestran el corte más abundante, que también coincide con al corte proximal del miARN/miARN*. Flechas en negro muestran otros cortes con al menos 5% de abundancia del número total de cortes, mientras que otros cortes minoritarios se muestran con una flecha gris. Con gris se resalta el stem de arriba del miR156 y miR160. El miARN se indica en color rojo y el miARN* en color azul. La imagen de la derecha muestra el patrón de corte típico detectado en la biblioteca de SPARE para estos precursores.

procesamiento de precursores con un mecanismo de loop a base. Con la excepción de dos miARNs (miR396a y miR162b) estos precursores no tienen una estructura obvia debajo del duplex miARN/miARN* (Figura 5.6). Estos precursores tienen una región terminal estructurada (Figura 5.6), que tiene un tamaño homogéneo de 42nt que incluye un loop corto en contraste con la misma región en los precursores que se procesan de base a loop donde es más variable (Figura 5.5 y 5.6).

En esta segunda parte del proyecto de tesis presentamos un estrategia y realizamos un análisis sistemático para la identificación de la biogénesis de precursores de miARNs desde un punto de vista genómico. De esta manera pudimos encontrar la dirección de procesamiento de la mayoría de los precursores de miARNs en *A. thaliana*.

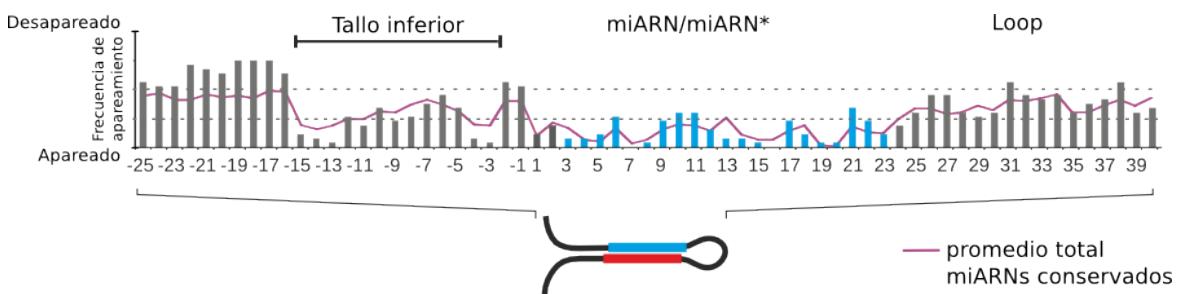


Figura 5.5 Estructura secundaria de precursores detectados que se procesan en dirección base a loop. Los matches en cada posición los consideramos como 0, mientras que "bulges" y "mismatches" fueron considerados como 1. La estructura secundaria considerando todos los miARNs conservados se indica con color violeta.

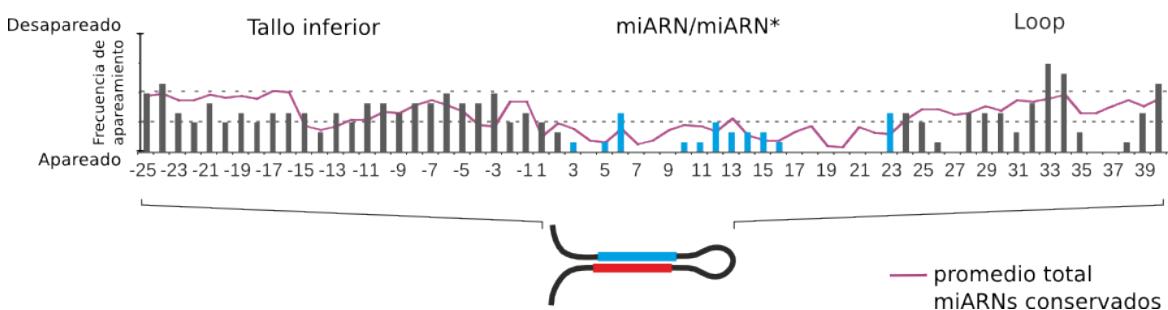


Figura 5.6 Estructura secundaria de precursores detectados que se procesan en dirección loop a base. Los matches en cada posición los consideramos como 0, mientras que "bulges" y "mismatches" fueron considerados como 1.

Estos resultados fueron publicado en la revista Genome Research [21]. En este mismo artículo se pudo demostrar que los precursores de miARNs en plantas, se pueden agrupar por cuatro mecanismos de procesamiento con distintas características (Figura 5.7).

- En los precursores con un mecanismo **corto de base a loop**, un loop interno seguido por un tallo inferior de ~15nt especifica la posición del primer corte. Esta estructura se encuentra en la mayoría de familias de miARNs [90, 125, 144]. A pesar de que el tallo puede contener bulges, la transición de un loop interno (simple hebra) al tallo inferior es bastante marcada, y tres pares de bases apareadas generalmente definen el comienzo del tallo inferior del precursor [21]. El segundo corte procede a una distancia fija de ~21 nt desde la posición del primer corte.
- En los precursores con un mecanismo **secuencial de base a loop** (ej: familia del miR169), el primer corte procede como en los cortos de base a loop, pero luego son necesarios dos cortes más para liberar el miARN, generando en el proceso niveles bajos de RNA pequeños adicionales [21].
- En los precursores con un mecanismo **cortos de loop a base** (ej: familia del miR156 y miR160), el procesamiento es guiado por un tallo superior, y son necesarios dos cortes para liberar el miARN maduro. La región terminal de estos precursores tienen una largo conservado de ~42 donde incluye un loop pequeño [21].
- En los precursores con un mecanismo **secuencial de loop a base** (ej: familia del miR319 y miR159), cuatro cortes secuenciales por DCL1 son los encargados de procesar los precursores de miARNs. En general muestran un tallo largo superior, del cual otros ARNs pequeños son generados [4, 19, 21]

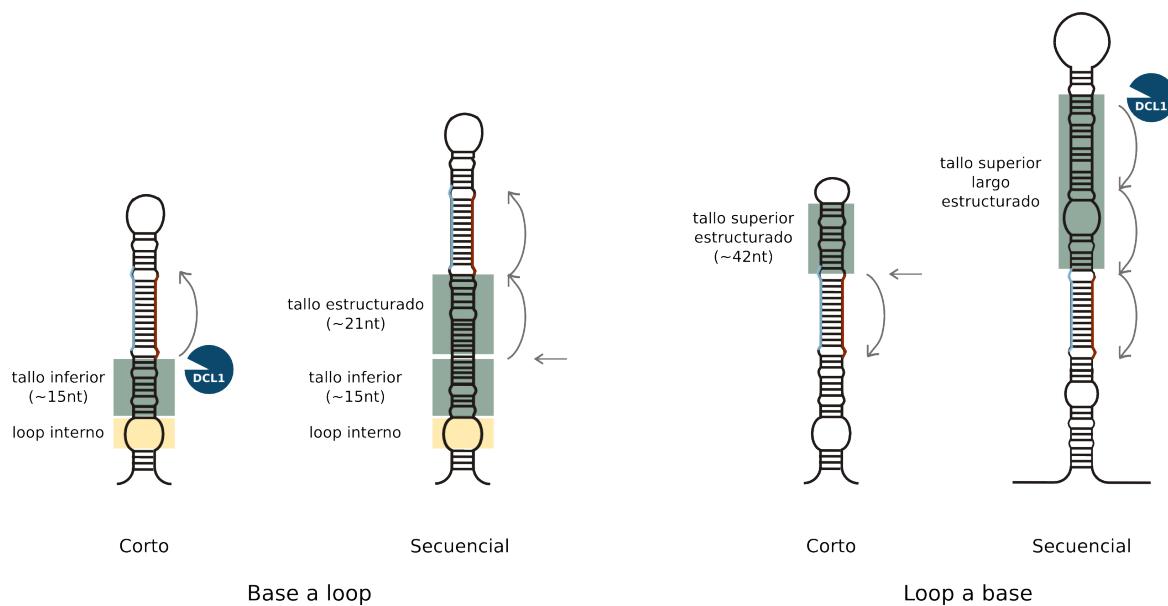


Figura 5.7 Distintos mecanismos de procesamientos de miARNs en plantas

5.2.2 Identificación de intermediarios de procesamiento en plantas mutantes en proteínas de procesamiento

La técnica de SPARE (del inglés Specific Parallel Analisys of 5'RNA Ends) fue desarrollada con el objetivo de caracterizar el procesamiento de los precursores de miARNs de *A. thaliana*. Para esto se identifican los intermediarios de procesamiento de cada uno de dichos precursores, y uniendo la información brindada por los diferentes fragmentos es posible dilucidar tanto la dirección, como el número de cortes requeridos para la biogénesis de cada miARN (Figura 5.8).

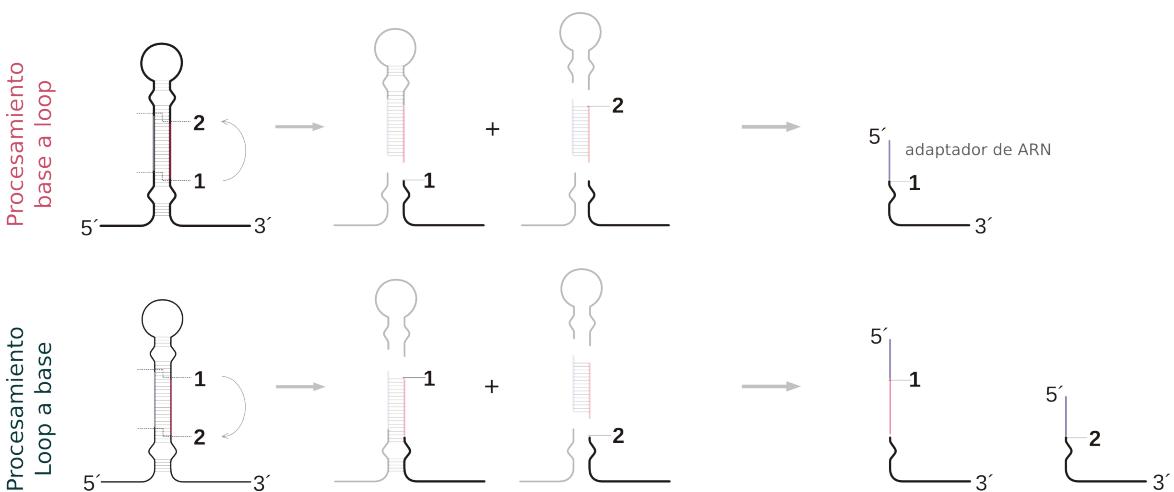


Figura 5.8 Técnica de SPARE para diferenciar mecanismos de procesamiento de precursores. Detalle donde se muestra a nivel molecular como la técnica permite diferenciar entre dos direcciones de procesamiento opuestas.

Brevemente la técnica consiste en emplear los extremos 5'P dejados por la maquinaria de procesamiento en los precursores luego del clivaje para realizar una ligación ARN-ARN entre los intermediarios de procesamiento y un oligo de ARN de secuencia conocida. A continuación se emplean oligos específicos para retro-transcribir los intermediarios de cada precursor en particular, con una cola común a todos. En este punto todos los fragmentos generados durante el procesamiento presentan los mismos extremos: en el 5' la secuencia del oligo de ARN y el en el extremo 3' la secuencia común incorporada durante la retrotranscripción. Esta información es empleada para diseñar oligos (FW y RV) que se emplean en una reacción de PCR a continuación. Finalmente los fragmentos son secuenciados.

En la sección 5.2.1 mostramos los resultados obtenidos de esta técnica para determinar la dirección de procesamiento de los precursores de miARNs en plantas. Esta nueva secuenciación se realizó debido que anteriormente no se pudieron determinar la dirección de procesamiento de algunos precursores. Además, en esta nueva tanda de secuenciación de

precursores se intentó identificar intermediarios de procesamiento en plantas mutantes en proteínas de procesamiento. Para esta parte del trabajo se hicieron algunas modificaciones de la técnica. Se ubicaron los oligos a 100 nt del último corte 3' esperado. En total se diseñaron 121 oligos para retrotranscribir el conjunto de precursores conservados y jóvenes validados hasta el momento. Además, anteriormente el procesamiento de los datos a partir de los datos crudos se hicieron en el mismo laboratorio que se realizaron las secuenciaciones.

Construcción de bibliotecas multiplex

El paso final consiste en secuenciar las bibliotecas construidas, sin embargo la secuenciación es costosa haciendo económicamente difícil secuenciar varias bibliotecas en paralelo para comparar entre diferentes condiciones (siendo este nuestro objetivo con las plantas mutantes en proteínas de procesamiento). En los últimos años se han desarrollado métodos que hacen posible secuenciar más de una biblioteca por calle. Esto consiste en marcar los fragmentos provenientes de cada biblioteca con secuencias específicas (detallado en Materiales y Métodos) y separar informáticamente cada secuencia en la biblioteca de la cual provino a luego de secuenciación. Técnicamente esto se hace agregando una PCR final donde los índices son incorporados en todas las secuencias de cada biblioteca.

Diseño experimental

Con el objetivo de construir bibliotecas de SPARE crecimos plántulas de *Arabidopsis thaliana* durante 10 días con luz continua (independizándonos de este modo la variación asociada al ritmo circadiano a la hora de la colecta de muestra). Se realizó un experimento con plantas mutantes en proteínas de procesamiento.

Plantas mutantes en proteínas de procesamiento

Hyl1, Se y Fiery. Siendo sus controles plantas silvestres de *Arabidopsis* ecotipo Col-O.

Herramienta web para el análisis de intermediarios de procesamiento en plantas

De la técnica de SPARE se obtienen una gran cantidad de datos arrojados por el secuenciador para cada biblioteca secuenciada. Estos datos se conocen como datos crudos y son fragmentos de secuencia que luego tienen que ser mapeados contra el genoma de datos de interés. Además, estos datos crudos cuentan con información de las secuencias, lo que permite mediante programas informáticos hacer un control de calidad de las secuencias obtenidas. En total se obtuvieron cerca de 150 millones de secuencias (Figura 5.9).

Esos fragmentos son agrupados y mapeados contra las secuencias únicas de los precursores de *A. thaliana*. Nos quedamos con 7158 secuencias únicas, que luego las filtramos por largo de la secuencia (Figura 5.9). Esa información ya procesada es la que utilizamos en la herramienta para analizar y comparar los intermediarios de procesamiento para plantas silvestre y mutantes de procesamiento.

Para facilitar el análisis de los datos obtenidos por la técnica de SPARE, creamos una herramienta web disponible en la red interna del laboratorio. La herramienta permite seleccionar un precursor en particular de los ~100 precursores que se detectaron fragmentos por la técnica de SPARE. Luego se selecciona una o varias mutantes y controles que se quieren analizar. Además la herramienta permite modificar otros parámetros como la longitud y abundancia de los fragmentos secuenciados a considerar.

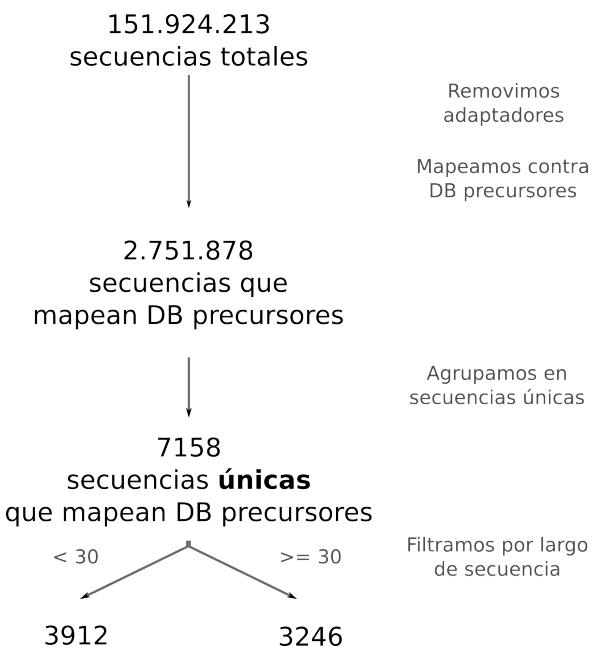


Figura 5.9 Análisis informático utilizado para procesar los datos de SPARE. Las secuencias únicas mapeadas contra los precursores son las que se utilizan en la herramienta para analizar los intermediarios de procesamiento.

Primero vemos la salida de la herramienta para el precursor del miR165a, analizando distintas mutantes de procesamiento al mismo tiempo (Figura 5.10). La posición final del duplex miARN/miARN* fue considerada como la posición 0, y las posiciones positivas son bases hacia arriba del dúplex mientras que las posiciones negativas son bases hacia abajo del dúplex. Los fragmentos se muestran en forma porcentual tomando cada biblioteca de forma independiente y en distintos colores se representan las distintas bibliotecas. Además se muestra una tabla con los valores absolutos de cada fragmento y con flechas verdes se

marcan las posiciones con los fragmentos de mayor abundancia en promedio de todas las bibliotecas (Figura 5.10).

En esta figura se muestra un precursor que es procesado con un mecanismo corto de base a loop, donde según se muestra en la Figura 5.8, los fragmentos detectados corresponden únicamente al primer corte por DCL1 (Figura 5.10). Se pueden observar también cortes en las posiciones ± 2 con respecto a las esperadas por actividad "sloopy" (poco rigurosa) de DCL1 [119]. Además se observan en la posición -5, que aparecen fragmentos en todas las bibliotecas aunque en muy baja abundancia (Figura 5.10).

Mediante esta herramienta no sólo se puede identificar los precursores procesados de base a loop o de loop a base, sino que también permite identificar a los precursores que denominamos secuenciales, que son los que requieren de más de dos cortes por DCL1 para liberar al miARN maduro. Por ejemplo, mostramos al precursor del miR169d que es procesado de forma secuencial de base a loop. En estos precursores, los sitios de cortes se localizan 21 nt por debajo del lado proximal del dúplex miARN/miARN* y luego DCL1 sigue cortando para procesar al precursor. En este caso se detecta un fragmento mayoritario en la posición -21, es decir 21nt por debajo del dúplex miARN/miARN* como era esperado para precursores secuenciales de base a loop 5.11.

El precursor del miR169d tiene una estructura particular donde presenta un loop terminal ramificado. Se ha estudiado este tipo de precursores, con esta estructura, donde se demostró que pueden ser procesados por DCL1 de manera bidireccional de base a loop o de loop a base, resultando en un procesamiento productivo y abortivo respectivamente [161]. Con la herramienta desarrollada por nosotros, se puede observar que este precursor particular presenta fragmentos que provienen del corte secuencial de la base al loop que corresponde a la posición -21 en el gráfico. Pero además, en la mutante de *fiery* se pueden observar fragmentos que corresponden a un corte abortivo por DCL1 a partir del reconocimiento de este loop ramificado 5.11. La herramienta permite discriminar este tipo de procesamiento en distintas mutantes de procesamiento.

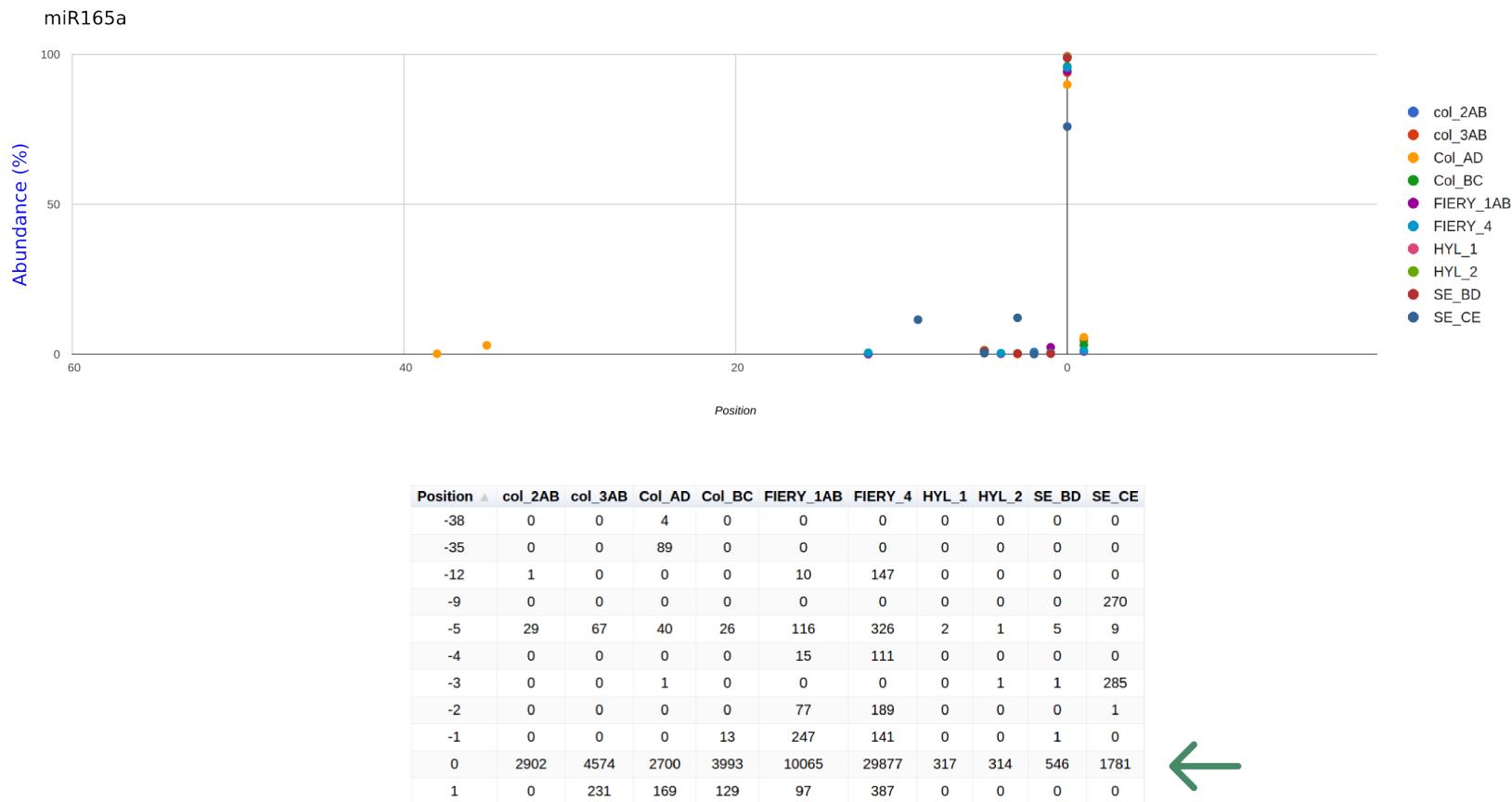


Figura 5.10 Captura de pantalla de la herramienta de SPARE para el miR165a. Porcentaje de fragmentos del precursor (abundancia relativa de los fragmentos en esa posición dividido la abundancia total de los fragmentos en el precursor). La posición final del duplex miARN/miARN* fue considerada como la posición 0. La tabla muestra los valores absolutos de cada fragmento. Las flechas verdes marcan las posiciones del precursor con los fragmentos de mayor abundancia.

