

UNIVERSIDAD DE ALMERIA

ESCUELA SUPERIOR DE INGENIERÍA

**Estudio de asociación (GWAS) en
cucurbitáceas para caracteres de interés
agronómico. Selección de marcadores con
valor agronómico.**

Curso 2020/2021

Alumno/a:

López Ferrer, Alba

Director/es:

Cecilia Martínez Martínez
Manuel Jamilena Quesada



Agradecimientos:

Antes de comenzar, me gustaría agradecer a todas las personas que han hecho posible que hoy esté finalizando mi etapa en este máster.

Quisiera expresar mi agradecimiento a mis tutores, Manuel Jamilena y Cecilia Martínez, por despertar en mí el interés en el campo de la investigación y por la dirección durante la redacción y elaboración de este trabajo.

En particular, quiero darle las gracias a Cecilia por todas las horas que me ha dedicado, enseñándome a utilizar numerosos programas informáticos y por la paciencia que ha tenido durante la coordinación y seguimiento del proyecto. Gracias por confiar en mí.

Dedicatoria:

A mis padres, hermanos y abuelos que me inspiran e impulsan a seguir preparándome.

A mi pareja, por apoyarme, confiar en mí y enseñarme tanto estos últimos años. Gracias por tanto en tan poco.

UNIVERSIDAD DE ALMERÍA

ESCUELA SUPERIOR DE INGENIERÍA



MÁSTER EN INGENIERÍA AGRONÓMICA

**“ESTUDIO DE ASOCIACIÓN (GWAS) EN
CUCURBITÁCEAS PARA CARACTERES DE INTERÉS
AGRONÓMICO. SELECCIÓN DE MARCADORES CON
VALOR AGRONÓMICO.”**

Alumna:

López Ferrer, Alba

Directores:

Martínez Martínez, Cecilia

Jamilena Quesada, Manuel

Almería, julio de 2021



INDICE

1. INTERÉS Y OBJETIVOS	4
1.1. Importancia económica de las cucurbitáceas	5
1.2. Interés del material empleado y de los caracteres analizados	10
1.3. Objetivos	11
2. REVISIÓN BIBLIOGRÁFICA	12
2.1. Características de las cucurbitáceas	13
2.1.1. Características de <i>Cucurbita</i> spp.	13
2.1.1.1. Origen y domesticación	13
2.1.1.2. Clasificación taxonómica	14
2.1.1.3. Usos y valor nutricional	15
2.1.1.4. Descripción morfológica	16
2.1.1.5. Morfotipos de las especies cultivadas	18
2.1.2. Características de la especie <i>Citrullus</i> spp.	20
2.1.2.1. Origen y domesticación	20
2.1.2.2. Clasificación taxonómica	22
2.1.2.3. Usos y valor nutricional	22
2.1.2.4. Descripción morfológica	23
2.1.2.5. Morfotipos de las especies cultivadas	25
2.2. Variabilidad natural: conservación de recursos fitogenéticos	27
2.3. Análisis de la diversidad genética a escala genómica	28
2.3.1. Análisis de asociación del genoma (GWAS)	28
2.4. Caracteres de mejora en cucurbitáceas	30
3. MATERIAL Y MÉTODOS	33
3.1. Material vegetal	34
3.2. Datos genómicos	34
3.2.1. Genomas de referencia	34
3.2.2. Secuenciación de accesiones y obtención de variantes	35
3.3. Toma de datos y procedimientos utilizados	38
3.3.1. Caracterización de las diferentes accesiones de <i>Cucurbita</i> spp. y <i>Citrullus</i> spp.	38
3.3.2. Análisis de asociación del genoma (GWAS)	39
3.3.3. Análisis de variantes asociadas	41
3.5. Análisis estadístico	44
4. RESULTADOS Y DISCUSIÓN	45
4.1. Estudio de la variación de caracteres de semilla en los géneros <i>Citrullus</i> y <i>Cucurbita</i> (CUCURBITACEAE)	46
4.2. Estudio de asociación del genoma (GWAS) para caracteres de interés en semilla	54
4.2.1. Peso de la semilla	54
4.3. Estudio del impacto de las variantes y de regiones del genoma con asociación para los caracteres de la semilla en <i>Citrullus</i> y <i>Cucurbita</i>	60
4.3.1. Localización de las variantes seleccionadas e identificación de genes	60
4.3.2. Impacto en la proteína de los SNPs localizados en regiones exónicas	63
5. CONCLUSIONES	67
6. BIBLIOGRAFÍA	71

Índice de figuras

Figura 1. Superficie en miles de hectáreas y producción en millones de toneladas mundial entre 1999 y 2019 (Elaboración propia a partir de FAOSTAT, 2021).	5
Figura 2. Producción en millones de toneladas en el año 2018, de los principales países productores europeos de: A. Sandías B. Calabazas y calabacines (Elaboración propia a partir de FAOSTAT, 2021).	6
Figura 3. Principales países a nivel mundial: A. exportadores de sandía (miles de toneladas), B. exportadores de calabazas y calabacines (miles de toneladas). (Elaboración propia a partir de FAOSTAT, 2021).	7
Figura 4. Distribución de la producción y superficie de sandía y calabacín en Andalucía por provincias. A. Producción de sandía. B. Superficie de sandía. C. Producción de calabacín. D. Superficie de sandía. (Elaboración propia a partir de MAPA, 2021).	9
Figura 5. Valor nutricional y mineral de <i>C. pepo</i> . (Elaboración propia a partir de Moreiras et al., 2013 y Valero et al., 2018).	16
Figura 6. Partes de una planta de <i>Cucurbita</i> : A. Hojas. B. Flor masculina. C. Frutos. D. Semillas. E. Flor femenina. (Elaboración propia).	18
Figura 7. Diferentes morfotipos para las subespecies <i>pepo</i> y <i>texana</i> comestibles. (Elaboración propia a partir de López, 2019).	19
Figura 8. Valor nutricional y mineral de <i>C. lanatus</i> . (Elaboración propia a partir de Moreiras et al., 2013 y Valero et al., 2018).	23
Figura 9. Partes de una planta de <i>Citrullus</i> : A. Parte externa del fruto. B. Parte interna del fruto. C. Semillas. D. Hoja pinado-partida. E. Flor femenina. F. Flor masculina. G. Zarcillos. (Elaboración propia).	25
Figura 10. Especies cultivadas del género <i>Citrullus</i> . (Elaboración propia a partir de Renner et al., 2019; 2017).	26
Figura 11. Partes del fruto de Charleston Gray. (Esquema de elaboración propia, imagen (Seed gallery, 2021)).	35
Figura 12. Distribución geográfica de 1.365 accesiones de <i>Citrullus</i> spp. del USDA. El diámetro del círculo es proporcional al número de accesiones de cada país (Wu et al., 2019).	37
Figura 13. Representación de gráficos QQ que se pueden encontrar en el análisis GWAS con los métodos GLM y MLM y que permitieron elegir entre estos dos métodos de análisis. Para ellos se eligió el método con mejor ajuste en el gráfico QQ: A. Datos muy bien ajustados. B. Sobreestimación del impacto de los SNPs sobre el carácter. C. Subestimación del impacto de las variantes sobre el carácter.	40
Figura 14. Regiones genómicas de los genomas de referencia de Cucurbitgenomic database (Elaboración propia a partir de Cucurbitgenomic database, 2021).	41
Figura 15. Esquema representativo del procedimiento seguido en este TFM para la realización de estudios de asociación en caracteres de semilla en cucurbitáceas. Se indica los pasos dados y los programas empleados para realizar el estudio de asociación, determinar la posición de las variantes con significación, establecer que alelos aparecen y valorar su impacto, en el caso de tener posición exónica, en la proteína. (Elaboración propia).	43
Figura 16. Cada caja de bigotes representa la mediana y los cuartiles de (A) peso de semilla. (B) Color de las semillas.	47

Figura 17. Color de testa de semilla de las diferentes accesiones. A. <i>C. lanatus</i> . B. <i>C. amarus</i> . C. <i>C. maxima</i>	48
Figura 18. Presencia-ausencia de hilio en las diferentes accesiones de <i>Citrullus</i> spp. en: A. <i>Citrullus</i> spp. B. <i>C. lanatus</i> . C. <i>C. mucosospermus</i>	49
Figura 19. Porcentaje de accesiones con presencia y ausencia de hilio. A. Peso semilla en gramos. B. Color de testa de semilla.	50
Figura 20. Porcentaje de accesiones que tienen hilio. A. Distribución mundial de semillas con hilio. B. Distribución mundial de los colores de semilla predominantes observados en <i>Citrullus</i> spp.	51
Figura 21. Gráficos Manhattan de estudio de asociación para el peso de semilla. A. <i>Citrullus</i> spp. B. <i>C. maxima</i> . C. <i>C. pepo</i>	55
Figura 22. Gráficos Manhattan obtenidos en los estudios de asociación del color de semilla. A. Resultados GWAS para <i>Citrullus</i> spp. B. Resultados GWAS para <i>C. maxima</i>	58
Figura 23. Gráfico Manhattan con la asociación de SNPs a la presencia y ausencia de hilio en <i>Citrullus</i> spp.	59

Índice de tablas

Tabla 1. Producción en toneladas en el año 2018, de los principales países productores de sandías y Calabazas y calabacines (Elaboración propia a partir de FAOSTAT, 2021).....	6
Tabla 2. Superficie (hectáreas) y producción (toneladas) de sandía y calabacín para 2018. Se muestra su distribución por comunidades. Elaboración propia a partir de MAPA, 2021.....	8
Tabla 3. Taxonomía de la sandía	14
Tabla 4. Taxonomía de la sandía	22
Tabla 5. Variantes obtenidas del llamado de variantes entre las accesiones secuenciadas por GBS (Proyecto CucCAP).....	37
Tabla 6. Distribución de las accesiones genotipadas por GBS de CuGenDB (Proyecto CucCAP).	38
Tabla 7. Se presentan los valores fenotípicos observados para los caracteres color y presencia-ausencia de hilio.....	39
Tabla 8. Resumen de las variantes destacadas como relevantes en este trabajo para los tres caracteres y análisis realizados. Se resaltan en amarillo, de cada análisis, las variantes significativas. Se indica el carácter, especies implicadas en el análisis, p-valor, identificador del SNP (SNP ID), localización, posición en la región (intergénica, exónica o intrónica), gen en el que se localizan en caso de estar dentro de la secuencia de un gen anotado (Gen ID) y descripción del gen.	62
Tabla 9. Impacto de las variantes exónicas sobre la proteína. Se indica cuando la variante causa una sustitución de aminoácido y si es así el impacto estimado en la proteína. En amarillo se muestran las dos variantes significativas en el análisis GWAS.	64

INTERÉS Y OBJETIVOS



1. INTERÉS Y OBJETIVOS

1.1. Importancia económica de las cucurbitáceas

Aunque se asume que su origen está en Asia, la familia de las cucurbitáceas está ampliamente diversificada y se encuentra distribuida en todo el mundo (Esteras et al., 2012). Esta familia incluye importantes especies cultivadas como *Citrullus lanatus* (sandía), *Cucumis melo* (melón), *Cucumis sativus* (pepino) y *Cucurbita pepo*, *C. maxima*, *C. moschata* y *C. argyrosperma* (calabazas, calabacines) y otras especies de importancia local (por ejemplo: *C. amarus*, *C. mucospermus* y *C. colocynthis*) y taxones silvestres. Las especies más importantes se consumen por sus frutos, y distintas especies de la familia se consideran además valiosas por sus usos medicinales u ornamentales (Carnide y Barroso, 2006; Simpson, 2018; Martínez et al., 2021).

De acuerdo con FAOSTAT (2021), desde 1994, la superficie y producción mundial de sandía experimentó un incremento (Figura 1). Mientras que en 1994 la superficie era de 2.214.549 ha y la producción de 38.396.781 t, en 2003 se incrementó hasta 3.444.890 ha y 88.569.263 t. Lo que representa un incremento del 36% de la superficie y 57% de la producción. En los últimos años, la superficie de sandía ha ido disminuyendo ligeramente hasta situarse en 3.120.446 ha, mientras que la producción ha seguido incrementándose hasta 100.458.926 t en 2018, suponiendo una pérdida del 9% de la superficie y un aumento del 22% de la producción respecto al año 2003.

Según la misma fuente (FAOSTAT, 2021), la superficie y producción mundial de calabazas y calabacines, que incluye especies *C. pepo*, *C. maxima* y *C. moschata*, también se incrementó (Figura 1) de 1.086.272 ha y 13.123.803 t, en 1994, a 2.070.832 ha y 27.678.883 t, en 2017. Esto supone un incremento del 48% y 53%, respectivamente. Sin embargo, para el año 2018 tanto la superficie como la producción disminuyeron un 26% y 18% respecto al año 2017.

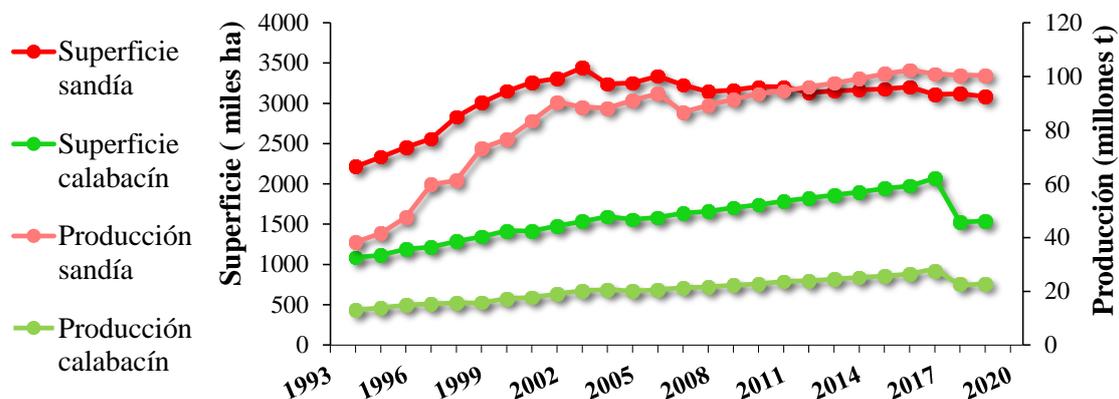


Figura 1. Superficie en miles de hectáreas y producción en millones de toneladas mundial entre 1999 y 2019 (Elaboración propia a partir de FAOSTAT, 2021).

Actualmente Asia es el principal continente productor de sandías, con más del 80% de la producción mundial. Sobresale China con una producción de 61.536.900 t, mientras que España ocupa el quinceavo lugar con 1.092.400 t (Tabla 1). En cuanto a la producción de calabazas y calabacines China lidera la producción mundial con 8.198.544 t, seguido de Ucrania y Rusia con 1.338.000 t y 1.191.538 t, respectivamente (Tabla 1). España ocupa el quinto lugar a nivel mundial con 717.650 t, lo que evidencia la gran importancia de la producción del cultivo en nuestro país.

Tabla 1. Producción en toneladas en el año 2018, de los principales países productores de sandías y Calabazas y calabacines (Elaboración propia a partir de FAOSTAT, 2021).

País	Producción sandía (t)	País	Producción de calabazas y calabacines (t)
China, Continental	61.536.900	China, Continental	8.198.544
Turquía	4.031.174	Ucrania	1.338.000
India	2.520.000	Federación de Rusia	1.191.538
Brasil	2.244.001	México	776.073
Argelia	2.095.757	España	717.650
Federación de Rusia	1.969.954	Estados Unidos	698.850
Uzbekistán	1.836.959	Turquía	616.777
Estados Unidos	1.771.051	Bangladesh	600.141
México	1.472.459	Italia	548.440
Irán	1.459.698	Malawi	480.233
Egipto	1.431.919	Indonesia	454.001
Kazajstán	1.248.613	Cuba	452.264
Vietnam	1.200.104	Egipto	430.503
Senegal	1.172.772	Argelia	388.875
España	1.092.400	República de Corea	310.218

Del continente europeo, Rusia y España lideran la producción de sandía con 1.969.954 t y 1.092.400 t, respectivamente (Figura 2A). Por su parte, los principales productores de calabazas y calabacines son Ucrania y Rusia, seguidos de España con 717.650 t e Italia con 548.440 t (Figura 2B) (FAOSTAT, 2021).

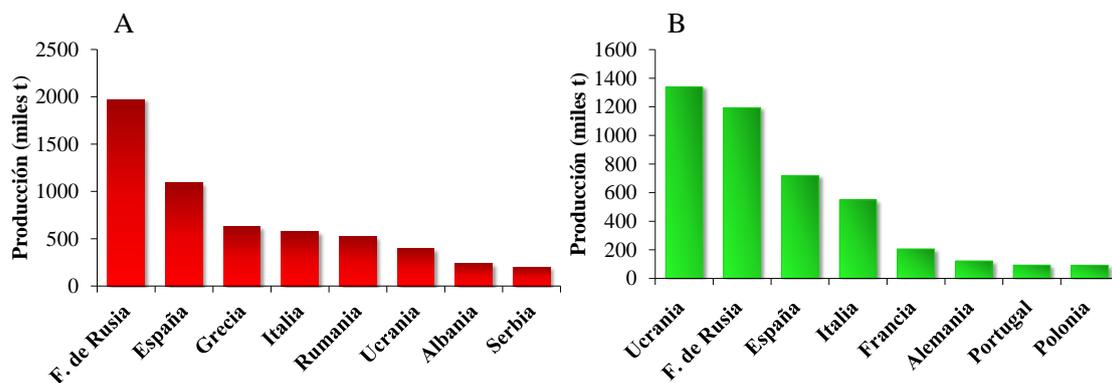


Figura 2. Producción en millones de toneladas en el año 2018, de los principales países productores europeos de: **A.** Sandías **B.** Calabazas y calabacines (Elaboración propia a partir de FAOSTAT, 2021).

En cuanto a las exportaciones mundiales de estas cucurbitáceas, España y México se situaron en primer lugar con 911.316 t para sandía (Figura 3A) y 538.038 t para calabacín (Figura 3B), respectivamente. Siendo los principales importadores de sandía Estados Unidos con 723.346 t y Alemania con 509.877 t. Mientras que Estados Unidos y Francia son los principales importadores de calabazas y calabacines con 508.216t y 156.307 t, respectivamente (FAOSTAT, 2021).

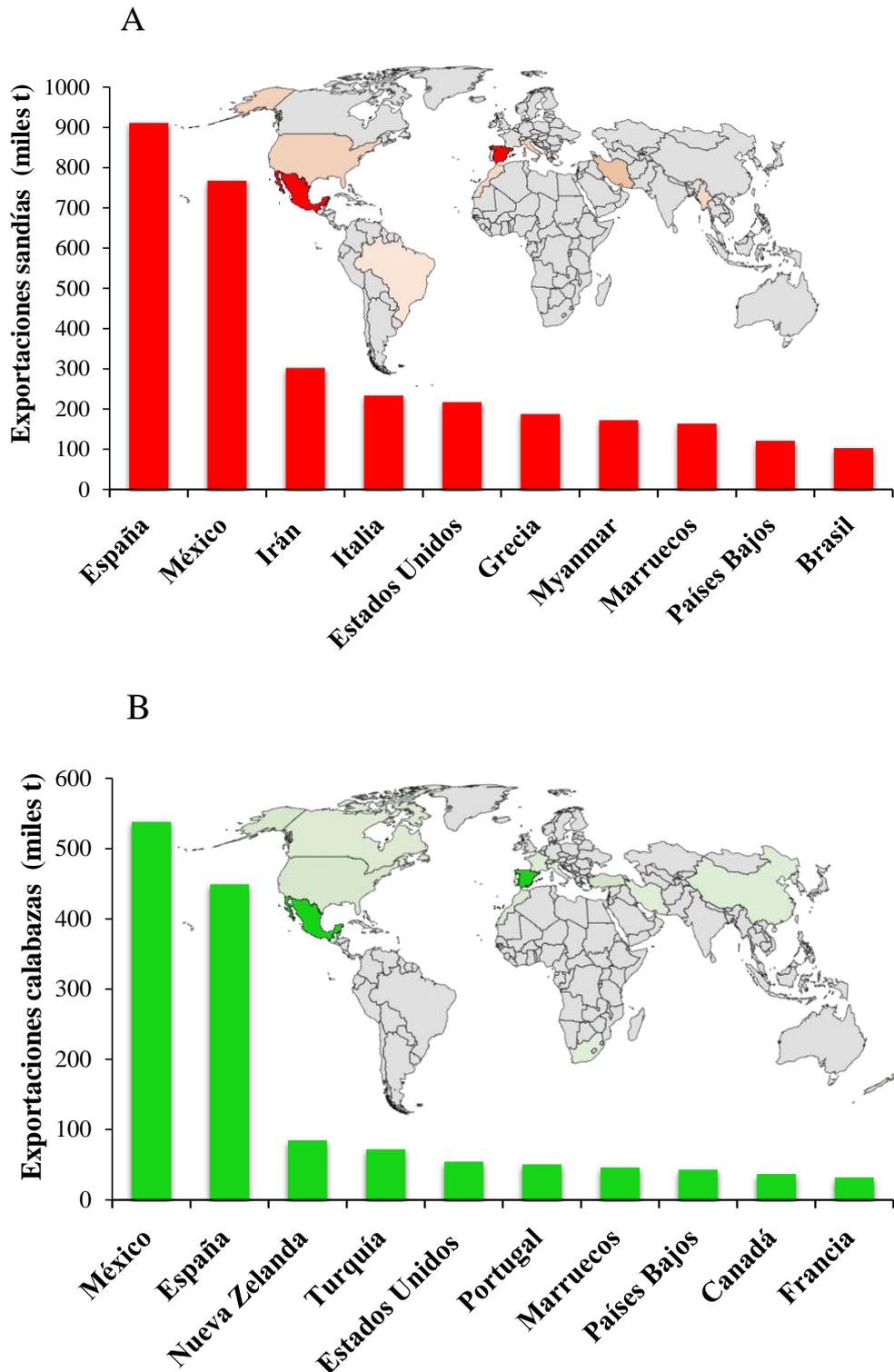


Figura 3. Principales países a nivel mundial: **A.** exportadores de sandía (miles de toneladas), **B.** exportadores de calabazas y calabacines (miles de toneladas). (Elaboración propia a partir de FAOSTAT, 2021).

Por lo tanto, desde el punto de vista económico el calabacín y la sandía son dos cultivos de mayor importancia en España, siendo uno de los productores y exportador principales a nivel mundial. En el año 2018 (Tabla 1), se cultivaron en España unas 21.460 hectáreas y se produjeron más de un millón de toneladas de sandía, Andalucía

lidera tanto en superficie cultivada (56% del total) como en producción (59% del total) (Tabla 2) (MAPA, 2021). De forma similar, Andalucía destaca en la producción de calabacín en invernadero con una superficie de 8168 ha y una producción de 471677 t (Tabla 2) (MAPA, 2021).

Tabla 2. Superficie (hectáreas) y producción (toneladas) de sandía y calabacín para 2018. Se muestra su distribución por comunidades. Elaboración propia a partir de MAPA, 2021.

Comunidad Autónoma	Sandía				Calabacín			
	Invernadero		Aire libre		Invernadero		Aire libre	
	Sup. (ha)	Prod. (t)	Sup. (ha)	Prod. (t)	Sup. (ha)	Prod. (t)	Sup. (ha)	Prod. (t)
Andalucía	7904	403981	4240	192602	8168	471677	630	22666
Aragón	-	-	11	638	-	-	52	1120
Baleares	60	2560	230	6992	49	2040	105	2153
Canarias	164	7135	129	3882	323	21555	217	7684
Cataluña	-	-	264	7800	-	-	260	8146
C. la Mancha	-	-	2426	177680	1	40	34	980
Castilla y León	-	-	42	1383	13	455	39	948
C. Valenciana	-	-	1484	73088	219	11226	124	3573
Extremadura	157	8126	53	2434	1	75	118	5677
Galicia	-	-	-	-	93	5893	-	-
La Rioja	-	-	3	63	3	180	25	1150
Madrid	-	-	41	1517	-	-	18	558
Navarra	-	-	-	-	1	95	122	8741
País Vasco	-	-	-	-	2	105	16	419
P. de Asturias	-	-	-	-	13	585	23	575
R. de Murcia	89	5242	2803	193687	235	13748	111	3164

El principal productor en invernadero de la comunidad andaluza es Almería. Actualmente el campo almeriense destaca por la agricultura intensiva bajo plástico y, más concretamente, por la producción de gran parte de las hortalizas a nivel nacional, destacando entre ellas el tomate (*Solanum lycopersicum* L.), la berenjena (*Solanum melongena* L.), el calabacín (*Cucurbita pepo* L.), el pepino (*Cucumis sativus* L.), el pimiento (*Capsicum annuum* L.), la judía verde (*Phaseolus vulgaris* L.), la sandía (*Citrullus lanatus* L.) y el melón (*Cucumis melo* L.). Aproximadamente cuenta con 32.364 hectáreas de cultivos protegidos (Gallegos, 2021). De ellas 9.860 ha se dedican a la producción de sandía y 7.860 ha a calabacín, siendo el tercer y cuarto cultivo más importante dentro de la provincia, representado el 81% (Figura 4C) y 89% (Figura 4D) de la superficie cultivada nacional, respectivamente. Con estas superficies se consigue cosechar 512.742 toneladas de sandía y 456.045 toneladas de calabacín, lo que supone el 86% (Figura 4A) y 92% (Figura 4B) de la producción española (MAPA, 2021).

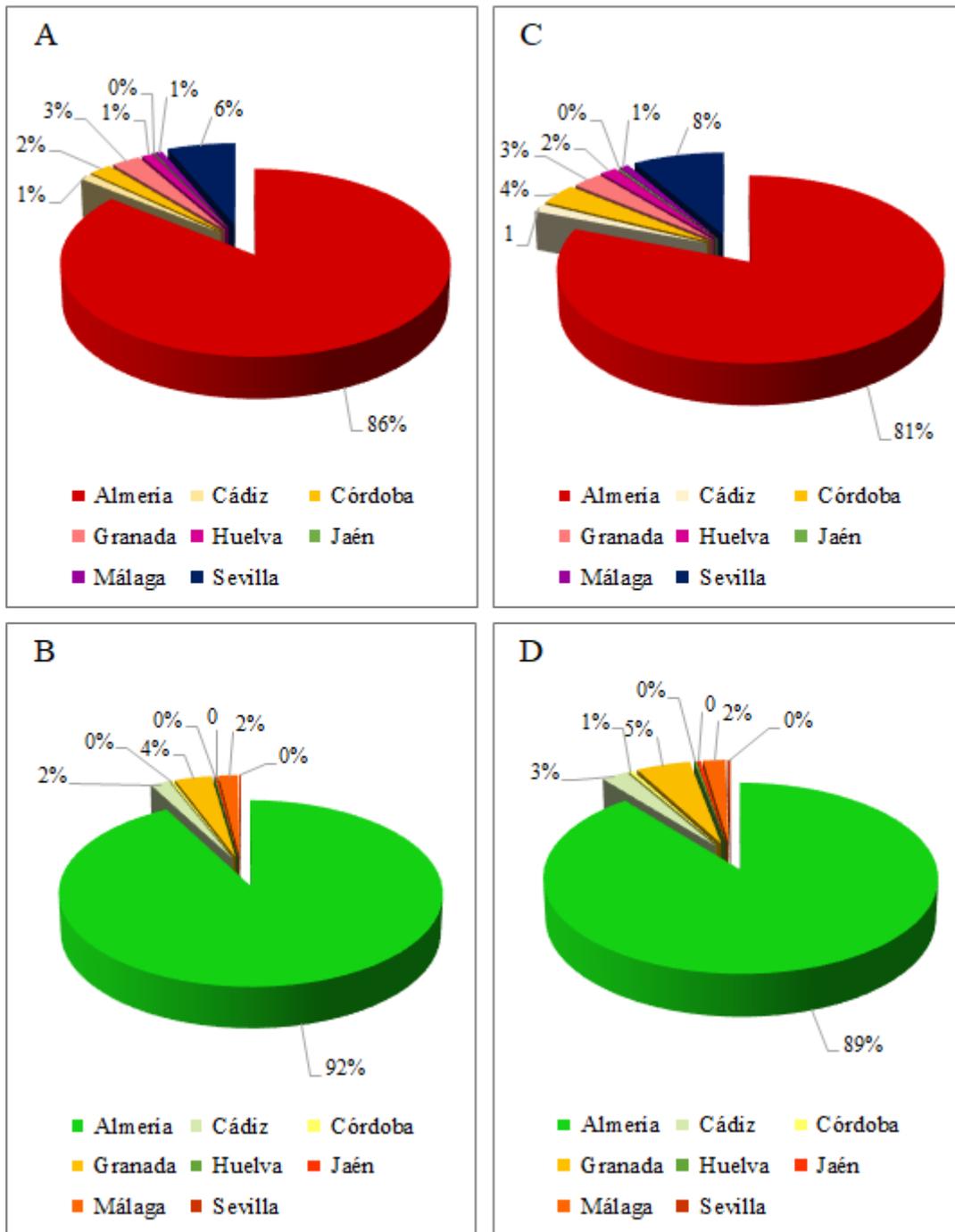


Figura 4. Distribución de la producción y superficie de sandía y calabacín en Andalucía por provincias. **A.** Producción de sandía. **B.** Superficie de sandía. **C.** Producción de calabacín. **D.** Superficie de sandía. (Elaboración propia a partir de MAPA, 2021).

Por su parte el valor de la producción en la provincia ha crecido un 5,2% y ha contabilizado 2.319,5 millones de €, de los que 2.291,6 millones de € provienen de la horticultura intensiva: en este caso, la sandía protagoniza la mayor subida de ingresos (+32,3%) y el calabacín (+5,8%) (CAJAMAR, 2020).

1.2. Interés del material empleado y de los caracteres analizados

Las cucurbitáceas son generalmente apreciadas por sus frutos, pero también por las semillas que son una buena fuente de aceite vegetal y proteínas. Por ejemplo, las semillas de sandía (Baboli y Kordi, 2010). El tamaño y el peso de la semilla de las cucurbitáceas cultivadas varían considerablemente según los usos previstos, en el caso de la sandía y de la calabaza las semillas tienen usos duales y el tamaño de la semilla es un objetivo importante para la mejora genética (Guo et al., 2020). Para el desarrollo de la planta, una semilla de calidad contribuye a mayor eficiencia de la producción, ya que es capaz de emerger de manera rápida y uniforme, bajo diferentes condiciones ambientales (Crang et al., 2018). Para los usos que implican la extracción de compuestos de valor de la semilla, tanto el contenido en aceites o antioxidantes, como el tamaño pueden ser de interés para la mejora. Por lo que la calidad de la semilla es un concepto basado en la valoración de diferentes atributos.

La importancia socioeconómica de las cucurbitáceas se traduce en relevancia desde el punto de vista de la mejora genética, por lo que se han conservado gran número de accesiones del germoplasma silvestre, asilvestrado, de razas autóctonas, cultivares antiguos e híbridos modernos en bancos de germoplasma de todo el mundo (Ferriol y Pico, 2008; Levi et al., 2017a). Estas accesiones, aunque menos productivas, favorecen el mantenimiento de la variabilidad genética de cualquier especie y constituyen una fuente de caracteres altamente valiosos para la mejora genética (Gutiérrez-Díez et al., 2009; García et al., 2018). De hecho con frecuencia son incorporadas a los programas de mejora para generar poblaciones variables sobre las que realizar selección para diferentes caracteres. La caracterización de estas accesiones de la variabilidad natural es fundamental para evaluar su interés para los programas de mejora. En este marco se sitúa este TFM que pretende la puesta en valor de estas accesiones desde el punto de vista de la semilla.

Una de las mejores colecciones de cucurbitáceas se conserva en el banco de germoplasma del Departamento de Agricultura de Estados Unidos (USDA), con 1.766 accesiones de *Cucurbita* spp. y 1.453 accesiones de *Citrullus* spp (<https://npgsweb.ars-grin.gov/gringlobal/search>). El trabajo de conservación realizado se coordina con el trabajo de re-secuenciación de numerosas accesiones de los dos géneros realizados en el marco del *Cucurbit Coordinated Agricultural Project* (CucCAP). Este proyecto ha derivado en la re-secuenciación por Genotyping by Sequencing (GBS) de 1.516 accesiones de *Cucurbita* spp, de las cuales 830 pertenecen a *Cucurbita pepo* y 372 a *C. maxima* (datos no publicados, variantes disponibles en <http://cucurbitgenomics.org/>), y de 1.365 del género *Citrullus* (Wu et al., 2019). Esta estrategia ha permitido la detección de numerosas variantes de utilidad para los estudios de asociación y con potencial valor para la selección de caracteres valiosos en programas de mejora genética de estas especies.

1.3. Objetivos

El principal objetivo del presente Trabajo Fin de Máster (TFM) es localizar posibles regiones candidatas con asociación a características físicas de calidad de las semillas como el peso, el color o la presencia de hilio. Empleando para ello una caracterización exhaustiva realizada en 1.518 accesiones de los géneros *Cucurbita* y *Citrullus* y los datos genómicos disponibles en CuGenDB (<http://cucurbitgenomics.org/>).

