

¿CÓMO IDENTIFICO MICRORNAs EN MI PLANTA?*

Mayra Liliana López-Valle¹, Svetlana Shishkova^{1,2}

¹Departamento de Biología Molecular de Plantas, Instituto de Biotecnología, Universidad Nacional Autónoma de México. Cuernavaca, Morelos, México

²Autor de correspondencia. Correo electrónico: svetlana.shishkova@mail.ibt.unam.mx

RESUMEN

Los microRNAs (miRNAs) son RNAs pequeños no codificantes de 20-24 nucleótidos que regulan la expresión génica mediante silenciamiento. La identificación *de novo* de miRNAs se basa en las particularidades de su biogénesis. A grandes rasgos, los criterios considerados son: la identificación de un transcrito que se pliega en una estructura secundaria de tallo-asa, así como las secuencias del miRNA y del miRNA* encontradas en los resultados de secuenciación de RNAs pequeños. Estas secuencias deben ser casi perfectamente complementarias, mapearse sobre el tallo de dicha estructura secundaria con un desfase de 2 nucleótidos y representar la mayoría de las lecturas que se mapean sobre la estructura de tallo-asa. Este artículo presenta información sobre la base de datos de miRNAs, miRBase, miRNAs conservados y linaje-específicos, haciendo hincapié en los miRNAs de plantas. Se describen los procedimientos de identificación *de novo* de miRNAs por predicción de la estructura de tallo-asa. Para esto se utilizan datos de secuenciación masiva de RNAs pequeños y de genoma o transcriptoma. Se describen también los métodos de predicción y validación de transcritos-blancos de miRNA.

ABSTRACT

MicroRNAs (miRNAs) are small non-coding RNAs 20 to 24 nucleotides long that silence gene expression post-transcriptionally. *De novo* identification of miRNAs is based on features of its biogenesis. In general, the criteria considered are the identification of a transcript that folds into a stem-loop secondary structure, as well as the miRNA and miRNA* sequences resulting from small RNA-sequencing. These sequences should possess nearly perfect complementarity, map onto the stem part of the secondary structure with a displacement of 2 nucleotides and represent most of the reads mapped onto the stem-loop structure. In this article, the miRBase database, the conserved and lineage-specific miRNA with an emphasis on plant miRNAs, and the procedures of *de novo* identification of miRNAs based on prediction of the stem-loop structure are described. These bioinformatic methods employ data obtained by next generation sequencing techniques, that is, small RNA-seq, coupled with RNA-seq for *de novo* transcriptome assembly, or available genome sequences. The methods of bioinformatic prediction and experimental validation of miRNA targets are also described.

INTRODUCCIÓN

Los microRNAs (miRNAs) son moléculas de cadena sencilla de RNA entre 20 y 24 nucleótidos (nt) de longitud, frecuentemente de 21 nt, las cuales mediante silenciamiento modulan la expresión génica a nivel postranscripcional o traduccional. Estas moléculas están presentes en todos los órganos y tejidos de plantas y animales. Los miRNA son reguladores muy importantes del desarrollo de organismos multicelulares; en plantas su función también es

necesaria para la regulación de la plasticidad fenotípica, la interacción con organismos simbióticos y la respuesta a diferentes tipos de estrés biótico y abiótico (1, 2, 3).

Existen varios tipos de RNAs pequeños de entre 20 y 30 nt de longitud que participan en la regulación de la expresión génica. Entre los más abundantes se encuentran los miRNAs y siRNAs (del inglés "small interfering RNA": RNAs pequeños de interferencia). Podemos distinguir estos dos grupos de acuerdo con un detalle particular de su biogénesis: mientras

PALABRAS

CLAVE

miRNA, RNAs pequeños, estructura secundaria tallo-asa, plantas, secuenciación masiva

KEY WORDS

miRNA, small RNAs, stem-loop secondary structure, plants, next generation sequencing

que un miRNA deriva de una molécula de RNA de una sola cadena con una estructura secundaria particular, un siRNA deriva de molécula de RNA de doble cadena (4).

En este artículo se describirán los procedimientos que se tienen que utilizar para la adecuada identificación de miRNAs, utilizando datos de secuenciación de RNAs pequeños y genoma o transcriptoma ensamblado *de novo*. Dichos procedimientos se basan en la formación de una estructura secundaria de tallo-asa por la molécula de RNA que originará a un miRNA. Esta metodología puede aplicarse a especies vegetales no modelo, que no cuentan con genoma secuenciado.

Biogénesis y modo de acción de los miRNAs

Un miRNA se origina a partir de un gen *MIR* que se transcribe por la RNA polimerasa II, dando lugar a un transcrito primario (pri-miRNA). Generalmente, los genes *MIR* de plantas se encuentran entre otros genes (a veces se dice que se encuentran en regiones intergénicas), pero también pueden ser parte de intrones e incluso exones de otros genes (5), en este caso se les denomina como intragénicos. A su vez, los transcritos primarios pueden tener intrones, los cuales se eliminan por el proceso de "splicing", o empalme. Además, de manera similar a otros transcritos de RNA polimerasa II, se modifican por adición del cap y la cola de poliadenina. Como ya se mencionó, los pri-miRNAs son transcritos de cadena sencilla, pero adquieren una estructura secundaria de tallo-asa, también llamada tallo-bucle (Fig. 1a). Estas secuencias forman doble cadena correspondiente al tallo de dicha estructura. Este segmento de doble cadena incluye tanto la secuencia del miRNA maduro como una secuencia casi complementaria denominada miRNA estrella (miRNA*), representadas en azul y verde, respectivamente, en la figura 1a. De manera canónica, el pri-miRNA se procesa (es decir, es cortado) en dos pasos. En plantas, a diferencia de animales, ambos pasos se realizan en el núcleo por una sola enzima RNasa tipo III de la familia DICER-LIKE (DCL), generalmente DCL1, según la nomenclatura para *Arabidopsis thaliana*. Primero el pri-miRNA se procesa, convirtiéndose en un precursor del miRNA (pre-miRNA, Fig. 1b) y luego en un dúplex miRNA/miRNA* (Fig. 1c) generalmente con hebras de 20 a 22 nt, aunque se han reportado algunos

miRNA funcionales de 23 o 24 nt (revisado en 3, 8). Las RNA-endonucleasas DCL cortan el RNA de doble cadena, dejando dos nucleótidos sobresalientes en los extremos 3', por lo que este dúplex presenta extremos colgantes, también llamados cohesivos, de dos nt.

El miRNA maduro se separa del miRNA* y se asocia a una proteína de la familia ARGONAUTA (AGO), generalmente AGO1, para formar parte del complejo de silenciamiento inducido por RNA (RISC, por sus siglas en inglés: RNA-induced silencing complex) (Fig. 1d). Aunque anteriormente se pensaba que la formación del complejo RISC de plantas ocurre en el citoplasma, publicaciones recientes sugieren que este proceso podría llevarse a cabo en el núcleo (revisado en 3). El miRNA dirige al complejo RISC hacia el transcrito blanco (Fig. 1e). Los miRNAs de plantas, a diferencia de lo que pasa en animales, poseen generalmente una alta complementariedad de bases con sus RNA blancos y por eso a menudo pueden tener como blanco un transcrito único, mientras que los miRNAs de animales bilaterales pueden tener numerosos blancos.

Generalmente, los miRNAs de plantas regulan la expresión de transcritos que codifican para proteínas, aunque también pueden actuar sobre transcritos de RNAs largos no codificantes. Un caso específico es de los genes *TAS* que no codifican para proteínas: los transcritos *TAS* son reconocidos por miRNAs y cortados por el complejo RISC. Después de ser cortados se convierten a moléculas de doble cadena por la actividad de una enzima RNA polimerasa dependiente de RNA; y luego se cortan en varias moléculas de 21 nt para generar un tipo de siRNAs denominadas tasiRNAs (del inglés "trans-acting" siRNA) (8). Los tasi son un grupo particular de phasiRNAs (del inglés "phased siRNAs") que son reclutados a un complejo RISC y dirigen el silenciamiento postranscripcional de otros transcritos no relacionados con los genes *TAS*. Los phasiRNAs, pueden originarse de genes que codifican o no para proteínas por el mecanismo arriba descrito para tasiRNAs (9).