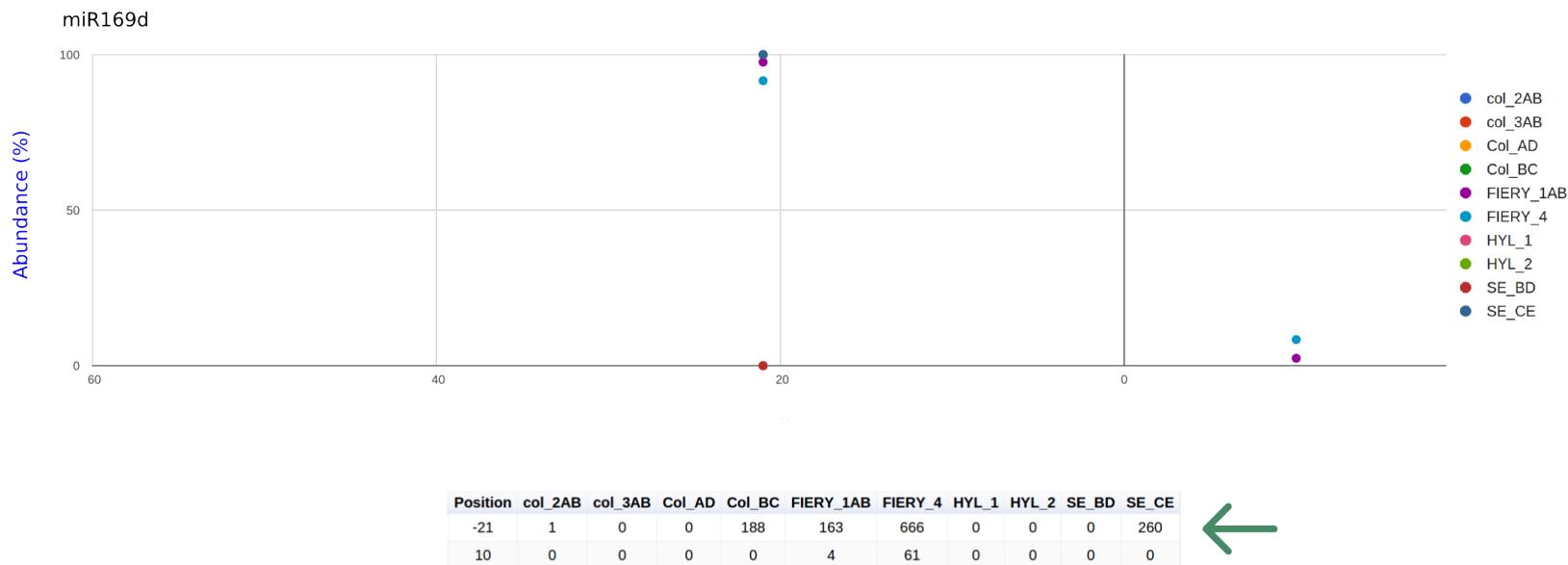


Figura 5.11 Capturntalla de la herramienta de SPARE para el miR169d. Porcentaje de fragmentos del precursor (abundancia relativa de los fragmentos eosición dividido la abundancia total de los fragmentos en el precursor). La posición final del duplex miARN/miARN* fue considerada como la posición 0. La tabla muestra los valores absolutos de cada fragmento. Las flechas verdes marcan las posiciones del precursor con los fragmentos de mayor abundancia.

Para el caso de los precursores cortos de loop a base, los fragmentos detectados pueden ser más de uno según muestra la técnica de SPARE (Figura 5.8). Vemos el precursor del miR156a, donde los cortes más abundantes corresponden a las posiciones 22 y 0 (Figura 5.12). Estos fragmentos corresponden a los cortes del precursor por DCL1, donde el primer corte se da en la parte distal del dúplex y el segundo en la parte proximal del mismo. Además se pueden observar otros fragmentos con abundancias relativas, cercanas a estas posiciones. En algunas bibliotecas se puede observar que acumulan más el primer corte (por ejemplo la biblioteca de Col_AD) y en otras el segundo es el más abundante (biblioteca de Fiery_1AB) (Figura 5.8). Esto podría ser interesante para estudiar en un futuro. Además se puede observar que en *Hyl* y *Serrate* hay otros fragmentos con abundancia relativa en otras posiciones que las esperadas (posición 3 y posición 14 respectivamente), sugeriendo que tal vez que los cortes podrían estar afectados en estas mutantes de procesamiento (Figura 5.8).

Por último en el caso del miR159, y al igual que todos los precursores que son procesados secuenciales de loop a base los cuatro fragmentos más abundantes corresponden a los cuatro cortes realizados por DCL1 que son requeridos para procesar este precursor (Figura 5.13). El primer corte se da en la posición 71, que de los cuatro cortes es el que tiene fragmentos menos abundantes, y luego DCL1 corta a 21 nucleótidos liberando otros ARN pequeños. Luego se pueden observar los fragmentos correspondientes al tercer (posición 21) y cuarto corte (posición 0), que son los necesarios para liberar el miARN maduro (Figura 5.13). En este caso se pueden ver otros fragmentos que son productos de la degradación del precursor, aunque nunca son tan abundantes como los fragmentos de los cortes por DCL1. Por cuestiones de simplicidad, en la tabla solo se muestra los fragmentos con mayor abundancia en promedio de todas las bibliotecas.



Figura 5.12 Captura de pantalla de la herramienta de SPARE para el miR156a. Herramienta para analizar los datos de la técnica SPARE. Porcentaje de fragmentos del precursor (abundancia relativa de los fragmentos en esa posición dividido la abundancia total de los fragmentos en el precursor). La posición final del duplex miARN/miARN* fue considerada como la posición 0. La tabla muestra los valores absolutos de cada fragmento. Las flechas verdes marcan las posiciones del precursor con los fragmentos de mayor abundancia.

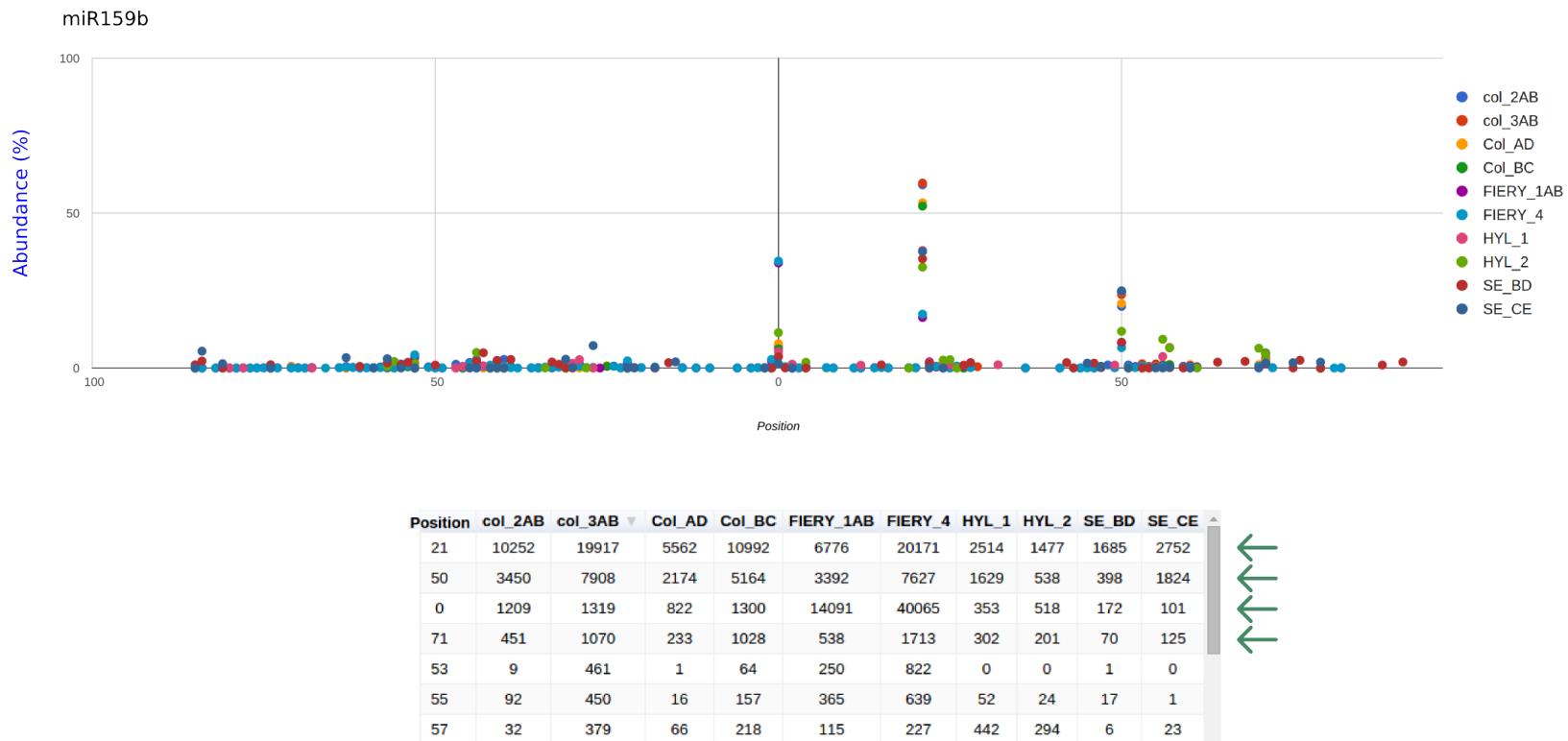


Figura 5.13 Captura de pantalla de la herramienta de SPARE para el miR159b. Porcentaje de fragmentos del precursor (abundancia relativa de los fragmentos en esa posición dividido la abundancia total de los fragmentos en el precursor). La posición final del duplex miARN/miARN* fue considerada como la posición 0. La tabla muestra los valores absolutos de cada fragmento. Por cuestiones de simplicidad, solo se muestran en la tabla los fragmentos con mayor abundancia en promedio de todas las bibliotecas. Las flechas verdes marcan las posiciones del precursor con los fragmentos de mayor abundancia.

Capítulo 3

Enfoque bioinformático para el estudio de la evolución y biogénesis de miARNs en plantas

6.1 Introducción

En el capítulo anterior vimos que los precursores de plantas pueden ser procesados por cuatro mecanismos con distintas características (Figura 5.7). Esto contribuye a la complejidad del sistema de procesamiento, respecto al de animales.

Los estudios que dan origen a estos modelos de procesamiento se basan principalmente en datos de secuenciación masiva de ARN en los que se detectan los intermediarios de procesamiento de los distintos precursores. Como se mostró anteriormente, a partir de los intermediarios de procesamiento es posible definir cuál es el mecanismo al que es sometido cada precursor.

En estudios previos que consideran a los precursores de miARNs en plantas como un solo grupo de moléculas, se han encontrado pocos elementos comunes en sus estructuras [90]. Aprovechando los datos genómicos generados por nuestro laboratorio a partir del estudio del procesamiento de miARNs en plantas, agrupamos los precursores por mecanismo de procesamiento y estudiamos en detalle a cada uno por separado. Para esto diseñamos un enfoque bioinformático para realizar un detallado análisis de la conservación de la secuencia de precursores en distintas especies de plantas.

6.2 Resultados y Discusión

Que los precursores de miARNs en plantas sean procesados de diferentes maneras [21], nos llevó a especular que su patrón de evolución también puede ser diferente y podrían estar vinculados a su mecanismo de procesamiento. Es por esto que elaboramos una estrategia bioinformática para caracterizar la relación entre la evolución de los precursores de miARNs en plantas y los mecanismos de procesamiento determinados previamente.

6.2.1 Estrategia

La metodología utilizada se detalla en la sección 3.10 y primero la utilizamos para estudiar precursores de miARNs conservados en dicotiledóneas. La misma consta de los siguientes pasos:

- Búsqueda de ortólogos para cada miembro de cada familia de *Arabidopsis* de precursores de miARNs conservados en plantas
- Extensión de las secuencias
- Alineamiento de secuencia primaria
- Alineamiento de secuencia teniendo en cuenta estructura secundaria
- Búsqueda de motivos conservados
- Representación gráfica de los precursores

6.2.2 Búsqueda de ortólogos de precursores de miARNs

Las secuencias de los precursores de miARNs en miRBASE rara vez son validadas experimentalmente, de hecho muchas de ellas son el resultado de predicciones computacionales de estructuras en forma de tallo y burbuja que incluye el miARN maduro, pero que no se extiende muchas bases más allá de el mismo [70]. Esto lleva a anotaciones erróneas de los precursores de miARNs. Además para cada miembro de cada familia de *A. thaliana* no es trivial asignarle un ortólogo en otra especie teniendo en cuenta la anotación de miRBASE.

Por esto, primero realizamos una búsqueda de ortólogos para cada miembro de cada familia de *A. thaliana* y empezamos nuestro análisis con una definición arbitraria de los precursores de plantas incluyendo 150 nt fuera del par miARN/miARN*. Para la búsqueda de ortólogos, extensión de la secuencia y el posterior análisis de los precursores, utilizamos datos genómicos de plantas extraídos de Phytozome.

Para la mayoría de los miembros de cada familia de precursores pudimos detectar ortólogos en otras especies. Teniendo en cuenta las 30 dicotiledóneas utilizadas en este estudio, en promedio, se encontraron ortólogos para cerca de 20 especies para cada precursor de miARN de *A. thaliana* (Figura 6.1).

Algunas familias de miARNs se han diversificado durante la evolución, donde en algunas especies un miARN puede encontrarse como un único gen y en otras especie el mismo miARN puede encontrarse en múltiples miembros. Por ejemplo, el miR396 en *Arabidopsis* está constituido por dos miembros, miR396a y miR396b mientras que en el caso de

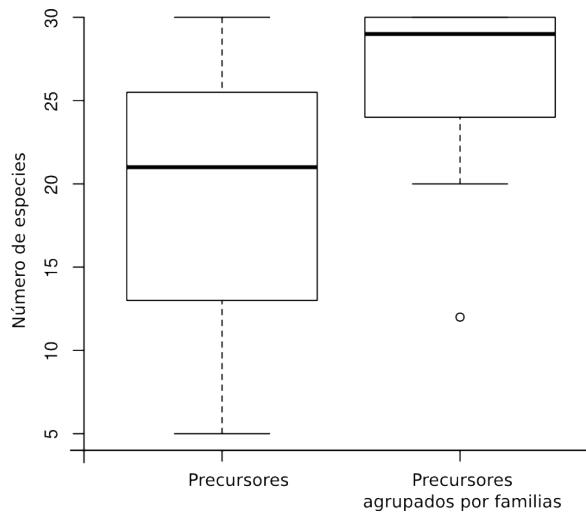


Figura 6.1 Cantidad de especies detectadas. Esta figura muestra la distribución de la cantidad de especies donde se encuentran ortólogos de precursores de miARNs. A la izquierda se muestra la distribución de especies por miembro de familias de precursores de miARNs donde se detecta el ortólogo de *A. thaliana*. A la derecha se agruparon todos los miembros de una misma familia y se muestra la distribución de la cantidad de especies donde se detecta un ortólogo de *A. thaliana*

arroz existen 8 miembros, nombrados del miR396a al miR396h. Por esto, agrupamos a los miembros por familias y vimos la distribución de la cantidad de especies donde se detecta un ortólogo de los precursores de miARNs pertenecientes a *A. thaliana* y en este caso vimos un alto número de especies detectadas para la mayoría de las familias de miARNs (Figura 6.1).

6.2.3 Alineamiento de los precursores en base a su secuencia primaria

Comenzamos con los alineamientos múltiple de cada miembro de las familias de precursores de miARNs conservadas en dicotiledóneas. Primero, realizamos el alineamiento en base a su secuencia primaria utilizando la herramienta T-Coffee [99]. En la Figura 6.2 se muestra el alineamiento del precursor del miR172a en distintas especies coloreado en base a la conservación evolutiva de la secuencia primaria. Se puede observar que el miR172a maduro y el miR172a* están muy conservados en las distintas especies, pero además se puede ver una cola de conservación que podría corresponder al tallo inferior del precursor (Figura 6.2).

Para poder corroborar esto y obtener información adicional sobre la conservación de los precursores, realizamos la predicción de estructura secundaria utilizando la herramienta RNAfold [83] (Figura 6.3). Además realizamos nuevamente los alineamientos pero considerando la estructura secundaria de los precursores, utilizando la herramienta R-Coffee [141].

En la Figura 6.3 se puede observar que existe un patrón que comparten los precursores en la región debajo del miARN/miARN*. No es simple reconocer patrones similares mirando los precursores de esta manera. Además la cantidad de precursores estudiados en esta parte del trabajo de Tesis dificulta aún más este tipo de análisis.

6.2 Resultados y Discusión

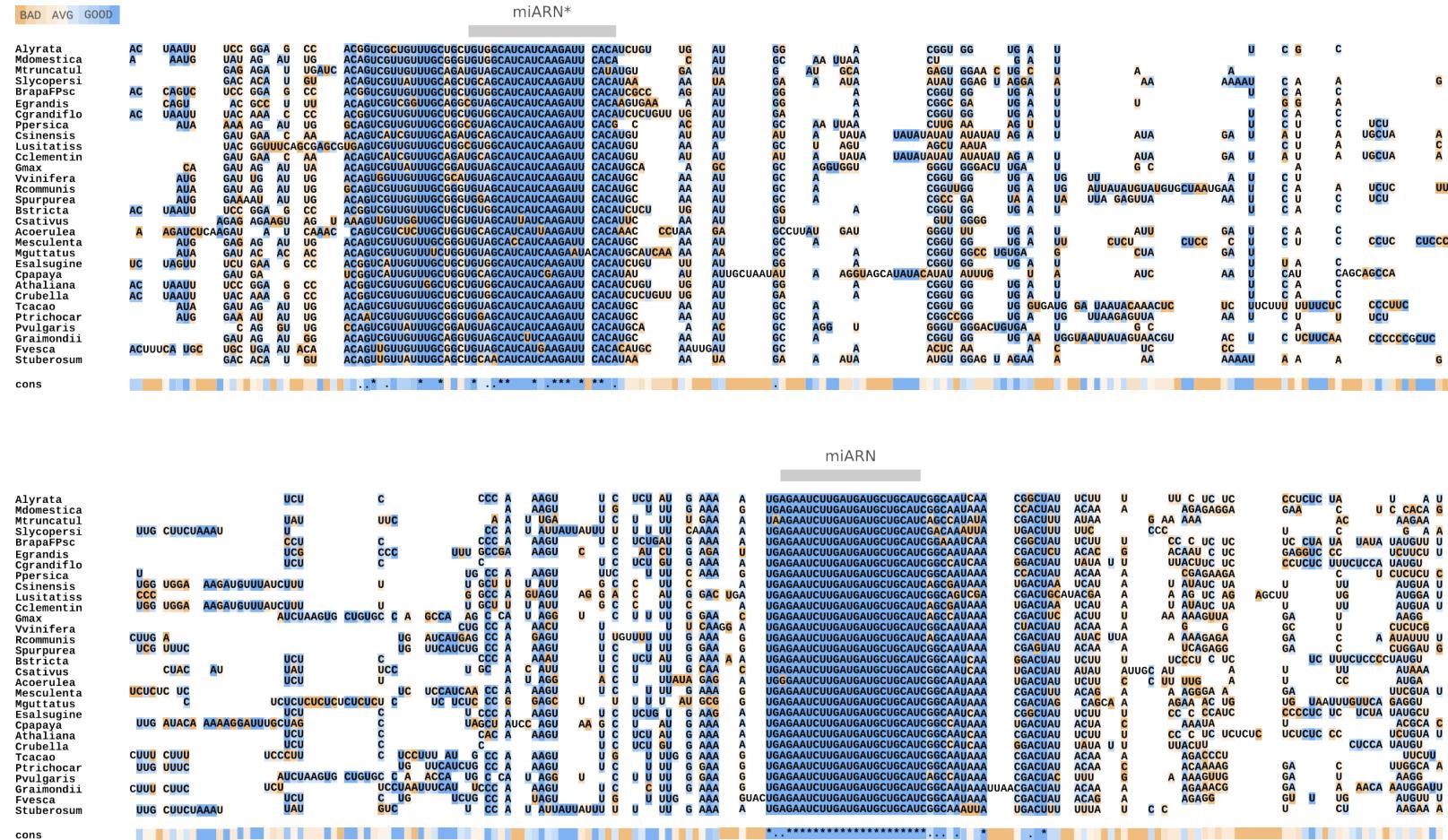


Figura 6.2 Alineamiento del precursor del miR172a. Se muestra el alineamiento del precursor del miR172a en distintas especies. En colores se muestra la conservación de la secuencia primaria, donde azul más oscuro denota mayor conservación y naranja menor conservación. Se muestran ~60 bases por debajo del dúplex miARN/miARN* y los alineamientos están separados en dos líneas para una mejor representación.

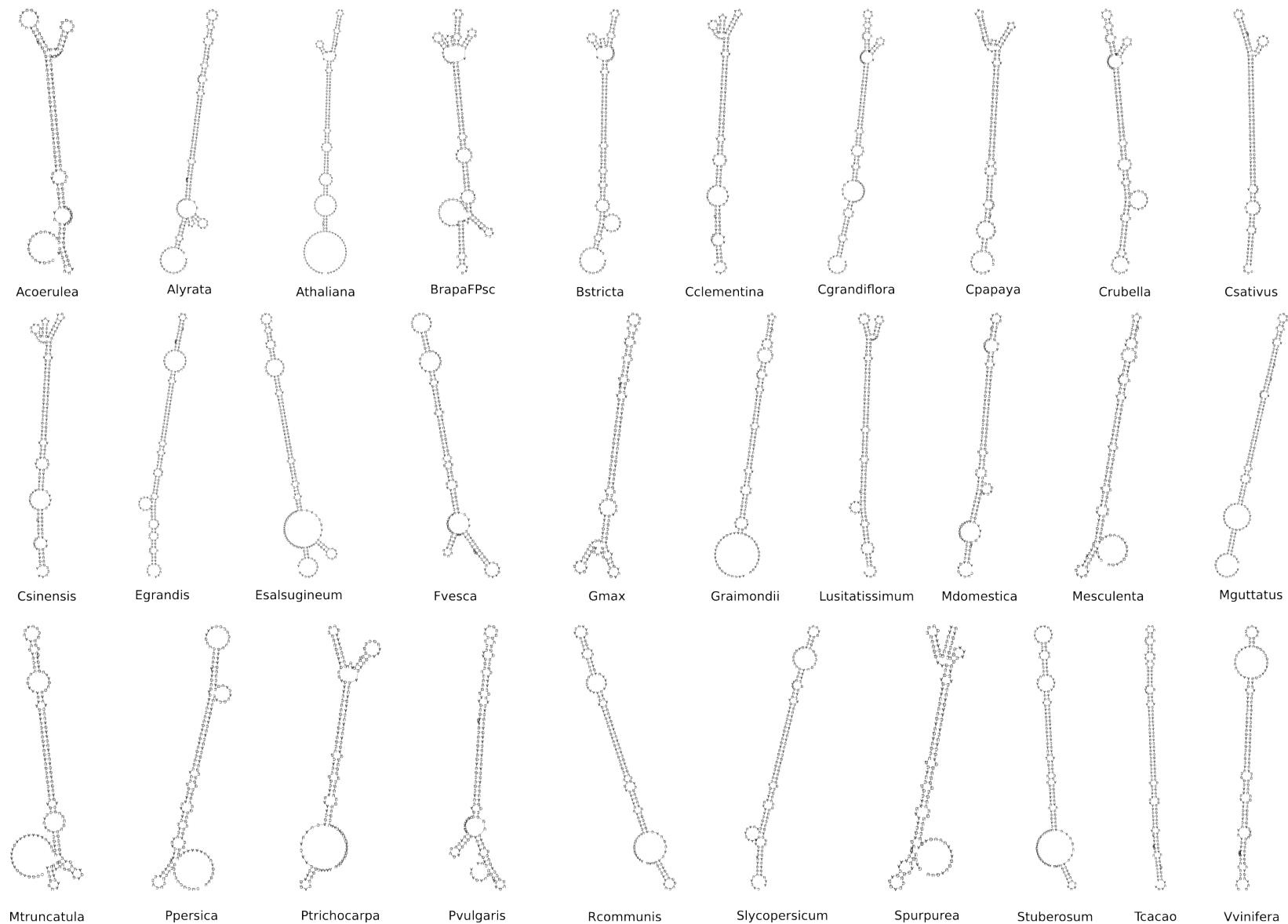


Figura 6.3 Estructura secundaria del miR172a en distintas especies. Se muestra la estructura secundaria de los precursores del miR172a en distintas especies, calculados con la herramienta RNAfold.

