

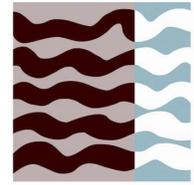


Universidad
Politécnica
de Cartagena



UPCT

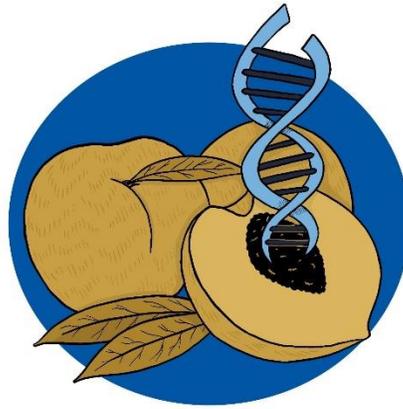
Escuela Técnica Superior de
Ingeniería Agronómica



ETSIA

*Grado en Ingeniería Agroalimentaria
y de Sistemas Biológicos*

**USO DE TÉCNICAS DE GENOTIPADO
EN EL MANEJO Y MANTENIMIENTO DE
COLECCIONES DE GERMOPLASMA
PARA LA MEJORA GENÉTICA DEL
MELOCOTONERO**



Autora: Dña María del Mar Gómez
Abajo

Dirección: Dr Juan Esteva Pascual

Codirección: Dr Pedro José Martínez García

Cartagena, 5 de diciembre de 2020

La investigación desarrollada en el presente trabajo está encuadrada en el proyecto “*PostPeachBreed*” (RTI2018-094176-R) de la comunidad Autónoma de la Región de Murcia. Ha sido desarrollado por el departamento de Mejora Genética de Frutales del Centro de Edafología y Biología Aplicada del Segura, CEBAS-CSIC, bajo la dirección de Pedro José Martínez García. Es un proyecto coordinado con 3 subproyectos:

- **SUBPROYECTO 1. Coordinador:** Celia M. Cantín. **Nombre:** VARIABILIDAD EN FISIOPATÍAS POSTCOSECHA EN DOS COLECCIONES NACIONALES DE GERMOPLASMA DE MELOCOTONERO Y PATRONES PRUNUS DE NUEVA GENERACIÓN (RTI2018-094176-R-C31).
- **SUBPROYECTO 2. Coordinador:** Pedro J. Martínez García. **Nombre:** ANÁLISIS DE LA VARIACIÓN Y DEL CONTROL GENÉTICO DE LAS FISIOPATÍAS POSTCOSECHA EN MELOCOTONERO USANDO HERRAMIENTAS GENÓMICAS (RTI2018-094176-R-32).
- **SUBPROYECTO 3. Coordinador:** M. Ángeles Moreno. **Nombre:** MEJORA DE PATRONES PRUNUS Y VARIEDADES DE MELOCOTONERO, ENFOCADO EN ENFERMEDADES POSTCOSECHA (RTI2018-094176-R-C33).

Declaración de Honestidad Académica

La alumna Dña. **María del Mar**, con DNI **48655410Z**,

como autora del TFE de título **Uso de Técnicas de Genotipado en el Manejo y Mantenimiento de Colecciones de Germoplasma para la Mejora Genética del Melocotonero**

dirigido por D. **Juan Esteva Pascual**

para la obtención del título

- Grado en Ingeniería Agroalimentaria y de Sistemas Biológicos
- Máster Universitario en Ingeniería Agronómica
- Máster Universitario en Técnicas Avanzadas en Investigación y Desarrollo Agrario y Alimentario

DECLARA:

- Que el mencionado TFE es íntegramente de su autoría.
- Que se trata de un trabajo original e inédito en el que no existe plagio.
- Que en todo momento se respeta la propiedad intelectual y en ningún caso se han utilizado como propios resultados ni materiales obtenidos o generados por otros autores.
- Que los resultados y materiales realizados por otros autores han sido debidamente identificados en la memoria.
- Que se ha aplicado al texto íntegro del TFE el control antiplagio que establece la *Normativa de Trabajos Fin de Estudios en la ETSIA*, y acompaña esta declaración de las páginas primera y última del informe obtenido de Turnitin a través de Aul@Virtual.
- Qué los directores del TFE conocen y han dado el visto bueno a los resultados del control antiplagio y, en su caso, han informado en la forma que indica el documento *Política de Calidad y Código de Buenas Prácticas*.

Y para que así conste, firma la presente declaración en,

Cartagena, a 5 de diciembre de 2020

Fdo. **María del Mar Gómez Abajo**

Agradecimientos

Me gustaría agradecer en primer lugar a mi tutor en el CEBAS, Pedro José Martínez García, por su paciencia e interés y por todo lo que me ha enseñado, tanto en las prácticas como durante la realización de este trabajo. Sin su ayuda y sin todos los conocimientos que me ha aportado con gran dedicación, este trabajo no habría sido posible.

A todo el grupo de trabajo del CEBAS por la buena acogida durante mi estancia allí, siempre otorgándome todos los recursos que he necesitado para aprender y ayudándome a aplicar todos los conocimientos adquiridos. En especial a Federico Dicenta, por darme esta oportunidad y a Jorge y a Ángela por su ayuda y por su tiempo.

A mi tutor Juan Esteva Pascual, por darme la oportunidad de realizar este trabajo y por fijar la base de todos los conocimientos adquiridos durante la realización de este.

A mi familia, por apoyarme siempre y sobre todo durante esta etapa académica, especialmente a mi hermana Coté, por sus buenos consejos.

Por último, agradecer a mis amigos de toda la vida, a todos los que me llevo de la Residencia Alberto Colao y en especial a los de la carrera, Anaís, Begoña, Diego y Jose, vosotros me habéis dado el apoyo y la energía necesaria para afrontar tan bien esta fase.

Índice de Contenido

1. Resumen y Abstract	1
2. Objetivos.....	2
3. Introducción	3
3.1. Importancia del cultivo del melocotonero.....	3
3.1.1. <i>Origen y expansión del cultivo de melocotonero</i>	3
3.1.2. <i>Usos del melocotón</i>	4
3.2. Superficie y producción en España	5
3.3. Mejora genética del melocotonero.....	6
3.3.1. <i>Grupos pomológicos</i>	6
3.3.2. <i>Programas de mejora genética del melocotonero</i>	7
3.4. Principales problemas postcosecha	7
3.5. Importancia y conservación de la diversidad genética	9
3.5.1. <i>Herramientas para estudiar la diversidad genética: Marcadores moleculares</i> .	10
3.5.2. <i>Herramientas para estudiar la diversidad genética: Programas estadísticos</i> ...	15
4. Material y Métodos.....	18
4.1. Material Vegetal	18
4.2. Extracción de ADN	21
4.3. Medidas de calidad del ADN	22
4.4. Chip de genotipado y procesamiento de datos	22
4.4.1. <i>Chip de genotipado</i>	23
4.4.2. <i>Procesamiento de datos</i>	23
4.4.3. <i>Análisis de diversidad genética</i>	26
4.4.4. <i>Relaciones de parentesco</i>	28
5. Resultados y Discusión.....	30
5.1. Resultados de la extracción y medida de calidad del ADN	31
5.2. Resultados obtenidos con GenomeStudio.....	31
5.3. Resultados obtenidos con ASSIsT	32
5.4. Estudio de la diversidad genética	33
5.5. Relaciones de parentesco identificadas.....	37

5.6. Resultados obtenidos con Structure	40
6. Conclusiones	43
6.1. Conclusiones Generales	43
6.2. Futuros pasos y posibles líneas de investigación	43
7. Referencias.....	44
Anexo A. Programas de Mejora Genética en España	49
Anexo B. <i>Workflow</i> de las diferentes Tareas realizadas en este Trabajo	53
Anexo C. Clasificación de las Variedades según su Origen.....	54
Anexo D. Frutos de algunas de las variedades presentes en este trabajo	59
Anexo E. Procedimiento y Material necesario para la Extracción de ADN	62
Anexo F. Resultados obtenidos en <i>NanoDrop</i> y <i>Qubit</i>	64
Anexo G. <i>Call Rate</i> obtenido en <i>GenomeStudio</i>	68
Anexo H. Resultados obtenidos en <i>PLINK</i>	73
Anexo I. Resultados de <i>fastStructure</i>	79
Anexo J. Resultados obtenidos en <i>Structure</i>	81

Índice de Tablas

Tabla 1: Destino y producción del Melocotón en España	4
Tabla 2: Relación de Parentesco Padre-Hijo según los Parámetros de PLINK.....	29
Tabla 3: N.º de Variedades según su Origen.....	30
Tabla 4: Clasificación de los SNPs en PLINK.....	33
Tabla 5: Medidas de Diversidad Genética según el Origen de los Individuos.....	34
Tabla 6: Datos Obtenidos de Fst en función del Origen	35
Tabla 7: Datos Obtenidos de Gst en función del Origen.....	36
Tabla 8: Medidas de Diferenciación Absoluta en función del Origen	36
Tabla 9: Programas de Mejora Genética en España	49
Tabla 10: Clasificación de las Variedades según su Código y Origen	54
Tabla 11: Resultados de las Medidas tomadas con Qubit y Nanodrop.....	64
Tabla 12: CallRate obtenido en GenomeStudio.....	68
Tabla 13: Sinonimias obtenidas en PLINK	73
Tabla 14: Grupos de Sinonimias	76
Tabla 15: Probabilidad Marginal de cada Clúster	80
Tabla 16: Clúster y N.º Asociado en Structure para cada Variedad.....	81

Índice de Figuras

Figura 1: Producción de Melocotón y Nectarina en España (2008-2018).....	5
Figura 2: Superficie de Melocotón y Nectarina en España (2008-2018).....	6
Figura 3: Clasificación de los Marcadores Moleculares.....	11
Figura 4: Ejemplo de SNP en 2 cadenas de Nucleótidos	12
Figura 5: Detección de los SNPs mediante Fluorescencia.....	13
Figura 6: Gráfica de GenomeStudio para un determinado SNP	14
Figura 7: Representación Gráfica de un Clúster (K=3).....	17
Figura 8: Orígenes de las Variedades Españolas presentes en el Trabajo	18
Figura 9: CallRate de los SNPs	32
Figura 19: Workflow de las Tareas realizadas en este Trabajo	53
Figura 10: Variedad La Escola.....	59
Figura 11: Variedad Maluenda 2375	59
Figura 12: Variedad Rojo del Rito	60
Figura 13: Variedad Paraguayo Jota.....	60
Figura 14: Variedad Sudanell 2.....	61
Figura 15: Variedad Miraflores.....	61
Figura 16: Gráfico de Probabilidad Marginal	79
Figura 17: Representación Gráfica de cada Clúster.....	85
Figura 18: Representación Gráfica de cada Clúster con Orígenes.....	86

1. Resumen y Abstract

Resumen

La conservación de colecciones vegetales en los bancos de germoplasma, es una labor importante para asegurar el mantenimiento de la diversidad genética. En el presente trabajo, se ha realizado el genotipado de 90 accesiones de la colección de melocotonero del banco de germoplasma de Zaragoza. Estas accesiones corresponden a material autóctono cuyas características pueden resultar útiles para la mejora genética de la especie. El genotipado de estas accesiones, también ha permitido encontrar errores en el etiquetado de las muestras y distinguir individuos duplicados, optimizando dicha colección y facilitando el trabajo del mejorador. Otro de los objetivos de este trabajo es validar el funcionamiento de la nueva versión del chip de genotipado de melocotonero *Illumina Infinium BeadChip 18K* (de 18000 SNPs). Esta nueva versión del chip de melocotonero, no ha sido utilizada en otros trabajos de investigación, lo que recalca la novedad de los resultados obtenidos. La utilización de este chip, no solo ha servido para establecer las relaciones de parentesco de las diferentes accesiones, sino también para el análisis de la diversidad genética. Esto ha permitido conocer la estructura poblacional, destacando las variedades murcianas, cuyo fondo genético parece diferente al resto de accesiones utilizadas. Esto sugiere un posible intercambio de material genético solamente entre las zonas del norte de España. La información obtenida en este trabajo será de utilidad para para la obtención de nuevas variedades de melocotonero resistentes a enfermedades emergentes cada año más problemáticas como Corky Spot.

Palabras clave: melocotón, banco de germoplasma, chip de genotipado, SNP, diversidad genética, fisiopatías, *Illumina Infinium BeadChip 18K*, material autóctono.

Abstract

Conservation of plant collections in germplasm banks is important to ensure the maintenance of genetic diversity. In this work, the genotyping of 90 accessions of the peach collection of the germplasm bank of Zaragoza has been carried out. These accessions correspond to native material containing important characteristics that can be use to breed this species. The genotyping of these accessions has also made it possible to find errors in the labelling of samples and to discover duplicate individuals, optimizing the collection and simplifying the work of the breeder. Another objective of this work is to validate the effectiveness of the *Illumina Infinium BeadChip 18K* array (18000 SNPs). This new version of the *Illumina* peach array (9000 SNPs) hasn't been used until now, which highlights the novelty of the results obtained. The use of this tool has not only served to establish the genetic relationships of the different accessions, but also the analysis of genetic diversity. This fact, has also allowed us to know the population structure, highlighting that varieties from Murcia, whose genetic background seems different from the rest of accessions. This suggest a possible exchange of genetic material between the north areas of Spain. The information obtained in this work will be useful to develop new cultivars with resistance to new and emerging diseases such as Corky Spot, which are more problematic each year.

Key words: peach, germplasm bank, SNP array, SNP, genetic diversity, physiopathy, *Illumina Infinium BeadChip 18K*, native material.

2. Objetivos

El objetivo principal del presente proyecto fue ampliar y clarificar la base documental del banco nacional de germoplasma de melocotonero para la obtención de nuevas variedades resistentes a daños en la conservación postcosecha.

Se han dividido en 4 las fases llevadas a cabo para la consecución del objetivo expuesto:

- Fase 1. Extracción de ADN de las muestras de los individuos con el genotipo seleccionado y la medición de la calidad de dichas muestras, asegurando que cumplen los estándares establecidos para su posterior utilización en genotipado masivo.
- Fase 2. Utilización del chip de genotipado comercial de Illumina para la realización del genotipado.
- Fase 3. Clasificación y filtrado de los datos de genotipado obtenidos con el chip y determinación final del genotipo de los individuos.
- Fase 4. Análisis de la diversidad genética y estructura poblacional.

Este trabajo ha integrado una gran diversidad de métodos modernos y análisis genéticos para la generación de nuevo conocimiento para el melocotonero. La información obtenida durante el desarrollo del presente trabajo, facilitará un manejo más eficiente de las variedades de melocotonero del banco nacional de germoplasma. Así, los resultados finales ayudarán a conocer qué individuos aportan valor y cuáles no, ahorrando costes de mantenimiento y espacio. A su vez, dicha información proporcionará a los mejoradores una descripción genética detallada a la hora de seleccionar los materiales más idóneos para sus programas de mejora.

3. Introducción

El género *Prunus* es uno de los principales exponentes de la familia *Rosaceae*, incluso dentro de las Angiospermas. Esto se debe al gran número de especies que lo componen, así como a la importancia comercial de estas (Pérez-Zabala, 2007). En particular el melocotonero (*Prunus Persica* (L.) Batsch), es un frutal de gran importancia en nuestro país. Sin embargo, la producción y la calidad de este frutal se está viendo afectada por la aparición de nuevas enfermedades, patógenos y fisiopatías, cada vez más emergentes como consecuencia de los efectos derivados del cambio climático. En este sentido, los objetivos de la mejora del melocotonero se centran en la resistencia a estas enfermedades y fisiopatías y en la adaptabilidad a inviernos con temperaturas medias más altas y resistencia a la sequía (IVIA, 2018).

Las variedades modernas de melocotonero, resultado de la mejora de esta especie, son en general más productivas que las variedades tradicionales, sin embargo, estas nuevas variedades pueden tener una riqueza genética más reducida que las hace más vulnerables a los nuevos estreses bióticos y abióticos. Para evitar la pérdida de la variabilidad intraespecífica debido a la sustitución de las variedades tradicionales es necesario investigar la variación genética en las colecciones nacionales presentes en los bancos de germoplasma, cuya misión consiste en la conservación de los recursos genéticos con el objetivo de evitar esta pérdida (CITA, 2020). De esta manera, el conocimiento de la diversidad genética disponible en los bancos de germoplasma para su uso en los programas de mejora, permitirá enfrentar los estreses derivados del cambio climático. Sin embargo, actualmente no existe una caracterización genética exhaustiva de las colecciones de melocotonero del banco de germoplasma nacional localizado en Zaragoza (CITA y EEAD-CSIC).

3.1. Importancia del cultivo del melocotonero

3.1.1. Origen y expansión del cultivo de melocotonero

El melocotón (*Prunus persica*) cuyo origen se sitúa en China (Wang y Zhuang, 2001), con indicios de su cultivo que datan del Neolítico (Chen, 1994) se empezó a domesticar a partir de ejemplares silvestres.

Mediante las antiguas rutas comerciales de la seda, el melocotón llegó a Persia, tomando la denominación de fruta pérsica, por lo que a menudo se ha pensado, de manera errónea, que era originario de dicho país. El melocotón llega a Grecia en el siglo IV a.C en torno al año 330, expandiéndose al resto de Europa hasta la Edad Media (Hancock y col., 2008).

En el siglo XVI, variedades de melocotón procedentes de España y Portugal llegaron a América, desde donde se extendieron al norte del continente (Skorza y Okie, 1990).

Finalmente, en el siglo XX se desarrolla la hibridación y algunos métodos de propagación de semillas de variedades seleccionadas (Iglesias y Alegre, 2014).

3.1.2. Usos del melocotón

Con respecto a los usos del melocotón, en España, el 86% de la producción total de melocotón se destina a las ventas para el mercado fresco y el 14% restante a la venta de este producto para su transformación en conservas, zumos y secado (MAPA, 2018). Este porcentaje apenas ha variado en los últimos 10 años, pues, aunque los gustos del consumidor fluctúan, la transformación de esta fruta, ofrece un amplio abanico de productos. El consumo propio, tanto para alimentación animal como humana, aunque está contabilizado, no representa ni un 1% del total de destino de la producción de melocotón (MAPA, 2018). En la siguiente tabla se presentan la producción de melocotón destinada a cada uso:

Tabla 1: Destino y producción del Melocotón en España

	Destino de la producción	Producción (Tn)
Ventas	Consumo en fresco	1234705
	Transformación	197281
Consumo propio	Alimentación animal	1117
	Alimentación humana	4441

Nota. Fuente: Elaboración propia a partir de los datos del MAPA (2018).

Si bien actualmente los usos del melocotón se centran en su consumo, se están comenzando a abrir nuevas vías de aprovechamiento del melocotón y los subproductos que este fruto genera. Una propuesta interesante, es el proyecto de Agropack, que plantea el aprovechamiento de los subproductos del melocotón y otros frutales de hueso para el desarrollo de bioenvases y otros envoltorios de uso agrícola. El desarrollo del proyecto se centra en la Región de Murcia, proponiendo una solución a los residuos producidos en el destrío de la fruta de hueso, que se aprovechan para obtener un valor añadido. Las repercusiones sociales, ambientales y económicas de este proyecto son notables, ya que

estos residuos suponen entre un 10% y un 28% de la producción total de frutales de hueso (Agropack, 2020).

3.2. Superficie y producción en España

España es el tercer productor a nivel mundial de melocotón, situándose detrás de China e Italia (FAO, 2018). Según la FAO, la producción de melocotón en España en 2018 fue de 903809 toneladas anuales con una superficie que abarcaba 49868 hectáreas de cultivo. Respecto a la producción, que se puede observar en la Figura 1, en los últimos 10 años había experimentado un crecimiento más o menos progresivo, pero en 2018 presenta un gran descenso. Los motivos de esta caída, se deben a varios factores, entre los que se encuentran las condiciones meteorológicas adversas que afectaron a prácticamente todas las zonas productoras de melocotón (MAPA, 2018). Las previsiones para 2020 han ido disminuyendo conforme ha avanzado el año. En la actualización del Boletín de fruta de hueso del MAPA, a principios de septiembre de 2020, se plantean valores de producción muy inferiores a los de los últimos 7 años. Una de las razones que se atribuyen a esta caída, son las tormentas que se dieron en Aragón y Cataluña (MAPA, 2020).

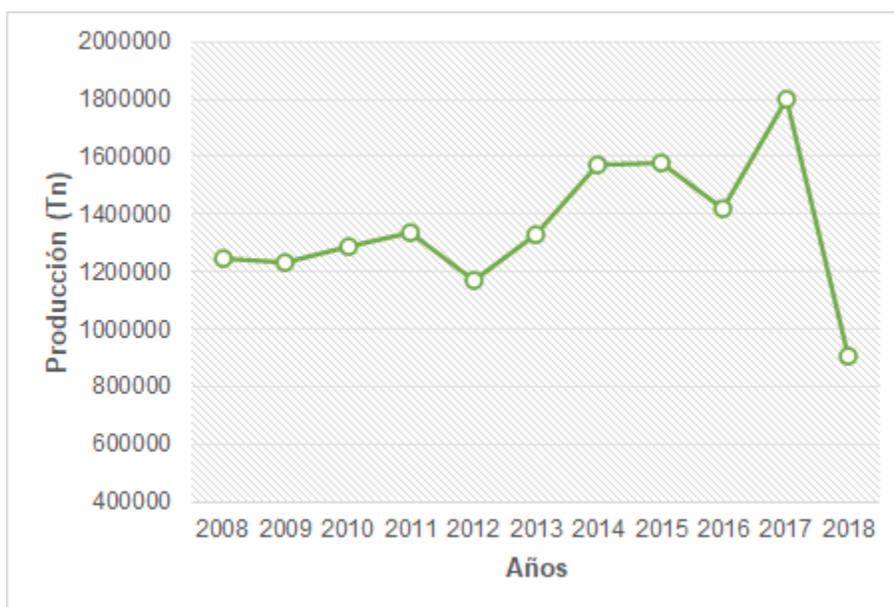


Figura 1: Producción de Melocotón y Nectarina en España (2008-2018)

Nota. Fuente: Elaboración propia a partir de datos recogidos de la FAO (2018).

La superficie dedicada a este cultivo, representada en la Figura 2, presentaba un crecimiento progresivo hasta el 2015, pero comenzó a decaer en 2016 y en 2018 alcanzó el valor más bajo de los últimos 10 años. La superficie sólo aumentó unas 32 hectáreas en Castilla-La Mancha, mientras que las Comunidades Autónomas con mayor reducción de

superficie fueron Cataluña con 825 hectáreas menos, Andalucía, que perdió 579 hectáreas y Aragón, reduciendo su superficie unas 388 hectáreas (MAPA, 2019). Las razones por las que se ha producido la disminución en la superficie de cultivo, no son necesariamente negativas; en los últimos años, se han estado desarrollando variedades mucho más productivas que permiten intensificar el sistema de cultivo obteniendo la misma producción e incluso mayor en una superficie más pequeña. No obstante, una de las razones más importantes era la de equilibrar la oferta y la demanda del sector (MAPA, 2019).

