Facultad de Ciencias Bioquímicas y Farmacéuticas.
Universidad Nacional de Rosario.

Carrera de Doctorado en Ciencias Biológicas.

Resumen de Tesis

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

Doctorando: Lic. Uciel P. Chorostecki.

Director: Dr. Javier Palatnik.

Tutores: Dr. Esteban Serra y Dr. Lucas Daurelio.

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 de aproximadamente 21, 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. Algunos tienen relación con distintos tipos de Cáncer y otros están relacionados con enfermedades cardíacas donde los niveles de expresión de miARNs específicos cambian en el corazón humano cuando están presentes dichas enfermedades. 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 complementariedad con sus genes blanco y su predicción por lo general se basa en el uso de parámetros empíricos deducidos de interacciones conocidos del par miARN-gen blanco.
En la primer parte del proyecto presentado aquí desarrollamos una estrategia para la identificación de genes blanco regulados por miARNs 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 proyecto 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 *Arabidopsis thaliana*.

Introducción.

Los miARNs son ARN no codificantes que regulan la expresión génica post-transcripcionalmente en animales y plantas, implicados en procesos biológicos muy

variables, como el desarrollo, la diferenciación y el metabolismo (1). Estos pequeños ARNs de ~21 nucleótidos (nts) reconocen secuencias parcialmente complementarias en los ARNm blanco y los guían a su degradación o los inhiben traduccionalmente.

En general, la biogénesis de estos ARN pequeños comienza con la transcripción por la RNA polimerasa II a partir de unidades transcripcionales propias distribuidas en el genoma (2). Los transcriptos primarios, llamados pri-microARNs, 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 y burbuja (stem-loop) (2).

En animales, los precursores de miARNs presentan estructuras de tallo-burbuja homogéneas. En cambio, en plantas se observan estructuras de forma y tamaño muy variable. En todos los casos conocidos, el procesamiento de precursores de miARNs es llevado adelante por ribonucleasas del tipo III, sin embargo existen importantes diferencias en la biogénesis de estos ARN pequeños entre animales y plantas.

En animales, el procesamiento comienza en el núcleo por DROSHA y finaliza en el citoplasma por la acción de DICER. En plantas, los precursores son procesados completamente en el núcleo a través de la acción de una ribonucleasa llamada DCL1 (del inglés DICER LIKE 1) en asociación con el cofactor proteico de unión a ARN de doble hebra HYL1 (del inglés HYPONASTIC LEAVES 1) y la proteína SERRATE. Al parecer es la estructura secundaria por sobre la secuencia primaria del precursor la más importante en la determinación del correcto procesamiento del mismo. 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. Esta especie madura se asocia a un complejo de tipo RISC (del inglés RNAi Silencing Complex) el cual incluye proteínas de la familia Argonauta como AGO1. Complejos RISC similares se encuentran presentes en células 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 señalandolos para su degradación (2). En cualquier caso, es el miARN el que proporciona la especificidad contra las moléculas de ARN blanco (3).

Los miARNs están codificados por familias de genes de 1 a 32 miembros que dan lugar a miARNs maduros idénticos o muy similares. 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 (2), estando la mayoría de ellas conservadas entre mono y dicotiledóneas (4).

Los avances tecnológicos en secuenciación de ADN observados en los últimos 5 años han cambiado notablemente el enfoque de los estudios de expresión génica. En particular, los nuevos métodos de secuenciación de alto rendimiento tales como MPSS (Massively Parallel Signature Sequencing), 454 sequencing (Roche Company) y SBS (Sequencing-By-Synthesis) (16) permiten obtener en detalle cuantiosa información sobre ARN pequeños en plantas. En

particular, la técnica SBS, desarrollado por Illumina®, permite la lectura de más de 2.000.000 de secuencias por corrida. Aunque la longitud de secuencias (~30-35 nucleótidos - nt) en Illumina® es menor que la obtenida por la tecnología 454 (~400 nt), es suficiente para el análisis de ARN pequeños por lo que es la tecnología elegida en estudios de miARNs (5,6).
Esta estrategia está siendo utilizada para identificar nuevos miARNs en distintas especies, ya que miembros de una misma familia de miARNs que varían en un solo nt pueden ser diferenciados con exactitud. Además, la frecuencia de clonado de los distintos ARN pequeños puede reflejar los niveles de expresión de los mismos (7,8).

En paralelo, se ha desarrollado una estrategia denominada PARE (Parallel Analysis of ARN Ends). La misma consiste en una combinación de la estrategia modificada 5´RACE, que permite confirmar si un ARNm es regulado por un determinado miARN, con las nuevas tecnologías de secuenciación. PARE permite la identificación de nuevos ARNm blancos en escala genómica (9). De la misma manera, a través de PARE se puede obtener información sobre los intermediarios de procesamiento de precursores de miARNs (9).

En *Arabidopsis thaliana* se conocen alrededor 200 miARNs (miRBase 15.0). Muchos de estos pequeños ARNs han aparecido recientemente en la evolución y no está claro si tienen algún rol biológico (10). Sin embargo, existen 21 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 (11). Estos últimos miARNs cumplen funciones esenciales para la biología de las plantas (2).

La identificación de los genes blanco regulados por miARNs es uno de los temas más importantes en el campo. En plantas, las estrategias genómicas más utilizadas han sido basadas en el estudio del degradoma, a través de la técnica PARE. Mediante esta estrategia, es posible identificar los fragmentos en los que se degradan los ARNm de la célula, lo que permite encontrar aquellos que son degradados mediante miARNs (14). En cambio se ha explorado muy poco la búsqueda de targets de miARNs mediante la conservación de las secuencias blanco. Herramientas de este tipo, han sido mucho más desarrolladas para encontrar genes regulados en animales (12,14). Uno de los aspectos propuestos en este proyecto es justamente analizar la posibilidad de encontrar nuevos genes regulados utilizando herramientas bioinformáticas basadas en la conservación de los sitios blancos en plantas, al menos para las 21 familias de miARNs altamente conservadas.

A diferencia de los estudios realizados en animales, poco se sabe sobre la regulación del procesamiento de miARNs en plantas. Resultados publicados indican que ciertos precursores de estos ARNs pequeños dependen preferencialmente de ciertos componentes de la maquinaria de procesamiento. De esta manera, mutaciones en HYL1 o SERRATE afectan los niveles de miARNs en forma disímil, mientras que la sobreexpresión de HYL1 puede aumentar solo los niveles de algunos miARNs. Estos resultados están de acuerdo con la marcada heterogeneidad que presentan en sus formas y tamaños los precursores de plantas, que contrastan con la homogeneidad observada en los precursores de animales. El análisis de los perfiles de expresión de los genes involucrados en el procesamiento de miARNs en microarreglos de acceso público muestra que los mismos no están co-regulados, habiendo ciertos tejidos donde se favorece la expresión de uno u otro componente. Estos resultados sugerirían que la composición de la maquinaria de procesamiento podría ser flexible y

depender de las células donde se analice. Finalmente, toda esta evidencia indicaría que la expresión de los miARNs maduros en la planta podría estar regulada no solo a nivel transcripcional, sino también a nivel post-transcripcional mediante la regulación del procesamiento de los precursores.