Para poder reconocer fácilmente la región conservada dentro del precursor y poder analizar más en detalle estos patrones de conservación, decidimos agrupar toda esta información generada a partir de los alineamientos de secuencia primaria y estructura secundaria en un gráfico circular como se muestra en la figura 6.4. Estos gráficos se hicieron utilizando el paquete Circos [72].

En dicha figura se muestra el Circos del miR172a a modo de ejemplo. Para representar el precursor, se tomaron 60 nt por debajo del dúplex miARN/miARN*. En el anillo exterior se representa en colores el grado de conservación de la secuencia consenso a partir de los alineamientos en base a su secuencia primaria. Además se muestran las bases del precursor de *A. thaliana*. Luego en el anillo interior se representa, mediante un histograma, la frecuencia de bases apareadas y desapareadas para cada posición del precursor en las distintas especies. Con líneas de distinto espesor se muestra la interacción de las bases del precursor considerando la estructura secundaria. Fuera del anillo exterior se marca la secuencia del miARN y miARN*.

Realizamos los Circos para visualizar de manera simple los precursores de miARNs en distintas especies de plantas. Y los utilizamos para caracterizar la relación entre los patrones de conservación y los mecanismos de procesamiento determinados previamente.

Por último, para cada precursor en distintas especies, realizamos una búsqueda de motivos conservados con la herramienta MEME [83] (Figura 6.5). Hicimos esto para poder ver de manera visual cuáles eran los patrones comunes entre las distintas especies y donde se localizaban dentro del precursor. Esta búsqueda de motivos la hicimos de diferentes maneras, la primera permitiendo encontrar hasta 2 motivos conservados en los precursores de distintas especies.

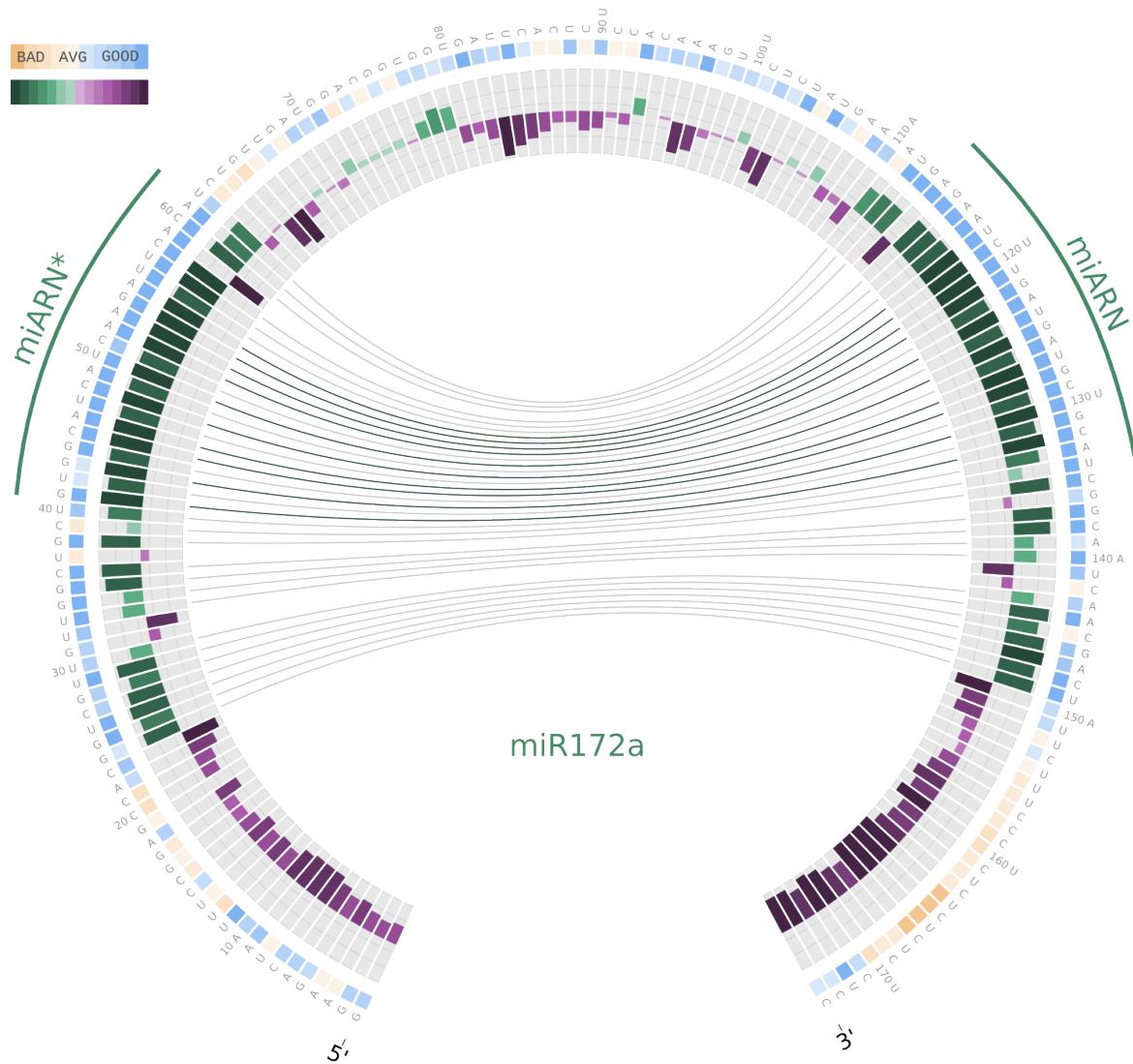


Figura 6.4 Circos del miR172a. En el anillo exterior se representa en colores el grado de conservación de la secuencia consenso a partir de los alineamientos en base a su secuencia primaria. Además se muestran las bases del precursor de *A. thaliana*. Luego en el anillo interior se representa, mediante un histograma, la frecuencia de bases apareadas y despareadas para cada posición del precursor en las distintas especies. Con líneas de distinto espesor se muestra la interacción de las bases del precursor considerando la estructura secundaria. Fuera del anillo exterior se marca la secuencia del miARN y miARN*.



Figura 6.5 **MEME del miR172a**. Se muestra el MEME donde se pueden observar los motivos conservados dentro de los precursores del miR172a en distintas especies. En color rojo el primer motivo conservado comprende parte del miR172 (logo 1). El segundo motivo más conservado comprende parte del miR172* (logo 2).

6.2.4 Mecanismo de base a loop cortos

En los precursores con un mecanismo corto de base a loop, un loop interno seguido por un tallo inferior de ~15nt especifica la posición del primer corte. A pesar de que el tallo puede contener bulges, la transición de un loop interno (simple hebra) al tallo inferior es bastante marcada, y tres pares de bases apareadas generalmente definen el comienzo del tallo inferior del precursor [21, 90]. El segundo corte procede a una distancia fija de ~21 nt desde la posición del primer corte.

Para la mayoría de los precursores procesados con este mecanismo pudimos observar el mismo patrón de conservación de estructura primaria y similares patrones de estructura secundaria. Para los precursores del miR172a (Figura 6.4 y el miR390a 6.6), mediante la estrategia pudimos encontrar precursores en las 30 especies utilizadas. En ambos casos, se pudo observar que la región con mayor conservación concuerda con el duplex miARN/miARN* pero además se puede ver una región conservada por debajo del dúplex, que coincide con el tallo inferior.

Otra ventaja de esta forma de representar a los precursores con la conservación entre especies, es que visualmente se puede reconocer fácilmente las posiciones donde existen "mismatches" conservados. Por ejemplo en el Circos del miR172, se puede observar un mismatch en la primera posición del miARN con la posición 19 del miARN*. Esta interacción es A-A y se mantiene sin variaciones en todas las especies (Figura 6.4). Esto podría ser interesante estudiarlo en detalle donde se podría generar un precursor mutante donde esas bases en particular estén apareadas y luego ver si esto afectó o no el procesamiento del precursor. Además notamos que este patrón de conservación se puede observar independientemente si el miARN maduro está en la hebra 5' (Figura 6.4) o si está en la hebra 3' (Figura 6.6).

A diferencia de los precursores que tienen un mecanismo de loop a base, donde el tallo superior es fundamental para su procesamiento, los precursores que se procesan cortos de base a loop no tienen el tallo superior conservado. Y mediante la búsqueda de motivos conservados, pudimos observar que el tamaño del tallo superior y del loop es muy variado en un mismo precursor en distintas especies, donde los mismos puede ir desde los 40nt hasta 115nt (Figura 6.5).

6.2.5 Mecanismo de base a loop secuenciales

En estos precursores, con un mecanismo secuencial de base a loop, el primer corte procede como en los cortos de base a loop, pero luego son necesarios dos cortes más para liberar el miARN, generando en el proceso niveles bajos de RNA pequeños adicionales [21]. Este es el caso de algunos miembros de la familia del miR169, que tiene en total 14 miembros

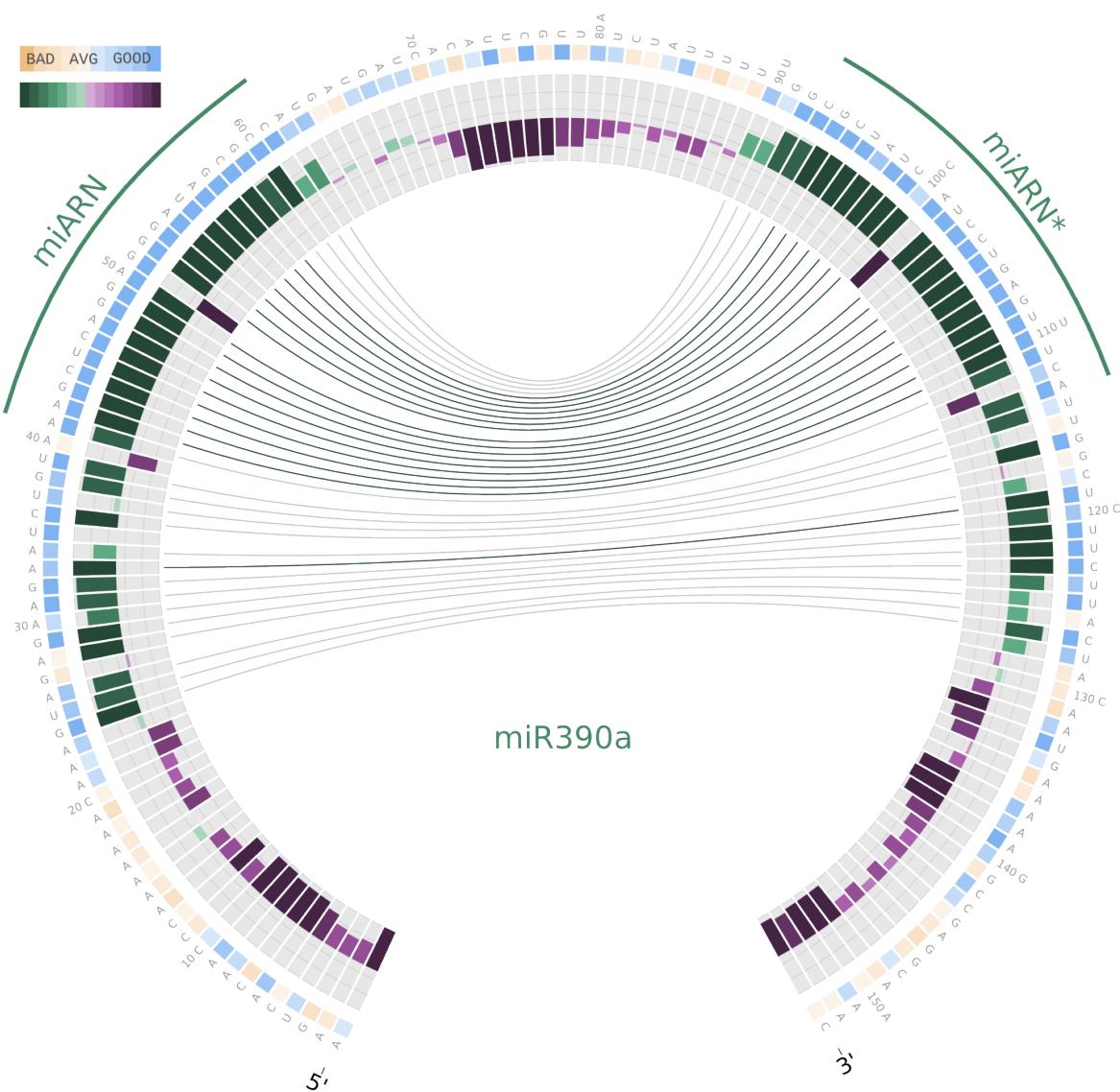


Figura 6.6 Circos del miR390a.

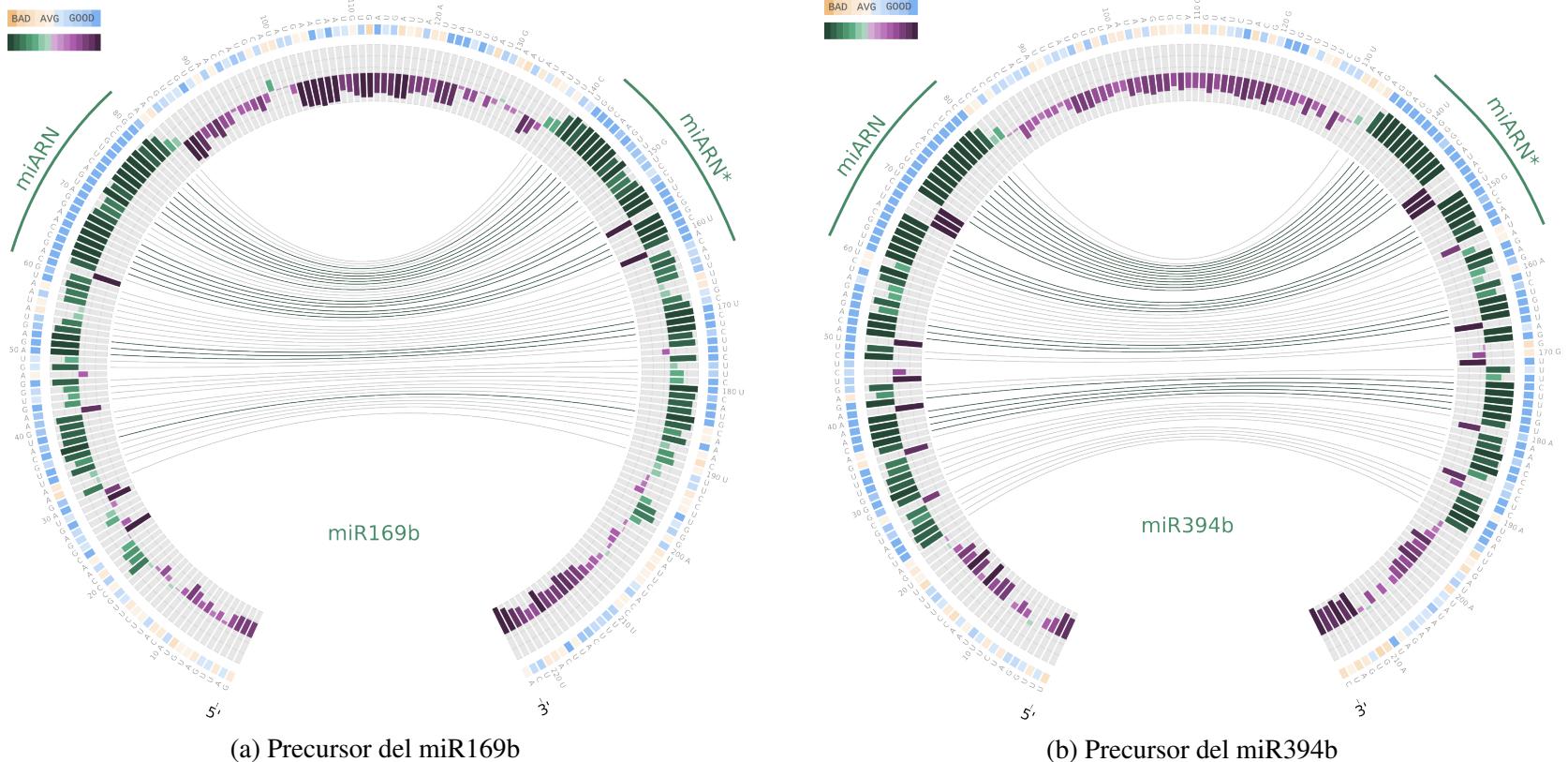
siendo la familia más grande de *A. thaliana*. Un estudio en detalle de los precursores del miR169b/d/e/f/g mostró que los sitios de cortes del precursor estaban localizados 21 nt por debajo del lado proximal del dúplex miARN/miARN*. Esta es la distancia esperada entre dos cortes de DCL1, sugiriendo que estos precursores eran procesados por más de dos cortes de la encima [21].

La familia del miR394 también es procesada de la misma manera y comparten similares patrones de estructura secundaria, donde se ve el tallo inferior largo bien estructurado (Figura 6.7). Si observamos el patrón de conservación de secuencia primaria podemos ver que la región del dúplex miARN/miARN está muy conservada, pero además podemos observar que debajo del dúplex existe otra región conservada que coincide con el tallo inferior largo que presentan estos precursores (Figura 6.7).

6.2.6 Mecanismo de loop a base cortos

En los precursores con un mecanismo **cortos de loop a base**, el procesamiento es guiado por un tallo superior, y son necesarios dos cortes para liberar el miARN maduro. La región terminal de estos precursores tienen una largo conservado de ~42, donde incluye un loop pequeño [21] a diferencia de los precursores procesados dee base a loop, donde esta región es variable en tamaño.

En este caso mostramos los Circos para el precursor del miR157a, que fue encontrado en 30 especies, y para el precursor del miR160a también encontrado en 30 especies. En los precursores que se procesan cortos de base a loop observamos que presentan una región conservada que corresponde al tallo superior además de la región que comprende al dúplex miARN/miARN* (Figura 6.7). A diferencia de los precursores que se procesan de la base al loop, estos precursores no presentan el tallo inferior ni estructurado ni conservado y esto se puede observar en ambos casos; en el miR157a (Figura 6.8a) y del miR160a (Figura 6.8b).



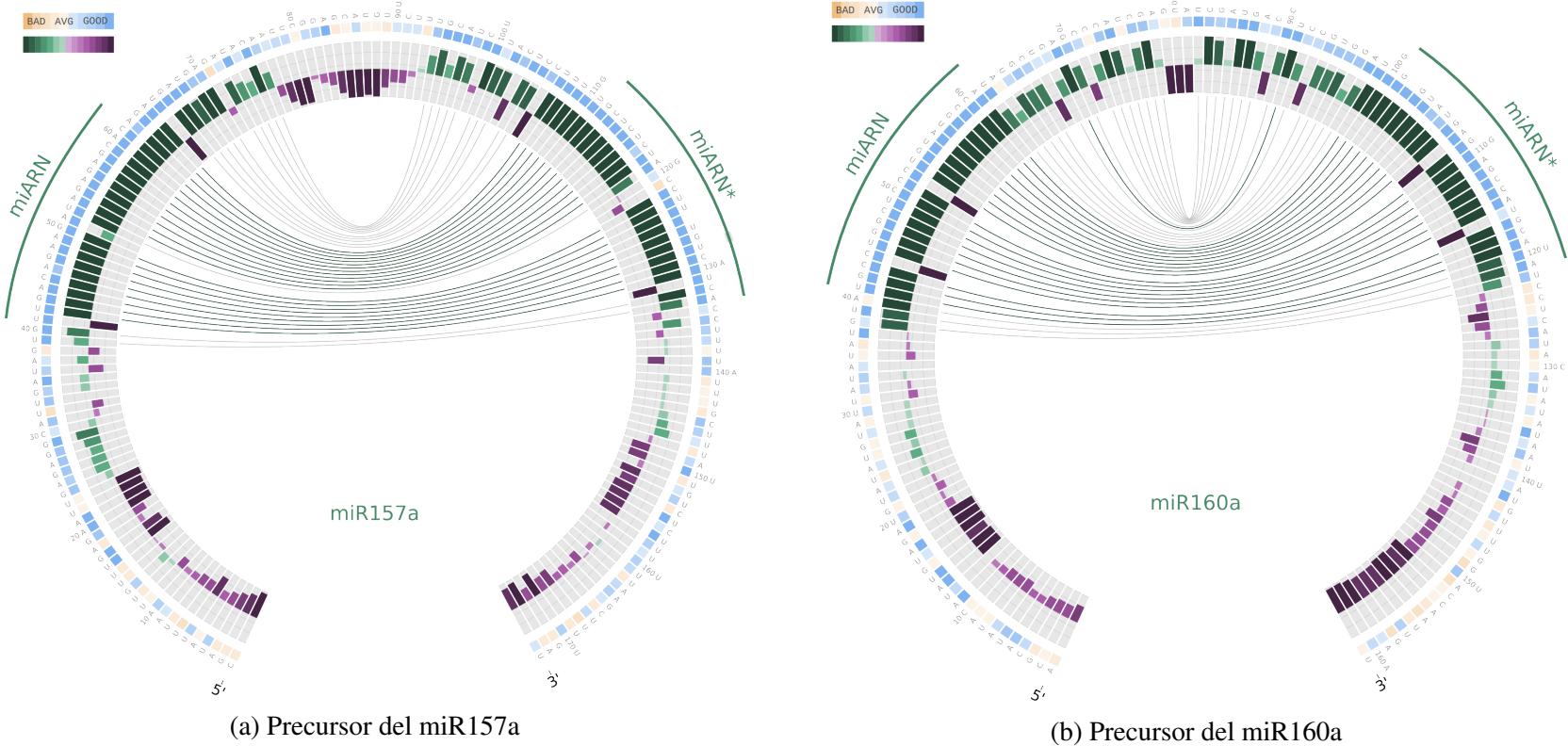


Figura 6.8 Circos de precursores con mecanismos de procesamientos cortos de loop a base

6.2.7 Mecanismo de loop a base secuenciales

En los precursores con un mecanismo **secuencial de loop a base**, cuatro cortes secuenciales por DCL1 son los encargados de procesar los precursores de miARNs. En general, estos precursores muestran un tallo largo superior, del cual otros ARNs pequeños son generados [4, 19, 21].

En este caso estudiamos la familia del miR319 y del miR159 y en particular mostramos al miR319a (Figura 6.9a) y al miR159b (Figura 6.9b). Pudimos reconocer el tallo superior conservado y estructurado. Además se pudo observar que los ARNs pequeños, que se originan del procesamiento de estos precursores, están conservados de la misma manera que el miARN y el miARN* (Figura 6.9).

