



RIDUNAJ
Repositorio Institucional
Digital UNAJ



Universidad Nacional
ARTURO JAURETCHE

Tesinas de Grado

Walter Andrés Bosich

Caracterización de marcadores moleculares involucrados en caracteres de calidad en manzano (*Malus × domestica* Borkh.)

2024

Instituto de Ingeniería y Agronomía

Carrera: Licenciatura en Ciencias

Agrarias



Esta obra está bajo una Licencia Creative Commons.

Atribución – No comercial – Compartir igual 4.0

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Documento descargado de RID - UNAJ Repositorio Institucional Digital de la Universidad Nacional Arturo Jauretche

Cita recomendada:

Bosich, W. A. (2024). *Caracterización de marcadores moleculares involucrados en caracteres de calidad en manzano (*Malus × domestica* Borkh.)* [Trabajo final de grado, Universidad Nacional Arturo Jauretche].

<https://rid.unaj.edu.ar/handle/123456789/2879>

Universidad Nacional Arturo Jauretche
Instituto de Ingeniería y Agronomía
Centre for Research in Agricultural Genomics
Institut de Reserca i Tecnologia Agroalimentàries



Trabajo Final

Modalidad: Trabajo de Investigación

**CARACTERIZACIÓN DE MARCADORES MOLECULARES
INVOLUCRADOS EN CARACTERES DE CALIDAD EN
MANZANO (*Malus × domestica*. Borkh.)**

Estudiante: Walter Andrés Bosich

Legajo: 35.984

D.N.I: 28.365.062

e-mail: walter.bosich@gmail.com

Directora: María José Aranzana (IRTA-CRAG)

Co-directora: Nicole Prada (CRAG)

Tutora: Vanesa Marín Viegas (UNAJ)

Índice

Resumen.....	3
Palabras claves.....	4
Introducción.....	4
Situación del cultivo de manzana en España.....	5
Situación del cultivo de manzana en Argentina.....	6
Uso de marcadores moleculares.....	7
Marcadores genéticos moleculares.....	8
Mapas de ligamientos.....	9
<i>Breeding</i> Molecular.....	11
Genome-Wide Association Studies (GWAS).....	12
Banco de Germoplasma.....	13
Germoplasma de manzanos.....	14
Forma del fruto.....	15
Descriptores de los caracteres morfológicos del fruto.....	19
<i>Deep Learning</i> y <i>Maching Learning</i>	21
Objetivos generales.....	23
Objetivos específicos.....	23
Materiales y métodos.....	24
Parte I.....	24
Parte II.....	28
Resultados y discusión.....	29
Parte I: Análisis estadístico descriptivo del fenotipado de caracteres.....	29
Starch Pattern Index (SPI).....	30
Días de cosecha.....	31
Visual <i>Wax Content</i>	34
Sólidos Solubles Totales (SST).....	37
Acidez Titulable (TA).....	39
RATIO SST: TA.....	41
Parte II: Asociación fenotipo – genotipo para la forma de fruto.....	43
Forma de los frutos.....	43
Análisis de regiones cromosómicas.....	45
Análisis de SNP AX-115482211 del cromosoma 2.....	45
Análisis de los SNPs AX-115327898 y AX-115327900 del cromosoma 11.....	48
Análisis de SNP AX-115355048 en el cromosoma 13.....	53
Conclusiones.....	56
Parte I: conclusiones del fenotipado de caracteres.....	56

Parte II.....	58
Análisis de SNPs para el carácter forma de fruto	58
Bibliografía.....	59
Material suplementario	62
Escaneo y reconocimiento del software <i>Shape Analyzer</i>	63
Tablas de desequilibrio de ligamiento	63

Resumen

En los últimos años, los usos de herramientas moleculares desarrolladas en estudios de biología molecular han cobrado importancia en la investigación de características de interés comercial en especies vegetales; así como también han pasado a ser herramientas de relevancia a la hora de evaluar características de nuevos cultivares antes que estos entren en producción. Dentro de las herramientas que más se han desarrollado en los últimos años se encuentran los marcadores moleculares, principalmente los *Single Nucleotides Polymorphisms* (SNPs). Los SNPs son marcadores moleculares nucleotídicos de muy alta especificidad que permiten establecer relaciones genotípicas-fenotípicas de forma muy precisa.

La familia de las *Rosáceas* incluye una gran parte de las especies de frutas de consumo masivo a nivel mundial, por ejemplo: manzana (*Malus × domestica B.*) y pera (*Pyrus communis*). El uso de marcadores moleculares en manzana y pera ha permitido establecer relaciones entre determinados marcadores y características fenotípicas para muchas variedades comerciales; en particular para evaluar parámetros de calidad de fruta y resistencia a enfermedades. El uso de marcadores moleculares permite una evaluación de estos parámetros más fácilmente, y a la vez expandir los programas de mejoras de frutos e investigaciones académicas.

En el presente trabajo se evaluarán características de importancia comercial asociadas a calidad de fruta en variedades de manzano. Las variedades que se estudiarán son cultivares y variedades españolas locales y nacionales, e internacionales. A partir de los resultados de fenotipado obtenidos de las citadas variedades se hará un análisis genético para correlacionar los potenciales SNPs asociados a calidad de frutas.

Palabras claves

Manzana (*Malus × domestica*, Borkh), SNP, Mapeo genético, Fenotipo, Genotipo, Calidad de fruta.

Introducción

El cultivo de manzana (*Malus × domestica*, Borkh.) es uno de los cultivos de mayor importancia comercial de las regiones templadas (FAO, 2022). La producción de manzano se ubica entre los mayores cultivos frutales del mundo, solamente superado por el tomate (*Solanum lycopersicum*. L) y la uva (*Vitis vinifera* L.) (Migicovsky et al., 2021).

El manzano posee un número cromosómico diploide $2n=34$, un genoma altamente duplicado con 651 Megabases (Mbs) de tamaño, y unos 63.541 estimado de genes anotados (Daccord et al., 2017). La alta diversidad genética de los cultivares de manzana y su alta heterocigosis conllevan una amplia diversidad fenotípica (Peace et al., 2019).

Los cultivares de manzana heredaron su gran diversidad de calidades y tamaños de fruto de sus ancestros silvestres (Cornille et al., 2014). Durante la domesticación, los frutos de manzana han aumentado de tamaño y ganado otras cualidades de la

fruta, reflejando cada vez más las preferencias de los consumidores humanos (Cornille et al., 2014). Desde la domesticación del manzano en Asia Central, esta se convirtió en una especie casi panmítica separada de otras especies, debido en gran parte, a la polinización abierta aún y presentar autoincompatibilidad (los individuos no son autofecundables). A lo largo de los siglos, muchos de estos cultivares se han propagado de forma vegetativa por medio de técnicas de injerto; junto con la selección de portainjertos enanizantes, facilitando la selección y propagación de genotipos superiores derivados de la polinización abierta (Cornille et al., 2014). A pesar de la enorme cantidad de cultivares que existe, solamente una mínima fracción del total se cultivan de forma comercial (Michaela Jung et al., 2020). El uso de un número reducido de cultivares en mejoramiento genético ha reducido la diversidad genética, especialmente en cultivares comerciales (Jung et al., 2022; Migicovsky et al., 2021). Para obtener una agricultura intensiva más sostenible y adaptada a nuevas condiciones agroclimáticas de este cultivo es necesario producir nuevos cultivares con caracteres diversificados, pero sin dejar de lado los gustos y preferencia de los consumidores (Jung et al., 2022). Uno de los grandes desafíos a los que se enfrentan las industrias dependientes del cultivo de manzana es obtener nuevos cultivares que tengan características de calidad deseadas por parte de los consumidores así también nuevas características fenotípicas como, por ejemplo, resistencia a enfermedades, sequia o similares (Migicovsky et al., 2021).

Situación del cultivo de manzana en España.

En 2023, la producción de manzana en España ronda las 617.095 toneladas métricas (Tm) que se distribuyen en: 516.339 Tm de manzana de mesa y 100.756 Tm de manzanas de sidra. Las principales regiones productoras de manzana de

mesa son: Cataluña (51.7 %); Aragón (16 %); Galicia (24%); y Castilla y León (2.4%) (Ministerio de Agricultura, 2023). Las principales regiones productoras de manzanas para sidra son: Galicia (72 %); Asturias (15,4%) y País Vasco (11,5%) (Ministerio de Agricultura, 2023).

En 2023 la superficie ocupada por el cultivo de manzano se encuentra en torno 29.500 hectáreas (ha), situándose en zonas de llanura, principalmente del Valle de Ebro, y en condiciones climáticas caracterizadas por elevadas temperaturas y baja humedad ambiental en las estaciones estivales. Las principales variedades que se cultivan son Golden (53%), Gala (23%) y Red Delicious (11%) y Fuji (5%) (Bonany, 2017).

En las condiciones anteriormente citadas, caracterizadas por elevadas temperaturas y baja humedad ambiental en el período estival, la mayoría de las variedades presentan una deficiente adaptación. Ello se traduce en una menor coloración en el caso de variedades como Gala, Delicious o Fuji, mientras que en las del grupo 'Golden' la falta de firmeza constituye una limitación, como también lo son las pérdidas de cosecha por golpes de sol en 'Fuji' o 'Granny Smith' (Iglesia Castellarnau, 2017). Dichos factores limitantes, han ocasionado en las dos últimas décadas, una pérdida importante de competitividad frente a las importaciones de países con climas más apropiados como Francia o Italia (Hemisferio Norte), Chile, Brasil o Nueva Zelanda (Hemisferio Sur) (Bonany, 2017). Por estas razones es importante el desarrollo de nuevos cultivares con características adaptables a las condiciones climáticas de estas regiones de cultivo (Iglesia Castellarnau, 2017).

Situación del cultivo de manzana en Argentina.

El área productiva en Argentina representa 38.000 hectáreas implantadas con frutales, 79,5% en Río Negro, 13% en Neuquén, y 7,5 % en Mendoza (Secretaría de Agricultura, 2022). La superficie involucrada en las tres principales provincias

alcanza unas 38.450 hectáreas y en lo que respecta al número de productores, se estiman unos 1.600 totales. Las provincias de Río Negro y Neuquén suman 1.676 productores, la superficie media por explotación de ambas provincias se ubica en torno a las 21,2 hectáreas (ha). En el caso de Mendoza, si bien no se cuenta con datos directos, según el Censo Nacional Agropecuario (CNA, 2018), existirían unas 205 explotaciones para manzano (Secretaría de Agricultura, 2022).

La producción argentina es de 1 millón de toneladas de manzanas. Se exportan anualmente un promedio 100.000 toneladas de manzanas, debido a las enormes exigencias sanitarias y de calidad que debe cumplir la fruta para ingresar a mercados extranjeros. Las exportaciones de manzana del Alto Valle representan la mitad de las exportaciones de fruta fresca del país.

El 65% de la cosecha nacional de manzana corresponde a 'Red Delicious' y sus clones; el 15% corresponde a 'Gala' y sus clones, coincidiendo este porcentaje con el de 'Granny Smith'. El 5% restante se reparte entre 'Pink Lady', 'Rome Beauty', 'Golden Delicious', 'Fuji' y 'Braeburn' (Secretaría de Agricultura, 2022; Toranzo, 2016).

En Argentina se consumen en promedio 6 Kg. de manzana por habitante y por año, valores bajos comparados con China y la Unión Europea (Secretaría de Agricultura, 2022).

Uso de marcadores moleculares

Los programas de mejora genética de manzana han considerado tradicionalmente la calidad y la productividad de la fruta dentro de sus principales objetivos, sin embargo, la necesidad de nuevas variedades adaptadas a los efectos provocados por el cambio de las condiciones climáticas (escasez de agua, temperaturas más altas, enfermedades, entre otros factores) exige estrategias de mejora genética

novedosas y más eficientes. Dichas estrategias incluyen métodos novedosos de fenotipado, así como el uso de marcadores moleculares (Laurens et al., 2018). Las herramientas moleculares cobran importancia para alcanzar estos objetivos, dado que brinda información valiosa para la selección de nuevos cultivares con características fenotípicas deseadas. (Migicovsky et al., 2021).

Marcadores genéticos moleculares

Un marcador molecular genético es una secuencia de ADN de posición conocida en el genoma, que permite identificar y monitorear a nivel individual o poblacional las diferencias o similitudes entre individuos relacionados genéticamente. Los marcadores simplemente son indicadores genéticos asociados a un gen que al expresarse produce el fenotipo estudiado (Montoya Jaramillo, 2016).

Los diferentes tipos de marcadores moleculares ofrecen diversas posibilidades según el tipo de estudio. Los más utilizados actualmente son aquellos que permiten identificar en el individuo la porción genética que se hereda de ambos progenitores, en particular dos de los más utilizados son los microsatélites (*Simple Sequence Repeats* (SSRs)) y de Single Nucleotide Polymorphisms (SNPs). Como métodos para detectar marcadores moleculares se emplean tres técnicas según el método de detección: a) hibridación de la molécula de ADN; b) reacción en cadena de polimerasa (PCR); y c) secuenciación de ADN, siendo las dos últimas las más utilizadas actualmente (Montoya Jaramillo, 2016).

Los marcadores genéticos representan diferencias genéticas entre organismos o especies individuales conocidas como polimorfismos genéticos. El ligamiento es la tendencia de genes o segmentos de ADN a heredarse juntos debido a su ubicación cercana dentro del mismo cromosoma. Cuanto más cerca están dos genes o secuencias uno del otro en un cromosoma, mayor es la probabilidad de que se

hereden juntos. Los marcadores moleculares no representan a los propios genes diana, sino que actúan como indicadores. Cabe mencionar que los marcadores nunca afectan al genotipo del rasgo de interés porque sólo están situados cerca o, inclusive, dentro de los genes ligados (Collard et al., 2005).

Mapas de ligamientos

Los mapas de ligamiento son series ordenadas de genes y marcadores genéticos localizados en un cromosoma. Indican la posición y las distancias genéticas relativas entre marcadores a lo largo de los cromosomas (Collard et al., 2005; Montoya Jaramillo, 2016). Los genes o marcadores que estén próximos o estrechamente ligados se transmitirán juntos de progenitor a progenie con más frecuencia que los genes o marcadores que se encuentran más separados (Montoya Jaramillo, 2016).

La construcción de un mapa de ligamiento requiere una población de plantas segregantes. En una población de segregantes, existe una mezcla de genotipos parentales y recombinantes (Collard et al., 2005). Los progenitores seleccionados para la población de mapeo diferirán en uno o más rasgos de interés. Por lo general, los marcadores segregarán de forma mendeliana, aunque pueden encontrarse ratios de segregación distorsionados (Collard et al., 2005). El tamaño de las poblaciones utilizadas en los estudios preliminares de cartografía genética suele oscilar entre 50 y 250 individuos, pero para cartografías de alta resolución se necesitan poblaciones más grandes. Si el mapa se va a utilizar para estudios de Quantitative Trait Loci (QTL), la población a mapear debe evaluarse fenotípicamente antes del mapeo posterior de QTL (Collard et al., 2005)

La etapa final de la construcción de un mapa de ligamiento consiste en interpretar los datos de cada marcador de ADN en cada individuo de una población y realizar un análisis de ligamiento mediante programas informáticos. Los marcadores enlazados se agrupan en "grupos de enlace", que representan segmentos cromosómicos o cromosomas enteros. La precisión de la medición de la distancia genética y de la determinación del orden de los marcadores está directamente relacionada con el número de individuos estudiados en la población de mapeo. Cuanto mayor sea la distancia entre marcadores, mayor será la probabilidad de que se produzca una recombinación durante la meiosis. La distancia a lo largo de un mapa de ligamiento se mide en términos de la frecuencia de recombinación entre marcadores genéticos (Collard et al., 2005)

El análisis de QTLs confiere la asociación entre el fenotipo y el genotipo de los marcadores localizados en una región concreta del genoma o "locus". Los marcadores se utilizan para dividir la población cartográfica en diferentes grupos genotípicos basados en la presencia o ausencia de un locus marcador particular y para determinar si existen diferencias significativas entre los grupos con respecto al rasgo que se está midiendo (Collard et al., 2005; Montoya Jaramillo, 2016).

Tres métodos muy utilizados para detectar QTLs son el análisis de marcador único, el mapeo de intervalo simple y el mapeo de intervalo compuesto (Collard et al., 2005). El análisis de marcador único (también denominado "análisis de punto único") es el método más sencillo para detectar QTLs asociados a fenotipos únicos. Los métodos estadísticos utilizados para el análisis de marcador único incluyen las pruebas t , el análisis de la varianza (ANOVA) y la regresión lineal. Esta última es la más utilizada porque el coeficiente de determinación (R^2) del marcador explica la variación fenotípica derivada del QTL ligado al marcador. Este método no requiere

un mapa de ligamiento completo y puede realizarse con programas de software estadístico (Collard et al., 2005).

El uso más importante de los mapas de ligamiento es la identificación de localizaciones cromosómicas que contienen genes y QTLs asociados a rasgos de interés. La cartografía de QTLs se basa en el principio de que los genes y marcadores segregan a través de la recombinación cromosómica (denominada entrecruzamiento) durante la meiosis, lo que permite el análisis de la progenie (Collard et al., 2005).

Breeding Molecular

Las asociaciones entre marcador – rasgo y su aplicación en la selección de nuevas líneas de cultivares se utilizan en diversas especies de interés agronómico. El mapeo de marcadores y su análisis permite rápidamente definir el potencial fenotípico de la especie analizada, en función de variables fenotípicas como ser resistencia a enfermedades, forma del fruto, acidez titulable, almidón, entre otros (Montoya Jaramillo, 2016). Sin embargo, en el caso del manzano sigue estando restringido a un número limitado de rasgos asociados a un solo gen o un grupo de ellos asociados a un QTL determinado forma, y color de fruto, mientras que acidez y firmeza hay más de tres QTLs asociados (Jung et al., 2022; Michaela Jung et al., 2020; Shen et al., 2019).

