



UNIVERSIDAD
POLITÉCNICA
DE MADRID



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

 **INIA**
Instituto Nacional de Investigación
y Tecnología Agraria y Alimentaria



UPM – CRF – INIA

**EN COLABORACIÓN CON EL MASTER OFICIAL EN MEJORA GENÉTICA VEGETAL DE
LA UNIVERSIDAD POLITÉCNICA DE VALENCIA**



**Selección y validación de la colección
nuclear de variedades tradicionales
españolas de Trigo blando**

PRESENTAN EL TRABAJO FIN DE MÁSTER

ALUMNO

Mario Fernández González

TUTOR ACADÉMICO

Joaquín Cañizares Sales

TUTORAS EXTERNAS

Laura Pascual Bañuls

Magdalena Ruiz Valcárcel

CURSO ACADÉMICO

2018 – 2019

VALENCIA, septiembre de 2019



UNIVERSIDAD
POLITÉCNICA
DE MADRID



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

 **INIA**
Instituto Nacional de Investigación
y Tecnología Agraria y Alimentaria



UPM – CRF – INIA

**EN COLABORACIÓN CON EL MASTER OFICIAL EN MEJORA GENÉTICA VEGETAL DE
LA UNIVERSIDAD POLITÉCNICA DE VALENCIA**



**Selección y validación de la colección
nuclear de variedades tradicionales
españolas de Trigo blando**

PRESENTAN EL TRABAJO FIN DE MÁSTER

ALUMNO

Mario Fernández González

TUTOR ACADÉMICO

Joaquín Cañizares Sales

TUTORAS EXTERNAS

Laura Pascual Bañuls

Magdalena Ruiz Valcárcel

CURSO ACADÉMICO

2018 – 2019

VALENCIA, septiembre de 2019

Agradecimientos

Agradezco a la Dra. M. Ruiz del Centro Nacional de Recursos Fitogenéticos su ayuda para la realización de las colecciones nucleares de trigo, a la Dra. L. Pascual su ayuda y tiempo compartido en la toma de datos de fenotipado y análisis de marcadores, y al Dr. J. Cañizares por su disponibilidad para actuar de tutor académico de este trabajo.

Agradezco al Centro Nacional de Recursos Fitogenéticos la cesión de los datos necesarios para la realización de este trabajo.

Este trabajo de investigación ha sido respaldado por los proyectos RFP2015-00008-C04-01 y AGL2016-77149-C2-1-P del Ministerio de Economía, Industria y Competitividad, y los proyectos RF2007-00001-C05 y AT2016-006 del Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria.

Finalmente me gustaría agradecer enormemente a todos los integrantes de la Unidad de Genética: Paco, Laura, Efraín, José María, Patri, Juan, Mati, Paqui e Isa de la Escuela Técnica superior de Ingeniería Agronómica, Alimentaria y de Biosistemas de la Universidad Politécnica de Madrid la oportunidad de realizar este trabajo en sus instalaciones bajo su supervisión, y el tiempo de risas y la paciencia y apoyo incondicional que han mostrado conmigo en los últimos días de redacción. También agradezco notablemente a Magdalena su disposición incondicional para enseñarme a partir de su experiencia en la toma de datos fenotípicos y la conservación de Recursos Fitogenéticos en el CRF.

Índice General

1. Introducción	1
1.1. Los Recursos Fitogenéticos.	1
1.2. Conservación de los Recursos Fitogenéticos. Los bancos de germoplasma y su labor.....	3
1.2.1. Iniciativas para la conservación de los Recursos Fitogenéticos.	5
1.2.2. La conservación de los Recursos Fitogenéticos en España.....	6
1.2.3. Manejo de colecciones de germoplasma.	7
1.3. El trigo y sus especies. Importancia económica.	9
1.3.1. Mejora del trigo	10
1.3.2. Recursos Fitogenéticos en trigo. El CRF y las colecciones de trigo españolas	12
2. Objetivos.....	13
3. Material y métodos.....	14
3.1. Material vegetal.	14
3.2. Obtención de los datos de pasaporte, agromorfológicos y moleculares de partida.....	14
3.3. Métodos de creación de las colecciones nucleares.	15
3.3.1. Determinación del tamaño de la colección nuclear.	15
3.3.2. Métodos de selección empleados.	16
3.4. Validación de la colección nuclear.....	17
3.4.1. Pasaporte.	17
3.4.2. Datos agromorfológicos cualitativos.....	17
3.4.3. Datos agromorfológicos cuantitativos.....	17
3.4.4. Datos genéticos.	18
4. Resultados.....	19
4.1. Selección del subconjunto inicial.	19
4.2. Creación de colecciones nucleares.	20
4.2.1. Determinación del tamaño.	20

4.2.2. Selección de las entradas.....	21
4.3. Validación de las colecciones.....	22
4.3.1. Datos de pasaporte.....	22
4.3.2. Caracteres agromorfológicos cualitativos y cuantitativos.	23
4.3.3 Marcadores bioquímicos y alelos del gen <i>Vrn-A1</i>	26
4.3.4. Validación con marcadores SNP.....	26
5. Discusión.....	29
5.1. Validación de las colecciones nucleares.....	30
6. Conclusiones.....	33
7. Bibliografía.....	34

Índice de Tablas

Tabla 1. Comparación de datos geográficos de pasaporte en el conjunto de partida y en el subconjunto inicial (SC).....	19
Tabla 2. Nombres locales presentes en el conjunto de partida y el subconjunto inicial (SC). ...	19
Tabla 3. Comparación de clases fenotípicas de los caracteres agromorfológicos en el conjunto de partida y en el subconjunto inicial (SC).	20
Tabla 4. Diversidad genética (Ht), y número de accesiones de cada población en el SC y seleccionadas.	21
Tabla 5. Rangos de datos geográficos en el subconjunto inicial SC, y colecciones.	22
Tabla 6. Nombres locales habituales en las variedades locales españolas según Gadea (1954) en el subconjunto inicial SC, y colecciones.	22
Tabla 7. Comparación de clases fenotípicas de los caracteres agromorfológicos cualitativos en el subconjunto inicial (SC), y colecciones.	23
Tabla 8. Medias, varianzas, coeficientes de variación y rangos de los caracteres agromorfológicos cuantitativos en el subconjunto inicial (SC), y las colecciones.	24
Tabla 9. Alelos de gluteninas de alto peso molecular en el subconjunto inicial (SC), y colecciones.	26
Tabla 10. Distribución de marcadores SNP en el subconjunto inicial (SC), y colecciones.	28
Tabla 11. Diversidad de Nei, distancia de Gower media y distancias E-NE y A-NE en el subconjunto inicial (SC), y colecciones.	28

Índice de Figuras

Figura 1. Representación del número de loci polimórficos respecto al número de entradas.	21
Figura 2. Histograma de frecuencias de las clases fenotípicas de los datos agromorfológicos y alelos de vernalización en subconjunto inicial (SC), y colecciones.	25
Figura 3. Análisis PCoA basado en los marcadores moleculares SNP.	27

Resumen

Uno de los objetivos más importantes de la mejora del siglo XXI es el desarrollo de cultivares que permitan mantener los rendimientos actuales en un escenario climático variable. En este contexto, las variedades tradicionales españolas presentan un gran potencial como fuente de caracteres adaptativos debido a la diversidad agroclimática presente en España.

La colección más numerosa y antigua de variedades tradicionales españolas de trigo se conserva en el Centro Nacional de Recursos Fitogenéticos (CRF) del INIA. La colección activa está formada por 3.722 accesiones, de las que 1.551 son variedades españolas. De éstas, 763 son variedades tradicionales españolas de trigo blando (*Triticum aestivum* subsp. *vulgare* (Vill.)). Estas accesiones presentan una amplia variabilidad genética y resultan de gran interés en investigación y mejora. El manejo y utilización de una colección con un alto número de accesiones, en la que además existen algunos duplicados, es difícil debido al volumen de material que se debe analizar y el espacio y recursos que consume. Como forma de facilitar el acceso a la variabilidad existente e incentivar su uso en mejora, se crean colecciones nucleares, que deben incluir la mayor diversidad posible evitando redundancias, y, por lo general, tener un tamaño inferior al 20% de la colección original.

A partir de la colección completa de variedades locales de trigo blando, en el CRF-INIA se realizó un cribado seleccionando 522 accesiones donde se excluyeron duplicados. Posteriormente, dentro de este grupo se seleccionaron 189 accesiones con objeto de cubrir la variabilidad geográfica y morfológica existente. Este subconjunto de variedades se ha caracterizado a nivel fenotípico y molecular con caracteres agromorfológicos (cuantitativos y cualitativos) y marcadores bioquímicos, y genotipado a nivel genómico empleando la tecnología DArTseq. Este conjunto de datos constituye el punto de partida del presente trabajo donde se ha abordado la creación de la colección nuclear española de trigo blando.

En primer lugar, se ha verificado la representatividad de la variabilidad existente en las 189 accesiones con datos geográficos y morfológicos. Posteriormente, tras estimar el tamaño más adecuado para la colección nuclear, se ha realizado la selección de las entradas utilizando cuatro métodos diferentes: 1) muestreo al azar de las accesiones, 2) selección mediante datos geográficos y morfológicos, 3) selección para maximizar la heterocigosidad estimada con 58K marcadores moleculares, y 4) selección

mediante muestreo estratificado en poblaciones genéticas teniendo en cuenta la diversidad de cada población. Para validar las cuatro colecciones obtenidas, se ha comparado su diversidad con la existente en las 189 accesiones de partida. Para ello se han empleado datos geográficos, fenotípicos, bioquímicos y marcadores SNP. Los resultados obtenidos han mostrado la ventaja que ofrece utilizar datos de caracterización molecular para complementar la información geográfica y fenotípica. De los métodos utilizados, tanto el método de estratificación como el de maximización de la heterocigosidad han mostrado su utilidad para crear colecciones nucleares. En cambio, la selección basada exclusivamente en datos geomorfológicos ha sido menos eficiente a la hora de cubrir la diversidad existente.

El presente trabajo ha permitido proponer una selección de 94 accesiones de variedades locales de trigo blando, que podrían ser la futura colección nuclear española conservada en el CRF – INIA facilitando el empleo de estos recursos genéticos en programas de mejora.

1. Introducción

1.1. Los Recursos Fitogenéticos.

El proceso evolutivo en la tierra ha originado una enorme diversidad de especies e individuos, que mediante los procesos de selección permanente se han adaptado a las diferentes condiciones terrestres. Esta variabilidad genética acumulada resulta esencial para el equilibrio del sistema y constituye lo que se denomina germoplasma del planeta (Martín, 2001).

Los Recursos Fitogenéticos comprenden la diversidad genética correspondiente al mundo vegetal, que se considera poseedora de un valor para el presente o el futuro, incluyendo normalmente las categorías siguientes: variedades de especies cultivadas, tanto tradicionales como comerciales; especies silvestres o asilvestradas afines a las cultivadas o con un valor actual o potencial, y materiales obtenidos en trabajos de mejora genética (Esquinas- Alcázar, 1993).

La diversidad del germoplasma es consecuencia del proceso evolutivo, incluyendo la especiación y la aparición de variabilidad nueva entre los genotipos ya existentes (mutaciones espontáneas, acción de la ventaja biológica de la selección natural en distintos ambientes, etc.). Desde el origen de la agricultura los seres humanos han intervenido en los procesos evolutivos naturales dando lugar a la domesticación de las plantas cultivadas. La domesticación y los procesos posteriores han generado mayor variabilidad fenotípica a lo largo del tiempo que la que se puede encontrar en sus parientes silvestres hoy en día. La adaptación a un nuevo ambiente, tanto en los lugares de donde es endémica una especie, por cambios en el clima, como en los que no, por las condiciones de cultivo, genera nueva diversidad. Esta diversidad que en las especies silvestres puede no prosperar y llevar a la desaparición de una población, en las especies cultivadas puede resultar beneficiosa o atractiva para el agricultor y ser seleccionada (Martín, 2001).

En ocasiones, erróneamente, se asocian los Recursos Fitogenéticos únicamente a las especies silvestres emparentadas con las cultivadas. La acción constante del hombre y la evolución natural en los entornos agrícolas ha permitido, allá donde se llevaba el cultivo, generar diversidad. Las poblaciones humanas transportaban genotipos por intereses nutricionales o culinarios, por lo que la diversidad potencial, debida al cultivo en nuevos nichos agroclimáticos, era muy amplia. Estos cultivares heterogéneos dieron lugar a las denominadas variedades locales, que se definen como un cultivar, mezcla de genotipos, con una alta capacidad de tolerar estreses bióticos y abióticos

con una alta estabilidad del rendimiento, manteniendo unos rendimientos intermedios en sistemas agrarios con bajos insumos (Zeven, 1998).

A diferencia de los procesos que dieron lugar a las variedades locales, el cambio de modelo agrícola ocurrido en el último siglo ha generado pérdida de diversidad a causa de la denominada erosión genética (Van de Wouw, 2009). Los primeros cambios hacia esta agricultura se dieron desde la revolución agraria inglesa ocurrida desde finales del siglo XVII hasta bien entrado el siglo XVIII (Allen, 2002). Las primeras casas de semillas del siglo XVIII (Cubero, 2000) buscaban los mejores fenotipos con el objetivo fundamental de aumentar las producciones. Esta selección incrementaba la homogeneidad causando pérdida de variabilidad. Además, la aparición de variedades más productivas que fueron desbancando a las variedades cultivadas tradicionalmente por los agricultores, causó un abandono de las variedades locales, estrechó la base genética de los cultivos, y generó pérdida de diversidad. Esta erosión genética producida por la mejora no tuvo gran repercusión hasta mediados del siglo pasado con la llegada de la Revolución verde.

La Revolución verde ocurrida en la segunda mitad del siglo pasado, ha supuesto uno de los hitos más importantes en la mejora vegetal. A su vez, es el desencadenante de la mayor pérdida de diversidad acontecida hasta nuestros días (Díez & Valcárcel, 2010). Durante este proceso se partió de una base genética estrecha de genotipos seleccionados, con los que se generaron gran cantidad de variedades que compartían en gran parte, el mismo fondo genético. A la vez que aumentaban los rendimientos de forma exponencial y se alimentaba a una población creciente con necesidades básicas, se erosionaban genéticamente los cultivos debido a la elección de muchos agricultores por la opción más rentable, dejando de lado a los cultivares tradicionales, muchos de los cuales se han perdido (Pérez de la Vega, 2010). La mejora actual dirigida a la uniformidad y el desarrollo de ideotipos varietales fijos demandados por los consumidores ha agravado la situación. Se ha llegado a un punto en el que las variedades del mercado, que son diferentes en al menos algún carácter fenotípico unas de otras, comparten una alta proporción de su fondo genético. Estas variedades modernas son mucho más productivas pero mucho menos resilientes ante las alteraciones climáticas y la aparición de nuevas enfermedades. El agotamiento de la diversidad genética es un problema grave, por lo que ha sido necesario recurrir a las variedades tradicionales y otros Recursos Fitogenéticos como fuente de variabilidad donde encontrar genes de resistencia y de adaptabilidad a las condiciones cambiantes. Un ejemplo de lo peligroso que es el agotamiento de la diversidad y la homogenización de los cultivos es el caso del tizón del maíz (*Helminthosporium*

maydis) en Estados Unidos en 1970. Este patógeno arrasó toda la producción de maíz híbrido. Posteriormente se descubrió que el citoplasma Texas era susceptible a la raza T de tizón, ya que asociado a este estaba una mutación en los receptores que permitía la interacción con la toxina de esta raza. Para superar éste problema fue necesario recurrir a los Recursos Fitogenéticos para encontrar nuevos mecanismos de androesterilidad (Vares, 1984).