6.2.8 Procesamiento mixto de miembros de la familia del miR170/miR171

En general, se sabe que diferentes miembros de una misma familia de miARNs comparten la misma vía de biogénesis. Esta observación no sorprende ya que se cree que las familias de miARNs se expanden por eventos de duplicación de un gen ancestral [6]. Sin embargo, ciertos el procesamiento de miembros de ciertas familias pueden variar de uno a otro [21]. Esto sucede, por ejemplo, en la familia del miR171 donde en *A. thaliana* existen 4 miembros. El precursor del miR171a es procesado de base a loop, mientras que los precursores del miR171b y miR171c son procesados de loop a base.

Nos pusimos a analizar esta familia en detalle para ver si los patrones de conservación y estructura secundaria eran diferentes o no. Observamos que el miR171a, que es procesado corto de base a loop, tiene un patrón de conservación similar a los precursores procesados de base a loop, donde el tallo inferior es estructurado y conservado (Figura 6.10b). Por el contrario el miR171c, que es procesado corto de loop a base, muestra conservación en el tallo superior, que además es estructurado, y no así el tallo inferior (Figura 6.10a).

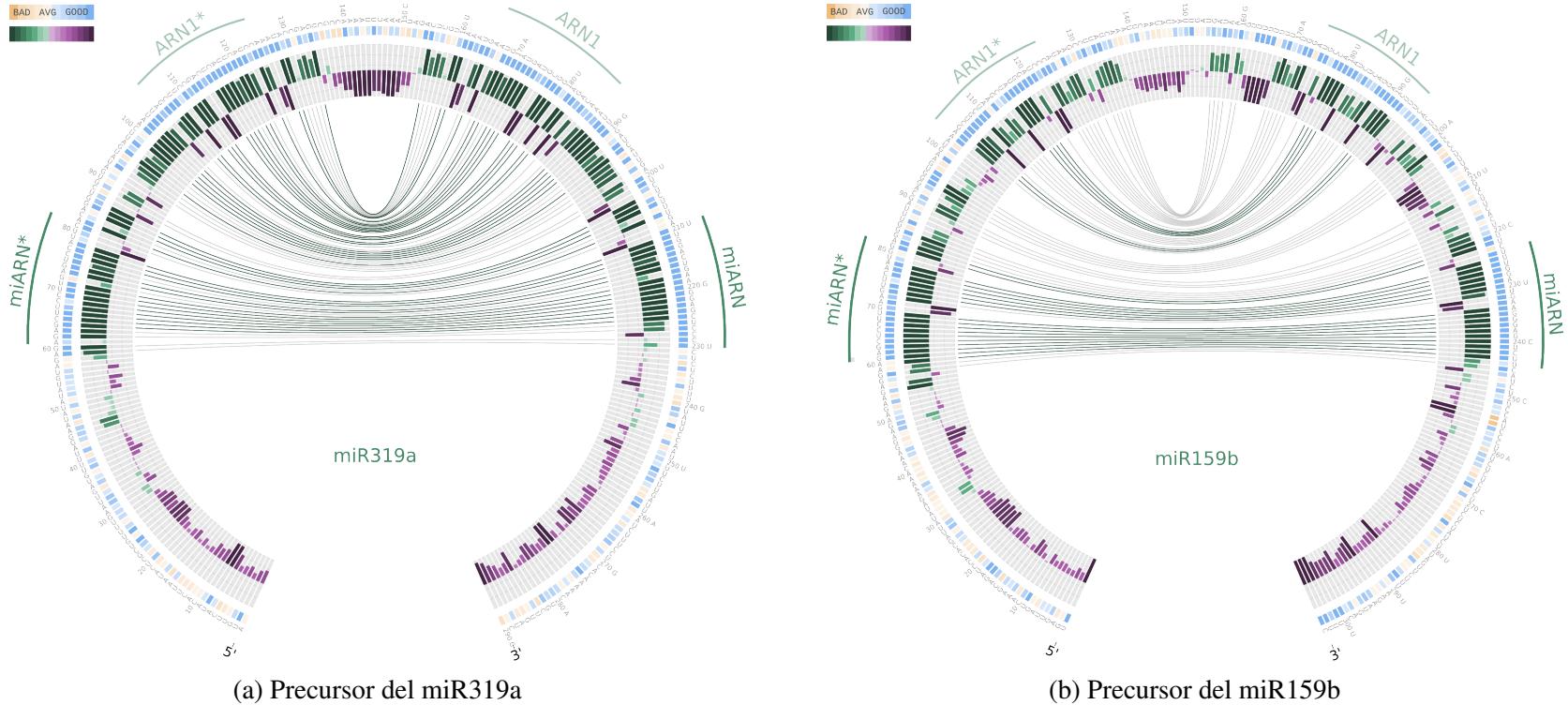


Figura 6.9 Circos de precursores con mecanismos de procesamientos secuenciales de loop a base

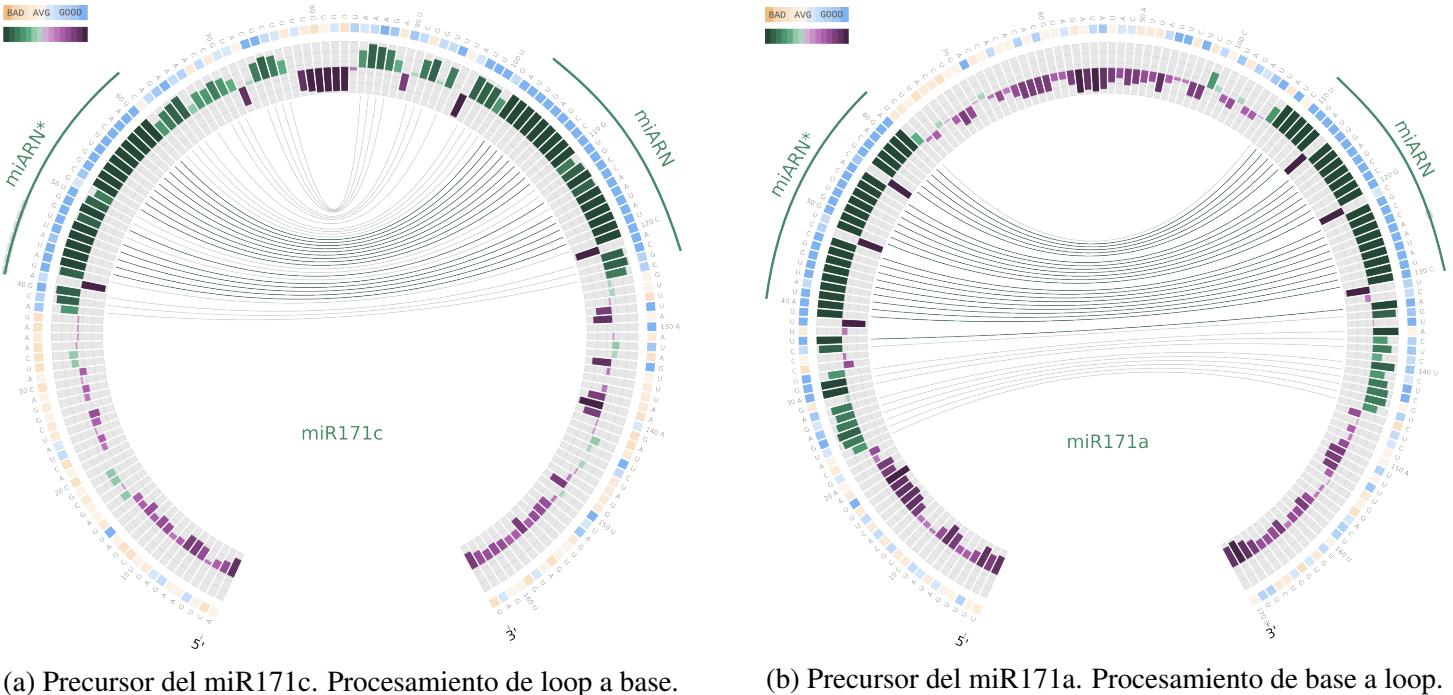


Figura 6.10 Procesamiento mixto de miembros de la familia del miR170/miR171

6.2.9 Mutaciones puntuales afectan el procesamiento de miARNs en plantas

Se ha demostrado recientemente que un polimorfismo de origen natural que ocurre en el gen del miR164a afecta a la forma de la hoja y la arquitectura del vástagos en *A. thaliana*, con los efectos de ser modificados por loci adicionales en el genoma [135]. Una sustitución única en un par de bases en la secuencia complementaria del miARN altera la estabilidad predicha del dúplex miARN/miARN*. Se reduce con ello, en gran medida, la acumulación del miARN maduro, probablemente porque interfiere en el procesamiento del precursor. Además se demostró en ese mismo artículo que no es una rara excepción y que las cepas naturales de *A. thaliana* albergan decenas de polimorfismos similares que afectan al procesamiento de una amplia gama de precursores de miARNs.

Es por esto que nos pusimos a estudiar en detalles al precursor del miR164a. A nivel de secuencia los alelos de Col-0 y C24 son diferentes sólo por un par de polimorfismos simples. Uno de ellos afecta a una C en el miARN* que aparea con una G en la posición dos del miARN (nos referimos a esta posición como *2) (Figura 6.14).

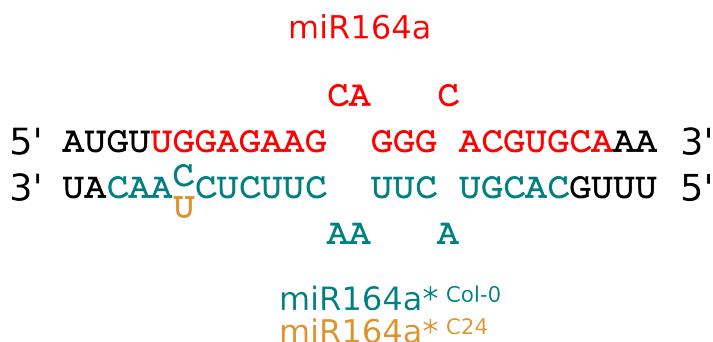


Figura 6.11 Apareamiento de bases en el precursor del miR164a. El miARN se muestra en rojo y el miARN* en azul. En la posición *2, el alelo alternativo C24 se muestra en naranja.

Se puede ver que el remplazo de una G:C en el miR164a^{Col-0} por un par no canónico G:U en el miR164a^{C24}, incrementa ligeramente la energía libre del plegado del precursor que cambia de -41.58 kcal/mol a -39.28 kcal/mol, pero no se modifica la estructura secundaria del miR164a (Figura 6.12).

Realizamos el Circos del miR164a en distintas especies para ver como era la conservación de la base en la posición 2* que hace que el precursor no pueda ser procesado de manera correcta. Lo que pudimos observar es que dicha posición del miR164a* está conservada en dicotiledóneas y se mantiene sin cambios en todas las especies (Figura 6.13). Además esa

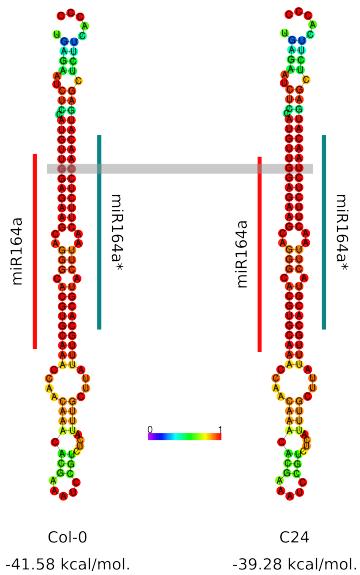


Figura 6.12 Estructura secundaria predicha del precursor del miR164a. La barra gris horizontal indica el polimorfismo, un G:C en Col-0 y un G:U en C24. Los colores indican probabilidad de pares de bases.

base está siempre apareada con la base correspondiente del miARN maduro, formando un G:C (Figura 6.13 a la derecha).

También se puede observar que en la secuencia del miR164a* existen mutaciones en distintas bases en distintas especies, pero no en la base que estamos estudiando. Esto sugiere que esa base en particular es importante para la estabilidad del precursor y su buen procesamiento.

En otro trabajo, se estudió un alelo mutante llamado miR394b-1, que tiene un mismatch en el stem inferior del precursor del miR394b [69]. La mutante incrementa dramáticamente la terminación del meristema apical. En la figura 6.14 se muestra el apareamiento de bases en el precursor del miR394b y miR394b-1, donde el miARN se muestra en rojo y el miARN* en azul. Y en naranja se muestra la posición de la mutación del precursor del miR394b-1 donde la G es reemplazada por una A con respecto al miR394b. En este caso se mostró que los niveles de maduro en el precursor del miR394b-1 se ven fuertemente disminuidos aunque no completamente ausentes [69].

Para varios precursores de plantas hemos visto anteriormente que esa región que corresponde al stem inferior es crucial para su procesamiento. En este caso mostramos el Circos correspondiente al miR394b, para poder visualizar que es lo que sucede con esta mutación puntual fuera del dúplex miARN/miARN* (Figura 6.15). Vemos, como en el caso del miR164a, esta mutación que afecta al procesamiento del precursor y a la acumulación del miARN maduro, está conservada en todas las especies estudiadas. Esto sugiere que esta

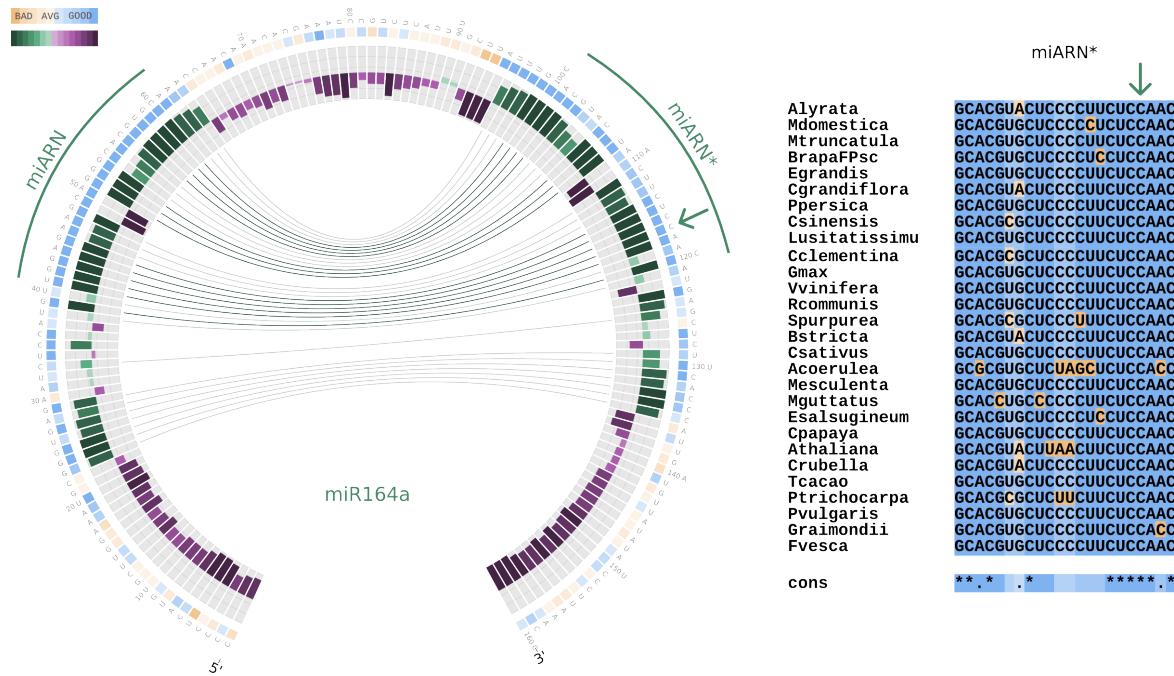


Figura 6.13 **Circos del miR164a.** A la izquierda se representa el Circos del miR164a. A la derecha el alineamiento del miR164a en distintas especies. Con una flecha verde se destaca la posición *2 a estudiar, en el Circos y en el alineamiento.



Figura 6.14 **Apareamiento de bases en el precursor del miR394b.** El miARN se muestra en rojo y el miARN* en azul. En naranja se muestra la posición de la mutación G por A en el stem inferior del precursor del miR394b-1.

base es importante para el procesamiento y que no sólo mutaciones de una base dentro del dúplex miARN/miARN* pueden afectar el procesamiento, sino que mutaciones simples en el precursor puede afectar el reconocimiento de DCL1.

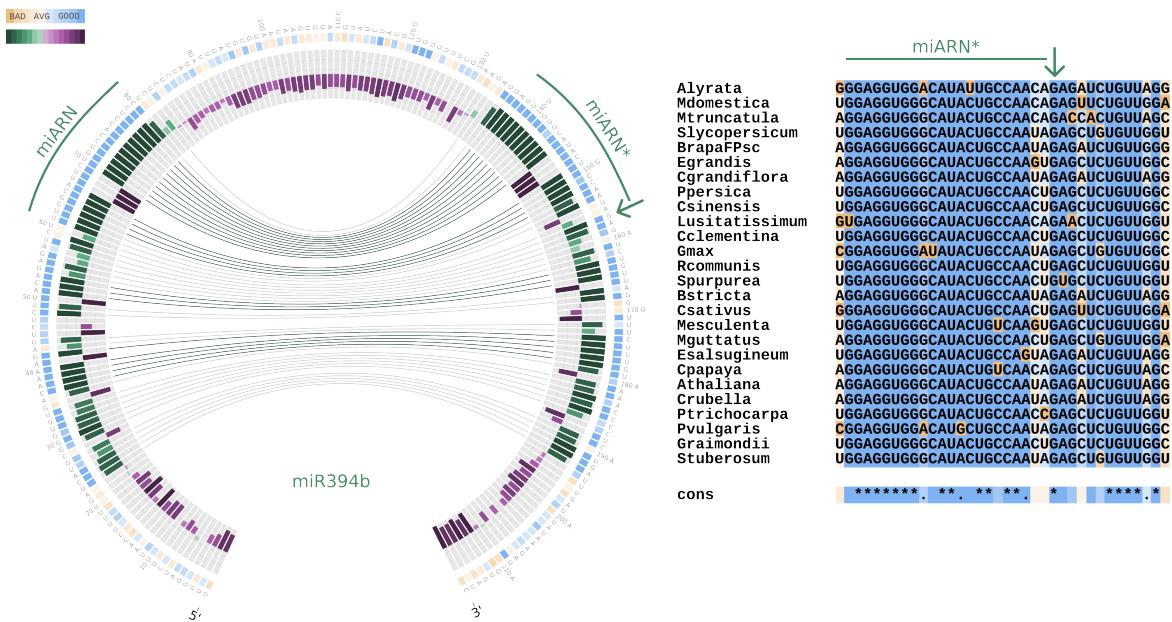


Figura 6.15 Circos del miR394b. A la izquierda se representa el Circos del miR394b. A la derecha el alineamiento del miR394b en distintas especies. Con una flecha verde se destaca la posición a estudiar, en el Circos y en el alineamiento.

La metodología empleada y la visualización por medio de los Circos permite deducir cuales son las bases o apareamientos importantes para el procesamiento del precursor y la liberación del miARN maduro. La misma podría ser utilizada para ayudar en el diseño de miARNs artificiales en distintas especies.

6.2.10 Precursores de plantas conservados en monocotiledóneas

Se conocen muchas familias de precursores conservados en plantas, algunos de ellos conservados desde musgos hasta dicotiledóneas [12, 13, 35]. Si bien la mayoría de los precursores de plantas utilizados hasta esta parte del trabajo están conservados en dicotiledóneas y monocotiledóneas, muchos fueron evolucionando encontrándose nuevos miembros de algunas familias en especies más recientes y perdiéndose otros. Por ejemplo, existe una variante del miR396 que fue detectada por primera vez en arroz [131] y estudios posteriores indicaron que la misma sólo es específica de monocotiledóneas [132, 158]. Es por esto que separamos a los precursores de plantas en dicotiledóneas y monocotiledóneas, y en esta parte del trabajo

nos enfocamos en estas últimas. Las especies utilizadas de monocotiledóneas se pueden ver en la Figura 3.6 de Materiales y Métodos.

Utilizamos para estudiar esto, la del miR396 de monocotiledóneas que es hiperactiva contra los GRFs. Los cambios observados en la secuencia del miR396 en diferentes especies se localizan en la región 5' del mismo. Fue demostrado que los cambios en esta región tienen efectos importantes sobre la actividad de los miARNs.

El sitio complementario al miR396 de los GRFs (genes blanco del miR396), posee un nucleótido, que en el apareamiento entre el miARN y el ARNm genera un bulge. Es muy interesante que la G extra presente en la variante de monocotiledóneas elimina este bulge en el apareamiento con el sitio blanco de los GRFs. Graficamos el Circos del miR396e de Arroz que estaba presente en todas las plantas monocotiledóneas estudiadas en nuestro trabajo (Figura 6.16). Vemos que ese nucleótido extra está conservado en todas las especies.

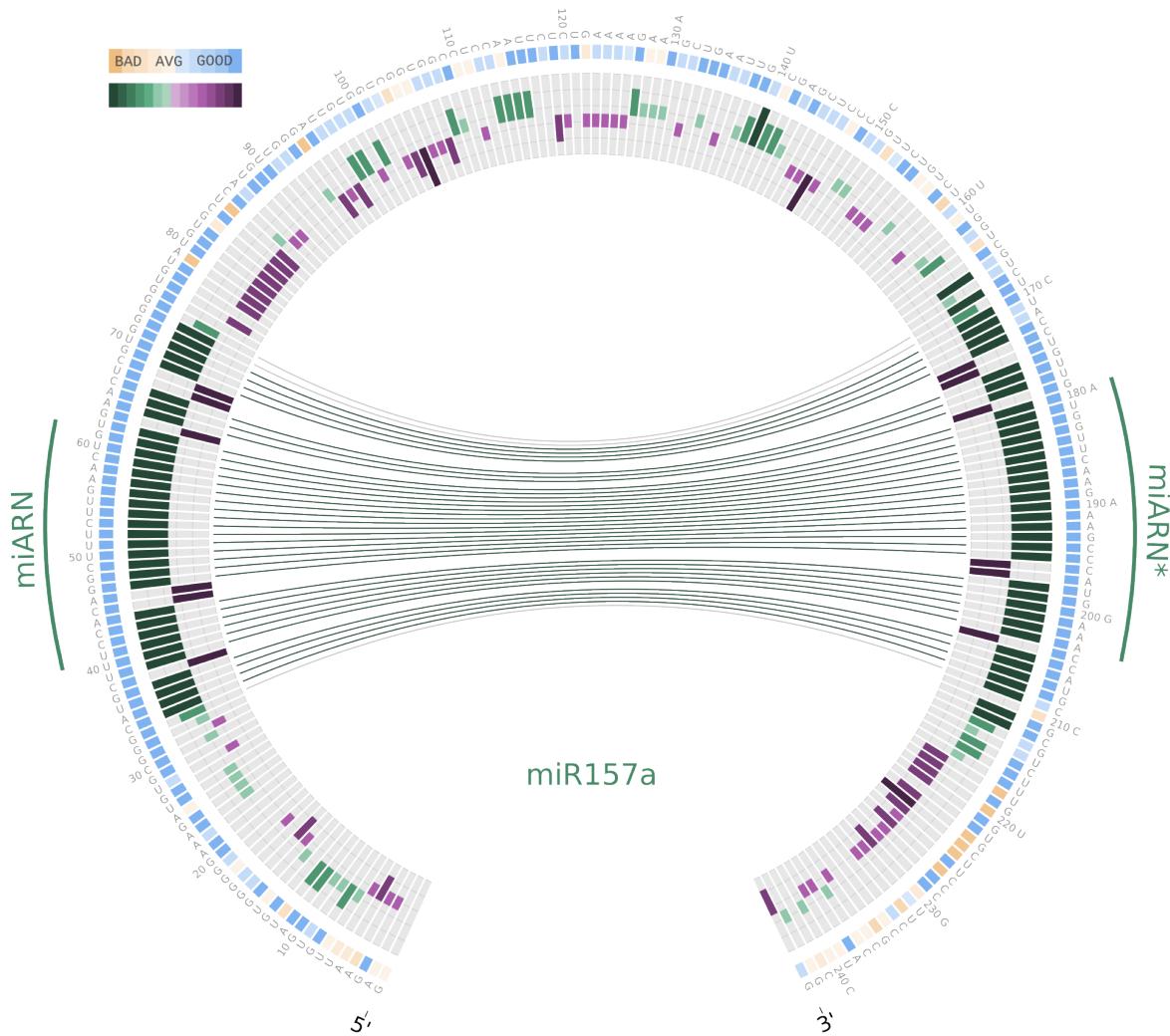


Figura 6.16 Circos del miR396e de monocotiledóneas

6.2.11 Precursores de animales

En animales existe un gran número de miARNs que están conservados en un gran rango de especies, incluso se han encontrado algunos miARNs conservados en humanos hasta gusanos [106]. Si bien los miARNs en plantas y en animales comparten componentes básicos, su biogénesis y su modo de acción es diferente. Una de las grandes diferencias con las plantas radica en que en animales el procesamiento de los precursores ocurre en dos pasos separados en tiempo y espacio (Figura 1.2), mientras que en plantas ocurre íntegramente en el núcleo (Figura 1.3).