Los miRNAs pueden desencadenar diferentes mecanismos de silenciamiento. En plantas, el silenciamiento génico mediante el corte del transcrito blanco (Fig. 1f) y su posterior degradación por la exorribonucleasa XRN4 es el que se ha reportado con mayor

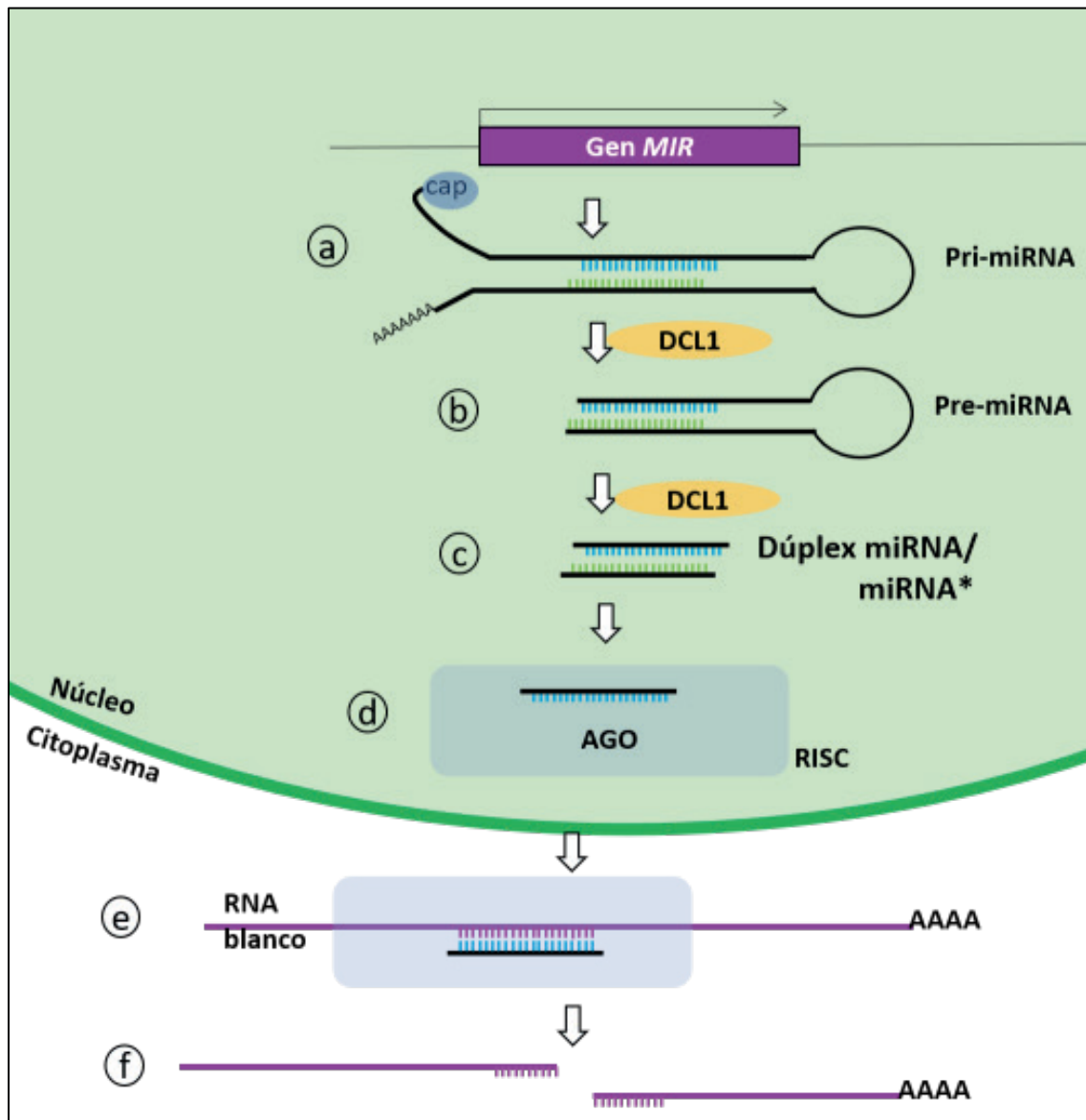


Figura 1. Biogénesis de miRNAs en plantas. (a) El gen MIR se transcribe generando un transcrito primario, o pri-miRNA, el cuál es procesado por la enzima DICER LIKE 1 (DCL1) en dos pasos; primero a precursor de miRNA, o pre-miRNA (b), y posteriormente a dúplex miRNA/miRNA* (c). Las hebras del miRNA y miRNA* tienen dos nucleótidos desapareados en su extremo 3'. (d) El miRNA maduro se asocia con la proteína ARGONAUTA (AGO), generalmente AGO1, como parte del complejo de silenciamiento inducido por RNA (RISC, por sus siglas en inglés: RNA-induced silencing complex). (e) El miRNA maduro dirige al complejo RISC hasta su transcrito blanco. (f) El modo más común de silenciamiento por miRNAs en plantas, el cual se indica en la figura, es el corte endonucleolítico del transcrito blanco por la proteína AGO. Para hacer esta imagen se utilizaron como referencia las revisiones 3, 6 y 7.

frecuencia. Sin embargo, cada día se acumulan más evidencias de inhibición de la traducción por miRNAs en plantas, como sucede, por ejemplo, con miR172, miR171, miR398 y miR156 que tienen como blancos a *APETALA 2*, *SCARECROW-LIKE PROTEIN 4*, *COPPER/ZINC SUPEROXIDE DISMUTASE 2* y *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 3*, respectivamente, así como sus parálogos. Los miRNA mencionados pueden regular a sus blancos por ambos mecanismos arriba

descritos (revisado en 6). Además, existen pocos reportes recientes de silenciamiento transcripcional mediante el mecanismo de metilación de DNA dependiente de miRNA. En un artículo se reporta en arroz la regulación de metilación de DNA mediante miRNAs de 24 nt, los cuales son generados por DCL3 y que se asocian a un complejo de silenciamiento que contiene a la proteína AGO4. En otro artículo se encontró que en *Arabidopsis* la complementariedad entre miR165/166 y sus

transcritos blanco *PHABULOSA* y *PHAVOLUTA* es necesaria para la metilación de estos dos genes (revisado en 3).

Familias de miRNAs

Los miRNAs maduros se agrupan en familias de acuerdo con sus secuencias. Los miembros de una familia comparten una similitud mayor al 84%, es decir, tienen al menos 17 nt en común y no más de 3 cambios de nucleótidos. Además, los miRNAs derivados del mismo precursor pueden tener diferentes longitudes por presentar o no un nucleótido extra en uno o ambos extremos del miRNA maduro; esto por el procesamiento diferencial por DCL, por particularidades del precursor o, en algunos casos por modificación después del procesamiento (10, 11).

Existen familias de miRNAs presentes en diversas especies vegetales que comparten blancos de la misma familia génica. Se refieren a ellas como familias conservadas de miRNAs. Por ejemplo, las plantas dicotiledóneas comparten 29 familias de miRNAs (12). No obstante, en las especies vegetales, son en realidad pocas las familias de las que se ha encontrado evidencia de conservación a lo largo de la historia evolutiva. Existen 9 familias de miRNAs que se han identificado desde el ancestro común de todas las embriofitas hasta las plantas con flores (10, 13). En cambio, la mayoría de los miRNAs que se han identificado son específicos de una sola especie. Se ha reportado que aproximadamente la mitad de los miRNA de cada especie no se conservan en otras taxa (10). Por ende, cada vez que se caracterizan los miRNAs de una nueva especie vegetal se descubren nuevos miRNAs.