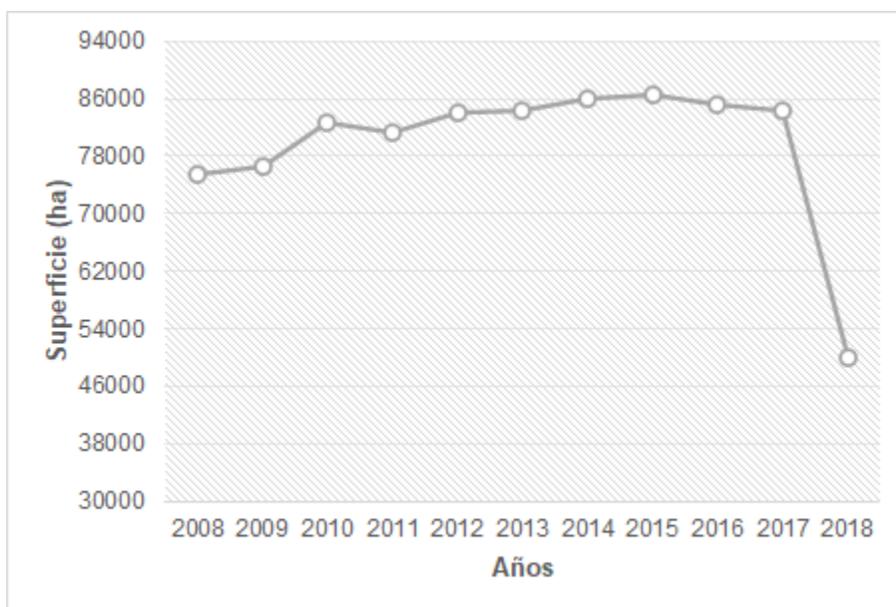


Figura 2: Superficie de Melocotón y Nectarina en España (2008-2018)

Nota. Fuente: Elaboración propia a partir de datos recogidos de la FAO (2018).

3.3. Mejora genética del melocotonero

3.3.1. Grupos pomológicos

El melocotonero, es un frutal diploide y su número de cromosomas es $2n=2x=16$ (Hancock y col., 2008).

Debido a la gran variabilidad de caracteres que presenta este fruto, existen clasificaciones en grupos pomológicos en función de sus caracteres morfológicos que nos permiten identificarlos fácilmente. No hay un consenso establecido respecto a estas clasificaciones, así en ocasiones no se establece una diferencia clara entre el melocotón y el durazno. Otras fuentes agrupan la nectarina y el bruñón en diferentes grupos, al igual que ocurre con los paraguayos y las platerinas. A continuación, se encuentran los cuatro grupos pomológicos que Rafael Cambra describió en 1984 para “Denominaciones de Melocotón”:

1. Durazno: La denominación de durazno se aplica al conjunto de frutos de epidermis vellosa y pulpa dura adherida al hueso. Las variedades tradicionales españolas son duraznos (Cambra, 1984).
2. Melocotón: Son de pulpa más blanda, tanto de color amarillo como blanco y poseen el hueso libre o semi-libre. Su epidermis también es vellosa (Cambra, 1984).
3. Nectarina: Son frutos de piel lisa, no pubescente y pulpa fibrosa (Cambra, 1984).
4. Paraguayos: Caracterizados por su forma achatada, pueden ser de piel vellosa o lisa, conociendo a estos últimos como platerinas (Cambra, 1984).

3.3.2. Programas de mejora genética del melocotonero

El melocotonero, es una de las especies frutales que mayor dinamismo varietal presenta, por lo que cada año aumenta el número de variedades comercializadas. En los años 70, se registraron cerca de 200 variedades nuevas, y la cifra fue aumentando, pues en la década de los 90 ya se registraban más de 1000 variedades nuevas de este fruto (Llácer y col., 2009).

Actualmente, en el mundo se encuentran activos 59 programas de mejora genética de melocotonero, de los cuales 12 se desarrollan en España. Estos programas de mejora pueden tener financiación privada, pública o ambas, y proceden tanto de obtentores privados, como de Universidades e Institutos de Investigación (Cos-Terrer, s.f). En el Anexo A, se exponen en una tabla todos los programas de mejora de melocotón activos, así como el tipo de financiación de estos.

3.4. Principales problemas postcosecha

Existen una serie de factores que influyen en la postcosecha del melocotón y que es importante tener en cuenta si se busca la calidad óptima del fruto. Estos factores, son los siguientes:

- El grado de madurez en el que se recolecta el fruto, que condiciona tanto el sabor como la calidad y vida útil del fruto. Los melocotones que se recolectan inmaduros suelen ablandarse de manera irregular y desarrollan una textura anómala. Por el contrario, los melocotones que se recolectan demasiado maduros, tienen una vida postcosecha mucho más corta ya que se ablandan muy rápido volviéndose más sensibles a daños mecánicos e infecciones (Crisosto y Costa, 2008).
- Concentración de etileno en el almacén, una hormona vegetal que juega un papel crucial en la maduración del fruto. El melocotón es una fruta climatérica ya que su respiración aumenta durante la maduración. La exposición a altos niveles de etileno acelera la maduración del fruto, por lo que es necesario controlar dichos niveles (Oria-Almudí, 2001).

- Problemas de deterioro, que se centran principalmente en los daños internos y los daños por frío. Ambos fenómenos están ligados a la temperatura de almacenamiento. Los daños internos vuelven la carne seca y harinosa, mientras que, en los daños por frío, es frecuente la pérdida de sabor antes de que se manifiesten sus efectos (Crisosto y Labavitch, 2002).
- Si las temperaturas son demasiado bajas, la fruta se expone a daños por frío y otras deficiencias. Estos daños incluyen incapacidad para madurar correctamente cuando la fruta vuelve a temperatura ambiente, pardeamiento de la carne y lo que se conoce como sangrado (Ramina y col., 2008).
- El riego, ya que algunos de los efectos negativos que produce el estrés hídrico son el aumento de defectos en la fruta, como suturas profundas y la formación de frutos dobles. Además, la producción final también se ve reducida (Crisosto y Costa, 2008).
- Nutrición mineral, un importante factor en la calidad y en la vida útil del fruto. Los principales nutrientes minerales a tener en cuenta por su influencia son (Crisosto y Costa, 2008):
 - El Nitrógeno en exceso causa un excesivo crecimiento vegetativo, produciendo sombreado y muerte de las fructificaciones inferiores, además de retrasar la madurez del fruto. También hace al fruto más susceptible a lesiones y podredumbre. Por el contrario, la deficiencia de este nutriente da lugar a árboles poco productivos, de frutos pequeños y de poco sabor (Crisosto y Costa, 2008).
 - El Calcio está involucrado en numerosos procesos bioquímicos y morfológicos. A este nutriente, aunque no está del todo claro, se le atribuyen fisiopatías como la mancha vitrescente, que se relaciona con el metabolismo del calcio y el desarrollo del fruto (Crisosto y Costa, 2008). Otra fisiopatía que actualmente ocasiona importantes problemas en los cultivos es Corky Spot, una alteración caracterizada por grandes áreas de piel marrón. Los daños comienzan siendo externos y evolucionan hacia el interior del fruto. En las primeras etapas se produce russetting en la piel y el fruto se va deshidratando, la pulpa se vuelve oscura y con una textura corchosa, que da nombre a la fisiopatía (Val-Falcón, 2017). La campaña de 2019-2020 en la Región de Murcia, ha contado con la presencia de esta fisiopatía a niveles preocupantes, provocando pérdidas, según la información aportada por la empresa Frutas Esther de Abarán. Su presencia se detectó hace unos 12-14 años en variedades de media estación, sobre todo en años de primaveras lluviosas y frías. La mayor incidencia se observa sobre árboles jóvenes y los síntomas se detectan en cámara tras la recolección cuando el daño es más leve, o en el campo antes de la recolección si los daños son más graves. En los últimos dos años se ha observado que el problema es más acusado en Murcia y Extremadura que en otras zonas como Lérica. Controlar el vigor, aportar Boro/Calcio, hacer podas de verano parecen ser acciones efectivas para su control. Otras propuestas son el uso de sondas de succión, analíticas, control de riego con cámara de presión y sondas de humedad entre otras (J. Gómez, Director Técnico Frutas Esther, comunicación personal) se están desarrollando para acabar con este problema. El control genético de esta

enfermedad y las posibles fuentes de resistencia no se han determinado todavía en melocotonero. Siendo este uno de los objetivos principales del proyecto nacional (RTI2018-094176-R-C32) en el cual se ha enmarcado este trabajo.

- El Potasio es el nutriente mineral más abundante en el melocotón, y su presencia en niveles óptimos da lugar a un alto nivel fotosintético y una mejora de la calidad de la fruta por reasignación de azúcares y ácidos orgánicos (Crisosto y Costa, 2008).
- El Hierro absorbido por la fruta, cuando existe deficiencia de este, afecta a los rendimientos y a la calidad del fruto. Además, el color del fruto también puede verse afectado (Crisosto y Costa, 2008).
- Patógenos fúngicos, los cuales generalmente penetran en el fruto mediante esporas o conidios que se introducen a través de heridas extendiendo la infección. Suele producirse por altas temperaturas de almacenamiento. Destacan los siguientes (Adaskaveg y col., 2008):
 - Podredumbre marrón (*M. fructicola*, *M. laxa* y *M. fructigena*) y podredumbre gris (*B. cinerea*): la podredumbre marrón y la gris son las enfermedades postcosecha más comunes de los frutales de hueso. La causada por *B. cinerea*, suele presentar un micelio algodonoso, mientras que *Monilia spp.* a menudo aparece en tonos muy oscuros y sin esporulación (Adaskaveg y col., 2008).
 - Pudrición por *Rhizopus*, *Gilbertella* y *Mucor*: las frutas infectadas por *Rhizopus* desarrollan un micelio algodonoso con diminutos esporangios negros. En las infecciones por *Gilbertella* también se desarrolla un micelio algodonoso, pero más pequeño que el de *Rhizopus* y con esporangios oscuros y terminales. En la infección por *Mucor* los esporangióforos son más largos y los esporangios son de color marrón (Adaskaveg y col., 2008).
 - Pudrición ácida (*Geotrichum candidum*): sus síntomas son líquido en descomposición, micelio blanco y descomposición que llega a consumir el fruto por completo. Provoca un olor característico (Adaskaveg y col., 2008).

3.5. Importancia y conservación de la diversidad genética

Los programas de mejora genética, como los comentados anteriormente, se ven afectados por la baja diversidad genética intraespecífica presente en melocotonero. Esto se debe principalmente a la auto-compatibilidad de esta especie homocigótica y al bajo número de genotipos introducidos y usados en los programas de mejora genética (Badenes y col., 2015).

Es evidente, que la diversidad genética proporciona la base fundamental para la resistencia de los frutales a nuevas enfermedades o plagas y a la adaptación a cambios en el clima. Por tanto, la principal herramienta para los programas de mejora son los recursos fitogenéticos. Estos recursos fitogenéticos se han conservado generalmente en los bancos de germoplasma de cada país, como en España. Este tipo de conservación fuera del hábitat

natural de una especie se conoce como conservación ex situ y puede desempeñar un papel importante en la protección contra la pérdida de la biodiversidad.

Claramente, los programas de conservación “ex situ” complementan la conservación “in situ” (protección y gestión de ecosistemas), ya que preservan a largo plazo el material genético de accesiones locales o especies olvidadas. El objeto principal es conservar la mayor cantidad de variabilidad genética en el menor número de individuos posible. Esto nos permite tener un mayor conocimiento de las características fenotípicas, fisiológicas y bioquímicas del material que se almacena en el banco de germoplasma y, además, proporciona material vegetal para la utilización en programas de mejora genética de especies cultivadas y en planes de reforzamiento, reintroducción o introducción (McNeely y col., 1990).

La gestión de los recursos fitogenéticos comprende los siguientes pasos principales: su conservación, que consiste en la adquisición de material vegetal (mediante la recolección o protección de reservas in situ, o mediante el intercambio de material ex situ), su mantenimiento (como almacenamiento y propagación), su caracterización (basada tanto en el genotipo como en el fenotipo) y su utilización para la investigación, los programas de reproducción o la producción (Bretting y Widrlechner, 1995).

Con respecto a su caracterización, los distintos genotipos de las especies frutales tradicionalmente se han clasificado mediante descriptores bien definidos para cada una de las especies. Se entiende como descriptor a una característica cuya expresión (fenotipo) se puede medir, registrar y evaluar de manera sencilla. Algunos ejemplos en frutales son: el color del fruto, su tamaño, la fecha de maduración, la fecha de floración, o la resistencia a los diferentes problemas postcosecha señalados en el apartado anterior.

Sin embargo, esta caracterización fenotípica no siempre captura la diversidad y estructura genética de una población, ya que la forma y la apariencia de un individuo no solo dependen de una variable, sino que son bastantes las variables que influyen, como las medioambientales y geográficas entre muchas otras.

Así, en ciertas ocasiones la documentación y la información sobre las colecciones, son escasas o imprecisas y la caracterización a nivel genético inexistente. En la actualidad, las herramientas moleculares contribuyen a la gestión de los recursos fitogenéticos (Wambugu y col., 2018).

3.5.1. Herramientas para estudiar la diversidad genética: Marcadores moleculares

Como se ha comentado anteriormente, los marcadores moleculares nos están ayudando a conocer y conservar mejor la variabilidad genética de las especies cultivadas y las silvestres afines (Bretting y Widrlechner, 1995; Lavi y col., 1994; Lee, 1995).

Al genotipar, los marcadores moleculares, nos muestran las diferencias que hay en el genotipo de diferentes individuos utilizando ensayos biológicos, normalmente su ADN (InnoVine, 2016). Entre los usos de los marcadores genéticos se encuentran:

1. Conocimiento y uso del sistema reproductivos.
2. Selección asistida en mejora
3. Desarrollo de plantas transgénicas
4. Identificación varietal
5. Uso en la agroindustria
6. Conservación y uso de los recursos genéticos.

El último de los mencionados “Conservación y uso de los recursos genéticos”, es el que claramente nos atañe en este trabajo. Los marcadores de ADN han sido utilizados en este trabajo con el fin de documentar y describir la colección del banco de germoplasma de Zaragoza, así los análisis de genotipado masivo han ocupado gran parte de este trabajo.

3.5.1.1. Clasificación de los marcadores moleculares

Según Nuez y col. (2002), los marcadores de ADN, pueden clasificarse en dos grandes grupos: los de restricción/hibridación y los basados en la reacción en cadena de la polimerasa (PCR), tal y como vemos en la Figura 3:

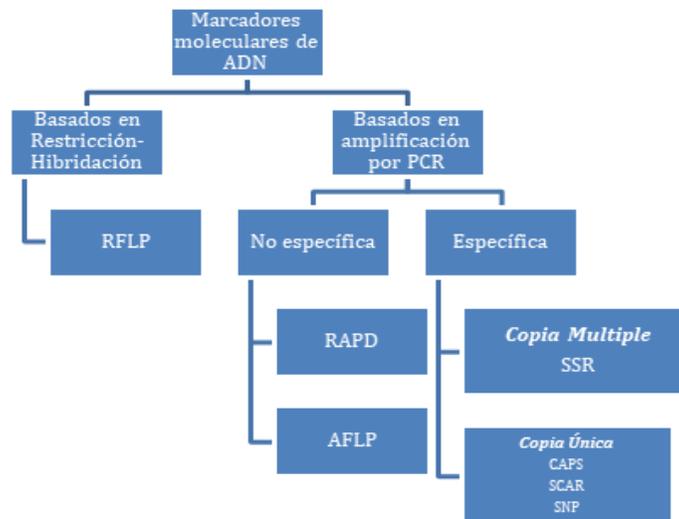


Figura 3: Clasificación de los Marcadores Moleculares

Nota: Fuente: Nuez y col. (2002).

Los marcadores basados en la PCR, fueron una nueva generación con una elaboración más sencilla. Dentro de este tipo de marcadores, se clasifican en base a si la amplificación es con cebadores, con secuencias específicas (SSR, CAPS, SNPs, etc.) o con no

específicas (aleatorios) (AFLP y RAPD) (Nuez y col., 2002). Los marcadores moleculares amplificados por PCR con cebadores específicos, se clasifican a su vez en marcadores de copia múltiple y en marcadores de copia única. Dentro del grupo de marcadores de copia única se encuentran los CAPS, los SCARs y los SNPs (polimorfismos de nucleótido único, en inglés *Single Nucleotide Polimorphisms*), que han sido descritos a continuación de manera más detallada al ser los marcadores utilizados en este trabajo.

3.5.1.2. Marcadores SNPs

Los avances en la secuenciación de alto rendimiento (del inglés *Next Generation Sequencing* o NGS) del ADN, los programas informáticos y la bioinformática han hecho que se haya extendido el uso de los SNPs (Heaton y col., 2002). Los SNPs son mutaciones resultantes de la sustitución de un solo nucleótido. Algunos autores sólo consideran como SNP una sustitución de base, mientras que otros también consideran SNP a la inserción o delección de una base (Martínez y col., 2010).

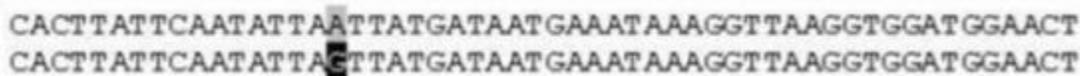


Figura 4: Ejemplo de SNP en 2 cadenas de Nucleótidos

Nota: Fuente: Martínez y col. (2010).

Los SNPs son por tanto de naturaleza bialélica, siendo cada alelo una pareja de nucleótidos; es decir, un alelo puede ser Adenina-Timina o puede ser Citosina-Guanina. Las ventajas de los SNPs son una mayor abundancia (Heaton y col., 2005), la estabilidad genética (Markovtsova y col., 2000), una nomenclatura más simple e idoneidad para el análisis automatizado y la interpretación de datos (Wang y col., 1998). Todo ello unido a un menor coste, permite un genotipado de cientos o miles de SNPs que de forma combinada proporcionan una capacidad de asignación de parentesco superior a los paneles de menos de 20 microsatélites utilizados habitualmente (Fernández y col., 2013).

3.5.1.3. Genotipado masivo

Las técnicas de genotipado masivo son aquellas que permiten genotipar un gran número de marcadores en una gran cantidad de muestras de manera automatizada. Habitualmente, estos sistemas requieren un alto coste en infraestructuras, por lo que los ensayos suelen realizarse en servicios centralizados (Bañuls, 2019). Existen diversos tipos de sistemas de genotipado masivo, como los sistemas fijos de arrays, los sistemas flexibles (Taqman,

KASP) o los sistemas basados en NGS como *Genotyping by Sequencing* (GBS) (Bañuls, 2019), siendo los sistemas fijos de *arrays* o chips de genotipado los que han sido utilizados en este trabajo.

El chip se diseña utilizando la tecnología de hibridación, basándose en el principio bioquímico de que un nucleótido se une con su complementario formando parejas de bases. El chip contiene cientos de miles de sondas (oligonucleótidos fijos) con secuencias únicas, por lo que cada sonda está diseñada con el objetivo de que se le una una secuencia de ADN específica, tal y como se muestra en la Figura 5:

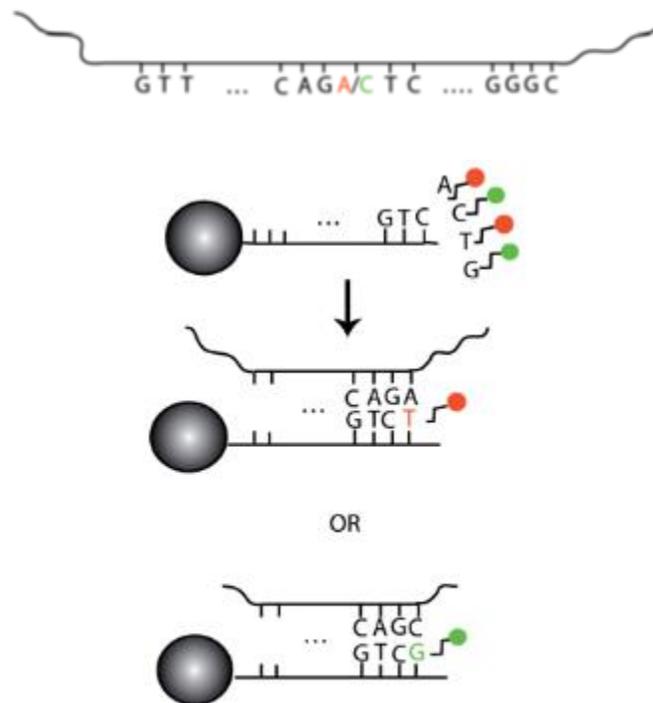


Figura 5: *Detección de los SNPs mediante Fluorescencia*

Nota. Fuente: LaFramboise (2009).

Los SNPs se detectan porque los oligonucleótidos diseñados para cada locus no coinciden entre sí en un sitio específico, que sería en el del SNP (nucleótidos individuales). Al hibridar con los fragmentos de ADN genómico, los oligonucleótidos se unen solo con sus complementarios, y los que no complementan se lavan. El resultado de esa unión puede tomarse como sistema de detección, marcando con distintos tipos de fluorescencia los nucleótidos (Martínez y col., 2010).

Los chips de genotipado, se diseñan secuenciando muchos individuos, para obtener una gran variabilidad y por tanto muchos SNPs. Una vez obtenido el chip se realiza el genotipado de los individuos a estudiar, lo que nos permite saber qué alelos tiene ese SNP

en cada individuo. Así, podemos conocer cuántos SNPs tiene cada individuo y si son homocigóticos o heterocigóticos para un alelo u otro.

Posteriormente se debe realizar un procesamiento informático de los datos generados con los chips de genotipado, para poder interpretarlos y poder exportarlos para su uso en diferentes análisis a posteriori. Este proceso conlleva desechar aquellos marcadores SNPs que nos pudieran llevar a una interpretación errónea del genotipo de un individuo para dicho marcador. Para llevar a cabo esta importante tarea existen múltiples *programas*, entre ellos:

GenomeStudio

Es un programa con el que se pueden visualizar y analizar los datos generados por los sistemas de *Illumina*, como *iScan*. Admite el análisis de genotipado de datos de chips de ADN, también conocidos como “*microarrays*”. Las herramientas, de rendimiento óptimo y la interfaz gráfica de cómodo manejo, permiten convertir los datos en resultados significativos de una forma fácil y sencilla (Illumina, 2020).

La función principal de *GenomeStudio* en este trabajo, es la transformación de la lectura del chip (lectura de la fluorescencia para cada SNP e individuo), en información genotípica que nos diga qué alelos tiene cada individuo para cada SNP.