Además de las diferencias en la biogénesis, otra de las grandes diferencias entre plantas y animales consiste en las estructuras de los precursores de miARNs. Los precursores de animales se caracterizan por presentar una estructura relativamente homogénea, que consta de un tallo imperfecto de una longitud de ~65 nt, flanqueado por dos regiones despareadas. Estas dos regiones son un bucle terminal en un extremo y ARNsh en el otro. En animales, la especificidad en la selección de la secuencia del miARN dentro del precursor está dada por el primer corte realizado por el complejo microprocesador. Por otro lado, los precursores de miARNs de plantas presentan tamaños que varían desde los 50 a 900 nt, lo que se traduce en un amplio abanico de estructuras diferentes (Figura 6.17).

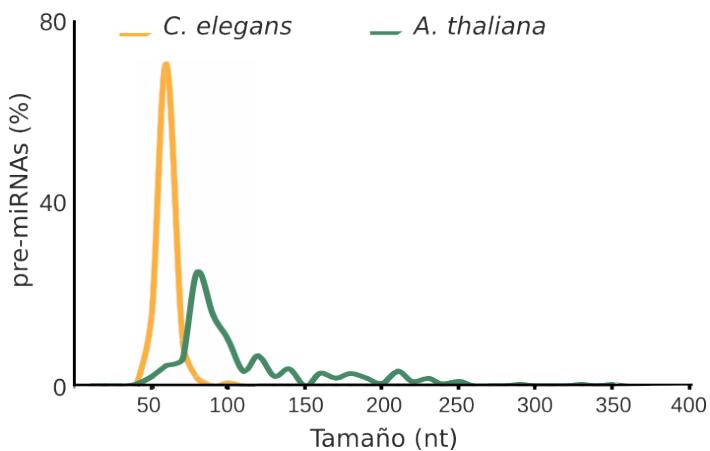


Figura 6.17 **Tamaños de precursores de miARNs.** En línea naranja se muestra la distribución de tamaños de pre-miARNs en *C. elegans* y en verde en *A. thaliana*.

Para esta parte del trabajo de Tesis, utilizamos genomas de distintos especies de *Metazoa* entre ellos humano, mono, rana, vaca y pez (Tabla 6.6). Y estudiamos que sucede con los precursores de miARNs conservados en animales. Para esto partimos de los precursores de humanos definidos en miRBASE y realizamos la búsqueda de ortólogos en otras especies al igual que lo hicimos en plantas.

Table 6.6 Especies de *Metazoa* utilizadas

Animales
<i>Bos taurus</i>
<i>Canis familiaris</i>
<i>Equus caballus</i>
<i>Gallus gallus</i>
<i>Gorilla gorilla</i>
<i>Homo sapiens</i>
<i>Macaca mulatta</i>
<i>Monodelphis domestica</i>
<i>Mus musculus</i>
<i>Ornithorhynchus anatinus</i>
<i>Petromyzon marinus</i>
<i>Sus scrofa</i>
<i>Xenopus tropicalis</i>

En la Figura 6.18 se muestra el Circos del precursor del miARN hsa-let-7a-1, un miembro de la familia let-7. El precursor del miARN let-7 fue identificado en un estudio en *C. elegans* y luego se mostró que formaba parte de una clase más grande de ARNs no codificantes, los miARNs. Actualmente ha sido predicho y validado en un gran número de especies.

En un artículo publicado en Nat. Struct. Mol. Biol., se encontró una ribonucleoproteína (hnRNP A1) que actúa como regulador negativo de let-7-a [93]. HnRNP A1 se une al loop terminal del precursor de let-7-a e inhibe su procesamiento por Drosha. En el Circos se puede observar que este loop terminal está conservado en las especies utilizadas (6.18).

La variación de los precursores de miARNs de plantas se debe en gran parte a la variación en el tamaño del loop. Observamos el tamaño del loop terminal en precursores de plantas procesados cortos de base a loop o de loop a base y lo comparamos con el tamaño del loop terminal en precursores conservados en animales (Figura 6.19). Se puede notar que el tamaño del loop en los precursores de plantas cortos de base a loop son más variables además de más grande. En cambio el tamaño del loop terminal de los precursores de humano son pequeños y similares al tamaño del loop terminal de precursores de plantas loop a base (Figura 6.19).

En animales, la estructura secundaria del loop terminal de precursores de miARNs, es irrelevante durante el procesamiento e inclusive se demostró que la presencia del mismo es innecesaria [59]. Aunque el mecanismo de procesamiento en animales es similar al mecanismo de procesamiento de los precursores de plantas cortos de base a loop, podemos observar que estructuralmente los precursores en animales son más similares a los precursores de plantas cortos de loop a base (Figura 6.20). Se puede observar que el loop terminal es pequeño y bien conservado y el stem inferior no está estructurado (salvo algunas pocas bases en los precursores de animales). Esta metodología de búsqueda de ortólogos, alineamientos

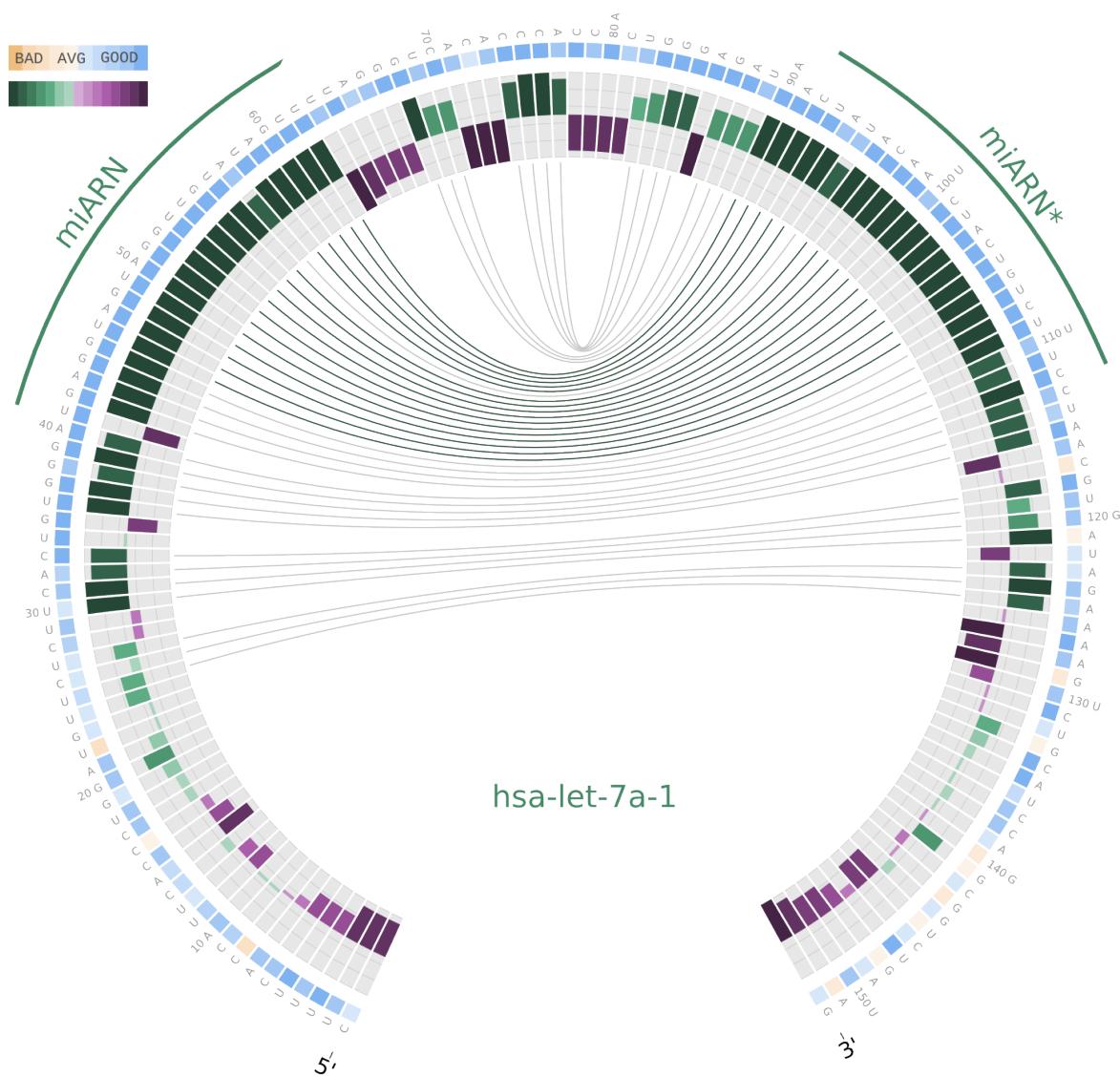


Figura 6.18 Circos del miARN hsa-let-7a-1.

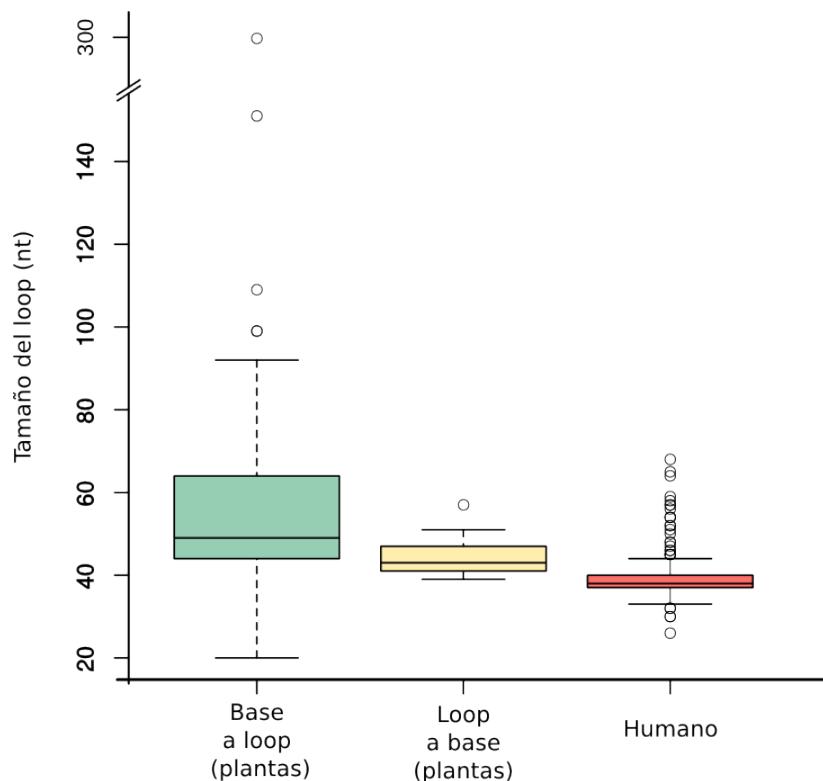


Figura 6.19 **Tamaño del loop terminal en precursores de distintos reinos.** Se muestra la distribución del tamaño del loop terminal (nt) de precursores de plantas cortos de base a loop, en cortos de loop a base y de humanos. Por simplicidad se utilizaron precursores de humanos conservados en 8 especies según nuestro criterio de búsqueda de ortólogos.

y posterior visualización con Circos permite comparar precursores de miARNs no solo de distintas familias, sino que también precursores de miARNs de distintos reinos.

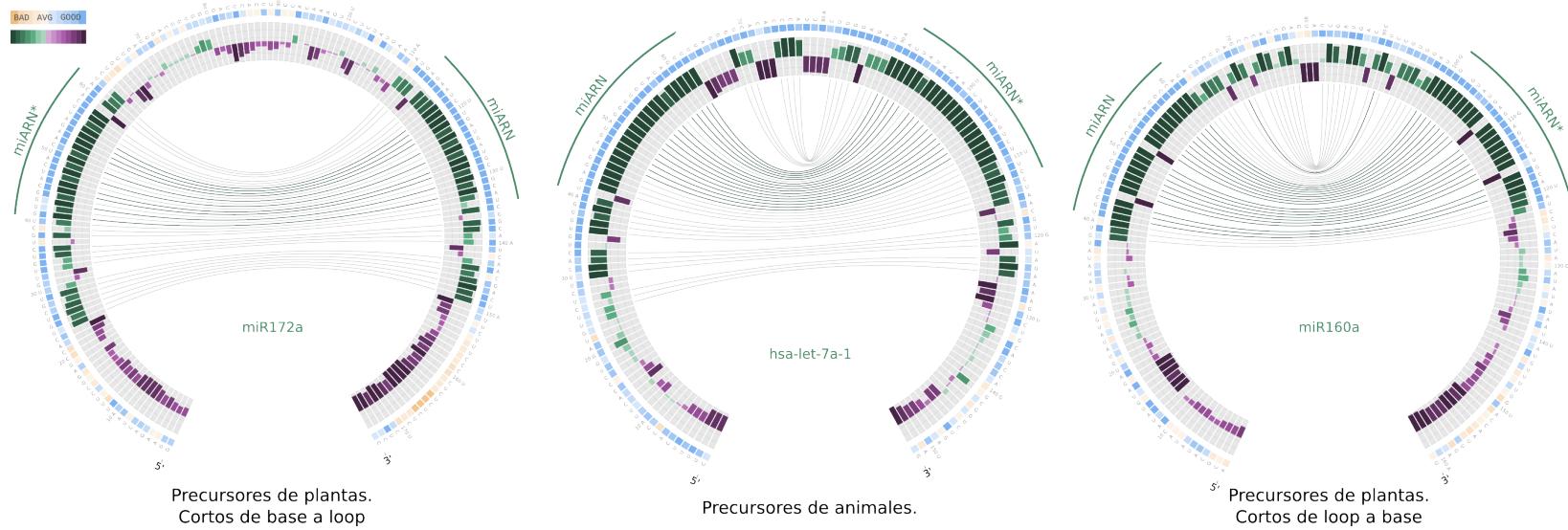


Figura 6.20 **Circos de animales vs plantas.** A la izquierda el Circos del miR172a, un precursor conservado en plantas con mecanismo corto de base a loop. En el medio el Circos del precursor de miARN de humano let-7a-1. A la derecha el Circos del miR160a, un precursor conservado en plantas con mecanismo corto de loop a base.

Capítulo 4

Estudio bioinformático sobre ARN pequeños y su actividad en *A. thaliana*.

7.1 Introducción

En plantas existen distintos tipos de ARNs pequeños con distintos tamaño que varían de 20 a 25 nucleótidos, origen y mecanismo de acción. Basado en las diferencias en su biogénesis y su modo de acción, los ARN pequeños de plantas con funciones regulatorias en la expresión génica han sido agrupados en cuatro clases diferentes

- siARNs (small interfering RNAs) que es la clase más abundante.
- miARNs (microRNAs). De 20 a 22 nt de longitud.
- ta-siARN (trans-acting siARNs)
- nat-siARNs y nat-miARNs (natural antisense siRNAs y miRNAs)

El cambio climático es uno de los temas más importantes para la ciencia en este momento. Debido a que las plantas no regulan la temperatura, es posible que una modificación en la temperatura del ambiente pueda cambiar el procesamiento de los miARNs. Para analizar esto, se realizaron secuenciaciones de alto rendimiento de ARN pequeños de plantas crecidas a diferentes temperaturas. En estas mismas plantas se secuenciaron bibliotecas de PARE [49] para la identificación de genes blanco de miARNs que permite realizar estudios del degradoma de ARN, mediante técnicas de alto rendimiento. Esto lo hicimos para estudiar como afecta la temperatura a los genes blanco predichos anteriormente y si existen nuevos genes blanco regulados por miARNs a bajas y altas temperaturas.

7.2 Resultados y Discusión

7.2.1 Biogénesis y actividad de ARN pequeños de plantas en distintas temperaturas

Mapeo y ARN pequeños en condiciones de estrés dependientes de la temperatura

Primero queríamos estudiar si se modifica la biogénesis y actividad de los ARN pequeños en distintas temperaturas. Para esto realizamos secuenciación masiva para bibliotecas de plantas de ARNm de *A. thaliana* a baja temperatura (8°C), a temperatura normal (22°C) y a alta temperatura (37°C).

Brevemente, un total de ~ 26 millones, ~ 18 millones y ~ 23 millones de lecturas en promedio fueron secuenciadas para baja temperatura, normal y alta. Las lecturas fueron filtradas para remover los adaptadores y antes de mapearlas contra el genoma de *A. thaliana*, observamos la distribución del tamaño de las secuencias totales (Figura . Las lecturas están normalizadas para poder hacer una comparación cuantitativa de las lecturas.

Observamos una distribución consistente del largo de las secuencias únicas a lo largo de las bibliotecas. La distribución varía entre 18nt y 30 nt. También observamos que la mayoría de las lecturas muestran un pico en largos entre 20 y 24 nt, consistente con reportes previos. El enriquecimiento para las todas las bibliotecas, sin importar la temperatura, está en los de 21 nt y 24 nt (Figura 7.1).

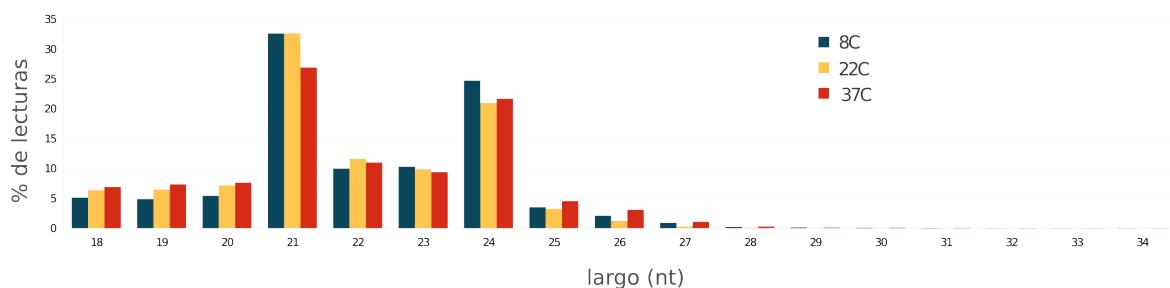


Figura 7.1 Distribución del tamaño de las lecturas de ARN pequeños Distribución del tamaño de las lecturas totales de ARN pequeños de las 3 bibliotecas secuenciadas (8°C, 22°C y 37°C).

Luego estás secuencias fueron mapeadas contra los elementos del genoma de *A. thaliana*. Y por último las lecturas redundantes fueron agrupadas en un dataset y el número de copias de cada lectura fue asignado a una lectura única (etiqueta). Esto dio como resultado a un

conjunto de lecturas no redundantes de ~ 2 millones para 8°C , $\sim 1,3$ millones para 22°C y $\sim 1,6$ millones para 37°C , de etiquetas únicas (Tabla 7.7).

Table 7.7 Bibliotecas de ARN pequeños

Código	Muestras de plantas	Secuencias totales	Lecturas que mapean contra el genoma	Lecturas únicas que mapean contra el genoma
col_8c_1s	Muestra 1 a 8°C	26594034	15235935	2181244
col_8c_2s	Muestra 2 a 8°C	24503495	14064726	2073756
col_22c_1s	Muestra 1 a 22°C	13922338	6805121	952466
col_22c_2s	Muestra 2 a 22°C	23037685	12430935	1699845
col_37c_1s	Muestra 1 a 37°C	22273823	9109248	1388169
col_37c_2s	Muestra 2 a 37°C	24307772	12748308	1867528

Una vez mapeadas las lecturas únicas contra el genoma, observamos nuevamente la distribución del tamaño de las lecturas en las distintas bibliotecas. Y observamos que la clase más abundante a distintas temperaturas son las secuencias de 21nt (Figura 7.3), aunque se ve una diferencia poco significativa a 8°C , 22°C y 37°C .

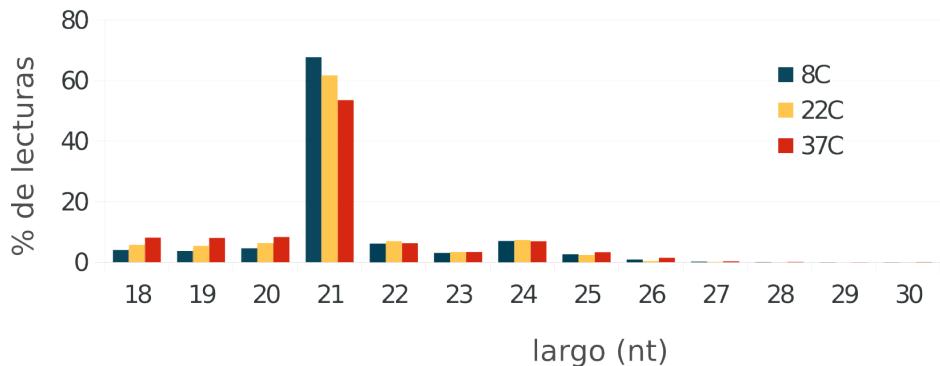


Figura 7.2 Distribución del tamaño de las lecturas de ARN pequeños que mapean contra el genoma de *A. thaliana*. Distribución del tamaño de las lecturas de ARN pequeños que mapean contra el genoma de *A. thaliana*. Se muestra la distribución para las 3 bibliotecas secuenciadas (8°C , 22°C y 37°C).

Perfil comparativo de ARN pequeños a distintas temperaturas y asignación a través de las características genómicas de *A. thaliana*

Para las lecturas únicas que mapean contra el genoma de *A. thaliana*, observamos la distribución de estos ARN pequeños a lo largo del mismo (Figura 7.3). Observamos que la clase más abundante para las 3 temperaturas era la de ARN ribosomal, donde es mayoritaria en

casi el 50% de las lecturas mapeadas. Luego aparecen los miARNs y ARN pequeños que pegan contra genes que codifican para proteínas. Otra de las clases más abundantes fueron los transposones y en menor porcentaje otros ARNs entre ellos ARN pequeños nucleares (Figura 7.3).