MiRBase, la principal base de datos de miRNAs

MiRBase (<http://www.mirbase.org/>) es la principal base de datos de secuencias de miRNAs de todos los organismos en los que han sido identificados. Esta base de datos es pública y reúne no solo las secuencias de miRNAs maduros, sino también de sus precursores, así como información acerca del método por el cual fueron identificados y de su acumulación en diferentes órganos, etapas de desarrollo y bajo distintos tratamientos. En marzo del 2018 esta base de datos se actualizó a la versión 22, aumentando en más de un tercio su contenido

en comparación con la versión 21 liberada 4 años antes. Para la versión 22 de miRBase, se homologaron y curaron nombres asignados a los miRNA de distintas especies tanto vegetales, como animales.

La nomenclatura actual de miRBase para miRNAs de plantas se establece de la siguiente forma: las 3 primeras letras hacen referencia al organismo en el que fueron identificados, la primera letra al género, segunda y tercera letras a la especie; después de un guion se escribe el sufijo miR, el número de la familia y una letra para designar cada una de las isoformas, o variantes de este miRNA. Finalmente, con el sufijo -5p o -3p se indica si el miRNA maduro proviene del brazo 5' o 3' de la estructura tallo-asa del precursor. Por ejemplo: ath-MIR156c-5p se refiere a la isoforma "c" de la familia 156 de *Arabidopsis thaliana* proveniente del brazo 5' del tallo-asa. Actualmente se han publicado varios trabajos que reportan que ambas cadenas del dúplex son funcionales (14). Por esta razón, en las últimas versiones de miRBase ya no se aplica la terminología miRNA/miRNA* para distinguir el miRNA que se expresa más, sino la de 5p y 3p, para ambas cadenas del dúplex (revisado en 15). Esta nomenclatura desde hace tiempo está establecida para miRNAs de animales, y se ha ido aceptando para miRNAs de plantas. Sin embargo, en este trabajo utilizamos la terminología anterior, miRNA/miRNA*, para fines prácticos.

miRBase 22 tiene registros de miRNAs provenientes de solamente 82 especies vegetales. Cabe mencionar, que el número de especies de plantas es mucho más alto: tan solo de angiospermas se conocen más de 369,400 especies (16). Entre ellas, el número de especies con secuencias genómicas ensambladas y publicadas es bastante más bajo: para noviembre del 2019 habían sido 351 especies. Muchos de estos 351 genomas públicamente accesibles se encuentran en estado de borrador todavía (<https://www.plabipd.de>). Por otro lado, datos transcriptómicos están disponibles para un mayor número de especies, por ejemplo, recientemente fueron publicados datos de transcriptoma de más de mil especies de embriofitas, es decir, plantas terrestres (17). En los últimos años la cantidad de especies vegetales con genomas y transcriptomas

disponibles ha aumentado drásticamente, lo que ha posibilitado la identificación por predicción *in silico* de genes *MIR* y miRNAs de estas especies, por ejemplo, de *Ipomoea batatas L* (camote) y *Arundo donax L* (caña) (18, 19). Para identificar genes *MIR* y miRNAs, así como comprobar la existencia de miRNAs, se requiere realizar la secuenciación de RNAs pequeños descrita abajo. A su vez, esta comprobación depende de las muestras analizadas, ya que la presencia de diferentes miRNAs puede variar entre los órganos o condiciones que se analizan (20).

Uso de microarreglos para la descripción de miRNAs

Los microarreglos de miRNAs, también llamados chips ("microarray" en inglés), se emplean principalmente para analizar la expresión diferencial de miRNAs (21), pero también se han utilizado algunas veces para describir miRNAs conservados en especies no modelo, como se hizo para *Opuntia ficus-indica* (22). Para este último fin se aprovecha la homología entre los miRNAs conservados que ya han sido caracterizados en especies modelo, para identificar a estos mismos en especies no modelo.

Los microarreglos de miRNAs son una colección de sondas, es decir, de moléculas pequeñas de DNA de una sola cadena que representan a cada una de las secuencias de miRNA descritos y depositados en la base de datos miRBase, o a su secuencia complementaria, dependiendo de la técnica del marcaje de muestras de RNA (23). Todas estas sondas están fijadas en una superficie sólida, por ejemplo, en vidrio o plástico. En los ampliamente utilizados microarreglos Affymetrix, actualmente vendidos por la compañía ThermoFisher (<https://www.thermofisher.com/mx/es/home/life-science/microarray-analysis/affymetrix.html>), las secuencias de las moléculas-sondas son complementarias a miRNAs conocidos. Al hibridarse estas sondas por complementariedad de bases con los miRNAs de las muestras de RNA marcados de algún modo, generalmente con un fluoróforo, se pueden identificar los miRNAs de secuencia conocida presentes en muestras de distintos órganos, etapas de desarrollo o tratamientos. El microarreglo de Affymetrix puede ser usado para evaluar la expresión de miRNAs tanto vegetales, como animales, ya que incluye sondas correspondientes a todos

los miRNA depositados a la base de datos miR-Base. En el caso de muestras de RNA que se utilizan para hibridar con los chips de Affymetrix, se marcan las mismas moléculas de RNA por poliadenilación y ligación de fragmentos de DNA marcados con biotina. En otros tipos de microarreglos, cuando las muestras se marcan en el proceso de la transcripción reversa, las sondas tendrán la secuencia de los miRNAs, y no su secuencia complementaria. Sin embargo, el uso de los microarreglos no asegura que las moléculas que se hibridan tengan exactamente la misma secuencia que la molécula de referencia utilizada en el microarreglo. En muchos de los casos los miRNAs detectados podrían representar una mezcla de varias isoformas que pertenecen a la familia de la secuencia usada como referencia, teniendo uno, dos o hasta tres nucleótidos distintos, por lo que no todas estas isoformas coinciden exactamente con la secuencia que se encuentra en el microarreglo. Esta tecnología sigue disponible y puede brindar un panorama muy general sobre las posibles familias de miRNA conservadas presentes y las diferencias en su acumulación en los órganos y condiciones estudiados. Una ventaja del uso de microarreglos para medir la acumulación diferencial de miRNA es que no requiere de un gran poder de cómputo y análisis bioinformáticos. Actualmente, en el mercado se encuentran disponibles microarreglos no solamente de miRNAs maduros de la versión 20 de miRBase, ("Affymetrix® miRNA 4.0 array"), sino también de secuencias de precursores (los cuales entre distintas especies pueden variar notablemente, por lo que se recomienda su uso para evaluar la acumulación diferencial en la especie, para la cual fue creado el microarreglo). Por otra parte, el abaratamiento de las tecnologías de secuenciación aunado a las ventajas que estas proveen ha desplazado el uso de microarreglos.