El nivel de intensidad de fluorescencia de cada sonda, representa la solidez de la señal para cada alelo. Tras medir los niveles de fluorescencia de las 2 sondas, procedentes de muchos individuos, *GenomeStudio* aplica un algoritmo según los niveles de fluorescencia, que permite distinguir a los individuos en clústeres o grupos según su clase genotípica AA, AB y BB (Zhao y col., 2018). *GenomeStudio* denomina alelo A, a la pareja Adenina-Timina, y el alelo B a la pareja Citosina-Guanina (Vanderzande y col., 2018). Estos clústeres, se pueden ver representados gráficamente en Figura 6 obtenida de *GenomeStudio*:

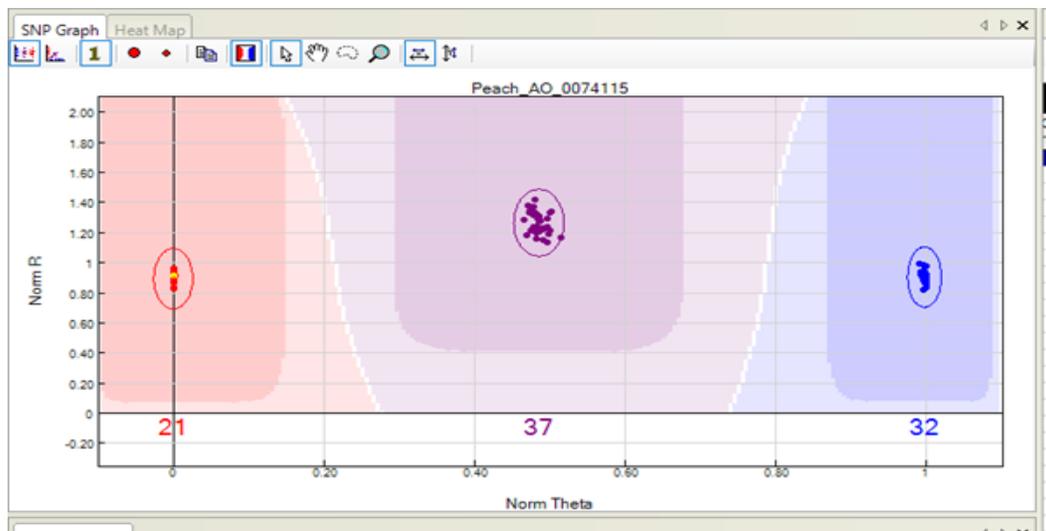


Figura 6: Gráfica de *GenomeStudio* para un determinado SNP

Nota: Fuente: *GenomeStudio*.

En la Figura 6, podemos ver la gráfica obtenida para un SNP. El eje Y, *Norm R*, representa la totalidad de la fluorescencia, y el eje X, *Norm Theta*, representa el ratio de la intensidad alélica, correspondiente al alelo B. Cada punto representa un individuo: el clúster rojo, representa el genotipo AA, el clúster morado representa el genotipo AB y el clúster azul representa el genotipo BB (Staaf y col., 2008; Zhao y col., 2018). Los números que aparecen bajo la gráfica (21,37 y 32 en este caso), nos indican el número de individuos que se ha asignado a cada clase genotípica.

Assist

ASSIsT (de su definición en inglés *Automatic SNP Scoring Tool*) es una herramienta que permite asignar, filtrar y clasificar los SNP del chip de *Illumina Infinium* de manera más rápida y eficiente que *GenomeStudio*. Este programa está específicamente diseñado para chips de genotipado masivo. Según se describe en el manual, *ASSIsT* requiere las siguientes matrices de datos exportadas de *GenomeStudio*, como:

1. *Final Report* de *GenomeStudio*: incluyendo los valores de *Theta*, *R* y *GenTrain Score* para cada SNP e individuo.
2. *DNA Report* de *GenomeStudio*: incluyendo valores de *Call Rate*, *No Calls*, Frecuencias de alelos A y B, para cada individuo.
3. Pedigrí: consiste en una hoja de cálculo en el que se indica los parentales de cada individuo. En este caso, al no conocer los parentales, se dejó en blanco.
4. Mapa físico: consiste en una hoja de cálculo en la que se indica el cromosoma y posición física de cada SNP.

Se puede emplear en el análisis de diferentes tipos de población (F2, germoplasma o F1). El resultado final se puede exportar en diferentes formatos según el análisis posterior, como el mapeo genético o la asociación de caracteres y marcadores (Di guardo y col., 2015). Es necesario como etapa final del genotipado, incorporar esta nueva información generada por *ASSIsT* al programa *GenomeStudio* para una visualización final y más detallada de cada SNP.

3.5.2. Herramientas para estudiar la diversidad genética: Programas estadísticos

Los avances en las tecnologías de genotipado masivo como los comentados anteriormente y la producción de gran cantidad de datos a un ritmo sin precedentes, ha requerido el desarrollo y el uso de nuevos programas informáticos. Estos programas permiten manejar y extraer toda la información incorporada mediante análisis clásicos y más sofisticados en el estudio de la diversidad genética dentro de las especies. Una descripción completa de algunos de estos programas como *Genepop*, *Fstat*, *Arlequin*, o *MEGA*, entre otros, ha sido recogida en el trabajo de Excoffier y Heckel (2016). Además, en los últimos años se han desarrollado gran cantidad de paquetes estadísticos dentro del entorno de *R* como *PopGenReport*, *hierfstat*, *poppr*, *SpadeR*, *StrataG*. Algunos de los programas más

destacados (y utilizados en este trabajo) en el análisis de la diversidad han sido descritos a continuación.

DiveRsity

Esta herramienta nos da la información necesaria para realizar un buen análisis de diversidad genética y poder describir la diversidad presente en cada especie. *DiveRsity* es un paquete con múltiples funciones y entorno de programación *R*, que fue creado para calcular medidas de diversidad genética en poblaciones por diversos investigadores (Keenan y col., 2013). Esta infraestructura de trabajo que nos provee *DiveRsity* nos permite obtener estadísticas de diferenciación genética, asignación de ascendencias, subdivisión entre poblaciones, así como parámetros de población básicos como frecuencias alélicas, heterocigosidad observada/esperada y pruebas de Hardy-Weinberg.

Plink

Este programa permite analizar grandes bases de datos, como gran cantidad de marcadores moleculares usados en muchos individuos, lo que permite manipularlos y analizarlos en su totalidad (Pizarro, 2018). Es un programa de código abierto y gratuito que se ha diseñado para realizar estos análisis a gran escala de manera muy eficiente. Está enfocado puramente en el análisis de datos genotipo/fenotipo (Brigham & Women's Hospital, 2017). Uno de los análisis que nos permite PLINK 1.07 es el estudio de las relaciones de parentesco. Para ello se calcula la identidad por descendencia (en inglés *Identity-By-Descent* (IBD)).

Structure

El software *Structure* es un programa que sirve para investigar la estructura de una población con información genotípica obtenida con marcadores comunes como SSR o SNPs, basándose en algoritmos iterativos Bayesianos. Su uso incluye la deducción de distintas poblaciones dentro de la muestra, asignando individuos a poblaciones y estudiando zonas híbridas (Pritchard y col., 2000).

La nueva versión de *Structure*, denominada *fastStructure*, ha mejorado el modelo de inferencia de este utilizando un marco bayesiano variacional, lo que le convierte en una versión más eficiente ante un elevado número de datos de genotípicos (como los generados en este trabajo), estando escrito en *Python 2*.

Con estos programas podemos situar varios escenarios, marcando un número *K*, de poblaciones o clúster, viendo que población asigna a cada individuo, o si lo asigna a dos o más poblaciones. Estas poblaciones se diferencian entre sí, por un conjunto de frecuencias alélicas para cada locus (Pritchard y col., 2010).

Cada individuo, por tanto, tiene un coeficiente estimado de afiliación (o coeficiente de mezcla) a cada población definido como *q*, con un valor máximo de 1, de forma que ese valor para cada individuo puede repartirse en una o varias poblaciones, según su genotipo

y el análisis llevado a cabo por *Structure* (Pritchard y col., 2010). Cada escenario con un número K de clústeres, se puede ver representado en una gráfica, de forma que en el eje Y tenemos el valor de q, y en el eje X, nos encontramos a cada individuo representado por una barra vertical. Esta barra vertical, estará coloreada con un color representativo de cada clúster, en proporción al valor de q de cada individuo para cada clúster (Pritchard y col., 2010). En la Figura 7 podemos ver un ejemplo:



Figura 7: Representación Gráfica de un Clúster (K=3)

Nota. Fuente: *Structure*.

4. Material y Métodos

Todas estas metodologías empleadas y los diferentes materiales usados en este trabajo se describen a continuación. En el Anexo B se observa una descripción esquematizada de todos los métodos y herramientas utilizadas en este trabajo.

4.1. Material Vegetal

Para la elaboración de este trabajo, se ha contado con 90 accesiones (84 nacionales y 6 extranjeras) de melocotonero procedentes de la colección nacional de frutales del Centro de Investigación y Tecnología Agroalimentaria de Aragón (CITA) y de la Estación Experimental de Aula Dei (EEAD-CSIC). En la selección, encontramos variedades representativas tanto de melocotón de carne dura amarilla como de melocotón de carne dura blanca y en menor medida de paraguayo y nectarina. En cuanto al origen de las variedades de durazno, predominan las variedades españolas y americanas.

Las accesiones de variedades españolas, provienen principalmente de la Región de Murcia, de Aragón, tanto de Huesca, como de Zaragoza y Teruel, también de Lérida, de Navarra, y algunas accesiones que proceden de La Rioja y Valencia (Figura 8). Debemos recordar que el fenotipo concreto de cada accesión no ha sido objeto de estudio y evaluación en este trabajo. Toda la información ha sido recabada de diferentes fuentes como diversos trabajos de investigación publicados.



Figura 8: Orígenes de las Variedades Españolas presentes en el Trabajo

Nota: Elaboración propia.

No todas las variedades presentes en este trabajo son variedades comerciales o tradicionales, de las que se conocen ampliamente sus caracteres morfológicos. Algunas de estas variedades están apenas documentadas, por lo que será la información genética, que es el fundamento de este trabajo, la que nos proporcionará dichos datos para ampliar la información de estas variedades en el banco de germoplasma.

Antes de proceder a la descripción de las variedades documentadas, es necesario aclarar una serie de términos en relación con las características de estas variedades. Es frecuente encontrar el fruto designado como *“melting”* o *“non-melting”*. Este término indica la textura del fruto, que puede ser fundente y no fundente. La textura fundente o *“melting”* indica que el fruto pierde firmeza durante la etapa postcosecha, mientras que una textura no fundente o *“non-melting”* señala un mejor mantenimiento de la firmeza durante dicha etapa. Los *“non melting”* son los típicos melocotones de Calanda, o melocotones aptos para conserva. También es frecuente señalar la adherencia del hueso a la carne como un carácter de interés, siendo *“freestone”* frutos de hueso libre y *“clingstone”* frutos de hueso adherido. Por último, el color de la pulpa va a ser otro de los datos a tener en cuenta. Aunque en la guía descriptiva de la Unión Internacional para la Protección de las Obtenciones Vegetales (UPOV) son 7 los colores descritos, las variedades presentes en este trabajo se clasifican únicamente como variedades de pulpa blanca o de pulpa amarilla.

A continuación, podemos dividir en 2 grandes grupos las variedades tratadas en el presente trabajo, cuya clasificación por orígenes se puede consultar en el Anexo C:

1. Variedades con origen y características conocidas

Entre las accesiones de variedades documentadas de melocotón procedentes de Aragón, encontramos sobre todo variedades de frutos gruesos o medio gruesos (Campillo Rocho) y redondeados, que presentan colores de piel que van desde el rojo intenso (Zaragozano y San Lorenzo), al amarillo con chapa rojiza (Maluenda 2375) (Ayala Delgado, 1984). Entre estas accesiones, encontramos variedades tipo *“melting”* o de textura fundente, como Blanco Tardío y las variedades de paraguayo Villamayor y San Mateo, todas procedentes de Zaragoza. Estas variedades además coinciden en el color de la pulpa, que es blanca. Paraguayo Jota también es una variedad de paraguayo de pulpa blanca procedente de Zaragoza, sin embargo, no se dispone de información acerca del tipo de textura que presenta. No obstante, es más frecuente que las variedades tengan una textura no fundente, como Gallur, Miraflores, Miraflores Serapio, Escolapio, Borracho de Jarque, L2-B y Montamar en Zaragoza y Pigat, Fulla, Comodin, Moret y San Lorenzo en Huesca. Todas estas variedades poseen además la pulpa amarilla. También son variedades aragonesas de pulpa amarilla y textura no fundente Amarillo Temprano (Ebro), Duraznillo 42B, Sunmel 1, Sunmel 2 y Risol. Sin embargo, no todas las variedades de pulpa amarilla son de textura no fundente ni todas las de pulpa blanca de textura fundente. La mayoría de las variedades de Huesca de las que se posee información, son de pulpa blanca y textura no fundente como Binaced, Fraga B.D., Pigat Susagna y Pomar. Las variedades zaragozanas Sudanell 3099 B.D., Montañana y Villamayor B.D. también poseen la pulpa blanca y una textura no fundente. Además, una de las variedades de Teruel, Valdeltormo B.D., también es de pulpa

blanca y textura no fundente. Calanda Sonrosado, la otra accesión de Teruel, es también de textura no fundente, pero tiene la pulpa amarilla (Bouhadida y col., 2011; Font i Forcada y col., 2014; Badenes y Zuriaga, 2016).

La única accesión procedente de Valencia, Utiel B.D., es una variedad de pulpa blanca y textura no fundente (Badenes y Zuriaga, 2016).

Las accesiones de las variedades de melocotón documentadas y originarias de Navarra, son 9 variedades de textura no fundente, 6 de pulpa blanca, Tambarria B.D, Deja-1, Deja-2, Deja-3, Montes de Cierzo y Valentin B.D, y 3 variedades de pulpa amarilla, Campiel Montes de Cierzo, Rojo de Azagra y Rojo de Tudela (Bouhadida y col., 2011; Font i Forcada y col., 2014; Badenes y Zuriaga, 2016).

Las accesiones de variedades de melocotón que proceden de Cataluña, y en concreto de Lérida son muy variadas. Entre estas se halla el grupo de las Sudanell (S), S. 1 y S. 2, de textura no fundente y pulpa amarilla. Este grupo, de gran importancia, se caracteriza por ser de frutos gruesos y redondeados, con piel amarilla y una pequeña chapa roja. La diferencia principal entre estas variedades se debe a su etapa de maduración (Ayala Delgado, 1984). Además, se sabe que la variedad S. 2 es de un vigor algo superior a S.1, y que S.1 presenta una producción algo menor respecto a esa variedad (Navarro, 1995). También son variedades de pulpa amarilla y “*non-melting*” procedentes de Lérida las accesiones Rojo del Rito, Tipo Rojo del Rito, Rojo-Amarillo Septiembre, Amarillo Oct., Roig de Aitona, La Escola y San Jaime, destacando esta última por su forma ligeramente alargada. Una de las variedades de Lérida, Buisan, es de pulpa blanca y textura fundente (Bouhadida y col., 2011; Font i Forcada y col., 2014; Badenes y Zuriaga, 2016).

La única variedad procedente de La Rioja, la variedad Generoso, es de textura no fundente y de pulpa amarilla (Badenes y Zuriaga, 2016).

Dentro de las accesiones de variedades descritas procedentes de Murcia, encontramos principalmente variedades tradicionales. En un grupo tenemos los distintos Calabaceros (C), C. Deleite, C. Soto, C. Rincón C. Rancho y C. Candelo. Estos en general presentan una forma redondeada, aunque a veces presentan formas poco asimétricas (C. Soto) o irregulares (C. Deleite). La epidermis de los calabaceros varía desde una epidermis amarilla anaranjada y una leve chapa roja (C. Deleite), una epidermis con bastante vello y color amarillo anaranjado con una leve chapa rojiza (C. Soto) o una epidermis con algunos tonos rojos (C. Rincón). El calibre de los calabaceros varía desde medio (C. Rincón y C. Soto) a buen tamaño (C. Deleite). Otro grupo son los Jeronimos (J), J. Copia, J. Ortiz, J. Prasio, J. Espuña y J. Balate. Tienen un calibre generalmente mediano, aunque alguna de las variedades presenta un menor tamaño (J. Copia). Suelen ser de forma redondeada, sin embargo, alguna variedad presenta una forma más achatada y asimétrica (J. Prasio). Son variedades de piel vellosa con tonalidades desde el amarillo al naranja pálido (J. Copia y J. Prasio). Por lo general presentan una chapa de color rojo intenso o algo más leve en otros casos (J. Prasio). Por último, el grupo de las Marujas (M), M. Tradicion, M. Alquibla, M.

Perfeccion, M. Tejar y M. Argos. Este grupo presenta variedades de piel muy vellosa, con una coloración desde el amarillo más dorado (M. Tradicion y M. Alquibla) al amarillo-anaranjado (M. Perfeccion y M. Tejar) y usualmente con chapa rojiza. El calibre de los frutos va desde mediano (M. Alquibla y M. Perfeccion) a algo más pequeño (M. Tradicion) y tienen formas muy variadas, hay variedades de frutos más ovalados y ápice bastante marcado (M. Tradicion y M. Alquibla), variedades de fruto más redondeado (M. Tejar) o de forma esférica-aplanada (M. Perfeccion) (Durán, 1993). Otra de las variedades de procedencia murciana, pulpa amarilla y textura no fundente es Brasileño Elipe (Badenes y Zuriaga, 2016).

Por último, no todas las variedades presentes en este trabajo son variedades de procedencia española. Dos de las accesiones de variedades de melocotón de las que se tienen datos, son variedades tradicionales, de pulpa amarilla y textura no fundente, procedentes de Estados Unidos. Entre ellas encontramos Andora, una variedad tradicional poco vigorosa y de forma redonda, tipo pavía, al igual que Vivian, de tamaño medio y algo más achatada, con piel amarilla y de tonos rojizos (Ayala Delgado, 1984). Otra de las accesiones americanas documentada, es Aniversario, variedad de nectarina de pulpa amarilla, con piel de tono rojo y un calibre de medio a grande. Esta variedad es originaria de Argentina (Valentini y Daorden, 2011).

El color y la forma de los frutos de algunas de las variedades descritas en este apartado puede verse en el Anexo D.

2. Variedades con origen conocido, pero características desconocidas

Algunas de las accesiones de este trabajo no presentan una descripción evidente y establecida en la literatura. Así, las variedades de las que no se posee información y que proceden de Aragón son: CD-RCH, Paraguayo Almudi, Paraguayo Almunia, Paraguayo Niqui, Pavía Amarilla de Tolosa y Pavía Blanca y Rincal. Igualmente, en el caso de la accesión Paraguayo T. Robert, con origen de Lérida, no presenta una clara descripción fenotípica (aún sin documentar). Entre las accesiones de variedades americanas que se encuentran sin documentar, encontramos Aurelio, procedente de Estados Unidos y Pepita, que es una variedad brasileña.

4.2. Extracción de ADN

Para conocer la diversidad genética y realizar cualquier estudio molecular, es necesaria la extracción del ADN de las especies a estudiar. En el presente trabajo, para la extracción de ADN se utilizó el método clásico de extracción de ADN genómico. Este método fue publicado por Doyle y Doyle en 1987 y se conoce como “CTAB *miniprep*”. El material necesario y el procedimiento para esta metodología están descritos en el Anexo E.

4.3. Medidas de calidad del ADN

Para realizar las medidas de calidad del ADN y realizar la cuantificación de este, se han utilizado dos métodos distintos con tecnologías diferentes: *Nanodrop* y *Qubit*.

La primera medida se realiza con *NanoDrop*. El modelo de instrumento utilizado es *Thermo Scientific™ NanoDrop One™*, fabricado por *Thermo Fisher Scientific*. Es un espectrofotómetro UV-Visible compacto y autónomo que se utiliza para analizar microvolúmenes de ácidos nucleicos y una gran cantidad de proteínas purificados. Para la medida de ácidos nucleicos, se debe calibrar el aparato con un blanco, que es una solución que evita que el tampón donde está resuspendido el analito contribuya a la absorbancia de la muestra. Es importante medir el blanco antes de cada serie de muestras. El blanco puede ser agua, pero también TE, como en este caso, en el cual este compuesto fue el utilizado. Una vez calibrado, se pipetea 1-2 μl de la solución de muestra en el pedestal, se baja el brazo y al finalizar la medición, la pantalla ofrece el espectro de medición y los valores correspondientes. El espectro típico de medición de los ácidos nucleicos es de 260 nm, y entre 230 y 280 nm se realiza una estimación de pureza de la muestra, midiendo la absorbancia a esas longitudes de onda para detectar la presencia de proteínas u otros compuestos (Thermo Fisher Scientific, 2016).

Cuando las muestras alcanzan los estándares exigidos en los ratios de *NanoDrop* (un ratio A_{260}/A_{230} entre 1,7 y 2,2 y un ratio A_{260}/A_{280} entre 1,7 y 1,9), y se encuentran dentro del espectro típico de los ácidos nucleicos, se realiza la medición con *Qubit*, que proporciona unas medidas más precisas de concentración de ADN. El modelo de instrumento utilizado es *Qubit® 3.0* de *Thermo Fisher Scientific*. Es un espectrofotómetro que usa fluorocromos específicos para medir ácidos nucleicos y proteínas. Se diferencia del *Nanodrop* porque puede cuantificar cadenas simples o dobles de ADN, lo que ofrece una medida más precisa (SCT UAB, 2020). Para comenzar a utilizarlo, se deben usar una serie de reactivos que vienen con el instrumento, preparando en primer lugar una solución stock de trabajo para las muestras y una para los estándares. Una vez preparado todo, se etiquetan los tubos para las muestras y para cada estándar. Antes de la medición se calibra el instrumento con los estándares y la solución de trabajo. A cada tubo que contiene la solución stock se le añade 1-2 μl de la muestra y se homogeneiza en el vortex. Con el equipo ya calibrado y los estándares leídos, se coloca el tubo con la muestra en la cavidad correspondiente para ser leído (RAI, 2016).

4.4. Chip de genotipado y procesamiento de datos

A continuación, se describe el chip de genotipado utilizado en el proyecto y los dos programas complementarios de genotipado que se han usado en este trabajo y que nos permiten un análisis detallado de cada SNP: *GenomeStudio* y *ASSIsT*.

