



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Escola Tècnica Superior
d'Enginyeria Agronòmica i del Medi Natural

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Agronómica
y del Medio Natural

Caracterización de los genes involucrados en la
pigmentación verde del fruto de berenjena (*Solanum
melongena*)

Trabajo Fin de Grado

Grado en Biotecnología

AUTOR/A: Valero Tebar, Juan

Tutor/a: Vilanova Navarro, Santiago

Director/a Experimental: ALONSO MARTIN, DAVID

CURSO ACADÉMICO: 2021/2022



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Escuela Técnica Superior
de Ingeniería Agronómica
y del Medio Natural



Instituto de Conservación y Mejora
de la Agrobiodiversidad Valenciana

Universitat Politècnica de València

Escola Tècnica Superior d'Enginyeria Agronòmica i del Medi Natural (ETSIAMN)

Grado en Biotecnología

Caracterización de los genes involucrados en la pigmentación verde del fruto de berenjena (*Solanum melongena*)

Curso Académico 2021/2022

Trabajo de Fin de Grado

Alumno: Juan Valero Tebar

Tutor: Santiago Vilanova Navarro

Tutor experimental: David Alonso Martín

Título: Caracterización de los genes involucrados en la pigmentación verde del fruto de berenjena (*Solanum melongena*).

Autor: Juan Valero Tebar

Localidad y fecha: Valencia, julio de 2022

Tutor académico: Santiago Vilanova Navarro

Tutor experimental: David Alonso Martín

Resumen

La berenjena (*Solanum melongena*) es un cultivo perteneciente al género *Solanum* de gran relevancia en todo el mundo. Sin embargo, la mayor parte de investigación se realiza en otros cultivos modelo del género como el tomate o la patata, siendo necesario una mayor investigación en este cultivo. Uno de los aspectos que requiere una mayor investigación es la pigmentación del fruto.

Tras el análisis GWAS sobre una población MAGIC realizado por el COMAV, se identificaron dos genes candidatos para la síntesis de clorofila: el gen APRR2 situado en el cromosoma 8 y el gen GLK2 situado en el cromosoma 4. El estudio reveló dos variantes en el gen APRR2 causantes de la ausencia de pigmentación por clorofila, también se detectó la presencia de una delección en el genoma que también provocaba la ausencia de pigmentación. Del gen GLK2 no se detectó ninguna variante de gran impacto que provocará la presencia de netting.

El proyecto tiene como objetivo la validación de la existencia de la delección en el gen APRR2 y la búsqueda de variantes en el gen GLK2 que provoquen la diferencia de expresión. Para el estudio del gen APRR2 se emplearon las variedades 67/3, que presenta la delección e IVIA y A0416 que presentan un SNP que provoca la ausencia de pigmentación. Tras realizar amplificaciones del gen se comprobó que la variedad 67/3 presenta una delección en el gen APRR2, este resultado fue comprobado mediante la amplificación de muestras de la colección nuclear de berenjena establecida bajo el marco del proyecto europeo G2P-SOL. Para el estudio del gen GLK2, se realizó el procesado de la resecuenciación de los parentales de poblaciones ILs (MEL1, MEL3, MEL5, INS1, DAS y ELE), no se detectó ninguna variante de gran impacto en el gen, pero sí se observó la presencia de una inserción en la región 5'UTR en las variedades con netting. Tras realizar amplificaciones de esta región, se vio su presencia en dos de los parentales con netting (DAS y ELE), mientras que los otros parentales con netting (MEL3 y INS1) presentaban una delección contrariamente a lo esperado, siendo necesario más investigación en esta región genómica. También se comprobó si existía diferencia de expresión del gen entre los parentales amplificando el cDNA, dando como resultado que los parentales con netting, presentaban una expresión ligeramente superior a los parentales que no poseían netting. Sin embargo, todavía es necesario la optimización de protocolos para poder observar esta diferencia de expresión con más claridad.

Palabras clave: Berenjena (*Solanum melongena*), color del fruto, color genético, alelos, SNPs

Abstract

The aubergine (*Solanum melongena*) is a crop belonging to the *Solanum* genus of great relevance throughout the world. However, most of the research is carried out on other model crops of the genus such as tomato or potato, and further research on this crop is necessary. One of the aspects that requires further investigation is the pigmentation of the fruit.

After GWAS analysis on a MAGIC population carried out by COMAV, two candidate genes for chlorophyll synthesis were identified: the APRR2 gene located on chromosome 8 and the GLK2 gene located on chromosome 4. The study revealed two variants in the APRR2 gene causing the absence of pigmentation by chlorophyll, the presence of a deletion in the genome that also caused the absence of pigmentation was also detected. On the GLK2 gene, no high-impact variants were detected that would cause the presence of netting.

The project aims to validate the existence of the deletion in the APRR2 gene and search for variants in the GLK2 gene that cause the difference in expression. For the study of the APRR2 gene, varieties 67/3 were used, which presents the deletion, also IVIA and A0416, which present a SNP that causes the absence of pigmentation. After performing gene amplifications, it was found that the 67/3 variety has a deletion in the APRR2 gene. This result was verified by amplifying samples from the nuclear collection of aubergines established under the framework of the European project G2P-SOL. For the study of the GLK2 gene, the processing of the resequencing of the parental ILs populations (MEL1, MEL3, MEL5, INS1, DAS and ELE) was carried out, no variant of great impact was detected in the gene, but it was observed the presence of an insertion in the 5'UTR region in the netted varieties. After carrying out amplifications of this region, its presence was seen in two of the netted parents (DAS and ELE), while the other netted parents (MEL3 and INS1) presented a deletion contrary to expectations, requiring more research in this area. genomic region. It was also checked if there was a difference in gene expression between the parents by amplifying the cDNA, resulting in the parents with netting having a slightly higher expression than the parents who did not have netting. However, protocol optimization is still necessary to be able to observe this difference in expression more clearly.

Key words: Aubergine (*Solanum melongena*), fruit color, genetic color, alleles, SNPs

Agradecimientos:

En primer lugar, quisiera agradecer a Santi y a David por toda la ayuda que me han ofrecido cada día durante la realización de este proyecto.

También darle las gracias a Marina y Paula por hacer más amenos todos los días que hemos pasado en el laboratorio.

No puedo olvidarme de mis compañeros de carrera, por hacerme disfrutar estos cuatro años que no podré olvidar.

Y, por último, a mi familia, por cuidarme y apoyarme durante todos estos años, sin ellos nada de esto habría sido posible.

ÍNDICE

1. INTRODUCCIÓN	1
1.1 Importancia de la berenjena	1
1.2 Origen y evolución.....	2
1.3 Especies utilizadas en el proyecto.....	3
1.4 Importancia de las nuevas tecnologías	5
1.5 Líneas de mejora	5
1.6 Pigmentación de la berenjena.....	6
1.7 Genes involucrados en la síntesis de pigmentos.....	8
1.7.1 Gen APRR2.....	9
1.7.2 Gen GLK2	9
2. OBJETIVOS.....	10
3. MATERIALES Y MÉTODOS.....	11
3.1 Material vegetal	11
3.2 Extracción DNA.....	12
3.3 Diseño de cebadores.....	12
3.4 Amplificación DNA.....	13
3.5 Resecuenciación de los Parentales de las ILs.....	13
3.6 Procesado de la resecuenciación, mapeado y búsqueda de polimorfismos	13
3.6.1 Filtrado por calidad	14
3.6.2 Mapeo de lecturas.....	14
3.6.3 Búsqueda y anotación de variantes	15
3.6.4 Visualización.....	16
3.7 Extracción RNA	16
3.8 Síntesis cDNA.....	17
3.9 Amplificación cDNA.....	17
3.10 Amplificación de la delección	18
4. RESULTADOS Y DISCUSIÓN	19
4.1 Gen APRR2.....	19
4.1.1 Validación de la delección detectada <i>in silico</i>	19
4.1.2 Comprobación de la delección en otras variedades	22
4.2 Gen GLK2	23
4.2.1 Resecuenciación del genoma completo y mapeado.....	24
4.2.2 Variantes genómicas y distribución	25
4.2.3 Análisis de la diferencia de expresión entre parentales	30
4.2.4 Validación de la inserción detectada <i>in silico</i>	31

4.2.5 Comprobación de la inserción en otras variedades	33
5. CONCLUSIÓN	35
6. BIBLIOGRAFÍA.....	36

Índice de Figuras

Figura 1: Tendencia de producción de berenjena en los países que más producen.

Figura 2: Distribución global del complejo berenjena según la clasificación de Lester.

Figura 3: Representación esquemática de las relaciones taxonómicas de las especies del género *Solanum*.

Figura 4: Estructura química de cis-nasunina (1) y trans-nasunina (2).

Figura 5: Estructura química de la clorofila.

Figura 6: Manhattan plot obtenido tras el GWAS, mostrando picos en el cromosoma 4 y en el cromosoma 8.

Figura 7: Diagrama de flujo del procesado informático de las lecturas obtenidas tras la resecuenciación.

Figura 8: Esquema del genoma de las muestras empleadas para la validación de la detección detectada *in silico*.

Figura 9: Amplificación de las variedades de berenjena con variantes en el gen APRR2 empleando la pareja de cebadores Indel-R APRR2 y APRR2-R.

Figura 10: Amplificación de las variedades de berenjena con variantes en el gen APRR2 con la pareja de cebadores Indel-R APRR2 y Indel-F APRR2-R y la pareja Indel-F APRR2 y A0416-F.

Figura 11: Variedades de la colección nuclear de berenjena establecida bajo el marco del proyecto europeo: G2P-SOL empleadas para la comprobación de los cebadores.

Figura 12: Amplificación de las muestras de la colección G2P-SOL con la pareja de cebadores Indel-R APRR2 y Indel-F APRR2-R y la pareja de cebadores Indel-F APRR2 y A0416-F.

Figura 13: Parentales de las líneas ILs desarrolladas en el COMAV.

Figura 14: Visualización en IGV tras la anotación de los parentales MEL1 y MEL5.

Figura 15: Visualización en IGV tras la anotación de los parentales MEL3 y INS1.

Figura 16: Visualización en IGV tras la anotación de los parentales DAS y ELE.

Figura 17: Amplificación de los parentales ILs con la pareja cebadores CAC-F y CAC-R.

Figura 18: Amplificación de los parentales ILs con la pareja cebadores GLK2-F y GLK2-R.

Figura 19: Muestras de los parentales de las poblaciones empleadas para la validación de la inserción.

Figura 20: Esquema del genoma de las muestras empleadas para la validación de la detección detectada *in silico*.

Figura 21: Amplificación de los parentales con la pareja de cebadores GLK2-Indel-F y GLK2-Indel-R.

Figura 22: Amplificación de las muestras de la colección G2P-SOL con la pareja de cebadores GLK2-Indel-F y GLK2-Indel-R.

Índice de Tablas

Tabla 1: Lista del subconjunto cultivares empleado para la comprobación de los genes.

Tabla 2: Lista de cebadores empleados durante los experimentos.

Tabla 3: Lista de componentes empleados durante la amplificación del DNA.

Tabla 4: Lista de componentes empleados durante la amplificación del cDNA.

Tabla 5: Lista de variantes presentes en los genomas de berenjena y el efecto que causan.

Tabla 6: Estadísticas de las lecturas tras el filtrado.

Tabla 7: Estadísticas tras el mapeo de las lecturas.

Tabla 8: Estadísticas tras la anotación de variantes de las lecturas.

Abreviaturas

MEL: *Solanum melongena* L.

INS: *Solanum insanum*

DAS: *Solanum dasyphyllum*

ELE: *Solanum eleagnifolium*

IL: Línea de introgresión

MAGIC: Cruzamiento multiparental de generación avanzada

ROS: especies reactivas de oxígeno

GWAS: estudio de asociación del genoma completo

PCR: reacción en cadena de la polimerasa

BGI: Instituto de Genómica de Beijing

SNP: polimorfismo de nucleótido único

MNP: polimorfismo de nucleótidos múltiples

1. INTRODUCCIÓN

1.1 Importancia de la berenjena

La berenjena (*Solanum melongena* L., $2n = 2x = 24$) es un miembro de la gran familia de las Solanáceas, que contiene alrededor de 3000 especies ([Toppino et al., 2020](#)), dentro de la cual se encuentran especies dedicadas para el consumo como el tomate (*Solanum lycopersicum*), la patata (*Solanum tuberosum*) o el pimiento (*Capsicum annuum*), también se encuentran especies con propiedades médicas como el tabaco (*Nicotiana sp.*) o la belladona (*Atropa belladonna*) así como plantas ornamentales como la petunia (*Petunia hybrida*) o el género Licuantes ([Hirakawa et al., 2014](#)). Dentro de esta diversa familia, el género *Solanum* el más grande y el más importante ([Ghatak et al., 2017](#)).

Las propiedades nutritivas de la berenjena son destacadas y es considerada una de las hortalizas más sanas por poseer un bajo recuento calórico y un considerable contenido en vitaminas, minerales y compuestos bioactivos potencialmente beneficiosos para la salud humana ([Taher et al., 2017](#)). Dentro de estos compuestos bioactivos encontramos fenoles, alcaloides, saponinas, terpenos, flavonoides, cumarinas y carotenoides ([Kaunda et al., 2019](#)), destacando los fenoles, particularmente el ácido clorogénico en la carne del fruto y las antocianinas en la piel ([Plazas et al., 2014](#)). No obstante, la biodisponibilidad de estos compuestos suele ser baja y esto puede ocultar sus efectos potenciales ([Velderrain-Rodríguez et al., 2014](#)). El potencial bioactivo que presenta la berenjena ha sido probado, pero todavía requiere de investigación en su digestión y absorción para conseguir los efectos deseados ([Elizalde-Romero et al., 2021](#)).

Por todos estos motivos, la berenjena es un cultivo de gran importancia en Asia, Oriente Medio y Próximo, el sur de Europa y África ([Hirakawa et al., 2014](#)). La producción global ha ido aumentando en los últimos años, especialmente en la última década, alcanzando el año 2020 una producción de 56,46 millones de toneladas ([Figura 1](#)). En ese año, China fue el país que más producción tuvo (36,56 millones de toneladas) seguido de la India (12,78 millones de toneladas), superando claramente al resto de países. Dentro de los países europeos Italia (304,69 miles de toneladas) y España (282,20 miles de toneladas) fueron los que más produjeron. Sin embargo, España fue el país que más exportó, con un valor de 198.9 millones de dólares, teniendo este valor una tendencia al alza durante el último lustro ([Tridge, 2022](#)).

En los últimos años, ha ocurrido una revolución en la mejora vegetal y en la genómica que ha generado grandes cantidades de información útiles para la mejora de muchos cultivos ([Hurtado et al., 2013](#)). No obstante, la mayoría de la investigación se ha realizado en otros cultivos

pertenecientes a las solanáceas como el tomate o el pimiento, quedando la berenjena relegada a un segundo lugar ([Gramazio et al., 2018](#)).

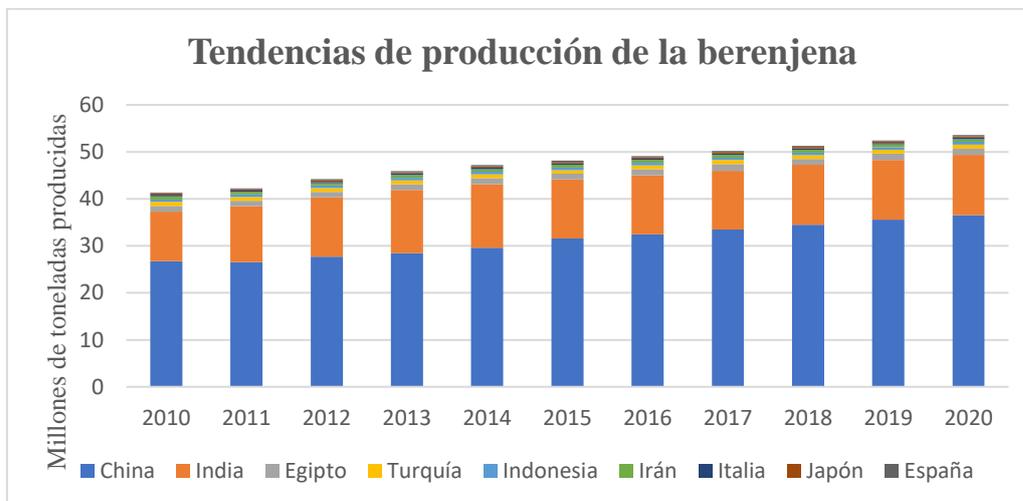


Figura 1: Tendencia de producción de berenjena en los países que más producen, datos obtenidos en ([Tridge, 2022](#)).