Por causa de las alteraciones climáticas y el agotamiento de recursos, es necesario aumentar el rango de adaptación de los cultivos para hacer frente al futuro incierto de la agricultura mundial. El ser humano de forma inconsciente, ha iniciado un proceso difícil de revertir. La modificación del medio por acción antrópica, alteraciones del clima (aumentos de temperaturas, cambios estacionales, etc.) y otros factores abióticos como la desertización y salinización de los suelos, dificultan el proceso de la producción agraria mundial. A esto se suma el cambio de la distribución geográfica de estreses bióticos debido a la globalización. Los reservorios de diversidad disponible permiten encontrar variabilidad genética donde hallar soluciones para superar los retos de la producción agraria mundial: adaptación a condiciones cambiantes, una agricultura sostenible donde se dé un mejor aprovechamiento de los recursos con un menor impacto ambiental, y una mejor resistencia a enfermedades. Esto permitirá aumentar la resiliencia de los cultivos, mediante la identificación y explotación de genotipos mejor adaptados a cada ambiente (Díez *et al.*, 2018).

También dentro de esta variabilidad genética es donde se pueden identificar materiales con características organolépticas, nutraceuticas y otros caracteres relacionados con la calidad (Uauy *et al.*, 2006). A causa de esto, se debe considerar a los Recursos Fitogenéticos y la variabilidad genética que contienen, como un bien muypreciado.

1.2. Conservación de los Recursos Fitogenéticos. Los bancos de germoplasma y su labor.

Como ya se ha indicado, la conservación de los Recursos Fitogenéticos es vital para mantener y recuperar la variabilidad genética. Se pueden conservar organismos completos o parte de los mismos.

Una forma ideal de conservación de la variabilidad existente sin frenar el desarrollo, ni aumentar la pérdida de diversidad genética a largo plazo, es el cultivo continuado de las variedades en las comunidades y áreas donde se desarrollaron. Es decir, la conservación *in situ* de la diversidad, que consiste en preservar las variedades o

poblaciones vegetales en sus hábitats originales (Martín, 2001). La conservación de la diversidad silvestre, se realiza cuidando los hábitats naturales en donde reside dicha diversidad. La conservación de los hábitats está en auge con el aumento de la protección de áreas naturales, como creaciones de parques naturales o reservas de la biosfera. Por desgracia, esto no garantiza que no se destruyan por intereses humanos o desastres naturales asociados o no al cambio climático.

De forma paralela, la conservación *in situ* de la diversidad cultivada se realiza mediante el cultivo “en finca”, es decir, el cultivo de materiales en sus zonas de origen y con las técnicas agrícolas tradicionales (Maxted *et al.*, 1997). Este tipo de conservación se lleva a cabo tanto por entidades públicas y privadas de investigación, como por los agricultores. En países en vías de desarrollo donde por lo general se agrupan los centros de origen y diversidad de las especies, la falta de recursos humanos y económicos dedicados al fomento y mejora del sector rural, ha permitido que los Recursos Fitogenéticos como las variedades tradicionales y cultivares obsoletos se mantengan gracias a los pequeños agricultores que practican el cultivo “en finca”. Sin embargo, en los últimos años, con la aparición de semillas mejoradas en los centros de origen y diversidad y el aumento de la tecnificación en dichos lugares, muchos cultivares tradicionales y materiales semidomesticados han sido desplazados. El resultado pone en peligro la conservación de los Recursos Fitogenéticos, que hasta ese momento no había supuesto una política prioritaria. En casos aislados, el paso a un sistema de cultivo en ecológico ha permitido el mantenimiento “en finca” de variedades gracias a las subvenciones de recuperación rural, como puede ser el caso del cultivar “pèsol negre” apoyado por la Fundació Miquel Agustí en Cataluña (<http://fundaciomiquelagusti.com/es/2016/06/02/el-cultiu-de-pesol-negre-al-bergueda/>).

Las variedades autóctonas adaptadas a sus áreas tradicionales de cultivo y con alta calidad organoléptica, son susceptibles de ser mejoradas. La introgresión por retrocruzamiento de algunos genes de resistencia, promovería el aumento del rendimiento de estas variedades. Así, la mejora genética puede ayudar a afianzar la permanencia de estos recursos en los campos de cultivo.

Debido a los diversos factores presentes en los ecosistemas y en el mundo rural, no se puede depender únicamente de la conservación *in situ*. De forma complementaria se desarrolla conservación *ex situ* de los Recursos Fitogenéticos. La conservación *ex situ* es aquella que se realiza fuera de los ambientes originales, normalmente en bancos de germoplasma. Un banco de germoplasma es un depósito de Recursos

Fitogenéticos (Rao *et al.*, 2006) y presenta la ventaja de concentrar en un mismo lugar gran cantidad de Recursos Fitogenéticos adaptados a gran variabilidad de ambientes.

La forma más fácil de conservación *ex situ* es conservando las semillas. Cuando la conservación de semillas no se pueda realizar, bien porque la reproducción favorable es la de otro material de propagación o porque las semillas sean recalcitrantes para su conservación, se procederá a otros tipos de conservación *ex situ*. La conservación *in vitro*, que es una forma de conservación *ex situ*, se realiza mediante el cultivo de órganos o tejidos en condiciones controladas con el sustento de un medio nutritivo de cultivo (Martín, 2001). La conservación *ex situ* de plantas vivas de variedades y/o patrones de frutales es otra alternativa que también gestionan los bancos de germoplasma (Llácer & Badenes, 2010).

Las labores de los centros de conservación *in situ* y *ex situ* son principalmente recolectar y mantener la diversidad tanto silvestre como cultivada. En estos centros se conservan todo tipo de especies, no solo de uso alimentario o industrial, sino también ornamentales y de valor medioambiental debido al principio de precaución.

1.2.1. Iniciativas para la conservación de los Recursos Fitogenéticos.

La pérdida de biodiversidad debido al remplazo de las variedades tradicionales (Díez *et al.*, 2018), impulsó la necesidad de coordinar esfuerzos en la conservación de los Recursos Fitogenéticos. La conferencia técnica de FAO/IBP de Roma en 1967 inició una actuación coordinada a nivel global sobre la conservación *in situ* y *ex situ* de los Recursos Fitogenéticos. Varios años después se creó el IBPGR (International Board for Plant Genetic Resources) y posteriormente el IPGRI (International Plant Genetic Resources Institute) que colaboraron en la organización y aplicación de medios para la realización de colectas de germoplasma.

La posibilidad de almacenar germoplasma fuera de los lugares de origen llevó a ciertas praxis que se tuvieron que reglamentar. En un principio, en 1988 se creó un comité de expertos en diversidad biológica como respuesta a esta necesidad. En 1993 con la última ratificación del convenio sobre la Diversidad Biológica, se ha reconocido el valor de la conservación de la biodiversidad, sentándose las bases para la participación justa y equitativa de los beneficios generados derivados del uso lucrativo de la utilización del germoplasma (CBD, 2019). El convenio aporta derechos y obligaciones a los países poseedores de la diversidad biológica y a los terceros que quieran hacer uso de los mismos. Los países poseedores han de identificar, controlar, conservar y proteger su diversidad biológica, además de restaurar, preservar y

promover las áreas que la contengan. Los países terceros que quieran hacer uso de estos recursos deben de reconocer la soberanía nacional de los mismos y acordar un consentimiento fundamentado previo con el país de origen, de forma que el uso comercial de los conocimientos y recursos ha de repercutir económicamente en el país de origen de los mismos, protegiéndolos y favoreciendo su conservación. En el año 2003, como complemento, entró en vigor el protocolo de Cartagena (FAO & OMS, 2003) sobre seguridad biotecnológica que busca la protección de la diversidad frente a riesgos desencadenados por la utilización de organismos vivos genéticamente modificados (MITECO, 2019).

1.2.2. La conservación de los Recursos Fitogenéticos en España.

En España, con anterioridad a los años 70, algunos centros de investigación, jardines botánicos y universidades habían hecho acopio de diversos materiales, creando colecciones “en finca”, y *ex situ* tanto de semilla como de planta viva. También los mejoradores españoles habían recopilado materiales y fueron pioneros en la conservación de Recursos Fitogenéticos. Posteriormente, en el año 1977 se fundó el primer banco de germoplasma público de especies cultivadas en la finca “El Encín”, en Alcalá de Henares (Madrid), del entonces Instituto Nacional de Investigaciones Agrarias (INIA). El ministerio de Agricultura, en colaboración con el INIA y el apoyo del IBPGR gestionaron las primeras expediciones sistemáticas de germoplasma en nuestro territorio. En el INIA se recogieron tanto materiales provenientes de expediciones, como de colecciones ya existentes del Instituto de Cerealicultura Español, por lo que se conservan materiales previos a la Revolución verde.

En la actualidad, según la European Search Catalogue for Plant Genetic Resources (EURISCO, 2019), hay 40 centros con bancos de germoplasma en España. Todos ellos tienen accesiones registradas en la Base de datos de la EURISCO en la última actualización de 2019. Diecinueve de ellos contienen accesiones dentro de la Red nacional del Programa para la conservación y uso sostenible de los Recursos Fitogenéticos (Díez *et al.*, 2018).

Los principales centros de importancia con bancos de germoplasma son: el Centro de Investigación y Tecnología Agroalimentaria (CITA) de Zaragoza y el Instituto Universitario de Conservación y Mejora de la Agrodiversidad Valenciana (COMAV) en cultivos hortícolas, la Misión Biológica de Galicia-Consejo Superior de Investigaciones Científicas (MBG-CSIC) en Brásicas y la Fundació Miquel Agustí (FMA) de Cataluña (Díez *et al.*, 2018).

Especial relevancia tiene el Centro Nacional de Recursos Fitogenéticos (CRF) del INIA, que “tiene como objetivo principal contribuir a evitar la pérdida de diversidad genética de las especies, variedades y ecotipos vegetales autóctonos y cultivares en desuso cuyo potencial genético sea susceptible de ser empleado en agricultura y alimentación” (CRF-INIA, 2019). Es el primer encargado de promover la conservación y utilización sostenible de los Recursos Fitogenéticos. Además, el CRF tiene la responsabilidad de conservar un duplicado en el banco base de todas las colecciones españolas activas de semillas de la Red nacional, así como de mantener las colecciones activas de leguminosas, cereales y otras especies de uso industrial, y el Inventario Nacional de la Red de Colecciones de Recursos Fitogenéticos del Programa Nacional. Finalmente, el CRF como entidad pública, se encarga de la coordinación nacional del Programa cooperativo europeo de Recursos Fitogenéticos (ECPGR/Biodiversity), interactuar con el Sistema integrado europeo de bancos de germoplasma (AEGIS), y participar en diferentes comisiones sobre los Recursos Fitogenéticos (CRF-INIA, 2019).

1.2.3. Manejo de colecciones de germoplasma.

Los bancos de germoplasma son capaces de mantener colecciones de germoplasma en diferentes formas dependiendo del material: semillas, material vivo (estacas, tubérculos, etc.) o plantas en campo (ej. frutales). Los Recursos Fitogenéticos conservados deben cumplir unos requisitos como la asociación de una serie de datos de pasaporte, que aportan información sobre la fecha y lugar de la recolección, e incluso de prácticas agronómicas asociadas. La cumplimentación de estos datos no evita la posibilidad de que existan redundancias en las colecciones debido a sinonimias y homonimias de los materiales con gran dispersión geográfica, como en el caso de plantas leñosas o con reproducción clonal, adquiridos durante las colectas de germoplasma.

Las accesiones recogidas se tratan de forma diferente según el tipo de colección en la que se conserven y el uso que se vaya a hacer de las mismas. Si van a conservarse a largo plazo (indefinidamente) será una colección base; en cambio, si van a suministrar material a usuarios, pertenecerán a una colección activa. La conservación de las accesiones de ambos tipos de colección se lleva a cabo en cámaras a baja temperatura y condiciones de humedad baja, aunque más estrictas para la colección base. Los recipientes utilizados en colecciones activas permitirán la extracción fácil de los materiales, mientras que en las colecciones base donde las semillas no van a extraerse de la colección, el material tiende a introducirse en contenedores sellados

herméticamente. En este último caso es imprescindible la realización de test rutinarios de viabilidad de la semilla que alerten de la necesidad de regenerar las entradas.

Para el mantenimiento de las colecciones activas es necesario regenerar y multiplicar las entradas. Esto implica el crecimiento de un número significativo de plantas de forma aislada por accesión, la recolección de frutos y extracción de semillas que mantengan en todo lo posible la variabilidad genética de la accesión. El mantenimiento, por tanto, es un trabajo y esfuerzo enorme. El tamaño de las colecciones mantenidas es también fundamental a la hora de manejarlas, tanto dentro del banco como para los investigadores que las soliciten. El desarrollo de subconjuntos reducidos con diferentes fines a partir de la colección inicial es un elemento clave y una herramienta útil. Permite la optimización del tiempo y trabajo enfocado a la investigación básica y la mejora, ayudando al análisis específico de uno o varios caracteres en las colecciones de un banco de germoplasma.

1.2.4. Colecciones nucleares.

Conociendo que la diversidad genética es la base de la mejora, y sin variabilidad no hay nada que seleccionar, los bancos de germoplasma forman parte indispensable como depósitos de la diversidad conocida. En ocasiones, y con más frecuencia cuando las colecciones son grandes, existen redundancias en cuanto a variabilidad genética entre las entradas. Esto implica un mayor esfuerzo para seleccionar materiales de la colección, y una utilización menos eficiente de la misma. Una posible solución que permite un mejor uso de las colecciones es la creación de una “colección nuclear”. “En el caso de un banco de germoplasma, una colección nuclear consta de un número limitado de las accesiones de una colección, que se escogen para representar el espectro genético de una colección” (Brown, 1989b). Se trata por tanto de un grupo de muestras representativas de la variabilidad genética de la colección total (Brown, 1995). Van Hintum *et al.* en 2003, la redefine como: “un número limitado de entradas seleccionadas para representar la diversidad genética de toda la especie cultivada y sus parientes silvestres”.

Una colección nuclear, por tanto, se puede crear a partir de una colección, de una parte de una colección, o de un conjunto de colecciones de diferentes especies relacionadas. En el ámbito de la investigación y la mejora, un subconjunto representativo de la mayor diversidad genética posible facilita el manejo de la diversidad de la colección total. Brown (1989a) sugiere tres criterios genéticos para conocer si una colección es de calidad. Primero, ha de incluir la mayor diversidad geográfica y de menor taxón posible. Segundo, hacer énfasis en los alelos de amplia

adaptación frente a los especializados. Y finalmente, maximizar la diversidad genética determinada por el número de alelos por locus en la colección nuclear. En este contexto, resulta muy útil el empleo de marcadores moleculares como herramienta para calcular la diversidad.

