

Capítulo 10

Expresión genética en plantas; transcriptomas y miARNs

Carolina Enríquez González
Mario Alejandro Mejía Mendoza
*Departamento de Química del Centro de Ciencias Básicas
Universidad Autónoma de Aguascalientes*

Resumen

Los genes son unidades moleculares que almacenan la información hereditaria y permiten transmitirla a la descendencia. Mediante el uso de la información contenida en cada gen estos son capaces de codificar una proteína o parte de ella; a este proceso se le llama expresión genética y es realizada por todos los seres vivos. Si bien existen ciertas diferencias entre como lo hacen procariontes y eucariontes, en ambos casos, el procedimiento se basa en la lectura del gen en grupos de 3 pares de base llamados *codones*. Cada codón corresponde a uno de los 20 diferentes aminoácidos usados para construir proteínas. La expresión genética de un organismo es vital para su metabolismo, supervivencia, adaptación y defensa contra estreses bióticos y abióticos.

En plantas al ser eucariontes, su expresión genética se da en el núcleo mediante la síntesis de ARN mensajero (ARNm) a partir de ADN por medio de la enzima ARN Polimerasa II. Posteriormente los codones del ARNm son leídos por los ribosomas que también empiezan la construcción de la proteína *de novo*. En este capítulo profundizaremos acerca de los procesos expresión génica en plantas, generalidades, inducción, regularización y métodos para medir los niveles de expresión de genes de interés.

Introducción

La expresión genética se define como el uso de la información contenida en un gen para la síntesis de una proteína. Para que esto pase, primero deben de suceder una serie de pasos que permitan transformar el ADN en aminoácidos. De estos pasos se distinguen dos principales: la transcripción, que es el proceso mediante el cual el ADN es transformado a ARNm, y la traducción, que se define como la codificación de codones a aminoácidos. Estos dos procesos suceden tanto en procariontes como eucariontes y son llevados a cabo mediante una serie de enzimas que trabajan de manera conjunta. La traducción y transcripción pueden ser ilustrados mediante el *Dogma central* que explica de manera gráfica la síntesis de proteínas a nivel citoplasmático a partir del ADN (Fig. 10.1). El *Dogma central* fue propuesto en 1958 por Francis Crick, y a partir de esa fecha ha sido modificado para lograr representar de manera más exacta el como ocurre la expresión genética en todos los seres vivos.

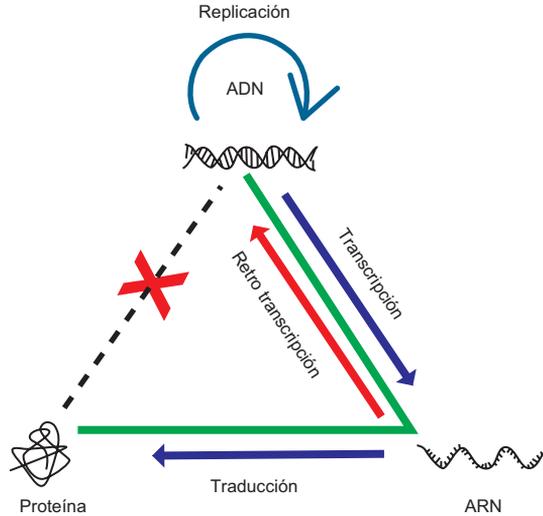


Figura 10.1. Dogma central de la biología. La línea punteada con la cruz roja indica que no es posible ir desde proteína a ADN.

El genoma de organismos vegetales está contenido en cada una de sus células, se compone de ADN (ácido desoxirribonucleico), el cual comprende toda la información y las instrucciones necesarias para llevar a cabo todos los procesos biológicos: para que el ADN pueda llevar cabo todos estos mecanismos, necesita transcribirse y traducirse, como resultado del primer proceso se obtendrá el ARN (ácido ribonucleico).

Existen varios tipos de ARN, sin embargo, al que le podemos atribuir cierta relevancia es al ARNm, también llamado como transcrito, la producción y regulación de transcritos, es el resultado de la respuesta de la célula ante estímulos intra o extra celulares, estos activan vías de cascadas de señalización para indicar cuales son los genes que se van a expresar, o bien, si así lo requiere la célula, que el gen se apague de acuerdo con el aliciente inicial. El primer paso de la expresión génica es la transcripción que se define como el proceso mediante el cual se genera ARNm a partir de la secuencia de ADN de un gen dentro del núcleo.

El proceso de transcripción comienza con la apertura del locus (sitio donde se ubica un gen), de la cromatina que contiene el gen de interés con ayuda de la enzima principal de este proceso: la ARN Polimerasa II (Pol II),

que además de separar las hebras de ADN, utiliza una de las cadenas como molde para sintetizar la cadena complementaria de ARN. Este proceso consta de tres etapas: iniciación, elongación y terminación (Figura 10.2).

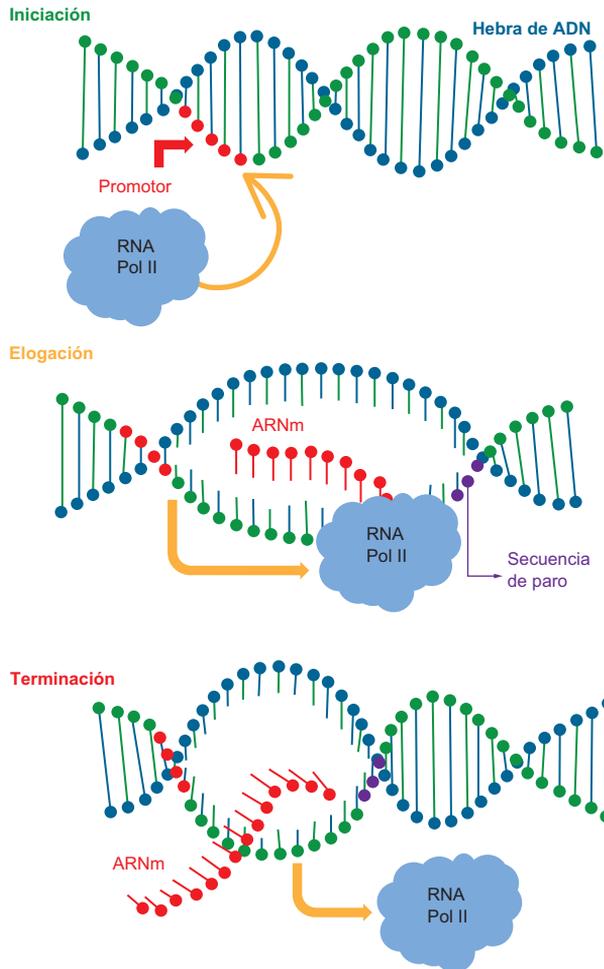


Figura 10.2. Proceso de transcripción. Principales etapas del proceso de transcripción de un fragmento de ARNm a partir de una hebra de ADN genómico.

- **Iniciación:** La Pol II, se une a la secuencia de ADN del gen de interés en una región llamada *promotor* al inicio del mismo. Posteriormente la Pol II procede a separar las cadenas de ADN y se pega a una de ellas para usarla como molde.
- **Elongación:** La Pol II comienza a “leer” la cadena molde del ADN y empieza a sintetizar su cadena complementaria en sentido 5’ a 3’ con la única diferencia de que contiene la base Uracilo en lugar de Timina. La cadena resultante se le llama pre-ARNm o transcrito ya que aun contiene *intrones* (secuencias no codificantes para aminoácidos) que deben de ser removidas antes de su salida del núcleo.
- **Terminación.** Cuando Pol II ha llegado al final de la secuencia de ADN del gen, encontrara un *terminador* que al contrario del *promotor* le indica que debe de terminar de “leer” la secuencia y separarse. Las bandas de ADN vuelven a unirse por afinidad de bases.

