Secuenciación del genoma de las especies vegetales: implicaciones y perspectivas*

Genome Sequencing in Plant Species: Implications and Perspectives

CITACIÓN: Montoya, C., Ávila, K., Reyes, P., Navia, M. y Romero, H.M. (2014). Secuenciación del genoma de las especies vegetales: implicaciones y perspectivas. *Palmas, 35*(3), 11-22.

PALABRAS CLAVE: palma de aceite, epigenética, interacciones planta-patógeno, marcadores moleculares, mejoramiento vegetal.

KEY WORDS: Oil palm, epigenetics, plant-pathogen interactions, molecular markers, plant breeding.

RECIBIDO: mayo de 2014.

APROBADO: junio de 2014.

* Artículo de investigación e innovación científica y tecnológica.

Carmenza Montoya
Programa de Biología y Mejoramiento de la Palma de Aceite, Cenipalma.

Kelly J. Ávila
Programa de Biología y Mejoramiento de la Palma de Aceite, Cenipalma. Programa de Doctorado en Biotecnología, Universidad Nacional de Colombia.

Paola Reyes
Programa de Biología y Mejoramiento de la Palma de Aceite, Cenipalma.

Mónica Navia
Programa de Biología y Mejoramiento de la Palma de Aceite, Cenipalma.

Hernán Mauricio Romero
Programa de Biología y Mejoramiento de la Palma de Aceite, Cenipalma. Departamento de Biología, Universidad Nacional de Colombia.
hromero@cenipalma.org

Resumen

En 2000 se completó la secuencia del genoma de la primera planta, *Arabidopsis thaliana*, especie que se ha establecido como modelo de investigación. Desde entonces, numerosos genomas se han secuenciado. A finales de 2013 se hizo pública la secuencia del genoma completo de la palma de aceite, por parte del MOPB (*Malaysian Oil Palm Board*) centro de investigaciones del mayor exportador de aceite de palma, Malasia. Con esta revisión se busca poner a la comunidad palmera colombiana al corriente de las investigaciones de vanguardia que se están realizando con el genoma. Se presentan las principales ventajas de la obtención de la secuencia del genoma completo de la palma, así como los nuevos retos que surgen a partir de este logro. Se tocan aspectos del potencial de utilización de estas secuencias como, por ejemplo, la dilucidación del mecanismo genético de herencia del grosor del endocarpio (cuesco), responsable de las variantes fenotípicas *dura, tenera* y *pisifera*. También se describe el potencial de uso de la información de la secuencia completa del genoma de la palma para

estudios de diversidad, evolución, metabolismo, fisiología, etc. Así mismo, se muestran mecanismos de regulación de respuestas bioquímicas y fisiológicas a través de la epigenética y cómo se pueden obtener marcadores moleculares específicos que ayuden en el mejoramiento vegetal de la palma. Con la secuencia genoma de la palma se pueden abrir nuevas líneas de investigación, como el de las interacciones planta-patógeno, que llevan al desarrollo de herramientas novedosas de diagnóstico de enfermedades y métodos de manejo de las principales enfermedades de la palma.

Abstract

In 2000 the first plant whole genome sequence, of *Arabidopsis thaliana* was completed. This is a model species for plant biology research. In the following years several plant genomes were sequenced, and at the end of 2013 the oil palm whole genome sequence was made public by the Malaysian Oil Palm Board (MOPB) the research institute of the number one palm oil exporter in the world, Malaysia. This review aims to bring Colombian oil palm community up to date in the new scientific discoveries related to plant genomes, especially regarding the molecular biology, genetics, biochemistry and physiology. The Advantages derived from deciphering the oil palm whole genome sequence and also the new challenges that come from this scientific achievement are discussed. Aspects of the potential use of the sequences are presented, such as, the understanding of the genetic mechanism behind the shell thickness which involves one gene with punctual mutations responsible of the *dura, tenera* and *pisifera* phenotypes. The potential use of the whole genome sequence for diversity, evolution, metabolism and physiology studies, is also highlighted. It is clear that genome sequence studies open new lines for research; for instance in plant-pathogen interactions it would be possible to develop new tools for disease diagnostics and management.

Introducción

En el ámbito de las ciencias agrícolas y más específicamente en el mejoramiento de especies vegetales de importancia económica tener a mano la información del genoma completo de una especie con importancia económica otorga el poder de conocer estos organismos como nunca antes se había hecho. Pero ¿qué clase de ventaja se puede obtener de esta información? ¿Cómo podemos desde nuestros respectivos campos de acción utilizar ese conocimiento para resolver problemas concretos?

Desde la publicación de la estructura de la molécula de ácido desoxirribonucleico (ADN) y sus implicaciones biológicas (Watson y Crick, 1953) ha quedado demostrado que las células que conforman un individuo se diferencian en sus funciones, pero todas contienen la misma información genética la cual se almacena en los cromosomas que contiene el núcleo celular. De manera simplificada, en la mayoría de organismos cada cromosoma es una doble cadena de ADN constituida por cuatro nucleótidos: Adenina (A), Citosina (C), Guanina (G) y Timina (T).

Existen secuencias ordenadas en el ADN que determinan de manera muy precisa la proteína a formar. Esto es lo que conocemos como un gen o unidad básica de expresión genética y herencia (Krebs *et al.*, 2009). También existen secuencias no-codificantes que representan una fracción importante del genoma y cumplen una función de regulación en la expresión (Lockton y Gaut, 2005).

El flujo de la información genética está coordinada de tal manera que la molécula de ADN es recodificada en ácido ribonucleico (ARN) mensajero, otra molécula intermedia que guarda información de forma pasajera, y esta información es decodifica-

da por los llamados ribosomas para conformar una secuencia de aminoácidos que se constituirán en proteínas, las cuales tienen función biológica. Esto lo conocemos como el Dogma Central de la Biología (Crick, 1970) y es uno de los pilares en la investigación de los sistemas biológicos.