En este proyecto se propone también analizar la composición de ARN pequeños en distintos tejidos y en plantas mutantes en distintos genes de la maquinaria de procesamiento, mediante técnicas de secuenciación de alto rendimiento. Los resultados permitirán generar un conocimiento general e integrado de la regulación del procesamiento de los miARNs en las plantas. Además, el cambio climático es uno de los temas más importantes para la ciencia en este momento. Las plantas no regulan la temperatura, por lo que los cambios ambientales influyen sobre su crecimiento y desarrollo y muchos procesos biológicos se ven afectados. Es por esto que se estudiará plantas crecidas a distintas temperaturas.

Dado que la manipulación de los niveles de expresión de miARNs es un elemento clave para el desarrollo de aplicaciones biotecnológicas en plantas, las aplicaciones potenciales de estos estudios podrían ser inmediatas.

Objetivos

Objetivo general

El objetivo general propuesto pretende identificar genes blanco regulados por miARNs en plantas mediante un enfoque bioinformático basado en la conservación evolutiva del par miARN-gen blanco. Además pretende entender el preciso mecanismo por el cual el precursor de miARN es reconocido y procesado por la maquinaria de procesamiento en plantas.

Objetivo específicos

- Identificar genes blanco regulados por miARNs.
- Identificación de genes blanco específicos de un grupo determinado de especies de plantas.
- Crear una herramienta online para la predicción de genes blanco regulados por miARNs en plantas.
- Realizar un análisis sistemático para obtener pautas generales sobre la biogénesis de miARNs plantas.
- Determinar la dirección de procesamiento de miARNs en *Arabidopsis thaliana.*

Resultados I)

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 (17,18). 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 (18).

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 familia de miARNs en esas tres especies.

Diseñamos una estrategia para identificar nuevos pares miARN-gen blanco principalmente basada en la conservación evolutiva de la secuencia del gen blanco (Figura 1). 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[1]" 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*[2] y *Oryza Sativa*[3]. Utilizando las secuencias de 18nt y permitiendo 3 mismatches (errores), la búsqueda de genes blanco dio como resultado 38.597 genes distribuídos en las 43 especies (Figura 1, bin 1).

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.

La estrategia permite la selección de los mejores candidatos basándose en la presencia de los genes blanco en un número de especies. Utilizando 4 especies (con una buena especificidad) dio como resultado 3.781 genes que corresponden a 533 tags diferentes (Figura 1, 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. De los 38.597 candidatos iniciales, 9.375 pasan estos filtros (Figura 1, bin 4). Combinando filtros de energía y filtro de conservación, la búsqueda arrojó como resultado 563 candidatos correspondientes a 146 tags (Figura 1, bin 5).

[1] http://compbio.dfci.harvard.edu/tgi/
[2] http://arabidopsis.org/
[3] http://rice.plantbiology.msu.edu/

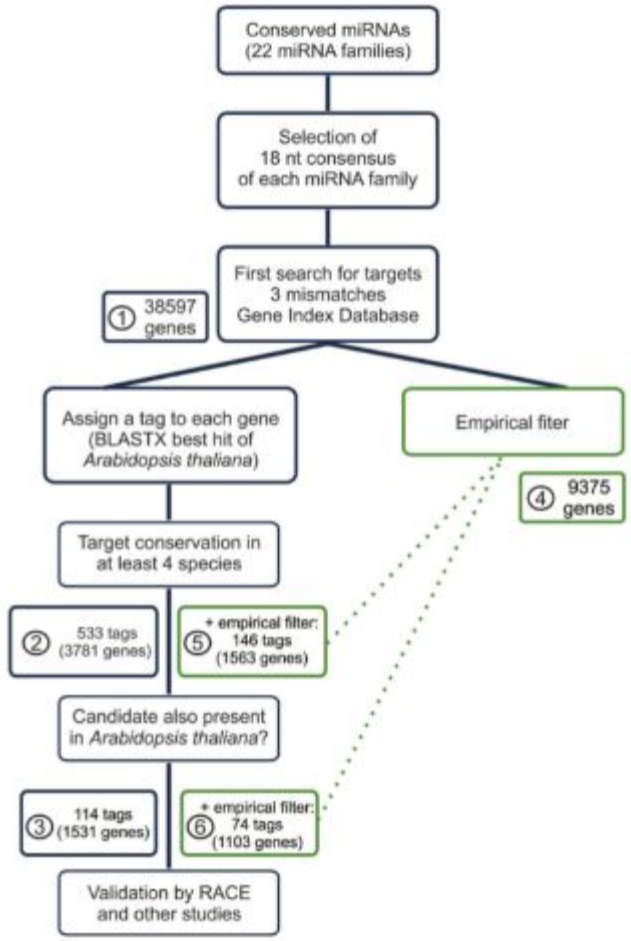


Figura 1: Esquema de la estrategia para la identificación de nuevos genes blanco. El número de genes blanco está identificado en cada paso. Luego de aplicar el análisis de conservación, todos los genes que tienen el mismo hit en Arabidopsis, fueron considerados como un solo gen blanco. El recuadro verde indica 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.

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. [e.g.,(21,22,23)]. En primer lugar, analizamos la detección de genes blanco validados previamente en *A. thaliana* [basado en (24)] usando nuestra estrategia y encontramos que el 84% de ellos estaban presentes en al menos 4 especies (Figura 3B). Consideramos esto como un buen resultado ya puede ser que no todos los genes blanco de *Arabidopsis* estén conservados evolutivamente.

Nos enfocamos únicamente en los potenciales genes blanco conservados en 4 especies, donde una de ellas es *A. thaliana* (Figura 1, bin 1). De esta manera identificamos 114 potenciales genes que satisfacen este criterio. Donde 76 de ellos son genes validados anteriormente o genes muy relacionados (Figura 3C). 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 2) e intentamos validarlos utilizando 5’ RACE PCR modificada (25,26).

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 (27) (Figura 4A). La validación experimental reveló fragmentos de ARNm compatible con este último sitio de corte (Figura 4A).

Otro candidato era At3g22110 que codifica para PROTEASOME ALPHA SUBUNIT C1 (PAC1) presente en 20 especies. Por medio de 5’ RACE PCR demostramos que este gen es gen blanco del miR408 (Figura 4A). Curiosamente la interacción del par miARN-gen blanco tiene 3 mismatches en la región 5’ que 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.

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 4C 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. El otro potencial gen blanco era At4g27330, conocido como NOZZLE/SPOROCYTLESS. Este factor de trascripción, que participa en desarrollo del estambre y óvulo, fue también validado por 5’ RACE PCR (Figura 4E). 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 4A-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. De esta manera el sitio de NOZZLE, donde cambia la secuencia en diferentes especies (Figura 4E), pudo ser detectado por este enfoque.

