



Herramientas genómicas para la identificación, autenticación y mejora del material vegetal

Victoria Ibáñez, Javier Terol y Manuel Talón
Centro de Genómica (IVIA)

1. Análisis de la situación actual

1.1. Las herramientas genómicas

Los cítricos son organismos diploides, es decir, tienen una dotación doble de 9 cromosomas distintos en sus células; por tanto, estas contienen un total de 18 cromosomas. Uno de los conjuntos procede del parental femenino y otro del masculino, como ocurre en los humanos. Los cromosomas paternos y maternos son prácticamente idénticos y por ello se denominan cromosomas homólogos. Los cromosomas están constituidos por proteínas y ácido desoxirribonucleico (ADN). A su vez, el ADN está formado por 4 unidades básicas: los nucleótidos adenina (A), citosina (C), timina (T) y guanina (G). El orden de los nucleótidos o bases compone la secuencia genética de un individuo. Las diferencias en la secuencia de las bases configuran las principales diferencias entre los seres vivos. Al conjunto de cromosomas se le denomina genoma. El primer genoma secuenciado de cítricos fue el de la clementina, que presenta una longitud de casi 300 millones de bases (Wu *et al.*, 2014), gracias a un consorcio público-privado en el que participó el Instituto Valenciano de Investigaciones Agrarias.

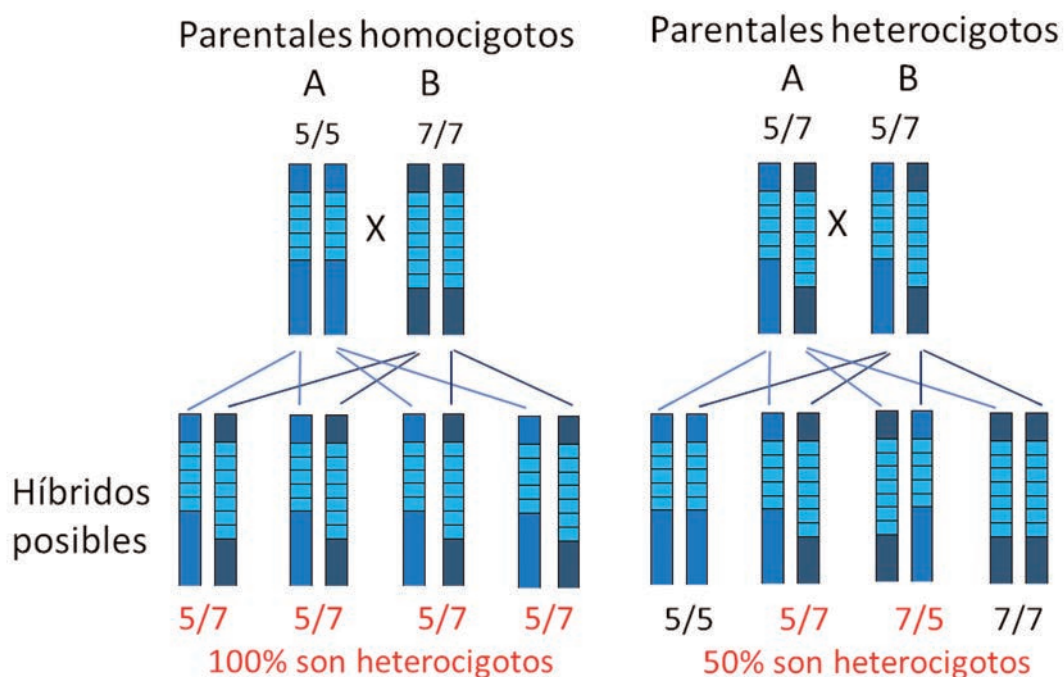
Los marcadores genéticos o moleculares son segmentos de ADN con una ubicación física identificable (locus) en un cromosoma. En las últimas décadas se han desarrollado distintos tipos de marcadores moleculares, RFLP, AFLP, RAPD o SCAR con el inconveniente de ser dominantes, es decir, no distinguen el individuo homocigoto (presenta, en una determinada posición del genoma, la misma secuencia en el cromosoma materno y en el paterno) del heterocigoto (presenta una secuencia distinta). Sin embargo, también se han desarrollado marcadores de tipo codominante, capaces de distinguir entre homocigotos y heterocigotos, como los microsatélites o SSR (*single sequence repeat*). Los microsatélites son repeticiones en tándem de unas pocas bases, generalmente entre 2 y 6,