La selección asistida por marcadores moleculares o SAM (en inglés, *Marker Assisted Selection*, (MAS)) es una estrategia usada en mejoramiento vegetal en la cual la detección y selección de secuencias de ADN (alelos y/o genes) mediante marcadores permite seleccionar el material vegetal más promisorio (Montoya Jaramillo, 2016).

SAM permite a los obtentores predecir rasgos fenotípicos basados en una asociación con uno o varios marcadores genéticos. El método SAM conduce a la eliminación temprana de plántulas con caracteres no deseados mucho antes de observar los fenotipos de las mismas, reduciendo el tiempo de persistencia en salas de cultivo, contenedores, o viveros, además de reducir costos de insumos y mano de obra (Montoya Jaramillo, 2016).

En los últimos años, varios trabajos han tenido como objetivo la identificación de polimorfismos de ADN asociados a rasgos del manzano. Por ejemplo, para el carácter de forma en manzano, se ha presentado una lista de 128 SNPs asociados para su validación en un panel de accesiones, que incluía variedades comerciales, selecciones avanzadas y plántulas. Algunos de los SNPs estaban altamente asociados con la forma del fruto por lo que se pueden utilizar de manera eficiente para el mejoramiento molecular (Dujak & Aranzana, 2023).

Genome-Wide Association Studies (GWAS)

Los estudios de asociación de genoma completo (*Genome-wide association studies* (GWAS)) pretenden identificar asociaciones de genotipos con fenotipos mediante pruebas de diferencias en la frecuencia alélica de variantes genéticas entre individuos ancestralmente similares, pero fenotípicamente diferentes. Los GWAS pueden tener en cuenta variantes en el número de copias o variaciones en la secuencia del genoma humano, aunque las variantes genéticas más estudiadas en los GWAS son los SNPs. Los GWAS suelen informar de bloques de SNPs correlacionados que muestran una asociación estadísticamente significativa con el rasgo de interés, conocidos como loci de riesgo genómico (Uffelmann et al., 2021).

El trabajo experimental de un GWAS implica varios pasos, incluida la recopilación de ADN e información fenotípica de un grupo de individuos. El genotipado de cada individuo utilizando las matrices GWAS disponibles o estrategias de secuenciación; el control de calidad; la imputación de variantes no tipificadas utilizando el ajuste de haplotipos y poblaciones de referencia; la realización de la prueba estadística de asociación; opcionalmente, la realización de un metaanálisis; la búsqueda de una réplica independiente; y la interpretación de los resultados mediante la realización de múltiples análisis posteriores al GWAS. En cada paso, pueden introducirse en el estudio posibles sesgos y errores, por lo que se requiere una planificación cuidadosa al establecer un GWAS, y se aconseja la adhesión a protocolos de análisis y control de calidad estandarizados (Uffelmann et al., 2021).

El genotipado de los individuos se realiza normalmente utilizando microarrays para las variantes comunes o métodos de secuenciación de nueva generación. Este tipo de genotipado es el método más utilizado para obtener datos útiles para estudios de GWAS debido al coste actual de la secuenciación de nueva generación (conocido en inglés como *High Throughput technologies* (Uffelmann et al., 2021).

A partir de los datos obtenidos, la generación de resultados fiables a partir de GWAS requiere un cuidadoso control. Herramientas como PLINK se han diseñado específicamente para analizar datos genéticos y pueden utilizarse para llevar a cabo muchos de estos pasos de control (Uffelmann et al., 2021).

Banco de Germoplasma

El germoplasma es el material genético de un individuo que puede transmitirse, sexual o somáticamente, de una generación a otra. En sentido general, el germoplasma puede representar una especie, población, variedad local, híbrido o

cultivar. La conservación del germoplasma puede adoptar muchas formas, generalmente se clasifica como "*in situ*" en áreas o granjas naturales gestionadas, o "*ex situ*" en bancos de semillas, depósitos de tejidos o jardines botánicos (Offord, 2017). El material genético debe recopilarse, almacenarse y gestionarse de manera que mantenga su utilidad (es decir, viabilidad, cantidad y diversidad) para el uso previsto. Específicamente, están compuestos por la diversidad encontrada en cualquier especie de planta moderna cultivada, domesticada y sus cultivares (actualmente en uso u obsoletos), variedades más antiguas cultivadas tradicionalmente o 'variedades locales' y las especies silvestres relacionadas, que pueden ser los parientes evolutivos inmediatos o más lejanos de las formas cultivadas (Offord, 2017).

La conservación de germoplasma en plantaciones a campo implica la recolección de material de un lugar y su transferencia y siembra en un segundo lugar. Estas técnicas de conservación implican el mantenimiento de variedades o sistemas de cultivo tradicionales por parte de agricultores o jardineros dentro de sistemas agrícolas tradicionales. Por ejemplo, las variedades locales se siembran y cosechan y, de vez en cuando, el agricultor guarda una porción de las semillas cosechadas para volver a sembrarla en temporadas posteriores. En este caso, es el agricultor quien intencionalmente conserva involuntariamente germoplasma (Offord, 2017).

Germoplasma de manzanos

Los bancos de germoplasma y colecciones de germoplasma son colecciones muy diversas de recursos fitogenéticos de una o varias especies. Su objetivo general es la conservación a largo plazo y la accesibilidad del germoplasma vegetal para los fitomejoradores, investigadores y otros usuarios. Los recursos fitogenéticos constituyen el material de partida para el mejoramiento de cultivos, y su conservación (FAO, 2014). Un ejemplo de banco de germoplasma que podemos

encontrar en España es el banco de germoplasma de la Universidad de Lleida (UdL), el cual contienen un total de 118 variedades de manzano y 170 de pera.

Una herramienta importante para este fin en manzano es la REFerence Population (REFPOP), una colección europea de cultivares de manzano formada por 534 genotipos (accesiones y progenies) replicada en seis países, que representan el germoplasma de mejora europeo actual (M. Jung et al., 2020). Los datos genotípicos de esta colección se obtuvieron por medio de SNP-Arrays de alta densidad y evaluada fenotípicamente durante varios años (Dujak & Aranzana, 2023). Uno de los principales objetivos de la REFPOP es el estudio de los efectos del medio ambiente y la resistencia de los diversos genotipos de manzana que se encuentran replicados entre los seis países (Michaela Jung et al., 2020).

Utilizando los datos fenotípicos y genotípicos de REFPOP, realizaron estudios de asociación de genoma completo (GWAS) y selección genómica (GS). Este estudio permitió identificar QTL relevantes asociados a diferentes caracteres que deben validarse para su uso en mejora genética.

Forma del fruto

El tamaño y la forma de los frutos son rasgos externos importantes de calidad en los cultivos frutales comerciales. El tamaño del fruto suele cuantificarse por el peso medio y las determinaciones de la longitud y el diámetro del fruto. La forma del fruto se puede describir utilizando atributos morfológicos, como el índice de forma del fruto (FSI; relación longitud/diámetro), redondez, compactación, conicidad, ángulos inferior y superior (Lateur et al., 2022). Dependiendo del mercado, los consumidores son proclives a elegir diversos tamaños de frutos (Toranzo, 2016). En el caso de las manzanas, los consumidores prefieren frutos con una extensión longitudinal relativamente mayor y un diámetro menor (Chang et al., 2014). El desarrollo de la

manzana consta de dos procesos distintos pero entrelazados; la división celular y el aumento de tamaño celular (alargamiento y agrandamiento), y fisiológicamente, la tasa de división celular determina posteriormente el tamaño del fruto final (Zadravec et al., 2013). Las condiciones tempranas de la temporada influyen fuertemente en el tamaño potencial de las manzanas en el momento de la cosecha. Los factores medioambientales también pueden afectar a la forma de la manzana, concretamente la temperatura y la humedad. Se observó que la longitud de los frutos era mayor cuando las temperaturas eran más bajas tras la floración (Chang et al., 2014).

La forma del fruto es uno de los rasgos que los consumidores utilizan para diferenciar las variedades y elegir para su consumo, así como también se utiliza para describir los cultivares cuando se solicitan derechos de variedad vegetal o registros para cultivares (Currie et al., 2000). Las categorías de forma tales como achatada, globosa, cónica y oblonga se han utilizado ampliamente para describir la forma del fruto en registros de cultivares de manzanas o anuncios de nuevos cultivares (Dapena de la Fuente, 2009). Los gráficos de formas desarrollados por la Junta Internacional de Recursos Fitogenéticos (I.B.P.G.R.) y por la *Union pour la Protection des Obtentions Vegetales* (U.P.O.V.) también se basaron en dichas categorías de formas (Currie et al., 2000). Sin embargo, no se pudieron identificar formas intermedias y no se clasificaron las categorías, lo que dificultó la selección de formas. La proporción de aspecto de la fruta fue el primer rasgo cuantitativo de la forma de la manzana aplicado por los mejoradores. Spinks (1936) y Brown (1960) informaron que la relación de aspecto se heredaba cuantitativamente en la manzana. Los rasgos de forma pueden verse influenciados por factores genéticos y ambientales. Los mejoradores (*Breeders*) pueden seleccionar rasgos que se heredan genéticamente, pero se necesitaba un análisis de componentes de

varianza para seleccionar rasgos con un componente genético significativo (Currie et al., 2000).

La tasa de crecimiento de las manzanas, el tiempo de maduración, el tamaño potencial del fruto y el rendimiento potencial son factores genéticamente controlados que pueden verse alterados por las prácticas de cultivo y las condiciones ambientales sólo hasta cierto punto (Zadravec et al., 2013).

El peso de cosecha de las manzanas está estrechamente correlacionado con el número de semillas (incluidas las semillas abortadas), y el aumento del peso del fruto es atribuido al aumento del número de células (Patricia Denne, 1963). El crecimiento continuo del fruto, desde la división celular hasta la maduración, se asocia principalmente a la expansión celular relacionada con la auxina (Chang et al., 2014).

La herencia cuantitativa de la forma del fruto en la manzana fue postulada ya en 1960. La forma del fruto como puntuación subjetiva de la redondez del fruto apareció bajo un fuerte control genético. La descripción de la forma del fruto como una relación entre la longitud y el diámetro del fruto dio como resultado el descubrimiento de QTL para longitud/diámetro en múltiples grupos de enlace (Keller et al., 2023). Se han identificado varios QTL para el tamaño del fruto de manzana en diferentes poblaciones de mapeo, incluyendo 'Fiesta' × 'Discovery', 'Telamon' × 'Braeburn', 'Royal Gala' × 'Braeburn' y 'Starkrimson' × 'Granny Smith' (Chang et al., 2014). Además, se han identificado regiones génicas que potencialmente contribuyen al aumento del tamaño del fruto durante la especiación de la manzana antes de su domesticación. Se ha encontrado un locus de rasgo cuantitativo para el tamaño del fruto en el cromosoma (Chr) 11 (Keller et al., 2023). Además, un locus en Chr 15 se asoció con el diámetro vertical del fruto de accesiones de *Malus* silvestres y cultivadas, y se encontró que dos loci en Chr 8 y

15 se asociaron significativamente con el peso del fruto en *M. × domestica* (Devoghalaere et al., 2012). En las mismas dos regiones genéticas, se informaron barridos selectivos en *M. × domestica* en comparación con *M. sieversii* y *M. sylvestris* (Keller et al., 2023). La base genética del tamaño del fruto en *M. × domestica* se ha estudiado más a fondo en rasgos como el peso, el diámetro y la longitud del fruto, se han reportado muchos loci de rasgos cuantitativos (QTL) en todos los cromosomas de la manzana (Keller et al., 2023).

La forma del fruto también se ve influida por la acción de proteínas OVATE FAMILY PROTEINS (OFPs), una clase proteica que actúa como reguladores vegetales. Las OFPs están ampliamente distribuidos en los reinos vegetales, sus funciones biológicas en las plantas siguen siendo en gran parte desconocidas. Sin embargo, estudios limitados en varias especies vegetales diferentes incluyendo tomate (*Solanum lycopersicum*), *Arabidopsis thaliana*, pimiento (*Capsicum annuum*) y plátano (*Musa acuminata*) han revelado que las proteínas OFP intervienen en la regulación del crecimiento de las plantas y su desarrollo. Más allá de las solanáceas, se ha descubierto que varios QTLs controlan la forma del fruto del melón (*Cucumis melo*) son homólogos de la OFP (*CmOFP*). Como factores de transcripción específicos de las plantas, las proteínas de la familia OVATE (OFP) están implicadas en la regulación del crecimiento, el desarrollo y la respuesta al estrés de las plantas, y la respuesta al estrés. Aunque se han descrito proteínas OFP en algunas especies, se sabe poco sobre su evolución, estructura y niveles de expresión bajo estrés biológico y abiótico, y las interacciones entre los miembros de OFP en manzana (Li et al., 2019).

A pesar de la gran abundancia de asociaciones conocidas con rasgos relacionados con el tamaño del fruto, el descubrimiento, la validación y la consolidación

adicionales de las asociaciones identificadas podrían facilitar la aplicación futura en el mejoramiento asistido por genómica (Keller et al., 2023).

Descriptores de los caracteres morfológicos del fruto

La evaluación de la forma es objetiva, utilizando listas de descriptores para la morfología del fruto. Otro tipo de descriptores son subjetivos, mediante la comparación de la manzana con una lista dando lugar a sesgos y desacuerdos entre distintos evaluadores de la calidad, como por ejemplo el sabor o la palatabilidad.

En la figura 1 se puede observar cómo se define la forma del fruto, principalmente en función de las relaciones altura/diámetro y anchura de la cubeta ocular/anchura de la cubeta peduncular, así como la posición del diámetro máximo (Dapena de la Fuente, 2009).

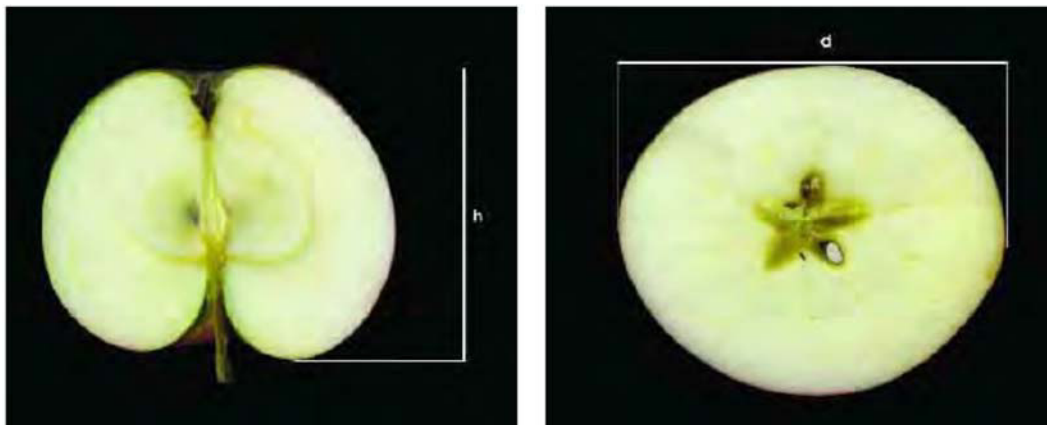


Figura 1. A) Corte longitudinal, detalle de determinación de la altura máxima, h =altura. B) Figura 4. Corte transversal, detalle de determinación del diámetro máximo, d =diámetro. Tomado de Dapena de la Fuente et al. 2009.

Relación altura/diámetro: es la relación que se obtiene al dividir la altura máxima entre el diámetro máximo (Figura 2). Clases: 1) aplanada ($\leq 0,75$), 3) bastante aplanada (0,76-0,85), 5) intermedia (0,86-0,95), 7) bastante elevada (0,96-1,05) y 9) elevada ($> 1,05$) (Dapena de la Fuente, 2009; Lateur et al., 2022).

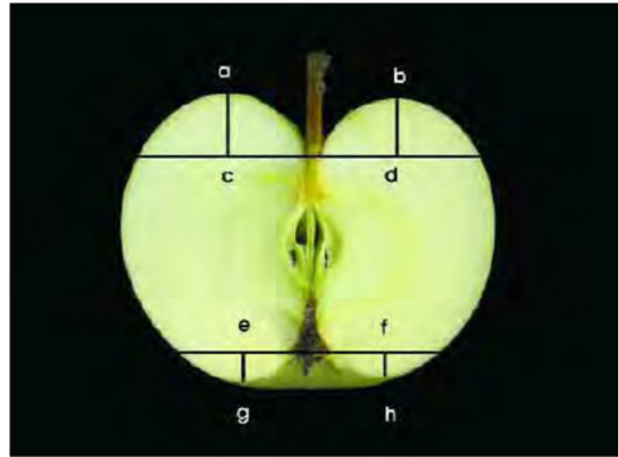


Figura 2.A) Corte longitudinal, detalle de determinación de la altura máxima, h=altura. B) Figura 4. Corte transversal, detalle de determinación del diámetro máximo, d=diámetro. Tomado de Dapena de la Fuente et al. 2009.

Forma de fruto: Se define principalmente en función de las relaciones altura/diámetro y anchura de la cubeta ocular/anchura de la cubeta peduncular, así como la posición del diámetro máximo (Figura 3). Clases: 1) aplanada, 2) aplanada troncónica, 3) aplanada cónica, 4) aplanada globulosa, 5) truncada cónica, 6) aplanada globulosa cónica, 7) globulosa, 8) globulosa troncónica, 9) globulosa cónica, 10) oblonga, 11) oblonga troncónica, 12) oblonga cónica, 13) oval, 14) oval troncónica, 15) cónica, 16) oblonga entallada y 17) oval entallada (Dapena de la Fuente, 2009; Lateur et al., 2022).

