

RNA-seq: herramienta transcriptómica útil para el estudio de interacciones planta-patógeno

Johana Carolina Soto Sedano¹ y Camilo Ernesto López Carrascal²

¹ Universidad Nacional de Colombia, sede Bogotá, Carrera 45 No. 26-85, Facultad de Agronomía-Grupo de Investigación en Fitopatología Molecular, Facultad de Ciencias, jcsotos@unal.edu.co

² Universidad Nacional de Colombia, sede Bogotá, Carrera 45 No 26-85 - Edificio Uriel Gutiérrez. Grupo de investigación en Fitopatología Molecular. Facultad de Ciencias, departamento de Biología celopezc@unal.edu.co

RESUMEN

El conocimiento del transcriptoma y su regulación es fundamental para la interpretación articulada de los diversos constituyentes moleculares que integran la red de respuesta génica ante un determinado evento inductor, como los que se presentan en interacciones planta-patógeno. La actual tecnología de secuenciación ha llevado al desarrollo del RNA-seq como herramienta transcriptómica que permite el secuenciamiento masivo de ADNc o ARN, y hace posible obtener perfiles de expresión génica de las respuestas de defensa, lo cual ofrece grandes posibilidades para profundizar en el entendimiento de los mecanismos que se activan durante las respuestas inmunes en plantas. El RNA-seq ha cambiado la manera de cómo se estudian los transcriptomas, y ha permitido identificar y cuantificar nuevos y conocidos transcriptos relacionados con defensa vegetal. Aquí se presenta el principio, aplicaciones y ventajas del RNA-seq; además se discuten trabajos recientes que revelan la importancia y utilidad de esta herramienta en estudios de interacciones planta patógeno.

Palabras claves: secuenciación, ADNc, fitopatología

ABSTRACT

The knowledge gained from transcriptome and its regulation is essential to articulate the constituents that make up the molecular network of response gene induction for a certain event, such as those occurring in plant pathogen interactions. The current sequencing technology has led to the development of RNA-seq as a tool that enables mass sequencing of cDNA or RNA, making possible to obtain gene expression profiles for defense responses, which offers great potential to deepen the understanding of mechanisms that are activated during immune responses in plants. RNA-seq is changing the way of how we used to study the transcriptome and has helped identify and quantify new and known plant defense related transcripts. Here, it has present the basis, applications and advantages of RNA-seq, also are discussed recent studies that have revealed the importance and usefulness of this tool for studies of plant pathogen interactions.

Key words: cDNA, sequencing, phytopathology

INTRODUCCIÓN

Celularmente la información genética cifrada en el ADN y contenida en los genes se expresa a través de los mecanismos de transcripción y traducción, a partir del cual se producen moléculas de ARNm y proteínas, respectivamente. Eventos celulares tales como la replicación, la diferenciación y la división celular y otros caracteres macroscópicos tales como rasgos fenotípicos, morfológicos, funcionales y de respuesta ante estímulos son producto de la expresión diferencial de genes. En plantas, el control de la respuesta frente a estados de estrés biótico y abiótico está dado por la actividad

transcripcional de activación o represión de genes [Proudfoot *et al.*, 2002]. La transcripción es un proceso nuclear cuya activación depende de estímulos intra o extracelulares que activan cascadas de señalización para determinar cuáles genes deben expresarse o reprimirse de acuerdo con el tipo de estímulo inicial.

La regulación de la transcripción depende de la unión de activadores o represores en los elementos del promotor ubicados en la región 5' de la secuencia codificante. Los activadores o represores dictaminan la tasa de síntesis de ARNm que debe producir la ma-

quinaria basal de transcripción, la cual está constituida por los factores de transcripción generales (GTFII) y la ARN polimerasa II [Proudfoot *et al.*, 2002].

La cantidad de moléculas producidas de determinado ARNm depende de la función que este tenga en un proceso celular específico. Así, cuando se requiera dar respuesta a una condición determinada en la cual un gen tiene una participación importante, más moléculas de este transcrito se producirán. De manera similar, bajo ciertas circunstancias particulares hay genes que permanecen apagados, pero un estímulo hace que se expresen y se inicia entonces la transcripción. De esta manera la determinación de dónde, cómo y cuándo es generado un transcrito, bajo una condición dada, es fundamental para el entendimiento de la actividad biológica de un gen. Más aún los niveles de ARNm pueden dar no solo una visión clara de patrones de expresión, sino también cuantificaciones altamente correlacionadas entre cambios en la abundancia de ARNm con cambios en la abundancia de proteínas [Lockhart y Winzeler, 2000]. En conjunto, todos los transcritos derivados de genes que se producen en una célula en un momento y bajo una condición fisiológica determinada se denomina *transcriptoma*.

El estudio y análisis del transcriptoma es esencial para el entendimiento de la función de genes. De manera general se puede establecer que si un gen se expresa en una condición o célula determinada es porque cumple allí una función. El estudio global del transcriptoma permite también establecer patrones de regulación génica coordinada, lo que contribuye no solo a dilucidar la función y agrupamiento de varios genes bajo un estímulo o condición específica, sino también a identificar elementos promotores comunes a varios genes. En la década de los noventa, los *northern blots*, los microarreglos de ADNc (ADN complementario obtenido por transcriptasa inversa a partir de ARNm), los cDNA-AFLP y el análisis serial de expresión de genes SAGE (del inglés *serial analysis of gene expression*), entre otras técnicas, permitieron el desarrollo y generación de conocimiento en transcriptómica, al estudiar la expresión de genes relacionados con respuestas a estímulos o a condiciones particulares, así como para determinar cambios en los patrones de expresión génica en tratamientos y cinéticas de expresión [Shalon *et al.*, 1996; Schena *et al.*, 1998; Meyers *et al.*, 2004; Marguerat y Bahler, 2010]; sin embargo, estas estrategias resultan limitantes al estar basadas en hibridación, tener baja cobertura, y en algunos casos necesitar algún conocimiento previo

de la secuencia del genoma para su implementación [Ward *et al.*, 2012].

Actualmente, y gracias a los avances en las técnicas de secuenciación del ADN, a través de tecnologías de nueva generación, NGS (del inglés Next Generation Sequencing), se han revolucionado campos como los de la genómica y la transcriptómica. Estas tecnologías han permitido no solo generar información con altos rendimientos y a bajo costo, sino también abrir nuevos horizontes para el entendimiento detallado y global de procesos de expresión génica [Mochida y Shinozaki, 2011; Schneeberger y Weigel, 2011; Ward *et al.*, 2012].

La caracterización completa y el análisis global de la expresión génica en una célula o tejido, aun sin ninguna información genómica previa, es ahora posible a través de la implementación de la secuenciación de ADNc, o más recientemente de la secuenciación directa de ARN, tecnología conocida como RNA-seq [Wang *et al.*, 2009; Garber *et al.*, 2011; Egan *et al.*, 2012; Ward *et al.*, 2012]. Esta herramienta transcriptómica cambia la manera de cómo se analizan y comprenden los transcriptomas [Wang *et al.*, 2009]. Además, el RNA-seq da una cobertura completa de transcritos, genera información no solo de la secuencia, sino también de la estructura de exones y posibles eventos de *splicing* alternativo [Lister *et al.*, 2009; Gullledge *et al.*, 2012]. La información de esta manera puede ser integrada e interpretada, y se constituye de gran utilidad para vislumbrar procesos biológicos y mecanismos de coexpresión.

En el poco tiempo que esta tecnología se encuentra disponible se han desarrollado un grupo relativamente amplio de investigaciones dirigidas a caracterizar y a cuantificar transcriptomas, así como a comprender los mecanismos de la variación de la expresión génica. Las aplicaciones de RNA-seq se han llevado a cabo en especies eucariotas tales como *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster*, el ratón y el humano [Nagalakshmi *et al.*, 2008; Mortazavi *et al.*, 2008; Maher *et al.*, 2009; Pickrell *et al.*, 2010; Gan *et al.*, 2010; Daines *et al.*, 2011; Peng *et al.*, 2012], lo que ha demostrado la alta aplicabilidad que el RNA-seq ha tenido en estudios de especies modelo.

Recientemente la tecnología de RNA-seq también se ha implementado para estudios de transcriptomas vegetales. Los estudios de RNA-seq en plantas ha permitido, por ejemplo, la identificación de genes expresados frente a tratamientos de vernalización y respuesta a

giberelinas en remolacha azucarera [Mutasa-Gottgens *et al.*, 2012] y en respuesta a estrés hídrico y calor [Gulledge *et al.*, 2012]. La información generada por RNA-seq también se ha explotado para la identificación de SNP (del inglés *single nucleotide polymorphism*) en uva y arroz [Zenoni *et al.*, 2010; Lu *et al.*, 2010] y para la detección de variantes de *splicing* alternativo durante el desarrollo del fruto de uva y en *Arabidopsis* [Zenoni *et al.*, 2010; Gulledge *et al.*, 2012]. También se ha constituido en una herramienta fundamental para ayudar en la anotación de genes [Lu *et al.*, 2010]. Lo anterior muestra la utilidad y posibilidades que esta herramienta transcriptómica tiene y su aplicación bajo diferentes enfoques investigativos.