Este objetivo general se aborda desde los siguientes objetivos específicos:

- Realizar una caracterización de al menos 6 semillas por accesión en 805 accesiones de *Cucurbita* spp. y 713 accesiones de *Citrullus* spp. Los caracteres de interés serán: tamaño, peso y color de las semillas.
- Empleando el programa Tassel5 (<https://www.maizegenetics.net/tassel>), contrastar los datos fenotípicos de las 1.518 accesiones con los datos genómicos para realizar estudios de asociación (GWAS). Por medio de los métodos *General Lineal Model* (GLM) y *Mixed Lineal Model* (MLM) se designarán variantes con potencial asociación con el carácter en estudio.
- En caso de detectar asociación, comparar los resultados en los distintos análisis con estudios previos con el fin de determinar la relevancia de las distintas regiones detectadas.
- Empleando las anotaciones del genoma disponibles para *C. pepo*, *C. maxima* y *C. lanatus* comprobar el impacto de las variantes exónicas en la proteína.

REVISIÓN BIBLIOGRÁFICA



2. REVISIÓN BIBLIOGRÁFICA

2.1. Características de las cucurbitáceas

La familia de las Cucurbitáceas constituye uno de los grupos taxonómicos más grandes y variables de plantas, podemos encontrar hasta 825 especies, agrupadas en 118 géneros (Zomlefer, 2004). Otros autores indican que hay 130 géneros con 800 especies (Jeffrey, 2005; Jeffrey and De Wilde, 2006). Los géneros más grandes son *Cayaponia* con 60 especies, *Momordica* con 45 y *Gurania*, *Syclos* y *Trichosanthes* con 40 especies cada uno (Jeffrey, 1990). No obstante, destacan por su valor económico los géneros: *Citrullus*, *Cucurbita* y *Cucumis*, que incluyen especies como la sandía (*Citrullus lanatus*), el calabacín (*Cucurbita pepo*), los diferentes tipos de calabaza (*Cucurbita pepo*, *Cucurbita moschata*, *Cucurbita maxima*, *Cucurbita argyrosperma*), el pepino (*Cucumis sativus*) y el melón (*Cucumis melo*).

En base a estudios morfológicos de sus órganos vegetativos y florales, a lo largo de la historia, la clasificación de las cucurbitáceas ha sufrido modificaciones (Jones, 1989; Takhtajan, 1997). Con base en estudios genéticos de ADN, en 2003 el grupo APG II las clasificó en el orden Cucurbitales, el cual contiene, además de la familia *Cucurbitaceae*, las familias *Anisophylleaceae*, *Begoniaceae*, *Coriaceae*, *Corynocarpaceae*, *Datisceae* y *Tetramelaceae* (APG II, 2003).

Los caracteres de diagnóstico más importantes para los géneros y tribus de las cucurbitáceas provienen de la morfología del androecio y del gineceo (número de estilos libres, fusión de filamentos y/o anteras), tipo de zarcillo, estructura del polen y cubierta de la semilla. La familia contiene dos subfamilias, *Nhandioboideae* o *Zanoniodeae* (Jeffrey, 1990), con 19 géneros y unas 60 especies sin valor económico, y *Cucurbitoideae*, con 111 géneros and 740 especies (Esteras et al., 2012). Esta subfamilia incluye las tribus y géneros más importantes de las cucurbitáceas: *Benincaseae* (*Benincasa*, *Citrullus*, *Coccinia*, *Lagenaria*, *Cucumeropsis*, *Cucumis*), *Luffeae* (*Luffa*), *Cucurbiteae* (*Cucurbita*), *Sicyeae* (*Cyclanthera*, *Sechium*), *Joliffieae* (*Momordica*, *Telfairia*), *Schizopeponeae* (*Schizopepon*), *Trichosantheae* (*Hodgsonia*, *Trichosanthes*), *Bryonieae* (*Ecballium*), *Herpetospermae* y *Coniandreae* (Jeffrey, 1990; Rubatzky y Yamaguchi, 2012; Jeffrey, 2005). La subfamilia *Cucurbitoideae* se caracteriza por tener estilos fusionados, polen con exina reticulada y zarcillos simples, bífidos o con la parte basal no espiralada. Estos caracteres se correlacionan bastante bien con la filogenia obtenida con los datos moleculares. De todas ellas son objeto de interés en este TFM la tribu *Benincaseae* que incluye el género *Citrullus* y la tribu *Cucurbiteae* que incluye el género *Cucurbita*.

2.1.1. Características de *Cucurbita* spp.

2.1.1.1. Origen y domesticación

En el género *Cucurbita* hay veintidós especies salvajes y cinco especies cultivadas (Decker, 1988). Las especies cultivadas son *C. pepo*, *C. moschata*, *C. maxima*, *C. ficifolia* y *C. argyrosperma* (Nee, 1990), siendo las tres primeras *C. pepo*, *C. moschata* y *C. maxima*, las más importantes económicamente y las más ampliamente distribuidas.

Witaker y Cutler (1971) informaron del hallazgo de semillas y pedúnculos pertenecientes a frutos cultivados de *C. pepo* subsp. *pepo* en el Valle de Oaxaca con una edad superior a 9000 años y en las Cuevas Ocampo en Tamaulipas, México, databan del 7000 a 5000 a.C. Aunque los registros arqueológicos no han revelado el origen de las

Cucurbitáceas cultivadas, proporcionan información sobre la dispersión geográfica y la diversificación de las especies domesticadas. (Decker, 1988).

Ford (1980), basándose en la distribución arqueológica de la especie, propuso que esta fue domesticada al menos dos veces, en el 5000 a.C. y en el 2000 a.C. La domesticación se produjo simultáneamente en la zona de Mesoamérica (México) y Norte América (EEUU-Canadá), teniendo a *C. fraterna* y *C. texana* como antepasados silvestres, respectivamente. Esta idea se vio reforzada por hallazgos arqueológicos que ponen de manifiesto la domesticación de *C. pepo* desde hace más de 4000 años en el sureste y noroeste de México y en el este de Estados Unidos (Decker, 1988; Paris, 2001; Ferriol et al., 2008).

Decker (1988) informó que las primeras evidencias de los morfotipos de calabaza redonda y calabacín de *Cucurbita pepo* tienen sus raíces en variedades de la zona de México. La introducción en Europa del calabacín data del siglo XVI (Decker, 1988). Por otra parte, el calabacín que comemos hoy día se corresponde con el morfotipo Zucchini en estado inmaduro. La palabra Zucchini proviene del diminutivo en plural de la voz italiana “zucca” que significa calabaza de verano. París (2001) afirmó que es el morfotipo más reciente de *C. pepo*, ya que se diversificó en Italia más tarde que las otras variedades y posteriormente se introdujo en EEUU desde Italia en el siglo XX (París, 2001).

Cucurbita maxima Duch. Ex Lam. es una de las principales especies cultivadas del género *Cucurbita*. Tiene su origen en América del Sur, y su centro de diversidad dentro del rango de distribución de su ancestro silvestre, *C. maxima* ssp. *andreana* (Naud.) Filov que se encuentra distribuidas en zonas templadas de Argentina, en las llanuras centrales de Bolivia, en Uruguay y Paraguay (Nee, 1990).

2.1.1.2. Clasificación taxonómica

De acuerdo con el *Integrated Taxonomic Information System* (ITIS, 2021) la taxonomía de la especie *C. pepo* y *C. maxima* es de la siguiente manera:

Tabla 3. Taxonomía de la sandía

Reino:	<i>Plantae</i>
Subreino:	<i>Viridiplantae</i>
Infrareino:	<i>Streptophyta</i>
Superdivisión:	<i>Embryophyta</i>
División:	<i>Tracheophyta</i>
Subdivisión:	<i>Spermatophytina</i>
Clase:	<i>Magnoliopsida</i>
Superorden:	<i>Rosanae</i>
Orden:	Cucurbitales
Familia:	<i>Cucurbitaceae</i>
Género:	<i>Cucurbita</i> L.
	<i>Cucurbita pepo</i> L.
Especies:	<i>Cucurbita maxima</i> Duchesne

2.1.1.3. Usos y valor nutricional

El uso del calabacín, *C. pepo* subsp. *pepo* Zucchini, está destinado a la alimentación humana exclusivamente. El consumo del fruto inmaduro del calabacín es muy popular en todo el mundo y especialmente en los países desarrollados (Paris, 2008). En España es muy común comerlo rebozado, frito o cocinado de otras formas. También se emplea en potajes y guisos en Mallorca, en el zarangollo murciano, en la escudella catalana o en los potajes y pucheros canarios. Aunque mucho menos frecuente, también se ha consumido crudo en ensaladas, como se ha registrado en la comarca almeriense del Cabo de Gata-Níjar. Por otra parte, las flores se han consumido fritas en Asturias, Cataluña y Mallorca en España, así como en otras partes del mundo. También se han preparado rellenas de arroz o harina en Mallorca o rellenas de queso y rebozadas con huevo y pan rallado en Cataluña, donde se denomina a este plato *carn de pobre* (carne de pobre). La flor masculina se ha usado para elaborar tortillas en varias comarcas catalanas y potajes en Gran Canaria. También, se han empleado como condimento de otros platos en El Hierro y en Mallorca (Tardío et al., 2018).

C. maxima constituye, dentro del género *Cucurbita*, una de las especies cultivadas más diversas con numerosas variedades locales, cultivares comestibles y ornamentales. El uso más común es el consumo de su pulpa madura, que es una de las de mejor calidad dentro del género por su sabor y riqueza en carotenoides (Itle y Kabelka, 2009; Della Gaspera y Rodríguez, 2013). En Argentina, algunos cultivares de frutos planos con hábito de crecimiento arbustivo se cultivan para el consumo de sus frutos jóvenes, conocidos localmente como zapallitos (Paris, 2017). El zapallito, es un grupo de *C. máxima* que es utilizado por sus semillas ya que sirven para contrarrestar los vómitos y mareos en las embarazas, como laxante natural y efecto antiinflamatorio. Además, se consume culinariamente para la preparación de sopas, cremas, estofados, dulces cuando está maduro y preparaciones saladas cuando esta inmaduro. (Grumet et al., 2017).

Las calabazas maduras de *C. pepo* y *C. maxima* tienen una considerable importancia monetaria, como ornamentales, en Estados Unidos, Europa y Australia. Las calabazas maduras de *C. pepo* con semilla, para la extracción de aceite, son económicamente importantes en Europa central. La producción de calabazas y calabazas ornamentales, especialmente de *C. pepo* y *C. maxima* para las actividades de Halloween y Oktoberfest, es una industria grande y creciente (Paris, 2017).

En cuanto a su valor nutricional, son una rica fuente natural de proteínas, triterpenos, lignanos, fitoesteroles, ácidos grasos poliinsaturados, compuestos fenólicos antioxidantes, carotenoides, tocoferol y minerales. Valores aproximados de vitaminas y minerales de *C. pepo* se muestran en la Figura 5. Valores similares presenta *C. máxima*, superando a *C. pepo* en calcio (29 mg), magnesio (10 mg), fósforo (19 mg), tiamina (0,16 mg) y vitamina A (75 µg) (Moreiras et al., 2013; Valero et al., 2018).

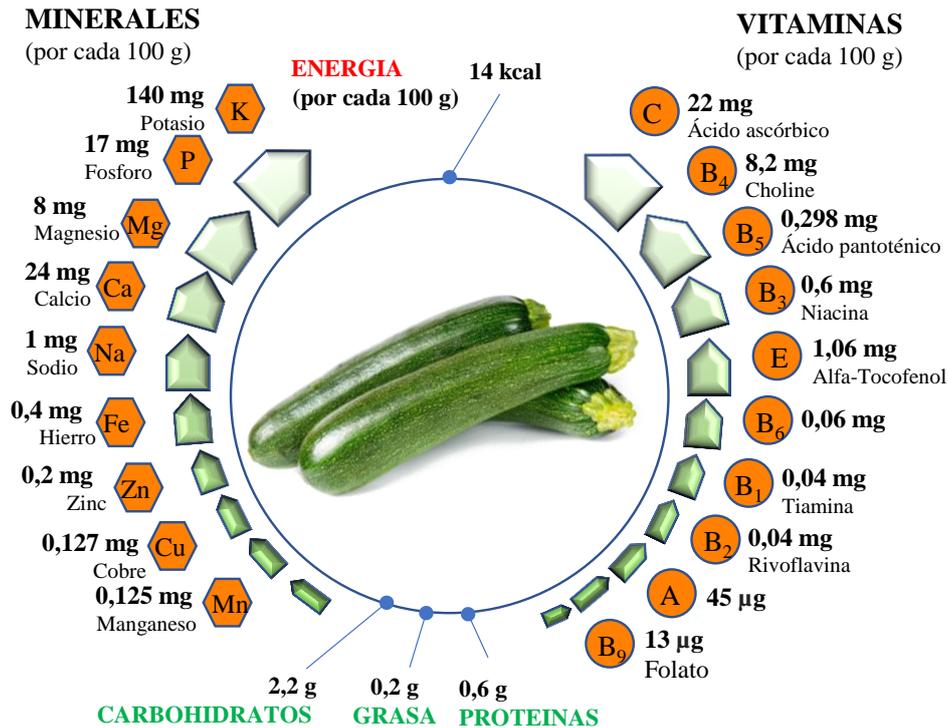


Figura 5. Valor nutricional y mineral de *C. pepo*. (Elaboración propia a partir de Moreiras et al., 2013 y Valero et al., 2018).

2.1.1.4. Descripción morfológica

Dentro de la subespecie *C. pepo* subsp. *pepo*, el morfotipo Zucchini se caracteriza por sus hojas grandes y palmeadas, sus flores amarillo-anaranjadas productoras de néctar y por sus frutos grandes, duros, esféricos e indehiscentes. No obstante este morfotipo varía con respecto al resto principalmente en la forma del fruto y color, pero también en otros caracteres vegetativos. Sus plantas son herbáceas, anuales, monoicas y arbustivas (Camacho, 2003; Martínez, 2003; Vicente, 2014). A continuación, se definen las características más relevantes de los distintos órganos de las plantas de esta especie.

El sistema radicular del calabacín está constituido por una raíz principal axonomorfa que alcanza un gran desarrollo en comparación con las raíces secundarias las cuales se extienden superficialmente. En terrenos enarenados el desarrollo radicular puede alcanzar 25-30 cm de profundidad y en terrenos desnudos el desarrollo radicular alcanza entre 50-80 cm. Cuando el tallo está en contacto con la tierra húmeda se pueden desarrollar raíces adventicias en los entrenudos de este (Martínez, 2003; Reche, 2000).

El calabacín posee un tallo principal herbáceo, grueso y cilíndrico con atrofia de brotes secundarios según el morfotipo. Los entrenudos suelen ser cortos (2-5 cm: aunque hay variedades que tienen entrenudos de hasta 30-40 cm) de donde parten las flores, hojas, frutos y zarcillos delgados que nacen junto al pedúnculo del fruto. El tamaño varía en función de las variedades, pero normalmente en la zona de Almería se cultivan variedades de morfotipo Zucchini y de porte arbustivo debido a su facilidad en el manejo (Martínez, 2003; Reche, 2000). En el caso de *C. maxima* los tallos son engrosados, redondeados o ligeramente angulosos (Lira et al., 2009), mientras que los zarcillos suelen tener de 3 a 5 ramificaciones (Lira, 1995). La principal diferenciación respecto a *C. pepo* se encuentra en la forma del fruto (Figura 6C).

Las hojas son palmeadas, con un gran limbo (llegan hasta los 50 cm de ancho y largo) y se sujetan al tallo por alargados peciolo alternándose de forma helicoidal. Es característica la presencia de vellosidades y espinas cortas y finas distribuidas a lo largo del mismo, además están huecos por dentro. El haz de las hojas es suave al tacto y el envés áspero, recubierto de espinas cortas a lo largo de las nerviaciones (Martínez, 2003; Delgado, 1999; Reche, 2000). *C. maxima* presenta hojas ligeramente o nada lobuladas, con márgenes enteros a denticulados y ambas superficies pubescentes (Villanueva, 2007) (Figura 6A).

Las flores del calabacín parten de la axila de la hoja, siendo solitarias, de un color amarillo-anaranjado vistoso, de gran tamaño y forma acampanada (Figura 6B y E). Los lóbulos son agudos a acuminados, obtusos y márgenes ondulados. Las plantas de calabacín son monoicas por lo que presenta tanto flores femeninas como masculinas en el mismo pie de planta. La diferenciación de las flores masculinas y femeninas es relativamente fácil, ya que esta última posee un ovario ínfero. Además, la flor masculina tiene un pedúnculo largo del que carece la flor femenina. La apertura de las flores tiene lugar a primera hora de la mañana y el cierre se produce a medio día (la flor está abierta aproximadamente de 6:00 am a 12:00 pm), momento en el que se aprovecha para la realización de la polinización ya sea entomófila o cruzada (Martínez, 2003; Delgado, 1999; Reche, 2000). En *C. maxima* las flores son aromáticas, de color amarillo brillantes, lóbulos anchos, reflexos y bordes crispados (Lira, 1995; Lira et al., 2009).

El fruto es una baya carnosa, unilocular, voluminosa sin cavidad central, de color verde o amarillo. Según el morfotipo la piel es lisa y puede ser muy sensible a rozaduras. En *C. maxima* los frutos presentan diversas formas (globoso, ovoide, cilíndrico o piriforme), con pesos de hasta 100 kg, piel lisa o con costillas y piel rígida o suave (Lira, 1995). Los frutos de algunas calabazas de exposición de *C. maxima* son los frutos más grandes del mundo, algunos de los cuales pesan más de 600 kg (McCreight et al., 2013). Además de estas calabazas de exposición, *C. maxima* contiene muchos cultivares que tienen formas peculiares de turbante o de plátano, así como cultivares que tienen forma de tambor, de copa, fusiforme u oblonga. El color exterior de los frutos tiene una bonita gama que va desde el rojo-anaranjado hasta el rosa-anaranjado, pasando por el casi blanco, el gris azulado verde intenso y negro-verde (Lira, 1995). El color de la carne puede ser amarillo anaranjado intenso dependiendo del cultivar, La mayoría de ellos tienen forma de tambor y son de color verde oscuro o negro-verde, y son muy apreciadas por su calidad (Grumet et al., 2017) (Figura 6C).

Las semillas son de color blanco amarillento, lisas al tacto y de forma oval, alargada y puntiagudas en el extremo superior. Además, presentan un surco longitudinal paralelo al borde exterior, de 1,5 centímetros de longitud, 0,6-0,7 centímetros de anchura y 0,1-0,2 centímetros de grosor. La temperatura óptima del suelo para que la semilla pueda germinar transcurridos de 2 a 5 días oscila entre los 20 y 25°C. Sin embargo, temperaturas superiores a 40°C o menores a 15°C afectan gravemente a la semilla provocando la disminución de la germinación (Reche, 2000). En *C. maxima* las semillas son gruesas, elípticas, inflados o tumescente. Los colores varían de blanco a pardo claro y suelen ser lisas o un poco estriadas. Bordes redondeados diferenciados del color del centro (Villanueva, 2007). (Lira et al., 2009). Las semillas suelen ser blancas, anaranjadas en algunos cultivares, y tienden a ser bastante grandes, y algunos cultivares se cultivan para el consumo de semillas (Grumet et al., 2017) (Figura 6D).

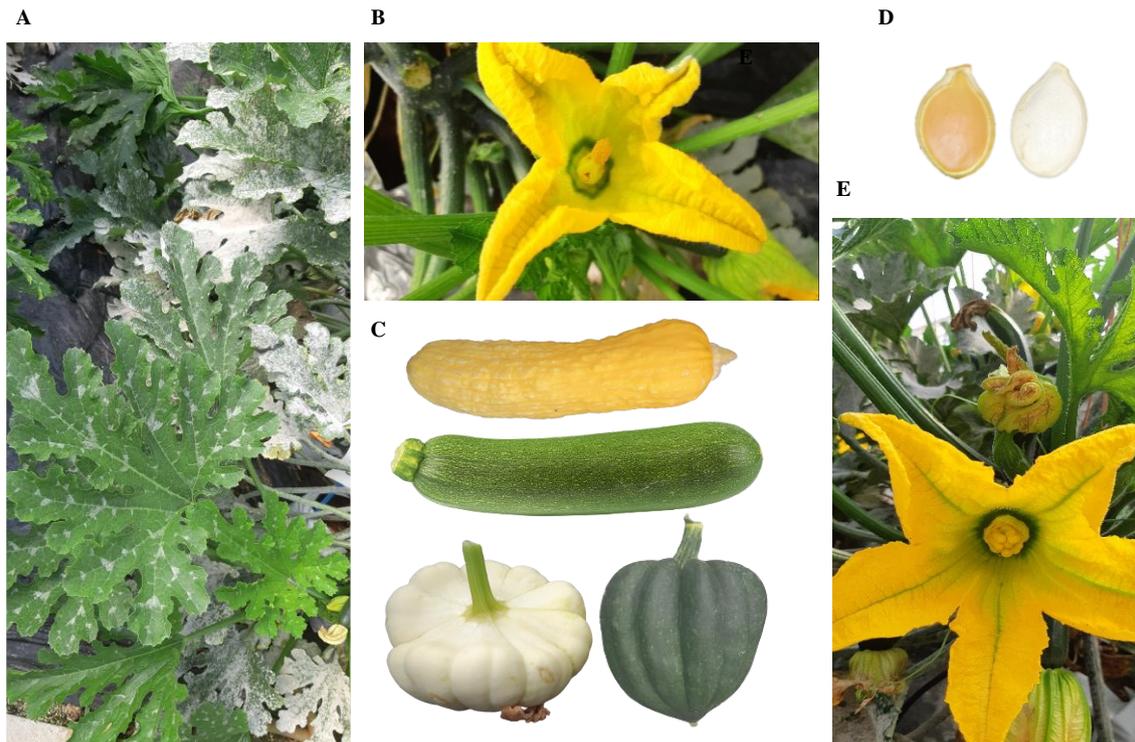


Figura 6. Partes de una planta de *Cucurbita*: **A.** Hojas. **B.** Flor masculina. **C.** Frutos. **D.** Semillas. **E.** Flor femenina. (Elaboración propia).

2.1.1.5. Morfotipos de las especies cultivadas

La especie *C. pepo* incluye 10 morfotipos comestibles o variedades botánicas diferentes y varios tipos silvestres, agrupados en 3 subespecies: spp. *pepo* (Cocozele, Vegetable Marrow, Zucchini, Pumpkin y 2 morfotipos ornamentales), spp. *ovifera* (Crookneck, Straightneck, Scallop, Acorn y un tipo ornamental) y spp. *fraterna* (incluye solo tipos silvestres) (Martínez, 2003; Vicente, 2014) (Figura 7).

Paris y colaboradores (1986;2001;2008) establecieron los diferentes morfotipos para las subespecies *pepo* y *texana* comestibles (Lebeda et al., 2007; Paris et al., 1986;2001;2008) en:

- *Cucurbita pepo* subsp. *pepo*
 - **Pumpkin:** cultivado por sus frutos maduros y semillas. Fruto esférico, corto, cónico y cilíndrico. De color naranja o amarillo verdoso. Corteza lignificada. Su distribución geográfica se encuentra en EEUU-Canadá, Europa, Asia Menor y México-Guatemala.
 - **Zucchini:** cultivado por sus frutos inmaduros. Fruto de forma cilíndrica y uniforme, de color verde oscuro y amarillo, con la corteza tierna y sensible a los golpes o arañazos. Se cultiva a nivel mundial.
 - **Cocozele:** cultivado por sus frutos inmaduros y flores femeninas. Fruto alargado y bulboso en la parte basal, de color verde rayado o verde claro sin rayas. Su distribución geográfica se encuentra en Europa (Italia), Lejano Oeste, Turquía y Yugoslavia.

- **Vegetable marrow:** cultivado por sus frutos inmaduros. Fruto corto de forma cónica y cilíndrica, con diversas tonalidades de verde claro. Corteza no lignificada. Su distribución geográfica se encuentra en Este Medio y norte de África.
- *Cucurbita pepo* subsp. *ovifera* (también conocida como *texana*)
 - **Acorn:** cultivado por sus frutos maduros. Fruto surcado, turbinado de color verde. Corteza no lignificada. Su disposición geográfica se encuentra en EEUU-Canadá.
 - **Scallop:** cultivado por sus frutos inmaduros. Los frutos son planos, festoneados de colores amarillo, blancos verdeceos. Su disposición geográfica se encuentra en Australia, EEUU y Europa.
 - **Straighneck:** cultivado por sus frutos inmaduros. Fruto de cuello corto y constreñido de color amarillo. La corteza está lignificada en la madurez. Su disposición geográfica se encuentra en EEUU y Europa.
 - **Crookneck:** cultivado por sus frutos inmaduros. Frutos de cuello estrecho y curvado. Normalmente tiene presencia de verrugas. La corteza es muy gruesa y lignificada en la madurez. Su disposición geográfica se encuentra en el Sudoeste de EEUU.

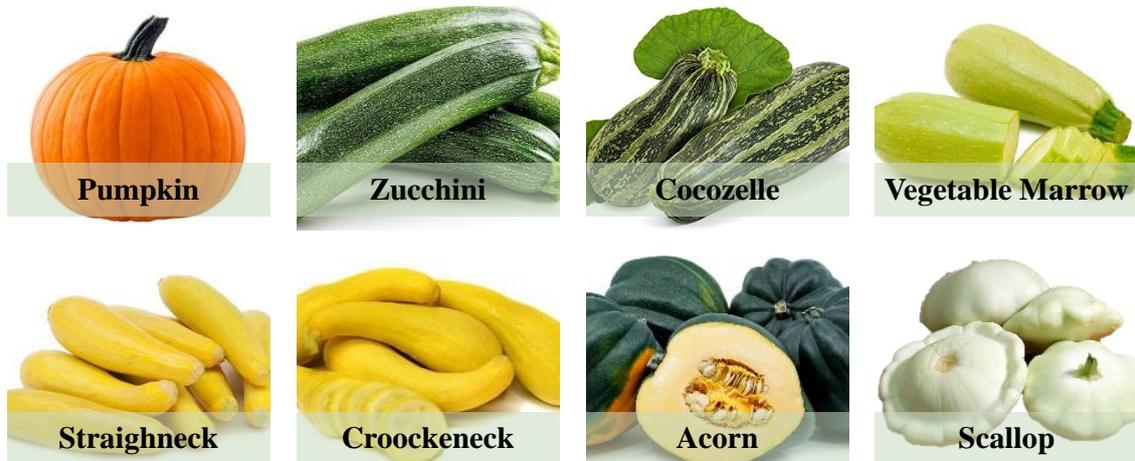


Figura 7. Diferentes morfotipos para las subespecies *pepo* y *texana* comestibles. (Elaboración propia a partir de López, 2019).

Esta clasificación se ajusta a los tipos morfológicos, pero también corresponde con los resultados moleculares obtenidos con isoenzimas y marcadores de DNA (Paris et al., 2003). Además, estos autores definieron tres variedades de calabazas silvestres (no comestibles) pertenecientes dos de ellas a la subespecie *pepo* y la otra a la subespecie *texana*. Las variedades Orange Gourd y Orange Wartyed, pertenecientes a la subespecie *pepo*, se caracterizan por su forma esférica y achatamiento en los polos diferenciándose entre ellas por la presencia de verrugas. La variedad Peara Gourd, perteneciente a la subespecie *texana*, se caracteriza por su forma ovalada-piriforme y lisa (Ruiz, 2017).

La especie *C. maxima* presenta dos subespecies, *C. maxima* subsp. *maxima*, que agrupa las variedades cultivadas y *C. maxima* subsp. *andreana*, la subespecie silvestre

(Filov, 1982). Algunos autores describen los morfotipos de la subespecie cultivadas (Grumet et al., 2017):

- Show Pumpkin: redondo, aunque a menudo desigual.
- Banana: fusiforme alargado.
- Deliciosos: forma de cornete.
- Queensland: oblato.
- Hubbard: fusiforme corto.
- Buttercup: forma de tambor.
- Zapallo: pepónide de forma y tamaño variada (ovalado, ovalalargada, piriforme, esférico, etc.), con epicarpio delgado de color verde, de superficie lisa y mesocarpio verde más claro y poco grosor. La pulpa es de color amarillo y el fruto es de 2,5-3 kg. Su distribución geográfica se encuentra en Argentina y Uruguay. Utilizado para decoración, como trampas para insectos, como alimentación por su pulpa tanto en inmaduro y en maduro, sus semillas flores
- Turban: extremo del estípote abultado, resto del fruto plano.

2.1.2. Características de la especie *Citrullus* spp.

El género xerofítico *Citrullus* Schrad. ex Eckl. & Zeyh ($2n = 22$) tiene su origen en África, donde las poblaciones silvestres, asilvestradas y autóctonas de este género prosperan en regiones áridas y semiáridas del continente (Levi et al., 2017b). El género *Citrullus* incluye además de *C. lanatus*, otras seis especies diploides, tres de las cuales son también de importancia regional (Jarret et al. 1997). La sandía egusi (*C. mucosospermus* (Fursa) Fursa), la sandía cidra (*C. amarus* Schrad.) y *C. colocynthis* (L.) Schrad. Otras especies próximas son la anual *C. rehmii* (De Winter 1990; Jarret and Newman 2000) y las perennes *C. ecirrhosus* (Cogn.) y *C. naudinianus* Sond. (Meeuse, 1962).

2.1.2.1. Origen y domesticación

La sandía (*Citrullus lanatus* (Thunb.) Matsum. et Nakai.) es una planta herbácea monoica cuyo origen está documentado en África tropical y subtropical, se expandió por el norte de África y el Próximo Oriente sobre el IV milenio a.C., donde aún hoy crece en forma silvestre (Giacconi, 1989). También existen registros de que este fruto llegó de manera temprana a Asia o que en este continente existió algún centro de origen o diversificación de plantas cercanas (Maroto, 2002; Del Grosso, 2013).

Se conoce su cultivo desde hace 5.000 años en Egipto, cultivada por su sabor y belleza. Desde Egipto se difundieron las semillas de sandía en las rutas del Mediterráneo, introducida por los musulmanes durante el período de dominación árabe de la península Ibérica entre los siglos VIII y XV. Su nombre marca esta procedencia ya que proviene del árabe hispánico sandíyya, y éste del árabe clásico *sindiyyah*, de Sind, una región del actual Pakistán (Crawford, 2017).

Estudios recientes han propuesto cuatro hipótesis para el origen de la sandía; Primero, que desciende del *C. colocynthis* del norte de África (McCreight et al., 2013). En segundo lugar, que deriva del melón cidro del sur de África, *Citrullus amarus* (Rubatsky, 2001). En tercer lugar, que proviene de *C. mucosospermus* de África occidental subsahariana (Guo et al., 2013; Chomicki y Renner, 2015; Levi et al., 2017a), y cuarto, que se originó y fue domesticado en el noreste de África (París, 2015). Un análisis filogenético de *Citrullus* (Chomicki y Renner, 2015), junto con datos genéticos, arqueológicos e

históricos (París, 2015) rechaza las dos primeras hipótesis, dejando África Occidental y África Nororiental. En un trabajo publicado recientemente se indica que el análisis filogenético de secuencias de genes nucleares sugiere que el melón de carne blanca de Sudán Kordophan es el pariente más cercano de la sandía domesticada (Renner et al., 2019). Los datos filogenómicos también sugieren que el melón egusi de África occidental, *C. mucospermus*, cultivado por sus semillas nutritivas y ricas en aceite (Achigan-Dako et al., 2015), fue domesticado a partir de un acervo genético diferente al de la sandía.

Citrullus amarus, también denominado melón cidro del sur de África, melón forrajero, tsamma, melón en conserva, entre otros, se cultiva por su pulpa que se cocina o encurte (Bush, 1978). Este taxón comprende cultivares antiguos y cultivares de melones de conservación y melones forrajeros (Jeffrey, 2001). Las formas primitivas crecieron en Sudáfrica, mientras que algunos cultivares se cultivan en los EEUU, en el sur de las repúblicas de la CEI (Comunidad de Estados Independientes) y las islas de la Polinesia (Hammer y Hammer, 2002). En Oriente Próximo, Asia Menor, el sur de Rusia y desde Florida hasta Texas y hacia el norte hasta Carolina del Norte (EE. UU.) se encuentra como mala hierba en los campos de melón, campos viejos, naranjales y bordes de carreteras (Stephens, 2003). *Citrullus amarus* cruza fácilmente con *C. lanatus* y con *C. mucospermus*. Sin embargo, existen grandes diferencias en las secuencias del genoma de *C. amarus* cuando se compara con *C. lanatus* o *C. mucospermus* (Guo et al., 2013). Las poblaciones genéticas derivadas de estos cruces con *C. amarus* con estas especies producen segregaciones sesgadas (no mendelianas) para la mayoría de las regiones genómicas (Levi et al., 2002), sugiriendo una mayor distancia genética con respecto a las dos anteriores.