Las anteriores reacciones en cadena nos presentan los tres principales niveles de expresión en un genoma: en primer nivel el genoma, que corresponde a la secuencia base de ADN, el cual es trascrito a ARN mensajero y finalmente es traducido a proteína, una molécula con actividad biológica.

El descubrimiento de las estructuras moleculares de los genes permite precisar la relación entre la secuencia nucleotídica y la función misma, todo esto haciendo parte de una complejidad biológica que comprende diferentes niveles (Figura 1). A manera de metáfora, una revisión sobre el tema nos compara los genes con subrutinas de lo que sería ese gran sistema operativo que es el genoma (Gerstein *et al.*, 2007).

Un gen tiene una estructura definida pero al mismo tiempo versátil que le permite expresarse y producir una secuencia proteica con la actividad biológica requerida. De manera simplificada, en un gen podemos identificar tres tipos de segmentos de ADN que son: secuencias que codifican (exones), secuencias que no codifican (intrones) y secuencias reguladoras (promotores) (Krebs *et al.*, 2009).

Lo que finalmente se observa, el fenotipo, viene de una reacción en cadena bastante compleja en la cual se decide cuáles genes han de expresarse y dentro del mismo gen, cuáles han de ser las secuencias codificadoras que participarán en la formación de la nueva proteína. Es de resaltar que generalmente un complejo proteico consta de "grupos" de proteínas, cada una codificada por genes distintos. Así que podemos comenzar a intuir el nivel de complejidad y organización que hay detrás de una sola proteína y qué decir de caracteres como el tamaño de la pulpa, contenido de aceite en fruto, número de racimos por palma o producción de aceite.

Los casos anteriormente expuestos pueden estar controlados por la expresión simultánea de varios genes y, a su vez, influenciados por factores medioambientales. Adicionalmente existen los denominados factores de transcripción, proteínas que actúan uniéndose a secuencias concretas del ADN para activar o reprimir

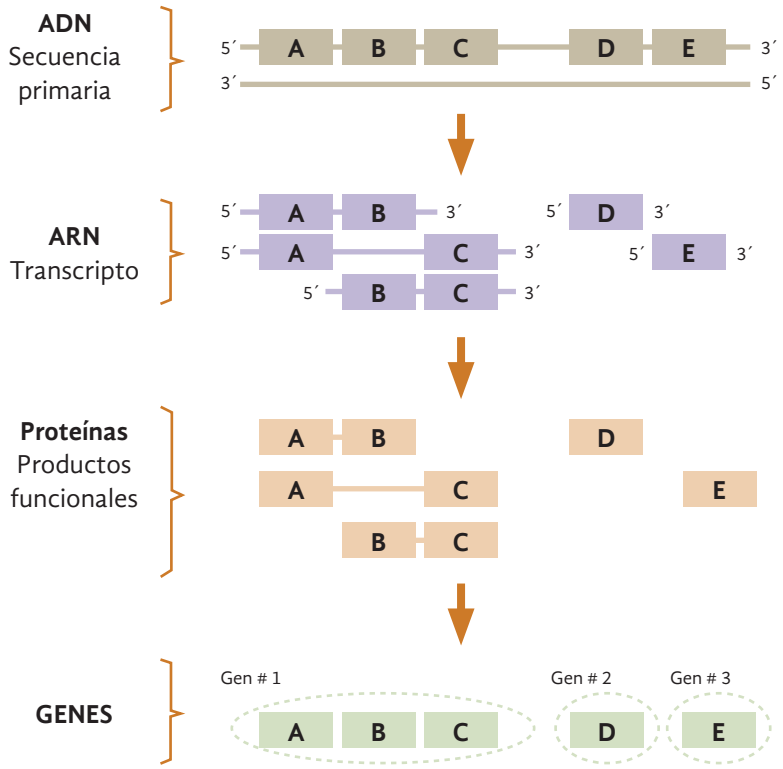

Figura 1. Representación de los niveles de organización y expresión génica a partir de las secuencias de ADN en el genoma.

A manera de ilustración, en un primer nivel las secuencias de ADN agrupadas en segmentos especiales (A, B, C, D y E). Estos segmentos serán transcritos (recodificados) a una molécula de ARN, no necesariamente siguiendo el mismo orden lineal observado en el ADN. Los transcritos pasarán a la "fábrica de proteínas" (los ribosomas) y se constituirán en productos con una actividad biológica específica. En el último nivel, se pueden observar los diferentes genes asociados a las proteínas producidas y su relación con la secuencia primaria de ADN. Diagrama con modificaciones tomado de Gerstein *et al.*, (2007).

la expresión de diversos genes. Es tan importante el rol de estos factores que errores en su funcionamiento pueden generar la alteración del ciclo celular y el crecimiento descontrolado de células; es decir, generar patologías como el cáncer (Colburn y Kensler, 2008).

Considerando lo anterior, el contar con la secuencia base del ADN de un ser vivo brinda una herramienta muy poderosa de trabajo ya que se cuenta con un punto muy certero de partida para indagar: ¿Por qué unos genes se expresan y otros no? ¿Qué determina que una enzima sea funcional? ¿Cómo influye el medio ambiente en la expresión de ciertos caracteres? ¿Por qué ciertos patógenos son más agresivos en determinadas especies? ¿Cómo se agrupan y regulan los genes involucrados en defensa a patógenos? Todas las preguntas planteadas han sido objeto de estudio desde hace mucho tiempo y la ciencia conoce algunas de las respuestas, pero aún falta un total dominio de la complejidad de las interacciones biológicas para, por ejemplo, "diseñar" una planta que responda eficazmente a las necesidades de una industria en particular.

Esta revisión tiene como objetivo general describir las aplicaciones básicas de la información obtenida a partir de genomas secuenciados en la resolución de problemas agronómicos específicos. De esta manera, se busca tener un contexto más robusto para poder hacer uso de la información, entender la importancia de estas herramientas, conocer su disponibilidad y futuras aplicaciones en las áreas requeridas.