Tecnología RNA-seq

El RNA-seq es una herramienta transcriptómica actual que está fundamentada en la secuenciación de ADNc basada en los desarrollos NGS. En esta tecnología se captura el ARN total o ARNm, el cual se fragmenta y convierte en una librería de ADNc. Uno de los pasos fundamentales es la obtención de un ARN de buena calidad que represente todos los transcritos que se producen en la condición y tejido de estudio. Para el aislamiento del ARN con frecuencia se emplean kits de extracción de ARNm, que aplican la captura a partir de la cola poly(A) [Ward *et al.*, 2012]. La fragmentación del ARN o del cDNA se realiza o bien por nebulización, por digestión con enzimas de restricción o a través del uso de cationes divalentes bajo condiciones de presiones elevadas [Wang *et al.*, 2009]. Generalmente el fraccionamiento se realiza posteriormente a la síntesis de ADNc. Esta síntesis se realiza con procedimientos estándares bien establecidos para la mayoría de organismos que hacen uso de la enzima transcriptasa reversa.

Una vez obtenido el ADNc se ligan adaptadores de tal forma que cada fragmento generado contendrá un adaptador ligado en sus extremos 3' y 5'. Las secuencias de estos adaptadores se conocen y serán necesarias para que cada fragmento pueda ser secuenciado, y en algunos casos pueden emplearse para diferenciar otros grupos de fragmentos obtenidos a partir de muestras de ADNc diferentes; sin embargo, no en todos los casos se requiere la ligación de adaptadores, lo cual dependerá de la plataforma de secuenciación a emplear. Los adaptadores se pueden ligar directamente a la muestra de ARN, previa síntesis de ADNc [Core *et al.*, 2008; Marguerat y Bahler, 2010], o alternativamente se pue-

den adicionar directamente a la cadena sencilla de ADNc [Maher *et al.*, 2009; Marguerat y Bahler, 2010].

En cuanto a la cantidad y concentración del ARNm que se requiere para la tecnología RNA-seq, el rango está actualmente entre 5 y 10 µg, con una concentración alrededor de 500 ng/µl [Ryan Kim, UC Davis Genome Center; Nong Chen, Business Development Director, BGI Americas, comunicación personal].

Por otro lado, dentro de las aplicaciones y ventajas que tiene la tecnología RNA-seq está que da una cobertura completa de transcriptos, genera información tanto de la secuencia como de la estructura de exones y sitios de *splicing* alternativo [Lister *et al.*, 2009; Gulledge *et al.*, 2012]. Asimismo, los datos arrojados por RNA-seq tienen una alta precisión con respecto a los niveles de expresión génica que se obtienen a través de PCR (del inglés *polimerase chain reaction*) cuantitativa (qPCR) [Nagalakshmi *et al.*, 2008; Wang *et al.*, 2009; Ward *et al.*, 2012]. Además, también se ha mostrado que los resultados son altamente reproducibles [Wang *et al.*, 2009].

Plataformas y estrategias de secuenciación para RNA-seq

La tecnología RNA-seq actualmente está disponible comercialmente en las compañías Roche/454, Solexa/Illumina, SOLiD/Life Technologies y Helicos/BioSciences; sin embargo, de las tecnologías de NGS disponibles las más aplicadas son Roche/454 y Solexa/Illumina [Strickler *et al.*, 2012]. No obstante, estas compañías y otras no cesan en la búsqueda de mayores rendimientos de secuenciación, obtención de lecturas más largas que se lleven a cabo en tiempo real y cada vez a costos más bajos [Metzker, 2010; Mardis, 2011].

Roche/454

Roche/454 es una tecnología que emplea el secuenciador 454-Genome-Sequencer FLX, desarrollado en 2005. Este fue el primer sistema comercial de NGS. Su principio se basa en la piro-secuenciación o detección de pirofosfato descrita en la década de los ochenta [Nyrén y Lundin, 1985]. En esta tecnología se incorporan adaptadores a los extremos de los fragmentos de cadena simple de ADN o ADNc, los cuales serán adheridos a microperlas que contienen en la superficie oligonucleótidos complementarios a las secuencias de los adaptadores. Posteriormente se lleva a cabo una PCR en emulsión para la amplificación de los fragmentos. El objetivo de esta amplificación es la obtención de un

gran número de moléculas idénticas que producirán altas intensidades de señal en cada lectura (*Fig.*) [Ansorge, 2009; Mardis, 2011; Egan *et al.*, 2012].

Al finalizar la amplificación se lleva a cabo una denaturación, y las microperlas son transferidas a pozos en un chip de fibra óptica, en donde se incuban con las enzimas ADN polimerasa, ATP sulfurilasa, luciferasa y apirasa, así como con los sustratos luciferina y adenosin-5-fosfosulfato (APS). Posteriormente sobre el chip se adiciona un deoxinucleótido particular (dNTP); así cuando la ADN polimerasa incorpore el dNTP a la cadena naciente se liberará pirofosfato. Este pirofosfato proviene de la formación del enlace fosfodiéster y es convertido a adenosin trifosfato (ATP), en presencia de la APS. El ATP así producido reaccionará con la enzima luciferasa en presencia de luciferina para generar oxiluciferina y producirse luz en proporciones equivalentes a las cantidades de ATP producidas. La emisión de luz será detectada por una cámara CCD (dispositivo de carga acoplada, del inglés *charge coupled device*). Finalmente, la apirasa removerá el ATP y dNTPs no incorporados. Una vez realizado esto, se repite el ciclo con un nuevo dNTP [Nyrén y Clifton 2007; Ansorge, 2009; Egan *et al.*, 2012].

Actualmente Roche/454 ofrece el servicio de secuenciación para RNA-seq con su más reciente secuenciador, el sistema SG FLX+. Con esta tecnología es posible obtener lecturas de hasta 1000 nucleótidos; sin embargo, la cobertura es baja, alrededor de 2,5 millones de lecturas o *reads* por corrida (454/Roche *sequencing contact*, comunicación personal).

Illumina

Por su parte, Solexa/Illumina se basa en el principio de amplificación en puente y el uso de marcaje por fluorescencia de nucleótidos modificados como terminadores reversibles (*Fig.*) [Metzker, 2010]. En esta tecnología, uno de los adaptadores de los extremos de los fragmentos de ADN o ADNc se liga complementariamente a oligonucleótidos adheridos a una superficie sólida o *flow cell*. Estos oligonucleótidos harán las veces de *primers* o cebadores sentido o antisentido, y crean puentes que favorecen la amplificación. Los amplicones permanecerán adheridos y luego de una denaturación formarán otro puente para permitir la amplificación. Estos pasos se repiten hasta generar millones de grupos o *clusters* de un fragmento determinado. Una vez formados estos grupos se desnaturalizarán nuevamente para iniciar la polimerización o

síntesis de cada fragmento, y se introduce esta vez en la mezcla de síntesis cuatro nucleótidos marcados 32-O-azidometil o terminadores reversibles [Ju *et al.*, 2006; Guo *et al.*, 2008; Metzker, 2010].

Los terminadores reversibles (dideoxinucleótidos) detendrán la síntesis de ADN una vez la ADN polimerasa integre a la cadena naciente el nucleótido correspondiente. Seguido de la síntesis, los fluoróforos de los terminadores reversibles integrados a la cadena naciente son activados por un láser. La emisión de luz será diferencial de acuerdo con el nucleótido incorporado. La información será registrada y almacenada. Una vez hecha la detección, un lavado retira los dideoxinucleótidos no integrados y enzimáticamente es cortado el terminador para que así un nuevo ciclo permita la incorporación del siguiente nucleótido [Metzker, 2010; Egan *et al.*, 2012].