1.2 Origen y evolución

La diversidad es muy importante para los programas de mejora, por eso el estudio de como evolucionó la berenjena y de las especies filogenéticamente más cercanas es un aspecto de vital importancia ([Taher et al., 2017](#)). La mayoría de las especies dentro del género *Solanum* tienen su origen en América. Sin embargo; las tres especies de berenjena cultivada pertenecientes al subgénero *Leptostemonum*: *Solanum melongena* L. (berenjena), *Solanum aethiopicum* L. (berenjena escarlata) y *Solanum macrocarpon* L. (berenjena gboma) y sus familiares silvestres presentan su origen en el “Viejo mundo” ([Doganlar et al., 2002](#)).

Las relaciones entre las especies silvestres, semidomesticadas y cultivadas de *Solanum melongena* son controversiales ([Weese and Bohs, 2010](#)). Hoy en día, es aceptado que la berenjena actual presenta su domesticación a partir de la especie *Solanum incanum*, originaria de África, la cual migró hacia los países asiáticos, concretamente en la zona entre la India y China ([Frery et al., 2007](#)).

El complejo berenjena incluye la especie *Solanum melongena* y sus parientes más cercanos ([Knapp et al., 2013](#)). Este complejo se divide en ocho grupos diferentes ([Figura 2](#)), según la clasificación de Lester ([Daunay et al., 2001](#)), cuatro de estos grupos A, B, C y D, representan a la especie *Solanum incanum*. Los grupos E, H, G y H, representan a la especie *Solanum melongena*. El grupo A, B y D comprenden especies taxonómicamente cercanas a *Solanum incanum*.

Comprendiendo el grupo C, *Solanum incanum* en sentido estricto, desde este grupo *S. incanum* migró hacia Asia, donde se produce la domesticación de la berenjena. El grupo F, repartido por toda Asia, comprende cultivos más primitivos que se fueron domesticando, generando nuevos grupos: el grupo E, que comprende *Solanum insanum* y el grupo G, que comprende cultivares primitivos de *Solanum melongena*. Finalmente, el grupo H comprende los cultivares modernos repartidos por todo el mundo, no incluyéndose en la figura 2 (Weese and Bohs, 2010).

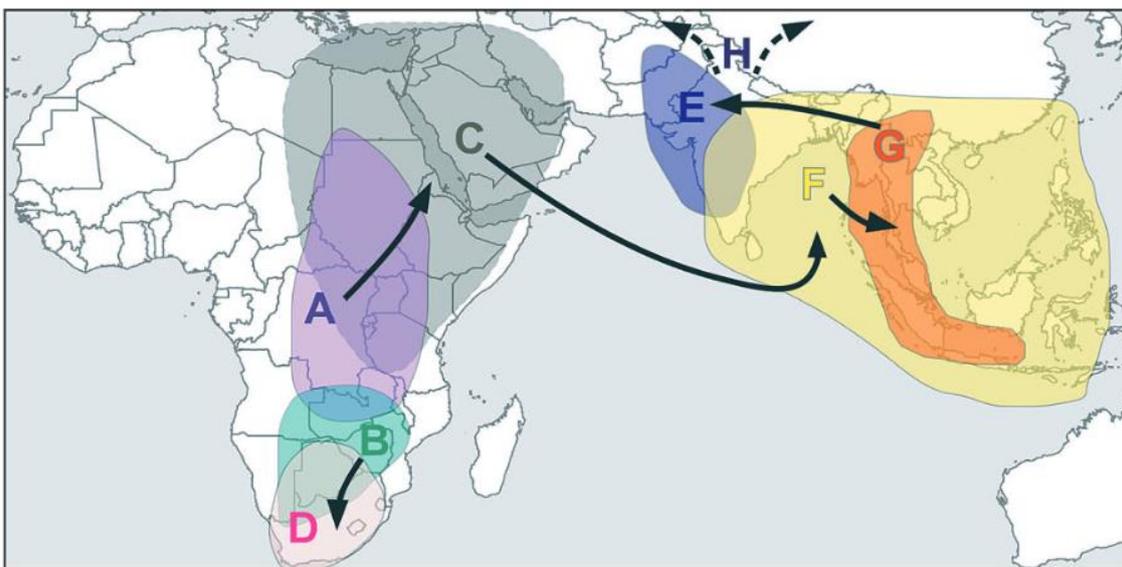


Figura 2: Distribución global del complejo berenjena según la clasificación de Lester (Weese and Bohs, 2010).

1.3 Especies utilizadas en el proyecto

Como se ha visto, el origen y la evolución de la berenjena es un proceso complejo que se ha dado a lo largo de muchos años. En este proyecto se ha trabajado con diferentes especies pertenecientes al género *Solanum* y al subgénero *Leptostemonum* (Figura 3) (Taher et al., 2017).

***Solanum melongena* (MEL):** berenjena común, perteneciente al grupo del viejo mundo. Es la especie de berenjena más cultivada en el mundo y sobre la que más estudio se han realizado (Taher et al., 2017). Descrita con detalle en el apartado Importancia de la berenjena.

***Solanum insanum* (INS):** perteneciente al grupo del viejo mundo, es un progenitor silvestre de la berenjena cultivada (*S. melongena*), se encuentra distribuido de forma natural en el sudeste asiático, Madagascar y las Islas Mauricio (Knapp et al., 2013). Ha sido sujeto de estudio en muchas investigaciones debido a su gran proximidad taxonómica con *S. melongena*, que favorece su empleo en la mejora vegetal. Esta especie ha sido empleada en una gran variedad de

preparaciones culinarias y médicas dentro de la cultura asiática, revelando estudios actuales su potencial debido a sus propiedades fitoquímicas ([Ranil et al., 2017](#)).

***Solanum dasyphyllum* (DAS):** perteneciente al grupo del viejo mundo, al grado de *anguivi*. Es común en las tierras altas de África ([Vorontsova et al., 2013](#)). Esta especie se caracteriza por ser rica en fitoquímicos bioactivos como glicoalcaloides, glicósidos, polifenoles y flavonoides ([Adesina, 1985](#)). Ha sido empleada tradicionalmente como un tratamiento para contra la diarrea, para tratar heridas, como antiespasmolítico, anticonvulsivo y por su actividad neuroprotectora. Actualmente, todas estas propiedades están siendo evaluadas ([Ilesanmi et al., 2020](#)).

***Solanum eleagnifolium* (ELE):** perteneciente al grupo del nuevo mundo. Originaria de las Américas, presentan una distribución disyuntiva, ocurriendo en los desiertos y zonas secas del hemisferio norte en el suroeste de Estados Unidos y México y en el sur de Argentina y Chile. Hoy en día, se ha extendido a regiones tropicales y subtropicales de todo el mundo ([Mekki, 2007](#)). Es una planta tóxica para el ganado, que presenta muchas dificultades para ser controlada, ya que se regenera con mucha facilidad. Es considerada por muchos países como una especie invasora, siendo un problema especialmente grave en Australia, en los países de la región EPPO, dentro de los cuales se encuentra España, todavía se encuentra en fase de establecimiento, siendo el riesgo de dispersión significativo ([Cunningham et al., 2004](#)).

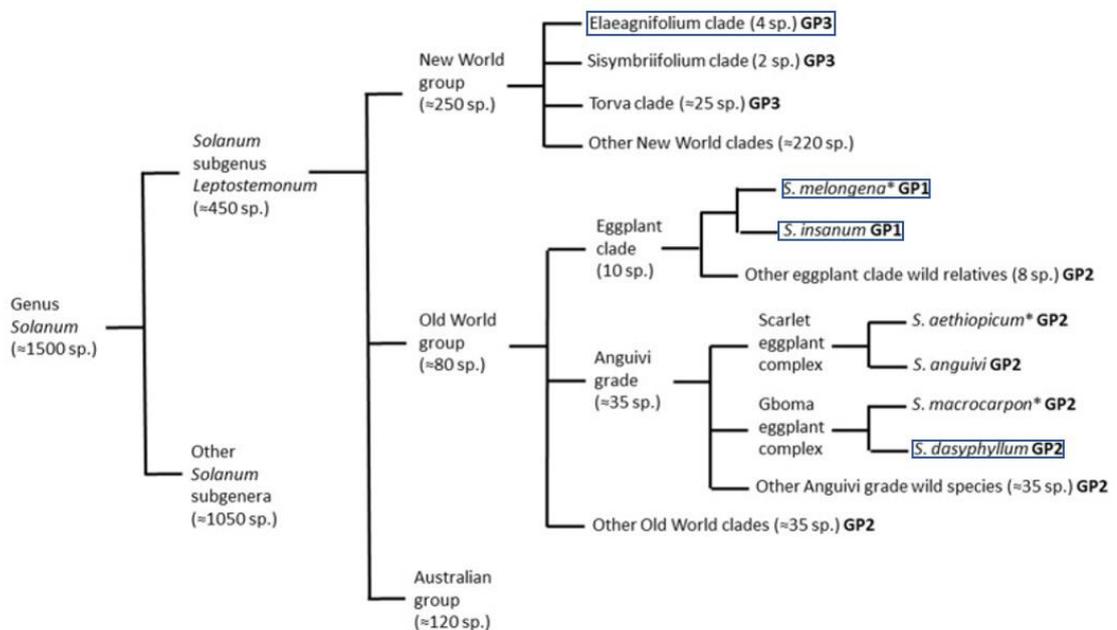


Figura 3: Representación esquemática de las relaciones taxonómicas de las especies del género *Solanum* ([Taher et al., 2017](#)).

1.4 Importancia de las nuevas tecnologías

La secuenciación genómica es una herramienta muy importante en el campo de la genética y genómica de plantas, siendo el genoma de *Arabidopsis thaliana*, el primero en ser secuenciado en el año 2000 ([Wei et al., 2020](#)). Desde ese momento se han secuenciado el genoma de múltiples plantas, habiéndose realizado la mayoría de los estudios genómicos en el género *Solanum* en especies modelo como el tomate o la patata ([Kevei et al., 2015](#)). No obstante, los recientes esfuerzos realizados en los estudios genómicos de la berenjena han reducido la brecha que había con las otras especies modelo ([Yang et al., 2014](#)). Las mejoras en las nuevas tecnologías de secuenciación y la reducción en los costes de los estudios genómicos asociados a estas mejoras han contribuido a esta mejora ([Gramazio et al., 2018](#)).

La primera secuenciación del genoma de *Solanum melongena* fue publicado en 2014, prediciendo 85446 genes, pero no llegó al nivel de cromosoma y estaba muy fragmentado, cubriendo sólo el 74% del genoma ([Hirakawa et al., 2014](#)). Otros genomas de referencia disponibles para el estudio de la berenjena son: el genoma de HQ-1315, el cuál es una variante sin pigmentación verde en el fruto ([Wei et al., 2020](#)). También están disponibles genomas de referencia que si muestran pigmentación en el fruto como son el de Nakate-Shinkuro ([Hirakawa et al., 2014](#)) y GUIQIE-1 ([Li et al., 2021](#)). La versión más actual del genoma de berenjena pertenece a la línea de berenjena “67/3” (versión 4), cuyo ensamblaje ha sido mejorado mediante el uso de HI-C, el mapeo óptico de Bionano y su integración con las lecturas cortas de Illumina ([Barchi et al., 2021](#)). La resecuenciación del genoma completo de la berenjena ha permitido obtener información de gran relevancia para la genética, el cultivo y la evolución de los cultivos, en este caso de la berenjena ([Gramazio et al., 2019](#)).

1.5 Líneas de mejora

La domesticación de las especies silvestres y la preferencia de variedades mejoradas respecto a las variedades tradicionales ha supuesto un proceso de pérdida de la diversidad genética ([Smýkal et al., 2018](#)), siendo perjudicial para los mejoradores, ya que pierden diversidad y herramientas potencialmente útiles para la mejora ([Carrillo, 2016](#)). Sin embargo; en los últimos años, se han realizado grandes esfuerzos para conservar la diversidad genética dentro de las variedades modernas, empleando variedades locales y silvestres en los programas de mejora ([Sotomayor Cantos et al., 2016](#)). En el caso de las silvestres, presentan algunos inconvenientes como la introgresión de caracteres no deseados que deben ser eliminados o problemas relacionados con la incompatibilidad sexual y las barreras reproductivas. A pesar de todos estos factores, al ser genéticamente tan variables son un gran punto de interés ([Bradshaw, 2017](#)). La eficacia de los procesos de mejora depende en gran parte de la disponibilidad de materiales de élite con un fondo

genético adaptado, en el caso de la berenjena estos aportes son muy limitados. Gracias al empleo de marcadores moleculares se han desarrollado poblaciones útiles para la mejora de los cultivos ([Sotomayor Cantos et al., 2016](#)), dentro de las cuales destacan:

- Líneas de introgresión (ILs): Una colección completa de ILs representa el genoma completo de la especie donante, la cual posee genes de interés. En estas líneas la mayoría del fondo genético es de la especie cultivada menos un pequeño fragmento de la especie donante ([Zamir, 2001](#)).

- Cruzamiento multiparental de generación avanzada (MAGIC): Poblaciones que presentan un conjunto de parentales elevados (8 o 16), siendo algún parental de una especie silvestre con genes de interés, que se cruzan entre ellos, originando un híbrido cuádruple que cuando se autofecunde, originará una población ultrasegregante ([Prohens et al., 2017](#)).

1.6 Pigmentación de la berenjena

Dentro de las cualidades de los frutos, el color de la piel es uno de los más importantes. No solamente, por las preferencias visuales de los consumidores, sino también porque los pigmentos que confieren color influyen en los valores nutricionales y de sabor ([Brand et al., 2014](#)).

El color del fruto de la berenjena es muy variable, tanto en aspectos cuantitativos como cualitativos. ([Li et al., 2018](#)). El color del fruto de la berenjena se puede reducir a la ausencia y presencia de dos pigmentos principales: las antocianinas y la clorofila. Las antocianinas son responsables del color púrpura en la piel del fruto y la clorofila le confiere el color verde, este color verde es visible cuando las antocianinas no están presentes o cuando su concentración es muy baja. Si las antocianinas están presentes, la clorofila refuerza el color púrpura del fruto ([Daunay et al., 2004](#)). Actualmente, las berenjenas púrpuras son las más demandadas en el mercado ([Li et al., 2018](#)).

Las antocianinas son pigmentos polifenólicos que pertenecen al grupo de los flavonoides responsable de la mayoría de los colores azules, púrpuras, naranjas y rojos presentes en los órganos de las plantas como frutos, hojas y flores ([Wallace and Giustiu, 2015](#)). Diferentes tipos de antocianinas han sido identificadas en la piel de berenjena, siendo la nasunina ([Figura 4](#)) como la antocianina con mayor presencia, con un rango del 69,1% al 87,7% ([Todaro et al., 2009](#)). Estos compuestos destacan por su capacidad antioxidante e antihiperlipidémica, la cual está relacionada con su estructura química ([Basuny et al., 2012](#)).

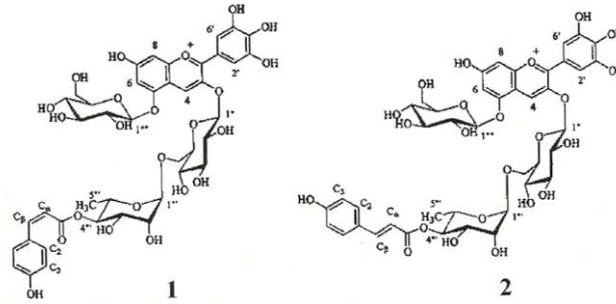


Figura 4: Estructura química de cis-nasunina (1) y trans-nasunina (2) (Gallo et al., 2014).

Las clorofilas es el pigmento natural más importante debido a su rol esencial en la fotosíntesis. Son una clase de tetrapirroles (Figura 5) y su función principal reside en los centros fotosintéticos de los cloroplastos, donde actúan como el centro de reacción principal, como pigmentos accesorios y como fotoprotectores (Senge et al., 2014). También actúan como fotosensibilizadores, produciendo especies reactivas de oxígeno (ROS) en situaciones donde hay intensidades de luz saturadas (Foyer, 2018). Además, hay evidencia de que el metabolismo de la clorofila no sólo tiene un impacto significativo en la maquinaria fotosintética, sino que también influye en otros procesos como la muerte celular programada, la comunicación cloroplasto-núcleo y el fenómeno “stay-green” (Tanaka and Tanaka, 2006).