El genoma y la agricultura

La secuenciación del genoma completo de un organismo dado permite obtener el catálogo completo de los genes de la especie, de los elementos regulatorios que controlan sus funciones y, adicionalmente, brinda un marco de trabajo para entender la variabilidad intra e interespecífica de diferentes especies dentro del mismo género (Feuillet *et al*., 2011). Es así como en la actualidad, la secuenciación de genomas ha generado un cambio de enfoque en cómo se entiende la biología de las plantas y cómo se ejecuta el mejoramiento vegetal. Ante los retos que representa una obtención sostenible de comida, fibras y combustible en el mundo moderno, la genómica es una herramienta fundamental que debe ser desarrollada (Feuillet *et al*., 2011).

En el mundo de los cultivos oleaginosos, las especies más relevantes a nivel económico cuentan con la secuencia del genoma de interés. Actualmente es posible acceder al genoma de la soya (*Glycine max*) que tiene un tamaño de 1,1 gigabases (Gb) (Schmutz *et al*., 2010), del girasol (*Helianthus annuus* L.) con 3,5 Gb (Staton *et al*., 2012) o del maíz (*Zea mays ssp. mays* L.) con 2,3 Gb (Schnable *et al*., 2009). Por su parte, el genoma de la canola (*Brassica napus*) fue anunciado en 2009 por Bayer Crop Science (Alemania), pero hasta el momento no ha sido publicado (Snowdon y Iniguez Luy, 2012). Sin embargo, es posible acceder al genoma de la especie cercana *Brassica rapa* (Wang *et al*., 2011). En lo concerniente a palma de aceite (género *Elaeis*), en 2009 el MPOB (Malaysian Palm Oil Board) y el grupo Orion Genomics anunciaron la obtención de la secuencia del genoma de las especies *E. guineensis* y *E. oleifera*. Finalmente, en julio de 2013 este grupo publicó en la revista *Nature* el artículo científico donde revelan los hallazgos más significantes en torno a lo que constituirá el genoma de referencia para el género *Elaeis* (Singh *et al*., 2013b). Con el nuevo conocimiento, ahora se sabe que este genoma tiene un tamaño de 1,8 Gb, con una predicción de 34.802 genes.

La Corporación Centro de Investigación en Palma de Aceite (Cenipalma) desde 2009 hace parte de una iniciativa de investigación conocida como el OPGP (Oil Palm Genome Project), que cuenta con la participación de 16 entidades públicas y privadas coordinadas por un grupo de trabajo perteneciente al CIRAD (Francia). Este proyecto consiste básicamente en explorar y comparar los patrones de expresión de los genes asociados a los caracteres agronómicos con más relevancia para la industria de la palma de aceite. Entre los principales objetivos se encuentra generar bases de datos que muestren los patrones de expresión genética de las palmas bajo diferentes condiciones de estrés biótico y abiótico, por ejemplo, la respuesta a patógenos como *Ganoderma boninense*. Adicionalmente, generar mapas genéticos de la población de estudio y encontrar las zonas del genoma responsables de la variación de algunos caracteres cuantitativos (QTL, Quantitative Trait Loci, por sus siglas en ingles), como la variación en la composición de ácidos grasos. El proyecto en mención ha tenido en cuenta las dos especies del género *Elaeis* así como diferentes orígenes

al interior de cada especie. La publicación del genoma de la palma aporta a este proyecto en la medida que contribuye al reservorio de base de datos (un genoma de referencia) para cotejar y examinar la información mediante análisis bioinformáticos.

Lo novedoso y los retos detrás de la secuenciación del genoma

El desarrollo de tecnologías de secuenciación cada vez más potentes y la disponibilidad de la información obtenida ha permitido que los genomas de plantas sean más completos y accesibles para investigaciones a pequeña escala. Las tecnologías de secuenciación de nueva generación pueden producir desde un borrador de secuencia de un genoma con una significante reducción de costos y tiempo, hasta genomas completos que además pueden ser usados para evaluar la estructura de genes y su expresión (Berkman *et al*., 2012; Feuillet *et al*., 2011). Sin embargo, el principal cuello de botella consiste en convertir la información en conocimiento que pueda ser aplicado a programas de mejoramiento en especies vegetales cultivables, sin olvidar que el análisis de estos datos que representa por su cantidad y complejidad es un reto en sí mismo (Feuillet *et al*., 2011). Al respecto, podemos evocar los alcances de las tecnologías de transformación genética (transgénesis), las cuales con casi 30 años de desarrollo aún presentan dificultades en lograr la correcta inserción del gen de interés en el lugar exacto. Más allá de eso, lograr el perfecto equilibrio entre promotores/silenciadores y la regulación de la expresión "en concierto" con todo el genoma. Actualmente, las nuevas tecnologías apuntan a sintetizar *de novo* segmentos de la molécula de ADN, empaquetarla en el equivalente a un minicromosoma artificial e introducirlo al genoma de interés (Liu *et al*., 2013). En este punto, se está ante otro de los retos detrás de la secuenciación de un genoma, ¿qué se espera en el campo de la transgénesis?

