



VNIVERSITAT
DE VALÈNCIA

ivia
instituto valenciano
de investigaciones agrarias

OBTENCIÓN DE MARCADORES MOLECULARES PARA LA IDENTIFICACIÓN DE VARIEDADES DE NARANJAS NAVEL

Juan Fernández Carrillo

TRABAJO FIN DE GRADO EN BIOTECNOLOGÍA

Tutores:

Victoria Ibáñez González
Manuel Talón Cubillo

Tutor académico:

Joan Ferré Manzanero

Burjassot, Julio de 2018

Resumen

La mayor parte de variedades de naranja dulce que se cultivan en la actualidad proceden de mutaciones espontáneas de una variedad materna. Este fenómeno determina que las variedades derivadas sean en la práctica cuasi clones de la variedad materna, lo que desemboca en una dificultad para la identificación varietal tradicional mediante caracteres morfológicos. Una manera de mejorar la identificación varietal basada en caracteres morfológicos es complementarla con una identificación varietal basada en el uso de marcadores moleculares. Este trabajo consiste en la utilización de herramientas genómicas desarrolladas por el consorcio público-privado Citrusseq-Citrusgenn en el Centro de Genómica del IVIA (Instituto Valenciano de Investigaciones Agrarias) para la obtención de marcadores moleculares que permitan la identificación inequívoca del grupo varietal de naranjas dulce tipo navel. Para ello, primeramente, se efectuó la exploración comparada de las secuencias genómicas de distintas variedades para la detección *in silico* de variaciones genómicas estructurales. Posteriormente se han validado marcadores moleculares mediante la utilización de PCR a tiempo real y secuenciación Sanger del producto obtenido.

Palabras clave: identificación varietal, marcadores moleculares, cítricos, naranjo dulce.

Índice

1. Introducción. Orígenes y domesticación del género <i>Citrus</i>	5
1. 1. Origen y domesticación del género <i>Citrus</i>	5
1. 2. Origen de la naranja.....	6
2. Impronta de la citricultura en el mundo y en la Comunidad Valenciana.....	7
3. Taxonomía de los cítricos.....	9
3. 1. Origen de la variación genética.....	9
3. 2. Variedades de naranjo dulce.....	10
4. Identificación varietal a partir de marcadores moleculares.....	11
4. 1. Elección de marcadores moleculares.....	13
5. Marco teórico de la detección de variaciones estructurales.....	14
5. 1. Antecedentes en la filogenia de las variedades de naranjo.....	14
5. 2. Secuenciación de genomas.....	15
5. 3. Manejo del software Integrative Genomics Viewer.....	16
6. Objetivo.....	19
7. Material y métodos.....	19
7. 1. Genoma de referencia.....	19
7. 2. Genomas de las variedades y material vegetal.....	19
7. 3. Búsqueda de marcadores moleculares <i>in silico</i>	20
7. 4. Diseño de oligonucleótidos.....	20
7. 5. Extracción de ADN genómico.....	20
7. 6. Cuantificación de cantidad y calidad del ADN extraído.....	21
7. 7. Reacciones de PCR en tiempo real.....	21
7. 8. Comprobación de los marcadores moleculares.....	22
8. Resultados.....	22
8. 1. Detección de variaciones estructurales del genoma.....	22
8. 1. 1. Deleción de 1.937 pb en el cromosoma 1 de Washington Navel.....	22
8. 1. 2. Deleción de 104.916 pb en el cromosoma 1 de Washington Navel.....	26
8. 2. Comprobación de la deleción mediante BLASTN.....	32
8. 3. Obtención de secuencia objetivo.....	32
8. 4. Cuantificación de la cantidad y calidad del ADN extraído.....	38
8. 5. Cribado de oligonucleótidos.....	38
8. 6. Confirmación de marcadores mediante PCR en tiempo real.....	42
8. 7. Secuenciación del marcador molecular evaluado.....	43
9. Discusión.....	43
10. Conclusiones.....	45
11. Bibliografía.....	47

1. Introducción. Orígenes y domesticación del género *Citrus*

1. 1. Origen y domesticación del género *Citrus*

Dilucidar el germen del género *Citrus* es un problema de difícil acometida debido a la facilidad de hibridación interespecífica y a la apomixis facultativa espontánea asociada a poliembrionía. Estos fenómenos son muy frecuentes en el género *Citrus* y en ellos se tiende a enmascarar la variación genética debido a la dominancia del genotipo materno¹, lo que genera una población genéticamente uniforme y altamente heterocigota².

El centro originario del género *Citrus* ha sido fuente de debate desde el siglo pasado. En un principio se pensó que podría residir entre el noreste de India y el norte de Birmania^{3,4}. Más tarde, otros autores se amparan en estudios poblacionales para considerar que los límites entre India, Birmania y el sur de China (concretamente la provincia china de Yunnan) podrían ser la cuna de los cítricos^{2,5}.

Un hito importante en la búsqueda del origen de los cítricos se produjo al encontrar fósiles de *Citrus linczangensis* datados del Mioceno tardío en el punto caliente de Yunnan⁶, el cual presenta características propias de la mayoría de las especies de *Citrus*, lo que refleja la existencia de un ancestro común de cítricos en Yunnan aproximadamente hace 8 millones de años⁶. Este punto caliente ha sido confirmado por estudios recientes⁷, donde se determinó que el origen del género *Citrus* se da en el piedemonte sudeste de los montes Himalayas, en una región que incluye la región este de Assam, norte de Birmania y oeste de Yunnan⁷. El ancestro común de los cítricos sufrió una especiación repentina durante el Mioceno tardío, un periodo que coincide con una atenuación de los monzones, pasando de climas húmedos a secos⁸. Esta transición está relacionada con la radiación a partir de un ancestro común^{7,9}.

A partir de aquí se produce una diversificación mediante los fenómenos de selección natural y domesticación, que daría lugar a diversas especies⁷.

La domesticación de cítricos comenzó probablemente con la identificación de especies interesantes para el ser humano para su selección y reproducción asexual, probablemente a partir de individuos híbridos y no a partir de selección recurrente desde una población de partida^{10,11}.

La diversidad subsiguiente se obtuvo a partir de mutaciones somáticas espontáneas, hibridaciones y fenómenos apomicticos; este hecho determina que los enfoques tradicionales usados para la identificación de otras especies de importancia agronómica no puedan ser aplicados en los cítricos¹².

El origen del naranjo dulce no está claro, aunque en la bibliografía se dan trabajos que determinan que surgió a través de la hibridación interespecífica entre especies ancestrales de pummelo (*Citrus maxima*) y mandarina (*Citrus reticulata*)^{7,13-16}. Algunas evidencias avalan estos estudios, como el hecho de que el ADN del cloroplasto de la naranja dulce provenga del pummelo. La hibridación entre especies ancestrales pudo haberse dado en China

meridional, Birmania y regiones de la India al sur del Himalaya. Es de sobra conocido que el naranjo dulce ha sido cultivado en China durante siglos, alcanzando en este país un gran desarrollo y perfeccionamiento antes de haber sido conocido en Europa¹⁷.

1. 2. Origen de la naranja

La primera referencia a los cítricos se da en el año 2200 a.C en China, bajo el liderazgo de Yu el Grande¹⁸. En esta época los cítricos, especialmente mandarinas y pomelos chinos, se reservaban para la corte imperial^{18,19}.

Posteriormente, se hace referencia a la naranja por primera vez en el año 500 a. C., en el segundo libro de los Cinco Clásicos, una serie de libros en los que se recoge la doctrina de Confucio²⁰.

El cidro, la naranja amarga y el limón ya eran conocidos por los romanos²¹. Siguiendo a las legiones romanas en sus viajes, los cítricos fueron plantados en las partes más cálidas de Europa y África del Norte. Posteriormente los comerciantes podrían haber llevado la fruta a otras partes de Europa.

Una etapa importante de la domesticación de los cítricos fue la invasión de la Península Ibérica por parte de los musulmanes⁵. Los musulmanes trajeron a la península nuevas técnicas de irrigación y especies vegetales novedosas como olivos, viñas, trigo y cítricos. La citricultura floreció en numerosas ciudades españolas como Sevilla, Valencia, Granada y Zaragoza. En Valencia concretamente, en los libros *Regiment de la cosa pública* de Francesc Eiximenis (1383) y *Viaje por España y Portugal* de Hieronymus Münzer (1498) se evidencian extensas plantaciones de cítricos^{22,23}.

A lo largo de la Edad Media los limones y naranjas fueron extendiéndose por todo el sur de Europa, considerándose allá donde iban como delicias. Después de la Toma de Constantinopla (1453) por el Imperio Otomano, los navegantes portugueses encontraron una nueva ruta hacia las Indias²⁴: rodear África por el cabo de Buena Esperanza, llegando hasta Malindi (actual Kenia), y desde ahí a Calicut (India). El viaje de los cítricos puede evidenciarse en los diarios del navegante Vasco da Gama, Gaspar Corrêa y Andrea Corsali²⁵. Esto desencadenó la llegada de naranjas dulces, interesantes desde el punto de vista alimenticio²⁶ desde India a Europa.

En el siglo XVI, la naranja dulce había adquirido una gran importancia comercial en Europa. La mandarina china fue traída a Inglaterra a principios del siglo XIX, donde se extendió a Malta, Sicilia e Italia. Los cítricos también se expandieron hacia el Nuevo Mundo de mano de Colón, que trajo las primeras semillas de cítricos a Haití en su segundo viaje en 1493²⁷. En el siglo XVI la naranja se expandió por México y Florida, y en el siglo XVIII ya se encontraba en California y Arizona¹⁷. A lo largo de la Edad Contemporánea los cítricos se han ido extendiendo a lo largo del globo, obteniéndose diversas especies y variedades adaptadas al cultivo local.

2. Impronta de la citricultura en el mundo y en la Comunidad Valenciana

La citricultura es el conjunto de técnicas agronómicas orientadas a la producción de especies vegetales principalmente del género *Citrus*, aunque también se engloba alguna otra especie de menor interés agronómico como son *Fortunella* y *Poncirus*²⁸.

A nivel mundial los cítricos son la fruta con mayores niveles de producción (Figura 1). La suma de producción de los grupos de frutas consideradas cítricos (naranja, tangerina, limón, lima y pomelo) alcanza los 132,4 millones de toneladas, situándose en el grupo de mayor producción, por encima de la sandía o el plátano a nivel mundial.

A pesar de que dentro de la categoría de “cítricos” se engloban diferentes especies (limones, pomelos, cidros, mandarinas, tangerinas...) las naranjas suponen el principal aporte a la producción mundial de cítricos. La distribución de la producción de cítricos en el mundo se centra en el hemisferio norte (alrededor del 70%), en particular en países mediterráneos y los Estados Unidos, aunque el principal productor es China, seguido del gigante austral Brasil (38,15 y 18,92 millones de toneladas, respectivamente). España se encuentra en la sexta posición del ranking (6,1 millones de toneladas), detrás de México (Figura 2).

Durante la última década, en nuestro país la producción anual ha sido superior a cinco millones de toneladas. Respecto a las exportaciones, España es el primer exportador mundial de fruta en fresco, el primer exportador mundial de naranja, limón y mandarina y el sexto de pomelo. En el ámbito de la Unión Europea, España es el mayor productor de cítricos.

En España, la Comunidad Valenciana es la principal región citrícola, ocupando aproximadamente 182.000 hectáreas y generando casi cuatro millones de toneladas de frutos (3.943.213 t) en la temporada 2016/2017, siendo la naranja dulce la más preponderante (50%), seguido de cerca por la mandarina (45%)²⁹. A nivel varietal, hoy en día se cultivan aproximadamente 35 variedades diferentes de cítricos, entre las que destacan las naranjas Navelate, Navelina y Valencia Navel y la mandarina Clemenules.

A pesar de esta numerosa variedad, el sector citrícola demanda nuevas variedades de cítricos con mejoras agronómicas, con una mayor amplitud en el calendario de recolección y que satisfagan las exigencias de los consumidores (ausencia de semillas, facilidad en el pelado, fundencia...).

Producción mundial de frutas* (2016)

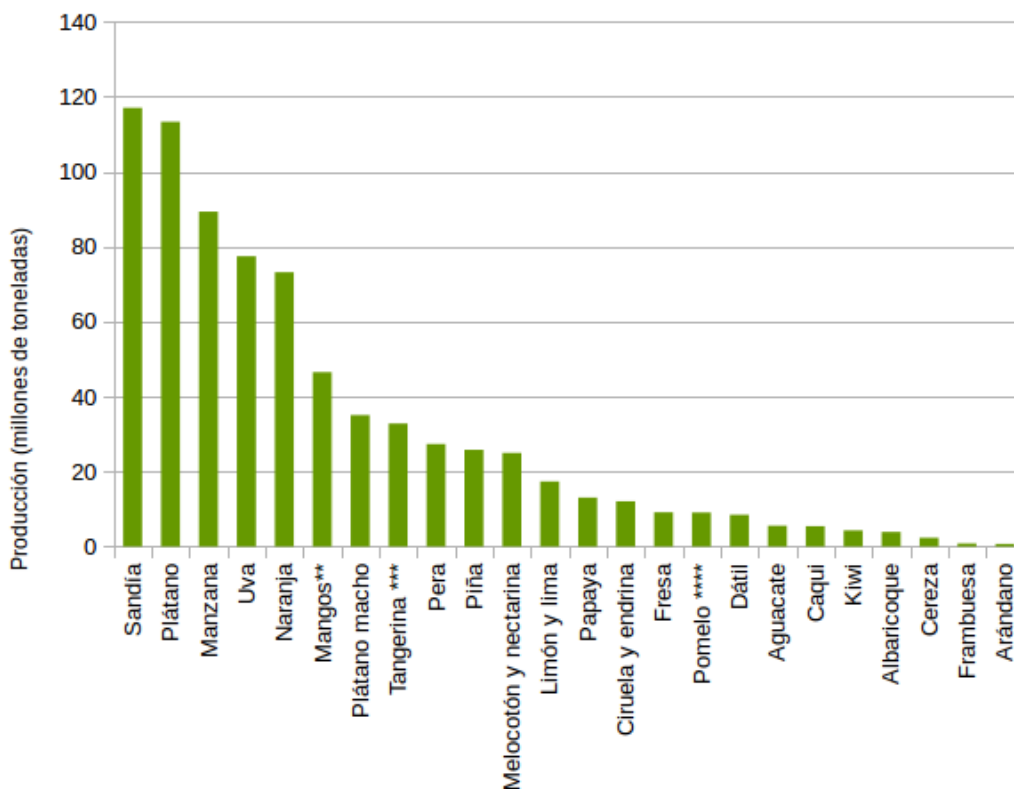


Figura 1. Producción mundial de frutas en el año 2016 en millones de toneladas. Si se consideran los cítricos como un todo (naranja, tangerina, limón y lima, pomelo) la producción de estos desbansa a las frutas más producidas como la sandía o el plátano, siendo de 132,4 millones de toneladas. Los datos han sido redondeados para permitir un menor manejo. *Sin incluir al melón. **Mango, mangostino y guayaba. ***Tangerina, mandarina, clementina y satsuma. ****Pomelo. Elaboración propia. Fuente: FAO, 2016.

Los diez países con mayor producción de cítricos (2015)

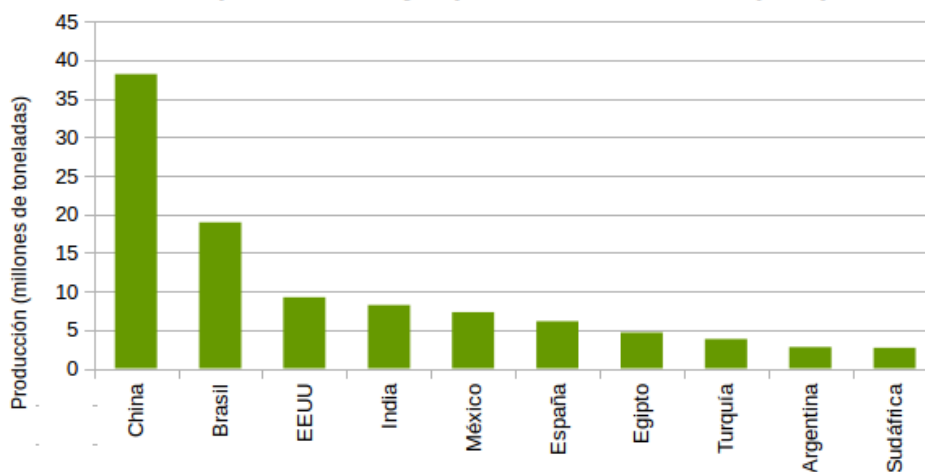


Figura 2. Los diez países del mundo con mayor producción de cítricos en el mundo en 2015. Elaboración propia. Fuente: FAO, EST Citrus Fruit.

3. Taxonomía de los cítricos

El género *Citrus* fue establecido por Carlos Linneo en el 1753, nombrando cinco especies: cidro (*Citrus medica*), limón (*Citrus medica* var. *lima*), naranja amarga (*Citrus aurantium*), naranja dulce (*Citrus aurantium* var. *sinensis*) y pomelo (*Citrus grandis*)³⁰.

Actualmente, los cítricos están incluidos en el Dominio *Eukaryota*, Reino *Plantae*, División *Magnoliophyta*, Clase *Magnoliopsida*, Subclase *Rosidae*, Orden *Sapindales*, Familia *Rutaceae*, y Subfamilia *Citroideae*. El género *Citrus* cuenta con más de 145 especies.

Las especies del género *Citrus* tienen una gran facilidad de reproducción interespecífica, generando híbridos fértiles³¹, y también sufren mutaciones espontáneas usualmente, por lo que tradicionalmente ha habido mucha confusión en la taxonomía de muchas especies.

Este hecho ha producido el desarrollo de diversos sistemas taxonómicos, como el de Swingle³² (1943), el de Tanaka³³ (1977) y posteriormente el de Mabberley³⁴ (1997), limitado a los cítricos cultivados, excluyendo el resto de especies del género. Swingle se centró en la descripción de taxones naturales, obteniendo 16 especies de cítricos. Por otro lado, Tanaka buscó englobar la mayor diversidad de especies de cítricos, alcanzando describir 162 especies.

La comunidad científica se opuso al sistema de Tanaka, arguyendo que la mayoría de especies identificadas eran híbridos. Actualmente el sistema de Swingle tampoco es utilizado. Estudios recientes basados en información genómica han permitido revisar las clasificaciones taxonómicas previas, proponiéndose una revisión y reestructuración de la misma^{7,15}.

3. 1. Origen de la variación genética

Tradicionalmente las especies de cítricos han sido obtenidas mediante cruces de diferentes especies, generando híbridos: ya sea el ejemplo del limón (*Citrus limon*), que es cruce del naranjo amargo (*Citrus aurantium*) y el cidro (*Citrus medica*); el naranjo dulce (*Citrus sinensis*), cruce de pummelo (*Citrus maxima*) y una mandarina ancestral (*Citrus reticulata*); y la clementina (*Citrus clementina*), resultado del cruce entre la mandarina común (*Citrus deliciosa*) y el naranjo dulce (*Citrus sinensis*)^{13,35}.

En cambio, la mayor parte de variedades de naranjo dulce han sido obtenidas a partir de mutaciones sobre una variedad. La mutación es un cambio en la secuencia del ADN cuyo origen puede ser natural o inducido mediante la aplicación de agentes mutágenos físicos o químicos. La irradiación, por ejemplo, ha sido utilizada comúnmente en programas de mejora de cítricos para obtener nuevas variedades³⁶⁻³⁹. Las mutaciones son, de este modo, divergencias de un genoma ancestral preestablecido que provocan cambios en la filogenia de las poblaciones actuales que derivan de un ancestro común.

Las variedades existentes son obtenidas a través de la selección de material vegetal generado mediante cruce o mediante mutación. En el caso de las naranjas dulces, estas

han sido seleccionadas a partir de mutaciones sucesivas a partir de una supuesta variedad ancestral¹⁵, generándose un amplio abanico de variedades de naranjo dulce con una variabilidad intraespecífica mínima.