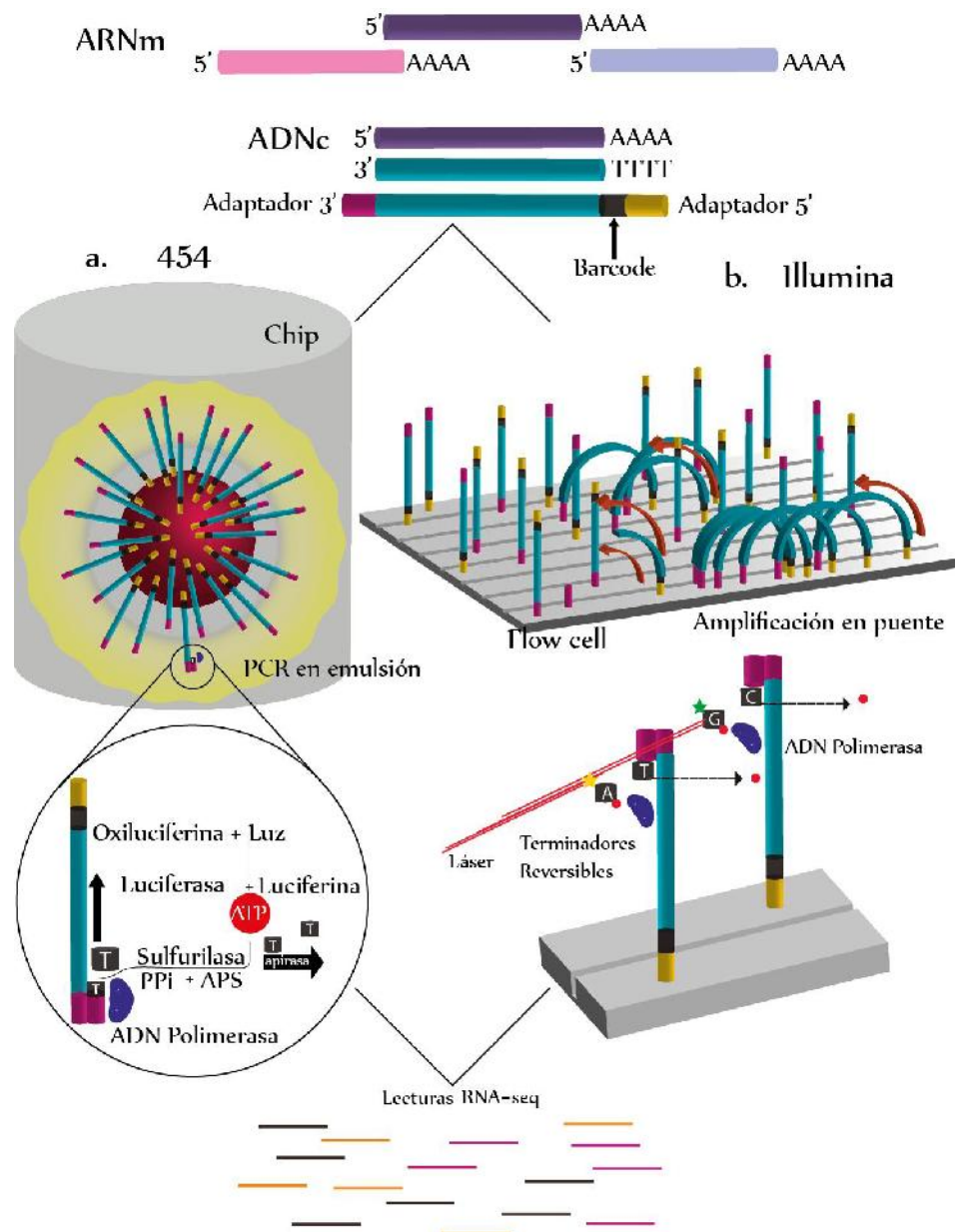
Cada *flow cell* de Illumina contiene ocho carriles o *lanes*. Cada uno de ellos en la actualidad tiene un rendimiento de 150 millones de lecturas o *reads*. La longitud de las lecturas generadas es pequeña, del rango de 50-100 nucleótidos, las cuales pueden ser lecturas sencillas desde un solo extremo, SE (del inglés *single end sequencing reads*) o bien lecturas desde ambos extremos PE (del inglés *pair end sequencing reads*) [Wang *et al.*, 2009; Garber *et al.*, 2011]. Aunque las longitudes sean cortas, el alto número de lecturas generadas incrementarán la cobertura y permitirá extender la secuencia hasta poder, en algunos casos, obtener la secuencia de todo el transcripto. Además, cada *lane* tiene una capacidad para 24 muestras o librerías de ADNc; sin embargo, entre mayor es el número de muestras que se ubiquen en cada *lane* de Illumina, los millones de lecturas por muestra disminuirán. En el caso en el que se desee potencializar la técnica a través de la obtención de un mayor número de lecturas por muestra, el uso de réplicas biológicas es adecuado [Ryan Kim, UC Davis Genome Center y Veridiana Cano, Illumina, comunicación personal].

El número de millones de lecturas deseables dependerá de varios factores. Uno de ellos es la disponibilidad de un genoma de referencia que facilitará el ensamblaje de las lecturas. Otro factor importante es la información sobre el número de genes de la especie, pues lo que se busca es tener representado en miles de lecturas cerca de la totalidad de genes expresados.

Secuenciación directa de ARN-Tecnología Helicos

Helicos/BioScience fue el primer sistema comercial en disponer de una tecnología de secuenciamiento por síntesis capaz de secuenciar una molécula sencilla [Harris *et al.*, 2008; Thompson y Milos, 2011]. Más aún, esta compañía ha desarrollado un método de secuenciación directo de ARN o DRS (del inglés *direct RNA sequencing*), sin necesidad de la previa conversión de ARN a ADNc, método que fue reportado por primera vez en transcritos de *Saccharomyces cerevisiae* y que aún está en perfeccionamiento [Ozsolak *et al.*, 2009; Lipson *et al.*, 2009; Levin, *et al.*,

2010; Ozsolak y Milos, 2011]. La secuenciación directa de ARN podría disminuir alguno de los inconvenientes que se pueden presentar durante la conversión de ARN a ADNc. Uno de los problemas más frecuentes es la obtención de ADNc quimérico, en donde la cadena naciente se puede disociar del molde ARN y posteriormente realinear a una cadena diferente de ARN con una secuencia similar, e iniciar la síntesis de nuevo [Mardis, 2011; Ozsolak y Milos, 2011]; sin embargo, hasta que la tecnología DRS no se establezca ampliamente, la obtención de librerías ADNc para RNA-seq seguirá de líder.



Secuenciamiento masivo de ADNc, RNA-seq, por las tecnologías NGS Illumina y 454. a: Tecnología de secuenciación 454, basada en el principio de PCR en emulsión y pirosecuenciación; b: Tecnología de secuenciación Illumina basado en el principio de amplificación en puente y uso de marcaje por fluorescencia de terminadores reversibles.

Estrategias y consideraciones para experimentos RNA-seq

Independiente de la plataforma de secuenciación a usar, es recomendable que cada muestra o librería contenga patrones de reconocimiento llamados *códigos de barra* o *barcodes*, lo que permitirá maximizar el número de muestras a ubicar en un solo *lane*. Esta estrategia es conocida como *barcoding* o *multiplex*. Cada *barcode* es una secuencia corta, alrededor de cinco o seis nucleótidos, que caracteriza cada muestra o librería de ADNc. Esta secuencia generalmente está ubicada contigua a la secuencia del adaptador en el fragmento de ADNc [Strickler *et al.*, 2012]. Los *barcodes* son muy útiles, ya que permiten no solo diferenciar las muestras, sino también que hacen más eficiente el uso de cada *lane*.

Los *barcodes* deben diseñarse uno por cada muestra o librería de ADNc. Aunque las compañías de secuenciación ofrecen secuencias prediseñadas y probadas, el diseño debe ser cuidadoso. Algunas de las recomendaciones es que los dos primeros nucleótidos no pueden ser iguales de un *barcode* a otro; además, un *barcode* no puede ser la secuencia reversa complementaria del otro, ni tampoco pueden ser secuencias palíndromas. Al hacer uso del *barcoding* se debe tener en cuenta que el tamaño de cada lectura de los transcriptos se reducirá en aproximadamente 18 nucleótidos, pues en el ensamblaje de estas se eliminarán tanto las secuencias de los adaptadores como las de los *barcodes* [Strickler *et al.*, 2012].

Una vez revisadas de manera general las tecnologías de secuenciación para RNA-seq, surge la pregunta de cómo escoger la plataforma de secuenciación apropiada para un proyecto RNA-seq. A partir de la premisa de similitud en cuanto a rendimiento, precisión y capacidad de las lecturas desde ambos extremos del ADNc (3' y 5'), y que actualmente las tecnologías con mayor demanda son Roche/454 y Solexa/Illumina, el punto crítico para la respuesta dependerá de cuál es la longitud de las lecturas que se desea obtener, así como si se cuenta o no con un genoma referencia e indudablemente del presupuesto con el que se disponga, pues en términos generales, el costo por base es menor con la tecnología Illumina. Un *lane* de Solexa/Illumina para RNA-seq en la actualidad tiene un costo cercano a los 3000 dólares.

Como se mencionó anteriormente, Roche/454 produce lecturas más largas con respecto a Solexa/Illumina; sin embargo, cuando el propósito es realizar ensamblaje *de Novo*, es decir, sin genoma de referencia, lo más reco-

mendable es obtener lecturas más largas con el fin de facilitar el ensamblaje. En este caso Roche/454 sería de gran utilidad; no obstante, es necesario considerar que el número de lecturas obtenidas no será muy alto. Si se dispone de un presupuesto cuantioso se puede realizar un elevado número de lecturas, lo que facilitará el ensamblaje. Alternativamente, si se cuenta con un genoma referencia, las lecturas cortas obtenidas con Illumina pueden ser la mejor opción, ya que no es necesario ensamblar el transcriptoma completo y el alto número de lecturas que se generan darán mayor confiabilidad a los datos para la cuantificación de la expresión y para otros fines como la detección de polimorfismos [Ozsolak y Milos, 2011].

Para el ensamblaje de las lecturas existe una amplia diversidad de programas bioinformáticos disponibles. Hay programas especializados en la eliminación de *barcodes*, eliminación de ruido, etc. Si se trata de un ensamblaje *de Novo*, las lecturas serán ensambladas a partir de *contigs*, es decir, en secuencias sobrelapadas, de allí la importancia de tener secuencias más largas. Algo importante de este tipo de ensamblaje es que es posible encontrar transcriptos que pueden no estar representados en el genoma, como por ejemplo eventos de *splicing* alternativo o genes que no se expresan bajo la condición de estudio [Strickler *et al.*, 2012].

Dentro de los programas bioinformáticos aplicados al ensamblaje y análisis de RNA-seq se encuentra por ejemplo Scriptutur, desarrollado en el Massachusetts Institute of Technology. Este se ha empleado para el estudio del transcriptoma de ratones [Guttman *et al.*, 2010]. Trinity es otro programa ampliamente usado, el cual fue desarrollado por Broad Institute y la Hebrew University of Jerusalem, que permite la reconstrucción de transcriptos, reconoce eventos de *splicing* alternativo y es especializado en análisis de muestras que no tienen genoma de referencia [Grabherr *et al.*, 2011]. El programa R ya ha diseñado un paquete estadístico, denominado DEGseq, dirigido al análisis de la expresión diferencial entre muestras y tratamientos [Wang *et al.*, 2010]. Además de estos, actualmente existen muchos otros programas útiles, probados y de acceso gratuito [Ward *et al.*, 2012].