Identificación *de novo* de miRNAs utilizando secuenciación masiva

Se denomina secuenciación masiva a la secuenciación de moléculas de ácidos nucleicos en la que se determinan en paralelo miles o millones de secuencias nucleotídicas. En inglés este tipo de secuenciación se llama "Deep sequencing" o NGS ("next generation sequencing": secuenciación de la siguiente generación). La secuenciación masiva de RNAs pe-

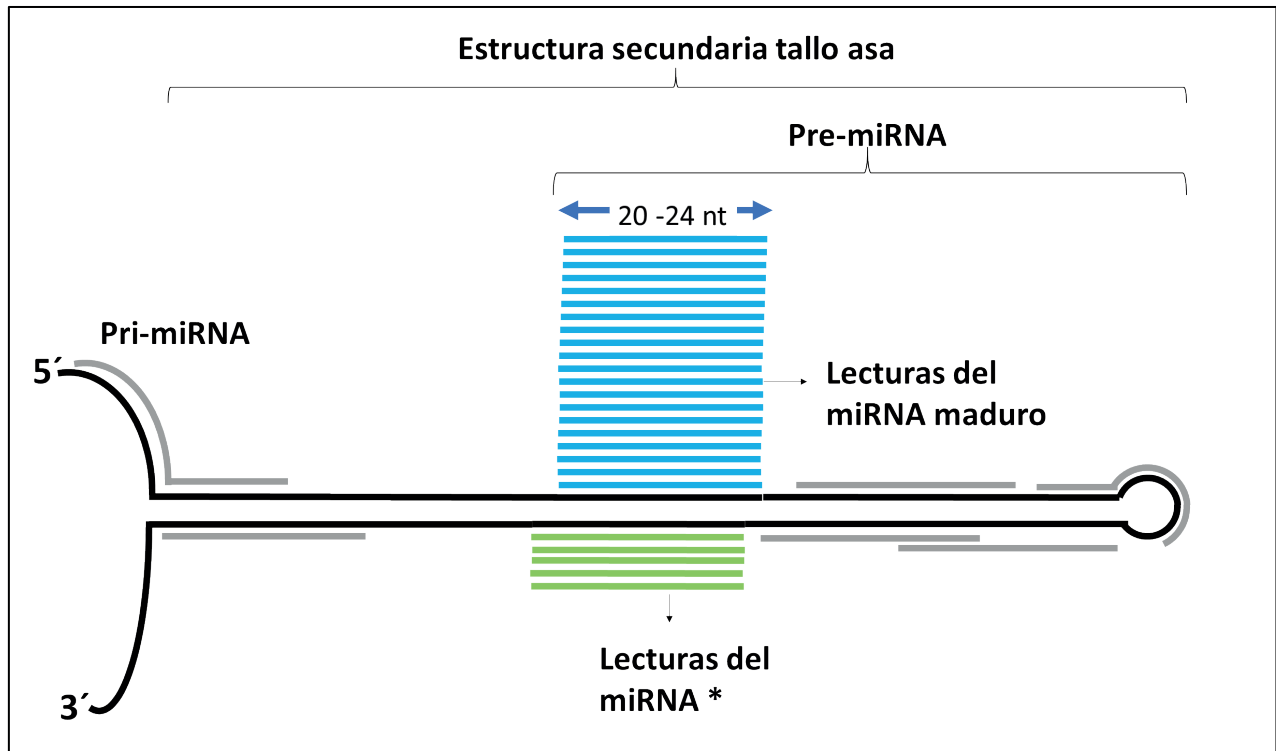


Figura 2. Análisis de posibles precursores de miRNAs. El miRNA primario (pri-miRNA) se representa con una línea negra. La estructura secundaria tallo-asa formada por el pri-miRNA y la sección que corresponde al precursor de miRNA (pre-miRNA) después del primer paso del procesamiento se indica con llaves. Con líneas grises se muestran las lecturas de la secuenciación de RNAs pequeños mapeados sobre el precursor de miRNA, la estructura tallo-asa, o el miRNA primario. En azul se representan las lecturas correspondientes al miRNA maduro y en verde las del miRNA*. La distribución de estas lecturas de RNAs pequeños es la que se espera por las características de la biogénesis de miRNAs (figura basada en la referencia 25).

queños ("small RNA-seq" en inglés, ver abajo) permite la identificación de probables miRNAs sin conocer previamente su secuencia. Técnicamente, cuando decimos secuenciación de RNA, en realidad se secuencian las moléculas de DNA complementario a este RNA. Como se mencionó arriba, para varias especies vegetales ya está disponible la secuencia de su genoma, y para muchas más están disponibles secuencias de transcriptomas ensamblados *de novo*. Si la especie de su interés no cuenta con genoma ni transcriptoma previamente reportado, para identificar miRNAs habría que realizar dos secuenciaciones distintas: una de RNA pequeños y otra de transcritos poliadenilados; esta última seguida por un ensamblado *de novo*. Más detalles acerca de la secuenciación y ensamblado *de novo* de transcriptomas se pueden consultar en una revisión de esta misma revista (24). El procedimiento para identificación *de novo* de miRNAs se basa en las particularidades de su biogénesis. A grandes rasgos, los criterios considerados, que se

explicarán detalladamente más adelante, son: las secuencias del miRNA y del miRNA* encontradas en los resultados de secuenciación como evidencia de su expresión, que forman un dúplex con un desfase de 2 ± 1 nt, como evidencia de corte por una RNAsa tipo III, además de la identificación de un precursor que cumple con características específicas de estructura secundaria (Fig. 2) (13).

En los últimos años se ha vuelto muy común el uso de secuenciación masiva de RNAs pequeños para la identificación de miRNAs. Se ha generado una gran cantidad de información, pero dicha información debe ser verificada utilizando datos transcriptómicos o genómicos de la especie de interés para averiguar el origen de los RNAs pequeños y comprobar si se trata o no de miRNAs. Solamente después del análisis abajo descrito, se deben depositar las secuencias de miRNAs identificados en bases de datos. Sin embargo, actualmente gran parte de la información depositada en miRBase no ha sido verificada más

allá de la secuenciación de RNAs pequeños. Esto ha provocado la anotación equivocada de otro tipo de RNAs pequeños, mayormente de siRNAs, como si fuesen miRNAs. Un análisis de los miRNAs de plantas de la versión 20 de miRBase mostró que el 75% de ellos no contaban con evidencia suficiente para ser anotados como miRNAs (26). Las imprecisiones en la identificación de miRNAs pueden resultar en su anotación errónea, lo que dificulta el estudio del papel de los miRNAs en la regulación génica.

A continuación, se describe la secuenciación de RNAs pequeños y los pasos de su posterior análisis necesario para una identificación exitosa de miRNAs.

Secuenciación masiva de RNAs pequeños

En un procedimiento de secuenciación masiva de RNAs pequeños se determina la secuencia de nucleótidos que corresponde a moléculas de RNA de entre 18 y 30 nt de longitud, aproximadamente, como su nombre lo define. Para ello, primero a partir de RNA total se aísla la fracción de moléculas de RNA de esta longitud que es utilizada en la construcción de bibliotecas para su secuenciación. En la purificación de RNAs pequeños, la muestra incluye moléculas cortas de RNA de varios tipos: además de los RNAs pequeños de interés, como miRNAs y siRNAs, también hay fragmentos de moléculas de RNAs más grandes, entre ellos los de RNAs mensajeros, RNAs de transferencia, RNAs ribosomales, entre otros.

En los extremos de estas moléculas de RNA se ligan fragmentos cortos de RNA denominados adaptadores. El adaptador ligado primero al extremo 3' de cada molécula de RNA contiene una secuencia conocida, y el adaptador ligado al extremo 5' tiene otra secuencia. Luego sobre el templado de las moléculas de RNA con adaptadores se sintetiza la primera cadena de cDNA en una reacción de transcripción reversa. Esta cadena de cDNA se utiliza para obtener varias moléculas de DNA de doble cadena por el proceso de PCR (reacción en cadena de la polimerasa, por sus siglas en inglés). Uno de los adaptadores se usará para iniciar la secuenciación. El método empleado en la elaboración de las bibliotecas que se secuenciarán es clave en la obtención de resultados, pues puede generar un ses-

go debido a que la incorporación de ciertos adaptadores puede verse desfavorecida. Se ha evidenciado que el uso de adaptadores con secuencias degeneradas, es decir, el uso de un conjunto de adaptadores con secuencias variables en el extremo en que se ligan a los RNAs pequeños, permite evitar este sesgo. Sin embargo, para el uso de adaptadores con secuencias degeneradas es preferible contar con un genoma de referencia para poder distinguir dentro de las lecturas de secuenciación los nucleótidos que fueron añadidos (adaptadores) y las secuencias propias de la especie de interés (27, 28).