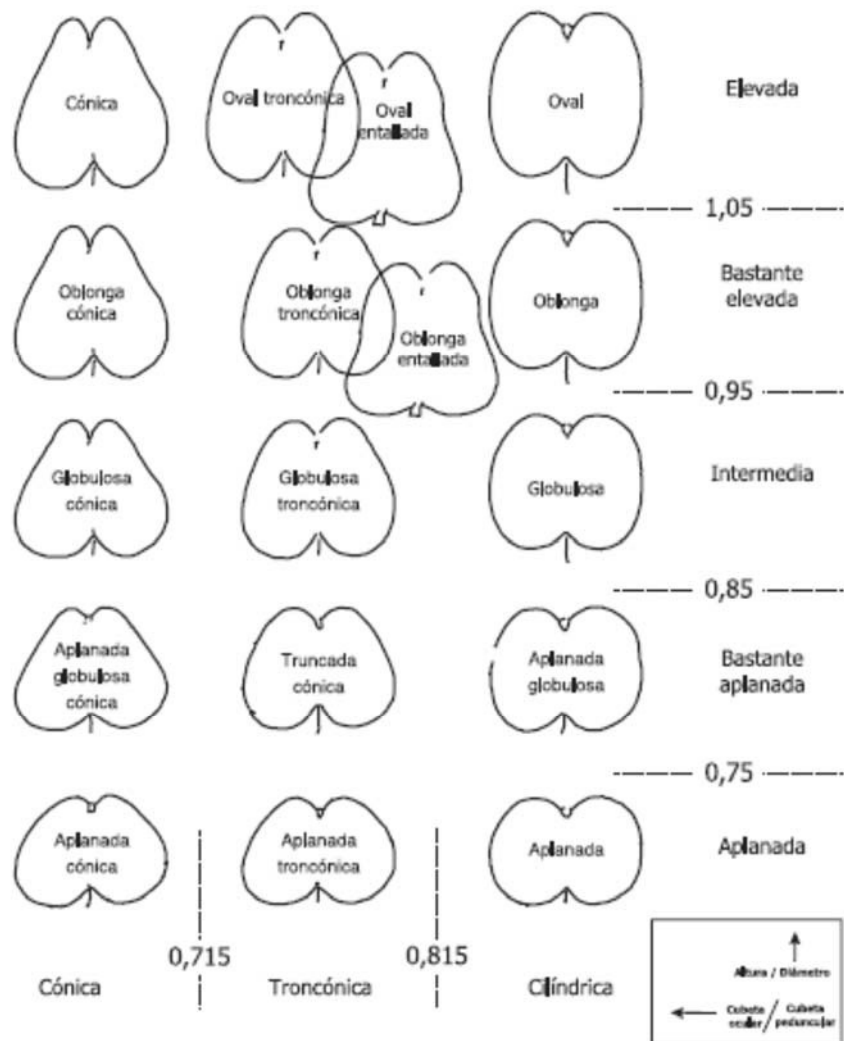


Figura 3. Forma del fruto. Tomado de Dapena de la Fuente et al. 2009.

Deep Learning y Maching Learning

Investigaciones recientes han desarrollado algoritmos de procesamiento de imágenes para medir objetivamente las características externas de los productos hortícolas, esta características de las manzanas incluía la presencia de defectos, el tamaño, la forma y el color (Paulus & Schrevens, 1999). No obstante, la mayoría de los algoritmos de forma cuantifican la redondez, la rectangularidad, o la elongación del producto calculando la relación entre el área proyectada y la anchura del producto. (Paulus & Schrevens, 1999).

El desarrollo de algoritmos depende principalmente del método responsable de distinguir las frutas de otros objetos mediante el procesamiento de imágenes (Risdin et al., 2020).

El *Deep Learning* es el subcampo del Machine Learning dentro de los estudios basados en Inteligencia Artificial. Es una colección de técnicas que modelan abstracciones de alto nivel en datos (Thi Phuong Chung & Van Tai, 2019). El *Deep Learning* es un campo específico de aprendizaje automático basado en la representación sucesiva de capas de datos. En los últimos años se ha acelerado la aplicación de técnicas de *Deep Learning* en diferentes campos de la agricultura y se desarrollaron diversos algoritmos de reconocimiento de frutas apropiados para algunos tipos de aplicaciones hortícolas, cada uno con diferentes niveles de precisión (Saedi & Khosravi, 2020)

En el *Deep Learning*, estos modelos pueden alcanzar una precisión de última generación, superando en ocasiones el rendimiento a nivel humano. Los modelos se entrenan utilizando un gran conjunto de datos etiquetados y arquitecturas de redes neuronales que contienen muchas capas en términos de precisión (Thi Phuong Chung & Van Tai, 2019)

La técnica es que cada capa está formada por un conjunto de neuronas, donde cada capa está completamente conectada a todas las neuronas de la capa anterior. Finalmente, hay una arquitectura, es decir, una última capa completamente conectada (la capa de salida) que representa las predicciones. Las capas están organizadas en 3 dimensiones: ancho, alto y profundidad. Las neuronas de una capa no se conectan con todas las neuronas de la siguiente capa, sino sólo con una pequeña región de esta. El resultado final del sistema se reduce a un único vector de puntuaciones de probabilidad, organizado a lo largo de la dimensión de profundidad (Risdin et al., 2020; Sa et al., 2016).

El laboratorio en concreto está desarrollando un software de análisis de imágenes llamado *Shape Analyzer* para la obtención de las medidas descriptoras morfológicas en manzanas (Jurado-Ruiz et al., in prep.).

Objetivos generales

El objetivo de este trabajo es obtener datos fenotípicos de forma del fruto, acidez titulable, azúcares totales, contenido en cerinas (*wax content*) y porcentaje de almidón de dos colecciones de manzano: el banco de germoplasma de la Universidad de Lleida (UdL) y la población de referencia REFPOP. A partir de los datos recogidos se realizarán los análisis estadísticos de los mismos a nivel fenotípico para ambas colecciones.

Consecuentemente, los datos de forma de fruto de las variedades de la REFPOP serán analizados y asociados a datos genotípicos previamente obtenidos dentro del grupo para asociar el carácter de forma de fruto a marcadores moleculares de interés. Paralelamente, los datos de forma y tamaño se emplearán para el entrenamiento de un software de reconocimiento de imagen que permita obtener las medidas de las manzanas.

Objetivos específicos

Los objetivos específicos de este trabajo son:

- Obtener datos fenotípicos de parámetros de calidad de frutas en el momento de maduración específica para el consumo: días de cosecha, sólidos solubles totales, acidez titulable, *wax content* visual, SPI, russetting y resistencia a enfermedad a *scab* (*Venturia inaequalis*) y oídio (*Powdery mildew*).

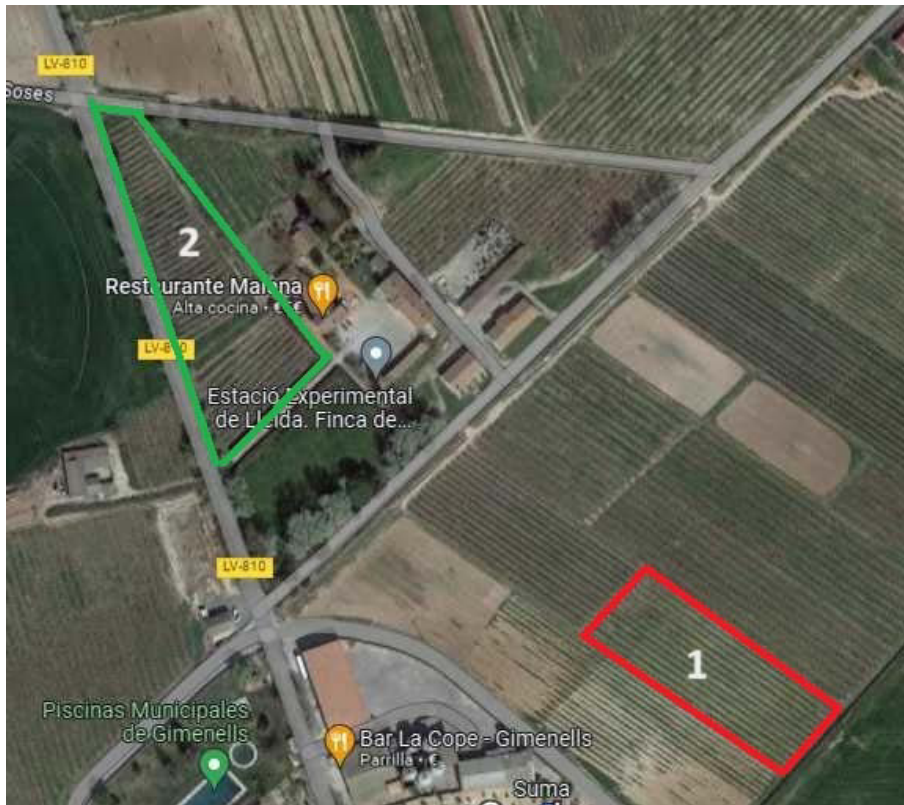
- Realizar los análisis estadísticos (estadígrafos) de cada parámetro de calidad de los frutos recolectados.
- Escanear las manzanas recolectadas para el entrenamiento de software de inteligencia artificial para reconocimiento de imagen.
- Determinar potenciales QTL ligados a genes implicados en la forma de la manzana y marcadores moleculares SNP significativamente asociados.

Materiales y métodos

Parte I

En esta primera parte se realizaron las determinaciones fenotípicas de las dos colecciones de manzanos: REFPOP y UdL. Estas determinaciones permitirán asociar caracteres fenotípicos con distintos marcadores moleculares asociados a caracteres genotípicos. Los caracteres fenotípicos analizados son: **porcentaje de almidón (SPI)**, **días de cosecha**, **visual wax o contenido en pruinas**, **Sólidos solubles totales (SST)**, y **acidez titulable (TA)**.

Material vegetal: se tomaron muestras de frutas de manzanos pertenecientes a dos colecciones: A) REFPOP, implantada en la granja experimental del IRTA (41°39'22.0"N 0°23'21.7"E) en Gimenells (Lleida). La colección está compuesta por 534 genotipos plantados en seis países europeos por 269 accesiones y 265 descendientes de 27 combinaciones parental (Michaela Jung et al., 2020). Las condiciones de tratamiento del campo fueron: poda, aclareo, riego por goteo a diario. La colección no fue tratada con hormonas. B) El banco de germoplasma de la UdL de especies locales y nacionales de manzanos que pertenece a la Universidad de Lleida ubicada en la misma finca experimental.



Fotografía 1. fotografía satelital de la Finca experimental de Gimeneles. 1) Ubicación de la colección REFPOP. 2) Ubicación del banco de germoplasma UdL.

Toma de muestra de los frutos: los árboles pertenecientes al banco de germoplasma UdL fueron evaluados semanalmente para determinar la madurez de las frutas. Las muestras se tomaron en días que las plantas mostraban una maduración visual del 50% del total de frutos en la planta, que presentan el color determinado color de fondo del fruto, decrecimiento de la firmeza, o cambio de color de las semillas a marrón oscuro. La madurez generalizada de la variedad se analizó sobre 3 frutos representativos por prueba de almidón (en inglés, *Starch Pattern Index*, SPI) mediante el uso de 'Lugol', solución de yodo molecular I₂ y yoduro potásico KI, que tiñe el almidón del fruto. Se escogieron tres frutos que se encontrasen entre valores 4-6 de la escala Cornell (Blanpied & Silsby, 1992). Se apuntó la fecha de cosecha y para realizar un correcto análisis de los mismo se utilizaron los números de días en lugar de las fechas de cosecha. Posteriormente se estimó visualmente el nivel de cerinas (pruina) presente en la piel de la fruta.

Para las mediciones en laboratorio se tomaron 6 frutas del total colectado y el resto se descartaron.

En el caso de la REFPOP, se procedió de la misma manera que con las muestras de la UdL, se registraron los pesos y la fecha de cosecha. Además, se tomaron datos de *russeting*, y las enfermedades *oídio* y *scab*. De los 20 frutos se tomaron 6 frutos para las pruebas de laboratorio.

Visual Wax Content: El visual *Wax content* se estipula de forma visual en función de la superficie cubierta de la manzana (Tabla 1).

Wax	% de superficie cubierta
0	0-20
1	20-40
2	40-60
3	60-80
4	80-100

Tabla 1. valores de visual wax content dependiendo de la superficie cubierta del fruto.

Como se puede apreciar la variable visual *wax content* es una función ordinal y arbitraria. Pudiendo elegir otra escala a conveniencia.

Determinación de índice de patrón de almidón (*Starch Patters Index (SPI)*):

Para la determinación de índice de patrón de almidón se toman 3 muestras de frutas del mismo árbol, se procede a realizar un corte radial de las mismas. Se sumerge la mitad del extremo peduncular en una placa de Petri parcialmente llena con solución de triyoduro, asegurando que la superficie expuesta quede completamente sumergida en la solución. Observar el resultado según el índice Cornell de Almidón (Blanpied & Silsby, 1992).

Estado de maduración fisiológico	índices SPI (Cornell's scale)
Verdes, manzanas almidonosas	1-3
Manzanas maduras para comer	4-6
Manzanas sobremaduras	7-8

Tabla 2. Correspondencia entre el Índice SPI y el estado de maduración.

La evolución de la maduración en las diferentes variedades, sumado a su vez a condiciones climáticas, sanitarias o regionales, que pueden hacer variar el patrón de acumulación o degradación del almidón. A continuación, se transcribe una tabla con los momentos óptimos de cosecha para cada variedad (Gergoff Grozeff, 2019):

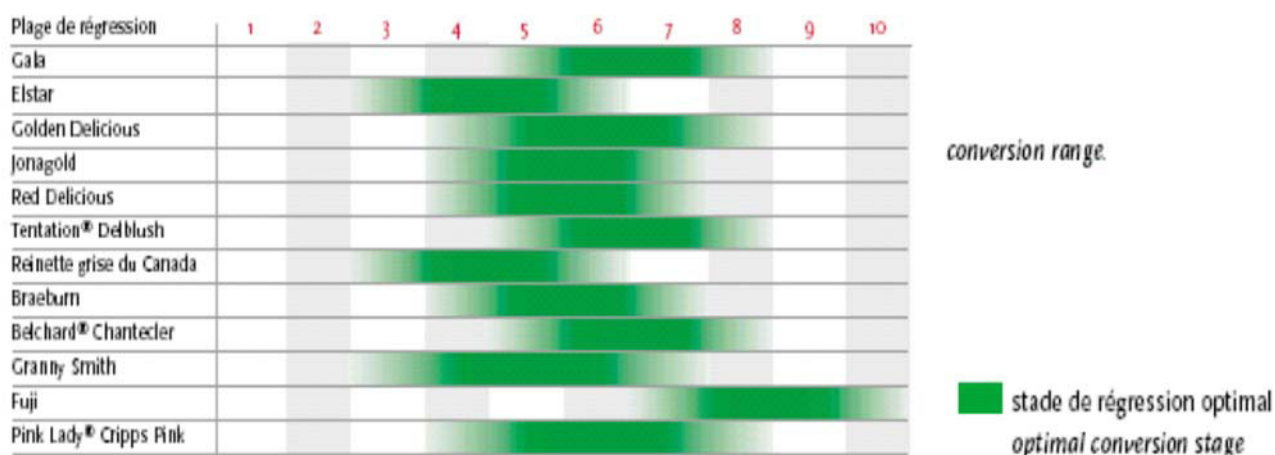


Tabla 3. Índices de maduración para variedades de manzana. Las columnas indican los índices de almidón ideales al momento de cosecha. Tomado de Georgoff Grozeff (2019).

Obtención de muestras de jugo: las muestras de jugo se obtuvieron de procesando la mitad del cáliz de las tres manzanas por medio de un extractor de jugo doméstico para realizar las determinaciones de sólidos solubles totales (*Soluble Solids Content* - SSC) y acidez titulable (*Titrateable acidity* - TA).

Determinación de Sólidos solubles totales (SST): los sólidos solubles totales presentes en el jugo de las manzanas se determinaron por medio de un

refractómetro electrónico (Atago Pocket Pal-1. L113963). La equivalencia entre la concentración de sólidos solubles y ° Brix es: 1 gramo de sólido soluble cada 100 gramos de solución equivalen a 1° Brix.

Determinación de acidez titulable (TA): el análisis de acidez titulable se realizó por medio de un equipo titulador automático de muestras de jugos (Hanna Instruments- Hi 84532), con agente titulante de bajo rango (Tritiant Low Range solution LR-Hi84532-50).

Ratio (SST/TA): El RATIO es el cociente entre el % de sólidos solubles (expresado como ° Brix = % S.S.) y la acidez titulable, también en % (expresada en gramos de ácido cada 100 ml. de jugo). Para obtener la TA en % se debe dividir el valor hallado en la titulación por 10. El resultado del valor obtenido de RATIO es adimensional, ya que expresa la cantidad de azúcares por unidad de ácido.

Análisis de resultados: Para los análisis estadísticos y gráficos se utilizaron los software *Excel*, *RStudio* e *Infostat (versión estudiantil)*(Di Rienzo et al., 2011).

Parte II

En la segunda parte de este trabajo se analizaron los resultados proporcionados por el software de reconocimiento de imágenes *Shape Analyzer* que se encuentra en proceso de desarrollo por miembros del equipo de investigación (Jurado-Ruiz et al., in prep).

Fenotipado de tamaño y forma: el escaneo de las muestras se llevó a cabo cortando las manzanas por el eje transversal. Las mitades de cada manzana son colocadas de forma que el mesocarpio este en contacto con el escáner (Mustek A3 2400S panel V2.0). Las fotos se realizaron con una definición de 600 dpi. Las imágenes obtenidas se emplearon para entrenar un software *Shape Analyzer*

basado en *Machine Learning* (Jurado-Ruiz et al., in prep) (ver anexo I). El citado programa permite obtener diversos parámetros del fruto como longitud, anchura, diversas longitudes, área, y perímetro, de manera automatizada. Estos parámetros permiten calcular los FSI internos y FSI externo de los frutos.