La publicación del genoma de la palma de aceite señala que la evidencia filogenética muestra una divergencia entre la palma datilera (*Phoenix dactylifera*) y palma de aceite de 65 millones de años (Ma), mientras que el tiempo de divergencia entre *E. guineensis* y *E. oleifera* es de 51 Ma. Lo anterior implica que *E. guineensis* diverge evolutivamente en el "nuevo mundo" dando origen a la palma de aceite americana. Estos tiempos son coherentes con aquellos estimados en el área de la geociencia, ya que de 116 a 120 millones de años es el tiempo de separación entre el continente africano y lo que corresponde hoy a Suramérica (Torsvik *et al*., 2009). Es tal la importancia de este hallazgo, que es interesante resaltar que la publicación en sí pertenece al ámbito de la biología evolutiva y que su título no se devela como la publicación del genoma del género *Elaeis* sino como: "La secuencia del genoma de la palma de aceite revela la divergencia de las especies interfértiles en el viejo y nuevo mundo" (-*Oil palm genome sequence reveals divergence of interfertile species in old and new worlds*-). Otro dato relevante es el porcentaje de homología genética entre palma de aceite y palma datilera de 94,4 %, mientras que con *Arabidopsis thaliana* es del 80,2 %. Esta última especie es la planta modelo en la investigación vegetal y cuenta con estudios muy completos en torno a la síntesis de lípidos y sus factores de regulación (Baud y Lepiniec, 2009; To *et al*., 2012). En la publicación de Singh *et al*., (2013b), podemos encontrar los niveles de expresión de los principales genes involucrados en síntesis de lípidos antes y durante la fase de máxima acumulación de aceite en pulpa y almendra. Lo anterior, utilizando como base el genoma de *A. thaliana* para tener certeza de la identidad de los genes implicados.

De los nuevos conocimientos ofrecidos en la publicación, se encuentra la comparación de los mapas genéticos (mapas definidos mediante unidades de frecuencia de recombinación entre cromosomas) obtenidos previamente, con el mapa físico de cada cromosoma y los respectivos marcadores moleculares, o sea, aquellas secuencias de referencia para una región específica de ADN (Maheswaran, 2004). Ya que la identificación de genes asociados a caracteres agronómicos claves puede ayudar a acelerar los procesos de mejoramiento vegetal, contar con un mapa genético saturado, es decir, donde cada punto de referencia está a 0,1 o 0,5 centimorgans (cM), entendido como la unidad de distancia en un cromosoma entre dos posiciones, entre las cuales hay 0,01 eventos de recombinación genética en una generación, facilita esta tarea (Blenda *et al*., 2012). Lo anterior abre la puerta para hacer un seguimiento a variedades con mejor rendimiento, calidad, tolerancia a condiciones ambientales desfavorables y resistencia a enfermedades (Edwards y Batley, 2010). En lo concerniente a palma

de aceite, ahora conocemos el tamaño de cada cromosoma (pares de bases y longitud en cM) y se ha determinado, por ejemplo, que el gen responsable del grosor de cuesco (gen *SHELL*) se encuentra físicamente en el cromosoma 7 (Singh *et al*., 2013a). Entonces, otro objetivo del grupo de investigación del MPOB fue dilucidar a nivel molecular las diferencias del gen SHELL (*Sh*) entre los frutos *dura*, *tenera* y *pisifera*, utilizando la información del genoma de *Elaeis*. De manera simplificada y según la información encontrada en Singh *et al*., (2013a) este gen es homólogo del gen *SEEDSTICK*, el cual es responsable del desarrollo del óvulo y la semilla en *A. thaliana*. Cuando este gen se expresa normalmente (frutos tipo *dura*) es el responsable de la pared gruesa con alto contenido de lignina (lo que se llama cuesco) alrededor de la almendra. Ahora, en los frutos tipo *pisifera*, la substitución de una sola base de ADN resulta en el cambio de una lisina por una asparagina lo cual desemboca en un cambio drástico en la proteína, impidiendo la formación del cuesco alrededor de la almendra. Al contrario, los frutos tipo *tenera* tienen una versión normal de este gen y la otra contiene la mutación; esta condición a nivel fenotípico se traduce en un cuesco de grosor intermedio y mayor cantidad de mesocarpo en fruto. De los ocho exones que componen el gen SHELL, una mutación puntual en la región del exón 1 es la responsable del cambio. Con esta información, usando herramientas de biología molecular, sería posible identificar en etapas tempranas (vivero) cuál será el tipo de fruto y así decidir qué materiales se conservarán con fines comerciales o de mejoramiento vegetal.

Otra característica importante estudiada a nivel molecular es el color del fruto a lo largo de su maduración. En los cultivos de palma de aceite comercial hay dos tipos de patrones de coloración: el *nigrescens* en el cual el fruto va de violeta oscuro a negro, con un cambio de color mínimo en su proceso de maduración y el *virescens*, que pasa del color verde en frutos inmaduros a un rojo intenso al final de la maduración. En el caso de *nigrescens*, la identificación de la madurez en un racimo se da por el número de frutos desprendidos del mismo que caen al suelo, llevando a una posible pérdida de material y tiempo. En *virescens*, el color rojo indica el momento adecuado de la cosecha, lo que la hace una característica deseable para optimizar la recolección.

Singh *et al*., 2014, describieron que el fenotipo *virescens* corresponde a la presencia de cinco tipos de mutaciones alélicas de carácter dominante, específicas del gen *VIR* involucrado en la síntesis de antocianinas, pigmento responsable de la coloración oscura del fruto. El fenotipo *nigrescens* es un carácter recesivo y se da solamente cuando la secuencia silvestre se presenta en los dos cromosomas. Aunque la mayoría de palmas cultivadas presenta este último fenotipo, el conocer la naturaleza genética de este carácter, permitirá la selección desde la semilla para establecer programas de mejoramiento que aumenten la frecuencia del fenotipo *virescens* en los materiales cultivables, facilitando la optimización de la cosecha (Singh *et al*., 2014).