Para la anotación de miRNAs se deben incluir resultados de secuenciación de al menos dos bibliotecas independientes. Otro de los objetivos importantes de la secuenciación de RNAs pequeños es el monitorear su expresión en diferentes órganos, etapas de desarrollo o condiciones. Para poder realizar el análisis estadístico de la expresión diferencial y resaltar resultados reproducibles, además de tener muestras de condiciones diferentes, es necesario incluir dos o tres replicas biológicas, es decir, muestras de una misma condición provenientes de experimentos independientes.

Procesamiento de los datos crudos de la secuenciación

Como resultado de la secuenciación se obtienen archivos en formato FastQ de cada una de las bibliotecas secuenciadas. Este es un formato basado en texto que contiene la secuencia de nucleótidos, así como valores de calidad de secuenciación para cada nucleótido. Al tratarse de secuencias cortas, los resultados de secuenciación son generalmente de alta calidad. Existen diversas herramientas que generan reportes de calidad a partir de este tipo de archivos; una de ellas es FastQC (29). Después de verificar la calidad de las secuencias obtenidas, se procede al procesamiento de resultados de secuenciación, es decir, la remoción de los adaptadores. Existen varias herramientas, como Atropos, cutadapt, Trimmomatic o BBDuk que permiten eliminar los adaptadores de secuencias conocidas no variables (Revisado en 30). Tras el procesamiento, se debe analizar la distribución de longitud de las secuencias procesadas, lo cual sirve tanto como control de calidad de la purificación del RNA, de la preparación de

bibliotecas y de la secuenciación, como para confirmar la ausencia de nucleótidos provenientes de adaptadores. Generalmente los RNAs pequeños más abundantes son siRNAs de 24 nt de longitud, seguidos por miRNAs de 21 nt (30).

Mapeo sobre genoma o transcriptoma de las lecturas procesadas

La secuencia de nucleótidos de cada una de las moléculas obtenida como resultado de secuenciación masiva de RNA se denomina lectura. Para identificar genes *MIR*, las secuencias exactas de las lecturas procesadas de secuenciación de los RNAs pequeños se buscan en el genoma de referencia o el transcriptoma ensamblado *de novo*. A este proceso se le denomina mapeo y se basa en algoritmos de alineamiento de ácidos nucleicos. Para posteriores análisis, deberán considerarse únicamente las lecturas que pudieron mapearse. Con este paso, se discriminan posibles artefactos generados por secuenciación o por contaminación. La longitud de las estructuras tallo-asa de los pri-miRNAs de plantas varía de entre menos de 100 y hasta 900 nt (10). Por esta razón, en caso de no contar con un genoma de referencia y mapear las lecturas de RNAs pequeños sobre el transcriptoma, deben considerarse los transcritos de este rango de longitud.

Conteo de lecturas

En el caso de los RNAs pequeños el número de lecturas con una misma secuencia (secuencia única) que se obtiene como resultado de secuenciación masiva indica directamente el número de veces que se secuenció dicha molécula en la biblioteca analizada. Después de que se han contabilizado las lecturas, deben conservarse solo las lecturas que aparecen en al menos dos bibliotecas: de dos muestras provenientes cada una de condiciones distintas o de dos muestras de una condición provenientes de dos experimentos independientes (replicas biológicas). Estos datos pueden usarse para la identificación de miRNAs y el número de lecturas puede normalizarse, como se explica más adelante, para posterior análisis de acumulación de miRNAs.

Otra alternativa es agrupar las lecturas con sus variantes posicionales (Fig. 3). Los miRNAs pueden tener variantes posicionales

5'-UGACAGAAGAGAGUGAGCAC-3'
 UGACAGAAGAGAGUGAGCA.
 .GACAGAAGAGAGUGAGCAC
 UGACAGAAGAGAGUGAGCACA
 AUGACAGAAGAGAGUGAGCAC

Figura 3. Variantes posicionales del microRNA de *Arabidopsis thaliana* *ath-MIR156c-5p* de acuerdo con la secuencia de su precursor y de la variante más abundante del miRNA maduro disponible en la base de datos miRBase (<http://www.mirbase.org/>).

que ocurren como resultado de corte por enzimas Dicer/DCL en posiciones ligeramente variables (31), o modificaciones en el extremo 3', como es el caso de miR1510 de soya que al monouridilarse pasa de 21 a 22 nt (11). Actualmente la variación de un nucleótido de longitud en un extremo de la molécula es aceptada como una variante posicional de un mismo miRNA, lo cual permite agrupar 5 variantes de un mismo miRNA (Fig. 3). Algunos algoritmos para el análisis de RNAs pequeños utilizan esta estrategia. Sin embargo, debe tenerse especial cuidado en que las secuencias agrupadas sean idénticas a las secuencias sobre las que se mapean.

Cuando se agrupan las variantes posicionales, para validar la presencia de un miRNA, se suma el número total de lecturas de las variantes posicionales tanto para el miRNA, como para el miRNA* y se calcula el porcentaje de estas lecturas respecto al número total de lecturas mapeadas al locus o al transcrito, la suma de ambos o alguno de los dos debe ser de al menos el 75%; este valor está sugerido por Axtell & Meyers en (13).

Predicción de precursores de miRNAs

En la siguiente etapa de la identificación *de novo* de miRNAs, se analiza la posibilidad de que las secuencias del genoma o transcritos a los que se mapearon las lecturas obtenidas por la secuenciación de RNAs pequeños correspondan a genes *MIR* o a precursores de miRNAs, respectivamente. Para ello se recaban, o minan, hasta 150 nt hacia cada lado, las secuencias que están flanqueando los sitios en donde se mapeó una lectura de

la secuenciación de RNAs pequeños (13). Se modela la estructura secundaria de estas secuencias recabadas para probar si pueden formar la estructura tallo-asa característica de los transcritos de genes *MIR*. Existen varios programas computacionales para la predicción de la estructura secundaria de RNA que evalúan también la probabilidad de que dichas secuencias sean precursores de miRNAs. Entre ellos destacan: miRDeep-P2 (32), ShortStack (33), miR-PREFeR (34), MIRENA (35). Estos algoritmos son de código abierto y de libre acceso. ShortStack, además, está disponible como una herramienta en línea que puede ser accedida desde <https://de.cyverse.org/de/> y no requiere mayor poder de cómputo. Aún al utilizar cualquiera de estos programas, es necesario revisar que los resultados predichos cumplen con los parámetros establecidos para la identificación *de novo* de miRNAs.

Como se ejemplifica en la figura 2, después de la predicción de la estructura secundaria, todo el conjunto de lecturas de secuenciación de RNAs pequeños se mapea sobre las secuencias minadas capaces de formar la estructura tallo-asa (mostrada en la figura 2 en color negro). Se contabiliza, cuántas lecturas se mapean a la secuencia exacta que corresponde al miRNA maduro (líneas en color azul), al miRNA* (en verde) y al resto de la estructura tallo-asa (en gris). Si la proporción de lecturas correspondientes al posible miRNA maduro es mayor a la del posible miRNA* y también se mapean unas pocas lecturas en otras regiones de la secuencia en cuestión, es alta la probabilidad de que la secuencia corresponda a un precursor de miRNA (32). Si en cambio, las lecturas mapeadas se traslapan entre sí a lo largo de la estructura tallo-asa, es probable que las lecturas que se mapean sobre esta estructura secundaria sean fragmentos de RNA generados como productos de degradación de otros tipos de RNA que pueden presentarla, como RNAs ribosomales, RNAs de transferencia, RNAs pequeños nucleares y RNAs pequeños nucleolares, entre otros, y no se trate de un precursor de miRNAs.