4.4.1. Chip de genotipado

El chip que se ha utilizado para el genotipado de las muestras, es la segunda versión del chip comercial de genotipado de melocotonero *Illumina Infinium BeadChip 18K* (18000 SNPs). Esta segunda versión presenta una mayor densidad de marcadores que el primer chip, *Illumina Infinium BeadChip 9K* (9000 SNPs) (Verde y col., 2012). Tanto las muestras de ADN de cada accesión como el chip de genotipado de *Illumina*, fueron enviados al servicio de genotipado del Centro de Investigación en Agrigenómica (CRAG) de Barcelona. Las características exigidas por el servicio de genotipado respecto a la calidad de las muestras de ADN fueron las siguientes:

1. No contener inhibidores de PCR.
2. Un ratio A260/230 entre 1,7 y 2,2.
3. Un ratio A260/280 entre 1,7 y 1,9.
4. Intacto según lo visualizado por electroforesis en gel.
5. Que no haya estado expuesto a temperaturas de más de 60°C, ya que por encima de estas se puede producir degradación.

Este servicio cuenta con una gran experiencia en genotipado masivo, y diferentes sistemas para dicha tarea, como el Sistema *iScan* (innovador lector de matrices que permite adquirir de forma rápida, sensible y precisa imágenes de *BeadChips* de *Illumina* para obtener unos resultados de análisis genéticos excepcionales) que fue utilizado para el análisis de nuestras muestras. Una vez concluido dicho genotipado, los datos en bruto (sin limpiar) fueron devueltos a nuestro grupo de investigación. Seguidamente se procedió al procesamiento de los datos mediante el *software GenomeStudio* y los estudios posteriores.

4.4.2. Procesamiento de datos

El genotipo de cada marcador para cada individuo se determinó a través del *software GenomeStudio*, transformando la información en bruto recibida desde el servicio de genotipado en la representación de los genotipos AA, AB y BB para cada marcador. Se determinó un umbral mínimo de *GenCall* de 0,15. El genotipo de los SNPs que tienen un valor de *GenCall* menor a 0,15, se determina como ausente, dado que se considera que no es fiable la información que dan, como se describe en Illumina, 2014.

GenomeStudio calcula gran cantidad de parámetros para facilitar y ayudar al procesamiento y visualización de la información genotípica, así como a su interpretación. Algunos de ellos son (Illumina, 2014):

- **GenCall:** Este parámetro es una medida para cada marcador en cada individuo, que toma valores entre 0 y 1, indicándonos la distancia que hay en la gráfica obtenida para cada SNP, entre el punto de un SNP de un individuo y el centro del clúster al

cual este punto está asociado. El valor de *GenCall* decrece cuanto más lejos esté el punto del centro del clúster. *GenomeStudio* establece un umbral mínimo estándar en el proyecto de 0,15 para que el SNP de dicho individuo sea un dato fiable para establecer un genotipo.

- **Calls:** número total de individuos para un SNP que tienen un *GenCall* superior al valor establecido en el proyecto como umbral mínimo de *GenCall*.
- **No Calls:** número total de individuos para un SNP que tienen un *GenCall* inferior al valor establecido en el proyecto como umbral mínimo de *GenCall*.
- **Call Rate:** es calculado para cada individuo en relación a todos los SNP muestreados como:

$$\text{Calls}/(\text{No Calls}+\text{Calls})$$

- **Call Freq:** es calculado para cada SNP en relación a todos los individuos muestreados como:

$$\text{Calls}/(\text{No Calls}+\text{Calls})$$

- **AA, AB, BB Freq:** para cada SNP, la proporción de cada genotipo entre el total de *calls*:

$$\text{AA Genotype Calls}/\text{Calls}$$

$$\text{AB Genotype Calls}/\text{Calls}$$

$$\text{BB Genotype Calls}/\text{Calls}$$

- **GenTrain Score:** es un parámetro que se calcula para cada SNP. Toma valores de 0 a 1 y depende de la forma de los clústeres y la distancia relativa entre ellos. Este parámetro está diseñado para imitar las evaluaciones visuales hechas por expertos. También tiene en cuenta penalizaciones al aparecer hechos como baja intensidad de fluorescencia (Illumina, 2005).

A través de *GenomeStudio* se realizaron exportaciones de informes para su posterior uso en *ASSisT*. Posteriormente, se realizó un filtrado de los SNPs con el *software ASSisT*. Las diferentes categorías y subcategorías de *ASSisT* son:

Aprobados

1. Robustos: todos los SNP genotipados con éxito, en los cuales la segregación sigue los principios mendelianos y el porcentaje de individuos en que ese SNP no está presente o no ha funcionado la prueba (*NoCall*) es menor al máximo permitido (se establece un 5%) (Di Guardo y col., 2015).

2. *OneHomozygRare_HWE*: son los SNPs cuya frecuencia de un clúster homocigótico, es menor del umbral situado de frecuencias alelo raro (5%) pero las proporciones de las 3 clases genotípicas respetan el equilibrio de Hardy-Weinberg (Di Guardo y col., 2015).
3. *OneHomozygRare_NotHWE*: son los SNPs cuya frecuencia de un clúster homocigótico, es menor del umbral situado de frecuencias alelo raro (5%) pero las proporciones de las 3 clases genotípicas no respetan el equilibrio de Hardy-Weinberg (Di Guardo y col., 2015).
4. *DistorsionadoYNoesperaSegre*: cuando la segregación en familias enteras muestra una severa inclinación hacia una clase genotípica, cuando una clase genotípica no aparece en una población de germoplasma o cuando una clase genotípica no sería posible según el genotipo parental (por ejemplo, porque no se ha determinado un alelo en un parental) (Di Guardo y col., 2015).

Descartados

1. *Monomórficos*: son aquellos que solo presentan una clase genotípica, es decir no son polimórficos, y su frecuencia es mayor que el umbral de alelos raros situado en 5% (Di Guardo y col., 2015).
2. *Fallos*: aquellos SNPs en los que el porcentaje de individuos en que ese SNP no está presente o no haya funcionado la prueba (*NoCall*) es muy alto, los que tienen una señal de intensidad muy baja o los que no se pueden clasificar correctamente debido a que no forman clústeres por clase genotípica (Di Guardo y col., 2015).
3. *Cambiado-Homo*: en una población, ocurre cuando una de las 2 clases homocigóticas no está presente (Di Guardo y col., 2015).
4. *AleloNulo-Fallo*: son los SNP cuya frecuencia de *HomozygousNull* (cuando la intensidad de la señal luminosa para dicho SNP es menor que el umbral establecido para declarar como nulo el SNP) es mayor que el umbral de frecuencia de alelos raros situado por defecto en 5% (Di Guardo y col., 2015).

Siendo lo más restrictivos posibles, se usarán para los análisis de diversidad y estructura genética, aquellos SNPs clasificados dentro de la categoría *Aprobados*. Es posible realizar un análisis más profundo de estos SNPs con el *software GenomeStudio* y recuperar algunos de los SNPs, por ejemplo, los de la categoría *Cambiado-Homo*, pero no se ha realizado para este trabajo.

4.4.3. Análisis de diversidad genética

Para una descripción detallada del panel de accesiones estudiado en este trabajo se estimaron las siguientes medidas de diversidad genética a través del *software DiveRsity*.

Riqueza alélica (en inglés *Allelic richness*) (A_r): es una medida de diversidad genética indicativa del potencial para persistir y adaptarse a largo plazo de una población. Se define como el número medio de alelos por locus en una población. Una reducción de la riqueza alélica, podría provocar una reducción del potencial de la población para adaptarse a cambios ambientales, ya que la diversidad es el material de partida de la evolución para la selección natural (Greenbaum y col., 2014).

Heterocigosidad Observada (en inglés *Observed heterozygosity*) (H_o): proporción dentro de un conjunto de individuos que son heterocigóticos para un locus determinado (de Vicente y col., 2003; Peakall y Smouse, 2009). Haciendo la media para todos los locus, podemos obtener el valor medio para la subpoblación. Cuanto mayor sea, más heterocigóticos habrá en la población.

$$H_o = \frac{\text{N}^\circ \text{ de Heterocigóticos}}{\text{N}^\circ \text{ de Individuos}}$$

Heterocigosidad esperada (en inglés *Expected heterozygosity*) (H_e): es la proporción de heterocigosidad esperada bajo cruzamientos al azar y p es la frecuencia de cada alelo en la subpoblación (de Vicente y col., 2003; Peakall y Smouse, 2009). El valor de H_e será máximo para un locus cuando los alelos se encuentren en las mismas proporciones en la población y cuanto mayor sea el número de alelos disponibles.

$$H_e = 1 - \sum_{i=1}^h p_{i,s}^2$$

- p : frecuencia de i alelos en la subpoblación S .
- h : N.º de alelos en la población.
- s : subpoblación.

Índice de Fijación F_{ST} : esta medida nos proporciona una medida de diferenciación genética entre subpoblaciones, pero no en un sentido absoluto, si no relativo. Es decir, nos indica la fijación que hay en relación a la fijación total posible (Jost y col., 2018). En *DiveRsity R package*, lo calculan según *Weir and Cockerham's fixation index estimator* (Ma y col., 2015; Keenan, 2017).

$$\theta = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_I^2 + \sigma_G^2}$$

Donde σ_p^2 , es la suma de los componentes de la varianza para las poblaciones, σ_I^2 , es la suma de los componentes de la varianza para individuos dentro de las poblaciones y σ_G^2 , es la suma de los componentes de la varianza para los alelos de los individuos.

Nei's g_{ST} (g_{ST}): Es una medida análoga a F_{ST} , nos proporciona una medida de diferenciación genética entre subpoblaciones, pero no en un sentido absoluto, sino relativo, nos indica la fijación que hay en relación a la fijación total posible. Explica la cercanía a la homocigosis o fijación de los alelos (Jost y col., 2018; Ma y col., 2015; Keenan, 2017).

$$g_{ST} = \frac{D_{ST}}{H_T} = \frac{H_T - H_S}{H_T}$$

- D_{ST} es la media de la diversidad genética entre subpoblaciones.
- H_T se define como la Total heterozygosity, la probabilidad de tener un genotipo con la unión de dos diferentes formas alélicas del total de la población.

$$H_T = 1 - J_T$$

- H_T también se puede descomponer linealmente en $H_T = H_S + D_{ST}$.
- J_T es la probabilidad de tener un genotipo con la unión de dos formas alélicas idénticas (homocigosis), donde p es la frecuencia alélica en la población.

$$J_T = \sum p_i^2$$

H_S es la media de las diversidades genéticas dentro de cada subpoblación, también puede ser expresada en una subpoblación como:

$$H_S = 1 - J_S$$

J_S se expresaría como (Nei, 1973):

$$J_S = \frac{\sum i p_i^2}{k}$$

D_{JOST}: Jost creó esta medida para evitar algunos inconvenientes de las anteriores G_{ST} , basando su medida en el número efectivo de alelos, tratando de seguir una linealidad y que la estimación de la diversidad fuera más intuitiva. Una desventaja de este método es que depende del tamaño de la muestra por lo que se necesita aplicar un factor de corrección para estandarizar el tamaño de las muestras. Además, esta medida mide la diferenciación genética, absoluta, en el sentido de que su valor es 1 cuando los alelos de las poblaciones no se comparten, son privados, y 0 cuando las poblaciones tienen los mismos alelos a las mismas frecuencias. Es decir, D tomará su valor máximo valor, 1, si y sólo si, las poblaciones no comparten alelos con el resto de poblaciones. En el caso de los marcadores SNP, solo hay dos posibles alelos, por lo que cuando haya dos poblaciones, podrá alcanzar el valor máximo 1, pero cuando haya más, el valor máximo posible a alcanzar será menor a 1, ya que no puede haber alelos privados para todas las poblaciones si solo hay 2 posibles alelos. (Jost y col., 2018; Meirmans y Hedrick, 2011).

$$D = \frac{k}{k-1} \times \frac{H_T - H_S}{1 - H_S}$$

- k : Número de poblaciones.

Jost y col. (2018), indican que en el caso de marcadores SNP, para las comparaciones por parejas entre poblaciones, las medidas de diferenciación relativa o de Fijación, F_{ST} y G_{ST} , dan una respuesta similar a las de diferenciación absoluta, como D .

En cambio, advierten, que al comparar con SNPs entre varias poblaciones, las medidas relativas y absolutas, nos aportan informaciones diferentes, ya que las relativas nos dan información acerca de la fijación mientras que las absolutas nos informan sobre la no compartición de alelos.

4.4.4. Relaciones de parentesco

Las medidas realizadas con el chip de ADN, nos dicen el genotipo de numerosos loci, así podemos comparar los individuos entre sí en términos de IBD por cada locus (SNP). En estos loci los individuos pueden tener 0, 1 o 2 alelos idénticos. Por lo que se realizó el cálculo por parejas de individuos de la proporción de Identidad por descendencia (IBD) a través del *software PLINK*, para detectar sinonimias o clones. Así, a través del comando “- genome”, *PLINK* nos proporciona el cálculo de la IBD (parámetro \hat{P}_i) entre todos los individuos (por parejas). La proporción IBD por tanto se puede calcular como (Purcell, 2012):

$$\text{Proporción IBD} = \frac{N.^{\circ} \text{ de marcadores con IBD} = 2 + 0,5 \times N.^{\circ} \text{ de marcadores con IBD} = 1}{N.^{\circ} \text{ de marcadores } non - missing \text{ (sin pérdidas de información)}}$$

Según (Vanderzande y col., 2018) valores de \hat{P}_i superiores a 0,97 suponen que los individuos son clones o individuos idénticos. Otros parámetros como Z_0 , Z_1 y Z_2 también se obtuvieron con dicho análisis. Son las proporciones de compartir 0 alelos, 1 alelo o 2 alelos. Esto es útil, ya que además de poder detectar sinonimias con *PLINK*, podemos detectar relaciones de parentesco, de forma que se verá una relación padre-hijo, cuando los valores de Z_0 y Z_2 sean 0, y el valor de Z_1 sea 1. Cuando 2 individuos sean sinónimos, los valores de Z_0 y Z_1 serán 0, y el valor de Z_2 sería 1. Podemos ver un ejemplo de una relación padre-hijo entre dos individuos en la Tabla 2:

Tabla 2: Relación de Parentesco Padre-Hijo según los Parámetros de *PLINK*

	Individuo 1	Individuo 2	IBD
Locus 1	AA	AB	1
Locus 2	AA	AB	1
Locus 3	AA	AB	1
<hr/>			
Pi_hat		0.5	
Z0		0	
Z1		1	
Z2		0	

Nota. Elaboración propia.

5. Resultados y Discusión

Antes de comenzar los análisis de los datos en este trabajo, fue necesario determinar el origen de algunas de las accesiones utilizadas en este trabajo. Para ello se mantuvieron conversaciones con investigadores encargados del banco de germoplasma de Zaragoza y se consultaron los libros de registro de este banco. Al tratarse de accesiones muy antiguas, estos orígenes, en algunos casos, eran desconocidos por los propios investigadores, como en el caso de Buisan (C.M. Cantín, investigadora CITA, comunicación personal). Así, entre las accesiones estudiadas, la región de Aragón es la que más accesiones aporta a nuestro panel, con un total de 43 accesiones, siendo la localidad de Zaragoza la más representada con 29 accesiones en el estudio. Tras esta, la siguiente en representación es Murcia con un total de 18 accesiones. Por otro lado, tanto en el caso de La Rioja como Valencia solamente una accesión de estas zonas ha sido utilizada en este estudio. La Tabla 3, muestra cuántas variedades existen para cada origen.

Tabla 3: *N.º de Variedades según su Origen*

Origen de las variedades	N.º
Extranjeras	6
Huesca	12
La Rioja	1
Lérida	12
Murcia	18
Navarra	9
Teruel	2
Valencia	1
Zaragoza	29

Nota. Elaboración propia.

5.1. Resultados de la extracción y medida de calidad del ADN

Para realizar el genotipado masivo y obtener buenos resultados, era necesario que las muestras alcanzaran unos estándares de calidad. Para ello, se procedió a la cuantificación del ADN mediante la utilización de *NanoDrop* y *Qubit*. Los resultados de estas medidas pueden consultarse en el Anexo F. En general todas las muestras alcanzaron los estándares necesarios para su correcto genotipado por el servicio de genotipado del CRAG. El promedio de concentración de ADN medido por *Qubit* fue de 191,9 ng/μL y el promedio del ratio A260/A280 de 2,088, siendo algo mayor que los estándares definidos, lo cual no supuso un problema. El ratio A230/260, sí se encontraba dentro del marco establecido, con un valor promedio de 1,709. Se tuvieron que extraer de nuevo 13 de las muestras antes de ser enviadas para su genotipado ya que no alcanzaban las concentraciones requeridas. Las accesiones sometidas a una segunda extracción fueron: Sudanell 3099 B.D. (5001), Zaragozano (5004), Deja-2 (5051), Fulla (5073), Pigat Susagna (5147), Tambarria B.D. (5186), Sudanell 1 (2211) (5175), Vivian (5266) (5206), Calabacero Deleite (5260), Calabacero Rincón (5262), Jeronimo Balate (5314), Sudanell 2 (2349) (5176), Sunmel 2 (5504).

5.2. Resultados obtenidos con GenomeStudio

En el Anexo G, podemos observar el *CallRate* obtenido por *GenomeStudio* para los 90 individuos. Los valores de todos los individuos son muy próximos siendo el menor el de Paraguay Villamayor (0,9543584) y el mayor el del individuo San Lorenzo (0,9680758). La media de todos los individuos es de 0,964531454. De acuerdo al servicio de genotipado, un *CallRate* mayor de 0,95 para cada individuo demuestra una idoneidad de las muestras. Esto fue comparado en nuestro experimento de genotipado, ya que se incluyeron seis muestras de almendro (e híbridos de almendro) que presentaron unos *CallRate* menores entre 0,84 y 0,93, lo que es evidente al tratarse de una especie diferente.

En cuanto al *GenTrain Score*, uno de los valores más importante en el genotipado masivo sino el más importante, se obtienen valores para los SNPs desde 0,0725 hasta 0,9638. La media de *GenTrain Score* para todos los SNPs es de 0,765445. En el gráfico de la Figura 9 podemos ver la representación de todos los SNPs con su respectivo valor de *GenTrain Score*, siendo ordenados de menor a mayor valor. En nuestro caso un total de 15190 SNPs tuvieron un *GenTrain Score* mayor de 0,5.

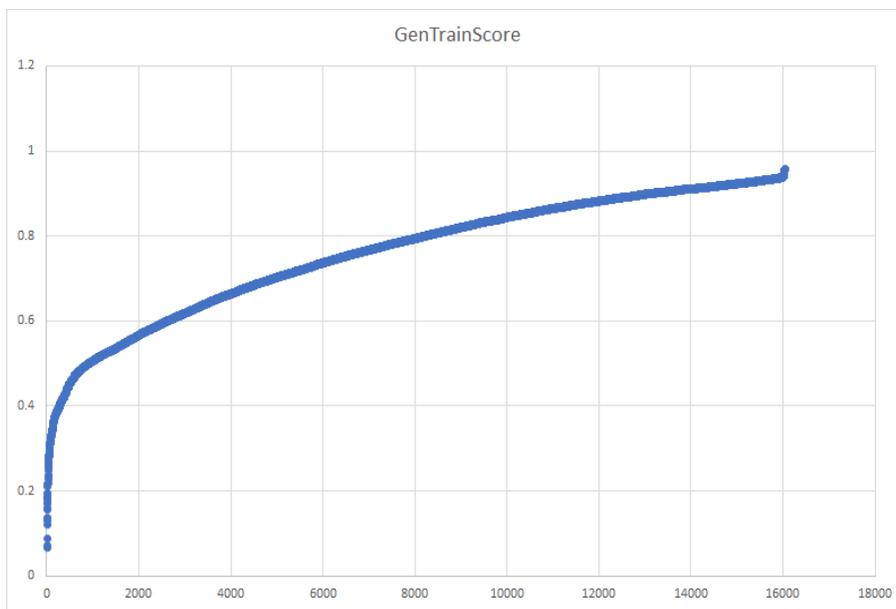


Figura 9: CallRate de los SNPs

Nota: Fuente: *GenomeStudio*.

Martínez-García y col. 2012, usaron la primera generación de chip de genotipado de *Illumina, GoldenGate® Genotyping*, con 1,536 SNPs, para el mapeo genético de 2 poblaciones de melocotonero. Estos autores establecieron como aceptables un total de 1400 SNPs, ya que tenían un *GenTrain Score* mayor de 0,5, un exceso de heterocigicidad mayor o igual de 0,3 y una frecuencia de alelos menores mayor o igual a 0,08. En nuestro caso, el número de SNPs aceptables usando estos parámetros fue de 5404. Sin embargo, este tipo de filtrado, aunque se ajusta a los criterios de *Illumina*, puede resultar demasiado restrictivo y favorece la pérdida de información valiosa para análisis posteriores. Recientemente se han desarrollado nuevos programas para el filtrado de los datos que consiguen una clasificación más detallada y eficiente de estos SNPs. Uno de estos métodos ha sido utilizado en este trabajo para mejorar nuestra elección de los SNPs como veremos a continuación.

5.3. Resultados obtenidos con ASSIsT

El cribado de los datos con *ASSIsT* da para un total de 16038 SNPs y 90 individuos, 8955 SNPs aprobados y 7083 SNPs descartados. Es decir, un 55,84% de los SNPs o lo que es lo mismo, algo más de la mitad, son útiles y aportan información, y el porcentaje restante queda descartado. De los SNPs aprobados, el 20,25% (3247) fueron clasificados como robustos y el 18,26% (2928) fueron descritos como distorsionados y con una segregación inesperada.

De los SNPs descartados, un 16,98% (2724) han sido rechazados por ser monomórficos y no presentar polimorfismo, un 6,33% (1016) se clasifican como fallidos, un 19,8% (3175) no tiene una de las dos clases homocigóticas presentes (Cambiado-Homo). Los datos del cribado de SNPs con ASSIsT se exponen en la siguiente tabla:

Tabla 4: Clasificación de los SNPs en PLINK

SNP Clasificación	Nº	%
Aprobados	8955	55.84%
Robustos	3247	20.25%
OneHomozygRare_HWE	1406	8.77%
OneHomozygRare_NotHWE	1374	8.57%
DistorsinadoYNoesperaSegre	2928	18.26%
Descartados	7083	44.16%
Monomórficos	2724	16.98%
Fallos	1016	6.33%
Cambiado-Homo	3175	19.80%
AleloNulo-Fallo	168	1.05%

Nota: Fuente: ASSIsT.

5.4. Estudio de la diversidad genética

De acuerdo a los resultados de *DiveRsity*, el valor global de F_{ST} fue de 0,0653, lo que indica que casi el 93% de la diversidad genética existente se produce dentro de las 6 (La Rioja y Valencia fueron excluidas por tener un solo miembro, así como las extranjeras) poblaciones utilizadas en este trabajo. Valores similares fueron encontrados en nogal, donde el valor global de F_{ST} fue de 0,07, indicando que el 93% de la diversidad genética se produjo dentro

de las 7 poblaciones iraníes de nogal estudiadas (Arab y col., 2019). De acuerdo a estos autores, estos resultados reflejan la dispersión mediada por humanos de especies, como el nogal y posiblemente el melocotón, en el espacio y el tiempo, lo que resulta en un reensamblaje y homogeneización de la diversidad genética de estos frutales.