que pueden variar entre cromosomas homólogos. Como ejemplo, podemos pensar en una variedad A cuyos dos cromosomas tienen la secuencia «CGA» repetida 5 veces y en una variedad B con la secuencia repetida 7 veces. La variedad híbrida H, cruce de la variedad A y B tendrá un cromosoma con 5 repeticiones procedente de A y su cromosoma homólogo tendrá 7 repeticiones, procedente de B. Las ventajas de los microsatélites son la codominancia y su abundancia en el genoma porque generalmente existen miles de secuencias repetidas en tándem. Sin embargo, según las leyes de Mendel, la probabilidad de que un híbrido entre A y B, siendo ambos homocigotos pero con distinto número de repeticiones, tenga la combinación de alelos 5 y 7 es del 100 %. Es decir, el total de los descendientes serían iguales para ese microsatélite. En caso de parentales heterocigotos, la probabilidad de aparición de dicha combinación se reduciría a la mitad, porcentaje inferior pero todavía claramente insuficiente para una identificación varietal (Figura 1).

Figura 1.

Segregación de distintos alelos en hibridaciones de parentales homocigotos (izda.) y heterocigotos (dcha.) con un microsatélite repetido 5 y 7 veces. Cuando los padres son homocigotos, el 100 % de los posibles híbridos serán heterocigotos. Si los parentales son heterocigotos, el 50 % de los posibles híbridos serán heterocigotos tal y como establecen las leyes de Mendel



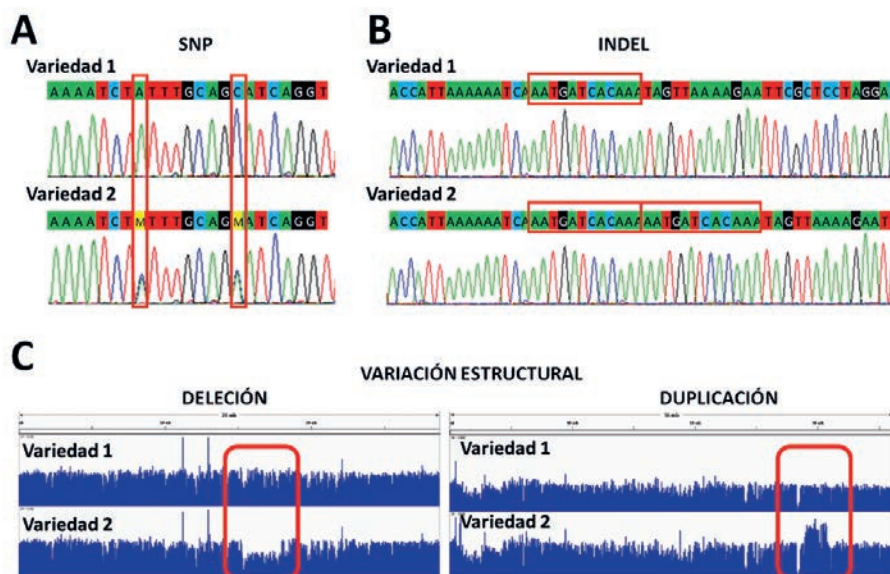
Las características de los microsatélites hacen necesario, por tanto, el empleo de un conjunto de ellos para ser empleado en identificación varietal. Una combinación de varios microsatélites puede ser altamente eficaz para discernir entre grupos varietales o para identificar una variedad híbrida. Sin embargo, el empleo de microsatélites en la identificación de variedades procedentes de mutación tanto naturales como inducidas es poco eficaz. La probabilidad de que una mutación afecte al microsatélite escogido (al número de repeticiones) es bajísima. La poca efectividad de discriminación de los microsatélites entre variedades procedentes de mutación cobra especial relevancia en aquellos cultivos que se han generado mediante este mecanismo. En los cítricos, alrededor de la mitad de

las variedades comerciales actuales del registro español proceden de mutación. Las variedades de mandarinas y naranjas más conocidas se han generado de manera espontánea por mutaciones naturales y en algún caso por mutaciones inducidas. Un claro ejemplo lo encontramos en las variedades de clementina como 'Clemenules', 'Hernandina', 'Oronules' o 'Oroval' que proceden de mutación natural de 'Clementina Fina'. En el caso de las naranjas, variedades como 'Navelate', 'Fukumoto' o 'Barnfield' surgieron como mutaciones naturales de yemas de 'Washington Navel'.