Las características con las cuales debe cumplir un posible precursor de miRNA para la anotación *de novo*, propuestas en el artículo 13, son las siguientes:

- El dúplex miRNA/miRNA* debe presentar extremos 3' colgantes con dos nucleótidos libres

- El dúplex miRNA/miRNA* puede tener hasta 5 bases no apareadas, solamente 3 de ellas formando protuberancias
 - Más del 75% de las lecturas mapeadas sobre el posible precursor deben corresponder al miRNA maduro o al miRNA*
 - La región de la estructura del tallo que forma el dúplex miRNA/miRNA* no debe tener tallos secundarios largos
 - La longitud del precursor debe limitarse a 300 nt
 - En general deben anotarse miRNAs de 20 a 22 nt
 - La anotación de miRNAs de 23 a 24 nt requerirá de fuerte evidencia que la respalde.
- Algunas de estas características se ilustran en pre-miRNAs de plantas en la figura 4.

Anotación de miRNAs

Finalmente, entre los miRNAs predichos *de novo* deben estar presentes miRNAs conservados en varias especies vegetales. Distinguir los miRNAs conservados de los nuevos, no descritos anteriormente, sirve como un control del correcto proceso de identificación *de novo*. Los miRNAs predichos pueden compararse contra las bases de datos públicos de miRNAs, como miRBase. Para realizar estas comparaciones pueden usarse herramientas para mapeo como seqmap o bowtie (36, 37). Generalmente se considera un miRNA como nuevo cuando difiere en más de 4 nt de los miRNAs conocidos y se consideran como isoformas de miRNAs conocidos aquellos que difieran entre 1 y 4 nt. Una gran parte de los miRNAs conservados actúan en procesos de desarrollo y de respuesta a estrés (1, 2, 3), por lo que se esperaría encontrar este tipo de miRNAs. El número de familias conservadas que se esperaría encontrar varía de acuerdo con el clado al que pertenece la especie vegetal que se está analizando. Por ejemplo, en el caso de una planta dicotiledónea se esperaría encontrar varias de las 29 familias que han sido anotadas como conservadas entre dicotiledóneas (13). Al ser los miRNAs de 21 y 22 nt los más representados en las bases de datos disponibles (12), se espera que la mayoría de los miRNAs conservados que se identifiquen tengan esa misma longitud.

Al hacer esta anotación, es importante considerar que la historia evolutiva de los miRNAs muestra una divergencia entre miRNAs de plantas y animales. Hasta hoy

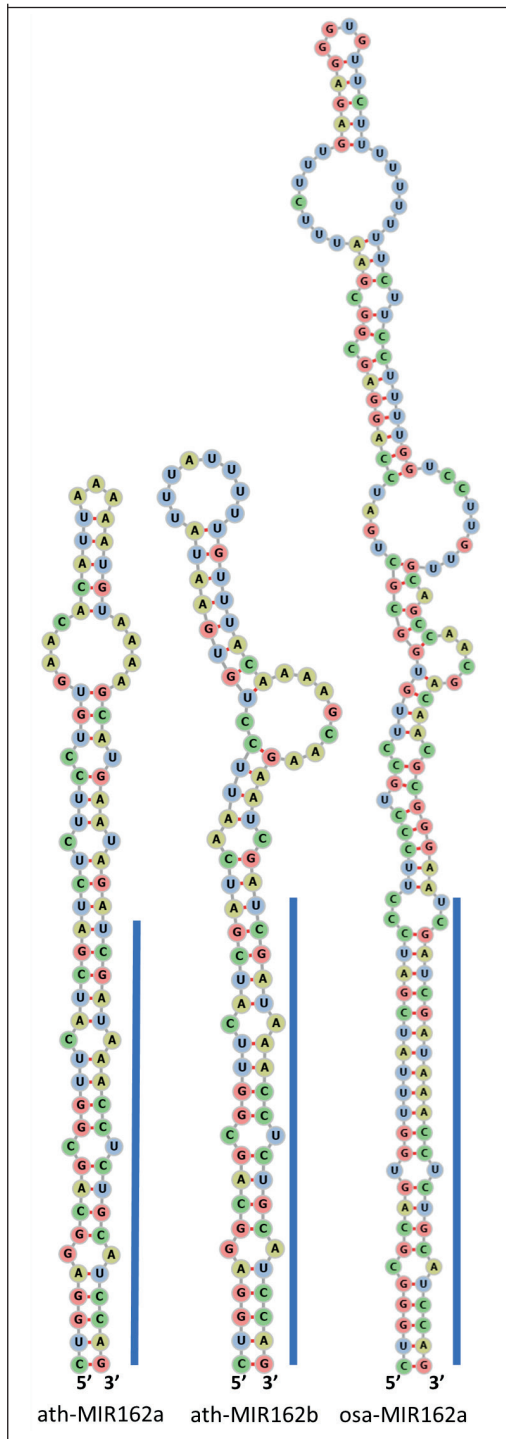


Figura 4. Estructura secundaria predicha de algunos precursores de miRNAs de *Arabidopsis thaliana* y *Oryza sativa*. La línea azul indica los nucleótidos de los miRNAs maduros. Imagen creada con <http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>, basada en la referencia 36. La parte de la estructura secundaria tallo-asa de pri-miRNA que se cortó durante el primer paso del procesamiento no se muestra.

en día son muy escasos los reportes sobre miRNAs que comparten homología en plantas y animales y que tienen el mismo blanco; por

ejemplo, en un artículo se describe que el miRNA con la misma secuencia en humanos y planta *Atropa belladonna* tiene un mismo blanco en humanos (39). Cuando en plantas se llegan a identificar varios miRNAs animales, lo más probable es que la muestra se haya contaminado en alguna etapa del proceso, por lo que sería recomendable no utilizar muestras contaminadas para la identificación de los miRNAs.

Los miRNAs identificados *de novo* en un procedimiento que cumple con todos los parámetros antes descritos, en el caso de tener menos del 84% de identidad con los miRNAs que pertenecen a una familia anteriormente reportada, se consideran como miRNAs nuevos, o linaje-específicos.

La normalización

Con el fin de poder comparar el nivel de acumulación de un miRNA entre diferentes condiciones estudiadas, es decir, entre las bibliotecas (muestras) analizadas, es necesario normalizar el número de lecturas, por lo que brevemente describimos las formas de normalizar las lecturas y las herramientas que se pueden utilizar para esto. Aunque las metodologías y parámetros para normalizar lecturas procesadas de secuenciación masiva de RNAs poliadenilados están más establecidos, estos métodos no siempre son los más apropiados para RNAs pequeños. En particular, la acumulación de distintos miRNAs sigue una distribución exponencial, es decir, la gran mayoría de lecturas totales provienen de unos cuantos miRNAs, mientras que el resto de miRNAs tienen bajo número de lecturas (40).

Se han implementado distintos métodos para normalizar los datos de secuenciación de RNAs pequeños. Dichos métodos pueden clasificarse en dos categorías: de escala, es decir, la aplicación de operaciones matemáticas lineales como en los métodos TMM (del inglés "Trimmed Mean of M-values"), RPTM ("Reads per ten million") o LOWESS ("locally weighted scatterplot smoothing") (41) y estadísticos, como cuantil, estabilización de varianza, método invariante o el algoritmo DEseq (42).