Uno de los objetivos al emplear la tecnología de RNA-seq no es solo identificar la presencia de transcriptos, sino la de cuantificar el nivel de expresión de cada uno. En este sentido aquellas lecturas que se encuentren en alta proporción representarán niveles altos de expre-

sión de determinado gen y aquellos transcritos ausentes, o con un bajo número de lecturas, serán aquellos que o no se expresan o lo hacen a niveles muy bajos [Schenk *et al.*, 2012]. En algunos casos se realizan normalizaciones químicas en las librerías ADNc con el fin de igualar su abundancia. De esta manera aquellos transcritos altamente expresados no serán los únicos para los que se obtengan lecturas en la secuenciación [Ward *et al.*, 2012]. Para este fin existen comercialmente diferentes kits, y por lo general están asociados con cada plataforma de secuenciación. En otras situaciones lo que se desea es comparar el perfil de expresión génica entre diferentes muestras que pueden corresponder a tratamientos, tejidos o condiciones.

Para los análisis de las lecturas generadas por RNA-seq es muy frecuente el uso de varios parámetros estadísticos. Uno de ellos es el RPKM (del inglés *reads per kilobase of exon per million mapped reads*). Con este parámetro es posible cuantificar niveles de transcritos, y facilita la comparación entre muestras [Mortazavi *et al.*, 2008]. Otro parámetro útil y muy usado es el *fold change* de las lecturas que corresponde a la división del número de lecturas generadas para un gen particular en una muestra *vs.* el de la otra. Al estimar este parámetro es posible correlacionar la expresión de un gen en dos condiciones distintas, así como establecer radios de expresión génica diferencial entre tratamientos [Auer y Doerge, 2010].

RNA-seq y sus aplicaciones enfocadas al estudio de interacciones planta-patógeno

Dentro de los retos de la fitopatología están el detallar cómo los patógenos son reconocidos por sus hospederos y cómo se establecen interacciones de resistencia y susceptibilidad. En este sentido la biología molecular y la bioinformática, así como el estudio de la expresión génica en eventos de patogenicidad han contribuido de manera importante [Verhage *et al.*, 2010; Lodha y Basak, 2012; Schenk *et al.*, 2012].

Actualmente el entendimiento molecular que se tiene de las interacciones planta-patógeno ha permitido desarrollar un modelo de la función y evolución de la inmunidad vegetal [Jones y Dangl, 2006]. Según este modelo denominado *zigzag*, el primer evento o fase de la respuesta de resistencia consiste en la capacidad de reconocimiento de moléculas conservadas en los microorganismos conocidas como PAMP (del inglés *pathogen associated molecular patterns*) o MAMP (del inglés *microbe-associated molecular patterns*). Dentro de

las moléculas reconocidas de este tipo, las más estudiadas son la flagelina (flg22) [Felix *et al.*, 1999; Chinchilla *et al.*, 2006], el factor de elongación Tu (EF-Tu) [Zipfel *et al.*, 2006], los lipopolisacáridos (LPS) y la quitina [Kaku *et al.*, 2006]. El reconocimiento de PAMP depende de proteínas denominadas PRR (del inglés *pathogen recognition receptors*), las cuales son proteínas de reconocimiento ubicadas generalmente en la membrana plasmática de la célula vegetal [Gómez-Gómez y Boller, 2000; He *et al.*, 2007; Boller y Yang He, 2009; Zipfel, 2009; Thomma *et al.*, 2011].

El primer PRR identificado y más estudiado es el receptor de flg22, denominado FLS2, el cual es una proteína transmembranal con un dominio extracelular rico en repeticiones de leucinas (LRR, *leucine rich repeats*) y un dominio serina/treonina kinasa intracelular, posiblemente encargado de la comunicación de la señal. La proteína FLS2 es clasificada como un receptor con actividad kinasa (RLK, Receptor-Like Kinase) [Gómez-Gómez y Boller, 2000; Asai *et al.*, 2002]. No fue hasta el 2006 cuando se hizo posible la clonación del receptor EFR que reconoce al PAMP EF-Tu, el cual también es de tipo RLK [Zipfel *et al.*, 2006]. Esta primera respuesta basada en la interacción entre PAMP-PRR es denominada como *inmunidad mediada* por PAMP o PTI (*PAMP triggered immunity*). Este tipo de inmunidad generalmente detiene la infección antes de que el microorganismo comience su multiplicación y es suficientemente efectiva contra patógenos potenciales no adaptados [Chisholm *et al.*, 2006].

Una vez que se da la percepción del ligando por parte de los PRR se desencadena una cascada de señalización mediada principalmente por MAP Kinasas (del inglés *mitogen associated protein kinase*) [Gohre y Robatzek, 2008; Colcombet y Hirt, 2008; Beckers *et al.*, 2009], que finaliza con la activación de factores de transcripción, lo que conlleva a una reprogramación en la expresión génica. Por ejemplo, en *Arabidopsis* se ha establecido que la inducción de la expresión de los genes *FRK*, *At2g17740* y *WRKY22/29* es un criterio diagnóstico de la activación de la PTI [Asai *et al.*, 2002; Shan *et al.*, 2008]. Adicionalmente estas respuestas están asociadas con eventos de apertura de canales iónicos en la membrana plasmática, producción de especies reactivas de oxígeno y fortificación de paredes celulares a través de la deposición de callosa [Buchanan *et al.*, 2000; Zipfel, 2008; 2009].

Durante la evolución, un grupo particular de patógenos desarrolló un tipo especial de proteínas denominadas

efectoras, que al ser inyectadas al interior de las células de las plantas hospederas bloquean la PTI y logran una colonización exitosa [Jones y Dangl, 2006]. Este tipo de interacciones se denominan *compatibles*, y el resultado para la planta es el desarrollo de la enfermedad. Según el modelo zigzag propuesto, este tipo de interacciones cae bajo el concepto de ETS (del inglés *effector triggered susceptibility*); sin embargo, las plantas frente a esta situación desarrollaron una segunda rama de la inmunidad basada en el reconocimiento de proteínas efectoras del patógeno, tercera fase del modelo zigzag. Esta inmunidad es denominada como ETI (del inglés *effector triggered immunity*). La ETI depende de la presencia de proteínas de resistencia que pueden reconocer, de manera directa o indirecta, efectores. Este reconocimiento desencadena una reacción de incompatibilidad o resistencia [Chisholm *et al.*, 2006; Jones y Dangl, 2006; Dodds y Rathjen, 2010]. De manera similar a la PTI, parte de las respuestas desencadenadas durante la ETI incluyen la reprogramación de la expresión génica. El estudio de esta reprogramación permite conocer los mecanismos moleculares de respuesta, y es susceptible de ser estudiada mediante la técnica RNA-seq.

La inducción de varios genes importantes para la defensa vegetal se ha identificado durante las respuestas ETI. Producto de la expresión de estos genes se encuentran las proteínas PR (del inglés *pathogenesis-related proteins*) [Loon y Strien, 1999]. Algunas proteínas codificadas por estos genes PR tienen actividad quitinasa y glucanasa. En la cascada de señalización que lleva a la transcripción de estos genes, el ácido salicílico SA (del inglés *salicylic acid*) y el etileno desempeñan una función reguladora y sinérgica. Asimismo, se ha encontrado que la activación de grupos de genes PR puede ser mediada por patógenos a través de un mecanismo llamado *resistencia sistémica adquirida* o SAR (del inglés *systemic acquired resistance*). Para que este mecanismo ocurra, la infección inicial debe resultar en una lesión necrótica, como parte de una respuesta hipersensible o HR (del inglés *hypersensitive response*). La activación de SAR produce una reducción de los síntomas de la enfermedad ante un siguiente ataque del patógeno e incluso algunos otros tipos de patógenos no relacionados con el de la primera infección [Durrant y Dong, 2004].

Las proteínas PR son una gran familia, y aunque muchas de ellas fueron inicialmente encontradas en tabaco y en *Arabidopsis*, su inducción por patógenos se ha reportado en diversas especies vegetales [Loon y Strien,

1999]. Más aún, su reconocida actividad de defensa ante eventos de patogénesis ha convertido a estos genes PR en genes marcadores de defensa, y ha hecho que la medición de los niveles de su expresión sea ampliamente utilizada en estudios de interacciones planta-patógeno [Slaughter *et al.*, 2012; Zhang *et al.*, 2012].

La reprogramación de la expresión génica durante las reacciones de PTI, ETS y ETI ha sido bastante estudiada en diferentes patosistemas a través de la aplicación de microarreglos o cDNA-AFLP (del inglés *amplified fragment length polymorphism*). En el modelo *Arabidopsis*, a través del estudio de perfiles de expresión de ARNm con microarreglos y frente a interacciones con *P. syringae*, fue demostrado que las respuestas entre interacciones compatibles (bacteria virulenta), incompatibles (bacteria avirulenta) y no hospedero (*P. syringae* pv. *phaseolicola*), son similares, y que las diferencias en expresión encontradas son cuantitativas en cuanto a la intensidad y en la rapidez con la que se producen; sin embargo, la respuesta vegetal en la interacción incompatible es robusta en términos de altos niveles de expresión de los genes de resistencia RPS2 y RPM1, los cuales activan la respuesta mediada por los genes *avr* *avrRpt2* y *avrB*, *avrRpm1* de *P. syringae* [Tao *et al.*, 2003]. Asimismo, el empleo de microarreglos contribuyó a la identificación del receptor EFR que reconoce el PAMP Ef-Tu [Zipfel *et al.*, 2006]. También recientemente esta técnica permitió profundizar en el entendimiento de la actividad de las citoquininas en la inmunidad vegetal, y puso en evidencia que este fitorregulador regula positivamente el incremento de las respuestas de defensa mediada por ácido salicílico y más aún, que este último tiene una retroalimentación negativa al inhibir la señalización de citoquininas [Argueso *et al.*, 2012].