Citrullus mucospermus, el melón egusi, crece en África Occidental, donde se encuentran formas silvestres, semicultivadas y cultivadas por el aceite de sus semillas (Jeffrey, 2001; Achigan-Dako et al., 2006; Dahl Jensen et al., 2011; Akusu y Kiin-Kabari, 2015), ha sido previamente clasificado como *C. lanatus* var. *lanatus* o como *C. lanatus* subsp. *mucospermus* (Fursa 1972). A pesar de la descripción de Fursa (1981, 1983), aún existen lagunas en el conocimiento de la morfología de *C. mucospermus*. Adjakidjè (2006) en “Flore du Benin” no pudo distinguir entre melón egusi y sandía. Esto también resulta de la inconsistencia notada en varias ediciones de Flora en África Tropical Occidental (Hutchinson y Dalziel, 1954). De hecho, *Citrullus mucospermus* y *Citrullus lanatus* comparten secuencias genómicas similares (Guo et al., 2013) y normalmente hibridan entre sí para producir una progenie altamente fértil (Levi et al., 2011). Sin embargo, se han descrito excepciones dependiendo de la dirección del cruce (Fursa, 1983; Gusmini et al., 2004). Otras especies relacionadas son la anual xerofítica *C. rehmii* (De Winter, 1990; Jarret y Newman, 2000) y las plantas perennes *C. ecirrhosus* (Cogn.) y *C. naudinianus* Sond. (Meeuse, 1962), pero quizás la cuarta especie más importante del género sea *C. colocynthis*. También conocida como manzana amarga, se cultiva por sus numerosas propiedades medicinales y el aceite de sus semillas (Hussain et al., 2014). *C. colocynthis* es originaria de los desiertos y las regiones semiáridas del norte de África y del suroeste y Asia central, incluidas las islas del Mediterráneo hacia el este hasta Afganistán, Pakistán e India (Jeffrey, 1967; Burkill, 1985; Dane y Liu, 2007; Paris, 2015).

Como resultado de muchos años de domesticación y selección por la calidad deseable del fruto, la mayoría de los cultivares modernos de sandía comparten una estrecha base genética (Levi et al., 2017a;b). *C. mucospermus* y *C. amarus* son fuentes valiosas de resistencia a muchas enfermedades para *C. lanatus*, (Levi et al., 2017b). Por lo tanto, estas especies tienen el potencial de proporcionar alelos valiosos para ampliar la base genética

de la sandía y aumentar la resistencia a enfermedades y plagas en cultivares de élite, pero también para la introgresión de alelos para otros caracteres de valor agronómico. Las colecciones de germoplasma de sandía mantenidas en distintos lugares del mundo como China, Sudáfrica, Zimbabue y el Sistema Nacional de Germoplasma Vegetal (NPGS) del Servicio de Investigación Agrícola (ARS) del Departamento de Agricultura de los Estados Unidos (USDA), son altamente valiosas para la mejora de *C. lanatus* y la puesta en valor de las especies de cultivo regional que pueden aportar valores añadidos como el contenido en antioxidantes (Martínez et al., 2021).

2.1.2.2. Clasificación taxonómica

De acuerdo con el *Integrated Taxonomic Information System* (ITIS,2021) la taxonomía de la especie *C. lanatus*, *C. mucosuspermus* y *C. amarus* es de la siguiente manera:

Tabla 4. *Taxonomía de la sandía*

Reino:	<i>Plantae</i>
Subreino:	<i>Viridiplantae</i>
Infrareino:	<i>Streptophyta</i>
Superdivisión:	<i>Embryophyta</i>
División:	<i>Tracheophyta</i>
Subdivisión:	<i>Spermatophytina</i>
Clase:	<i>Magnoliopsida</i>
Superorden:	<i>Rosanae</i>
Orden:	Cucurbitales
Familia:	<i>Cucurbitaceae</i>
Género:	<i>Citrullus</i> Schrad. ex Eckl. & Zeyh. <i>Citrullus lanatus</i> (Thunb.) Matsum. & Nakai
Especies:	<i>Citrullus amarus</i> Schrad. <i>Citrullus mucosuspermus</i> (Fursa) Fursa

2.1.2.3. Usos y valor nutricional

Como ya se ha mencionado, además del fruto de *C. lanatus*, distintas especies del género son cultivadas por sus frutos y por sus semillas. De la sandía se aprovechan principalmente sus frutos que son ricos en azúcar, muy refrescantes y de bajo valor calórico. Las semillas, como las de otras cucurbitáceas, son ricas en aceite comestible o de usos industriales rico en ácido linoleico (40-60%), oleico (10-20%) y palmítico (0-10%). Estas semillas pueden consumirse tostadas y son una buena fuente de proteínas y minerales (Robinson y Decker, 1997). En *C. amarus*, los frutos se utilizan como pienso para el ganado y ocasionalmente también como pienso para cerdos, con menos frecuencia para la preparación de cáscara de cidra o para la producción de pectina (Laghetti y Hammer, 2007). También se encuentra como malas hierbas en pinares y otros campos (Grichar et al., 2001; Hall et al., 2004). *C. colocynthis* y *C. mucosuspermus* son cultivados por los aceites de sus semillas (Levi et al., 2017a). Valores aproximados de vitaminas y minerales de *C. lanatus* se muestran en la Figura 8 (Moreiras et al., 2013; Valero et al., 2018).

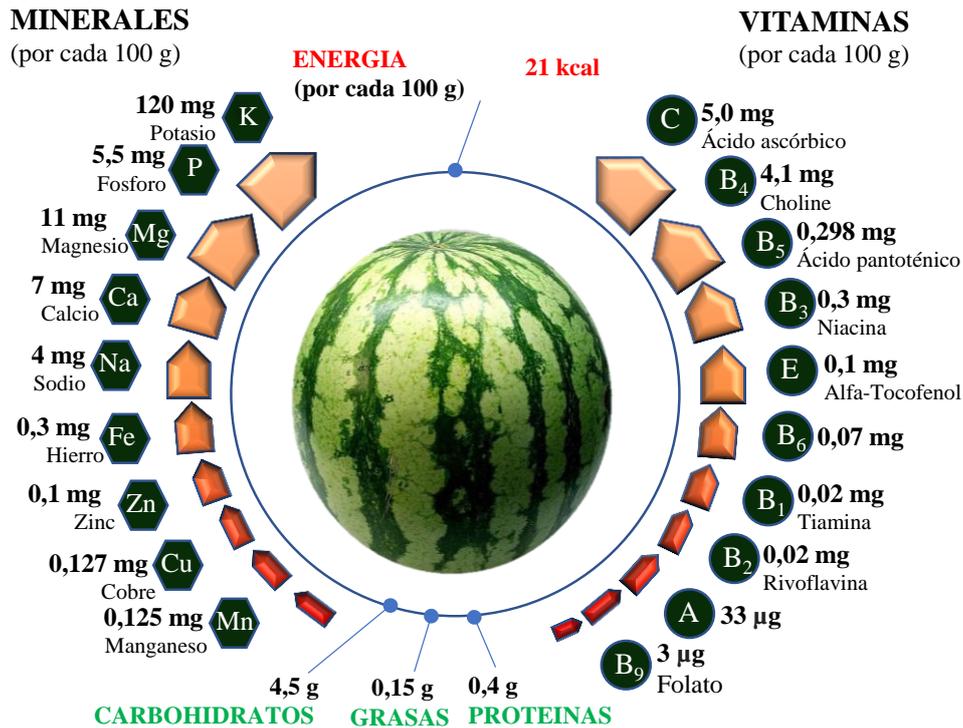


Figura 8. Valor nutricional y mineral de *C. lanatus*. (Elaboración propia a partir de Moreiras et al., 2013 y Valero et al., 2018).

2.1.2.4. Descripción morfológica

C. lanatus es una especie anual, monoica, herbácea, sin tronco, de tallos o guías tiernas, blandos, flexibles, rastreros que pueden alcanzar 4 a 6 metros de largo, provistos de zarcillos bífidos o trifidos (Figura 9G), por medio de los cuales puede tener hábito trepador. El desarrollo aéreo de la planta parte del brote principal y no es hasta que aparecen de 5 a 8 hojas bien desarrolladas cuando emergen otros brotes. Cuando la planta se ha desarrollado completamente, emergen los brotes terciarios y así sucesivamente hasta alcanzar superficies de 4-5 m². Las hojas son pinnado-partidas (Figura 9D) y están divididas en tres-cinco lóbulos y estos a su vez se dividen en segmentos redondeados. El haz de la hoja es liso y el envés áspero y recubierto de pilosidades (Reche, 2000).

El sistema radicular de la sandía presenta una raíz principal mucho más desarrollada que las secundarias que se localizan superficialmente de modo amplio. Este órgano de la planta carece de importancia hoy día ya que la parte área de la sandía es injertada sobre patrón *C. máxima* x *C. moschata* (Camacho, 2003).

La flor de la sandía es de color amarillo, solitaria, pedunculada y axilar que atrae a los insectos por su color, aroma y néctar. La planta de sandía es monoica, es decir coexisten en el mismo pie las flores masculinas o estaminadas (Figura 9F) y las femeninas o pistiladas (Figura 9E). La diferenciación de sexos es relativamente fácil ya que la flor femenina posee un ovario ínfero recubierto de pilosidades. La flor femenina aparece en el brote principal y en los brotes secundarios y terciarios. La primera flor femenina aparece en la axila de la 7-10 hoja del brote principal y luego por cada 4-5 flores masculinas aparece una flor femenina (Camacho, 2003; Reche, 2000). Además, existe correlación entre el número de tubos polínicos germinados y el tamaño del fruto (Maroto et al., 2002). Existen cultivares trimonoicos de sandía en los que conviven flores masculinas, femeninas y hermafroditas (Manzano et al., 2016; Aguado et al., 2020). Además *C.*

naudinianus se ha descrito como especie dioica (Renner et al., 2017). El fruto es una baya globosa u oblonga en pepónide, con placenta carnosa y epicarpio quebradizo, con colores tanto de pulpa como de corteza muy variados (Figura 9A y B). Hay variedades con corteza de color verde oscuro, verde claro, amarillo o franjas de color amarillo o verde sobre fondos verdes. La pulpa, por su parte, adquiere pigmentaciones rojas, rosadas o amarillas (Camacho, 2003; Reche, 2000). El tamaño del fruto oscila desde los dos a los veinte kilogramos, aunque la tendencia actual es hacer frutos más pequeños que puedan ser consumidos de una vez. El tamaño del fruto está directamente relacionado por el número de frutos por planta, debido a que todos los fotoasimilados de la misma deben ser repartidos entre todos los frutos (Mejías, 2004). El fruto es climatérico, con gran contenido de agua, mayor a un 90%, y sabor dulce (10-14°Brix (Maynard, 2001; Wehner et al., 2001; Crawford, 2017). Las variedades autóctonas primitivas de *C. lanatus* tienen un bajo contenido de azúcar y podrían ser una valiosa fuente de compuestos bioactivos, como la citrulina, flavonoides y carotenoides, como el α - y el β -caroteno o el licopeno (Davis et al., 2007).

Las semillas (Figura 9C) son de tamaño variable en función de la variedad y suelen ser de color blanco, negro o marrón. Están distribuidas por toda la pulpa, a diferencia del melón en el que se encuentran dispuestas en la cavidad central. Las semillas son aplastadas, ovoides, duras y con el doble de anchura que de longitud (Camacho, 2003; Reche, 2000). Por último, el tiempo de germinación será menor si las semillas son diploides, seguidas de las tetraploides y por último las triploides. *C. mucospermus*, el melón 'egusi' se cultiva para el consumo de las semillas, la única porción comestible; la pulpa de la fruta en estos cultivares suele ser demasiado amarga para el consumo humano (Burkill, 1995; van der Vossen et al., 2004; Achigan Dako et al., 2008a).

Achigan-Dako et al. (2015) estudiaron la diversificación de *Citrullus*, incluyendo *C. mucospermus*, *C. lanatus*, *C. amarus* y *C. colocynthis*. Los marcadores morfológicos que utilizaron para caracteres de la hoja, el fruto y de la semilla (tamaño de la hoja y del del fruto, el color y el sabor de la pulpa, el tamaño de la semilla y el color de la semilla) confirman la gran diversidad inter e intraespecífica. Aunque se observa variación en la longitud y anchura de la hoja, no hay una diferenciación clara entre las especies. Sí se observa variación en el peso del fruto intraespecífica, pero también entre especies. Los frutos grandes son típicos de *C. lanatus*, mientras que *C. mucospermus* presenta un tamaño de fruto entre pequeño y mediano. El sabor de la carne es casi siempre amargo en *C. mucospermus* y *C. colocynthis*, mientras que *C. lanatus* es dulce o neutro en el caso de accesiones de carne no roja. El color de la carne es, en la mayoría de los casos, rojo en *C. lanatus* y puede volverse amarillo-naranja y, raramente, blanco (Maggs-Kolling et al., 2000; Solmaz y Sarı 2009; Solmaz et al., 2010; van der Vossen et al., 2004). Sin embargo, en las demás especies predomina el color blanco de la carne (Achigan-Dako et al., 2015) (Figura 9B). También encontraron variación en el tamaño y forma de las semillas. El peso y el área de las semillas varió entre especies, con mayor valor en *C. mucospermus*. En *C. lanatus* el tamaño de las semillas varía de pequeño a mediano. En cuanto al color de las semillas, en *C. lanatus* va de negro a amarillo-naranja, con un bajo porcentaje de color oscuro. *C. mucospermus* tiene mayoritariamente semillas con tegumento amarillo con un borde negro o blanquecino y en *C. colocynthis* el color predominante fue amarillo-naranja, seguido de gris-marrón y naranja-gris. Por último, las semillas de *C. amarus* presentan color amarillo-naranja o grisáceo.

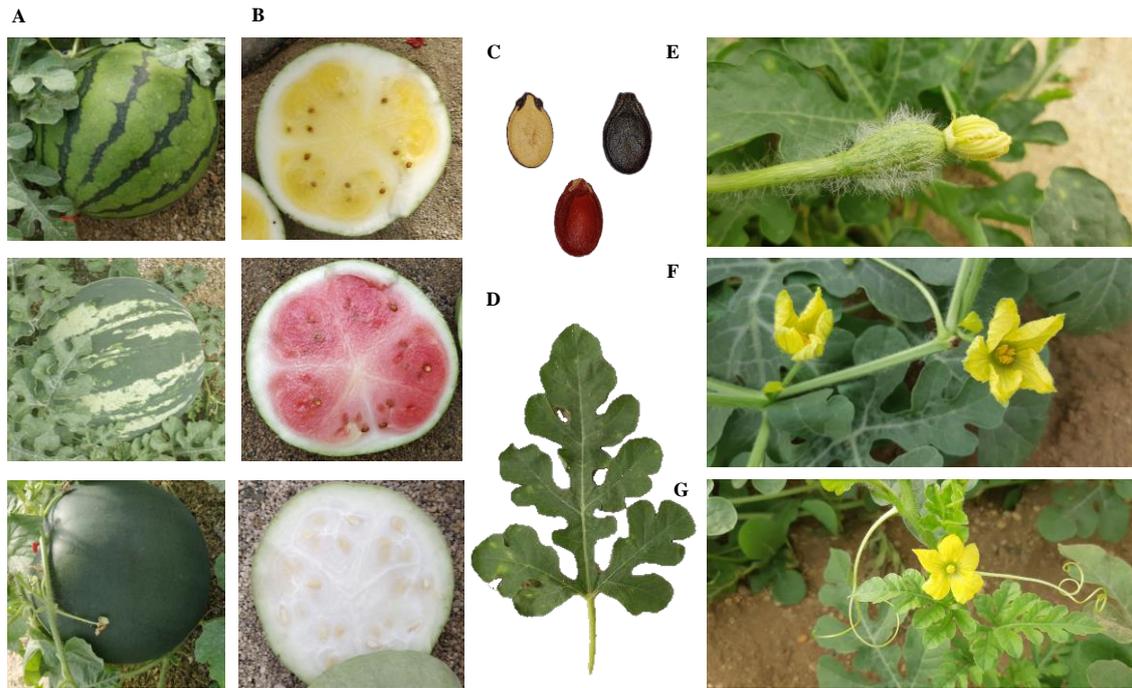


Figura 9. Partes de una planta de *Citrullus*: **A.** Parte externa del fruto. **B.** Parte interna del fruto. **C.** Semillas. **D.** Hoja pinado-partida. **E.** Flor femenina. **F.** Flor masculina. **G.** Zarcillos. (Elaboración propia).

2.1.2.5. Morfotipos de las especies cultivadas

La sandía es la especie más cultivada del género, no obstante existen otras de importancia local o regional. Las especies cultivadas son (Figura 10):

- (1) *Citrullus lanatus* ((Thunb.) Matsum y Nakai,).
- (2) *Citrullus colocynthis* (L.) Schard
- (3) *Citrullus amarus*. Clasificada durante años como *C. lanatus* var. *citroides* (Levi et al., 2017b).
- (4) *Citrullus mucospermus*. Considerada por algunos autores como subespecie de *C. lanatus* (Levi et al., 2017b).

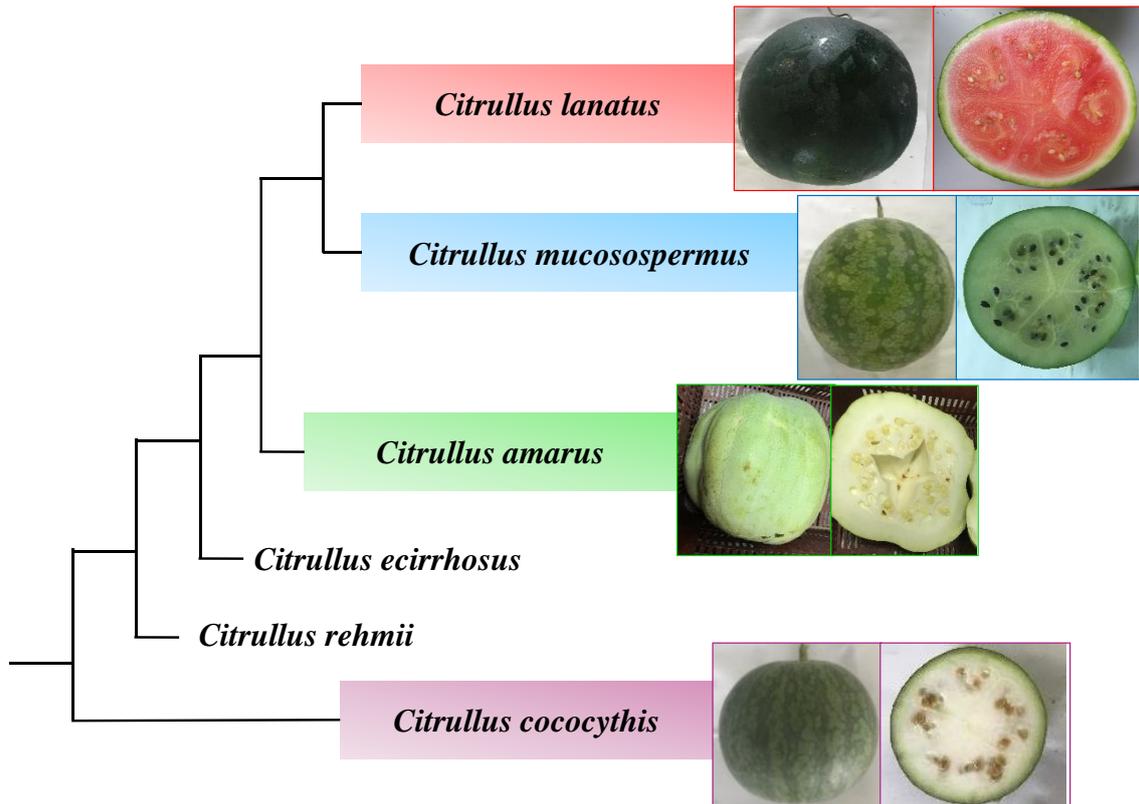


Figura 10. Especies cultivadas del género *Citrullus*. (Elaboración propia a partir de Renner et al., 2019; 2017).

La mejora en sandía se ha centrado en características morfológicas (tamaño y forma del fruto, color de la pulpa, color y patrón de la cáscara, grosor y firmeza de la cáscara y tamaño de semilla), en su calidad (perfil de azúcares) y en la resistencia a enfermedades, y plagas desarrollándose variedades e híbridos (F1) con las características deseadas. El gran surtido de variedades comercializadas hoy en día, son híbridos F1 obtenidos por mejora genética tras el cruzamiento y selección de materiales de interés. Los híbridos F1 derivan del cruce de dos variedades distintas. Se consigue así que las semillas resultantes de la planta madre que da lugar a una descendencia homogénea únicamente durante el primer año (F1). Esta homogeneidad se rompe los años siguientes por recombinación genética (en la generación segregante F2).

Diversas empresas de semillas nos ofrecen aproximadamente sesenta y cinco variedades diferentes de sandía, en el Sureste Peninsular se cultivan aproximadamente ocho-diez variedades (Camacho, 2003). Para estas diez variedades podemos hacer grupos:

- Variedades de corteza verde oscuro “tipo Sugar Baby”.
- Variedades de corteza rayada “tipo Crimson”.

Las sandías minis sin pepitas son una innovación que ha ido ganando mercado en los últimos años debido a la comodidad de su tamaño.

Genéticamente las sandías pueden ser:

- Diploides. Son las variedades más tradicionales. Producen semillas oscuras de consistencia leñosa.

- Triploides. No producen semillas viables. Derivan del cruce entre una variedad tetraploide (4n) y otra diploide (2n) dando como resultado una semilla híbrida (triploide, 3n) estéril que no produce semillas. Por este motivo deben cultivarse junto a variedades diploides, que actúan como polinizadoras, para que las flores puedan ser fecundadas y den producción.

Según la forma de sus frutos encontramos:

- Frutos redondos. Con corteza de color verde oscuro o negro, son los ejemplares más cultivados. Destacan: Sugar Baby, Crimson Sweet, Resistent, Dulce Maravilla, Sweet Marvell y Early Star, entre las más conocidas y cultivadas en la cuenca Mediterránea.

- Frutos alargados. Con corteza verde con bandas claras u oscuras. En España apenas se cultivan. Destacan los tipos Charleston Gray, Klondike, y Dumara, o combinaciones entre ellos.

Predomina cada vez más el cultivo de variedades triploides. Tienen semillas tiernas de color blanco que pasan desapercibidas al comer. Destacan las de piel rayada como Reina de Corazones, Boston, Motril, Trix Paula, Estel Deluxe, Reina Linda, Torera etc. dan lugar a frutos de gran calibre. En los últimos años se ha incrementado la producción de sandías de piel negra triploides predominando Fashion.

2.2. Variabilidad natural: conservación de recursos fitogenéticos

Los recursos fitogenéticos están constituidos por colecciones de variedades cultivadas y variedades recién obtenidas, variedades en desuso, razas locales y primitivas, especies silvestres y estirpes genéticas especiales (Cubero, 2013). Existen diversas posibilidades de conservación de estos recursos. Uno de los mejores sistemas de conservación *ex situ* son los bancos de germoplasma. En la actualidad, gran parte de estos bancos están destinados a la conservación de especies con interés agroalimentario. Los bancos de germoplasma pueden ser colecciones de semillas, de plantas vivas, de tejidos vegetales cultivados *in vitro*, de polen, o de ADN. Los bancos de germoplasma y las colecciones son sistemas de conservación *ex situ* consistentes en un conjunto de diferentes accesiones o entradas de una especie (para designar a cada muestra diferente dentro de una colección se utiliza el término “entrada”, aunque es frecuente también el empleo del anglicismo “acesión”). Se consideran bancos de germoplasma a aquellas colecciones de cierta envergadura, que sean referente de la especie y generalmente localizadas en organismos públicos, donde se garantice la continuidad y la disponibilidad para su investigación y uso. La mayoría de los bancos de germoplasma tiene el mandato de distribuir germoplasma a los usuarios (Kameswara, 2007). Existen alrededor de 1600 bancos de germoplasma, algunos de los más importantes son el banco del Instituto Vavilov, en San Petersburgo (N.I. Vavilov Institute of Plant Industry, VIR), el del Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK), en Gatersleben, la red de información de los recursos de germoplasma (GRIN), del departamento de agricultura de los Estados Unidos (USDA) y los del Grupo Consultivo para la Investigación Agrícola Internacional (GICAI). En España existe el CFR (Centro de Recursos Filogenéticos) creado en 1993 como centro de conservación de colecciones base de semillas y como centro de documentación de los recursos fitogenéticos de la red, realizando además las actividades correspondientes a la gestión del germoplasma depositado.

Dentro de las alternativas que tienen los bancos de germoplasma para la conservación de los recursos fitogenéticos, la técnica más extendida es la de los bancos de semillas,

esto es debido a la facilidad de almacenamiento y manipulación, pues en poco espacio pueden conservarse cantidades enormes de individuos vivos durante prolongados espacios de tiempo (Maxted et al., 1997). Por un lado, las semillas son unidades adaptadas a la dispersión en el tiempo y, por tanto, capaces en la mayoría de los casos de permanecer viables, de forma natural, durante largos períodos de tiempo (Chin, 1994). En segundo lugar, el pequeño tamaño de las semillas, unido a la posibilidad de que cada una de ellas posea una constitución genética diferente, asegura la conservación de una gran diversidad genética en un espacio reducido (Iriondo y Pérez, 1999). Normalmente los bancos están organizados en dos grandes colecciones: una primera de conservación a largo plazo, denominada colección base, y una segunda para el trabajo cotidiano del banco, la colección activa. Aunque el almacenamiento en bancos de germoplasma está universalmente aceptado como parte integral de los programas de conservación de plantas silvestres, las muestras de semillas de plantas silvestres almacenadas en bancos de germoplasma suponen menos del 2 % del total de germoplasma almacenado, destinado esencialmente a plantas cultivadas (Iriondo, 2001).

En este TFM se empleará los recursos de conservación procedentes del USDA-GRIN. Esta colección contiene alrededor de 1.453 accesiones de las siete especies de *Citrullus*, con numerosas variedades autóctonas y primitivas recogidas a lo largo de más de 75 años (Levi et al., 2017a). La colección de *Citrullus* ha sido evaluada en cuanto a su resistencia a las principales enfermedades y plagas de la sandía. Muchas accesiones PIs (Plant Introduction) han sido seleccionadas para la mejora genética de distintos caracteres y han demostrado ser útiles en estudios genéticos para estudiar la herencia de distintos caracteres y también para el estudio de la taxonomía y la evolución de *Citrullus* (Reddy et al., 2014a, b; Levi et al. 2013; 2017a). La colección de *Cucurbita* es también destacable y cuenta con 1.766 accesiones y representan la variabilidad presente en los cinco continentes (ver Material y Métodos). Un valor añadido de esta colección es son los trabajos de re-secuenciación sobre distintas accesiones que se está realizando en los últimos años (Wu et al., 2019; Guo et al., 2019) y que suponen una fuente de información con la que contrastar los datos fenotípicos. Estos resultados apoyan los datos genómicos disponibles previamente, tanto *C. lanatus*, como *C. pepo* y *C. maxima*, especies objeto de interés en este TFM, cuentan con genomas de referencia (Guo et al., 2013; Montero-Pau et al., 2017; Sun et al., 2017).

2.3. Análisis de la diversidad genética a escala genómica

2.3.1. Análisis de asociación del genoma (GWAS)

Los avances en las tecnologías de secuenciación de próxima generación y los sistemas de genotipado de alto rendimiento han revolucionado el campo de la genómica de plantas, lo que ha llevado a una fácil disponibilidad de miles de polimorfismos de un solo nucleótido (SNP) para la mejora genética y estudio de especies cultivadas (Lado, 2012). Los estudios de asociación de todo el genoma (GWAS), son una herramienta que permite utilizar gran número de accesiones y los SNPs de todo el genoma para estudiar la asociación entre caracteres y regiones del genoma que las regulan. El GWAS ha surgido como un enfoque alternativo para el mapeo de QTL biparental en varios cultivos. GWAS parte de un enfoque de genética de población para realizar análisis de asociación entre la variación fenotípica y los SNPs de todo el genoma en accesiones, con mayor resolución que los análisis de poblaciones biparentales (Rafalski, 2010; Zhang et al., 2010). GWAS puede identificar los factores genéticos que determinan las variaciones fenotípicas del conjunto genético de múltiples accesiones, lo que indica que se pueden identificar

múltiples regiones relacionadas con varios mecanismos moleculares implicados en las variaciones fenotípicas (Nakano, 2020).

En contraste con el mapeo genético que usa progenie biparental, no hay necesidad de desarrollar una población de mapeo (como una F2) para GWAS, porque esta estrategia es detectar las regiones genómicas asociadas usando los eventos histórico-naturales de recombinación entre las accesiones (Nakano, 2020). Además, una vez que se dispone de la información genotípica de las accesiones, esta se puede usar para realizar análisis GWAS para distintos caracteres, lo que lleva a una disección rápida del control de distintos caracteres con valor agronómico. Por tanto, si bien los avances en la tecnología de secuenciación para generar marcadores SNPs distribuidos en todo el genoma han facilitado la implementación de GWAS, el fenotipado de un gran número de accesiones sigue siendo un cuello de botella para GWAS en plantas (Korte y Farlow, 2013).

Mientras que el análisis de QTLs se basa en el desequilibrio de ligamiento (LD) detectado a través del genotipado y fenotipado de diversos individuos de una misma población. Generalmente, un GWAS infiere estas asociaciones a través de una prueba de hipótesis con diferentes pruebas estadísticas. Estas pruebas o métodos estadísticos se han implementado para controlar factores que pueden alterar los resultados. Los principales son la estructura de la población y las relaciones crípticas entre accesiones (Marees et al., 2018). La estructura de la población ocurre cuando se observa estratificación de la población debido a la existencia de subpoblaciones dentro de la población principal, y las distintas subpoblaciones tienen asociado un fenotipo en particular. Dicho de otra manera, el fenotipo no varía dentro de las accesiones de una subpoblación. La estructura de la población y las relaciones crípticas conducen a la identificación de falsos positivos en el análisis GWAS. Mientras que las relaciones fuertes entre individuos también conducen a estimaciones sesgadas debido al error de los efectos de cada SNP (Marees et al., 2018). Un ejemplo de la importancia de la estructura población es presentado en el análisis de las 1.234 accesiones de la colección de pepino del USDA. Las accesiones del este de Asia se encontraban claramente separadas de las accesiones del sur de Asia e India, mientras que una gran parte de las accesiones procedentes de Europa, América del Norte, África, Turquía y Asia Central/Occidental mostraron una mezcla genética, mientras que la mayoría de las accesiones de Asia Oriental tenían un fondo genético homogéneo (Wang et al., 2018). Para afrontar estos problemas se estiman las matrices de distancia (Multidimensional Scalling, MDS) y *Kinship*.

El desarrollo de la técnica GWAS ha conducido, en los últimos años, al desarrollo de métodos de análisis que incluyen estas correcciones, el método GLM (*General Lineal Model*) estima las distancias entre las accesiones empleando la matriz de genotipado para ponderar la estratificación de la población y el método MLM (*Mixed Lineal Model*) estima además la matriz *Kinship* que pondera las relaciones entre accesiones. La selección de un método u otro para el análisis GWAS se puede realizar a través del ajuste del gráfico QQ (se presenta a continuación). En general, se utilizan tres gráficos para la visualización de resultados (Koh, 2015):

- Gráfico de Manhattan: gráfico de dispersión que muestra los p-valores en la escala $-\log_{10}(p)$ frente a la posición genómica de los SNP y sus números de cromosomas. Los picos más altos corresponden a valores p pequeños (y mayor significación estadística) e indican que la región genómica correspondiente tiene una fuerte asociación con el carácter.

- Gráfico de cuantiles-cuantiles (QQ): gráfico que se utiliza para evaluar la bondad del modelo utilizado (GLM o MLM) en el GWAS. Compara para cada SNP el p-valor observado y el esperado. Es más probable que los SNP en la sección superior derecha del gráfico se desvíen de la diagonal estén asociados con el carácter de estudio. De forma que si el número de SNPs que se desvían positivamente es muy alto puede existir sobreestimación de los SNPs asociados, y si el ajuste de los SNP se sitúa por debajo de la línea diagonal puede indicar una subestimación (ver Material y Métodos). El análisis del gráfico QQ conduce a la selección del método adecuado de análisis y, por tanto, a los SNP con mejor asociación al carácter.
- Gráfico de componentes principales (PC): método para estimar el efecto de la estructura de la población mediante el análisis de datos multivariados en términos de estructura de covarianza de los datos.

El primer análisis GWAS fue realizado en humanos con gran éxito (Klein et al., 2005), posteriormente se extendió su uso a animales y plantas. Un gran número de análisis GWAS han sido realizado en los principales cultivos de importancia mundial asociados a características agronómicas de importancia, como resistencia a enfermedades y tolerancia a estrés biótico y abiótico (Brachi y Borevitz, 2011). La identificación de asociaciones marcador-carácter proporciona información sobre la ubicación de las regiones genómicas asociadas con el control genético de los caracteres de interés (Camargo, 2017). Se han realizado numerosos análisis con éxito en cucurbitáceas como la sandía, para caracteres del fruto (Guo et al., 2019), para caracteres de determinación del sexo (Aguado et al., 2020) y también en pepino para la identificación de regiones de interés para la resistencia a enfermedades (Wang et al., 2018)

2.4. Caracteres de mejora en cucurbitáceas

Uno de los factores que preocupa a los mejoradores en *Cucurbita* es el control genético de la floración. Ha sido identificado un gen único para la expresión sexual ginoica (*G*) en *C. foetidissima*, aunque por el momento no se ha transferido a los calabacines cultivados (Kole, 2011). Por otra parte, hay cuatro genes que controlan la esterilidad masculina, dos para *C. pepo* y dos para *C. máxima*. Estos genes no pueden ser utilizados en la actualidad ya que no se dispone de marcadores moleculares ligados. Además, hay dos genes de esterilidad total, uno en *C. máxima* y otro en *C. pepo* (Kole, 2011). Además de estos genes, el sistema de determinación del sexo, para la producción de flores femeninas, masculinas y hermafroditas se conoce bastante bien en melón, pepino y cada vez más en calabacín y sandía. En estas cuatro especies el etileno juega un papel fundamental en la inducción de la producción de flores femeninas (Boualem et al., 2008; 2009; 2015; Chen et al., 2016; Martínez et al., 2014; Martin et al., 2009; Manzano et al., 2016; García et al., 2020a;b).

