



POLITÉCNICA



E.T.S. DE INGENIERÍA AGRONÓMICA,
ALIMENTARIA Y DE BIOSISTEMAS

UNIVERSIDAD POLITÉCNICA DE MADRID
ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA AGRONÓMICA,
ALIMENTARIA Y DE BIOSISTEMAS
GRADO EN BIOTECNOLOGÍA

DEPARTAMENTO DE BIOLOGÍA – BIOTECNOLOGÍA VEGETAL

***Identificación y análisis de regiones genómicas
que controlan la arquitectura radicular en
Triticum aestivum***

TRABAJO FIN DE GRADO

Autor: Alejandro Sánchez Cano

Tutora: Patricia Giraldo Carbajo

Julio de 2021



POLITÉCNICA



E.T.S. DE INGENIERÍA AGRONÓMICA,
ALIMENTARIA Y DE BIOSISTEMAS

UNIVERSIDAD POLITÉCNICA DE MADRID
ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA
AGRONÓMICA, ALIMENTARIA Y DE BIOSISTEMAS

GRADO DE BIOTECNOLOGÍA

IDENTIFICACIÓN Y ANÁLISIS DE REGIONES GENÓMICAS
QUE CONTROLAN LA ARQUITECTURA RADICULAR EN
TRITICUM AESTIVUM

TRABAJO FIN DE GRADO

Alejandro Sánchez Cano

MADRID, 2021

Directora: Patricia Giraldo Carbajo
Dpto. de Biotecnología – Biología Vegetal



POLITÉCNICA



E.T.S. DE INGENIERÍA AGRONÓMICA,
ALIMENTARIA Y DE BIOSISTEMAS

IDENTIFICACIÓN Y ANÁLISIS DE REGIONES GENÓMICAS QUE CONTROLAN LA ARQUITECTURA RADICULAR EN *TRITICUM AESTIVUM*

**Memoria presentada por Alejandro Sánchez Cano para la obtención del título
de Graduado en Biotecnología por la Universidad Politécnica de Madrid**

Fdo: Alejandro Sánchez Cano

VºBº Tutora y directora del TFG

**Dra. Patricia Giraldo Carbajo
Dpto. de Biotecnología – Biología Vegetal
ETSIAAB – Universidad Politécnica de Madrid**

Madrid, 25, Junio, 2020

AGRADECIMIENTOS



e gustaría empezar agradeciendo este trabajo a la Dra. Magdalena Ruiz, responsable de las colecciones de trigo del CRF-INIA y al Dr. Juan M. González de la UAH por el préstamo de los datos fenotípicos.

Sin duda alguna, también me gustaría expresar mi más profundo agradecimiento al laboratorio de genética por el excelente trato recibido. Gracias a la Paqui por soportarme en la M-30 y tener un corazón que no te cabe en el pecho. Espero haber sido un buen esbirro. A Mario por hacer más llevaderos los días en el laboratorio y en el campo. A Mati por ser un ser de luz que me ha sacado de más atolladeros de los que puedo contar y siempre con una adorable amabilidad. A Paco porque las conversaciones contigo son risas continuas. A Naima porque nosotros éramos los reyes de la pecera. A Alba por ser la mejor DJ de toda la universidad. A mis trillizas Sandra, Ale e Iván porque vosotros habéis sido la clave para que esta temporada haya sido la mejor de la carrera, nadie ha tenido unos compañeros de laboratorio mejores y me jugaría la pierna izquierda. Especial mención para mi tutora Patricia, por guiarme en todo este trayecto con una mano y con la otra mostrarme tu devoción por la ciencia y la vida. He aprendido muchísimo con vosotros y espero haber dejado huella. Corazones de Paqui para todos.

A mis queridos padres, por ser un constante faro iluminando el buen camino todos estos años. Siento un profundo orgullo de ser una mezcla perfecta de los dos y, aunque vaya a estar un par de añitos fuera, tranquilo Pocho que os llamaré mucho, haré mucha calistenia y beberé kéfir todos los días. A la gochaca de mi hermana, por ser un pilar fundamental en mi vida a la que quiero con locura y a la que seguiré molestando cuando tenga 80 años y esté aburrido de jugar al ajedrez. A todos mis 5 abuelos, que sin ellos no habría sido posible porque ellos han definido mi personalidad, me han dado las ganas de aprender y ser un sabihondo, de seguir dándolo todo, de disfrutar de la vida y del campo. Y no me olvido del Guengli, a la que yo también me vuelvo loco cuando la veo.

Lo mejor de haber entrado en esta carrera son, sin lugar a una mínima duda, mis amigos de Gente Depravada, con los que he compartido momentos y experiencias inolvidables. Porque vosotros erais la razón por la que cogía el metro para ir a clase todos los días y veros vuestras carusas. En todo momento me habéis brindado un soporte emocional inestimable, desinteresado y cándido por el que siempre estaré con vosotros en las malas aportando un hombro en el que echarse y en las buena celebrando vuestros éxitos, que me juego la pierna derecha a que serán muchos.

Por último, todo esto no habría sido posible sin que Chyanne se hubiera levantado 20 años y decidiera que iba a revolucionar el mundo de la música con Salomé y Torero.

ÍNDICE DE CONTENIDO

ÍNDICE DE FIGURAS.....	vi
ÍNDICE DE TABLAS.....	vii
LISTA DE ABREVIATURAS	viii
SUMMARY.....	x
CAPÍTULO 1. INTRODUCCIÓN	1
1.1 Importancia del trigo	1
1.2 Importancia de la variabilidad genética y las variedades locales en los programas de mejora	2
1.3 Arquitectura radicular.....	2
1.4. Genómica aplicada a la mejora de especies vegetales.....	3
1.5. Objetivos.....	5
CAPÍTULO 2. MATERIAL Y MÉTODOS.....	6
2.1 Material vegetal.....	6
2.2 Datos fenotípicos	6
2.3 Datos genotípicos	7
2.4 Análisis de los datos fenotípicos.....	8
2.4.1 Normalidad y homogeneidad de varianza.....	8
2.4.2 Análisis de correlación.....	8
2.4.3 ANOVA.....	9
2.4.4 PCA.....	9
2.5 Análisis de asociación genómica.....	9
2.6 Análisis de asociación de marcadores individuales	10
2.7 Análisis de marcadores asociados	10
CAPÍTULO 3. RESULTADOS Y DISCUSIÓN	12
3.1 Material vegetal.....	12
3.2 Análisis de datos fenotípicos	13
3.2.1 Análisis de correlación	14
3.2.2 Normalidad y homogeneidad de varianzas	16
3.2.3 ANOVA.....	18
3.2.4 PCA.....	19
3.3 Análisis de asociación genómica.....	20
3.4 Análisis de asociación de marcadores individuales	24
3.5 Análisis de marcadores asociados	27
CAPÍTULO 4. CONCLUSIONES.....	28
CAPÍTULO 5. BIBLIOGRAFÍA	29

ÍNDICE DE FIGURAS

Figura 1. Distribución de las principales áreas de cultivo de trigo	1
Figura 2. Ensayo in vitro de fenotipado de la arquitectura radicular en plántulas	7
Figura 3. A Gráfico STRUCTURE; B Lugar de origen de las variedades que conforman la colección de 189 accesiones	8
Figura 4. Análisis PCoA comparativo	12
Figura 5. Diagramas de cajas y bigotes resultado de la estadística descriptiva para los 10 caracteres analizados	13
Figura 6. Matriz triangular de correlación	15
Figura 7. Diagramas de densidad	17
Figura 8. Diagramas de cajas y bigotes resultado de la comparación de poblaciones genéticas	19
Figura 9. Representación del PCA en función de la población genética	20
Figura 10. Representación del PCA en función del tipo de floración	21
Figura 11. Manhattan plots de los estudios de asociación genómica (GWAS) para los caracteres MxRA (A) y SDSS (B).	22
Figura 12. QQPlots de los caracteres analizados	24