Genotipado de la REFPOP: Los datos genotípicos de la REFPOP utilizados para este estudio se obtuvieron del genotipado de la REFPOP mediante un chip array de 480K para un total de 1356 genotipos únicos a partir de estudios previos. El conjunto de datos del marcador SNP incluyó todas las accesiones de REFPOP de manzana (incluidos 13 padres del grupo de progenie de REFPOP de manzana) y genotipos únicos adicionales utilizados posteriormente como conjunto de referencia para la imputación de marcadores (M. Jung et al., 2020).

Análisis de estadístico de los SNPs

Se seleccionaron tres regiones, localizadas en los chr 2, 11 (dos regiones), y 13, respectivamente. Las regiones fueron escogidas a partir trabajos previos del grupo (Dujak & Aranzana, 2023). En estas regiones se encontraron 4 SNPs asociados a la morfología y tamaño de la manzana. Las regiones donde se localizan estos SNPs para poder hacer la asociación haplotípica de fenotipo-genotipo.

Los análisis estadísticos se realizaron sobre regiones específicas del genoma en regiones no recombinantes conocidas como **haplobloques**.

Para el análisis estadístico de los SNPs solamente se tuvieron en cuenta los haplotipos que presentaban $n \geq 2$.

Resultados y discusión.

Parte I: Análisis estadístico descriptivo del fenotipado de caracteres

Starch Pattern Index (SPI)

Gráficos de distribución de datos para SPI

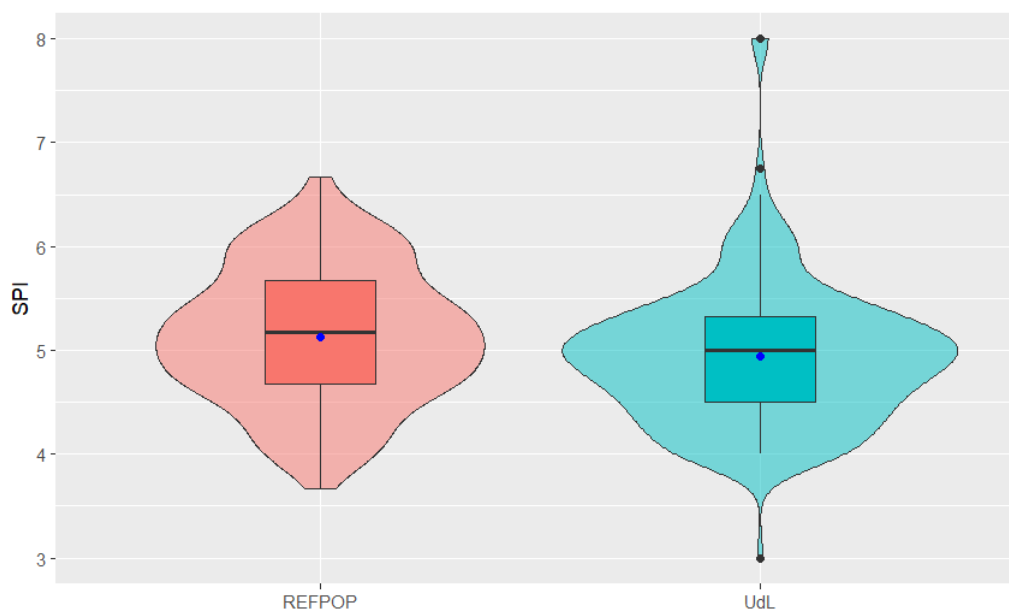


Gráfico 1. Diagramas de violín para SPI. La línea negra indica la mediana de los datos y el punto azul indica la media.

Resumen de estadígrafos para SPI

SPI								
Colección	n	Media	D.E.	Var(n-1)	CV	Mediana	Q1	Q3
REFPOP	130	5,13	0,65	0,42	12,61	5,17	4,67	5,67
UdL	224	4,94	0,68	0,46	13,79	5,00	4,50	5,33

Tabla 4. Resumen de los estimadores estadísticos para los datos de colecciones referente al SPI. En la tabla se detallan el número de individuos (n), la media, el desvío estándar (D.E.), la varianza muestral (Var (n-1)), coeficiente de variación (CV), la mediana y los cuartiles primero y tercero (Q1 y Q3)

Pruebas de normalidad para el SPI

Pruebas de normalidad							
		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Colección	Estadístico	gl ^b	p-value	Estadístico	gl	p-value
SPI	REFPOP	0,190	130	0,000	0,914	130	0,000
	UdL	0,158	224	0,000	0,939	224	0,000

a. Corrección de significación de Lilliefors

b. Grados de libertad

Tabla 5. Resumen de los estimadores estadísticos para la prueba de normalidad para los datos referentes a SPI.

En los estadígrafos que se detallan en la tabla 4 se puede observar que la medias y las medianas están en valores cercanos, esto indica que las muestras se tomaron en fechas de maduración fisiológicas aptas para consumo, por lo que los otros valores fenotípicos se encuentran en consonancia con el punto de maduración óptimo de la fruta (Gergoff Grozeff, 2019).

Los datos mostrados en el gráfico 1 y los resultados de la tabla 5, sugieren a simple vista que hay una distribución normal y con datos que se distribuyen uniformemente a ambos lados de la media. La prueba de normalidad Kolmogorov-Smirnov y Shapiro Wilk para SPI mostraron valores p-value de las dos colecciones muestran valores menores al nivel de significancia de 0,05; por lo tanto, se infiere que los datos no siguen una distribución normal, a pesar de la apariencia. Por esta razón se decidió no realizar pruebas de ANOVA u otro tipo de pruebas paramétricas.

Días de cosecha.

Los resultados para los días de cosecha se muestran a continuación.

Gráficos de distribución de datos para días de cosecha

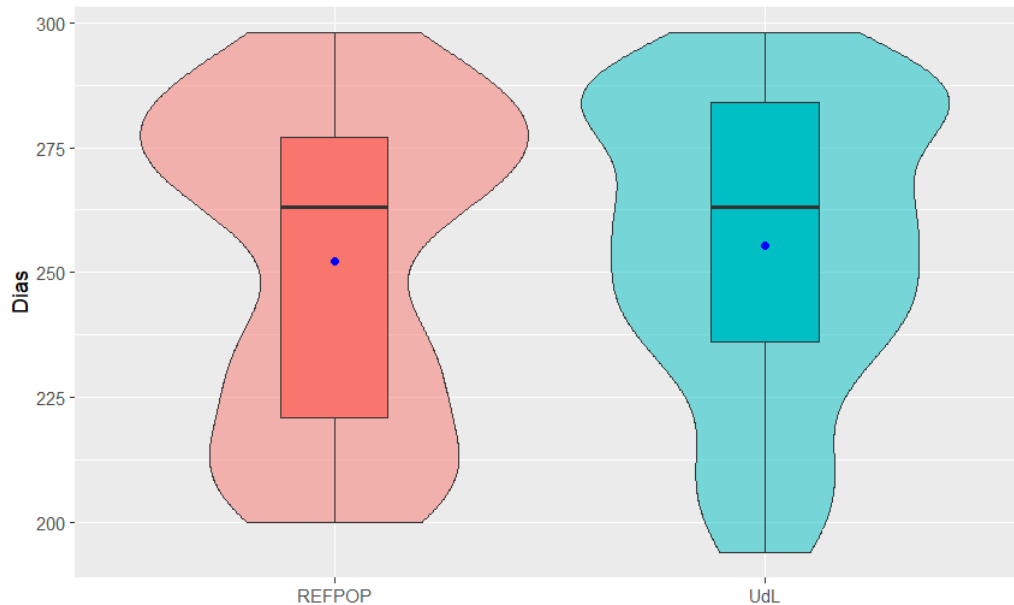


Gráfico 2: Diagramas de violín para días de cosecha. La línea negra indica la mediana de los datos y el punto azul indica la media.

Resumen de los estadígrafos para días de cosecha

Días de cosecha								
Colección	n	Media	D.E.	Var(n-1)	CV	Mediana	Q1	Q3
REFPOP	130	252,23	31,45	989,28	12,47	263	221	277
UdL	224	255,46	28,97	839,06	11,34	263	236	284

Tabla 6: Resumen de los estimadores estadísticos para los datos referente a los días de cosecha. En la tabla se detallan el número de individuos (n), la media, el desvío estándar (D.E.), la varianza muestral (Var (n-1)), coeficiente de variación (CV), la mediana y los cuartiles primero y tercero (Q1 y Q3)

Pruebas de normalidad para días de cosecha

Pruebas de normalidad							
Días	Colección	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Estadístico	gl ^b	p-value	Estadístico	gl	p-value
	REFPOP	0,188	130	0,000	0,900	130	0,000
	UdL	0,128	224	0,000	0,937	224	0,000

a. Corrección de significación de Lilliefors

b. Grados de libertad

Tabla 7. Resumen de los estimadores estadísticos para la prueba de normalidad Kolmogorov-Smirnov y Shapiro-Wilk para los datos referentes a los días de cosecha.

Los resultados de los estadígrafos muestran que las medias (\bar{X}) de las dos colecciones de datos son similares. Teniendo en cuenta los valores de desvío estándar (D.E.) se debería utilizar una prueba de ANOVA para confirmar si las medias presentan diferencias significativas. El inconveniente es que los datos recogidos para días de cosecha no presentan una distribución normal, tal como se puede observar en los gráficos de violín (gráfico 2). Solamente, los datos obtenidos de la UdL muestran que los datos se concentran en torno a la media (\bar{X}) mientras que la REFPOP muestra una concentración de datos hacia los extremos; lo cual muestra una tendencia distinta a una distribución normal.

Para confirmar se empleó la prueba de Kolmogorov- Smirnov que muestran valores (tabla 7) de *p-value* menores a 0.05, que corresponde al valor de significancia. Por lo tanto, realizar un estudio de ANOVAS no brindaría información certera sobre las medias; se debería realizar un test no paramétrico.

Los coeficientes de variación (C.V.) para los dos conjuntos de datos muestran valores menores a 25%, valor que se estipula como máximo aceptable para los datos a campo.

Los resultados para los días de cosecha muestran que la colección REFPOP muestra que los datos se concentran en los primeros días y últimos días, es decir, que una parte de los ejemplares son de maduración temprana y otra parte marcadamente tardías. En cambio, la colección UdL indica que una mayor preponderancia a maduraciones tardías e intermedias. Los datos de días de cosechas reportados por Jung et al. (2020) para 30 individuos al azar de la REFPOP para la península ibérica están en el orden de los 200 a 280 días como límites. En el caso de los valores colectados, se puede apreciar que esos valores corresponden a los reportados en el mencionado trabajo, tanto la REFPOP como la UdL se encuentran en los mismo valores límites. Se puede inferir en este caso,

que las variables climáticas de la zona peninsular, juega un papel relevante en el día de cosecha, ya que difieren para otras zonas como Italia o Francia donde ejemplares llegan hasta el día 300 para su cosecha (M. Jung et al., 2020).

Visual Wax Content

Gráficos de distribución de datos para visual wax content

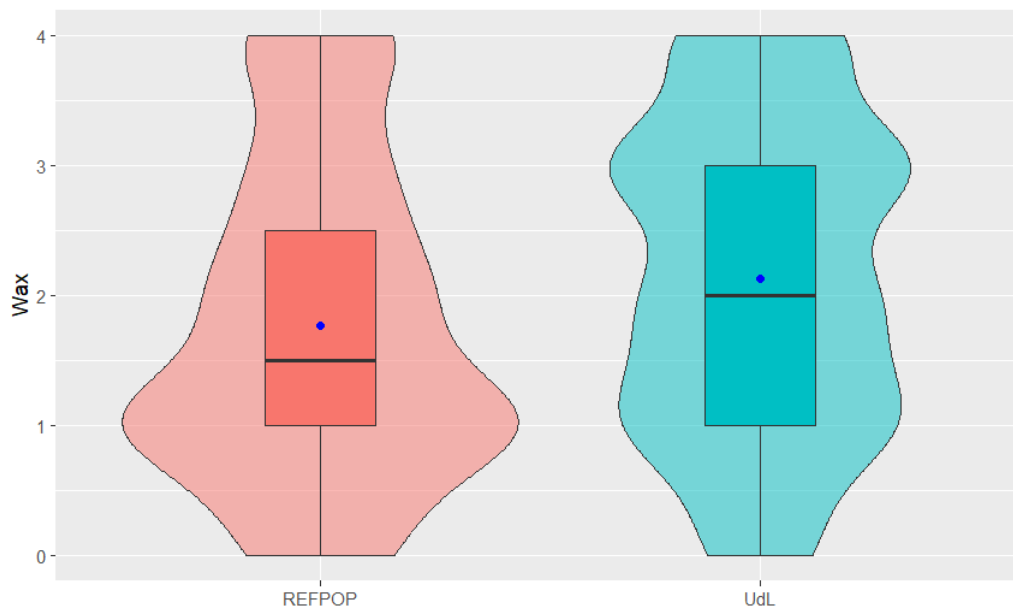


Gráfico 3. Diagramas de violín para visual wax content. La línea negra indica la mediana de los datos y el punto azul indica la media.

Resumen de los estadígrafos para visual wax content

Visual Wax content								
Colección	n	Media	D.E.	Var(n-1)	CV	Mediana	Q1	Q3
REFPOP	130	1,77	1,21	1,47	68,56	1,5	1	2,5
UdL	224	2,14	1,18	1,4	55,18	2	1	3

Tabla 8. Resumen de los estimadores estadísticos para los datos referente a visual wax content. En la tabla se detallan el número de individuos (n), la media, el desvío estándar (D.E.), la varianza muestral (Var (n-1)), coeficiente de variación (CV), la mediana y los cuartiles primero y tercero (Q1 y Q3)

Pruebas de normalidad para visual wax content

		Pruebas de normalidad					
		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Colección	Estadístico	Gl ^b	p-value	Estadístico	gl	p-value
Visual	REFPOP	0,190	130	0,000	0,914	130	0,000
Wax	UdL	0,158	224	0,000	0,939	224	0,000
Content							

a. Corrección de significación de Lilliefors

b. Grados de libertad

Tabla 9. Resumen de los estimadores de la prueba de Kolmogorov-Smirnov y Shapiro-Wilks para cada conjunto de datos de visual wax content.

En el caso del parámetro de visual *wax content* - que son las ceras presentes en forma de polvillo en la superficie del fruto - se observa que los valores de CV que se muestran en la tabla 5 son sensiblemente mayores que el 25 % de valor límite que se acepta para determinaciones a campo. El valor del coeficiente de variación muestra una gran dispersión de los datos con respecto a las medias.

Los gráficos de violín (Gráfico 3) muestran que los datos no tienen una distribución normal. Se puede observar que los datos de la colección REFPOP tiene más individuos con una propensión a producir poca pruina; mientras que las variedades locales y nacionales presentan una heterogeneidad de individuos, se puede observar que los datos de la UdL muestran una repartición de datos entre las cotas mínimas y máximas una distribución más homogénea. La prueba de Kolmogorov-Smirnov, confirma que los datos de la REFPOP como de la UdL no siguen una distribución normal; por lo tanto, realizar una prueba de ANOVA u otros test paramétricos no es factible.

Los datos muestran que los individuos de la colección REFPOP son más proclive a producir menos *wax content* que las variedades locales presentes en el banco de

germoplasma UdL, esto puede deberse a que las variedades locales y nacionales están más adaptadas al clima mediterráneo de Cataluña; el clima durante los meses de verano es más seco y con niveles altos de radiación solar que afectarían al fruto.

Los individuos de la colección REFPOP, al ser productos de mejoras genéticas, presentan una población minoritaria proclive a producir pruinas, mientras que la mayoría de los individuos presentan niveles bajos de las mismas, esto posiblemente se deba a que estas variedades están seleccionadas para climas que presenten veranos no tan rigurosos como los de Cataluña. Una comprensión de los componentes y las cantidades de cera de la fruta es importante para obtener una mejor calidad de la fruta, mejorar la resistencia a las enfermedades y desarrollar estrategias de tratamiento postcosecha. Lo que es más importante, la estructura, composición y concentración de la cera del fruto varían entre cultivares de la misma especie; por lo tanto, el conocimiento de los rasgos de la cera cuticular de diferentes germoplasmas también será útil para la selección de progenitores de nuevas variedades. Por lo tanto, la investigación de la cera cuticular a nivel de germoplasma es de valor significativo tanto para el estudio genético como para el programa de mejoramiento de frutas (Belding et al., 1998; Verardo et al., 2003).

Sólidos Solubles Totales (SST)

Gráficos de distribución de datos para sólidos solubles totales

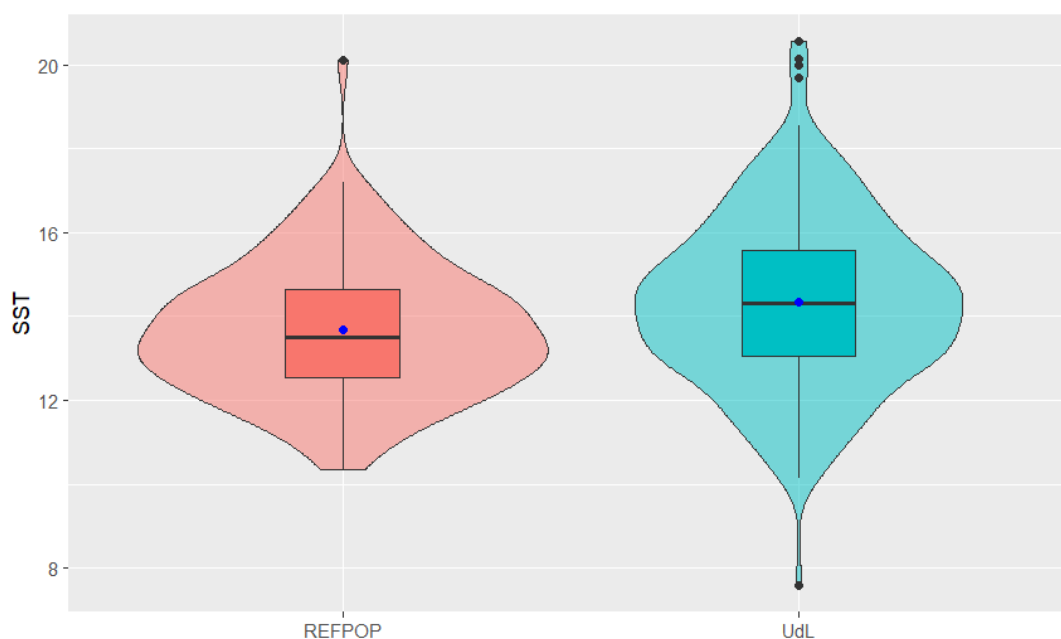


Gráfico 4. Diagramas de violín para Sólidos Solubles Totales. La línea negra indica la mediana de los datos y el punto azul indica la media.