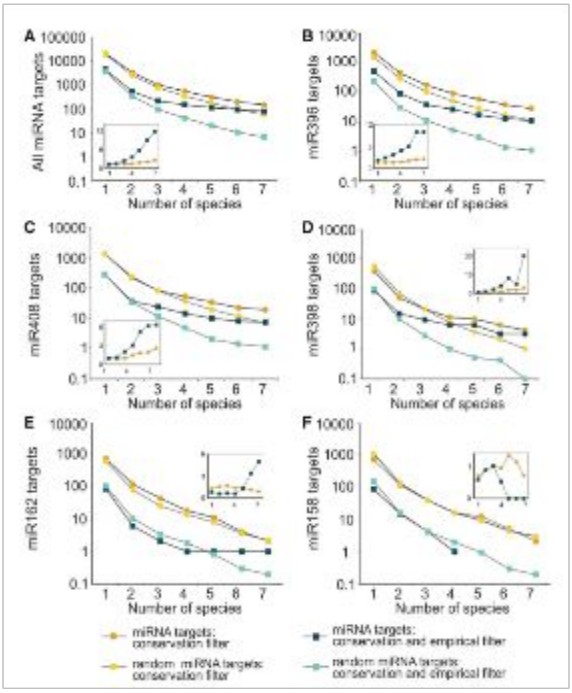


Figura 2. Conservación de potenciales genes blanco en distintas especies. Todos los miARNs **(A),** miR396 **(B),** miR408 **(C),** miR398 **(D),** miR162 **(E),** miR158 **(F)**. Puntos naranja representan los genes blanco de miARNs usando filtro evolutivo. Puntos amarillos representan los genes blanco 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.

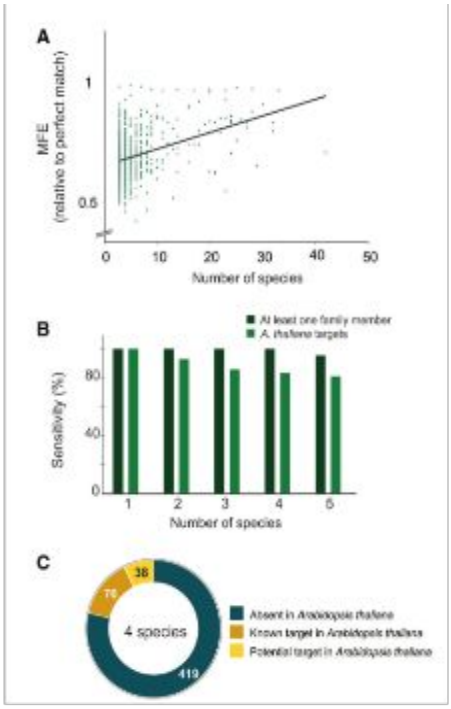


Figura 3. Selección de genes blanco por conservación evolutiva de la secuencia. **(A)** Relación entre MFE y el número de especies en donde cada gen blanco fue detectado. **(B)** Sensibilidad de la estrategia, analizado de dos modos distinto. Verde clarito: evaluando la presencia de genes validados en *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.

Identificación de nuevos genes blanco permitiendo interacciones G-U.

Los genes blanco identificados en este trabajo 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 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 G-U.

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 5A). Encontramos 125 potenciales genes blanco en *A. thaliana* teniendo en cuenta este criterio (Figura 5A) 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 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 5B y C), que está involucrado en el control de niveles libre de auxina.

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 5B y C). La técnica de 5' RACE PCR confirmó que el gen realmente era gen blanco del miR167 (Figura 5C).

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 6 especies *Solanaceae*) fueron aplicados (Figura 6A),. Curiosamente, al aplicar ambos filtros dio como resultado una relación señal/ruido por encima de 6 (Figura 6A), 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 blanco del miR398, mientras que MT2B, homólogo de este gen, fue detectado presente en 5 especies (Figura 6B-D).

Aprovechamos las plantas transgénicas de tabaco que contienen un transgén *35S.mir398* (A.F., N. y J.F. 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. Curiosamente, observamos que tanto MT2A como MT2B disminuyeron sus niveles de transcripción >5 veces en estas plantas (Figura 6E). Estos resultados concuerdan con la regulación de MT2A y MT2B por el miR398, aunque no necesariamente demuestra una interacción directa. Estos resultados demuestran que los genes blanco presentes en un grupo específico de especies pueden ser encontrados utilizando esta estrategia.

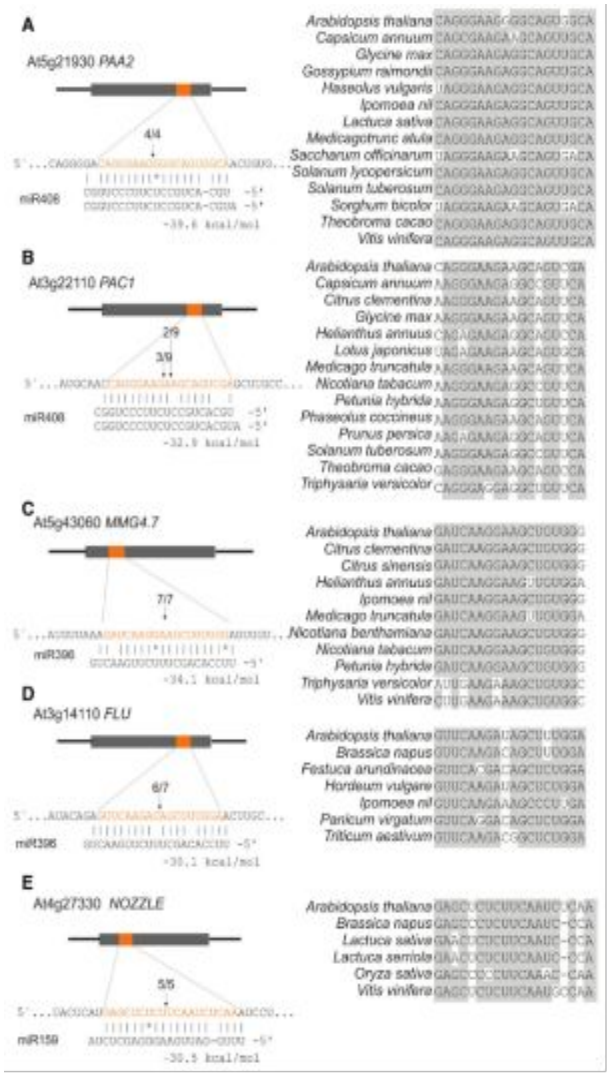
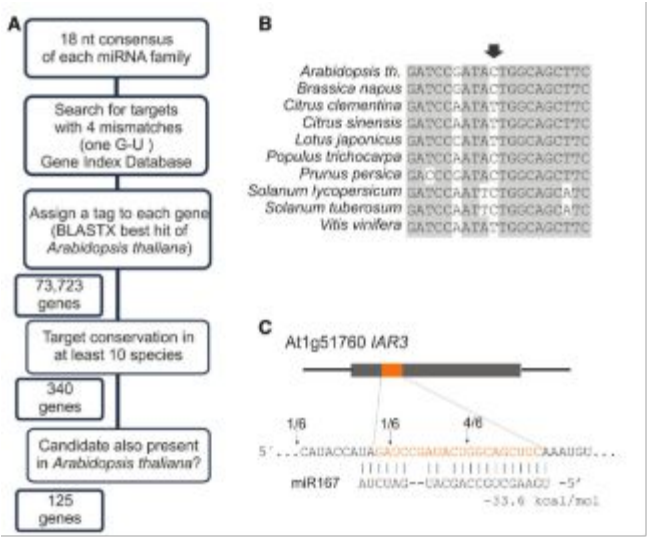