Una apropiada normalización facilitará la identificación de miRNAs diferencialmente expresados y, además, permitirá distinguir y descartar los datos de secuenciación cuya reproducibilidad no es confiable. Es recomen-

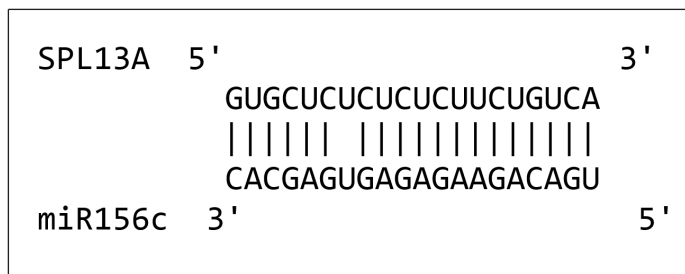


Figura 5. Sitio de unión de un miRNA y su transcrito blanco. Se muestra el apareamiento casi perfecto de miR156c de *Arabidopsis thaliana* con SPL13A (del inglés "Squamosa Promoter-Binding Protein Like 13A"), las líneas verticales representan los enlaces formados entre ambas moléculas. Imagen generada con la herramienta RNAhybrid de BiBiServ2 (<https://bibiserv.cebitec.uni-bielefeld.de/rnahybrid>).

dable el uso del software UEA workbench (42), ya que presenta una interfase amigable con el usuario para la aplicación de diferentes métodos de normalización de lecturas de RNAs pequeños. Esto facilita la elección del método más apropiado para las muestras utilizadas. Por otra parte, para utilizar esta plataforma se requiere de un genoma de referencia, lo que restringe su uso.

La expresión diferencial de un miRNA puede indicar la importancia de regulación de expresión génica por este miRNA durante un proceso en particular. Para poder descifrar el papel de un miRNA, es importante identificar los transcritos sobre los cuáles puede tener acción.

Predicción y validación de blancos de miRNA

Existen programas bioinformáticos para predecir los blancos sobre los que un miRNA podría actuar, uno de los más utilizados es psRNATarget (<http://plantgrn.noble.org/psRNATarget/>). Este tipo de programas mediante alineamientos identifican los sitios donde un miRNA se uniría a un blanco por complementariedad de bases. En la figura 5 se muestra un ejemplo de la predicción del apareamiento entre un miRNA y su blanco. Se recomienda en el caso de miRNAs de plantas considerar la región semilla (la región que debe ser complementaria con el transcrito) desde el nucleótido 2 al 13 del miRNA, permitiendo hasta dos nucleótidos desapareados en esta región.

La identificación de pares miRNA posibles blancos que previamente han sido reporta-

dos en otras especies sirve como control del correcto proceso. Es importante considerar que estos procedimientos realizados *in silico* requieren de comprobación experimental.

Por otra parte, el efecto de un miRNA que regula un transcrito mediante corte puede observarse en la disminución de la acumulación del transcrito. Por lo tanto, cuando se cuentan con datos de expresión de transcritos, pueden ser comparados cambios en acumulación de un miRNA y de su blanco. Una correlación inversa de la acumulación de miRNA y su blanco sugiere el papel del miRNA en la regulación de su blanco por corte endonucleolítico y posterior degradación. Sin embargo, no siempre se observa esta correlación inversa de acumulación del miRNA y su blanco, ya que en algunos casos puede requerirse de un umbral de acumulación para la activación de la regulación por corte endonucleolítico. Además, la regulación puede ser específica de algunos tipos celulares, por lo que podría no reflejarse en la acumulación de miRNA y su blanco en muestras de órganos completos que incluyen células pertenecientes a diferentes tejidos y tipos celulares (43). También debe considerarse que en el caso de inhibición de traducción mediada por miRNAs no se observará un efecto sobre la acumulación de transcrito.

De manera experimental la regulación de un miRNA mediante corte puede ser probada utilizando la técnica modificada de RACE (del inglés "Rapid Amplification of cDNA Ends": amplificación rápida de los extremos de cDNA). La modificación consiste en un paso previo de ligar un adaptador a la molécula de RNA, por lo que se denomina RLM-RACE ("RNA ligase-mediated RACE"). La acción de corte de AGO sobre un transcrito blanco de miRNA genera un monofosfato 5' en el extremo que ha sido cortado. La técnica de RLM-RACE aprovecha esta característica para ligar el adaptador de RNA en el extremo 5' (entonces, en este caso es 5' RLM-RACE). Luego, utilizando un oligonucleótido específico para el gen de interés y el otro homólogo a la secuencia del adaptador, se amplifica una parte del transcrito cortado río abajo del sitio de corte. Secuenciando el fragmento amplificado se evidencia el sitio exacto en donde ocurrió el corte (44). Esta técnica fue adaptada para analizar mediante secuenciación masiva todos

los sitios de corte por AGO en los transcritos de una muestra biológica, a esta técnica adaptada se le denomina degradoma o análisis en paralelo de los extremos de RNA, PARE por sus siglas en inglés (45, 46). Por la gran cantidad de información que el análisis de predicción de blancos o de su confirmación por PARE puede arrojar, es recomendable realizar una ontología de los genes y realizar un análisis de las categorías ontológicas a las que pertenecen los genes, cuya expresión se regula por miRNAs, y así identificar los procesos globales en los que pueden participar. El procedimiento de identificación y clasificación de miRNAs de plantas ha tenido grandes mejoras en los últimos años gracias a la aplicación de técnicas de secuenciación masiva y al arduo trabajo de diferentes grupos de investigación para refinar los procesos mediante los cuales se analizan dichas secuencias. Por un lado, se han ido refinando los parámetros de identificación de miRNAs, y por el otro lado, se han ido desarrollado diferentes herramientas que hacen más accesible su aplicación al requerir menos poder de cómputo y ser más amigables con el usuario. Actualmente existen plataformas y programas gratuitos, algunos de ellos mencionados arriba, que facilitan

el proceso descrito de identificación de miRNAs. La regulación genética es uno de los detonadores de la diversidad y la plasticidad en especies vegetales. La identificación de miRNAs en distintas especies vegetales ha permitido a su vez identificar genes que son regulados por miRNAs específicos en especies particulares. Pero, además, la recopilación y análisis de los datos que se van generando, permite visualizar parte del panorama de regulación genética que se comparte o no entre clados taxonómicos, influyendo en la diversidad de las especies vegetales.

Agradecimientos. El trabajo de las autoras sobre identificación de miRNAs y sus blancos en el cardón *Pachycereus pringlei*, una especie vegetal no modelo, ha sido parcialmente financiado por proyecto PAPIIT-UNAM IN201318. Agradecemos mucho todos los consejos y sugerencias del Dr. Damien Formey, Centro de Ciencias Genómicas - UNAM, a lo largo de realización de este proyecto de investigación, y las sugerencias de tres revisores anónimos que permitieron mejorar este artículo. A Mayra Liliana López-Valle se le otorgó una beca doctoral del CONACyT (registro 288069).



REFERENCIAS

1. D'Ario M, Griffiths-Jones S, Kim M (2017) Small RNAs: big impact on plant development. *Trends Plant Sci* 22:1056-1068.
2. Tang J, Chu C (2017) MicroRNAs in crop improvement: fine-tuners for complex traits. *Nature Plants* 3:17077.
3. Song X, Li Y, Cao X, Qi Y (2019) MicroRNAs and their regulatory roles in plant-environment interactions. *Annu Rev Plant Biol* 70:489-525.
4. Morris KV, Mattick, JS (2014) The rise of regulatory RNA. *Nature Reviews Genetics* 15:423-437.
5. Colaiacovo M, Lamontanara A, Bernardo L, Alberici R, Crosatti C, Giusti L, Cattivelli L, Faccioli P (2012) On the complexity of miRNA-mediated regulation in plants: novel insights into the genomic organization of plant miRNAs. *Biology Direct* 7:15.
6. Yu Y, Jia T, Chen X (2017) The 'how' and 'where' of plant microRNAs. *New Phytologist* 216:1002-1017.
7. Chen X (2009) Small RNAs and their roles in plant development. *Annu Rev Cell Dev Biol* 25:21-44.
8. Rogers K, Chen X (2013) Biogenesis, turnover, and mode of action of plant microRNAs. *Plant Cell* 25:2383-2399.
9. Deng P, Muhammad S, Cao M, Wu L (2018) Biogenesis and regulatory hierarchy of phased small interfering RNAs in plants. *Plant Biotechnol J*. 16:965-975.
10. Cuperus JT, Fahlgren N, Carrington JC (2011) Evolution and functional diversification of MIRNA genes. *Plant Cell* 23:431-442.
11. Fei Q, Yu Y, Liu L, Zhang Y, Baldrich P, Dai Q, Chen X, Meyers BC (2018) Biogenesis