Una vez se ha llevado a cabo la transcripción, al pre-ARNm se le deben hacer una serie de modificaciones para asegurar la eliminación de los *intrones* y asegurar su integridad una vez este salga del núcleo. Estas modificaciones son la adición de las moléculas *cap* y *polyA* a los extremos del transcrito que ayudaran a protegerlo de nucleasas (Enzimas que degradan ácidos nucleicos) nativas de la célula. Por otra parte, tenemos el proceso de separación de *intrones* o *splicing* con el fin de que al momento de la traducción se genere una proteína funcional.

La primera modificación consta de la adición de un nucleótido de guanina al cual se le ha agregado un grupo metilo (CH_3) en la posición del carbono 7 mediante un proceso llamado metilación. Esta guanina se une al extremo 5’ de la cadena de pre-ARNm, mediante un enlace trifosfato 5’-5’. Posteriormente la enzima Poli A polimerasa, comienza a añadir en el extremo 3’ del pre-ARNm de 100 a 200 residuos de adenina que ayudan a proteger el ARNm de la degradación causada por nucleasas encontradas en el citoplasma.

La segunda modificación es la remoción de los *intrones* mediante el *splicing*, el cual comienza con el reconocimiento de los *intrones* por un sistema enzimático llamado *Spliceosoma* y posterior corte y eliminación de estas secuencias “basura”. Terminado este paso tenemos un ARNm maduro que puede ser transportado al citoplasma de la célula y seguir con el proceso de traducción.

La traducción como tal, se define como el proceso en el cual el ARNm maduro es leído en codones por los ribosomas en el citoplasma y traducido a una

secuencia de aminoácidos. Esta traducción está definida por el código genético que establece la relación que existe entre los codones de una secuencia y el aminoácido específico para el que codifica. Este mecanismo se lleva a cabo de igual manera en 3 pasos que son: iniciación, elongación y terminación (Figura 10.3).

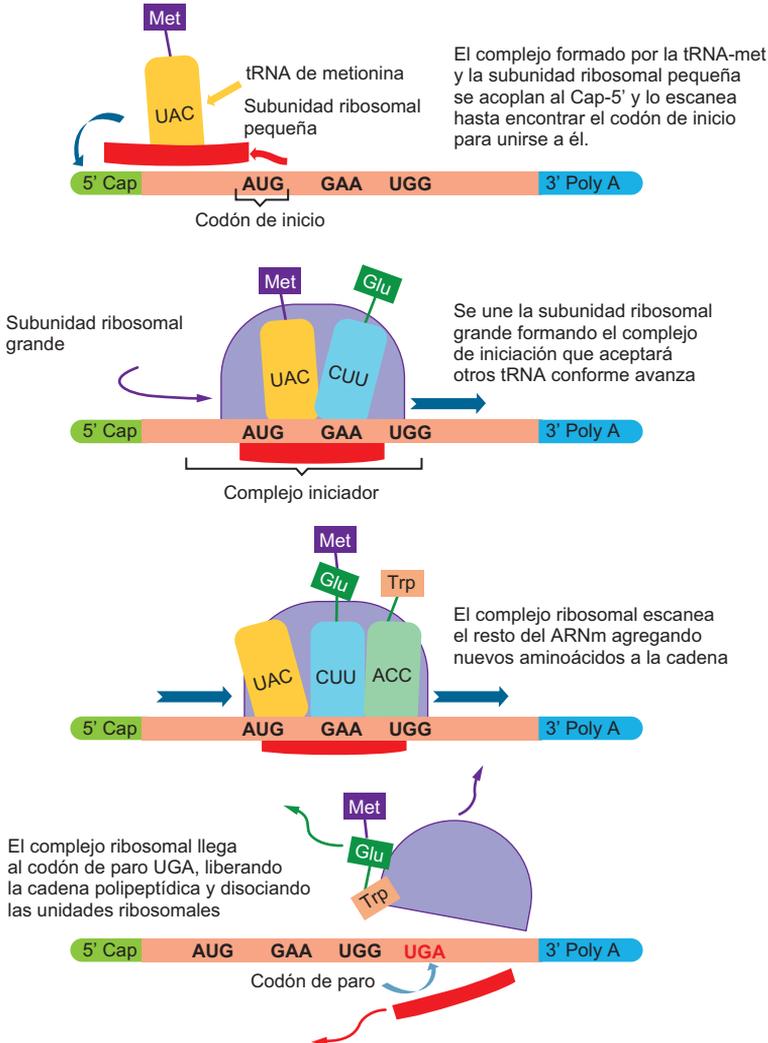


Figura 10.3. Proceso de traducción. Principales etapas de procesamiento de un ARNm maduro a proteína (Adaptación de Berg J. *et al.* (2002)).

- **Iniciación:** los ribosomas se acoplan alrededor del ARNm y comienzan a leerlo a partir del primer codón llamado de *iniciación* (AUG en eucariontes), posteriormente con la ayuda de una molécula de ARN de transferencia (ARNt) que tiene como función llevar el aminoácido codificante al codón leído, que en este caso siempre se trata de metionina (en eucariontes) y así comenzar con la traducción.
- **Elongación:** durante esta etapa los ribosomas van leyendo codón por codón y agregando con ayuda de los ARNt los aminoácidos correspondientes a la cadena de péptidos que inició con metionina.
- **Terminación:** al acercarse el final de la cadena de ARNm, los ribosomas encontrarán uno de los siguientes codones: UAA, UAG, UGA llamados *codones paro* que les indican a los ribosomas que ha llegado el final del transcrito y por tanto deben de detenerse y liberar la cadena de aminoácidos (polipéptido).

Métodos para analizar la expresión del gen

Existen múltiples herramientas moleculares que en conjunto ayudan a comprender y visualizar que genes y cuanto llegan a expresarse en diferentes órganos y/o bajo determinadas condiciones. La primera técnica de expresión fue el Northern blot; que consiste en la hibridación de una cadena de ARNm con un gen o fragmento de ADN específica llamado sonda, el cual está marcado radioactivamente. Esta técnica consiste en una electroforesis del ARNm en un gel desnaturalizante y posteriormente es transferido a una membrana de nitrocelulosa y posteriormente se hibrida con la sonda (gen marcado) sondas que se unen al ARN y por último se coloca una película de rayos X y posteriormente se revela la película y se observa en esta las manchas. Otra de las técnicas que fueron más utilizadas –y aún se siguen utilizando– es el RT-PCR. Esta técnica consiste en la extracción de ARNm y posteriormente es convertido a ADN complementario (ADNc) por la enzima transcriptasa reversa. El ADNc es utilizado como molde para la PCR donde se utilizan oligonucleótidos específicos del gen de interés.

En la actualidad tanto el RT-PCR como el Northern blot son poco utilizadas ya que existen más y mejores técnicas gracias a la secuenciación de nueva generación tales como: los transcriptomas, las librerías de ADNc y los análisis de PCR en tiempo real (qPCR). Esta última es una de las más importantes e indis-

pensables para el análisis de expresión ya que ya que permite con certeza medir que tanto se expresa un gen en un momento, condiciones y órgano específico. La teoría de esta técnica radica en la fluorescencia emitida cada que se lleva a cabo la síntesis de una nueva cadena de ADN, la cual es detectada por foto sensores incluidos en el equipo de qPCR como ya se mencionó en capítulos anteriores.