El genoma de un organismo posee mucha información; no obstante, para que esta sea útil debe estar en un contexto determinado. Solo mediante la comparación con otros genomas se puede entender cómo estos organismos han evolucionado, cuáles genes han surgido a lo largo de la historia y pueden ser determinantes de características fenotípicas particulares, lo que se conoce como genómica comparativa. Sin embargo, muchas veces la presencia o ausencia de un gen no dice tanto como los patrones de expresión del mismo. Un ejemplo claro de esto es la síntesis de ácidos en *E. guineensis* y *P. dactylifera* (palma datilera). Al comparar los genomas de la palma aceitera con la palma datilera solo 5,6 % de los genes que posee la palma de aceite difieren en función de los de *P. dactylifera*. Adicionalmente, la cantidad de genes involucrada en la síntesis de triglicéridos de las dos palmas es similar (Singh *et al*., 2013b). Sin embargo, la palma de aceite acumula hasta 90 % de aceite en su mesocarpo mientras que la palma datilera acumula casi exclusivamente azúcares. ¿A qué puede deberse esta diferencia? Análisis de expresión diferencial en el mesocarpio de las dos palmas muestran que la expresión de genes relacionados con lípidos son similares, pero la expresión del subgrupo de genes implicados en síntesis de ácidos grasos en el plastidio se encuentra más elevada en *E. guineensis*, especialmente aquellos genes relacionados con transporte, enzimas glicolíticas y un factor de transcripción que juega un rol importante en controlar la cantidad de aceite, conocido como WRI1 (Bourgis *et al*., 2011). Estos resultados nos muestran que la regulación de la expresión génica en la fase de síntesis de lípidos es determinante para el alto contenido de aceite en los frutos de *E. guineensis*,

más que el ensamblaje de las moléculas en forma de triglicéridos (Bourgis *et al*., 2011; Singh *et al*., 2013b). Todo lo anterior no hubiera podido ser demostrable basándose solo en datos genómicos.

¿Qué más se puede hacer con la información de un genoma secuenciado?

La identificación de todos los genes de una especie abre la posibilidad de entender cómo importantes caracteres agronómicos son controlados, conocimiento que puede ser utilizado para mejorar la producción de dicha especie. La secuenciación completa del genoma es el método más robusto para realizar estudios de diversidad genética en una población y para entender mejor la relación entre el genoma heredado y los caracteres hereditarios observados (Edwards y Batley, 2010).

Diversidad, evolución y metabolismo

Leer el "manual de instrucciones" de un organismo, compararlo con los individuos a nivel intra e interespecie e identificar las diferencias permite establecer las relaciones evolutivas entre estas. Reconstruir la evolución de las especies y entender su similitud cobra importancia en el estudio de los híbridos interespecíficos, por ejemplo, en el género *Elaeis* ya que permite entender cómo dos especies distintas pueden reproducirse sexualmente. Esta información también permite desarrollar lo que se conoce como: estudios de asociación de genoma completo. De manera simple, estudios que buscan la asociación entre el genotipo y el locus (posición del gen en el cromosoma) responsable de un carácter agronómico específico (Edwards y Batley, 2010). Por ejemplo, un estudio por Huang *et al*., (2010) en arroz (*Oriza sativa*) utilizando el genoma de referencia de la especie logró identificar las posiciones del genoma (loci) que explican aproximadamente el 36 % de la variación de 14 caracteres agronómicos de importancia comercial.

Epigenética y regulación génica

La epigenética (del griego *epi*, en o sobre, y -genética) es el estudio de cambios heredables en la función génica que se producen sin un cambio en la secuencia base del ADN y que pueden atribuirse a varios mecanismos que favorecen la expresión génica o la reprimen, generando lo que se conoce como silenciamiento génico (Bird, 2002; Cavagnari, 2012). En palma de aceite, la producción de clones *in vitro* se ve afectada por una anomalía floral en la etapa reproductiva conocida como *mantled* para la cual existen evidencias de que corresponde a una alteración epigenética (Jaligot *et al*., 2011, 2000). Antes de tener disponible el genoma, varios estudios se habían enfocado en la búsqueda de marcadores moleculares que permitieran identificar tempranamente el problema, sobre todo con la generación de marcadores moleculares asociados a la expresión (EST, acrónimo del inglés *Expressed Sequence Tag*). Como lo expresó Beulé *et al*., (2010) mientras que el genoma de la especie no esté disponible, la generación de EST ayuda a enriquecer la disponibilidad de recursos genéticos para dar solución al problema. Actualmente, con la liberación del genoma de la palma de aceite se podría facilitar el mapeo genético de las regiones directamente implicadas en esta anomalía (Singh *et al*., 2013b). En el caso de *Arabidopsis,* diferentes técnicas de secuenciación han sido diseñadas exclusivamente para ADN con patrones de metilación (condición epigenética) y se han utilizado en el estudio de la epigenética de esta especie (Lister *et al*., 2009). Una mayor comprensión de la modificación epigenética de los genomas y del impacto sobre la expresión de genes permite a futuro el desarrollo de programas de mejoramiento de una manera más rápida y precisa (Edwards y Batley, 2010)

Marcadores moleculares y mejoramiento vegetal

Uno de los usos más importantes para los mejoradores vegetales es el desarrollo de marcadores moleculares que faciliten y aceleren el proceso de mejora vegetal. Según la revisión presentada por Feuillet *et al*., (2011) en el caso del arroz, un resultado directo de este esfuerzo fue el incremento en el mapeo de QTL relacionados a caracteres de importancia agronómica. Por ejemplo, se han diseñado marcadores basados en la secuenciación de la región del gen de la resistencia a añublo del arroz (denominado *pi21*) usado para identificar recombinantes entre *pi21* y otro gen

ubicado a 37 kb de distancia que confiere baja calidad nutritiva. Con esta información, fue posible por primera vez en décadas seleccionar variedades de arroz que combinaran resistencia al añublo y alta calidad nutritiva. Otro ejemplo se encuentra en el sorgo (*Sorghum bicolor*), según Mace y Jordan (2010) se acopló la información de diferentes mapas genéticos a la secuencia del genoma publicado (Paterson et al., 2009) logrando la integración de 35 genes de efectos mayores en un mapa único, generando la base para que los mejoradores la relacionen fácilmente con los estudios previos de QTL y facilitar ciertos pasos en los programas de mejoramiento. Especial énfasis se hace en este artículo en la publicación de un mapa genético conceso que se puede convertir en una de las herramientas de trabajo para laboratorios o países en vía de desarrollo donde no es fácil generar tales resultados.

Las secuencias de genomas de referencia para varias especies de plantas cultivables están ahora disponibles y esta información permite la rápida identificación de genes candidatos a través de análisis bioinformáticos y a través de la comparación de la referencia con datos de secuencias de diferentes cultivares (Edwards y Batley, 2010).