No se observan grandes diferencias en la clasificación de los ARN pequeños más abundantes, aunque cuanto más baja es la temperatura parecería que la clase de elementos transponibles aumentan. Lo opuesto pasa con los ARNs ribosomales (Figura 7.3).

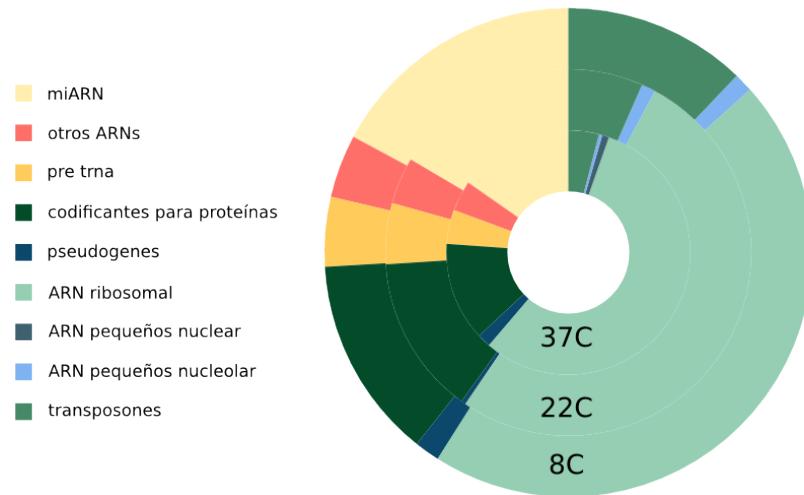


Figura 7.3 Distribución de ARN pequeños en el genoma de *A. thaliana*

Entre ellos los más abundantes son miembros de la familia del miR165/166, el miR396 (siendo el miR396b más abundante que el miR396a), el miR398b/c, el miR158, el miR159, el miR156/157, el miR403 y el miR408. En cuanto a los ARN pequeños identificados como miARNs, se detectaron una gran cantidad de miARNs conocidos. La tabla completa para miARNs que tienen más de 100 lecturas, sumando todas las bibliotecas juntas, se muestra en la tabla del anexo 9.10.

En total nos quedamos con 1271 ARN pequeños de plantas crecidas a 8°C, con 1305 ARN pequeños de plantas crecidas a 22°C y con 1129 ARN pequeños de plantas crecidas 37°C. Si bien la mayoría de los ARN pequeños encontrados son los mismos en las 3 bibliotecas, existen algunos que sólo se detectaron en dos bibliotecas o en una sola. Estas diferencias se debe en gran parte al corte arbitrario del número de lecturas que utilizamos para el conjunto de ARN pequeños final.

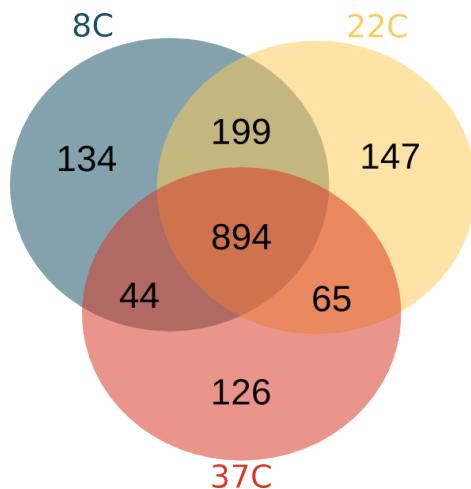


Figura 7.4 Diagrama de Venn de ARN pequeños más abundantes a diferentes temperaturas

7.2.2 Pipeline automatizado para el análisis genómico de ARN pequeños y su actividad in vivo en plantas

Además de las bibliotecas de ARN pequeños, utilizamos estas mismas plantas sometidas a estrés en distintas temperaturas para realizar bibliotecas de PARE[49]. La misma permite realizar estudios del degradoma de ARN e identificar fragmentos cortados por miARNs u otros ARN pequeños que causen cortes endonucleolíticos de los ARNm blanco. Para esto utilizamos una estrategia y desarrollamos un pipeline automatizado para la predicción de genes blanco de ARN pequeños de *A. thaliana* (Figura 7.5).

Utilizamos CleaveLand [3], que es una herramienta computacional para la detección de cortes de miARNs en datos de degradoma. Cleaveland toma como entrada datos de degradomas, ARN pequeños y una base de datos de ARNm y devuelve genes blanco de ARN pequeños. En nuestro caso utilizamos todos los ARN pequeños más abundantes mapeados de las bibliotecas a distintas temperaturas que mostramos anteriormente (Figura 7.5).

Los resultados de este pipeline automatizado se muestran en la Figura 7.6. De los 1271 ARN pequeños a 8°C, se encontraron 78 ARNm con señal por PARE de los cuales son cortados por ARN pequeños que provienen de distintos ARNs: 10 de ARN ribosomal, 1 de genes que codifican para proteínas, 65 por pre-miARNs y 2 por otros ARN pequeños. A 22 °C, se encontraron 67 ARNm con señal por PARE de los cuales son cortados por ARN pequeños que provienen de distintos ARNs: 16 de ARN ribosomal, 1 de genes que codifican para proteínas, 49 por pre-miARNs y 1 por otros ARN pequeños. Y a 37 °C, se encontraron 65 ARNm con señal por PARE de los cuales son cortados por ARN pequeños que provienen

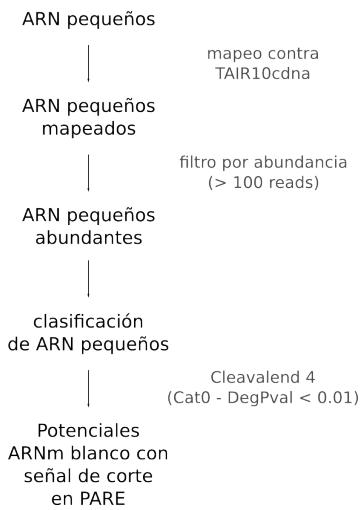


Figura 7.5 Estrategia para el análisis de bibliotecas de ARN pequeños de *A. thaliana*. Las bibliotecas de ARN pequeños fueron filtradas para remover los adaptadores y luego fueron mapeadas contra el genoma de *A. thaliana*. Nos quedamos con los ARN pequeños más abundantes (>100 reads en ambas muestras) que luego fueron clasificados. Con estos ARN pequeños hicimos una búsqueda sistematizada de potenciales ARNm blanco que tienen señal de corte por PARE.

de distintos ARNs: 11 de ARN ribosomal, 1 de genes que codifican para proteínas, 51 por pre-miARNs y 2 por otros ARN pequeños (Figura 7.6).

Estos resultados son utilizando parámetros de CleaveLand bien estrictos por simplicidad. Donde por ejemplo solo se consideran los genes blanco con señal en el degradoma que están en la categoría 0 de CleaveLand que significa que tiene más de una lectura y además es la más abundante del transcripto y además es la única mayoritaria. Esto se hizo para eliminar falsos positivos, pero se podrían relajar los parámetros para obtener mayor cantidad de candidatos y luego estudiarlos en mayor detalle.

Esta misma estrategia puede ser utilizada para encontrar potenciales genes blanco de ARN pequeños en regiones intergénicas o en intrones.

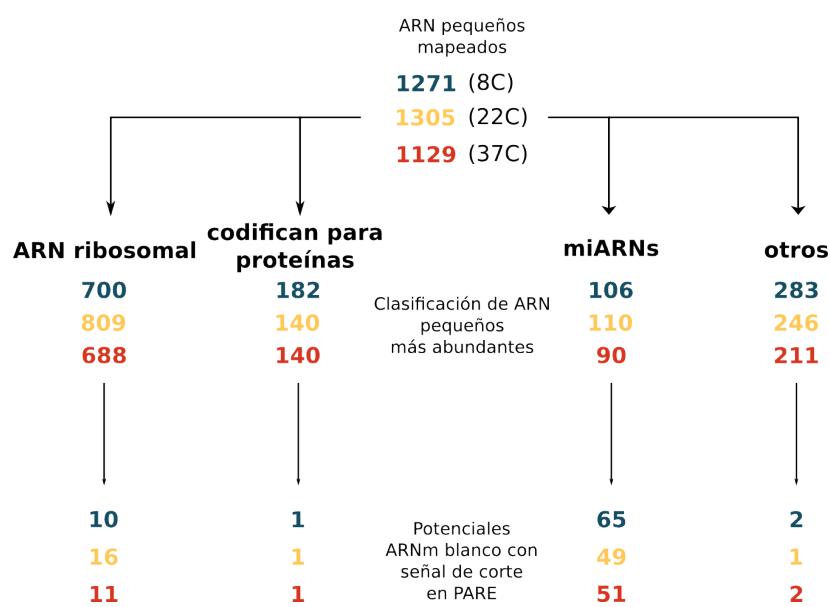


Figura 7.6 ARNm blanco con señal de corte en PARE a distintas temperaturas En color azul se muestran los resultados para los ARN pequeños y potenciales genes blanco en plantas crecidas a 8°C. En amarillo para plantas crecidas a 22°C. En rojo para plantas crecidas a 37°C.

Conclusiones

8.1 Aplicaciones bioinformáticas para el estudio de interacciones miARN-gen blanco

En cuanto a la primera parte de la Tesis y mediante diferentes estrategias y estudios, hemos alcanzado las siguientes conclusiones:

- Diseñamos una estrategia para identificar genes blanco regulados por miARNs en plantas, basado en la conservación evolutiva del par microARN-gen blanco.
- El enfoque requiere que la interacción miARN-gen blanco, pueda ocurrir en el contexto de un conjunto mínimo de parámetros que interactúan en diferentes especies. Pero la secuencia del gen blanco en sí, no necesariamente tiene que estar conservada.
- Además, nuestro enfoque permite ajustar el número de especies requeridas como un filtro para realizar la búsqueda con diferentes sensibilidades y relaciones señal/ruido.
- Utilizando esta estrategia identificamos y validamos experimentalmente nuevos genes blanco en *A. thaliana*, a pesar de que este sistema ya había sido estudiado en detalles en distintos enfoques genómicos a gran escala ([2, 5, 50, 63, 108, 123]).
- Tres de los nuevos genes blanco validados tienen bulges. Parámetros empíricos usualmente le otorgan una gran penalidad a ellos, que puede llegar a ser el doble que un mismatch regular [63], sin embargo es probable que genes blancos con bulges asimétricos sean más frecuente de lo que se pensaba previamente en plantas.
- El enfoque ofrece una estrategia alternativa a otras predicciones que se basan en parámetros empíricos del par miARN-gen blanco [5, 35, 44, 63].
- Una ventaja de la estrategia presentada es que las interacciones miARN-gen blanco conservadas probablemente participen en procesos biológicos relevantes.

- Además, esta estrategia puede ser fácilmente modificada para incorporar datos de otras bibliotecas, y/o para realizar la búsqueda de genes blanco presentes en un grupo específico de especies de plantas.

Esta parte del trabajo de Tesis fue publicado en la revista Nucleic Acid Research [32] y en la revista Bioinformatics [33].

8.2 Estudios genómicos sobre la biogénesis de miARNs

En la segunda parte de esta Tesis, desarrollamos un enfoque genómico a gran escala para estudiar el procesamiento de miARNs en plantas y determinamos, de esta manera, el mecanismo de procesamiento de la mayoría de los miARNs de *A. thaliana* conservados evolutivamente [21]. Encontramos que los miARNs en plantas pueden ser procesados por cuatro mecanismos, dependientes de la dirección secuencial de la maquinaria de procesamiento y del número de cortes requeridos para liberar el miARN (Figura 5.7) [21]. Precursores procesados en el mismo mecanismo comparten determinantes estructurales, explicando el gran rango de tamaño y forma observado en precursores de miARNs en plantas [21]. Además desarrollamos una estrategia bioinformática y luego una herramienta para estudiar a los intermediarios de procesamiento en plantas mutantes en proteínas de procesamiento. Con esta herramienta se puede determinar la dirección de procesamiento y se puede observar diferencias en los cortes determinados por DCL1 y precisión de la maquinaria en mutantes de procesamiento.

Realizamos un enfoque bioinformático para el estudio de la evolución y biogénesis de miARNs en plantas. De esta manera estudiamos la evolución en plantas con distintos mecanismos de procesamiento. Y pudimos caracterizar la relación entre patrones de conservación y mecanismos de procesamiento determinados previamente. Además, mostramos que esta técnica se puede utilizar también para estudiar precursores en animales.

Por otro lado generamos un pipeline para el análisis a gran escala de ARN pequeños que se puede aplicar a distintas condiciones de estrés y en distintas especies de plantas. También utilizamos la técnica de PARE para identificar genes regulados por ARN pequeños de forma sistematizada.

Bibliografía

- [1] Addo-Quaye, C., Eshoo, T. W., Bartel, D. P., and Axtell, M. J. (2008). Endogenous siRNA and miRNA targets identified by sequencing of the *Arabidopsis* degradome. *Curr. Biol.*, 18(10):758–762.
- [2] Addo-quaye, C., Eshoo, T. W., Bartel, D. P., and Axtell, M. J. (2009). Endogenous siRNA and microRNA targets identified by sequencing of the *Arabidopsis* degradome. *NIH Public Access*, 18(10):758–762.
- [3] Addo-Quaye, C., Miller, W., and Axtell, M. J. (2009a). CleaveLand: a pipeline for using degradome data to find cleaved small RNA targets. *Bioinformatics*, 25(1):130–131.
- [4] Addo-Quaye, C., Snyder, J. A., Park, Y. B., Li, Y. F., Sunkar, R., and Axtell, M. J. (2009b). Sliced microRNA targets and precise loop-first processing of MIR319 hairpins revealed by analysis of the *Physcomitrella patens* degradome. *RNA*, 15(12):2112–2121.
- [5] Allen, E., Xie, Z., Gustafson, A. M., and Carrington, J. C. (2005). microRNA-Directed Phasing during Trans-Acting siRNA Biogenesis in Plants. *Cell*, 121(2):207–221.
- [6] Allen, E., Xie, Z., Gustafson, A. M., Sung, G. H., Spatafora, J. W., and Carrington, J. C. (2004). Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. *Nat. Genet.*, 36(12):1282–1290.
- [7] Altschup, S. F., Gish, W., Pennsylvania, T., and Park, U. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology*, pages 403–410.
- [8] Anders, G., Mackowiak, S. D., Jens, M., Maaskola, J., Kuntzagk, A., Rajewsky, N., Landthaler, M., and Dieterich, C. (2012). doRiNA: a database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Res.*, 40(Database issue):D180–186.
- [9] Arazi, T., Talmor-Neiman, M., Stav, R., Riese, M., Huijser, P., and Baulcombe, D. C. (2005). Cloning and characterization of micro-RNAs from moss. *The Plant journal : for cell and molecular biology*, 43(6):837–48.
- [10] Aukerman, M. J. and Sakai, H. (2003). Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. *Plant Cell*, 15(11):2730–2741.
- [11] Axtell, M. J. (2008). Evolution of microRNAs and their targets: are all microRNAs biologically relevant? *Biochimica et biophysica acta*, 1779(11):725–34.

- [12] Axtell, M. J. and Bartel, D. P. (2005). Antiquity of microRNAs and their targets in land plants. *Plant Cell*, 17(6):1658–1673.
- [13] Axtell, M. J. and Bowman, J. L. (2008). Evolution of plant microRNAs and their targets. *Trends in Plant Science*, 13(7):343–349.
- [14] Axtell, M. J., Westholm, J. O., and Lai, E. C. (2011). Vive la différence: biogenesis and evolution of microRNAs in plants and animals. *Genome Biol.*, 12(4):221.
- [15] Bailey, T. L. and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol*, 2:28–36.
- [16] Bartel, D. P., Lee, R., and Feinbaum, R. (2004). MicroRNAs: Genomics , Biogenesis , Mechanism , and Function Genomics : The miRNA Genes. *Cell*, 116:281–297.
- [17] Baulcombe, D. (2004). RNA silencing in plants. *Nature*, 431(7006):356–363.
- [18] Bo, X. and Wang, S. (2005). TargetFinder: a software for antisense oligonucleotide target site selection based on MAST and secondary structures of target mRNA. *Bioinformatics*, 21(8):1401–1402.
- [19] Bologna, N. G., Mateos, J. L., Bresso, E. G., and Palatnik, J. F. (2009). A loop-to-base processing mechanism underlies the biogenesis of plant microRNAs miR319 and miR159. *The EMBO journal*, 28(23):3646–56.
- [20] Bologna, N. G., Schapire, A. L., and Palatnik, J. F. (2012). Processing of plant microrna precursors. *Briefings in Functional Genomics*.
- [21] Bologna, N. G., Schapire, A. L., Zhai, J., Chorostecki, U., Boisbouvier, J., Meyers, B. C., and Palatnik, J. F. (2013). Multiple RNA recognition patterns during microRNA biogenesis in plants. *Genome research*, 23(10):1675–89.
- [22] Bonnet, E., He, Y., Billiau, K., and Van de Peer, Y. (2010). TAPIR, a web server for the prediction of plant microRNA targets, including target mimics. *Bioinformatics*, 26(12):1566–1568.
- [23] Borsani, O., Zhu, J., Verslues, P. E., Sunkar, R., and Zhu, J. K. (2005). Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell*, 123(7):1279–1291.
- [24] Boutet, S., Vazquez, F., Liu, J., Beclin, C., Fagard, M., Gratias, A., Morel, J. B., Crete, P., Chen, X., and Vaucheret, H. (2003). *Arabidopsis* HEN1: a genetic link between endogenous miRNA controlling development and siRNA controlling transgene silencing and virus resistance. *Curr. Biol.*, 13(10):843–848.
- [25] Bowman, J. L. (2004). Class III HD-Zip gene regulation, the golden fleece of ARGONAUTE activity? *Bioessays*, 26(9):938–942.
- [26] Brodersen, P., Sakvarelidze-Achard, L., Bruun-Rasmussen, M., Dunoyer, P., Yamamoto, Y. Y., Sieburth, L., and Voinnet, O. (2008). Widespread translational inhibition by plant miRNAs and siRNAs. *Science*, 320(5880):1185–1190.

- [27] Carrington, J. C. and Ambros, V. (2003). Role of microRNAs in plant and animal development. *Science*, 301(5631):336–338.
- [28] Chapman, E. J. and Carrington, J. C. (2007). Specialization and evolution of endogenous small RNA pathways. *Nat. Rev. Genet.*, 8(11):884–896.
- [29] Chekanova, J. A., Gregory, B. D., Reverdatto, S. V., Chen, H., Kumar, R., Hooker, T., Yazaki, J., Li, P., Skiba, N., Peng, Q., Alonso, J., Brukhin, V., Grossniklaus, U., Ecker, J. R., and Belostotsky, D. A. (2007). Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the *Arabidopsis* transcriptome. *Cell*, 131(7):1340–1353.
- [30] Chen, H. M., Chen, L. T., Patel, K., Li, Y. H., Baulcombe, D. C., and Wu, S. H. (2010). 22-Nucleotide RNAs trigger secondary siRNA biogenesis in plants. *Proc. Natl. Acad. Sci. U.S.A.*, 107(34):15269–15274.
- [31] Chen, X. (2004). A microRNA as a translational repressor of APETALA2 in *Arabidopsis* flower development. *Science*, 303(5666):2022–2025.
- [32] Chorostecki, U., Crosa, V. A., Lodeyro, A. F., Bologna, N. G., Martin, A. P., Carrillo, N., Schommer, C., and Palatnik, J. F. (2012). Identification of new miRNA-regulated genes by conserved targeting in plant species. *Nucleic Acids Research*.
- [33] Chorostecki, U. and Palatnik, J. F. (2014). comTAR: a web tool for the prediction and characterization of conserved microRNA targets in plants. *Bioinformatics (Oxford, England)*, 30(14):2066–7.
- [34] Cuperus, J. T., Carbonell, A., Fahlgren, N., Garcia-Ruiz, H., Burke, R. T., Takeda, A., Sullivan, C. M., Gilbert, S. D., Montgomery, T. A., and Carrington, J. C. (2010). Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in *Arabidopsis*. *Nat. Struct. Mol. Biol.*, 17(8):997–1003.
- [35] Cuperus, J. T., Fahlgren, N., and Carrington, J. C. (2011). Evolution and functional diversification of miRNA genes. *The Plant cell*, 23(2):431–442.
- [36] Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K., and Scheible, W. R. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiol.*, 139(1):5–17.
- [37] Dai, X. and Zhao, P. X. (2011). psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res.*, 39(Web Server issue):W155–159.
- [38] Davies, R. T., Goetz, D. H., Lasswell, J., Anderson, M. N., and Bartel, B. (1999). IAR3 Encodes an Auxin Conjugate Hydrolase from *Arabidopsis*. *Plant Cell*, 11(March):365–376.
- [39] Debernardi, J. M., Rodriguez, R. E., Mecchia, M. A., and Palatnik, J. F. (2012). Functional Specialization of the Plant miR396 Regulatory Network through Distinct MicroRNA–Target Interactions. *PLoS Genet*, 8(1):e1002419.

- [40] Denli, A. M., Tops, B. B., Plasterk, R. H., Ketting, R. F., and Hannon, G. J. (2004). Processing of primary microRNAs by the Microprocessor complex. *Nature*, 432(7014):231–235.
- [41] Ding, Y., Chan, C. Y., and Lawrence, C. E. (2004). Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res.*, 32(Web Server issue):W135–141.
- [42] Dugas, D. V. and Bartel, B. (2008). Sucrose induction of *Arabidopsis* miR398 represses two Cu/Zn superoxide dismutases. *Plant Mol. Biol.*, 67(4):403–417.
- [43] Fabian, M. R., Sonenberg, N., and Filipowicz, W. (2010). Regulation of mRNA Translation and Stability by microRNAs. *Annual Review of Biochemistry*, 79(79):351–79.
- [44] Fahlgren, N. and Carrington, J. (2010). mirna target prediction in plants. In Meyers, B. C. and Green, P. J., editors, *Plant MicroRNAs*, volume 592 of *Methods in Molecular Biology*, pages 51–57. Humana Press.
- [45] Fahlgren, N., Jogdeo, S., Kasschau, K. D., Sullivan, C. M., Chapman, E. J., Laubinger, S., Smith, L. M., Dasenko, M., Givan, S. a., Weigel, D., and Carrington, J. C. (2010). MicroRNA gene evolution in *Arabidopsis lyrata* and *Arabidopsis thaliana*. *The Plant cell*, 22(4):1074–89.
- [46] Floyd, S. K. and Bowman, J. L. (2004). Gene regulation: ancient microRNA target sequences in plants. *Nature*, 428(6982):485–486.
- [47] Friedman, R. C., Farh, K. K., Burge, C. B., and Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, 19(1):92–105.
- [48] Gasciolli, V., Mallory, A. C., Bartel, D. P., and Vaucheret, H. (2005). Partially redundant functions of *Arabidopsis* DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. *Curr. Biol.*, 15(16):1494–1500.
- [49] German, M. A., Luo, S., Schroth, G., Meyers, B. C., and Green, P. J. (2009). Construction of Parallel Analysis of RNA Ends (PARE) libraries for the study of cleaved miRNA targets and the RNA degradome. *Nat Protoc*, 4(3):356–362.
- [50] German, M. a., Pillay, M., Jeong, D.-H., Hetawal, A., Luo, S., Janardhanan, P., Kannan, V., Rymarquis, L. a., Nobuta, K., German, R., De Paoli, E., Lu, C., Schroth, G., Meyers, B. C., and Green, P. J. (2008). Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nature biotechnology*, 26(8):941–6.
- [51] Giegerich, R., Rehmsmeier, M., Steffen, P., and Ho, M. (2004). Fast and effective prediction of microRNA / target duplexes. *RNA*, 100(2003):1507–1517.
- [52] Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., and Rokhsar, D. S. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic acids research*, 40(Database issue):D1178–86.
- [53] Gregory, R. I., Yan, K. P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. (2004). The Microprocessor complex mediates the genesis of microRNAs. *Nature*, 432(7014):235–240.