En la Tabla 5, podemos ver las medidas de diversidad genética que nos describen las subpoblaciones según el origen de los individuos.

Tabla 5: Medidas de Diversidad Genética según el Origen de los Individuos

Subpoblación	A_r	Nº Individuos	H_o	H_e
Zaragoza	1.524	29	0.195	0.341
Murcia	1.386	18	0.258	0.204
Huesca	1.503	12	0.235	0.306
Lérida	1.5	12	0.254	0.284
Navarra	1.318	9	0.098	0.236
Teruel	1.391	2	0.235	0.218

Nota. Fuente: *DiveRsity*.

Respecto a A_r , Zaragoza obtiene el valor más alto con 1,524, seguido de cerca por Lérida y Huesca, mientras que Navarra tiene el menor valor con 1,318. Esto nos indica que en los individuos de Zaragoza podemos encontrar ambos alelos en un mayor número de loci respecto a otras subpoblaciones como Navarra. Un aspecto a tener en cuenta, es el número de individuos, ya que, por ejemplo, Teruel solo cuenta con 2 individuos, por lo que su diversidad alélica puede encontrarse influenciada y por tanto reducida por este hecho.

En cuanto a la H_o , Murcia obtiene el mayor valor (0,258), seguido de cerca por Lérida con 0,254. En último lugar, vuelve a aparecer Navarra, con 0,098. Esto nos indica que el porcentaje de heterocigóticos para los distintos loci en esta población, es inferior respecto al resto. En general estos valores son inferiores a los obtenidos por Badenes y col. (2015) y por Chavez y col., (2014), con una media de $H_o=0,38$ y $H_o=0,4$ respectivamente. Micheletti y col., (2015) por su parte, genotiparon 1240 accesiones de melocotonero con 4271 marcadores SNPs , obteniendo una H_o de 0,286, superior a nuestros valores pero algo más cercana. Una posible explicación es el mayor número de alelos por locus de los SSRs con respecto a los SNPs (Micheletti y col., 2015). Es preciso tener en cuenta la naturaleza y características de los marcadores SNP y SSR a la hora de comparar los

resultados de diversidad genética. Las tasas de mutación de los SSRs son varias órdenes de magnitud superiores a las de los SNPs, y mucho más variables debido a su fundamento como marcador. Con los marcadores SSRs se muestrean pequeñas fracciones del genoma, en cambio al utilizar grandes cantidades de SNPs, se suele cubrir una mayor parte del genoma y los resultados suelen ser mejores según algunos autores (Fischer y col., 2017). Sin embargo, también hay que tener en cuenta, que puede producirse un sesgo con los SNPs, como consecuencia de una sobrerrepresentación de SNPs comunes (Fischer y col., 2017).

La medida H_e obtiene su valor más alto en la población de Zaragoza (0,341), y Huesca (0,306), siendo Murcia con 0,204, la población que menor valor tiene para esta medida. En general estos valores son inferiores a los obtenidos por Micheletti y col, (2015) ($H_e = 0.39$) y a los que obtuvieron Chavez y col., (2014) ($H_e = 0.49$).

En la Tabla 6, están los datos obtenidos de F_{ST} , que nos indican la diferenciación entre las distintas subpoblaciones. Los mayores valores de diferenciación entre todas las subpoblaciones se dan entre Murcia con Teruel (0,2365), con Navarra (0,1956) y con Huesca (0,1624). También encontramos entre las subpoblaciones de Lérida-Navarra (0,0799) y Huesca-Navarra (0,0655), diferencias moderadas. Los valores inferiores, se observan entre Zaragoza con Teruel (-0,0577), Lérida (-0,0041) y Huesca (-0,0028).

Tabla 6: Datos Obtenidos de F_{ST} en función del Origen

F_{ST}	Zaragoza	Murcia	Huesca	Lérida	Navarra	Teruel
Murcia	0.1085					
Huesca	-0.0028	0.1624				
Lérida	-0.0041	0.093	0.023			
Navarra	0.0438	0.1956	0.0655	0.0799		
Teruel	-0.0577	0.2365	-0.0027	0.0103	0.0085	

Nota. Fuente: *DiveRsity*.

En la Tabla 7, podemos ver los resultados obtenidos con las medidas de diferenciación genética relativa de subpoblaciones Nei's g_{ST} . comparadas por parejas según su origen.

Tabla 7: Datos Obtenidos de G_{ST} en función del Origen

g_{ST}	Zaragoza	Murcia	Huesca	Lérida	Navarra	Teruel
Murcia	0.0639					
Huesca	0.005	0.0848				
Lérida	0.0037	0.0463	0.016			
Navarra	0.0353	0.1069	0.0453	0.0507		
Teruel	0.0074	0.103	0.0226	0.0177	0.0449	

Nota. Fuente: *DiveRsity*.

Los mayores valores de diferenciación genética g_{ST} , se observan en la subpoblación de Murcia con Navarra (0,1069), Teruel (0,103), Huesca (0,0848) y Zaragoza (0,0639). También encontramos un valor alto en la diferenciación entre las poblaciones de Lérida y Navarra (0,0507). Zaragoza, en cambio, es la que más parecido genético guarda con el resto de subpoblaciones, con Lérida, con 0,0037, Huesca con 0,005 y con Teruel 0,0074. En la Tabla 8, observamos la medida de diferenciación absoluta D , las mayores diferenciaciones son con la subpoblación de Murcia, que tiene de diferenciación con Huesca (0,017), con Zaragoza (0,0168) y Navarra (0,16). Los menores valores son entre Zaragoza-Teruel (0), Zaragoza-Lérida (0), Huesca-Teruel (0,0001) y Teruel-Lérida (0,0001).

Tabla 8: Medidas de Diferenciación Absoluta en función del Origen

D	Zaragoza	Murcia	Huesca	Lérida	Navarra	Teruel
Murcia	0.0168					
Huesca	0.0002	0.017				
Lérida	0	0.0063	0.0011			
Navarra	0.0063	0.016	0.0062	0.0068		
Teruel	0	0.0081	0.0001	0.0001	0.0011	

Nota. Fuente: *DiveRsity*.

Las 3 estimaciones F_{ST} , g_{ST} y D , parecen indicar que Murcia tiene un fondo genético diferente al resto de orígenes, mientras que Zaragoza tiene un fondo genético compartido con el resto de provincias del norte.

5.5. Relaciones de parentesco identificadas

Tras el filtrado de los datos, se ha realizado el estudio de las relaciones de parentesco en este panel. Como se ha comentado en el apartado de material y métodos, *PLINK* nos ofrece información sobre los individuos que son iguales o clones (sinonimias) mediante el parámetro Pi_Hat ($\text{Pi_Hat}=1$). Como resultado del comando “*--indep-pairwise*” de *PLINK*, el número total de SNPs usados en el análisis de parentesco fue de 257. Este paso es necesario (y recomendado en el manual de *PLINK*) para quitar los SNPs que estén en desequilibrio de ligamiento, seleccionando solo SNPs que estén en equilibrio de ligamiento (es decir que segregan al azar). En nogal, estos métodos han ayudado a conocer las relaciones de parentesco en un panel de 95 genotipos representando gran parte de la diversidad genética en poblaciones de nogal de Irán, el centro de origen de esta especie frutal (Arab y col., 2019). En este trabajo se han identificado 15 grupos de duplicados o clones que se han dispuesto en el Anexo H.

El Grupo 1, nos muestra que las accesiones denominadas como Calabacero originarias de Murcia, son individuos idénticos o clones. Recientemente, Pérez y col. (2020) usando 10 microsatélites genotiparon 89 accesiones de melocotonero, establecieron como individuos sinónimos a Calabacero Rancho y Calabacero Soto, pero no al resto. Además, afirman una relación de sinonimia entre Jeronimo Espuña, Calabacero Rincon, Jeronimo Balate, Jeronimo Prasio, Maruja y Maruja Argos. Los resultados del presente trabajo junto a las descripciones documentadas, indican que podría tratarse de clones de la misma variedad de Calabacero.

El Grupo 2, establece que las accesiones procedentes de Navarra Deja-1 (5050) y Deja-2 (5051) son duplicados o clones, pero no presentan esta semejanza con Deja-3. Contrariamente, Pérez y col. (2020) por su parte, no las establece como clones, pero si sitúa a Deja-3 como variedad genéticamente lejana, al igual que en nuestro caso. Según los resultados que hemos obtenido y la información recabada sobre su morfología, es posible que Deja-1 y Deja-2 sean clones de la misma variedad Deja.

En el Grupo 3, dos variedades de Zaragoza, Miraflores Serapio (5122) y Miraflores (5213), se establecen como individuos sinónimos. Ambas variedades se encuentran descritas con las mismas características morfológicas, lo que lleva a pensar que estas 2 accesiones son una única accesión.

En el Grupo 4, Jeronimo Ortiz (1) (5265), Maruja Perfeccion (5269), Jeronimo Prasio (5266) y Jeronimo Copia (5263), 4 variedades procedentes de Murcia, son sinonimias. Los resultados de Pérez y col. (2020) también difieren, estableciendo que Maruja Perfeccion, Jeronimo Copia y Jeronimo Ortiz no presentan relaciones de sinonimia y como ya hemos comentado anteriormente Jeronimo Prasio es sinónimo de varios individuos. Según los

resultados establecidos en nuestro proyecto y la información bibliográfica sobre sus características morfológicas, podría tratarse de clones de Jeronimo.

El Grupo 5 nos indica que Tipo Rojo de Rito (5233) y Rojo del Rito (5162), con origen en Lérida, son dos variedades que presentan sinonimia. Su denominación, unida a una idéntica descripción según la información recabada, sugieren que ambas son una única accesión.

El Grupo 6, incluye las cuatro variedades murcianas Maruja Alquibla (5268), Maruja Tradicion (5267), Maruja Argos (5311) y Maruja Tejar (5270) y una variedad de Lérida, San Jaime (5006), siendo estas cinco variedades individuos sinónimos. Contrariamente a estos resultados, Pérez y col., (2020) observa que Maruja Argos presenta sinonimia con las variedades citadas en otros párrafos (Jerónimo Espuña, Calabacero Rincon, etc.) y establece que Maruja Tradicion y Maruja Tejar no presenta sinonimias. Sin embargo, sí que establece como individuos sinónimos a Maruja Alquibla y San Jaime. Según las descripciones consultadas y los resultados de este trabajo, es probable que todas las accesiones sean clones de la variedad Maruja. El clon de San Jaime es posible que se etiquetara con una denominación diferente en la zona en la que fue obtenido.

En el Grupo 7, tres variedades de paraguayo procedentes de Zaragoza, Paraguayo Almudi (5138), Paraguayo San Mateo (5143) y Paraguayo Villamayor (5259), son sinonimias. Según Pérez y col., (2020), aunque estas 3 variedades de paraguayo son genéticamente muy cercanas, no son completamente idénticas en términos genéticos. La información obtenida de fuentes bibliográficas sobre su fenotipo y los clones establecidos en este trabajo, señalan que podría tratarse de clones de la misma variedad de paraguayo etiquetados con denominaciones distintas.

El Grupo 8, indica que las tres variedades denominadas Pomar, Pomar 1 (5149), Pomar 2 (5150) y Pomar 3 (5151), originarias de Huesca, son sinonimias. Al contrario de lo expuesto en este trabajo, Pérez y col., (2020) no las identifican como clones. Sin embargo, según nuestros resultados y la información consultada respecto a su morfología, es posible que se trate de 3 clones de la misma variedad Pomar.

En el Grupo 9, las variedades de Zaragoza Sunmel 1 (5503) y Sunmel 2 (5504) son individuos sinónimos. Ambas poseen las mismas características morfológicas según lo especificado en fuentes bibliográficas. Esto unido a los resultados obtenidos, sugieren que se trata de clones de la misma variedad Sunmel.

El Grupo 10, muestra las variedades Escolapio (5231) y Zaragozano (5004), ambas procedentes de Zaragoza, como sinonimias. Pérez y col. (2020) no establece esta relación de sinonimia, sin embargo, indica que Zaragozano y Borracho de Jarque son clones, algo que no ha sido confirmado en este trabajo. Tanto Escolapio como Zaragozano son descritos con textura no fundente y pulpa amarilla, por lo que es probable que se trate de clones de la misma variedad etiquetados con una denominación distinta.

En el Grupo 11, tenemos 3 variedades, Sudanell 3099 B.D. (5001) de Zaragoza, La Escola (5340) y Sudanell 2 (2349) (5176), de Lérida, establecidos como clones. Sin embargo, la variedad Sudanell 3099 B.D. es de pulpa blanca mientras que las otras 2 variedades son de pulpa amarilla. Estos resultados complican la identificación de la verdadera denominación de los clones. Es importante destacar, que el programa *PLINK* no detecta

las 2 accesiones denominadas La Escola como clones o individuos idénticos. El valor de Pi_{hat} entre estas dos accesiones es igual a 0, por lo que no se trata de la misma variedad. El Grupo 12, presenta 3 variedades con sinonimia, 2 de ellas originarias de Huesca, Fulla (5073) y Pigat (5146) y una de ellas procedente de Lérida, Rojo-Amarillo Septiembre (5159). Por su parte, Pérez y col., (2020) no afirman una relación de sinonimia entre Fulla y Rojo-Amarillo Septiembre. Según sus características, las 3 accesiones son variedades de pulpa amarilla y de textura no fundente, esto resultados podrían confirmar que se trata de una única accesión con diferentes nombres según la zona de origen.

En el Grupo 13, 2 variedades de paraguay, se indican como individuos sinónimos, Paraguay Jota (5141) procedente de Zaragoza y Paraguay Niqui (5359), que es originaria de Huesca. Estos resultados son contrarios a los encontrados por Pérez y col., (2020), ya que según estos autores ambas accesiones no son clones. Sabemos que Paraguay Jota es una variedad de pulpa blanca, y aunque no disponemos de información acerca de la morfología de Paraguay Niqui, es probable que se trate de clones de la misma variedad de paraguay que han sido etiquetados con nombres distintos.

El Grupo 14, incluye dos variedades navarras, Tambarria B.D (5186) y Rojo de Tudela (5161) como sinonimias. Esto supone un problema al banco de germoplasma, ya que las dos variedades presentan características morfológicas diferentes. Así, según la información consultada, ambas son de textura no fundente, pero Tambarria B.D. es una variedad de pulpa blanca y Rojo de Azagra de pulpa amarilla.

El Grupo 15, por último, establece que la variedad Borracho de Jarque (5421) que procede de Zaragoza y las variedades Moret (5126) y Comodin (5044), originarias de Huesca, son sinonimias. En el caso de Borracho de Jarque, Pérez y col., (2020) concluyen que es sinonimia de Zaragozano, algo que como hemos dicho anteriormente, no ha sido confirmado en este trabajo. Sin embargo, estos autores sí que establecen que Moret y Comodin son clones. Los clones obtenidos en este proyecto junto con la bibliografía que describe a las 3 variedades como variedades de pulpa amarilla y textura no fundente, indican que es probable que se trate de clones de la misma variedad a los que se les han atribuido distintos nombres.

Las diferencias observadas entre nuestro trabajo y los resultados obtenidos por Pérez y col., (2020) podrían ser debidas al tipo de marcadores moleculares usados entre ambos estudios. De acuerdo con Fernández i Martí y col., (2012) la disponibilidad y estabilidad de los SNPs en comparación con los SSRs ofrecen mejores perspectivas para la identificación de variedades y la evaluación de la diversidad genética. Es evidente que es necesario corregir, clarificar y unificar estas diferencias que han sido el resultado de pequeños errores durante la adquisición o mantenimiento (o propagación) de estos materiales. Como se ha comentado anteriormente, los resultados de este trabajo serán presentados y estarán a disposición del personal investigador del banco de germoplasma.

5.6. Resultados obtenidos con Structure

El análisis de la estructura poblacional se realizó con *fastStructure*, cuyos resultados pueden consultarse en el Anexo I. De acuerdo con la mejor elección del algoritmo (*chooseK.py*) de *fastStructure*, la K más probable oscila entre 5 y 7. La selección de un valor de "K" puede ser facilitada por métodos estadísticos, sin embargo, estos 'nunca' tienen en cuenta la cuestión biológica de los datos (Durand y col., 2009). Así en nuestro caso una vez obtenidos los resultados de *fastStructure* vamos a seleccionar K=7, como número de subpoblaciones que mejor representan nuestro panel de accesiones. En el Anexo J, también se puede consultar el gráfico donde vemos representado para cada escenario de K poblaciones/clústeres, a todos los individuos en barras verticales, que según su color nos indican su coeficiente de adhesión o membresía a una población/clúster, que se ve representado en el eje y de 0 a 1.

En el clúster 1 (color azul), encontramos las accesiones Blanco Tardío y Buisan, variedades de pulpa blanca y fundente. Atendiendo a esta información este grupo podría englobar a las accesiones de textura fundente. Sin embargo, las variedades Aniversario (Argentina) y Rojo de Azagra (Navarra), presentan pulpa amarilla y textura no fundente, por lo que es complicado atribuir unas características morfológicas a Aurelio (0,92), única variedad de este grupo sin descripción previa. Tanto Aurelio como Aniversario comparten afiliación con el grupo 6. Por otra parte, Buisan (0,56), también presenta afinidad con el clúster 5, 4 y 3. De acuerdo a otros autores, un coeficiente de mezcla de 0,7 (en el caso de SNPs) es suficiente para determinar que una accesión es considerada como una mezcla genética (Bernard y col., 2020).

El clúster 2 (color verde), aúna las accesiones de textura no fundente, la mayoría de pulpa amarilla. Así encontramos, a Miraflores Serapio y su clon Miraflores, Sudanel 3099 B.D. de pulpa blanca (y sus sinonimias Sudanel 2 y La Escola), Sudanel 1 y Generoso (La Rioja) con un alto coeficiente de mezcla (0,99). Otras variedades con coeficientes de afiliación más bajos como Pigat Susagna (0,60) y Fraga (0,45), de pulpa blanca, presentan sobre todo afinidad por el grupo 6, grupo muy extenso y variado donde se encuentran una gran cantidad de variedades de Aragón y Navarra, y se encuentran la mayoría de variedades de pulpa blanca, dando sentido a ese porcentaje. Destaca la presencia de La Escola (número de accesión 5340), que no aparece en el mismo clúster que su duplicada. Además, *PLINK* afirma una relación de sinonimia entre esta variedad, Sudanel 3099 y Sudanel 2, como se ha indicado anteriormente, por lo que esta accesión podría ser en realidad Sudanel. En este clúster existen variedades tanto de pulpa blanca como amarilla, pero todas son de textura no fundente, por lo que es muy probable que la variedad Rincal, presente en este clúster y una de las variedades sin documentar, sea de textura no fundente también.

En el clúster 3 (color naranja), se agrupan variedades tradicionales murcianas del grupo de los calabaceros (C). Son las accesiones Calabacero, C. Rincon, C. Deleite, C. Soto, C. Rancho y C. Candelo, consideradas clones según los resultados de *PLINK*. En este clúster también encontramos la variedad de Zaragoza Campillo Rocho, la cual también presenta

bastante afinidad (0,37) por el clúster 7, donde también se encuentran variedades murcianas del grupo de Marujas y Jeronimos, y alguna otra variedad aragonesa. Esto resulta interesante, ya que Campillo Rocho parece estar genéticamente más cerca de las variedades murcianas que de las aragonesas.

El clúster 4 (color morado), se caracteriza por agrupar todas las variedades de paraguayos presentes en este trabajo. Además, son casi en su totalidad variedades aragonesas. La única que no procede de Aragón, es el Paraguayo T Robert, de Lérida, con un coeficiente de mezcla más bajo (0,53). Sin embargo, muestra afinidad destacable por el clúster 6, lo cual es lógico, ya que en ese clúster se agrupan la mayoría de variedades catalanas. Existen accesiones de textura fundente y no fundente y de pulpa amarilla y blanca, por lo que lo destacable en este clúster sería la presencia predominante de paraguayos, lo que podría indicar que se represente la forma plana (en inglés *flat shape*) del fruto en este clúster. En este clúster se encuentran las accesiones Paraguayo Almudi y Paraguayo Almunia, de las que no tenemos descripción morfológica. Sin embargo, es probable que presenten características morfológicas parecidas a las de las variedades de paraguayos documentadas.

En el clúster 5 (color azul grisáceo), se agrupan variedades de orígenes variados, destacando la variedad brasileña Pepita, de la que no hay descripción previa. Sin embargo, todas las variedades documentadas de este clúster, son de pulpa amarilla y textura no fundente, lo que puede significar que Pepita también presenta estas características morfológicas. La única accesión que presenta un coeficiente de afiliación por debajo de 0,9 es Brasileño Elipe, variedad murciana. Muestra también una afinidad notable (0,39) por el clúster 6, donde solo se encuentra una variedad murciana, Jerónimo Espuña, la cual, a su vez, muestra algo de afinidad por este clúster. En este grupo se incluye la que parece ser la accesión verdadera de la variedad La Escola (con número de accesión 5093).

El clúster 6 (color amarillo), es el más extenso y diverso encontrado en este trabajo. Todas las variedades documentadas incluidas en él son no fundentes, la mayoría aragonesas, pero en este grupo encontramos la totalidad de variedades de Navarra en el presente trabajo, y la mayor parte de las variedades catalanas de este proyecto. La mayoría de las variedades de pulpa blanca documentadas y presentes en este trabajo, pertenecen a este clúster. En general, las variedades que presentan mayor coeficiente de afiliación en este grupo, son las variedades de pulpa blanca y no fundente, como Valentín B.D., los clones de Pomar o Tambarria B.D. entre otras. En este clúster también se encuentran 2 variedades americanas, Andora (0,65) y Vivian (0,46), que presentan una afinidad destacable por el clúster 7 (0,25 y 0,37, respectivamente), donde se agrupan variedades tradicionales murcianas. El porcentaje de mezcla es diferente entre las variedades de Zaragoza, así tenemos variedades con coeficientes de afiliación muy altos (Escolapio, Risol y Maluenda) y otras variedades con coeficientes menores y con afinidad notable con el clúster 2 (más de 0,3) como Gallur, Villamayor B.D. o CD-RCH, entre otras, siendo el clúster 2 donde se agrupan la mayoría de variedades de esa zona. Dos de las variedades, Pavía Blanca y CD-RCH, se encuentran sin documentar, sin embargo, las características que reúnen las

variedades de este clúster, indican que podría tratarse de variedades de textura no fundente.