En cuanto a la resistencia a enfermedades víricas y fúngicas, el oídio es la más ampliamente distribuida y devastadora en calabacín. Aunque hay cultivares que presentan alguna resistencia a la enfermedad (*C. moschata* y *C. pepo*) la fuente de resistencia procede de una transferencia interespecífica inicial desde *C. okechobeensis* en *C. moschata* y posteriormente a *C. pepo*. Algunos estudios han demostrado que el gen *Pm-1* es el único que confiere resistencia al odio y es por ello por lo que se han identificado marcadores moleculares tipo SSR (*Simple Sequence Repeats*) ligados a ambos lados del gen localizados a distancias de 5,2 y 8,7 cM respectivamente (Kole, 2011; Zhang, 2013) que están siendo utilizados en programas de mejora.

Por otra parte, la podredumbre de cuello, raíces y frutos causadas por el humedecimiento de la planta causado por *Phytophthora capsici* parece estar controlada por tres genes dominantes. En cuanto a otras enfermedades y virus, desde especies silvestres han sido transferidas resistencias a diversos virus (Kole, 2011). Una fuente de resistencia para múltiples enfermedades víricas de *C. pepo* ha sido *C. moschata*, múltiples accesiones de la variabilidad natural han resultado resistentes, entre ellas se han detectado resistencias al virus ToLCNDV, que ha generado muchas pérdidas en esta especie en los últimos años (Sáez et al., 2020; Romero-Masegosa et al., 2021).

En cuanto a la forma, color y aspecto general de fruto, los mejoradores continuamente buscan genes que controlen estos caracteres con el objetivo de obtener variedades que duren en buen estado más tiempo y que tengan la forma, sabor y color del calabacín comercial. Existen ya casas de semillas que han conseguido cultivares en los que se ha reducido los tricomas en tallos y peciolo evitando así la abrasión del fruto (Kole, 2011). Los mejoradores buscan genes en variedades silvestres que controlen el brillo, el tiempo de duración de la calidad del fruto: la ausencia de pelos finos y duros en los tallos, todos los esfuerzos por ofrecer un fruto de mayor calidad, que aguante más tiempo y sea más cómodo tanto para el manejo del agricultor como para el transporte a grandes distancias. También es interesante la partenocarpia por la dificultad de polinización y fecundación (Kole, 2011).

El tamaño de las semillas es un objetivo de selección en la mejora de las sandías, calabacines y calabazas, tanto si se desea una semilla bien formada para favorecer el desarrollo de la planta, como si el objetivo es obtener semillas de las que extraer aceites (Prothro et al., 2012; Meru et al., 2018; Achigan-Dako et al., 2008b; Mogotlane et al., 2018; Fruhwirth et al., 2007; Seo et al., 2005; Paris y Nelson, 2003). El color de las semillas puede estar relacionado con el contenido en compuestos antioxidantes, tal y como se ha observado para la pulpa de los frutos en *Cucurbita* spp. y *C. lanatus* (Murkovic et al., 2002; Yoo et al., 2012). De hecho, de ambos grupos taxonómicos se han realizado trabajos que demuestran la importancia de la semilla en relación al contenido en antioxidantes (Mogotlane et al., 2018; Seo et al., 2005).

Hasta la fecha se han detectado hasta 61 *Quantitative trait loci* (QTLs) en poblaciones segregantes de sandía asociados al tamaño de la semilla (Guo et al., 2020). Mientras que en *Cucurbita* spp. se han detectado 16 QTLs en 7 grupos de ligamiento (Tan et al., 2013; Wang et al., 2020). De igual forma se han detectado QTL de interés para caracteres de la semilla en melón y en pepino (Guo et al., 2020). Entre los distintos QTLs, cuatro de sandía (*CISS2.2*, *CISS6.1*, *CISS6.2* y *CISS8.2*), dos de pepino (*CsSS4.1* y *CsSS5.1*) y uno de melón (*CmSS11.1*) eran QTLs mayores para el tamaño y el peso de las semillas. El análisis de sintenia indica que estos QTL de efecto mayor se localizaron en regiones sintéticas en los genomas de las distintas especies, lo que indica conservación estructural y funcional de algunos genes importantes para el control del tamaño de las semillas. La anotación indica que familias de proteínas zinc finger y ubiquitina-proteína ligasa E3 son importantes para el tamaño de las semillas (Guo et al., 2020).

También se ha estudiado el color de la semilla tanto en *Citrullus* spp. (principalmente *C. lanatus*), como en *C. maxima* (Paudel et al., 2019; Li et al., 2020; Shi et al., 2021). El control genético del color de la semilla es conocido mejor en sandía, en esta especie se han identificado cuatro genes *R/r*, *T/t*, *W/w* y el modificador *D/d*, que sólo actúa cuando alelos dominantes de los otros genes están presentes (Paudel et al., 2019). La combinación de estos loci da lugar a la formación de un color u otro de testa: negro (*RTWD*), negro

punteado (*RTWd*), verde (*rTW*), bronce (*RtW*), mezcla (*RTw*), rojo (*RTw*), blanco de punta bronce (*Rtw*) y blanco punta rosa (*rtw*) (Paudel et al., 2019). Los avances realizados en secuenciación en los últimos años y estudios de mapeo realizados previamente han permitido determinar distintas regiones genómicas en las que se localizan estos loci. Los genes *R/r*, *T/t*, *W/w* y *D/d* han sido mapeados en los cromosomas 3, 5, 6, y 8 del genoma de sandía respectivamente (Paudel et al., 2019; Li et al., 2020). Asimismo, una región situada en el cromosoma 15 del genoma de *C. maxima* ha sido detectada en el análisis de poblaciones segregantes (Shi et al., 2021).

La disponibilidad de información genómica generada en los últimos años en estos grupos taxonómicos permite un nuevo enfoque para detectar regiones del genoma implicadas en los caracteres de interés. La caracterización de accesiones para el peso, color y presencia de hilio de la semilla permitirá, junto a los datos genómicos disponibles, realizar estudios de asociación (*Genome Whole Association Study*, GWAS) útiles para descubrir regiones de interés, marcadores y accesiones con valor para la mejora genética.

MATERIAL Y MÉTODOS



3. MATERIAL Y MÉTODOS

3.1. Material vegetal

El Departamento de Agricultura de EE. UU (USDA) mantiene una colección de 1.453 accesiones de *Citrullus* spp. y más de 1.766 de *Cucurbita* spp. Entre las accesiones de *Citrullus* spp. destacan 1.239 de *Citrullus lanatus* (Thunb.) Matsum. & Nakai, 132 accesiones de *C. amarus* Schrad., 52 de *Citrullus mucosospermus* (Fursa) Fursa, además conservan diferentes accesiones de las especies *Citrullus colocynthis* (L.) Schrad., *Citrullus ecirrhosus* Cogn., *Citrullus naudinianus* (Sond.) Hook. f. y *Citrullus rehmi* De Winter. Las especies del género *Cucurbita* mejor representadas en la colección son: *Cucurbita pepo* L. (724), *Cucurbita maxima* Duchesne (528), *Cucurbita moschata* Duchesne (278), *Cucurbita argyrosperma* C. Huber (135 accesiones), *Cucurbita foetidissima* Kunth (30), *Cucurbita lundelliana* L. H. Bailey (9), *Cucurbita okechobeensis* (Small) L. H. Bailey (8), además la colección incluye otras especies como *Cucurbita pedatifolia* L. H. Bailey, *Cucurbita cordata* S. Watson y *Cucurbita digitata* A. Gray. Estas accesiones representan la diversidad de ambos géneros, tanto en su representación por especie como por su origen geográfico. Las accesiones de *Citrullus* spp. proceden de los cinco continentes con máxima representación de su origen de diversificación (África) y centros secundarios o regiones con importancia económica, como Turquía, China y EEUU (USDA, 2021). De igual manera, las accesiones de *Cucurbita* spp. de este banco de germoplasma han sido recolectadas en diferentes regiones geográficas principalmente en Centroamérica, origen principal de diversificación, seguida por Sur y Norteamérica, Europa (principalmente España), Asia, África y Australia. Esta variabilidad se traduce en una gran diversidad fenotípica y genética (Levi et al., 2013; Nimmakayala et al., 2014) y representa un recurso valioso para la mejora genética de múltiples caracteres de los cultivares de sandía y calabacín actuales (Wu et al., 2019).

Para el desarrollo del TFM se han seleccionado accesiones de estas dos colecciones de cucurbitáceas del banco de germoplasma del Departamento de Agricultura de los Estados Unidos (USDA), con un total 623 accesiones de *Cucurbita pepo*, 182 *Cucurbita maxima*, 53 *Citrullus amarus*, 621 *Citrullus lanatus* y 39 *Citrullus mucosospermus*. La selección se realizó en base a la información disponible en GRIN-GLOBAL (USDA, 2021), priorizando la disponibilidad de datos genómicos de la accesión y atendiendo a características disponibles de la semilla (tamaño y color). Se seleccionaron accesiones de *Citrullus* spp. y *Cucurbita* spp. que incluyeran diferentes tamaños, pesos y color para poder estudiar las regiones del genoma implicadas en estos caracteres.

3.2. Datos genómicos

3.2.1. Genomas de referencia

El desarrollo de las tecnologías de secuenciación masiva ha permitido la generación de genomas de referencia de utilidad para los análisis de asociación, QTL, selección de marcadores, etc. en el marco de la mejora de caracteres de interés agronómico. Entre las cucurbitáceas cultivadas se han desarrollado genomas de referencia para *C. melo*, *C. sativus*, *C. lanatus*, *C. pepo*, *C. maxima*, *C. moschata*, además de para otras especies de menor importancia económica. En el caso de *C. lanatus* se dispone de dos genomas de referencia de las variedades 97103 (de origen chino) y Charleston Gray (Figura 11), principal cultivar estadounidense (Levi et al., 2001) caracterizado por frutos de gran tamaño, forma elipsoidal, endocarpio rojo rosado y epicarpio verde claro que junto con el mesocarpio le confiere resistencia a las roturas y una larga vida postcosecha (Yong y

Gou, 2016). Además, esta variedad cuenta con resistencias a enfermedades radiculares, así como al marchitamiento por *Fusarium* y antracnosis. Charleston Gray es el genoma de referencia empleado para el mapeo de resultados derivado de las 1.365 accesiones secuenciadas en el que se basa el desarrollo de este TFM. Este genoma ha sido recientemente re-secuenciado y ensamblado *de novo* (Wu et al., 2019). Esta secuenciación se realizó empleando tecnología Illumina y dio lugar a un genoma de 396.4 Mb, que representa el 94.6% del genoma, anclado a 12 pseudomoléculas y 23.440 genes que codifican para proteína (Guo et al., 2013).

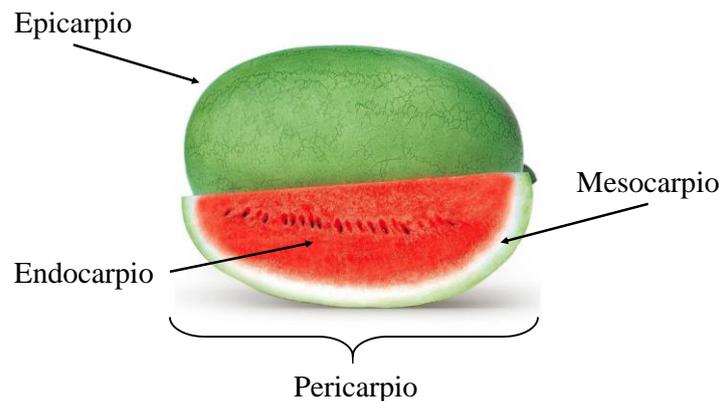


Figura 11. Partes del fruto de Charleston Gray. (Esquema de elaboración propia, imagen (Seed gallery, 2021)).

Las especies del género *Cucurbita* cuentan con genomas duplicados y los dos subgenomas que han conservado un número similar de genes (Sun et al., 2017). El genoma de *C. maxima*, secuenciado sobre la variedad Rimu, tiene un tamaño de ~387 Mb ancladas en 20 pseudomoléculas y 32.076 genes codificantes (Sun et al., 2017). El genoma de *C. pepo* ha sido desarrollado sobre la variedad MUCU16, una variedad del morfotipo Zucchini de color verde. El ensamblado dio lugar a una secuencia de 263 Mb, que se estima cubre el 93% del genoma, anclado en 20 pseudomoléculas. En este genoma se han identificado por técnicas bioinformáticas 28.870 genes condificantes (Montero-Pau et al., 2018).

3.2.2. Secuenciación de accesiones y obtención de variantes

Este trabajo pretende emplear la información genómica disponible de estos dos géneros para estudiar qué regiones genómicas son relevantes en caracteres de interés de la semilla. Los datos genómicos provienen de CuGenDB (<http://cucurbitgenomics.org/>), teniéndose 1.516 accesiones de *Cucurbita* spp. de las cuales 830 pertenecen a *C. pepo* y 372 a *C. maxima* y 1.365 accesiones del género *Citrullus* genotipadas por *Genotyping by Sequencing* (GBS), todas ellas pertenecientes al banco de germoplasma del Departamento de Agricultura de Estados Unidos (USDA).

El genotipado por secuenciación (GBS, *Genotyping by Sequencing*) se basa en la Secuenciación de Nueva Generación (NGS) de datos genómicos obtenidos a partir de los lugares de corte de enzimas de restricción (RE). Por lo que el primer paso consiste en la creación de bibliotecas de secuenciación. Para ello el ADN genómico se somete al corte por enzimas de restricción. Se emplean una o dos enzimas con el propósito es reducir la complejidad del genoma y eliminar regiones repetidas o poco informativas, obteniendo así información suficiente a la vez que se reducen costes (Elshire et al., 2011). Por tanto,

la cobertura del genoma y las variantes que se detectan por GBS no cubren el 100% del genoma.

El ADN de cada una de las accesiones que se quieren analizar se somete a este procedimiento. El siguiente paso consiste en incorporar un adaptador a cada muestra, que actúa como código de barras permitiendo que las muestras (accesiones) puedan ser identificadas de manera individual y posibilitando el análisis simultáneo. Posteriormente se realiza una amplificación y secuenciación, obteniéndose las secuencias de ADN que serán analizadas mediante técnicas bioinformáticas. Esta técnica es adecuada para estudios de población, caracterización de germoplasma, selección genómica y mapeo genético (Elshire et al., 2011).

Las secuencias resultantes se someten a tratamiento informático para eliminar los adaptadores y mapear las lecturas derivadas contra un genoma de referencia. En el caso de las accesiones de *Citrullus* spp. el genoma de Charleston Grey (Wu et al., 2019). En el caso de las accesiones de *Cucurbita* spp. se han empleado los genomas de referencia, para las accesiones de *Cucurbita pepo* la v4 genoma de referencia sobre MUCU16 (Montero-Pau et al., 2018), para *C. máxima* el genoma es Rimu (v1) (Sun et al., 2017). Estos genomas están disponibles en <http://cucurbitgenomics.org/>. El siguiente paso es la obtención de variantes tipo *Single Nucleotide Polymorphism* (SNP). El llamado de variantes consiste en la comparación de las lecturas de cada accesión contra el genoma de referencia permite determinar si, en una posición dada, tienen el alelo del genoma de referencia o algún alelo alternativo. Esa información se vuelca en un archivo tipo *Variant Call Format* (vcf) en que se indican para cada accesión implicada en el análisis la posición de las variantes y el alelo (referencia o alternativo) y características de calidad. Este archivo incluye todas las variantes observadas entre las accesiones de análisis con su posición física en pares de bases (pb) y localización en los cromosomas. El resultado de este análisis se empleó en este TFM para estudiar variantes, regiones del genoma y genes implicadas en el control de distintos caracteres de la semilla.

El número de variantes obtenidas se presenta en la Tabla 5. El número de variantes depende del número de accesiones en análisis y de la variabilidad genética entre ellas. *Citrullus lanatus*, el genoma de referencia del análisis de *Citrullus* spp. Cuenta con 11 cromosomas, mientras que *C. pepo* y *C. maxima* cuentan con 20 cromosomas. En los tres casos se incluye un cromosoma adicional (cromosoma 0) en el análisis, resultado de de secuencias de difícil ensamblaje de los genomas. El mayor número de variantes se observa en *C. pepo* de las que se analizaron 830 accesiones, seguido por *Citrullus* spp. de las que se analizaron 1.365 accesiones y cuenta con menor número de cromosomas, y finalmente para *C. maxima* se observan 5.600 variantes que reflejan la variabilidad genética de 372 accesiones. El análisis inicial de las accesiones de sandía, de las especies *C. lanatus*, *C. mucospermus* y *C. amarus*, reveló una relación próxima entre *C. lanatus* y *C. mucospermus*, mientras que *C. amarus* se diferencia de las dos especies anteriores (Elshire et al., 2011; Guo et al., 2019).

Tabla 5. Variantes obtenidas del llamado de variantes entre las accesiones secuenciadas por GBS (Proyecto CucCAP).

Especie/Género de análisis	Variantes tipo SNP
<i>Citrullus</i> spp.	25.739
<i>C. maxima</i>	5.600
<i>C. pepo</i>	47.544

El trabajo de genotipado de las accesiones se está realizando en el marco del *Cucurbit Coordinated Agricultural Project* (CucCAP) y representa la variabilidad de estas cucurbitáceas contenida en el banco de germoplasma USDA, que incluye accesiones de distintas partes del mundo. En la Figura 12 se representa la distribución geográfica de las 1.365 accesiones de *Citrullus* spp. Como se muestra existe una amplia representación de accesiones de África, centro de origen del género (Wu et al., 2019), seguida por Asia oriental, el subcontinente indio (centros secundarios de diversificación), Europa occidental y Norteamérica.

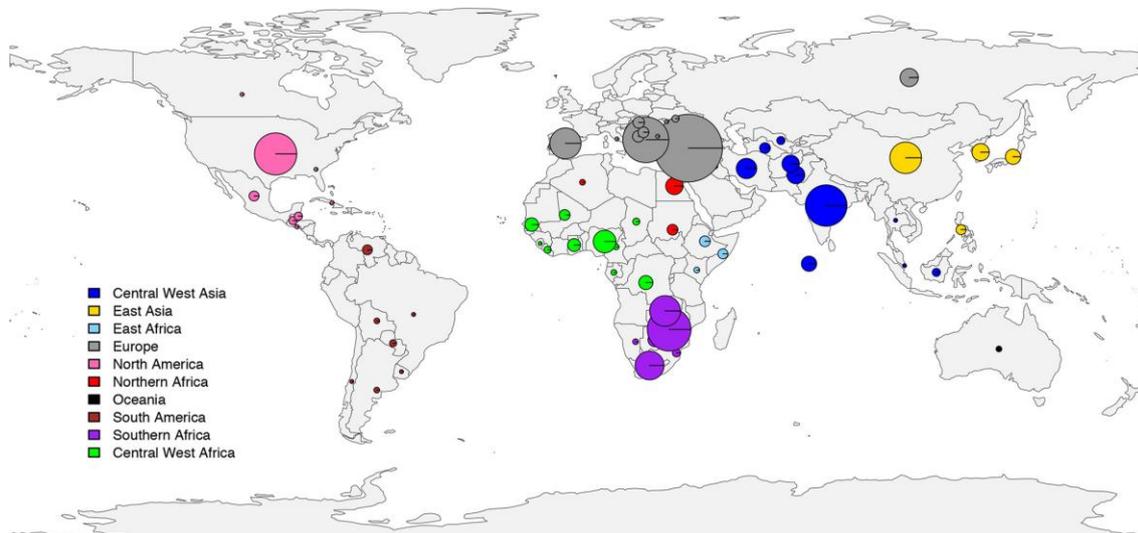


Figura 12. Distribución geográfica de 1.365 accesiones de *Citrullus* spp. del USDA. El diámetro del círculo es proporcional al número de accesiones de cada país (Wu et al., 2019).

Las accesiones de *Cucurbita* spp. proceden en su mayoría del continente americano, centro de origen de este género (Esteras et al., 2012). Tanto para *C. pepo* como para *C. maxima*, la distribución de las accesiones genotipadas por continente (Tabla 6) muestra una amplia representación de accesiones americanas, seguidas de Asia, Europa y África.

Tabla 6. Distribución de las accesiones genotipadas por GBS de CuGenDB (Proyecto CucCAP).

Continentes	<i>C. pepo</i>	<i>C. maxima</i>
África	34	68
América del Norte	106	50
América del Sur	11	60
Centroamérica	106	1
Europa	238	70
Asia	331	123

3.3. Toma de datos y procedimientos utilizados

3.3.1. Caracterización de las diferentes accesiones de *Cucurbita* spp. y *Citrullus* spp.

Se fenotiparon al menos 6 semillas por accesión atendiendo a caracteres descritos por diferentes autores, tal y como se muestra en la Tabla 7. Se evaluaron los caracteres peso, color de testa y presencia de hilio, únicamente en las accesiones de *Citrullus* spp. El único parámetro cuantitativo analizado, el peso, fue evaluado tres veces usando seis semillas cada vez y se obtuvo su media aritmética. El peso fue evaluado con una balanza de precisión, mientras que la longitud y ancho de semilla fue evaluado empleando el programa ImagenJ que permite transformar los píxeles de una imagen en unidades métricas una vez determinada la escala.

Por otra parte, el color se analizó asignando un valor numérico diferente a cada tonalidad encontrada. En el caso de *Citrullus* spp., en cada estudio, los autores desarrollan su propia metodología de fenotipado, lo que dificulta la comparación de resultados entre experimentos (Weetman, 1937; Poole et al., 1944; Sachan y Nath, 1976; Nath y Khandelwal, 1978). Dado que la descripción fenotípica desarrollada por Poole et al. (1944) es el más detallado entre los estudios realizados anteriormente, nos hemos basado en ella. En cuanto a la descripción fenotípica de *Cucurbita* spp. no existen muchos estudios debido a la pequeña variabilidad de color de testa, aun así, algunos autores hacen diferencias en *C. máxima* entre color blanco y crema o amarillo (Shi et al., 2021).

Finalmente, el hilio fue valorado como ausencia o presencia, tal y como se muestra en la Tabla 7 (Poole et al., 1944). Las accesiones en las que se encontró mucha heterogeneidad para algún carácter fueron eliminadas del estudio para dicho carácter para evitar datos que falsen los resultados.

Tabla 7. Se presentan los valores fenotípicos observados para los caracteres color y presencia-ausencia de hilio.

Carácter	Especie	Escala de valor					
Color	<i>Citrullus spp.</i>						
		1 Crema	2 Marrón	3 Oscuro	4 Naranja	5 Rojo	6 Verde
	<i>C. maxima</i>						
			1 Blanco		2 Crema		
Hilio	<i>Citrullus spp.</i>						
			0 Ausencia		1 Presencia		

3.3.2. Análisis de asociación del genoma (GWAS)

Analizado el fenotipo y con los datos de genotipo de cada accesión disponibles (sección 1.2), se realizó un análisis GWAS mediante la utilización de un modelo lineal general (GLM) que considera la estructura poblacional en el análisis de asociación o un modelo lineal mixto (MLM) que considera tanto la estructura poblacional (matriz Q, MDS) como la relación (matriz de parentesco, *kinship*) para cada carácter y género/especie utilizando el software Tassel 5 (Bradbury et al., 2007). Ambos métodos estiman el p-valor de cada SNP individual en el análisis y se compara con el p-valor esperado. La elección de uno u otro modelo para análisis posteriores vino determinado por el mejor ajuste de la distribución de los p-valores esperados y observados representado en un gráfico tipo QQ. En la Figura 13 se puede observar tres gráficos QQ con los tres casos considerados para la evaluación de los datos ajustado, subestimado y subestimado. Un resultado subestimado deriva en la identificación de menos variantes de las esperadas asociadas a un carácter (falsos negativos) y un resultado sobreestimado en la inclusión de falsos positivos en el análisis.

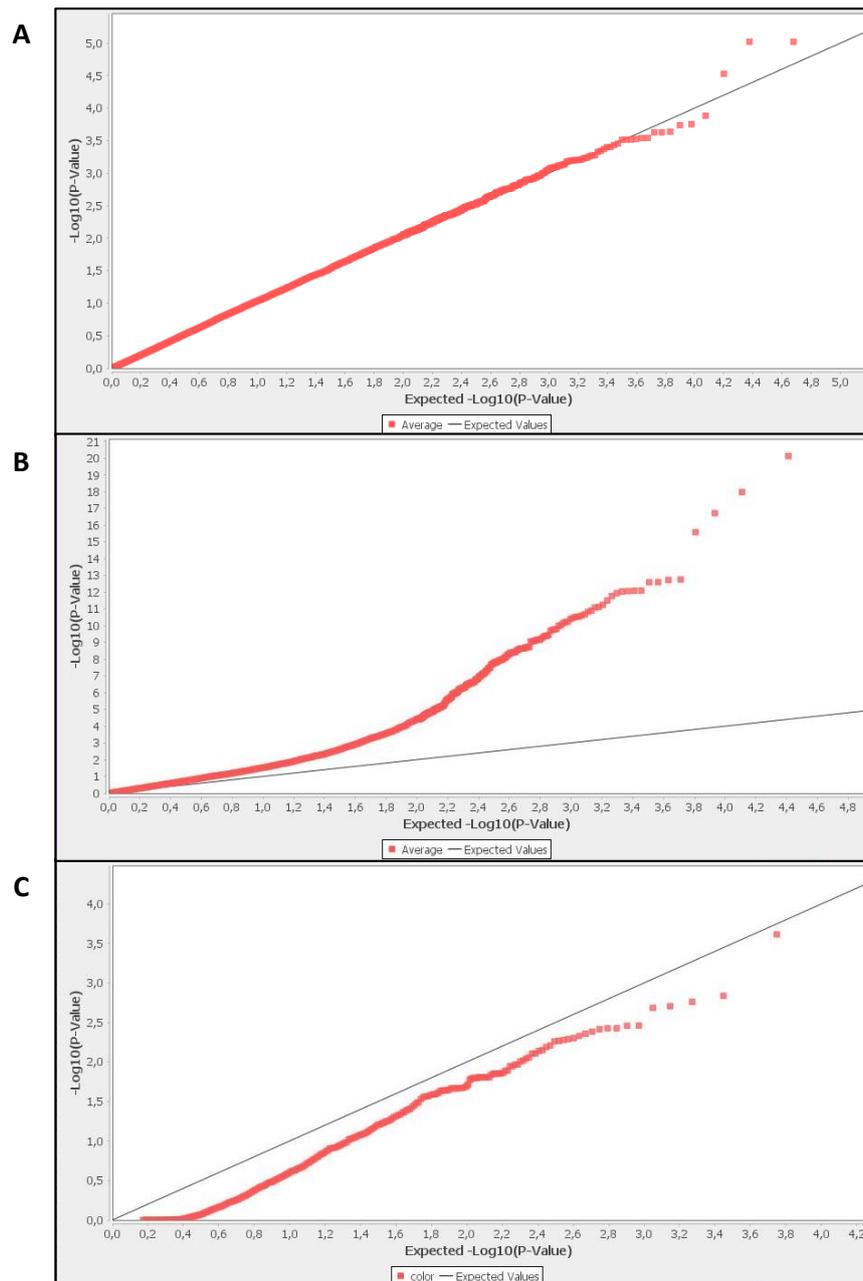


Figura 13. Representación de gráficos QQ que se pueden encontrar en el análisis GWAS con los métodos GLM y MLM y que permitieron elegir entre estos dos métodos de análisis. Para ellos se eligió el método con mejor ajuste en el gráfico QQ: **A.** Datos muy bien ajustados. **B.** Sobreestimación del impacto de los SNPs sobre el carácter. **C.** Subestimación del impacto de las variantes sobre el carácter.

El GWAS es una prueba estadísticas múltiple (para cada SNP se realiza una prueba independiente de asociación) por lo que la significación de este análisis alcanza una nueva dimensión en comparación con las pruebas estadísticas individuales en las que se suele rechazar la hipótesis nula H_0 con un valor $p < 0,05$ (Bonhomme y Jacquet, 2019). Los límites de significación se estiman según el método de Bonferroni como α/N , donde α es el nivel de significación y N es el número de test independientes que corresponde con el número de SNPs en análisis.

En este TFM se asume el p-valor sin ajuste y se estima, en base al ajuste observado en el gráfico QQ, qué método, GLM o MLM, presenta un mejor ajuste. Posteriormente,

se representan los límites de significación ($-\log_{10}(0.05/N)$ y $-\log_{10}(0.01/N)$) para el 95 y 99% (Marees et al., 2018). Para ello los resultados de cada análisis realizado con Tassel 5 se exportan en formato tabla (tabulada). Estos resultados, asociados a un carácter, y un análisis GLM o MLM, incluyen para cada variante tipo SNP, el valor de p-valor observado, y los valores de aditividad (a), dominancia (d) y varianza explicada (R^2). El software R cuenta con diferentes paquetes de análisis para GWAS. El paquete “qqman” posibilita representar el resultado del análisis de asociación en un gráfico Manhattan, con los límites de tolerancia de cada análisis. El Manhattan es un gráfico de dispersión que representa el $-\log_{10}$ pvalor calculado para cada variante (SNP). En el eje X se representa la posición de cada SNP en los distintos cromosomas y el eje Y indica el nivel de asociación de cada SNP al carácter.

3.3.3. Análisis de variantes asociadas

Las variantes asociadas a cada uno de los caracteres analizados fueron empleadas para identificar regiones del genoma y genes con posible implicación en el carácter. Puesto que los datos genéticos proceden de GBS, que reduce la complejidad del genoma y no presenta una cobertura del 100%, es posible que las variantes con asociación no sean las responsables del carácter. En este caso presentarían asociación por su proximidad con las variantes que regulan el carácter, con las que presentarían desequilibrio de ligamiento.

Para estudiar el impacto de los SNPs con asociación se han seguido el siguiente procedimiento:

a. Localización de posibles genes candidatos

La localización de las variantes se realizó utilizando la información disponible en Cucurbitgenomic database (ftp://cucurbitgenomics.org/pub/cucurbit/GBS_SNP/watermelon/). En primer lugar, se buscó en el archivo tipo vcf la posición de los diferentes SNPs. Con la herramienta “Genome Browser” de cucurbitgenomics se localizó la posición de esa variante, que muestra asociación al carácter, dentro del genoma de referencia. Se analizó si los SNPs caían en región intergénica o génica y en tal caso si estaban localizados en región exónica, intrónica o en una región no trasladada (5'UTR o 3'UTR) (Figura 14).

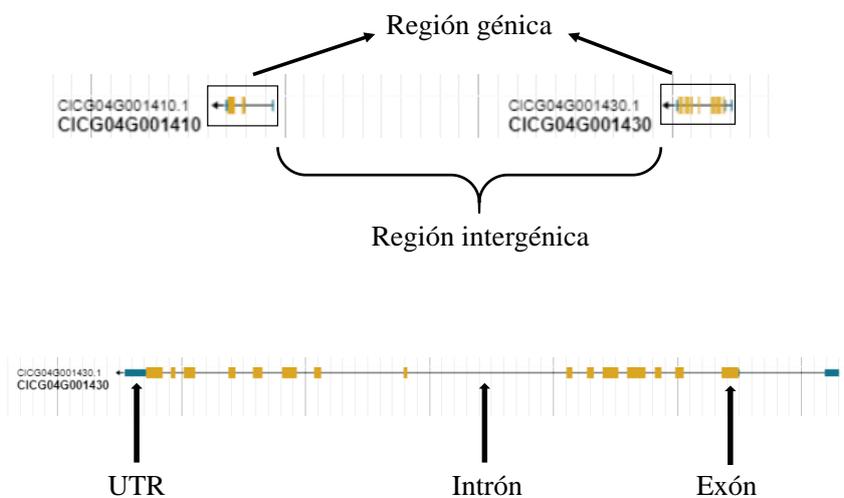


Figura 14. Regiones genómicas de los genomas de referencia de Cucurbitgenomic database (Elaboración propia a partir de Cucurbitgenomic database, 2021).

b. Evaluación del interés de las regiones detectadas

Para aquellas variantes situadas en regiones exónicas que pueden generar potencialmente un cambio de aminoácido se comprobó el impacto de la posible sustitución empleando los programas SNAP2 (<https://rostlab.org/services/snap2web/>) y Protein Variation Effect Analyzer (PROVEAN) (<http://provean.jcvi.org/index.php>). Ambos programas estiman el impacto en la proteína de las posibles sustituciones de aminoácido que variarán en función de la posición en la secuencia de la proteína y el tipo de aminoácido y determinan si la mutación es neutra o tiene impacto asignando un valor. Puesto que el archivo vcf, indica para cada variante los posibles alelos, y se conoce la posición de la variante respecto al genoma de referencia, es posible realizar la sustitución en la secuencia de ADN y comprobar, traduciendo a proteína la secuencia codificante, si se produce una sustitución de aminoácido. Para traducir a proteína la secuencia del gen se emplea el programa Expasy translate (<https://web.expasy.org/translate/>) y la información disponible en Cucurbitgenomic database (ftp://cucurbitgenomics.org/pub/cucurbit/GBS_SNP/watermelon/) que nos permitió obtener la región codificante de cada gen.