Figura 4. Nuevos genes blanco validados en Arabidopsis. thaliana. El alineamiento entre el miARN y los nuevos genes blanco identificados se muestran en la izquierda. La conservación evolutiva de la secuencia del sitio reconocido por el miARN en las especies seleccionadas se muestra a la derecha. La figura 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 clonadao de cada fragmento .



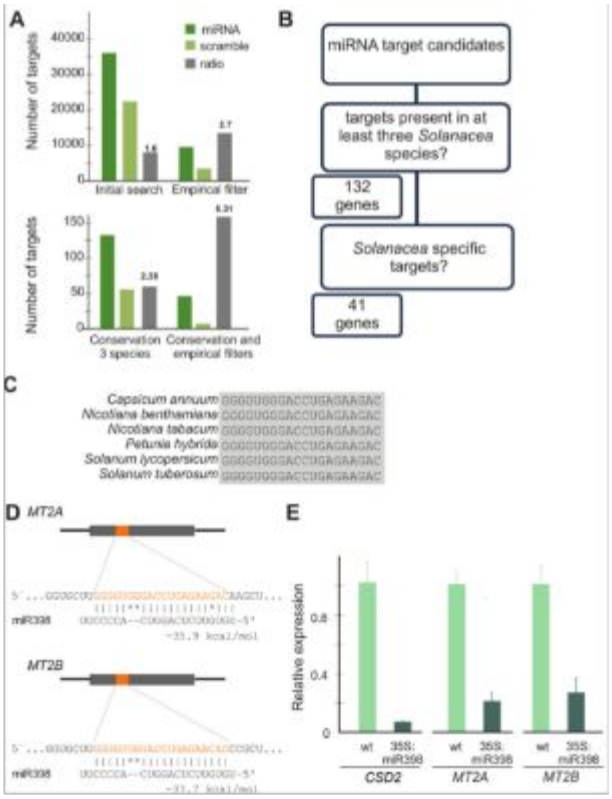


Figura 5. Identificación de un nuevo gen blanco de miARN, relajando los parámetros de interacción pero incrementando el parámetro de conservación evolutiva. **(A)** Esquema de la estrategia modificada para identificar genes blanco de miARNs. **(B)** Conservación del sitio blanco reconocido por el miARN en distintas especies. La flecha indica una variación de G-C o G-U con el miARN dependiendo de la especie. **(C)** Alineamiento en Arabidopsis 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.

Figura 6. Identificación de genes blanco de miARN, específicos de Solanaceae. **(A)** Predicción de genes blanco de miARN en 5 espeices solanaceas. 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.

Discusión I)

En esta primera parte realizamos diferentes análisis bioinformáticos para la predicción de genes blancos regulados por miARNs en plantas. Mediante un enfoque basado en la conservación evolutiva del par miARN-gen blanco pudimos encontrar y validar experimentalmente nuevos genes blanco en *Arabidopsis thaliana*. El gen PAA2 estaba presente en 22 especies distintas incluido monocotiledóneas y dicotiledóneas y fue validado como gen blanco del miR408, al igual que PAC1 presente en 20 especies. Luego por el mismo método pudimos detectar y validar genes blanco regulados por el miR396 que son FLU y MMG4.7. En contraste con el miR408 y miR396, donde tienen varios potenciales genes blanco, obtuvimos un bajo número de potencial genes blanco para el miR159 y uno de ellos conocido como NOZZLE fue posible validarlo experimentalmente. Además notamos 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.

Luego realizamos la búsqueda de genes blanco pero permitiendo interacciones no canónicas del tipo G-U. De esta manera encontramos que el miR167 regula a un gen denominado IAR3 que está involucrado en el control de niveles libre de auxina. 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.

Por último utilizamos la estrategia para encontrar genes blanco presentes específicos de la familia de *Solanaceae*. El gen blanco más común fue la metalotioneína MT2A, presente en las 6 *Solanaceae,* como potencial gen blanco del miR398. Encontramos que MT2A y MT2B, homólogo de este gen, disminuyen sus niveles de transcripción al igual que el gen blanco conocido del miR398, CSD2 sugiriendo la regulación de MT2A y MT2B por el miR398.

Resultados II)

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

A partir de la estrategia descripta anteriormente, que fue utilizada para encontrar y validar experimentalmente genes blanco regulados por miARNs en *Arabidopsis thaliana* (35), desarrollamos una herramienta web denominada comTAR (29) (rnabiology.ibrconicet.gov.ar/comtar) 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.

MiARN y transcriptos.

Como las secuencias del maduro del miARN puede variar en distintas especies, especialmente en la posición 1, 20 y 21 (Chorostecki et al., 2012), 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 (secuencias consenso de miARN). El usuario además puede realizar las búsqueda de nuevos ARNs pequeños teniendo en cuenta esta consideración. Los datos correspondiente a secuencias de transcriptos de plantas fueron obtenidos de bibliotecas del proyecto Phytozome[4] formados por archivos de nucleótidos en formato FASTA de transcriptos de ARNm (UTR, exones) con variantes de splicing.

Búsqueda de genes blanco.

Para la búsqueda de secuencias utilizamos una herramienta libre llamada Patmatch (15) e integramos otras herramientas y scrpits hechos por nosotros para potenciar la herramienta y hacerla más específica para la búsqueda de genes blanco de miARNs conservados en plantas. Brevemente,

- Para el alineamiento del miARN y el gen blanco, implementamos en Perl una versión modificada del algoritmo de Needleman-Wunsch (30).
- Integramos la herramienta RNAHybrid (31), mediante scripts para encontrar la menor energía de hibridación del duplex miARN-gen blanco para cada candidato.

[4] http://phytozome.jgi.doe.gov/pz/portal.html

- Las secuencias candidatas fueron etiquetadas con el mejor hit del locus ID del Arabidosis TAIR10, utilizando los archivos de anotación de Phytozome, y lo utilizamos como “TAG” (etiqueta). De esta manera agrupamos genes de diferente especies que tienen el mismo TAG.
- Cada TAG de *Arabidopsis* fue indexado con una breve descripción funcional y computacional obtenida del TAIR10. Además los genes blanco candidatos fueron ordenados por familias teniendo en cuanta la clasificación de familias del TAIR10.