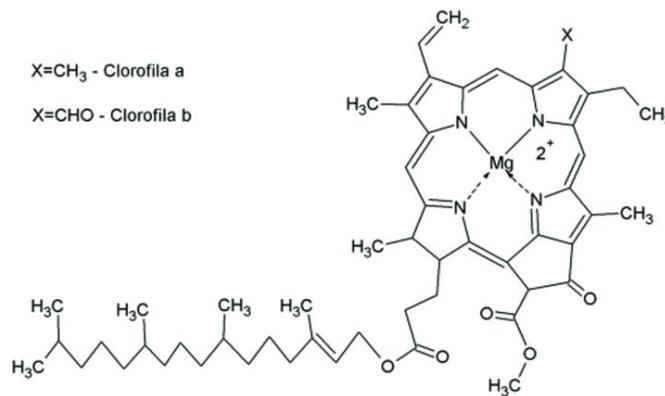


Figura 5: Estructura química de la clorofila (Vallejo et al., 2018).

Las hojas son el mayor órgano fotosintético de las plantas, aunque los frutos verdes en desarrollo son capaces también de realizar la fotosíntesis. Varios frutos en desarrollo, los cuales presentan cloroplastos funcionales capaces de realizar la fotosíntesis, esto lleva a la acumulación de metabolitos asociados con la calidad nutricional del fruto (Jia et al., 2020).

Por todos estos motivos, entender los mecanismos a través de los cuales se sintetizan las clorofilas en la piel del fruto de la berenjena es de gran relevancia para la mejora genética de la berenjena (Jia et al., 2020).

1.7 Genes involucrados en la síntesis de pigmentos

Los genes candidatos para la síntesis, distribución y acumulación de antocianinas en los frutos ya han sido propuestos, habiendo localizado dos clusters de genes. Uno localizado en el cromosoma 10 relacionado con la producción y acumulación de antocianinas en los órganos vegetativos, este cluster está formado por un conjunto de cinco genes predichos como ANS, que incluyen factores de transcripción tipo MYB y fosfoinositida fosfatasa. El otro cluster está localizado en el cromosoma 5 relacionado con la tonalidad en la corola de las flores, siendo el gen candidato AAT, acetil-CoA-benzialcohol-acetiltransferasa ([Topinno et al., 2020](#)). En otro estudio más reciente, se confirmó el gen SmMYB113 como responsable en la producción de antocianinas ([Yang et al., 2022](#)). Sin embargo, los genes causantes de la presencia de clorofila en la piel del fruto de la berenjena todavía no han sido clarificados.

Estudios previos realizados en el COMAV desarrollaron una población MAGIC con seis parentales que presentaban pigmentación por clorofila y otros dos que no. Sobre esta población se realizó un GWAS (estudio de asociación del genoma completo) y su análisis combinando la información genotípica y fenotípica reveló dos asociaciones importantes, una más importante en el cromosoma 8 y otra de menor relevancia en el cromosoma 4 ([Figura 6](#)).

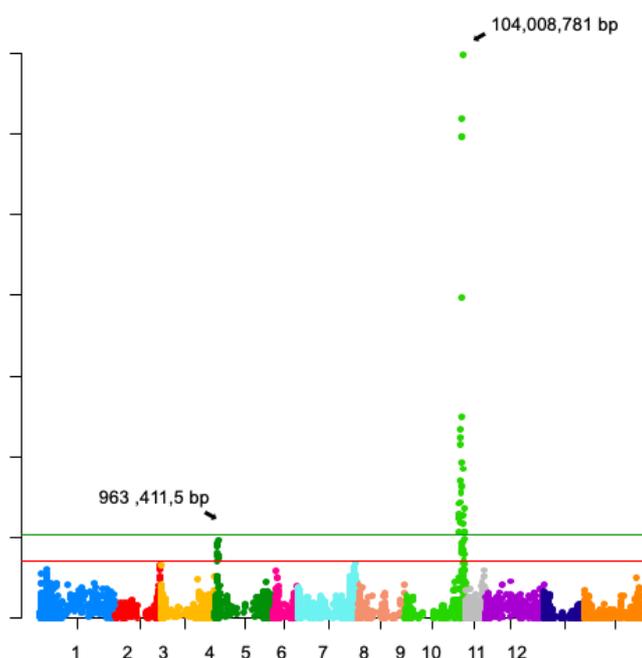


Figura 6: Manhattan plot obtenido tras el GWAS, mostrando picos picos en el cromosoma 4 y en el cromosoma 8.

El pico del cromosoma 8 revelado por el GWAS se correspondía a una región cercana a un gen similar a APRR2. Se predijeron dos variantes de este gen, las cuales tienen un gran impacto, las mutaciones provocan un cambio en la pauta de lectura. TCTCCG es sustituido por TG en la

posición 14.010.718 del exón 3 en la primera variante y ACT es sustituido por AT en la posición 104.011.141 del exón 6 en la segunda variante, siendo asociadas ambas con la ausencia de pigmentación de clorofila en la piel del fruto. También se identificó *in silico* una gran deleción de 5,27 kb en dos de los genomas de referencia de las accesiones que no presentaban pigmentación en la piel del fruto. Mientras que, en la región asociada al pico menor, presente en el cromosoma 4, se detectó la existencia de un gen candidato, el gen GLK2. De este gen no se predijeron variantes que causarían un gran impacto a nivel de exón.

1.7.1 Gen APRR2

El gen APRR2 (*Arabidopsis* pseudo response regulator2-like) perteneciente a la familia de reguladores de respuesta de dos componentes similar a CheY, codifica el regulador de pseudorespuesta 2 (APRR2) que interactúa con un sensor de calcio CML9 ([String, 2022](#)). Este gen es el causante de la diferencia cualitativa entre la corteza verde oscura y clara en los cultivos. Se ha observado que, en tomate, una sobreexpresión de este gen causa un aumento del número de plastidios, el área y el contenido de pigmento aumentando los niveles de clorofila en frutos inmaduros y carotenoides en frutos rojos maduros ([Pan et al., 2013](#)). En pimiento, se encontró una fuerte asociación entre CαAPRR2 y la acumulación de clorofila en tejidos de frutos en desarrollo ([Jeong et al., 2020](#)). Además, una mutación que inactiva el gen APRR2 ha sido relacionada con alteraciones en la señalización del ácido abscísico involucrada con el desarrollo de los plástidos causando una baja densidad de cloroplastos y contenido en clorofila ([Pan et al., 2013](#)).

1.7.2 Gen GLK2

El gen GLK2 es un factor de transcripción del tipo Golden2-like (GLK), pertenece a la subfamilia GARP de los factores de transcripción MYB ([Li et al., 2018](#)). Este factor de transcripción induce la expresión de genes relacionados con la diferenciación de los cloroplastos y la activación de genes implicados en la fotosíntesis ([Siddiqui et al., 2018](#)). En tomate existen dos copias de este gen, SIGLK1 y SIGLK2, cada uno con patrones de expresión diferentes. Mientras que el primero predomina en las hojas, el segundo se expresa en frutos, en la región del pedicelo, formando un patrón reticulado de las clorofilas, también llamado netting en el hombro de los frutos ([Lupi et al., 2019](#)). La pérdida de la función de los genes GLK provoca la formación deficiente de tilacoides y la reducción en la acumulación de productos de genes fotosintéticos que son necesarios para la captación de luz y para la síntesis de clorofila ([Bravo-García et al., 2009](#)).

2. OBJETIVOS

- 1.- Validar la delección identificada *in silico* del gen APRR2 en varias accesiones.
- 2.-Analizar la resecuenciación de los parentales de las líneas de introgresión que difieren en el carácter netting.
- 3.- Buscar variantes del gen GLK2 que se asocien al fenotipo netting.
- 4.- Examinar potenciales diferencias de expresión del gen GLK2 en la piel del fruto de variedades con y sin netting.

3. MATERIALES Y MÉTODOS

3.1 Material vegetal

Para el estudio del gen APRR2, se emplearon diversas variedades cultivadas de *S. melongena*: IVIA y A0416 que no presentan pigmentación por clorofila, que se utilizaron como parentales de la población MAGIC con la que se realizó el GWAS, y 67/3, que es la variedad del genoma de referencia, la cual no presenta pigmentación por clorofila, pero sí por antocianinas. Estas tres variedades se utilizaron para la búsqueda de cebadores adecuados en la validación de la delección.

Para el estudio del gen GLK2 se emplearon diferentes especies del género *Solanum*, dentro de las cuales había especies que presentan netting y otras que no. *S. dasyphyllum* (DAS), *S. eleagnifolium* (ELE) y *S. insanum* (INS) son tres especies silvestres que presentan netting. Dentro de *S. melongena* (MEL) se emplearon tres variedades diferentes: MEL 1 y MEL5 no presentan netting en su fenotipo, mientras que MEL3 sí que presenta netting. Estas especies se utilizaron como parentales de tres ILs en desarrollo: la primera IL: MEL1 x DAS, la segunda IL: MEL3 x ELE y la tercera IL: MEL5 x INS.

Por último, para confirmar los resultados obtenidos en ambos genes, se empleó un subconjunto de cultivares tradicionales, seleccionados por la presencia o ausencia de pigmentación verde en el fruto, pertenecientes a la colección nuclear de berenjena establecida bajo el marco del proyecto europeo G2P-SOL(<http://www.g2p-sol.eu/>) (Tabla 1).

Tabla 1: Lista del subconjunto cultivares empleado para la comprobación de los genes.

ID	Código G2P-SOL	Fenotipo
B2_46	GPE001970	Referencia 67/3 Morada blanca
B1_154_1	GPE014270	Morada blanca sin netting
B1_203_1	GPE021160	Blanca listada sin netting
B1_185_1	GPE019980	Blanca sin netting
B1_228_1	GPE022930	Blanca sin netting
B1_377_1	GPE045970	Blanca sin netting
B1_184_1	GPE019870	Blanca sin netting
B1_376_1	GPE045840	Verde con netting
B1_405_1	GPE053050	Verde con netting
B1_383_1	GPE047650	Verde sin netting
B1_354_1	GPE040230	Verde sin netting
B1_281_1	GPE031380	Verde con netting

3.2 Extracción DNA

Para el estudio del gen *APRR2* se extrajo el DNA genómico de las hojas de diferentes variedades, cogiendo aproximadamente alrededor de 50 mg de material vegetal. Se empleó el protocolo de extracción basado en CTAB-sílex desarrollado por [\(Vilanova et al., 2020\)](#). Una vez extraído el DNA, se cuantificó empleando un espectrofotómetro Nanodrop ND1000, teniendo en cuenta los ratios 260/230 y 260/280, comprobando que fuera mayor que 1,8 para asegurar una correcta calidad de la extracción sin presencia de ningún contaminante.

También se realizó un gel de agarosa al 0,8% para comprobar la integridad del ADN. Las muestras se cargaron en el gel junto al tampón de carga Blue sin Red (6X DNA Loading Dye de la casa comercial ThermoFisher). La electroforesis corrió durante 20 minutos a 120V. Se realizó una tinción con bromuro de etidio y se visualizaron los resultados en un transiluminador de luz UV Gel Doc empleando el software Quantity One (BioRad). Se emplearon los marcadores moleculares Lambda *HindIII* de la casa comercial Fermentas y 100bp de la casa comercial Nippon Genetics para determinar el tamaño de los fragmentos.

3.3 Diseño de cebadores

Los cebadores empleados durante los procesos de amplificación fueron diseñados mediante el software Primer3 [\(Kõressaar et al., 2018\)](#). Para un correcto diseño se tuvo en cuenta que la longitud de los cebadores fuera entre 18 y 23 pares de bases, que tuvieran una temperatura de melting (T_m) mínima de 52°C y máxima de 64°C, y un contenido CG alrededor del 40% [\(Singh and Kumar, 2001\)](#). Se utilizaron diferentes cebadores para el estudio de los genes [\(Tabla 2\)](#).

Tabla 2: Lista de cebadores empleados durante los experimentos.

Cebador	Secuencia	Temperatura melting
A0416_F	TGACCAAGACAGTGGTGAGC	62°C
APRR2_R	ACACATGGCTCACTCTGTGC	62°C
APRR2_INDEL_R	GATGAAAACCAGGCACCATT	60°C
APRR2_INDEL_F	TTTCTTTCCCTCAATGTGGG	58°C
CAC_R	GGCAGAGCTTGATCCTCTTT	62°C
CAC_F	CCAGGAACCACTCAAGATACAG	62°C
GLK2_R	ATGCCTGGGACCATAAGAGG	60°C
GLK2_F	GGATTGGACTCCGGAGCTAC	63°C
GLK2_INDEL_R	TGGAAGGAATTGGAAGTTTCT	55°C
GLK2_INDEL_F	TGTGAAGATTTAGTTTCCATGT	54°C

3.4 Amplificación DNA

La amplificación del DNA extraído se realizó empleando el protocolo Taq 2X Master Mix de la casa comercial PCR Biosystems siguiendo las indicaciones del fabricante, en la [tabla 3](#) se indican los volúmenes empleados para la reacción de PCR. Previamente, el DNA fue diluido a una concentración de 50 ng/μL, empleando 100 ng para la amplificación. Las condiciones de amplificación en el termociclador fueron: una desnaturalización inicial de 92°C, luego se realizaron 35 ciclos con una fase de desnaturalización a 92°C durante 15 segundos, una fase de alineamiento a 60°C durante 20 segundos y una fase de extensión a 72°C durante 2 minutos, finalmente se realizó una elongación final a 72°C durante 10 minutos.

Tabla 3: Lista componentes empleados durante la amplificación del DNA.

Reactivo	Reacción 25μL	Concentración Final
10 μM Forward Primer	0,5 μL	0,2 μM
10 μM Reverse Primer	0,5 μL	0,2 μM
Taq 2X Master Mix	12,5 μL	1X
Nuclease-free water	9,5 μL	-
Template DNA	2 μL	< 1000 ng

La visualización de las muestras amplificadas se realizó según lo descrito en el apartado 3.2.

3.5 Resecuenciación de los Parentales de las ILs

El DNA fue aislado de una planta de cada una de las seis parentales descritos en el apartado 3.1 y fue enviado al Beijing Genomics Institute (BGI Genomics) para que fueran resecuenciadas. El protocolo que utilizaron se basó en la preparación de las librerías de paire-ends(2x150pb) con un tamaño de inserto aproximado de 300pb y fueron secuenciados con la plataforma DNBseq.

3.6 Procesado de la resecuenciación, mapeado y búsqueda de polimorfismos

Una vez obtenidas las lecturas estas fueron procesadas siguiendo el diagrama de flujo mostrado en la [figura 7](#). Se realizó el diagrama de flujo completo para cada una de las diferentes muestras.

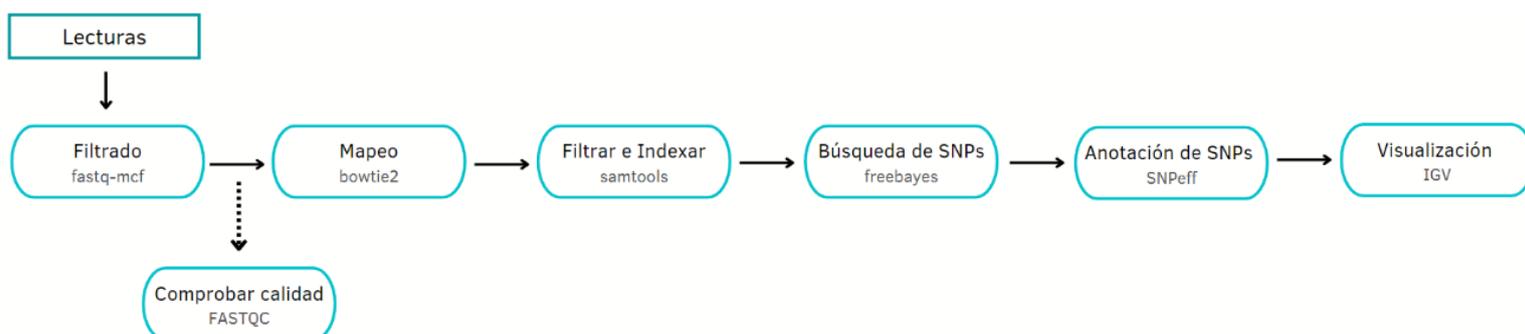


Figura 7: Diagrama de flujo del procesamiento informático de las lecturas obtenidas tras la resecuenciación.

A continuación, se detallan cada uno de los pasos.