De manera más general los microarreglos han permitido el estudio de la reprogramación génica en *Arabidopsis* en respuesta a virus [Whitham *et al.*, 2003], a hongos [Ramonell *et al.*, 2002] e incluso a nematodos [Puthoff *et al.*, 2003]. También se han empleado microarreglos de papa para el estudio de las interacciones incompatibles con *Phytophthora* [Avrova *et al.*, 1999], pero también durante la interacción compatible [Restrepo *et al.*, 2005]. En arroz los perfiles de expresión por microarreglos se han estudiado frente a infección por patógenos como *Xanthomonas oryzae* pv. *oryzae* y *Magnaporthe oryzae* [Li *et al.*, 2006].

En cuanto al uso de cDNA-AFLP para análisis de transcriptomas vegetales en patosistemas, esta ha sido

aplicada fundamentalmente para la búsqueda de perfiles polimórficos, relacionados con la expresión de genes frente al ataque por patógenos [Birch y Kamoun, 2000; Durrant *et al.*, 2000]. También a través de cDNA-AFLP se han estudiado perfiles de expresión relacionados con respuesta hipersensible en tabaco y tomate [Vandenaabee *et al.*, 2003; Gabriëls *et al.*, 2006], así como en resistencia sistémica adquirida en *Arabidopsis* [Maleck *et al.*, 2000]; sin embargo, como se mencionó antes, estas técnicas o bien requieren de pasos de hibridación como los microarreglos, o poseen una baja cobertura de transcritos. Esto hace que la alternativa del uso de RNA-seq sea llamativa y muy prometedora.

Antecedentes del uso de RNA-seq en interacciones planta-patógeno

Los primeros trabajos de secuenciación de ADNc en plantas se llevaron a cabo en *Medicago truncatula*, *Arabidopsis thaliana* y *Zea mays* con la tecnología 454. En estos reportes se identificaron más de 17 000 genes de *Arabidopsis*, 25 000 secuencias genómicas en maíz que no estaban anotadas ni tenían similitud alguna con otras especies, y 400 SSR (del inglés *simple sequence repeats*) en *M. truncatula* [Cheung *et al.*, 2006; Emrich *et al.*, 2007; Weber *et al.*, 2007]. Desde entonces el RNA-seq mostró ser altamente sensible y prometedor para el análisis profundo de transcriptomas en plantas. Dado lo novedoso de la técnica, los estudios en interacciones planta-patógeno, en donde se hace uso de RNA-seq, son escasos. A la fecha se han reportado estudios en *Arabidopsis thaliana*, algodón y soya, pero en otras especies como yuca y maíz las investigaciones están en desarrollo.

Uno de los trabajos más recientes enfocado al análisis de múltiples genomas y transcriptomas en *Arabidopsis* con el uso de RNA-seq reveló que la variación en los niveles de expresión génica es mucho mayor en genes que están involucrados en respuestas a estrés biótico. Asimismo, que los genes de resistencia de las subfamilias NB-LRR (del inglés *nucleotide binding leucine rich repeat*), *coiled-coil*, receptores Toll interleuquina-1 y genes relacionados con defensa codifican para proteínas más variables que aquellas codificadas por genes del metabolismo basal [Xiangchao *et al.*, 2011].

En el patosistema algodón-*Verticillium dahliae* se reportó inicialmente, mediante microarreglos, cambios transcriptómicos en las respuestas de defensa de 211 genes, así como la activación de la respuesta de defensa basal correspondiente a la PTI y a la rápida produc-

ción de fitoalexinas, terpenoides y felipropanoides [Cui *et al.*, 2000]. No obstante este conocimiento, el entendimiento de las respuestas de defensa del algodón ante *V. dahliae* era limitado. Recientemente, a través de la aplicación de RNA-seq y la plataforma Illumina, se logró obtener el primer análisis global del transcriptoma de defensa en algodón. En este estudio se pudo monitorear los perfiles de expresión en raíces a las 4, 12, 24 y 48 h posinoculación (hpi), se detectó expresión diferencial en más de 3000 genes, lo que permitió enriquecer el entendimiento de cómo los genes involucrados en actividad enzimática, especialmente en la ruta metabólica fenilpropanoide, están involucrados en eventos de respuesta [Xu *et al.*, 2011]. También se reportaron niveles de lignificación y actividad enzimática contrastante, así como expresión génica diferencial en plantas resistentes y susceptibles al hongo.

En soya, con RNA-seq, fueron mapeados más de 43 000 genes con el genoma referencia de esta especie, y se estudiaron expresiones diferenciales de más de 1000 genes en líneas casi isogénicas a las 0, 6 y 12 hpi con *Xanthomonas axonopodis* pv. *glycines* (*Xag*), agente causal de la BPL (del inglés *bacterial leaf pustule*) de la soya. Asimismo, bajo este enfoque, se detectó la sobreexpresión de PRR y genes que son inducidos por estos en líneas resistentes a *Xag* [Kim *et al.*, 2011]. En esta investigación se unieron tres réplicas biológicas de cada tiempo 0, 6 y 12 hpi, se corrió un *lane* de Illumina para cada uno de ellos y se obtuvo más de 125 millones de *reads*. También en soya y con la tecnología Illumina, pero en un estudio de la roya, otra de las enfermedades más limitantes de este cultivo, recientemente se analizaron patrones de expresión de genes con el objetivo de dilucidar los eventos moleculares que ocurren tras la infección por parte del hongo *Phakopsora pachyrhizi* [Tremblay *et al.*, 2011]. En plantas susceptibles y en etapas avanzadas de infección, un alto porcentaje de genes involucrados en el metabolismo de síntesis de aminoácidos, carbohidratos y lípidos fueron detectados con regulación negativa. Por el contrario, muchos otros genes relacionados con rutas metabólicas implicadas en defensa se sobreexpresaron en etapas iniciales de la infección. De acuerdo con los autores, muchos de los genes encontrados en este trabajo han dado luces para el desarrollo de un programa de mejoramiento genético enfocado a lograr resistencia amplia contra la roya de la soya a través de sobreexpresión o silenciamiento génico [Tremblay *et al.*, 2011].

El RNA-seq también ha sido aplicado para análisis de perfiles transcriptómicos en fitopatógenos. Es el caso

de *Phytophthora phaseoli*, agente causal del mildio veloso en *lima bean*, *Phaseolus lunatus*. Poco se conoce de la base molecular de la interacción de este oomycete con su hospedero. En esta investigación con la tecnología Illumina para RNA-seq se compararon los transcriptomas de tres tratamientos: el primero *P. phaseoli* creciendo bajo condiciones de cultivo, el segundo *P. phaseoli* infectando a *P. lunatus* tres días posinoculación (dpi), y el tercero a los seis dpi. Se trabajaron dos réplicas biológicas por tratamiento y un total de 150 millones de *reads* fueron obtenidos. Los resultados mostraron similitud en 10 427 genes de *P. phaseoli* con el genoma de *P. infestans*, de los cuales 318 son genes putativos de virulencia, y que mostraron ser sobreexpresados en los transcriptomas del patógeno en interacción con la planta [Kunjetti *et al.*, 2011]. Este reporte muestra que esta herramienta también tiene un gran potencial en investigaciones direccionadas hacia el estudio de transcriptomas en fitopatógenos.

Sin duda la implementación de RNA-seq está dentro de las actuales y futuras proyecciones de investigación en diversos patosistemas. A modo de ejemplos está el maíz (*Zea mays-Aspergillus flavus*) y la yuca (*Manihot esculenta-Xanthomonas axonopodis* pv. *manihotis* (*Xam*)). En maíz se planea encontrar genes expresados diferencialmente entre líneas susceptibles y resistentes, así como integrar esta información del transcriptoma con la presencia de QTL (del inglés *quantitative trait loci*), ya reportados de resistencia a la acumulación de aflatoxinas producidas por el patógeno [Kelley *et al.*, 2010].

En yuca actualmente se desarrolla una estrategia en donde se combinará la información arrojada por RNA-seq de líneas resistentes y susceptibles a *Xam* y el genoma referencia, con el fin de asignar funciones putativas a muchos de los 35 000 genes de la yuca. Asimismo se busca identificar los genes que se inducen o reprimen frente al ataque de la bacteria, y finalmente realizar identificación de SNP con el fin de establecer relaciones entre estos marcadores y QTL para resistencia [López y Bernal, 2012]. Este enfoque contribuirá de manera determinante no solo en el mejor entendimiento de la interacción molecular de este patosistema, sino también aportará información para incrementar el número de marcadores presentes en el mapa genético de yuca, lo que facilitará la clonación de genes R contra la bacteriosis vascular de yuca.

De igual forma, hoy por hoy es factible la posibilidad de profundizar en preguntas como ¿cuáles son aquellos

genes que son expresados en el hospedero ante el ataque de un fitopatógeno?, incluso, ¿cuáles son aquellos genes que son expresados en una planta dentro de un sistema heterólogo no hospedero?, ¿cuáles son aquellos genes que se expresan en el patógeno frente a la interacción con su hospedero?, ¿cuál es el nivel de esta expresión?, ¿en qué momento después de iniciada la interacción ocurre?, ¿de manera temprana o tardía?, ¿cuáles son esos nuevos transcriptos que no pudieron ser identificados con previas técnicas para análisis de transcriptoma?, ¿cómo en determinado patosistema la estructura génica en cuanto a exones y sitios de *splicing* alternativo están relacionados con activaciones de respuestas de defensa? Las respuestas a estos y otros cuestionamientos pueden encontrarse bajo el enfoque de RAN-seq.