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 y los datos que surgen de ese análisis fueron almacenados en una base de datos en MySQL[5] . El back-end es el encargado de realizar la búsqueda de secuencias, además ahí es donde se integraron las herramientas y scripts para aumentar la especificidad y sensibilidad de la herramienta y es el encargado de generar los resultados finales. Mientras el front-end es el responsable de mostrar los resultados (Figura 7A).

El TAG del mejor hit en Arabidopsis es el que determina el número de especies donde un hit está presente, y el mínimo número de especie es un parámetro que es definido por el usuario. El usuario puede realizar la búsqueda de genes blanco de nuevos miARNs y además tiene acceso por pantalla del alineamiento del miARN-gen blanco, la energía de hibridación y los filtros empíricos de interacciones conocidas del par miARN-gen blanco (Figura 7B). 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. Además los usuarios pueden poner un locus TAG en particular (tanto de *Arabidopsis* como el 'gene ID' del Phytozome) y comTAR identifica las especies donde este gen en particular puede ser un potencial gen blanco de algún miARN.

[5] https://www.mysql.com/

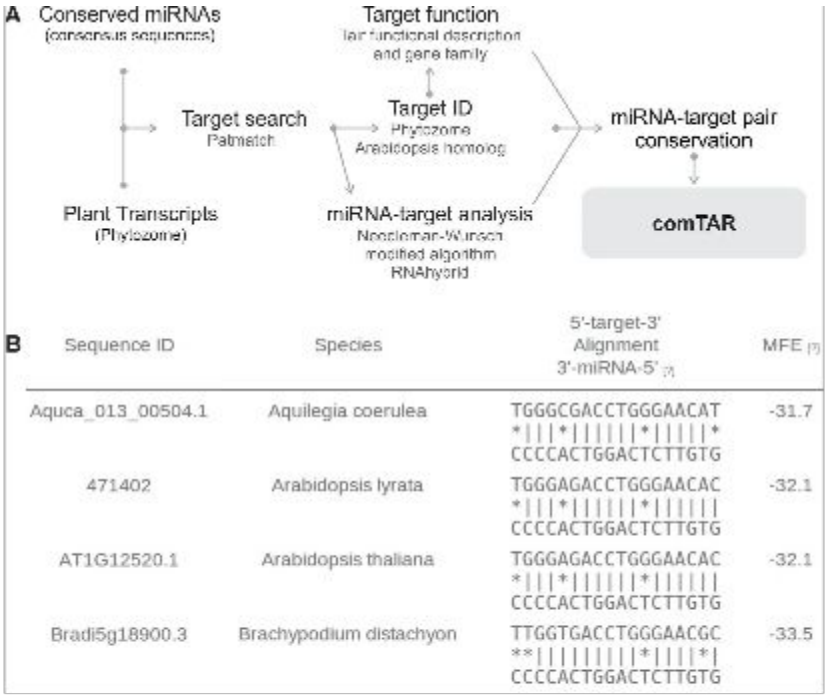


Figura 7: comTAR **(A)** Diagrama de flujo que describe la herramienta. **(B)** Salida de comTAR mostrando el par miR398/SOD1 (At1g12520) en diferentes especies.

Discusión II)

En esta etapa del proyecto desarrollamos comTAR, que permite la caracterización de interacciones entre miARN-gen blanco en distintas especies de plantas. Una ventaja de esta herramienta es que la interacciones par miARN-gen blanco conservadas probablemente participen en procesos biológicos relevantes para la planta. Este enfoque requiere que el reconocimiento del gen blanco ocurra en el contexto de un conjunto mínimo de parámetros que interactúan en distintas especies.

Resultados III)

Estudios genómicos sobre la biogénesis de miARN en plantas.

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. En esta parte del proyecto y en el marco de la colaboración con el grupo del Dr. Blake Meyers[6] (Delaware,USA), el cual se especializa en secuenciación y análisis de ARN pequeños, nos propusimos entender cómo se procesan los precursores de miARNs plantas.

[6] http://deedee.dbi.udel.edu/meyers_lab

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 Arabidopsis 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 (32) (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 miRNAs 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 (Bologna et al., 2013).

Análisis de datos y precursores detectados.

Mediante la cantidad de cortes detectados la técnica de SPARE permite definir si el mecanismo es “base to-loop” o “loop-to-base” y esta técnica arroja una gran cantidad de datos producto de la secuenciación de alto rendimiento. 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. Además desarrollamos una herramienta web para visualizar estos datos utilizando MySQL como base de datos para almacenar los datos procesados (Figura 8).

Un precursor fue considerado como “detectado” si más de tres lecturas corresponden a la secuencia de ese precursor. De esta manera encontramos fragmentos de RNA que corresponden a 129 precursores, 71 de ellos de miARNs conservados y 58 de miARNs jóvenes (Figura 9). Y con la ayuda de la herramienta pudimos definir la dirección de procesamiento de los cuales 32 de ellos fueron definidos como base-to-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 el caso del miR168a, miR172b y el miR395b (Figura 10). Encontramos 16 precursores de miARNs conservados con cortes detectados (>5%) en el lado distal del miARN/miARN* los cuales fueron definidos como loop-to-base (Figura 11).