3. 2. Variedades de naranjo dulce

Las variedades de naranjo dulce (*Citrus sinensis* (L.) Osbeck) se agrupan en tres grupos principales (Figura 3):

- **Naranjas blancas.** Dentro de este grupo encontramos variedades finas o selectas, con pocas o ninguna semilla, y variedades blancas comunes, con un número variable de semillas. Los naranjos de esta variedad son árboles robustos, con ramas erguidas espinosas, formando una copa elevada, con chupones hacia el interior de esta. Sus hojas son lanceoladas de un color verde oscuro con un peciolo vigoroso. Sus flores son medianamente grandes, con gran cantidad de polen. Los frutos son pequeños redondos, algo achatados, con una coloración intermedia entre amarillo anaranjado a naranja, dulce, y sutilmente ácido. Variedades que pertenecen a esta división: Blanca Comuna, Salustiana y Cadenera.
- **Naranjas navel.** Reciben el nombre de navel por presentar sus frutos una protrusión en forma de ombligo en la zona opuesta al pedúnculo, debido al desarrollo parcial de un segundo fruto. Los árboles son de tamaño mediano a grande, redondeado. Tienden a crecer de manera desparramada, con ramas con espinas pequeñas. Las flores son grandes, apareciendo solitarias o en racimo. Las anteras son blancas o de color crema, sin granos de polen. El fruto es grande y redondo, llegando a pesar hasta 400 gramos, de piel fina de color naranja intenso. Presenta pulpa naranja con un elevado porcentaje de zumo, sin semillas (partenocarpia). Entre este grupo se encuentran las variedades Washington Navel, Navelina, Chislett, Navelate.
- **Naranjas sanguinas**, rojas o pigmentadas. Los naranjos tipo sangre se caracterizan por dar frutos pigmentados. Esta pigmentación es debida a antocianinas (infrecuentes en cítricos) de color rojo hidrosolubles (principalmente cianidina-3-O-glucósido) que colorean la pulpa y la corteza. El árbol es de un tamaño pequeño a medio con ligera dominancia apical. Presenta hojas pequeñas y ovaladas, con flores grandes y anteras amarillas. El estilo permanece durante bastante tiempo en el fruto. Algunas variedades importantes dentro de este grupo son la Moro, Sanguinelli y Tarocco.

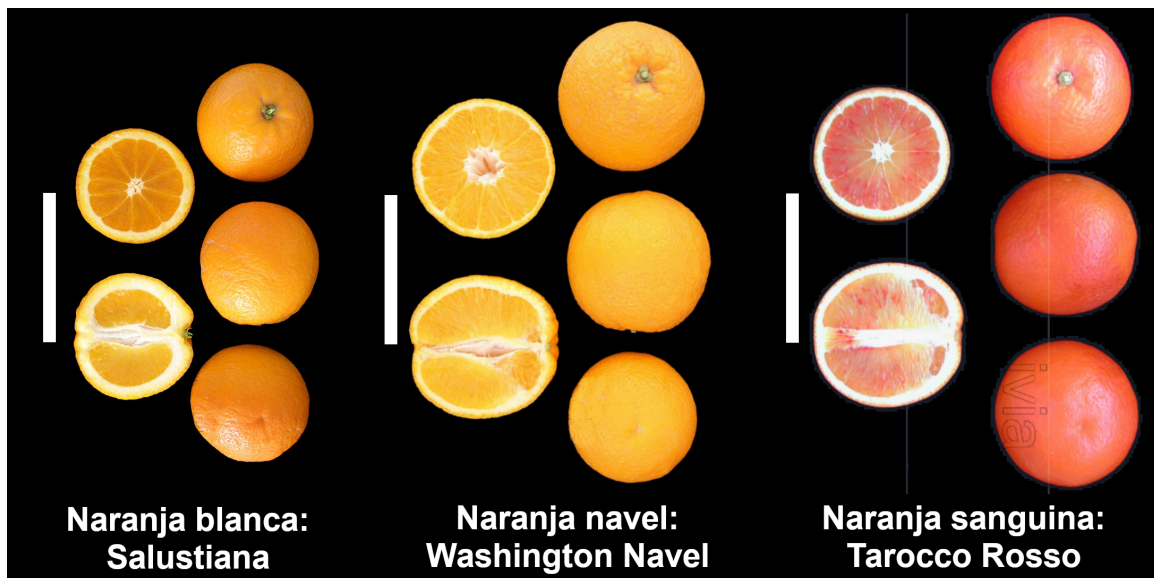


Figura 3. Naranjas blancas, navel y sanguinas. Como ejemplo de naranja blanca tenemos la variedad Salustiana. La variedad Washington Navel, del grupo de las navel, presenta en su parte distal un pequeño fruto parcialmente desarrollado, lo que provoca la aparición de una morfología umbilical. La variedad Tarocco Rosso, perteneciente a las naranjas sanguinas, muestra pulpa y cáscara pigmentada. Observar la diferencia de tamaño entre la naranja Salustiana y la Washington Navel. Barra de escala: 10 centímetros. Imágenes modificadas obtenidas de: <http://www.ivia.gva.es/es/naranjos>

4. Identificación varietal a partir de marcadores moleculares

La obtención de nuevas variedades con mejoras agronómicas presenta numerosas ventajas. La más evidente quizás es una mejora en el incremento del rendimiento frente a variedades ya conocidas, que se traduciría en mayores ingresos. Este hecho tiene una importancia económica fundamental en los países exportadores. Otras características agronómicas de especial interés son la mayor tolerancia a estreses abióticos y bióticos y la generación de variedades con caracteres distintivos deseados por los consumidores, como la ausencia de semillas, facilidad de pelado, aspecto y forma atractiva que incrementa la demanda y por consiguiente el precio del producto si la producción está controlada y no llega a saturar el mercado.

El proceso de mejora genética vegetal en frutales es especialmente largo y costoso comparado con especies herbáceas. Por ello, las nuevas variedades son registradas para proteger los derechos de explotación de la misma por parte de los obtentores.

Según el Convenio Internacional para la Protección de las Obtenciones Vegetales de la Unión Internacional para la Protección de Obtenciones Vegetales (UPOV) de 1991 la concesión del Título de Obtención Vegetal determina que el obtentor (textualmente: *la persona que haya creado o descubierto y puesto a punto una variedad*) de la variedad vegetal tiene la potestad de limitar toda actividad sobre el material vegetal de la variedad protegida a terceros, siendo necesaria la autorización del obtentor para llevar a cabo dicha actividad. La concesión de la autorización se denomina licencia, que permite a la vez la explotación comercial de la variedad por parte de terceros durante un periodo de tiempo

estipulado y la conservación de los derechos sobre la obtención vegetal del obtentor. La existencia de tales reglamentos pone de manifiesto la necesidad de identificar variedades de manera certera.

La determinación de variedades en una población heterogénea tradicionalmente se ha llevado a cabo mediante la selección de caracteres morfológicos, fisiológicos y etológicos, como pueden ser hábitos determinados de crecimiento, modalidades de dispersión de semilla, resistencia a ciertas plagas y enfermedades, diferencias en la productividad... Estos caracteres se recogen por diferentes organismos y colectivos como la UPOV, en sus exámenes DHE (examen de la distinción, la homogeneidad y la estabilidad), y el Instituto Internacional de Recursos Fitogenéticos (IPGRI), que recopila descriptores de los cítricos⁴⁰. En estos exámenes se evalúan características de la variedad vegetal, ya sean características del árbol, de la hoja o del fruto (vigor y porte del árbol, forma de la hoja, peso del fruto...) capaces de ser reconocidas en un análisis visual⁴¹. La evaluación de estos caracteres puede verse dificultada debido a que son limitados en número, pueden estar influenciados por el ambiente y se deben medir en cierta etapa del crecimiento (por ejemplo, cuando los individuos son adultos), entre otros motivos; lo que determina que la identificación por fenotipos no sea muy eficaz.

Es por ello por lo que los marcadores moleculares son una herramienta tan importante en la identificación varietal. Los marcadores moleculares son segmentos de ADN con una ubicación física identificable en un cromosoma y cuya herencia se puede rastrear en una familia. Por lo tanto, no están afectados por el ambiente.

Los marcadores moleculares se basan en la variabilidad del ADN. En principio se pueden considerar polimorfismos ilimitados en el ADN, lo que conlleva un número ilimitado de marcadores moleculares. Estos marcadores son insensibles al ambiente y su análisis es posible en cualquier momento del desarrollo del individuo, con lo que suplen las limitaciones del análisis visual. Este hecho, junto al desarrollo exponencial de las técnicas biotecnológicas y el abaratamiento de las tecnologías de Next-Generation Sequencing, han permitido el desarrollo de nuevos marcadores moleculares para la identificación varietal de un gran abanico de variedades de cítricos. Por ello, la identificación varietal basada en marcadores moleculares debe ser un complemento a la identificación basada en caracteres botánicos sobre la cual se sustenta el registro de una variedad vegetal.

La mayor parte de los marcadores moleculares presentan una serie de características comunes: son altamente polimórficos (permiten encontrar diferentes modalidades de una misma variación), discriminantes (detectan diferencias entre individuos estrechamente relacionados), tienen una alta densidad (están distribuidos en gran cantidad y uniformemente en el genoma) y son reproducibles.

Así, los marcadores moleculares constituyen una herramienta rápida y eficaz para la identificación varietal, los estudios de variabilidad, los análisis de pedigrís y para la selección asistida por marcadores moleculares. La asociación de un marcador molecular con un carácter de interés puede ser aprovechado en agronomía para seleccionar de manera temprana y eficaz individuos. Es de remarcar que en los cítricos la fase juvenil dura

varios años. La selección de caracteres adultos mediante marcadores moleculares puede realizarse en estadíos juveniles, antes de la aparición de frutos, mientras que utilizando la selección fenotípica sería necesario esperar a que el individuo creciera, floreciera y diera frutos. En el caso de su uso en la identificación de variedades vegetales esta puede ser utilizada para amparar los derechos del obtentor, asegurar una utilización lícita del material vegetal y otorgar identidad en el proceso de trazabilidad del producto. Por todo ello, la identificación mediante marcadores moleculares debería ir conjunta a la identificación basada en marcadores morfológicos o bioquímicos.

4. 1. Elección de marcadores moleculares

Es posible encontrar una enorme diversidad de categorías de marcadores moleculares. La cuestión radica en qué tipo de marcadores elegir. Para ello, se exponen sucintamente las bases de los marcadores más utilizados.

Los SSRs (*Simple Sequence Repeats* o microsatélites; también denominados STR, *Short Tandem Repeats*) son repeticiones en tándem de ADN en forma de di, tri y tetranucleótidos. La distinción entre individuos se da por diferente número de repeticiones. Son marcadores codominantes, permitiendo distinguir entre híbridos.

Los RAPDs (*Random Amplification of Polymorphic ADN*) se basan en la amplificación de fragmentos de manera aleatoria utilizando un único oligonucleótido de pequeño tamaño que consigue hibridar con numerosas regiones del genoma. En el caso de que los oligonucleótidos hibriden en dirección opuesta y a una distancia adecuada, se generarán productos de PCR. Los fragmentos se resuelven en un gel y se distribuyen en función de su tamaño. En esta ocasión el patrón de bandas (presencia o ausencia de banda o bandas) permite distinguir diferentes individuos.

Los RFLPs (*Restriction Fragment Length Polymorphisms*) son polimorfismos obtenidos por la fragmentación diferencial de genomas llevada a cabo por enzimas de restricción. Los fragmentos generados son resueltos en un gel de electroforesis, transferidos a una membrana e incubados con una sonda que permite su revelado. La diferencia en la distribución de las bandas durante el revelado permite discriminar distintos genomas.

La técnica de AFLPs (*Amplified Fragment Length Polymorphism*) utiliza una digestión con enzimas de restricción y una posterior amplificación con cebadores específicos. De nuevo se genera un patrón de bandas determinado en cada uno de los individuos, de modo que la presencia o ausencia de cierta banda o bandas permitirá la discriminación entre individuos.

Como ya se ha expuesto previamente, la génesis de una nueva variedad de naranjo se encuentra en la mutación espontánea de la variedad original. Para obtener marcadores moleculares discriminantes entre los individuos que permitan diferenciarlos, para cada tipo de marcador expuesto, sería necesario, en el supuesto de los SSRs, que las mutaciones espontáneas afecten al número de repeticiones en tándem; en el caso de los RAPDs al ADN molde donde hibridarían los oligonucleótidos; y en el caso de los RFLPs y AFLPs, las mutaciones tendrían que afectar a los sitios de restricción de los enzimas a utilizar durante

la digestión. Todas estas dianas posibles de la mutación son posibles, pero son altamente improbables, con lo que estos marcadores no son adecuados para variedades que proceden de mutaciones espontáneas; sin embargo, son aptos para distinguir híbridos, especies alejadas, o para ver la diversidad genética del género *Citrus*.

Este cribado nos deja con los marcadores SNPs, INDELS, y las variaciones estructurales. Los SNPs (*Single Nucleotide Polymorphisms*) son cambios en una sola base. Los INDELS son pérdidas o ganancias de pequeñas regiones del genoma (inserciones o deleciones). Las variaciones estructurales pueden subdividirse en variaciones estructurales cuantitativas y cualitativas. Las variaciones estructurales cuantitativas son aquellas que implican pérdida o ganancia del número de copias génicas (delección o duplicación). Las variaciones estructurales cualitativas no implican cambios en el número de copias pero sí en la estructura del genoma (translocación o inversión).

En este trabajo se han empleado los INDELS y las variaciones estructurales para la identificación de grupos y variedades de naranjo dulce.

5. Marco teórico de la detección de variaciones estructurales

5. 1. Antecedentes en la filogenia de las variedades de naranjo

Como guía para la identificación de naranjas se utilizó la filogenia de las variedades de naranjo dulce descrita en la literatura^{42,43}. En principio, una variedad que presente una variación estructural transmitirá esta variación a las variedades que descienden de ella, ya que estas han sido originadas a partir de mutaciones espontáneas sobre la variedad madre. Por ejemplo, si se presenta una delección concreta en la variedad Washington Navel esta misma delección se encontrará en las variedades Barnfield, Navelate, Ricalate, Lane Late, Cara Cara, Navelina, Powell, Fukumoto y Chislett ya que estas variedades descienden de Washington Navel (Figura 4).

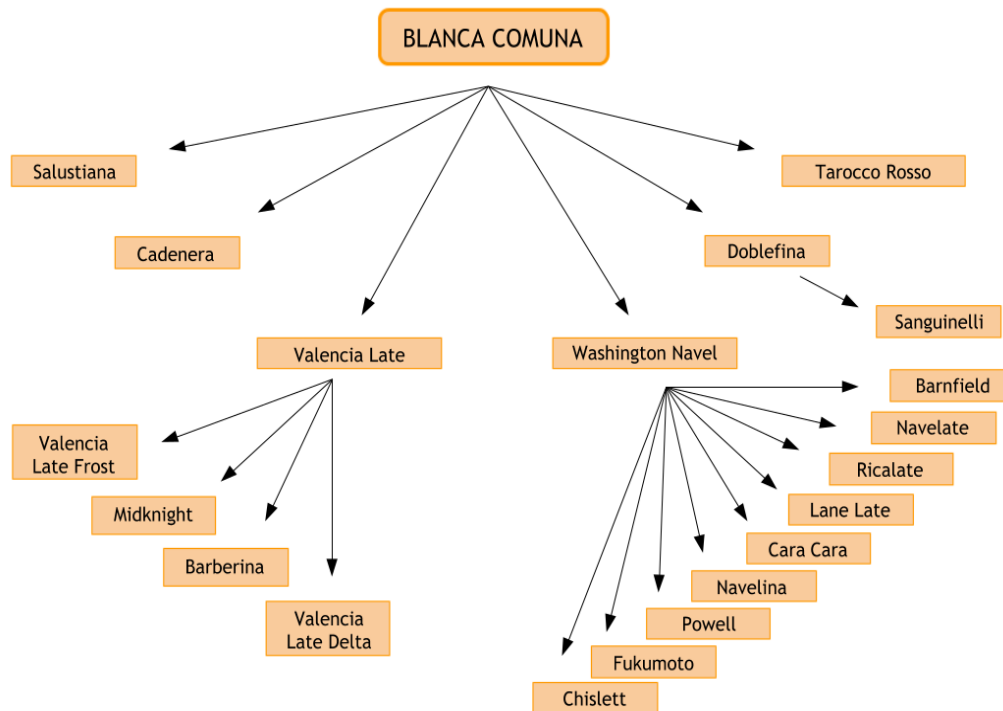


Figura 4. Filogenia de las variedades de naranjo dulce según se describe en la bibliografía. En principio, el ancestro común de todas las variedades sería la variedad Blanca Comuna. A partir de esta variedad se produce una diversificación en la que comienzan a aparecer los distintos grupos de variedades de naranjo dulce. Se distinguen los grupos de naranjas blancas (Blanca Comuna, Salustiana, Cadenera), naranjas navel (Washington Navel y las variedades derivadas a partir de esta), naranjas late (Valencia Late y las variedades derivadas a partir de esta) y naranjas sanguinas (Tarocco Rosso, Sanguinelli). Las variedades derivadas de otra variedad deberían, en principio, presentar al menos los mismos marcadores moleculares que la variedad madre.

5. 2. Secuenciación de genomas

Para la búsqueda de polimorfismos que pudieran ser utilizados para discriminar entre distintas variedades de naranjo dulce se partió del trabajo realizado por el consorcio Citrusseq-Citrusgenn, conformado por instituciones públicas y privadas. Este grupo realizó la secuenciación de distintas variedades de cítricos (entre ellas, naranjos) mediante la tecnología Illumina Paired-End Sequencing⁴⁴.

La secuenciación Illumina Paired-End (Figura 5) implica la fragmentación del ADN genómico a secuenciar. Los fragmentos de ADN obtenidos, de un tamaño aproximado de 500 pb, son secuenciados por los dos extremos, obteniendo lecturas emparejadas forward y reverse (de ahí el nombre de paired-ends, o paired reads). Tras la obtención de la secuencia de los extremos, los fragmentos son alineados respecto a una secuencia de referencia, en este caso *Citrus clementina* Hort. ex Tan¹⁵, permitiendo obtener el orden de los fragmentos y con ello la secuencia del genoma secuenciado y las posibles variaciones estructurales del genoma. Debido a que la distancia entre las lecturas de los extremos es conocida, los algoritmos de alineamiento permiten la utilización de esta información para mapear las lecturas de una manera más precisa.

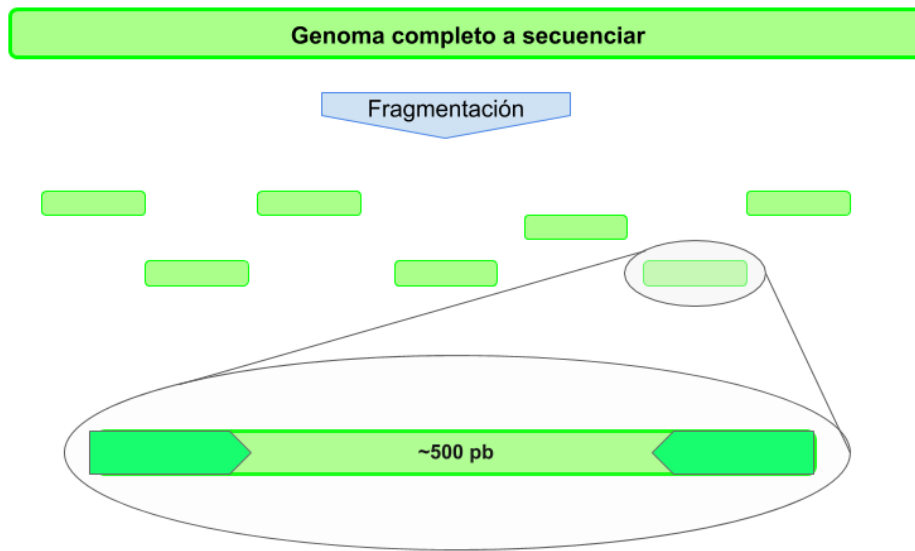


Figura 5. Secuenciación mediante la tecnología Illumina Pair End Sequencing. El genoma a secuenciar es fragmentado en pequeñas moléculas de ADN con una longitud aproximada de 500 pb. Ambos extremos de los fragmentos generados son secuenciados, obteniendo una librería de lecturas emparejadas (paired reads).

5. 3. Manejo del software Integrative Genomics Viewer

El alineamiento de las lecturas de una variedad de naranjo dulce sobre el genoma de referencia genera una distribución específica de cada variedad. El alineamiento queda almacenado en una serie de ficheros, denominados BAM (*Binary Alignment Map*). El visualizador Integrative Genomics Viewer o IGV^{45,46} permite, leyendo archivos BAM, ver la distribución y el alineamiento de las lecturas del genoma evaluado respecto al genoma de referencia. Un alineamiento "correcto", es decir, un alineamiento de lecturas cuya secuencia corresponde con la secuencia del genoma de referencia se visualizará con una etiqueta de color gris (Figura 6). Un alineamiento "erróneo" se produce cuando las lecturas no son capaces de alinear correctamente respecto al genoma de referencia (debido a que la lectura y el genoma de referencia no tienen la misma secuencia en una posición concreta del genoma), total o parcialmente, generando lecturas con etiquetas de distintos colores. Estas incongruencias en el alineamiento, como por ejemplo, la obtención de distancias distintas a las esperadas entre las lecturas forward y reverse (lecturas emparejadas o paired reads), o la presencia de lecturas emparejadas con extremos alineados en diferentes cromosomas determinará la presencia de una variación estructural.

El barrido comparativo de los cromosomas de distintas variedades permite visualizar las diferencias entre sus genomas. Este barrido puede llevarse a cabo de manera manual, explorando las distintas posiciones de cada uno de los cromosomas de la variedad en cuestión mediante el software IGV, o de manera automatizada: mediante distintas herramientas bioinformáticas que detectan variaciones estructurales.