Conclusiones, retos y perspectivas futuras

- El secuenciamiento masivo de ADNc permite ampliar los rangos de detección de transcriptos que se lograron en su momento bajo la herramienta de microarreglos. Es evidente el gran potencial que tiene el RNA-seq en cuanto a reconocimiento de genes relacionados con defensa, factores de transcripción involucrados en activaciones de respuesta génica y determinación de expresión diferencial, incluso cuando el conocimiento del genoma es escaso o incluso nulo.
- Asimismo, el conocimiento generado por el RNA-seq contribuirá de manera importante dentro de programas de fitomejoramiento. Con la información generada será posible la identificación de nuevos genes blancos para su uso en transformación genética, en búsqueda de expresión génica que se traduzca en el desarrollo de plantas resistentes a patógenos. Con seguridad, en el futuro cercano la literatura enfocada a interacciones planta-patógeno contendrá un sinnúmero de reportes alrededor de esta herramienta transcriptómica. Más aún, las tecnologías de NGS avanzarán, y en esa medida la herramienta se volverá cada vez más poderosa, mejorarán sus rendimientos, se podrán obtener lecturas más largas en un menor tiempo, y por supuesto a menores costos.
- El secuenciamiento directo de ARN es uno de los mayores retos que presenta la técnica; sin embargo, las investigaciones, desarrollos y comprobaciones al respecto están en marcha, y pronto la alta oferta de esta tecnología será una realidad [Ozsolak *et al.*, 2009; Ozsolak y Milos, 2011]. Asimismo, el reto computacional es grande, el análisis de millones de lecturas

que generan archivos de tamaños enormes requiere del desarrollo de herramientas bioinformáticas cada vez con mayor capacidad de análisis, rendimientos más altos, procesos de ensamblaje más sencillos, así como con capacidad de caracterización precisa de estructura y dinámica de transcriptoma. Estos son otros de los retos que tiene esta poderosa herramienta transcriptómica para los próximos años.

- El reto para la fitopatología es explotar la utilidad de esta herramienta transcriptómica y a través de ella profundizar en el entendimiento de las complejas interacciones planta-patógeno, así como revelar eslabones moleculares involucrados que hasta hoy no habían podido ser revelados.

REFERENCIAS

- Ansorge, W. J.: «Next-generation DNA Sequencing Techniques», *New Biotechnology* 25: 195-203, Suiza, 2009.
- Argueso, C. T.; J. F. Ferreira; P. Epple; J. P. To; C. E. Hutchison; G. E. Schaller; J. L. Dangl; J. J. Kieber: «Two-Component Elements Mediate Interactions Between Cytokinin and Salicylic Acid in Plant Immunity», *PLoS Genetics* 8: 1-13, EE.UU., 2012.
- Asai, T.; G. Tena; J. Plotnikova; M. R. Willmann; W. Chiu; L. Gómez-Gómez; T. Boller; F. Ausubel; J. Sheen: «MAP Kinase Signalling Cascade in Arabidopsis Innate Immunity», *Nature* 415: 977-983, EE.UU., 2002.
- Auer, P. L.; R. W. Doerge: «Statistical Design and Analysis of RNA Sequencing Data», *Genetics* 185: 405-416, EE.UU., 2010.
- Avrova, A. O.; H. E. Stewart; W. D. De Jong; J. Heilbronn; G. D. Lyon; P. R. Birch: «A Cysteine Protease Gene Is Expressed Early in Resistant Potato Interactions with *Phytophthora infestans*», *Molecular Plant-Microbe Interactions* 12: 1114-1119, Inglaterra, 1999.
- Beckers, G. J.; M. Jaskiewicz; Y. Liu Y.; W. R. Underwood; S. Y. He; S. Zhang; U. Conrath: «Mitogen-Activated Protein Kinases 3 and 6 Are Required for Full Priming of Stress Responses in *Arabidopsis thaliana*», *Plant Cell* 21: 944-953, Alemania, 2009.
- Birch, P. R.; S. Kamoun: «Studying Interaction Transcriptomes: Coordinated Analyses of Gene Expression During Plant Microorganism Interactions», *Trends in Plant Pathology* 77-82, Inglaterra, 2000.
- Boller, T.; S. Y. He: «Innate Immunity in Plants: An Arms Race Between Pattern Recognition Receptors in Plants and Effectors In Microbial Pathogens», *Science* 324: 742-744, Suiza, 2009.
- Buchanan, B.; W. Gruissem.; R. Jones: *Biochemistry and Molecular Biology of Plants*, American Society of Plants Biologists, Maryland, EE. UU., 2000.
- Cheung, F.; B. Haas; S. Goldberg; G. May; Y. Xiao.; C. Town: «Sequencing *Medicago truncatula* Expressed Sequenced Tags Using 454 Life Sciences Technology», *BMC Genomics* 6: 7-272, Suecia, 2006.
- Chinchilla, D.; Z. Bauer; M. Regenass; T. Boller; G. Felix: «The Arabidopsis Receptor Kinase FLS2 Binds flg22 and Determines the Specificity of Flagellin Perception», *Plant Cell* 18 (2): 465-47, Suiza, 2006.
- Chisholm, S.; G. Coaker; B. Day; B. Staskawicz: «Host-Microbe Interactions: Shaping the Evolution of the Plant Immune Response», *Cell* 124: 803-814, EE.UU., 2006.
- Colcombet, J.; H. Hirt: «Arabidopsis MAPKs: A Complex Signaling Network Involved in Multiple Biological Processes», *Biochemical Journal* 413: 217-226, Francia, 2008.
- Core, L.; J. Waterfall; J. Lis: «Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters», *Science* 322: 1845-1848, EE.UU., 2008.
- Cui, Y.; A. Bel; O. Joost; C. Magill C: «Expression of Potential Defense Response Genes in Cotton», *Physiological and Molecular Plant Pathology* 56: 25-31, EE.UU., 2000.
- Daines, B.; H. Wang; L. Wang; Y. Li; Y. Han; D. Emmert; W. Gelbart; X. Wang; W. Li; R. Gibbs; R. Chen: «The *Drosophila melanogaster* Transcriptome by Paired-End RNA Sequencing», *Genome Research* 21: 315-324, EE.UU., 2011.
- Dodds, P.; J. Rathjen: «Plant Immunity: Towards an Integrated View of Plant-Pathogen Interactions», *Nature Reviews, Genetics* 11: 539-548, EE.UU., 2010.
- Durrant, W.; X. Dong: «Systemic Acquired Resistance», *Annual Review of Phytopathology* 42: 185-209, EE.UU., 2004.
- Durrant, W.; O. Rowland; P. Piedras; K. Hammond-Kosack; J. Jones: «cDNA-AFLP Reveals a Striking Overlap in Race-Specific Resistance and Wound Response Gene Expression Profiles» *The Plant Cell* 12: 963-977. Reino Unido. 2000.
- Egan, A.; J. Schluter; D. Spooner: «Applications of Next-Generation Sequencing in Plant Biology», *American Journal of Botany* 99 (2): 175-185, EE.UU., 2012.
- Emrich, S.; W. Barbazuk; L. Li; P. Schnable: «Gene Discovery and Annotation Using LCM-454 Transcriptome Sequencing», *Genome Research* 17 (1): 69-73, EE.UU., 2007.
- Felix, G.; J. Duran; S. Volko; T. Boller: «Plants Recognise Bacteria Through the Most Conserved Domain of Flagellin», *Plant Journal* 18: 265-276, Inglaterra, 1999.
- Gabriels, S.; F. Takken; J. Vossen; C. de Jong; Q. Liu; S. Turk; L. Wachowski; Y. Peters; H. Witsenboer; P. de Wit; M. Joosten «cDNA-AFLP Combined with Functional Analysis Reveals Novel Genes Involved in the Hypersensitive Response», *Molecular Plant-Microbe Interactions* 19: 567-576, Holanda, 2006.
- Gan, Q.; I. Chepelev; G. Wei; L. Tarayrah; K. Cui; K. Zhao; X. Chen: «Dynamic Regulation of Alternative Splicing and Chromatin Structure in *Drosophila* Gonads Revealed by RNA-seq», *Cell Research* 20: 763-783, EE.UU., 2010.
- Garber, M.; M. Grabherr; M. Guttman; C. Trapnell: «Computational Methods for Transcriptome Annotation and Quantification Using RNA-seq», *Nature Methods* 8: 469-478, EE.UU., 2011.
- Gómez-Gómez L.; T. Booller: «FLS2: an LRR Receptor-Like Kinase Involved in the Perception of the Bacterial Elicitor Flagellin in *Arabidopsis*», *Molecular Cell* 5 (6):1003-1011, Suiza, 2000.
- Gohre, V.; S. Robatzek: «Breaking the Barriers: Microbial Effector Molecules Subvert Plant Immunity», *Annual Review of Phytopathology* 46: 189-215, Alemania, 2008.
- Grabherr, M.; B. Haas; M. Yassour; J. Levin; D. Thompson; I. Amit; X. Adiconis; L. Fan; R. Raychowdhury; Q. Zeng; Z. Chen; E. Mauceli; N. Hacohen; A. Gnirke; N. Rhind; F. Palma; B. Birren; C. Nusbaum; K. Lindblad-Toh; N. Friedman; A. Regev: «Full-Length Transcriptome Assembly from RNA-Seq Data Without a Reference Genome», *Nature Biotechnology* 29: 644-652, EE.UU., 2011.
- Gulledge, A.; A. Roberts; H. Vora; K. Patel; A. Loraine: «Mining *Arabidopsis thaliana* RNA-Seq Data with Integrated Genome Browser Reveals Stress-Induced Alternative Splicing of the Putative Splicing Regulator SR45A», *American Journal of Botany* 99 (2): 219-231, EE.UU., 2012.
- Guo, J.; N. Xu; Z. Li; S. Zhang; J. Wu; D. Kim; M. Marma; Q. Meng; H. Cao; X. Li; S. Shi; L. Yu; S. Kalachikov; J. Russo; N. Turro; J. Ju:

- «Four-Color DNA Sequencing with 32 -O-Modified Nucleotide Reversible Terminators and Chemically Cleavable Fluorescent Dideoxynucleotides», *PNAS* 105 (27): 9145-9150, EE.UU., 2008.
- Guttman, M.; M. Garber; J. Levin; J. Donaghey; J. Robinson; X. Adiconis; L. Fan; M. Koziol; M. Gnirke; C. Nusbaum; J. Rinn; E. Lander; A. Regev: «Ab Initio Reconstruction of Cell Type-Specific Transcriptomes in Mouse Reveals the Conserved Multi-Exonic Structure of lincRNAs», *Nature Biotechnology* 28: 503-510, EE.UU., 2010.
- Harris, T.; P. Buzby; H. Babcock; E. Beer; J. Bowers; I. Braslavsky; M. Causey; J. Colonell; J. Dimeo; J.W. Efcavitch; E. Giladi; J. Gill; J. Healy; M. Jarosz; D. Lapen; K. Moulton; S.R. Quake; K. Steinmann; E. Thayer; A. Tyurina; R. Ward; H. Weiss; Z. Xie: «Single-Molecule DNA Sequencing of a Viral Genome», *Science* 320: 106-109, EE.UU., 2008.
- He, P.; L. Shan; J. Sheen: «Elicitation and Suppression of Microbe-Associated Molecular Pattern-Triggered Immunity in Plant-Microbe Interactions», *Cellular Microbiology* 9: 1385-1396, EE.UU., 2007.
- Jones, J.; J. Dangl: «The Plant Immune System», *Nature* 444: 323-329, EE.UU., 2006.
- Jones, K.; J. Kadonaga; P. Rosenfeld; T. Kelly; R. Tjian: «A Cellular DNA-Binding Protein That Activates Eukaryotic Transcription and DNA Replication», *Cell* 48: 79-89, EE.UU., 1987.
- Ju, J.; D. Kim; L. Bi; Q. Meng; X. Bai; Z. Li; X. Li; M. Marma; S. Shi; J. Wu; J. Edwards; A. Romu; N. Turro: «Four-Color DNA Sequencing by Synthesis Using Cleavable Fluorescent Nucleotide Reversible Terminators», *PNAS* 103: 19635-19640, EE.UU., 2006.
- Kaku, H.; H. Nishizawa; N. Ishii-Minami; C. Akimoto-Tomiyama; N. Dohmae; K. Takio; E. Minami; N. Shibuya: «Plant Cells Recognize Chitin Fragments for Defense Signaling Through a Plasma Membrane Receptor», *PNAS* 103: 11086-11091, Japón, 2006.
- Kelley, R.; C. Gresham; J. Harper; S. Bridges; M. Warburton; L. Hawkins; O. Pechanova; B. Peethambaran; T. Pechan; D. Luthe; J. Mylroie; A. Ankala; S. Ozkan; W. Henry; W. Williams: «Integrated Data Base for Identifying Candidate Genes for *Aspergillus flavus* Resistance in Maize», *BMC Bioinformatics* 11: 1-9, Inglaterra, 2010.
- Kim, K.; Y. Kang; D. Kim; M. Yoon; J. Moon; M. Kim; K. Van; S. Lee: «RNA-Seq Analysis of a Soybean Near-Isogenic Line Carrying Bacterial Leaf Pustule-Resistant and Susceptible Alleles», *DNA Research* 18: 483-497, Inglaterra, 2011.
- Kunjeti, S. G.; T. Evans; A. Marsh; A. Gregory; S. Kunjeti; B. Meyers; V. Kalavacharla; N. Donofrio: «RNA-Seq Reveals Infection-Related Global Gene Changes in *Phytophthora phaseoli*, the Causal Agent of Lima Bean Downy Mildew», *Molecular Plant Pathology* 24/11/2011, <http://www.ncbi.nlm.nih.gov/pubmed/22112294>.
- Levin, J.; M. Yassour; X. Adiconis; C. Nusbaum; D. Thompson; N. Friedman; A. Gnirke; A. Regev: «Comprehensive Comparative Analysis of Strand-Specific RNA Sequencing Methods», *Nature Methods* 7 (9): 709-715, EE.UU., 2010.
- Li, Q.; F. Chen; L. Sun; Z. Zhang; Y. Yang; Z. He: «Expression Profiling of Rice Genes in Early Defense Responses to Blast and Bacterial Blight Pathogens Using cDNA Microarray», *Physiological and Molecular Plant Pathology* 68 :51-60, China, 2006.
- Lipson, D.; T. Raz; A. Kieu; D. Jones; E. Giladi; E. Thayer; J. Thompson; S. Letovsky; P. Milos; M. Causey: «Quantification of the Yeast Transcriptome by Single Molecule Sequencing», *Nature Biotechnology* 27: 652-65, EE.UU., 2009.
- Lister, R.; B. Gregory; J. Ecker: «Next Is Now: New Technologies for Sequencing of Genomes, Transcriptomes and Beyond», *Current Opinion in Plant Biology* 12 (2): 107-118, EE.UU., 2009.
- Lockhart, D.; E. Winzeler: «Genomics, Gene Expression and DNA Arrays», *Nature* 415 :827-836, EE.UU., 2000.
- Lodha, T.; J. Basak: «Plant-Pathogen Interactions: what Microarray Tells About it?», *Molecular Biotechnology* 50 (1): 87-97, India, 2012.
- Loon, V.; V. Strien: «The Families of Pathogenesis-Related Proteins, Their Activities, and Comparative Analysis of PR-1 Type Proteins», *Physiological and Molecular Plant Pathology* 55: 85-97, Holanda, 1999.
- López, C.; A. Bernal: «Cassava Bacterial Blight: Using Genomics for the Elucidation and Management of an Old Problem», *Tropical Plant Biology*. 21-1-2012, Colombia, 2012.
- Lu, G.; D. Fan; C. Zhu; W. Li; Q. Zhao: «Function Annotation of the Rice Transcriptome at Single-Nucleotide Resolution by RNA-Seq», *Genome Research* 20: 1238-49, China, 2010.
- Maher, C.; C. Kumar-Sinha; X. Cao; S. Kalyana-Sundaram; B. Han; X. Jing; L. Sam; T. Barrette; N. Palanisamy; A. Chinnaiyan: «Transcriptome Sequencing to Detect Gene Fusions in Cancer», *Nature* 458: 97-101, EE.UU., 2009.
- Maleck, K.; A. Levine; T. Eulgem; A. Morgan; J. Schmid; K. Lawton; J. Dangl; R. Dietrich: «The Transcriptome of *Arabidopsis thaliana* During Systemic Acquired Resistance», *Nature Genetics* 26: 403-410, EE.UU., 2000.
- Mardis, E.: «A Decade's Perspective on DNA Sequencing Technology», *Nature* 470: 198-203, EE.UU. 2011.
- Marguerat, S.; J. Bahler: «RNA-seq: from technology to biology» *Cellular and Molecular Life Sciences* 67: 569-579. Reino Unido. 2010.
- Metzker, M.: «Sequencing technologies the next generation» *Nature Reviews Genetics* 11: 31-46. EE.UU., 2010.
- Meyers, B.; D. Galbraith; T. Nelson; V. Agrawal: «Methods for Transcriptional Profiling in Plants. Be Fruitful and Replicate», *Plant Physiology*, 135: 637-652, EE.UU., 2004.
- Mochida, K.; K. Shinozaki: «Advances in Omics and Bioinformatics Tools for Systems Analyses of Plant Functions», *Plant and Cell Physiology* 52 (12): 2017-2038, Japón, 2011.
- Mortazavi, A.; B. Williams; K. McCue; L. Schaeffer; B. Wold: «Mapping and Quantifying Mammalian Transcriptomes by RNA-Seq», *Nature Methods* 5: 621-628, EE.UU., 2008.
- Mutasa-Gottgens, E.; A. Joshi; H. Holmes; P. Hedden; B. Gottgens: «A New RNASeq-Based Reference Transcriptome for Sugar Beet and Its Application in Transcriptome-Scale Analysis of Vernalization and Gibberellin Responses», *BMC Genomics* 13: 99, Inglaterra, 2012.
- Nagalakshmi, U.; Z. Wang; K. Waern; C. Shou; D. Raha; M. Gerstein; M. Snyder: «The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing», *Science* 320: 1344-1349, EE.UU., 2008.
- Nowrousian, M.: «Next-Generation Sequencing Techniques for Eukaryotic Microorganisms: Sequencing-Based Solutions to Biological Problems», *Eukaryotic Cell* 9: 1300-1310, Alemania, 2010.
- Nyrén, P.; A. Lundin: «Enzymatic Method for Continuous Monitoring of Inorganic Pyrophosphate Synthesis», *Analytical Biochemistry* 151: 504-509, EE.UU., 1985.
- Nyrén, P.; N. J. Clifton: «The History of Pyrosequencing», *Methods in Molecular Biology*, 373: 1-14, Suecia, 2007.
- Ozsolak, F.; A. Platt; D. Jones; J. Reifengerger; J. Sass; P. McInerney; J. Thompson; J. Bowers; M. Jarosz; P. Milos: «Direct RNA Sequencing», *Nature Letters* 461: 814-819, EE.UU., 2009.
- Ozsolak, F.; P. Milos: «RNA Sequencing: Advances, Challenges and Opportunities», *Nature Reviews Genetics* 12 (2): 87-98, EE.UU., 2011.
- Peng, Z.; Y. Cheng; B. Chin-Ming; L. Kang; Z. Tian; Y. Zhu; W. Zhang; Y. Liang; X. Hu; X. Tan; J. Guo; Z. Dong Z; Z. Liang; L. Bao; J. Wang: «Comprehensive analysis of RNA-seq data reveals extensive RNA editing in a human transcriptome» *Nature Biotechnology* 1-9. China. 2012.