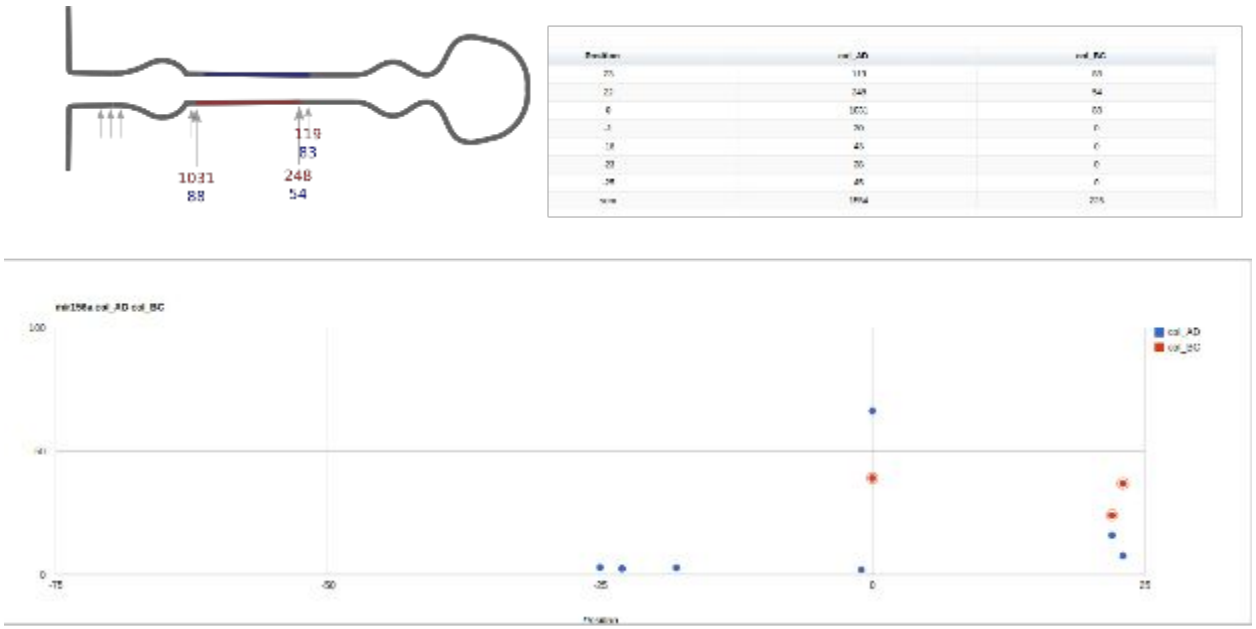


Figura 8: Herramienta para analizar los datos de la técnica SPARE. Porcentaje de cortes del miR156a (abundancia relativa de los cortes en esa posición dividido la abundancia total de los cortes en el precursor). La posición final del duplex miARN/miARN* fue considerada como la posición 0.

Estructura secundaria de los precursores.

Determinamos la estructura secundaria de precursores detectados que se procesan en dirección baseto-loop (Figura 12) y los que se procesan loop-to-base (Figura 13). Obtuvimos las estructuras secundarias para cada precursor calculada a partir de la herramienta mfold (33) con los parámetros por default a 37°C de temperatura. Definimos a un match 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 12 y 13).

Base-to-loop.

Consideramos la estructura secundaria de 32 precursores analizados en esta parte del trabajo que se procesan base-to-loop y todos ellos tienen un claro 'lower stem' de 15 nt (Figura 12). Además este stem se pudo ver tanto para los precursores validados experimentalmente que se procesan base-to-loop como para todos los precursores conservados (Figura 12 en violeta). Pero pudimos observar que las bases inmediatamente debajo del duplex miARN/miARN* (posición -2 y -1) tienden a estar desapareadas (Figura 12). Además las posiciones -3 y -4 y las 3 últimas posiciones del lower stem (-13,-14 y -15) están apareadas casi siempre (Figura 12). En general, nuestros resultados muestran que los precursores procesados en una dirección base-to-loop son más uniformes de lo que se pensaba previamente y que al menos algunos de los precursores no detectados como base-to-loop probablemente tengan otros determinantes específicos de ARN.

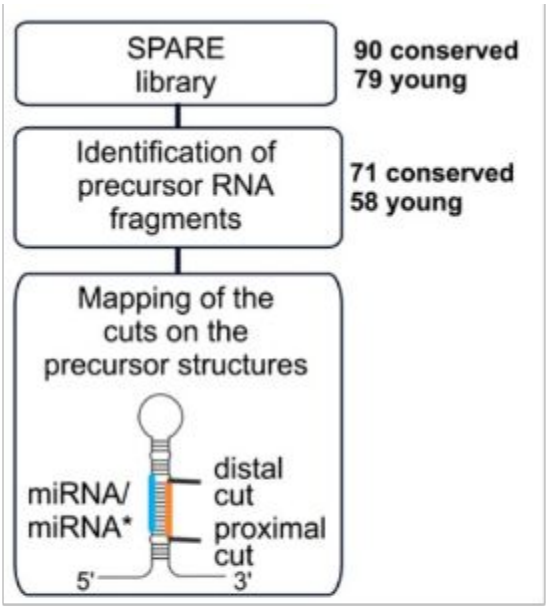


Figura 9: Esquema del procedimiento para analizar los datos de sPARE.

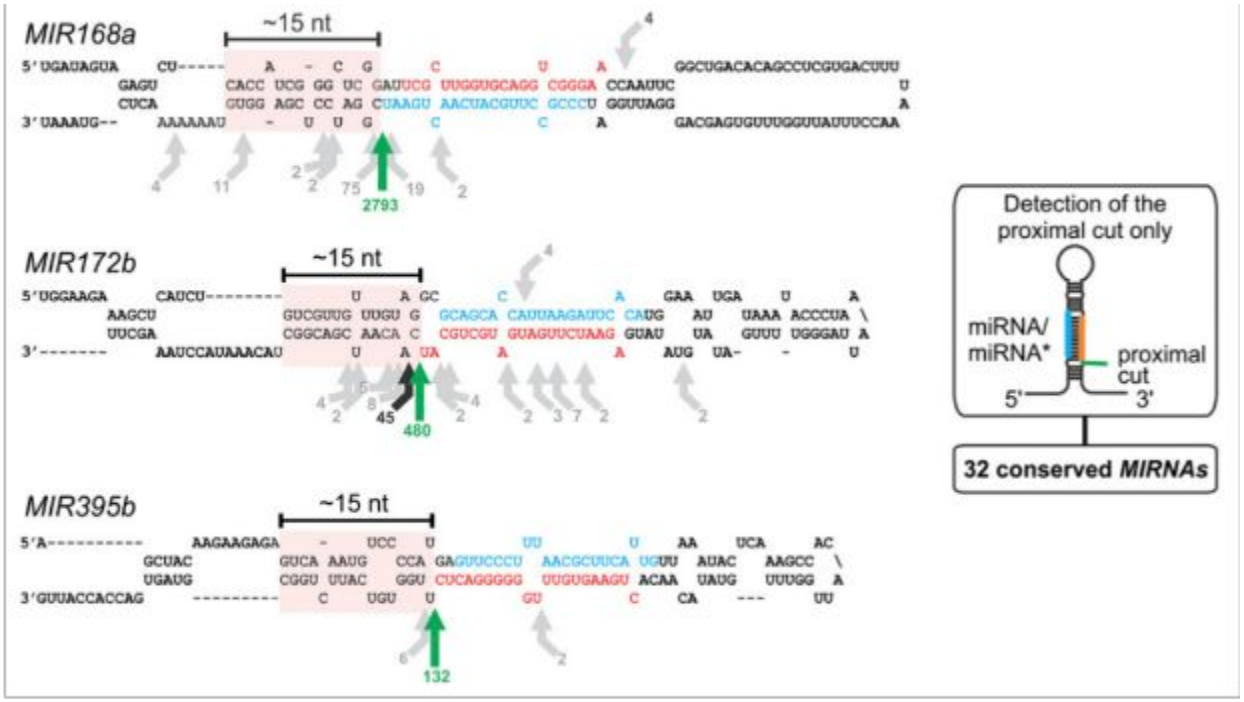


Figura 10: Identificación y caracterización de precursores de miARNs procesados "base-to-loop". Esquema donde se muestra la estructura secundaria del miR168a, miR172b y miR395b. Las flechas indican la posición y número de lecturas de los cortes del precursor identificado. Flechas en verde muestra el 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. El recuadro de la derecha muestra el patrón de corte típico detectado en la biblioteca de SPARE para estos precursores.