Existen distintos tipos de marcadores moleculares que permiten la identificación de variedades procedentes de mutación. Algunos afectan a una o unas pocas bases dentro de un mismo cromosoma (SNP e INDEL) mientras que otros afectan a muchas bases y pueden implicar a distintos cromosomas (variaciones estructurales). Los SNP (*single nucleotide polymorphisms*) son variaciones en la secuencia de ADN que afectan a una sola base de una secuencia en el genoma. Los INDEL son pérdidas (delección) o ganancias (inserción) de un fragmento de ADN de un cromosoma que afectan a unas pocas bases. Las variaciones estructurales afectan a miles o millones de bases y pueden ser de tipo cuantitativo: deleciones o duplicaciones donde se pierde o gana material genético o de tipo cualitativo como las translocaciones o inversiones donde se reestructura el genoma, pero no hay ganancia o pérdida de ADN. La secuenciación del genoma de una variedad es clave para identificar SNP, INDEL y variaciones estructurales (Figura 2).

Figura 2.

Variantes genéticas de tipo SNP, INDEL y variación estructural entre dos variedades de cítricos. Fragmento de las secuencias consenso de los dos cromosomas de las variedades 1 y 2. La variedad 1 muestra la secuencia consenso homocigota de referencia, mientras que la variedad 2 muestra dos variaciones de tipo SNP encuadradas en rojo y marcadas con una M, lo que indica que un cromosoma posee la una base distinta a su cromosoma homólogo (A). Fragmento de las secuencias consenso de los dos cromosomas de las variedades 1 y 2. La variedad 1 muestra la secuencia consenso homocigota de referencia, mientras que la variedad 2 muestra una variación de tipo INDEL enmarcada en rojo, concretamente una inserción (B). Representación del número de lecturas en los cromosomas 1 (izquierda) y 2 (derecha) en las variedades 1 y 2. En la imagen de la izquierda, en la variedad 2, se observa una delección (pérdida de lecturas) enmarcada en rojo, mientras que en la imagen de la derecha, se observa también en la variedad 2 y en rojo, una duplicación (C)





1.2. El uso de herramientas genómicas para la mejora varietal

En los cítricos se busca la generación de nuevos cultivares con frutos de mayor calidad, más productivos, con períodos de maduración más amplios y resistentes a estreses abióticos (salinidad, sequía) y bióticos (plagas y enfermedades exóticas), con el objetivo de competir en un mercado globalizado y en un ambiente cambiante debido al calentamiento global.

La mejora convencional de plantas se basa en la selección por el fenotipo de los individuos de interés por alguna característica distintiva, entre los individuos de progenies segregantes resultado de la hibridación. La obtención de nuevos cultivares por esta vía precisa no menos de 8 o 10 años y en ocasiones no se garantiza la obtención de un cultivar mejorado. Hoy se perfila como insuficiente para poder dar respuesta a los problemas que afronta la citricultura debido a sus elevados costes y a la gran cantidad de tiempo necesario para generar una nueva variedad.

La selección asistida por marcadores permite hacer mucho más eficientes las estrategias de selección, al hacer posible la determinación de las características genéticas de las plantas y seleccionar por el genotipo, en lugar de por el fenotipo. La selección de individuos en sus primeras fases de desarrollo permite reducir los costes derivados del personal e infraestructura para cultivar el material vegetal hasta manifestar el carácter elegido. La selección se realiza en estadio de plántula, eliminando aquellas líneas que no tienen el marcador ligado al carácter de interés.