Resumen de estadígrafos para SST

SST								
Colección	n	Media	D.E.	Var(n-1)	CV	Mediana	Q1	Q3
REFPOP	130	13,67	1,6	2,55	11,67	13,5	12,5	14,65
UdL	224	14,33	2,04	4,17	14,25	14,3	13,05	15,6

Tabla 10. Resumen de los estimadores estadísticos para los datos referentes a los sólidos solubles totales (SST). En la tabla se detallan el número de individuos (n), la media, el desvío estándar (D.E.), la varianza muestral (Var (n-1)), coeficiente de variación (CV), la mediana y los cuartiles primero y tercero (Q1 y Q3)

Pruebas de normalidad para SST

Pruebas de normalidad							
		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Colección	Estadístico	gl ^b	p-value	Estadístico	gl	p-value
SST	REFPOP	0,061	130	0,200*	0,975	130	0,017
	UdL	0,048	224	0,200*	0,993	224	0,427

*. Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors

b. Grados de libertad

Tabla 11. Resumen de los estimadores de la prueba de Kolmogorov-Smirnov y Shapiro-Wilks para cada conjunto de datos de sólidos solubles totales.

En los diagramas de violín para las determinaciones de sólidos solubles totales (SST), se puede observar que los datos muestran una posible distribución normal (Gráfico 4), con datos simétricos centrado en la media, lo cual puede asumir que los datos siguen una distribución normal.

Los estadígrafos (tabla 10) muestran que los valores de las medias aritméticas de las colecciones son similares. Los coeficientes de variación (CV) muestran valores menores a 25% que es la tolerancia para experimentos a campo.

La prueba de normalidad de Kolmogorov-Smirnov confirma que los datos presentan una distribución normal debido a que el *p-value* muestran valores mayores a $\alpha = 0.05$ como valor de significancia; de esta manera se cumple una de las condiciones para realizar una prueba de ANOVA o prueba T, además se cumple la condición de ser muestras independientes. Para el último supuesto se realiza una prueba de homogeneidad de varianza (homocedasticidad) de datos, para ello se emplea la prueba de Levene.

		Prueba de homogeneidad de varianza			
		Estadístico de Levene	gl1	gl2	p-value
SST	Basada en la media	7,412	1	357	0,007
	Basada en la mediana	7,565	1	357	0,006
	Se basada en la mediana y con gl ajustado	7,565	1	341,991	0,006
	Basada en la media recortada	7,517	1	357	0,006

Tabla 12. Resumen de los estimadores de la prueba de Levene.

Como se aprecia en la tabla 12 el valor de p -value de la media es menor al valor de $\alpha = 0.05$; por esto se rechaza la hipótesis nula que indica que los datos presentan varianzas iguales. Los datos al no presentar homocedasticidad, no se pueden realizar test paramétricos como ANOVA o prueba de T- Student para saber si hay diferencias significativas en las medias entre las dos colecciones de datos.

Acidez Titulable (TA).

Gráficos de distribución de datos para Acidez titulable (TA)

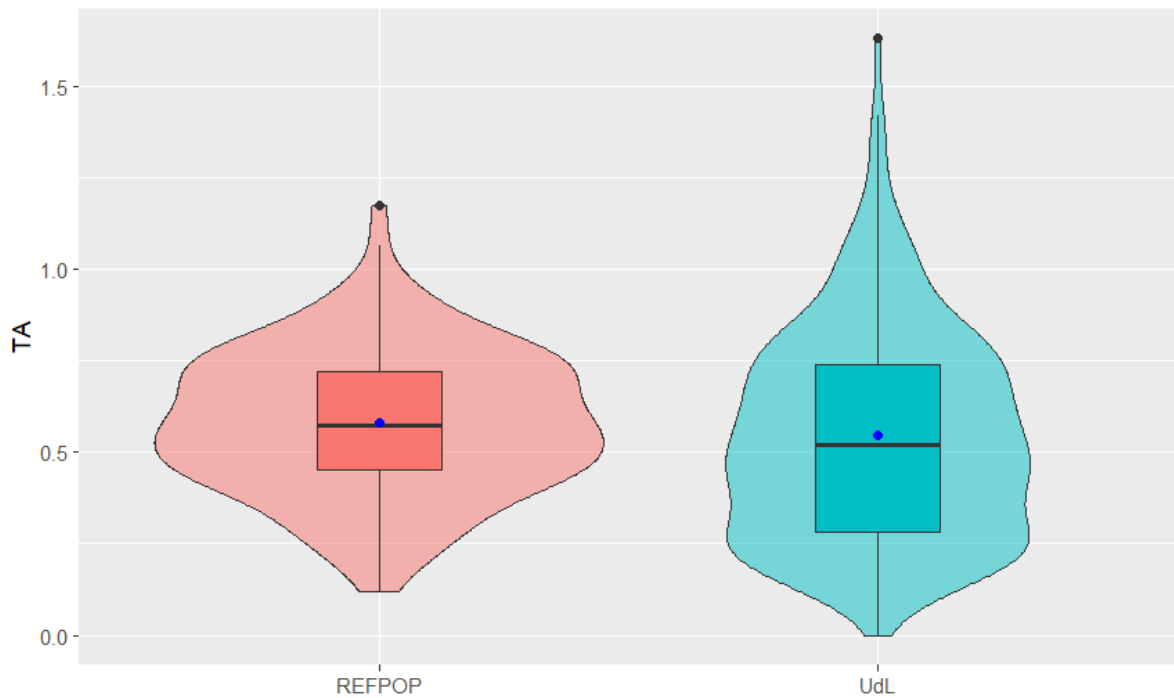


Gráfico 5. Diagramas de violín para TA. La línea negra indica la mediana de los datos y el punto azul indica la media.

Resumen de estadígrafos para acidez titulable.

TA								
Colección	n	Media	D.E.	Var(n-1)	CV	Mediana	Q1	Q3
REFPOP	130	0,58	0,2	0,04	33,88	0,57	0,45	0,72
UdL	224	0,56	0,35	0,12	63,07	0,52	0,28	0,74

Tabla 13. Resumen de los estimadores estadísticos para los datos referente al TA. En la tabla se detallan el número de individuos (n), la media, el desvío estándar (D.E.), la varianza muestral (Var (n-1)), coeficiente de variación (CV), la mediana y los cuartiles primero y tercero (Q1 y Q3)

Pruebas de normalidad para la variable TA.

		Pruebas de normalidad					
		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Coleccion	Estadístico	gl ^b	p-value	Estadístico	gl	p-value
TA	REFPOP	0,044	130	0,200*	0,994	130	0,866
	UdL	0,105	224	0,000	0,795	224	0,000

*. Esto es un límite inferior de la significación verdadera.

- a. Corrección de significación de Lilliefors
- b. Grados de libertad

Tabla 14. Resumen de los estimadores de la prueba de Kolmogorov-Smirnov y Shapiro-Wilks para cada conjunto de datos referente a TA

Los resultados obtenidos para TA muestran datos no presentan una distribución normal, según se puede observar en el gráfico de violín (Gráfico 5). Solamente presenta una distribución normal la colección REFPOP, al poseer individuos con mayor variabilidad en este parámetro cuantitativo, en cambio la UdL tiene un predominio de individuos con bajo nivel de ácido málico, pero algunos de ellos, presentan niveles de ácido málico singularmente altos. Los datos recogidos de la UdL presentan una distribución de datos similar a una distribución normal, pero con datos que agrupan a uno de los lados de la media, tal como puede apreciarse en el gráfico de violines. Los coeficientes de variación (CV) mostrados en la tabla 12 muestran valores elevados dando una noción de que hay una gran variabilidad de los mismos en los individuos que componen cada colección. Inclusive, la colección REFPOP muestra una variabilidad mayor del 25 %, pero la UdL muestra valores elevados, lo que sugiere que hay individuos con niveles muy bajos y altos de ácido málico.

Se confirma por el test de normalidad (tabla 14) que solamente la REFPOP sigue una distribución normal ($p\text{-value} > 0.05$) y por tal motivo no podría hacerse una prueba paramétrica para evaluar diferencias en las medias.

Se podría inferir que los valores de TA de la REFPOP pueden responder a que la REFPOP es una colección de individuos seleccionados y producto de mejoramientos genético para obtener frutos cuyos parámetros de calidad organolépticos sean acordes a las demandas de los consumidores en la UE.

RATIO SST: TA

Gráficos de distribución de datos para Ratio

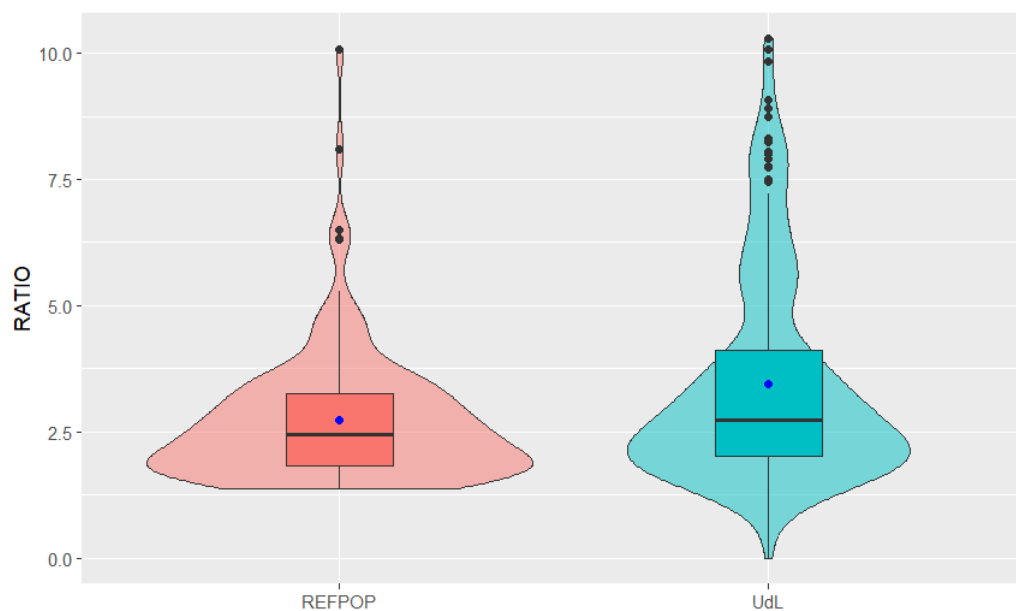


Gráfico 6. Diagramas de violín para RATIO. La línea negra indica la mediana de los datos y el punto azul indica la media.

Resumen de estadígrafos para RATIO.

RATIO								
Colección	n	Media	D.E.	Var(n-1)	CV	Mediana	Q1	Q3
REFPOP	130	2,73	1,32	1,74	48,42	2,43	1,81	3,26
UdL	224	3,45	2,08	4,31	60,14	2,73	2,01	4,16

Tabla 15. Resumen de los estimadores estadísticos para los datos referente al RATIO. En la tabla se detallan el número de individuos (n), la media, el desvío estándar (D.E.), la varianza muestral (Var (n-1)), coeficiente de variación (CV), la mediana y los cuartiles primero y tercero (Q1 y Q3)

Pruebas de normalidad para RATIO.

Pruebas de normalidad							
		Kolmogorov-Smirnov ^b			Shapiro-Wilk		
	Colección	Estadístico	gl	Sig.	Estadístico	gl	Sig.
RATIO	REFPOP	0,153	130	0,000	0,787	130	0,000
	UdL	0,166	224	0,000	0,853	224	0,000

b. Corrección de significación de Lilliefors

Tabla 16. Resumen de los estimadores de la prueba de Kolmogorov-Smirnov y Shapiro-Wilks para cada conjunto de datos referido al RATIO.

La variable RATIO presenta una dispersión de datos considerable como puede observarse en el valor de coeficiente de variación en los estadígrafos de la tabla 15. Estos valores están muy por encima de los valores de CV aceptable para ensayos a campo y en laboratorios (CV >25 %) Lo cual demuestra una gran variabilidad de los datos.

En el gráfico de violines se observa la dispersión de datos, principalmente en datos que sobrepasan los valores de ratio de 5. En los datos de la UdL se puede observar que hay valores de ratio altos (ratio > 6); esto puede deberse a que los individuos de la UdL que conforman el banco de germoplasma tienen mayor variabilidad debido a que muchos ejemplares no son producto de una mejora genética con fines comerciales, a diferencia de los individuos de la REFPOP, los cuales son producto de mejoras genéticas, tal y como ocurre también en el caso de visual *wax content* o días de cosecha.

Las variedades mejor aceptadas en la mayoría de los países europeos, según bibliografía consultada, se caracterizan por tener firmeza, dulzor y una acidez media o inferior a la media. Consumidores de Alemania, Países Bajos y Suiza prefieran

variedades ácidas, jugosas y crujientes. 'Golden Delicious' se incluyó en el grupo de variedades con mayor puntuación en España y Polonia (Bonany, 2017). Los valores muy altos de TA y firmeza, característicos de esta variedad, dieron como resultado un sabor muy ácido y una textura firme; esto puede explicar el bajo nivel de aceptación de esta variedad por parte de un gran grupo de consumidores en muchos países (Bonany et al., 2013). Finalmente, Alemania fue el país que más se diferenciaba del resto. la preferencia Alemania preferencia es un sabor más ácido, frutas jugosas y crujientes (Bonany, 2017).

Los resultados obtenidos de los ratios de la REFPOP muestran que los valores de acidez son los aceptados por la mayoría de países de la UE, ya que el ratio (promedio) con SST altos y una acidez media. Además, se observa que los valores de ratios se agrupan entre 2 – 3.5, lo que demuestra que son producto de una selección más exhaustiva. Las manzanas que componen la UdL son manzanas que presentan más variación en las ratios, se pueden observar valores bajos de acidez y altos valores de SST, haciendo que la ratio presente valores más dispersos que en el caso de la REFPOP, respondiendo a un consumo más acorde al consumidor local, tal como sugiere Bonany et al. (2013)

Parte II: Asociación fenotipo – genotipo para la forma de fruto

Forma de los frutos

En los siguientes gráficos se muestran los datos obtenidos por el Software Shape Analyzer para el fenotipado de las formas de los frutos de cada una de las colecciones.

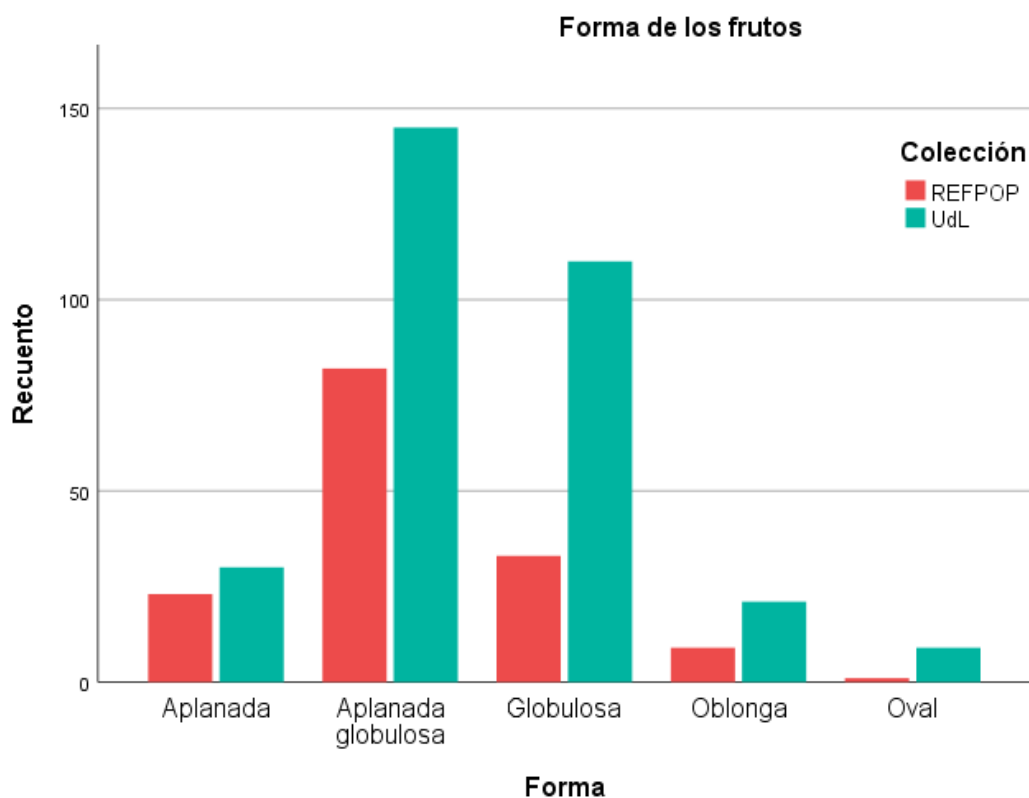


Gráfico 7: Forma de los frutos por árbol para ambas colecciones.

Los datos obtenidos por el software permiten obtener parámetros morfométricos de forma automatizada, efectiva y sencilla. A partir de los datos se construyeron los gráficos 7 y 8. Los árboles de la UdL, presentan una mayor variación en las formas de los frutos, esto es esperable ya que la colección de variedades locales son productos de cruzamientos aleatorios o injertos, realizados por los agricultores.