- of a 22-nt microRNA in *Phaseoleae* species by precursor-programmed uridylation. *Proc Natl Acad Sci* 115:8037–8042.
12. Chávez Montes RA, de Fátima Rosas-Cárdenas F, De Paoli E, Accerbi M, Rymarquis LA, Mahalingam G, Marsch-Martínez N, Meyers BC, Green PJ, de Folter S (2014) Sample sequencing of vascular plants demonstrates widespread conservation and divergence of microRNAs. *Nat Commun* 5:3722.
 13. Axtell M, Meyers B (2018) Revisiting criteria for plant microRNA annotation in the era of big data. *Plant Cell* 30: 272–284.
 14. Liu W W, Meng J, Cui J, Luan Y S (2017) Characterization and function of MicroRNAs* in plants. *Frontiers in Plant Science* 8:2200.
 15. Kozomara A, Griffiths-Jones S (2014) MiRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research* 42: D68–D73.
 16. Lughadha E N, Govaerts R, Belyaeva I, Black N, Lindon H, Allkin R, Magill R E, Nicolson N (2016) Counting counts: revised estimates of numbers of accepted species of flowering plants, seed plants, vascular plants and land plants with a review of other recent estimates. *Phytotaxa* 272:82–88.
 17. One Thousand Plant Transcriptomes Initiative (2019) One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574:679–685.
 18. Dehury B, Panda D, Sahu J, Sahu M, Sarma K, Barooah M, Sen P, Modi M (2013) *In silico* identification and characterization of conserved miRNAs and their target genes in sweet potato (*Ipomoea batatas* L.) expressed sequence tags (ESTs). *Plant Signal Behav* 8:12.
 19. Jike W, Sablok G, Bertorelle G, Li M, Varotto C (2018) *In silico* identification and characterization of a diverse subset of conserved microRNAs in bioenergy crop *Arundo donax* L. *Scientific Reports* 8:16667.
 20. Cui J, You C, Chen X (2017) The evolution of microRNAs in plants. *Curr Opin Plant Biol.* 35:61–67.
 21. Thomson JM, Parker JS, Hammond SM (2007) Microarray analysis of miRNA gene expression. *Methods Enzymol* 427:107–22.
 22. Rosas-Cárdenas Fde F, Caballero-Pérez J, Gutiérrez-Ramos X, Marsch-Martínez N, Cruz-Hernández A, de Folter S (2015) microRNA expression during *prickly pear* cactus fruit development. *Planta* 241:435–48.
 23. Liu C, Calin G, Volinia S, Croce CM (2008) MicroRNA expression profiling using microarrays. *Nat Protoc* 3:563–578.
 24. Rodríguez-Alonso G, Shishkova S (2018) Estudio del transcriptoma mediante RNA-seq con énfasis en las especies vegetales no modelo. *Rev Educ Bioquímica* 37:75–88.
 25. Friedländer MR, Chen W, Adamidi C, Maaskola J, Einspanier R, Knespel S, Rajewsky N (2008) Discovering microRNAs from deep sequencing data using miRDeep. *Nature Biotechnology* 26:407–415.
 26. Taylor RS, Tarver JE, Hiscock SJ, Donoghue PC (2014) Evolutionary history of plant microRNAs. *Trends Plant Sci* 19:175–82.
 27. Sorefan K, Pais H, Hall AE, Kozomara A, Griffiths-Jones S, Moulton V, Dalmay T (2012) Reducing ligation bias of small RNAs in libraries for next generation sequencing. *Silence* 3:4.
 28. Dard-Dascot C, Naquin D, d'Aubenton-Carafa Y, Alix K, Thermes C, van Dijk E (2018) Systematic comparison of small RNA library preparation protocols for nextgeneration sequencing. *BMC Genomics* 19:118.
 29. Andrews S (2010) FastQC a Quality Control Tool for High Throughput Sequence Data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
 30. Chávez Montes RA, Jaimes-Miranda F, de Folter S (2019) Bioinformatic analysis of small RNA sequencing libraries. *Methods in Molecular Biology* 1932:51–63.
 31. Bologna NG, Schapire AL, Zhai J, Chorostecki U, Boisbouvier J, Meyers BC, Palatnik JF (2013) Multiple RNA recognition patterns during microRNA biogenesis in plants. *Genome Res* 23:1675–1689.
 32. Kuang Z, Wang Y, Li L, Yang X (2019) miR-Deep-P2: accurate and fast analysis of the microRNA transcriptome in plants. *Bioinformatics* 35:2521–2522.
 33. Johnson NR, Yeoh JM, Coruh C, Axtell MJ (2016) Improved placement of multi-mapping small RNAs. *G3* 6:2103–2111.
 34. Lei J, Sun Y (2014) miR-PREFeR: an accurate, fast and easy-to-use plant miRNA prediction tool using small RNA-Seq data. *Bioinformatics* 30: 2837–2839.

35. Mathelier A, Carbone A (2010) MIRENA: finding microRNAs with high accuracy and no learning at genome scale and from deep sequencing data. *Bioinformatics* 26:2226–2234.
36. Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP (2002) MicroRNAs in plants. *Genes Dev* 16:1616–1626.
37. Jiang H, Wong WH (2008) SeqMap: mapping massive amount of oligonucleotides to the genome. *Bioinformatics* 24:2395–2396.
38. Langmead B, Trapnell C, Pop M, Salzberg S L (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10: R25.
39. Avsar B, Zhao Y, Li W, Lukiw WJ (2019) *Atropa belladonna* expresses a microRNA (aba-miRNA-9497) highly homologous to *Homo sapiens* miRNA-378 (hsa-miRNA-378); both miRNAs target the 3'-untranslated region (3'-UTR) of the mRNA encoding the neurologically relevant, zinc-finger transcription factor ZNF-691. *Cell Mol Neurobiol* 40:179–188.
40. Barquist L, Vogel J (2015) Accelerating discovery and functional analysis of small RNAs with new technologies. *Annu Rev Genet* 49:367–394.
41. Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* 11:R25.
42. Beckers M, Mohorianu I, Stocks M, Applegate C, Dalmay T, Moulton V (2017) Comprehensive processing of high-throughput small RNA sequencing data including quality checking, normalization, and differential expression analysis using the UEA sRNA Workbench. *RNA* 23:823–835.
43. He J, Xu M, Willmann MR, McCormick K, Hu T, Yang L, Starker CG, Voytas DF, Meyers BC, Poethig RS (2018) Threshold-dependent repression of SPL gene expression by miR156/miR157 controls vegetative phase change in *Arabidopsis thaliana*. *PLoS Genetics* 14:4.
44. Wang C, Fang J (2015) RLM-RACE, PPM-RACE, and qRT-PCR: an integrated strategy to accurately validate miRNA target genes. *Methods Mol Biol* 1296:175–186.
45. Addo-Quaye C, Eshoo TW, Bartel DP, Axtell MJ (2008) Endogenous siRNA and miRNA targets identified by sequencing of the *Arabidopsis* degradome. *Current Biology* 18:758–762.
46. German MA, Luo S, Schroth G, Meyers BC, Green PJ (2009) Construction of Parallel Analysis of RNA Ends (PARE) libraries for the study of cleaved miRNA targets and the RNA degradome. *Nat Protoc.* 4:356–62.