Un ejemplo sencillo de detección de una deleción sería el siguiente (Figura 7): se obtiene un alineamiento del genoma a evaluar respecto al genoma de referencia donde las lecturas emparejadas están separados por una distancia mucho mayor a 500 pb (que sería la distancia a la que alinearía si el genoma que estamos evaluando fuera igual al genoma de referencia), por ejemplo a 10.000 pb. Estas lecturas emparejadas lejanamente aparecen de color rojo. Esto determina que el genoma evaluado, donde la distancia entre los extremos de las lecturas es de 500 pb aproximadamente, es discordante con el genoma de referencia, donde la distancia entre estos mismos extremos es de 10.000 pb.

El análisis de las lecturas discordantes puede permitir dilucidar la posición de roturas (el lugar donde comienza y acaba la deleción). También, debido a que las lecturas que incluyen el sitio de ruptura presentan los dos flancos de la deleción, estas se representarán además de color rojo, con barras verticales coloreadas, simbolizando las bases desapareadas respecto al genoma de referencia. Las Figuras 7 y 8 esquematizan las lecturas alineadas con un genoma de referencia, parte de las cuales coincide en parte con el genoma de referencia y otra parte con 10.000 posiciones downstream a la ruptura).

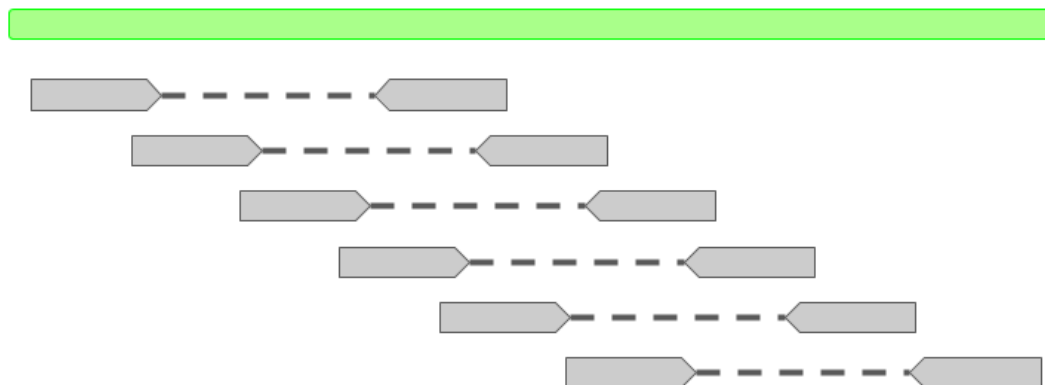


Figura 6. Alineamiento de los fragmentos generados durante la secuenciación (pentágonos unidos entre sí por una línea discontinua) respecto al genoma de referencia (verde). Las lecturas emparejadas se muestran en color gris debido a que su alineamiento es coherente con el genoma de referencia: la secuencia de ambos extremos se encuentran separadas por 500 pb, aproximadamente la longitud de los fragmentos generados durante la secuenciación, al igual que en el genoma de referencia. El alineamiento de los fragmentos con extremos secuenciados (tecnología Illumina Paired-End Sequencing) permite a los algoritmos de alineamiento utilizar esta información para mapear las lecturas de una manera más precisa.

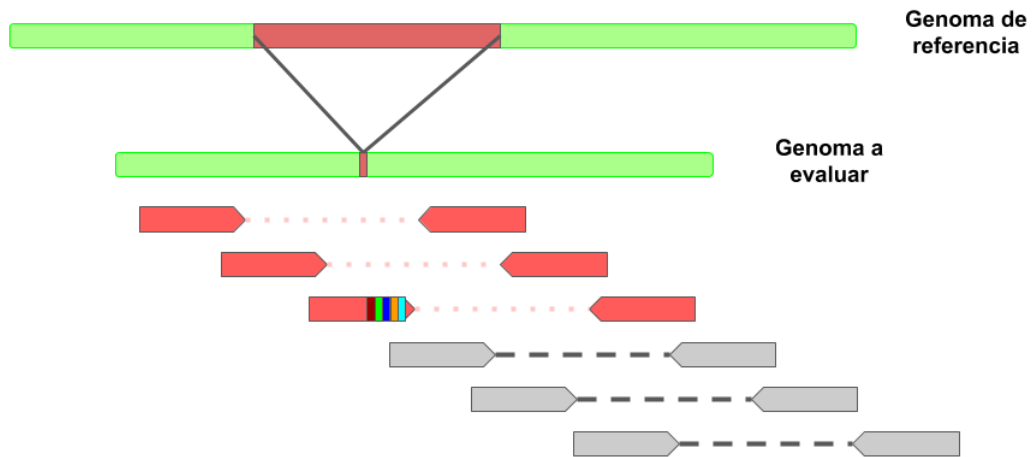


Figura 7. Detección de una deleción. Un cromosoma del genoma de referencia (parte superior) ha sufrido una deleción, perdiendo el fragmento de color rojo, obteniendo un cromosoma en el genoma a evaluar que no presenta esa región del genoma (se representa la ruptura de con una pequeña sección en color rojo). En el alineamiento del genoma evaluado se obtienen zonas donde no se da un alineamiento canónico: paired-reads que deberían estar separadas por aproximadamente 500 pb están separadas por una distancia mucho mayor (10.000 pb en el ejemplo del texto; se muestran las paired-reads con alineamiento no canónico y canónico como pentágonos de color rojo y gris, respectivamente): estas secuencias alinean en regiones distintas del cromosoma (separadas por 10.000 pb aproximadamente, en el ejemplo del texto). El genoma evaluado, donde la distancia entre los extremos es de 500 pb aproximadamente, es distinto al genoma de referencia, donde la distancia entre estos mismos extremos es de 10.000 pb. Otro indicio de esta deleción es la presencia de paired-reads que alinean exactamente en el sitio de ruptura: estas lecturas presentan los dos flancos de la deleción (representadas de color rojo y barras verticales coloreadas, representando las bases desapareadas respecto al genoma de referencia, véase Figura 8).

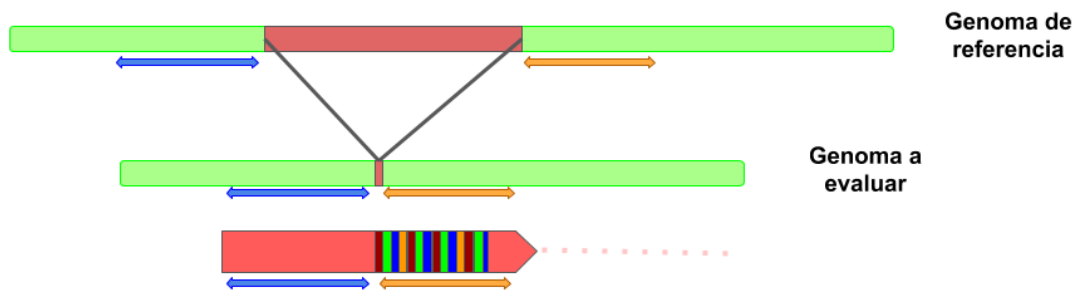


Figura 8. Alineamiento no canónico de una lectura (representada como un pentágono rojo) sobre un sitio de ruptura de una deleción. Se representan con flechas azules y naranjas fragmentos del cromosoma con la misma secuencia en el genoma de referencia, en el genoma evaluado y en la lectura del genoma evaluado. Debido a que la secuencia de la lectura incluye las secuencias flanqueantes a la deleción esta presentará una zona cuya secuencia es la misma de la del genoma de referencia y otra zona cuya secuencia no concuerda con el genoma de referencia, que corresponde a la secuencia del flanco downstream de la deleción, y se representa con barras verticales coloreadas (bases desapareadas respecto al genoma de referencia).

El presente trabajo realizado en el Centro de Genómica del IVIA se basa en la detección *in silico* de marcadores moleculares mediante la comparación de genomas de diferentes variedades de naranjo dulce, previamente secuenciadas en el Centro de Genómica del IVIA bajo el proyecto Citrusseq-Citrusgenn. En principio, cualquier discrepancia que se obtenga en el análisis de los genomas de las distintas variedades puede ser utilizada para discriminarlas.

6. Objetivo

El objetivo principal de este trabajo es la obtención de marcadores moleculares que permitan la identificación de grupos o variedades de naranjo dulce mediante comparación genómica. De tal modo, se desglosan aquí los objetivos que permitirían llegar a tal propósito:

- Detección de variaciones genómicas *in silico* de tipo deleción, translocación, inversión o duplicación mediante comparación genómica de las variedades de naranjas registradas hasta la fecha en la base de datos público-privada del Centro de Genómica del IVIA.
- Desarrollo y acotación de marcadores moleculares que permitan diferenciar e identificar inequívocamente ciertos grupos varietales de naranjas dulces.
- Validación *in vitro* de los marcadores moleculares detectados que permitan diferenciar e identificar inequívocamente ciertos grupos varietales de naranjas dulce.

7. Material y métodos

7. 1. Genoma de referencia

El genoma de referencia empleado es el primer genoma de referencia de cítricos que se obtuvo mediante secuenciación con tecnología Sanger¹⁵ a partir de un individuo haploide de la variedad de clementina Clementina de Nules (*Citrus clementina* Hort. ex Tan.) obtenido por ginogénesis *in vitro*⁴⁷.

7. 2. Genomas de las variedades y material vegetal

Se han utilizado los genomas de las variedades comerciales de *Citrus sinensis* (L.) (Osb.): Blanca Comuna, Salustiana, Cadenera, Tarocco Rosso, Doblefina, Sanguinelli, Valencia Late Frost, Valencia Late Delta, Midnight, Barberina, Washington Navel, Barnfield, Navelate, Ricalate, Lane Late, Cara Cara, Navelina, Powell, Fukumoto, Chislett secuenciadas en el consorcio Citrusseq-Citrusgenn y registradas en el Centro de Genómica IVIA. El material vegetal de las mismas fue obtenido del banco de germoplasma del IVIA y de parcelas comerciales y privadas.

En el IVIA se disponía previamente a mi incorporación de muestras de ADN de las variedades Barnfield, Cadenera, Doblefina, Navelina, Sanguinelli y Chislett. El material vegetal necesario para extraer ADN de las variedades Washington Navel, Blanca Comuna, Salustiana y Valencia Late Frost procedió del banco del germoplasma del IVIA.

7. 3. Búsqueda de marcadores moleculares *in silico*

La búsqueda de marcadores moleculares *in silico* se basó en la comparación genómica de la secuenciación genómica mediante tecnología Illumina Pair-End.

La comparación de genomas se realizó empleándose el visualizador de genomas Integrative Genomics Viewer^{45,46}. El visualizador muestra la distribución del alineamiento de las lecturas del genoma evaluado sobre el genoma de referencia almacenados en los ficheros BAM (*Binary Alignment Map*). En este proceso se utilizaron tanto las comparaciones visuales de los ficheros BAM como las combinaciones de pasos bioinformáticos (scripts) desarrollados por el equipo del Centro de Genómica basados en algoritmos que apoyándose en diferentes parámetros permiten encontrar zonas del genoma con altas probabilidades de presentar variaciones estructurales.

7. 4. Diseño de oligonucleótidos

El diseño de oligonucleótidos se realizó mediante el software en línea Primer3⁴⁸, empleándose como secuencia molde el fragmento genómico reconstruido *in silico* a partir de la interpretación de las variantes genómicas. La visualización del mismo se realizó mediante el software BioEdit Sequence Alignment Editor⁴⁹.

Los oligonucleótidos diseñados oscilaron en un rango de tamaños entre 18 y 23 pb, una temperatura de hibridación entre 59 °C y 61 °C y un contenido en GC entre 45% y 55% con valores óptimos de tamaño, temperatura y % GC de 20 pb , 60 °C y 50%, respectivamente. Se seleccionaron preferentemente aquellos cuyo extremo 3' terminaba en citosina o guanina, sin tener más de cinco citosinas o guaninas y se descartaron los oligonucleótidos con más de tres nucleótidos repetidos en tándem. El fragmento del producto amplificado estuvo comprendido entre 150 y 900 pb.

La posición de al menos uno de los oligonucleótidos de cada pareja se escogió a una distancia mínima de 50 pares de bases respecto al límite de la delección para su correcta visualización en el electrofenograma del producto de PCR secuenciado mediante tecnología Sanger.

7. 5. Extracción de ADN genómico

El protocolo de extracción de ADN genómico a partir de material vegetal incluyó la lisis celular y una inactivación de endonucleasas.

Se trituraron 0.1 g de hoja fresca congelada a -80 °C empleándose un TissueLyser (QIAGEN). Se le adicionó 800 µl CTAB 2%, con 2-mercaptoetanol al 0.2%, precalentado a

65 °C, incubándolo en agitación suave durante 15 minutos a 65 °C. Tras la incubación se añadió 800 µl de cloroformo: alcohol isoamílico (24:1), homogeneizando por inversión y centrifugándolo 15 minutos a 13000 rpm, 4 °C. Se adicionó 1 volumen de isopropanol al sobrenadante obtenido dejándolo precipitar durante 1 hora a -80 °C. El ADN precipitado fue centrifugado durante 15 minutos a 13000 rpm, 4 °C eliminándose el sobrenadante tras la centrifugación. El ADN obtenido fue lavado adicionando 400 µl de etanol 70% y posterior centrifugado durante 5 minutos a 12000 rpm, 4 °C. El ADN obtenido fue resuspendido en 50 µl de tampón TE 1X al cual se añadió 1 µl de RNAsa, incubándose a 37 °C durante 30 minutos.

7. 6. Cuantificación de cantidad y calidad del ADN extraído

La cuantificación del ADN se realizó mediante espectrofotometría con un instrumento Nanodrop (Thermo Fisher) que proporciona las relaciones de absorbancia A_{260}/A_{280} y A_{260}/A_{230} , indicadores de la calidad del ADN extraído, siendo el óptimo de la relación A_{260}/A_{280} un valor de 1,8 en el caso de ADN y el óptimo de la relación A_{260}/A_{230} un valor entre 1,8 y 2,2.

Teniendo en cuenta que si la relación A_{260}/A_{280} tiene un valor menor al óptimo puede indicar la presencia de proteínas, fenol u otros contaminantes con fuerte absorbancia a 280 nm; y si la relación A_{260}/A_{230} tiene un valor menor a 1,8 puede indicar la presencia de contaminantes como sales, glúcidos o fenoles; la extracción de ADN cuyas relaciones de calidad no fueron óptimas se descartaron hasta obtener extracciones con valores de relaciones de absorbancia adecuados.

7. 7. Reacciones de PCR en tiempo real

La amplificación mediante PCR en tiempo real se llevó a cabo en el *LightCycler® 2.0 Real-Time PCR System* (Roche Life Science, a partir de ahora LightCycler) que mediante la detección de fluorescencia permite la cuantificación del producto de PCR a medida que este se sintetiza. La reacción de PCR en tiempo real se realizó empleándose la mezcla enzimática preparada en hielo de 2 ng de ADN, 0.05 µl de cada par de oligonucleótidos a 5 mM y 2 µl de la mezcla del kit *LightCycler® FastStart ADN Master SYBR Green I* (Roche Life Science) que contiene en disolución el fluoróforo SYBR Green I, deoxinucleótidos trifosfato, DNA polimerasa Taq FastStart y $MgCl_2$ siendo el volumen final de reacción de 10 µl. La reacción de PCR en tiempo real se realizó en tres etapas: una primera etapa de incubación a 95 °C durante 10 minutos; seguida de una segunda etapa de 45 ciclos que comprendía 10 segundos a 95 °C (fase de desnaturalización), 5 segundos a 60 °C (fase de hibridación) y entre 40 y 60 segundos a 72 °C (fase de elongación), en función del tamaño de fragmento amplificado siendo determinada, en segundos, como la relación entre el número de pares de bases del amplicón y 25; y una fase final de 30 segundos a 72 °C (fase final de elongación). En cada reacción de PCR en tiempo real se incluyó un control negativo al cual no se le añadió ADN.

7. 8. Comprobación de los marcadores moleculares

El producto de PCR en tiempo real obtenido fue secuenciado mediante tecnología Sanger para validación de los marcadores moleculares obtenidos.

Para ello, se seleccionó la reacción de PCR positiva (en la que se evidenció amplificación utilizando oligonucleótidos forward y reverse que hibridaban en posiciones upstream y downstream, respectivamente, del punto de rotura de la deleción) y se mezcló con 10 μ L de agua DEPC y 2 μ L de uno de los oligonucleótidos utilizado en la reacción de PCR en tiempo real. La elección de los oligonucleótidos determinó la dirección de la secuenciación, en función de si la secuenciación parte de la región upstream o downstream de la rotura. De modo que se seleccionó el oligonucleótido cuya posición de hibridación hasta la posición de cambio no contuviera regiones poli-A o poli-AT (como máximo 7 repeticiones) ni INDELS.

8. Resultados

8. 1. Detección de variaciones estructurales del genoma

8. 1. 1. Deleción de 1.937 pb en el cromosoma 1 de Washington Navel

Mediante el barrido comparativo de los ficheros BAM de las variedades Blanca Comuna y Washington Navel se detectó una deleción en el cromosoma 1, entre las posiciones 25.588.663 pb y 25.590.600 pb. Al acudir a estas posiciones, en la variedad Washington Navel se encontraron numerosas lecturas emparejadas de color rojo, separadas por una distancia de 2.000 pb aproximadamente, lo que fue el indicio de una deleción (Figura 9).

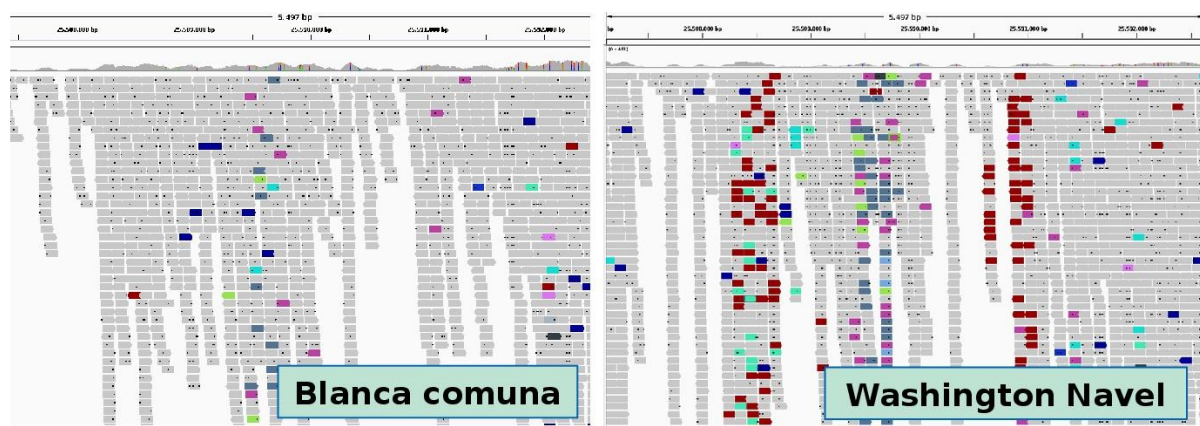


Figura 9. Vista de la región de la deleción, en el cromosoma 1, entre las posiciones 25.587.000 pb y 25.592.700 pb. La variedad de la izquierda es Blanca Comuna. La variedad de la derecha es Washington Navel, que muestra gran cantidad de lecturas emparejadas de color rojo. La franja roja de la izquierda en la variedad Washington Navel esta formada por lecturas que emparejan con lecturas de la franja derecha a una distancia aproximada de 2.000 pb, y viceversa, lo que indica la presencia de una deleción de este tamaño entre estas franjas.

Para confirmar que en esta región del cromosoma 1 se había dado una deleción en la variedad Washington Navel fue necesario acercarse a cada uno de los supuestos flancos de la deleción (las regiones en las que predominan las lecturas de color rojo). Al ampliar cada una de estas zonas, las posiciones 25.588.663 pb y 25.590.600 pb (Figura 10), se

apreciaron lecturas con bases no concordantes con el genoma de referencia. Estas bases no concordantes con el genoma de referencia en estas posiciones correspondían al flanco opuesto de la deleción, lo que confirmó de forma *in silico* esta variación.

Véase la Figura 10 en la que se recuadran del mismo color secuencias iguales en distintas posiciones. En la posición 25.588.663 pb (Figura 10A, flanco upstream de la deleción) se encontraron lecturas con una parte gris, región que concordaba con el genoma de referencia (secuencia recuadrada de naranja), y una parte donde aparecen bases desapareadas, ya que se trata de una secuencia que no concordaba con el genoma de referencia (secuencia recuadrada de azul). Esta secuencia discordante es la secuencia del flanco downstream de la deleción (véase el recuadro azul en la posición 25.590.600 pb, correspondiente al genoma de referencia, Figura 10B).

De igual modo, en la posición 25.590.600 pb (Figura 10B, flanco downstream de la deleción) se encontraron lecturas con parte gris, correspondientes a la región que concordaba con el genoma de referencia (secuencia recuadrada de azul) y una parte con bases desapareadas, una secuencia que no concordaba con el genoma de referencia (secuencia recuadrada de naranja). Esta secuencia coincide con la secuencia del flanco upstream de la deleción (Figura 10A).