3.6.1 Filtrado por calidad

Primeramente, las lecturas se procesaron con el paquete Ea-utils ([Aronesty, 2013](#)), utilizando la herramienta fastq-mcf, esta herramienta sirve para la eliminación de adaptadores Illumina presentes, para un filtrado según la calidad y para obtener las estadísticas de la resecuenciación (Q20, %GC). Las lecturas con una longitud menor a 50 pares de bases (-l) y con un Phred score menor a 20 (-q) fueron eliminadas. Se empleó el comando:

```
fastq-mcf -q 20 -l 50 -o readR.fq -o readL.fq
```

Para comprobar que este paso se había realizado de forma correcta y que las lecturas eran adecuadas para los siguientes pasos se utilizó el software FASTQC (versión 0.11.9), el cual es una herramienta de control de calidad para datos de secuencia de alto rendimiento ([Wingett and Andrews, 2018](#)). Se empleó el comando:

```
fastqc readR_Q30_L50.fq readL_Q30_L50.fq
```

3.6.2 Mapeo de lecturas

Una vez obtenidas las lecturas de alta calidad, estas se mapearon contra el genoma de referencia “67/3”, ensamblado combinando la tecnología de la secuenciación Illumina y el mapeo óptico ([Barchi et al., 2021](#)). Para el mapeo se empleó la versión 4.1 de este genoma, que es la versión más actualizada y se diferencia de la anterior en la reorientación de la secuencia de algunos cromosomas ([Solgenomics, 2022](#)). Para mapear las lecturas se utilizó el mapeador Bowtie2 (versión 2.4.5) ([Li, 2014](#)), este software se eligió ya que logra la eficiencia, la flexibilidad y velocidad de los algoritmos de programación logrando una elevada velocidad, sensibilidad y precisión ([Langmead and Salzberg, 2012](#)). El paquete SAMtools (versión 1.15.1) se utilizó para convertir los archivos salientes en archivos bam, para filtrar las lecturas no mapeadas y para obtener las estadísticas del mapeo ([Li et al., 2009](#)). Se utilizaron los parámetros por defecto para el mapeo, -p32 indica el número de procesadores que se utilizan, -x se utiliza para indicar el genoma de referencia, Illumina_Adapters.fa es una lista de adaptadores empleados por Illumina que el mapeador debe tener en cuenta, samtools view -bS se utiliza para generar un archivo bam como salida y samtools sort se usa para ordenar las lecturas. Se empleó el comando:

```
bowtie2 -p32 -x Indexed_eggplant_genome_V4_1  
Illumina_Adapters.fa -1 readR_Q30_L50.fq -2 readL_Q30_L50.fq |  
samtools view -bS - | samtools sort -o mapped_reads
```

En el caso de lecturas de *S. eleagnifolium*, al ser una especie más alejada filogenéticamente, al emplear los parámetros por defecto mapearon pocas lecturas. Por este motivo se ajustaron los parámetros de mapeo, se utilizó la opción `-very-sensitive-local`, de esta forma se relajaron los parámetros y se consiguió un mayor porcentaje de lecturas mapeadas ([Poullet et al., 2020](#)). Se empleó el comando:

```
bowtie2          -very-sensitive-local          -p32          -x
Indexed_eggplant_genome_V4_1.fa      Illumina_Adapters.fa      -1
readR_Q30_L50.fq -2 readL_Q30_L50.fq | samtools view -bS - |
samtools sort -o mapped_reads
```

Una vez mapeadas y ordenadas las lecturas, utilizando el paquete SAMtools para filtrar los mapeados empleando el comando:

```
samtools view -F4 mapped_reads.bam > mapped_sorted_reads.bam
```

y se indexaron con el comando:

```
samtools index mapped_filtered_reads.
```

El indexado de las lecturas es necesario para el último paso del diagrama donde se visualizarán los resultados ([Li et al., 2009](#)).

Para ver que todo el proceso de mapeado se había efectuado correctamente se obtuvieron las estadísticas, el número de lecturas, antes y después del filtrado, con el objetivo de asegurar que el mapeado había sido exitoso ([Li et al., 2009](#)). Se emplearon los comandos:

```
samtools view -c mapped_reads.bam y samtools view -c
mapped_filtered_reads.bam
```

3.6.3 Búsqueda y anotación de variantes

Una vez terminado el mapeado, se empezó la búsqueda de SNPs utilizando el software bayesiano FreeBayes (versión 1.3.6), este software está diseñado para buscar pequeños polimorfismos, SNPs (single-nucleotide polymorphisms), indels (inserciones y deleciones), MNPs (multi-nucleotide polymorphisms) y eventos más complejos (inserciones compuestas y eventos de sustitución) más pequeños que la longitud del tamaño de lectura. El software genera un archivo vcf como output. ([Garrison and Marth, 2012](#)). Se ajustaron los parámetros para buscar SNPs con una profundidad mínima de 5, `-C`, una calidad de mapeado mínima de 20, `-m` y una calidad de

base mínima de 20, -q. -f se utiliza para introducir el genoma de referencia y -b para generar el archivo de salida. Se empleó el comando:

```
freebayes -f Indexed_eggplant_genome_V4_1.fa -m 20 -q20 -C 5 -b mapped_filtered_reads.bam > freebayes_reads.vcf
```

Para la anotación del genoma se utilizó el software SNPeff (versión 5.1), el cual es un software que realiza la anotación de variantes y la predicción del efecto de estas variaciones, ofreciendo desde anotaciones muy simples, a que gen está afectando la variación, hasta muy complejas, si la variación no codificante afectará a la expresión de algún gen ([Cingolani et al., 2012](#)). -Xmx32g indica el número de procesadores a utilizar y -v el archivo de salida. Se empleó el comando:

```
java -Xmx32g -jar snpEff.jar -v freebayes_reads.vcf > snpEff_reads.vcf
```

3.6.4 Visualización

Para visualizar los resultados, se empleó el software IGV (Integrative Genomics Viewer) (versión 2.12.3), este software permite la visualización de toda la información obtenida anteriormente. Se visualizará la región del gen GLK2 para ver si hay algún polimorfismo candidato que explique las diferencias en la expresión del gen ([Robinson et al., 2020](#)), concretamente en el cromosoma 4 en la región 4017032-4020680. Se empleará el genoma de referencia de berenjena, descargado a través de la web ([Solgenomics, 2022](#)), un archivo gff3 con la anotación de los genes conocida, el archivo bam generado por samtools tras el filtrado y el archivo vcf generado por snpEff.

3.7 Extracción RNA

Para el estudio del gen GLK2 se extrajo el RNA de la piel de los parentales de las ILs, cogiendo alrededor de 75 mg de material vegetal. Se empleó el protocolo de extracción basado en EXTRAzol de la casa comercial EM300 (Blirt DNA) siguiendo las indicaciones del fabricante. El primer y el segundo paso fueron modificados, empleando el protocolo normal, las muestras no eran fragmentadas correctamente y al final de la extracción presentaban una elevada contaminación por fenoles. En vez de moler la muestra con el machacador Qiagen TissueLyser II, se utilizó un mortero y un mazo, previamente enfriados en el congelador, para moler las muestras empleando nitrógeno líquido, después se añadieron perlas de metal y se molieron durante tres minutos en el machacador Qiagen TissueLyser II. Los primeros pasos son críticos para la correcta extracción del RNA y hay que asegurar que el material vegetal no se descongele en ningún momento para evitar la degradación del RNA.

Se eligió este protocolo debido a que proporciona una rápida extracción de alta calidad del material genético y además el material extraído es ideal para procesos posteriores como síntesis de cDNA o para su utilización en una PCR.

También se realizó un gel de agarosa al 0,8% para comprobar la integridad y calidad de las extracciones. Las muestras se cargaron en el gel junto al tampón de carga Blue sin Red. La electroforesis corrió durante 20 minutos a 120V. Se realizó una tinción con bromuro de etidio y se visualizaron los resultados en un transiluminador de luz UV Gel Doc empleando el software Quantity One (BioRad). Se emplearon los marcadores moleculares Lambda *Hind*III y 100bp para determinar el tamaño de los fragmentos.

3.8 Síntesis cDNA

Primeramente, se estandarizó la cantidad de RNA de las muestras a 300 ng/ μ L. De esta forma, se facilitaron los cálculos posteriores del proceso de síntesis, simplificándolo.

A continuación, se realizó un tratamiento con DNasa (DNase I, RNase-free, Thermo Scientific) para eliminar el DNA genómica presente en la muestra y obtener el RNA aislado. Para el tratamiento se empleó 1 μ L de 10X reaction buffer con $MgCl_2$; 1 μ L de DNasa (1U); 4,67 μ L agua ultrapura y 3,33 μ L del RNA de la muestra (1 μ g). Se incubó a 37°C durante 30 minutos, se añadió 1 μ L 50 mM de EDTA (Ácido etilendiaminotetraacético) y se incubó a 65°C durante 10 minutos. El RNA se hidroliza durante el calentamiento con cationes divalentes en ausencia de agentes quelantes, este es el motivo de adición de EDTA. Alternativamente, también se puede realizar una extracción con fenol/cloroformo. Después de este tratamiento, se puede realizar una PCR con los genes de referencia. Si aparece una banda, es que hay DNA genómico remanente.

Finalmente, se realizó la síntesis de cDNA. Se empleó el protocolo basado en Fastgene Scriptase de la casa comercial Nippon Genetics Europe siguiendo las indicaciones del fabricante. En este paso es necesario emplear un control negativo para saber si la Scriptasa está contaminada, se utiliza una muestra con todos los componentes sustituyendo el material genético por agua. Se realiza una amplificación y si aparece alguna banda significa que el producto se encuentra contaminado.

3.9 Amplificación cDNA

Para la amplificación del cDNA se cambió el protocolo de amplificación, ya que empleando el protocolo descrito en el apartado 3.4 no ocurría la amplificación y al revelar el gel se observaba degradación en las muestras. Por eso se cambiaron los volúmenes iniciales de la reacción ([Tabla 4](#)) y el método de tinción del gel. Se utilizó 100 ng de cDNA para la amplificación. Para

comprobar que la retrotranscripción había funcionado y validar los resultados obtenidos amplificando con los otros cebadores, se empleó el gen de referencia CAC (Clathrin adaptor complex medium subunit) en todas las muestras ([Kanakachari et al., 2016](#)).

Tabla 4: Lista componentes empleados durante la amplificación del cDNA.

Reactivo	Reacción 12,5µL	Concentración Final
10 µM Forward Primer	0,24 µL	0,2 µM
10 µM Reverse Primer	0,24 µL	0,2 µM
Taq 2X Master Mix	5,98 µL	1X
Nuclease-free water	5,3 µL	-
Template DNA	1 µL	< 1000 ng

Una vez amplificadas las muestras, se realizó un gel de agarosa al 2% para comprobar que la amplificación había funcionado y que los fragmentos presentaban los tamaños esperados. Las muestras se cargaron en el gel junto al tampón de carga Blue sin Red. La electroforesis corrió durante 20 o 45 minutos a 120V. Se realizó una tinción con Midori Green y se visualizaron los resultados en un transiluminador de luz UV Gel Doc empleando el software Quantity One (BioRad). Se emplearon los marcadores moleculares Lambda *Hind*III y 100bp para determinar el tamaño de los fragmentos.

3.10 Amplificación de la delección

Para la amplificación de la delección se utilizó el protocolo descrito en el apartado 3.9, con las mismas concentraciones en los reactivos. Las condiciones de amplificación en el termociclador fueron: una desnaturalización inicial de 92°C, luego se realizaron 35 ciclos con una fase de desnaturalización a 92°C durante 15 segundos, una fase de alineamiento a 50°C durante 20 segundos y una fase de extensión a 72°C durante 2 minutos, finalmente se realizó una elongación final a 72°C durante 10 minutos. La visualización de las muestras se realizó según lo descrito en el apartado 3.9.

4. RESULTADOS Y DISCUSIÓN

4.1 Gen APRR2

Como se ha comentado la introducción, anteriormente se realizó un estudio GWAS a partir de la población MAGIC donde se detectó la asociación entre el gen APRR2 situado en el cromosoma 8 y la pigmentación con clorofila en el fruto. Estudios posteriores detectaron variantes de este gen entre diferentes variedades de *Solanum melongena*. En el genoma de referencia de Nakate-Shinkuro (Hirakawa et al., 2014) y GUIQIE-1 (Li et al., 2021), las cuales presentan pigmentación por clorofila, presentan una estructura del gen con 11 exones. Entre las variedades blancas, se encontraban las variedades A0416, IVIA y 67/3. El genoma de A0416 también presenta un gen APRR2 con 11 exones. Sin embargo, en el exón 3 presenta una variante que provoca un cambio en la pauta de lectura. El genoma de IVIA presenta 11 exones y también presenta una variante en el exón 6. Ambas variantes se asocian con la ausencia de pigmentación de clorofila en el fruto. Por último, el genoma de referencia 67/3 (Barchi et al., 2021) presentaba una estructura del gen APRR2 diferente, se detectó *in silico* la presencia de una gran deleción de los últimos 5 exones (Tabla 5).

Tabla 5: Lista de variantes presentes en los genomas de berenjena y el efecto que causan.

Genotipo	Exón 3	Exón 6	Exón 7-11	Fenotipo
Nakate-Shinkuro	TCTCCG	ACT	Completo	Morada con verde
GUIQIE-1	TCTCCG	ACT	Completo	Morada con verde
67/3	TCTCCG	ACT	Deleción	Morada sin verde
A0416	T---G	ACT	Completo	Blanca
IVIA	TCTCCG	A-T	Completo	Blanca listada

4.1.1 Validación de la deleción detectada *in silico*

Para validar la existencia de la deleción presente en algunos genomas de referencia que se había detectado *in silico*, se realizaron amplificaciones con distintas combinaciones de cebadores en tres muestras diferentes (Figura 8). Las tres muestras que se emplearon fueron: IVIA y A0416, las cuales son variedades de *Solanum melongena* que tienen un SNP en su genoma que hace que el gen APRR2 no se exprese, presentando la berenjena un color blanco. 67/3 es una variedad que presenta la deleción en su genoma afectando al gen APRR2 impidiendo su expresión, de forma que no presenta la pigmentación verde en la piel.

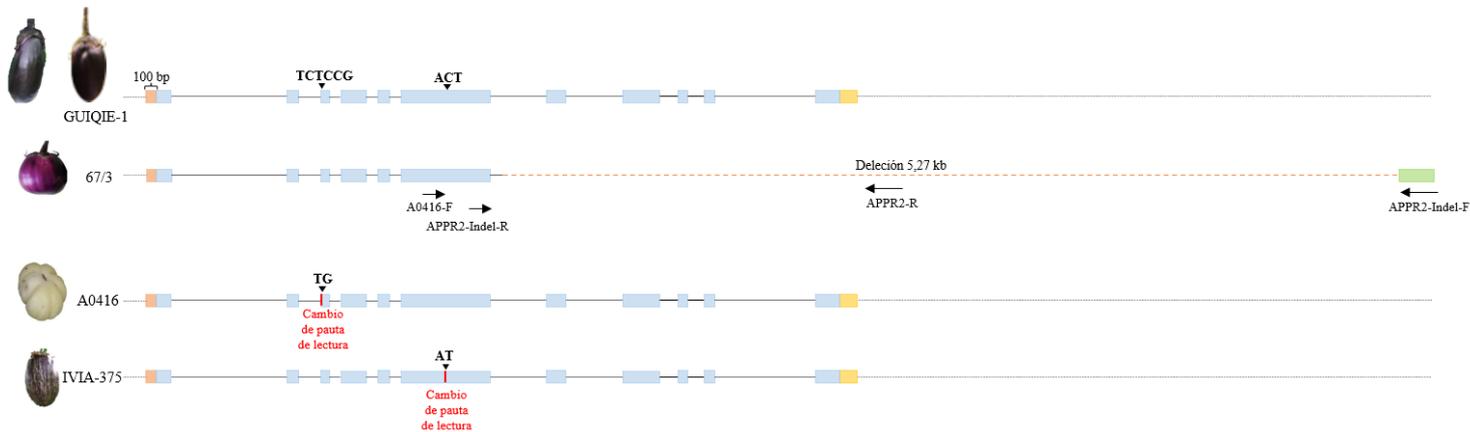


Figura 8: Esquema del genoma de las muestras empleadas para la validación de la detección detectada *in silico*.

La extracción del DNA de las muestras se realizó según el protocolo basado en CTAB-sílex (Vilanova et al., 2020). Una vez extraído el DNA, se intentó amplificar con cebadores situados en las regiones flanqueantes de la deleción (Figura 10, pocillo 2-4), con los cebadores Indel-R APRR2 y Indel-F APRR2, pero como el fragmento era demasiado grande, más de 5Kb, no amplificaba correctamente. Por esto, se emplearon distintas combinaciones de cebadores (Figura 9 y Figura 10).