ÍNDICE DE TABLAS

Tabla 1. Resultado del <i>Single Marker Analysis</i> (SMA) de los marcadores candidatos comunes con un estudio GWAS en trigo duro.....	25
Tabla 2. Resultado de la comparación de las frecuencias alélicas entre el 10% de variedades de fenotipo extremo en los marcadores candidatos compartidos con un estudio realizado en trigo duro	26

LISTA DE ABREVIATURAS

ANOVA (*ANalysis Of VAriance*): Análisis de la varianza

BLAST (*Basic Local Alignment Search Tool*): Herramienta básica de búsqueda de alineamientos locales

cm: centímetros

CRF: Centro de Recursos Fitogenéticos

CV: Coeficiente de Variación

D: Diámetro medio de las raíces

DARtseq (*Diversity Arrays Technology sequencing*): Secuenciación mediante la tecnología de matrices de diversidad

g: gramos

GBS (*Genotyping By Sequencing*): Genotipado por secuenciación

GLM (*General Linear Model*): Modelo lineal general

GO (*Gene Ontology*): Ontología genética

GWAS (*Genome-Wide Association Study*): Estudio de asociación genómica

INIA: Instituto Nacional de Investigación y tecnología Agraria y alimentaria

L: longitud media de las raíces

LD (*Linkage Disequilibrium*): Desequilibrio de ligamiento

LPR: Longitud de la raíz primaria

MAF (*Minimum Allele Frequency*): Frecuencia del alelo menos frecuente

MAS (*Marker-Assisted Selection*): Selección asistida por marcadores

Mb: Megabases

MLM (*Mixed Linear Model*): Modelo lineal mixto

MRA: Ángulo promedio de las raíces respecto a la vertical

MnRA: Ángulo mínimo de las raíces respecto a la vertical

MxRA: Ángulo máximo de las raíces respecto a la vertical

NCBI (*National Center for Biotechnology Information*): Centro nacional para la información biotecnológica

NGS (*Next Generation Sequencing*): Tecnologías de secuenciación de nueva generación

NR: Número de raíces

PCA (*Principal Component Analysis*): Análisis de componentes principales

PCoA (*Principal Coordinates Analysis*): Análisis de coordenadas principales

PEPC (*PhosphoEnolPyruvate Carboxylase*): FosfoEnolPiruvato Carboxilasa

QQPlot (*Quantile-Quantile Plot*): Gráfico cuantil-cuantil

QTL (*Quantitative Trait Loci*): Locus de Rasgo Cuantitativo

S: Superficie media de las raíces

SDSS (*Sodium Dodecyl Sulphate Sedimentation*): Sedimentación con dodecil sulfato sódico

SMA (*Single Marker Analysis*): Análisis de un único marcador

SNP (*Single Nucleotide Polymorphism*): Polimorfismo de nucleótido único

TASSEL (*Trait Analysis by aSSociation, Evolution and Linkage*): Análisis de rasgos por asociación, evolución y vinculación

V: Volumen medio de las raíces

W: Peso de la parte aérea

SUMMARY

Root system architecture is known to play a crucial role in plant development, tolerance against abiotic stresses and nutrient uptake, especially in cereals, where the phenological stage of grain filling is undoubtedly the most relevant one. Nonetheless, the knowledge of its genetic control has been hampered by the complex, repetitive and polyploid genome of wheat species and the inherent difficulties that imply phenotyping the root system architecture-related traits. All in all, the limited information gathered makes the root an ideal subject for future explorations, which could lead to the huge leap in the enhancement of wheat traits of interest like abiotic stress tolerance necessary to solve the current demographic escalation problems and climate change issues.

When compared to other germplasm collections, the Spanish wheat landraces conserved at the National Plant Genetic Resources Center (CRF-INIA) showed a huge genetic diversity in prior studies. In the present work, a representative set from Spanish bread wheat landraces (85 accessions) has been analyzed. First and foremost, the phenotypic variability for several quantitative traits related to root architecture was studied, and then the population structure using genotype data from a set of 4713 molecular markers was investigated. The wide range of variation exhibited by the root traits and the existence of low genetic clustering supported the use of this collection to perform a GWAS study, where phenotype-genotype associations were pursued. After testing several methodologies to correct the analysis, no significant association was detected, so a quality trait with a well-known genetic control was added to the analysis as a positive control. The appearance of significant associations regarding this trait revealed the necessity of phenotyping more landraces to build a larger collection with which to repeat the analysis. Nevertheless, a molecular marker identified from previous studies was proven associated by SMA analysis. Together with markers from this work that were not associated but looked promising, we were able to detect QTLs (Quantitative Trait Loci) – genomic regions potentially responsible for the studied traits. These regions were previously described in the literature, but we explored them in search of possible genes responsible for the association, which could undergo future studies to assess its potential as molecular markers for MAS (Marker Assisted Selection) in future plant breeding programs.

CAPÍTULO 1. INTRODUCCIÓN

1.1 Importancia del trigo

El trigo es el nombre común con el que se designan los cereales pertenecientes al género *Triticum*, encuadrado dentro de la familia de las gramíneas (*Poaceae*). Desde su temprana domesticación en la región de Asia Menor entre los ríos Nilo y Tigris y Éufrates denominada como Creciente Fértil (Figura 1), el trigo ha adquirido un papel fundamental en la alimentación mundial, erigiéndose actualmente como el cultivo que satisface el 35% de los requerimientos nutricionales mundiales (Demir *et al.*, 2015). En 2019, el trigo fue el cultivo con mayor superficie cultivada mundial (216 millones de ha) y el tercer cultivo en términos de producción (765 millones de toneladas) por detrás de la caña de azúcar y el arroz (FAOSTAT, 2019).

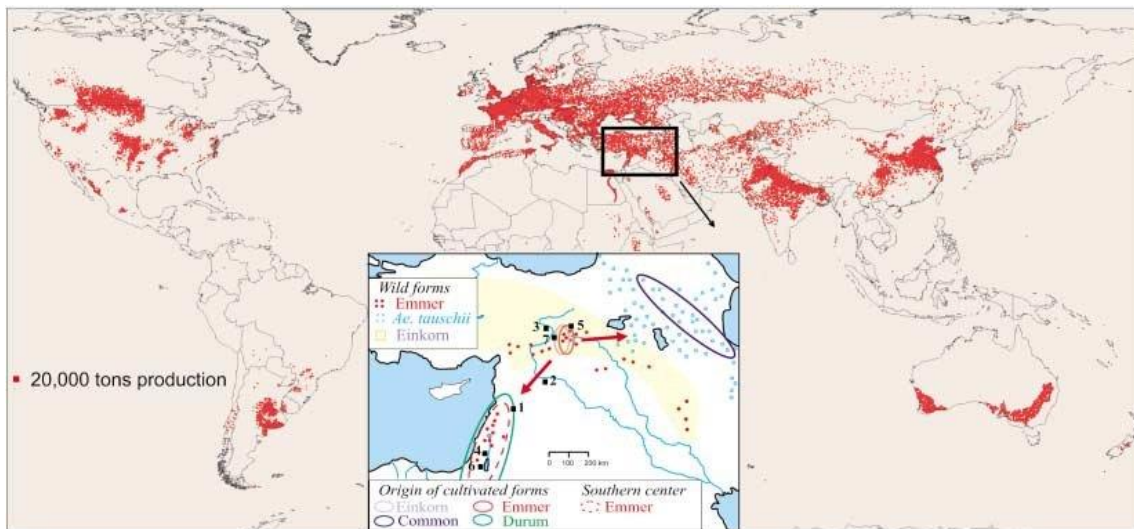


Figura 1. Distribución de las principales áreas de cultivo de trigo. La región ampliada representa el Creciente Fértil y se señalan los orígenes de las especies actuales y las silvestres de las que procedieron. Foto tomada de Dubcovsky & Dvorak, (2014).