1.3. El uso de herramientas genómicas para la identificación varietal

Los marcadores genéticos se aplican a cualquier parte de la planta y en cualquier estado de desarrollo de la misma, y permiten identificar los cambios en el ADN, una molécula que está presente en todas las células y que no está condicionada a los cambios ambientales. Estos marcadores genéticos facilitan la identificación varietal de variedades procedentes de mutación, que constituyen más del 50 % de las variedades comerciales españolas de cítricos, y su desarrollo requiere la secuenciación del genoma. El coste de secuenciación ha decrecido drásticamente en los últimos años, haciendo hoy día accesible esta metodología al sector agrícola. En este sentido, el Centro de Genómica del IVIA tiene establecido un protocolo de identificación de SNP, INDEL y reorganizaciones cromosómicas que está siendo aplicado para la identificación de variedades procedentes de mutaciones tanto espontáneas como inducidas (Ibanez-Gonzalez *et al.*, 2018; Ibáñez *et al.*, 2012). El registro de una variedad de cítricos y su identificación están basados en caracteres botánicos, que pueden ser influenciados por el ambiente. Para las variedades cuyas diferencias estriban principalmente en características del fruto, su discriminación solo puede realizarse en un estrecho margen temporal. La identificación empleando marcadores moleculares basados en repeticiones en tándem, como los microsatélites o polimorfismos en el sitio de restricción (AFLP, RAPD...).

El espectacular desarrollo de las tecnologías de identificación rápida de las variantes del genoma ha permitido su aplicación a los programas de mejora para la obtención de nuevas variedades, patrones y cultivos, incluidos los árboles frutales (Iwata *et al.*, 2016). Las mejoras en el rendimiento, combinadas



con una disminución global de los costes de la secuenciación, están permitiendo el mapeo y la caracterización de rasgos de elevado interés agronómico (Salgotra *et al.*, 2014). Los SNP ya son los marcadores de genotipado más ampliamente utilizados, debido a su abundancia en el genoma y la relativa facilidad para determinar su frecuencia de una manera rentable (Bauchet *et al.*, 2017).

1.4. El uso de herramientas genómicas para la identificación de los orígenes de las variedades

Las herramientas genómicas también pueden emplearse para la identificación de los parentales de variedades. Los microsatélites y los SNP son los marcadores moleculares más adecuados para esta tarea. Un abordaje completo inequívoco incluiría la secuenciación de los posibles parentales y del híbrido, permitiendo el análisis simultáneo de miles de SNP en todos los cromosomas. En la práctica, unas decenas de microsatélites pueden discriminar fácilmente las variedades híbridas de sus parentales.

2. Escenario futuro

2.1. El genotipado por secuenciación y la selección genómica

En el genotipado por secuenciación (GBS), las diferencias de secuencia entre los genomas se utilizan directamente como marcadores. No se requiere información de la secuencia previa y todos los marcadores se originan en la población que se va a genotipar. Los abordajes mediante estos desarrollos ya se han usado con éxito en especies agrícolas de gran relevancia, incluso en leñosas. La llegada del GBS tiene ya un impacto profundo en las estrategias de mapeo, que se benefician de una distribución de gran densidad de marcadores por todo el genoma.

El análisis mediante GBS se complementa con los estudios de asociación del genoma completo o *Genome Wide Association Studies* (GWAS) y la selección genómica o *Genomic Selection* (GS). Estos estudios de asociación que abarcan todo el genoma o GWAS utilizan la diversidad natural presente en una población para proporcionar una resolución más alta, para mapear los caracteres de interés que las que proporcionan los análisis de poblaciones de mapeo por ligamiento (Wang *et al.*, 2012). Estas estrategias se están comenzando a implementar ampliamente para la mejora genética de plantas (Figura 3). Es de interés destacar también que estas variaciones pueden ser en sí mismas funcionales y estar relacionadas con la fisiología del carácter (Yundaeng *et al.*, 2015) pero en la mayoría de los casos se utilizan para el mapeo y la ubicación de los verdaderos *loci* relevantes. Existen múltiples ejemplos donde los estudios de GWAS han sido útiles para identificar marcadores asociados a caracteres agronómicos de gran interés (por ejemplo, en el arroz y en el maíz, o en frutales como la manzana o el melocotón). En la Figura 3 se ha ejemplificado un proceso de selección de frutales mediante la aplicación del GBS y el GWAS, mientras que en la Figura 4 se muestra el resultado del análisis estadístico de la asociación del carácter espinosidad con las variables genéticas detectadas en una población segregante de cítricos.