La base de la expresión génica y su regulación se basa principalmente en una serie de cambios y modificaciones que afectan directamente al ARNm, en diferentes etapas de su desarrollo. Estos cambios pueden ocurrir durante el *splicing* de los intrones, en las etapas traducción y postraduccional y sobre todo son motivo de pequeñas variaciones inducidas en el medio ambiente del organismo, como lo son cambios en la luz, temperatura, hidratación, pH del suelo, etc, etc.

Sin embargo, estos cambios pueden llegar a ser imperceptibles a simple vista, es por ello que se han desarrollado diferentes técnicas bioinformáticas como la secuenciación de nueva generación y la construcción de librerías de ADNc y transcriptomas que permite monitorear estos cambios.

Otra manera de analizar la expresión de un gen de interés es mediante la aparición y correlación con un microARN (miARN) característico, esto debido a que los miARN son elementos reguladores específicos de genes. Con esto en mente, se puede saber si un gen se expresa o no, analizando si el miARN que es específico para ese gen está presente o no. Incluso la relación que existe entre los niveles de expresión de un miARN y su gen objetivo ha sido utilizada para tener conocimiento en qué etapas, estreses o bien periodos de desarrollo del organismo un gen este encendido y en cuales debe ser apagado.

Transcriptomas

Un transcriptoma se define como la identificación y cuantificación total de moléculas de ARNm que están presentes en una célula, tejido, órgano u organismo bajo una condición y un momento dado, a los cuales se les conocen como transcritos, y la relación de estos con la expresión diferencial de cada uno de los genes que hacen posible dicha expresión.

En esencia un transcriptoma es la colección de ARNm producidos por una planta, en un tejido bajo tiempos y condiciones específicas.. De alguna forma se podría decir que es una foto exacta de los genes expresándose en ese momento determinado. Ahora bien, esto tiene más aplicaciones que saber

simplemente la producción total de ARNm de la planta, por ejemplo, al inducir un tratamiento deseado a otra planta de la misma especie y generar otro transcriptoma, ambos se comparan y se observan los cambios en los genes expresados. A este tipo de análisis se les conoce como: expresión diferencial de genes, el cual indica, si los genes expresados son los mismos o se expresaron otros genes o por el contrario dejaron de expresarse e incluso determinar que tanto se expresa un gen con respecto a su parte control. El transcriptoma también es una herramienta importante y casi indispensable para la obtención de datos valiosos correspondientes a la parte de regulación en la expresión de un gen determinado, puesto que no solamente se puede inducir cambios en la expresión de un gen con estímulos ambientales, sino también se puede utilizar una serie de herramientas moleculares para silenciar o regular la expresión de un gen y comparar estos niveles con los originales de la planta control. Estos mecanismos se basan principalmente del silenciamiento de genes mediante mutaciones con químicos, Knock Out por ARN antisentido o bien ARNs pequeños de interferencia (siARN), este último siendo una de las técnicas más populares y específicas que existen.

Se debe tomar en cuenta que no todos los genes se transcriben al mismo tiempo, sino que, cada gen se regula de forma individual, por lo que sólo se van a transcribir los genes que tienen relevancia en la respuesta a cierto estímulo y bajo cierta circunstancia (Fig. 10.4). También existen genes que pueden permanecer apagados y al haber otro estímulo o cambio de situación se va a requerir que se enciendan, iniciando así su propia transcripción. La colección de todos estos transcritos presentes en la célula se le llama Transcriptoma.

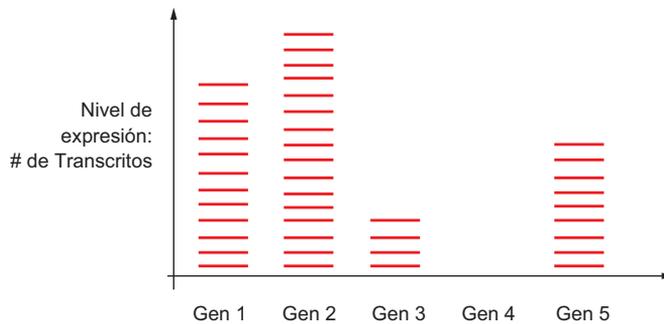


Figura 10.4. Ejemplificación de la diferencia de expresión entre distintos genes de un mismo individuo, bajo un mismo estímulo.

En la actualidad a través de la secuenciación de nueva generación (NGS, Next Generation Sequencing) y gracias a la implementación de la técnica de la secuenciación directa del ARN (ARN-seq), ha sido posible la caracterización y análisis completo de una célula sin ninguna información genómica previa, a lo que se le conoce como transcriptoma *De novo*, a un bajo costo y con muy altos rendimientos.

Pasos para la construcción de un transcriptoma

Para llevar a cabo la secuenciación de nueva generación con el fin de obtener un transcriptoma, se deben de seguir una serie de metodologías desde la extracción de ARN hasta el ensamble de las secuencias; la siguiente información se describe de manera sintetizada en la Figura 10.5.

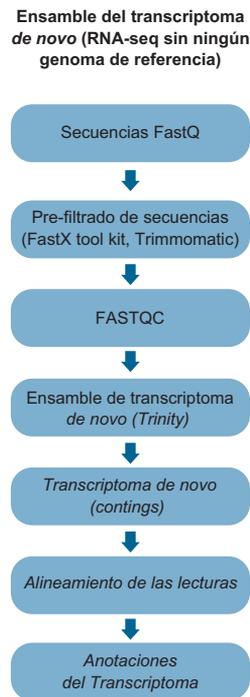


Figura 10.5. Diagrama en el que se muestran la metodología a seguir para realizar un transcriptoma *de novo*.

En este texto se describe sólo una de las aplicaciones de ARN-seq, el transcriptoma *De novo*, exponiendo algunas metodologías utilizadas y de las cuales se han obtenido resultados favorables (descripción de la Figura 10.5), el material biológico utilizado es de origen vegetal de la familia *Cactaceae*.

Paso 1: extracción de ARN total

El ARN total del material vegetal se extrajo con el reactivo TRIzol (Invitrogen), al final cada pastilla se resuspendió en 30 mL de agua DEPC. La pureza e integridad de las muestras se analizaron en el equipo Agilent 2100 Bioanalyzer ARN 6000 Nano Assay chip (Agilent Technologies, Stockport). Se sintetizó el ADNc y con este se realizaron las bibliotecas.

Paso 2: Generación de ARN-seq

Existen varias plataformas y estrategias para la secuenciación directa del ARN (ARN-seq), entre los cuales se pueden clasificar en tres tipos principales: 1) Secuenciación por Síntesis; donde realiza una librería de ARN, se ligan adaptadores a cada una de las secuencias, se amplifican para mejorar la señal fluorescente y colocan en celdas de flujo para su posterior secuenciación. Secuenciadores más utilizados en esta técnica: *Roche 454 pyrosequencing*, *Ion Torrent* y la más utilizada *Illumina*. 2) Secuenciación por Ligación: Este método usa oligonucleótidos en presencia de fluorescencia; como ejemplo esta la técnica de *SOLiD* y *polonator*, y 3) Secuenciación de molécula única: Produce una señal de oligonucleótidos detectada por quimioluminiscencia, entre las técnicas más comunes se encuentran: *Helicos* y *Pacific Biosceinces*.

Paso 3.- Secuenciación en plataformas Illumina

En las plataformas *Illumina* las bibliotecas de ADNc son diluidas y cargadas en una celda de flujo, aquí se lleva a cabo una amplificación por PCR-puente. Esta PCR se realiza con el fin de formar un *cluster* anclado a una base sólida, se lleva a cabo por nucleótidos marcados por fluorescencia, estos nucleótidos hacen el papel de cebadores creando puentes facilitando la amplificación elongando la cadena, el nucleótido marcado es excitado y su manifestación lumínica es captada por un sistema óptico. Los resultados obtenidos de cada

cluster son procesados para formar un *calling* y así producir las lecturas, esta metodología se realiza en ambos sentidos *forward* y *reverse* conocidas como secuencias *paired-end* (Fig. 10.6).