En la figura 9, se emplearon la pareja de cebadores Indel-R APRR2 y APRR2-R. El cebador APRR2-R se encuentra dentro de la región delecionada en el genoma de 67/3. Por eso, en las muestras IVIA y A0416, pocillos 2-3, se observa que ha amplificado un fragmento de tamaño cercano a 3Kb que era lo esperado según el genoma anotado y en la muestra 67/3, pocillo 4, la cual presenta la deleción, no hay ninguna amplificación ya que no posee la región del genoma en la cual se une el cebador APRR2-R.

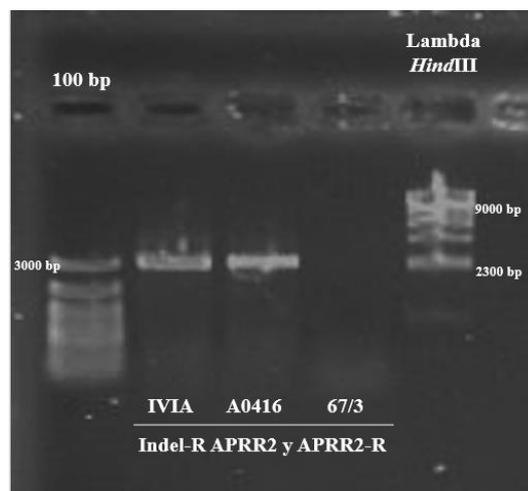


Figura 9: Amplificación empleando como marcadores moleculares 100bp situado en el primer pocillo y Lambda *HindIII* en el quinto. En los pocillos 2-4, se encuentran las muestras IVIA, A0416 y 67/3 amplificadas con la pareja de cebadores Indel-R APRR2 y APRR2-R.

En la [figura 10](#), se emplearon dos parejas diferentes de cebadores: como primera pareja: Indel-F APRR2 y Indel-R APRR2, pocillos 2-4. Esta combinación de cebadores no consiguió amplificar ninguna de las tres muestras. Como segunda pareja: Indel-F APRR2 y A0416-F en los pocillos 5-7. El cebador Indel-F APRR2 se encuentra en la región flanqueante de la deleción, mientras que el cebador A0416-F se encuentra aguas abajo del gen, en una posición más alejada. Por eso en las muestras IVIA y A0416, pocillos 5-6, el fragmento resultante era muy grande y no se observaba ningún fragmento durante la visualización del gel. Sin embargo, en el pocillo 7, que contenía la muestra 67/3, sí que se observa que ha amplificado un fragmento más pequeño, ya que, al presentar la deleción en el genoma, la región entre los cebadores es de menor tamaño y sí que se consigue una correcta amplificación.

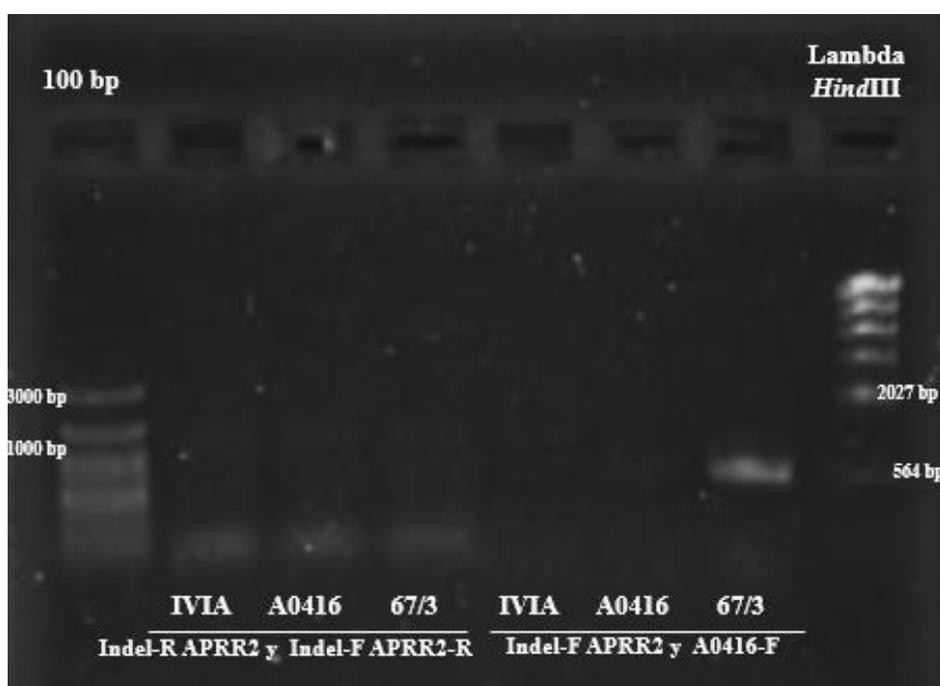


Figura 10: Amplificación empleando como marcadores moleculares 100bp situado en el primer pocillo y Lambda *Hind*III en el octavo. En los pocillos 2-4, se encuentran las muestras IVIA, A0416 y 67/3 amplificados con la pareja de cebadores Indel-R APRR2 y Indel-F APRR2-R. En los pocillos 5-7, se encuentran las muestras IVIA, A0416 y 67/3 amplificadas con la pareja de cebadores Indel-F APRR2 y A0416-F.

Como conclusión de estas primeras amplificaciones, obtuvimos dos parejas de cebadores que apoyan la existencia de la deleción en el genoma afectando al gen APRR2. La primera de ellas, Indel-R APRR2 y APRR2-R, en las muestras con SNP (IVIA y A0416) amplifica un fragmento con un tamaño aproximado de 1000pb y en la muestra con la deleción (67/3) no se produce la amplificación, ya que el cebador APRR2 no se puede unir. La segunda pareja, Indel-F APRR2 y A0416-F, en las muestras con SNP, al amplificar genera un fragmento demasiado grande y no se puede visualizar en el gen, mientras que en la muestra con la deleción (67/3) amplifica un fragmento de un tamaño aproximado de 800pb.

4.1.2 Comprobación de la delección en otras variedades

Una vez realizada las amplificaciones y encontrar parejas de cebadores que confirmaban la presencia de la delección, se procedió a comprobar si existe una relación entre la delección y la ausencia de pigmentación verde en el fruto. Para ello, se extrajo el DNA de distintas variedades de berenjena y se amplificaron con las parejas de cebadores que habían funcionado anteriormente. Para esta parte del experimento se escogieron las variedades presentadas en la [figura 11](#), se escogieron tanto variedades con pigmentación verde en la piel del fruto como otras que no la presentaban.



Figura 11: Variedades de la colección nuclear de berenjena establecida bajo el marco del proyecto europeo: G2P-SOL empleadas para la comprobación de los cebadores

Después de realizar la amplificación, se observó que las variedades verdes (muestras 6-8, 10-11) no presentan la delección, ya que amplifican con la primera combinación de cebadores, Indel-R APRR2 y APRR2-R, y con la segunda pareja, Indel-F APRR2 y A0416-F, no amplifican. Dentro de las variedades blancas (muestras 1-5 y 9), hay dos situaciones diferentes. La primera, la presentan las muestras 1, 2, 4 y 9, amplifican con la primera pareja, pero no con la segunda. Estas variedades no presentan la delección en su genoma, pero podrían tener otras variantes que les confiere el color blanco. La segunda, la presentan las muestras 3 y 5, las cuales no amplifican con la primera pareja, pero sí con la segunda. Tanto la muestras 3 como la 5 presentan un fruto blanco a causa de la delección en el gen APRR2 ([Figura 12](#)).

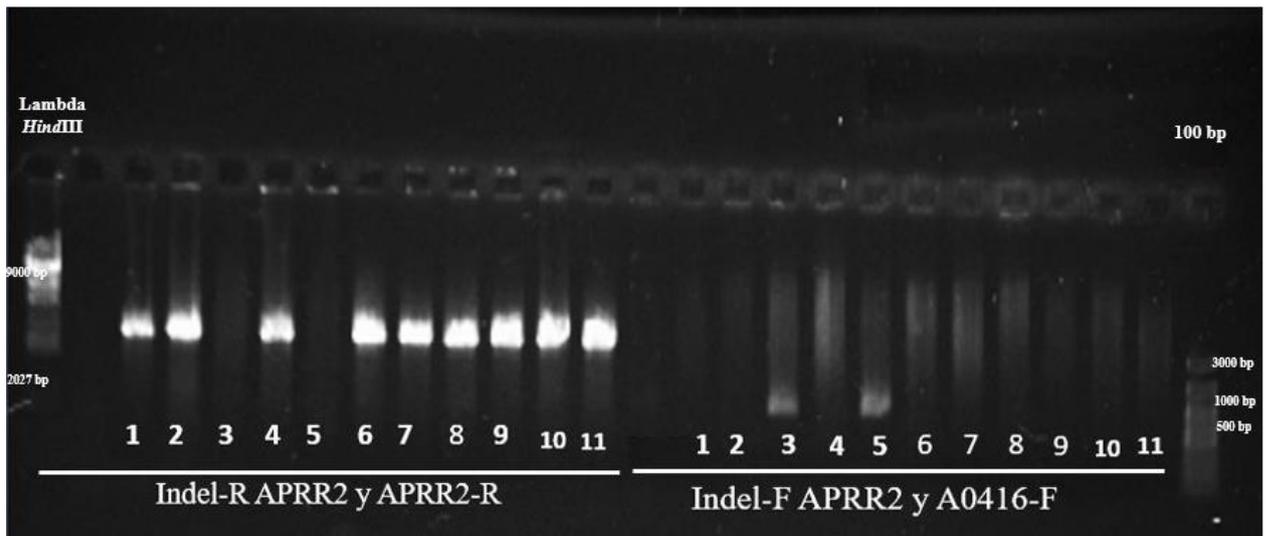


Figura 12: Amplificación empleando como marcadores moleculares Lambda *HindIII* situado en el primer pocillo y 100bp en el 26. En los pocillos 3-13, se encuentran las muestras mencionadas en la figura 11 amplificadas con la pareja de cebadores Indel-R APRR2 y Indel-F APRR2-R. En los pocillos 15-25, se encuentran las muestras IVIA, A0416 y 67/3 amplificadas con la pareja de cebadores Indel-F APRR2 y A0416-F.

Tras realizar todos los experimentos y observar los resultados, se puede ver una relación entre el gen APRR2 y la pigmentación de clorofila del fruto. Análisis posteriores del COMAV que llevaron a cabo una secuenciación de las variedades empleadas en la comprobación confirmaron que las variedades verdes presentaban una estructura del gen completo. Dentro de las variedades blancas se encontraban las que presentaban la delección y el resto de ellas se confirmó que presentaban o la variante descrita en IVIA o en A0416.. Se intuye una gran relación entre la estructura del gen APRR2 y la pigmentación verde. Esta relación también ha sido descrita en otros cultivos, como por ejemplo el pimiento, donde se encontró una fuerte asociación entre $C\alpha$ APRR2 y la acumulación de clorofila en frutos en desarrollo (Jeong et al., 2020). En tomate se ha asociado una sobreexpresión de este gen con un aumento en la producción de clorofila (Pan et al., 2013). También en melón y sandía tras realizar proyectos de secuenciación en poblaciones segregantes se ha detectado si los genes CmAPRR2 y ClAPRR2 sufren mutaciones se asocian a un color de cáscara más clara y se ha correlacionado la expresión de CmAPRR2 con genes vinculados al desarrollo de plástidos (Oren et al., 2019).

4.2 Gen GLK2

El GWAS también reveló que el gen GLK2 situado en el cromosoma 4 también estaba relacionado con la pigmentación verde del fruto, implicado con el netting. Aunque no se detectó ninguna variante que causara un gran impacto en la regulación y síntesis de clorofila. Por ello, se llevó a cabo un proyecto de resecuenciación para obtener más información de este gen.

4.2.1 Resecuenciación del genoma completo y mapeado

Para el análisis del gen GLK2, se trabajó con seis parentales de una población de ILs, presentando cuatro de ellos netting (MEL3, INS1, DAS y ELE) y dos de ellos no (MEL1 y MEL5) (Figura 13).

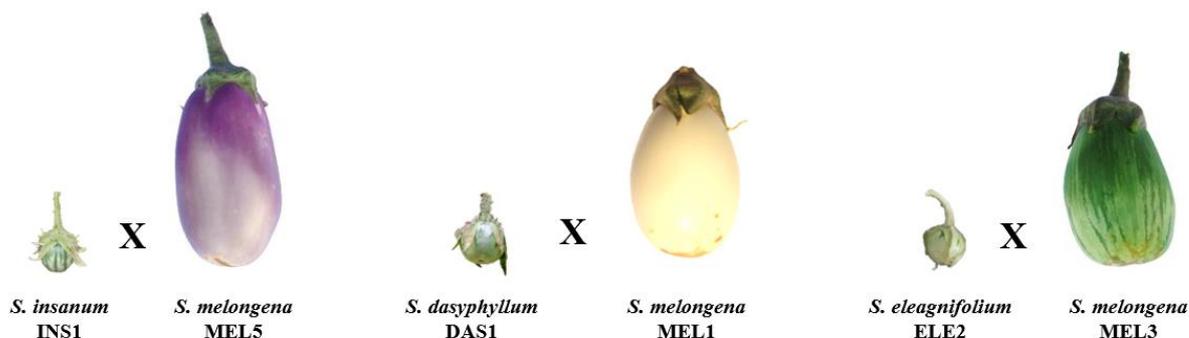


Figura 13: Parentales de las líneas ILs desarrolladas en el COMAV.

Gracias al trabajo previo del laboratorio, se partió con lecturas obtenidas de la reseuenciación de cada uno de los seis parentales. La reseuenciación de los seis parentales generó alrededor de 600 millones de lecturas de paire-ends. Una vez obtenidas las lecturas, se realizó el primer paso de filtrado para eliminar adaptadores y lecturas de mala calidad presentados en la tabla 6.

Tabla 6: Estadísticas de las lecturas tras el filtrado.

Muestra	Nº lecturas	Nº bases (millones)	Longitud lectura	Q20%	GC(%)
DAS	96.314.255	28.894	150	97,54	36,62
ELE	96.218.455	28.866	150	97,61	36,83
INS1	88.190.715	26.457	150	97,46	36,55
MEL1	96.225.770	28.868	150	97,46	36,38
MEL3	82.138.112	24.641	150	97,50	36,30
MEL5	96.185.459	28.856	150	97,64	36,64

Estas lecturas fueron mapeadas con bowtie2, versión 2.4.5 (Li, 2014) contra el genoma de referencia 67/3, versión 4.1 (Barchi et al., 2021), obteniendo todas las muestras una gran cantidad de lecturas mapeadas, con una media de 185092272 lecturas por muestra. Seguidamente, para ver el número de lecturas mapeadas correctamente, se utilizó el paquete de herramientas SAMtools (Li et al., 2009). El porcentaje de lecturas fue bastante elevado en casi todas las muestras, consiguiendo INS1, MEL1, MEL3 y MEL5 porcentajes casi del 100%. DAS y ELE al ser especies más alejadas de *S. melongena* obtuvieron un menor porcentaje de lecturas mapeadas casi del 80%. En el caso de ELE, se tuvo que ajustar los parámetros del mapeado para que fueran más permisivos y se consiguiera aumentar el número de lecturas mapeadas, pasando de un 36,15% de

lecturas a un 76,65% (Tabla 7). Tras el filtrado, se utilizó el software IGV (Robinson et al., 2020) para visualizar el archivo bam generado y comprobar que el genoma estaba bien mapeado y con una buena cobertura, Se observó con especial atención la región genómica del gen GLK2 situada en el cromosoma 4.

Tabla 7: Estadísticas tras el mapeo de las lecturas.

	DAS	ELE	INS1	MEL1	MEL3	MEL5
Lecturas mapeadas	192.628.510	192.436.910	176.381.430	192.451.540	164.276.224	192.379.018
Lecturas mapeadas correctamente	150.819.699	147.496.224	171.372.534	190.765.575	162.995.801	190.572.857
% lecturas	78,30	76,65	97,16	99,12	99,22	99,06

4.2.2 Variantes genómicas y distribución

Tras la búsqueda y la anotación de SNPs, se detectaron un total de 54 millones de polimorfismos entre las 6 muestras resecuenciadas, perteneciendo 3,95 millones al cromosoma 4. DAS, ELE, INS1 y MEL3 son los parentales que presentan una mayor cantidad de variaciones en el cromosoma 4, MEL1 y MEL5 presentan una menor cantidad. Dentro de las variaciones la mayor cantidad se localizaban en regiones intergénicas, seguido de la región aguas arriba y aguas debajo de los genes, habiendo muy poca cantidad de polimorfismos en los exones (Tabla 8).