En caso de producirse sustitución, se analizó el impacto en la proteína con los programas SNAP2 Y PROVEAN. Ambas herramientas predicen si una sustitución de aminoácidos o un indel tiene impacto en la función biológica de una proteína (Choi et al., 2012; Choi y Chan, 2015). En el caso de PROVEAN, se considera significativo el impacto por debajo de -2,5, SNAP2 establece una escala de -100 a +100, y contra más positivo es el valor más impacto tendrá la sustitución de aminoácido.

En la Figura 15 se presenta un esquema del procedimiento para la localización en el genoma, anotación de su posición y determinación potencial del impacto en la proteína.

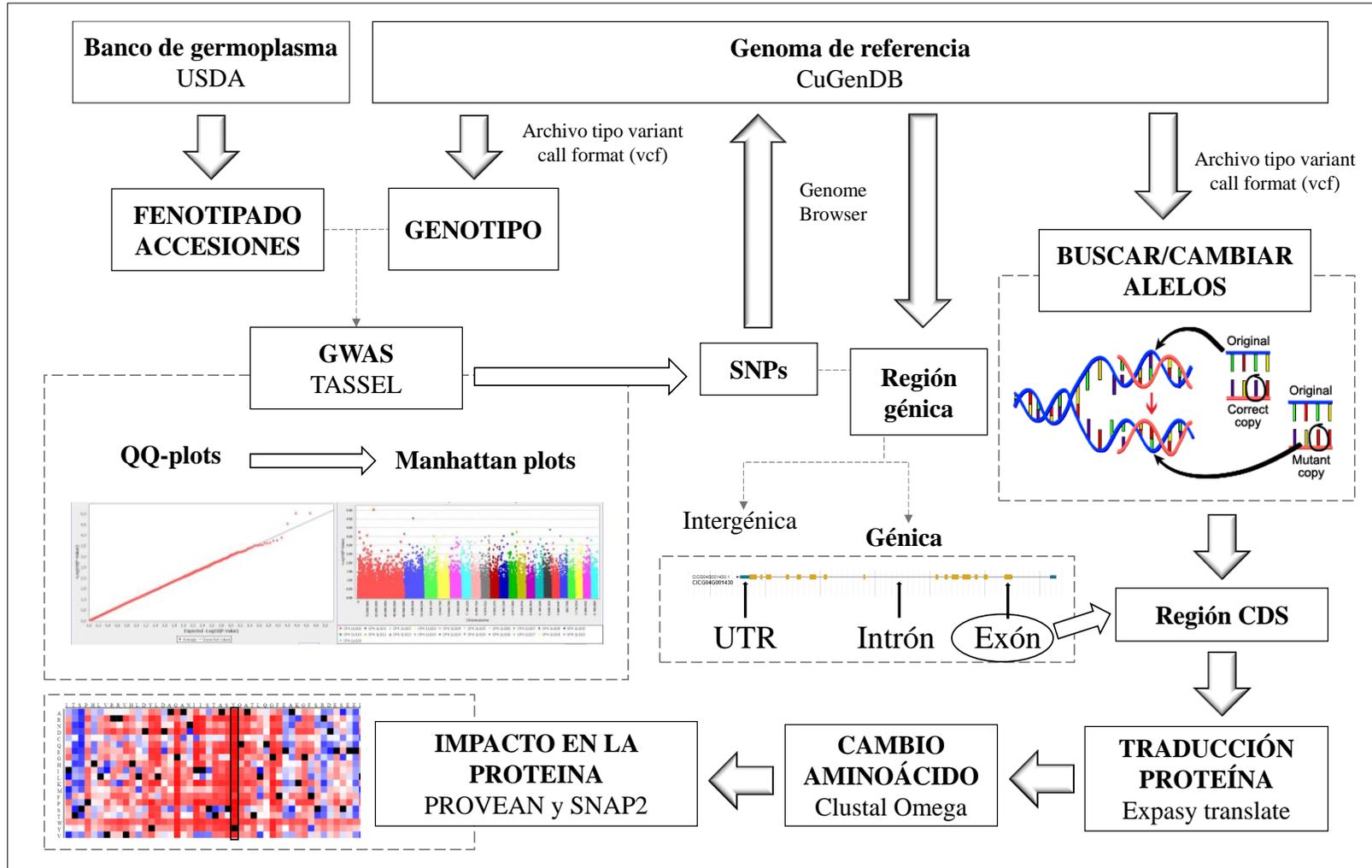


Figura 15. Esquema representativo del procedimiento seguido en este TFM para la realización de estudios de asociación en caracteres de semilla en cucurbitáceas. Se indica los pasos dados y los programas empleados para realizar el estudio de asociación, determinar la posición de las variantes con significación, establecer que alelos aparecen y valorar su impacto, en el caso de tener posición exónica, en la proteína. (Elaboración propia).

3.5. Análisis estadístico

Para el estudio de los resultados se empleó el paquete estadístico Statgraphics Centurión XVI que permite realizar comparaciones entre las diferentes especies. Se estudió la normalidad de los datos y en caso de no cumplir con la hipótesis de normalidad se realizó una transformación de estos para lograr que se ajustasen a una distribución normal. Una vez se cumplió con la hipótesis de normalidad, se realizaron análisis de varianza (ANOVA). Todos los datos se clasificaron por carácter y cada grupo se sometió a un análisis de único factor. La comparación de grupos fue realizada mediante un test de estadístico, Test de Rango Múltiple, de comparación de medias LSD (mínimas diferencias significativas de Fisher) con un nivel de significación $p < 0,05$ en el que se determinaron las medidas que eran significativamente diferentes.

RESULTADOS Y DISCUSIÓN



4. RESULTADOS Y DISCUSIÓN

4.1 Estudio de la variación de caracteres de semilla en los géneros *Citrullus* y *Cucurbita* (CUCURBITACEAE)

El peso de semillas es una estimación del tamaño. Este carácter es de importancia desde distintos puntos de vista, puede resultar fundamental para desarrollar una buena planta, por lo que se desea semillas grandes y bien formadas y con peso o, por el contrario, se puede seleccionar para tener semillas pequeñas, que apenas molesten durante el consumo, como es el caso de la sandía (Guo et al., 2020). El análisis de las 1518 accesiones para el carácter peso muestra que existe variación dentro de las especies y entre las distintas especies de cucurbitáceas estudiadas. La media del peso varió mucho entre especies siendo *C. maxima* la que mostró el mayor peso medio de semilla (3,01g/6 semillas), mientras que *C. pepo* presentó un peso medio intermedio (0,94 g/6 semillas) y las semillas con menor peso se observaron en las tres especies estudiadas del género *Citrullus* (Figura 6A). La mayor variabilidad en este carácter se observa en *C. maxima*, especie en la que la accesión PI_357905 alcanza un tamaño de 3,01 g/6 semillas y en la que la accesión con la semilla más pequeña PI_271771 tiene un tamaño de 0,10 g/6 semillas. A pesar de no encontrar diferencias significativas para el peso de la semilla en *Citrullus* spp. sí se observa variabilidad entre las especies por lo que es adecuado para análisis posteriores de GWAS.

Además del peso de las semillas, las 1518 accesiones fueron evaluadas para el color. Este fenotipado se realizó asignando un valor a cada tono de color de la testa (Tabla 6). Este carácter no es variable para *C. pepo* o *C. mucosospermus*, pero si se observó variación en *C. maxima*, *C. lanatus* y *C. amarus* (Figura 16B). Las semillas de *C. pepo* y *C. mucosospermus*, presentaban un color crema a blanquecino sin variación entre las accesiones.

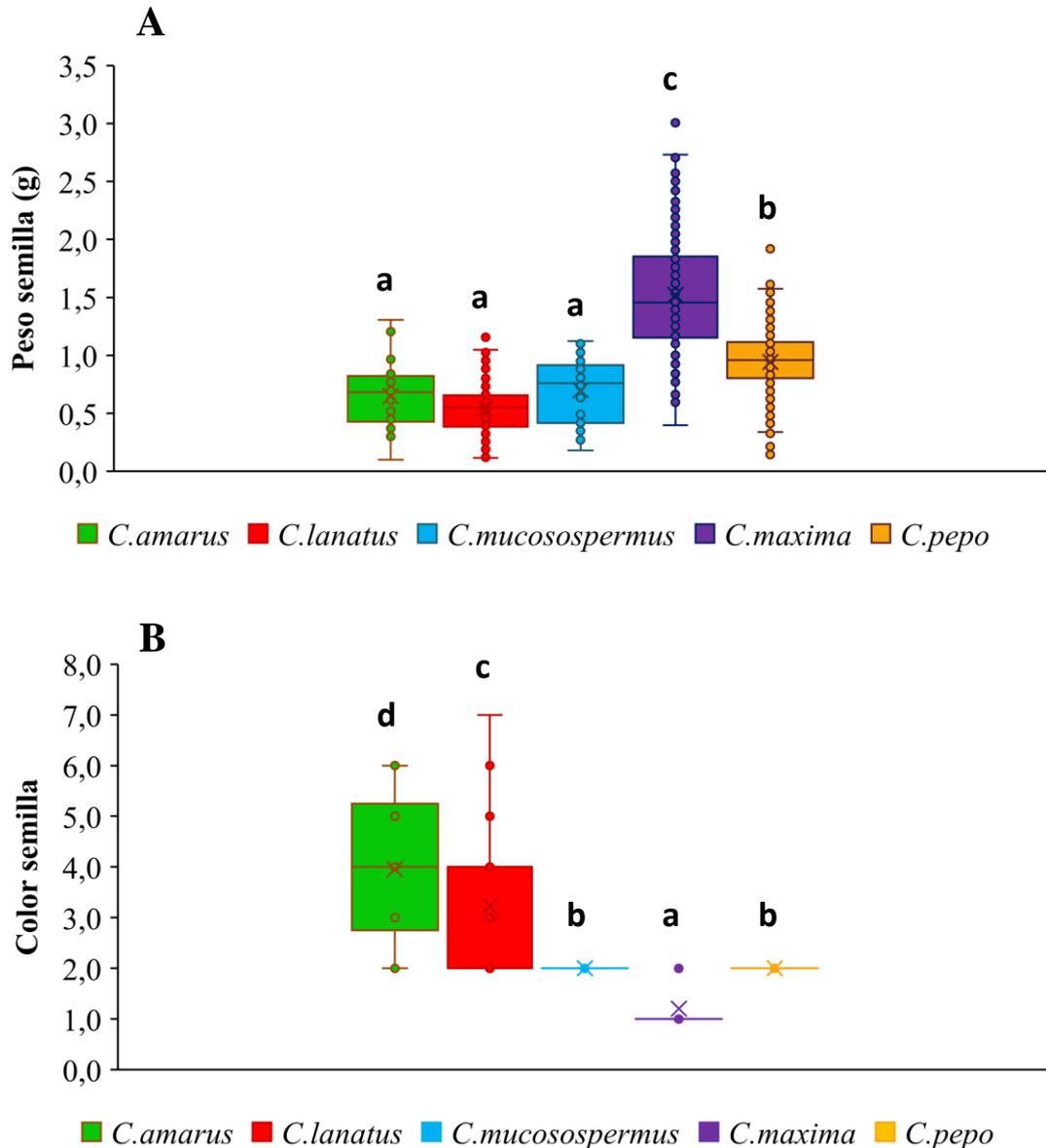


Figura 16. Cada caja de bigotes representa la mediana y los cuartiles de (A) peso de semilla. (B) Color de las semillas.

Letras diferentes indican diferencias significativas ($p \leq 0.05$).

En la figura 17 se presentan los colores de semilla observados tanto en *C. lanatus*, como en *C. amarus* y *C. maxima*, las tres especies que varían para el color. Si nos centramos en el color de testa, en la especie *C. lanatus* (Figura 17A) podemos observar 7 categorías de color descritas, esta especie presenta 6 de ellas. Siendo el color marrón oscuro (263 accesiones) el más representado, seguido del color crema (205 accesiones) y el marrón (83 accesiones). Sin embargo, los colores más anómalos son el naranja (2 accesiones), verde (5 accesiones) y rojo (13 accesiones). Además, de las 620 accesiones de *C. lanatus*, 49 presentan variación dentro de la accesión, por lo que han sido eliminadas del análisis GWAS.

En *C. amarus* (Figura 17B), de las siete categorías de color descritas, presenta 5 de ellas. Entre los colores predominantes destaca el color crema (10 accesiones) y el color marrón oscuro (12 accesiones). Por otra parte, los colores menos frecuentes son el marrón,

rojo y verde (6, 4 y 4 accesiones respectivamente). Además, de las 53 accesiones de *C. amarus* descritas, 17 presentan diferentes categorías y fueron eliminadas del análisis GWAS para este carácter.

En *C. maxima* (Figura 17C), solo se presentan dos categorías de colores, blanca (23 accesiones) y crema (89 accesiones). Además, es la especie que más variaciones de color presenta dentro de una misma accesión, de 182 accesiones, 70 presentan diferentes categorías de color de testa y fueron eliminadas del análisis GWAS.

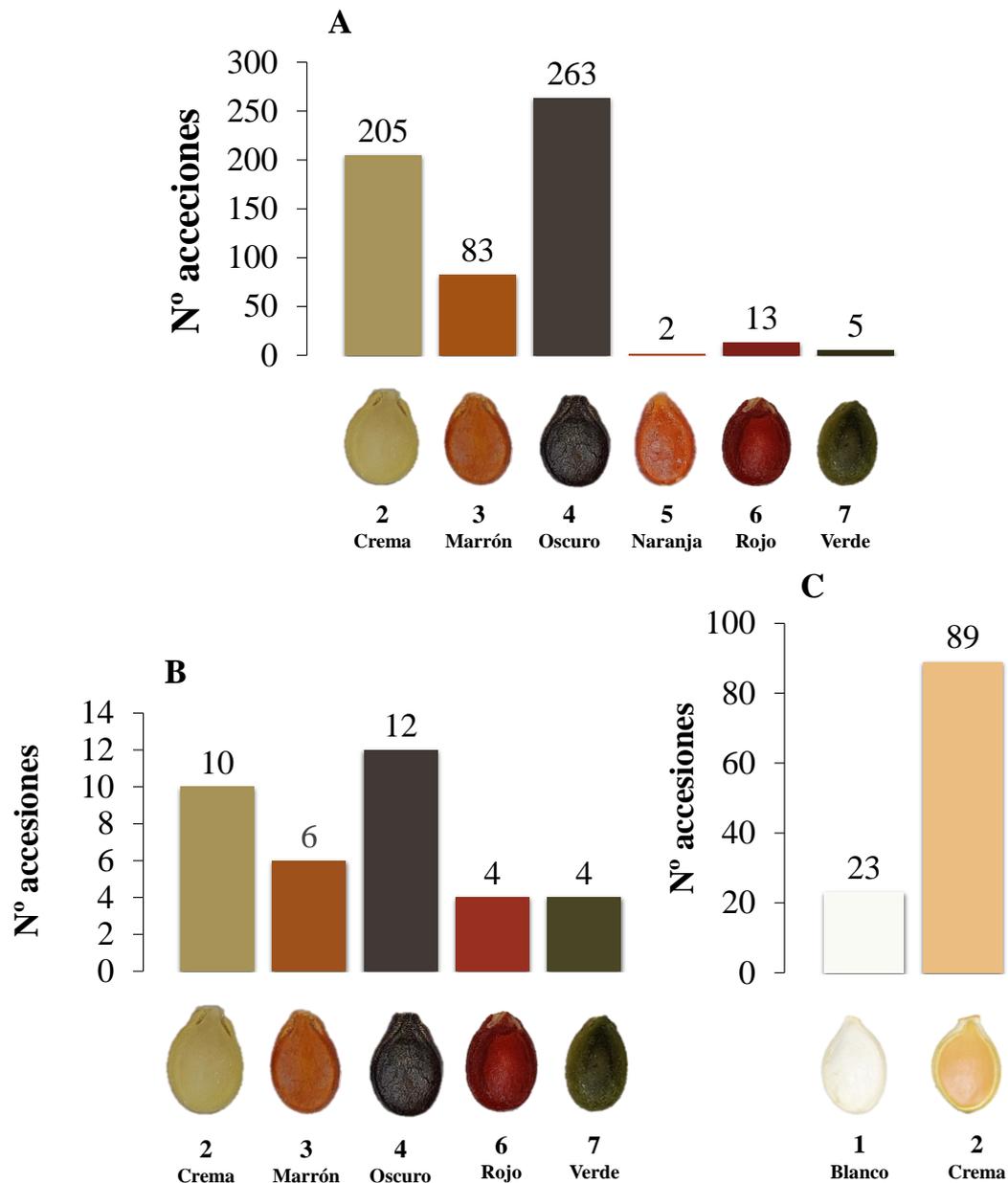


Figura 17. Color de testa de semilla de las diferentes accesiones. **A.** *C. lanatus*. **B.** *C. amarus*. **C.** *C. maxima*.

Además del peso, y el color de la semilla se ha analizado el hilio, sólo presente en semillas de las especies del género *Citrullus* (*C. lanatus* y *C. mucosospermus*). El hilio es una cicatriz en la cubierta de la semilla que se produce por la fijación durante su formación a la pared del ovario, es decir, muestra el lugar de sujeción al fruto, presente en algunas especies y se cree que puede tener funciones relacionadas con la desecación de la semilla

o su rehidratación previa a la germinación (Crang et al., 2018). Este carácter se ha evaluado como presencia-ausencia, por lo que no se muestran datos de variabilidad entre las especies del género, ya que en las dos especies había accesiones con presencia y con ausencia de hilio. En la figura 18A se muestran los resultados del número de accesiones con presencia y ausencia en las especies de *Citrullus*. De las 713 accesiones de *Citrullus*, 592 no presentan hilio y 63 sí lo presentan, siendo *C. lanatus* y *C. mucosospermus* las únicas especies que lo presentan. En el caso de *C. lanatus* (Figura 18C) predominan las semillas sin hilio en un 94% (585 de 621 accesiones) y solo un 5% presentan hilio (31 de 621 accesiones). En *C. mucosospermus* (Figura 18B) el 82% de las accesiones presenta hilio (32 de 39 accesiones) y tan solo el 18% no lo presenta (7 accesiones de 39 accesiones).

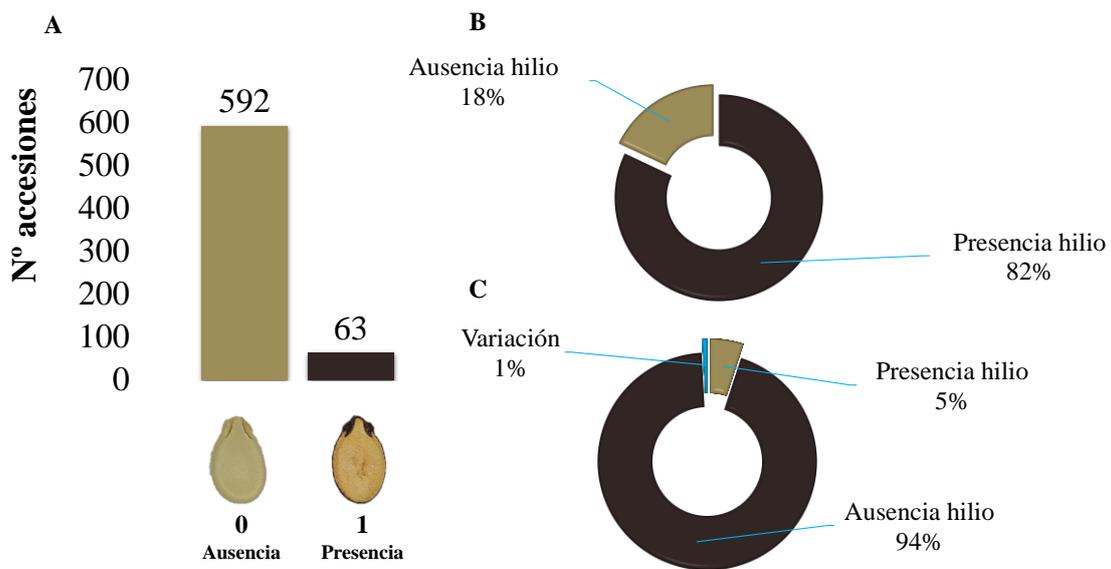


Figura 18. Presencia-ausencia de hilio en las diferentes accesiones de *Citrullus* spp. en: A. *Citrullus* spp. B. *C. lanatus*. C. *C. mucosospermus*.

En la Figura 19A se presenta el estudio de la correlación de los distintos caracteres para *Citrullus* spp. En particular de la presencia-ausencia de hilio con el peso y con el color de la semilla. La presencia de hilio no está condicionada por el peso de la semilla, al menos en las categorías de 0,6, 0,9 y 1,2 g, sí en la categoría de tamaño más pequeño (0,3 g) en la que todas las accesiones muestran ausencia de hilio. Sin embargo, si correlacionamos la ausencia o presencia de hilio con el color de semilla (Figura 19B), podemos observar una clara tendencia de color crema de testa en las semillas que presentan hilio y una variación de colores en las semillas en las que no está presente esta característica.

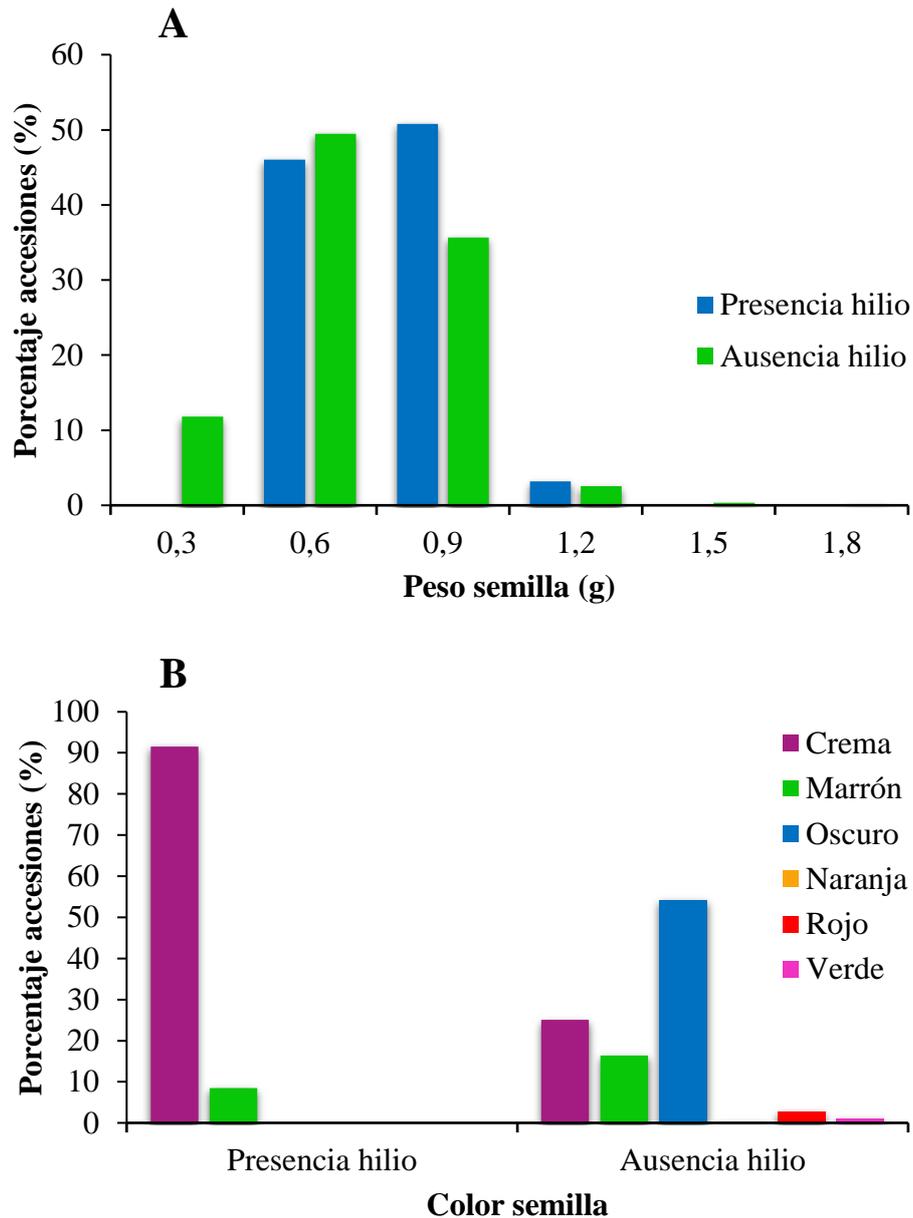


Figura 19. Porcentaje de accesiones con presencia y ausencia de hilio. **A.** Peso semilla en gramos. **B.** Color de testa de semilla.

Si por otra parte correlacionamos la presencia o ausencia de hilio con el lugar de origen, encontramos que las semillas con presencia de esta característica se localizan en su mayoría en Oriente medio, Sur de Europa y Sur de África (Figura 20A) representando un 85% de las accesiones. Además, el color de estas semillas en los diferentes lugares es predominantemente color crema, en todos excepto en América del Sur donde son de color marrón (Figura 20B). Las semillas de color marrón también se observan en accesiones procedentes de África Occidental y Oriente medio, 33,3% y un 6,06% respectivamente.

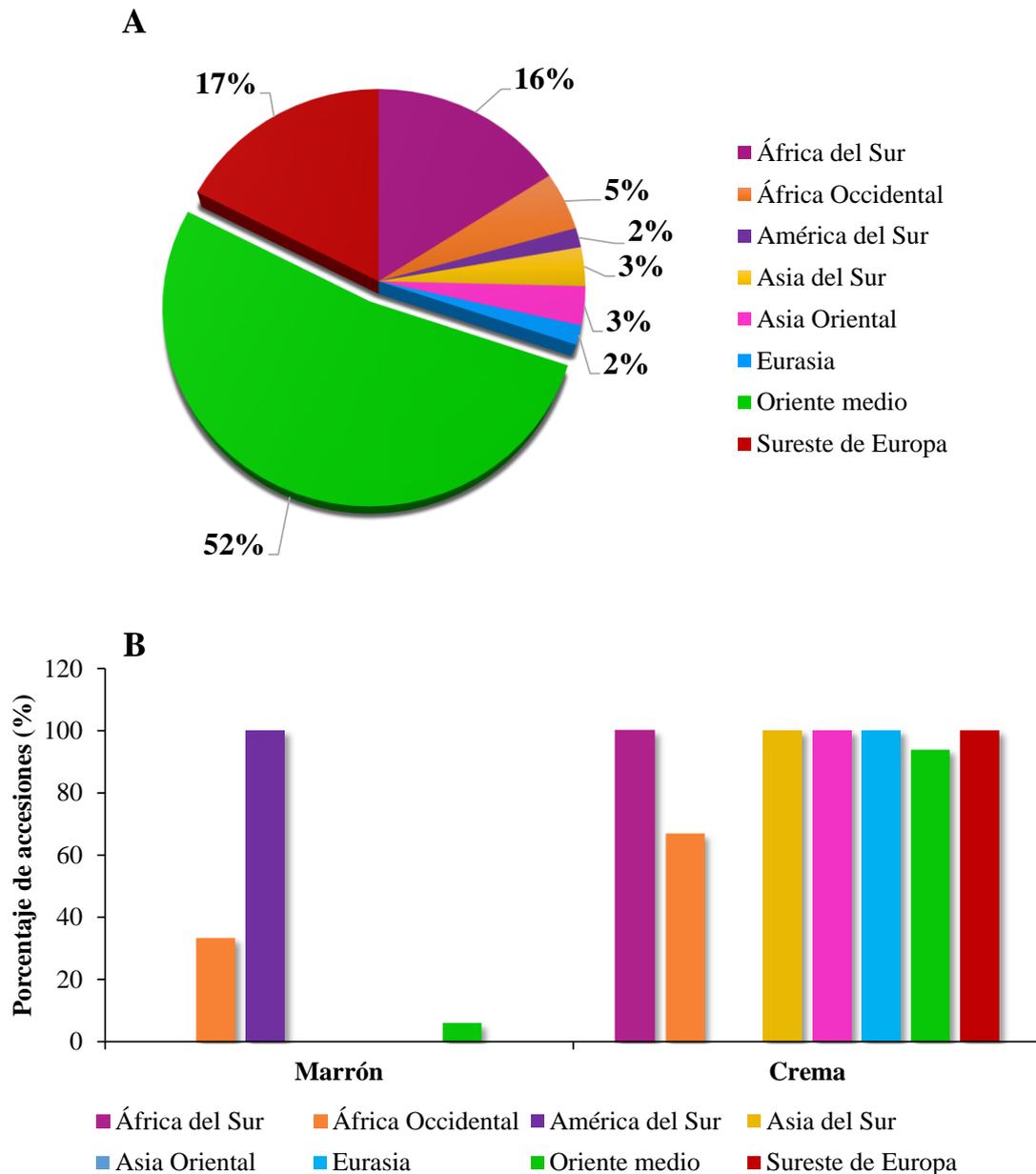


Figura 20. Porcentaje de accesiones que tienen hilos. **A.** Distribución mundial de semillas con hilos. **B.** Distribución mundial de los colores de semilla predominantes observados en *Citrullus spp.*

En este trabajo se han caracterizado un total de 9108 semillas, de 713 accesiones del género *Citrullus* y 805 accesiones del género *Cucurbita*. Dentro de *Citrullus spp.* se han caracterizado semillas de *C. lanatus*, *C. mucosospermus* y *C. amarus*. *C. lanatus*, es la especie más conocida, cultivada mundialmente y originada en Sudán y Egipto. Incluye formas silvestres, asilvestradas y cultivadas (Paris, 2015). La selección realizada en esta especie se ha orientado principalmente a la calidad del fruto, no obstante, existe interés por la semilla. En la selección por el tipo de semilla, además de por el color, se ha trabajado por el tamaño, se desean semillas de mayor tamaño cuando la orientación es la producción de plantas, y semillas de poco tamaño o peso, como las microsemillas, si se desean para consumo (Guo et al., 2020). Además de esta especie, *Citrullus* incluye otras seis especies diploides, entre ellas la sandía egusi (*C.mucosospermus*) que tiene importancia regional es originaria de África occidental subsahariana y se cultiva por el

aceite de sus semillas (Jarret et al., 1997; Achigan-Dako et al., 2006; Dahl-Jensen et al., 2011; Akusu y Kiin-Kabari, 2015). *C. amarus* (cidro, tsamma) es originario del sur de África y se cultiva por su pulpa comestible y dura, que suele cocinarse o encurtirse (Bush 1978). Este género despliega una gran variabilidad en distintos aspectos como el color y la presencia del hilio.

El género *Cucurbita* diversificó en el continente americano, dando lugar a múltiples especies. Es uno de los géneros más diversos de la familia. La mejora genética de *Cucurbita* se ha centrado principalmente en el rendimiento, la resistencia a las enfermedades y la mejora de la calidad de los frutos maduros e inmaduros (Paris, 2017). Sin embargo, *Cucurbita* spp. han demostrado valor en la producción de semilla para aceite y con contenido antioxidante (Teppner, 2004; Lelley et al., 2010; Martínez et al., 2021). En este trabajo se muestra variación en el tamaño de la semilla en *C. pepo* y *C. maxima*. De hecho, aunque no se ha podido correlacionar el tamaño con los morfotipos existentes en las dos especies, debido a no tener esa información, es posible que exista una relación. Sin embargo, los morfotipos en *C. pepo* han sido establecidos principalmente debidos a caracteres del fruto que están entre los más variables de las plantas (Duchesne, 1786; Naudin, 1856; Paris, 2017). Aunque no se ha encontrado variación en el color de la semilla de *C. pepo*, su tamaño si varía y se ha descrito que varía desde 8 mm a 25 mm en longitud (Paris, 2001; Paris et al., 2015). La mejora, principalmente realizada en la subespecie *pepo* ha inducido un incremento del tamaño de semilla, sobre todo comparado con la subespecie *texana* (Paris, 2017). *C. maxima* también es muy variable, en el pasado restringida a Sudamérica antes de la llegada de los europeos (Cutler y Whitaker, 1961; Merrick, 1995; Ferriol y Pico, 2008), se introdujo en Europa a principios del siglo XVI (Paris, 2017). Al contrario de lo observado en *C. pepo*, además del peso de la semilla se ha observado variación en el color en las semillas de esta especie. Estos resultados coinciden con lo observado anteriormente, y aunque las semillas suelen ser blancas, son anaranjadas en algunos cultivares e incluso marrones en cultivares con consumo de semilla (Paris, 2017). Además, tienden a ser bastante grandes (Paris, 2017), tal y como hemos observado en este TFM.