Los conservadores de germoplasma tratan dos cuestiones fundamentales antes de crear una colección nuclear: la optimización del número de accesiones para mantener una proporción aceptable de alelos en la colección y el método que van a utilizar para seleccionar las entradas (Ruiz *et al.*, 2013). Existen diferentes métodos para crear colecciones nucleares en función de los medios de que se disponen: selección aleatoria de las entradas, basarse en caracteres agromorfológicos o en la diversidad genética de la población. Tras crear una colección nuclear se ha de verificar el trabajo mediante la validación de la misma, que consiste en comparar la colección nuclear con el conjunto de partida.

La creación de una colección nuclear se debe basar en caracteres con poca influencia ambiental. Por eso, caracteres agronómicos como el rendimiento no son adecuados. En resumen, se trata de una herramienta que permite racionalizar recursos para el manejo, caracterización y posterior análisis de las plantas, favoreciendo el avance científico. Especialmente, permite incrementar la eficiencia del fenotipado de los materiales, ya que este constituye el factor limitante en las evaluaciones. Por tanto, es un elemento a considerar en cualquier banco de germoplasma.

1.3. El trigo y sus especies. Importancia económica.

El trigo es un cereal de la familia Poaceae que designa los cereales del género *Triticum*. Se domesticó en el Creciente Fértil en torno al 7000 a.c. (Peng *et al.*, 2011), centro de diversidad de los géneros *Triticum* y *Aegilops*. Consta de varios centros secundarios de diversidad en el Himalaya occidental para trigo blando y Etiopía para trigo duro. Las especies emparentadas con el trigo cultivado son principalmente las del género *Triticum* y las 22 especies del género *Aegilops*, y de forma más lejana, algunas especies de *Agropyron*.

El trigo se clasifica en función del número básico de cromosomas, ya que las especies del género *Triticum* tienen probablemente un origen monofilético. Se distinguen dos grandes grupos: trigo duro o semolero, *Triticum turgidum* (L.) Thell, ($2n=4x=28$) de endospermo muy duro y vítreo, y trigo blando, *Triticum aestivum* (L.) em. Thell ($2n=6x=42$) de endospermo de dureza variable y textura harinosa. Dentro del trigo

blando se encuentran las cuatro subespecies cultivadas *vulgare*, *spelta*, *macha* y *sphaerococcum*.

Respecto al papel del trigo (blando y duro) en alimentación humana, con el paso de los siglos sus usos se han diversificado y su consumo no ha parado de crecer. Es el cereal de mayor superficie cultivada y el segundo por detrás del maíz en cuanto a producción mundial con 771 millones de toneladas según los datos de FAOSTAT en 2017 (FAOSTAT, 2019). La organización de las Naciones Unidas para la alimentación y la agricultura (FAO) espera un aumento del 60% en la producción para 2050, ya que provee en torno al 20% de las proteínas y calorías de la dieta. Debido a su importancia y dispersión, cualquier mínimo avance en investigación y/o mejora, puede llegar a ser relevante. Por esto se introdujo dentro del marco del Climate Change, Agriculture and Food Security (CCAFS). El consorcio de centros de investigación del CGIAR (Consultative Group on International Agricultural Research) es el encargado de desarrollar, entre otros cultivos, variedades de trigo con resiliencia que permitan afrontar los retos del cambio climático y asegurar el alimento a la población del futuro. Una necesidad que empuja a continuar con la investigación y la mejora del cultivo. Paralelamente, reconoce la importancia de mantener la variabilidad existente como mejor forma de enfrentarse a las incertidumbres del porvenir.

En cuanto al papel del trigo en nuestro país, a nivel comercial España es importador de cereales, ya que no cubre la demanda. En 2017 se sembraron de trigo el 12,21% de las tierras cultivables españolas (2.059.224 ha), de las cuales, el 80% son de trigo blando según el Anuario de Estadística (MAPA 2019). Con un rendimiento medio comercial de trigo blando de los últimos 10 años de 3.250 kg/ha (MAPA, 2019), el trigo blando en España representa un mercado importante. El mercado del trigo blando tiende al alza en las producciones en los últimos años, lo que aporta buenas perspectivas al sector según la Asociación Española de Técnicos Cerealistas (AETC). En lo que respecta al trigo espelta, solo un 0,071% de la semilla certificada de trigos blandos en España es de esta subespecie, por lo que su valor es residual ya que solo se cultiva en pequeños enclaves de la mitad norte de España.

1.3.1. Mejora del trigo

A nivel mundial el cultivo del trigo es uno de los cultivos más extendidos y antiguos. En el caso de España el cultivo del trigo viene desde antiguo, traído por romanos y árabes, se cultivó de forma continuada hasta nuestros días, generando la gran cantidad de variedades locales que se encuentran repartidas por la geografía peninsular. Con la llegada de las casas de semillas, la mejora del trigo ejercida por los

mejoradores vegetales previa a la Revolución verde, se basaba principalmente en la selección en variedades heterogéneas. Ejemplo de ello fue la obtención de la variedad italiana *Senatore Capelli* en trigo duro, de alto valor agronómico, obtenida de la población argelina *Jenah Rhetifah* en 1915 (Di Fonzo *et al.*, 2005). En España una referencia sobre trigo blando es la obtención de la variedad *Aragón 03* a partir de la variedad tradicional *Catalán de Monte*. Fue una selección ejercida sobre las diversas poblaciones de esta variedad en Aragón. El proceso de selección se llevó a cabo en la estación de investigación situada en Ejea de los Caballeros (Zaragoza) por D. Manuel Gadea en los años 40 del siglo pasado (Sánchez-García *et al.*, 2012).

La mejora del trigo tras la Revolución verde, introdujo un nuevo ideotipo, el trigo de menor estatura (semienanos), más resistente al encamado, enfermedades, y con mayor productividad. Un avance que cambió la forma de manejar el cultivo de muchos cereales, dirigiéndolos hacia la mecanización y el uso de insumos. Estas variedades eran tan productivas que no había prácticamente variedades tradicionales que pudieran competir. En contraposición a su éxito, supuso un cuello de botella en la diversidad cultivada del trigo a nivel mundial. Esto es debido a que unas pocas variedades semienanas desarrolladas por Norman Borlaug en el Centro Internacional de Mejora de Maíz y Trigo (CIMMYT) de Méjico, fueron el punto de partida de la mayor parte de nuevas variedades de trigo, uniformizándose el fondo genético de las variedades posteriores a la década de los 70.

Actualmente, la mejora se ayuda de toda clase de herramientas biotecnológicas como los marcadores moleculares para los trabajos de selección tanto de parentales como de líneas avanzadas. El trigo blando es una especie hexaploide, su genoma de 17 GB (Appels *et al.*, 2018) es mucho más grande que el de otras especies de uso alimentario, lo que explica que el avance en esta especie sea más lento. Por suerte, la importancia económica ha impulsado la investigación, avanzando en la secuenciación y anotación de un genoma cada vez más completo. En Agosto de 2018 se publicó finalmente el genoma completo de trigo blando por Appels *et al.* (2018), en el que participaron 75 instituciones. La publicación del genoma posibilitará el uso de herramientas biotecnológicas en el trigo, como la secuenciación y el desarrollo de marcadores moleculares. Toda esta información permitirá la visualización de la variabilidad de la especie y realizar una mejora más precisa y eficiente.

1.3.2. Recursos Fitogenéticos en trigo. El CRF y las colecciones de trigo españolas

En España la conservación *in situ* no resulta relevante puesto que pocos agricultores mantienen la diversidad a nivel local. En los últimos años se ha detectado un interés creciente por la recuperación de las variedades tradicionales, por lo que la conservación *ex situ* de los materiales si ha resultado provechosa. El CRF contiene una colección muy completa del género *Triticum* con 3.337 entradas, de las cuales 2.005 son de *T. aestivum*, casi todas de la subespecie *vulgare* (solo 99 son de la subespecie *spelta* y de origen español), 1.260 pertenecen a *T. turgidum* y el resto pertenecen a otras especies como *T. monococcum*, *T. timopheevii* o *T. urartu* (EURISCO, 2019; CRF-INIA, 2019).

El CRF ya posee una colección nuclear de trigo duro. Contiene 94 accesiones de tres subespecies de *T. turgidum*, “creada usando un método basado en la estructura de las poblaciones, agrupando las accesiones por subespecies y asignando el número de genotipos entre poblaciones de acuerdo a la diversidad de marcadores microsatelites (Single Sequence Repeats SSR)” (Ruiz *et al.*, 2013). Esta colección está validada geográfica y morfológicamente y cubre el 97% de los alelos SSRs presentes en la colección inicial. También se ha evaluado comparándola con colecciones nucleares creadas con otras estrategias analizando los caracteres cualitativos, cuantitativos y moleculares (alelos de gliadinas y marcadores SSRs) incluidos en cada colección nuclear.

Dentro de la colección de trigo blando del CRF, hay 763 accesiones españolas (*T. aestivum* subsp. *vulgare* (Vill.)), la mayoría, variedades locales colectadas antes de la década de los 60. Muchas de ellas poseen información agro-morfológica recogida en diferentes publicaciones (Gadea 1954; Sánchez-Monge 1957), y por el personal del CRF durante las multiplicaciones del material. Una gran parte de las variedades están representadas en el Espigario conservado en el CRF, iniciado en los años 50 por los investigadores del INIA. Gracias a esta colección se ha conservado una amplia diversidad de genotipos adaptados a diferentes ambientes de gran interés para la investigación y la mejora. Sin embargo, el tamaño de la colección dificulta su utilización, especialmente para aquellos trabajos que requieren mucho personal y recursos (como pueden ser las evaluaciones). Para fomentar su uso, en el presente trabajo se plantea crear una colección nuclear de variedades locales españolas de *T. aestivum* subsp. *vulgare* (Vill.) a partir de la colección existente en el CRF. Esto permitirá un mejor manejo y reducirá los costes de evaluación de los materiales.

2. Objetivos

El objetivo principal de este trabajo es la creación de una colección nuclear de variedades locales españolas de trigo blando (*Triticum aestivum* subsp. *vulgare* (Vill.)), a partir de 189 entradas preseleccionadas del Centro de Nacional de Recursos Fitogenéticos (CRF-INIA) partiendo del siguiente esquema:

- ❖ Establecer el tamaño más adecuado de la colección nuclear.
- ❖ Creación de diferentes colecciones nucleares usando metodologías distintas.
- ❖ Validar cada colección nuclear evaluando la representatividad del conjunto de partida.

3. Material y métodos

3.1. Material vegetal.

El material de partida es la colección de variedades tradicionales de trigo blando (*T. aestivum* subsp. *vulgare* (Vill.)) del banco de germoplasma del CRF-INIA, que se compone de un total de 763 accesiones. Con los datos de pasaporte de las entradas (nombre local y origen geográfico), que dispone la base de datos del INIA (http://webx.inia.es/web_coleccionescrf/), se realizó un primer cribado excluyendo las entradas sin datos de pasaporte, redundancias y errores en la clasificación taxonómica o identidad varietal (detectados durante la multiplicación/caracterización agromorfológica de las accesiones). A partir de este segundo conjunto de entradas se realizó una selección posterior en función de datos de pasaporte y agromorfológicos (descritos a continuación). De esta última selección se desarrolló una línea pura de cada accesión mediante siembra espiga–surco durante tres ciclos de multiplicación. Estas líneas serán denominadas a partir de ahora el subconjunto inicial (SC). En las líneas se realizaron comprobaciones de la identidad varietal siguiendo las descripciones de Gadea (1954) y los materiales del espigario que se encuentra actualmente en el CRF-INIA.

3.2. Obtención de los datos de pasaporte, agromorfológicos y moleculares de partida.

Los datos de pasaporte se han obtenido de las bases de datos CRF-INIA. Incluyen datos de nombre local, provincia y municipio de colecta de la accesión. Estos datos contienen coordenadas geográficas de latitud y longitud del municipio, así como su altitud.

Los datos agromorfológicos incluyen: caracteres cualitativos y cuantitativos. Los datos cualitativos se tomaron en Alcalá de Henares durante la multiplicación del material gestionado por el banco de germoplasma del CRF-INIA. Se tomaron 8 caracteres cualitativos siguiendo el descriptor para trigo (IBPGR, 1985): hábito de crecimiento, longitud de las barbas, rugosidad de barba, color de la barba, color de la gluma, vellosidad de la gluma, densidad de la espiga y color de la semilla. Todos ellos descritos en el Anejo I.

El subconjunto inicial SC se seleccionó en un proyecto previo coordinado (RF2007-00001-C05-05), basándose en los datos de pasaporte y agromorfológicos cualitativos, tomados en el propio proyecto. La selección del SC se ha validado en el presente

trabajo confirmando que los rangos de los datos geográficos y todas las clases de los datos cualitativos de la colección de partida estaban representados en el SC.

Durante la campaña 2016 – 2017, el SC se sembró en los campos del CRF-INIA en Alcalá de Henares y se caracterizó para los 8 caracteres cualitativos descritos previamente, y para 5 caracteres cuantitativos: días a espigado, días a maduración, altura de planta, longitud de espiga y número de espiguillas por espiga. Todos ellos descritos en el Anejo I.

También se dispone de datos de marcadores genéticos y moleculares del SC. Los datos moleculares se obtuvieron de un trabajo previo (Pascual *et al.*, en revisión). Todas las accesiones del SC fueron analizadas mediante genotipado masivo con la tecnología DArTseq GBS (Kilian *et al.*, 2012). Esta tecnología permite obtener un alto número de marcadores dominantes de presencia/ausencia de fragmentos de ADN, denominados a partir de ahora DArT, así como un alto número de marcadores de polimorfismo de un único nucleótido (SNP). En el caso de los marcadores DArT se seleccionaron 59.276 marcadores tras excluir aquellos con idéntico perfil alélico en todas las variedades y/o con más de un 10% de datos faltantes. En el caso de los SNP se seleccionaron 14.830 marcadores excluyendo aquellos heterocigotos, con idéntico perfil alélico en todas las variedades, y/o con más de un 10% de datos faltantes. Se empleó la metodología de filtrado descrita en Pascual *et al.* (en revisión). Además se realizó el análisis de la estructura poblacional con marcadores DArT que clasificó las entradas del SC en 4 poblaciones (Pascual *et al.*, en revisión). De este mismo estudio, se extrajeron los datos disponibles de genotipado para el gen de vernalización *Vrn-A1* relacionado con la transición del crecimiento vegetativo a la etapa reproductiva (Chen *et al.*, 2013).

Además, se obtuvieron datos de perfil electroforético de proteínas del endospermo (Pascual *et al.*, en revisión) para determinar los alelos de gluteninas de alto peso molecular. Estas proteínas se analizaron mediante electroforesis en geles de poliacrilamida (SDS-PAGE) al 12% en presencia de dodecilsulfato sódico, siguiendo el protocolo de Payne *et al.* (1980).