Figura 3.

Selección de individuos a partir de los estudios de asociación (GWAS) sobre una población segregante. A partir de los datos de genotipo y fenotipo de la población se obtiene un panel de marcadores (SNP) asociados a caracteres de interés. Este panel de marcadores se puede usar para seleccionar segregantes en nuevos cruces, de manera que solo aquellos que presenten el genotipo asociado al rasgo de interés pasarán a ser fenotipados, descartando aquellos que no porten los alelos deseados

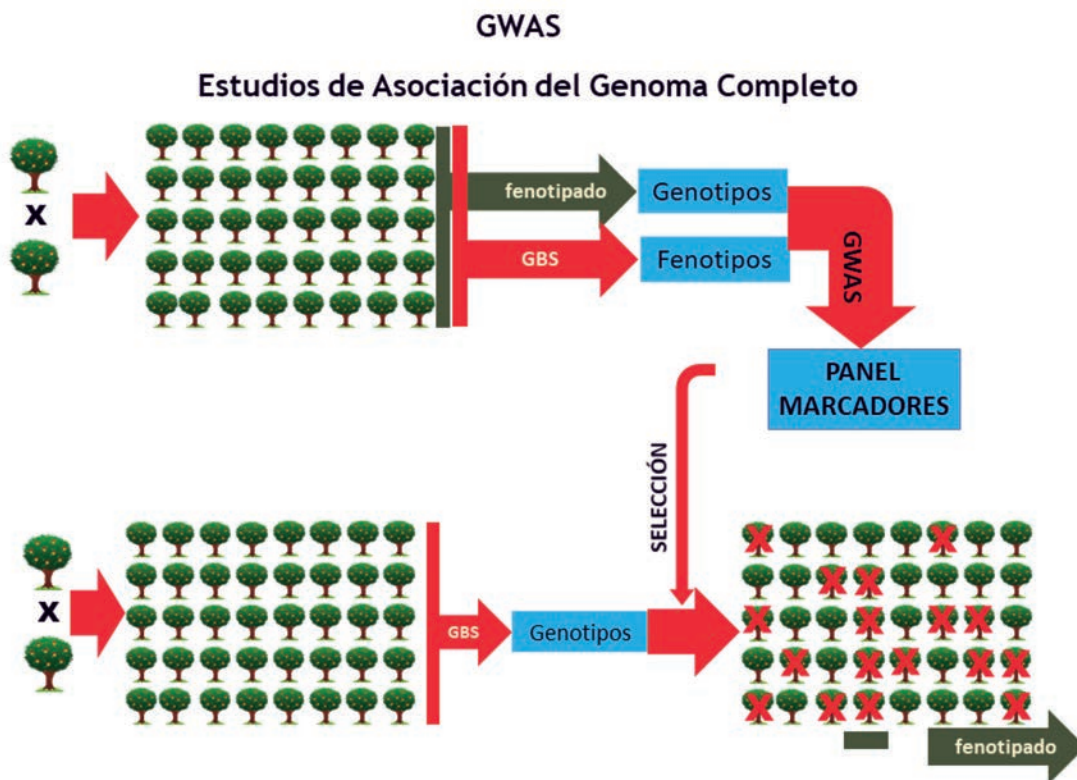


Figura 4.

Gráfico de Manhattan que ilustra la asociación entre los marcadores SNP y el carácter de espinosidad en cítricos