Tabla 8: Estadísticas tras la anotación de variantes de las lecturas.

	DAS	ELE	INS1	MEL1	MEL3	MEL5
Variantes cromosoma 4	572.700	845.716	841.679	443.828	807.676	433.734
Aguas abajo	19,31%	20,43%	13,36%	15,37%	7,23%	12,84%
Exón	2,33%	4,39%	0,27%	0,14%	0,09%	0,08%
Intergénica	50,22%	40,25%	72,66%	69,24%	77,78%	73,99%
Intrón	8,66%	13,61%	2,69%	4,78%	2,11%	3,44%
Aguas arriba	18,23%	19,51%	10,91%	10,47%	12,78%	9,61%

Una vez anotadas todas las variaciones presentes en el cromosoma 4, se utilizó el software IGV (Robinson et al., 2020) para visualizar la región genómica del cromosoma 4 y ver si había alguna variante que provocara una diferencia en la expresión del gen GLK2 (SMEL4.1_04g003340.101) (Figuras 14, 15 y 16). En primer lugar, nos centramos en la búsqueda de variantes dentro del gen. En las muestras MEL1 y MEL5, las cuales no presentan netting, solamente se encontró un SNP en cada caso en la región de los intrones que no tenía ningún impacto en la expresión (Figura 14). En las muestras MEL3 y INS1, se detectaron un mayor número de variantes, 16 variantes en cada muestra, aunque todas estaban situadas en intrones y ninguna afectaba a los exones, sin causar ningún efecto significativo para la expresión del gen (Figura 15). En las muestras DAS y ELE, al ser especies más alejadas filogenéticamente de *Solanum melongena*, el número de variantes era

muy elevado, la gran mayoría afectaba a los intrones, pero había alguna en los exones. Sin embargo, ninguna de estas variaciones tenía un efecto significativo en el gen ([Figura 16](#)).

Tras el primer análisis del gen y no ver ninguna variante significativa que afectará al gen GLK2. Por eso, decidimos buscar en las regiones 5'UTR y 3'UTR del gen para ver si había alguna variante que pudiera afectar al gen, para ello analizamos estas regiones del gen ortólogo en tomate al estar este genoma mejor caracterizado que en berenjena. Se empleó la herramienta Jbrowse disponible en la web Solgenomics ([Solgenomics, 2022](#)) para visualizar el gen GLK2 en tomate y ver la secuencia de las regiones 5'UTR y 3'UTR, estas secuencias se utilizaron para hacer un blast contra el genoma de referencia de berenjena 67/3 versión 4.1 en la web Solgenomics ([Solgenomics, 2022](#)). El blast dio hits en el genoma de berenjena y la región genómica de estos hits se analizó en el visualizador IGV ([Robinson et al., 2020](#)). En la región 3'UTR, en ninguna de las seis muestras se observó alguna variante que pudiera causar una diferencia en la expresión del gen. Sin embargo, en la región 5'UTR sí que se observó diferencias entre las muestras que presentaban netting (MEL3, INS, DAS y ELE) y las que no (MEL1 y MEL5). Las muestras sin netting mapeaban correctamente contra el genoma de referencia y no presentaban ninguna variante ([Figura 14](#)), mientras que las variantes con netting presentaban varios polimorfismos e indels, siendo uno de estos bastante interesante. En la región 5'UTR, estas muestras presentaban una inserción de un tamaño aproximado de 300pb, en el caso de MEL3 y INS1 era de un tamaño menor en comparación con la delección presentan en DAS y ELE ([Figura 15](#) y [16](#)).

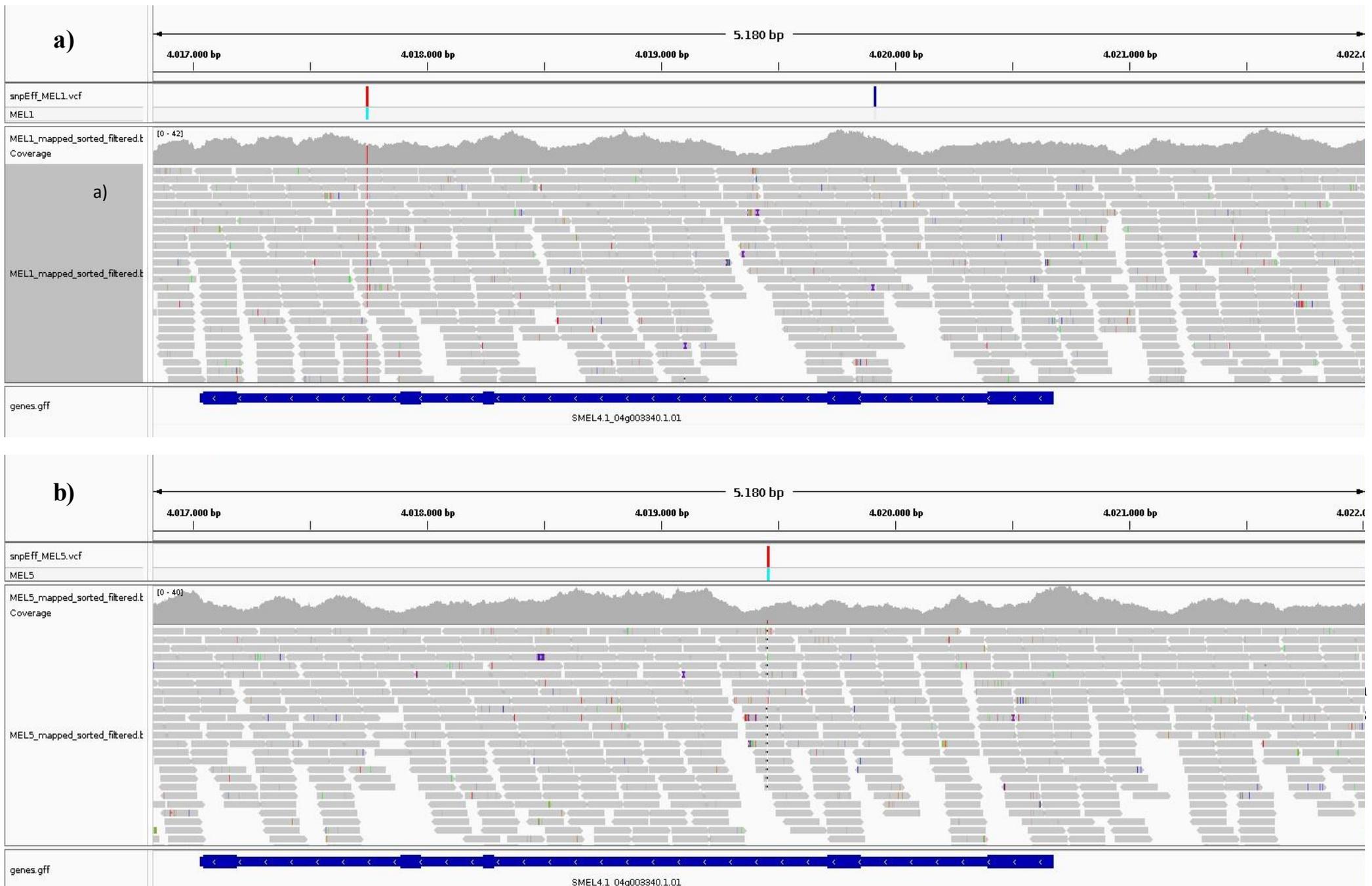


Figura 14: a) Visualización en IGV del archivo snpEff y del archivo bam del mapeado de la muestra MEL1 y del archivo gff3 de los genes de berenjena en la región genómica del gen GLK2 en el cromosoma 4. b) Visualización en IGV del archivo snpEff y del archivo bam del mapeado de la muestra MEL5 y del archivo gff3 de los genes de berenjena en la región genómica del gen GLK2 en el cromosoma 4

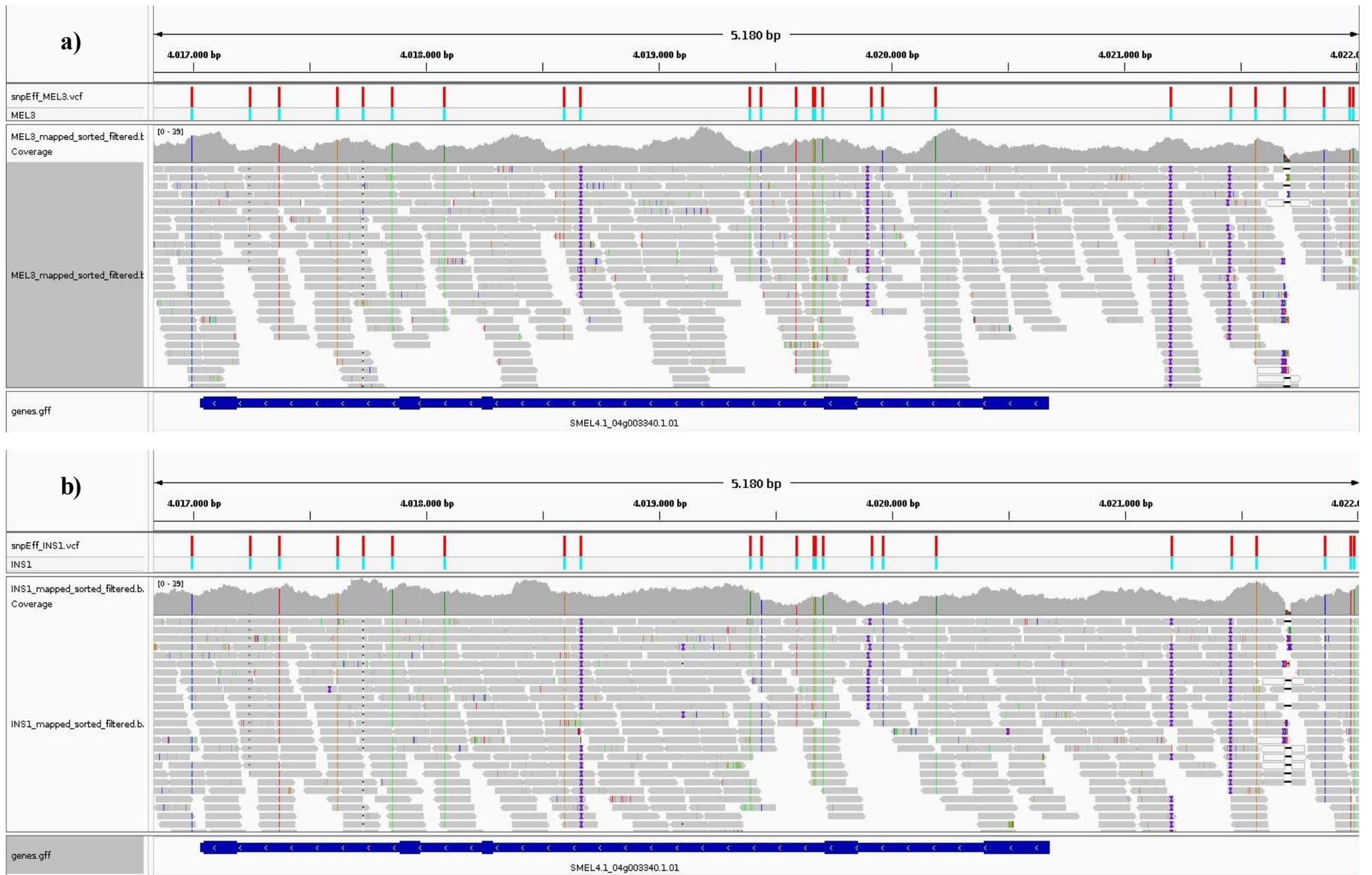


Figura 15: a) Visualización en IGV del archivo snpEff y del archivo bam del mapeado de la muestra MEL1 y del archivo gff3 de los genes de berenjena en la región genómica del gen GLK2 en el cromosoma 4. b) Visualización en IGV del archivo snpEff y del archivo bam del mapeado de la muestra MEL5 y del archivo gff3 de los genes de berenjena en la región genómica del gen GLK2 en el cromosoma 4.

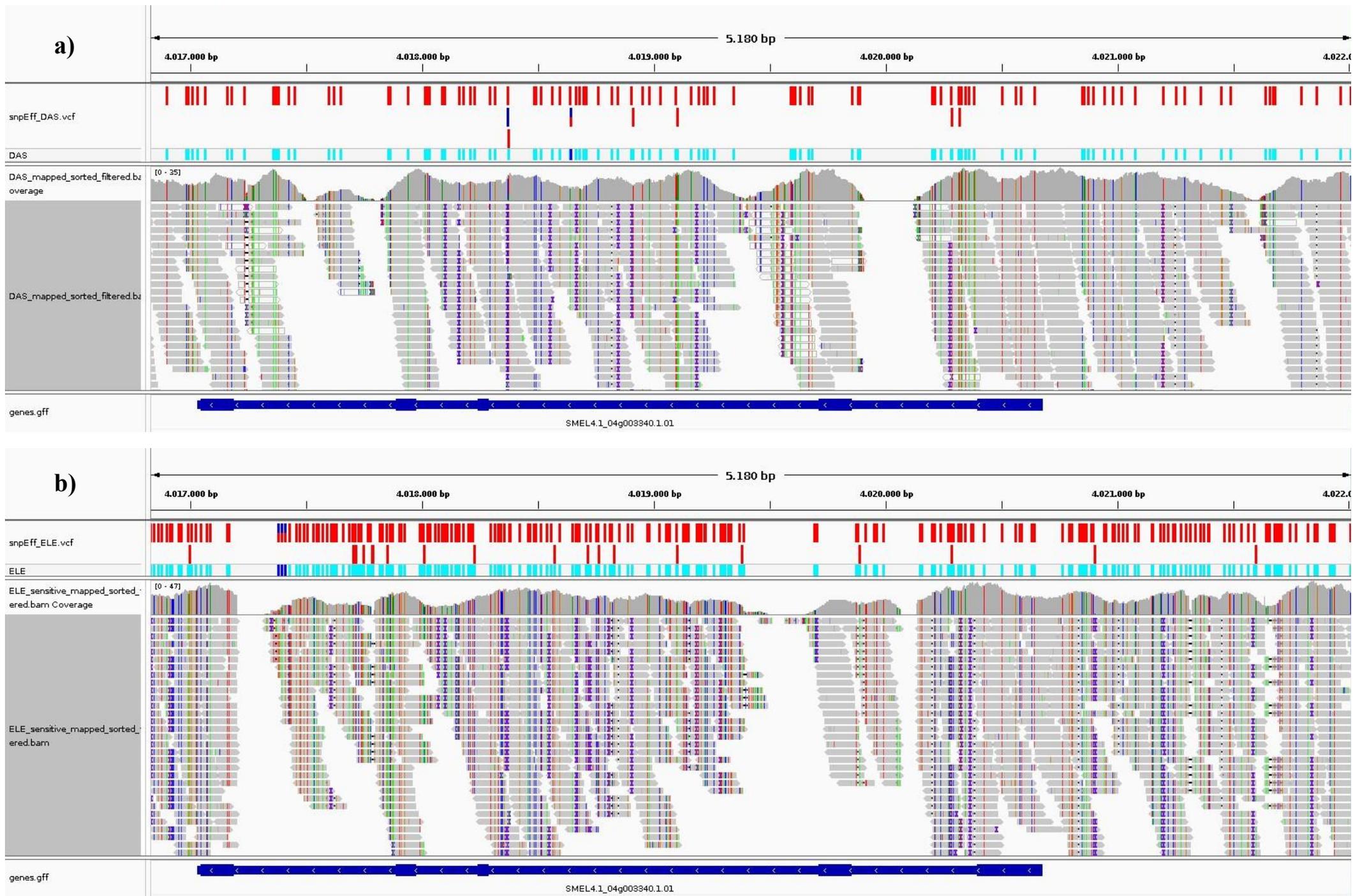


Figura 16: a) Visualización en IGV del archivo snpEff y del archivo bam del mapeado de la muestra DAS y del archivo gff3 de los genes de berenjena en la región genómica del gen GLK2 en el cromosoma 4. b) Visualización en IGV del archivo snpEff y del archivo bam del mapeado de la muestra ELE y del archivo gff3 de los genes de berenjena en la región genómica del gen GLK2 en el cromosoma 4.