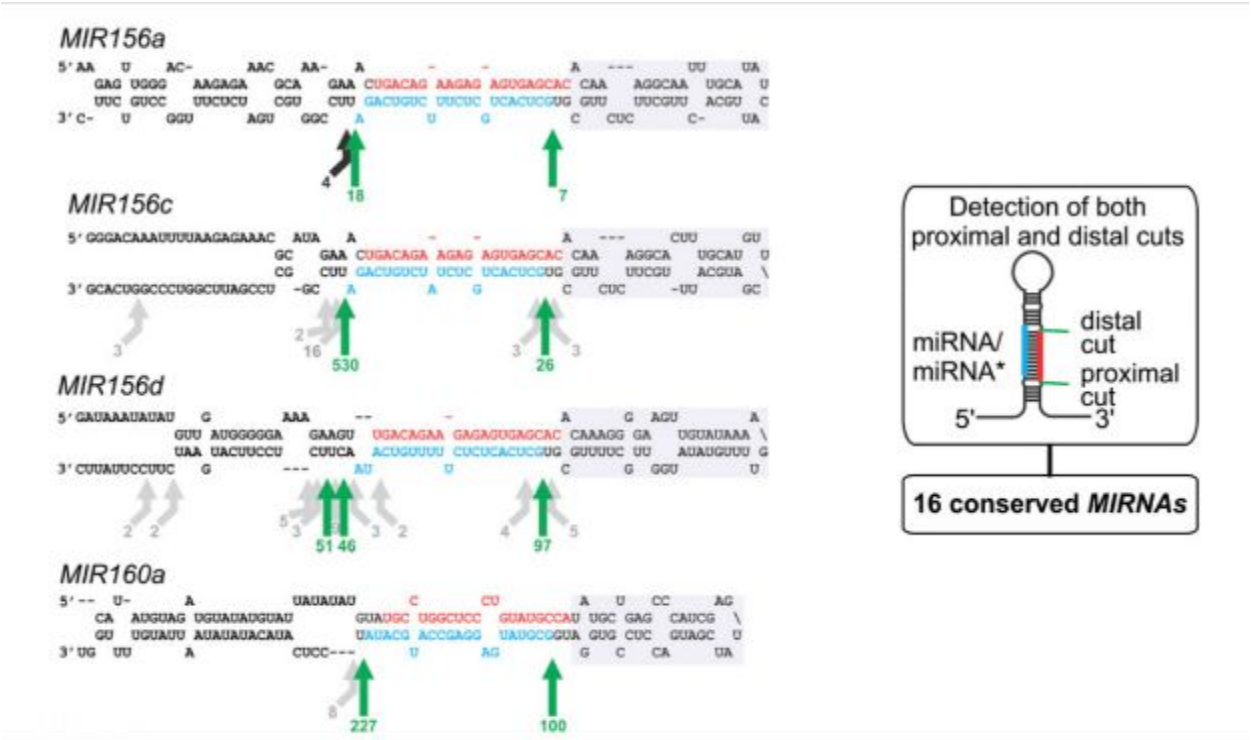


Figura 11: Identificación y caracterización de precursores de miARNs procesados "loop-to-base". Esquema donde se muestra la estructura secundaria del miR156a, miR156c, miR156d y miR160a. Las flechas indican la posición y número de lecturas de los cortes del precursor identificado. Flechas en verde muestra 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. El recuadro de la derecha muestra el patrón de corte típico detectado en la biblioteca de SPARE para estos precursores.

Loop-to-base.

Estos precursores que tienen un procesamiento loop-to-base tienen un corte mayoritario que se puede detectar en nuestras bibliotecas, que es el esperado en la dirección de procesamiento en un mecanismo loopto-base. Con la excepción del dos miARNs (MIR396a y MIR162b) estos precursores no tienen una estructura obvia debajo del duplex miARN/miARN* (Figura 13). Estos precursores tienen una región terminal estructurada (Figura 13), que tiene un tamaño homogéneo de ~42nt que incluye un loop corto en contraste con la misma región en los precursores que se procesan base-to-loop donde es más variable.

Discusión III)

En esta segunda parte del proyecto 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 precursores de miARNs en *Arabidopsis thaliana*.

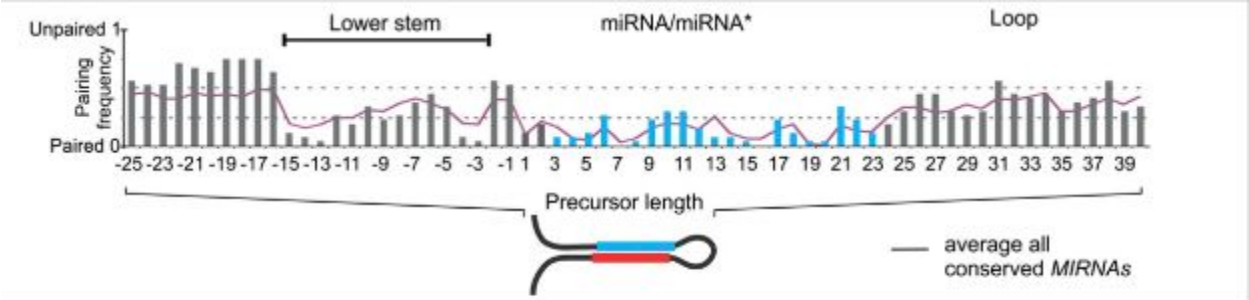


Figura 12: Estructura secuendaria de precursores detectados que se procesan en dirección base-to-loop. Los matches en cada posición los consideramos como 0, mientras que bulges y mismatches fueron considerados como 1. La estructura secundaria considerando todos los miARNs conservados se indica con color violeta.

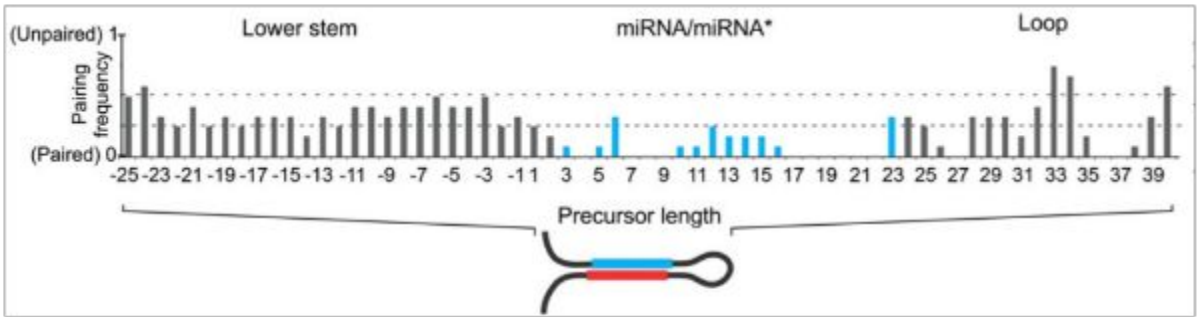


Figura 13: Estructura secuendaria de precursores detectados que se procesan en dirección loop-to-base. Los matches en cada posición los consideramos como 0, mientras que bulges y mismatches fueron considerados como 1.