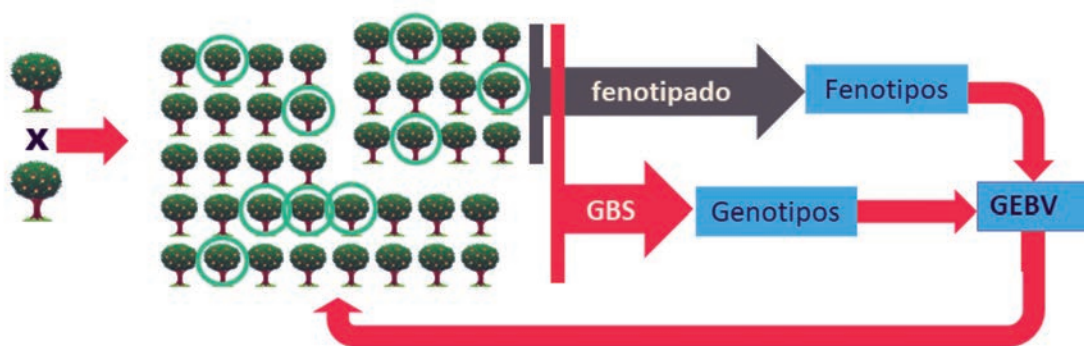


El método de Selección Genómica (GS) se basa en calcular el valor estimado de mejora (GEBV, *Genomic Estimated Breeding Value*) para un carácter de interés de los individuos en una subpoblación de entrenamiento. Para el cálculo del GEBV se usan los SNP de todo el genoma de cada ejemplar obtenidos mediante GBS y los datos del fenotipado realizado en la subpoblación. Una vez establecido el modelo se puede estimar el GEBV de toda la población únicamente con los datos de genotipado, sin necesidad de fenotipar el total de los individuos. La gran ventaja de la Selección Genómica consiste precisamente en que se evita la parte más costosa de los programas de mejora, que es el fenotipado de todos los individuos de una población, ya que mediante el fenotipado de una subpoblación de

entrenamiento se puede estimar el valor genómico de mejora de todos ellos y proceder a su selección. Esta técnica, por tanto, tiene un gran potencial para acelerar la ganancia genética de un programa de mejora y ya se ha aplicado con éxito en plantas anuales como el trigo, la cebada, el maíz, la avena y el arroz, pero también en leñosas, acortando en varios años el tiempo para la obtención de nuevas variedades.

Figura 5.

Selección genómica realizada sobre una población segregante proveniente de un cruce. Una fracción reducida de la población es fenotipada con detalle y en función del genotipo se calcula el GEBV para un carácter determinado. El resto de la población es únicamente genotipada y, a partir de los valores de GEBV obtenidos, se puede realizar la selección de individuos (círculos verdes) sin necesidad de fenotiparlos



2.2. La autenticación y la propiedad varietal

La Oficina Española de Variedades Vegetales cuenta en su registro con más de 230 entradas del género *Citrus*: 52 % mandarinos, 24 % naranjos, 8 % limones, 3 % pomelos, 2 % pummelos y un 11 % patrones.

El registro de variedades comerciales español, acorde a los protocolos europeos de la oficina comunitaria de variedades, comprende exámenes técnicos de campo y de laboratorio para dictaminar si un material vegetal es distinto, estable y homogéneo (DEH) respecto al resto de las variedades de la colección de referencia. Estos exámenes están basados en caracteres fenotípicos, botánicos o agronómicos, muchos de los cuales solo se observan en algunas partes de la planta, como las hojas o el fruto, y todos ellos están sujetos más o menos a la variabilidad debida a las condiciones ambientales.

La oferta varietal de cítricos ha crecido drásticamente en los últimos años y es previsible que siga haciéndolo gracias a los programas de mejora, que están realizándose por entidades públicas y privadas, dada la necesidad de obtener variedades mejor adaptadas a las nuevas condiciones climáticas, debido al cambio climático actual, y para satisfacer las exigencias del mercado cítrico.



La secuenciación de genomas comerciales para la obtención de marcadores moleculares de identificación se comienza a percibir como una necesidad adicional en el proceso de registro y protección, y también como un parámetro para dar garantía de autenticidad a las variedades comerciales. La identificación varietal de una muestra vegetal mediante marcadores moleculares está empezando a ser empleada y se prevé su generalización por ser un sistema de identificación objetivo, rápido y aplicable en toda la cadena productiva. De hecho, la autenticación molecular aportará seguridad en todos los eslabones de la cadena: a los agricultores, al garantizarles la identidad del material vegetal adquirido en vivero; al distribuidor, al que le asegura la identidad de la fruta para su comercialización; y al consumidor, al que le certificará la autenticidad del producto adquirido para su consumo. El sistema, por tanto, puede llegar a permitir la trazabilidad del producto de forma objetiva.

La aplicación de la identificación molecular a las variedades de cítricos de forma objetiva conlleva la necesidad de instaurar una unidad de certificación oficial, solicitada por los distintos integrantes de la cadena citrícola, por lo que parece lógico que sea una realidad en breve.