Estudios de interacción planta-patógeno

Las plantas carecen de un sistema inmune específico que les permita responder de manera sistémica ante el ataque de un organismo. Por el contrario, la respuesta a los patógenos es de cierta manera dada por cada célula vegetal de manera individual. En la práctica, los mecanismos moleculares que se tejen alrededor de la respuesta de la planta son altamente complejos, dinámicos y están en un proceso de coevolución constante debido a la presión que ejerce el patógeno. El patógeno también debe contar con una gran diversidad de mecanismos de infección para colonizar una planta y así poder sobrevivir. En el nivel más amplio, la resistencia vegetal se ha dividido en dos ramas: una compuesta por barreras preformadas (pared, cutículas, etc.) y la resistencia química la cual depende de las proteínas que sintetiza la planta, es decir, depende de los genes con que cuenta la especie.

La resistencia en plantas se ha descrito desde la aproximación de lo que se conoce como sistemas heterólogos, en los cuales la especie modelo ha sido *A. thaliana*. Teniendo en cuenta que cada especie es diferente y que dentro de la misma especie su respuesta puede cambiar según el tipo de patógeno y condiciones ambientales, muchas de las respuestas que se observan en *A. thaliana* no son fácilmente extrapolables a otra especie vegetal, de allí la importancia de tener el genoma de la especie blanco.

En el campo de la fitopatología molecular, un objetivo de estudio ha sido la identificación de los genes de resistencia (Genes R), no solo para una especie en particular sino también desde la perspectiva de sus diferencias entre plantas monocotiledóneas y dicotiledóneas. Cuando se estudian los genes R (clúster de resistencia) se deben ver desde la perspectiva de la secuencia, tipo de gen y organización genómica. Además de su implicación en la evolución de los mecanismos de respuesta de la planta y la interacción con el efector, una proteína que contribuye a la virulencia y hace efectivo el ataque de parte del patógeno (Kim *et al.*, 2012).

Sin embargo, con todos los hallazgos que se han realizado aún no hay conclusiones definitivas respecto a la inmunidad vegetal. Las vías de resistencia de las plantas son múltiples y diversas. Esta falta de consenso informativo no tiene por qué desanimar, por el contrario, es una invitación para que los investigadores diseñen modelos biológicos precisos y acordes a sus necesidades con el fin que puedan responder preguntas de investigación.

Las tecnologías de secuenciación de última generación han abierto nuevas posibilidades para un análisis completo del transcriptoma (aquella porción del genoma que es expresada) en la especie de interés (Edwards y Batley, 2010). Entender el transcriptoma es esencial para la interpretación de los elementos funcionales de un genoma bajo ciertas condiciones de desarrollo o enfermedad. Por lo tanto, la transcriptómica ha revolucionado la forma de estudiar los procesos biológicos (Wang *et al.*, 2009) y en especial las interacciones planta-patógeno.

Se hace necesario entonces, generar análisis transcriptómicos en múltiples tiempos durante la infección, que ofrezcan la caracterización de la respuesta molecular que se desencadena en el hospedero, permitiendo entender cómo cambian las redes genéticas en el tiempo y en el proceso de infección (Sturdevant *et al.*, 2010). En palma de aceite se pue-

den encontrar dos ejemplos recientes de este tipo de investigación, concretamente del tipo de respuesta que se genera durante la infección con *Ganoderma boninense,* el hongo patogénico responsable de la Pudrición basal de estípite (Tee *et al*., 2012; Yeoh *et al*., 2013). Esta información contribuye al entendimiento de los mecanismos de defensa y plantea como perspectiva el desarrollo de marcadores moleculares que permitan tamizar materiales por su tolerancia a *G. boninense*.

De su parte, Cenipalma, en el área de la transcriptómica, ha desarrollado un proyecto de investigación encaminado a dilucidar los patrones de expresión génica en respuesta al ataque de *Thielaviopsis paradoxa*, para identificar los principales genes y/o rutas metabólicas activadas durante el proceso de defensa (Riaño-Pachón *et al*., 2013). Analizar condiciones de metabolismo bajo ciertas condiciones de desarrollo o de estrés, con tecnologías de secuenciación se ha convertido en una herramienta para describir procesos moleculares altamente complejos y contar con la secuencia de un genoma de referencia permite asignar un nombre específico a ese gen que interesa por su patrón de expresión. Lo anterior, a través de procesos bioinformáticos que comparan las homologías de las secuencias de ADN entre el genoma de referencia y aquellas secuencias expresadas que constituyen una incógnita al inicio de la investigación.

Sin duda alguna, en el futuro se podrá tener información valiosa para el gremio palmicultor, la cual depende del esfuerzo y del trabajo en equipo en donde cada disciplina aporte su experticia y conocimiento para integrar resultados con un fin común. Finalmente y, de acuerdo con lo sugerido por Tripathy *et al*., (2012) y Werner, (2010), la información que se obtiene de estas plataformas no se puede quedar solamente en el gran volumen de información o en el avance bioinformático, esta debe tener una utilidad biológica/agronómica. Para terminar, en la Figura 2 se esquematiza de forma general el camino desde la secuenciación hasta las posibles aplicaciones de la información obtenida.

Conclusiones y perspectivas

Con la publicación del genoma de la palma de aceite, las investigaciones serán más precisas y dirigidas a un objetivo en particular, sin decir con esto que todos los problemas serán solucionados, ya que no hay que olvidar que la información del genoma secuenciado es apenas una puerta de entrada a un mundo de investigación que no ha sido explorado y que requiere ir generando conocimiento de base, en el cual en un futuro no lejano permita diseñar planes de mejoramiento y estrategias de control eficientes.