- [54] Grimson, A., Farh, K. K., Johnston, W. K., Garrett-Engele, P., Lim, L. P., and Bartel, D. P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell*, 27(1):91–105.
- [55] Guil, S. and Caceres, J. F. (2007). The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nat. Struct. Mol. Biol.*, 14(7):591–596.
- [56] Guo, H., Ingolia, N. T., Weissman, J. S., and Bartel, D. P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, 466(7308):835–840.
- [57] Hamilton, A. J. and Baulcombe, D. C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science*, 286(5441):950–952.
- [58] Han, J., Lee, Y., Yeom, K. H., Kim, Y. K., Jin, H., and Kim, V. N. (2004a). The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev.*, 18(24):3016–3027.
- [59] Han, J., Lee, Y., Yeom, K. H., Nam, J. W., Heo, I., Rhee, J. K., Sohn, S. Y., Cho, Y., Zhang, B. T., and Kim, V. N. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell*, 125(5):887–901.
- [60] Han, M.-H., Goud, S., Song, L., and Fedoroff, N. (2004b). The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. *Proceedings of the National Academy of Sciences of the United States of America*, 101(4):1093–8.
- [61] Heo, I., Joo, C., Cho, J., Ha, M., Han, J., and Kim, V. N. (2008). Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol. Cell*, 32(2):276–284.
- [62] Jha, A. and Shankar, R. (2011). Employing machine learning for reliable miRNA target identification in plants. *BMC Genomics*, 12:636.
- [63] Jones-Rhoades, M. W. and Bartel, D. P. (2004). Computational identification of plant microRNAs and their targets, including a stress-induced mirna. *Molecular Cell*, 14(6):787 – 799.
- [64] Jones-Rhoades, M. W., Bartel, D. P., and Bartel, B. (2006). MicroRNAs and their regulatory roles in plants. *Annual review of plant biology*, 57:19–53.
- [65] Kakrana, A., Hammond, R., Patel, P., Nakano, M., and Meyers, B. C. (2014). sPARTA: a parallelized pipeline for integrated analysis of plant miRNA and cleaved mRNA data sets, including new miRNA target-identification software. *Nucleic Acids Res.*, 42(18):e139.
- [66] Kasschau, K. D., Xie, Z., Allen, E., Llave, C., Chapman, E. J., Krizan, K. a., and Carrington, J. C. (2003). P1/HC-Pro, a Viral Suppressor of RNA Silencing, Interferes with Arabidopsis Development and miRNA Function. *Developmental Cell*, 4(2):205–217.
- [67] Katiyar-Agarwal, S., Morgan, R., Dahlbeck, D., Borsani, O., Villegas, A., Zhu, J. K., Staskawicz, B. J., and Jin, H. (2006). A pathogen-inducible endogenous siRNA in plant immunity. *Proc. Natl. Acad. Sci. U.S.A.*, 103(47):18002–18007.

- [68] Kim, V. N., Han, J., and Siomi, M. C. (2009). Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.*, 10(2):126–139.
- [69] Knauer, S., Holt, A. L., Rubio-Somoza, I., Tucker, E. J., Hinze, A., Pisch, M., Javelle, M., Timmermans, M. C., Tucker, M. R., and Laux, T. (2013). A protodermal miR394 signal defines a region of stem cell competence in the *Arabidopsis* shoot meristem. *Dev. Cell*, 24(2):125–132.
- [70] Kozomara, A. and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic acids research*, 42(Database issue):D68–73.
- [71] Krüger, J. and Rehmsmeier, M. (2006). Rnahybrid: microrna target prediction easy, fast and flexible. *Nucleic Acids Research*, 34(suppl 2):W451–W454.
- [72] Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., and Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.*, 19(9):1639–1645.
- [73] Kurihara, Y. and Watanabe, Y. (2004). Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc. Natl. Acad. Sci. U.S.A.*, 101(34):12753–12758.
- [74] Lai, E. C. (2004). Predicting and validating microRNA targets. *Genome Biol.*, 5(9):115.
- [75] Lanet, E., Delannoy, E., Sormani, R., Floris, M., Brodersen, P., Crete, P., Voinnet, O., and Robaglia, C. (2009). Biochemical evidence for translational repression by Arabidopsis microRNAs. *Plant Cell*, 21(6):1762–1768.
- [76] Laufs P, Peaucelle A, M. H. and J, T. (2004). MicroRNA regulation of the CUC genes is required for boundary size control in *Arabidopsis* meristems. *Development*, 12(5):622–627.
- [77] Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, 75(5):843–854.
- [78] Lekprasert, P., Mayhew, M., and Ohler, U. (2011). Assessing the utility of thermodynamic features for microRNA target prediction under relaxed seed and no conservation requirements. *PLoS ONE*, 6(6):e20622.
- [79] Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, 120(1):15–20.
- [80] Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P., and Burge, C. B. (2003). Prediction of mammalian microRNA targets. *Cell*, 115(7):787–798.
- [81] Llave, C., Xie, Z., Kasschau, K. D., and Carrington, J. C. (2002). Cleavage of Scarecrow-like mRNA Targets Directed by a Class of *Arabidopsis* miRNA. *Science*, 297(September):2053–2056.
- [82] Lobbes, D., Rallapalli, G., Schmidt, D. D., Martin, C., and Clarke, J. (2006). SERRATE: a new player on the plant microRNA scene. *EMBO reports*, 7(10):1052–8.

- [83] Lorenz, R., Bernhart, S. H., Honer Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., and Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms Mol Biol*, 6:26.
- [84] Lund, E., Guttinger, S., Calado, A., Dahlberg, J. E., and Kutay, U. (2004). Nuclear export of microRNA precursors. *Science*, 303(5654):95–98.
- [85] Mallory, A. C., Reinhart, B. J., Jones-rhoades, M. W., Zamore, P. D., Kathryn, M., and Bartel, D. P. (2004). MicroRNA control of PHABULOSA in leaf development : importance of pairing to the microRNA 5'0 region. *Embo journal*, 23(16):3356–3364.
- [86] Manavella, P. A., Koenig, D., and Weigel, D. (2012). Plant secondary siRNA production determined by microRNA-duplex structure. *Proc. Natl. Acad. Sci. U.S.A.*, 109(7):2461–2466.
- [87] Maragkakis, M., Alexiou, P., Papadopoulos, G. L., Reczko, M., Dalamagas, T., Giannopoulos, G., Goumas, G., Koukis, E., Kourtis, K., Simossis, V. A., Sethupathy, P., Vergoulis, T., Koziris, N., Sellis, T., Tsanakas, P., and Hatzigeorgiou, A. G. (2009). Accurate microRNA target prediction correlates with protein repression levels. *BMC Bioinformatics*, 10:295.
- [88] Maragkakis, M., Vergoulis, T., Alexiou, P., Reczko, M., Plomaritou, K., Gousis, M., Kourtis, K., Koziris, N., Dalamagas, T., and Hatzigeorgiou, A. G. (2011). DIANA-microT Web server upgrade supports Fly and Worm miRNA target prediction and bibliographic miRNA to disease association. *Nucleic Acids Res.*, 39(Web Server issue):W145–148.
- [89] Markham, N. R. and Zuker, M. (2005). DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res.*, 33(Web Server issue):W577–581.
- [90] Mateos, J. L., Bologna, N. G., Chorostecki, U., and Palatnik, J. F. (2010). Identification of microRNA processing determinants by random mutagenesis of Arabidopsis MIR172a precursor. *Current Biology*, 20(1):49–54.
- [91] Maunoury, N. (2011). AGO1 and AGO2 Act Redundantly in miR408-Mediated Plantacyanin Regulation. *Plos One*, 6(12).
- [92] Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C., Chen, S., Hannon, G. J., and Qi, Y. (2008). Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. *Cell*, 133(1):116–127.
- [93] Michlewski, G. and Caceres, J. F. (2010). Antagonistic role of hnRNP A1 and KSRP in the regulation of let-7a biogenesis. *Nat. Struct. Mol. Biol.*, 17(8):1011–1018.
- [94] Millar, A. a. and Gubler, F. (2005). The Arabidopsis GAMYB-like genes, MYB33 and MYB65, are microRNA-regulated genes that redundantly facilitate anther development. *The Plant cell*, 17(3):705–21.
- [95] Montgomery, T. A., Howell, M. D., Cuperus, J. T., Li, D., Hansen, J. E., Alexander, A. L., Chapman, E. J., Fahlgren, N., Allen, E., and Carrington, J. C. (2008a). Specificity of ARGONAUTE7-miR390 Interaction and Dual Functionality in TAS3 Trans -Acting siRNA Formation. *Cell*, pages 128–141.

- [96] Montgomery, T. A., Yoo, S. J., Fahlgren, N., Gilbert, S. D., Howell, M. D., Sullivan, C. M., Alexander, A., Nguyen, G., Allen, E., Ahn, J. H., and Carrington, J. C. (2008b). AGO1-miR173 complex initiates phased siRNA formation in plants. *Proc. Natl. Acad. Sci. U.S.A.*, 105(51):20055–20062.
- [97] Needleman, S. B. and Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48(3):443 – 453.
- [98] Niyogi, K. K., Pilon, M., Shikanai, T., Abdel-ghany, S. E., and Mu, P. (2005). Two P-Type ATPases Are Required for Copper Delivery in *Arabidopsis thaliana* Chloroplasts. *Plant Cell*, 17(April):1233–1251.
- [99] Notredame, C., Higgins, D. G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.*, 302(1):205–217.
- [100] Obernosterer, G., Leuschner, P. J., Alenius, M., and Martinez, J. (2006). Post-transcriptional regulation of microRNA expression. *RNA*, 12(7):1161–1167.
- [101] Okamura, K., Hagen, J. W., Duan, H., Tyler, D. M., and Lai, E. C. (2007). The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell*, 130(1):89–100.
- [102] Ossowski, S., Schwab, R., and Weigel, D. (2008). Gene silencing in plants using artificial microRNAs and other small RNAs. *Plant J.*, 53(4):674–690.
- [103] Palatnik, J. F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J. C., and Weigel, D. (2003). Control of leaf morphogenesis by microRNAs. *Nature*, 425(6955):257–263.
- [104] Palatnik, J. F., Wollmann, H., Schommer, C., Schwab, R., Boisbouvier, J., Rodriguez, R., Warthmann, N., Allen, E., Dezulian, T., Huson, D., Carrington, J. C., and Weigel, D. (2007). Sequence and expression differences underlie functional specialization of *Arabidopsis* microRNAs miR159 and miR319. *Developmental cell*, 13(1):115–25.
- [105] Paraskevopoulou, M. D., Georgakilas, G., Kostoulas, N., Vlachos, I. S., Vergoulis, T., Reczko, M., Filippidis, C., Dalamagas, T., and Hatzigeorgiou, A. G. (2013). DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic Acids Res.*, 41(Web Server issue):W169–173.
- [106] Pasquinelli, A. E., Reinhart, B. J., Slack, F., Martindale, M. Q., Kuroda, M. I., Maller, B., Hayward, D. C., Ball, E. E., Degnan, B., Muller, P., Spring, J., Srinivasan, A., Fishman, M., Finnerty, J., Corbo, J., Levine, M., Leahy, P., Davidson, E., and Ruvkun, G. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*, 408(6808):86–89.
- [107] Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.*, 29(9):e45.
- [108] Rajagopalan, R., Vaucheret, H., Trejo, J., and Bartel, D. P. (2006). A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes & development*, 20(24):3407–25.

- [109] Ramachandran, V. and Chen, X. (2008). Small RNA metabolism in Arabidopsis. *Trends Plant Sci.*, 13(7):368–374.
- [110] Rampey, R. A., Leclere, S., Kowalczyk, M., Ljung, K., Bartel, B., Biology, C., and Texas, R. A. R. (2004). A Family of Auxin-Conjugate Hydrolases That Contributes to Free Indole-3-Acetic Acid Levels during Arabidopsis Germination 1. *Plant Physiol.*, 135(June):978–988.
- [111] Rehmsmeier, M., Steffen, P., Hochsmann, M., and Giegerich, R. (2004). Fast and effective prediction of microRNA/target duplexes. *RNA*, 10(10):1507–1517.
- [112] Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B., and Bartel, D. P. (2002). MicroRNAs in plants. *Genes Dev*, pages 1616–1626.
- [113] Ren, G., Xie, M., Zhang, S., Vinovskis, C., Chen, X., and Yu, B. (2014). Methylation protects microRNAs from an AGO1-associated activity that uridylates 5' RNA fragments generated by AGO1 cleavage. *Proc. Natl. Acad. Sci. U.S.A.*, 111(17):6365–6370.
- [114] Rhoades, M. W., Reinhart, B. J., Lim, L. P., Burge, C. B., Bartel, B., and Bartel, D. P. (2002). Prediction of plant microrna targets. *Cell*, 110(4):513 – 520.
- [115] Rodriguez, R. E., Mecchia, M. A., Debernardi, J. M., Schommer, C., Weigel, D., and Palatnik, J. F. (2010). Control of cell proliferation in Arabidopsis thaliana by microRNA miR396. *Development*, 112:103–112.
- [116] Rubio-Somoza, I., Cuperus, J. T., Weigel, D., and Carrington, J. C. (2009). Regulation and functional specialization of small RNA-target nodes during plant development. *Curr. Opin. Plant Biol.*, 12(5):622–627.
- [117] Rubio-Somoza, I. and Weigel, D. (2011). MicroRNA networks and developmental plasticity in plants. *Trends Plant Sci.*, 16(5):258–264.
- [118] Ruby, J. G., Jan, C. H., and Bartel, D. P. (2007a). Intronic microRNA precursors that bypass Drosha processing. *Nature*, 448(7149):83–86.
- [119] Ruby, J. G., Stark, A., Johnston, W. K., Kellis, M., Bartel, D. P., and Lai, E. C. (2007b). Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. *Genome Res.*, 17(12):1850–1864.
- [120] Schapire, A. L., Bologna, N. G., Moro, B., Zhai, J., Meyers, B. C., and Palatnik, J. F. (2013). Construction of Specific Parallel Amplification of RNA Ends (SPARE) libraries for the systematic identification of plant microRNA processing intermediates. *Methods (San Diego, Calif.)*, 64(3):283–91.
- [121] Schauer, S. E., Jacobsen, S. E., Meinke, D. W., and Ray, A. (2002). DICER-LIKE1: blind men and elephants in Arabidopsis development. *Trends Plant Sci.*, 7(11):487–491.
- [122] Schiefthaler, Balasubramanian, Sieber, Chevalier, Wisman, and Schneitz (1999). Molecular analysis of NOZZLE , a gene involved in pattern formation and early sporogenesis during sex organ development in Arabidopsis thaliana. *Proc. Natl Acad. Sci.*, 96(September):11664–11669.

- [123] Schwab, R., Palatnik, J. F., Riester, M., Schommer, C., Schmid, M., and Weigel, D. (2005). Specific effects of micrornas on the plant transcriptome. *Developmental Cell*, 8(4):517 – 527.
- [124] Shenoy, A. and Blelloch, R. H. (2014). Regulation of microRNA function in somatic stem cell proliferation and differentiation. *Nat. Rev. Mol. Cell Biol.*, 15(9):565–576.
- [125] Song, L., Axtell, M. J., and Fedoroff, N. V. (2010). RNA secondary structural determinants of miRNA precursor processing in Arabidopsis. *Curr. Biol.*, 20(1):37–41.
- [126] Song, L., Han, M. H., Lesicka, J., and Fedoroff, N. (2007). Arabidopsis primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body. *Proc. Natl. Acad. Sci. U.S.A.*, 104(13):5437–5442.
- [127] Souret, F. F., Kastenmayer, J. P., and Green, P. J. (2004). AtXRN4 degrades mRNA in Arabidopsis and its substrates include selected miRNA targets. *Mol. Cell*, 15(2):173–183.
- [128] Srivastava, P. K., Moturu, T. R., Pandey, P., Baldwin, I. T., and Pandey, S. P. (2014). A comparison of performance of plant miRNA target prediction tools and the characterization of features for genome-wide target prediction. *BMC Genomics*, 15:348.
- [129] Stark, A., Brennecke, J., Bushati, N., Russell, R. B., and Cohen, S. M. (2005). Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. *Cell*, 123(6):1133–1146.
- [130] Stark, A., Brennecke, J., Russell, R. B., and Cohen, S. M. (2003). Identification of Drosophila MicroRNA targets. *PLoS Biol.*, 1(3):E60.
- [131] Sunkar, R., Girke, T., Jain, P. K., and Zhu, J. K. (2005). Cloning and characterization of microRNAs from rice. *Plant Cell*, 17(5):1397–1411.
- [132] Sunkar, R. and Jagadeeswaran, G. (2008). In silico identification of conserved microRNAs in large number of diverse plant species. *BMC Plant Biol.*, 8:37.
- [133] Takeda, A., Iwasaki, S., and Watanabe, T. (2008). The Mechanism Selecting the Guide Strand from Small RNA Duplexes is Different Among Argonaute Proteins. *Plant Cell Physiol*, 49(4):493–500.
- [134] Thadani, R. and Tammi, M. T. (2006). MicroTar: predicting microRNA targets from RNA duplexes. *BMC Bioinformatics*, 7 Suppl 5:S20.
- [135] Todesco, M., Balasubramanian, S., Cao, J., Ott, F., Sureshkumar, S., Schneeberger, K., Meyer, R. C., Altmann, T., and Weigel, D. (2012). Natural variation in biogenesis efficiency of individual *Arabidopsis thaliana* microRNAs. *Curr. Biol.*, 22(2):166–170.
- [136] Trabucchi, M., Briata, P., Garcia-Mayoral, M., Haase, A. D., Filipowicz, W., Ramos, A., Gherzi, R., and Rosenfeld, M. G. (2009). The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. *Nature*, 459(7249):1010–1014.
- [137] Vaucheret, H. (2006). Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes Dev.*, 20(7):759–771.

- [138] Vaucheret, H. (2008). Plant ARGONAUTES ′. *Cell.*
- [139] Vazquez, F., Cre, P., and Bartel, D. P. (2004a). The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. *Genes Dev*, pages 1187–1197.
- [140] Vazquez, F., Gasciolli, V., Crété, P., and Vaucheret, H. (2004b). The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. *Current biology : CB*, 14(4):346–51.
- [141] Viswanathan, S. R., Daley, G. Q., and Gregory, R. I. (2008). Selective blockade of microRNA processing by Lin28. *Science*, 320(5872):97–100.
- [142] Voinnet, O. (2001). RNA silencing as a plant immune system against viruses. *Trends Genet.*, 17(8):449–459.
- [143] Voinnet, O. (2009). Origin, Biogenesis, and Activity of Plant MicroRNAs. *Cell*, 136(4):669–687.
- [144] Werner, S., Wollmann, H., Schneeberger, K., and Weigel, D. (2010). Structure determinants for accurate processing of miR172a in *Arabidopsis thaliana*. *Curr. Biol.*, 20(1):42–48.
- [145] Xie, F. and Zhang, B. (2010). Target-align: a tool for plant microRNA target identification. *Bioinformatics*, 26(23):3002–3003.
- [146] Xie, Z., Allen, E., Fahlgren, N., Calamar, A., Givan, S. A., and Carrington, J. C. (2005). Expression of *Arabidopsis* MIRNA Genes . *Plant Physiology*, 138(August):2145–2154.
- [147] Xie, Z., Kasschau, K. D., and Carrington, J. C. (2003). Negative Feedback Regulation of Dicer-Like1 in *Arabidopsis* by microRNA-Guided mRNA Degradation. *Current Biology*, 13(9):784–789.
- [148] Xie, Z. and Qi, X. (2008). Diverse small RNA-directed silencing pathways in plants. *Biochim. Biophys. Acta*, 1779(11):720–724.
- [149] Yamasaki, H., Abdel-Ghany, S. E., Cohu, C. M., Kobayashi, Y., Shikanai, T., and Pilon, M. (2007). Regulation of copper homeostasis by micro-RNA in *Arabidopsis*. *The Journal of biological chemistry*, 282(22):16369–78.
- [150] Yan, T., Yoo, D., Berardini, T. Z., Mueller, L. A., Weems, D. C., Weng, S., Cherry, J. M., and Rhee, S. Y. (2005). Patmatch: a program for finding patterns in peptide and nucleotide sequences. *Nucleic Acids Research*, 33(suppl 2):W262–W266.
- [151] Yang, L., Liu, Z., Lu, F., Dong, A., and Huang, H. (2006). SERRATE is a novel nuclear regulator in primary microRNA processing in *Arabidopsis*. *Plant J.*, 47(6):841–850.
- [152] Yang, W.-c., Ye, D., Xu, J., and Sundaresan, V. (1999). The SPOROCYTELESS gene of *Arabidopsis* is required for initiation of sporogenesis and encodes a novel nuclear protein. *Genes Dev.*, pages 2108–2117.

- [153] Yi, R., Qin, Y., Macara, I. G., and Cullen, B. R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.*, 17(24):3011–3016.
- [154] Yoshikawa, M., Peragine, A., Park, M. Y., and Poethig, R. S. (2005). A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. *Genes Dev.*, 19(18):2164–2175.
- [155] Zeng, Y. and Cullen, B. R. (2005). Efficient processing of primary microRNA hairpins by Drosophila requires flanking nonstructured RNA sequences. *J. Biol. Chem.*, 280(30):27595–27603.
- [156] Zhai, J., Arikit, S., Simon, S. A., Kingham, B. F., and Meyers, B. C. (2014). Rapid construction of parallel analysis of RNA end (PARE) libraries for Illumina sequencing. *Methods*, 67(1):84–90.
- [157] Zhang, B., Pan, X., Cannon, C. H., Cobb, G. P., and Anderson, T. A. (2006). Conservation and divergence of plant microRNA genes. *Plant J.*, 46(2):243–259.
- [158] Zhang, L., Chia, J. M., Kumari, S., Stein, J. C., Liu, Z., Narechania, A., Maher, C. A., Guill, K., McMullen, M. D., and Ware, D. (2009). A genome-wide characterization of microRNA genes in maize. *PLoS Genet.*, 5(11):e1000716.
- [159] Zhang, W., Gao, S., Zhou, X., Xia, J., Chellappan, P., Zhou, X., Zhang, X., and Jin, H. (2010). Multiple distinct small RNAs originate from the same microRNA precursors. *Genome Biol.*, 11(8):R81.
- [160] Zheng, Y., Li, Y. F., Sunkar, R., and Zhang, W. (2012). SeqTar: an effective method for identifying microRNA guided cleavage sites from degradome of polyadenylated transcripts in plants. *Nucleic Acids Res.*, 40(4):e28.
- [161] Zhu, H., Zhou, Y., Castillo-Gonzalez, C., Lu, A., Ge, C., Zhao, Y. T., Duan, L., Li, Z., Axtell, M. J., Wang, X. J., and Zhang, X. (2013). Bidirectional processing of pri-miRNAs with branched terminal loops by Arabidopsis Dicer-like1. *Nat. Struct. Mol. Biol.*, 20(9):1106–1115.
- [162] Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, 31(13):3406–3415.