3. Recomendaciones a corto y medio plazo

La nueva revolución de los secuenciadores automáticos, con la irrupción de las técnicas de secuenciación de ADN de nueva generación (NGS), ha puesto a nuestra disposición el mejor marcador posible: la secuencia completa de los genomas. En la actualidad ya es posible la secuenciación masiva de genomas completos a un precio razonable y ya es una herramienta habitual para la mejora y el uso de los marcadores moleculares, por tanto, es pieza clave tanto para el desarrollo de nuevos programas de mejora genética como para su uso en la identificación varietal.

La utilización de marcadores moleculares para la identificación y autenticación de material vegetal se prevé que se generalizará a lo largo de toda la cadena de comercialización, desde el material de origen procedente de los viveros hasta el producto obtenido por el consumidor en destino, fruta, zumos y derivados. Mediante una certificación molecular, el propietario de la variedad podrá garantizar a los agricultores la identidad del material vegetal adquirido. La incorporación de estos desarrollos a los distribuidores primarios y secundarios muy probablemente incrementará su modelo de calidad, al aportar valor añadido al producto, un aspecto que será convenientemente valorado por el consumidor final.

La aplicación de herramientas genómicas como el GWAS y MAS en los programas de mejora de variedades y patrones previsiblemente se generalice y pase a ser un elemento necesario y rutinario de los mismos.

Agradecimientos

Este trabajo se ha realizado en el marco de los proyectos RTI2018-097790-R-I00 del Ministerio de Ciencia, Innovación y Universidades y del proyecto 51915 del Instituto Valenciano de Investigaciones Agrarias, que cuenta con cofinanciación de fondos FEDER.



Referencias bibliográficas

BAUCHET, G. *et al.* (2017): «Identification of Major Loci and Genomic Regions Controlling Acid and Volatile Content in Tomato Fruit: Implications for Flavor Improvement»; *New Phytologist* 215(2); pp. 624-41. Disponible en: <http://doi.wiley.com/10.1111/nph.14615>.

IBANEZ-GONZALEZ, V.; BORREDÁ, C.; PÉREZ-ROMÁN, E. y TALÓN, M. (2018): «Identificación y autenticación de variedades de cítricos mediante detección de variaciones estructurales cromosómicas»; *Levante Agrícola* 441(2); pp. 84-88. Disponible en: <http://redivia.gva.es/handle/20.500.11939/6132>.

IBAÑEZ, V. *et al.* (2012): «Identificación inequívoca de variedades de cítricos mediante comparación genómica 1»; *Levante Agrícola* 413; pp. 309-317.

IWATA, H. *et al.* (2016): «Genomics-Assisted Breeding in Fruit Trees»; *Breeding Science* 66(1); pp. 100-115. Disponible en: https://www.jstage.jst.go.jp/article/jsbbs/66/1/66_100/_article.

SALGOTRA, R. K.; GUPTA, B. B. y STEWART, C. N. (2014): «From Genomics to Functional Markers in the Era of Next-Generation Sequencing»; *Biotechnology Letters* 36(3); pp. 417-26. Disponible en: <http://link.springer.com/10.1007/s10529-013-1377-1>.

WANG, M. *et al.* (2012): «Genome-Wide Association Mapping of Agronomic and Morphologic Traits in Highly Structured Populations of Barley Cultivars»; *Theoretical and Applied Genetics* 124(2); pp. 233-46. Disponible en: <http://link.springer.com/10.1007/s00122-011-1697-2>.

WU, G. A. *et al.* (2014): «Sequencing of Diverse Mandarin, Pummelo and Orange Genomes Reveals Complex History of Admixture during Citrus Domestication»; *Nature Biotechnology* 32(7); pp. 656-62. Disponible en: <http://www.nature.com/articles/nbt.2906>.

YUNDAENG, C. *et al.* (2015): «A Single Base Substitution in BADH/AMADH Is Responsible for Fragrance in Cucumber (*Cucumis Sativus* L.) and Development of SNAP Markers for the Fragrance»; *Theoretical and Applied Genetics* 128(9); pp. 1881-92. Disponible en: <http://link.springer.com/10.1007/s00122-015-2554-5>.