3.3. Métodos de creación de las colecciones nucleares.

3.3.1. Determinación del tamaño de la colección nuclear.

Para la determinación del tamaño de la Colección Nuclear se utilizaron los marcadores DArT descritos previamente. Se realizaron simulaciones de selecciones de diversos tamaños, partiendo del 10% del SC. Las simulaciones se llevaron a cabo con el

software BIO-R (Pacheco *et al.*, 2016) desarrollado en el CIMMYT, que emplea el algoritmo Core Hunter para realizar la selección (Thachuk *et al.*, 2009). Al realizar las simulaciones se maximizó la heterocigosidad (parámetro $H_e=1$), para el resto de parámetros se mantuvieron los valores predeterminados en el software. Para cada una de las simulaciones se calcularon los marcadores polimórficos. El tamaño seleccionado fue aquel a partir del cual comenzaba a saturar la cantidad de marcadores polimórficos.

Para los análisis de los marcadores, así como los análisis estadísticos posteriores se utilizó el software R versión 3.5.2 (R core team, 2018). En concreto para determinar los marcadores polimórficos se empleó la función “*gl.report.monomorphs*” de la librería “dartR” (Gruber *et al.*, 2018a; Gruber *et al.*, 2018b).

3.3.2. Métodos de selección empleados.

Colección Geográfica-morfológica (GM): Obtenida en el proyecto RF2007-00001-C05-05. La colección se seleccionó basándose en los datos de pasaporte y la caracterización agromorfológica de los caracteres cualitativos ya descritos, tomados en el proyecto.

Colección al Azar (AZ): Tomada al realizar una selección aleatoria de entradas del subconjunto inicial SC. Para seleccionar las entradas, estas se ordenaron al azar y se seleccionaron comenzando por la primera hasta llegar al tamaño óptimo determinado anteriormente para la colección nuclear.

Colección sin estratificación de las entradas (SE): Basada en el uso del algoritmo Core Hunter, empleando los marcadores DArT. Al igual que en el caso de las simulaciones, se maximizó la heterocigosidad ($H_e=1$). Se fijó el número de entradas seleccionadas en base al tamaño óptimo de la colección nuclear determinado previamente.

Colección estratificada (CE): Basada en selecciones realizadas dentro de cada una de las 4 poblaciones identificadas en el SC por Pascual *et al.* (en revisión). Primero se determinó la diversidad genética (H_t) (Nei, 1973) dentro de cada una de las poblaciones utilizando los marcadores DArT. Para ello se usó la función “*basic.stats*” de la librería “hierfstat” (Goudet, 2005). Después, teniendo en cuenta el tamaño y diversidad se determinó el número de entradas a seleccionar en cada una de las poblaciones, de forma que el número de accesiones seleccionadas de cada población fuera proporcional a su diversidad (en caso de haber menos accesiones en la población que el mínimo calculado se tomaron todas). Por último, se realizó la

selección dentro de cada población empleando los marcadores DArT, el algoritmo Core Hunter y maximizando la diversidad ($H_e=1$), como se ha descrito previamente.

3.4. Validación de la colección nuclear.

3.4.1. Pasaporte.

Se analizó la presencia de los nombres genéricos de trigos más comunes (Candeal, Chamorro, Jeja, etc.) recogidos por Gadea (1954) en las distintas colecciones seleccionadas. También se analizó la representación de los rangos geográficos de latitud, longitud y altitud.

3.4.2. Datos agromorfológicos cualitativos.

Se llevó a cabo un examen de presencia/ausencia de las clases fenotípicas de cada carácter. En paralelo, se efectuó el cálculo de las frecuencias de las clases fenotípicas de los diferentes caracteres. Para detectar diferencias significativas entre el SC y cada colección nuclear, se realizó un test exacto de Fisher ($p\text{-valor}<0,05$) sobre los conteos de las clases fenotípicas (Fisher, 1954). Para ello se usó la función “*fisher.test*” de la librería “stats” (R core Team, 2018). Se empleó el mismo procedimiento para comparar los alelos de las proteínas gluteninas de alto peso molecular y los alelos de vernalización.

3.4.3. Datos agromorfológicos cuantitativos.

Para cada carácter cuantitativo se calcularon las medias, varianzas, coeficientes de variación y cuartiles 0, 50 y 100 que definen el rango de variación del carácter. Para analizar las diferencias significativas entre el SC y cada colección, se utilizó el test de T de Student para las medias y el test de homogeneidad de Fisher para las varianzas ($p\text{-valor}<0,05$). Además, siguiendo la propuesta de Hu *et al.* (2000), se calculó el porcentaje de medias diferentes (MD), el porcentaje de varianzas diferentes (VD), el ratio del coeficiente de variación ($VR\% = 1/m \sum_{j=1}^m CV_C / CV_I \times 100$) y el ratio de coincidencia del rango ($CR\% = 1/m \sum_{j=1}^m R_C / R_I \times 100$) de los caracteres en cada colección frente al subconjunto SC. El CR y VR han sido diseñados para evaluar las propiedades de una colección nuclear respecto de la colección inicial, donde R_C es el rango de la colección nuclear, R_I es el rango de la colección inicial, CV_C es el coeficiente de variación de la colección nuclear, CV_I es el coeficiente de variación de la colección inicial y m es el número de caracteres (Hu *et al.*, 2000). De acuerdo con la propuesta de Hu *et al.* (2000) una colección nuclear es representativa para un MD $\leq 20\%$ y un CR $\geq 80\%$.

3.4.4. Datos genéticos.

Para validar las colecciones nucleares se utilizaron los marcadores SNP del trabajo de Pascual *et al.* (en revisión). Por un lado, se calculó la diversidad genética H_t con marcadores SNP en el conjunto SC y en las colecciones nucleares como se ha descrito previamente. Por otro lado, se estimó para cada uno de los marcadores SNP la frecuencia del alelo menos frecuente (MAF) dentro del SC. Los marcadores se clasificaron en función del MAF en las siguientes categorías $>0,1$; $\geq 0,05$; $\geq 0,03$; $>0,01$ y $\leq 0,01$. Finalmente, se determinó el número de marcadores fijados (monomórficos) en cada una de las colecciones nucleares dentro de cada categoría.

Para visualizar la variabilidad genética contenida en las colecciones, se realizó un análisis de coordenadas principales (PCoA) con la función “*gl.pcoa*” de la librería “*dartR*” (Gruber *et al.*, 2018b). En dicho análisis se incluyeron todas las accesiones del SC y los marcadores SNP con un $MAF > 0,05$. Posteriormente se representó la variabilidad contenida en las coordenadas PCoA1 y PCoA2, resaltando las accesiones en función de si habían sido seleccionadas en las diferentes colecciones nucleares.

Por último, se estimó el coeficiente de similitud de Gower o distancia de Gower (Gower, 1971), que determina el grado de semejanza o similitud entre individuos. Este coeficiente permite simultáneamente emplear variables cualitativas, cuantitativas y dicotómicas (Chávez *et al.*, 2010). Para su cálculo se empleó la función “*daisy*” seleccionando el parámetro métrico “*gower*”, de la librería *cluster* (Maechler *et al.*, 2018) partiendo de los datos de marcadores SNP, alelos de gluteninas, alelos del gen *Vrn-A1*, y los caracteres agromorfológicos cualitativos y cuantitativos. A partir del coeficiente de Gower se determinaron dos distancias para cada colección nuclear. La distancia A-NE que estima la distancia entre cada genotipo del SC y el más cercano en la colección nuclear, y la distancia E-NE que estima la distancia entre cada genotipo de la colección nuclear y el genotipo más próximo dentro de la misma colección. Dichas distancias se estimaron como se describe en Odong *et al.* (2013).

4. Resultados

4.1. Selección del subconjunto inicial.

El estudio partió de la colección de 763 accesiones españolas recolectadas en España y conservadas en el CRF-INIA de Alcalá de Henares, Madrid. Se realizó un cribado de estas accesiones en el que se eliminaron accesiones redundantes, las que en los datos de pasaporte no tuvieran origen geográfico ni nombre local, y los errores taxonómicos (detectados con datos de control interno del banco y perfiles electroforéticos de proteínas). El conjunto resultante constaba de 522 accesiones con datos de pasaporte y agromorfológicos de caracteres cualitativos recabados durante la multiplicación rutinaria de las accesiones.

Partiendo de estas 522 entradas se redujo el número de accesiones, evitando la pérdida de variabilidad, para poder abordar su genotipado a gran escala. Para ello, en un proyecto previo, se utilizaron datos geográficos de pasaporte y datos agromorfológicos cualitativos descritos en el Anejo 1. En total se seleccionaron 189 accesiones denominadas subconjunto inicial (SC) que representaran el conjunto de las 522 accesiones para diferentes tipos de caracteres (Tabla 1, 2 y 3).

Tabla 1. Comparación de datos geográficos de pasaporte en el conjunto de partida y en el subconjunto inicial (SC).

	Latitud max.	Latitud min.	Longitud Oeste	Longitud Este	Altitud
522	433310N, Asturias	281820N, Sta.C. Tenerife	174928W, Sta.C. Tenerife	41559E, Baleares	10 - 1610
SC	433310N, Asturias	281820N, Sta.C. Tenerife	174928W, Sta.C. Tenerife	41559E, Baleares	10 - 1610

Los resultados indican que los datos de pasaporte del SC incluyen los rangos de los datos geográficos del conjunto inicial (Tabla 1). Además, todos los nombres locales genéricos están también incluidos en el SC (Tabla 2).

Tabla 2. Nombres locales presentes en el conjunto de partida y el subconjunto inicial (SC).

Nombres locales	522	SC
Barbilla	15	5
Basto	5	1
Blanco	13	6
Blanquillo	12	5
Cabezorro	8	4
Candéal	34	6
Catalán	10	1
Chamorro	17	9
Hembrilla	11	2
Jeja, Xexa, Xeixa	40	8
Marceño, Marceño, Tremesino	13	1
Mocho	23	7
Negrete, Negrillo	43	14
Rojo	33	11

Tabla 3. Comparación de clases fenotípicas de los caracteres agromorfológicos en el conjunto de partida y en el subconjunto inicial (SC).

	Hábito de crecimiento	Rugosidad Barbas	Color Barbas	Longitud Barbas	Densidad Espiga	Vellosidad Gluma	Color Gluma	Color semilla
522	3-5-7	0-1	0-1-2-4	1-2-3-4	3-5-7-9	1-2-3	1-2-3	1-2
SC	3-5-7	0-1	0-1-2-4	1-2-3-4	3-5-7-9	1-2-3	1-2-3	1-2

Las diferentes clases fenotípicas de los 8 caracteres agromorfológicos también se mantienen al realizar la selección del SC (Tabla 3).

Las 189 entradas pasaron por varios ciclos de autofecundación para obtener líneas altamente homocigóticas. Posteriormente se realizó la comprobación de la identidad varietal. Durante el examen se detectaron errores mediante comparación con las descripciones varietales publicadas por Gadea (1954) y Sánchez-Monge (1957), con el espigario, y con perfiles electroforéticos de proteínas prolaminas con granos procedentes del espigario y/o del Banco del CRF. Los 26 errores de identidad varietal (EIVs) detectados en el conjunto SC, fueron descartados para su selección dentro de cualquier colección nuclear.

4.2. Creación de colecciones nucleares.

Para la realización de las colecciones nucleares primero se ha determinado el tamaño idóneo y luego se ha realizado la selección de las entradas mediante los métodos descritos en Material y Métodos.

4.2.1. Determinación del tamaño.

Para la determinación del tamaño idóneo de la colección nuclear se partió de 59.276 marcadores DArT polimórficos en el SC.

Para proceder a la determinación del tamaño, primero se realizaron simulaciones seleccionando distintos tamaños de colección (desde el 10% al 90% del tamaño del subconjunto SC a intervalos de un 10-5%). Posteriormente se representó el número de marcadores que seguían siendo polimórficos en cada selección respecto al número de accesiones seleccionadas. En la Figura 1 se observa que a partir de 80 o 90 accesiones seleccionadas, el porcentaje de marcadores polimórficos acumulados no aumentaba con la misma intensidad. A partir de estos resultados se decidió que el tamaño óptimo para la colección nuclear era 94 accesiones, un 18% de la colección inicial de 522 accesiones.

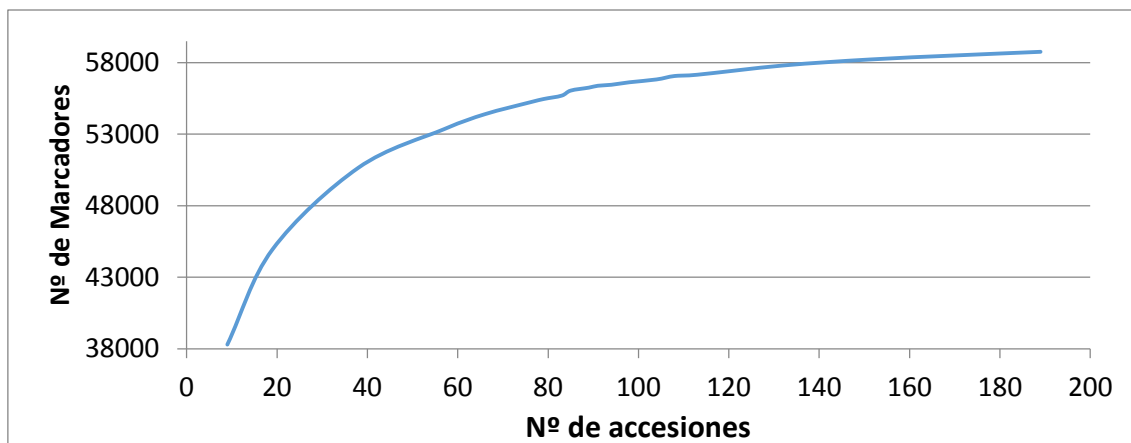


Figura 1. Representación del número de loci polimórficos respecto al número de entradas.

4.2.2. Selección de las entradas.

La existencia de algún error de identidad varietal (EIV) en el subconjunto inicial SC obligó a eliminar esas entradas de la colección nuclear. Por ello, las entradas marcadas con error de identidad y seleccionadas en primera instancia para formar parte de la colección nuclear fueron sustituidas por otras con identidad correcta.

La colección basada en parámetros geográficos y agromorfológicos (GM) incluida en este trabajo, y que consta de 84 accesiones, fue seleccionada en el proyecto (RF2007-00001-C05-05) y descrita por Aparicio *et al.* (2010). En el presente trabajo, con el descubrimiento de los EIV, se eliminaron las accesiones identificadas como tal, quedando un tamaño final de 74 entradas distribuidas por 34 provincias.

La colección al azar (AZ) se creó seleccionando aleatoriamente 94 entradas sin incluir entradas con EIV.

La colección sin estratificar (SE) se creó empleando el algoritmo Core Hunter como se ha descrito en el Material y Métodos, seleccionándose 94 entradas, 13 de las cuales, correspondían a EIV y se remplazaron por la entrada más cercana de acuerdo a la distancia de Gower.

La colección estratificada (CE) se basó en las 4 poblaciones descritas por Pascual *et al.* (en revisión). Para cada población se analizó la diversidad de genética (Ht) usando marcadores DArT (Tabla 4).