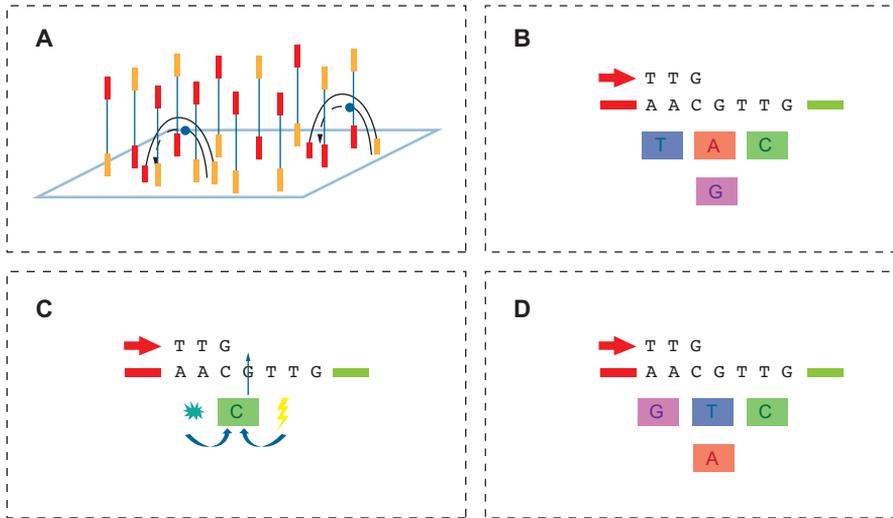


Figura 10.6. Secuenciación por síntesis masivo ARN-seq, en la plataforma *Illumina*. El proceso utiliza el principio de amplificación en puente y el marcaje por fluorescencia de terminadores reversibles. A) Los adaptadores de los extremos de las secuencias ligándose por complementariedad a cebadores adheridos a una “Flow cell”, los amplicones permanecen de esta manera hasta la próxima desnaturalización en la cual permitirán la formación de un nuevo puente y así dar pie a la amplificación. B) el amplicón se genera a partir de oligonucleótidos únicos para los adaptadores, en cada flujo los cebadores con fluorescencia elongan la cadena. C) se unen las bases complementarias a la cadena por una fuente lumínica excitando el fluoróforo del nucleótido añadido. D) el resto de los nucleótidos son eliminados y el ciclo se repite nuevamente.

Paso 4.- Control de calidad de las secuencias obtenidas

A) Análisis con el programa FASTQC

Se realiza el análisis de la secuencia obtenida con la plataforma *Illumina* mediante el programa *FastQC*, el cual permite analizar las calidades de las secuencias obtenidas.

B) Fastq_quality_filter

Posteriormente se realiza un filtrado de las secuencias con la herramienta `Fastq_quality_filter`, esto a partir del gráfico de la “calidad de secuencia por base”, que nos muestra el promedio y el rango de los valores de calidad de secuencia a lo largo de la lectura, esto por medio del comando:

```
$fastq_quality_filter -v Q 64 -p 32 -q 80 -i *MB1R1_trimmer.Fastq -o
*MB1R1_filter.fastq
```

Donde:

- i: input file
- o: output file
- v: report number of sequences
- Q 64: determines the input quality ASCII offset (in this case 64).
- q 32: the quality value required.
- p 80: the percentage of bases that have to have that quality value

A las secuencias obtenidas se les vuelve a verificar la calidad mediante el programa `FASTQC`

C) Ensamble de las secuencias

Para el ensamble de las secuencias a partir de los datos de ARN-seq sin un genoma de referencia, se usó el programa `Trinity`, el cual se compone de 3 módulos: 1) `Inchworm` el cual reconstruye eficientemente secuencias específicas consecutivas unas a otras y superpuestas (contigs) de transcritos lineales, 2) `Chrysalis` que superpone los contigs en conjuntos de componentes conectados, y construye gráficos completos de `Bruijn` para cada componente y 3) `Butterfly` que reconstruye transcritos aceptables de longitud completa al comparar los gráficos individuales de `Bruijn` generados en el módulo anterior con las lecturas originales y a partir de esto reconstruir transcripciones distintas para empalmar isoformas y genes parálogos (Figura10.7).

relacionados con el estrés, el de *Hylocereus undatus*, se realizó un ensamble *de novo* y un análisis diferencial en genes relacionados con la inducción floral y en *Mammillaria bombycina* se realizó un ensamble *de novo* para dilucidar algunos genes implicados en la homeostasis de la cactácea bajo condiciones de estrés abiótico.

Los miARN y la regulación genética

El mecanismo de silenciamiento o regulación genética mediante el uso de ARNs pequeños (ARNp) se ha vuelto increíblemente popular debido a que ha sido identificado como uno de los mecanismos naturales de regulación mayormente conservado entre los eucariontes. Los principales tipos de ARN que han tenido mayor auge son los ARN de interferencia (ARNi), los ARN pequeños de interferencia (ARNsi) y micro ARN (miARN).

Los miARN son fragmentos de ARN de entre 18 a 22 nucleótidos, no codificante que actúan de manera específica regulando la expresión de genes, estos se encuentran en el genoma de todas las células eucariontes. Estos se sintetizan a partir de regiones en forma de bucle provenientes de precursores de transcritos endógenos, estas regiones se conocen como pre-miARN, las cuales posteriormente forman un complejo dúplex de ARN por la acción de una endoribonucleasa III (Dicer-like1). El miARN maduro se une a la proteína argonauta (AGO1), formando un complejo proteico de silenciamiento de genes inducido por complementariedad de ARN llamado RISC que causa la degradación de un ARNm o bien la represión transcripcional (Fig.10.8).

La manera en que actúa este sistema proteico llamado RISC, por sus siglas en inglés de *ARN-induced silencing complex*, comienza al unirse a la secuencia de miARN y llevarla con precisión al gen objetivo, una vez ahí tiene dos mecanismos de acción: el primero se basa en la regulación postranscripcional, en donde el complejo RISC se une al gen objetivo y este lo corta como una enzima de restricción, y el segundo que es mediante el bloqueo durante la transcripción de ADN a ARNm, ya que el complejo RISC se pega a la cadena de ADN evitando que la Pol II actúe.