El clúster 7 (color rosa), se caracteriza por agrupar variedades tradicionales murcianas del grupo de los Jeronimos (J.) y las Marujas (M.). En este clúster encontramos accesiones como M. Argos y sus clones (M. Alquibla, San Jaime, M. Tradición y M. Tejar) y J. Ortiz y sus clones (J. Copia, J. Prasio y M. Perfeccion). Estas accesiones de origen murciano, presentan un coeficiente de afiliación muy alto a este clúster, lo que pone de manifiesto una diferencia genética notable con el grupo de los Calabaceros, que también presentan un coeficiente de afiliación alto para su clúster. Destaca sin embargo J. AD, que sí que presenta afinidad destacable (0,37) por el clúster 3, donde se agrupan los Calabaceros. Por último, llama la atención la presencia de las variedades Roig de Aitona (Lérida) y Pavía Amarilla de Tolosa (Zaragoza), las cuales, aunque muestran un coeficiente de afiliación para este clúster más bajo que las variedades murcianas, parecen estar genéticamente más cerca de estas que de las de sus lugares de origen. Pavía Amarilla de Tolosa muestra bastante afinidad (0,47) por el clúster 6, algo normal, ya que allí se agrupan gran parte de las variedades de Zaragoza. Además, es la única variedad sin documentar de este grupo, por lo que es muy probable que también sea una variedad de pulpa amarilla y de textura no fundente. Roig de Aitona por su parte, muestra algo de afinidad por el clúster 5, donde encontramos solo otra variedad de Lérida (La Escola).

Los resultados obtenidos en el análisis de la estructura poblacional, manifiestan ciertas diferencias con los orígenes geográficos considerados al inicio de este trabajo. En el trabajo de Pérez y col., (2020) los grupos iniciales se mantuvieron casi iguales tras los análisis de estructura de la población. Por su parte, Aranzana y col., (2010), estudiaron la estructura poblacional de 224 variedades de melocotonero genotipadas con microsátélites y encontraron tres grupos principales basados principalmente en las características del fruto de estas variedades. Posteriormente, Font i Forcada y col., (2012) utilizando 94 variedades de melocotonero de la EEAD (algunas de ellas utilizadas en este trabajo) y 40 marcadores SSR encontraron solamente 2 subpoblaciones, variedades locales y variedades modernas, con mezcla genética dentro de ambos grupos. En otro trabajo más reciente, esas 2 subpoblaciones fueron observadas usando el mismo set de muestras y una gran cantidad (4.558) de SNPs (Font i Forcada y col., 2019). Las diferencias con nuestros resultados se deben principalmente a las diferentes accesiones utilizadas y, probablemente, al mayor número de SNPs proporcionados por la versión segunda del chip de melocotonero.

Finalmente, la información obtenida en este trabajo servirá para la disección de caracteres importantes como la resistencia a enfermedades en estudios de asociación de genoma completo (en inglés *Genome-Wide Association Studies* (GWAS)). La presencia de estructura poblacional en la población de mapeo en los estudios de asociación puede conducir a falsos positivos en las asociaciones entre marcadores y caracteres (Eltaher y col., 2018). Por lo tanto, probar la estructura de la población subyacente es el primer paso para realizar GWAS para identificar una verdadera asociación entre los marcadores y caracteres y los genes subyacentes que controlan estos caracteres.

6. Conclusiones

6.1. Conclusiones Generales

Tras los análisis realizados sobre la diversidad genética y la estructura poblacional de nuestro conjunto de accesiones se pueden sacar varias conclusiones.

1. La nueva versión del chip de genotipado de melotocontero es una herramienta eficiente, en términos de coste por muestra e información obtenida, para los estudios genéticos en esta especie. Así el 55.84% de los SNPs del *Illumina Infinium BeadChip* 18K se han mostrado como útiles en este proyecto.
2. Se ha obtenido una descripción detallada de la diversidad genética presente en nuestro panel de accesiones, que no se habían estudiado anteriormente con este nuevo chip.
3. Todo parece indicar que existen gran cantidad de individuos idénticos o clones en el banco de germoplasma nacional lo que representa un gran coste de mantenimiento y una mala gestión del espacio disponible. Se han detectado 44 individuos que se podrían condensar en 15 individuos al ser clones entre muchos de ellos.
4. Los errores en el registro de algunas accesiones parecen evidentes, habiendo accesiones con nombres duplicados. Se han detectado 3 errores de registro, viendo que individuos que se consideraban duplicados, no lo son en realidad (La Escola), y que individuos de variedades con distintas características morfológicas según la información bibliográfica (estas características deben ser comprobadas en el banco de germoplasma), han resultado ser clones (Tambarria B.D. y Rojo de Azagra).
5. Los resultados han mostrado un mayor intercambio de materiales entre las zonas del norte del país, y un menor con la zona de Murcia, que mantiene un fondo genético diferente al resto, e incluso con las accesiones extranjeras.
6. La estructura genética de la población parece indicar unas diferencias basadas en las características del fruto, no estando correlacionada con los orígenes geográficos de las accesiones.

6.2. Futuros pasos y posibles líneas de investigación

1. El 19,80% de los SNPs del *Illumina Infinium BeadChip* 18K clasificados como Cambiado-Homo pueden ser de utilidad realizando un análisis más profundo de ellos.
2. Reorganizar la base de datos del banco de germoplasma y el mismo banco, con la nueva información de clones y errores de registro detectados.
3. La estructura de la población obtenida y descrita, sirve como base para la realización de estudios de asociación de caracteres de interés agronómico, como resistencia a enfermedades como Corky Spot.
4. Profundizar más en el origen de las accesiones debido al parecido genético entre accesiones del norte y las extranjeras y determinar que accesiones sería eficiente adquirir para aumentar la variabilidad del banco de germoplasma.

7. Referencias

- Adaskaveg J.E., Schnabel G. y Förster H. (2008).** Diseases of Peach Caused by Fungi and Fungal-like Organisms: Biology, Epidemiology and Management. En D. R. Layne y D. Bassi (Eds.), *The Peach: botany, production and uses* (pp. 353-403). CABI.
- Agropack (2020).** *Aprovechamiento de residuos de fruta de hueso para su valorización como subproductos para la fabricación de packaging y envoltorios de uno agrícola y otros.* Recuperado de <https://agropack.es/>
- Arab, M. M., Marrano, A., Abdollahi-Arpanahi, R., Leslie, C. A., Askari, H., Neale, D. B. y Vahdati, K. (2019).** Genome-wide patterns of population structure and association mapping of nut-related traits in Persian walnut populations from Iran using the Axiom J. regia 700K SNP array. *Scientific reports*, 9(1). <https://doi.org/10.1038/s41598-019-42940-1>
- Aranzana, M. J., Abbassi, E. K., Howad, W. y Arús, P. (2010).** Genetic variation, population structure and linkage disequilibrium in peach commercial varieties. *BMC genetics*, 11(1), 69. <https://doi.org/10.1186/1471-2156-11-69>
- Ayala Degado, J.M. (1984).** Variedades de Melocotonero. *Hojas Divulgadoras del Ministerio de Agricultura, Pesca y Alimentación*, 17(84).
- Badenes, M. L., Cambra, M., López, M. M., Batlle, I., Iglesias, I., Aranzana, M. J., ... y Rubio-Cabetas, M. J. (2015).** A peach germplasm collection for increasing the genetic diversity in European breeding programs. <https://doi.org/10.17660/ActaHortic.2015.1084.15>
- Badenes, M. L., Zuriaga E. (2016).** Las variedades locales en la mejora genética de plantas. En J.I. Ruiz de Galarreta, J. Prohens, y R. Tierno. (Eds.) *Las variedades locales en la mejora genética de plantas* (pp. 239-259). Vitoria-Gasteiz, España: Gobierno Vasco.
- Bañuls, L. (31 de enero de 2019).** Marcadores Moleculares. *Curso Nacional de Mejora Genética de Plantas.* Universidad Almería, Almería, España.
- Bernard, A., Marrano, A., Donkpegan, A., Brown, P. J., Leslie, C. A., Neale, D. B., ... y Dirlwanger, E. (2020).** Association and linkage mapping to unravel genetic architecture of phenological traits and lateral bearing in Persian walnut (*Juglans regia* L.). *BMC genomics*, 21(1). <https://doi.org/10.1186/s12864-020-6616-y>
- Bouhadida, M., Moreno, M. Á., Gonzalo, M. J., Alonso, J. M., & Gogorcena, Y. (2011).** Genetic variability of introduced and local Spanish peach cultivars determined by SSR markers. *Tree genetics & genomes*, 7(2), 257-270.
- Bretting, P. K. y Widrechner, M. P. (1995).** Genetic markers and plant genetic resource management.
- Brigham & Women's Hospital, Harvard Medical School. (25 de enero de 2017).** *PLINK... Whole genome association analysis toolset.* <http://zzz.bwh.harvard.edu/plink/>
- Cambra, R. (1984).** Denominaciones del melocotón. *Boletines de la Estación Experimental de Aula Dei Zaragoza*, (13), 79-81.
- Chavez, D. J., Beckman, T. G., Werner, D. J., & Chaparro, J. X. (2014).** Genetic diversity in peach [*Prunus persica* (L.) Batsch] at the university of florida: past, present and future. *Tree Genetics and Genomes*, 10(5). <https://doi.org/10.1007/s11295-014-0769-2>
- Chen, W. H. (1994).** *Illustrated Handbook of Ancient Relics of Chinese Agriculture.* Nanchang, China: Jiangxi Technology Press.
- CITA (Centro de Investigación y Tecnología Agroalimentaria de Aragón)(2020).** Banco de Germoplasma Hortícola de Zaragoza. <https://bghz.cita-aragon.es/>
- Cos-Terrer, J. (s.f).** Mejora Genética de Melocotón y Nectarina. *Instituto Murciano de I+D Agrario y Alimentario (IMIDA).* Murcia, España.

- Crisosto C.H. y Costa G. (2008).** Preharvest Factors Affecting Peach Quality. En D. R. Layne y D. Bassi (Eds.), *The Peach: botany, production and uses* (pp. 536-547). CABI.
- Crisosto, C.H. y Labavitch, J.M. (2002).** Developing a quantitative method to evaluate peach (*Prunus persica*) flesh mealiness. *Postharvest Biology and Technology*, 25(2), 151-158. [https://doi.org/10.1016/S0925-5214\(01\)00183-1](https://doi.org/10.1016/S0925-5214(01)00183-1)
- De Vicente, M. C., Lopez, C., y Fulton, T. (2004).** Genetic diversity analysis with molecular marker data: learning module. *International Plant Genetic Resources Institute (IPGRI)*.
- Di Guardo, M., Micheletti, D., Bianco, L., Koehorst-van Putten, H. J., Longhi, S., Costa, F., ... y van de Weg, E. W. (2015).** ASSIsT: an automatic SNP scoring tool for in-and outbreeding species. *Bioinformatics*, 31(23), 3873-3874.
- Doyle J.J. y Doyle J.L. (1987).** A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, (19), 11-15.
- Durán, S. (1993).** *Melocotoneros, nectarina y pavías. Portainjertos y variedades*. Barcelona, España: Fundación La Caixa.
- Durand, E., Jay, F., Gaggiotti, O. E. y François, O. (2009).** Spatial inference of admixture proportions and secondary contact zones. *Molecular biology and evolution*, 26(9). <https://doi.org/10.1093/molbev/msp106>
- Eltaher, S., Sallam, A., Belamkar, V., Emara, H. A., Nower, A. A., Salem, K. F., ... y Baenziger, P. S. (2018).** Genetic diversity and population structure of F3: 6 Nebraska winter wheat genotypes using genotyping-by-sequencing. *Frontiers in genetics*, 9, 76. <https://doi.org/10.3389/fgene.2018.00076>
- Excoffier, L. y Heckel, G. (2006).** Computer programs for population genetics data analysis: a survival guide. *Nature Reviews Genetics*, 7(10). (pp. 745-758).
- FAO (Organización de las Naciones Unidas para la Alimentación y la Agricultura). (2018).** *Principales países a nivel mundial según el promedio de producción de melocotón y nectarina (2008-2018)*. Recuperado de <http://www.fao.org/faostat/es/#data/QC/visualize>
- FAO (Organización de las Naciones Unidas para la Alimentación y la Agricultura). (2018).** *Producción de melocotón y nectarina en España (2008-2018)*. Recuperado de <http://www.fao.org/faostat/es/#data/QC>
- FAO (Organización de las Naciones Unidas para la Alimentación y la Agricultura). (2018).** *Área cosechada de melocotón y nectarina en España (2008-2018)*. Recuperado de <http://www.fao.org/faostat/es/#data/QC>
- Fernandez i Marti, A., Athanson, B., Koepke, T., Font i Forcada, C., Dhingra, A. y Oraguzie, N. (2012).** Genetic diversity and relatedness of sweet cherry (*Prunus avium* L.) cultivars based on single nucleotide polymorphic markers. *Frontiers in Plant Science*, 3, 116.
- Fernández, M. E., Goszczynski, D. E., Lirón, J. P., Villegas-Castagnasso, E. E., Carino, M. H., Ripoli, M. V., ... y Giovambattista, G. (2013).** Comparison of the effectiveness of microsatellites and SNP panels for genetic identification, traceability and assessment of parentage in an inbred Angus herd. *Genetics and molecular biology*, 36(2). (pp.185-191).
- Fischer, M. C., Rellstab, C., Leuzinger, M., Roumet, M., Gugerli, F., Shimizu, K. K., Holderegger, R., & Widmer, A. (2017).** Estimating genomic diversity and population differentiation - an empirical comparison of microsatellite and SNP variation in *Arabidopsis halleri*. *BMC Genomics*, 18(1). <https://doi.org/10.1186/s12864-016-3459-7>
- Font i Forcada, C. F., Gradziel, T. M., Gogorcena, Y., y Moreno, M. Á. (2014).** Phenotypic diversity among local Spanish and foreign peach and nectarine [*Prunus persica* (L.) Batsch] accessions. *Euphytica*, 197(2), 261-277. <http://dx.doi.org/10.1007/s10681-014-1065-9>
- Font i Forcada, C., Guajardo, V., Chin-Wo, S. R. y Moreno, M. A. (2019).** Association mapping analysis for fruit quality traits in *Prunus persica* using SNP markers. *Frontiers in plant science*, 9, 2005. <https://doi.org/10.3389/fpls.2018.02005>

- Font i Forcada, C., Oraguzie, N., Igartua, E., Moreno, M. A. y Gogorcena, Y. (2012). Population structure and marker-trait associations for pomological traits in peach and nectarine cultivars. *Tree genetics & genomes*, 9(2). <https://doi.org/10.1007/s11295-012-0553-0>
- Greenbaum, G., Templeton, A. R., Zarmi, Y., y Bar-David, S. (2014). Allelic richness following population founding events - A stochastic modeling framework incorporating gene flow and genetic drift. *PLoS ONE*, 9(12). <https://doi.org/10.1371/journal.pone.0115203>
- Guillot, G., Leblois, R., Coulon, A., y Frantz, A. C. (2009). Statistical methods in spatial genetics. *Molecular Ecology*, 18(23), 4734-4756.
- Hancock, J. F., Scorza, R., y Lobos, G. A. (2008). Peaches. En J.F Hancock (Ed.) *Temperate fruit crop breeding* (pp. 265-298). Dordrecht, Países Bajos: Springer.
- Heaton, M. P., Harhay, G. P., Bennett, G. L., Stone, R. T., Grosse, W. M., Casas, E., ... y Laegreid, W. W. (2002). Selection and use of SNP markers for animal identification and paternity analysis in US beef cattle. *Mammalian genome*, 13(5). <https://doi.org/10.1007/s00335-001-2146-3>
- Heaton, M. P., Keen, J. E., Clawson, M. L., Harhay, G. P., Bauer, N., Shultz, C., ... y Laegreid, W. W. (2005). Use of bovine single nucleotide polymorphism markers to verify sample tracking in beef processing. *Journal of the American Veterinary Medical Association*, 226(8). <https://doi.org/10.2460/javma.2005.226.1311>
- Iglesias, I. y Alegre, S. (2014). Melocotonero. En J.J. Hueso y J. Cuevas (Ed.), *La fruticultura del siglo XXI en España* (pp. 125-154). España: Cajamar Caja Rural.
- Illumina, Inc. (2005). Illumina GenCall Data Analysis Software. <https://www.illumina.com/>
- Illumina, Inc. (2014). Infinium Genotyping Data Analysis-Technical Note. <https://www.illumina.com/>
- Illumina, Inc. (2020). Convert array data into meaningful results. <https://www.illumina.com/>
- InnoVine. (2016). Glosario de Genética. <http://www.innovine.eu/es/glosario-de-genetica.html>
- IVIA (Institut Valencià d'Investigacions Agràries). (2018). *Memoria de actividades 2018*. <http://www.ivia.gva.es/va/>
- Jost, L., Archer, F., Flanagan, S., Gaggiotti, O., Hoban, S., & Latch, E. (2018). Differentiation measures for conservation genetics. *Evolutionary Applications*, 11(7). <https://doi.org/10.1111/eva.12590>
- Keenan, K. (2017). *DiveRsity v1.9.90 Help Manual* [Archivo PDF]. <https://cran.r-project.org/web/packages/diveRsity/vignettes/diveRsity.pdf>
- Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., y Prodöhl, P. A. (2013). DiveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, 4(8), 782–788. <https://doi.org/10.1111/2041-210X.12067>
- LaFramboise, T. (2009). Single nucleotide polymorphism arrays: A decade of biological, computational and technological advances. *Nucleic Acids Research*, 37(13). <https://doi.org/10.1093/nar/gkp552>
- Lavi, U., Akkaya, M., Bhagwat, A., Lahav, E. y Cregan, P. B. (1994). Methodology of generation and characteristics of simple sequence repeat DNA markers in avocado (*Persea americana M.*). *Euphytica*, 80(3). (pp. 171-177).
- Lee, M. (1995). DNA markers and plant breeding programs. *Advances in agronomy*, 55, (pp. 265-344). Academic Press.
- Llácer, G., Alonso, J. M., Rubio, M. J., Batlle, I., Iglesias, I., Vargas, F. J. y Badenes, M. L. (2009). Situación del material vegetal de melocotonero utilizado en España. *Itea*, (105), 67-83.
- Ma, L., Ji, Y. J. y Zhang, D. X. (2015). Statistical measures of genetic differentiation of populations: Rationales, history and current states. *Current Zoology*, 61(5). <https://doi.org/10.1093/czoolo/61.5.886>
- MAPA (Ministerio de Agricultura, Pesca y Alimentación) (2018). *Superficies y producciones de cultivos frutales de fruto fresco no cítricos: Resumen nacional del rendimiento, producción y destino, 2018*. Recuperado de <https://www.mapa.gob.es/es/estadistica/temas/estadisticas-agrarias/agricultura/superficies-producciones-anales-cultivos/>

- MAPA (Ministerio de Agricultura, Pesca y Alimentación). (2018).** *Fruta de Hueso: Análisis Campaña 2018* [Archivo PDF]. https://www.mapa.gob.es/es/agricultura/temas/producciones-agricolas/181219analisisexpostfdh2018_tcm30-500906.pdf
- MAPA (Ministerio de Agricultura, Pesca y Alimentación). (2019).** *Fruta de Hueso. Campaña 2019* [Archivo PDF]. https://www.mapa.gob.es/es/agricultura/temas/producciones-agricolas/190524frutadehuesoiniciocampaña2019_tcm30-510574.pdf
- MAPA (Ministerio de Agricultura, Pesca y Alimentación). (2020).** *Boletín de fruta de Hueso. Campaña 2020* (18) [Archivo PDF]. https://www.mapa.gob.es/es/agricultura/temas/producciones-agricolas/boletinfrutadehueso2020n18-2020_tcm30-543875.pdf
- Markovtsova, L., Marjoram, P. y Tavaré, S. (2000).** The effects of rate variation on ancestral inference in the coalescent. *Genetics*, 156(3).
- Martínez M. C., Helguera M. y Carrera A. (2010).** Marcadores moleculares. En G. Levitus, V. Echenique, C. Rubinstein, E. Hopp y L. Mroginski (Eds.), *Biotecnología y Mejoramiento Vegetal II* (pp.70-85). Buenos Aires, Argentina: Instituto Nacional de Tecnología Agropecuaria.
- McNeely, J. A., Miller, K. R., Reid, W. V., Mittermeier, R. A. y Werner, T. B. (1990).** *Conserving the world's biological diversity*. International Union for conservation of nature and natural resources.
- Meirmans, P. G., & Hedrick, P. W. (2011).** Assessing population structure: FST and related measures. *Molecular Ecology Resources*, 11(1). <https://doi.org/10.1111/j.1755-0998.2010.02927.x>
- Micheletti, D., Dettori, M. T., Micali, S., Aramini, V., Pacheco, I., Linge, C. D. S., ... y Lambert, P. (2015).** Whole-genome analysis of diversity and SNP-major gene association in peach germplasm. *PLoS one*, 10(9). <https://doi.org/10.1371/journal.pone.0136803>
- Morales, M. C. (2001).** Variedades y Calidad de las frutas de hueso en Aragón. Zaragoza, España: apeph.
- Navarro, J. R. (1995).** Variedades del melocotón de carne dura. *Vida rural*, 19, 62-66.
- Nei, M. (1973).** Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*, 70(12(I)). <https://doi.org/10.1073/pnas.70.12.3321>
- Nuez, F., Carrillo, J. M. y Lozano, R. (2002).** *Genómica y mejora vegetal*. Mundi-Prensa.
- Oria-Almudí, R. A. (2001).** Mantenimiento postcosecha de la calidad del Melocotón de Calanda. *Informaciones técnicas: Dirección General de Tecnología Agraria*, 96, 1-20.
- Peakall, R., y Smouse, P. (2009).** GenAIEx Tutorials-Part 1: Introduction to Population Genetic Analysis. *Australian National University, Australia*.
- Pérez, V., Larrañaga, N., Abdallah, D., Wünsch, A. y Hormaza, J. I. (2020).** Genetic Diversity of Local Peach (*Prunus persica*) Accessions from La Palma Island (Canary Islands, Spain). *Agronomy*, 10(4), 457.
- Pérez-Zabala, J. A. (2007, diciembre).** Studies on the genus *Prunus* (Rosaceae) in the Neotropics: taxonomic and nomenclature changes for Colombia. *Anales del Jardín Botánico de Madrid* (64),2. (pp. 177-190).
- Pizarro, M. G. (2018).** *Comparación de efectos epistáticos en interacción de polimorfismos (SNPs) y genes en enfermedades complejas* [Tesis de Maestría, Universitat Oberta de Catalunya]. <http://hdl.handle.net/10609/73206>
- Pritchard, J. K., Stephens, M. y Donnelly, P. (2000).** Inference of population structure using multilocus genotype data. *Genetics*, 155(2).
- Pritchard, J. K., Wen, X. y Falush, D. (2010).** Documentation for STRUCTURE software, version 2.3. Universidad de Chicago, Chicago, Estados Unidos.
- Purcell, S. (2012).** PLINK (1.07) Documentation.
- RAI (Red de Apoyo a la Investigación). (2016).** *Manual de uso del fluorómetro Qubit 3.0 Life Technologies Agosto 2016*. Recuperado de <http://rai.unam.mx/>
- Ramina, A., Tonutti, P. y McGlasson, W. (2008).** Ripening, Nutrition and Postharvest Physiology. En D. R. Layne y D. Bassi (Eds.), *The Peach: botany, production and uses* (pp. 550-567). CABI.