Tabla 4. Diversidad genética (Ht), y número de accesiones de cada población en el SC y seleccionadas.

	Población 1	Población 2	Población 3	Población 4
Nº de accesiones por población en SC	25	112	16	36
Diversidad de Nei (Ht)	0,21	0,32	0,13	0,23
Nº de accesiones a seleccionar	23	34	14	24
Nº de accesiones seleccionadas	21	37	10	26

En la tabla 4 se muestra el número de entradas por población en el SC y el número de accesiones por población a seleccionar en función de la diversidad genética. El N° de accesiones seleccionadas indica cuantas se han seleccionado finalmente debido a la presencia de EIV.

La selección de entradas de cada población se hizo de forma que se maximizara la heterocigosidad y que las entradas fueran lo más diferentes posible en cuanto a caracteres agromorfológicos y alelos de gluteninas de alto peso molecular (HMW-GS).

4.3. Validación de las colecciones.

La validación de las distintas colecciones se ha realizado comparando su variabilidad con la del subconjunto SC siguiendo la metodología descrita.

4.3.1. Datos de pasaporte.

Todas las colecciones cubren los rangos de latitud y longitud salvo la SE, que en longitud oeste solo llega hasta Gerona y no incluye ninguna entrada de las islas Baleares. La altitud queda cubierta prácticamente en su totalidad en las cuatro colecciones (Tabla 5).

Tabla 5. Rangos de datos geográficos en el subconjunto inicial SC, y colecciones al azar (AZ), sin estratificar (SE), geográfica-morfológica (GM) y estratificada (CE).

	Latitud max.	Latitud min.	Longitud Oeste	Longitud Este	Altitud
SC	433310N, Asturias	281820N, Sta.C. Tenerife	174928W, Sta.C. Tenerife	41559E, Baleares	10 – 1610
AZ	433310N, Asturias	281820N, Sta.C. Tenerife	162421W, Sta.C. Tenerife	31238E, Baleares	63 – 1540
SE	433310N, Asturias	281820N, Sta.C. Tenerife	174928W, Sta.C. Tenerife	14716E, Gerona	22 – 1540
GM	433310N, Asturias	281820N, Sta.C. Tenerife	174928W, Sta.C. Tenerife	41559E, Baleares	22 – 1610
CE	433310N, Asturias	281820N, Sta.C. Tenerife	174928W, Sta.C. Tenerife	41559E, Baleares	35 – 1540

Los nombres locales están recogidos en casi todas las colecciones nucleares (Tabla 6). Salvo en la colección GM que no recoge nombres como “Blanco” y derivados, y “Catalán”; y la colección SE que no contiene ninguna variedad del tipo “Hembrilla”.

Tabla 6. Nombres locales habituales en las variedades locales españolas según Gadea (1954) en el subconjunto inicial SC, y colecciones al azar (AZ), sin estratificar (SE), geográfica-morfológica (GM) y estratificada (CE).

Nombres locales	SC	Az	SE	GM	CE
Barbilla	5	2	3	0	3
Basto	1	1	1	1	1
Blanco	6	2	2	0	3
Blanquillo	5	1	3	0	3
Cabezorro	4	0	2	2	2
Candéal	6	1	3	3	3
Catalán	1	1	1	0	1
Chamorro	9	5	3	6	5
Hembrilla	2	1	0	2	1
Jeja, Xexa, Xeixa	8	4	6	5	6
Marceño, Marciño, Tremesino	1	2	3	1	3
Mocho	7	4	2	1	2
Negrete, Negrillo	14	7	7	10	10
Rojo	11	6	1	4	1

4.3.2. Caracteres agromorfológicos cualitativos y cuantitativos.

Las diferentes clases fenológicas descritas con anterioridad, están completamente representadas en las colecciones nucleares AZ, GM y CE. La colección SE no captura dos clases fenotípicas: longitud de barba [2] y densidad de espiga [9] (Tabla 7). Sobre estos datos se calcularon las frecuencias de las diferentes clases fenotípicas (Figura 2. A-H); no se detectaron diferencias significativas entre el subconjunto SC y las colecciones seleccionadas.

Tabla 7. Comparación de clases fenotípicas de los caracteres agromorfológicos cualitativos en el subconjunto inicial (SC), y colecciones al azar (AZ), sin estratificar (SE), geográfica-morfológica (GM) y estratificada (CE).

	Hábito de crecimiento	Rugosidad Barbas	Color Barbas	Longitud Barbas	Densidad Espiga	Vellosidad Gluma	Color Gluma	Color semilla
SC	3-5-7	0-1	0-1-2-4	1-2-3-4	3-5-7-9	1-2-3	1-2-3	1-2
AZ	3-5-7	0-1	0-1-2-4	1-2-3-4	3-5-7-9	1-2-3	1-2-3	1-2
SE	3-5-7	0-1	0-1-2-4	1-3-4	3-5-7	1-2-3	1-2-3	1-2
GM	3-5-7	0-1	0-1-2-4	1-2-3-4	3-5-7-9	1-2-3	1-2-3	1-2
CE	3-5-7	0-1	0-1-2-4	1-2-3-4	3-5-7-9	1-2-3	1-2-3	1-2

Los caracteres agromorfológicos cuantitativos en las colecciones AZ, SE, GM y CE no presentan diferencias significativas respecto al SC (Tabla 8). Los valores nulos obtenidos en todas las colecciones para los parámetros de evaluación del porcentaje de diferencias en medias (MD) y porcentaje de diferencias en varianzas (VD) indican que el SC está bien representado en las cuatro colecciones.

La colección SE presenta un coeficiente de variación mayor que otras colecciones en altura de planta y número de espiguillas por espiga. El ratio variable del coeficiente de variación (VR) y el ratio de coincidencia del rango (CR) son menores en la colección GM. El CR es superior al resto en las colecciones CE y SE (Tabla 8). Cabe destacar que el valor inferior del rango en el SC para los caracteres: días a espigado y días a maduración, se corresponde con EIVs que no pueden ser incluidos en ninguna de las colecciones nucleares. Hecho que puede llevar a sobrestimar la diferencia entre el SC y las colecciones. Al eliminar los EIVs, los rangos quedan en 155 – 188 días para los días a espigado, y 199 – 216 días para días a maduración.

Tabla 8. Medias, varianzas, coeficientes de variación y rangos de los caracteres agromorfológicos cuantitativos en el subconjunto inicial (SC), y las colecciones al azar (AZ), sin estratificar (SE), geográfica-morfológica (GM) y estratificada (CE).

Carácter	Media					Varianza					Coeficiente de variación					Rango				
	SC	AZ	SE	GM	CE	SC	AZ	SE	GM	CE	SC	AZ	SE	GM	CE	SC	AZ	SE	GM	CE
Días a espigado (d)	171,23	171,19	171,60	171,70	172,08	49,16	48,97	46,67	43,86	48,83	4,09	4,09	4,00	3,86	4,04	150 [†] -188	155-188	157-188	155-183	157-188
Días a maduración (d)	206,86	207	206,89	207,07	206,90	11,34	10,54	10,86	9,08	9,59	1,63	1,57	1,60	1,46	1,49	197 [†] -216	200-216	199-216	200-216	199-215
Altura de planta (cm)	88,27	88,19	87,74	87,39	85,97	137,87	159,70	152,93	119,42	137,73	13,30	14,33	14,17	12,50	13,58	53-119	53-119	53-115	68-115	53-114
Longitud de espiga (mm)	117,03	119	118,54	117,31	115,98	365,54	385,03	400,96	439,40	340,78	16,34	16,49	16,89	17,87	15,83	59-168	59-153	67-168	59-168	59-168
Espiguillas por espigas (n°)	19,11	19,28	19,29	19,08	19,02	4,16	3,51	4,62	4,19	4,22	10,68	9,72	11,14	10,72	10,74	14-24	14-23	14-24	15-24	14-24
	MD	0	0	0	0	VD	0	0	0	0	VR	99,18	102,07	97,47	97,92	CR	89,46	91,53	83,82	91,64

MD: Porcentaje de medias diferentes, VD: Porcentaje de varianzas diferentes, VR: Ratio del coeficiente de variación, CR: Ratio de coincidencia del rango.

Ninguna de las diferencias fue significativa (p -valor $<0,05$) de acuerdo con el T test de Student y el test de homogeneidad de Fisher.

[†] Valor correspondiente a una accesión con error de identidad varietal.

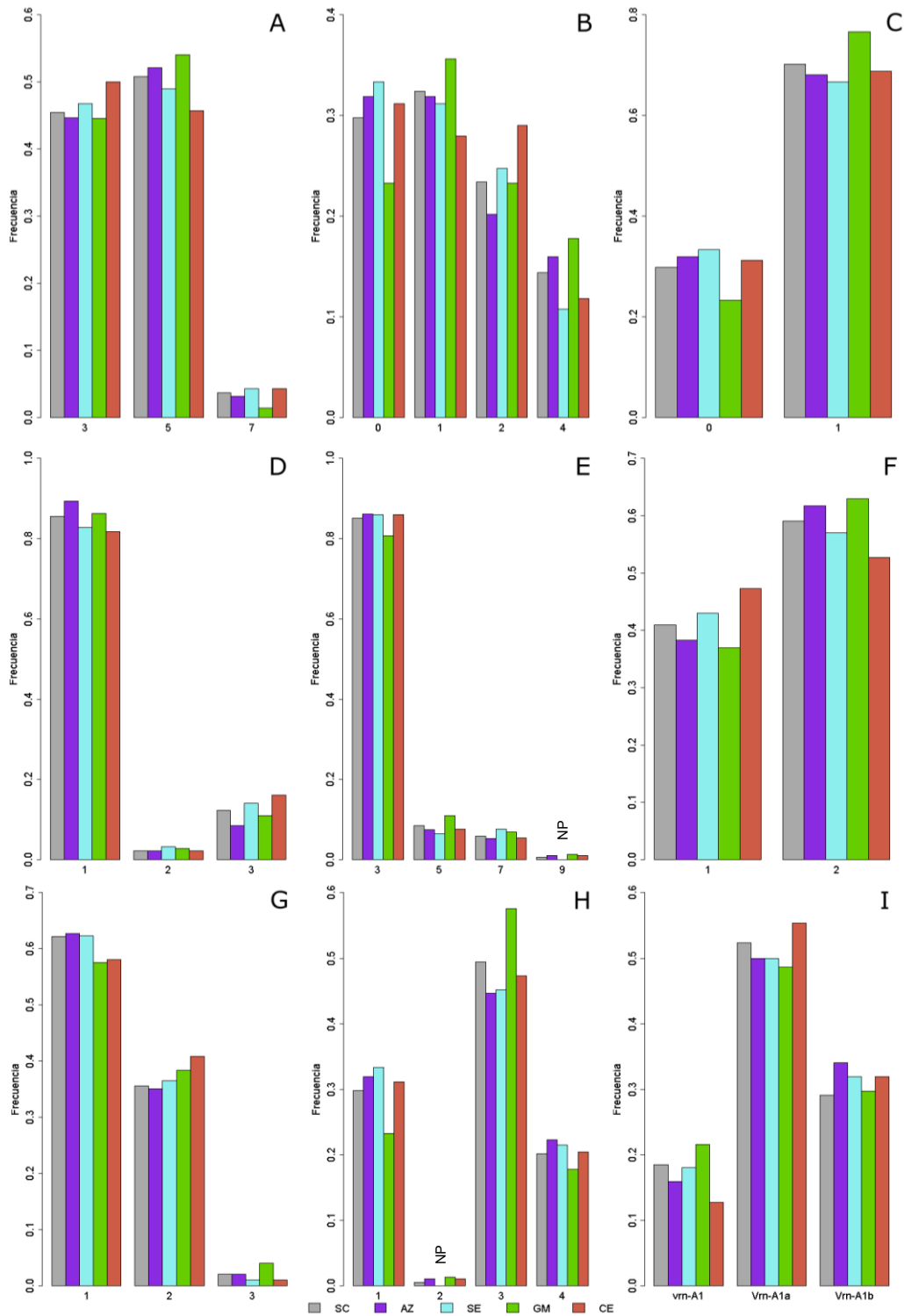


Figura 2. Histograma de frecuencias de las clases fenotípicas de los datos agromorfológicos y alelos de vernalización en subconjunto inicial (SC), y colecciones al azar (AZ), sin estratificar (SE), geográfica-morfológica (GM) y estratificada (CE).

SC en gris, AZ en morado, SE en azul claro, GM en verde y CE en rojo. (A) hábito de crecimiento, (B) color de la barba, (C) rugosidad de la barba, (D) vellosidad de la gluma, (E) densidad de la espiga, (F) color de la semilla, (G) color de la gluma, (H) longitud de barba, (I) alelos de vernalización. NP: no presente.

4.3.3 Marcadores bioquímicos y alelos del gen *Vrn-A1*.

Los alelos de gluteninas de alto peso molecular de los genomas A, B y D para los loci *Glu-A1*, *Glu-B1* and *Glu-D1*, respectivamente, han sido analizados en las 4 colecciones (Tabla 9). Todas las colecciones recogen los alelos del *Glu-A1*. La colección AZ y GM pierden alelos en los loci *Glu-B1* y *Glu-D1*. La colección SE solo pierde un alelo del *Glu-D1* y la colección CE es la única que contiene la totalidad de alelos presentes en el SC. Esto es excluyendo los alelos *n1* del *Glu-A1*, y *n7* y *u* del *Glu-D1* identificados en accesiones con EIV.

Tabla 9. Alelos de gluteninas de alto peso molecular en el subconjunto inicial (SC), y colecciones al azar (AZ), sin estratificar (SE), geográfica-morfológica (GM) y estratificada (CE).

Gluteninas	<i>Glu-A1</i>	<i>Glu-B1</i>	<i>Glu-D1</i>
SC	<i>a,b,c,n1[†],y</i>	<i>7⁺+9,a,al,am,aq,d,e,f,h,i,n2,n3,n4,n5,u</i>	<i>a,c,d,h,j,l,n6,n7[†],u[†]</i>
AZ	<i>a,b,c,y</i>	<i>7⁺+9,a,al,am,aq,d,e,f,h,i,n3,n4,n5,u</i>	<i>a,c,d,h,j,l</i>
SE	<i>a,b,c,y</i>	<i>7⁺+9,a,al,am,aq,d,e,f,h,i,n2,n3,n4,n5,u</i>	<i>a,c,d,j,l,n6</i>
GM	<i>a,b,c,y</i>	<i>7⁺+9,a,al,am,aq,d,e,f,h,i,n2,n3,u</i>	<i>a,c,d,h,j,l,n6</i>
CE	<i>a,b,c,y</i>	<i>7⁺+9,a,al,am,aq,d,e,f,h,i,n2,n3,n4,n5,u</i>	<i>a,c,d,h,j,l,n6</i>

[†] Errores de identidad varietal

Los tres alelos del gen de vernalización del locus *Vrn-A1*, están presentes en las cuatro colecciones nucleares seleccionadas (Figura 2. I).