Una vez obtenida toda la información del proceso de resecuenciación se llevaron a cabo dos experimentos: el primero para comprobar si hay una diferencia de expresión del gen GLK2 entre los parentales y el segundo para validar la existencia de la inserción detectada *in silico*.

4.2.3 Análisis de la diferencia de expresión entre parentales

Para comprobar si existe una diferencia de expresión del gen GLK2 entre las muestras con y sin netting dentro de los parentales de las poblaciones ILs. Se utilizaron dos réplicas de cada parental con netting: INS1, DAS, ELE y MEL3, y una réplica del parental sin netting: MEL1 (Figura 13).

Se extrajo el RNA de las muestras empleando el protocolo basado en EXTRAzol. Una vez extraído el RNA se retrotranscribió empleando el protocolo basado en Fastgene Scriptase. Para comprobar que todo el proceso se había realizado correctamente se realizó una amplificación del cDNA utilizando los cebadores del gen de referencia CAC (Kanakachari et al., 2016) (Figura 17). Se empleó un control negativo para confirmar que no se utilizó ningún reactivo contaminado durante el proceso de retrotranscripción y un control positivo, muestra DNA genómico, para poder validar los resultados. Al amplificar se puede ver que todas las muestras amplifican, presentando una banda del mismo tamaño.

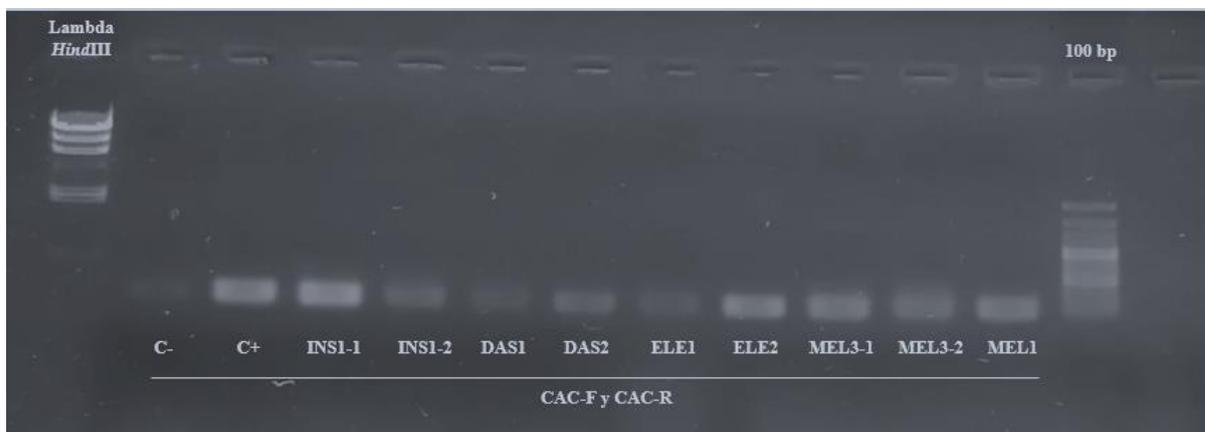


Figura 17: Amplificación empleando como marcadores moleculares Lambda *HindIII* situado en el primer pocillo y 100bp en el 13. En el pocillo se encuentra un control negativo, en el segundo se encuentra un control positivo, DNA genómico y en los pocillos 4-12, se encuentran las muestras amplificadas con la pareja de cebadores CAC-F y CAC-R.

Comprobado la calidad del cDNA, el siguiente que se realizó fue la amplificación del gen GLK2 en las muestras. Para ello, se empleó la pareja de cebadores GLK2 -F y GLK2-R (Figura 18). Al realizar la amplificación se utilizó un control negativo y dos muestras de DNA genómico como controles positivos. Todas las muestras amplificaron y mostraron una banda de aproximadamente 700 pb. Sin embargo, las muestras que presentaban netting netting (INS1-1, INS1-2, DAS1, DAS2, ELE1, ELE2, MEL3-1 y MEL3-2) tenían mayor intensidad que las muestras sin netting

(MEL1), viéndose la diferencia con mayor claridad en las muestras INS1-1 y DAS2, donde la intensidad es superior al resto de muestras. MEL1, la cual no presenta netting, tiene una banda de menor intensidad en comparación con el resto de las muestras.

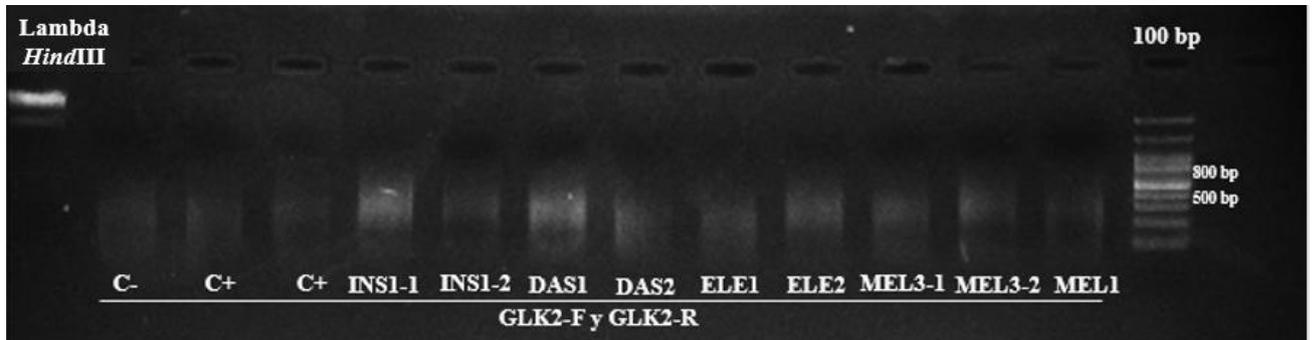


Figura 18: Amplificación empleando como marcadores moleculares Lambda *HindIII* situado en el primer pocillo y 100bp en el 14. En el pocillo se encuentra un control negativo, en el segundo se encuentra un control positivo, DNA genómico y en los pocillos 4-13, se encuentran las muestras amplificadas con la pareja de cebadores GLK2-F y GLK2-R.

Tras comprobar la amplificación se puede intuir que sí que hay diferencia de expresión entre las muestras con netting y las muestras sin netting. No obstante, todavía hay que optimizar el protocolo de amplificación para conseguir una visualización más clara de los resultados.

4.2.4 Validación de la inserción detectada *in silico*

Para validar la existencia de la inserción que se había detectado *in silico* en la región 5'UTR en el genoma de los parentales que presentaban netting, se diseñaron cebadores flanqueantes a la inserción mediante la web primer3 ([Kõressaar et al., 2018](#)). Los parentales sin netting (MEL1 y MEL5), que no presentan la inserción, amplifican un fragmento de 596 pb y los parentales con netting (INS1, DAS, ELE y MEL3), que presentan la inserción, amplifican un fragmento aproximadamente de 800 pb ([Figura 19](#)).

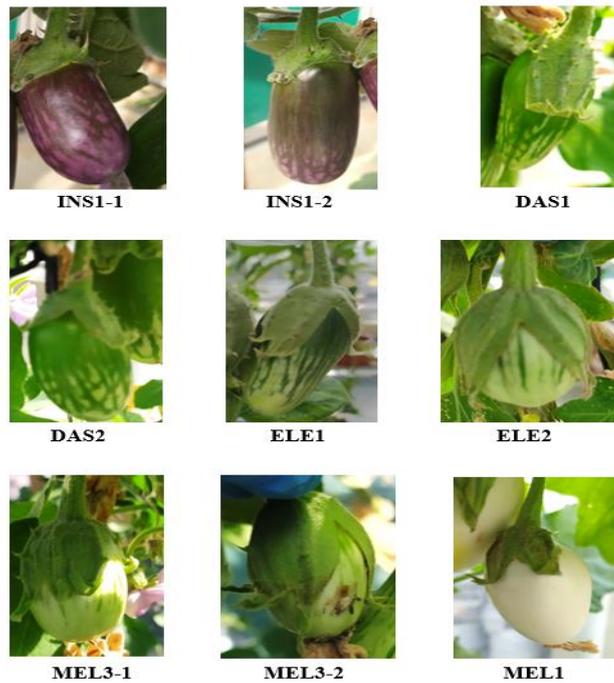


Figura 19: Muestras de los parentales de las poblaciones empleadas para la validación de la inserción.

Se utilizó el DNA de los parentales que fue extraído cuando se mandaron las muestras a resecuenciar. El DNA se amplificó con los cebadores GLK2-Indel-R y GLK2-Indel-F ([Figura 20](#)).

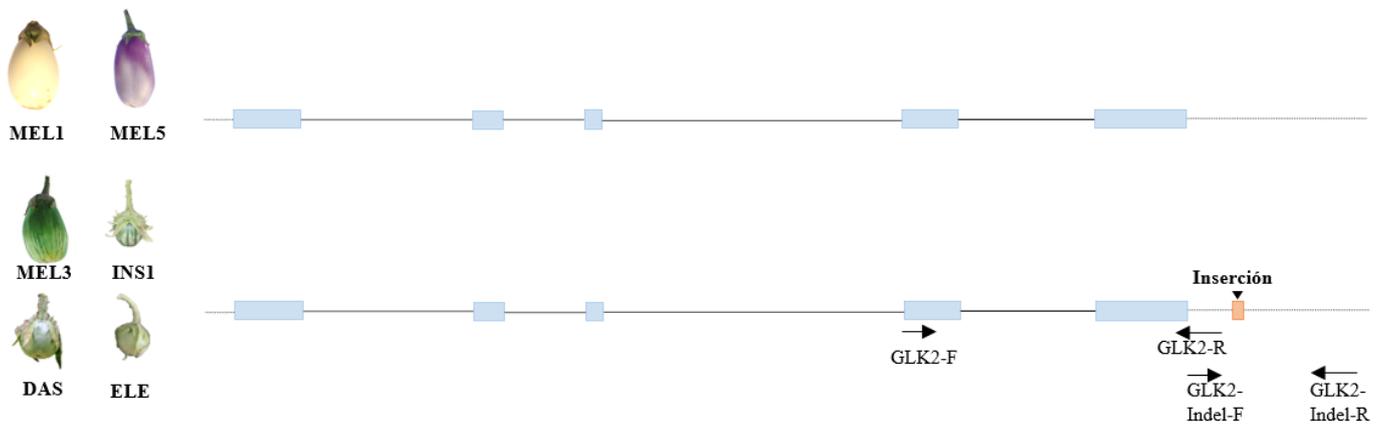


Figura 20: Esquema del genoma de las muestras empleadas para la validación de la detección detectada *in silico*.

Para una correcta visualización de las bandas se dejó correr las muestras durante 45 minutos en un gel de agarosa al 2%. Se utilizaron Lambda *Hind*III, pocillo 1, y 100bp, pocillo 8, como marcadores moleculares para conocer el tamaño de los fragmentos amplificados. Los parentales sin netting: MEL1 y MEL5, pocillos 2 y 3, que no presentan la inserción, amplificaron un fragmento cercano a 600 pb, lo cual coincidía con los resultados obtenidos tras el procesamiento de la

resecuenciación. Los parentales con netting: DAS y ELE, pocillos 6 y 7, que presentan la inserción, amplificaron un fragmento con un tamaño aproximado de 800 pb, coincidiendo con lo predicho. Sin embargo, los parentales MEL3 y INS1, los cuales presentan netting, amplificaron un fragmento con un tamaño cercano a 500 pb, presentando una delección respecto al genoma de referencia 67/3 (Barchi et al., 2021), siendo un resultado contrario a lo predicho (Figura 21). Se repitió la amplificación de estas muestras obteniendo los mismos resultados.



Figura 21: Amplificación empleando como marcadores moleculares Lamda *HindIII* situado en el primer pocillo y 100bp en el 8. En los pocillos 2-7, se encuentran los parentales amplificados con la pareja de cebadores GLK2-Indel-F y GLK2-Indel-R.

4.2.5 Comprobación de la inserción en otras variedades

Para comprobar la presencia de las variantes detectadas en los parentales con netting, tanto la inserción como la delección, se emplearon las variedades de berenjena descritas en la figura 11. El DNA de estas muestras fue amplificado con la pareja de cebadores GLK2-Indel-RF y GLK2-Indel-R. Después de realizar la amplificación, se observó que las variedades que no presentaban netting (muestras 1-5, 7, 9 y 10) amplificaban un fragmento de aproximadamente 600pb, coincidiendo con los resultados obtenidos en los parentales sin netting. Las muestras que presentaban netting (muestras 6, 8 y 11) amplificaron un fragmento con un tamaño cercano a 500pb coincidiendo con la delección que presentaban los parentales INS1 y MEL3.

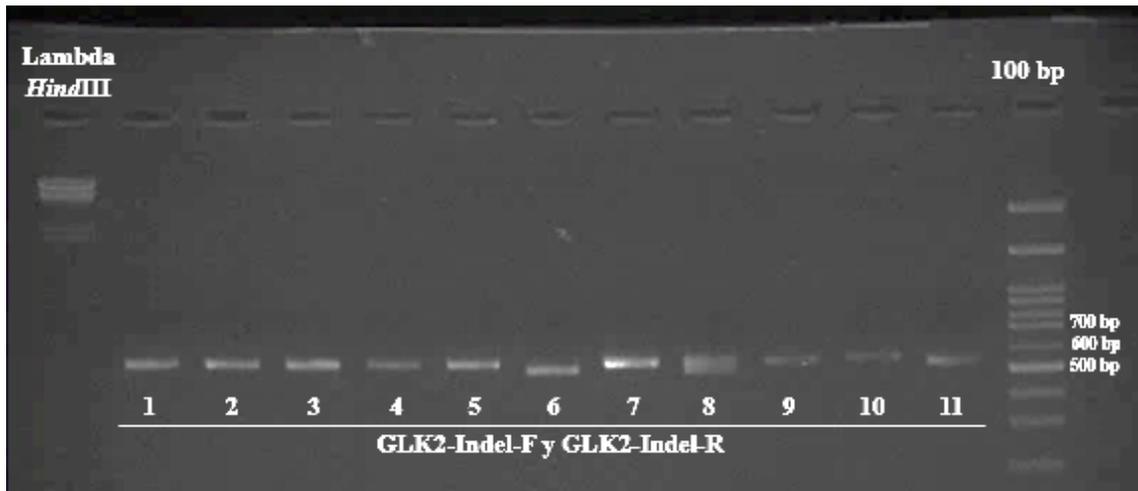


Figura 22: Amplificación empleando como marcadores moleculares Lambda *Hind*III situado en el primer pocillo y 100bp en el 13. En los pocillos 2-12, se encuentran las muestras mencionadas en la figura 11 amplificadas con la pareja de cebadores GLK2-Indel-F y GLK2-Indel-R.

Tras analizar los resultados obtenidos, nos dimos cuenta de que todas las muestras que no presentaban netting no tenían ninguna variante de gran importancia en comparación con el genoma de referencia 67/3 y dentro de las muestras con netting surgieron dos posibilidades: la primera, la inserción de alrededor 200pb, que coincidía con lo predicho tras el procesado de la resecuenciación, que se observó en los parentales DAS y ELE, pero no en ninguna de las muestras durante la comprobación. La segunda, una inserción se presentó en los parentales INS y MEL3, una delección de alrededor 100 bp que también aparecía en las muestras con netting durante la comprobación. No obstante, los resultados no son del todo precisos ya que se presentan diferencias entre lo predicho después del procesado de la resecuenciación y lo observado al amplificar, hay que seguir investigando y haciendo más pruebas de la zona candidata. Sin embargo, se puede intuir una posible relación en la expresión del netting cuando hay una presenta variante de la región 5'UTR del gen GLK2. En otros cultivos como el tomate se ha relacionado la pérdida de función de los genes GLK2 con la reducción en la acumulación de los productos fotosintéticos ([Bravo-García et al., 2009](#)).