- Pickrell, J.; J. Marioni; A. Pai; J. Degner; B. Engelhardt; E. Nkadori; J. Veyrieras; M. Stephens; Y. Gilad; J. Pritchard J: «Understanding Mechanisms Underlying Human Gene Expression Variation with RNA Sequencing», *Nature* 464: 768-772, EE.UU., 2010.
- Proudfoot, N.; A. Furger; M. Dye: «Integrating mRNA Processing with Transcription», *Cell* 108: 501-512, Inglaterra, 2002.
- Puthoff, D.; D. Nettleton; S. Rodermeil; T. Baum: «*Arabidopsis* Gene Expression Changes During Cyst Nematode Parasitism Revealed by Statistical Analyses of Microarray Expression Profiles», *The Plant Journal* 33: 911-921, EE.UU., 2003.
- Ramonell, K.; B. Zhang; R. Ewing; Y. Chen; D. Xu; G. Stacey; S. Somerville: «Microarray Analysis of Chitin Elicitation in *Arabidopsis thaliana*», *Molecular Plant Pathology* 3: 301-311, EE.UU., 2002.
- Ren, X.; O. Vorst; M. Fiers; W. Stiekema; J. Nap: «In Plants, Highly Expressed Genes Are the Least Compact», *Trends in Genetics* 22: 528-532, Holanda, 2006.
- Restrepo, S.; K. Myers; O. del Pozo; G. Martin; A. Hart; C. Buell; W. Fry; C. Smart: «Gene Profiling of a Compatible Interaction Between *Phytophthora infestans* and *Solanum tuberosum* Suggests a Role for Carbonic Anhydrase», *Molecular Plant-Microbe Interactions* 18: 913-922, EE.UU., 2005.
- Schena, M.; R. Heller; T. Theriault; K. Konrad; E. Lachenmeier; R. Davis: «Microarrays: Biotechnology's Discovery Platform for Functional Genomics», *Trends Biotechnology* 16: 301-306, EE.UU., 1998.
- Schenk, P.; L. Carvalhais; K. Kazan: «Unraveling Plant-Microbe Interactions: Can Multi-Species Transcriptomics Help?», *Trends in Biotechnology* 3: 177-84, Australia, 2012.
- Schneeberger, K.; D. Weigel: «Fast-Forward Genetics Enabled by New Sequencing Technologies», *Trends in Plant Science* 16: 282-288, Alemania, 2011.
- Severin, A.; J. Woody; Y. Bolon; B. Joseph; B. Diers; A. Farmer; G. Muehlbauer; R. Nelson; D. Grant; J. Specht; M. Graham; S. Cannon; G. May; C. Vance; R. Shoemaker: «RNA-Seq Atlas of *Glycine max*: A Guide to the Soybean Transcriptoma», *BMC Plant Biology* 10: 1-16, EE.UU., 2010.
- Shan, L.; P. He; J. Li; A. Heese; S. Peck; T. Nürnberger; G. Martin; J. Sheen: «Bacterial Effectors Target the Common Signaling Partner BAK1 to Disrupt Multiple MAMP Receptor-Signaling Complexes and Impede Plant Immunity», *Cell Host Microbe* 4: 17-27, EE.UU., 2008.
- Shalon, D.; S. Smith; P. Brown: «A DNA Microarray System for Analyzing Complex DNA Samples Using Two-Color Fluorescent Probe Hybridization», *Genome Research* 6: 639-645, EE.UU., 1996.
- Slaughter, A.; X. Daniel; V. Flors; E. Luna; B. Hohn; B. Mauch-Mani: «Descendants of Primed *Arabidopsis* Plants Exhibit Resistance to Biotic Stress», *Plant Physiology* 158 (2): 835-843, Suiza, 2012.
- Strickler, S.; A. Bombarely; L. Mueller: «Designing a Transcriptome Next-Generation Sequencing Project for a Nonmodel Plant Species», *American Journal of Botany* 99: 257-266, EE.UU., 2012.
- Tao, X.; X. Xie; W. Chen; J. Glazebrook; H. Chang; B. Han; T. Zhu; G. Zou; F. Katagiri: «Quantitative Nature of *Arabidopsis* Responses during Compatible and Incompatible Interactions with the Bacterial Pathogen *Pseudomonas syringae*», *The Plant Cell* 15: 317-330, EE.UU., 2003.
- Tremblay, A.; P. Hosseini; N. Alkharouf; S. Li; B. Matthews: «Gene Expression in Leaves of Susceptible *Glycine max* during Infection with *Phakopsora pachyrhizi* Using Next Generation Sequencing», *Sequencing* 1-14, EE.UU., 2011.
- Thomma, B.; T. Nurnberger; M. Joosten: «Of PAMPs and Effectors: The Blurred PTI-ETI Dichotomy», *The Plant Cell* 23: 4-15, Holanda, 2011.
- Thompson, J.; P. Milos: «The Properties and Applications of Single-Molecule DNA Sequencing», *Genome Biology* 12: 217, EE.UU., 2011.
- Vandenabeele, S.; K. Van Der Kelen; J. Dat; I. Gadjev; T. Boonefaes; S. Morsa; P. Rottiers; L. Slooten; M. Van Montagu; M. Zabeau; D. Inze; F. Van Breusegem: «A Comprehensive Analysis of Hydrogen Peroxide-Induced Gene Expression in Tobacco», *PNAS* 26: 16113-16118, Bélgica, 2003.
- Verhage, C.; C. Saskia; C. Pieterse: «Plant Immunity: It's the Hormones Talking, But What Do They Say?», *Plant Physiology* 154: 536-540, Holanda, 2010.
- Wang, Z.; M. Gerstein; M. Snyder: «RNA-Seq: a Revolutionary Tool for Transcriptomics», *Nature Reviews, Genetics* 10: 57-63, EE.UU., 2009.
- Wang, L.; Z. Feng; X. Wang; X. Wang; X. Zhang: «DEGseq: an R Package for Identifying Differentially Expressed Genes from RNA-Seq Data», *Bioinformatics* 26: 136-138, China, 2010.
- Ward, J.; L. Ponnala; C. Weber: «Strategies for Transcriptome Analysis in Nonmodel Plants», *American Journal of Botany* 99 (2): 267-276, EE.UU., 2012.
- Weber, A.; K. Weber; K. Carr; C. Wilkerson; J. Ohlrogge: «Sampling the *Arabidopsis* Transcriptome with Massively Parallel Pyrosequencing», *Plant Physiology* 144 :32-42, EE.UU., 2007.
- Whitham, S.; S. Quan; H. Chang; B. Cooper; B. Estes; T. Zhu; X. Wang; Y. Hou: «Diverse RNA Viruses Elicit the Expression of Common Sets of Genes in Susceptible *Arabidopsis thaliana* plants», *The Plant Journal* 33: 271-283, EE.UU., 2003.
- Xiangchao, G.; O. Stegle; J. Behr; J. Steffen; P. Drewe; K. Hildebrand; R. Lyngsoe; S. Schultheiss; E. Osborne; V. Sreedharan; A. Kahles; R. Bohnert; G. Jean; P. Derwent; P. Kersey; E. Belfield; N. Harberd; E. Kemen; C. Toomajian; P. Kover; M. Clark; G. Ratsch; R. Mott: «Multiple Reference Genomes and Transcriptomes for *Arabidopsis thaliana*», *Nature* 1-5, Inglaterra, 2011.
- Xu, L.; L. Zhu; L. Tu; L. Liu; D. Yuan; L. Jin; L. Long; X. Zhang: «Lignin Metabolism Has a Central Role in the Resistance of Cotton to the Wilt Fungus *Verticillium dahliae* As Revealed by RNA-Seq-Dependent Transcriptional Analysis and Histochemistry», *Journal of Experimental Botany* 62: 5607-5621, China, 2011.
- Zenoni, S.; A. Ferrarini; E. Giacomelli; L. Xumerle; M. Fasoli; G. Malerba; D. Bellin; M. Pezzotti; M. Delledonne: «Characterization of Transcriptional Complexity during Berry Development in *Vitis vinifera* Using RNA-Seq», *Plant Physiology* 152: 1787-1795, Italia, 2010.
- Zhang, Z.; Y. Wu; M. Gao; J. Zhang; Q. Kong; Y. Liu; H. Ba; J. Zhou; Y. Zhang: «Disruption of PAMP-Induced MAP Kinase Cascade by a *Pseudomonas syringae* Effector Activates Plant Immunity Mediated by the NB-LRR Protein SUMM2», *Cell Host & Microbe* 11: 253-263, China, 2012.
- Zipfel, C.; G. Kunze; D. Chinchilla; A. Caniard; J. Jones: «Perception of the Bacterial PAMP EF-Tu by the Receptor EFR Restricts Agrobacterium-Mediated Transformation», *Cell*, 125: 749-760, Suiza, 2006.
- Zipfel, C.: «Pattern-Recognition Receptors in Plant Innate Immunity», *Current Opinion in Immunology* 20: 10-16, Inglaterra, 2008.
- Zipfel, C.: «Early Molecular Events in PAMP-Triggered Immunity», *Current Opinion in Plant Biology* 12: 414-420, Inglaterra, 2009.