El tamaño y el color de la cubierta de la semilla son caracteres de mejora de la calidad de la semilla para varias especies cultivadas como el caupí, la colza, el lino y *Arabidopsis* y sandía (Paudel et al., 2019; Dana et al., 2008; Zhang et al., 2008; Debeaujon et al., 2000). El color de la cubierta definitivo se observa cuando la semilla alcanza la madurez fisiológica, y se ha demostrado que la pigmentación de la capa de la semilla desempeña un papel importante en la latencia y la germinación de la semilla (Debeaujon et al., 2000; Ochuodho y Modi, 2008). Además, en distintas especies, el color es un parámetro de calidad, como ocurre en la judía (Yang et al., 2010). Se ha encontrado asociación entre la producción y el color de la semilla en especies como el sésamo, la judía y la soja (Zhang et al., 2013; Yang et al., 2010; Tiryaki et al., 2016). También se ha observado que el tamaño de la semilla y el color de la cubierta son características importantes para distinguir entre las variedades de semilla dura y de semilla blanda de *Vicia sativa* y veza común, siendo el tamaño en las líneas de semilla dura menor (negras) que en semillas blandas (color marrón claro) (Büyükkartal et al., 2013). El color y el tamaño también muestran asociación con el contenido de aceite y antioxidantes como flavonoides y antocianos en distintas especies como *Brassica*, *Glycine max*, *Citrullus* ssp. y *Cucurbita* spp., *Arabidopsis* (Bagheri et al., 2013; Levi et al., 2017a; Paris, 2017; Yu, 2013, Li et al., 2020). De hecho, en judía las variedades de semilla blanca contienen más contenido en antocianinas y proantocianinas (Ranilla et al., 2007).

Por todo ello, conocer fuentes de variabilidad de estos parámetros, en especies cultivadas como las cucurbitáceas, es importante para la mejora genética y es interesante su estudio. En este trabajo se presentan resultados que muestran variabilidad intra e interespecífica en dos géneros de las cucurbitáceas, *Citrullus* y *Cucurbita*, para el peso y para el color de la cubierta, estos resultados pueden ser de interés para la mejora de esta especie de cara a estudiar cómo se correlaciona con la producción y con la germinación.

El hilio regula la salida y entrada de agua de la semilla durante el proceso de desecación en leguminosas (Crang et al., 2018). El hilio se forma por la unión de la semilla al funículo y facilita la entrada en la semilla de agua y reservas durante su formación. En leguminosas se abre en presencia de aire y se cierra en presencia de agua facilitando el secado de la semilla si las condiciones son favorables. Cuando la semilla se expone al aire las células de las paredes del hilio se encogen y el poro se abre; si el aire es húmedo las células de la pared se hinchan y el poro se cierra (Crang et al., 2018). La germinación se produce debido a la exposición constante a agua y la descomposición de la cubierta.

La presencia-ausencia de hilio observada entre las accesiones de *C. lanatus* y *C. mucospermus* y su correlación con el color indica que en nuestras accesiones el hilio predomina en las especies de color crema, y puede ser interesantes para seleccionar en relación a otros caracteres valiosos como la capacidad de germinación y emergencia en condiciones de estrés abiótico o normales. La estructura de la cubierta y su color afecta a la absorción del agua, influye en la latencia y en la tasa de germinación en distintas familias de plantas como las Papilionaceae, Anacardaceas, Brassicas y las Fabaceas (Souza y Marcos-Filho, 2001; Tiryaki y Topu, 2014; Debeaujon et al., 2000; Li et al., 1999). De hecho las semillas de achicoria con color claro germinan peor que las semillas oscuras (Pimpin et al., 2002).

Finalmente, la distribución mundial de estos caracteres, color e hilio, muestra diferencias entre distintas regiones geográficas. El hilio, asociado a colores crema de semilla, se encuentra predominantemente entre accesiones recolectadas en Oriente medio, Sureste de Europa y África del Sur, representando sólo las de Oriente medio el 52% de las accesiones. Por tanto, parece que la presencia de hilio está asociada a regiones cálidas. Puesto que tanto el hilio, como el color se ha encontrado en otras especies regulando el movimiento del agua, puede ser que en estas regiones la selección natural o artificial haya favorecido la presencia dominante de estos caracteres para favorecer una rápida absorción de agua durante la germinación o, una rápida desecación en el proceso de maduración fisiológica de la semilla. Por lo que sería necesario realizar una investigación más detallada, ya que no hemos encontrado en la bibliografía datos en este sentido. Por otro lado, la concentración de accesiones con semillas de color marrón en el Sur de América puede ser debido a una selección para el carácter oscuro en esa región, puesto que el color de la semilla es un carácter de importancia comercial (Li et al., 2020).

4.2. Estudio de asociación del genoma (GWAS) para caracteres de interés en semilla

El análisis GWAS permite la identificación de regiones del genoma que regulan caracteres (QTLs) empleando la variabilidad natural completa de una especie o un género (Zargar et al., 2015). En este caso se han realizado tres análisis GWAS independientes para el carácter peso de la semilla, en las especies *C. pepo*, *C. maxima* y en género *Citrullus* (incluyendo las especies *C. lanatus*, *C. mucosospermus* y *C. amarus*). Dos análisis para el color de la semilla, ya que no se observa variación respecto a este carácter en *C. pepo*; y finalmente, uno para el hilio de la semilla en las especies del género *Citrullus* (*C. lanatus* y *C. mucosospermus*). El gráfico Manhattan representa la posición en los cromosomas de cada SNP del análisis, esto es, se muestra la distribución de SNPs a lo largo de los cromosomas y, en el eje “y”, el p-valor estimado. El número de cromosomas es por tanto 12, en el caso de *Citrullus* spp., y 21, en el caso de *Cucurbita* spp., ya que se ha incluido el cromosoma 0 en el análisis. La selección del método para explicar los resultados, GLM o MLM, se establece, tal y como se recoge en Material y Métodos, en base a los resultados del gráfico QQ que compara el p-valor esperado y el observado. Se han elegido los métodos que procuran un mejor ajuste y evitan la sobreestimación o infraestimación de las variantes (SNPs) (ver Material y Métodos). Ambos métodos consideran el efecto de la estructura de la población, además, el método MLM emplea en su estimación la matriz *Kinship* que pondera las relaciones entre accesiones. En todos los estudios de asociación realizados, excepto para el estudio del color en la especie *C. maxima* que se eligió GLM, el método seleccionado ha sido MLM. Los límites calculados con el método de Bonferroni identifican SNPs con asociación estadística ($\alpha=0,05$ y $\alpha=0,01$).

4.2.1. Peso de la semilla

Tal y como se observa en el primer apartado de Resultados y Discusión, se trata de un carácter variable dentro de las especies y entre especies. Por lo que el panel de accesiones es potencialmente apto para el análisis de asociación del peso de la semilla. Las accesiones que mostraban variación en la misma accesión no fueron incluidas en el análisis. El estudio que más asociación muestra es el realizado en *Citrullus* spp. que indica asociación entre el peso de la semilla y regiones del cromosoma 1, 2, 5, 7, 8, 9, 10 y 11 ($\alpha=0,05$) (Figura 21), y con el cromosoma 2 (p-valor=7,413) (S02_27388336) con $\alpha=0,01$ (Figura 21A).

El análisis permitió detectar una región genómica para este carácter en el cromosoma 00 de *C. pepo* ($\alpha=0,05$), además de este SNP destaca un segundo SNP en el cromosoma 01 (S01_9916739) con un p-valor = 4,527, si bien queda por debajo de límite significación ($\alpha=0,05$) (Figura 21C). El análisis revela un SNP en el cromosoma 9 de *C. maxima* (Cma_Chr09_6128751) con un p-valor de 4,475 también por debajo del límite de $\alpha=0,05$ (Figura 21B).

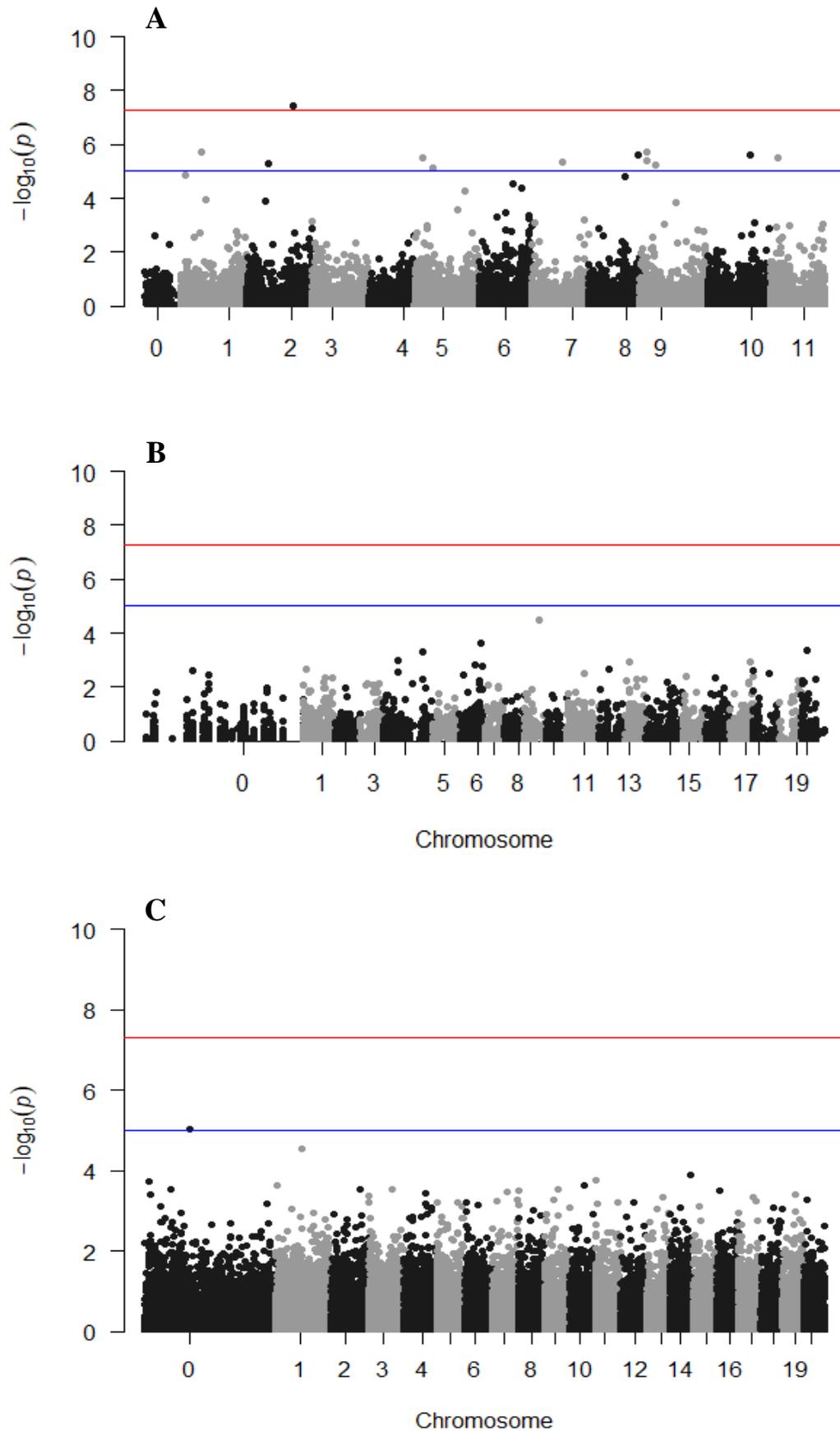


Figura 21. Gráficos Manhattan de estudio de asociación para el peso de semilla. **A.** *Citrullus* spp. **B.** *C. maxima*. **C.** *C. pepo*.

El peso y el tamaño de la semilla son caracteres importantes en sandía (Guo et al., 2020). Una semilla bien formada favorece una buena germinación y generación de planta, pero además puede tener otros intereses, como la producción de aceite o de antioxidantes (Paris, 2017; Martínez et al., 2021). En sandía, la selección se realiza en ese sentido y en el contrario. Para el consumo del fruto es deseable contar con semillas de pequeño tamaño o microsemillas que faciliten su deglución. Por lo que identificar regiones del genoma implicadas en el control del tamaño y peso de la semilla es de importancia para la mejora. De hecho, existen múltiples trabajos previos en este sentido.

El estudio de asociación realizado en las especies de *Citrullus*, *C. lanatus*, *C. mucospermus* y *C. amarus*, señala múltiples regiones, no obstante la más relevante se localiza en el cromosoma 2 (S02_27388336). Estudios previos realizados en *Citrullus lanatus* con poblaciones derivadas de cruces biparentales indican que existen un QTL en el cromosoma 2 para la longitud, la anchura y el peso de 20 semillas (Gao et al., 2020). Los tres caracteres parecen estar controlados por el mismo locus y, el QTL localizado explica el 35,5% de la varianza fenotípica observada en relación a la anchura de la semilla, el 43,5% respecto a la varianza de la longitud y el 50,1 % respecto a la variación en el peso. Además, en los tres casos, presentaban un *Logaritme of the odds* (LOD) superior a 15, siendo para el peso de 21,88 (Gao et al., 2020). En este mismo trabajo describen dos QTLs menores más para la anchura en los cromosomas 3 y 10, tres para la longitud en los cromosomas 5 y 7 y uno más para el peso de la semilla en el cromosoma 10 (Gao et al., 2020). Además de este trabajo, existen otros previos realizados con poblaciones segregantes derivadas de cruces biparentales. Estos estudios confirman la existencia de un QTL mayor en el cromosoma 2 para los tres caracteres. El trabajo de Prothro et al. (2012) estudia dos poblaciones e identifica hasta trece QTL relacionados con la longitud, anchura y peso de la semilla en los grupos de ligamiento (GL) 2, 4, 9 y 11. Tres QTL mayores para los tres caracteres se identifican el GL2 que corresponde al cromosoma 6 y un QTL con efecto mayor para la anchura de la semilla (25,6%) se localiza en el cromosoma 2 (GL9). La reciente revisión de Guo et al. (2020) indica que se han identificado un total de 61 QTLs para la longitud (20), la anchura (19), el espesor (3) y el peso de la semilla (19). El estudio de la posición cromosómica de estos QTLs ha derivado en 14 QTLs consenso que se posicionan en 1, 2, 3, 5, 6, 8, 11. Cuatro QTLs, localizados en los cromosomas 2, 6 (2) y 8 se han detectado en al menos dos estudios, los dos de cromosoma 6 se identificaron en poblaciones derivadas de sandías cultivadas y silvestres, pero el del 2 se detectó a partir de líneas cultivadas (Guo et al., 2020).

Además de en el cromosoma 2, en este TFM se han identificado SNPs asociados al peso de la semilla en el cromosoma 1, 2, 5, 7, 8, 9, 10 y 11. Lo que confirma que algunos de los QTLs identificados en poblaciones biparentales se detectan en la variabilidad natural, como los localizados en el 1, 2, 5, 8 y 11. A falta de un estudio más detallado, que escapa de este TFM, estos QTL podrían colocalizar con las regiones identificadas en GWAS. Es de destacar que la revisión de Guo et al. (2020) plantea dos QTLs claramente diferenciados al inicio y final del cromosoma 2, al igual que se identifica en este TFM dos SNPs en posiciones 12.482.692 y 27.388.336 pb del genoma de referencia Charleston Gray (Wu et al., 2019). Por lo que nuestros resultados confirman el potencial valor de los estudios de asociación para detectar regiones del genoma de interés para caracteres de semilla.

Al contrario de lo observado en sandía, en *C. maxima* no se detecta ningún SNP en relación al peso de la semilla con significación. No obstante, en el cromosoma 9 (posición 6.128.751 pb del genoma de referencia Rimu (Sun et al., 2017)) se identifica un SNP con

p-valor 4,475. Mientras que el análisis en *C. pepo* para este carácter indica que hay un SNP en el cromosoma 00 (S00_17194302) ($\alpha=0,05$) y un SNP en el cromosoma 1 (S01_9916739) con p-valor=4,527 por debajo del nivel de significación. A pesar de que la semilla no sólo afecta al rendimiento, sino que también puede alterar la calidad del fruto (Guo et al., 2020), la identificación de QTL para tamaño/peso de semilla en el género *Cucurbita* no ha sido tan extensa como en el caso de *Citrullus*. Se han identificado 4 QTL para la anchura de la semilla (Tan et al., 2013) y en *C. maxima* se empleó una población F2 para mapear 10 QTL localizados en 6 cromosomas (4, 5, 6, 8, 17 y 18), de ellos, 4 estaban implicados en la longitud, 4 en la anchura y 2 en el peso de la semilla. Los dos QTLs de mayor efecto se encontraron para la longitud y la anchura en el cromosoma 6, explicando el 38,6 y 28,9 % de la variabilidad (Wang et al., 2019). Ninguno de estos QTLs coincide con los SNPs localizados en el cromosoma 0 y 1 de *C. pepo* y en el 9 (para *C. maxima*) que hemos considerado más relevantes por su p-valor.

La falta de identificación de regiones asociadas puede ser debida, en el caso de *C. maxima* al bajo nivel de SNPs y la técnica de genotipado, pero también podría ser debido al panel de accesiones seleccionadas (182). Dado que en *C. pepo* la saturación de variantes es la mayor de los tres análisis, el número de SNPs no es limitante y puede ser que el panel de accesiones no sea lo suficientemente variable. Otra explicación es que la existencia de subpoblaciones o relaciones entre las accesiones impida la detección de variantes de interés. Fuera del marco de este TFM se harán análisis con distintos sets de accesiones y se comprobará si se identifican algún SNP. En este sentido indicar que el uso en *Citrullus* de distintas especies, *C. lanatus*, *C. mucosospermus* y *C. amarus*, puede suponer un efecto en la estructura de la población que dé lugar a falsos positivos. Para controlar este efecto, se ha seleccionado el método de análisis MLM. En un trabajo previo Wu et al. (2019) emplea indistintamente accesiones de *C. lanatus* y *C. mucosospermus* para GWAS en caracteres del fruto y resistencias a enfermedades con buenos resultados. Por último, la colocación de las regiones identificadas en este TFM para *Citrullus* spp., en particular la segunda del cromosoma 2, con trabajos previos sustentan la validez del análisis.

El color de la semilla es un carácter de interés para la mejora en algunas especies como la judía (Tiryaki et al., 2016; Zhang et al., 2013; Yang et al., 2010). En sandía hemos visto una clara diferenciación del color dependiendo de la región del mundo (las accesiones de Sur América son de color oscuro) y, además, el color de la semilla está determinado por el contenido de sustancias con valor antioxidantes como los flavonoides, y por el contenido en melatonina (Yu, 2013; Zhang et al., 2008). El color de la semilla de la sandía es importante, además de por su valor comercial, porque se asocia con la absorción de agua, la latencia de la semilla y la germinación (Mavi, 2010), también se asocia a características bioquímicas (El-Bramawy et al., 2008) y afecta a la apariencia de la carne del fruto (Wehner, 2008). El GWAS para el color de la semilla fue realizado en *Citrullus* spp. y *C. maxima*, ya que no hay variación de este carácter en *C. pepo*. Se detectó una variante de interés en el cromosoma 3 ($\alpha=0,05$) de *Citrullus* spp. (S03_60539010) (p-valor=6,015), si bien encontramos otra por debajo de $\alpha=0,05$ en el cromosoma 6 (S06_7303045) (p-valor=4,414) (Figura 22A). En cuanto al análisis realizado en la especie *C. maxima*, no se detecta ningún SNP con p-valor significativo. Las dos variantes (Cma_Ch11_10869324 y Cma_Ch11_10869735) con p-valor más alto del análisis (p-valor, 4,580 y 4,315, respectivamente) se localizan en la misma región del cromosoma 11 (Figura 22B).

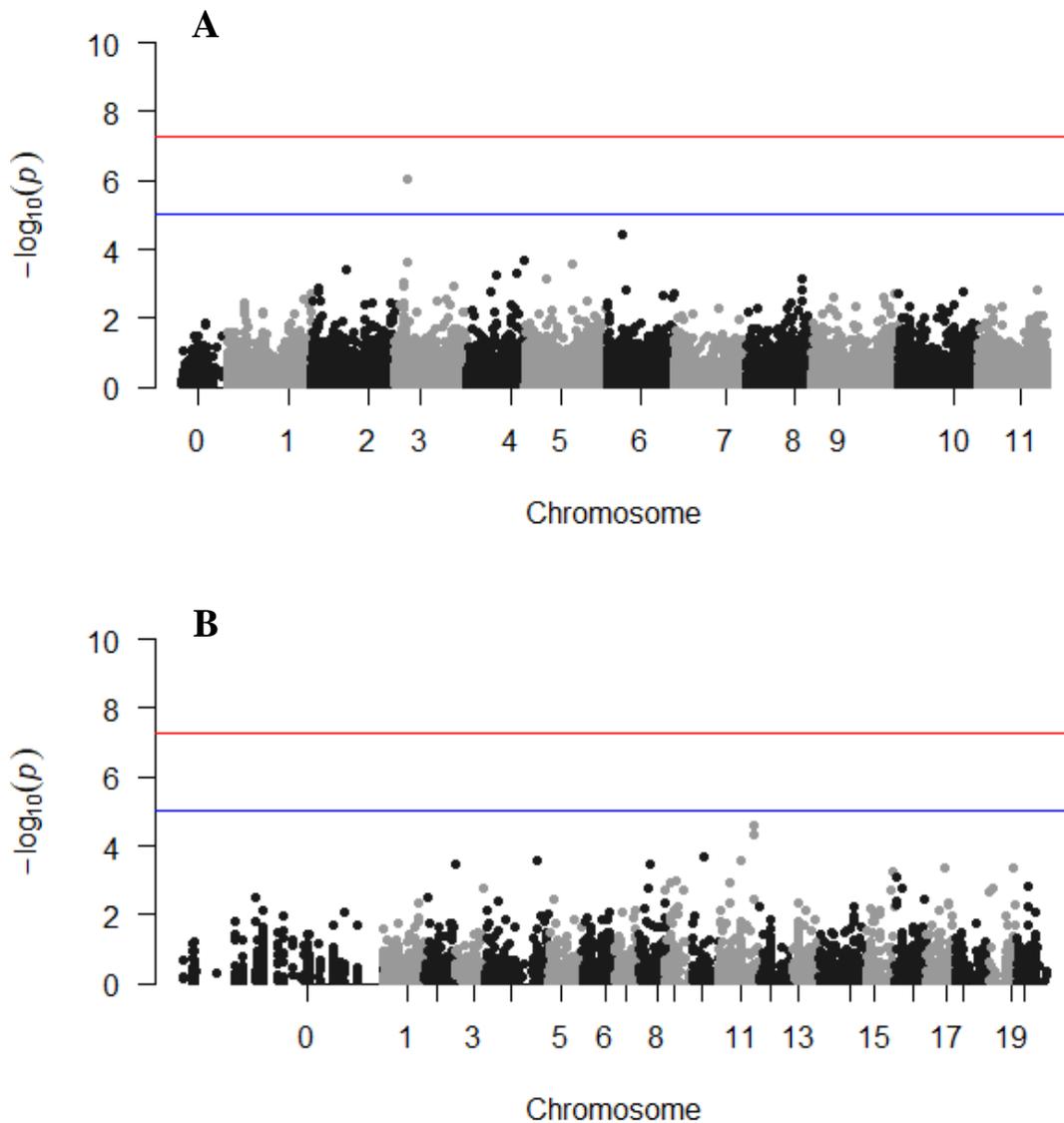


Figura 22. Gráficos Manhattan obtenidos en los estudios de asociación del color de semilla. **A.** Resultados GWAS para *Citrullus* spp. **B.** Resultados GWAS para *C. maxima*.

Se han realizado estudios genéticos del color de la cubierta de la semilla en melón y en sandía (Li et al., 2018; Li et al., 2020; Paudel et al., 2019; Ning et al., 2018). En el melón, el color blanco de la testa es dominante respecto al amarillo y el marrón (McKay, 1936). En la sandía, el color de la semilla depende de cuatro loci (*R/r*, *W/w*, *T/t* para rojo, blanco y bronceado, respectivamente; *D/d* para producir testa negra y punteada) que han sido identificados en los cromosomas 3, 5, 6 y 8 (Li et al., 2018; Li et al., 2020; Paudel et al., 2019). La interacción de estos genes da lugar a los seis colores base: negro (*RR TT WW*); variado (*RR TT ww*); bronceado (*RR tt WW*) blanco con punta bronceada (*RR tt ww*); rojo (*rr tt WW*); y blanco con rosa (*rr tt ww*) (Kanda, 1931; McKay, 1936; Poole, 1944).

En este TFM se ha identificado asociación del color con una única variante en *Citrullus* spp. ($\alpha = 0,05$) situada al inicio del cromosoma 3 (posición: 6.053.901 pb). Esta variante se encuentra próxima (297.356 pb) a la región de 70,1 kb identificada en una población de RILs entre las posiciones 5.686.151 y 5.756.365 pb (Li et al., 2020). Los

autores indican que está asociada a el color de la testa negro y encontraron en la región un gen candidato, *Polifenol oxidasa (polyphenol oxidase, PPO)*, implicado en la síntesis de melatonina. Se sugiere que una mayor cantidad de polifenoles favorecen una mayor síntesis de melatonina catalizada por la PPO dando lugar a un color negro de la cubierta (Li et al., 2020). Por otro lado, Paudel et al. (2019) mapearon los loci *R*, *T^l*, *W*, y *D* en los cromosomas 3, 5, 6, y 8, usando secuenciación por GBS y QTL-seq de dos grupos de plantas (uno con semilla de color negro punteado y otro de color verde) de una población segregante para el color de la semilla. Localizan al inicio del cromosoma 6 (marcador más próximo a 9 cM se sitúa en 7.076.766 pb) la posición del gen *W/w* responsable del color blanco. Lo que da relevancia al marcador identificado en nuestro panel de accesiones en el cromosoma 6, posición 7.303.045 pb, si bien este SNP no resulta significativo en nuestro análisis.

Si bien no es significativo, se han identificado dos SNPs en el cromosoma 11 de *C. maxima* con p-valor superior al resto de variantes del estudio. En esta especie el color de la cubierta de la semilla parece estar controlada por un único gen recesivo denominado *white seed coat (wsc)* (Shi et al., 2021). El estudio con una población F2 localiza la región candidata en el cromosoma 15 (Shi et al., 2021).

El tercer carácter analizado en este TFM es la presencia de manchas negras en la cicatriz del hilio se encontró en algunas accesiones de *Citrullus* spp. Tal y como se muestra en la sección 1 está asociado al color de la semilla. Trabajos en distintas especies indican que el hilio y el color de la semilla influyen la latencia y la germinación (Mavi, 2010; Mattiolo y Buscalioni, 1892; Corner, 1951). De hecho, el hilio parece estar implicado en el balance de humedad durante la formación de la semilla (Crang et al., 2018). Tras realizar el análisis de asociación del genoma (GWAS), se encontraron dos posiciones genómicas en los cromosomas 4 (S04_4876169) y 10 (S10.1_1183562) ($\alpha=0,01$) asociadas a este carácter (Figura 23). Además de estas dos variantes se encontraron otras con $\alpha=0,05$, localizadas en el cromosoma 2 (S02_2945381 y S02_38007899), en el cromosoma 5 (S05_12685196 y S05_30250707), en el cromosoma 8, dos SNP localizados en las posiciones 15.074.562 pb y 15.075.809 pb del genoma de referencia Charleston Gray (Wu et al., 2019), en el cromosoma 9 (S09_11607084) y en el cromosoma 10 (S10_20127639).

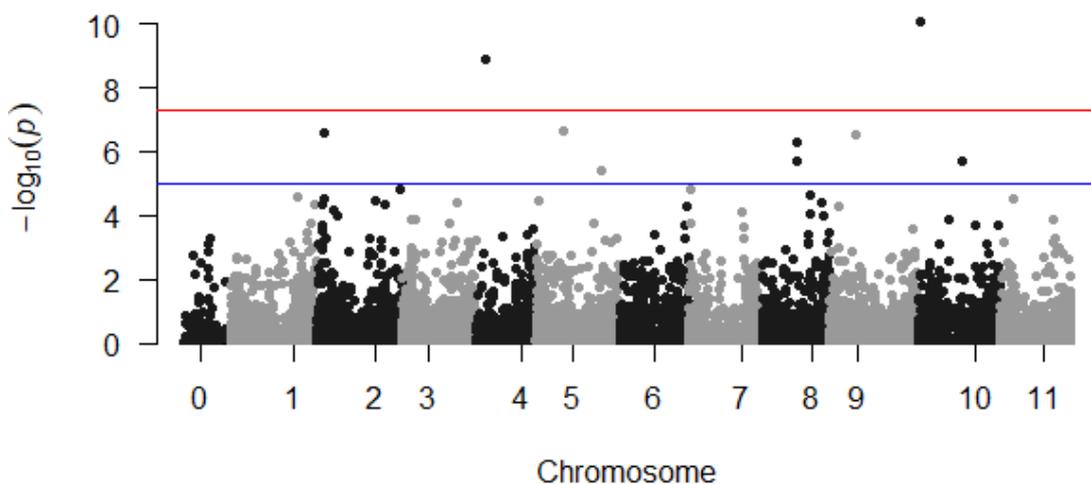


Figura 23. Gráfico Manhattan con la asociación de SNPs a la presencia y ausencia de hilio en *Citrullus* spp.

Aunque no hemos encontrado datos en referencia al estudio genético del hilio en *Citrullus* spp. este carácter se ha utilizado para hacer clasificaciones taxonómicas en sandía (Sahin y Polat, 2020). Existen trabajos en otras especies que apuntan a una función relacionada con el movimiento del agua durante la formación de la semilla o para facilitar el movimiento de gases y agua que conduce a la germinación y la emergencia de la radícula (Mattirolo y Buscalioni, 1892; Corner, 1951; Crang et al., 2018). Por lo que podría ser un carácter de interés para estudiar estreses abióticos durante la germinación. En la siguiente sección se discute el interés de este carácter desde el punto de vista de los genes en los que se han encontrado variantes.

4.3. Estudio del impacto de las variantes y de regiones del genoma con asociación para los caracteres de la semilla en *Citrullus* y *Cucurbita*

Las variantes detectadas en el análisis de asociación pueden ser las responsables de la variación del carácter, pero puede darse el caso de que la asociación observada se deba a que se sitúen próximas y ligadas a otros SNPs, que sean las mutaciones verdaderamente responsables de la variación del carácter. Esto es así aún más debido a la técnica de secuenciación empleada, GBS. Como se indica en la introducción el método de secuenciación *Genotyping By Sequencing* reduce la complejidad de los genomas usando enzimas de restricción y acotando la secuenciación a las regiones de corte de esas enzimas (Elshire et al., 2011). Por esta razón el genoma no está totalmente secuenciado y es posible perder mutaciones causales.

Teniendo en cuenta esta consideración, en esta sección se estudia las variantes con mayor asociación en cada uno de los análisis realizados. Puesto que las variantes con significación se han encontrado en los GWAS de *Citrullus*, se estudian los SNPs asociados al peso, color y presencia-ausencia de hilio en entre las accesiones de *Citrullus*. Con el objetivo de no realizar un trabajo engorroso y poco claro se han seleccionado únicamente las variantes con mayor p-valor de cada análisis, es decir, se han eliminado las variantes asociadas al peso y el hilio con nivel de significación por encima de $\alpha=0,05$, pero debajo de $\alpha=0,01$. Sí se ha añadido al análisis la variante del cromosoma 6 asociada al color por estar en una zona relevante según estudios previos (Paudel et al., 2019).

4.3.1 Localización de las variantes seleccionadas e identificación de genes

Tal y como se indica en Material y Métodos (Figura 16), las variantes encontradas asociadas con el programa Tassel5 al comparar genotipo y fenotipo se localizan en la tabla de resultados. En esta tabla se indica el cromosoma y posición del SNP respecto al genoma de referencia, el p-valor y otros datos como el porcentaje de la varianza explicada por ese SNP según el análisis. En la Tabla 8 se presentan los datos para cada variante analizada: alelos observados entre las accesiones, cromosoma, posición (pb), p-valor, localización y anotación. Con esta localización respecto al genoma de referencia (cromosoma y posición) se estudia en con la herramienta de cucurbitgenomics (<http://cucurbitgenomics.org/>) “Genome Browser” si la variante se sitúa en una región intergénica (entre genes) o en una región génica. Si se trata de una región génica, si se sitúa en la secuencia de un exón, un intrón o en 5' o 3'UTR (*untranslated región*). En caso de situarse en un exón se evalúa el putativo cambio de aminoácido en la proteína y el impacto de este cambio.