En contraposición, la colección REFPOP presenta mayoritariamente individuos que producen frutos con formas aplanadas globulosas. Esta diferencia se debe en principio que la colección REFPOP es un conjunto de individuos que fueron seleccionados para producción, y son variedades de manzanas que se comercializan en la UE.

Los datos morfométricos permiten establecer las relaciones entre los valores de FSI externo (FSIE) y FSI interno (FSII) de las caras escaneadas del fruto. De esta forma

se pudo obtener datos de FSIs de algunos ejemplares que componen la REFPOP y los individuos que componen el banco de germoplasma de UdL para posteriores análisis.

Los datos obtenidos por los GWAS de la REFPOP adquiridos en conjunto con los datos de FSI obtenidos por escaneo y analizados por el software *shape analyzer*, permitirán postular posibles SNPs implicados en las formas de las manzanas de algunos individuos de que componen la REFPOP.

La selección de la región comprendida entre los SNPs asociados a forma de fruto se seleccionó de tal forma que la cadena de SNPs no sea demasiado larga y por lo tanto tan específica que no permita agrupar haplotipos iguales, pero tampoco una cadena corta que sea una cadena inespecífica que agrupe varios haplotipos disimiles. Los SNPs asociados a forma y los SNPs que presentan desequilibrio de ligamiento permitieron conseguir una cadena de longitud adecuada para el análisis, debido no solamente al número de SNPs sino también a la longitud en pares de bases (bp) que los separa.

Análisis de regiones cromosómicas

Análisis de SNP AX-115482211 del cromosoma 2

El cromosoma 2 de manzana, contiene el SNP AX-115482211, asociado al fenotipo de forma.

Box plot de FSIE para los haplotipos del Chr 2

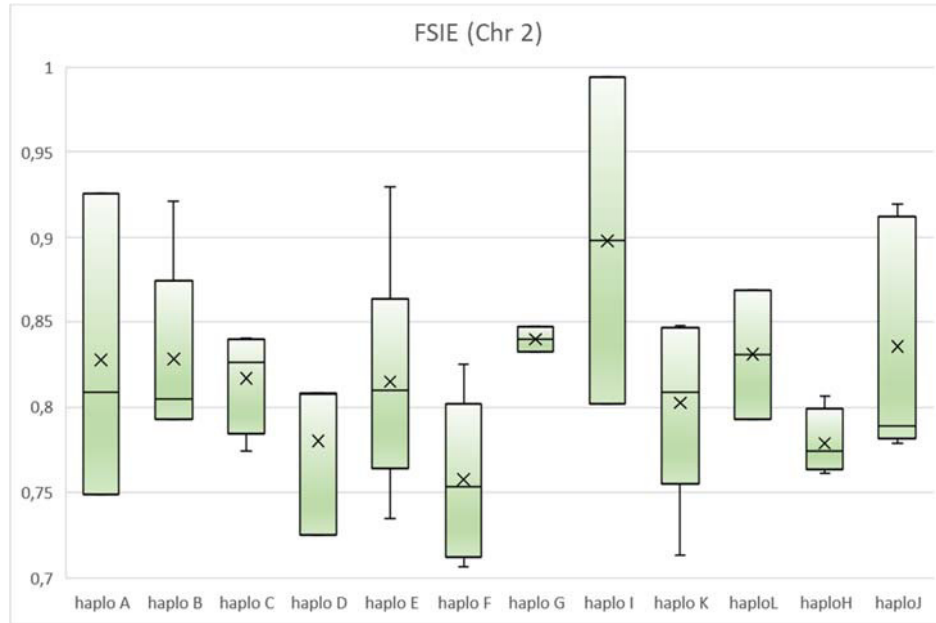


Gráfico 8: Diagrama Box - Plot de distribución de frecuencias de valores de FSI Externo para el Cromosoma 2. La x indica la media y la barra la mediana.

Descriptivos estadísticos de FSIE para los haplotipos del Chr 2

Haplotipo	haplo A	haplo B	haplo C	haplo D	haplo E	haplo F
$\bar{X} \text{ Ext.}$	0,8277	0,8281	0,8170	0,7804	0,8150	0,7578
$SD \text{ Ext.}$	0,0901	0,0542	0,0307	0,0476	0,0637	0,0456
$Error \text{ Ext.}$	0,0520	0,0242	0,0153	0,0275	0,0225	0,0186

Haplotipo	haplo G	Haplo H	haplo I	Haplo J	haplo K	haplo L
$\bar{X} \text{ Ext.}$	0,8397	0,7790	0,8982	0,8355	0,8026	0,8309
$SD \text{ Ext.}$	0,0105	0,0194	0,1360	0,0704	0,0544	0,0541
$Error \text{ Ext.}$	0,0075	0,0097	0,0961	0,0315	0,0243	0,0383

Tabla 17: Descriptores estadísticos para FSI externo para haplotipos del cromosoma 2.

Box plot de FSII para los haplotipos del Chr 2

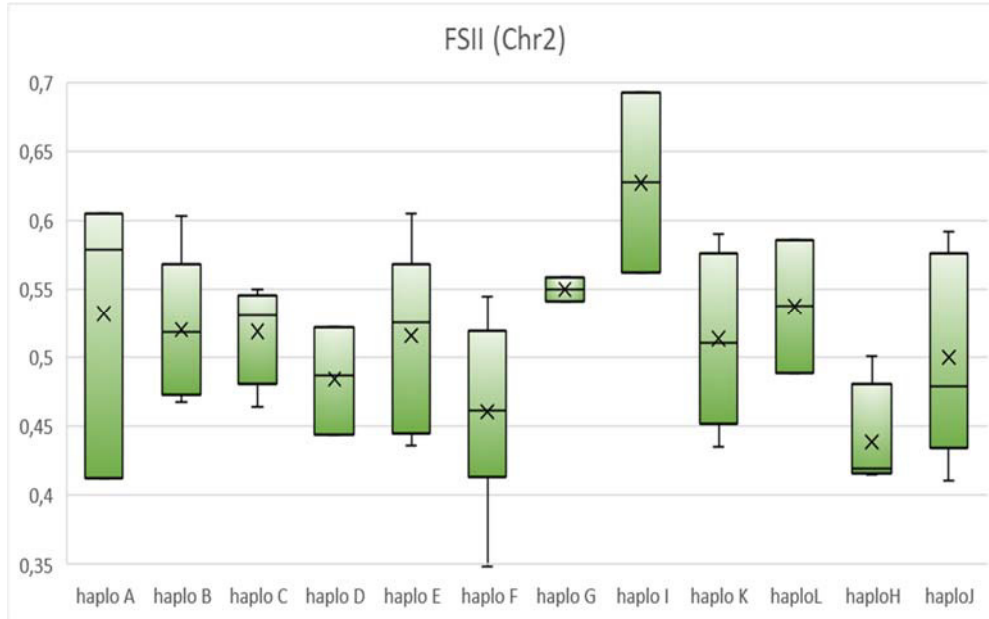


Gráfico 9 : Diagrama Box - Plot de distribución de frecuencias de valores de FSI Interno para el Cromosoma 2. La x indica la media y la barra la mediana.

Descriptivos estadísticos de FSIE para los haplotipos del Chr 2

haplotipo	haplo A	haplo B	haplo C	haplo D	haplo E	haplo F
$\bar{X} Int.$	0,5322	0,5201	0,5191	0,4843	0,5160	0,4605
$SD Int.$	0,1048	0,0539	0,0377	0,0388	0,0642	0,0688
$Error Int.$	0,0605	0,0311	0,0188	0,0224	0,0227	0,0281

haplotipo	haplo G	Haplo H	haplo I	Haplo J	haplo K	haplo L
$\bar{X} Int.$	0,5501	0,4390	0,6277	0,5002	0,5135	0,5375
$SD Int.$	0,0126	0,0412	0,0925	0,0745	0,0640	0,0686
$Error Int.$	0,0089	0,0206	0,0654	0,0333	0,0286	0,0485

Tabla 18: descriptores estadísticos para los haplotipos para los valores de FSI interno del cromosoma 2.

En el caso del cromosoma 2 la cadena de SNPs utilizada para el análisis constó de 12 caracteres.

Según los datos de la tabla 18 y el gráfico 10, se puede observar que para el cromosoma 2 muestran valores promedios de FSIE que se corresponden a formas “aplanada globulosa”. Se aprecia que, de los 12 haplotipos, solamente 1 haplotipo

(haplotipo) presenta un valor promedio de FSIE superior a 0.85 que corresponde a una forma “globulosa”.

Los valores promedios de FSII de los haplotipos analizados corresponden a frutos de forma “cónica”, los valores de FSII medidos para cada individuo también corresponden a valores de FSII de formas cónicas; esta inferencia se puede observar en los *boxplot* (Gráfico 10) que no superan los valores de 0.715. Además, no se observa dispersión de los datos, como se puede observar en los gráficos de *boxplot* (Gráfico 10) y las desviaciones estándar (SD) listadas en la tabla 18.

Análisis de los SNPs AX-115327898 y AX-115327900 del cromosoma 11

El cromosoma 11 de manzana, posee 2 haplobloques: Uno contiene el SNP AX-115327898, asignado con la letra (a) y otro que contiene el SNP AX-115327900, asignado con la letra b). Para el análisis de SNPs se eligieron cadenas que contengan SNPs que no se repiten en ambos haplobloques en este cromosoma.

Box plot de FSIE para los haplotipos del Chr 11 (a)

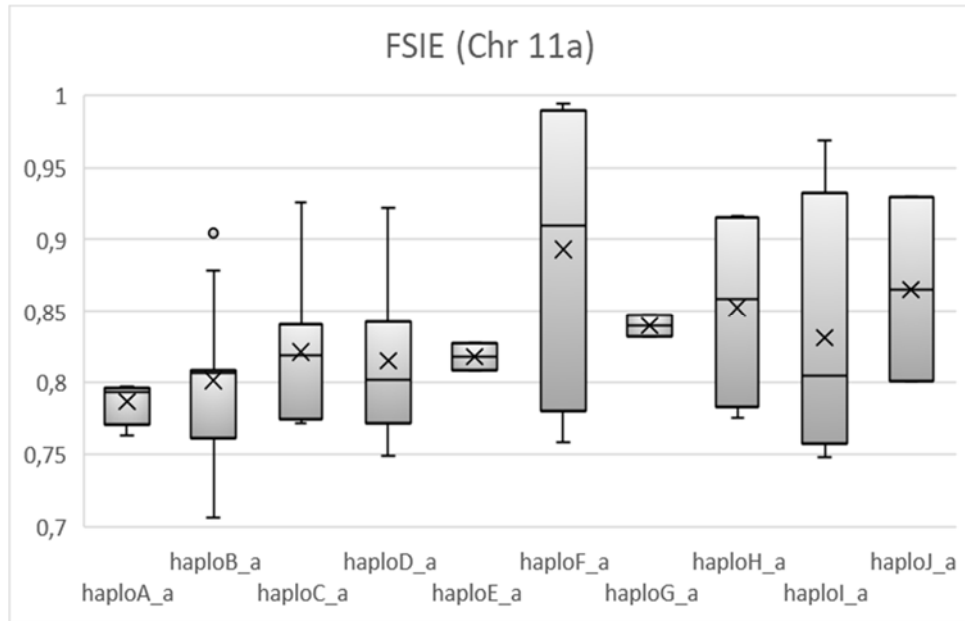


Gráfico 10: Diagrama Box - Plot de distribución de frecuencias de valores de FSI externo para el Cromosoma 11. La x indica la media y la barra la mediana.

Descriptivos estadísticos de FSIE para los haplotipos del Chr 11 (a)

Haplotipo	Haplo A_a	Haplo B_a	Haplo C_a	Haplo D_a	Haplo E_a	Haplo F_a	Haplo G_a	Haplo H_a	Haplo I_a	Haplo J_a
$\bar{X} Ext$	0,7869	0,8015	0,8215	0,8153	0,8183	0,8934	0,8400	0,8525	0,8315	0,8655
$SD Ext$	0,0159	0,0565	0,0550	0,0551	0,0133	0,1114	0,0110	0,0728	0,0965	0,0909
$Error Ext$	0,0079	0,0170	0,0208	0,0153	0,0094	0,0557	0,0078	0,0364	0,0482	0,0642

Tabla 19: Descriptivos estadísticos para cada haplotipo para el FSI externo del cromosoma 11a.

Box plot de FSII para los haplotipos del Chr 11 (a)

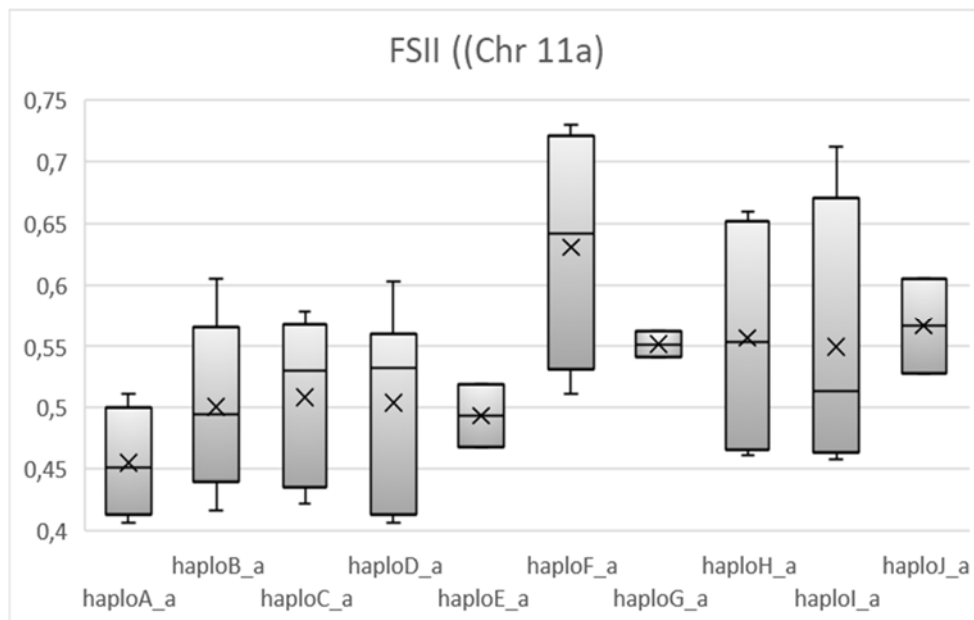


Gráfico 11: Diagrama Box - Plot de distribución de frecuencias de valores de FSI Interno para el Cromosoma 11. La x indica la media y la barra la mediana.

Descriptivos estadísticos de FSII para los haplotipos del Chr 11 (a)

Haplotipo	Haplo A_a	Haplo B_a	Haplo C_a	Haplo D_a	Haplo E_a	Haplo F_a	Haplo G_a	Haplo H_a	Haplo I_a	Haplo J_a
\bar{X}_{Int}	0,4552	0,5009	0,5085	0,5043	0,4936	0,6311	0,5516	0,5570	0,5494	0,5670
SD_{Int}	0,0449	0,0646	0,0662	0,0704	0,0354	0,0993	0,0148	0,1018	0,1146	0,0546
Error Int	0,0224	0,0195	0,0250	0,0195	0,0250	0,0496	0,0104	0,0509	0,0573	0,0386

Tabla 20: Descriptivos estadísticos para cada haplotipo para el FSI interno del cromosoma 11a.

En el caso del cromosoma 11 haplobloque (a) la cadena de SNPs utilizada para el análisis constó de 11 caracteres.

Los resultados de los análisis de haplotipos para el haplobloque (a) muestra una distribución de boxplot (Gráfico 11) centrados en valores promedios que corresponden a formas de manzana “aplanada globulosa”. De los 10 haplotipos analizados 7 de ellos presentan la forma más representativa de toda la REFPOP, que es aplanada globulosa según el gráfico 7. Los datos medidos para este valor presentan dispersión de datos que se observan en los boxplot (gráfico 11) y en la tabla 19 aunque la media y la mediana se encuentran por debajo del valor 0.85 para

los valores de FSIE. Se observa que 3 haplotipos presentan valores superiores a 0.85 que corresponde a una forma oblonga, pero se observa que hay individuos que poseen valores de FSIE medidos por debajo de ese valor.

El FSII promedio calculado corresponde a una forma de fruto cónica, los valores representados en el gráfico 12 se correlacionan con los valores calculados; esto se puede observar en los *boxplot* (gráfico 12) y en la tabla 20.

Box plot de FSIE para los haplotipos del Chr 11 (b)

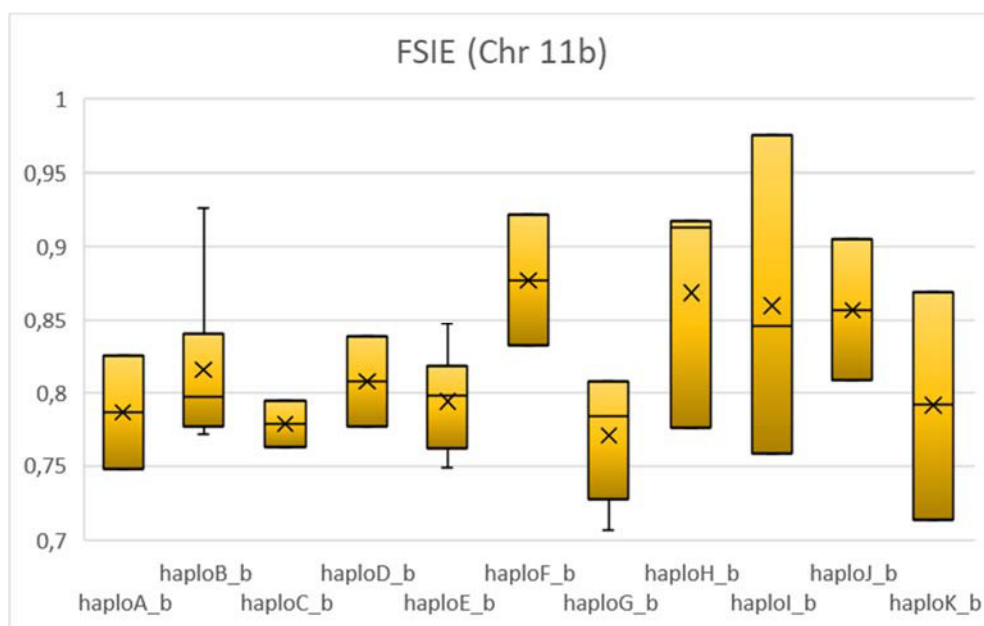


Gráfico 12: Diagrama Box - Plot de distribución de frecuencias de valores de FSI externo para el Cromosoma 11. La x indica la media y la barra la mediana.