4.3.4. Validación con marcadores SNP.

Se realizó un análisis de escalado multidimensional PCoA para tener una representación global de la variabilidad genética y comparar como está representada esta variabilidad en las colecciones nucleares. En la Figura 3 A y C se observa que las colecciones AZ y GM no contienen ninguna de las accesiones situadas en la zona central del PCoA. Además la colección GM no incluye las accesiones situadas en la parte superior.

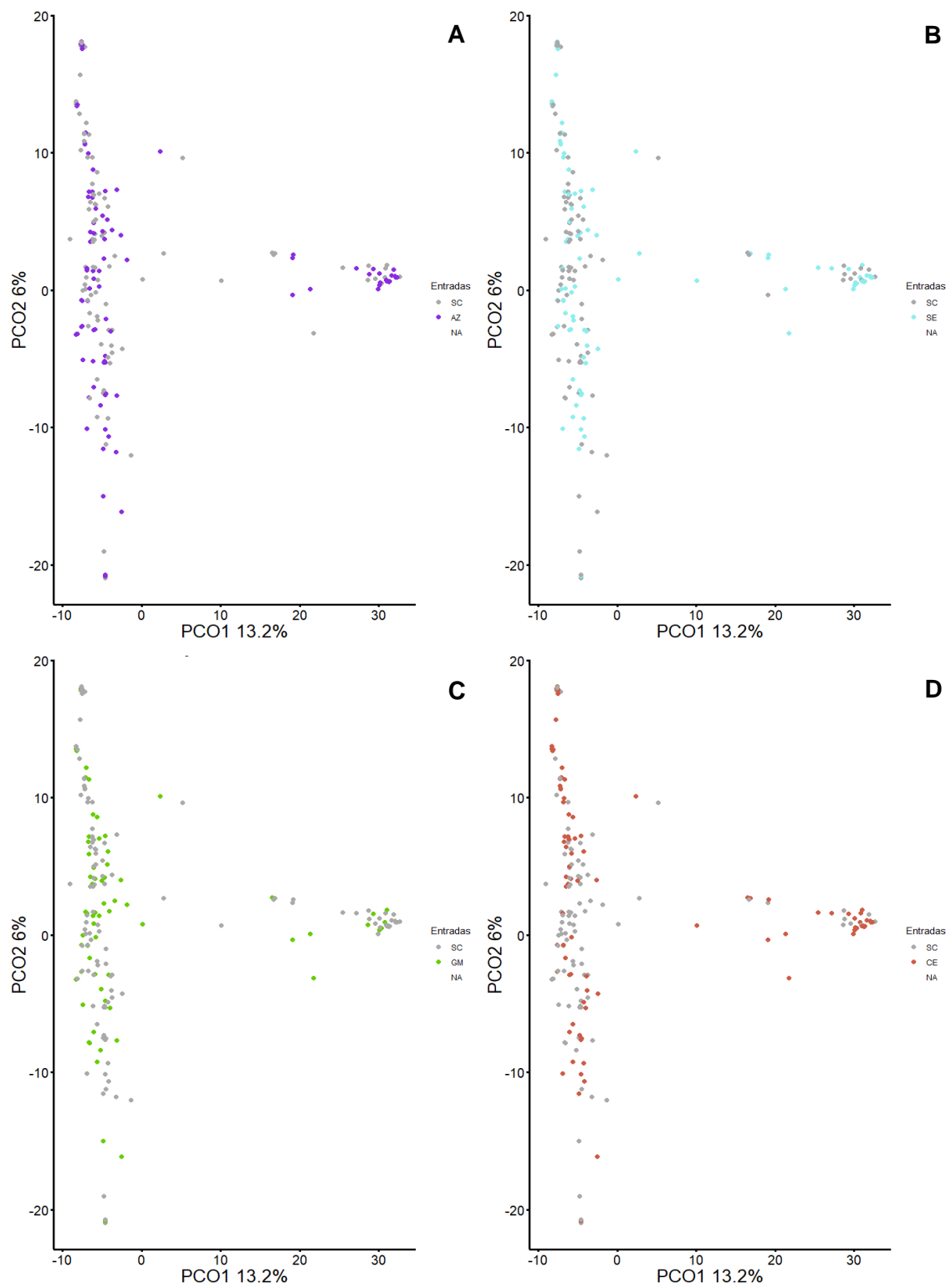


Figura 3. Análisis PCoA basado en los marcadores moleculares SNP.

En gris se representan las accesiones del subconjunto inicial SC y en color morado la colección al azar (A), azul claro la colección sin estratificar (B), verde la colección geográfica-morfológica (C) y rojo la colección estratificada (D).

Por otro lado se calculó el número de marcadores SNP fijados en cada una de las colecciones. Para ello se clasificaron los marcadores en función de la frecuencia del alelo menos frecuente (MAF) en el subconjunto inicial SC (Tabla 10). Los marcadores con alelos más frecuentes ($MAF > 0,1$) no se han fijado en ninguna colección. Los marcadores con alelos comunes ($MAF \geq 0,05$) son polimórficos en las colecciones SE y GM, mientras que se han fijado algunos en la AZ y CE. Para estos marcadores con $MAF \geq 0,05$, el alelo menos frecuente está presente en un mínimo de 9 accesiones. Los marcadores con alelos menos frecuentes ($MAF \geq 0,03$) que están presentes en al menos 6 accesiones, muy poco frecuentes ($MAF > 0,01$) que están presentes en al menos 2 accesiones, y raros ($MAF \leq 0,01$) que están presentes en 1 accesión, se han fijado en mayor o menor medida en todas las colecciones. En total, la colección SE es donde menos marcadores (6,3%) se han fijado, mientras que la colección GM es la que contiene el mayor número de marcadores fijados (10,1%).

Tabla 10. Distribución de marcadores SNP en el subconjunto inicial (SC), y colecciones al azar (AZ), sin estratificar (SE), geográfica-morfológica (GM) y estratificada (CE).

MAF	Nº de marcadores	Nº de marcadores fijados respecto al SC				
	SC	AZ	SE	GM	CE	
>0,1	6376	0	0	0	0	
$\geq 0,05$	8394	2	0	0	6	
$\geq 0,03$	10092	8	3	39	43	
>0,01	14002	586	546	976	685	
$\leq 0,01$	828	453	388	526	419	
Total	14830	1049	937	1541	1153	

El análisis de los marcadores SNP ha revelado que la diversidad genética es mayor en la colección SE (Tabla 11), mientras que la distancia de Gower media ha sido más baja en la colección GM. La distancia E-NE (distancia media entre cada genotipo y el más próximo de su conjunto) es mayor en todas las colecciones que en el SC, siendo la mayor la de la colección GM. La distancia A-NE (distancia media entre cada genotipo en el SC y el más próximo en la colección nuclear) es menor en las colecciones AZ y SE. (Tabla 11).

Tabla 11. Diversidad de Nei, distancia de Gower media y distancias E-NE y A-NE en el subconjunto inicial (SC), y colecciones al azar (AZ), sin estratificar (SE), geográfica-morfológica (GM) y estratificada (CE).

	Ht	Gw med.	E-NE	A-NE
SC	0,20	0,189	0,073	-
AZ	0,19	0,191	0,081	0,043
SE	0,20	0,192	0,088	0,042
GM	0,19	0,187	0,097	0,051
CE	0,19	0,190	0,081	0,051

5. Discusión

El objetivo principal del trabajo es el desarrollo de una colección nuclear española de variedades tradicionales de trigo blando (*T. aestivum* subsp. *vulgare* (Vill.)), que facilite el acceso a la variabilidad genética y fenotípica de la colección inicial a investigadores y mejoradores, y estimule las evaluaciones preliminares de germoplasma para caracteres de interés (Crossa *et al.*, 1993).

Esta colección debe representar las 522 variedades tradicionales españolas conservadas en el CRF para las cuales se dispone de datos de pasaporte y caracteres agromorfológicos cualitativos. Para crear la colección nuclear, el primer paso ha sido seleccionar un subconjunto inicial (SC) de 189 accesiones, representadas por un solo genotipo homocigótico, que incluían el rango geográfico (latitud, longitud y altitud), los nombres locales genéricos, y la variabilidad agromorfológica de las 522 variedades locales de partida. Este subconjunto inicial ha sido evaluado en un mismo ambiente para caracteres agromorfológicos cuantitativos y cualitativos, y su diversidad ha sido estudiada a nivel genómico con marcadores moleculares (DArT y SNP).

Estudios previos, con marcadores moleculares han permitido identificar cuatro poblaciones en el subconjunto inicial (SC) y han mostrado la alta variabilidad genética contenida en la colección de partida (Ruiz *et al.*, 2002; Pascual *et al.*, en revisión). Este hecho hace necesario determinar de manera precisa cual sería el tamaño adecuado para capturar la mayor parte de la diversidad presente. Para ello se emplearon simulaciones con el objeto de comprobar el porcentaje de marcadores polimórficos que serían retenidos en colecciones en función de su tamaño (Figura 1). Una vez se obtuvo este dato, para la elección final del número de entradas, se tuvieron en cuenta dos principios empleados habitualmente para seleccionar una colección nuclear (Zhang *et al.*, 2011). El primero está relacionado con la eficiencia de selección, por lo que hay que encontrar un nivel de compromiso entre obtener la máxima ganancia de diversidad genética y evitar acumular redundancia. El segundo se relaciona con la validación del muestreo, es decir se debe verificar que la variabilidad predominante esté incluida en la colección nuclear. En base al primer principio y a las simulaciones realizadas se estimó el tamaño óptimo en 94 entradas. Este tamaño supone un 18% del conjunto de partida (522 accesiones), y está dentro del rango (del 5 al 20%) que según Brown (1989a) permitiría retener al menos un 70% de la variabilidad. Este tamaño es similar a la colección nuclear española de trigo duro (Ruiz *et al.* 2013), y cuyo SC presenta una variabilidad genética similar al SC del presente trabajo ($H_t=0,32$ y $0,30$, para 190 y 189 genotipos de trigo duro y trigo blando, respectivamente),

estimada con los mismos marcadores moleculares por Pascual *et al.* (en revisión). Finalmente la calidad de las selecciones realizadas por 4 métodos diferentes fue evaluada verificando que la diversidad fenotípica y genética estuviera bien representada.

5.1. Validación de las colecciones nucleares.

La representatividad de los caracteres agromorfológicos y marcadores moleculares en las colecciones nucleares respecto a las colecciones de partida, es lo que permite evaluar la calidad de las mismas. La evaluación de la diversidad con marcadores genéticos es fundamental para validar una colección nuclear (Reyes-Valdés *et al.*, 2018). Un conjunto reducido de caracteres agromorfológicos, controlados por unos pocos genes, no son suficientes para estimar la variabilidad genética existente (Hu *et al.*, 2000; Balfourier *et al.*, 2007), además su expresión está influida por el ambiente. Sin embargo, es necesario tener en cuenta que para validar una colección los marcadores empleados deben ser diferentes a los usados para la selección (Odong *et al.*, 2013). En el presente trabajo se han empleado marcadores DArT para la creación de las colecciones, mientras que la evaluación de la variabilidad genética en cada colección se ha realizado con marcadores SNP. Las cuatro colecciones propuestas han sido capaces de capturar entre un 94 y un 90% de los polimorfismos tipo SNP, siendo la colección GM, seleccionada únicamente por caracteres agromorfológicos, la que presenta el mayor número de marcadores fijados (Tabla 10). Diferentes estudios realizados con marcadores SSR han demostrado que una retención de entre un 70 y un 97% de los polimorfismos presentes en la colección de partida garantiza una buena representación de la variabilidad genética (Balfourier *et al.*, 2007). De acuerdo con Crossa *et al.* (1993) y Zhang *et al.* (2011), una colección nuclear debe capturar todos los marcadores con alelos más comunes ($MAF > 0,1$), y la mayoría de los alelos comunes ($0,1 \geq MAF > 0,05$). Las cuatro colecciones analizadas incluyen la totalidad de los alelos de alta frecuencia, y el porcentaje de marcadores con alelos comunes que mantienen el polimorfismo varía entre el 100% (en SE y GM) y el 99,7% (en CE). Estos resultados indican que las cuatro colecciones tienen una buena representación de la variabilidad genética. Los marcadores con $MAF < 0,05$, incluyen alelos poco frecuentes y localizados, por lo que su inclusión requeriría aumentar en gran medida el tamaño de las colecciones nucleares y no suele merecer la pena (Marshall & Brown, 1975; Frankel *et al.*, 1995; Odong *et al.*, 2011; Zhang *et al.*, 2011). Las colecciones seleccionadas en el presente trabajo incluyen alrededor del 80% de estos alelos, excepto la GM (77%). Las colecciones SE y CE son las que capturan la mayor cantidad de polimorfismos raros ($MAF \leq 0,01$) debido a que las entradas se

seleccionaron maximizando la heterozigosidad (Tabla 10). Cuando analizamos la variabilidad genética en su conjunto, observamos que está mejor representada en las colecciones SE y CE (Figura 3), en concordancia con este hecho se encuentran los valores de diversidad genética H_t (Tabla 11).

Todas las colecciones incluyen la distribución geográfica del subconjunto inicial SC, excepto la SE que no incluye ninguna entrada de las islas Baleares (Tabla 5). Esto puede deberse a que las accesiones de las islas Baleares (Población 2) sean relativamente parecidas a otras accesiones provenientes de la Península y no aporten nueva diversidad, por lo que no habrían sido seleccionadas entre todas las del subconjunto SC. Los caracteres agromorfológicos cualitativos están bien representados, en general, en todas las colecciones. Sin embargo, la colección SE no incluye todas las clases fenotípicas de los caracteres longitud de barba y densidad de espiga. Aunque no se trata de diferencias significativas, puesto que las clases perdidas presentaban una frecuencia muy baja en el SC. Para evaluar los caracteres agromorfológicos cuantitativos se estimaron los parámetros propuestos por Hu *et al.* (2000). De acuerdo con estos autores una colección nuclear representa mejor la colección inicial cuanto más similares sean las medias de ambas colecciones (valores bajos de MD) y mayor sea la variación capturada (valores elevados de CR, VD y VR) en la colección nuclear. Según este criterio, las colecciones CE y SE son las que mejor representan la variabilidad inicial, mientras que la GM no incluye algunos valores extremos de los caracteres cuantitativos (Tabla 8).

Además de los marcadores moleculares SNP, se han evaluado otros marcadores moleculares que se consideran interesantes para los mejoradores. Los alelos de gluteninas de alto peso molecular tienen mucho interés en trigo blando por su asociación con la calidad panadera de la harina (Anjum *et al.*, 2007), aunque estas proteínas representan solo un 10% de la variabilidad de las proteínas de reserva (Branlard *et al.*, 2003). En este caso, la colección que representa mejor esta variabilidad alélica es la colección CE (Tabla 9). Los alelos de vernalización determinan el hábito de crecimiento invernal o de primavera en trigo blando, y por tanto están implicados en la adaptación del trigo a distintos ambientes (Distelfeld *et al.*, 2007). En este caso no se observaron diferencias entre el SC y las colecciones, por lo que no es un factor determinante para discriminar entre las colecciones (Figura 2 I).