Debido a que los sitios de ruptura de la deleción se encontraron en las posiciones 25.588.663 pb y 25.590.600 pb se determinó que la deleción tiene una longitud de 1.937 pb.

La presencia o de la deleción fue evaluada en diferentes variedades de naranjo dulce, obteniendo los siguientes resultados (Figura 11).

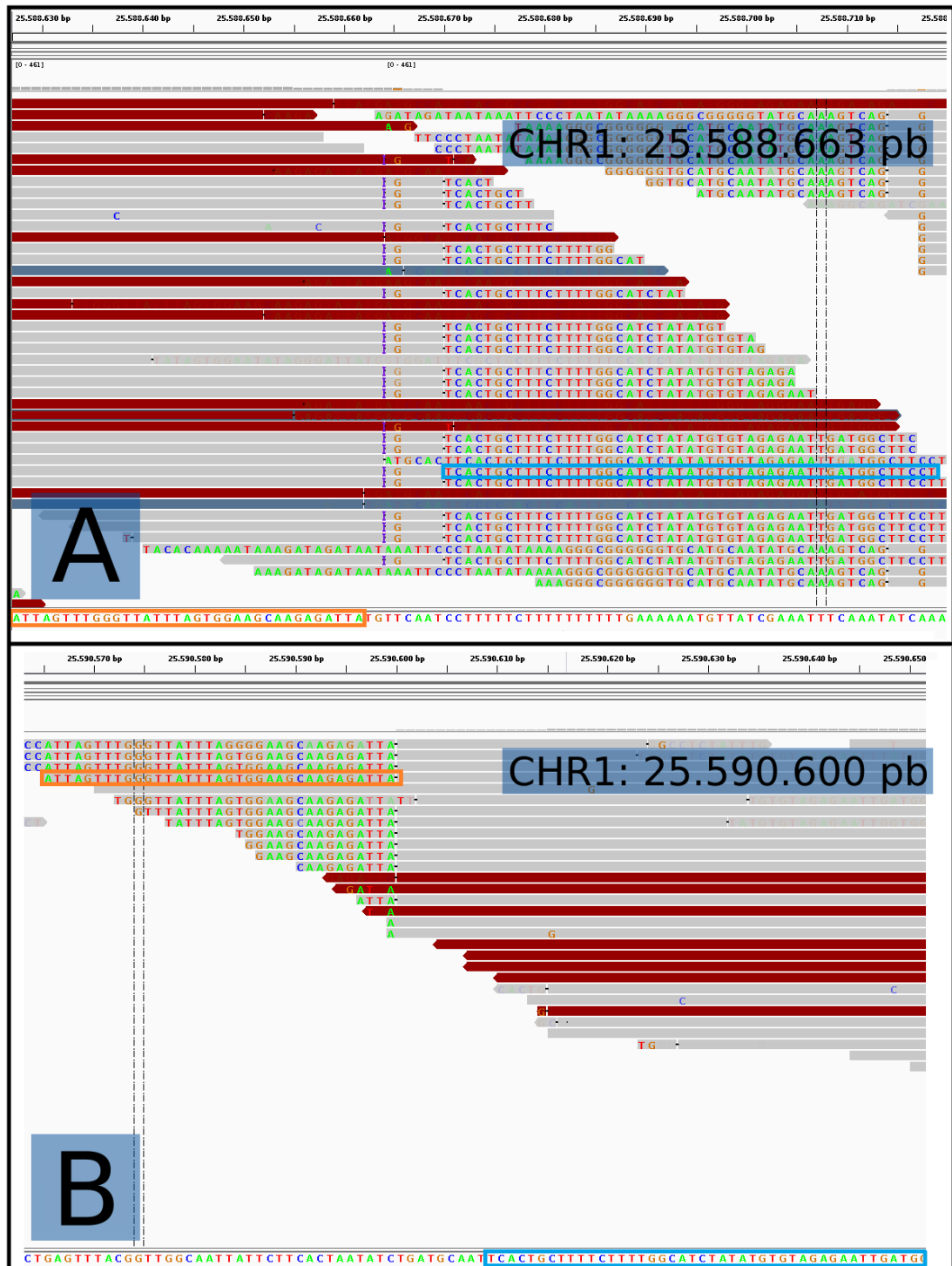


Figura 10. Alineamientos en las zonas de ruptura de la delección del cromosoma 1 desde la posición 25.588.663 a 25.590.600 en la variedad Washington Navel. A: flanco upstream. B: flanco downstream. Las secuencias recuadradas de color naranja son coincidentes en las dos posiciones, así mismo ocurre con las secuencias recuadradas de color azul, lo que evidenció la presencia de la delección. Las posiciones donde se encontró la ruptura son 25.588.663 y 25.590.600, resultando una delección de 1.937 pb.

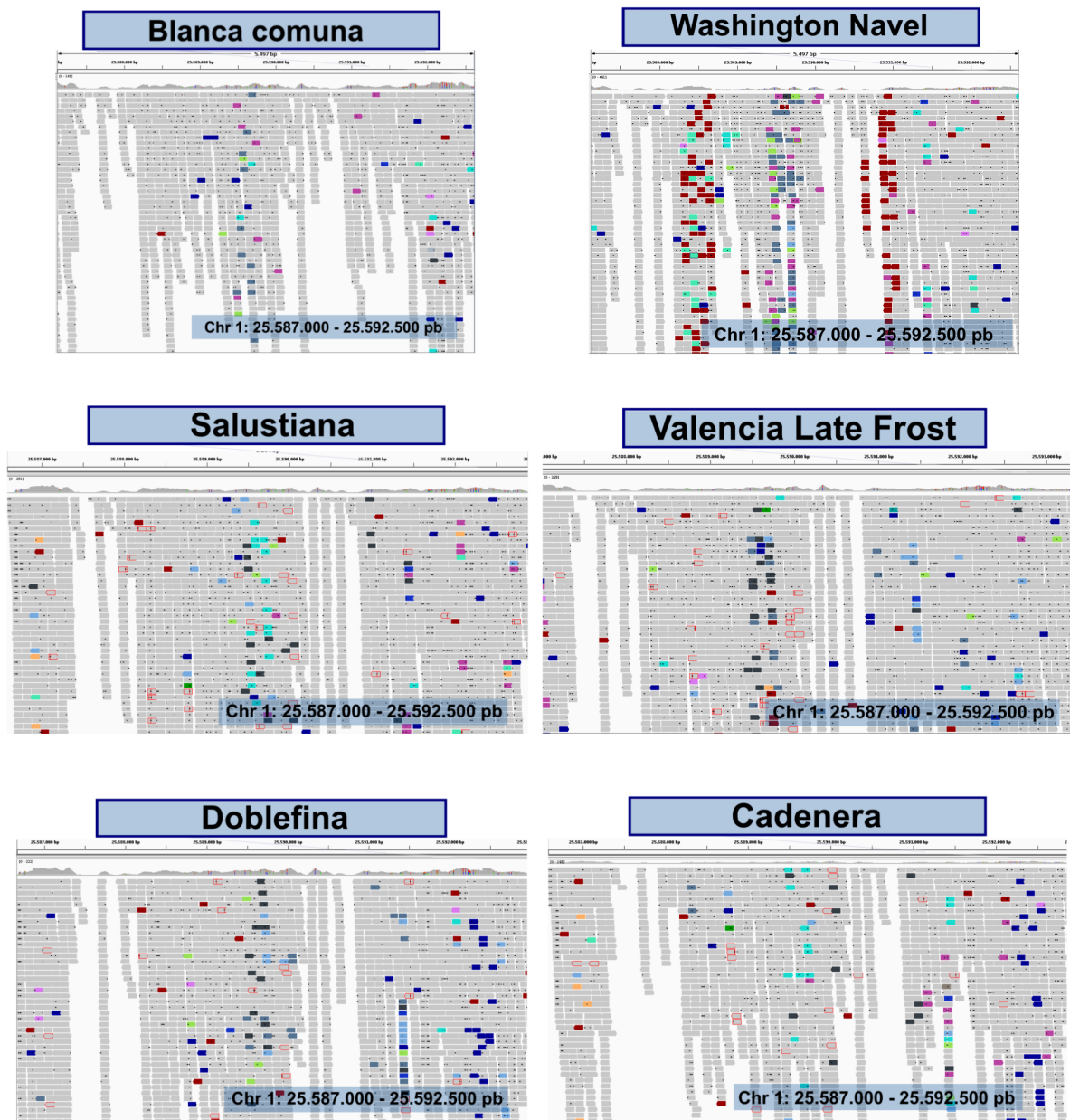


Figura 11. Evaluación de la deleción de 1.937 pb en el cromosoma 1 en las variedades de naranjo dulce Blanca Comuna, Washington Navel, Salustiana, Valencia Late Frost, Doblefina y Cadenera. Se muestran en las imágenes las posiciones 25.587.000 – 25.592.500 del cromosoma 1.

Entre las seis variedades, la deleción se encontró solamente en la variedad Washington Navel. Véase la variedad Washington Navel en la Figura 11: se muestran las lecturas en color rojo, flanqueando el lugar de la deleción.

La comparativa *in silico* de la deleción del cromosoma 1 desde la posición 25.588.663 pb a la posición 25.590.600 pb mostró que todas las variedades del grupo Navel analizadas presentaban dicha deleción, estando ausente en las variedades de los otros grupos de naranja (Tabla 1), apoyando la relación filogenética descrita previamente (Figura 4).

Tabla 1. Evaluación in silico mediante visualización de los alineamientos de secuenciación de las variedades estudiadas de naranjo dulce de la existencia (Presente ✓) o no (Ausente ✗) de la delección de 1.937 pb en el cromosoma 1.

Variedad de naranjo dulce	Presencia / ausencia de delección 1.937 pb cromosoma 1
Blanca comuna	Ausente ✗
Washington Navel	Presente ✓
Barnfield	Presente ✓
Navelate	Presente ✓
Ricalate	Presente ✓
Lane Late	Presente ✓
Cara Cara	Presente ✓
Navelina	Presente ✓
Powell	Presente ✓
Fukumoto	Presente ✓
Chislett	Presente ✓
Salustiana	Ausente ✗
Cadenera	Ausente ✗
Tarocco Rosso	Ausente ✗
Doblefina	Ausente ✗
Sanguinelli	Ausente ✗
Valencia Late Frost	Ausente ✗
Valencia Late Delta	Ausente ✗
Midnight	Ausente ✗
Barberina	Ausente ✗

8. 1. 2. Delección de 104.916 pb en el cromosoma 1 de Washington Navel

Mediante el barrido comparativo de los ficheros BAM de las variedades Blanca Comuna y Washington Navel se detectó una delección entre las posiciones 15.321.770 pb y 15.426.686 pb del cromosoma 1. Al examinar estas posiciones en la variedad Washington Navel se encontraron dos flancos de delección, marcados por numerosas lecturas emparejadas de color rojo, separadas por una distancia aproximada de 100 Kb (Figura 12).

Para confirmar la delección se ampliaron los presuntos flancos de la delección en la variedad Washington Navel: flanco upstream 15.321.000 pb – 15.323.000 pb; y flanco downstream 15.426.000 pb – 15.528.000 pb. Al ampliar ambos flancos se apreciaron lecturas con bases desapareadas; las bases de cada flanco correspondían al flanco opuesto de la delección, lo que corroboró la delección en Washington Navel (Figura 13).

En la Figura 13 se recuadran del mismo color secuencias idénticas en los distintos flancos de la delección. En la posición 15.321.770 pb (Figura 13A, flanco upstream de la delección) se encontraron lecturas con una región gris, lo que correspondió a una secuencia concordante con el genoma de referencia (secuencia recuadrada de naranja); y una región con bases desapareadas (secuencia recuadrada de azul). Estas bases no concordaban con el genoma de referencia, si no que coincidían con la secuencia del flanco downstream de la delección (véase el recuadro azul en la posición 15.426.686 pb, correspondiente al genoma de referencia, Figura 13B).

Análogamente, en la posición 15.426.686 pb (Figura 13B, flanco downstream de la delección), aparecían lecturas con una región gris, lo que correspondía a una secuencia que concuerda con el genoma de referencia (secuencia recuadrada de azul) y una región con bases desapareadas (secuencia recuadrada de naranja), ya que esta secuencia no coincidía con el genoma de referencia. Esta secuencia coincidió con la secuencia del flanco upstream de la delección (véase el recuadro naranja en la posición 15.321.770 pb, parte superior de la Figura 13).

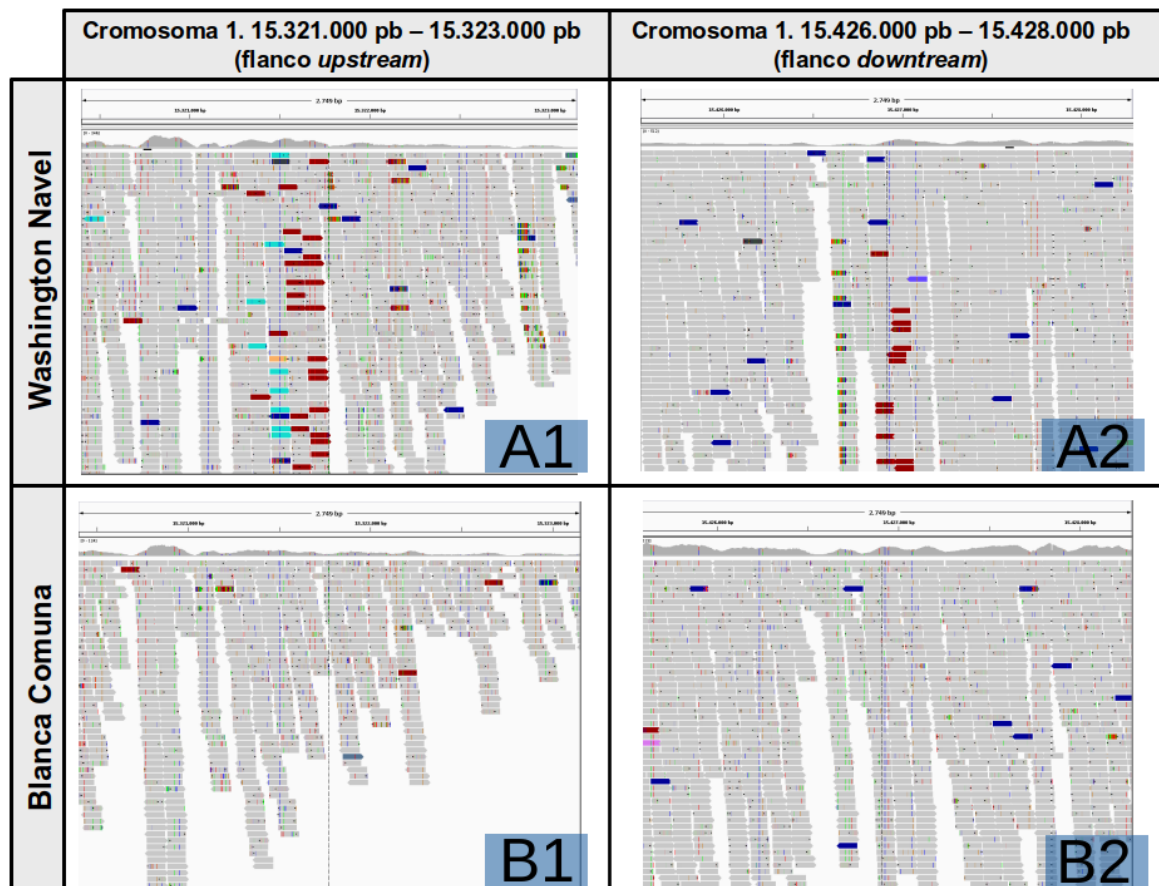


Figura 12. Vistas del alineamiento de lecturas emparejadas secuenciadas de las variedades Washington Navel y Blanca Comuna en las regiones del flanco upstream (15.321.000 pb – 15.323.000 pb) y downstream (15.426.000 pb – 15.528.000 pb) de la delección detectada. A1: flanco upstream en la variedad Washington Navel. A2: flanco downstream en la variedad Washington Navel. B1: flanco upstream en la variedad Blanca Comuna. B2: flanco downstream en la variedad Blanca Comuna. En el flanco upstream de la variedad Washington Navel (A1) se encontrarán numerosas lecturas de color rojo que emparejaban con las lecturas del flanco downstream (A2) y viceversa, a una distancia de aproximadamente 100 Kb, lo que indicó la presencia de una delección de un tamaño cercano a esta longitud.

Debido a que los sitios de ruptura de la deleción se encontraron en las posiciones 15.321.770 pb y 15.426.686 pb se determinó que la deleción tiene una longitud de 104.916 pb.

Se evaluó la presencia *in silico* de deleción en diferentes variedades de naranjo dulce, obteniendo los siguientes resultados (Figura 14 – 19, Tabla 2).

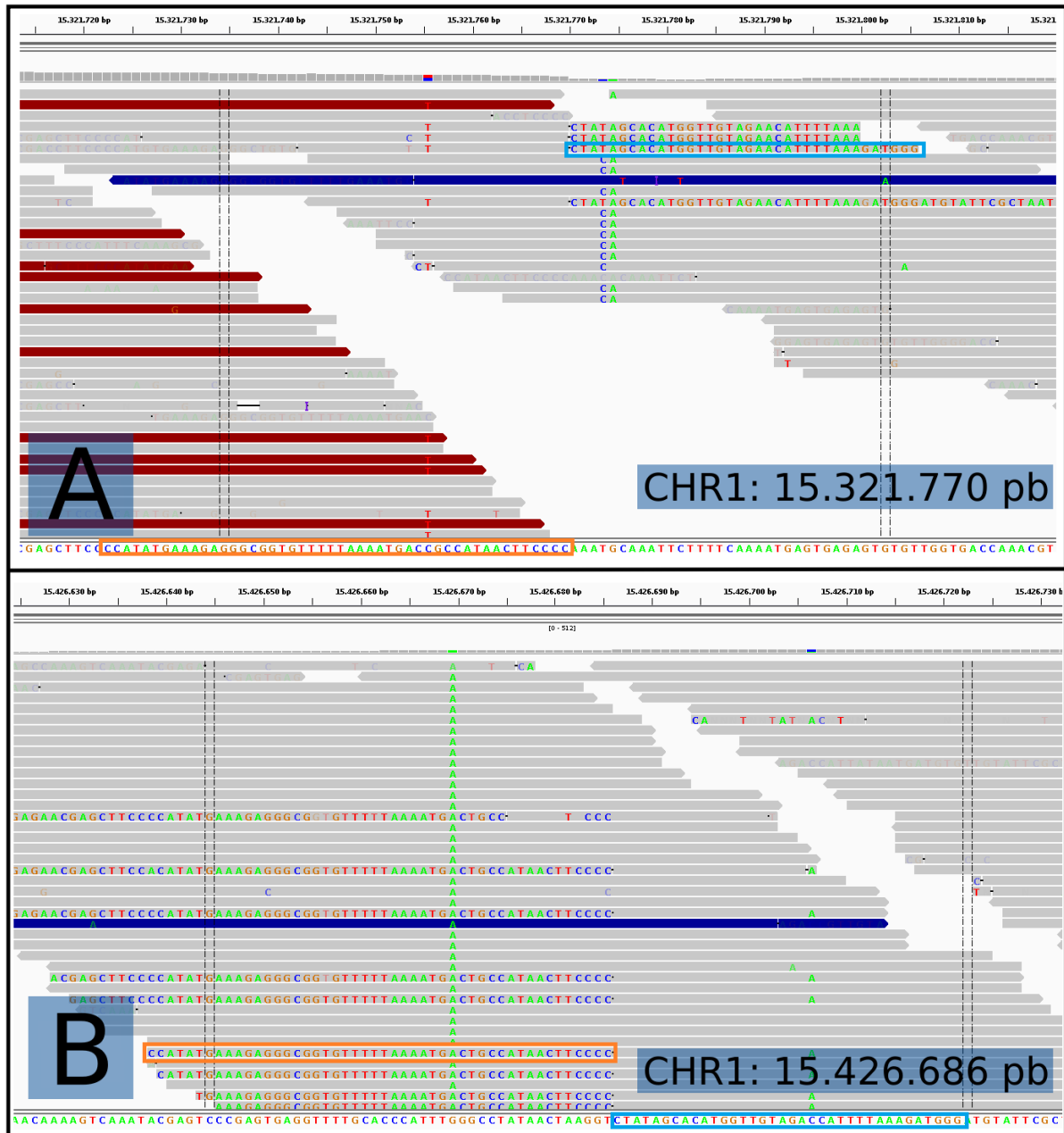


Figura 13. Alineamientos en las zonas de ruptura de la deleción (o flancos de la deleción) en la variedad Washington Navel. A: flanco upstream; B: flanco downstream. Las secuencias recuadradas de color naranja son coincidentes en las dos posiciones del cromosoma 1 (15.321.770 y 15.426.686, figuras A y B respectivamente), así mismo ocurre con las secuencias recuadradas de color azul, lo que evidencia la presencia de la deleción. Las posiciones donde se encuentra la ruptura son 15.426.686 y 15.321.770, resultando una deleción de 104.916 pb.