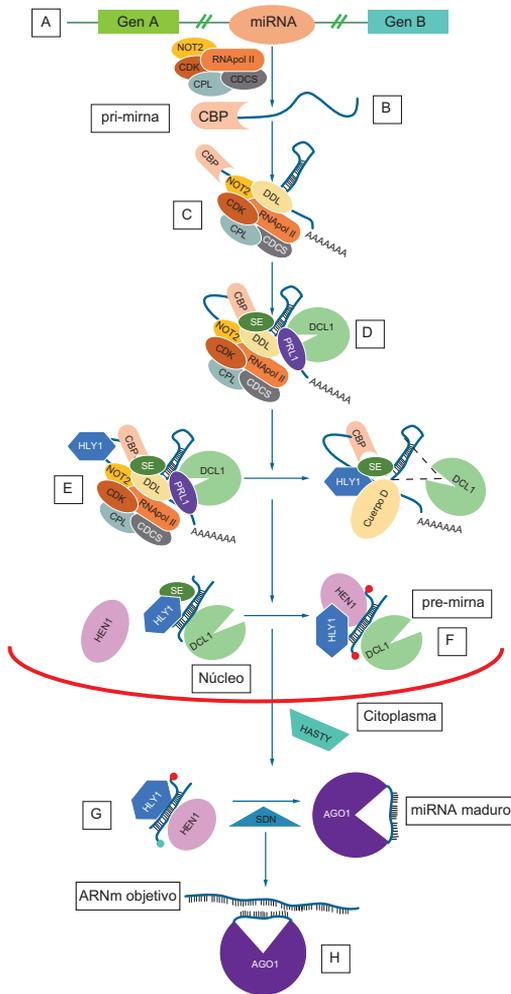


Figura 10.8. Biosíntesis de miARN en plantas. A) Los miARN de plantas son transcritos a partir de secuencias no codificantes entre genes por la *ARN polimerasa II* (ARNPol II) con ayuda de los cofactores NOT2, CDK, CPL y CDC5 generando un producto llamado pri-miARN. B) El pri-miARN es modificado en su extremo 5' por la *Cap Binding Protein* (CPB). C) El complejo ARNPoII pliega la hebra de ARN consigo misma atrayendo a la CPB y formando el bucle característico de los miARN. Esta estructura lleva el nombre de cuerpo D (D-bodies) y es estabilizada por la proteína DAWDLE (DDL). D) La reacción anterior genera el plegamiento de los factores PRL1 y SE que ayudan al plegamiento de la enzima *Dicer Like-1* (DCL1) al complejo ARNPoII. E) La enzima HLY1 fosforila los extremos del bucle del pri-miARN activando la acción de corte de DCL1 (líneas punteadas). F) Una vez cortado el bucle la estructura recibe el nombre de pre-miARN, el cual es metilado en sus extremos (puntos rojos) por la enzima HEN1 que desplaza al factor SE y deja al pre-miARN listo para su exportación al citoplasma con ayuda de la enzima HASTY. G) Uno de los extremos del pre-miARN pierde su grupo metilo y es degradado por la nucleasa *ARN Degrading Nuclease* (SDN). La hebra restante es miARN maduro que posteriormente se une a una proteína AGO1 formando el complejo de silenciamiento dirigido por ARN, RISC. H) El complejo RISC es llevado hasta el ARNm objetivo para su regulación (Adaptación de Voinnet O. (2009); Achkar N. *et al.* (2012))

En plantas se ha observado que las secuencias de miARN tienen una alta complementariedad con su gen objetivo, con máximo cinco diferencias entre las secuencias de miARN y gen objetivo y aun así tener actividad catalítica sobre el gen.

Como ejemplo, en el 2005 un grupo de investigadores liderado por el Dr. Schwab, demostró la actividad del miARN 159, sobre un gen objetivo con cinco diferencias entre sus secuencias, resultando en la regulación negativa del gen de interés. Además de esto en el 2015 un grupo de investigadores, demostró que, aunque no sea un objetivo específico y conservado evolutivamente entre especies, se da la regulación, lo que indica, en conjunto con el cambio evolutivo de ARNm, el miARN cambia a la par. Estos hallazgos resaltan el hecho de que miARN maduros, así como las secuencias de sus genes objetivos están altamente relacionadas y conservadas entre diferentes especies de plantas. Esto se ha logrado mediante la secuenciación de ARN, herramienta con la cual se pueden caracterizar e identificar los ARNp expresados en una planta.

En años recientes se han descubierto que muchas de las secuencias de miARN en plantas están conservadas entre especies, sobre todo los correspondientes a estrés hídrico. Sin embargo, en otros estudios también otros tipos de miARN se han descrito y encontrado entre especies por ejemplo, los responsables de la regulación de genes de maduración en frutos climátricos. Esto genera la posibilidad a que un perfil completo de miARN de una planta puede ser comparado con un transcriptoma de la misma planta y así encontrar posibles genes objetivos de los miARN.

En este capítulo se explicará a detalle la manera para construir y analizar una biblioteca y perfil completo de miARN en una planta, además de ejemplificar su desarrollo con material vegetal del fruto del guayabo.

Construcción de Bibliotecas de miARN

a) Cuantificación y verificación de la integridad de las muestras

Utilizando la metodología de extracción de ARN mencionada anteriormente, se trabajó con muestras de fruto de guayaba en estadio maduro. Se calcula la concentración de las muestras y se realiza una electroforesis en un gel desnaturalizante de poliacrilamida, para comprobar la integridad y calidad del ARN (Fig. 10.9).

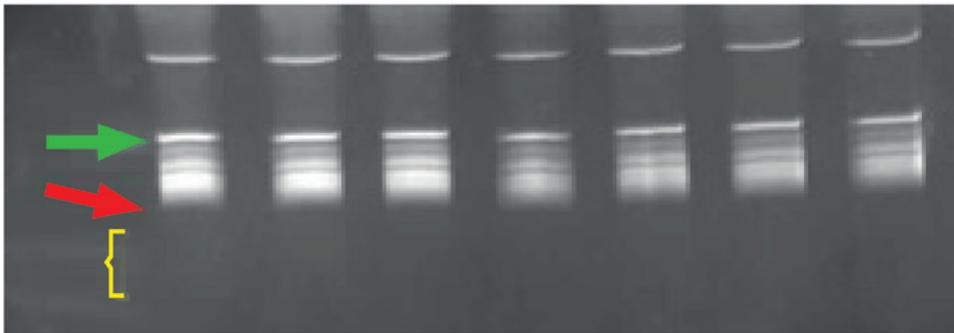


Figura 10.7. Extracción de ARN total. Flecha verde: banda correspondiente a la subunidad ribosomal 28 S. Flecha roja subunidad 18 S ribosomal. Corchete amarillo, localización de los ARNp, ARNt, miARN.

b) Aislamiento de microARNs a partir de muestras de ARN total

Se cortan las áreas del gel de poliacrilamida donde se encuentran las bandas correspondientes a 20 y 27 pares de base del marcador de peso molecular. Posteriormente los cortes se dividen en 6 microtubos y se les agrega 700 μL de agua tratada con DEPC, estéril y se dejan eluir durante toda la noche a 5 $^{\circ}\text{C}$.

Las muestras eluidas junto con la poliacrilamida se recuperan con una jeringa de 5 mL conectada a un filtro estéril el cual impide el paso de partículas no deseadas de poliacrilamida, pero sí la recuperación del sobrenadante. Las muestras fueron colectadas en un tubo nuevo de 1.5 mL, se afora con agua destilada estéril hasta un volumen de 500 μL por tubo. A continuación, los

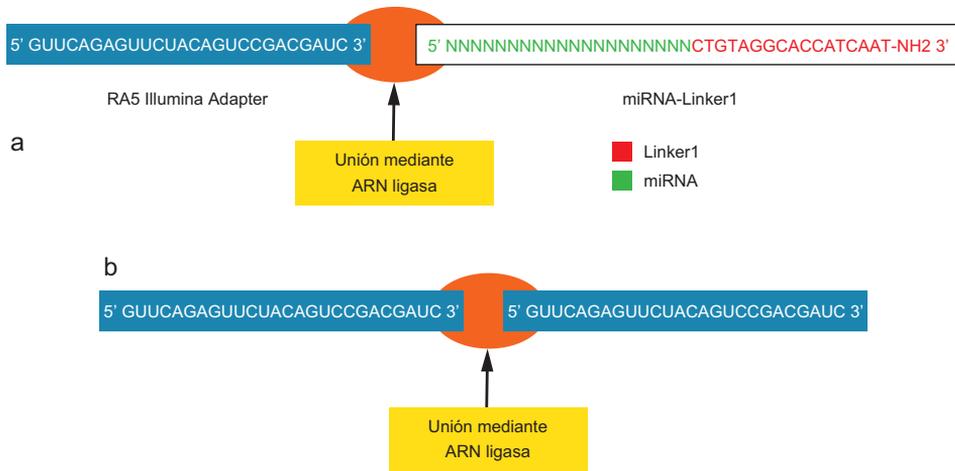


Figura 10.11. Ligación del adaptador 5'. a) Unión deseada entre la primera ligación y el RA5 en el extremo 5'. b) Producto vacío o no deseado, en donde dos adaptadores RA5, se pliegan en sus extremos 3' y 5'.

e) Síntesis de ADN complementario del producto de las ligaciones

Para este paso, se puede utilizar cualquier kit para síntesis de ADNc. Sin embargo, hay que tener en cuenta que se necesita introducir un oligonucleótido especialmente diseñado para ser la reversa complementaria exacta del adaptador 3', así como del adaptador 5', el cual, para propósitos demostrativos, elegimos el Illumina adapter, esto con la finalidad de tener un oligonucleótido que sea capaz de también transcribir esta parte a ADNc, como se muestra en la figura 10.12.