Anexo

Predicción de genes regulados por miARNs.

Listing 9.1 circos.conf

```
<<include etc/colors_fonts_patterns.conf>>
<<include etc/housekeeping.conf>>

<image>
    angle_offset* = 114.5
    <<include etc/image.conf>>
</image>

# Chromosome name, size and color definition
karyotype = data/karyotype.txt

<ideogram>
    <spacing>
        default = 0.001r
        <pairwise last first>
            spacing = 200r
        </pairwise>
    </spacing>

    label_format = eval( if(var(label) =~ /(\d*)(\w*)/){ if($2 eq 'x'){ sprintf "%d -",$1;} else{ sprintf "%d %s",$1,$2; } })
    radius = 0.80r
    thickness = 35p
    fill = yes
    stroke_color = dgrey
    stroke_thickness = 0p
    show_label = yes
    label_font = semibold
    label_radius = 1r + 15p
    label_color = grey
    label_size = 20
    label_parallel = no

</ideogram>

<<include data/links.conf>>

<highlights>
    <highlight>
        file = data/highlight.txt
        fill_color = 96,122,66
        r0 = 1.10r
        r1 = 1.12r
    </highlight>
</highlights>

<plots>
    type = text
    color = black
    label_font = default
    label_size = 40p
```

```
<plot>
  file = data/text_5p_3p.txt
  r0 = 1r+80p
  r1 = 1r+200p
  label_parallel = yes

  show_links = yes
  link_dims = 4p,4p,8p,4p,4p
  link_thickness = 2p
  link_color = black

</plot>
<<include data/histogram.conf>>

</plots >
```

Listing 9.2 histogram.conf

```
<plot>
  type = histogram
  file = data/histogram.txt
  r1 = 0.97r
  r0 = 0.8r
  max = 0.5
  min = -0.5
  stroke_type = outline
  thickness = 0
  color = vdgrey
  extend_bin = no

  <backgrounds>
    <background>
      color = 231,231,231
    </background>
  </backgrounds>

  <axes>
    <axis>
      spacing = 0.2r
      color = lgrey
      thickness = 1
    </axis>
  </axes>

  <rules>
    <rule>
      condition = 1
      fill_color = eval(sprintf("circos_hist-%d-div-%d", remap_int(var(value), -0.5, 0.5, 1, 14)))
    </rule>
  </rules>

</plot>
```

Listing 9.3 links.conf

```
<links>
  z = 10
  radius = 0.79r

  <link>
    bezier_radius = undef
    show = yes
    color = 197,197,197
    file = link.txt

    <rules>
      <rule>
```

```
condition = var(thickness) == 8
thickness = 3
color      = circos_hist -14-div -14
z          = 15
</rule>

<rule>
  condition = var(thickness) <= 4
  show      = no
</rule>

<rule>
  condition = 1
  thickness = 3
  z         = 10
</rule>
</rules>

</link>

</links>
```

Table 9.8 Especies y base de datos utilizadas para la búsquedas de genes blanco de miARNs conservados

Specie	Database
Allium_cepa	http://compbio.dfci.harvard.edu/tgi/
Aquilegia	http://compbio.dfci.harvard.edu/tgi/
Arabidopsis_thaliana	http://arabidopsis.org/
Beta_vulgaris	http://compbio.dfci.harvard.edu/tgi/
Brassica_napus	http://compbio.dfci.harvard.edu/tgi/
Capsicum_annuum	http://compbio.dfci.harvard.edu/tgi/
Citrus_clementina	http://compbio.dfci.harvard.edu/tgi/
Citrus_sinensis	http://compbio.dfci.harvard.edu/tgi/
Coffea_canephora	http://compbio.dfci.harvard.edu/tgi/
Euphorbia_esula	http://compbio.dfci.harvard.edu/tgi/
Festuca_arundinacea	http://compbio.dfci.harvard.edu/tgi/
Glycine_max	http://compbio.dfci.harvard.edu/tgi/
Gossypium	http://compbio.dfci.harvard.edu/tgi/
Gossypium_raimondii	http://compbio.dfci.harvard.edu/tgi/
Haseolus_vulgaris	http://compbio.dfci.harvard.edu/tgi/
Helianthus_annuus	http://compbio.dfci.harvard.edu/tgi/
Hordeum_vulgare	http://compbio.dfci.harvard.edu/tgi/
Ipomoea_nil	http://compbio.dfci.harvard.edu/tgi/
Lactuca_sativa	http://compbio.dfci.harvard.edu/tgi/
Lactuca_serriola	http://compbio.dfci.harvard.edu/tgi/
Lotus_japonicus	http://compbio.dfci.harvard.edu/tgi/
Malus_x_domestica	http://compbio.dfci.harvard.edu/tgi/
Medicago_truncatula	http://compbio.dfci.harvard.edu/tgi/
Mesembryanthemum_crystallinum	http://compbio.dfci.harvard.edu/tgi/
Nicotiana_benthamiana	http://compbio.dfci.harvard.edu/tgi/
Nicotiana_tabacum	http://compbio.dfci.harvard.edu/tgi/
Oryza_sativa	http://www.jcvi.org/
Panicum_virgatum	http://compbio.dfci.harvard.edu/tgi/
Petunia_hybrida	http://compbio.dfci.harvard.edu/tgi/
Phaseolus_coccineus	http://compbio.dfci.harvard.edu/tgi/
Populus	http://compbio.dfci.harvard.edu/tgi/
Prunus_persica	http://compbio.dfci.harvard.edu/tgi/
Saccharum_officinarum	http://compbio.dfci.harvard.edu/tgi/
Secale_cereale	http://compbio.dfci.harvard.edu/tgi/
Solanum_lycopersicum	http://compbio.dfci.harvard.edu/tgi/
Solanum_tuberosum	http://compbio.dfci.harvard.edu/tgi/
Sorghum_bicolor	http://compbio.dfci.harvard.edu/tgi/
Theobroma_cacao	http://compbio.dfci.harvard.edu/tgi/
triphysaria	http://compbio.dfci.harvard.edu/tgi/
Triphysaria_versicolor	http://compbio.dfci.harvard.edu/tgi/
Triticum_aestivum	http://compbio.dfci.harvard.edu/tgi/
Vitis_vinifera	http://compbio.dfci.harvard.edu/tgi/
Zea_mays	http://compbio.dfci.harvard.edu/tgi/

Table 9.9 Genes blanco validados de familias miARNs conservados en plantas

microRNA	Target	ID
miR156/miR157	SPL	At1g27370
miR156/miR157	SPL	At1g53160
miR156/miR157	SPL	At2g33810
miR156/miR157	SPL	At3g15270
miR156/miR157	SPL	At5g43270
miR156/miR157	SPL	At1g69170
miR156/miR157	SPL	At2g42200
miR156/miR157	SPL	At3g57920
miR156/miR157	SPL	At5g50670
miR159/miR319	TCP	At1g30210
miR159/miR319	TCP	At1g53230
miR159/miR319	TCP	At2g31070
miR159/miR319	MYB	At3g11440
miR159/miR319	TCP	At3g15030
miR159/miR319	TCP	At4g18390
miR159/miR319	MYB	At5g06100
miR159/miR319	MYB	At2g26950
miR159/miR319	MYB	At2g32460
miR159/miR319	MYB	At5g55020
miR160	ARF	At1g77850
miR160	ARF	At2g28350
miR160	ARF	At4g30080
miR162	DCL	At1g01040
miR164	NAC	At1g56010
miR164	NAC	At3g15170
miR164	NAC	At5g07680
miR164	NAC	At5g53950
miR164	NAC	At5g01430
miR164	NAC	At3g12977
miR164	NAC	At5g39610
miR166/miR165	HD-ZIPIII	At1g30490
miR166/miR165	HD-ZIPIII	At1g52150
miR166/miR165	HD-ZIPIII	At2g34710
miR166/miR165	HD-ZIPIII	At5g06990
miR166/miR165	HD-ZIPIII	At4g32880
miR167	ARF	At1g30330
miR167	ARF	At5g37020
miR168	AGO	At1g48410
miR169	HAP2	At1g17590
miR169	HAP2	At1g54160
miR169	HAP2	At1g72830
miR169	HAP2	At3g05690
miR169	HAP2	At3g20910
miR169	HAP2	At5g06510
miR170/miR171	SCL	At2g45160
miR170/miR171	SCL	At3g60630
miR170/miR171	SCL	At4g00150
miR172	AP2	At2g28550
miR172	AP2	At4g36920
miR172	AP2	At5g60120
miR172	AP2	At5g67180
miR172	AP2	At2g39250
miR172	AP2	At3g54990
miR390/miR391	TAS3	At3g17185
miR390/miR391	TAS3	At5g49615
miR390/miR391	TAS3	At5g57735
miR393	TIR1/AFB	At1g12820
miR393	bHLH	At3g23690
miR393	TIR1/AFB	At3g26810
miR393	TIR1/AFB	At3g62980
miR393	TIR1/AFB	At4g03190
miR394	F-Box	At1g27340
miR395	APS	At3g22890
miR395	AST	At5g10180
miR395	APS	At5g43780
miR395	APS	At4g14680
miR396	GRF	At2g22840
miR396	GRF	At2g36400
miR396	GRF	At2g45480
miR396	GRF	At4g24150
miR396	GRF	At4g37740
miR396	GRF	At5g53660
miR396	GRF	At3g52910
miR397	LAC	At2g29130
miR397	LAC	At2g38080
miR397	LAC	At5g60020
miR398	CSD	At1g08830
miR398	CSD	At2g28190
miR398	CytC oxidase	At3g15640
miR399	E2-UBC	At2g33770
miR399	E2-UBC	At2g33770
miR408	LAC	At2g30210
miR408	PLC	At2g02850
miR827	SPX	At1g02860

Table 9.10 Abundancia de miARN en bibliotecas de ARN pequeños

miRNA	miRNA sequence	Len	Sum	Average	Max	Min (>0)	col_8c_1s	col_8c_2s	col_22c_1s	col_22c_2s	col_37c_1s	col_37c_2s
ath-miR166a-3p	UCGGGACCAGGGCUUCAUCCCC	21	16802485	2800414	3628371	2364322	2984869	2750256	3628371	2629876	2364322	2444791
ath-miR166b-3p	UCGGGACCAGGGUUCAUCCCC	21	16802485	2800414	3628371	2364322	2984869	2750256	3628371	2629876	2364322	2444791
ath-miR166c	UCGGGACCAGGUUCAUCCCC	21	16802485	2800414	3628371	2364322	2984869	2750256	3628371	2629876	2364322	2444791
ath-miR166d	UCGGGACCAGGUUCAUCCCC	21	16802485	2800414	3628371	2364322	2984869	2750256	3628371	2629876	2364322	2444791
ath-miR166e-3p	UCGGGACCAGGUUCAUCCCC	21	16802485	2800414	3628371	2364322	2984869	2750256	3628371	2629876	2364322	2444791
ath-miR166f	UCGGGACCAGGUUCAUCCCC	21	16802485	2800414	3628371	2364322	2984869	2750256	3628371	2629876	2364322	2444791
ath-miR166g	UCGGGACCAGGUUCAUCCCC	21	16802485	2800414	3628371	2364322	2984869	2750256	3628371	2629876	2364322	2444791
ath-miR165a-3p	UCGGGACCAGGUUCAUCCCC	21	3851572	641929	814872	516339	656838	659292	814872	650253	516339	553978
ath-miR165b	UCGGGACCAGGUUCAUCCCC	21	3851572	641929	814872	516339	656838	659292	814872	650253	516339	553978
ath-miR396b-5p	UUCACACAGCUUUCUGAACU	21	750396	125066	162912	104133	145406	104133	121733	108197	162912	108015
ath-miR398b-3p	UGUGUUUCUAGGUACCCUG	21	722178	120363	149952	85791	149952	119184	126261	123989	117001	85791
ath-miR398c-3p	UGUGUUUCUAGGUACCCUG	21	722178	120363	149952	85791	149952	119184	126261	123989	117001	85791
ath-miR158a-3p	UCCCAAAGUAGACAAAGC	20	625066	104178	258737	45714	50958	56936	106138	106583	258737	45714
ath-miR159b-3p	UUUGGAUUGAAGGGAGCUU	21	282578	47096	61555	33890	47188	33890	40857	41733	57355	61555
ath-miR396a-5p	UCCCACAGCUUUCUGAACU	21	206932	34489	55844	21358	21358	37179	31758	55844	25437	35356
ath-miR159a	UUUGGAUUGAAGGGAGCUA	21	172986	28831	64058	19739	22580	19739	22848	20756	64058	23005
ath-miR156a-5p	UGACAGAAAGAGAGUGACAC	20	156140	26023	61447	10868	17217	18243	20258	28107	61447	10868
ath-miR156b-5p	UGACAGAAAGAGAGUGACAC	20	156140	26023	61447	10868	17217	18243	20258	28107	61447	10868
ath-miR156c-5p	UGACAGAAAGAGAGUGACAC	20	156140	26023	61447	10868	17217	18243	20258	28107	61447	10868
ath-miR156d-5p	UGACAGAAAGAGAGUGACAC	20	156140	26023	61447	10868	17217	18243	20258	28107	61447	10868
ath-miR156e	UGACAGAAAGAGAGUGACAC	20	156140	26023	61447	10868	17217	18243	20258	28107	61447	10868
ath-miR156f-5p	UGACAGAAAGAGAGUGACAC	20	156140	26023	61447	10868	17217	18243	20258	28107	61447	10868
ath-miR408-3p	AUGCACUGCCUUCCCCUGGC	21	136445	22741	29269	13654	20649	18436	25584	28853	29269	13654
ath-miR403-3p	UUAGAUUCACGCCAACACU	21	135840	22640	53642	12646	13685	14689	20091	21087	53642	12646
ath-miR157a-5p	UUGACAGAAAGUAGAGAGAC	21	128540	21423	36355	11539	14195	17493	22273	26685	36355	11539
ath-miR157b-5p	UUGACAGAAAGUAGAGAGAC	21	128540	21423	36355	11539	14195	17493	22273	26685	36355	11539
ath-miR157c-5p	UUGACAGAAAGUAGAGAGAC	21	128540	21423	36355	11539	14195	17493	22273	26685	36355	11539
ath-miR161.1	UGAAAAGGACUACAUCCGGG	21	97743	16291	27542	8800	18597	11783	15294	15727	27542	8800
ath-miR162a-3p	UCGAUAACCUCCUGCAUCC	21	69654	11609	24351	4406	5740	8390	10974	15793	24351	4406
ath-miR162b-3p	UCGAUAACCUCCUGCAUCC	21	69654	11609	24351	4406	5740	8390	10974	15793	24351	4406
ath-miR166a-5p	GGACUGUUGUCUGGUAGCG	21	69580	11597	29265	2941	10001	14915	8070	29265	4388	2941
ath-miR166b-5p	GGACUGUUGUCUGGUAGCG	21	69580	11597	29265	2941	10001	14915	8070	29265	4388	2941
ath-miR168a-5p	UCGUUUGGGUGCAGGUCCGG	21	69508	11585	15498	6714	15498	10380	15300	9355	12261	6714
ath-miR168b-5p	UCGUUUGGGUGCAGGUCCGG	21	69508	11585	15498	6714	15498	10380	15300	9355	12261	6714
ath-miR319a	UUGGACUGAAGGGAGCUCCU	21	44828	7471	8487	6440	7360	8400	7146	8487	6440	6995
ath-miR319b	UUGGACUGAAGGGAGCUCCU	21	44828	7471	8487	6440	7360	8400	7146	8487	6440	6995
ath-miR397b	UCAUUGAGUGCAUCUGUAG	21	24336	4056	5339	2134	3996	5339	3533	4328	2134	5006
ath-miR157c-3p	GCUCUCUAACUUCUGUAC	21	23303	3884	6164	2084	3140	2703	3766	6164	5446	2084
ath-miR824-3p	CCUUCUCAUCGAGGUUCA	21	22410	3735	6632	2449	2449	2764	3311	2571	6632	4683
ath-miR390a-5p	AAGCUUCAGGGAGGUAGCG	21	20341	3930	7207	1353	2386	2535	2696	4164	7207	1353
ath-miR390b-5p	AAGCUUCAGGGAGGUAGCG	21	20341	3930	7207	1353	2386	2535	2696	4164	7207	1353
ath-miR157a-3p	GCUCUCUAGCCUUCUGUAC	21	17051	2842	3653	1902	3326	2580	3102	3653	2488	1902
ath-miR157b-3p	GCUCUCUAGCCUUCUGUAC	21	17051	2842	3653	1902	3326	2580	3102	3653	2488	1902
ath-miR163	UUGAAAGGAGACUUGGAACU	24	16378	2730	5294	1594	2456	3328	1600	2106	1594	5294
ath-miR393b-3p	AUCAUGCGAUCUUCUGGAAU	21	11887	1981	3005	892	2235	2332	1644	1779	892	3005
ath-miR170-5p	UAUUGCCUUGGUACUCCAGA	21	10325	1721	2539	1043	1819	2191	1372	2539	1361	1043
ath-miR171a-5p	UAUUGCCUUGGUACUCCAGA	21	10325	1721	2539	1043	1819	2191	1372	2539	1361	1043
ath-miR167d	UGAAGCUGCCAGCAUGAUC	22	8342	1390	1759	1125	1458	1406	1355	1125	1759	1239
ath-miR167a-5p	UGAAGCUGCCAGCAUGAUC	21	7747	1291	2107	672	1535	1292	1218	923	2107	672
ath-miR167b	UGAAGCUGCCAGCAUGAUC	21	7747	1291	2107	672	1535	1292	1218	923	2107	672
ath-miR822-5p	UGCGGGAAAGCAUUUGCACAU	21	7416	1236	2462	772	1422	877	1080	803	2462	772
ath-miR399b	UGCCAAGGGAGAGUUGCCCG	21	7204	1201	1552	904	1552	1197	1481	1019	1051	904
ath-miR399c-3p	UGCCAAGGGAGAGUUGCCCG	21	7204	1201	1552	904	1552	1197	1481	1019	1051	904
ath-miR167a-3p	GAUCAUGUUCCGAGUUUCAC	21	6412	1069	1837	422	1206	1022	1002	1837	923	422
ath-miR169f-3p	GCAAGUUGACCUUGGGUCU	21	5647	941	2396	153	565	904	1358	2396	271	153
ath-miR846-3p	UGAUUAUGAAGUGCUUGAAU	21	5486	914	2459	335	639	957	539	557	335	2459
ath-miR157d	UGACAGAAAGAUAGAGACAC	20	4960	827	2253	373	469	555	373	807	2253	503
ath-miR394a	UUGGCAUUCUGGUCCACCU	20	4414	736	1118	466	631	764	600	1118	835	466

Table 9.11 Continuación: Abundancia de miARN en bibliotecas de ARN pequeños

miRNA	miRNA sequence	Len	Sum	Average	Max	Min (>0)	col_8c_1s	col_8c_2s	col_22c_1s	col_22c_2s	col_37c_1s	col_37c_2s
ath-miR394b-5p	UUGGCAUUCUGGUCCACCUCC	20	4414	736	1118	466	631	764	600	1118	835	466
ath-miR158a-5p	CUUUGUCUACAUUUUUGAAA	21	4352	725	1138	130	664	1099	736	1138	130	585
ath-miR319c	UUGGACUGAAGGGAGGUCCUU	21	4103	684	950	352	616	887	492	806	352	950
ath-miR775	UUCGAUGUCUAGCAGUGCCA	20	4026	671	1026	488	666	488	1026	513	709	624
ath-miR168a-3p	CCCGCCUUGCAUCAACUGAAU	21	3943	657	1044	359	1044	726	586	638	359	590
ath-miR399a	UGCCAAGGAGAUUUGCCUG	21	3912	652	910	344	910	597	727	480	344	854
ath-miR156b-3p	UGCUCACCUCUUUCUGUCAGU	23	3694	616	1446	275	364	336	447	826	1446	275
ath-miR824-5p	UAGACCAUUUGUGAGAACGGA	21	3078	513	1002	260	367	491	260	432	526	1002
ath-miR161.2	UCAAUGCAUUGAAAGUGACUA	21	2874	479	831	273	357	434	522	457	273	831
ath-miR5651	UUGUGCGGUUCAAAUAGUAAC	21	2563	427	735	273	320	451	354	430	273	735
ath-miR156c-3p	GCUCACUGCUCUAUCUGUCAGA	22	2501	417	504	283	460	482	392	504	283	380
ath-miR847	UCACUCCUCUUCUUCUUGAUG	21	2396	399	608	230	334	288	372	564	608	230
ath-miR779.2	UGAUUGGAAAUUCUGUUGACU	21	2309	385	738	126	503	738	226	342	126	374
ath-miR841a-5p	UACGAGGCCACUUGAAACUGAA	21	2239	373	584	147	344	459	284	421	584	147
ath-miR393a-5p	UCCAAAGGGGAUCGCAUUGAUCC	22	2083	347	483	251	288	376	251	483	433	252
ath-miR393b-5p	UCCAAAGGGGAUCGCAUUGAUCC	22	2083	347	483	251	288	376	251	483	433	252
ath-miR165a-5p	GGAAUGUUGUCUGGAUCGAGG	21	1796	299	622	78	622	186	316	406	188	78
ath-miR391-3p	ACGGUAUCUCUCCUACUGAC	21	1783	297	563	163	259	344	163	563	165	289
ath-miR1886.2	UGAGAUGAAAUUCUUUGAUGG	21	1699	283	419	184	184	283	237	419	258	318
ath-miR171b-5p	AGAUAUUAGUGGGGUUCAUC	21	1550	258	373	90	350	373	320	327	90	90
ath-miR173-5p	UUCGCUUGCAGAGAACAUAC	22	1522	254	372	180	186	274	206	372	304	180
ath-miR171a-3p	UGAUUGACCGCGCCAUAUC	21	1488	248	499	111	178	206	168	326	499	111
ath-miR447a.2-3p	UAUGGAAGAAAUAUGUAGUAU	21	1471	245	486	120	198	252	215	200	120	486
ath-miR408-5p	ACAGGGAACAAGCAGAGCAUG	21	1458	243	889	44	258	82	82	103	889	44
ath-miR396a-3p	GUUCAUAAAAGCUGUGGGAG	21	1443	241	449	141	297	150	165	241	449	141
ath-miR827	UUAGAUGACCAUCAACAAACU	21	1386	231	614	118	146	118	203	167	614	138
ath-miR5012	UUUUACUGCUACUUGUGUUC	21	1308	218	336	109	109	176	197	336	291	199
ath-miR400	UAUGAGAGUAUUAAGUCAC	21	1275	213	345	133	133	200	150	345	170	277
ath-miR823	UGGGUGGUGAUCAUAUAGAU	21	1241	207	292	117	265	137	245	117	292	185
ath-miR781a	UUAGAGUUUUUCUGGAUACUA	21	1064	177	300	61	190	243	157	113	61	300
ath-miR781b	UUAGAGUUUUUCUGGAUACUA	21	1064	177	300	61	190	243	157	113	61	300
ath-miR858a	UUUCGUGUGUCGUUCGACCU	21	1014	169	307	88	104	157	129	229	88	307
ath-miR171c-5p	AGAAUUAGGUGCGGUUCAUC	21	1000	167	235	119	140	178	134	235	194	119