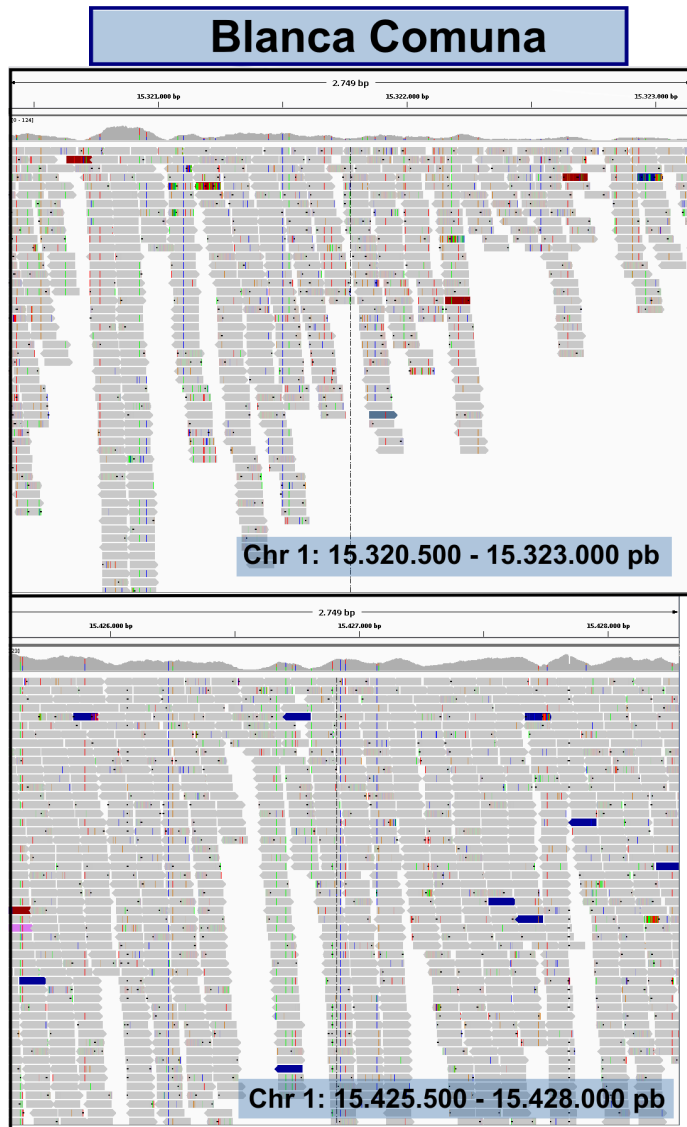


Figura 14. Evaluación de la delección de 104.916 pb en el cromosoma 1 en la variedad Blanca Comuna. Se muestran en la figura las posiciones 15.320.500 – 15.323.000 pb (superior) y 15.425.500 – 15.428.000 pb (inferior) del cromosoma 1. En estas regiones no se encontraron lecturas que pudieran evidenciar la presencia de una delección.

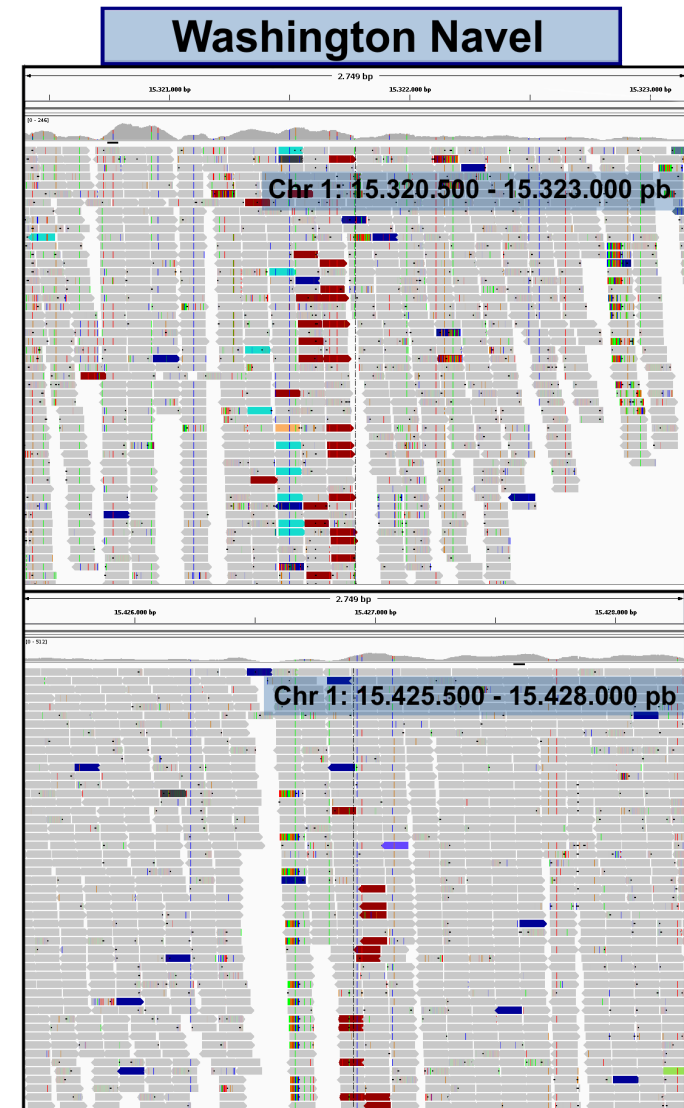


Figura 15. Evaluación de la delección de 104.916 pb en el cromosoma 1 en la variedad Washington Navel. Se muestran en la figura las posiciones 15.320.500 – 15.323.000 pb (superior) y 15.425.500 – 15.428.000 pb (inferior) del cromosoma 1. En estas regiones se encontraron lecturas emparejadas separadas por una distancia aproximada de 100 Kb, marcadas de color rojo, lo que evidenció la delección.

Salustiana

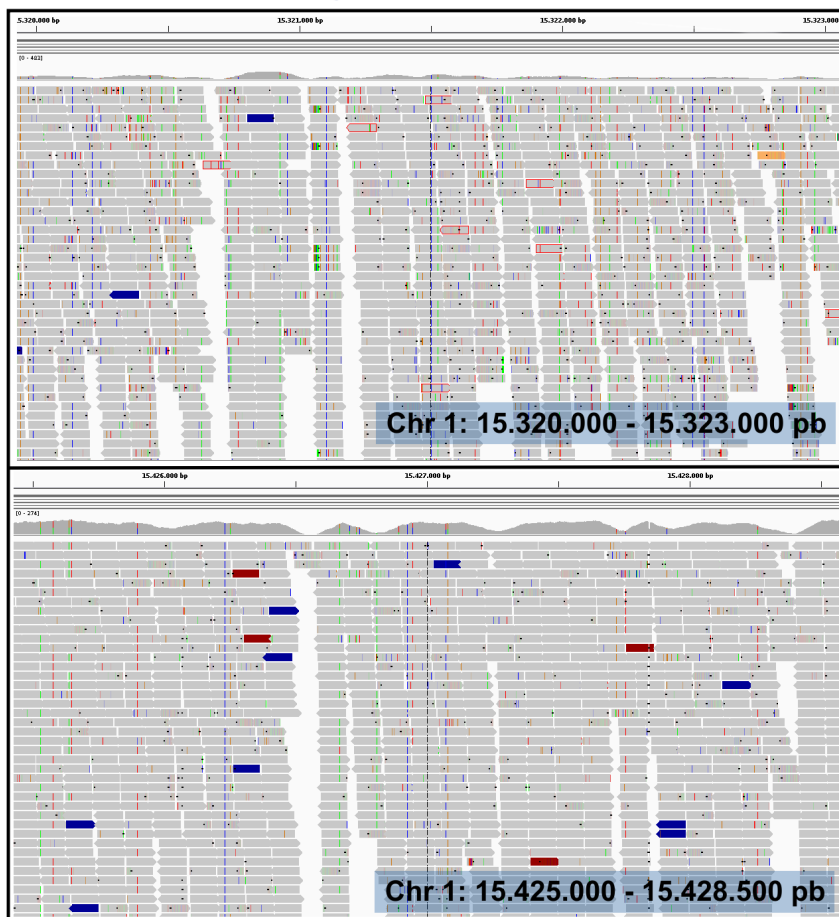


Figura 16. Evaluación de la deleción de 104.916 pb en el cromosoma 1 en la variedad Salustiana. Se muestran en la figura las posiciones 15.320.000 – 15.323.000 pb (superior) y 15.425.000 – 15.428.500 pb (inferior) del cromosoma 1. En estas regiones no se encontraron lecturas emparejadas que pudieran evidenciar la deleción.

Valencia Late Frost

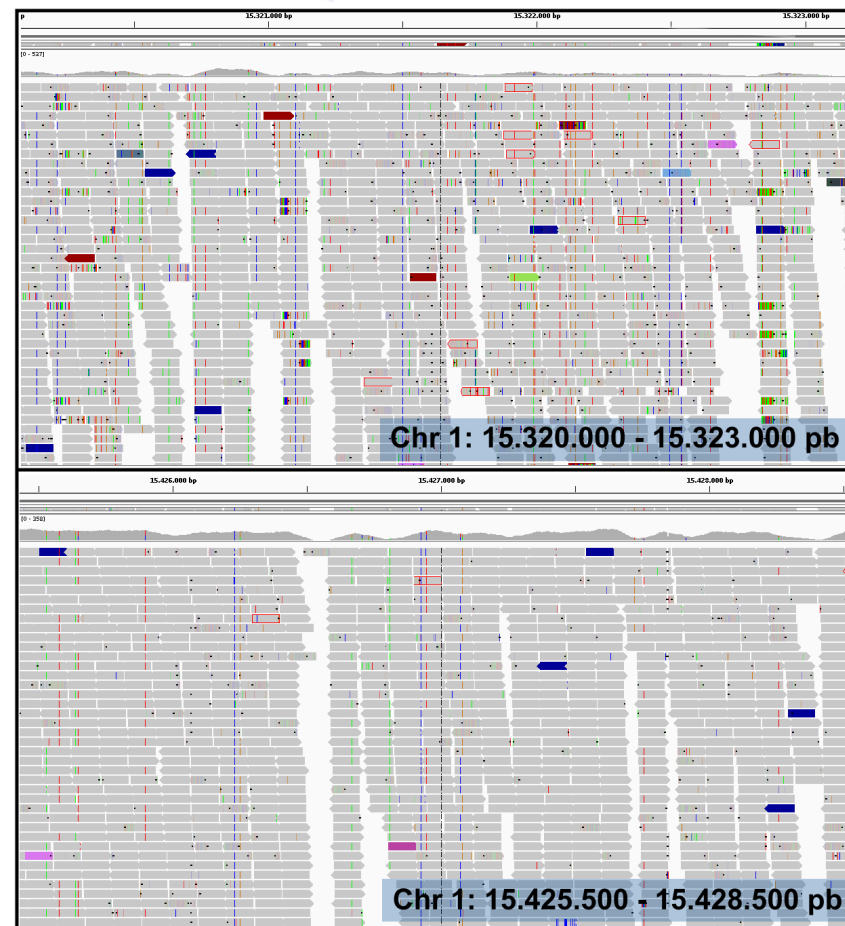
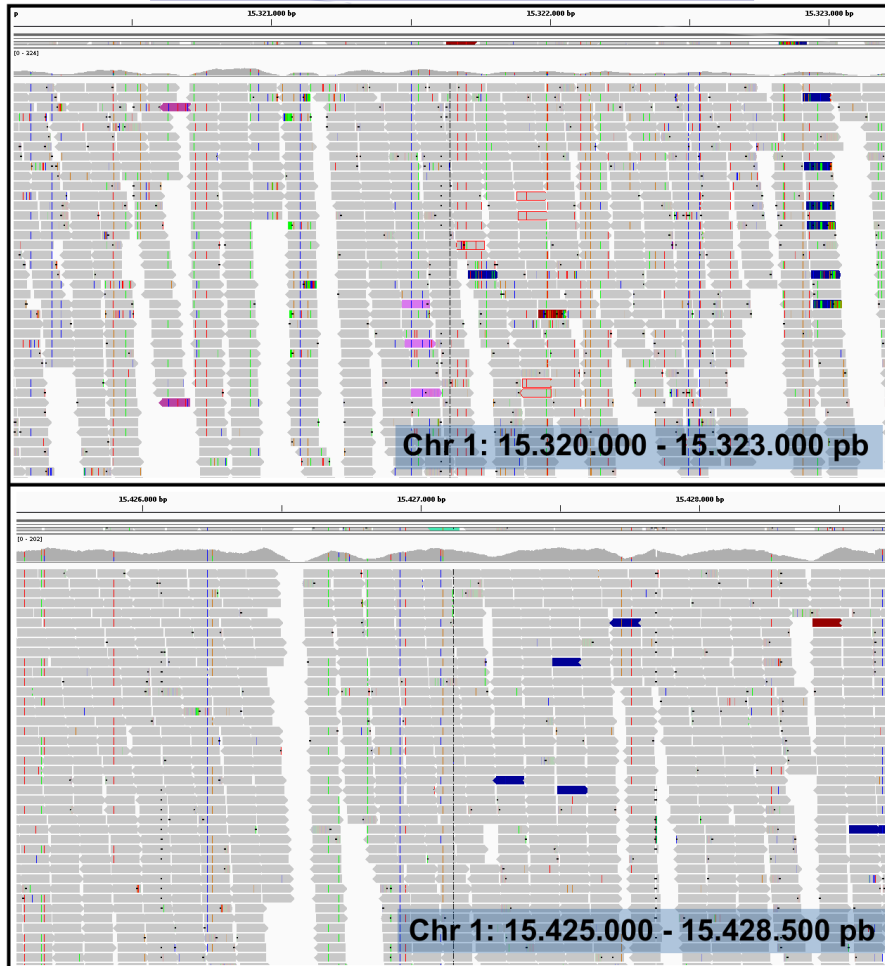


Figura 17. Evaluación de la deleción de 104.916 pb en el cromosoma 1 en la variedad Valencia Late Frost. Se muestran en la figura las posiciones 15.320.000 – 15.323.000 pb (superior) y 15.425.500 – 15.428.500 pb (inferior) del cromosoma 1. En estas regiones no se encontraron lecturas emparejadas que pudieran evidenciar la deleción.

Doblefina



Cadenera

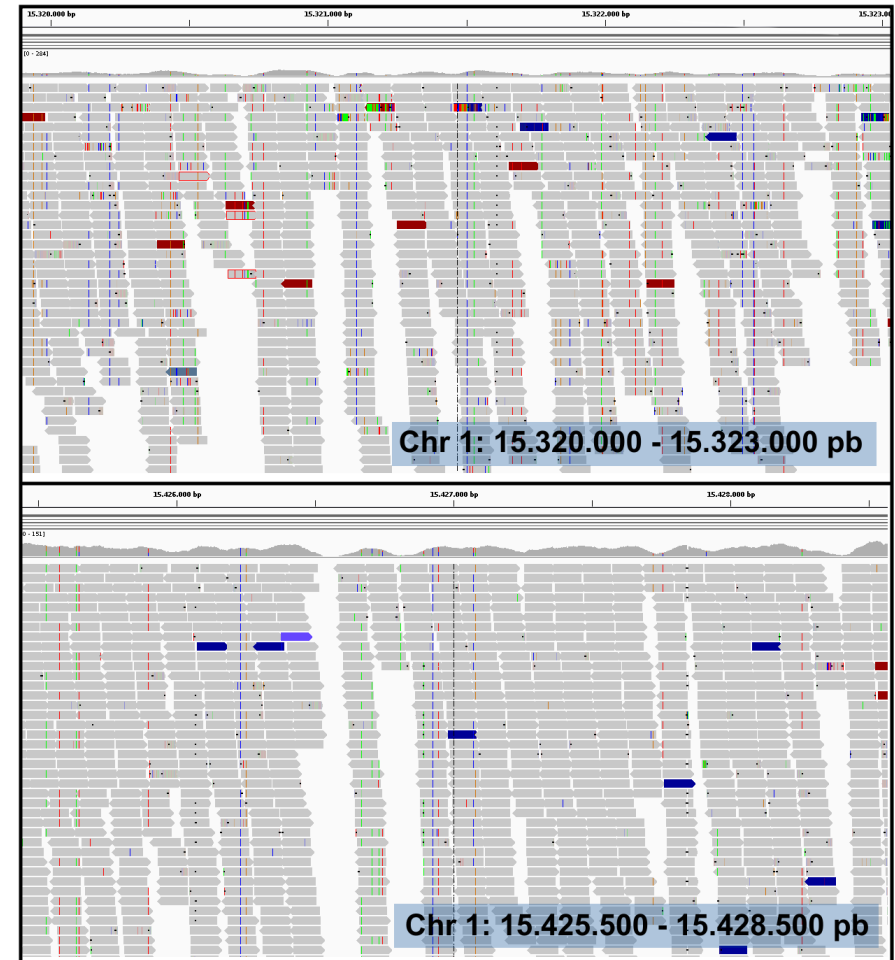


Figura 18. Evaluación de la deleción de 104.916 pb en el cromosoma 1 en la variedad Doblefina. Se muestran en la figura las posiciones 15.320.000 – 15.323.000 pb (superior) y 15.425.500 – 15.428.500 pb (inferior) del cromosoma 1. En estas regiones no se encontraron lecturas emparejadas que pudieran evidenciar la deleción.

Figura 19. Evaluación de la deleción de 104.916 pb en el cromosoma 1 en la variedad Cadenera. Se muestran en la figura las posiciones 15.320.000 – 15.323.000 pb (superior) y 15.425.500 – 15.428.500 pb (inferior) del cromosoma 1. En estas regiones no se encontraron lecturas emparejadas que pudieran evidenciar la deleción.

Tabla 2. Evaluación *in silico* mediante visualización de los alineamientos de secuenciación de las variedades estudiadas de naranjo dulce la existencia (Presente ✓) o no (Ausente ✗) de la delección de 104.916 pb en el cromosoma 1.

Variedad de naranjo dulce	Presencia / ausencia de delección 104.916 pb cromosoma 1
Blanca comuna	Ausente ✗
Washington Navel	Presente ✓
Barnfield	Presente ✓
Navelate	Presente ✓
Ricalate	Presente ✓
Lane Late	Presente ✓
Cara Cara	Presente ✓
Navelina	Presente ✓
Powell	Presente ✓
Fukumoto	Presente ✓
Chislett	Presente ✓
Salustiana	Ausente ✗
Cadenera	Ausente ✗
Tarocco Rosso	Ausente ✗
Doblefina	Ausente ✗
Sanguinelli	Ausente ✗
Valencia Late Frost	Ausente ✗
Valencia Late Delta	Ausente ✗
Midknight	Ausente ✗
Barberina	Ausente ✗

8. 2. Comprobación de la delección mediante BLASTN

Las delecciones también pudieron confirmarse mediante BLASTN (Tabla 3), utilizando como consulta las bases desapareadas de cada uno de los flancos de la delección (upstream y downstream) frente al genoma de referencia *Citrus clementina* Hort. ex Tan.

En cada una de las delecciones detectadas, al usar como consulta las bases desapareadas del flanco upstream se obtuvo como resultado posiciones asociadas con el flanco downstream, y viceversa, lo que mostró que en el genoma de referencia estos flancos están separados por la región deleccionada y la delección se da en las variedades en las que aparecen lecturas con bases desapareadas en los flancos upstream y downstream de la delección.

8. 3. Obtención de secuencia objetivo

Una vez detectadas las variaciones genómicas en variedades objetivo *in silico* se buscó confirmar dichas secuencias *in vitro* mediante PCR en tiempo real. Para ello, se fue necesario “reconstruir” el genoma de estas variedades a partir del genoma de referencia.

Tabla 3. Resultados del BLASTN utilizando como consulta las bases desapareadas de los flancos upstream y downstream de las dos deleciones detectadas (deleción de 1.937 pb y deleción de 104.916 pb). Al consultar el flanco upstream de cada una de las deleciones se obtiene como salida un alineamiento en las posiciones correspondientes al flanco downstream de la deleción y viceversa, evidenciando que ambos flancos están unidos, lo que corrobora las deleciones.

Leyenda: cromosoma (cromosoma en el que se produce el alineamiento); % identidad (grado de coincidencia entre dos secuencias); Long. Alin. (longitud del alineamiento); mismatch. (bases discordantes en el alineamiento); gaps (huecos en el alineamiento), Com. consul (comienzo del alineamiento en la secuencia consultada), Fin. consul. (final del alineamiento en la secuencia consultada), Com. seq. (comienzo del alineamiento en el genoma de referencia), Fin. seq. (final del alineamiento en el genoma de referencia), E value (número esperado de puntuaciones iguales o superiores a las de la secuencia utilizada como consulta atribuibles al azar), bit score (valores de puntuación normalizados con respecto al tamaño de la base de datos y el sistema de puntuación).