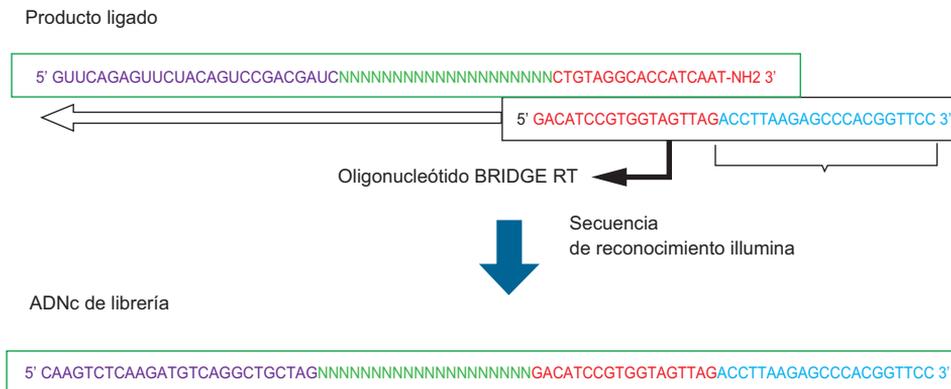


Figura 10.12. Síntesis de ADNc de la librería de miARN. Estructura del oligonucleótido BRIDGE RT y síntesis de ADN complementario (ADNc) de los productos ligados anteriormente.

f) Asignación de índices para las librerías de miARN

Los índices son secuencias específicas de 6 pares de base que se colocan cerca del extremo 3' de la construcción y se utilizan como identificación de cada librería en el momento de la secuenciación. Para este ejemplo, se utilizaron los índices de reconocimiento Illumina, del kit Illumina Adapter Sequences, TruSeq Small ARN, de la serie C, se usaron los números 2, 4, 5, 6, 11 y 12 (Tabla 10.1). Los índices, fueron agregados a la construcción mediante PCR, los pasos descritos se mencionan en el siguiente punto.

Tabla 10.1. Secuencias de los índices Illumina utilizados.

Índice	Secuencia
C2	CGATGT
C4	TGACCA
C5	GTGACA
C6	GCCAAT
C11	CTTGTA
C12	TACAAG

Esto concluye la construcción de nuestras librerías de miARN. Sin embargo, el trabajo apenas comienza, ya que después de esto se manda a secuenciar las librerías y posteriormente analizarlas bioinformática y estadísticamente. En esta parte, existen un gran número de programas de libre acceso y de paga que podrán facilitarnos estos pasos, uno de los más recomendados por los autores de este libro es el CLC Genomics Workbench, en cualquiera de sus versiones, debido a la gran cantidad de herramientas y amigable interfaz para utilizarlas.

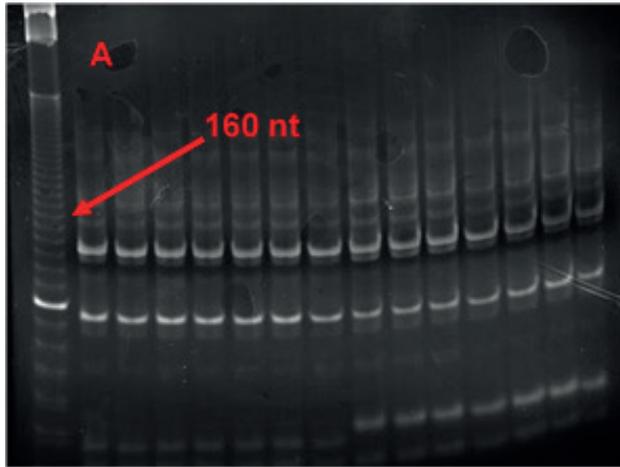


Figura 10.14. Electroforesis de RT-PCR de librerías de miARN. Flecha roja: banda de 160 nt correspondiente a una librería de miARN correcta.

Los miARN en el fruto de guayaba

Estudios actuales han identificado, analizado y perfilado los miARN expresados durante la maduración en el fruto de la guayaba mediante el uso de herramientas bioinformáticas y técnicas de secuenciación masiva. Para este caso, se realizó un análisis de miARN en el fruto de guayaba mexicana “Siglo XXI” variedad media china de la cual, a partir del aislamiento de ARN total, se construyeron 2 bibliotecas de miARN (utilizando el protocolo antes mencionado) una para miARNs encontrados en estadio rayado y otra para el estadio maduro. La secuenciación inicial de las librerías resultó en 16,729,248 secuen-

cias, a las que después de remover el adaptador y pasar por un filtro de tamaño de entre 18 y 22 pb, disminuyeron su número a 244,277. Estas secuencias fueron sometidas a un análisis de similitud con el genoma de guayaba y la base de datos de la página web miRBase obteniendo 3909 secuencias con similitudes, de las cuales solo se tomaron en cuenta las pertenecientes a miARN maduros, resultando en 77 secuencias de miARN. Posteriormente de estas secuencias, se identificaron 29 familias diferentes de miARN. En el estadio maduro se observó la expresión de 27 de estas familias, mientras que en rayado sólo 17. De las 29 familias la MIR166 fue la que expresó un mayor número de miembros con 14 en rayado y 9 en maduro, seguida de la familia MIR482 con 8 en maduro y 7 en rayado, y en tercer lugar la MIR 395 con 7 en maduro y 6 en rayado.

El tamaño de los miARN está entre 18 y 25 nt, siendo el de 21 nt (84 secuencias) el más abundante, seguido de 22 nt (24 secuencias), 20 nt (12 secuencias), del resto se encontró una secuencia de 25 nt, una de 23 nt, una de 19 nt y una de 18 nt.

Las familias 159, 160, 166, 171, 319 y 535 mostraron estar directamente relacionados al proceso de maduración (biosíntesis de etileno y modificación de pared celular), mientras que las familias 169, 395, 398, 528 y 2120 mostraron una relación indirecta (factores de crecimiento y enzimas secundarias).

Protocolo

Protocolo 10.1 Extracción de ARN

Se extrajo ARN total del tejido expuesto a condiciones de estrés utilizando el kit PureZol de BioRad. Se pesaron 100 mg de tejido fresco pulverizado y se le agregó 1 mL de PureZol, se homogenizó y se dejó incubar por 24 h a 4 °C, al término de ese tiempo se le agregaron 0.2 mL de cloroformo mezclándolo vigorosamente por 15 min, se dejó incubar a temperatura ambiente por 24 h, se centrifugó a 12000 rpm por 15 min a 4 °C, se transfirió la fase acuosa a un tubo nuevo y se le adicionaron 0.5 mL de alcohol isopropílico, se incubó por 5 min a temperatura ambiente por 5 min y se centrifugó a 12000 rpm por 15 min a 4 °C, se descartó el sobrenadante y se lavó el pellet adicionando 1 mL de etanol al 70 %, se agitó vigorosamente en vortex y se centrifugó a 7500 rpm.

por 5 min a 4 °C, se adicionaron 50 µL de agua libre de ARNs (agua tratada con DEPC) y se almacenó a -80 °C.