5. CONCLUSIÓN

- La visualización de las accesiones permitió validar la existencia de la delección del gen APRR2 presente en las muestras con ausencia de pigmentación con clorofila en la piel del fruto, que luego fue confirmada con la amplificación de las muestras de la colección G2P-SOL. Una correcta estructura del gen APRR2 es necesaria para la correcta síntesis de clorofila.
- Tras el análisis de la resecuenciación de los parentales de las líneas de introgresión, se comprobó que el gen GLK2 no presenta ninguna variante de gran impacto. No obstante, en la región 5'UTR se identificó una inserción en las variantes con netting que era un candidato para la expresión del netting.
- La amplificación de la región 5'UTR, donde se identificó la inserción, comprobó que los parentales sin netting no presentaban la inserción en su genoma. Dentro de los parentales con netting, DAS y ELE confirmaron la presencia de la inserción en su genoma, mientras que MEL3 y INS1 amplificaron un tamaño menor del genoma de referencia, presentando una delección en su genoma. Análisis posteriores de esta región serán necesarios para confirmar los resultados.
- Tras la amplificación del cDNA, se comprobó una leve diferencia de expresión entre los parentales que presentaban netting y aquellos que no. Sin embargo, es necesario una mejor optimización de los protocolos para poder confirmar estos resultados.

6. BIBLIOGRAFÍA

Toppino, L., Barchi, L., Mercati, F., Acciarri, N., Perrone, D., Martina, M., ... & Rotino, G. L. (2020). A new intra-specific and high-resolution genetic map of eggplant based on a RIL population, and location of QTLs related to Plant anthocyanin pigmentation and seed vigour. *Genes*, *11*(7), 745.

Hirakawa, H., Shirasawa, K., Miyatake, K. O. J. I., Nunome, T., Negoro, S., Ohyama, A. K. I. O., ... & Fukuoka, H. (2014). Draft genome sequence of eggplant (*Solanum melongena* L.): the representative solanum species indigenous to the old world. *DNA research*, *21*(6), 649-660.

Kaunda, J. S., & Zhang, Y. J. (2019). The genus solanum: an ethnopharmacological, phytochemical and biological properties review. *Natural products and bioprospecting*, *9*(2), 77-137.

Plazas, M., Prohens, J., Cuñat, A. N., Vilanova, S., Gramazio, P., Herraiz, F. J., & Andújar, I. (2014). Reducing capacity, chlorogenic acid content and biological activity in a collection of scarlet (*Solanum aethiopicum*) and gboma (*S. macrocarpon*) eggplants. *International Journal of Molecular Sciences*, *15*(10), 17221-17241.

Ghatak, A., Chaturvedi, P., Paul, P., Agrawal, G. K., Rakwal, R., Kim, S. T., ... & Gupta, R. (2017). Proteomics survey of Solanaceae family: current status and challenges ahead. *Journal of proteomics*, *169*, 41-57.

Taher, D., Solberg, S. Ø., Prohens, J., Chou, Y. Y., Rakha, M., & Wu, T. H. (2017). World vegetable center eggplant collection: origin, composition, seed dissemination and utilization in breeding. *Frontiers in Plant Science*, 1484.

Velderrain-Rodríguez, G. R., Palafox-Carlos, H., Wall-Medrano, A., Ayala-Zavala, J. F., Chen, C. O., Robles-Sánchez, M., ... & González-Aguilar, G. A. (2014). Phenolic compounds: their journey after intake. *Food & function*, *5*(2), 189-197.

Elizalde-Romero, C. A., Montoya-Inzunza, L. A., Contreras-Angulo, L. A., Heredia, J. B., & Gutiérrez-Grijalva, E. P. (2021). *Solanum* fruits: phytochemicals, bioaccessibility and bioavailability, and their relationship with their health-promoting effects. *Frontiers in nutrition*, *8*.

Tridge. Available online at: <https://www.tridge.com/intelligences/eggplant> (Accessed June 15, 2022).

Hurtado, M., Vilanova, S., Plazas, M., Gramazio, P., Herraiz, F. J., Andújar, I., & Prohens, J. (2013). Phenomics of fruit shape in eggplant (*Solanum melongena* L.) using Tomato Analyzer software. *Scientia Horticulturae*, *164*, 625-632.

Gramazio, P., Yan, H., Hasing, T., Vilanova, S., Prohens, J., & Bombarely, A. (2019). Whole-genome resequencing of seven eggplant (*Solanum melongena*) and one wild relative (*S. incanum*) accessions provides new insights and breeding tools for eggplant enhancement. *Frontiers in plant science*, 1220.

Doganlar, S., Frary, A., Daunay, M. C., Lester, R. N., & Tanksley, S. D. (2002). A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the Solanaceae. *Genetics*, *161*(4), 1697-1711.

- Weese, T. L., & Bohs, L. (2010). Eggplant origins: out of Africa, into the Orient. *Taxon*, 59(1), 49-56.
- Frary, A., Doganlar, S., & Daunay, M. C. (2007). Eggplant. In *Vegetables* (pp. 287-313). Springer, Berlin, Heidelberg.
- Knapp, S., Vorontsova, M. S., & Prohens, J. (2013). Wild relatives of the eggplant (*Solanum melongena* L.: Solanaceae): new understanding of species names in a complex group. *PLoS one*, 8(2), e57039.
- Daunay, M. C., Lester, R. N., & Ano, G. (2001). Cultivated eggplants. In *tropical plant breeding* Edited by: Charpier A, Jacquot M, Hamon S, Nicolas D.
- Ranil, R. H. G., Prohens, J., Aubriot, X., Niran, H. M. L., Plazas, M., Fonseka, R. M., ... & Knapp, S. (2017). *Solanum insanum* L. (subgenus *Leptostemonum* Bitter, Solanaceae), the neglected wild progenitor of eggplant (*S. melongena* L.): a review of taxonomy, characteristics and uses aimed at its enhancement for improved eggplant breeding. *Genetic resources and crop evolution*, 64(7), 1707-1722.
- Vorontsova, M. S., Stern, S., Bohs, L., & Knapp, S. (2013). African spiny *Solanum* (subgenus *Leptostemonum*, Solanaceae): a thorny phylogenetic tangle. *Botanical Journal of the Linnean Society*, 173(2), 176-193.
- Adesina, S. K. (1985). Constituents of *Solanum dasyphyllum* fruit. *Journal of Natural Products*, 48(1), 147-147.
- Ilesanmi, O. B., Efe, O., Odewale, T. T., Atanu, F. O., Adeogun, E. F., Akinmoladun, A. C., & Olaleye, T. M. (2020). Reversal effect of *Solanum dasyphyllum* against rotenone-induced neurotoxicity.
- Mekki, M. (2007). Biology, distribution and impacts of silverleaf nightshade (*Solanum elaeagnifolium* Cav.). *EPPO bulletin*, 37(1), 114-118.
- Cunningham, D. C., Barry, S. C., Woldendorp, G., & Burgess, M. B. (2004). A framework for prioritizing sleeper weeds for eradication1. *Weed Technology*, 18(sp1), 1189-1193.
- Wei, Q., Wang, J., Wang, W., Hu, T., Hu, H., & Bao, C. (2020). A high-quality chromosome-level genome assembly reveals genetics for important traits in eggplant. *Horticulture research*, 7.
- Kevei, Z., King, R. C., Mohareb, F., Sergeant, M. J., Awan, S. Z., & Thompson, A. J. (2015). Resequencing at ≥ 40 -fold depth of the parental genomes of a *Solanum lycopersicum* \times *S. pimpinellifolium* recombinant inbred line population and characterization of frameshift InDels that are highly likely to perturb protein function. *G3: Genes, Genomes, Genetics*, 5(5), 971-981.
- Yang, X., Cheng, Y. F., Deng, C., Ma, Y., Wang, Z. W., Chen, X. H., & Xue, L. B. (2014). Comparative transcriptome analysis of eggplant (*Solanum melongena* L.) and turkey berry (*Solanum torvum* Sw.): phylogenomics and disease resistance analysis. *BMC genomics*, 15(1), 1-13.
- Daunay, M. C., Aubert, S., Frary, A., Doganlar, S., Lester, R. N., Barendse, G., ... & Jullian, E. (2004, May). Eggplant (*Solanum melongena*) fruit color: pigments, measurements and genetics. In *Proceedings of the 12th EUCARPIA Meeting on Genetics and Breeding of Capsicum and Eggplant* (pp. 108-116).

- Wei, Q., Wang, W., Hu, T., Hu, H., Wang, J., & Bao, C. (2020). Construction of a SNP-based genetic map using SLAF-Seq and QTL analysis of morphological traits in eggplant. *Frontiers in genetics*, 11, 178.
- Li, D., Qian, J., Li, W., Yu, N., Gan, G., Jiang, Y., ... & Wang, Y. (2021). A high-quality genome assembly of the eggplant provides insights into the molecular basis of disease resistance and chlorogenic acid synthesis. *Molecular Ecology Resources*, 21(4), 1274-1286.
- Smýkal, P., Nelson, M. N., Berger, J. D., & Von Wettberg, E. J. (2018). The impact of genetic changes during crop domestication. *Agronomy*, 8(7), 119.
- Carrillo, J. M. (2016). Erosión genética y reemplazo de variedades tradicionales. In *Las variedades locales en la mejora genética de las plantas* (pp. 29-41). Servicio Central de Publicaciones= Argitalpen Zerbitzu Nagusia.
- Sotomayor Cantos, I. A., Gramazio, P., Vilanova Navarro, S., Plazas Ávila, M. D. L. O., & Prohens Tomás, J. (2016). Fenotipado de líneas de introgresión de *Solanum incanum* en el fondo genético de la berenjena. *Revista ESPAMCIENCIA (Online)*, 7(2), 117-126.
- Bradshaw, J. E. (2017). Plant breeding: past, present and future. *Euphytica*, 213(3), 1-12.
- Zamir, D. (2001). Improving plant breeding with exotic genetic libraries. *Nature reviews genetics*, 2(12), 983-989.
- Prohens, J., Gramazio, P., Plazas, M., Dempewolf, H., Kilian, B., Diez, M. J., ... & Vilanova, S. (2017). Introgressomics: a new approach for using crop wild relatives in breeding for adaptation to climate change. *Euphytica*, 213(7), 1-19.
- Brand, A., Borovsky, Y., Hill, T., Rahman, K. A. A., Bellalou, A., Van Deynze, A., & Paran, I. (2014). CaGLK2 regulates natural variation of chlorophyll content and fruit color in pepper fruit. *Theoretical and Applied Genetics*, 127(10), 2139-2148.
- Li, F., Song, X., Wu, L., Chen, H., Liang, Y., & Zhang, Y. (2018). Heredities on fruit color and pigment content between green and purple fruits in tomato. *Scientia Horticulturae*, 235, 391-396.
- Wallace, T. C., & Giusti, M. M. (2015). Anthocyanins. *Advances in Nutrition*, 6(5), 620-622.
- Todaro, A., Cimino, F., Rapisarda, P., Catalano, A. E., Barbagallo, R. N., & Spagna, G. (2009). Recovery of anthocyanins from eggplant peel. *Food Chemistry*, 114(2), 434-439.
- Gallo, M., Naviglio, D., & Ferrara, L. (2014). Nasunin, an antioxidant anthocyanin from eggplant peels, as natural dye to avoid food allergies and intolerances. *European Scientific Journal*, 10(9).
- Basuny, A. M., Arafat, S. M., & El-Marzooq, M. A. (2012). Antioxidant and antihyperlipidemic activities of anthocyanins from eggplant peels. *Journal of Pharma Research & Reviews*, 2(3), 50-57.
- Senge, M. O., Ryan, A. A., Letchford, K. A., MacGowan, S. A., & Mielke, T. (2014). Chlorophylls, symmetry, chirality, and photosynthesis. *Symmetry*, 6(3), 781-843.
- Vallejo, W. A., Diaz, C., Alvis, M., Cantillo, A., & Fajardo, C. (2018). Methylene blue degradation under visible irradiation on TiO₂ electrodes sensitized by dye chlorophyll extract from *Spinacia Olerá*/Degradación de azul de metileno bajo irradiación visible sobre electrodos de TiO₂ sensibilizados con pigmentos de clorofila. *Prospectiva*, 16(2), 7-12.

- Foyer, C. H. (2018). Reactive oxygen species, oxidative signaling and the regulation of photosynthesis. *Environmental and experimental botany*, 154, 134-142.
- Tanaka, A., & Tanaka, R. (2006). Chlorophyll metabolism. *Current opinion in plant biology*, 9(3), 248-255.
- Jia, T., Cheng, Y., Khan, I., Zhao, X., Gu, T., & Hu, X. (2020). Progress on understanding transcriptional regulation of chloroplast development in fleshy fruit. *International Journal of Molecular Sciences*, 21(18), 6951.
- Yang, G., Li, L., Wei, M., Li, J., & Yang, F. (2022). Smmyb113 is a key transcription factor responsible for compositional variation of anthocyanin and color diversity among eggplant peels. *Frontiers in Plant Science*, 13.
- String. Disponible en línea: <https://string-db.org> (Consultado Junio 7, 2022).
- Pan, Y., Bradley, G., Pyke, K., Ball, G., Lu, C., Fray, R., ... & Seymour, G. B. (2013). Network inference analysis identifies an APRR2-like gene linked to pigment accumulation in tomato and pepper fruits. *Plant Physiology*, 161(3), 1476-1485.
- Jeong, H. B., Jang, S. J., Kang, M. Y., Kim, S., Kwon, J. K., & Kang, B. C. (2020). Candidate gene analysis reveals that the fruit color locus C1 corresponds to PRR2 in pepper (*Capsicum frutescens*). *Frontiers in plant science*, 11, 399.
- Li, G., Chen, D., Tang, X., & Liu, Y. (2018). Heterologous expression of kiwifruit (*Actinidia chinensis*) GOLDEN2-LIKE homolog elevates chloroplast level and nutritional quality in tomato (*Solanum lycopersicum*). *Planta*, 247(6), 1351-1362.
- Lupi, A. C. D., Lira, B. S., Gramegna, G., Trench, B., Alves, F. R. R., Demarco, D., ... & Rossi, M. (2019). *Solanum lycopersicum* GOLDEN 2-LIKE 2 transcription factor affects fruit quality in a light-and auxin-dependent manner. *PLoS One*, 14(2), e0212224.
- Siddiqui, H., Hayat, S., & Bajguz, A. (2018). Regulation of photosynthesis by brassinosteroids in plants. *Acta Physiologiae Plantarum*, 40(3), 1-15.
- Bravo-Garcia, A., Yasumura, Y., & Langdale, J. A. (2009). Specialization of the Golden2-like regulatory pathway during land plant evolution. *New Phytologist*, 183(1), 133-141.
- Vilanova, S., Alonso, D., Gramazio, P., Plazas, M., García-Forte, E., Ferrante, P., ... & Prohens, J. (2020). SILEX: a fast and inexpensive high-quality DNA extraction method suitable for multiple sequencing platforms and recalcitrant plant species. *Plant Methods*, 16(1), 1-11.
- Köressaar, T., Lepamets, M., Kaplinski, L., Raime, K., Andreson, R., & Remm, M. (2018). Primer3_masker: integrating masking of template sequence with primer design software. *Bioinformatics*, 34(11), 1937-1938.
- Singh, V. K., & Kumar, A. (2001). PCR primer design. *Mol. Biol. Today*, 2(2), 27-32.
- Aronesty, E. (2013). Comparison of sequencing utility programs. *The open bioinformatics journal*, 7(1).
- Wingett, S. W., & Andrews, S. (2018). FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research*, 7.

Barchi, L., Rabanus-Wallace, M. T., Prohens, J., Toppino, L., Padmarasu, S., Portis, E., ... & Giuliano, G. (2021). Improved genome assembly and pan-genome provide key insights into eggplant domestication and breeding. *The Plant Journal*, *107*(2), 579-596.

Solgenomics. Disponible en línea: <https://solgenomics.net> (Consultado Junio 7, 2022).

Li, H. (2014). Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics*, *30*(20), 2843-2851.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, *9*(4), 357-359.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079.

Poulet, M., & Orlando, L. (2020). Assessing DNA sequence alignment methods for characterizing ancient genomes and methylomes. *Frontiers in Ecology and Evolution*, *8*, 105.

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907*.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., ... & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, *6*(2), 80-92.

Robinson, J. T., Thorvaldsdóttir, H., Turner, D., & Mesirov, J. P. (2020). igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *bioRxiv*.

Kanakachari, M., Solanke, A. U., Prabhakaran, N., Ahmad, I., Dhandapani, G., Jayabalan, N., & Kumar, P. A. (2016). Evaluation of suitable reference genes for normalization of qPCR gene expression studies in brinjal (*Solanum melongena* L.) during fruit developmental stages. *Applied biochemistry and biotechnology*, *178*(3), 433-450.

Oren, E., Tzuri, G., Vexler, L., Dafna, A., Meir, A., Faigenboim, A., ... & Gur, A. (2019). The multi-allelic APRR2 gene is associated with fruit pigment accumulation in melon and watermelon. *Journal of experimental botany*, *70*(15), 3781-3794.