Deleción	Flanco	Cromosoma	% identidad	Long. Alin.	Mismatch.	Gaps	Com. consul.	Fin. consul.	Com. seq.	Fin. seq.	E value	Bit score
Deleción de 1.937 pb	Upstream	scaffold_1 Cclementina	100	49	0	0	1	49	25590609	25590657	7,44e-18	89,7
	Downstr.	scaffold_1 Cclementina	100	37	0	0	1	37	25588625	25588661	1,22e-11	68,0
Deleción de 104.916 pb	Upstream	scaffold_1 Cclementina	97,222	36	1	0	1	36	15426686	15426721	1,80e-09	60,8
	Downstr.	scaffold_1 Cclementina	97,917	48	1	0	1	48	15321722	15321769	1,06e-15	824

En primer lugar, se buscaron los límites (o flancos) de la variación encontrada previamente en cierta variedad mediante el fichero BAM (Figura 20). En cada uno de los flancos de la deleción se encontraron lecturas que alineaban parcialmente con el genoma de referencia (Figura 20A y 20B): estas lecturas tenían una parte con la misma secuencia del genoma de referencia y otra parte cuya secuencia no coincidía con el genoma de referencia en esa posición, ya que se trataba de la secuencia del otro flanco de la deleción. En la Figura 20A (flanco upstream de la deleción) se obtuvieron lecturas que alinean en parte con la secuencia de referencia (aparecen en color gris, secuencia recuadrada de naranja en el genoma de referencia) y en parte tienen bases discordantes (simbolizadas en colores sobre la misma lectura, recuadradas de color azul). Esta secuencia coloreada corresponde a la secuencia del flanco downstream de la deleción (Figura 20B, secuencia recuadrada de azul). Del mismo modo, en la Figura 20B (flanco downstream de la deleción) aparecían lecturas que alinean de manera parcial; en la zona que alinean correctamente se vio la lectura de color gris (secuencia recuadrada de azul sobre el genoma de referencia), y en la zona donde no alinean con el genoma de referencia se vieron bases discordantes (coloreadas sobre la lectura y recuadradas de color naranja). Esta secuencia coloreada corresponde a la secuencia del flanco upstream de la deleción (Figura 20A, secuencia recuadrada de naranja).

Las secuencias del genoma de referencia upstream y downstream a la deleción (las que la flanquean) fueron anotadas, obteniéndose una secuencia teórica del genoma donde se ha producido esta deleción (Figura 20C). La utilización de esta secuencia como ADN molde junto con oligonucleótidos que hibridaban en ambos flancos permitió confirmar el marcador molecular *in vitro* mediante PCR en tiempo real.

La ADN polimerasa utilizada durante la PCR en tiempo real poseía una procesividad que le permitió sintetizar fragmentos de un tamaño máximo aproximado entre 900 y 1000 pb. Esto determinó que solo los oligonucleótidos que hibriden sobre el ADN molde en el que se haya dado la deleción, en cada uno de los flancos de la deleción, en distintas hebras del ADN y a una distancia adecuada generaron producto de PCR. Si no se hubiera dado esta deleción los oligonucleótidos hubieran hibridado a una distancia muy lejana, y debido a la procesividad de la ADN polimerasa no se podría haber dado amplificación.

Para realizar esta confirmación fue necesario obtener una secuencia más larga que incluyera la secuencia reconstruida previamente. Esta secuencia permitió la hibridación de los oligonucleótidos y en definitiva, la reacción de amplificación. Conocida la secuencia que flanquea la deleción, esta se buscó en el genoma de referencia mediante un programa de alineamiento (Figura 21). Así, se seleccionó 1 kb upstream a la deleción y 1 kb downstream a la deleción, obteniendo una secuencia de 2 kb que debería encontrarse en variedades en las que se ha producido la deleción.

Esta secuencia fue de nuevo revisada en el software IGV, en busca de posibles variaciones genómicas (INDELS o SNPs) que pudieran darse sobre el genoma evaluado, afectando a la hibridación de los oligonucleótidos o a la subsiguiente secuenciación. La secuencia se utilizó para diseñar oligonucleótidos que participarán en la PCR en tiempo real, lo que permitió finalmente evidenciar la variación detectada *in silico*. La estrategia expuesta

previamente fue utilizada para cada una de las deleciones detectadas.

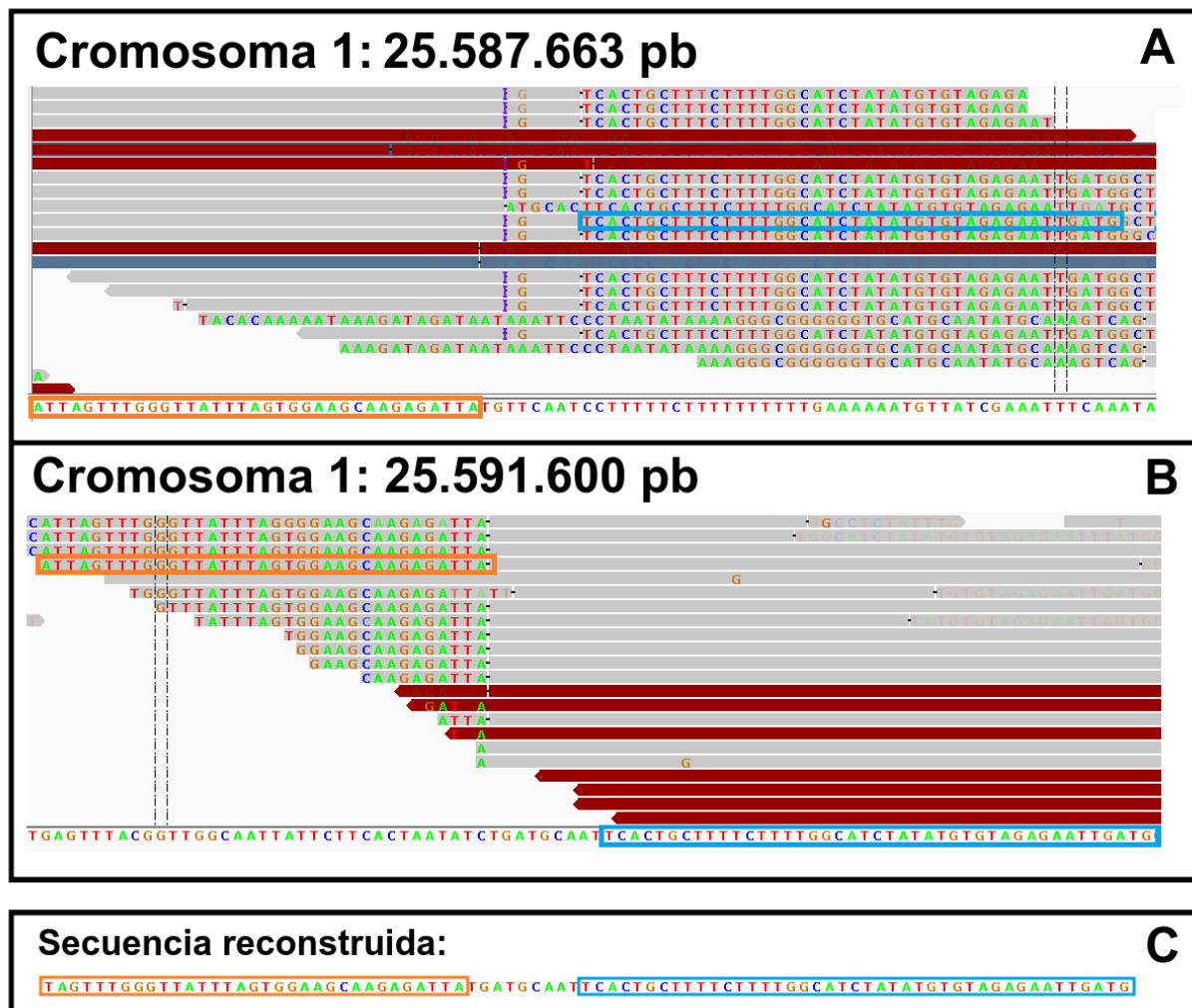


Figura 20. Ejemplo de reconstrucción de la secuencia del genoma en el que se produce la deleción de 1.937 pb. La posición 25.587.663 del cromosoma 1 (cuadro A) corresponde al flanco upstream de la deleción. La posición 25.591.600 del cromosoma 1 (cuadro B) corresponde al flanco downstream de la deleción. En el cuadro C se representa la secuencia reconstruida teórica de una variedad en la que se da esta deleción. Se recuadran de color naranja las secuencias upstream de la deleción y de color azul las secuencias downstream de la deleción. Las secuencias recuadradas del mismo color son coincidentes en los cuadros A, B y C, lo que evidencia la deleción.

En cada uno de los flancos de la deleción (cuadros A y B) se encontraron lecturas que alinean parcialmente con el genoma de referencia. En el cuadro A (flanco upstream de la deleción) las lecturas tienen parte de secuencia que es la misma de la del genoma de referencia (lecturas simbolizadas de color gris, secuencia recuadrada de color naranja) y parte cuya secuencia no concuerda con el genoma de referencia en esta posición (simbolizada con bases coloreadas y recuadrada de color azul): la secuencia discordante se trata de la secuencia del flanco downstream de la deleción (recuadrada de azul en cuadros A, B y C). En el cuadro B (flanco downstream de la deleción) las lecturas que alinean parcialmente tienen parte de la secuencia igual al genoma de referencia (lecturas en color gris, cuya secuencia se recuadra de color azul) y otra parte discordante, que no alinea con la secuencia de referencia (secuencia simbolizada con bases coloreadas y recuadrada de color azul): la secuencia discordante se trata de la secuencia del flanco upstream a la deleción (recuadrada de naranja en cuadros A, B y C). En definitiva, la secuencia con bases desapareadas corresponde al flanco opuesto de la deleción, lo que corroboró la presencia de la deleción en el genoma de la variedad evaluada y permitió reconstruir la secuencia teórica de la variedad donde se produce la deleción.



Figura 21. Ejemplo de obtención de la secuencia del ADN molde en el que se da la delección de 1.937 pb. A partir de la secuencia consenso del genoma de referencia de clementina, se seleccionaron 1000 pb upstream y downstream a los límites de la delección mediante el software BioEdit Sequence Alignment Editor.

A continuación se exponen las secuencias teóricas obtenidas del genoma de referencia de clementina donde se da la delección. Se muestran en color verde y en azul las secuencias upstream y downstream de la delección, respectivamente. Se simbolizan las variaciones respecto al genoma de referencia: con letra cursiva y en negrita las inserciones; y con letra subrayada y en negrita las sustituciones de bases.

- Delección 1.937 pb en el cromosoma 1.

TCAGCCACATTAATCCCTGGCGTCAACAACAGATTTGGAGAAGCGACCTTGCCTGCGAGCACTCAGTACTCCAGCG
TTTTGAGGTGGAGGTGCGGTGCAGAAGTCTTTGAGCAATGGAGCAGCAAAATGGGTTGAATGGGAACATCTTAGGC
CAACAACCTGTCGGTAATTTGATTGATGGCAGTGCACAAACAGGACCAATCTGGCCAATTTGGCCAATTTGGCCA
AGCAACAGGGGACCATAATCTCCGTGAGACAACCTTGGTACGCTTGTAGAGATGACCAGCACTTTTCAATATCT
TGAGTAGGCTGCTCTTGTGACAGTCTGAGGGGAACAGCATTTGTCGCGAACACAAACATCAGCACCATAAGAAGG
GTATTTGATGATTTTGCCATCGGATCAGCAAGATCACTGGAAGTCTTTGCACTGGTTTTGTTGATACAAGTTTTCC
TGTGATTTAAAAGGATACTTGTGGTCTGTGGAAGGATTAGGTAGGAAGAGAACACACTCTAAATACTGATTACG
GAAACGCTTCTGTAAATTTGACGAAACAGTTTCAATATGTAGCATTTACCATGAGTTATGTGCTAAGTGAGTTTTT

ATGGCATAATTGTAACGGAAAAATAATAGCTAGGATTCTTTTTTTTTTTTTTTAATTTTTAATTCTGTGTGAAAA
GCCTGCTATCGATATGATAGAACATAGATGGTGGTCTAATCTCCATTCAAAATTATAACTTATTTAAATTTAAATTT
TAAAATTTTCATTTTTTTTTATAAATAACATATTGTAAGAAAGTATGCACCGACATATACTATTTACGACATATCAT
TTAAGACTTTATTTACTGAATAACATGTCATACATTTTACAGTGTGATAAATTTTTTAAAGATTTTTTTTTTTTTTAC
GTTTTTCGTATTTTTTATTCAAATATTACCTTTGCATACATGTTCCATTAGTATACATATGGGCCCATTAGTTTTG
GGTTATTTAGTGAAGCAAGAGATTATGATGCAATTCCTGCTTTCTTTTGGCATCTATATGTGTAGAGAATTGA
TGGCTTCCTTGCTGGTTTTCTCGCTGTACTTAGCATTACTAGGGGAAACTTTTCACCACAATCACAGAAGCTGTC
ATTTTCTGTTACTAGAGTGGATGGGGATTGCTATTTACTCTATTATGCATCCATAGATATTGGATTTGGAGGGGG
GAGAGCTAGAGTAACCCACGGCTGTTGGATTTGGAGGGAGGAGAGAGAGCCGTGTGTTGTGAGGAAGAACAGA
GAGAGCGTGAGGAAGAACGTGAGGGGAGAGAGCAATGTGCTGTAGTTGAGAGCGTGTGTTGCAATTGTGCGATTTA
CAGAGAGAGCGTGAGAAACATTTCCAAAGTTGTTTCCCTTTACTGTGATTTGTCTACAAATTTTTTCGAGAGCATT
TCACCAGCGAAACAAGGATGATGATGGCCGGAGTGAGGGGGAGAAATGACCAGCTTTTCAATTCCTTTTTATTTTA
TTTTATTATTTTTTGAAGAATTATCAGAAATGACCATCGTTTTTTTTTTTTTAAATTTCTGTGGGTCCACGCAAGT
TTAAAAAAGAGCAAAATCTAACCGGTGAAACGGTAATTTAAAAATTACGTTTCTGTCGTTCTTTTATTTTTCATTGA
GTTAGTTTATCATTTTCTTTTCTTTTTTATGAAAAAAGGAGACCTACACTCTACACTCTTTAAGGTTTAGAGAAA
TGGTCTTTAGGTTTATGGAAATGAGGAAATAGTCGTAAGACGCTCAAAAAGATATCAATTTTTTTTTTAAATTACCA
TATCTTTTAAATTTTATAATTTTTTAAAAAAATAATTTTATAATTTTATAATAAATAAAAAATATGATATTTAAATAA
ATTTAATAATATATATATATTTTTAAATAATTTAATATTTAAATTTTTTAAATTTTTTAAATCTATTAGATTTA
TTTTATTATATTTAAATAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA

- Delección de 104.916 pb en el cromosoma 1.

TTGATTATCTTGAAATAGAGCAATTATAGCTGAATAAGACAATGGCGTAATAAGATGGAGTCACATTAATCACGAA
AATTTAACTGCAATCATATTCGGTAATAAAAAAAGGAGAACAATATCATCTCATAAAGTAAAAAACATAGAACAA
ACATCATCACATAAAGTAAAAAAAACATACTTATAAAACGATAGAACCAAGATCATTACATGATAATGTTATGTTA
TCTTAAAGAAATGGGGATCCTGAGATCACCTTGTCTATTAGAGACTTAGAGAATTGAAAAGTGCAACTAATCAG
TTTGGGGTACTCTAGGTACCCCGCATGATAAAGTACCCACATAAATAGATCAAATATCATAAACGCATAAGCATT
AAAGAAAAATAAATAAAGTAAATAGAAAGAGGTAAAGTGCTTTCTTTTATTAGTAGTACTTCCTCGATATGGCAG
CATTTACAAAGTGTTTTATAATTGTTCTTTTCATTTAGAGTAGCGTGTGTATTCATTCCTTTTCTTCCCTCGGCA
CCCGAGGATCCAGGGGTGCTCTAACTCTTCGATTAATGCTGAATTAGGACTTTGGGGTACCTCGGCTACATCAT
GCATAGCTGCATCTACCTTTCATTGGCCCTACTCTTTTTTGTTTTAGAGGACTAGTTGTCTATAACACTTTCCACC
GTGCAGCTCTTCATTGTATTGCAGCGCCTAATTAGTAAGGCAATAAATGTCAATTGCATGCTGCTGGGGTACTCGG
GGTACTCCAAAGAATATTCAGCCCTCTCCACGTGGGGTTGGACGAGGTGGGCTGTGTACCCGAGGTGCTCAAGAG
GGTACCGAAGGTACCTCGCTCTCTTGGGTTTCTCCTTGTCTCAAACTTTGACTTTGTCTTCTCCCTCCGCTTT
ATGTTACACTTCTCATCCATTTATATTTGTTGTTAGGGCTCGAGAACGAGCTTCCCCATATGAAAGAGGGCGGTGT
TTTTTAAATGACTGGCCATAACTTCCCCCTATAGCACATGGTTGTAGAAACATTTTTAAAGATGGGATGTATTCGCTA
ATCAAGAGGTGGCAGTTGAACTTTGAGTTGATTTTTTGCATGCATGTAGATATTGGTAGGACAAGGCCATAAGA
TAAAACATGAGTTAAATCCCAGTGCCAAGTGAGGTTTTATGAGTTCATTACAAGACCACATGAATAATTTTAGTA
TAATTAGTTAAAAAAAAGAACTTGTGGGTTGATGCATAACTCTTTGCAGGCACAGAAATTGGTGAAATGAGAC
TGCAAGATCAAAAAATAGTTGACTCTAAGGCATAGATGAGGTTTTATGCCATTTAGGCATATAATCGAAGGGTA
TAGCACTTGTCTAGGTAGTTTAAATGTGTATTCGACACTGAAGGTGCTGGGCATTGAATTTCTGTGTTAATTTT
TTGCATGCCTTTGACATTAGTGAGAAGACATCGCCAAAATCAAAACAAGAGCCGATTATGATGCCTTGGTGAAGTTT
TACGAGGTCATTACAAAACAACATGGATAATTTTTCTCATAATTTTCAAAATCAATTCCCAAAAATTTATGGAGAAC
CTATTACATTGATATAAACTCTTTAGAAGGCATGTATCTAGATATAAACCTATCACAACATTAATGAAAGTCG
ATTTCAAGGCTCAAGTGACGTTTTACGCCATTCAAGCCTAGAACCAGGCTATAGCACATGGTTTTAGATTTTTTCT
TATGCGCATATTCGCTATTGAAGAGGTTAAGTGTGAAATTTAAAGCTAATGGCAAGATGAGAACACAATATGAAA
TAAGAGTCAATTCCTTAGCCTAAGTGAGGTTTTACAAGGTTGTTCTGAGATGTTTGGAGAGCACAAGAATGATTT
TCACAAATCTTCAAAACATTCATAACAAAAATAAAGAAAACTTATCGAATTAATGCAAACTCTTTGGCCGATGT
AGAAATTGGTAAGCAAGGCCATCAATGAATAAGGCATAAAATTTTTGTGTTATG

Estas secuencias fueron utilizadas para diseñar oligonucleótidos mediante la herramienta Primer3 (Tabla 4 – 5). Para cada deleción se seleccionaron tres parejas de oligonucleótidos.

Tabla 4. Parejas de oligonucleótidos propuestas para la detección de la deleción de 1.937 pb.

Descriptor de la pareja	Oligonucleótido forward (5' → 3')	Oligonucleótido reverse (5' → 3')
A.1	5'-ACTTGTGGTTCTGTGGAAGGA - 3'	5'-TCACAACACACGGCTCTCTC-3'
B.1	5'-TCGCGAACACAACAATCAGC-3'	5'-TGACAGCTTCTGTGATTGTGG-3'
C.1	5'-TGTGGTTCTGTGGAAGGATTAGG-3'	5'-CAACAGCCGTGGGTTACTCT-3'

Tabla 5. Parejas de oligonucleótidos propuestas para la detección de la deleción de 104.916 pb.

Descriptor de la pareja	Oligonucleótido forward (5' → 3')	Oligonucleótido reverse (5' → 3')
A.2	5'-TGACTTTGTCTTCCTCCCGC- 3'	5'-CCTCACTTGGCACTGGGATT-3'
B.2	5'-TGTTGTAGGGCTCGAGAACG-3'	5'-AAACCTCACTTGGCACTGGG-3'
C.2	5'-CCCATATGAAAGAGGGCGGT-3'	5'-TGCCAACCTCTTGATTAGCGA-3'

8. 4. Cuantificación de la cantidad y calidad del ADN extraído

Se midió la absorbancia a una longitud de onda de 280, 260 y 230 nm (A_{280} , A_{260} y A_{230} , respectivamente) de las extracciones de ADN de material vegetal y se evaluó su cantidad y calidad (Tabla 6).

Tabla 6. Resultados de la cuantificación de la cantidad y calidad del ADN extraído.