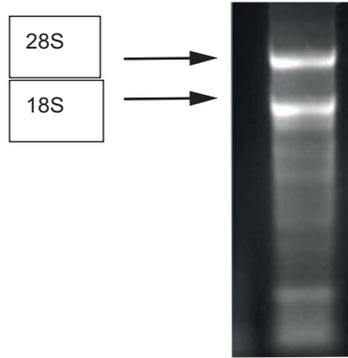


Figura 10.15. ARN total de *Suaeda edulis*. Se observa la presencia de dos bandas correspondientes a las subunidades 28S y 18S del ARNr de *Suaeda edulis*.

Extracción de ARN de guayaba mediante un método CTAB modificado.

Todas las muestras utilizadas (hoja, tallo, raíz y fruto en estadios de maduración verde, rayado, maduro y sobre maduro) siguieron la misma metodología para la extracción de ácidos nucleicos. La extracción fue con el método de CTAB de Doyle y Doyle (1987), con modificaciones. El buffer de lisis contuvo lo siguiente, CTAB 3 %, PVP 2 %, EDTA pH 8.0 al 25 mM, Tris-HCl pH 8.0 al 100 mM, NaCl 2 M y agua destilada hasta alcanzar un volumen de 10 mL. El material vegetal se pulverizó con nitrógeno líquido en mortero y se le adicionó inmediatamente 700 µL buffer de lisis precalentado a 65 °C, la solución se mezcló vigorosamente en vórtex y se incubó a 65 °C durante 5 min (agitando constantemente). Posteriormente se añadió 1 volumen de cloroformo–alcohol isoamílico 49:1, agitando vigorosamente durante 20 s y centrifugando a 13,000 rpm durante 10 min. Seguido se recuperó la parte acuosa y pasó a un microtubo nuevo, añadiendo 1 volumen de fenol–cloroformo 1:1, seguido de una centrifugación a 13,000 rpm durante 10 min, al finalizar se recuperó la fase acuosa en un tubo nuevo y se añadieron 2 volúmenes de etanol absoluto y 0.3 volúmenes de acetato de sodio 3 M, la mezcla se incubó a -20 °C durante toda la noche. Se continuó centrifugando por 15 min a 13,000 rpm, se decantó todo

el sobrenadante teniendo cuidado de no despegar la pastilla del fondo. Para los lavados de la pastilla se añadió 1 mL de etanol al 80 %, agitando vigorosamente, posteriormente se centrifugó a 10,000 rpm durante 5 min, se decantó el etanol y la pastilla se dejó secar parcialmente, finalmente se resuspendió en 30 μ L de agua tratada con DPEC y se guardó a -20 °C hasta su uso. Se comprobó la integridad y calidad del ARN mediante una electroforesis en geles de agarosa al 1 % a 90 volts durante 40 min utilizando TAE como buffer de carga.

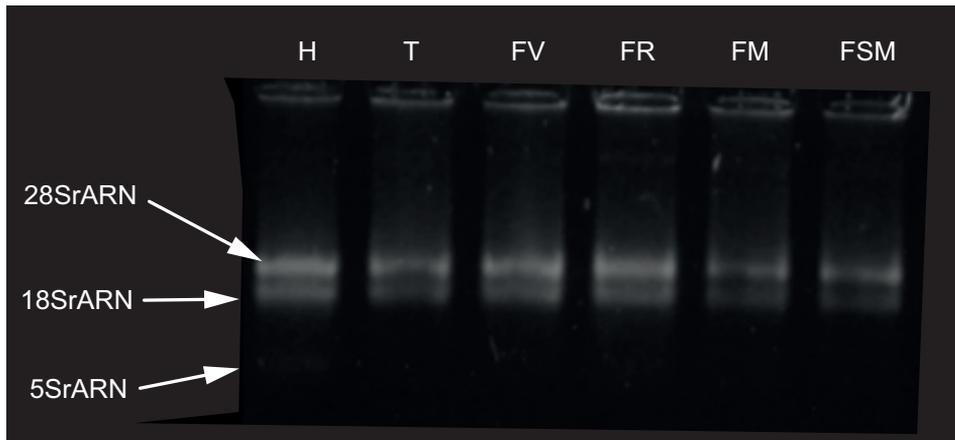


Figura 10.16. Muestras de ARN de diferentes tejidos de la guayaba. H) Hoja, T) Tallo, FV) Fruto verde, FR) Fruto rayado, FM) Fruto maduro, FSM) Fruto sobre maduro.

Glosario

Codón: grupo de 3 nucleótidos que codifican para un aminoácido específico.
ARNm: tipo de ARN de cadena única que participa en la síntesis de proteínas.
ARN Polimerasa II: Enzima encargada de “leer” la cadena de ADN y formar ARNm en el proceso de transcripción.

De novo: Que se origina

Cromatina: ADN enrollado en proteínas que conforma a los cromosomas

Promotor: Secuencia de ADN que da inicio al proceso de transcripción.

Terminador: Secuencia de ADN que finaliza el proceso de transcripción.

Polipéptido: Cadena de aminoácidos

Splicing: Proceso de remoción de intrones de un ARNm

Bucle: Que da una vuelta, tiene forma de broche de pelo.

Adaptador: Secuencia nucleotídica específica de reconocimiento para sistemas de secuenciación Illumina.

Adenilación: Proceso mediante el cual se agrega un grupo adenilo del adenosintrifosfato a una molécula aceptora.

Dideoxicitidina: Nucleósido obtenido a partir de la remoción de los grupos OH⁻ de una citosina.