Las dos especies de trigo de interés agrícola son el trigo blando o panadero (*Triticum aestivum* ssp. *vulgare* L., $2n = 6x = 42$) y el trigo duro o semolero (*Triticum turgidum* ssp. *durum* L., $2n = 4x = 28$). Ambas especies son resultado de eventos de alopoliploidización con especies silvestres durante las etapas tempranas de su domesticación (Marcussen *et al.*, 2014). Aparte de en su ploidía, también se diferencian en la textura del endospermo y la riqueza proteica, lo que afecta al uso que se le da a cada uno: mientras que el trigo blando se emplea en la producción de harina de panificación y repostería, el trigo duro se utiliza principalmente en la fabricación de pasta. Actualmente, alrededor del 95% del cultivo mundial de trigo está dedicado al trigo blando (Bibi *et al.*, 2009), lo que pone de manifiesto su gran relevancia y explica los severos impactos socioeconómicos que tienen los avances a nivel de mejora genética en este cultivo.

1.2 Importancia de la variabilidad genética y las variedades locales en los programas de mejora

Con el agravamiento de problemas medioambientales como el cambio climático, existe la imperiosa necesidad de continuar mejorando la tolerancia de los cultivos a estreses abióticos. Para ello, los programas de mejora requieren de la existencia de variabilidad genética para la selección de los individuos con alelos más beneficiosos (El-Beltagy & Madkour, 2012). Esta variabilidad puede venir preestablecida en el material de partida o generarse mediante mecanismos como los cruzamientos, mutagénesis y transgénesis. A consecuencia de la intensa selección llevada a cabo durante la Revolución Verde y la introducción a gran escala de cultivares modernos genéticamente muy productivos, pero genéticamente homogéneos, se produjo un fenómeno de erosión genética, el cual ha conllevado una pérdida de variabilidad. Esta tendencia a la convergencia genética en los cultivos repercute negativamente en su capacidad de adaptación a nuevas condiciones (Skovmand *et al.*, 2001).

Cuando se quiere disponer de un amplio rango de variabilidad para, por ejemplo, profundizar en la base genética de un carácter, se puede recurrir al análisis de variedades locales o indígenas. Estas variedades han experimentado un proceso de evolución en estrecho contacto con los estreses bióticos y abióticos presentes en su región de origen, por lo que mostrará una elevada adaptación (Lopes *et al.*, 2015). Por tanto, cada variedad se espera que contenga alelos geográficamente específicos que contribuyan a la construcción de un *pool* genético diverso que represente la variabilidad genética existente en las variedades locales. En este sentido, la indiscutible disparidad agroclimática a lo largo de todo el territorio hace de España un país idóneo para la búsqueda de variedades locales que sirvan de fuente de variabilidad genética de utilidad en el ámbito de la pre-mejora (Pascual *et al.*, 2020)

En el caso del trigo, la colección más numerosa y antigua de variedades locales se conserva en el Centro Nacional de Recursos Fitogenéticos (CRF) del INIA. La colección activa está compuesta por 3722 accesiones, de las cuales 1551 son de origen español (<https://bancocrf.inia.es/es/>).

1.3 Arquitectura radicular

Las principales características que definen los sistemas radiculares son la arquitectura y la morfología. La morfología radicular se define como las características superficiales de una raíz como órgano vegetal (e.g. pelos radiculares, caliptra, ondulaciones, senescencia cortical...). En cambio, la arquitectura radicular tiene en cuenta el conjunto del sistema radicular y describe su comportamiento topológico y geométrico (Boudiar *et al.*, 2020).

El estudio de la arquitectura radicular es fundamental para el desarrollo de nuevas variedades que manejen de forma eficiente la absorción de nutrientes y presenten una elevada tolerancia frente a estreses abióticos, lo que tendría un gran impacto positivo en el rendimiento (Manschadi *et al.*, 2006). Sin embargo, su estudio se ha visto obstaculizado debido a la dificultad de su fenotipado y a la enorme influencia ambiental que normalmente presentan los caracteres radiculares (Zhu *et al.*, 2011). No obstante, en los últimos años se han implementado plataformas de fenotipado automatizado no invasivas que se basan en la utilización de softwares de tratamiento de imágenes que trabajan con las raíces seminales del estado de plántula (Lobet *et al.*, 2013). Así, se permite acortar los ciclos de selección e incrementar la respuesta al realizar una medición directa en planta del carácter (Kuijken *et al.*, 2015). Es posible el fenotipado de individuos en estadios iniciales de su desarrollo dada la elevada correlación encontrada entre plántulas y plantas adultas para los caracteres relacionados con el desarrollo de la arquitectura radicular (Li *et al.*, 2020). Por ejemplo, el ángulo de la raíz seminal (MRA) se ha correlacionado con el ángulo de la raíz nodal (Alahmad *et al.*, 2019).

1.4. Genómica aplicada a la mejora de especies vegetales

La mejora genética vegetal se basa en la selección de individuos con características favorables con el fin de obtener variedades mejoradas que destaquen por la mayor o menor expresión de un carácter. En sus inicios, la mejora dependía de la detección y selección de fenotipos de interés, esperando que dicho fenotipo se correspondiera con un genotipo portador de alelos favorables que transmitir a la descendencia. Desde los años 1970s, el desarrollo de nuevas herramientas de análisis genético como el cartografiado de QTLs (*Quantitative Trait Loci*) o los más recientes estudios genómicos de asociación (GWAS, *Genome-Wide Association Study*), han permitido un cambio de paradigma en detrimento de la selección clásica por el fenotipo y a favor de las estrategias de selección asistida por marcadores (MAS, *Marker-Assisted Selection*) o selección genómica (Leng *et al.*, 2017). Estas aproximaciones posibilitan la identificación directa de genotipos de interés, por lo que constituyen avances tecnológicos más eficientes en tiempo y recursos. En concreto, los GWAS posibilitan reconocer regiones genómicas responsables de la expresión de un carácter cuantitativo. Además, esta metodología permite trabajar fácilmente con paneles de asociación formados por una población natural sin relaciones de parentesco entre los individuos que la conforman, en lugar de recurrir a poblaciones biparentales, cuyo desarrollo supone un gran coste de tiempo y dinero y albergan menor diversidad genética (Zargar *et al.*, 2015). En este tipo de estudios se aplican técnicas estadísticas que evalúan la relación fenotipo-genotipo midiendo el grado de asociación entre la variación fenotípica y cada uno de los polimorfismos genéticos analizadas mediante modelos lineales.

Los estudios genómicos de asociación se basan en el fenómeno de desequilibrio de ligamiento (LD, *Linkage Disequilibrium*). El desequilibrio de ligamiento se refiere a la propiedad de dos o más alelos de loci diferentes de segregar de manera conjunta en una población genética. En poblaciones naturales, el LD es el producto de la historia evolutiva de la evolución y, en general, se diluye rápidamente conforme aumenta la distancia entre los loci. Sin embargo, esta misma historia evolutiva de las poblaciones puede dar lugar a la existencia de loci que se encuentren ligados, pero alejados en el cromosoma o incluso en diferentes cromosomas. En estos casos, los estudios GWAS tienden a obtener asociaciones espurias. Para evitar esto, antes de realizar un estudio GWAS es altamente recomendable investigar la estructura genética de la población y corregir el análisis en base a ella (Korte & Farlow, 2013).