La posición física del SNP S02_27388336, asociado a la variación en el peso de la semilla de *Citrullus* spp., se localiza en 27.388.336 pb del cromosoma 2 del genoma de referencia Charleston Gray (Wu et al., 2019). En un intrón del gen anotado

CICG02G013450 (Tabla 8) que codifica para una Metal tolerance protein. Para el carácter peso de la semilla se localiza un SNP ($\alpha=0,05$) en el cromosoma 0 (S00_17194302) del genoma de referencia de *C. pepo* (Montero-Pau et al., 2018), este cromosoma es generado por la falta de consenso durante el ensamblaje del genoma y habitualmente es eliminado de los análisis. En cuanto al carácter color, en el estudio de asociación de este carácter en *Citrullus* spp. se han analizado dos variantes, situadas en el cromosoma 3 (posición 6.053.901 pb) y 6 (posición=7.303.045 pb) del genoma de referencia Charleston Gray (Wu et al., 2019). El SNP S03_60539010 (p-valor= 6,015) está situado en región intergénica y el SNP S06_7303045 (p-valor=4,414) en el segundo exón del gen *CICG06G006370* (Tabla 8) que codifica para la proteína Homocisteína S-metiltransferasa 3. Por último, las dos variantes más relevantes del estudio de asociación realizado en *Citrullus* spp. para el hilio ($\alpha=0,01$) se localizan en el cromosoma 4 y 10 del genoma de referencia. El SNP S04_4876169 (p-valor=8,888) está localizado en el decimocuarto exón del gen *CICG04G001430* y la variante S10_1183562 (cromosoma 10, posición 1.183.562, p-valor=10,055) se localiza en el tercer exón del gen *CICG10G001160*. Estos dos genes codifican para Glucose-6-phosphate 1-dehydrogenase y PHD finger family protein, respectivamente.

Tabla 8. Resumen de las variantes destacadas como relevantes en este trabajo para los tres caracteres y análisis realizados. Se resaltan en amarillo, de cada análisis, las variantes significativas. Se indica el carácter, especies implicadas en el análisis, p-valor, identificador del SNP (SNP ID), localización, posición en la región (intergénica, exónica o intrónica), gen en el que se localizan en caso de estar dentro de la secuencia de un gen anotado (Gen ID) y descripción del gen.

Carácter	Especie/grupo de especies	p-valor	SNP ID	Localización		Tipo de región	Genes involucrados	
				Cromosoma	Posición		Gen ID	Descripción
Peso	<i>Citrullus</i> spp.	7,413	S02_2738833 6	CG_CHR02	27.388.336	intrónica	<i>CICG02G013450</i>	Metal tolerance protein
Color	<i>Citrullus</i> spp.	6,015	S03_6053901	CG_CHR03	6.053.901	intergénica		
Color	<i>Citrullus</i>	4,414	S06_7303045	CG_CHR06	7.303.045	exónica	<i>CICG06G006370</i>	Homocysteine S-methyltransferase 3 family protein
Hilio	<i>Citrullus</i> spp.	8,888	S04_4876169	CG_CHR04	4.876.169	exónica	<i>CICG04G001430</i>	Glucose-6-phosphate 1-dehydrogenase
Hilio	<i>Citrullus</i> spp.	10,055	S10_1183562	CG_CHR10	1.183.562	exónica	<i>CICG10G001160</i>	PHD finger family protein

4.3.2. Impacto en la proteína de los SNPs localizados en regiones exónicas

Las posiciones de las variantes en los genomas de referencia y los alelos identificados en el llamado de variantes (previo a el análisis GWAS) se utilizaron para estudiar el impacto de los SNP en la función de las proteínas. Esto es, primero se analizó si la sustitución de un alelo por otro causaba un cambio de aminoácido en la proteína y, en caso de ser así, se estudió el impacto con los programas PROVEAN (<http://provean.jcvi.org/index.php>) y SNAP2 (<https://rostlab.org/services/snap2web/>) (Tabla 9). Sólo tres de las variantes analizadas (una para el color y dos para la presencia-ausencia del hilio) causan un cambio de aminoácido. La variante S06_7303045 (A/G) del cromosoma 6 asociada al color en *Citrullus* spp., resulta en una sustitución en el aminoácido 78 de la proteína de tirosina por cisteína (Y78C). Esta sustitución tiene impacto en la función de la proteína según las estimaciones realizadas por PROVEAN (-8,95) y SNAP2 (88). PROVEAN considera una mutación no neutra aquella con un valor estimado inferior a -2,5 y SNAP2 si supera el 0 (en una escala de -100 a +100). De los dos SNPs identificados con mayor p-valor y significación en el estudio de asociación a la presencia-ausencia del hilio, sólo uno causa sustitución de aminoácido. El SNP S04_4876169 (G/A) no causa sustitución en el gen *C1CG04G001430* (*Glucose-6-phosphate 1-dehydrogenase*) (Tabla 9), pero la variante del cromosoma 10 (S10_1183562), A/C, resulta en una sustitución en el gen *C1CG10G001160* (*PHD finger family protein*) de cisteína por triptófano en la posición 636. Este cambio tiene impacto en la proteína según se estima con los programas PROVEAN y SNAP2 (Tabla 9).

Tabla 9. Impacto de las variantes exónicas sobre la proteína. Se indica cuando la variante causa una sustitución de aminoácido y si es así el impacto estimado en la proteína. En amarillo se muestran las dos variantes significativas en el análisis GWAS.

Carácter	Especie	SNP ID	Localización		Alelos	Gen ID	Descripción	Cambio de aminoácido	Impacto en la proteína	
			Cromosoma	Posición					PROVEAN	SNAP2
Color	<i>Citrullus</i> spp.	S06_7303045	CG_CHR06	7.303.045	A/G	<i>CICG06G006370</i>	Homocysteine S-methyltransferase 3 family protein	Y78C	-8,95	88
Hilio	<i>Citrullus</i> spp.	S04_4876169	CG_CHR04	4.876.169	G/A	<i>CICG04G001430</i>	Glucose-6-phosphate 1-dehydrogenase	F454F	--	
Hilio	<i>Citrullus</i> spp.	S10_1183562	CG_CHR10	1.183.562	A/C	<i>CICG10G001160</i>	PHD finger family protein	C636W	-10,725	92

En este TFM se han caracterizado empleando la variabilidad natural de dos géneros, *Citrullus* y *Cucurbita*, caracteres de valor agronómico en semilla. A pesar de no encontrar una asociación significativa en algunos caracteres, como el peso y el color de la semilla en *Cucurbita*, se han detectado variantes tipo SNP relevantes para los tres caracteres en *Citrullus*. Las razones para la falta de asociación, ya indicadas, puede ser el número de variantes y el panel de accesiones empleadas. Por otro lado, los resultados en *Citrullus* resultan altamente interesantes ya que muchas de las variantes asociadas con mayor p-valor se localizan en regiones previamente descritas como relevantes para el carácter. Entre las génicas hay cuatro localizadas en genes relacionadas con el peso (1), color (1) y presencia de hilio (2). Los genes anotados son *Metal tolerance protein* (intrónica), *Homocisteína S-metiltransferasa 3* (exónica y causa cambio en la proteína), *Glucose-6-phosphate 1-dehydrogenase* (exónica, sin cambio en la proteína) y *PHD finger family protein* (exónica, sustitución en la proteína), respectivamente.

A pesar de encontrar la variante intrónica se ha revisado el papel del gen *Metal tolerance protein* (*MTP*). En muchos organismos, entre ellos plantas, se ha indicado que los intrones pueden regular la expresión génica incrementando la transcripción por distintos mecanismos (Shaul, 2017). Este gen está involucrado en el secuestro del exceso de zinc en el citoplasma en vacuolas para mantener la homeostasis del zinc, pero también actúan transportando manganeso, hierro, cadmio y níquel (Ricachenevsky et al., 2013). Existen múltiples genes *MTP* en plantas realizando diversas funciones. En *Oryza sativa* se ha identificado uno de ellos por análisis de QTLs (*OsMTP1*) probablemente implicado en incrementar las concentraciones de hierro y zinc en semilla (Anuradha et al., 2012). La Homocisteína S-metiltransferasa 3, cuya variante se ha encontrado asociada al color en *Citrullus* spp., cataliza la transferencia de grupos metilo de la S-metilmetionina a la adenosil-L-homocisteína (AdoMet). AdoMet es uno de los más importantes metabolitos que contienen azufre en las plantas y regulador clave de la vía del aspartato, un importante donante de grupos metilo para varias reacciones de metilación y un precursor de la biosíntesis de poliaminas y etileno (Watanabe et al., 2021). Tanto las poliaminas, como el etileno son reguladores del estrés en plantas (Watanabe et al., 2021), mientras que el etileno está implicado en la germinación en plantas, para romper la dormancia (Matilla, 2000; Botha et al., 1982). En sandía se ha estudiado la implicación del color en la germinación (Mavi et al., 2010). Estos resultados podrían sugerir un papel de este gen en la adquisición de humedad por lo que sería necesario evaluar las accesiones de *Citrullus* aquí analizadas respecto a este carácter y ver su asociación con el color. No obstante, indicar que, tal y como se discute en la sección correspondiente, esta variante se encuentra próxima a la región en la que se localiza el gen *W/w* responsable del color blanco de la semilla, por tanto esta variante no tendría por qué ser la mutación causal del cambio de color.

El SNP S04_4876169 localizado en el gen *CICG04G001430* (*Glucose-6-phosphate 1-dehydrogenase*) no causa cambio de aminoácido. No obstante, es interesante analizar la función de la *Glucose-6-phosphate 1-dehydrogenase* con respecto al hilio. Esta variante está asociada a la presencia-ausencia del hilio, relacionada en otras familias de plantas con el balance de agua y gases en la formación de la semilla y la germinación (Souza et al., 2001; Tiryaki y Topu, 2014; Debeaujon et al., 2000; Li et al., 1999; Crang et al., 2018; Mattiolo y Buscalioni, 1892; Corner, 1951). No se han encontrado datos sobre estudios genéticos en relación al carácter. Sin embargo, en *Arabidopsis* los mutantes nulos *g6pd5* y *g6pd6* y el doble mutante *g6pd5/6* deficientes en la función *Glucose-6-phosphate 1-dehydrogenase* germinan peor bajo estrés salino, y acumulan más especies

reactivas de oxígeno que el tipo salvaje (Yang et al., 2019), lo que sugiere un papel en la regulación de la respuesta a este estrés abiótico. Este resultado sería coherente con el papel indicado para el hilio en otras especies. La variante S10_1183562 asociada al hilio en *Citrullus* spp., causa la sustitución de aminoácido de mayor impacto entre las detectadas en este TFM. El cambio, en un miembro de la familia de proteínas PHD finger family protein, podría ser de interés para su estudio. Esta familia de proteínas cuenta con numerosos miembros en especies como *Arabidopsis*, donde se han descrito 45 PHD, o arroz con 44 PHD (Feng et al., 2004). La familia cuenta con numerosas funciones en el desarrollo de las plantas, entre las que se encuentra promover la germinación, regular la floración, el desarrollo del polen y la respuesta a estreses (Molitor et al., 2014; Lopez-Gonzalez et al., 2014; Fernandez-Gomez et al., 2014; Sun et al., 2017). Por lo que el análisis de las accesiones que muestran variación para este SNP puede ser interesante en relación para descubrir accesiones con caracteres agronómicos de interés, más allá de la germinación.

La revisión de la literatura respecto al control del peso de la semilla en cucurbitáceas, o el color, no menciona ninguno de los genes aquí citados, por lo que estos resultados, a pesar de prometedores, deben considerarse preliminares y los SNPs como variantes candidatas a estar asociadas. Será necesario una reevaluación de los caracteres en base a material de estas accesiones obtenido no en EEUU (origen del material vegetal) sino aquí, con el fin de ponderar el efecto del ambiente. Además, es necesario resaltar que las limitaciones de la técnica GWAS (existencia de subpoblaciones o relaciones) y de la técnica de secuenciación (GBS), pueden dar lugar, por un lado, a falsos positivos y/o falsos negativos, y por otro, a la no identificación de la mutación causal del carácter por no estar secuenciada. Por ello, más análisis son recomendables más allá este TFM.

CONCLUSIONES



5. CONCLUSIONES

PRIMERA.- Se ha caracterizado un total de 9.108 semillas de 713 accesiones de *Citrullus* spp. y 805 accesiones de *Cucurbita* spp. El panel de accesiones muestra variabilidad fenotípica respecto a los caracteres de semilla, peso, color e hilio, por lo que son de interés para realizar análisis de asociación para estos caracteres.

SEGUNDA.- El análisis de asociación GWAS realizado para el peso de la semilla en *Citrullus* spp., *C. maxima* y *C. pepo* permite identificar variantes asociadas. El panel de *Citrullus* spp. conduce a identificar hasta 9 regiones candidatas asociadas en los cromosomas 1, 2, 5, 7, 8, 9, 10 y 11. Una de ellas, en el cromosoma 2, presenta alta significación y colocaliza con un QTL mayor identificado en estudios realizados con poblaciones biparentales. Además, los SNPs del 2, 5, 7, 8 y 10 se localizan en regiones de QTLs menores identificados en el análisis de poblaciones biparentales para el peso, el ancho o el largo de la semilla.

TERCERA.- Al contrario del panel de *Citrullus* spp., el panel de accesiones de *C. maxima* no permite detectar ningún SNP significativamente asociado. No obstante, en el cromosoma 9 se identifica un SNP con p-valor más alto del análisis. Mientras que el análisis en el panel de *C. pepo* para este carácter indica que hay un SNP en el cromosoma 0 ($\alpha=0,05$) y un SNP en el cromosoma 1 con p-valor=4,527 por debajo del nivel de significación. Ninguno de estos SNPs se sitúan en regiones identificadas previamente ligadas a los caracteres de peso/tamaño de la semilla, no obstante podrían indicar regiones candidatas de interés para su análisis en distintos paneles de accesiones por GWAS, o en poblaciones biparentales.

CUARTA.- El análisis de asociación GWAS para el color de la semilla sólo se ha realizado en *Citrullus* spp. y *C. maxima*, ya que *C. pepo* no muestra variabilidad para este carácter. La diversidad es mayor en *Citrullus* spp. y se han identificado dos SNPs candidatos en el cromosoma 3 (significativo) y otro en el cromosoma 6 (no significativo) asociados al carácter. Estos dos SNPs colocalizan con QTLs descritos en estudios de poblaciones biparentales para el color de la semilla en sandía. El primero, en el cromosoma 2, se encuentra a aproximadamente 300 kb de la región del locus *D/d* que controla el color negro de la semilla. El segundo, en el cromosoma 6, se sitúa a aproximadamente 220 kb del marcador más próximo al locus *W/w* que regula la formación de una semilla de testa blanca. Dos SNPs no significativos localizados en el cromosoma 11 se han identificado en el panel de accesiones de *C. maxima*, podrían delimitar una nueva región candidata de interés para el color de la semilla.

QUINTA.- El estudio de asociación para el hilio de la semilla en el panel de *Citrullus* spp. permite identificar 2 SNPs candidatos con alta significación localizados en el cromosoma 4 y el 10. Además, se han identificado 7 variantes más en los cromosomas 2, 5, 8, 9 y 10 potencialmente implicadas en la regulación del carácter. No existen estudios previos en relación al carácter por lo que nuevos estudios serán de interés.

SEXTA.- El estudio de la localización de las variantes identificadas en relación al peso, color e hilio en los distintos paneles de accesiones, permite determinar que cuatro variantes son génicas, una intrónica y tres exónicas.

SÉPTIMA.- De las variantes exónicas analizadas, dos están asociadas al hilio, una no produce sustitución de aminoácido (cromosoma 6) y la otra sí (cromosoma 10). La variante sin impacto se sitúa en un gen *Glucose-6-phosphate 1-dehydrogenase* descrito

en *Arabidopsis* implicado en la tolerancia al estrés salino durante la germinación. La segunda variante (cromosoma 10) causa sustitución en un gen *PHD finger family protein*, con múltiples papeles en el desarrollo de plantas, entre ellas la germinación de la semilla.

BIBLIOGRAFÍA



6. BIBLIOGRAFÍA

- Achigan-Dako, E.G., Avohou, E.S., Linsoussi, C., Ahanchede, A., Vodouhe, R.S., y Blattner, F.R. (2015). Phenetic characterization of egusi-type (*C. mucosospermus*). *Genetic Resources and Crop Evolution*, 62:1159-1179.
- Achigan-Dako, E.G., Fagbemissi, R., Avohou, H.T., Vodouhe, R.S., Coulibaly, O., y Ahanchede, A. (2008a). Importance and practices of Egusi crops (*Citrullus lanatus* (Thunb.) Matsum. Y Nakai, *Cucumeropsis mannii* Naudin and *Lagenaria siceraria* (Molina) Standl. cv. Aklamkpa) in sociolinguistic area in Benin. *Biotechnology, Agronomy and Society and Environment*, 12(4):393-403.
- Achigan-Dako, E.G., Vodouhe, S.R., Sangare, A. (2008b). Caracterisation morphologique des cultivars locaux de *Lagenaria siceraria* (Cucurbitaceae) collectes au Benin et au Togo. *Belgian Journal of Botany*, 141(1):21-38.
- Achigan-Dako, E., Fanou, N., Kouke, A., Avohou, H., Vodouhe, S., y Ahanchede, A. (2006). Evaluation agronomique de trois espèces de Egusi (Cucurbitaceae) utilisées dans l'alimentation au Bénin et élaboration d'un modèle de prédiction du rendement. *Biotechnology, Agronomy and Society and Environment*, 10:121-129.
- Adjakidjé, V. (2006). Cucurbitaceae, p. 520-534. En: Akoégninou, A, Van derBurg, W.J., Van der Maesen, L.J.G. et al. (eds). Flore du Bénin. Backhuys Publishers, Cotonou and Wageningen.
- Aguado, E., García, A., Iglesias-Moya, J., Romero, J., Wehner, T.C., Gómez-Guillamón, M. L., ... y Jamilena, M. (2020). Mapping a partial andromonoecy locus in *Citrullus lanatus* using BSA-seq and GWAS approaches. *Frontiers in Plant Science*, 11:1243.
- Akusu, M.O., y Kiin-Kabari, D.B. (2015). Comparative studies on the physicochemical and sensory properties of watermelon (*Citrullus lanatus*) and melon (*Citrullus vulgaris*) seed flours used in "egusi" soup preparation. *Journal of Food Research*, 4(5):1-6.
- Anuradha, K., Agarwal, S., Rao, V., Rao, V., Viraktamath, C., y Sarla, N. (2012). Mapping QTLs and candidate genes for iron and zinc concentrations in unpolished rice of Madhukar× Swarna RILs. *Gene*, 508(2):233-240.
- Baboli, Z.M., y Kordi, A.S. (2010). Characteristics and composition of watermelon seed oil and solvent extraction parameters effects. *Journal of the American Oil Chemists' Society*, 87:667-671.
- Bagheri, H., Pino-del-Carpio, D., Hanhart, C., Bonnema, G., Keurentjes, J., y Aarts, M. (2013). Identification of seed-related QTL in *Brassica rapa*. *Spanish Journal of Agricultural Research*, 11(4):1085-1093.
- Bonhomme, M., y Jacquet, C. (2019). Genome-wide association mapping and population genomic features in *Medicago truncatula*. *The Model Legume Medicago truncatula*, 2:870-881.
- Botha, C., Small, C., y Grobbelaar, N. (1982). Seed germination in *Citrullus lanatus*. Part 2. The involvement of phytochrome and ethylene in controlling light sensitivity. *South African Journal of Botany*, 1(4):131-133.
- Boualem, A., Fergany, M., Fernandez, R., Troadec, C., Martin, A., y Morin, H. (2008). A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons. *Science*, 321:836-838.
- Boualem, A., Troadec, C., Camps, C., Lemhemdi, A., Morin, H., Sari, M.A., Fraenkel-Zagouri, R., Kovalski, I., Dogimont, C., Perl-Treves, R., y Bendahmane, A. (2015). A cucurbit androecy gene reveals how unisexual flowers develop and dioecy emerges. *Science*, 350:688-91.

- Boualem, A., Troadec, C., Kovalski, I., Sari, M.A., Perl-Treves, R., y Bendahmane, A. (2009). A conserved ethylene biosynthesis enzyme leads to andromonoecy in two *Cucumis* species. *PLoS One*, 4:e6144.
- Bradbury, J., Zhang, Z., Kroon, E., Casstevens, M., Ramdoss, Y., Buckler, S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23:2633-2635.
- Brachi, B., Morris, P., y Borevitz, O. (2011). Genome-wide association studies in plants: the missing heritability is in the field. *Genome biology*, 12(10):1-8.
- Burkill, H.M. (1985). The useful plants of west tropical Africa. 2 ed. *Royal Botanic Gardens*, Kew, Richmond, UK.
- Burkill, H.M. (1995). The useful plants of west tropical Africa. 2 ed (Vol. 1-3). *Royal Botanic Gardens*, Kew, Richmond, UK.
- Bush, A. (1978). Citron melon for cash and condiment. *Economic Botany*, 32:182-184.
- Büyükkartal, N., Çölgeçen, H., Pinar, M., y Erdoğan, N. (2013). Seed coat ultrastructure of hard-seeded and soft-seeded varieties of *Vicia sativa*. *Turkish Journal of Botany*, 37(2):270-275.
- CAJAMAR. (2020). Análisis de la campaña hortofrutícola de Almería. Campaña 2019/2020. *Cajamar Caja Rural*. (Consultado: 25 de marzo de 2021 en: <https://www.plataformatierra.es/detalle/analisis-campana-hortofruticola>).
- Camacho, F. (2003). Técnicas de producción en cultivos protegidos. Tomo 2. Ed. Caja Rural Intermediterránea, CAJAMAR, Almería, España.
- Camargo, Y. (2017). Asociación de marcadores AFLPs y caracteres de calidad y rendimiento en trigo candeal *Triticum turgidum* L. var. *Durum*. *Universidad Nacional del Sur*, Bahía Blanca, Argentina.
- Carnide, V., y Barroso, M. (2006). Las cucurbitáceas: bases para su mejora genética. *Horticultura Internacional*, 16-21.
- Chen, H., Sun, J., Li, S., Cui, Q., Zhang, H., Xin, F., y Huang, S. (2016). An ACC oxidase gene essential for cucumber carpel development. *Molecular plant*, 9(9):1315-1327.
- Chin, H.F. (1994). Seedbanks: conserving the past for the future. *Seed Science and Technology*, 22:385-400.
- Choi, Y., Chan, A.P. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, 31(16):2745-2747.
- Choi, Y., Sims, G.E., Murphy, S., Miller, J.R., Chan, A.P. (2012). Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS ONE*, 7(10)e46688.
- Chomicki, G., y Renner, S.S. (2015). Watermelon origin solved with molecular phylogenetics including L innaean material: another example of museomics. *New Phytologist*, 205(2):526-532.
- Corner, E.J.H. (1951). The leguminous Seed. *Phytomorphology*, 1:117-150.
- Crang, R., Lyons-Sobaski, S., y Wise, R. (2018). Plant anatomy: a concept-based approach to the structure of seed plants. *Springer*. Switzerland AG.
- Crawford, H. (2017). Manual de manejo agronómico para cultivo de sandía *Citrullus lanatus* (Thunb.) Matsum. et Nakai. Boletín (02). *Instituto de Investigaciones Agropecuarias (INIA)*, Santiago, Chile.
- Cubero, S.I. (2013). Introducción a la mejora genética vegetal. *Mundi-Prensa*. Madrid, España.
- Cutler, H.C., y Whitaker, T.W. (1969). A new species of *Cucurbita* from Ecuador. *Annals of the Missouri Botanical Garden*, 55:392-396.

- Dahl-Jensen, B., Ag-Hamattal, M., Aya-Touré, F., y Dolo-Nantoumé, A. (2011). Watermelons in the Sand of Sahara: cultivation and use of indigenous landraces in the Tombouctou Region of Mali. *Ethnobotany Research and Applications*, 9:151-62.
- Dana, W., y Ivo, W. (2008). Computer image analysis of seed shape and seed color for flax cultivar description. *Computers and Electronics in Agriculture*, 61(2):126-135.
- Dane, F., y Liu, J. (2007). Diversity and origin of cultivated and citron type watermelon (*Citrullus lanatus*). *Genetic Resources and Crop Evolution*, 54:1255-65.
- Davis, R., Levi, A., Tetteh, A., Wehner, T., Russo, V., y Pitrat, M. (2007). Evaluation of watermelon and related species for resistance to race 1W Powdery Mildew. *Journal of the American Society for Horticultural Science*, 132:790-795.
- De Winter, B. (1990). Notes on African plants. Cucurbitaceae. A new species of *Citrullus* (Benincaseae) from the Namib Desert, Namibia. *Bothalia*, 20:209-211.
- Debeaujon, I., Léon-Kloosterziel, M., y Koornneef, M. (2000). Influence of the testa on seed dormancy, germination, and longevity in *Arabidopsis*. *Plant Physiology*, 122(2):403-413.
- Decker, D.S. (1988). Origin(s), evolution, and systematics of *Cucurbita pepo* (Cucurbitaceae). *Economic Botany*, 42:4-15.
- Del Grosso, L. (2013). Colecciones públicas de recursos fitogenéticos de *Citrullus spp.* Trabajo Fin de Máster, *Universidad de Almería*, España.
- Delgado, G.J. (1999). El cultivo de calabacín en el Levante de Almería. Técnicas de producción de frutas y hortalizas en los cultivos protegidos. *Instituto la Rural*, 3:55-98.
- Della Gaspera, P., y Rodríguez, R.A. (2013). El género *Cucurbita*, 9-24. En: Della Gaspera, P. (Ed.): Manual del cultivo del zapallo anquito (*Cucurbita moschata* Duch.). *Instituto Nacional de Tecnología Agropecuaria*, Argentina.
- Duchesne, A.N. (1786). Essai sur l'histoire naturelle des courges. Panckoucke, Paris.
- El-Bramawy, S., El-Hendawy, S., y Shaban, A. (2008). Assessing the suitability of morphological and phenological traits to screen Sesame genotypes for *Fusarium* wilt and charcoal rot disease resistance. *Journal of Plant Protection Research*, 48(4):397-410.
- Elshire, J., Glaubitz, C., Sun, Q., Poland, A., Kawamoto, K., Buckler, S., y Mitchell, E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one*, 6(5)e19379.
- Esteras, C., Nuez, F., y Picó, B. (2012). Genetic Diversity Studies in Cucurbits Using Molecular Tools, p. 140-198. En: Wang, Y.H., Behera, T.K., Kole, C., (Eds.). Genetics, Genomics and Breeding of Cucurbits. *CRC Press*, New York, NY, USA.
- FAOSTAT. (2021). Datos sobre alimentación y agricultura. (Consultado: 4 de abril de 2021 en: <http://www.fao.org/faostat/es/#data/QC>).
- Feng, Y., Liu, P., y Xue, Z. (2004). Comparative phylogenetic analysis of the rice and *Arabidopsis* PHD-finger proteins. *Acta Genetica Sinica*, 31(11):1284-1293.
- Fernández-Gómez, J., y Wilson, A. (2014). A barley PHD finger transcription factor that confers male sterility by affecting tapetal development. *Plant Biotechnology Journal*, 12(6):765-777.
- Ferriol, M., y Picó, B. (2008). Pumpkin and winter squash, p. 317-349. En: Prohens, J., y Nuez, F. (Eds). Vegetables I. *Springer*, New York, USA.
- Filov. (1982). *Cucurbita maxima* subsp. *andreana* (Naudin). *Kul'turnaya Flora SSSR*, 21:177.

- Ford, R.I. (1980). Artifacts that grew their roots in Mexico. *Early Man*, 2:19-23.
- Fruhworth, G.O., y Hermetter, A. (2007). Seeds and oil of the Styrian oil pumpkin: Components and biological activities. *European Journal of Lipid Science and Technology*, 109(11):1128-1140.
- Fursa, T.B. (1972). K sistematike roda *Citrullus* Schrad. *Botanicheskii Zhurnal*, 57:31-41.
- Fursa, T.B. (1981). Intraspecific classification of watermelon under cultivation. *Die Kulturpflanze*, 29(1):297-300.
- Fursa, T.B. (1983). Novyi vid arbuza *Citrullus mucosospermus* (Fursa) Fursa. (A new species of watermelon *Citrullus mucosospermus* (Fursa) Fursa.). *Trudy Po Prikladnoi Botanike. Genetike i Seleksii*. 81:108-112.
- Gallegos, C.J. (2021). Improved design of containers and fertigation systems to increase production efficiency in soilless culture. Ph.D. Dissertation. *Almeria University*, Almeria, Spain.
- García, A., Aguado, E., Garrido, D., Martínez, C., y Jamilena, M. (2020a). Two androecious mutations reveal the crucial role of ethylene receptors in the initiation of female flower development in *Cucurbita pepo*. *The Plant Journal*, 103(4):1548-1560.
- García, A., Aguado, E., Martínez, C., Loska, D., Beltrán, S., Valenzuela, L., Garrido, D., y Jamilena, M. (2020b). The ethylene receptors *CpETR1A* and *CpETR2B* cooperate in the control of sex determination in *Cucurbita pepo*. *Journal of experimental botany*, 71(1):154-167.
- García, J., Pérez, O., Cos, J.E., Ruíz, L., y Sánchez, E. (2018). Influencia del cambio climático en la mejora genética de plantas. Sociedad Española de Ciencias Hortícolas y Sociedad Española de Genética. Consejería de Agua, Agricultura, Ganadería y Pesca. *Compobell, S.L.* Región de Murcia, España.
- Grichar, W.J., Besler, B.A., Brewer, K.D. (2001). Control de citronmelón (*Citrullus lanatus* var. Citroides) en maní de Texas (*Arachis hypogaea*) usando herbicidas de postemergencia. *Weed Technology*, 15(3):481-484.
- Grumet, R., Katzir, N., y Garcia-Mas, J. (2017). Genetics and genomics of Cucurbitaceae. *Springer International Publishing*.
- Guo, S., Zhang, J., Sun, H. et al. (2013). The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nature Genetics*, 45(1):51-58.
- Guo, S., Zhao, S., Sun, H. et al. (2019). Resequencing of 414 cultivated and wild watermelon accessions identifies selection for fruit quality traits. *Nature Genetics*, 51:1616-1623.
- Guo, Y., Gao, M., Liang, X., Xu, M., Liu, X., Zhang, Y., Liu, X., Liu, J., Gao, Y., Qu, S., y Luan, F. (2020). Quantitative trait loci for seed size variation in Cucurbits—a Review. *Frontiers in Plant Science*, 11(304), 01-12.
- Gusmini, G., Wehner, T., y Jarret, B. (2004). Inheritance of egusi seed type in watermelon. *Journal of Heredity*, 95:268-70.
- Gutiérrez-Díez, A., Martínez, J., García, E.A., Iracheta, L., Ocampo, J.D., y Cerda-Hurtado, I.M. (2009). Estudio de diversidad genética del aguacate nativo en Nuevo León, México. *Revista Fitotecnia Mexicana*, 32(1):09-18.
- Hall, D.W., Vandiver, V.V., Ferrel, J.A. (2004). Citron (Citron Melon), *Citrullus lanatus* (Thunb.) Mats y Nakai. En: Weeds in Florida, SP 37, Servicio de Extensión Cooperativa de Florida, Instituto de Ciencias Agrícolas y Alimentarias, *Universidad de Florida*. USA.
- Hammer, P., y Hammer, K. (2002). Pflanzenabbildungen auf Münzen. Kürbis, Kiwano und Co. *der Katalog zur Ausstellung*. Deutschland.

- Hussain, I., Rathmore, A., Sattar, Z., Chatha, A., Sarker, D., y Gilani, H. (2014). *Citrullus colocynthis* (L.) Schrad. (Bitter apple fruit): a review of its phytochemistry, pharmacology, traditional uses and nutritional potential. *Journal of Ethnopharmacol*, 8:54-66.
- Hutchinson, J., y Dalziel, J.M. (1954). Flora of west tropical Africa. 2 *Springer on behalf of the Governments of Nigeria, the Gold Coast, Sierra Leone, and the Gambia*.
- Iriondo, J.M. (2001). Conservación de germoplasma de especies raras y amenazadas (Revisión). *Investigación Agraria. Producción y Protección Vegetales*, 16(1):5-24.
- Iriondo, J.M., y Pérez, C. (1999). Propagation from Seeds and Seed Preservation, p. 46-57. En: Bowes, B.G. (ed.). *A Colour Atlas of Plant Propagation and Conservation*. *Manson Publishing*, London.
- ITIS (2021). Integrated Taxonomic Information System online database. (Consultado 5 junio de 2021 en: <https://www.itis.gov/>).
- Itle, R.A., y Kabelka, E.A. (2009). Correlation between L*a*b* color space values and carotenoid content in pumpkins and squash (*Curcubita* spp.). *HortScience*, 44:633-637.
- Jarret, R.L., Merrick, L.C., Holms, T., Evans, J. y Aradhya, M.K. 1997. Simple sequence repeats in watermelon (*Citrullus lanatus* (Thunb.) Matsum. & Nakai). *Genome*, 40: 433-441.
- Jarret, R.L., y Newman, M. (2000). Phylogenetic relationships among species of *Citrullus* and the placement of *C. rehmii* De Winter as determined by internal transcribed spacer (ITS) sequence heterogeneity. *Genetic Resources and Crop Evolution*, 47:215-222.
- Jeffrey, C. (1967). *Cucurbitaceae*. In: Milne, E., y Polhill, M. (Ed.). *Flora of Tropical East Africa*. Published under the Authority of the Minister for Overseas Development. London: Crown Agents for Oversea Governments and Administration.
- Jeffrey, C. (1990). Systematics of the Cucurbitaceae: an overview, p. 449-463. En: Bates, D.M., Robinson, R.W, Jeffrey, C (eds.). *Biology and utilization of the Cucurbitaceae*. *Cornell University Press*, Ithaca, New York, U.S.A.
- Jeffrey, C. (2001). *Cucurbitaceae*, p. 1510-1557. En: Hanelt, P. (ed). *Mansfeld' s encyclopedia of agricultural and horticultural crops*. *Springer*, Berlin.
- Jeffrey, C. (2005). A new system of Cucurbitaceae. *Botanicheskii Zhurnal*, 90(3):332-335.
- Jeffrey, C., y De Wilde, O. (2006). A review of the subtribe Thladianthinae (Cucurbitaceae). *Botanicheskii Zhurnal*, 91:766-776.
- Jones, S.B. (1989). Sistemática vegetal. 2da. Ed. *Mc. Grauw Hill*, México.
- Kameswara, R., Hanson, J., Ehsan, D., Ghosh, D., Larinde, M. (2007). Manual para el Manejo de Semillas en Bancos de Germoplasma. *Bioversity Internacional*. Roma, Italia.
- Kanda, T. (1931). The inheritance of seed-coat colouring in watermelon. *The Japanese Journal of Genetics*, 7:30-48.
- Klein, J., Zeiss, C., Chew, Y., Tsai, Y., Sackler, S., Haynes, C., Henning, K., SanGiovanni, P., Mane, M., Mayne, T. et al. (2005). Complement factor H polymorphism in age-related macular degeneration. *Science*, 308:385-389.
- Koh, J., Kwon, Y., y Thomson, M. (2015). Current technologies in plant molecular breeding. A guidebook of plant molecular breeding for researchers. *Springer*, Netherlands.
- Kole, C. (2011). Genetics, genomics and breeding of cucurbits. *Science Publishers*, Estados Unidos.