Variedad	Concentración (ng/μL)	A_{260} / A_{280}	A_{260} / A_{230}
Washington Navel	365,3	1,96	1,62
Blanca Comuna	2807,6	2,01	1,89
Salustiana	1475,9	2,06	1,81
Valencia Late Frost	46,4	1,65	2,91

8. 5. Cribado de oligonucleótidos

Para confirmar la ausencia de artefactos durante la PCR en tiempo real se realizó una reacción preliminar utilizando cada una de las parejas de oligonucleótidos (Figura 22 – 27). El análisis de las curvas de melting de dichas reacciones permitió seleccionar las combinaciones de oligonucleótidos que no produjeron amplificaciones inespecíficas (A.1, B.2 y C.2), mientras que las demás combinaciones fueron descartadas.

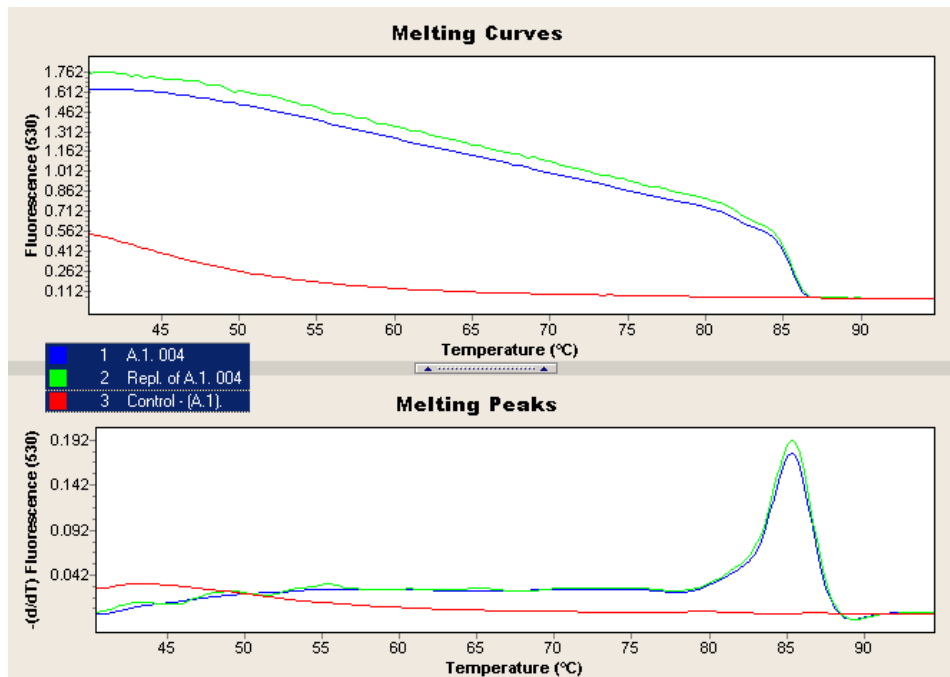


Figura 22. Curva de melting de PCR utilizando la pareja de primers A.1 y el genoma de la variedad Washington Navel como molde. En la parte superior se representa la intensidad de fluorescencia ($\lambda = 530$ nm) frente a la temperatura. En la parte inferior se representa la primera derivada negativa de la intensidad de fluorescencia ($\lambda = 530$ nm) frente a la temperatura.

En color verde y azul se representan reacciones de PCR en las que se obtuvo un mismo producto (verde es réplica del azul) y en rojo el control negativo. En la representación superior se ve la disminución de la intensidad de fluorescencia ($\lambda = 530$ nm) a medida se aumenta la temperatura. Cuando la temperatura llega a 85 °C la intensidad de fluorescencia baja bruscamente hasta 0 u.a., lo que coincide con el pico en la representación derivativa: la temperatura de melting (T_m) de este producto de PCR es de 85 °C; se trata de una amplificación específica.

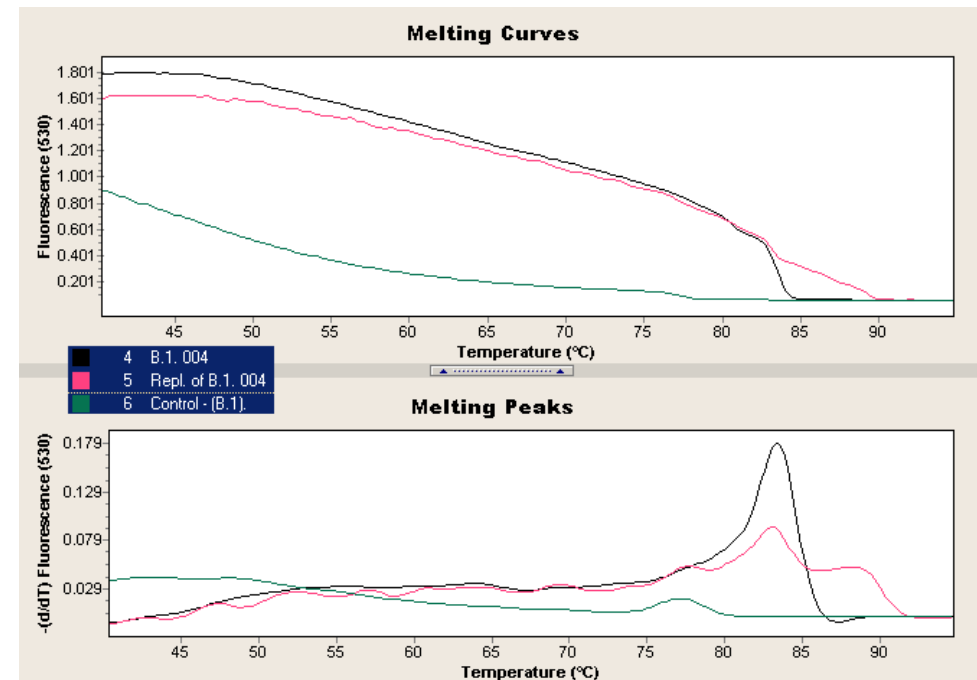


Figura 23. Curva de melting de PCR utilizando la pareja de primers B.1 y el genoma de la variedad Washington Navel como molde. En la parte superior se representa la intensidad de fluorescencia ($\lambda = 530$ nm) frente a la temperatura. En la parte inferior se representa la primera derivada negativa de la intensidad de fluorescencia ($\lambda = 530$ nm) frente a la temperatura.

En color negro y rosa se representan reacciones de PCR en las que se obtuvo un mismo producto (rosa es réplica del negro) y en aguamarina el control negativo. En la representación superior se ve la disminución de la intensidad de fluorescencia ($\lambda = 530$ nm) a medida se aumenta la temperatura. Cuando la temperatura llega a 83 °C la intensidad de fluorescencia baja bruscamente hasta 0 u.a. en el caso de la reacción en negro, mientras que en la reacción rosa existe una estabilización hasta llegar a 89 °C, donde la temperatura baja bruscamente. Debido a esta amplificación irregular esta pareja de oligonucleótidos fue descartada.

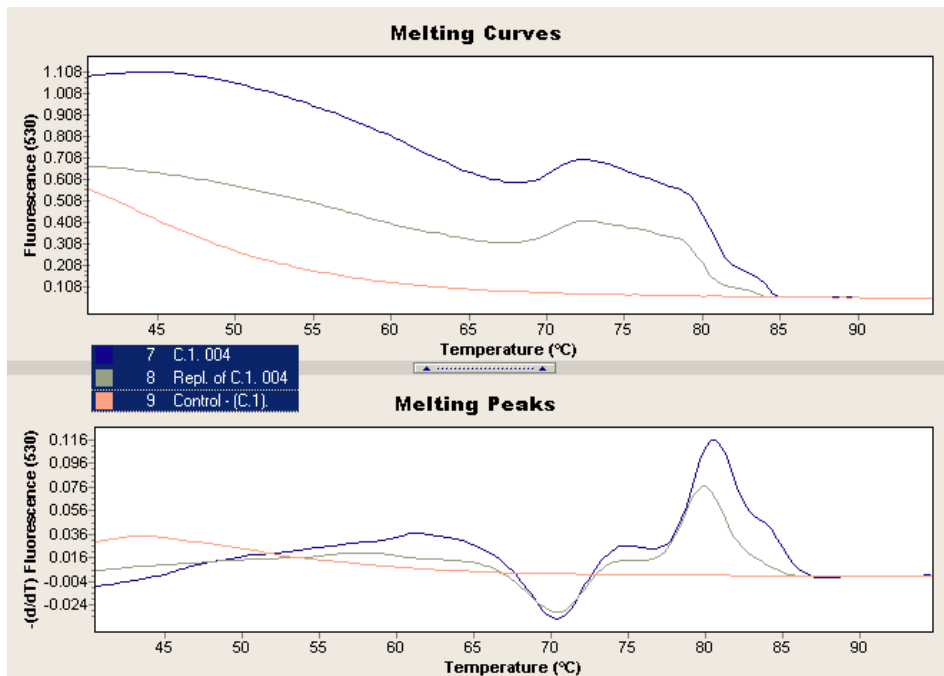


Figura 24. Curva de melting de PCR utilizando la pareja de primers C.1 y el genoma de la variedad Washington Navel como molde. En la parte superior se representa la intensidad de fluorescencia ($\lambda = 530 \text{ nm}$) frente a la temperatura. En la parte inferior se representa la primera derivada negativa de la intensidad de fluorescencia ($\lambda = 530 \text{ nm}$) frente a la temperatura.

En color azul y gris se representan reacciones de PCR en las que se obtuvo un mismo producto (gris es réplica del azul) y en rosa el control negativo. En la representación superior se ve inestabilidad en la disminución de la intensidad de fluorescencia ($\lambda = 530 \text{ nm}$) a medida se aumenta la temperatura, lo que se traduce en subidas y bajadas inestables en la representación derivativa. Debido a esta amplificación irregular esta pareja de oligonucleótidos fue descartada.

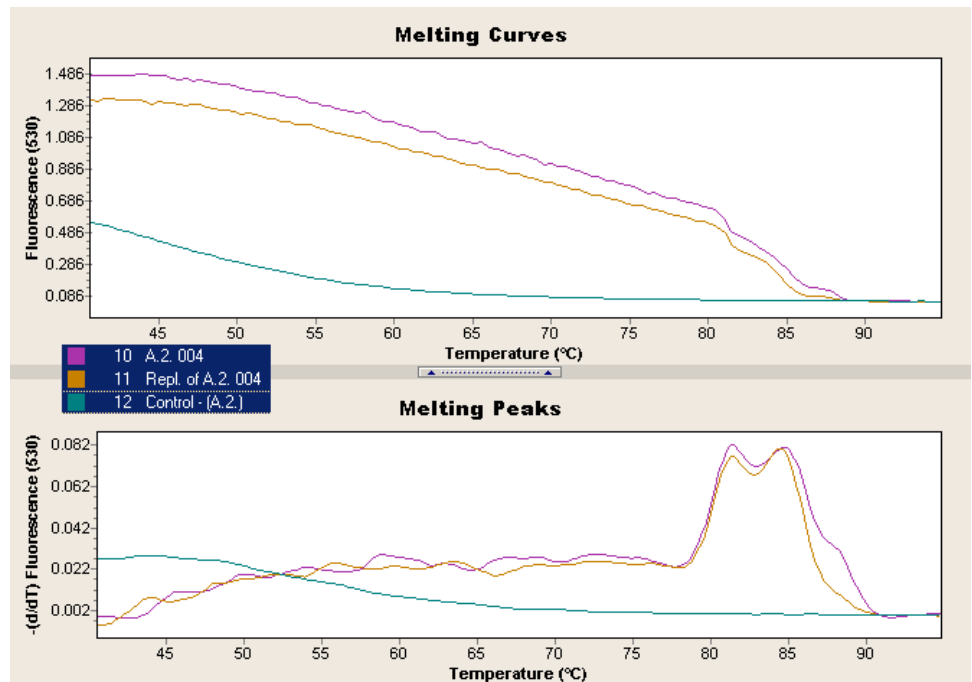


Figura 25. Curva de melting de PCR utilizando la pareja de primers A.2 y el genoma de la variedad Washington Navel como molde. En la parte superior se representa la intensidad de fluorescencia ($\lambda = 530 \text{ nm}$) frente a la temperatura. En la parte inferior se representa la primera derivada negativa de la intensidad de fluorescencia ($\lambda = 530 \text{ nm}$) frente a la temperatura.

Se representan en violeta y caqui dos reacciones que dieron un mismo producto de PCR (caqui es réplica del violeta) y en verde azulado el control negativo. En estas representaciones, a medida la temperatura aumenta se ve inestabilidad en la bajada de la intensidad de fluorescencia. Si nos fijamos en la representación derivativa se muestran claramente dos picos a 82 °C y 85 °C, además de una “joroba” en la curva púrpura alrededor de 87 °C. La aparición de dos picos y esta joroba informa de la síntesis de productos de amplificación independientes: se da una amplificación no específica, con lo que se determinó que estos oligonucleótidos no eran adecuados para llevar a cabo una reacción de amplificación y con ello confirmar el marcador molecular.

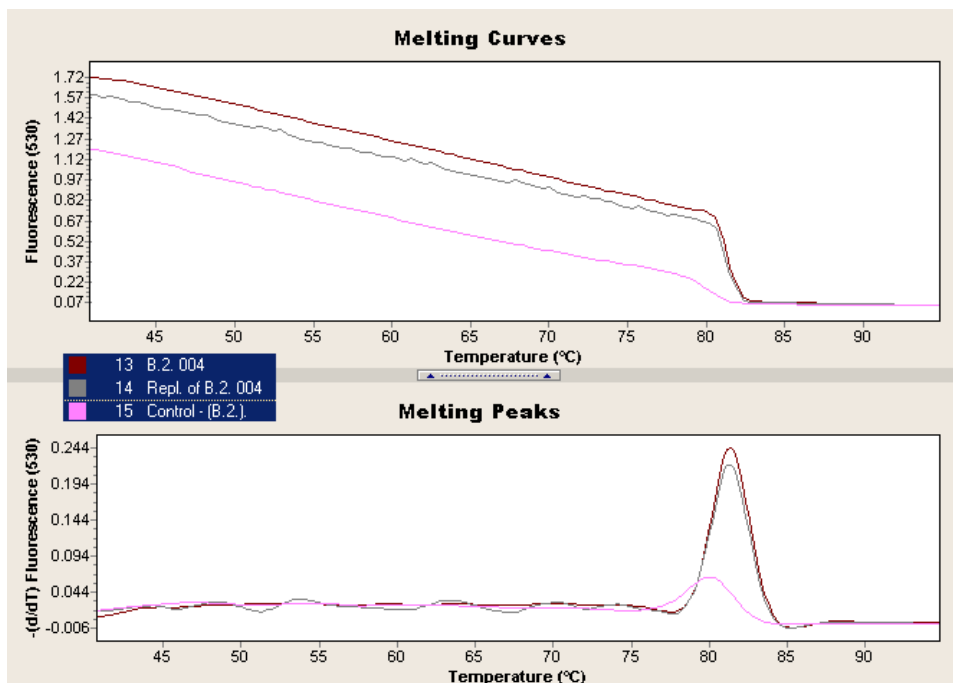


Figura 26. Curva de melting de PCR utilizando la pareja de primers B.2 y el genoma de la variedad Washington Navel como molde. En la parte superior se representa la intensidad de fluorescencia ($\lambda = 530 \text{ nm}$) frente a la temperatura. En la parte inferior se representa la primera derivada negativa de la intensidad de fluorescencia ($\lambda = 530 \text{ nm}$) frente a la temperatura.

En color granate y gris se representan reacciones de PCR en las que se obtuvo un mismo producto (gris es réplica del granate) y en fucsia el control negativo. En la representación superior se ve la disminución de la intensidad de fluorescencia ($\lambda = 530 \text{ nm}$) a medida se aumenta la temperatura. Cuando la temperatura llega a 82°C la intensidad de fluorescencia baja bruscamente hasta 0 u.a. , lo que coincide con el pico en la representación derivativa: la temperatura de melting (T_m) de este producto de PCR es de 82° ; se trata de una amplificación específica.

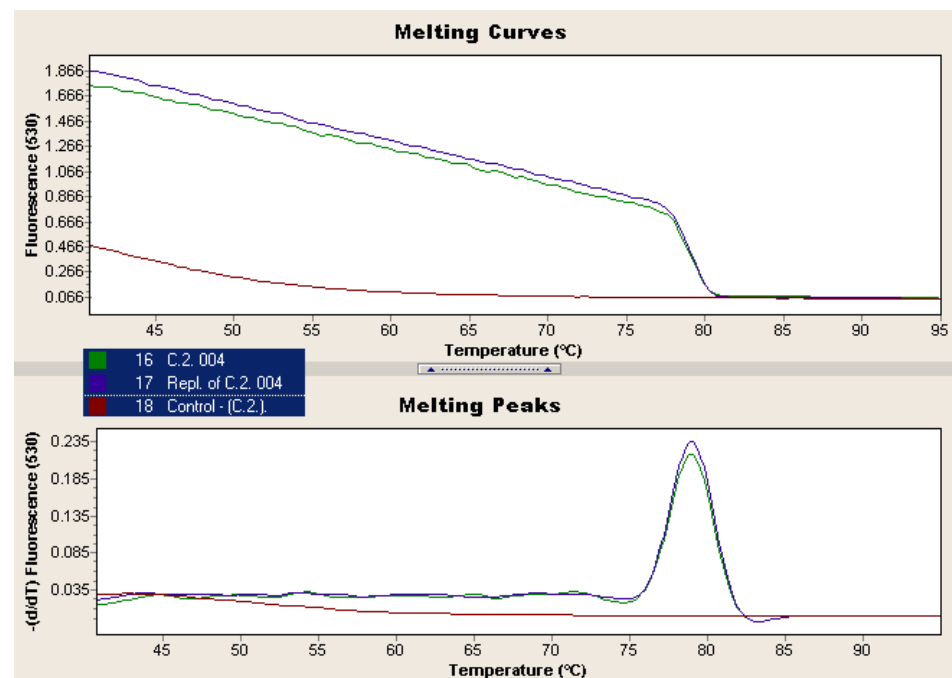


Figura 27. Curva de melting de PCR utilizando la pareja de primers C.2 y el genoma de la variedad Washington Navel como molde. En la parte superior se representa la intensidad de fluorescencia ($\lambda = 530 \text{ nm}$) frente a la temperatura. En la parte inferior se representa la primera derivada negativa de la intensidad de fluorescencia ($\lambda = 530 \text{ nm}$) frente a la temperatura.

En color verde y azul se representan reacciones de PCR en las que se obtuvo un mismo producto (azul es réplica del verde) y en granate el control negativo. En la representación superior se ve inestabilidad en la disminución de la intensidad de fluorescencia ($\lambda = 530 \text{ nm}$) a medida se aumenta la temperatura, lo que se traduce en subidas y bajadas inestables en la representación derivativa. Debido a esta amplificación irregular esta pareja de oligonucleótidos fue descartada.

8. 6. Confirmación de marcadores mediante PCR en tiempo real

En la reacción de PCR en tiempo real se obtuvo un producto de amplificación utilizando como DNA molde el DNA genómico de la variedad Washington Navel y la pareja de oligonucleótidos B.2. La aparición de un único pico de melting junto con el tamaño de los oligonucleótidos empleados son indicadores de la especificidad de la reacción (Figura 28).