Las líneas tradicionales de investigación de la palma cuentan ahora con una herramienta muy podero-

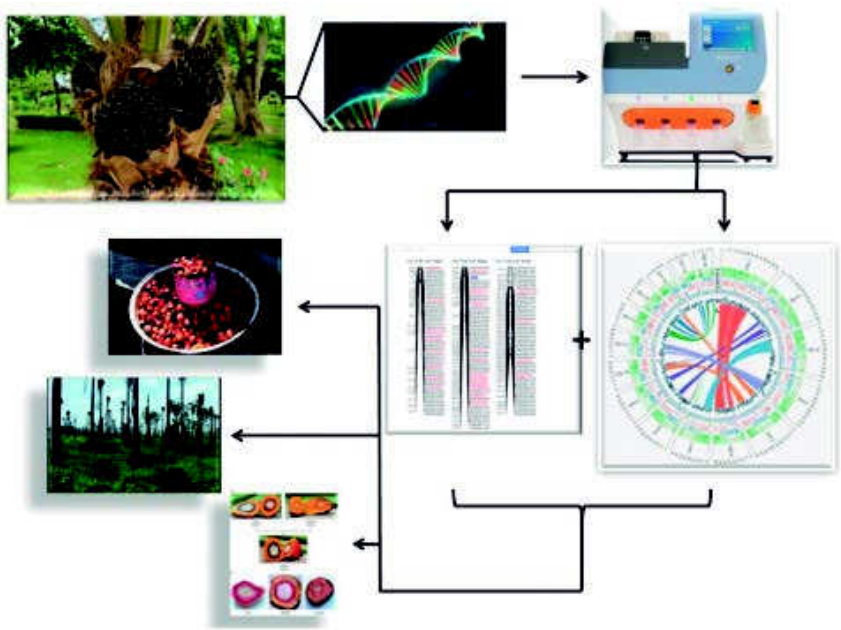

Figura 2. Esquema ilustrativo de los pasos y resultados esperados detrás de la secuenciación del genoma del género *Elaeis*.

En el sentido de las manecillas del reloj: la palma, su secuencia de ADN, los resultados principales como lo son el mapa físico con los 16 cromosomas y el mapa genético con los marcadores moleculares de interés. Para terminar, tres áreas que son prioridad en los estudios de genómica a venir: la producción de aceite, la resistencia a enfermedades y la selección temprana del tipo de fruto mediante protocolos de biología molecular. Algunas imágenes tomadas de Singh *et al.*, (2013a, 2013b).

sa que les permitirá trabajar con objetivos más claros y precisos y seguramente a una velocidad mucho mayor para obtener las soluciones que Colombia necesita para los problemas de la palma. Así mismo, estas herramientas permitirán el desarrollo de programas de mejoramiento que produzcan variedades con una tolerancia más duradera a las enfermedades y que serán obtenidas en menor tiempo en comparación con el mejoramiento clásico.

La publicación del genoma del género *Elaeis* marca un hito en la historia del mejoramiento vegetal de la palma de aceite. Queda pendiente iniciar el camino de compresión de las complejas interacciones entre genes, sus factores de regulación, ¿cómo hacerlo aplicable en el ámbito agronómico?, más todo aquello que es heredable por fuera de la secuencia del ADN, es decir, lo epigenético. Interesante el camino que aún falta recorrer.

Agradecimientos

Los autores agradecen a Fedepalma-Fondo de Fomento Palmero por el apoyo en la realización de esta investigación.

Bibliografía

Baud, S. & Lepiniec, L. (2009). Regulation of de novo fatty acid synthesis in maturing oilseeds of Arabidopsis. *Plant Physiol Biochem, 47,* 448–455.

Berkman, P.J., Lai, K., Lorenc, M.T. & Edwards, D. (2012). Next-generation sequencing applications for wheat crop improvement. *Am. J. Bot. 99,* 365–371.

Beulé, T., Camps, C., Debiesse, S., Tranchant, C., Dussert, S., Sabau, X., Jaligot, E., Alwee, S.S.R.S. & Tregear, J.W. (2010). Transcriptome analysis reveals differentially expressed genes associated with the mantled homeotic flowering abnormality in oil palm (*Elaeis guineensis*). *Tree Genet Genomes, 7,* 169–182.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev, 16,* 6–21.

Blenda, A., Fang, D.D., Rami, J.-F., Garsmeur, O., Luo, F. & Lacape, J.-M. (2012). A High Density Consensus Genetic Map of Tetraploid Cotton That Integrates Multiple Component Maps through Molecular Marker Redundancy Check. *PLoS One 7,* e45739.

Bourgis, F., Kilaru, A., Cao, X., Ngando-Ebongue, G.-F., Drira, N., Ohlrogge, J.B., Arondel, V. (2011). Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning. *Proc. Natl. Acad. Sci. U. S. A. 108,* 12527–32.

Cavagnari, B.M. (2012). Regulación de la expresión génica: cómo operan los mecanismos epigenéticos. *Arch Argent Pediatr, 110,* 132–136.

Colburn, N.H. & Kensler, T.W. (2008). Targeting transcription factors for cancer prevention--the case of Nrf2. *Cancer Prev Res 1*, 153–155.

Crick, F. (1970). Central Dogma of Molecular Biology. *Nature 227,* 561–563.

Edwards, D. & Batley, J. (2010). Plant genome sequencing: applications for crop improvement. *Plant Biotechnol. J. 8,* 2–9.

Feuillet, C., Leach, J.E., Rogers, J., Schnable, P.S. & Eversole, K. (2011). Crop genome sequencing: lessons and rationales. *Trends Plant Sci. 16,* 77–88.

Gerstein, M.B., Bruce, C., Rozowsky, J.S., Zheng, D., Du, J., Korbel, J.O., ... Snyder, M. (2007). What is a gene, post-ENCODE ? History and updated definition. *Genome Res 17,* 669–681.

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., ...Han, B. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet. 42,* 961–967.

Jaligot, E., Adler, S., Debladis, É., Beulé, T., Ilbert, P., Richaud, F., ... Beule, T., 2011. Epigenetic imbalance and the floral developmental abnormality of the in vitro-regenerated oil palm *Elaeis guineensis*. *Ann. Bot. 108,* 1453–1462.