Descriptivos estadísticos de FSIE para los haplotipos del Chr 11 (b)

Haplotipo	Haplo A_b	Haplo B_b	Haplo C_b	Haplo D_b	Haplo E_b	Haplo F_b	Haplo G_b	Haplo H_b	haplo I_b	Haplo J_b	Haplo K_b
$\bar{X} Ext$	0,7864	0,8156	0,7787	0,8075	0,7942	0,8770	0,7708	0,8685	0,8597	0,8566	0,7914
$SD Ext$	0,0545	0,0490	0,0218	0,0436	0,0334	0,0633	0,0436	0,0802	0,1088	0,0680	0,1100
$Error Ext$	0,0385	0,0163	0,0154	0,0309	0,0118	0,0448	0,0178	0,0463	0,0628	0,0481	0,0778

Tabla 21: Descriptivos estadísticos para cada haplotipo para el FSI externo del cromosoma 11b

Box plot de FSII para los haplotipos del Chr 11 (b)

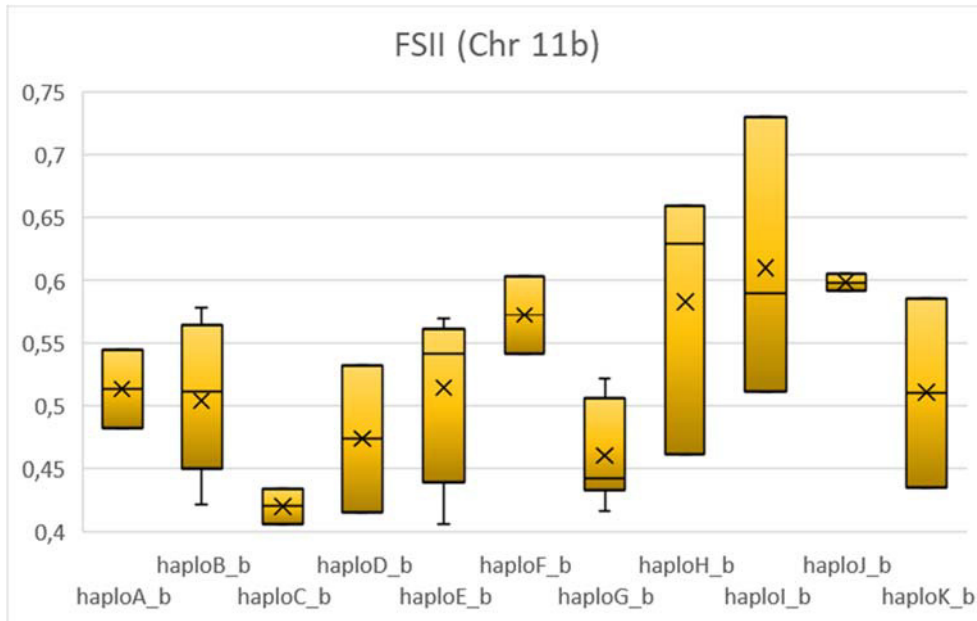


Gráfico 13: Diagrama Box - Plot de distribución de frecuencias de valores de FSI Interno para el Cromosoma 11. La x indica la media y la barra la mediana.

Descriptivos estadísticos de FSII para los haplotipos del Chr 11 (b)

Haplotype	Haplo A_a	Haplo B_a	Haplo C_a	Haplo D_a	Haplo E_a	Haplo F_a	Haplo G_a	Haplo H_a	Haplo I_a	Haplo J_a
\bar{X}_{Int}	0,4552	0,5009	0,5085	0,5043	0,4936	0,6311	0,5516	0,5570	0,5494	0,5670
SD_{Int}	0,0449	0,0646	0,0662	0,0704	0,0354	0,0993	0,0148	0,1018	0,1146	0,0546
$Error_{Int}$	0,0224	0,0195	0,0250	0,0195	0,0250	0,0496	0,0104	0,0509	0,0573	0,0386

Tabla 22: Descriptivos estadísticos para cada haplotipo para el FSI interno del cromosoma 11b

En el caso del cromosoma 2 la cadena de SNPs utilizada para el análisis constó de 7 SNPs. Los descriptivos estadísticos para los haplotipos que se analizaron para el haplobloque se listan en las tablas 21 y 22. También se puede observar en los gráficos de boxplot 13 y 14, FSIE y FSII.

Los valores de FSIE para el haplobloque (b) para 7 haplotipos presentan un FSIE promedio que corresponde a valores inferiores a 0.85 pero superiores a 0.75, esto corresponde a una forma aplanada globulosa.

Análisis de SNP AX-115355048 en el cromosoma 13

El cromosoma 13 de manzana, contiene el SNP AX-115482211, asociado al fenotipo de forma.

Box plot de FSIE para los haplotipos del Chr 13

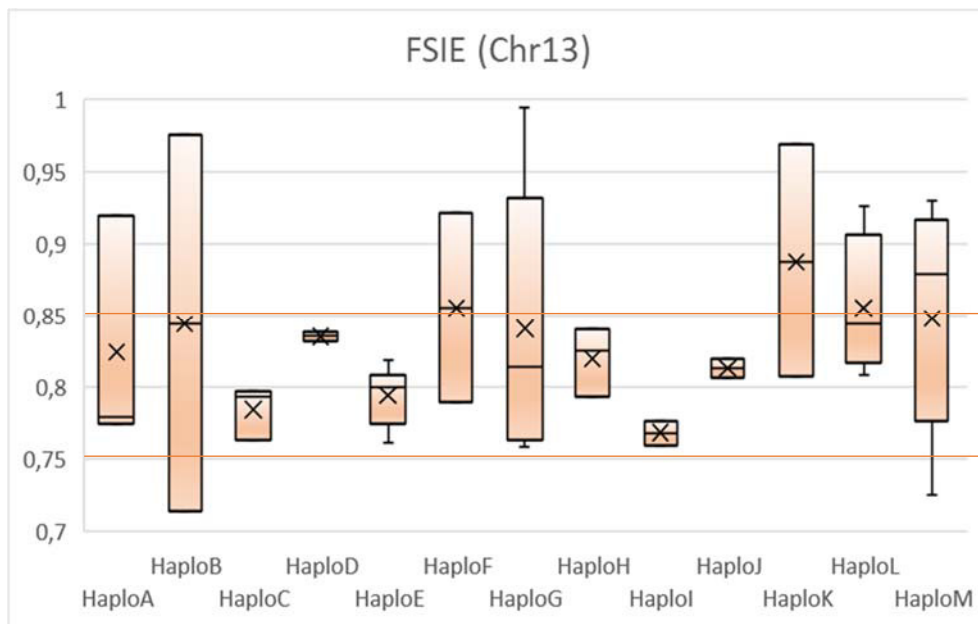


Gráfico 14: Diagrama Box - Plot de distribución de frecuencias de valores de FSI Externo para el Cromosoma 13. La x indica la media y la barra la mediana.

Descriptivos estadísticos de FSIE para los haplotipos del Chr 13

Haplotipo	HaploA	HaploB	HaploC	HaploD	HaploE	HaploF	HaploG
\bar{X} Ext.	0,824	0,844	0,784	0,835	0,794	0,855	0,841
SD Ext.	0,083	0,185	0,019	0,004	0,020	0,094	0,096
Error Ext	0,048	0,131	0,011	0,003	0,007	0,066	0,043

Haplotipo	HaploH	HaploI	HaploJ	HaploK	HaploL	HaploM
\bar{X} Ext.	0,820	0,768	0,813	0,888	0,855	0,848
SD Ext.	0,024	0,012	0,009	0,114	0,050	0,079
Error Ext.	0,014	0,008	0,006	0,081	0,025	0,030

Tabla 23: Descriptivos estadísticos para cada haplotipo para el FSI externo del cromosoma 13

Box plot de FSII para los haplotipos del Chr 13

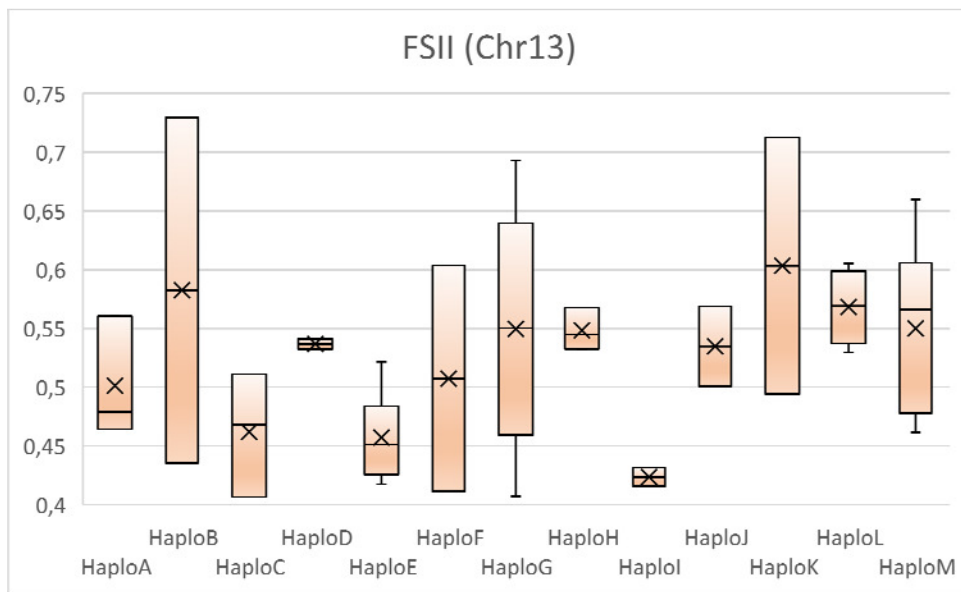


Gráfico 15: Diagrama Box - Plot de distribución de frecuencias de valores de FSI interno para el Cromosoma 13. La x indica la media y la barra la mediana.

Descriptivos estadísticos de FSII para los haplotipos del Chr 13

Haplotipo	HaploA	HaploB	HaploC	HaploD	HaploE	HaploF	HaploG
\bar{X} Int.	0,5012	0,5826	0,4619	0,5368	0,4571	0,5073	0,5495
SD Int.	0,0517	0,2081	0,0524	0,0062	0,0353	0,1360	0,1045
Error Int.	0,0299	0,1472	0,0303	0,0044	0,0125	0,0961	0,0468

Haplotipo	HaploH	HaploI	HaploJ	HaploK	HaploL	HaploM
------------------	---------------	---------------	---------------	---------------	---------------	---------------

Tabla 24: Descriptivos estadísticos para cada haplotipo para el FSI interno del cromosoma 13.

\bar{X} Int.	0,5484	0,4236	0,5348	0,6033	0,5683	0,5500
SD Int.	0,0178	0,0114	0,0483	0,1542	0,0319	0,0754
Error Int.	0,0103	0,0081	0,0341	0,1091	0,0160	0,0285

En el caso del cromosoma 13 la cadena de SNPs utilizada para el análisis constó de 13 caracteres.

Los haplotipos de los individuos evaluados para el cromosoma 13, indican una prevalencia de formas “aplanadas globulosas”, ya que los promedios se encuentran entre los valores 0.75 y 0.85 de FSIE (gráfico 15 y tabla 23). Se puede observar en los promedios que 10 haplotipos responden a la forma “aplanada globulosa” mientras que otros tres haplotipos responden a una forma “globulosa”.

Los valores de FSII promedios listados en la tabla 24, se corresponden a una forma cónica en las manzanas (figura 3).

El análisis de haplotipos se llevó a cabo con un número de haplotipos y de muestras pequeño, tanto de haplotipos seleccionados, es decir aquellos haplotipos que tenían un número de individuos mayor igual a 2 ($n \geq 2$) como así también el número de individuos seleccionados dentro la REFPOP, ya que solamente se seleccionaron un total de 41 individuos. Por lo tanto, en futuros estudios de la REFPOP se debería repetir este análisis con mayor número de muestras y en diversos años para evaluar si las asociaciones se mantienen.

Los SNPs asociados con forma de los cromosomas 2,11 y 13 mencionados en el anteriormente, y los SNPs que presentan *desequilibrio de ligamiento* debido a su cercanía física en un cromosoma se presentan juntos de manera más frecuente de que lo esperado por azar (ver anexo 2).

El análisis de los SNPs AX-115482211 del cromosoma 2, AX-115327898 y AX-115327900 presentes en el cromosoma 11, y el AX-115355048 en el cromosoma 13 mostraron que la forma de fruto responde al fenotipo “aplanado globuloso” para la mayoría de los haplotipos estudiados.

Los valores de FSIE promedio para los haplotipos estudiados guardan estrecha correlación con los FSIE medidos para los individuos seleccionados de la REFPOP.

La cadena de SNPs entre el SNP asociado a la forma y los SNPs en desequilibrio de ligamiento sugiere otros posibles SNPs candidatos asociados a la forma del fruto. Los haplotipos analizados muestran una tendencia a la forma “aplanada globulosa cónica” y hay haplotipos ligados a la forma “globulosa cónica” en los haplobloques seleccionados.

A futuro sería de gran importancia realizar un análisis similar para el banco de germoplasma de la UdL, ya que nos permitirá comprender las características de las variedades ya fenotipadas de dicha colección. Los datos de fenotipos obtenidos - descritos en la primera parte de este trabajo- servirán en el momento en el que estas variedades sean genotipadas.

Conclusiones

Parte I: conclusiones del fenotipado de caracteres

En la primera parte del presente trabajo se obtuvieron diversos parámetros de calidad de los frutos de las dos colecciones. Estos parámetros fenotípicos serán parte de un conjunto de datos que se obtienen año a año para poder determinar las características fenotípicas que poseen los individuos que conforman las

colecciones. A su vez, esto permitirá asociar las distintas características fenotípicas con marcadores moleculares, principalmente SNPs, para poder establecer correlaciones entre las distintas características y un gen o conjunto de genes candidatos y/o causales.

Los datos obtenidos en esta primera parte permitirán construir una base de datos fenotípicos de las diversas colecciones y se podrá hacer un seguimiento de estos parámetros a lo largo del tiempo. Estas bases de datos permitirán a futuro poder vincular otros potenciales SNPs de otros parámetros de calidad de fruta; por supuesto que entre más grande sea la base de datos de caracteres fenotípicos más eficiente será la asociación entre los SNPs y el carácter fenotípico asociado.

Por otra parte, la obtención de estos datos permitirá a futuro, poder diseñar y planear los distintos cruzamientos para poder identificar las características potenciales de nuevos cultivares, sin necesidad de esperar a que los nuevos cultivares obtenidos produzcan frutos – una media de 5 años- sino que obteniendo los mapas genéticos de los progenitores y/o el mapa genético de los nuevos cultivares se puede inferir de forma rápida las características que tendrá el nuevo cultivar, si es apto o no para un clima o para determinados mercados internacionales.

En el caso de las colecciones de la UdL estos ejemplares pueden a futuro brindar información de caracteres para mejoras en entornos regionales, por ejemplo: forma de frutos, sólidos solubles totales, entre otras características. Estas variedades si bien no presentan características de calidad óptima para la mayoría de los estándares de consumidores, son de consumo locales; además son importantes como fuente de información genética en estudios de características fenotípicas o origen de parentales, o fuente de material para mejoras genéticas.

Es importante, a futuro, poder obtener datos genotípicos y hacer estudios genotípicos de estas variedades, por ejemplo, un GWAS de estas colecciones y

que pueda correlacionarlos con los datos fenotípicos obtenidos a lo largo de un periodo de años.

Parte II

El desarrollo de un software para el escaneo y medición de las medidas de altura y anchura; como también de medidas internas permite obtener medidas más precisas y simplificar el trabajo de medición manual, reduciendo las probabilidades de errores en las mediciones. La tecnología de IA de *neuronal network* es una importante herramienta que permite evaluar frutos que no siguen simetrías definidas, o frutos que presentan deformidades.

El entrenamiento de este programa es un paso clave para el correcto desarrollo del software, por esta razón el procesamiento manual de las primeras imágenes de manzanas para el entrenamiento del programa fue un paso sumamente importante para i) entrenar el programa y que a posteriori el aprendizaje del programa permita reconocer las caras internas de los frutos y ii) obtener las medidas de altura, diámetro, cubetas internas de las frutas que se utilizaron para las medidas de FSII y FSIE.

Análisis de SNPs para el carácter forma de fruto

Como se mencionó en las discusiones de los resultados de la parte II es sumamente importante destacar que el número de muestras utilizadas en este trabajo fue bajo, tanto de individuos seleccionados dentro de la REFPOP como el número de haplotipos analizados. Un número mayor de muestras permitiría un mayor número de haplotipos distintos, y se vería afectada la variabilidad intra-haplotipo, lo que

conllevaría a un análisis más ajustado de la región analizada. Esto se reflejaría en la necesidad de modificar la longitud de la región analizado o inclusive seleccionar otra región, es decir, que la región seleccionada no sea tan específica que permita asociar los SNPs con el carácter fenotípico de forma de fruto.