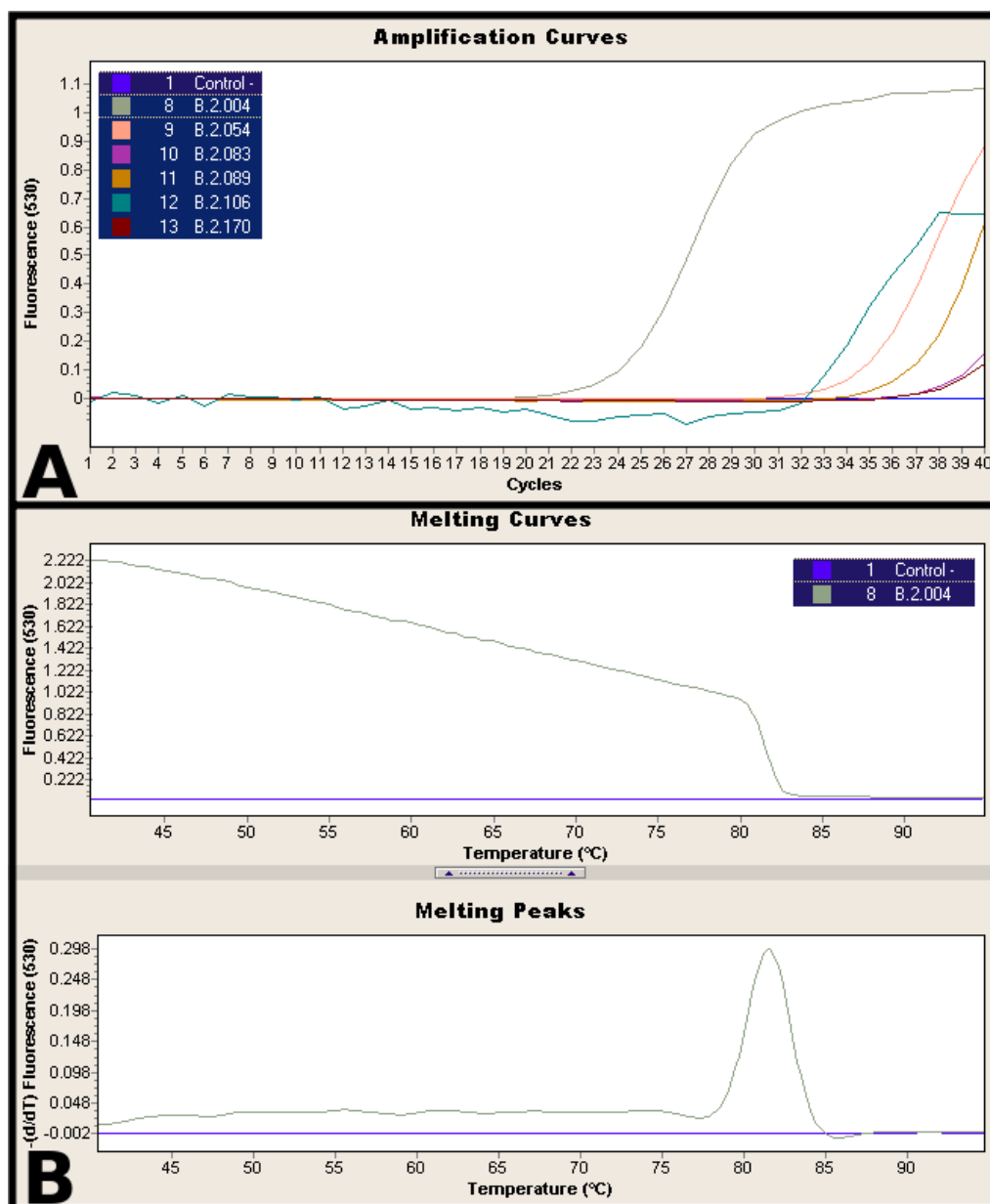


Figura 28. Reacción de PCR para la confirmación de la delección 104.916 pb en el cromosoma 1 mediante la utilización de la pareja de oligonucleótidos B.2. A) Representación de la intensidad de fluorescencia ($\lambda = 530 \text{ nm}$) durante la reacción de PCR. Se muestra amplificación de la variedad Washington Navel (gris) y ausencia de amplificación en las demás variedades. B) Curva de melting correspondiente a la reacción del genoma de Washington Navel como ADN molde (gris) y control negativo (azul), sin ADN molde. En la parte superior se representa la intensidad de fluorescencia frente a la temperatura y en la parte inferior se representa la primera derivada negativa de la intensidad de fluorescencia ($\lambda = 530 \text{ nm}$) frente a la temperatura. Leyenda: control negativo (azul oscuro), Washington Navel (gris), Doblefina (rosa), Salustiana (lila), Blanca Comuna (caqui), Cadenera (aguamarina), Valencia Late Frost (marrón).

8. 7. Secuenciación del marcador molecular evaluado

La comparación de la secuencia teórica directa reconstruida *in silico* de las variedades en las que se da la delección de 104.916 pb en el cromosoma 1 y la secuencia directa obtenida mediante la secuenciación del producto de amplificación del DNA genómico de la variedad Washington Navel con la pareja de oligonucleótidos B.2 (Figura 28, en gris) permitió corroborar que en esta variedad se produce dicha delección, lo que validó el marcador molecular (Figura 29).

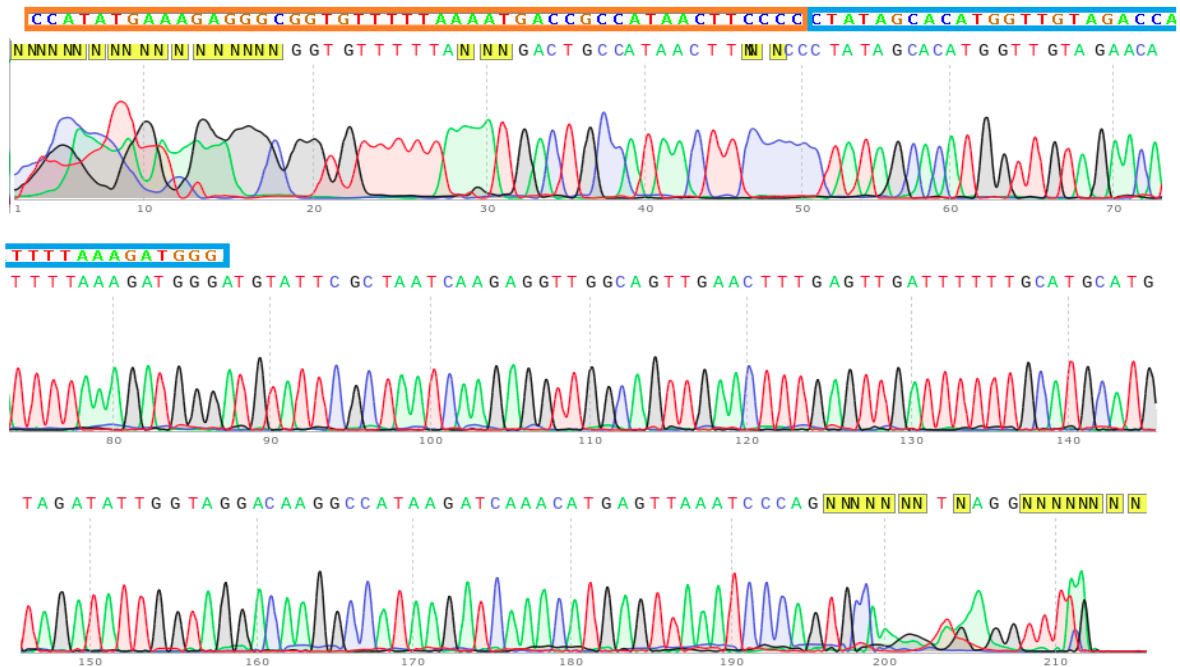


Figura 29. Electrofenograma obtenido tras la secuenciación Sanger del producto PCR en tiempo real utilizando la pareja de oligonucleótidos B.2. y DNA genómico de la variedad Washington Navel. Se muestra recuadrado en color naranja (flanco upstream) y azul (flanco downstream) la secuencia teórica reconstruida del genoma en el que se produce la delección. La posición 50 del electrofenograma corresponde con la rotura de la delección.

9. Discusión

Mediante la metodología de identificación de grupos varietales de naranjo dulce mediante marcadores moleculares que se expone en este trabajo se pretende auxiliar a la identificación varietal tradicional basada en la evaluación de caracteres fenotípicos que presentan los distintos grupos varietales del naranjo dulce. El apoyo de los marcadores moleculares ha sido ampliamente utilizado en la identificación varietal en las últimas décadas, sobre todo en frutales.

A lo largo de la historia del cultivo del naranjo dulce ha predominado la selección de individuos con fenotipos interesantes para el consumidor seguida de la propagación de estos individuos de manera asexual a través de esquejes, injertos, acodos y fenómenos apomícticos, generando una uniformidad en la constitución genética de la generación obtenida mediante estos métodos³¹. Las diferencias entre los distintos grupos varietales de

naranja dulce radican en mutaciones espontáneas. En vista de lo anterior resulta patente la dificultad de obtener marcadores moleculares en naranja dulce.

En los últimos veinte años se han dado enfoques que, mediante diferentes modalidades de marcadores moleculares, han permitido distinguir en mayor o menor medida diferentes cultivares de naranja dulce.

Orford y colaboradores no fueron capaces de encontrar polimorfismos en diferentes grupos varietales de naranja dulce aplicando minisatélites, aunque sí en otras especies del género *Citrus*⁵⁰.

Posteriormente, gracias a la aplicación de los ISSRs (*Inter-Simple Sequence Repeats*) se encontraron polimorfismos que permitieron diferenciar inequívocamente diferentes variedades de naranjas navel y naranjas sanguinas⁵¹.

El estudio de marcadores RAPDs en diferentes variedades de naranja dulce no logró encontrar polimorfismos que discriminaran estas variedades, con lo que se concluyó que los RAPDs no eran aplicables en esta diferenciación varietal y que se debía buscar algún método más sensible para la identificación varietal mediante marcadores moleculares⁵².

Existen varios estudios en la bibliografía en los que utilizando marcadores SSR se han subdividido poblaciones compuestas por distintos grupos varietales de naranja dulce^{53,54}. La técnica RAM (*Random Amplified Microsatellites*, microsatélites amplificados al azar) permitió agrupar las accesiones estudiadas en naranjas blancas, naranjas navel y naranjas sanguinas.

En este trabajo se han desarrollado y validado marcadores basados en la detección *in silico* de variaciones estructurales del genoma que permiten discriminar diferentes individuos generados a partir de mutaciones espontáneas. El desarrollo de este tipo de marcadores moleculares es laborioso, sin embargo, una vez validados su aplicación en la identificación varietal es sencilla, barata y reproducible.

Se han identificado dos marcadores moleculares *in silico* y se ha validado *in vitro* uno de ellos, lo que muestra el potencial de esta metodología para el desarrollo de nuevos marcadores moleculares.

La variación estructural detectada *in silico* pero no corroborada *in vitro* debe ser validada con el fin de poder ser utilizada en la identificación inequívoca de los grupos varietales de naranja dulce.

Los resultados obtenidos pueden ser utilizados para identificar variedades pertenecientes al grupo de naranjas navel. Por otro lado, no ha sido posible encontrar marcadores moleculares que permitan la distinción inequívoca de variedades pertenecientes a los grupos varietales de naranjas blancas y naranjas sanguinas.

10. Conclusiones

- A través de la comparación de genomas secuenciados por el consorcio Citrusseq – Citrusgenn se han detectado variaciones estructurales *in silico* en distintas variedades de naranja dulce. Estas variaciones estructurales son cuantitativas con pérdida de dosis génica, es decir, deleciones.
- Las variaciones estructurales localizadas y acotadas en el genoma de las variedades evaluadas han servido para desarrollar marcadores moleculares que han permitido la identificación del grupo de naranjos dulce tipo navel respecto a los otros grupos de naranja dulce.
- Se ha conseguido validar *in vitro* uno de los marcadores moleculares detectados: la deleción de 104.916 pb en el cromosoma 1. Este marcador está asociado al grupo de naranjas dulce tipo navel: Washington Navel, Barnfield, Navelate, Ricalate, Lane Late, Cara Cara, Navelina, Powell, Fukumoto y Chislett.

En este trabajo se ha descrito un método general para la búsqueda de variaciones estructurales cuantitativas y cualitativas *in silico* que pueden ser utilizadas como marcadores moleculares desarrollado por el Centro de Genómica del IVIA así como el procedimiento de validación de los mismos basado en la extracción de ADN genómico de material vegetal de la variedad problema, seguida de una PCR a tiempo real y secuenciación Sanger del amplicón de PCR obtenido.

La validación del marcador molecular detectado permite asegurar que la identificación de marcadores moleculares *in silico* y su posterior validación *in vitro* supone una metodología robusta que puede ser utilizada en el sector citrícola para la identificación inequívoca rápida y sencilla del grupo de variedades de naranja dulce tipo navel, todo ello utilizando una pequeña cantidad de material vegetal.

Como conclusión final se muestra la Tabla 7 en la que figuran los marcadores moleculares detectados.

Tabla 7. Resumen de la evaluación de las variaciones genómicas sobre las distintas variedades de naranjo dulce. Estas variaciones del genoma pueden ser utilizadas como marcadores moleculares. Se muestran las variedades en las que se encuentra la delección (Presente ✓) y en las que no se encuentra (Ausente ✗).

Variedad de naranjo dulce	Presencia / ausencia de delección 1.937 pb cromosoma 1	Presencia / ausencia de delección 104.916 pb cromosoma 1
Blanca comuna	Ausente ✗	Ausente ✗
Washington Navel	Presente ✓	Presente ✓
Barnfield	Presente ✓	Presente ✓
Navelate	Presente ✓	Presente ✓
Ricalate	Presente ✓	Presente ✓
Lane Late	Presente ✓	Presente ✓
Cara Cara	Presente ✓	Presente ✓
Navelina	Presente ✓	Presente ✓
Powell	Presente ✓	Presente ✓
Fukumoto	Presente ✓	Presente ✓
Chislett	Presente ✓	Presente ✓
Salustiana	Ausente ✗	Ausente ✗
Cadenera	Ausente ✗	Ausente ✗
Tarocco Rosso	Ausente ✗	Ausente ✗
Doblefina	Ausente ✗	Ausente ✗
Sanguinelli	Ausente ✗	Ausente ✗
Valencia Late Frost	Ausente ✗	Ausente ✗
Valencia Late Delta	Ausente ✗	Ausente ✗
Midnight	Ausente ✗	Ausente ✗
Barberina	Ausente ✗	Ausente ✗

11. Bibliografía

1. Aleza, P., Juárez, J., Ollitrault, P. & Navarro, L. Polyembryony in non-apomictic citrus genotypes. *Ann. Bot.* **106**, 533–545 (2010).
2. Frederick G. Gmitter, J. & Hu, X. The Possible Role of Yunnan, China, in the Origin of Contemporary Citrus Species (Rutaceae). *Econ. Bot.* **44**, 267–277 (1990).
3. Tanaka, T. Species problem in Citrus. (1954).
4. Ziegler, L. W. & Wolfe, H. S. *Citrus growing in Florida*. (University of Florida Press, 1975).
5. Scora, R. W. On the History and Origin of Citrus. *Bull. Torrey Bot. Club* **102**, 369–375 (1975).
6. Xie, S., Manchester, S. R., Liu, K., Wang, Y. & Sun, B. Citrus linczangensis sp. n., a leaf fossil of Rutaceae from the late Miocene of Yunnan, China. *Int. J. Plant Sci.* **174**, 1201–1207 (2013).
7. Wu, G. A. *et al.* Genomics of the origin and evolution of *Citrus*. *Nature* **554**, 311–316 (2018).
8. Jacques, F. M. *et al.* Quantitative reconstruction of the Late Miocene monsoon climates of southwest China: a case study of the Lincang flora from Yunnan Province. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **304**, 318–327 (2011).
9. Clift, P. D., Wan, S. & Blusztajn, J. Reconstructing chemical weathering, physical erosion and monsoon intensity since 25 Ma in the northern South China Sea: a review of competing proxies. *Earth-Sci. Rev.* **130**, 86–102 (2014).
10. Gaut, B. S., Díez, C. M. & Morrell, P. L. Genomics and the contrasting dynamics of annual and perennial domestication. *Trends Genet.* **31**, 709–719 (2015).
11. Miller, A. J. & Gross, B. L. From forest to field: perennial fruit crop domestication. *Am. J. Bot.* **98**, 1389–1414 (2011).
12. Hamblin, M. T., Buckler, E. S. & Jannink, J.-L. Population genetics of genomics-based crop improvement methods. *Trends Genet.* **27**, 98–106 (2011).

13. Moore, G. A. Oranges and lemons: clues to the taxonomy of Citrus from molecular markers. *TRENDS Genet.* **17**, 536–540 (2001).
14. Nicolosi, E. *et al.* Citrus phylogeny and genetic origin of important species as investigated by molecular markers. *Theor. Appl. Genet.* **100**, 1155–1166 (2000).
15. Wu, G. A. *et al.* Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nat. Biotechnol.* **32**, 656 (2014).
16. Xu, Q. *et al.* The draft genome of sweet orange (*Citrus sinensis*). *Nat. Genet.* **45**, 59 (2013).
17. Laszlo, P. *Citrus: A History*. (University of Chicago Press, 2008).
18. Webber, H. J. & Batchelor, L. D. The citrus industry. Vol. I. History, botany and breeding. *Citrus Ind. Vol Hist. Bot. Breed.* (1943).
19. Nagy, S. & Attaway, J. A. *Citrus nutrition and quality*. (ACS Publications, 1980).
20. McPhee, J. *Oranges*. (Farrar, Straus and Giroux, 2011).
21. Aschmann, H. Man's Impact on the Several Regions with Mediterranean Climates. in *Mediterranean Type Ecosystems* 363–371 (Springer, Berlin, Heidelberg, 1973). doi:10.1007/978-3-642-65520-3_21
22. Eiximenis, F. Regiment de la cosa pública. (1927).
23. Münzer, H. & Evangelista, F. C. *Viaje por España y Portugal, Reino de Granada*. (Tat, 1987).
24. Hodgson, R. W. Horticultural varieties of citrus. *Citrus Ind.* 431–591 (1967).
25. Dugo, G. & Giacomo, A. D. *Citrus: The Genus Citrus*. (CRC Press, 2003).
26. Hyman, C. *Oranges: A Global History*. (Reaktion Books, 2013).
27. Reitz, H. J. The world citrus crop. *Outlook Agric.* **13**, 140–146 (1984).
28. Agustí, M. *Cítricultura*. (2003).
29. Consellería de Agricultura, Medio Ambiente, Cambio Climático y Desarrollo Rural. Previsión de cosecha de cítricos en la campaña 2017/2018 en la Comunitat Valenciana.

30. Linnaeus, C. von. *Species plantarum*, vols 1, 2. 1957 *Carl Linnaeus Species Plant.* (1753).
31. Herrero, R., Asins, M., Carbonell, E. & Navarro, L. Genetic diversity in the orange subfamily Aurantioideae. I. Intraspecies and intragenus genetic variability. *Theor. Appl. Genet.* **92**, 599–609 (1996).
32. Swingle, W. T. The botany of Citrus and its wild relatives in the orange subfamily. *Citrus Ind.* **1**, 128–474 (1943).
33. Tanaka, T. Fundamental discussion of Citrus classification. *Stud. Citrol.* **14**, 1–6 (1977).
34. Mabberley, D. J. A classification for edible Citrus (Rutaceae). *Telopea* **7**, 167–172 (1997).
35. Grosser, J. W., Ollitrault, P. & Olivares-Fuster, O. Somatic hybridization in citrus: an effective tool to facilitate variety improvement. *Vitro Cell. Dev. Biol.-Plant* **36**, 434–449 (2000).
36. Bado, S. *et al.* Plant mutation breeding: current progress and future assessment. *Plant Breed. Rev.* **39**, 23–88 (2015).
37. Kafa, G., Seday, U., Uysal, O. & Polatoz, S. Effect of Budwood Irradiation on Seed Number of 'Clementine', 'Nova' and 'Robinson' Mandarin Mutants. *Acta Hortic.* 539–542 (2015).
38. Sutarto, I., Agisimanto, D. & Supriyanto, A. Development of promising seedless citrus mutants through gamma irradiation. *Induc. Plant Mutat. Genomics Era Food Agric. Organ. U. N. Rome* 306–308 (2009).
39. Ulukapi, K. & Nasircilar, A. G. Developments of gamma ray application on mutation breeding studies in recent years. in 31–34 (2015).
40. IPGRI, I. I. P. G. R. *Descriptores para los cítricos (Citrus spp.)*. (International Plant Genetic Resources Institute, 1999).
41. Aznar, J. S. *Reconocimiento de variedades de cítricos en campo*. (Generalitat Valenciana. Conselleria de Agricultura, Pesca y Alimentación, 1999).

42. UCR: Citrus Variety Collection. Disponible en: <http://www.citrusvariety.ucr.edu/>. (Accedido: 14 de junio del 2018)
43. Variedades de Naranjos - IVIA. Disponible en: <http://www.ivia.gva.es/es/naranjos>. (Accedido: 14 de junio del 2018)
44. Terol, J. *et al.* Sequencing of 150 citrus varieties: linking genotypes to phenotypes. in **61**, (2012).
45. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24 (2011).
46. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
47. Aleza, P. *et al.* Recovery and characterization of a Citrus clementina Hort. ex Tan.'Clemenules' haploid plant selected to establish the reference whole Citrus genome sequence. *BMC Plant Biol.* **9**, 110 (2009).
48. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. in *Bioinformatics methods and protocols* 365–386 (Springer, 2000).
49. Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. in **41**, 95–98 ([London]: Information Retrieval Ltd., c1979-c2000., 1999).
50. Orford, S., Scott, N. S. & Timmis, J. A hypervariable middle repetitive DNA sequence from citrus. *Theor. Appl. Genet.* **91**, 1248–1252 (1995).
51. Fang, D. & Roose, M. Identification of closely related citrus cultivars with inter-simple sequence repeat markers. *Theor. Appl. Genet.* **95**, 408–417 (1997).
52. Natividade Targon, M., Machado, M., Coletta Filho, H. & Cristofani, M. Genetic polymorphism of sweet orange (*Citrus sinensis* [L.] Osbeck) varieties evaluated by random amplified polymorphic DNA. in 51–54 (1998).
53. Novelli, V. M., Cristofani, M., Souza, A. A. & Machado, M. A. Development and characterization of polymorphic microsatellite markers for the sweet orange (*Citrus*

sinensis L. Osbeck). *Genet. Mol. Biol.* **29**, 90–96 (2006).

54. Coronado, A. C. M. *et al.* Caracterización de la diversidad genética en naranja y comparación del polimorfismo de microsatélites amplificados al azar (RAMs) usando electroforesis de poliacrilamida y agarosa. *Acta Agronómica* **58**, 234–244 (2009).