Como requisito para realizar un GWAS, es necesario el genotipado de un gran número de individuos con un elevado número de marcadores moleculares para cubrir la mayor extensión del genoma como sea posible (Korte & Farlow, 2013). Esto permitirá definir una región de asociación estrecha que facilite la identificación de los genes que explican la asociación. En base a su secuencia, se podrán diseñar marcadores moleculares de utilidad en programas de mejora. Los avances en biología molecular y el desarrollo de modernas tecnologías de genotipado masivo como el genotipado por secuenciación masiva (GBS, *Genotyping By Sequencing*) permiten genotipar de manera eficiente prácticamente cualquier especie vegetal con miles de marcadores. No obstante, el trigo presenta un genoma con ciertas características que han retrasado la obtención de datos de genotipado masivo adecuado para realizar estudios GWAS. En primer lugar, el 85% del extenso y complejo genoma del trigo está compuesto por secuencias de DNA repetidas. Además, los dos eventos de alopoliploidización del trigo blando han derivado en que su genoma esté conformado por tres genomios distintos de especies diferentes ($2n = 6X = 42$, AABBDD). No ha sido hasta hace pocos años que se ha conseguido construir genomas de referencia de calidad (Appels *et al.*, 2018). Aunque estas características dificultan el análisis de secuencias homeólogas y el genotipado masivo, cada vez son más los estudios que aplican técnicas de genotipado masivo en trigo para, en concreto, la caracterización de la variabilidad genética de variedades locales (Rasheed *et al.*, 2018).

No obstante, no es necesario tener repartidos abundantes marcadores por cada región del genoma. Se conoce que el LD depende principalmente de la distancia genética y de la tasa de recombinación de cada región cromosómica. Debido a que ocurre un menor número de eventos de recombinación en las zonas centroméricas, se crean en estas zonas grandes grupos de bloques de ligamiento. Por tanto, es necesario tener mayor resolución de mapeo en las regiones periféricas que en las regiones centroméricas (Wickland *et al.*, 2017).

En especies complejas como el trigo, resulta especialmente útil el genotipado mediante la tecnología DArTseq (*Diversity Arrays Technology sequencing*). Esto se debe a que combina la reducción de la complejidad del genoma mediante el método DArT y las nuevas tecnologías de secuenciación masiva (NGS, *Next Generation Sequencing*) para la obtención de marcadores moleculares polimórficos tipo SNP (*Single Nucleotide Polymorfism*). Una de las ventajas de esta tecnología es que, en el proceso de reducción de complejidad para separar las secuencias de bajo número de copias de la fracción repetitiva del genoma, se seleccionan las zonas del genoma que se corresponden predominantemente con regiones transcripcionalmente activas, lo que es especialmente interesante para estudios GWAS (Kilian *et al.*, 2012).

1.5. Objetivos

Teniendo en cuenta la importancia de la arquitectura radicular en la tolerancia al estrés hídrico y el potencial en el ámbito de la mejora de las variedades locales españolas por su gran diversidad, el objetivo general de este trabajo consiste en profundizar en la base genética de la arquitectura radicular mediante un estudio de asociación GWAS en una colección de 85 variedades locales españolas de trigo blando. Esto supondrá el punto de partida en el desarrollo de marcadores moleculares para la selección de genes implicados en el desarrollo de la raíz en trigo, útiles para programas de mejora.

Los objetivos parciales propuestos son:

1. Evaluación de la adecuación de la colección para la realización de un estudio de asociación genómica (GWAS) en base al análisis estadístico de la variabilidad fenotípica presente en la colección obtenida a partir de datos previos de fenotipado para la arquitectura radicular.
2. Realizar un GWAS para la identificación de regiones genómicas asociadas a los caracteres fenotípicos analizados.
3. Explorar las regiones genómicas de asociación e identificar genes candidatos implicados en la arquitectura radicular para el posible diseño de marcadores moleculares en programas de mejora.

CAPÍTULO 2. MATERIAL Y MÉTODOS

2.1 Material vegetal

El material vegetal inicial consiste en una colección de 831 variedades de *Triticum aestivum* L. procedente del CRF-INIA (Centro de Recursos Fitogenéticos – Instituto Nacional de Investigación y tecnología Agraria y alimentaria) que recoge la variabilidad genética, ecogeográfica y agromorfológica existente de las variedades locales de trigo españolas. Con el fin de reducir el número de variedades de trabajo y facilitar el fenotipado, a partir de esta colección fue realizada una selección de 189 variedades, de las cuales se han obtenido datos fenotípicos de 85 variedades. Estas accesiones constituyen el punto de partida del trabajo.

2.2 Datos fenotípicos

Los datos fenotípicos fueron proporcionados por la Dra. Magdalena Ruiz, responsable de las colecciones de trigo del CRF-INIA y el Dr. Juan M. González de la UAH. Ambos han formado parte del equipo investigador del proyecto AGL2016-77149.

Los datos fenotípicos se refieren a diez caracteres cuantitativos relativos a la arquitectura radicular: longitud total de la raíz en cm (L), área superficial radicular en cm^2 (S), volumen total radicular en cm^3 (V), diámetro promedio en cm (D), longitud de la raíz principal en cm (LPR), número total de raíces en unidades (NR) y ángulo promedio (MRA), mínimo (MnRA) y máximo (MxRA) de las raíces respecto a la vertical en grados ($^\circ$). Debido a la estrecha relación entre la raíz y la parte aérea, también se ha tenido en cuenta en el estudio el peso de la parte aérea en g (W). Para cada variedad se analizaron 12 granos en 3 réplicas técnicas. Estos datos se obtuvieron de acuerdo con la metodología descrita en Ruiz *et al.*, (2018) para trigo duro (Figura 2).

Además, como control de estudio, se dispuso de datos fenotípicos para la fuerza del gluten, carácter relacionado con la calidad harinera del trigo y estimado mediante la prueba de sedimentación SDSS (*Sodium Dodecyl Sulphate Sedimentation*) (López-Fernandez *et al.*, 2021).



Figura 2. Ensayo in vitro de fenotipado de la arquitectura radicular en plántulas. La escala gráfica indica 1 cm. En **a** se muestra cómo se calcularon los parámetros MRA (MAV) y MxRA (MxAV). La flecha verde apunta a la raíz principal. Figura tomada de Ruiz et al. (2018).

2.3 Datos genotípicos

Los datos genotípicos consistían en un conjunto de marcadores moleculares tipo SNP (*Single Nucleotide Polymorphism*) obtenidos utilizando el servicio DArTseq GBS (Genotyping by Sequencing) de Diversity Array Technology Pvt., Ltd. (Canberra, Australia). El total de 4856 marcadores moleculares obtenidos y mapeados se sometió a un filtrado con el fin de eliminar aquellos marcadores cuyo alelo menos frecuente (MAF, *Minimum Allele Frequency*) estuviera presente en menos del 5% de las accesiones y aquellos con más del 15% de datos faltantes. En total, se obtuvieron un total de 4.713 marcadores distribuidos principalmente por las regiones no centroméricas del genoma, como se apunta en Pascual *et al.*, (2020). En dicho estudio se utilizaron los datos genotípicos obtenidos mediante estos marcadores para estructurar la colección de 189 variedades en 4 grupos poblacionales genéticamente diferentes (Figura 3). La comparación de la estructura genética poblacional entre la colección de 189 variedades y la de 85 analizada en este trabajo se realizó mediante un PCoA.

Además, se disponía de datos de genotipado para los genes *Vrn-A1* y *Vrn-B1*, relacionados con los requerimientos de vernalización, y de un set de 10 marcadores multialélicos relacionados con el estrés hídrico en trigo duro (Mérida-García *et al.*, 2020). Estos marcadores mapeaban específicamente ciertas enzimas claves del metabolismo vegetal.