Finalmente, para las cuatro colecciones se calculó la distancia media de Gower, basada en los marcadores moleculares (SNP, loci *Glu-1* y *Vrn-A1*), y los caracteres fenotípicos (cualitativos y cuantitativos). Todas las colecciones mostraron una

distancia media mayor entre sus entradas que la del SC, excepto la GM (Tabla 11). A partir de la distancia de Gower se determinaron las distancias A-NE y E-NE que permiten analizar las colecciones en su conjunto (Odong *et al.*, 2013). El menor valor para la distancia A-NE observado en la SE indica que esta colección es la que mejor representa las entradas del SC (Tabla 11) y maximiza la diversidad (Odong *et al.*, 2013). Todas las colecciones presentaron un valor de distancia E-NE mayor que el SC, siendo mayores los de las colecciones GM y SE, lo que indica que las entradas de estas colecciones nucleares son las más diferentes entre sí (Odong *et al.*, 2013). En el caso de la colección GM, hay que tener en cuenta que el análisis PCoA muestra que dicha colección incluye un mayor número de entradas de grupos distantes (Figura 3 C), lo que puede explicar que la distancia entre las entradas sean mayores, a pesar de no cubrirse la totalidad de la diversidad existente.

Los resultados han demostrado que cuando no se tienen grandes medios, una colección seleccionada al azar puede ser útil si se parte de un subconjunto inicial bien elegido, como ha ocurrido en este trabajo. Una colección basada únicamente en datos geográficos y morfológicos es útil, pero puede perder parte de la diversidad que no está representada fenotípicamente por los caracteres evaluados.

El algoritmo Core Hunter ciertamente ha resultado de interés para el desarrollo de una colección nuclear en trigo blando, tanto para la creación de la colección SE como CE, que son las que han mostrado mejor representación de la variabilidad del subconjunto inicial SC. De hecho 75 de las 94 entradas son comunes en ambas colecciones. La colección CE, creada con el método estratificado y seleccionando las entradas de cada población de forma independiente, representa algo mejor los caracteres fenotípicos cualitativos, alelos de gluteninas y la variación geográfica. En cambio, la colección sin estratificar (SE) captura algo mejor la variabilidad genética y muestra una mayor distancia entre las entradas. Este resultado contrasta con el obtenido en la colección nuclear de trigo duro (Ruiz *et al.*, 2013), donde la colección CE capturaba mayor diversidad genética. Este hecho podría ser debido a las diferencias en la estructura poblacional entre las colecciones de ambas especies. En el caso de trigo blando las accesiones de la población 2 (la más diversa) presentan un alto grado de mezcla con las poblaciones 1 y 3 (Pascual *et al.*, en revisión), mientras que en trigo duro la colección incluía tres subespecies y poblaciones más diferenciadas. En resumen, teniendo en cuenta los resultados obtenidos, se propone que la colección española de variedades tradicionales de trigo blando se base en la colección SE o la CE seleccionadas en el presente trabajo.

6. Conclusiones

1. Empleando simulaciones basadas en el algoritmo Core Hunter se ha conseguido estimar el tamaño óptimo para la colección nuclear española de trigo blando, que resultó ser de 94 accesiones.
2. Los métodos de selección de colecciones nucleares aleatoriamente o con datos geográficos y morfológicos son menos eficientes a la hora de conseguir representar en su totalidad la diversidad genética contenida en el subconjunto inicial
3. A diferencia de lo ocurrido en el caso de la colección nuclear española de trigo duro, cuando trabajamos con trigo blando no existen grandes diferencias entre los dos métodos fundamentados en la maximización de la heterocigosidad empleados en la selección de la colección no estratificada y la colección estratificada.
4. El presente trabajo ha permitido diseñar cuatro colecciones nucleares y evaluarlas. En base a los resultados obtenidos, la colección española de variedades tradicionales de trigo blando (*T. aestivum* subsp. *vulgare* (Vill.)) debería escogerse entre las colecciones nucleares sin estratificar o estratificada.

7. Bibliografía.

- Allen R.C. (2002) Revolución en los campos: La reinterpretación de la Revolución Agrícola inglesa. Sociedad Española de Historia Agraria (SEHA). *Historia agraria: Revista de agricultura e historia rural*, nº 26, pp. 13-32.
- Anjum F.M., Khan M.R., Din A., Saeed M., Pasha I. & Arshad M.A. (2007) Wheat gluten: high molecular weight glutenin subunits--structure, genetics, and relation to dough elasticity. *J Food Sci.* 2007 Apr; 72(3): R56–R63. doi: 10.1111/j.1750-3841.2007.00292.x
- Aparicio N., Alvaro F., Sillero J., Ruiz M., López P., Catedra M. & Codesal, P. (2010). Bread wheat (*Triticum aestivum* L.) core collection based in Spanish landraces. En: Dzyubenko NI, editor. Proceedings 8th International wheat conference. June 1-4; St. Petersburg, Russia. N.I. Vavilov Research Institute of Plant Industry. p. 85
- Appels R, Eversole K, Feuillet C, Keller B, Rogers J, Stein N, *et al.* (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361(6403):eaar7191. DOI:10.1126/science.aar7191
- Asociación Española de Técnicos Cerealistas, AETC. Visto el 5 de abril de 2019. <http://www.aetc.es>
- Balfourier F., Roussel V., Strelchenko P., Exbrayat-Vinson F., Sourdille P., Boutet G., Koenig J., Ravel C., Mitrofanova O., Beckert M. & Charmet G.. (2007) A worldwide bread wheat core collection arrayed in a 384-well plate. *Theor. Appl. Genet.* 114:1265–1275. doi:10.1007/s00122-007-0517-1
- Branlard G., Dardevet M., Amiour N. & Gilberto I. (2003) Allelic diversity of HMW and LMW glutenin subunits and omegagliadins in French bread wheat (*Triticum aestivum* L.). *Genetic Resources and Crop Evolution.* 50: 669. <https://doi.org/10.1023/A:1025077005401>
- Brown, A.H.D. (1989a) Core collections: A practical approach to genetic resources management. *Genome* 31:818–824. doi:10.1139/g89-144
- Brown A.H.D. (1989b) The case for core collections. pp 136-156. En: Brown A.H.D., Frankel O.H., Marshall D.R. & Williams J.T. (Ed.) The use of plant Genetic Resources. *Cambridge University press*, Cambridge.
- Carrillo J.M., Díez M. J., Pérez de la Vega M., Nuez F. (2010) Recursos fitogenéticos en la mejora genética vegetal. pp 9-49. En: Carrillo J.M., Díez M.J., Pérez de la Vega M., Nuez F. (Ed.) MEJORA GENÉTICA Y RECURSOS FITOGENÉTICOS: "Nuevos avances en la conservación y utilización de los Recursos Fitogenéticos". Madrid, España: Ministerio de Medio ambiente y Medio Rural y Marino.

- Carrillo J.M., Vázquez J.F., Rodríguez M. & Ruiz M. (2006) Mejora de la calidad del trigo. pp 127-164. En: Llácer G., Díez M.J., Carrillo J.M. & Badenes M.L. (Ed.) MEJORA GENÉTICA DE LA CALIDAD EN PLANTAS. Valencia, España: Editorial de la Universidad Politécnica de Valencia.
- CBD. Convenio sobre la diversidad biológica. Visto el 26 de julio de 2019 en: <https://www.cbd.int/history/default.shtml>
- Chávez Esponda D., Miranda Cabrera I., Varela Nualles M & Fernández L. (2010) Utilización del análisis de cluster con variables mixtas en la selección de genotipos de maíz (*Zea mays*). *Revista Investigación Operacional*, Vol. 30, Nº. 3, 209 - 216
- Chen F., Gao M., Zhang J., Zuo A., Shang X. & Cui D. (2013) Molecular characterization of vernalization and response genes in bread wheat from the Yellow and Huai Valley of China. *BMC plant biology* 2013; 13(1):199.
- Chu C.G., Tan C.T., Yu G.T., Zhong S., Xu S.S. & Yan L. (2011) A Novel Retrotransposon Inserted in the Dominant Vrn-B1 Allele Confers Spring Growth Habit in Tetraploid Wheat (*Triticum turgidum* L.). *G3 (Bethesda)* 2011 Dec;1(7):637-645.
- CRF-INIA. Visto el 24 de abril de 2019 en: <http://wwwsp.inia.es/Investigacion/centros/crf/Paginas/CRF.aspx>
- Crossa J., Hernández C.M., Bretting P., Eberhart S.A. & Taba S (1993) Statistical genetic considerations for maintaining germplasm collections. *Theor. Appl. Genet.* 86:673–678. doi:10.1007/BF00222655
- Cubero J.I. (2000) Historia de la biotecnología vegetal. I Jornadas sobre productos transgénicos en agricultura. Córdoba.
- Di Fonzo N., Ravaglia S., De Ambrogio E., Blanco A. & Troccoli A. (2005) Durum wheat improvement in Italy. pp 825-881. En: Royo C., Nachit M.M., Di Fonzo N., Araus J.L., Pfeiffer W.H. & Slafer G.A. (Ed.) DURUM WHEAT BREEDING. "Current approaches and future strategies". Vol 2. Food Product Press. *Crop Science*.
- Díez, C.M., Imperato A., Rallo L., Barranco D. & Trujillo I. (2012) Worldwide core collection of olive cultivars based on simple sequence repeat and morphological markers. *Crop Sci.* 52:211–221. doi:10.2135/cropsci2011.02.0110
- Díez M.J. & Valcarcel J.V. (2010) La red de bancos de germoplasma. pp 273-294. En: Carrillo J.M., Díez M.J., Pérez de la Vega M., Nuez F. (Ed.) MEJORA GENÉTICA Y RECURSOS FITOGENÉTICOS: "Nuevos avances en la conservación y utilización de los Recursos Fitogenéticos". Madrid, España: Ministerio de Medio ambiente y Medio Rural y Marino.

- Goudet J. (2005) Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, 5(1):184-186.
- Gower J. (1971) A general coefficient of similarity and some of its properties. *Biometrics* 27, 857-872. 1971.
- Gruber B., Unmack P.J., Berry O.F. & Georges A. (2018 a) "dartR: an R package to facilitate analysis of SNP data generated from reduced representation genome sequencing". *Molecular Ecology Resources*, vol. 18(3):691-699
- Gruber B., Georges A., Unmack P.J., Clark L.V. & Berry O. (2018 b) "dartR: Importing and Analysing SNP and Silicodart Data Generated by Genome-Wide Restriction Fragment Analysis" R package version 1.0.5. Disponible en internet: <https://CRAN.R-project.org/package=dartR>, 2018, Consultado el 20-Mar-2019.
- Hu, J., Zhu J. & Xu H.M. (2000) Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. *Theor. Appl. Genet.* 101:264–268. doi:10.1007/s001220051478
- IBPGR. 1985. Revised descriptor list for wheat (*Triticum ssp*). *International Board for Plant Genetic Resources*, Rome, Italy.
- Jaccoud D., Peng K., Feinsein D. & Kilian A. (2001). Diversity Arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Research*, vol. 29, no. 4, p 25e
- Kilian A., Wenzl P., Huttner E., Carling J., Xia L., Blois H., Caig V., Heller-Uszynska K., Jaccoud D., Hopper C., Aschenbrenner-Kilian M., Evers M., Peng K., Cayla C., Hok P. & Uszynski G. (2012) Diversity arrays technology: a generic genome profiling technology on open platforms. *Springer*. pp. 67-89.
- Llácer G. & Badenes M.L. (2010) Conservación de plantas de multiplicación vegetativa. pp 173-210. En: Carrillo J.M., Díez M.J., Pérez de la Vega M., Nuez F. (Ed.) MEJORA GENÉTICA Y RECURSOS FITOGENÉTICOS: "Nuevos avances en la conservación y utilización de los Recursos Fitogenéticos". Madrid, España: Ministerio de Medio ambiente y Medio Rural y Marino.
- Maechler M., Rousseeuw P., Struyf A., Hubert M. & Hornik, K. (2018) cluster: Cluster Analysis Basics and Extensions. R package version 2.0.7-1.
- Mallor C., Igartua E. & Errea P. (2014) Las variedades tradicionales en el panorama actual de la mejora y la producción sostenible. pp 35-61. En: Socías R., Rubio M.J., Garcés A., Mallor C. & Álvarez J.M. (Ed.) LA OBTENCIÓN DE VARIEDADES: DESDE LA MEJORA CLÁSICA HASTA LA MEJORA GENÉTICA MOLECULAR. Centro de Investigación y Tecnología

Agroalimentaria, Sociedad Española de Ciencias Hortícolas, *Sociedad Española de Genética*. ISBN 978-84-8380-320-2.

MAPA. Anuario de Estadística. Visto el 20 de julio de 2019 en: <https://www.mapa.gob.es/es/estadistica/temas/publicaciones/anuario-de-estadistica/2017/>

Marshall D.R. & Brown A.H.D. (1975) Optimum sampling strategies in genetic conservation. En: O.H. Franked & J.G. Hawkes (Ed). *CROP GENETIC RESOURCES FOR TODAY AND TOMORROW*. *Cambridge Univ. Press*, Cambridge, UK. p. 53–80.

Martín I. (2001) Conservación de Recursos Fitogenéticos. Ministerio de Agricultura, Pesca y Alimentación. *Hojas divulgadoras*, 2114.

Maxted N., Ford-LLOYD B.V. & Hawkes J.G. (1997). *Plant Genetic Conservation. The in situ approach*. *Chapman & Hall*, Londres, UK.

MITECO, Ministerio para la transición ecológica. Visto el 20 de agosto de 2019. <https://www.miteco.gob.es/es/calidad-y-evaluacion-ambiental/temas/biotecnologia/organismos-modificados-geneticamente-omg-/protocolo-cartagena/>

Nei M. (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*. 70(12):3321.