Abreviaturas

ARN: ácido ribonucleico

ARNm: ácido ribonucleico mensajero

ARNp: ácido ribonucleico pequeño

ARNi: ácido ribonucleico de interferencia

ARNip: ácido ribonucleico de interferencia pequeño

miARN: ácido ribonucleico micro

ADN: ácido desoxirribonucleico

ADNc: ácido desoxirribonucleico complementario

Pol II: ARN Polimerasa II

tARN: ácido ribonucleico de transferencia

PCR: Reacción en cadena de la polimerasa

RT-PCR: Retro transcripción de la Reacción en cadena de la polimerasa

qPCR: Reacción en cadena de la polimerasa en tiempo real

AGO1: proteína argonauta 1

RISC: ARN-induced silencing complex

DEPC: dietil pirocarbonato

ATP: Adenosín trifosfato

RPM: revoluciones por minuto

ARNasa: ácido ribonucleico endonucleasa

CTAB: bromuro de cetrimonio

EDTA: ácido tetraacético etildiamina

HCl: ácido clorhídrico

NaCl: Cloruro de sodio

TAE: Tris-base-ácido acético-EDTA

Bibliografía

- Achkar, P., Natalia, Damián, A., Cambiagno, Pablo, A., Manavella. (2016). miRNA Biogenesis: A Dynamic Pathway. *Trends in Plant Science*, 21(12), 1034-1044.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. & Walter, P. (2002). Posttranscriptional controls. *Molecular Biology of the Cell* (4ta ed.). Nueva York, NY: Garland Science.
- Bai, Selinge, Yunyun, Tian, Chao, Tan, Shunbuer, Bai, Jinfeng, Hao, & Agula, Hasi. (2020). Genome-Wide Identification of MicroRNAs Involved in the Regulation of Fruit Ripening and Climacteric Stages in Melon (*Cucumis Melo*). *Horticulture Research*, 7(1).
- Berg, J., M., Tymoczko, J., L. & Stryer, L. (2002). Eukaryotic protein synthesis differs from prokaryotic protein synthesis primarily in translation initiation. In *Biochemistry*. (5th ed.). New York, NY: W. H. Freeman.
- Bi, Fangcheng, Xiangchun, Meng, Chao, Ma, & Ganjun, Yi. (2015). Identification of MiRNAs Involved in Fruit Ripening in Cavendish Bananas by Deep Sequencing. *BMC Genomics*, 16(1), 1–15.
- Carbone, Fabrizio, Leonardo, Bruno, Gaetano, Perrotta, Maria, B., Bitonti, Innocenzo, Muzzalupo, & Adriana, Chiappetta. (2019). Identification of MiRNAs Involved in Fruit Ripening by Deep Sequencing of *Olea europaea* L. Transcriptome. *PLOS ONE*, 14(8), 1–22.
- Cris, Kuhlemeier. (1992). Transcriptional and post-transcriptional regulation of gene expression in plants. *Plant Molecular Biology*, 19, 1-14.
- Egan, A., J., Schlueter & D., Spooner. (2012). Applications of Next-Generation Sequencing in Plant Biology. *American Journal of Botany*, 99(2), 175-185.
- Felipe, F., Nick, M., Seung, Y. & Rodrigo, G., (2008). The rules of gene expression in plants: Organ identity and gene body methylation are key factors for regulation of gene expression in *Arabidopsis thaliana*. *BMC Genomics*, 9, 438.
- Guo, Da Long, Qiong, Li, Wen, Q., L, Guo, H., Zhang, & Yi, He, Yu. (2018). MicroRNA Profiling Analysis of Developing Berries for ‘Kyoho’ and Its Early-Ripening Mutant during Berry Ripening. *BMC Plant Biology*, 18(1), 1–16.
- Han, Yuanfang, Guosong, Chen, Heying, Zhou, Qing, Zhang, Ling, Qin, & Yu, Xing. (2019). Identification of Novel and Conserved MicroRNAs Invol-

- ved in Fruit Development and Ripening in *Fragaria vesca*. *Acta Societatis Botanicorum Poloniae*, 88(2), 1–13.
- Hou, Yanming, Lulu, Zhai, Xuyan, Li, Yu, Xue, Jingjing, Wang, Pengjie, Yang, Chunmei, Cao, Hongxue, Li, Yuhai, Cui, & Shaomin, Bian. (2017). Comparative Analysis of Fruit Ripening-Related MiRNAs and Their Targets in Blueberry Using Small RNA and Degradome Sequencing. *International Journal of Molecular Sciences*, 18(12).
- Jannatul, F., Syed, S., Hussian & Bu-Jun, Shi. (2015). Role of microARN in plant drought tolerance. *Plant Biotechnology Journal*, 13, 293-305.
- Johana, C., Soto, S. & Camilo, E., López, C. (2012). ARN-seq: herramienta transcriptómica útil para el estudio de interacciones planta patógeno. *Fitosanidad*, 16(2), 101-113.
- Kim, V., Narry, & Jin, Wu, Nam. (2006). Genomics of MicroRNA. *Trends in Genetics*, 22(3), 165–73.
- Li, Fangfang, Weidi, Wang, Nan, Zhao, Bingguang, Xiao, Peijian, Cao, Xingfu, Wu, Chuyu, Ye, *et al.* (2015). Regulation of Nicotine Biosynthesis by an Endogenous Target Mimicry of MicroRNA in Tobacco. *Plant Physiology*, 169(2), 1062–71.
- Marguerat, S. & J., Bahler. (2010). ARN-seq: from technology to biology. *Cellular and Molecular Life Sciences*, 67, 569-579.
- Metzker, M. (2010). Sequencing technologies the next generation *Nature Reviews Genetics*, 11, 31-46.
- Millar, A., Allan, Lohe, & Gigi, Wong. (2019). Biology and Function of MiR159 in Plants. *Plants*, 8(8).
- Onik, Jakaria, Chowdhury, Xiaoja, Hu, Qiong, Lin, & Zhidong, Wang. (2018). Comparative Transcriptomic Profiling to Understand Pre- and Post-Ripening Hormonal Regulations and Anthocyanin Biosynthesis in Early Ripening Apple Fruit. *Molecules*, 23(8), 1–19.
- Park, Y., S., Kim, S., K., *et al.* (2019). The transcriptome analysis of the *Arabidopsis thaliana* in response to the *Vibrio vulnificus* by RNA-sequencing. *PLOS ONE*, 14(12), e0225976.
- Proudfoot, N., A., Furger & M., Dye. (2002). Integrating ARNm Processing with Transcription. *Cell*, 108, 501-512.
- S., Gharat & P., Shaw. (2016). Computational prediction and experimental validation of a novel miARN in *Suaeda maritima*, a halophyte. *Genetics and Molecular Research*, 15.

- Sarwar, M., B., Ahmad, Z., & Rashid, B. (2019). *De novo* assembly of *Agave sisalana* transcriptome in response to drought stress provides insight into the tolerance mechanisms. *Sci Rep*, 9, 396.
- Schwab, Rebecca, Javier, F., Palatnik, Markus, Riester, Carla, Schommer, Markus, Schmid, & Detlef, Weigel. (2005). Specific Effects of MicroRNAs on the Plant Transcriptome. *Developmental Cell*, 8(4), 517–27.
- Sun, L., Wang, J., Song, K., *et al.* (2019). Transcriptome analysis of rice (*Oryza sativa* L.) shoots responsive to cadmium stress. *Sci Rep*, 9, 10177.
- Tsuzuki, Masayuki, Ryuichi, Nishihama, Kimitsune, Ishizaki, Yukio, Kurihara, Minami, Matsui, John, L., Bowman, Takayuki, Kohchi, Takahiro, Hamada, & Yuichiro, Watanabe. (2016). Profiling and Characterization of Small RNAs in the Liverwort, *Marchantia Polymorpha*, Belonging to the First Diverged Land Plants. *Plant and Cell Physiology*, 57(2), 359–72.
- Voinnet, Olivier. (2009). Origin, Biogenesis, and Activity of Plant MicroRNAs. *Cell*, 136(4), 669–87.
- Wang, Z., M., Gerstein & M., Snyder. (2009). ARN-Seq: A Revolutionary Tool for Transcriptomics. *Nature Reviews, Genetics*, 10, 57–63.
- Wu, Juxun, Saisai, Zheng, Guizhi, Feng, & Hualin, Yi. (2016). Comparative Analysis of MiRNAs and Their Target Transcripts between a Spontaneous Late-Ripening Sweet Orange Mutant and Its Wild-Type Using Small RNA and Degradome Sequencing. *Frontiers in Plant Science*, 7, 1–17.
- Xin, Chengqi, Wanfei, Liu, Qiang, Lin, Xiaowei, Zhang, Peng, Cui, Fusen, Li, Guangyu, Zhang, *et al.* (2015). Profiling MicroRNA Expression during Multi-Staged Date Palm (*Phoenix Dactylifera* L.) Fruit Development. *Genomics*, 105(4), 242–51.
- Zeng, Shaohua, Yongliang, Liu, Lizhu, Pan, Alice, Hayward, & Ying, Wang. (2015). Identification and Characterization of MiRNAs in Ripening Fruit of *Lycium Barbarum* L. Using High-Throughput Sequencing. *Frontiers in Plant Science*, 6, 1–15.