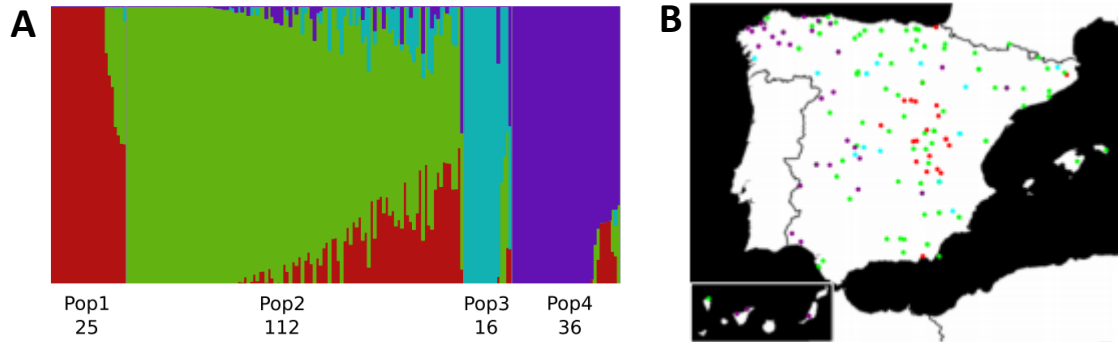


Figura 3. A Gráfico STRUCTURE que refleja la estructura poblacional en base al grado de alelos únicos o compartidos con otras variedades. El número debajo de la población indica el número de accesiones que agrupa cada población **B** Lugar de origen de las variedades que conforman la colección de 189 accesiones. Figuras obtenidas de Pascual et al. (2020).

2.4 Análisis de los datos fenotípicos

Para los análisis estadísticos y estudios posteriores se empleó el entorno de desarrollo integrado RStudio Desktop Pro v.1.4.1103 del lenguaje de programación R 4.1.0. (R Core Team, 2021). Se calcularon parámetros de estadística descriptiva como la media, desviación típica, coeficiente de variación y rango (diferencia entre el valor máximo y el valor mínimo) de los diez caracteres analizados.

2.4.1 Normalidad y homogeneidad de varianza

Muchas pruebas de inferencia estadística como el ANOVA (*ANalysis Of VAriance*) o el test de Tukey se basan en los supuestos de normalidad y homogeneidad de varianza de las variables analizadas. Para la evaluación de la normalidad de los distintos caracteres se empleó el test de normalidad de Shapiro-Wilk (Shapiro & Wilk, 1965), con un nivel de significación $\alpha = 0,05$. En el caso de que una determinada variable se desviara significativamente de la normalidad, se utilizaron diversas transformaciones de los datos (inverso, raíz cuadrada y logaritmo) con el fin de aproximarlos a una distribución normal.

La evaluación de la homogeneidad de varianza se llevó a cabo mediante el test de Levene (Levene, 1960), con un nivel de significación $\alpha = 0,05$.

2.4.2 Análisis de correlación

La correlación entre cada par de caracteres estudiados fue estimada mediante el cálculo del coeficiente de correlación de Pearson. Además, se estudió la correlación entre los caracteres de arquitectura radicular y variables edáfo-climáticas cuyos datos fueron proporcionados por la Dra. Magdalena Ruiz del CRF-INIA. Se representaron gráficamente las correlaciones en matrices triangulares de correlación.

2.4.3 ANOVA

Con el objetivo de encontrar diferencias significativas entre grupos de accesiones para cada uno de los caracteres estudiados se empleó la prueba ANOVA de análisis de la varianza, con un nivel de significación $\alpha = 0,05$. En el caso de que se encontraran diferencias significativas, se procedía a la comparación de todos los pares de medias mediante el test de comparaciones múltiples de Tukey (Tukey, 1949).

Los grupos de accesiones usadas como niveles del factor en el ANOVA fueron las poblaciones genéticas en las que se estructuró la colección empleando los datos genotípicos de marcadores SNP DArT (Pascual *et al.*, 2020), la composición alélica para los genes *Vrn-A1* y *Vrn-B1*, los marcadores relacionados con estrés hídrico y el tipo de floración de las variedades (i.e. invierno o primavera) en función de sus requerimientos de vernalización.

2.4.4 PCA

Para evaluar la estructura de la colección en base a los datos fenotípicos, se realizó un análisis PCA (*Principal Component Analysis*) un método estadístico que permite simplificar la complejidad de los espacios muestrales multidimensionales en pocas variables o componentes que tratan de explicar el mayor porcentaje de la variación.

2.5 Análisis de asociación genómica

Se llevó a cabo un estudio GWAS (*Genome-Wide Association Study*) con la finalidad de detectar las regiones genómicas responsables del control genético de los caracteres analizados. Para ello, se hizo uso del software TASSEL (*Trait Analysis by aSSociation, Evolution and Linkage*) (Bradbury *et al.*, 2007), que permite el estudio de asociaciones fenotipo-genotipo mediante la aplicación de métodos estadísticos: GLM (*General Linear Model*) y MLM (*Mixed Linear Model*). Ambos son modelos lineales que aplican una regresión de mínimos cuadrados y utilizan un PCoA (*Principal Coordinates Analysis*) (método alternativo al PCA para datos cualitativos) de los datos genotípicos para introducir la estructura genética poblacional como cofactor. De esta manera, se consigue reducir la influencia de la estructura genética en el análisis y evitar la aparición de asociaciones espurias.

Los resultados de la asociación se recogen en Manhattan plots, que representan la significancia estadística de la asociación de cada marcador a cada carácter en función de su posición en el cromosoma.

Los gráficos QQPlots representan los p-values observados en el análisis de asociación realizado con respecto a lo que se esperaría por azar de acuerdo con una distribución χ^2 teórica, es decir, si no hubiera asociaciones. Fueron obtenidos en el mismo software donde se llevó a cabo el estudio de asociación: TASSEL. Se utilizaron los QQPlots como método final de validación, ya que una mayor aproximación a la diagonal indica una menor estratificación poblacional y, por tanto, una menor aparición de falsos positivos (Ehret, 2010).

La realización de un GWAS implica aplicar un test de comparaciones múltiples con tantos test independientes como marcadores empleados. La minimización de la aparición de falsos positivos (debido al elevado número de marcadores moleculares) se llevó a cabo mediante la corrección de la significancia estadística de acuerdo con la desigualdad de Bonferroni (Abdi, 2007) y diversos métodos FDR (*False Discovery Rate*) contenidos en el paquete “multtest” de R Bioconductor (Pollard *et al.*, 2005).

Se consideró que un determinado marcador estaba asociado al carácter analizado cuando el p-valor corregido era inferior a 0,05.

2.6 Análisis de asociación de marcadores individuales

La evaluación de la asociación de marcadores individuales se llevó a cabo mediante un análisis SMA (*Single Marker Analysis*), consistente en comparar mediante un test t de Student la media de los datos fenotípicos de las variedades portadoras de cada uno de los dos alelos del marcador SNP candidato. En el caso de los marcadores multialélicos relacionados con el estrés hídrico, se empleó un ANOVA considerando los distintos alelos como niveles del factor. Este análisis se acompañó de la comparación mediante un test χ^2 de las frecuencias alélicas que presentaban el 10% de las variedades con fenotipo más extremo (Roselló *et al.*, 2019).

2.7 Análisis de marcadores asociados

Una vez hallados los marcadores asociados, se consideró que la región genómica que podría contener el QTL (*Quantitative Trait Locus*) abarcaba el segmento contenido entre los 2 marcadores adyacentes. La exploración de estas regiones en busca de genes candidatos se realizó en la versión actualizada del genoma de *Triticum aestivum* cv. Chinese Spring IWGSC RefSeq v1.0. El explorador de genomas utilizado fue el de la base de datos Ensembl Wheat (http://plants.ensembl.org/Triticum_aestivum/Info/Index). Se evaluaron los genes que se encontraban anotados en la base de datos como de “elevada confianza”. La anotación correspondiente a este ensamblado genómico carecía de información acerca de las funciones

génicas, por lo que la determinación de la función de los genes localizados en la región de asociación se dilucidó empleando BLAST (*Basic Local Alignment Search Tool*) (Altschul *et al.*, 1990) y usando el la base de datos proteica de *Arabidopsis thaliana* del NCBI (*National Center for Biotechnology Information*) como conjunto de búsqueda. Para ello, se diseñó un script propio de Python en el que se utilizaban paquetes bioinformáticos como Biopython (Cock *et al.*, 2009) y Goatools (Klopfenstein *et al.*, 2018) para extraer del archivo GenBank (contiene toda la información anotada relativa a dicha región genómica) las secuencias de los genes presentes en la región de asociación, sus identificadores y sus categorías GO (Gene Ontology). El resultado es un archivo multifasta con el que se procedió a realizar la búsqueda en BLASTp (alineamientos entre secuencias proteicas). Una vez obtenido el nombre del gen candidato, se utilizó la base de datos de proteínas Uniprot (<https://www.uniprot.org/>) para la determinación de la función y demás características que revelaran una posible relevancia en el control del carácter asociado.