Conclusiones

A partir de diferentes análisis y experimentos realizados durante mi trabajo de tesis obtuvimos las siguientes conclusiones acerca de los genes blanco de miARNs en plantas y sobre el procesamiento de los mismos:

- Pudimos encontrar y validar experimentalmente nuevos genes blanco en *Arabidopsis thaliana*. PAA2 y PAC1 como genes blanco del miR408, FLU y MMG4.7 como genes blanco del miR396 y por último NOZZLE como gen blanco del miR159.
- Validamos a IAR3 como gen blanco del miR167 realizando la búsqueda de forma más relajada permitiendo interacciones no canónicas.
- Utilizamos la estrategia para encontrar genes blanco presentes específicos de la familia de *Solanaceae*.
- Hemos desarrollado una herramienta online denominada comTAR que permite la caracterización de interacciones entre miARN-gen blanco en distintas especies de plantas.
- Construimos e implementamos un pipeline bioinformático utilizando para poder analizar los datos de las bibliotecas obtenidos a partir de la técnica de SPARE. Además desarrollamos una herramienta web para visualizar estos datos.

- Con la ayuda de dicha herramiento identificamos la dirección de procesamiento de precursores de miARNs en *Arabidopsis thaliana*.

Bibliografía.

1) 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: 257-263.
2) Jones-Rhoades, M. W.; Bartel, D. P.; and Bartel, B. (2006). MicroRNAs and their regulatory roles in plants. Annu. Rev. Plant Biol. 57: 19-53.
3) Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism and function. Cell 116: 281-297.
4) Griffiths-Jones,S., Bateman,A., Marshall,M., Khanna,A., and Eddy,S.R. (2003). Rfam: an RNA family database. Nucleic Acids Res. 31, 439-441.
5) Meyers BC, Ten SS, Vu TH, Haudenschild CD, Agrawal V, Edberg SB, Ghazal H, Decola S (2004) The use of MPSS for whole-genome transcriptional analysis in Arabidopsis. Genome Res 14:1641–1653
6) Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Law TF, Grant SR, Dangl JL, Carrington JC (2007) High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. PLoS ONE 2:e219
7) Lu C, Kulkarni K, Souret FF, MuthuValliappan R, Ten SS, Poethig RS, Henderson IR, Jacobsen SE, Wang W, Green PJ, Meyers BC (2006) MicroRNAs and other small RNAs enriched in the *Arabidopsis* RNA-dependent RNA polymerase- 2 mutant. Genome Res 16:1276–1288
8) Zhu QH, Spriggs A, Matthew L, Fan L, Kennedy G, Gubler F, Helliwell C (2008) A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. Genome Res 18:1456–1465
9) German MA, Pillay M, Jeong DH, Hetawal A, Lou S, Janardhanan P, Kannan V, Rymarquis LA, Nobuta K, German R, De Paoli E, Lu C, Schroth G, Meyers BC, Green PJ (2008) Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends.Nat Biotechnol 26:941–946
10) Axtell, M.J., and Bowman, J.L. (2008). Evolution of plant microRNAs and their targets. Trends Plant Sci. 13: 343–349.
11) Axtell, M.J. (2008). Evolution of microRNAs and their targets: Are all microRNAs biologically relevant? Biochim. Biophys. Acta. 1779, 725-734.
12) Bartel, D.P. (2009). MicroRNAs: Target recognition and regulatory function. Cell. 136 :215-233
14) Addo-Quaye C, Eshoo TW, Bartel DP, Axtell MJ. (2008). Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. Curr Biol. 20;18(10):758-62. Epub 2008 May 8.
15) 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.

16) Chen, Y.-J.; Roller, E.E.; Huang, X. DNA Sequencing by Denaturation: Experimental proof of concept with an integrated fluidic device. Lab on a Chip (2010) May 7;10(9):1153-9. Epub 2010 Feb 9.
17) Cuperus, J.T., Fahlgren, N. and Carrington, J.C. (2011) Evolution and functional diversification of MIRNA genes. Plant Cell, 23, 431-442.
18) 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, e1002419.
19) Xie, Z., Kasschau, K.D. and Carrington, J.C. (2003) Negative feedback regulation of Dicer-Like1 in Arabidopsis by microRNA-guided mRNA degradation. Curr Biol, 13, 784-789.
20) Vaucheret, H., Vazquez, F., Crete, P. and Bartel, D.P. (2004) The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. Genes Dev, 18, 1187-1197.
21) Allen, E., Xie, Z., Gustafson, A.M. and Carrington, J.C. (2005) microRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell, 121, 207-221.
22) Jones-Rhoades, M.W. and Bartel, D.P. (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol Cell, 14, 787-799.
23) Schwab, R., Palatnik, J.F., Riester, M., Schommer, C., Schmid, M. and Weigel, D. (2005) Specific effects of microRNAs on the plant transcriptome. Dev Cell, 8, 517-527.
24) 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. et al. (2010) MicroRNA gene evolution in Arabidopsis lyrata and Arabidopsis thaliana. Plant Cell, 22, 1074-1089.
25) 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, 2053-2056.
26) 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 unction. Dev Cell, 4, 205-217.
27) Maunoury, N. and Vaucheret, H. (2011) AGO1 and AGO2 act redundantly in miR408-mediated Plantacyanin regulation. PLoS ONE, 6, e28729.
28) Millar, A.A. and Gubler, F. (2005) The Arabidopsis GAMYB-like genes, MYB33 and MYB65, are microRNA-regulated genes that redundantly facilitate anther development. Plant Cell, 17, 705-721.
29) Chorostecki,U and Palatnik, J. (2014) comTAR: a web tool for the prediction and characterization of conserved microRNA targets in plants. Bioinformatics. 2014 Apr 12. [Epub ahead of print]
30) 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
31) Kruger, J. and Rehmsmeier, M. (2006). Rnahybrid: microrna target prediction easy, fast and flexible. Nucleic Acids Research, 34(suppl 2), W451–W454.
32) Construction of Specific Parallel Amplification of RNA Ends (SPARE) libraries for the systematic identification of plant microRNA processing intermediates. Schapire AL, Bologna

NG, Moro B, Zhai J, Meyers BC, Palatnik JF. Methods. 2014 May 1;67(1):36-44. doi: 10.1016/j.ymeth.2014.04.001. Epub 2014 Apr 13.
33) Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31: 3406–3415.
34) Multiple RNA recognition patterns during microRNA biogenesis in plants. 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.
35) Identification of new microRNA-regulated genes by conserved targeting in plant species. 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.