Jaligot, E., Rival, a., Beulé, T., Dussert, S. & Verdeil, J. L. (2000). Somaclonal variation in oil palm (*Elaeis guineensis* Jacq.): the DNA methylation hypothesis. *Plant Cell Rep. 19,* 684–690.

Kim, J., Lim, C., Lee, B.-W., Choi, J.-P., Oh, S.-K., Ahmad, R., ... Hur, C.-G., (2012). A genome-wide comparison of NB-LRR type of resistance gene analogs (RGA) in the plant kingdom. *Mol. Cells 33,* 385–392.

Krebs, J.E., Goldstein, E.S., Kilpatrick, S.T., 2009. Lewin's GENES X, 10th ed. Jones & Bartlett Learning, Sudbury, Massachusetts.

Lister, R., Gregory, B.D. & Ecker, J.R. (2009). Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. Curr. Opin. *Plant Biol. 12,* 107–18.

Liu, W., Yuan, J.S., Stewart Jr, C.N. & Stewart, C.N.J. (2013). Advanced genetic tools for plant biotechnology. *Nat. Rev. Genet. 14,* 781–793.

Lockton, S. & Gaut, B.S. (2005). Plant conserved non-coding sequences and paralogue evolution. *Trends Genet. 21*, 60–5.

Mace, E.S. & Jordan, D.R. (2010). Location of major effect genes in sorghum (Sorghum b(icolor (L.) Moench). *Theor. Appl. Genet. 121,* 1339–56.

Maheswaran, M. (2004). Molecular Markers: History, Features History, Features and Applications. *Adv. Biotech,*17–24.

Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, ... Rokhsar, D.S. (2009). The Sorghum bicolor genome and the diversification of grasses. *Nature 457,* 551–6.

Riaño-Pachón, D.M., Gaitán, S.L. & Romero, H.M., 2013. Sequencing and annotation of the transcriptome of Oil Palm during the interaction with Thielaviopsis paradoxa. In: *Segundo Congreso Colombiano de Biología Computacional Y Bioinformática. Manizales, Colombia.*

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., ... Jackson, S.A. (2010). Genome sequence of the palaeopolyploid soybean. *Nature, 463,* 178–183.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., ... Wilson, R.K. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science, 326,* 1112–5.

Singh, R., Low, E.-T.L., Ooi, L.C.-L., Ong-Abdullah, M., Nookiah, R., Ting, N.-C., ... Sambanthamurthi, R. (2014). The oil palm VIRESCENS gene controls fruit colour and encodes a R2R3-MYB. *Nat. Commun, 5,* 4106.

Singh, R., Low, E.-T.L., Ooi, L.C.-L., Ong-Abdullah, M., Ting, N.-C., Nagappan, J., ... Martienssen, R.A., (2013ª). The oil palm SHELL gene controls oil yield and encodes a homologue of SEEDSTICK. *Nature 0,* 1–7.

Singh, R., Ong-Abdullah, M., Low, E.-T.L., Manaf, M.A.A., Rosli, R., Nookiah, R., ... Sambanthamurthi, R., 2013b. Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature 500*, 335–339.

Snowdon, R.J. & Iniguez Luy, F.L. (2012). Potential to improve oilseed rape and canola breeding in the genomics era. *Plant Breed. 131,* 351–360.

Staton, S.E., Bakken, B.H., Blackman, B.K., Chapman, M. a, Kane, N.C., Tang, S., Ungerer, M.C., ... Burke, J.M. (2012). The sunflower (Helianthus annuus L.) genome reflects a recent history of biased accumulation of transposable elements. *Plant J. 72,* 142–53.

Sturdevant, D.E., Virtaneva, K., Martens, C., Bozinov, D., Ogundare, O.M., Castro, N., Kanakabandi, K. & Beare, P.A. (2010). Host–microbe interaction systems biology: lifecycle transcriptomics and comparative genomics. *Futur. Microbiol 5,* 205–219.

Tee, S.-S., Tan, Y.-C., Abdullah, F., Ong-Abdullah, M. & Ho, C.-L. (2012). Transcriptome of oil palm (*Elaeis guineensis* Jacq.) roots treated with Ganoderma boninense. *Tree Genet Gen 9,* 377–386.

To, A., Joubès, J., Barthole, G., Lécureuil, A., Scagnelli, A., Jasinski, S., Lepiniec, L. & Baud, S. (2012). WRINKLED transcription factors orchestrate tissue-specific regulation of fatty acid biosynthesis in Arabidopsis. *Plant Cell 24,* 5007–5023.

Torsvik, T.H., Rousse, S., Labails, C. & Smethurst, M. a. (2009). A new scheme for the opening of the South Atlantic Ocean and the dissection of an Aptian salt basin. *Geophys. J Int 177,* 1315–1333.

Tripathy, S., Deo, T. & Tyler, B.M., 2012. Oomycete Transcriptomics Database: a resource for oomycete transcriptomes. *BMC Genomics 13,* 303.

Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., ... Zhang, Z. (2011). The genome of the mesopolyploid crop species Brassica rapa. *Nat. Genet. 43,* 1035–1039.

Wang, Z., Gerstein, M. & Snyder, M. (2009). RNA-Seq : a revolutionary tool for transcriptomics. *Nat Rev Gene, 10,* 57–63.

Watson, J. & Crick, F.H. (1953). Molecular Structure of Nucleids Acids. *Nature 171,* 737–738.

Werner, T. (2010). Next generation sequencing in functional genomics. *Brief. Bioinform. 11,* 499–511.

Yeoh, K. A., Othman, A., Meon, S., Abdullah, F. & Ho, C. L. (2013). Sequence analysis and gene expression of putative oil palm chitinase and chitinase-like proteins in response to colonization of Ganoderma boninense and Trichoderma harzianum. *Mol Biol Rep, 40,* 147–158.