CAPÍTULO 3. RESULTADOS Y DISCUSIÓN

3.1 Material vegetal

La colección de 85 variedades locales de trigo blando españolas constituye un subconjunto de la colección de 189 variedades, que condensa la variabilidad genética, ecogeográfica y agromorfológica disponible en el CRF-INIA. Con el fin de comprobar el mantenimiento de estas propiedades en la colección reducida de 85 variedades, se llevó a cabo un análisis PCoA comparativo entre ambas colecciones (Figura 4). Para ello, se emplearon los mismos 4713 marcadores SNP DArT que se utilizarán posteriormente en el GWAS.

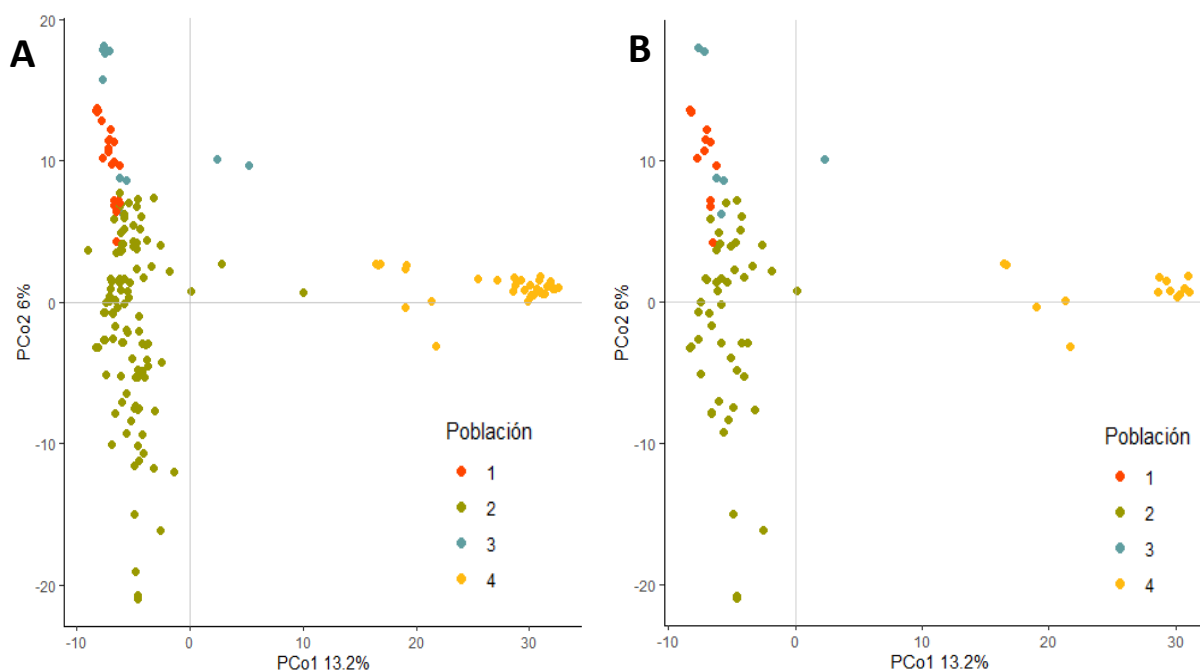


Figura 4. Análisis PCoA comparativo. Se representa la colección original de 189 variedades (A) y la colección de 85 variedades analizada en este trabajo (B) para las dos coordenadas principales. El color que presenta cada variedad hace referencia a la población genética a la que pertenece.

$n_1 = 13$, $n_2 = 50$, $n_3 = 7$, $n_4 = 15$

Como consecuencia de emplear un número elevado de marcadores, las dos coordenadas principales explicaron 19,2% del total de la variación contenida en la colección. La primera coordenada de los PCoA consiguió explicar el 13,2% de la variación y permite diferenciar la población 4 del resto de poblaciones. La segunda coordenada condensó el 6% de la variación y permite diferenciar entre las poblaciones 1, 2 y 3. Sin embargo, estas tres poblaciones muestran cierto solapamiento derivado del alto grado de mezcla de alelos o *admixture* presente en las variedades de la población 2 (Figura 3A). Por tanto, se pueden diferenciar 2 grupos de variedades de forma general: la población 4 y las poblaciones 1, 2 y 3.

La población 4 emerge del análisis como la más diferente genéticamente. Esto se podría explicar teniendo en cuenta que las variedades que conforman la población 4 proceden de regiones occidentales peninsulares, donde existe una prevalencia de suelos ácidos o neutros.

Estos resultados confirman que el set de 85 variedades reflejaba la variabilidad de la población de partida.

3.2 Análisis de datos fenotípicos

El cálculo de parámetros de dispersión y tendencia central de estadística descriptiva se resumen en la Figura 5. Este tipo de análisis también se ha realizado en Ruiz *et al.*, (2018), donde se analiza una colección de 94 variedades locales españolas de trigo duro como paso previo a un GWAS. Los valores medios, máximos y mínimos calculados no se ajustaban a lo obtenido en este estudio, lo que refleja la existencia de diferencias en los caracteres radiculares incluso en especies tan cercanas filogenéticamente como *T. aestivum* y *T. durum*. No obstante, se ha descrito que la mayor diversidad en arquitectura radicular se debe a la variabilidad de los genomas A y B, compartidos por ambas especies (Rufo *et al.*, 2020).

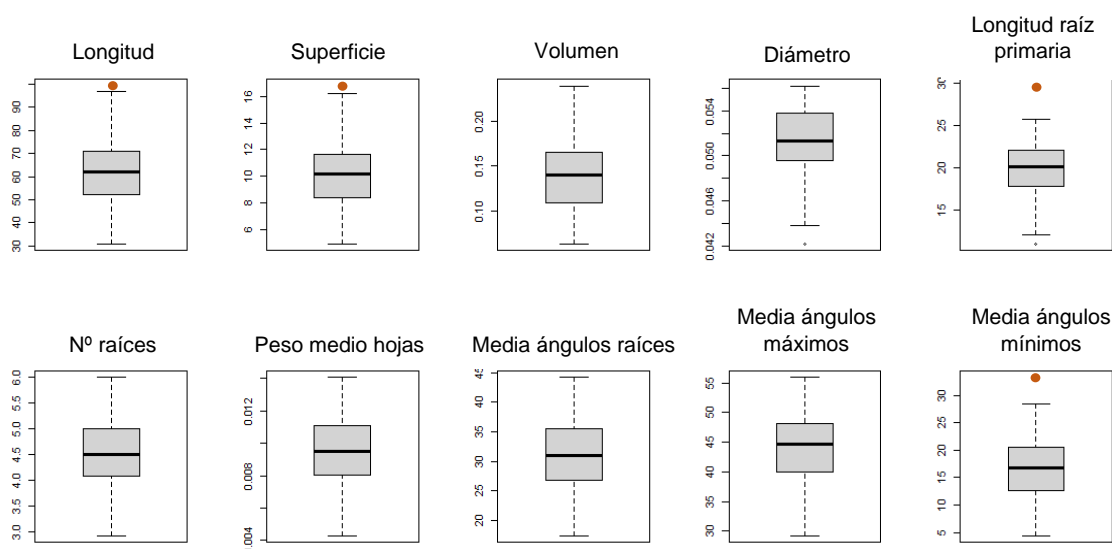


Figura 5. Diagramas de cajas y bigotes resultado de la estadística descriptiva para los 10 caracteres analizados. La línea negra que atraviesa la caja indica el valor medio de los datos fenotípicos. Los puntos rojos representan el valor de la accesión AGL-384 cuando constituye un valor atípico.