- Korte, A., y Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, 9:29.
- Lado, L.B. (2012). Identificación de SNPs mediante genotipado por secuenciación para el mejoramiento genético de trigo (*Triticum aestivum* L.). Tesis de Magister. *Universidad de la Republica*, Uruguay.
- Laghetti, G., y Hammer, K. (2007). The Corsican citron melon (*Citrullus lanatus* (Thunb.) Matsum. et Nakai subsp. *lanatus* var. *citroides* (Bailey) Mansf. ex Greb.) a traditional and neglected crop. *Genetic Resources and Crop Evolution*, 54(4):913-916.
- Lebeda, A., Widrlechner, M., Staub, J., Ezura, H., Zalapa, J., Křístková, E. (2007). Cucurbits (Cucurbitaceae: Cucumis spp., Cucurbita spp., Citrullus spp.). *Editorial NCRPIS, Genetic Resources, Chromosome Engineering, and Crop Improvement: Vegetable Crops*, 8:271-376.
- Lelley, T., Loy, B., Murkovic, M. (2010). Breeding oil seed pumpkins, p. 469-92. En: Vollmann, J., Rajcan, I., (Eds.). *Handbook of plant breeding*, vol 4: oil crops. *Springer*, New York.
- Levi, A., Jarret, R., Kousik, S., Wechter, W.P., Nimmakayala, P., Reddy, U. (2017a). Genetic resources of watermelon, p. 61-86. En: Grumet, R., Katzir, N., Garcia-Mas, J. (Eds). *Genetics and genomics of Cucurbitaceae*. *Springer*, New York, USA.
- Levi, A., Simmons, A., Massey, L., Coffey, J., Wechter, W., Jarret, L., Tadmor, Y., Nimmakayala, P., Reddy, K. (2017b). Genetic Diversity in *Citrullus colocynthis* and its relationship with *C. lanatus* and *C. ecirrhosus* as determined using high frequency oligonucleotide–targeting active gene (HFO–TAG) markers. *Horticultural Science*, USA.
- Levi, A., Thies, A., Wechter, P., Harrison, F., Simmons, L., Reddy, K., Nimmakayala, P., Fei, Z. (2013). High frequency oligonucleotides: targeting active gene (HFO-TAG) markers revealed wide genetic diversity among *Citrullus* spp. accessions useful for enhancing disease or pest resistance in watermelon cultivars. *Genetic Resources and Crop Evolution*, 60:427-40.
- Levi, A., Thomas, E., Joobeur, T., Zhang, X., Davis, A. (2002). A genetic linkage map for watermelon derived from a testcross population: (*Citrullus lanatus* var. *citroides* x *C. lanatus* var. *lanatus*) x *Citrullus colocynthis*. *Theoretical and Applied Genetics*, 105(4):555-63.
- Levi, A., Thomas, E., Wehner, C., y Zhang, X. (2001). Low genetic diversity indicated the need to broaden the genetic base of cultivated watermelon. *HortScience*, 36:1096-1101.
- Levi, A., Wechter, P., Massey, L., Carter, L., Hopkins, D. (2011). An extended genetic linkage map for watermelon based on a testcross and a BC2F2 population. *American Journal of Plant Sciences*, 2:93-110.
- Li, B., Lu, X., Dou, J., Aslam, A., Gao, L., Zhao, S., He, N., y Liu, W. (2018). Construction of a high-density genetic map and mapping of fruit traits in watermelon (*Citrullus lanatus* L.) based on whole-genome resequencing. *International journal of molecular sciences*, 19(10):3268.
- Li, B., Lu, X., Gebremeskel, H., Zhao, S., He, N., Yuan, P., Gong, C., Mohammed, U., y Liu, W. (2020). Genetic mapping and discovery of the candidate gene for black seed coat color in watermelon (*Citrullus lanatus*). *Frontiers in Plant Science*, 10:1689.
- Li, X., Baskin, M., y Baskin, C. (1999). Anatomy of two mechanisms of breaking physical dormancy by experimental treatments in seeds of two North American *Rhus* species (*Anacardiaceae*). *American Journal of Botany*, 86(11):1505-1511.

- Lira, R., Eguiarte, L.E., y Montes, H.S. (2009). Proyecto Recopilación y análisis de la información existente de las especies de los géneros *Cucurbita* y *Sechium* que crecen y/o se cultivan en México. *CONABIO*, México, D.F.
- Lira, R. (1995). Estudios taxonómicos y ecogeográficos de las Cucurbitaceae latinoamericanas de importancia económica. *International Plant Genetic Resources Institute*, Rome, Italia.
- López, A. (2019). Obtención y caracterización de mutantes en cucurbitáceas para el incremento de la base genética de la familia. Trabajo Fin de Grado, *Universidad de Almería*, Almería, España.
- López-González, L., Mouriz, A., Narro, L., Bustos, R., Martínez, M., Jarillo, A., y Piñeiro, M. (2014). Chromatin-dependent repression of the Arabidopsis floral integrator genes involves plant specific PHD-containing proteins. *The Plant Cell*, 26(10):3922-3938.
- Maggs-Kolling, L., Madsen, S., Christiansen, L. (2000). A phenetic analysis of morphological variation in *Citrullus lanatus* in Namibia. *Genetic Resources and Crop Evolution*, 47:385-93.
- Manzano, S., Aguado, E., Martínez, C., Megías, Z., García, A., y Jamilena, M. (2016). The Ethylene Biosynthesis Gene CitACS4 Regulates Monoecy/Andromonoecy in Watermelon (*Citrullus lanatus*). *PLoS One*, 11:e0154362.
- MAPA. (2021). Anuario de estadística. (Consultado: 5 de abril de 2021 en: <https://www.mapa.gob.es/es/estadistica/temas/publicaciones/anuario-de-estadistica/default.aspx>).
- Marees, T., De Kluiver, H., Stringer, S., Vorspan, F., Curis, E., Marie, C., y Derks, M. (2018). A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *International journal of methods in psychiatric research*, 27(2):e1608.
- Maroto, J.V., Miguel, A., y Pomares, F. (2002). El cultivo de la sandía. *Fundación Caja Rural Valencia y Mundi-Prensa*, España.
- Martin, A., Troadec, C., Boualem, A., Rajab, M., Fernandez, R., Morin, H., Pitrat, M., Dogimont, C., y Bendahmane, A. (2009). A transposon-induced epigenetic change leads to sex determination in melon. *Nature*, 461:1135-1138.
- Martínez, C. (2013). Involvement of ethylene biosynthesis and signaling genes in sex determination and parthenocarpic fruit set and development in Zucchini (*Cucurbita pepo* L.). PhD Dissertation, *Universidad de Almería*, Almería, España.
- Martínez, C., Manzano, S., Megías, Z., Barrera, A., Boualem, A., Garrido, D., Bendahmane, A, y Jamilena, M. (2014). Molecular and functional characterization of *CpACS27A* gene reveals its involvement in monoecy instability and other associated traits in squash (*Cucurbita pepo* L.). *Planta*, 239:1201-1215.
- Martínez, C., Valenzuela, J.L., y Jamilena, M. (2021). Genetic and pre-and postharvest factors influencing the content of antioxidants in cucurbit crops. *Antioxidants*, 10(6):894.
- Martínez, M.C. (2003). El cultivo protegido del calabacín. Técnicas de producción en cultivos protegidos. *Caja Rural Intermediterránea, Cajamar*, Almería, España.
- Matilla, A.J. (2000). Ethylene in seed formation and germination. *Seed science research*, 10(2):111-126.
- Mattirolo, O., y Buscalioni, L. (1892). Ricerche anatomofisiologiche sui tegumenti seminali delle Papilionacee. Serie 2.T., *Memorie della Reale Accademiadelle Scienze di Torino*, 42:223-445.

- Mavi, K. (2010). The relationship between seed coat color and seed quality in watermelon Crimson sweet. *Horticultural Science*, 37(2):62-69.
- Maxted, N., Ford-Lloyd, B.V., y Hawkes, J.G. (1997). Complementary Conservation Strategies, p. 15-39. En: Maxted, N., Ford-Lloyd, B.V., Hawkes, J.G. (eds.). *Plant Genetic Conservation. The In Situ Approach*. ed. *Chapman & Hall*, London.
- Maynard, D.N. (2001). An introduction to the watermelon, p. 9-20. En: Maynard, N. (Ed.). *Watermelon characteristics, production and marketing*. *ASHS Press*, Alexandria, VA.
- McCreight, J.D., Staub, J.E., Wehner, T.C., y Dhillon, N.P.S. (2013). Gone global: Familiar and exotic cucurbits have Asian origins. *HortScience*, 48:1078-1089.
- McKay, J.W. (1936). Factor Interaction in *Citrullus*. *Japanese Journal of Crop Science*, 27:110-112.
- Meeuse, A.D.J. (1962). The Cucurbitaceae of southern Africa. *Bothalia*, 8:1-111.
- Mejías, Z. (2004). Efecto del 1-(2-cloro-4-piridil)-3-fenilurea (CPPU) en el cultivo del calabacín. Trabajo fin de grado. *Universidad de Almería*. Almería, España.
- Meru, G., Fu, Y., Leyva, D., Sarnoski, P., y Yagiz, Y. (2018). Phenotypic relationships among oil, protein, fatty acid composition and seed size traits in *Cucurbita pepo*. *Scientia Horticulturae*, 233:47-53.
- Mogotlane, E.A., Mokwala, P.W., y Mangena, P. (2018). Comparative analysis of the chemical compositions of indigenous watermelon (*Citrullus lanatus*) seeds from two districts in Limpopo Province, South Africa. *African Journal of Biotechnology*, 17(32):1001-1006.
- Molitor, M., Bu, Z., Yu, Y., y Shen, H. (2014). Arabidopsis AL PHD-PRC1 complexes promote seed germination through H3K4me3-to-H3K27me3 chromatin state switch in repression of seed developmental genes. *PLoS Genetics*, 10(1):e1004091.
- Montero-Pau, J., Blanca, J., Bombarely, A., Ziarsolo, P., Esteras, C., Marti-Gomez, C., Ferriol, M., Gomez, P., Jamilena, M., Mueller, L., Pico, B., Canizares, J. (2018). De novo assembly of the zucchini genome reveals a whole-genome duplication associated with the origin of the *Cucurbita* genus. *Plant Biotechnol Journal*, 16:1161-1171.
- Moreiras-Varela, O., Carbajal, A., y Vives, C. (2013). Tablas de composición de alimentos: guía de prácticas. *Ediciones Pirámide, SA*. Madrid, España.
- Murkovic, M., Müllleder, U., y Neunteufl, H. (2002). Carotenoid content in different varieties of pumpkins. *Journal of Food Composition and Analysis*, 15(6):633-638.
- Nakano, Y., y Kobayashi, Y. (2020). Genome-wide association studies of agronomic traits consisting of field-and molecular-based phenotypes. *Reviews in Agricultural Science*, 8:28-45.
- Nath, P., y Khandelwal, R.C. (1978). Genetics of seed-coat colour in *Citrullus lanatus* (Thunb.) Mansf. *Theoretical and Applied Genetics*, 53(1):29-31.
- Naudin, C. (1856). Nouvelles recherches sur les caractères spécifiques et les variétés des plantes du genre *Cucurbita*. *Annales des Sciences Naturelles; Botanique*, 6:5-73.
- Nee, M. (1990). The domestication of *Cucurbita* (Cucurbitaceae). *Economic Botany*, 44(3suppl.):56-68.
- Nimmakayala, P., Levi, A., Abburu, L., Abburu, V.I., Tomason, Y.R., Saminathan, T., Vajja, V.G., Malkaram, S., Reddy, R., Wehner, T.C., Reddy, U.K. (2014). Single nucleotide polymorphisms generated by genotyping by sequencing used to characterize genome-wide diversity, linkage disequilibrium and selection sweep for worldwide cultivated watermelon. *BMC Genomics*, 15:767.
- Ning, X., Wang, X., Yu, Z., Lu, S., y Li, G. (2018). Inheritance and molecular mapping

- of tight-placenta gene and seed traits in Chinese Hami Melon 'Queen'. *Hortscience*, 53(6):788-794.
- Ochuodho, J.O., y Modi, A.T. (2008). Dormancy of wild mustard (*Sisymbrium capense*) seeds is related to seed coat colour. *Seed Science and Technology*, 36(1):46-55.
- Paris, H.S. (1986). A proposed subspecific classification for *Cucurbita pepo*. *Phytologia*, 61:133-138.
- Paris, H.S. (2001). Histoty of the cultivar-groups of *Cucurbita pepo*. *Horticultural Reviews*, Westport, New York, 25:71-170.
- Paris, H.S. (2008). Summer Squash, p. 351-379. Handbook of Plant Breeding. *Springer*, New York.
- Paris, H.S. (2015). Origin and emergence of the sweet dessert watermelon *Citrullus lanatus*. *Annals of Botany*, 116:133-148.
- Paris, H.S. (2017). Genetic Resources of Pumpkins and Squash, *Cucurbita* spp., p. 111-154. En: Jorgensen, R.A. (Ed.). Genetics and genomics of Cucurbitaceae. *Springer*, Cham.
- Paris, H.S., y Nerson, H. (2003). Seed dimensions in the subspecies and cultivar-groups of *Cucurbita pepo*. *Genetic Resources and Crop Evolution*, 50(6):615-625.
- Paris, H.S., Yonash, N., Portnoy, V., Mozes-Daube, N., Tzuri, G., Katzir, N. (2003). Assessment of genetic relationships in *Cucurbita pepo* (Cucurbitaceae) using AFLP, ISSR, and SSR markers. *Theoretical and Applied Genetics*, 106:971-978.
- Paudel, L., Clevenger, J., y McGregor, C. (2019). Chromosomal locations and interactions of four loci associated with seed coat color in watermelon. *Frontiers in Plant Science*, 10:788.
- Pimpin, F., Filippini, F., Sambo, P., y Gianquinto, G. (2002). The effect of seed quality (seed colour variation) on storability, germination temperature and field performance of radicchio. *Seed Science and Technology*, 30(2):393-402.
- Prothro, J., Sandlin, K., Gill, R., Bachlava, E., White, V., Knapp, J., y McGregor, C. (2012). Mapping of the Egusi seed trait locus (eg) and quantitative trait Loci associated with seed oil percentage in watermelon. *Journal of the American Society for Horticultural Science*, 137(5):311-315.
- Rafalski, J.A. (2010). Association genetics in crop improvement. *Current Opinion in Plant Biology*, 13(2):174-180.
- Ranilla, L.G., Genovese, M.I., y Lajolo, F.M. (2007). Polyphenols and antioxidant capacity of seed coat and cotyledon from Brazilian and Peruvian bean cultivars (*Phaseolus vulgaris* L.). *Journal of Agricultural and Food Chemistry*, 55:90-98.
- Reche, M.J. (2000). Cultivo de calabacín en invernadero. Colegio de Ingenieros Agrícolas de Almería. *Editorial del Ministerio de Agricultura, Pesca y Alimentación*. Hojas divulgadoras, Almería, España.
- Reddy, U.K., Abburi, L., Abburi, L., Saminathan, T., Cantrell, R., Vajja, G., Reddy, R., Tomason, R., Levi, A., Wehner, C., Nimmakayala, P. (2014a). A genome-wide scan of selective sweeps and association mapping of fruit traits using microsatellite markers in watermelon. *Journal of Heredity*, 106(2):166-176.
- Reddy, U.K., Nimmakayala, P., Levi, A., Abburi, V.L., Saminathan, T., Tomason, Y.R., Vajja, G., Reddy, R., Abburi, L., Wehner, T.C., Ronin, Y., y Korol, A. (2014b). High-resolution genetic map for understanding the effect of genome-wide recombination rate on nucleotide diversity in watermelon. *G3: Genes, Genomes, Genetics*, 4(11):2219-2230.
- Renner, S.S., Pérez-Escobar, O.A., Silber, M.V., Preick, M., Hofreiter, M., Chomicki,

- G. (2019). A 3500-year-old genome from a Pharaonic tomb reveals that New Kingdom Egyptians were cultivating domesticated watermelon. *BioRxiv*, 642785.
- Renner, S.S., Sousa, A., y Chomicki, G. (2017). Chromosome numbers, Sudanese wild forms, and classification of the watermelon genus *Citrullus*, with 50 names allocated to seven biological species. *Taxon*, 66(6):1393-1405.
- Ricachenevsky, K., Menguer, K., Sperotto, A., Williams, E., y Fett, P. (2013). Roles of plant metal tolerance proteins (MTP) in metal storage and potential use in biofortification strategies. *Frontiers in Plant Science*, 4:144.
- Robinson, R., y Decker, W. (1997). Cucurbits. *CAB International*, New York, USA.
- Romero-Masegosa, J., Martínez, C., Aguado, E., García, A., Cebrián, G., Iglesias-Moya, J., ... y Jamilena, M. (2021). Response of *Cucurbita* spp. to tomato leaf curl New Delhi virus inoculation and identification of a dominant source of resistance in *Cucurbita moschata*. *Plant Pathology*, 70(1):206-218.
- Rubatsky, V.E. (2001). Origin, distribution, and uses, p. 21-26. En: Maynard, D.N. (ed.). Watermelons: characteristics, production, and marketing. *ASHS Press*, Alexandria, VA.
- Rubatzky, V.E., y Yamaguchi, M. (2012). World vegetables: principles, production, and nutritive values. *Springer*. USA.
- Ruiz, J.J. (2017). Obtención y caracterización de mutantes EMS en *Cucurbita pepo*. Trabajo fin de Máster, *Universidad de Almería*, Almería, España.
- Sachan, P., y Nath, P. (1976). Inheritance of some qualitative characters in watermelon, *Citrullus lanatus* (Thunb.) Mansf, p. 99-105. In: *Proceedings Third International Symposium. Sub-tropical Tropical Horticulture*. Bangalore.
- Sáez, C., Martínez, C., Montero-Pau, J., Esteras, C., Sifres, A., Blanca, J., Ferriol, M., López, C. y Picó, B. (2020). A Major QTL Located in Chromosome 8 of *Cucurbita moschata* is responsible for resistance to Tomato Leaf Curl New Delhi Virus. *Front. Plant Science*, 11:207.
- Sahin, N., y Polat, S. (2018). Morphological characterizations of some watermelon seeds, p. 193-200. En: XXX International Horticultural Congress IHC2018: V International Symposium on *Plant Genetic Resources and International*, 1297.
- Seeds-Gallery. (2021). Semillas de sandía charleston gray. (Consultado 5 junio de 2021 en: <https://www.seeds-gallery.shop/es/inicio/semillas-de-sandia-charleston-gray.html>)
- Seo, J.S., Burri, B.J., Quan, Z., y Neidlinger, T.R. (2005). Extraction and chromatography of carotenoids from pumpkin. *Journal of Chromatography A*, 1073(1-2):371-375.
- Shaul, O. (2017). How introns enhance gene expression. *The international journal of biochemistry & cell biology*, 91:145-155.
- Shi, Y., Zhang, M., Shu, Q., Ma, W., Sun, T., Xiang, C., Wang, C., y Duan, Y. (2021). Genetic mapping and identification of the candidate gene for white seed coat in *Cucurbita maxima*. *International Journal of Molecular Sciences*, 22(6):2972.
- Simpson, G.M. (2018). Plant Systematics. *Academic press*, California, USA.
- Solmaz, I., y Sari, N. (2009). Characterization of watermelon (*Citrullus lanatus*) accessions collected from Turkey for morphological traits. *Genetic Resources and Crop Evolution*, 56:173-88.
- Solmaz, I., Sari, N., Aka-Kacar, Y., Yalcin-Mendi, Y. (2010). The genetic characterization of Turkish watermelon (*Citrullus lanatus*) accessions using RAPD markers. *Genetic Resources and Crop Evolution*, 57:763-71.
- Souza, D., y Marcos-Filho, J. (2001). The seed coat as a modulator of seed-environment

- relationships in *Fabaceae*. *Revista Brasileira de Botânica*, 24(4):365-375.
- Stephens, J.M. (2003). Citron *Citrullus lanatus* (Thunb.) Mansf. var. *citroides* (Bailey) Mansf. Publ. n. HS585, Serie of the Horticultural Sciences Department, *Institute of Food and Agricultural Sciences, University of Florida*, USA.
- Sun, H., Wu, S., Zhang, G., Jiao, C., Guo, S., Ren, Y. et al. (2017). Karyotype stability and unbiased fractionation in the paleo-allotetraploid *Cucurbita* genomes. *Molecular plant*, 10(10):1293-1306.
- Sun, M., Jia, B., Yang, J., Cui, N., Zhu, Y., y Sun, X. (2017). Genome-wide identification of the PHD-finger family genes and their responses to environmental stresses in *Oryza sativa* L. *International Journal of Molecular Sciences*, 18(9):2005.
- Takhtajan, A.L. (1997). Diversity and Classification of Flowering Plants, *Columbia University Press*, New York, USA.
- Tan, X.Z., Ge, Y., Xu, W.L., Cui, C.S., y Qu, P. (2013). Construction of genetic linkage map and QTL analysis for seed width in pumpkin (*Cucurbita maxima*). *Acta Botanica Boreali-Occidentalia Sinica*, 33:697-702.
- Tardío, J., Pardo, M., Morales, R., Molina, M., y Aceituno. (2018). Inventario español de los conocimientos tradicionales relativos a la biodiversidad agrícola. *Ministerio de Agricultura, Alimentación y Medio Ambiente*, Madrid, España.
- Teppner, H. (2004). Notes on *Lagenaria* and *Cucurbita* (Cucurbitaceae)—review and new contributions. *Phyton*, 44:245-308.
- Tiryaki, I., y Topu, M. (2014). A novel method to overcome coat-imposed seed dormancy in *Lupinus albus* L. and *Trifolium pratense* L. *Journal of Botany*, ID647469.
- van der Vossen, M., Denton, A., y El Tahir, M. (2004). *Citrullus lanatus* (Thunb.) Matsum. et Nakai, p. 185-191. En: Grubben, H., Denton, A. (eds). Plant resources of tropical Africa. Vegetables. *PROTA Foundation/Backhuys Publishers*, Wageningen.
- Valero, G.T., Rodríguez, A.P., Ruiz, M.E., Ávila, T.J., y Varela, M.G. (2018). La alimentación española: características nutricionales de los principales alimentos de nuestra dieta. *Imprenta ROAL, S.L.* Madrid, España.
- Vicente, D.N. (2014). Nuevos recursos para la mejora genética de la especie cucurbita pepo: colección de familias de mutantes y plataforma de selección Tilling. Tesis Doctoral. *Universidad de Almería*, Almería, España.
- Villanueva, C.V. (2007). Calabazas cultivadas. Identificación de especies, caracterización y descripción varietal. Tesis, *Universidad Autónoma Chapingo*, México.
- Wang, X., Bao, K., Reddy, U.K., Bai, Y., Hammar, S.A., Jiao, C., y Fei, Z. (2018). The USDA cucumber (*Cucumis sativus* L.) collection: genetic diversity, population structure, genome-wide association studies, and core collection development. *Horticulture research*, 5(1):1-13.
- Wang, Y., Wang, C., Han, H., Luo, Y., Wang, Z., Yan, C., Xu, W., y Qu, S. (2020). Construction of a high-density genetic map and analysis of seed-related traits using specific length amplified fragment sequencing for *Cucurbita maxima*. *Frontiers in Plant Science*, 10(1782):01-12.
- Weetman, L.M. (1937). Inheritance and correlation of shape, size and colour in watermelon, *Citrullus vulgaris* Schrad. Res. Bull. *Iowa Agriculture and Home Economics Experiment Station*, 228:223-56.
- Watanabe, M., Chiba, Y., y Hirai, M.Y. (2021). Metabolism and Regulatory Functions of O-Acetylserine, S-Adenosylmethionine, Homocysteine, and Serine in Plant Development and Environmental Responses. *Frontiers in Plant Science*, 12:593.
- Wehner, C., Shetty, V., y Elmstrom, W. (2001). Breeding and seed production, p. 21-73.

- En: Maynard, N. (Ed.). Watermelons. Characteristics, production, and marketing. *ASHS Press*, Alexandria, VA.
- Wehner, T.C. (2008). Watermelon, p. 381-418. En: Prohens, J., Nuez, F. (Eds.). *Vegetables I. Asteraceae, brassicaceae, chenopodiaceae, and cucurbitaceae. Springer*, New York, NY.
- Whitaker, T.W., y Cutler, H.C. (1971). Pre-historic cucurbits from the valley of Oaxaca. *Economic Botany*, 25(2):123-127.
- Wu, S., Wang, X., Reddy, U., Sun, H., Bao, K., Gao, L. et al. (2019). Genome of ‘Charleston Gray’, the principal American watermelon cultivar, and genetic characterization of 1,365 accessions in the US National Plant Germplasm System watermelon collection. *Plant Biotechnology Journal*, 17(12):2246-2258.
- Yamaguchi, M., Hughes, D.L., Yabumoto, K., Jennings, W.G. (1977). Quality of cantaloupe muskmelons – variability and attributes. *Science Horticulture*, 6:59-70.
- Yang, K., Jeong, N., y Moon, K. (2010). Genetic analysis of genes controlling natural variation of seed coat and flower colors in soybean. *Journal of Heredity*, 101(6):757-768.
- Yang, L., Wang, X., Chang, N., Nan, W., Wang, S., Ruan, M., Sun, L., Li, S., y Bi, Y. (2019). Cytosolic glucose-6-phosphate dehydrogenase is involved in seed germination and root growth under salinity in Arabidopsis. *Frontiers in Plant Science*, 10:182.
- Tiryaki, G., Cil, A., y Tiryaki, I. (2016). Revealing seed coat colour variation and their possible association with seed yield parameters in common vetch (*Vicia sativa* L.). *International Journal of Agronomy*, ID1804108.
- Yong, X., y Guo, S. (2016). The watermelon genome. En: Grumet, R., Katzir, N., y Garcia-Mas, J. (Eds.). *Genetics and genomics of the Cucurbitaceae. Springer Intl Pub AG*, New York.
- Yoo, K.S., Bang, H., Lee, E.J., Crosby, K., y Patil, B.S. (2012). Variation of carotenoid, sugar, and ascorbic acid concentrations in watermelon genotypes and genetic analysis. *Horticulture, Environment, and Biotechnology*, 53(6):552-560.
- Yu, C.Y. (2013). Molecular mechanism of manipulating seed coat coloration in oilseed Brassica species. *Journal of Applied Genetics*, 54(2):135-145.
- Zargar, M., Raatz, B., y Sonah, H. (2015). MuslimaNazir; Bhat, JA; Dar, ZA; Agrawal, GK; Rakwal, R. Recent advances in molecular marker techniques: Insight into QTL mapping, GWAS and genomic selection in plants. *Journal of Crop Science and Biotechnology*, 18:293-308.
- Zhang, H., Miao, H., Wei, L., Li, C., Zhao, R., y Wang, C. (2013). Genetic analysis and QTL mapping of seed coat color in sesame (*Sesamum indicum* L.). *PLoS ONE*, ID63898.
- Zhang, X.K., Chen, J., Chen, L., Wang, H.Z., y Li, J.N. (2008). Imbibition behavior and flooding tolerance of rapeseed seed (*Brassica napus* L.) with different testa color. *Genetic Resources and Crop Evolution*, 55(8):1175-1184.
- Zhang, Z., Ersoz, E., Lai, Q., Todhunter, J., Tiwari, K., Gore, A., Bradbury, J., Yu, J., Arnett, K., Ordovas, M., y Buckler, E. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics*, 42:355-360.
- Zomlefer, W.B. (2004). Guía de las familias de plantas con flor. *Acribia S.A.*, Zaragoza, España.

Páginas web de los softwares utilizados:

CuGenDB (<http://cucurbitgenomics.org/>).

Expasy translate (<https://web.expasy.org/translate/>).

Tassel5 (<https://www.maizegenetics.net/tassel>).

SNAP2 (<https://rostlab.org/services/snap2web/>).

PROVEAN (<http://provean.jcvi.org/index.php>).

Resumen

La familia de las cucurbitáceas es una de las más diversas en el reino vegetal. En esta familia se encuentran especies cultivadas a nivel mundial, pero también especies de importancia regional y local de los géneros *Citrullus* y *Cucurbita*. Esta relevancia desde el punto de vista natural y su valor socioeconómico ha permitido la conservación de numerosas accesiones en distintos bancos de germoplasma de todo el mundo. Una de las colecciones más importantes es la del Departamento de Agricultura de EEUU (USDA), que conserva más de 3.000 accesiones de las especies de estos dos géneros. Se trata de accesiones del germoplasma silvestre, asilvestrado, de razas autóctonas, cultivares antiguos e híbridos modernos que favorecen el mantenimiento de la variabilidad genética y constituyen una fuente de caracteres altamente valiosos para la mejora genética. El trabajo de conservación se coordina con el trabajo de re-secuenciación de numerosas accesiones de los dos géneros realizados en el marco del *Cucurbit Coordinated Agricultural Project* (CucCAP). Esta información genómica se encuentra disponible para la realización de estudios de asociación (*Genome Whole Association Study*, GWAS) que permitan localizar regiones genómicas relacionadas con las diferencias fenotípicas observadas para un carácter dado, empleando para ello la variabilidad de la diversidad natural. Aunque las especies más importantes de estos dos géneros, *C. lanatus* y *C. pepo*, se cultivan principalmente por sus frutos, el tamaño de la semilla de las cucurbitáceas cultivadas varían considerablemente según los usos previstos, en el caso de la sandía y de la calabaza las semillas tienen usos duales y el tamaño de la semilla es un objetivo importante para la mejora genética. Una semilla de calidad contribuye a mayor eficiencia de la producción por lo que se desea que sea de tamaño, e igual si su destino es la extracción de compuestos. Por el contrario, se desean semillas de pequeño tamaño si el objetivo es el consumo del fruto. En este TFM se han caracterizado el peso, color y presencia-ausencia de hilio de las semillas de 1.518 accesiones de las especies *C. lanatus*, *C. mucospermus*, *C. amarus*, *C. pepo* y *C. maxima*. Esta caracterización permite no sólo estudiar la variabilidad de la colección, si no también realizar estudios de asociación (GWAS) para designar regiones del genoma candidatas al control de estos caracteres de interés en la semilla.

Abstract

The cucurbit family is one of the most diverse in the plant kingdom. This family includes worldwide cultivated species, but also regionally and locally important species of the genera *Citrullus* and *Cucurbita*. This relevance from a natural point of view and its socio-economic value has allowed the conservation of numerous accessions in different germplasm banks over the world. One of the most important collections is that of the US Department of Agriculture (USDA), which conserves more than 3,000 accessions of the species of these two genera. These are accessions of wild and feral germplasm, landraces, old cultivars and modern hybrids that help to maintain genetic variability and are highly valuable source of traits breeding. The conservation work is coordinated with the re-sequencing work of numerous accessions of the two genera carried out in the Cucurbit Coordinated Agricultural Project (CucCAP). This genomic information is available for association studies (*Genome Whole Association Study*, GWAS) to locate genomic regions related to the phenotypic variation observed for a given trait, using the natural variability. Although the most important species of these two genera, *C. lanatus* and *C. pepo*, are mainly grown because of their fruits, seed size of cultivated cucurbits varies considerably according to their intended uses, in watermelon and squash seeds have dual uses and seed size is an important target for breeding. Quality seed contributes to greater production efficiency, so seed size is desired, and the same is true if it is intended for extraction of compounds. On the contrary, small seed size is desired if the objective is the consumption of the fruit. In this TFM, the weight, colour and presence-absence of hilum of the seeds of 1,518 accessions of the species *C. lanatus*, *C. mucospermus*, *C. amarus*, *C. pepo* and *C. maxima* have been characterised. This characterisation allows us, not only to study the variability of the collection, but also to carry out association studies (GWAS) to designate candidate genome regions for the control of these traits of interest in the seed.