- Scorza, R. y Okie, W.R. (1990).** Peaches (Prunus). En Moore, J.N. y Ballington, J.R. Jr (Eds.) *Genetic resources of temperate fruit and nut crops. Acta Horticulturae*, 290, (pp. 177–231).
- SCT UAB (Serveis Científicotècnics de la Universitat Autònoma de Barcelona). (2020).** *Qubit*. <https://sct.uab.cat/genomica-bioinformatica/es/content/qubit-0>
- Staaaf, J., Vallon-Christersson, J., Lindgren, D., Juliusson, G., Rosenquist, R., Höglund, M., Borg, Å. y Ringnér, M. (2008).** Normalization of Illumina Infinium whole-genome SNP data improves copy number estimates and allelic intensity ratios. *BMC Bioinformatics*, 9. <https://doi.org/10.1186/1471-2105-9-409>
- Thermo Fisher Scientific. (2016).** *NanoDrop One Guía del Usuario* [Archivo PDF]. <http://tools.thermofisher.com/content/sfs/manuals/3091-NanoDrop-One-User-Guide-v1.3-sw-SPANISH.pdf>
- Valentini, G. y Daorden, M. E. (2011).** *Actualización del panorama varietal de durazneros y nectarinas para el norte Bonaerense*. Buenos Aires, Argentina: Instituto Nacional de Tecnología Agropecuaria (INTA).
- Val-Falcón, J. (2017).** El calcio, determinante en el comportamiento poscosecha de fruta de hueso y pepita.
- Vanderzande, S., Howard, N. P., Cai, L., Da Silva Linge, C., Antanaviciute, L., Bink, M. C., ... y Van de Weg, E. (2019).** High-quality, genome-wide SNP genotypic data for pedigreed germplasm of the diploid outbreeding species apple, peach, and sweet cherry through a common workflow. *PLoS one*, 14(6). <https://doi.org/10.1371/journal.pone.0210928>
- Verde, I., Bassil, N., Scalabrin, S., Gilmore, B., Lawley, C. T., Gasic, K., ... y Main, D. (2012).** Development and evaluation of a 9K SNP array for peach by internationally coordinated SNP detection and validation in breeding germplasm. *PLoS one*, 7(4). <https://doi.org/10.1371/annotation/33f1ba92-c304-4757-91aa-555de64a0768>
- Wambugu, P. W., Ndjiondjop, M. N. y Henry, R. J. (2018).** Role of genomics in promoting the utilization of plant genetic resources in genebanks. *Briefings in functional genomics*, 17(3). (pp.198-206).
- Wang, D. G., Fan, J. B., Siao, C. J., Berno, A., Young, P., Sapolsky, R., ... y Kruglyak, L. (1998).** Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science*, 280(5366). (pp. 1077-1082).
- Wang, Z. H., y Zhuang, E. J. (2001).** *China Fruit Monograph-Peach Flora*. Beijing, China: China Forestry Press.
- Weir, B. S. y Cockerham, C. C. (1984).** Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6). <https://doi.org/10.1111/j.1558-5646.1984.tb05657.x>
- Zhao, S., Jing, W., Samuels, D. C., Sheng, Q., Shyr, Y. y Guo, Y. (2018).** Strategies for processing and quality control of Illumina genotyping arrays. *Briefings in Bioinformatics*, 19(5). <https://doi.org/10.1093/bib/bbx012>

Anexo A. Programas de Mejora Genética en España

Tabla 9: Programas de Mejora Genética en España

País	Financiación	Programa
España	Privada	ALM Frutaria Planasia PSB Producción Vegetal Proseplan Provedo SPS Royal Primor Gravilorenzo La Tiesa
	Pública	Estación Experimental de Aula Dei (EEAD-CSIC) Instituto Valenciano de Investigaciones Agrarias (IVIA)
	Privada y Pública	IMIDA-Novamed IRTA-ASF (Fruit Future)
Francia	Privada	A&L Maillard-ASF Europépinières Star Fruits-AC Monteux-Caillet
	Pública	INRA

Italia	Privada	A. Minguzzi CAV CIV Vivai G. Batisttini V. Ossani CRA: FRU Roma + FRU Forlí Neri
	Pública	DCA-Universidad de Bologna Universidad de Florencia Universidad de Pisa
	Privada y Pública	Universidad de Milán-New Plant
Australia	Pública	Universidad de Western Sidney
Brasil	Pública	EMBRAPA Instituto Agronómico
China	Pública	Agricultural Research Institute Beijing Institute of Pomology Forestry and Fruit Research Institute Zhengzou Fruit Research Institute Horticultural Research Institute (HRI)

Estados Unidos	Privada	Burchell Nursery G. Merrill Bradford Farm & B Q Genetics P.J. Friday Sun World International, Inc. Zaiger's Genetics Inc.
	Pública	Arkansas University Clemson University Florida University Michigan University North Carolina University Rutgers University USDA-ARS Texas A & M University
Hungría	Pública	Universidad de Budapest. Facultad Agraria de Corvinus
México	Pública	Centro de Fruticultura, Colegio de Universidad Autónoma de Queretaro
Nueva Zelanda	Pública	Plant Food Research
Polonia	Pública	Research Institute os Pomology and Floriculture
Rumanía	Pública	SCPP Fruit Research Station
Sudáfrica	Pública	ARC-Infrutec

Tailandia	Pública	Royal Project Fundation Royal Agricultural Station
Taiwán	Pública	Taiwan Agricultural Research Institute

Nota. Fuente: Cos-Terrer, (s.f).

Anexo B. *Workflow* de las diferentes Tareas realizadas en este Trabajo

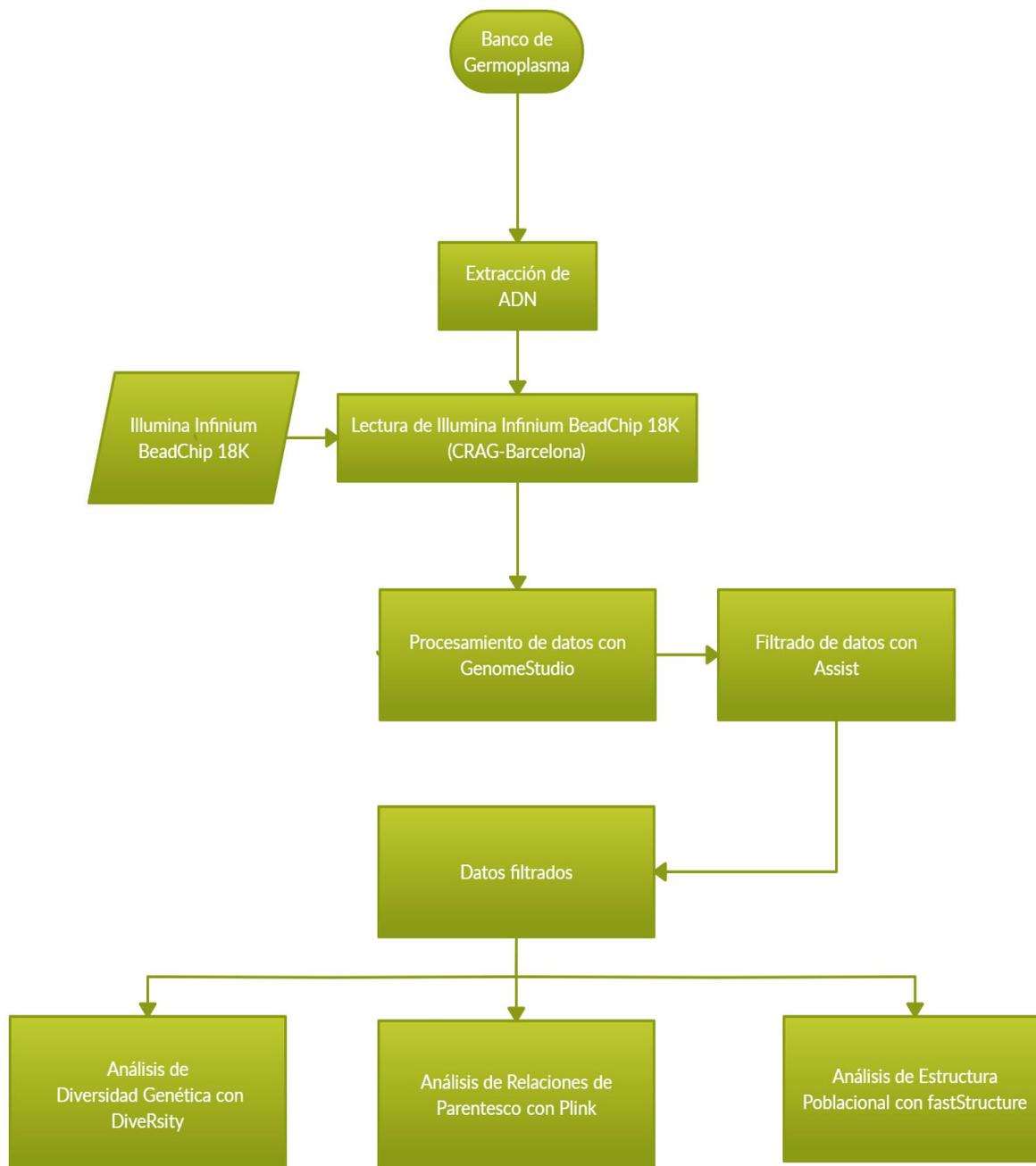


Figura 10: *Workflow* de las Tareas realizadas en este Trabajo

Nota. Fuente: Elaboración propia.

Anexo C. Clasificación de las Variedades según su Origen

Tabla 10: *Clasificación de las Variedades según su Código y Origen*

Nombre	Código	País	Origen
Sudanell 3099 B.D.	5001	España	Zaragoza
Zaragozano	5004	España	Zaragoza
San Jaime	5006	España	Lérida
Jeronimo	5009	España	Murcia
Amarillo Octubre	5017	España	Lérida
Binaced	5032	España	Huesca
Buisan	5034	España	Lérida
CD-RCH	5041	España	Zaragoza
Comodin	5044	España	Huesca
Deja-1	5050	España	Navarra
Deja-2	5051	España	Navarra
Deja-3	5052	España	Navarra
Fraga B.D.	5072	España	Huesca
Fulla	5073	España	Huesca
Gallur	5075	España	Zaragoza
Generoso	5079	España	La Rioja
La Escola	5093	España	Lérida

La Escola	5340	España	Lérida
L2-B	5101	España	Zaragoza
Maluenda (2375)	5104	España	Zaragoza
Miraflores Serapio	5122	España	Zaragoza
Montañana	5123	España	Zaragoza
Montes de Cierzo	5124	España	Navarra
Moret	5126	España	Huesca
Paraguayo Almudi	5138	España	Zaragoza
Paraguayo Almunia	5139	España	Zaragoza
Paraguayo Jota	5141	España	Zaragoza
Paraguayo San Mateo	5143	España	Zaragoza
Paraguayo T. Robert	5144	España	Lérida
Pigat	5146	España	Huesca
Pigat Susagna	5147	España	Huesca
Pomar 1	5149	España	Huesca
Pomar 2	5150	España	Huesca
Pomar 3	5151	España	Huesca
Roig de Aitona	5158	España	Lérida
Rojo-Amarillo Septiembre	5159	España	Lérida
Rojo de Azagra	5160	España	Navarra
Rojo de Tudela	5161	España	Navarra
Rojo del Rito	5162	España	Lérida

San Lorenzo	5168	España	Huesca
Sudanell 1 (2211)	5175	España	Lérida
Sudanell 2 (2349)	5176	España	Lérida
Tambarría B.D.	5186	España	Navarra
Utiel BD	5191	España	Valencia
Valdeltormo B.D.	5192	España	Teruel
Valentin B.D.	5193	España	Navarra
Miraflones	5213	España	Zaragoza
Escolapio	5231	España	Zaragoza
Tipo Rojo del Rito	5233	España	Lérida
Montamar	5234	España	Zaragoza
Campiel M.de Cierzo	5243	España	Navarra
Villamayor B.D.	5258	España	Zaragoza
Paraguayo Villamayor	5259	España	Zaragoza
Calabacero Deleite	5260	España	Murcia
Calabacero Soto	5261	España	Murcia
Calabacero Rincón	5262	España	Murcia
Jeronimo Copia	5263	España	Murcia
Jeronimo Ortiz (1)	5265	España	Murcia
Jeronimo Prasio	5266	España	Murcia
Maruja Tradición	5267	España	Murcia
Maruja Alquibla	5268	España	Murcia

Maruja Perfección	5269	España	Murcia
Maruja Tejar	5270	España	Murcia
Brasileño Elipe	5271	España	Murcia
Calabacero	5306	España	Murcia
Maruja Argos	5311	España	Murcia
Jeronimo Espuña	5312	España	Murcia
Jeronimo Balate	5314	España	Murcia
Campillo Rocho	5326	España	Zaragoza
Paraguayo Niqui	5359	España	Huesca
Blanco Tardío	5373	España	Zaragoza
Blanco Tardío	5373	España	Zaragoza
Amarillo Temprano (Ebro)/Maria Serena	5407	España	Zaragoza
Calanda Sonrosado	5410	España	Teruel
Borracho de Jarque	5421	España	Zaragoza
Duraznillo 42B	5491	España	Zaragoza
Sunmel 1	5503	España	Zaragoza
Sunmel 2	5504	España	Zaragoza
Risol	5550	España	Zaragoza
Pavia Blanca	5582	España	Zaragoza
Calabacero Rancho	5648	España	Murcia
Calabacero Candelo	5649	España	Murcia
Pavia Amarilla de Tolosa	5653	España	Zaragoza

Rincal	5664	España	Zaragoza
Andora	5018	Estados Unidos	-
Andora	5018	Estados Unidos	-
Aurelio	5023	Estados Unidos	-
Pepita	5145	Brasil	-
Vivian (5266)	5206	Estados Unidos	-
Aniversario	5369	Argentina	-

Nota. Fuente: C. M. Cantín, investigadora CITA, comunicación personal.

Anexo D. Frutos de algunas de las variedades presentes en este trabajo



Figura 11: *Variedad La Escola*

Nota. Fuente: Morales (2001).



Figura 12: *Variedad Maluenda 2375*

Nota. Fuente: Morales (2001).



Figura 13: *Variedad Rojo del Rito*

Nota. Fuente: Morales (2001).



Figura 14: *Variedad Paraguayo Jota*

Nota. Fuente: Morales (2001).



Figura 15: *Variedad Sudanell 2*

Nota. Fuente: Morales (2001).



Figura 16: *Variedad Miraflores*

Nota. Fuente: Morales (2001).

Anexo E. Procedimiento y Material necesario para la Extracción de ADN

Procedimiento

Para realizar la extracción se ha utilizado el método publicado por Doyle y Doyle (1987). Comienza por sumergir el micromortero en nitrógeno líquido para enfriar y proceder a la trituración de la muestra de hoja. Cuando se efectúa el triturado, se añade una pequeña cantidad de muestra en tubos eppendorf de 2 ml estériles, obteniendo tres eppendorf por cada muestra de hoja triturada. La extracción de tres copias por cada muestra es importante porque nos permite continuar con la metodología en el caso en el que una de las muestras se pierda o se contamine durante el proceso. Además, es recomendable la conservación de todas y cada una de ellas pues, aunque la muestra no esté contaminada, puede no cumplir los estándares de calidad exigidos posteriormente en las mediciones realizadas con *Nanodrop* y *Qubit*, en cuyo caso se procedería a medir la calidad de otro de los eppendorf con el ADN extraído de la misma muestra.

Cuando todas las muestras están trituradas, se añaden a cada tubo 750 μ l del tampón de extracción de CTAB con β -mercaptoetanol, que antes ha sido precalentado a 65°C en el baño termostático. Una vez incubado en el baño termostático a 65°C, se añaden los 750 μ l de cloroformo:isoamilalcohol (24:1) y se mezcla bien por inversión durante 2-3 minutos. Tras centrifugarlo, el sobrenadante es transferido a un tubo limpio con ayuda de una micropipeta. Al sobrenadante transferido, se le añaden 450 μ l de isopropanol frío para precipitar el ADN. Se mezcla de nuevo por inversión haciendo visible el ADN, como una masa algodonosa o pellet que flota en el interior del tubo. Se centrifuga, se elimina el sobrenadante y se lava el pellet de ADN añadiendo 500 μ l de etanol al 70% frío, invirtiendo con cuidado el tubo una o dos veces, permitiendo que el pellet flote en la solución. Se centrifuga por última vez y se elimina el sobrenadante con una pipeta con cuidado para no perder el pellet.

Tras dejar evaporar los restos de Etanol y secar el pellet, se resuspenden en 50 μ l de tampón TE de pH 8 (10 mM Tris HCl pH 8 + 1 mM EDTA) y se conservan a -4°C hasta su cuantificación. Finalizadas las extracciones, las muestras medidas con NanoDrop serán conservadas a -20°C y clasificadas en cajas para medirlas posteriormente en Qubit, que proporciona unas medidas más precisas.

Material

Para realizar la extracción de ADN según este método se necesita el siguiente material:

- a) Micromorteros de polipropileno autoclavados
- b) Tubos eppendorf 2 ml estériles, Tubos eppendorf 1,5 ml estériles y puntas de pipeta estériles.

La instrumentación utilizada consta de:

- a) Micropipetas
- b) Centrífuga de sobremesa

- c) Baño termostático a 65°C
- d) Campana de extracción de gases.

Para preparar el tampón de extracción CTAB se utilizan estos reactivos:

- a) Agua estéril filtrada (SIGMA)
- b) CTAB (Cetyltrimethylammonium Bromide) (SIGMA)
- c) Cloruro sódico
- d) EDTA (SIGMA)
- e) PVP40 (SIGMA)
- f) Tris (Trizma Base) (SIGMA)

El tampón está compuesto por:

- a) 2% CTAB
- b) 1,4 M NaCl
- c) 20 mM EDTA
- d) 100mM Tris pH 5
- e) 2% PVP 40 (opcional según Doyle y Doyle).

Posteriormente, se debe autoclavar y conservar a temperatura ambiente. Para su uso en la extracción, se añade un 1% de β -mercaptoetanol (0,2% según Doyle y Doyle) y una vez añadido este, el tampón debe ser utilizado el mismo día.

Los reactivos a utilizar en la extracción son:

- a) Nitrógeno líquido
- b) Cloroformo:Isoamilalcohol (24:1) (SIGMA)
- c) Isopropanol (SIGMA)
- j) Etanol 70%

Anexo F. Resultados obtenidos en *NanoDrop* y *Qubit*

Tabla 11: Resultados de las Medidas tomadas con *Qubit* y *Nanodrop*

	Placa	Eppendorf	Qubit	Unidad	A260/A280	A260/A230
A1	1	57	95,6	ng/μL	1,868	0,941
A2	2	61 r'	176	ng/μL	2,117	1,936
A3	3	120	158	ng/μL	2,13	2,14
A4	4	53	175	ng/μL	2,05	1,43
A5	5	96	244	ng/μL	2,057	1,492
A6	6	6-2	140	ng/μL	2,15	1,86
A7	7	50-2	95,4	ng/μL	2,044	1,265
A8	8	104.3	580	ng/μL	2,17	2,16
A9	9	114	220	ng/μL	2,154	1,879
A10	10	64 r	320	ng/μL	2,107	1,694
A11	11	123	83,8	ng/μL	2,082	1,719
A12	12	130	163	ng/μL	2,12	1,99
B1	13	33-1	136	ng/μL	2,112	1,764
B2	14	10.2	173	ng/μL	2,11	1,71
B3	15	110.2	278	ng/μL	2,08	1,47
B4	16	63 r	654	ng/μL	2,131	1,951
B5	17	99-2	200	ng/μL	1,997	1,357
B6	18	82	75,2	ng/μL	2,05	1,635
B7	19	19-2	197	ng/μL	2,09	1,575
B8	20	23	147	ng/μL	2,06	1,34
B9	21	105.3	292	ng/μL	2,08	1,57
B10	22	58.3	102	ng/μL	2,05	1,47
B11	23	13	184	ng/μL	2,13	1,85

B12	24	56	196	ng/μL	2,17	1,83
C1	25	113-2	144	ng/μL	2,09	1,764
C2	26	126	175	ng/μL	2,099	1,905
C3	27	136	111	ng/μL	2,141	1,805
C4	28	5.1	202	ng/μL	2,13	1,77
C5	29	134	114	ng/μL	2,09	1,95
C6	30	35-3	290	ng/μL	2,061	1,678
C7	31	18-2	87	ng/μL	2,091	1,654
C8	32	112-2	51,2	ng/μL	1,846	1,625
C9	33	75 r	190	ng/μL	2,121	1,765
C10	34	93	127	ng/μL	2,08	1,58
C11	35	127	142	ng/μL	2,15	1,96
C12	36	107	276	ng/μL	2,04	1,63
D1	37	109	173	ng/μL	2,08	1,64
D2	38	74	106	ng/μL	1,967	1,396
D3	39	36 r	164	ng/μL	2,14	2,09
E8	40	27 r'	272	ng/μL	2,119	1,684
D5	41	133,3	165	ng/μL	2,105	1,798
D6	42	60-2	98,4	ng/μL	2,097	1,598
D7	43	65 r	110	ng/μL	2,118	1,735
D8	44	97-2	396	ng/μL	2,017	1,742
D9	45	16	228	ng/μL	2,07	1,35
D10	46	85-2	134	ng/μL	2,113	1,519
D11	47	11-2	310	ng/μL	2,124	1,918
D12	48	117	204	ng/μL	2,15	1,99
E1	49	66 r	212	ng/μL	2,104	1,751
E2	50	101.2	410	ng/μL	2,115	1,973
E3	51	89.2	114	ng/μL	2,117	1,349

E4	52	67-3	123	ng/μL	2,101	1,904
E5	53	87-3	134	ng/μL	2,081	1,471
E6	54	119	150	ng/μL	2,12	1,95
E7	55	37-3	193	ng/μL	2,139	1,863
E8	56	42-1	174	ng/μL	2,021	1,477
E9	57	45 r'	290	ng/μL	2,157	1,918
E10	58	24	202	ng/μL	2,084	1,627
E11	59	69 r	167	ng/μL	2,079	1,446
E12	60	81.2	165	ng/μL	2,111	1,83
F1	61	129	156	ng/μL	2,13	2,09
F2	62	106,2	161	ng/μL	2,05	1,87
F3	63	80 r	200	ng/μL	2,092	1,874
F4	64	128	121	ng/μL	2,1	1,74
F5	65	102-2	404	ng/μL	2,045	2,054
F6	66	59.1		ng/μL	2,059	1,527
F7	67	131-3	140	ng/μL	2,07	1,511
F8	68	21	174	ng/μL	2,083	1,682
F9	69	132	126	ng/μL	2,06	1,58
F10	70	83 r'	197	ng/μL	2,07	1,76
F11	71	71 r	177	ng/μL	2,175	1,941
F12	72	95.2	360	ng/μL	2,083	1,749
G1	73	94 r'	152	ng/μL	2,101	1,56
G2	74	17	230	ng/μL	2,09	1,58
G3	75	4	139	ng/μL	2,02	1,33
G4	76	38	80,2	ng/μL	2,091	1,683
G5	77	115-3	185	ng/μL	2,085	1,897
G6	78	122	192	ng/μL	2,15	1,92
G7	79	118-3	304	ng/μL	2,034	1,447

G8	80	30-2	197	ng/μL	2,132	1,846
G9	81	20 r'	226	ng/μL	2,16	2,07
G10	82	9	308	ng/μL	2,15	1,89
G11	83	92.2	344	ng/μL	2,062	1,776
G12	84	54	396	ng/μL	2,07	1,54
H1	85	124-2	150	ng/μL	2,116	1,918
H2	86	7-2	177	ng/μL	2,039	1,356
H3	87	44	276	ng/μL	2,09	1,6
H4	88	116-3	258	ng/μL	2,108	1,744
H8	89	6-2	140	ng/μL	2,153	1,857
H9	90	38	80,2	ng/μL	2,091	1,683

Nota. Fuente: Datos obtenidos de *Qubit* y *NanoDrop*.