La colección analizada este estudio presentó valores medios inferiores en comparación con los de la colección de variedades locales de trigo duro para los caracteres L, S y V. En relación con las variables de dispersión, las variedades locales de trigo blando empleadas albergan una amplia diversidad fenotípica para los caracteres analizados, de forma análoga a lo descrito en los estudios mencionados anteriormente. De hecho, la colección estudiada presenta mayor variabilidad, como así lo refleja el coeficiente de variación medio calculado (21,06%), superior

en un 5% al obtenido en Ruiz *et al.*, (2018). Esta elevada variabilidad fenotípica, soportada por una, también elevada, variabilidad genotípica, es esencial para la realización de estudios GWAS y los objetivos futuros de este estudio: selección de variedades asistida con marcadores en programas de mejora.

La representación en diagrama de cajas y bigotes permite la identificación de valores atípicos, los cuales pueden distorsionar análisis posteriores. Se detectó que la accesión AGL-384 mostraba valores muy superiores al resto en 4 de los 10 caracteres estudiados (Figura 5). No obstante, se comprobó que su presencia no alteraba los resultados que se obtuvieron en los demás análisis, por lo que se decidió mantener dicha accesión en la colección.

3.2.1 Análisis de correlación

En primer lugar, se estudió la correlación entre los distintos caracteres de arquitectura radicular (Figura 6). Se detectó un grado de correlación casi absoluto ($r \approx 1$) entre las variables L, S y V. También mostraron un alto grado de correlación ($0,9 > r > 0,7$) las variables L, S y V con las variables LPR, NR, y W, y entre las variables MRA y MxRA, MnRA, lo cual es de esperar debido a que MRA se calcula a partir de MxRA y MnRA. Los pares de variables que mostraron una correlación moderada ($0,7 > r > 0,4$) fueron V-D y MxRA-MnRA. Este último resultado pone de manifiesto que el sistema radicular exhibe dos tipos de comportamientos: crecimiento superficial (valores de MxRA y MnRA elevados) y crecimiento en profundidad (valores de MxRA y MnRA bajos). En general, el carácter W mostró una correlación elevada con las variables relativas al tamaño radicular (L, S, V, D, LPR y NR), lo cual resulta lógico teniendo en cuenta que un sistema radicular más desarrollado es capaz de soportar un mayor peso de la parte aérea. Finalmente, se obtuvo una correlación nula ($r \approx 0$) entre los principales caracteres que determinan el tamaño radicular (L, S y V) y los caracteres relativos a la disposición espacial de las raíces (MRA, MxRA y MnRA), lo que refleja una total independencia entre las dos divisiones principales de la arquitectura radicular. Las únicas correlaciones negativas fueron bajas ($0 < r < -0,4$) y ocurrieron entre LPR y las variables MRA, MxRA y MnRA. Esto refleja que, de forma general, un crecimiento en profundidad de la raíz principal llevará asociado un crecimiento en profundidad del conjunto de raíces.

En Ruiz *et al.*, (2018) se estudiaron en trigo duro las correlaciones entre los mismos caracteres. Se mantuvo la correlación casi absoluta entre las variables L, S y V. No obstante, el resto de las correlaciones obtenidas se desviaban generalmente de las obtenidas en este trabajo, lo que pone de manifiesto, de nuevo, la existencia de diferencias en la arquitectura radicular entre dos

especies tan próximas filogenéticamente como *T. aestivum* y *T. durum*. Por ejemplo, no se obtuvieron unas correlaciones tan elevadas entre W y las variables relativas al tamaño radicular.

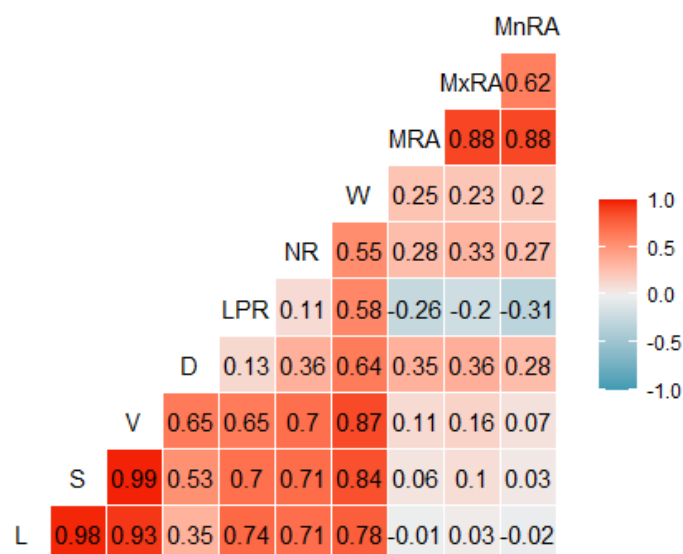


Figura 6. Matriz triangular de correlación. Cada cuadrado representa el coeficiente de correlación de Pearson (r) entre los caracteres estudiados mediante un valor numérico y un color en la escala de temperatura.

Además, en este trabajo las variables MRA, MxRA y MnRA mostraron un alto grado de correlación con los caracteres relacionados con el tamaño radicular, especialmente en las subespecies *dicoccon* y *turgidum*.

Puesto que, además de la variabilidad genética y agromorfológica, esta colección también representa la variabilidad ecogeográfica contenida en las variedades locales españolas de trigo blando, se ha estudiado la correlación existente entre las variables radiculares analizadas y distintas variables relacionadas con parámetros edafoclimáticos (Apéndices A, B). Las correlaciones entre dichos parámetros representaban relaciones triviales de poco interés para este trabajo. En relación con las correlaciones entre los caracteres de arquitectura radicular y los parámetros edafoclimáticos, se obtuvieron de forma general valores de r bajos ($0,35 > r > 0$). No obstante, se pudieron extraer varias correlaciones que aportaban información relevante:

- Todos los caracteres se mostraron correlacionados con dos variables que determinan la composición del suelo: fracción arcillosa ($r = 0,15$) y fracción arcillosa ($r = -0,25$). Este resultado revela que el crecimiento radicular en trigo blando parece estar favorecido en suelos arenosos respecto de suelos arcillosos. Esto está apoyado por otros estudios (Clark *et al.*, 2008).
- Los caracteres que definen la disposición espacial (MRA, MxRA y MnRA) manifestaron una correlación negativa ($r = -0,35$) con la temperatura media de octubre y los demás

caracteres una correlación positiva ($r = 0,35$) con la temperatura media de julio. Distintos estudios han encontrado correlaciones positivas entre la temperatura y el desarrollo radicular (Nagel *et al.*, 2009).

- Combinando parámetros de temperatura y precipitación se obtuvo la variable aridez, la cual se mostró correlacionada con los caracteres MnRA (0,25) y D (-0,25).

La utilidad de la correlación con variables edafoclimáticas reside en poder predecir caracteres de complicado fenotipado como la arquitectura radicular en base a variables de sencilla medición como la temperatura, la precipitación o la composición del suelo. En este caso, las correlaciones calculadas no son demasiado elevadas, pero apuntan a que la variabilidad fenotípica observada refleja una adaptabilidad de las distintas variedades locales a las condiciones eco-climáticas de sus lugares de origen. En un estudio previo en trigo duro (Ruiz *et al.*, 2018), se detectaron asociaciones entre los caracteres radiculares y las variables eco-climáticas analizadas, especialmente en las subespecies *dicoccum* y *turgidum*, donde la forma (y no el tamaño) de las raíces parecía tener un papel fundamental en la adaptabilidad a zonas áridas.