Este trabajo permitió el fenotipado de un banco de germoplasma y una colección para diferentes caracteres de importancia agronómica como: SST, días de cosecha, acidez total titulable, visual wax content, ratio, lo cual dio pie a un análisis estadístico descriptivo y asociación fenotipo-genotipo para la forma del fruto. Las regiones analizadas brindarán en un futuro información para seguir estudiando los SNPs asociados a las formas de los frutos tanto para la UdL como para la REFPOP, además de otras características agronómicas mencionadas anteriormente.

Bibliografía

- Belding, R. D., Blankenship, S. M., Young, E., & Leidy, R. B. (1998). Composition and Variability of Epicuticular Waxes in Apple Cultivars. *Journal of the American Society for Horticultural Science jashs*, 123(3), 348-356. <https://doi.org/10.21273/JASHS.123.3.348>
- Blanpied, G. D., & Silsby, K. (1992). Predicting Harvest Date Windows for Apples. *Information bulletin* 221, 221.
- Bonany, I. I. C. J. C. P. J. (2017). *Innovación varietal en manzanas: Situación y perspectivas de futuro (parte I)* (Revista Frutiocola, Issue.
- Bonany, J., Buehler, A., Carbó, J., Codarin, S., Donati, F., Echeverria, G., Egger, S., Guerra, W., Hilaire, C., Höller, I., Iglesias, I., Jesionkowska, K., Konopacka, D., Kruczyńska, D., Martinelli, A., Pitiot, C., Sansavini, S., Stehr, R., & Schoorl, F. (2013). Consumer eating

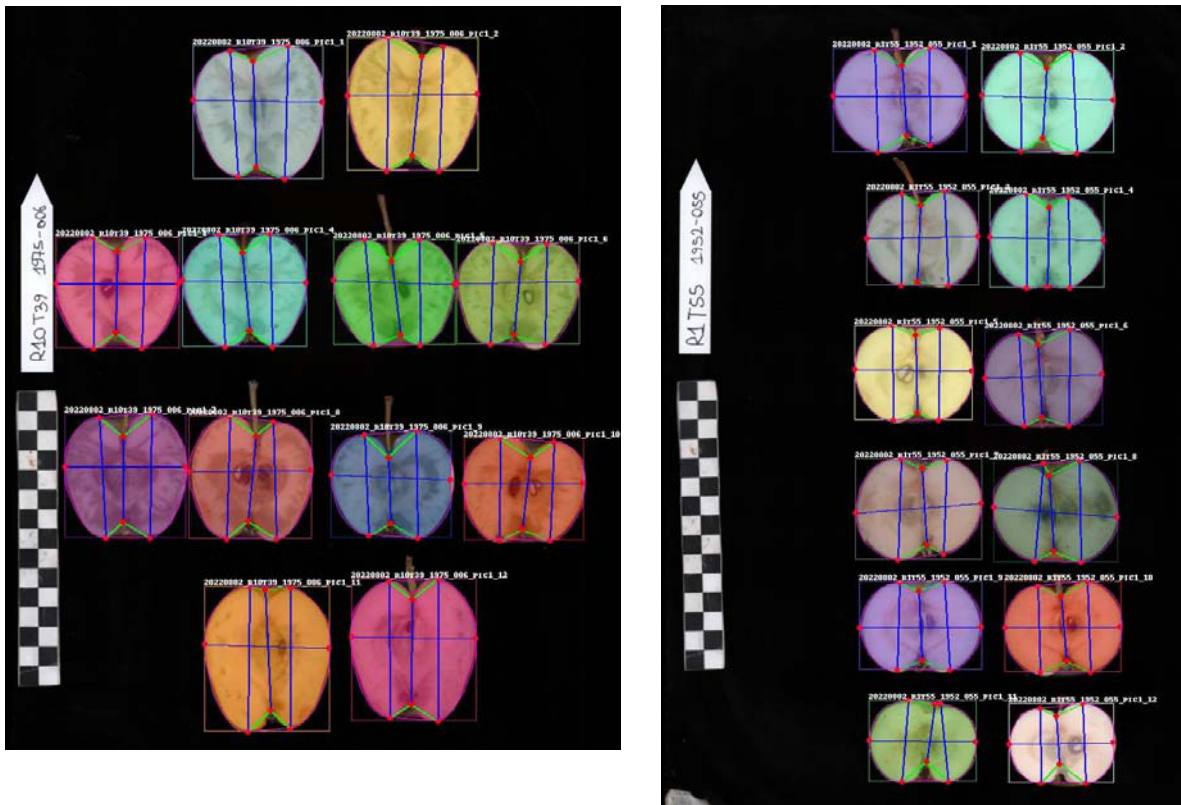
- quality acceptance of new apple varieties in different European countries. *Food Quality and Preference*, 30(2), 250-259.
<https://doi.org/https://doi.org/10.1016/j.foodqual.2013.06.004>
- Chang, Y., Sun, R., Sun, H., Zhao, Y., Han, Y., Chen, D., Wang, Y., Zhang, X., & Han, Z. (2014). Mapping of quantitative trait loci corroborates independent genetic control of apple size and shape. *Scientia Horticulturae*, 174, 126-132.
<https://doi.org/https://doi.org/10.1016/j.scienta.2014.05.019>
- Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., & Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica*, 142(1), 169-196.
<https://doi.org/10.1007/s10681-005-1681-5>
- Cornille, A., Giraud, T., Smulders, M. J., Roldan-Ruiz, I., & Gladieux, P. (2014). The domestication and evolutionary ecology of apples. *Trends Genet*, 30(2), 57-65.
<https://doi.org/10.1016/j.tig.2013.10.002>
- Currie, A. J., Ganeshanandam, S., Noiton, D. A., Garrick, D., Shelbourne, C. J. A., & Oraguzie, N. (2000). Quantitative evaluation of apple (*Malus × domestica* Borkh.) fruit shape by principal component analysis of Fourier descriptors. *Euphytica*, 111(3), 221-227.
<https://doi.org/10.1023/A:1003862525814>
- Daccord, N., Celton, J.-M., Linsmith, G., Becker, C., Choisine, N., Schijlen, E., van de Geest, H., Bianco, L., Micheletti, D., Velasco, R., Di Pierro, E. A., Gouzy, J., Rees, D. J. G., Guérif, P., Muranty, H., Durel, C.-E., Laurens, F., Lespinasse, Y., Gaillard, S., . . . Bucher, E. (2017). High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nature Genetics*, 49(7), 1099-1106. <https://doi.org/10.1038/ng.3886>
- Dapena de la Fuente, E. B. N., M.D. (2009). *Descripción de las variedades de manzana de la D.O.P. Sidra de Asturias*. Villaviciosa, Asturias, España: SERIDA
- Devoghalaere, F., Doucen, T., Guitton, B., Keeling, J., Payne, W., Ling, T. J., Ross, J. J., Hallett, I. C., Gunaseelan, K., Dayatilake, G. A., Diak, R., Breen, K. C., Tustin, D. S., Costes, E., Chagné, D., Schaffer, R. J., & David, K. M. (2012). A genomics approach to understanding the role of auxin in apple (*Malus x domestica*) fruit size control. *BMC Plant Biology*, 12(1), 7.
<https://doi.org/10.1186/1471-2229-12-7>
- Di Rienzo, J. A., Casanoves, F., Balzarini, M. G., Gonzalez, L., Tablada, M., & Robledo, C. W. (2011). *InfoStat*. Universidad Nacional de Córdoba. <http://www.infostat.com.ar/>
- Dujak, C., & Aranzana, M. J. (2023). Genomic analysis of fruit size and shape traits in apple: unveiling candidate genes through GWAS analysis. *bioRxiv*, 2023.2008.2021.554124.
<https://doi.org/10.1101/2023.08.21.554124>
- FAO. (2014). *Normas para bancos de germoplasma de recursos fitogenéticos para la alimentación y la agricultura*
- FAO. (2022). FAOSTAT. <https://www.fao.org/faostat/es/>
- Gergoff Grozeff, G. (2019). *Maduración e Índices de cosecha de frutos. Aspectos fisiológicos. Determinación de estados de madurez* [Guía de Estudios y Trabajos Prácticos]. U. N. d. L. Plara.
- Iglesia Castellarnau, I. C. P., Joaquim; Bonany, Joan. (2017). Innovación varietal en manzanas: Situación y perspectivas de futuro (I). *Revista Frutiorticola. Copefrut S.A.*, 39(2).
- Jung, M., Keller, B., Roth, M., Aranzana, M. J., Auwerkerken, A., Guerra, W., Al-Rifai, M., Lewandowski, M., Sanin, N., Rymenants, M., Didelot, F., Dujak, C., Forcada, C. F., Knauf, A., Laurens, F., Studer, B., Muranty, H., & Patocch, A. (2022). Genetic architecture and genomic predictive ability of apple quantitative traits across environments. *Hortic Res.*
<https://doi.org/10.1093/hr/uhac028>
- Jung, M., Roth, M., Aranzana, M. J., Auwerkerken, A., Bink, M., Denancé, C., Dujak, C., Durel, C.-E., Font i Forcada, C., Cantin, C. M., Guerra, W., Howard, N. P., Keller, B., Lewandowski, M., Ordidge, M., Rymenants, M., Sanin, N., Studer, B., Zurawicz, E., . . . Muranty, H. (2020). The apple REFPOP—a reference population for genomics-assisted breeding in apple. *Horticulture Research*, 7(1), 189. <https://doi.org/10.1038/s41438-020-00408-8>

- Jung, M., Roth, M., Aranzana, M. J., Auwerkerken, A., Bink, M., Denance, C., Dujak, C., Durel, C. E., Font, I. F. C., Cantin, C. M., Guerra, W., Howard, N. P., Keller, B., Lewandowski, M., Ordidge, M., Rymenants, M., Sanin, N., Studer, B., Zurawicz, E., . . . Muranty, H. (2020). The apple REFPOP-a reference population for genomics-assisted breeding in apple. *Hortic Res*, 7(1), 189. <https://doi.org/10.1038/s41438-020-00408-8>
- Keller, B., Jung, M., Bühlmann-Schütz, S., Hodel, M., Studer, B., Broggin, G. A., & Patocchi, A. (2023). The genetic basis of apple shape and size unraveled by digital phenotyping. *bioRxiv*, 2023.2008.2025.554767. <https://doi.org/10.1101/2023.08.25.554767>
- Lateur, M., Dapena, E., Szalatnay, D., Guyader, A., Hjalmarsson, I., Höfer, M., Militaru, M., Jiménez, C. M., Osterc, G., & Rondia, A. U. h. b. g. e. b. i. y. (2022). *ECPGR Characterization and Evaluation Descriptors for Apple Genetic Resources: Apple (Malus X Domestica)*. ECPGR - European Cooperative Programme for Plant Genetic Resources.
- Laurens, F., Aranzana, M. J., Arus, P., Bassi, D., Bink, M., Bonany, J., Caprera, A., Corelli-Grappadelli, L., Costes, E., Durel, C.-E., Mauroux, J.-B., Muranty, H., Nazzicari, N., Pascal, T., Patocchi, A., Peil, A., Quilot-Turion, B., Rossini, L., Stella, A., . . . van de Weg, E. (2018). An integrated approach for increasing breeding efficiency in apple and peach in Europe. *Horticulture Research*, 5(1), 11. <https://doi.org/10.1038/s41438-018-0016-3>
- Li, H., Dong, Q., Zhao, Q., & Ran, K. (2019). Genome-wide identification, expression profiling, and protein-protein interaction properties of ovate family proteins in apple. *Tree Genetics & Genomes*, 15(3), 45. <https://doi.org/10.1007/s11295-019-1354-5>
- Migicovsky, Z., Gardner, K. M., Richards, C., Thomas Chao, C., Schwaninger, H. R., Fazio, G., Zhong, G. Y., & Myles, S. (2021). Genomic consequences of apple improvement. *Hortic Res*, 8(1), 9. <https://doi.org/10.1038/s41438-020-00441-7>
- Ministerio de Agricultura, P. y. A. (2023). *Análisis campaña 2022-23 y perspectivas 2023-24*. Retrieved from https://www.mapa.gob.es/es/agricultura/temas/producciones-agricolas/230923pepitacampana2022-23yperspectivas2023-24sectorial26-9-23_tcm30-660773.pdf
- Montoya Jaramillo, C. (2016). ¿Qué se debe saber sobre los QTL (Quantitative Trait Loci) y su utilidad en el mejoramiento genético de palma de aceite? *Palmas*, 37(3), 65-75. <https://publicaciones.fedepalma.org/index.php/palmas/article/view/11807>
- Offord, C. A. (2017). Germplasm Conservation. In B. Thomas, B. G. Murray, & D. J. Murphy (Eds.), *Encyclopedia of Applied Plant Sciences (Second Edition)* (pp. 281-288). Academic Press. <https://doi.org/10.1016/B978-0-12-394807-6.00046-0>
- Patricia Denne, M. (1963). Fruit development and some tree factors affecting it. *New Zealand Journal of Botany*, 1(3), 265-294. <https://doi.org/10.1080/0028825X.1963.10428999>
- Paulus, I., & Schrevers, E. (1999). Shape Characterization of New Apple Cultivars by Fourier Expansion of Digitized Images. *Journal of Agricultural Engineering Research*, 72(2), 113-118. <https://doi.org/10.1006/jaer.1998.0352>
- Peace, C. P., Bianco, L., Troggio, M., van de Weg, E., Howard, N. P., Cornille, A., Durel, C.-E., Myles, S., Migicovsky, Z., Schaffer, R. J., Costes, E., Fazio, G., Yamane, H., van Nocker, S., Gottschalk, C., Costa, F., Chagné, D., Zhang, X., Patocchi, A., . . . Vanderzande, S. (2019). Apple whole genome sequences: recent advances and new prospects. *Horticulture Research*, 6, 59. <https://doi.org/10.1038/s41438-019-0141-7>
- Risdin, F., Mondal, P. K., & Hassan, K. M. (2020). Convolutional neural networks (CNN) for detecting fruit information using machine learning techniques. *IOSR Journal of Computer Engineering (IOSR-JCE)*, 22(2), 01-13.
- Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., & McCool, C. (2016). DeepFruits: A Fruit Detection System Using Deep Neural Networks. *Sensors*, 16(8).
- Saedi, S. I., & Khosravi, H. (2020). A deep neural network approach towards real-time on-branch fruit recognition for precision horticulture. *Expert Systems with Applications*, 159, 113594. <https://doi.org/10.1016/j.eswa.2020.113594>
- Secretaria de Agricultura, G. y. P. (2022). *Pera y Manzana. Informe 2021 y avance 2022*. República Argentina Retrieved from <https://www.argentina.gob.ar/sites/default/files/2021/09/perasymanzanas-sep2022-1.pdf>

- Shen, F., Huang, Z., Zhang, B., Wang, Y., Zhang, X., Wu, T., Xu, X., Zhang, X., & Han, Z. (2019). Mapping Gene Markers for Apple Fruit Ring Rot Disease Resistance Using a Multi-omics Approach. *G3 (Bethesda)*, 9(5), 1663-1678. <https://doi.org/10.1534/g3.119.400167>
- Thi Phuong Chung, D., & Van Tai, D. (2019). A fruits recognition system based on a modern deep learning technique. *Journal of Physics: Conference Series*, 1327(1), 012050. <https://doi.org/10.1088/1742-6596/1327/1/012050>
- Toranzo, J. (2016). *Producción mundial de manzanas y peras* (I. N. d. T. A. (INTA), Ed. 1era edición ed.).
- Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, 1(1), 59. <https://doi.org/10.1038/s43586-021-00056-9>
- Verardo, G., Pagani, E., Geatti, P., & Martinuzzi, P. (2003). A thorough study of the surface wax of apple fruits. *Analytical and Bioanalytical Chemistry*, 376(5), 659-667. <https://doi.org/10.1007/s00216-003-1945-7>
- Zadravec, P., Veberic, R., Stampar, F., Eler, K., & Schmitzer, V. (2013). Fruit size prediction of four apple cultivars: Accuracy and timing. *Scientia Horticulturae*, 160, 177-181. <https://doi.org/https://doi.org/10.1016/j.scienta.2013.05.046>

Material suplementario

Escaneo y reconocimiento del software *Shape Analyzer*



Fotografía N° 2. Escaneo y medidas obtenidas por el programa

Tablas de desequilibrio de ligamiento

Cromosoma 2

	CHR A	BP A	SNP A	CHR B	BP B	SNP B
109092	2	21759081	AX-115482211	2	21759630	AX-115482210

Cromosoma 11

	CHR A	BP A	SNP A	CHR B	BP B	SNP B
587952	11	4695971	AX-115327894	11	4699021	AX-115327898

587954	11	4695971	AX-115327894	11	4703926	AX-115327900
587957	11	4696478	AX-115327895	11	4699021	AX-115327898
587958	11	4696478	AX-115327895	11	4703926	AX-115327900
587961	11	4696865	AX-115327896	11	4699021	AX-115327898
587962	11	4696865	AX-115327896	11	4703926	AX-115327900
587964	11	4698983	AX-115327897	11	4699021	AX-115327898
587965	11	4699021	AX-115327898	11	4703926	AX-115327900
587966	11	4699021	AX-115327898	11	4704815	AX-115327902
587973	11	4702497	AX-115327899	11	4703926	AX-115327900
587979	11	4703926	AX-115327900	11	4704815	AX-115327902
587980	11	4703926	AX-115327900	11	4705073	AX-115327903
587981	11	4703926	AX-115327900	11	4705729	AX-115327904
587982	11	4703926	AX-115327900	11	4706289	AX-115327905
587983	11	4703926	AX-115327900	11	4706554	AX-115327908

Cromosoma 13

	CHR A	BP A	SNP A	CHR B	BP B	SNP B
703272	13	5172049	AX-115374982	13	5186503	AX-115355048
703282	13	5177212	AX-115374984	13	5186503	AX-115355048
703289	13	5184638	AX-115355050	13	5186503	AX-115355048
703294	13	5186503	AX-115355048	13	5189494	AX-115355043
703295	13	5186503	AX-115355048	13	5191036	AX-115355041
703296	13	5186503	AX-115355048	13	5194751	AX-115355040