Anexo G. *Call Rate* obtenido en *GenomeStudio*

Tabla 12: *CallRate* obtenido en *GenomeStudio*

Individuos	Call Rate
Paraguayo Villamallor	0.9543584
Paraguayo Almudi	0.9545454
Paraguayo San Mateo	0.9548572
Pepita	0.9557301
Montaced (Binaced)	0.9578501
Rojo de Azagra	0.9584113
Buisan	0.9589725
La Escola	0.9589725
Aurelio	0.9590348
Amarillo Temprano (Ebro)	0.9594089
Moret	0.9594713
Paraguayo Almunia	0.9604065
Borracho de Jarque	0.9604689
Comodin	0.9605936
Brasileño Elipe	0.9606559
Montes de Cierzo	0.9609677

Paraguay T. Robert	0.9610301
Blanco Tardío	0.9612171
Aniversario	0.9617159
Paraguay Niqui	0.9627135
Blanco Tardío Rep 1	0.9627759
Paraguay Jota	0.9628382
Valdeltormo B.D.	0.9629006
Roig de Aitona	0.9632747
Pavia Blanca	0.9637112
Utiel BD	0.9637735
Pavia Amarilla de Tolosa	0.9641476
L2-B	0.96421
Gallur	0.96421
Fraga B.D.	0.9644594
Campiel M.de Cierzo	0.9645841
Rojo del Rito	0.9645841
Deja-1	0.9646465
Tipo Rojo del Rito	0.9646465
Pomar 3	0.9647712

Pomar 2	0.9650829
Deja-2	0.9652076
Pomar 1	0.96527
Valentin B.D.	0.9653323
Calanda Sonrosado	0.9653947
Zaragozano	0.9653947
Escolapio	0.9653947
Deja-3	0.9655194
Pigat Susagna	0.9655817
Montamar	0.9655817
Jeronimo	0.9657688
Maluenda (2375)	0.9658312
Villamallor B.D.	0.9659559
Duraznillo 42B	0.9660805
Rincal	0.9661429
Tambarría B.D.	0.9662676
Maruja Tradición	0.96633
Campillo Rocho	0.96633
Rojo de Tudela	0.9663923

Jeronimo Balate	0.9663923
Risol	0.9664547
Maruja Tejar	0.966517
Montañana	0.9665794
San Jaime	0.9665794
Maruja Alquibla	0.9665794
Jeronimo Copia	0.9665794
Jeronimo Ortiz (1)	0.9666417
Vivian (5266)	0.9666417
Calabacero Rancho	0.9667041
Maruja Perfección	0.9667664
Amarillo Oct.	0.9668288
Jeronimo Prasio C-15	0.9668288
CD-RCH	0.9668288
Maruja Argos	0.9668911
Sudanell 1 (2211)	0.9668911
Jeronimo España	0.9669535
Calabacero Deleite	0.9671406
Generoso	0.9672652

Sunmel 1	0.9672652
Andora	0.9672652
Andora_Copia	0.9672652
Samel 2	0.9673276
Calabacero Candelo	0.9673899
Calabacero Rincón	0.9674523
Calabacero Soto	0.9675146
La Escola Rep 1	0.9675146
Calabacero	0.9677017
Miraflores	0.9677017
Miraflores Serapio	0.9677017
Rojo-Amarillo Septiembre	0.9678888
Pigat	0.9679511
Fulla	0.9679511
Sudanell 2 (2349)	0.9679511
Sudanell 3099 B.D.	0.9680758
San Lorenzo	0.9680758
Promedio	0.964531454

Nota. Fuente: *GenomeStudio*.

Anexo H. Resultados obtenidos en *PLINK*

Este Anexo incluye la tabla obtenida en *PLINK* con los resultados de las sinonimias y la tabla elaborada a partir de estos resultados con los grupos de individuos que se han establecido como clones.

Tabla 13: *Sinonimias obtenidas en PLINK*

Código	Nombre	Código	Nombre	RT	EZ	Z0	Z1	Z2	PI_HAT
5004	Zaragozano	5231	Escolapio	OT	0	0	0	1	1
5138	Paraguayo Almudi	5143	Paraguayo San Mateo	OT	0	0	0	1	1
5138	Paraguayo Almudi	5259	Paraguayo Villamayor	OT	0	0	0	1	1
5233	Tipo Rojo del Rito	5162	Rojo del Rito	OT	0	0	0	1	1
5311	Maruja Argos	5270	Maruja Tejar	OT	0	0	0	1	1
5311	Maruja Argos	5006	San Jaime	OT	0	0	0	1	1
5311	Maruja Argos	5268	Maruja Alquibla	OT	0	0	0	1	1
5311	Maruja Argos	5267	Maruja Tradición	OT	0	0	0	1	1
5050	Deja-1	5051	Deja-2	OT	0	0	0	1	1
5143	Paraguayo San Mateo	5259	Paraguayo Villamayor	OT	0	0	0	1	1
5262	Calabacero Rincón	5261	Calabacero Soto	OT	0	0	0	1	1
5262	Calabacero Rincón	5306	Calabacero	OT	0	0	0	1	1
5262	Calabacero Rincón	5648	Calabacero Rancho	OT	0	0	0	1	1
5262	Calabacero Rincón	5260	Calabacero Deleite	OT	0	0	0	1	1
5262	Calabacero Rincón	5649	Calabacero Candelo	OT	0	0	0	1	1

5149	Pomar 1	5150	Pomar 2	OT	0	0	0	1	1
5149	Pomar 1	5151	Pomar 3	OT	0	0	0	1	1
5261	Calabacero Soto	5306	Calabacero	OT	0	0	0	1	1
5261	Calabacero Soto	5648	Calabacero Rancho	OT	0	0	0	1	1
5261	Calabacero Soto	5260	Calabacero Deleite	OT	0	0	0	1	1
5261	Calabacero Soto	5649	Calabacero Candelo	OT	0	0	0	1	1
5306	Calabacero	5648	Calabacero Rancho	OT	0	0	0	1	1
5306	Calabacero	5260	Calabacero Deleite	OT	0	0	0	1	1
5306	Calabacero	5649	Calabacero Candelo	OT	0	0	0	1	1
5503	Sunmel 1	5504	Sunmel 2	OT	0	0	0	1	1
5270	Maruja Tejar	5006	San Jaime	OT	0	0	0	1	1
5270	Maruja Tejar	5268	Maruja Alquibla	OT	0	0	0	1	1
5270	Maruja Tejar	5267	Maruja Tradición	OT	0	0	0	1	1
5122	Miraflores Serapio	5213	Miraflores	OT	0	0	0	1	1
5006	San Jaime	5268	Maruja Alquibla	OT	0	0	0	1	1
5006	San Jaime	5267	Maruja Tradición	OT	0	0	0	1	1
5150	Pomar 2	5151	Pomar 3	OT	0	0	0	1	1
5268	Maruja Alquibla	5267	Maruja Tradición	OT	0	0	0	1	1
5265	Jeronimo Ortiz (1)	5269	Maruja Perfección	OT	0	0	0	1	1
5265	Jeronimo Ortiz (1)	5266	Jeronimo Prasio	OT	0	0	0	1	1

5265	Jeronimo Ortiz (1)	5263	Jeronimo Copia	OT	0	0	0	1	1
5001	Sudanell 3099 B.D.	5176	Sudanell 2 (2349)	OT	0	0	0	1	1
5001	Sudanell 3099 B.D.	5093	La Escola	OT	0	0	0	1	1
5648	Calabacero Rancho	5260	Calabacero Deleite	OT	0	0	0	1	1
5648	Calabacero Rancho	5649	Calabacero Candelo	OT	0	0	0	1	1
5176	Sudanell 2 (2349)	5093	La Escola	OT	0	0	0	1	1
5269	Maruja Perfección	5266	Jeronimo Prasio	OT	0	0	0	1	1
5269	Maruja Perfección	5263	Jeronimo Copia	OT	0	0	0	1	1
5266	Jeronimo Prasio	5263	Jeronimo Copia	OT	0	0	0	1	1
5018	Andora	5018	Andora	OT	0	0	0	1	1
5373	Blanco Tardío	5373	Blanco Tardío	OT	0	0	0	1	1
5073	Fulla	5146	Pigat	OT	0	0	0	1	1
5073	Fulla	5159	Rojo-Amarillo Septiembre	OT	0	0	0	1	1
5141	Paraguay Jota	5359	Paraguay Niqui	OT	0	0	0	1	1
5260	Calabacero Deleite	5649	Calabacero Candelo	OT	0	0	0	1	1
5146	Pigat	5159	Rojo-Amarillo Septiembre	OT	0	0	0	1	1
5186	Tambarría B.D.	5161	Rojo de Tudela	OT	0	0	0	1	1
5421	Borracho de Jarque	5126	Moret	OT	0	0	0	1	1
5421	Borracho de Jarque	5044	Comodin	OT	0	0	0	1	1
5126	Moret	5044	Comodin	OT	0	0	0	1	1

Nota. Cada fila incluye dos variedades que presentan sinonimia. Fuente: *PLINK*.

Tabla 14: *Grupos de Sinonimias*

Grupo Sinonimia	Individuo	Origen
Grupo Sin 1	Calabacero	Murcia
	Calabacero Rancho	Murcia
	Calabacero Deleite	Murcia
	Calabacero Candelo	Murcia
	Calabacero Rincón	Murcia
	Calabacero Soto	Murcia
Grupo Sin 2	Deja-1	Navarra
	Deja-2	Navarra
Grupo Sin 3	Miraflores Serapio	Zaragoza
	Miraflores	Zaragoza
Grupo Sin 4	Jeronimo Ortiz (1)	Murcia
	Maruja Perfección	Murcia
	Jeronimo Prasio	Murcia
	Jeronimo Copia	Murcia
Grupo Sin 5	Tipo Rojo del Rito	Lérida
	Rojo del Rito	Lérida
Grupo Sin 6	Maruja Alquibla	Murcia

	Maruja Tradición	Murcia
	Maruja Argos	Murcia
	Maruja Tejar	Murcia
	San Jaime	Lérida
Grupo Sin 7	Paraguayo Almudi	Zaragoza
	Paraguayo San Mateo	Zaragoza
	Paraguayo Villamayor	Zaragoza
Grupo Sin 8	Pomar 1	Huesca
	Pomar 2	Huesca
	Pomar 3	Huesca
Grupo Sin 9	Sunmel 1	Zaragoza
	Sunmel 2	Zaragoza
Grupo Sin 10	Zaragozano	Zaragoza
	Escolapio	Zaragoza
Grupo Sin 11	Sudanell 3099 B.D.	Zaragoza
	La Escola	Lérida
	Sudanell 2 (2349)	Lérida
Grupo Sin 12	Fulla	Huesca
	Pigat	Huesca

	Rojo-Amarillo Septiembre	Lérida
Grupo Sin 13	Paraguayo Jota	Zaragoza
	Paraguayo Niqui	Huesca
Grupo Sin 14	Tambarría B.D.	Navarra
	Rojo de Tudela	Navarra
Grupo Sin 15	Borracho de Jarque	Zaragoza
	Moret	Huesca
	Comodin	Huesca

Nota. Elaboración propia a partir de los datos obtenidos con *PLINK*.

Anexo I. Resultados de *fastStructure*

En este Anexo se exponen los resultados del gráfico de probabilidad marginal y sus valores. Según el gráfico de probabilidad marginal, el número más probable de subgrupos en nuestro panel es K5. Determinar el número de clústeres en los datos genéticos de poblaciones es un tema muy difícil. Realmente, no existe ningún método que pueda hacer esto de manera perfecta. Existen algunos trabajos como el de Guillot y col. (2009) como ejemplo de que el uso incorrecto de un criterio estadístico puede producir resultados erróneos y llevar a conclusiones engañosas, independientemente del programa utilizado. Por este motivo, creemos que nuestra elección de K=7 tiene en cuenta tanto el resultado estadístico como el sentido biológico de nuestro trabajo.

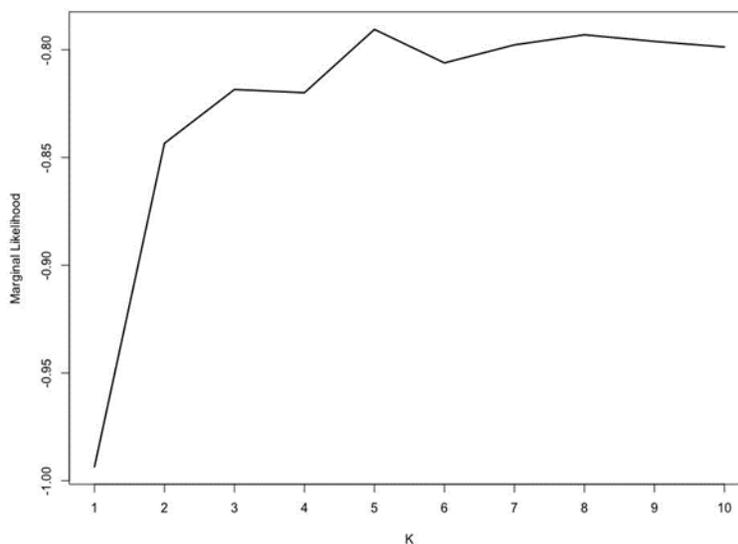


Figura 17: Gráfico de Probabilidad Marginal

Nota. Fuente: *fastStructure*.

Tabla 15: *Probabilidad Marginal de cada Clúster*

K	Probabilidad Marginal
1	-0.9935238047
2	-0.8433998401
3	-0.8184170103
4	-0.8199132350
5	-0.7905320264
6	-0.8060804662
7	-0.7977184489
8	-0.7930450285
9	-0.7961011195
10	-0.7986992079

Nota. Elaboración propia a partir de los datos obtenidos en *fastStructure*.

Anexo J. Resultados obtenidos en *Structure*

Este Anexo incluye una tabla con cada variedad asociada a su número en *Structure* y su clúster, y la representación gráfica de los clústeres.

Tabla 16: Clúster y N.º Asociado en *Structure* para cada Variedad

Nº en <i>Structure</i>	Variedad	Grupo
15	Aniversario	G1
17	Aurelio	G1
33	Rojo de Azagra	G1
36	Buisan	G1
51	Blanco Tardío	G1
61	Blanco Tardío	G1
16	Amarillo Octubre	G2
18	Miraflores Serapio	G2
26	Rincal	G2
28	Sudanell Blanco Duro	G2
31	Miraflores	G2
37	Sudanell 2 (2349)	G2
38	Generoso	G2
41	Fraga B.D.	G2
49	Duraznillo 42B	G2
54	Sudanell 1 (2211)	G2
58	Pigat Susagna	G2
69	La Escola	G2
90	Montamar	G2
8	Calabacero Rincón	G3

9	Calabacero Deleite	G3
10	Calabacero Soto	G3
11	Calabacero	G3
44	Calabacero Rancho	G3
45	Campillo Rocho	G3
76	Calabacero Candelo	G3
2	Paraguayo Almudi	G4
6	Paraguayo Almunia	G4
7	Paraguayo San Mateo	G4
21	Comodin	G4
32	Borracho de Jarque	G4
42	Binaced	G4
67	Paraguayo Niqui	G4
70	Paraguayo Jota	G4
73	Paraguayo Villamayor	G4
77	Paraguayo T Robert	G4
87	Moret	G4
47	La Escola	G5
79	Pepita	G5
80	Brasileño Elipe	G5
83	Amarillo Temprano (Ebro)/Maria Serena	G5
3	Tipo Rojo del Rito	G6
5	Deja-1	G6
12	L-2-B	G6
13	Sunmel 1	G6
14	Escolapio	G6

20	Calanda Sonrosado	G6
22	Deja-2	G6
23	Pomar 2	G6
27	Pavia Blanca	G6
29	Montes de Cierzo	G6
34	Maluenda (2375)	G6
35	Gallur	G6
39	Campiel M.de Cierzo	G6
48	Andora	G6
50	Pomar 3	G6
52	Fulla	G6
55	Risol	G6
56	Valdeltormo B.D.	G6
59	Rojo del Rito	G6
60	Villamayor B.D.	G6
62	Pomar 1	G6
63	Pigat	G6
65	CD-RCH	G6
66	Montañana	G6
68	Tambarría B.D.	G6
71	Rojo de Tudela	G6
72	Andora	G6
74	Deja-3	G6
75	Jeronimo Espuña	G6
81	Sunmel 2	G6
82	Vivian (5266)	G6

85	San Lorenzo	G6
86	Rojo-Amarillo Septiembre	G6
88	Valentin B.D.	G6
89	Utiel BD	G6
4	Maruja Argos	G7
19	San Jaime	G7
24	Maruja Alquibla	G7
25	Jerónimo Ortiz (1)	G7
30	Jerónimo AD	G7
40	Maruja Perfección	G7
43	Jerónimo Prasio	G7
46	Roig de Aitona	G7
53	Maruja Tradición	G7
57	Maruja Tejar	G7
64	Jerónimo Balate	G7
78	Pavía Amarilla de Tolosa	G7
84	Jerónimo Copia	G7

Nota. Elaboración propia a partir de los datos obtenidos con *Structure*.

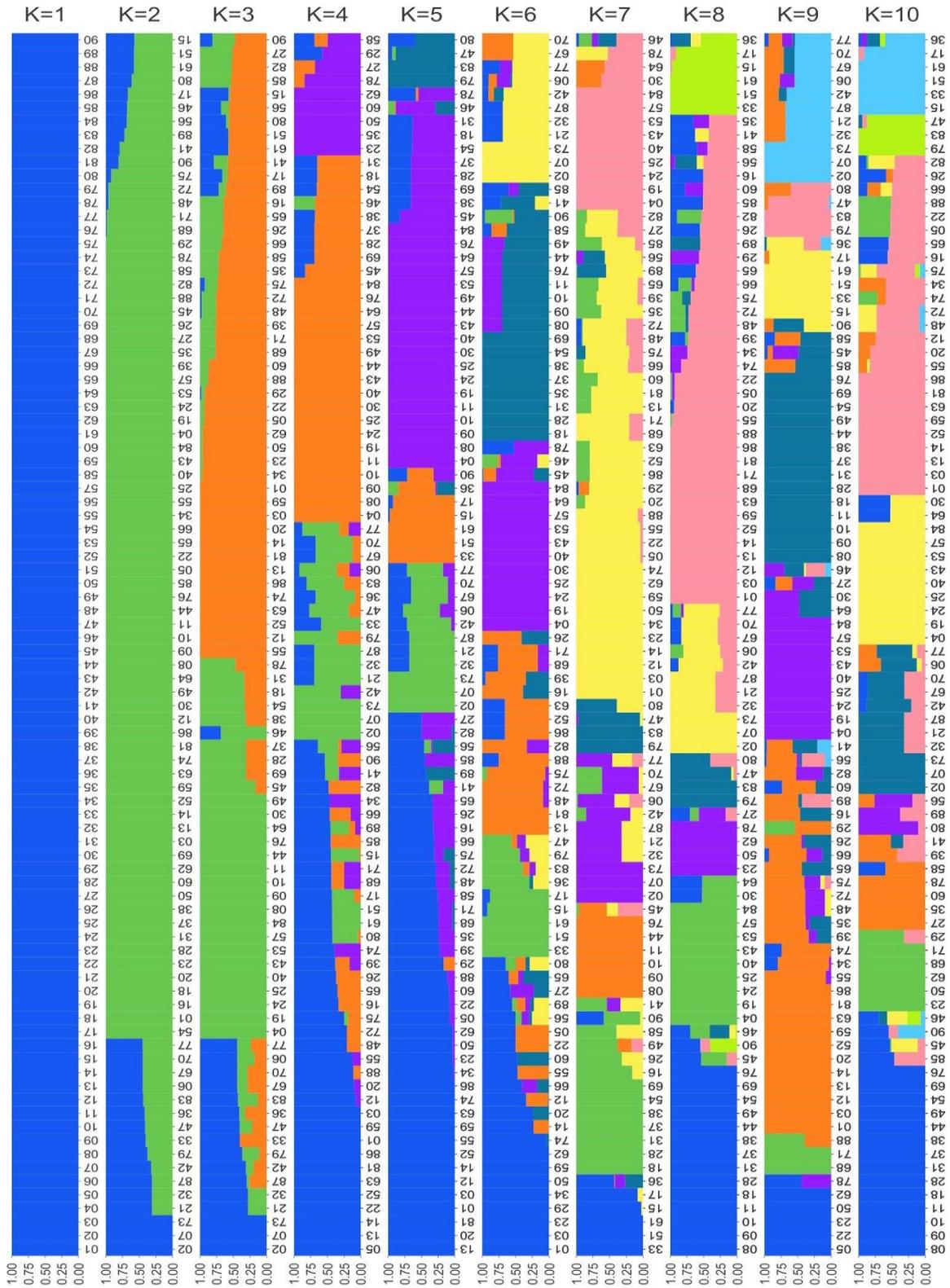


Figura 18: Representación Gráfica de cada Clúster

Nota. Fuente: Structure.