Odong T.L., Jansen J., Van Eeuwijk F.A. & Van Hintum T.J.L. (2013) Quality of core collections for effective utilisation of genetic resources review, discussion and interpretation. *Theor. Appl. Genet.* 126:289–305. doi:10.1007/s00122-012-1971-y

Odong, T.L., van Heerwaarden J., Jansen J., van Hintum T.J.L. & van Eeuwijk F.A. (2011) Statistical techniques for defining reference sets of accessions and microsatellite markers. *Crop Sci.* 51:2401–2411. doi:10.2135/cropsci2011.02.0095

Pacheco A., Alvarado G., Rodríguez F. & Burgueño, J. (2016) BIO-R (Biodiversity analysis with R for Windows) Version 2.0, hdl:11529/10820, CIMMYT *Research Data & Software Repository Network*, V8

Pascual L., Ruiz M., López-Fernández M., Pérez-Peña H., Benavente E., Vázquez J.F. & Giraldo P. (En revisión) Genomic analysis of Spanish wheat landraces reveals their huge potential for breeding. *BMC Genomics*. <https://doi.org/10.21203/rs.2.9775/v1>

Payne P.I., Law C.N. & Mudd E.E. (1980) Control by homoeologous group 1 chromosomes of the high-molecular-weight subunits of glutenin, a major protein of wheat endosperm. *Theoretical and Applied Genetics*. 58, 113–120

Peng J.H., Sun D. & Nevo E. (2011) Domestication evolution, genetics and genomics in wheat. *Molecular Breeding* 28: 281. <https://doi.org/10.1007/s11032-011-9608-4>

- Pérez de la Vega M. (2010) La domesticación de las plantas cultivadas y la base genética de los cultivos. pp 50-100. En: Carrillo J.M., Diez M.J., Pérez de la Vega M., Nuez F. (Ed.) MEJORA GENÉTICA Y RECURSOS FITOGENÉTICOS: "Nuevos avances en la conservación y utilización de los Recursos Fitogenéticos". Madrid, España: Ministerio de Medio ambiente y Medio Rural y Marino.
- Pérez Peña H. (2018). Caracterización a nivel genómico de variedades españolas de trigo duro mediante tratamiento computacional de datos de secuenciación Trabajo Final de Grado. ETSIAAB, Universidad Politécnica de Madrid.
- R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Raj A., Stephens M., & Pritchard JK. (2014) Variational inference of population structure in large SNP datasets. *Genetics*. 114.164350.
- Rao N.K., Hanson J., Dulloo M.E., Ghosh K., Nowell D. & Larinde M. (2006) Manual para el manejo de semillas en bancos de germoplasma. Manuales para Bancos de Germoplasma No. 8 *Bioversity International*. Rome (Italy):
- Reyes-Valdés M.H., Burgueño J., Singh S., Martínez O. & Sansaloni C.P. (2018) An informational view of accession rarity and allele specificity in germplasm banks for management and conservation. *PLoS One*. 13(2):e0193346. doi:10.1371/journal.pone.0193346
- Rosenberg NA. (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* 4(1):137-138.
- Ruiz M., Giraldo P., Royo C. & Carrillo J. M. (2013) Creation and validation of the Spanish durum wheat core collection. *Crop Science*, 53(6), 2530-2537.
- Ruiz M., Rodriguez-Quijano, M., Metakovsky E.V., Vazquez F. & Carrillo JM. (2002) Polymorphism, variation and genetic identity of Spanish common wheat germplasm based on gliadin alleles. *Field Crops Res.* 79: 185-196. Doi: 10.1016/S0378-4290(02) 00139-9
- Sánchez-García M., Fanny A., Martín-Sánchez J.A., Sillero J.C., Escribano J. & Royo C. (2012) Breeding effects on the genotype x environment interaction for yield of bread wheat grown in Spain during the 20th century. *Field Crops Res.*, 126: 79-86.
- Sánchez-Monge E. (1957) Catálogo genético de trigo españoles. Ministerio de Agricultura, Dirección General de coordinación, crédito y capacitación agraria. Madrid.
- Thachuk C., Crossa J., Franco J., Dreisigacker S., Warburton M. & Davenport, G. (2009) Core Hunter: An algorithm for sampling genetic resources based on multiple genetic measures. *BMC bioinformatics*. 10. 243. 10.1186/1471-2105-10-243.

- Uauy C., Distelfeld A., Fahima T., Blechl A. & Dubcovsky J. (2006) A NAC Gene Regulating Senescence Improves Grain Protein, Zinc, and Iron Content in Wheat. *Science*. Vol. 314, Issue 5803, pp. 1298-1301. DOI: 10.1126/science.1133649
- Van Hintum T.J.L., Brown A.H.D., Spillane C. & Hodgkin T. (2003) Colecciones nucleares de Recursos Fitogenéticos. *Boletín Técnico N° 3, Instituto internacional de Recursos Fitogenéticos (IPGRI)*. Roma.
- Van de Wouw M., Kik C., Van Hintum T.J.L., Van Treuren R., & Visser B. (2009) Genetic erosion in crops: concept, research results and challenges. *Plant Genetic Resources* 8, 1–15. doi: 10.1017/S1479262109990062
- Vares Megino F. (1984) Algunas enfermedades criptogámicas del maíz. *Hojas Divulgadoras del Ministerio de Agricultura, Pesca y Alimentación*. N° 8
- Zeven A.C. (1998) Landraces: A review of definitions and classifications. *Euphytica* 104: 127–139. Netherlands.
- Zhang H., Zhang D., Wang M., Sun J., Qi Y., Li J., Wei X., Han L., Qiu Z., Tang S. & Li Z. (2011) A core collection and mini core collection of *Oryza sativa* L. in China. *Theor. Appl. Genet.* 122:49–61. doi:10.1007/s00122-010-1421-7

ANEJO 1. DESCRIPTORES DE CARACTERIZACION DE TRIGO.

Caracterización realizada en el CRF, Finca la Canaleja, Alcalá de Henares, Madrid.

Responsable de la caracterización Rosario Fité.

Toma de datos: Rosario Fité, M^a Jesús de Tomás.

Coordinación: Magdalena Ruiz

Basados fundamentalmente en los descriptores para trigo publicados por el IBPGR (1985) y datos anteriores de otras evaluaciones en el INIA.

Número de banco. (NUMBAN).

Lugar de caracterización (LUGCAR): Lugar en el que se realizó la caracterización (acrónimo de la Institución).

Año en el que se realizó la caracterización (AÑOCAR): Lugar en el que se realizó la caracterización (acrónimo de la Institución).

Nombre de la persona responsable de la caracterización (NOMCAR).

Fecha de siembra (FE_SIE) = fecha.

Día-mes-año. Se tomará aquella en que a partir de la siembra el terreno reúna las condiciones adecuadas de humedad para la germinación de las semillas.

Faja. (FAJA):

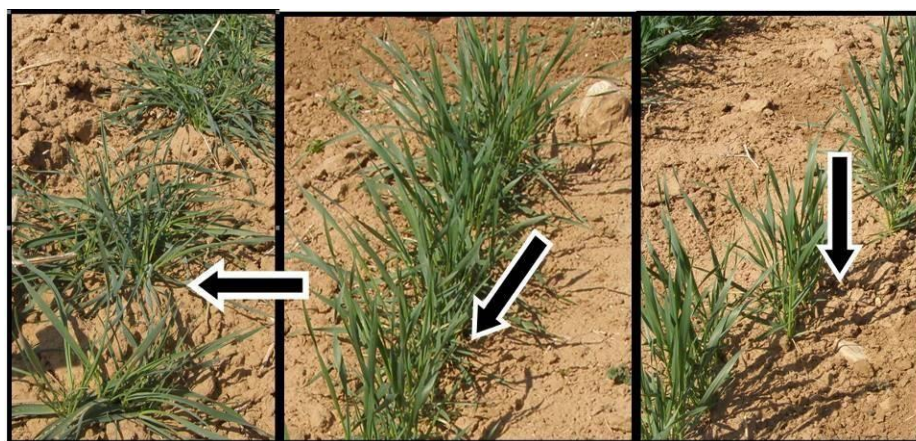
Nº de la faja en que fue sembrada la muestra en el campo.

Testigo (T).

Muestra testigo, señalada con una "T". Si hay dos testigos por faja se indica "T1" o "T2" Según su posición al principio o final de la faja

Hábito de crecimiento (HA_CRE). Evaluado durante el ahijamiento (4-6 hijuelos aproximadamente 60 días tras la nascencia).

- 3- postrado
- 5- intermedio
- 7- erecto



HA_CRE 3

HA_CRE 5

HA_CRE 7

Días hasta espigado (DI_ESP). Desde la fecha de siembra hasta que media espiga emerja de la vaina en el 50% de las plantas de la parcela.

Días hasta la maduración (DI_MAD). Desde fecha de siembra hasta que están maduras el 50% de plantas de la parcela.

Altura de la planta (cm) (AL_PLA). Tomada en la madurez desde el suelo a la parte final de la última espiga sin incluir las aristas. Valor medio de cinco plantas.



AL_PLA: 90 cm

Rugosidad de las barbas (RU_BAR). Medida en laboratorio en madurez. Este carácter se toma por tacto, no es necesario usar lupa.

- 1- ásperas
- 2- lisas

Color de las barbas (CO_BAR). Medida en laboratorio en madurez.

- 1- blancas
- 2- negras en la base
- 3- negras (gris/negras para ssp. spelta)
- 4- rojas a marrones



CO_BAR 1

Longitud de las barbas (LO_BAR). Medida en el tercio central de la espiga en laboratorio y madurez. Valor medio de 5 espigas.

- 1- ausencia. C si son caducas
- 2- semiaristadas: 1-3 cm
- 3- aristadas: 3-8 cm
- 4- aristas largas: >8 cm



Densidad de la espiga (DE_ESP). Medida en el tercio central de la espiga en laboratorio y madurez. Valor medio de 5 espigas:

- 3- laxa ($D < 20$)
- 5- media ($20 - 25$)
- 7- densa ($25 - 31$)
- 9- muy densa ($D > 31$)

$$D = \frac{(n-1) \cdot 100}{L}$$

n: nº de espiguillas

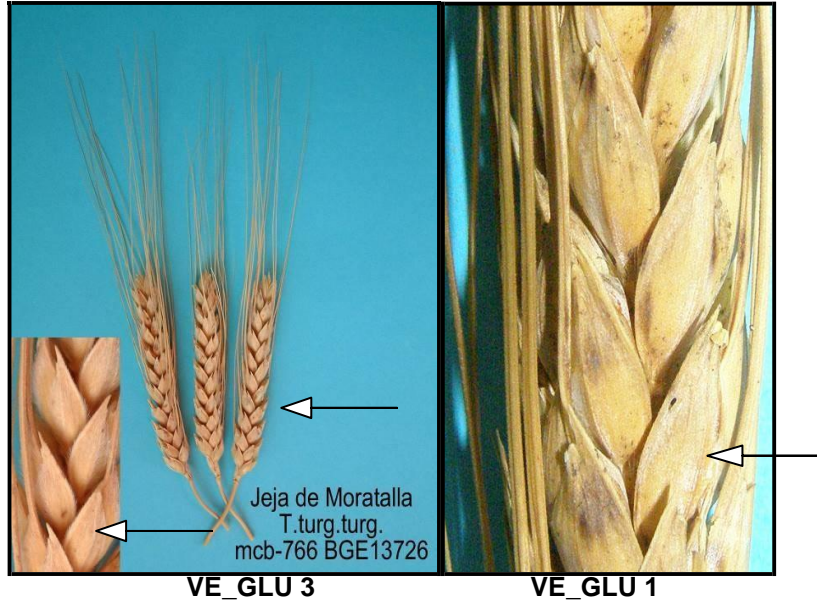
L: longitud del raquis (mm) [desde el inicio, incluidas las estériles, hasta la base de la espiguilla apical].



DE_ESP 9

Vellosidad de la gluma (VE_GLU). Medida en el tercio central de la espiga en laboratorio y madurez. Se toma de la gluma inferior del tercio central de la espiga.

- 1- lampiña : 1.1 brillante (sólo *T. monococcum*)
1.2 opaca (sólo en *T. monococcum*)
- 2- pubescente
- 3- vellosa (pelos muy largos)



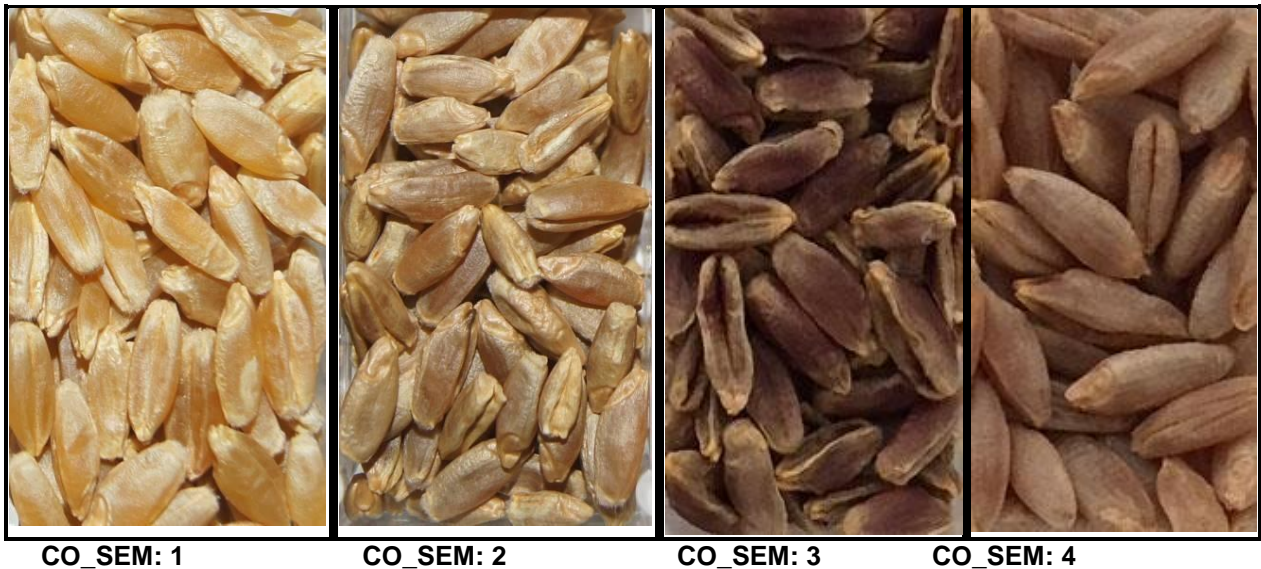
Color de la gluma (CO_GLU). Medida en el tercio central de la espiga en laboratorio y madurez.

- 1, blanco
- 2, rojo/marrón
- 3, púrpura a gris/negro



Color de la semilla (CO_SEM). Medida en el tercio central de la espiga en laboratorio y madurez.

- 1, blanco
- 2, rojo (rojo/marrón para ssp. spelta)
- 3, púrpura
- 4, gris (sólo ssp. spelta)



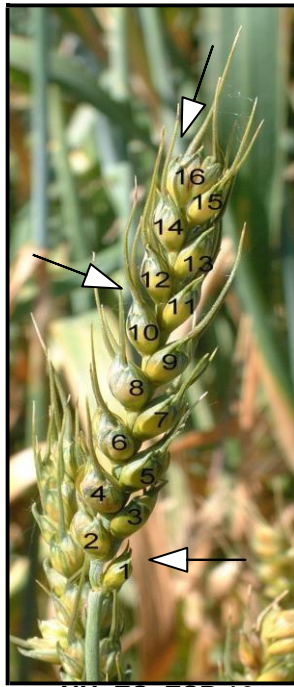
Si es difícil decidir, se puede usar la prueba del NaOH. Colocar los granos en una placa Petri+25 ml. de NaOH 5% durante 60-90 min. Los granos rojos se ponen marrón anaranjado oscuro, los blancos viran a amarillos.

Longitud de la espiga en mm (LO_ESP). Medida en el tercio central de la espiga en laboratorio y madurez, desde la base (sin incluir las espiguillas estériles), hasta el extremo de la espiguilla apical (sin incluir aristas). Valor medio de 5 espigas.



LO_ESP 90 mm

Número de espiguillas por espiga (NU_ES_ESP). Medida en el tercio central de la espiga en laboratorio y madurez. N° total de espiguillas excluyendo las basales estériles, si las hay. Valor medio de 5 espigas.



NU_ES_ESP 16

OBSERV: Observaciones.