Se comprobó la significancia estadística de las correlaciones previamente detalladas y todas las que se han destacado en este trabajo mostraban valores de p-values $< 0,01$ (**).

3.2.2 Normalidad y homogeneidad de varianzas

Aunque se ha demostrado que no es imprescindible, es preferible que los caracteres fenotípicos incluidos en estudios de asociación se distribuyan conforme a una distribución normal y presenten homogeneidad de varianzas, especialmente cuando se trabaja con tamaños muestrales bajos (Goh & Yap, 2009).

La obtención de p-values $> 0,05$ en el test de Levene ($\alpha = 0,05$) al comparar las varianzas de los distintos niveles de un factor reflejó la existencia de homogeneidad de varianzas en los datos.

La normalidad se estimó empleando el test Shapiro-Wilk ($\alpha = 0,05$). Todos los caracteres exhibieron un comportamiento que se ajustaba a una distribución normal salvo D, NR, MnRA (Figura 7). Cabe destacar que el caso del carácter NR (número de raíces) es especial, ya que se trata de un carácter cualitativo merístico en el que no es adecuado evaluar características de caracteres cualitativos continuos como la normalidad. Con objeto de ajustarlos a una distribución normal, los 3 caracteres fueron sujeto de varias transformaciones: inverso, raíz cuadrada y logaritmo). Ninguno de ellos mejoró su ajuste a una distribución normal, por lo que se conservaron los datos originales.

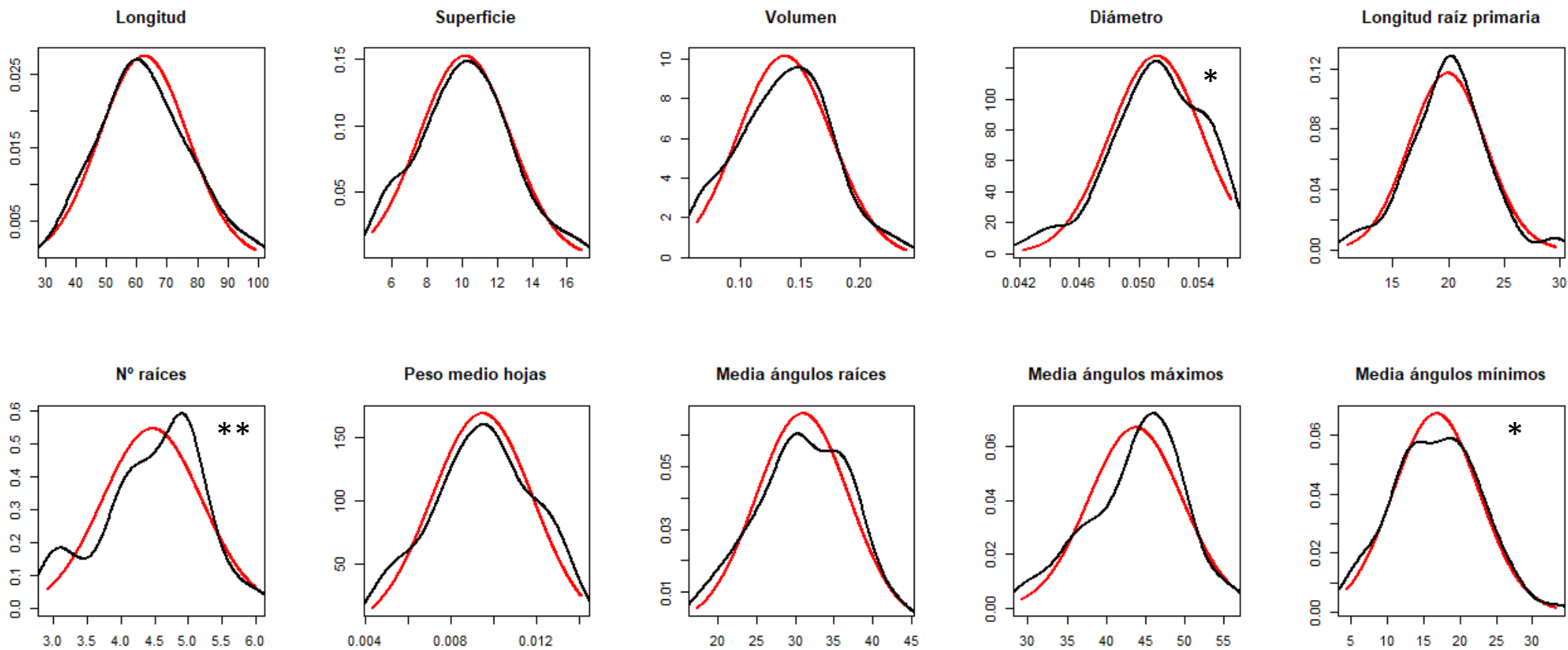


Figura 7. Diagramas de densidad. Superposición de la distribución de densidad de cada variable (negro) frente a la distribución normal teórica de media y varianza igual a la de los caracteres estudiados. La comparación se realizó mediante el test de normalidad Shapiro-Wilk ($\alpha=0,05$): p-value < 0,05 (*), p-value < 0,01 (**).

No obstante, pruebas como el ANOVA son muy robustas frente a cierto grado de la normalidad siempre que se cumpla el supuesto de homogeneidad de varianzas.

Este resultado es concordante con lo obtenido en diversos estudios realizados en trigo duro: en Ruiz *et al.*, (2018), por ejemplo, todos los caracteres se ajustaban aproximadamente a una distribución normal salvo NR y MnRA, al que se le aplicó la transformación de raíz cuadrada para mejorar el ajuste.

3.2.3 ANOVA

Se utilizó la prueba ANOVA para dilucidar la existencia de diferencias en los caracteres en función de la población genética y el tipo de floración (variedades de primavera que no presentan vernalización y no requieren acumular horas de frío para florecer y variedades de invierno que requieren atravesar un proceso de vernalización) en los que se agruparon las variedades gracias a marcadores moleculares (Pascual *et al.*, 2020) y ensayos experimentales en campo, respectivamente.

Considerando como factor el tipo de floración, la prueba ANOVA concluía que no existían diferencias significativas para ninguno de los caracteres analizados (p -values $> 0,05$). Aunque se ha descrito la existencia de una estrecha relación entre el desarrollo radicular y la vernalización (Voss-Fels *et al.*, 2018), los resultados de este trabajo no reflejaron dicha relación.

En el ANOVA realizado teniendo en cuenta como factor la población genética se obtuvieron diferencias significativas (p -values $< 0,05$) entre las 4 poblaciones genéticas en las que se estructuró la colección para 4 de los 10 caracteres analizados: L, S, NR y MnRA (Figura 8). Dado que la prueba ANOVA únicamente puede confirmar la existencia de diferencias significativas entre los niveles del factor, para evaluar las diferencias entre pares de niveles se utilizan pruebas *post hoc* como el test de Tukey. En los caracteres L, S y NR, la población 1 presentó valores significativamente más elevados que el conjunto o parte del resto de poblaciones genéticas. Esto se podría explicar teniendo en consideración que las variedades pertenecientes a la población 1 proceden de localidades de las provincias de Cuenca y Guadalajara, mientras que las variedades de las demás poblaciones genéticas presentan unos orígenes repartidos por toda España (Figura 3B). Más estudios podrían determinar si existe alguna propiedad específica de los suelos o el clima de esta región que propicie el desarrollo radicular. Por tanto, mientras que la población 4 emergía del análisis de los datos genotípicos como la más diferente genéticamente, el análisis reveló que la población 1 presentaba las mayores diferencias fenotípicas.

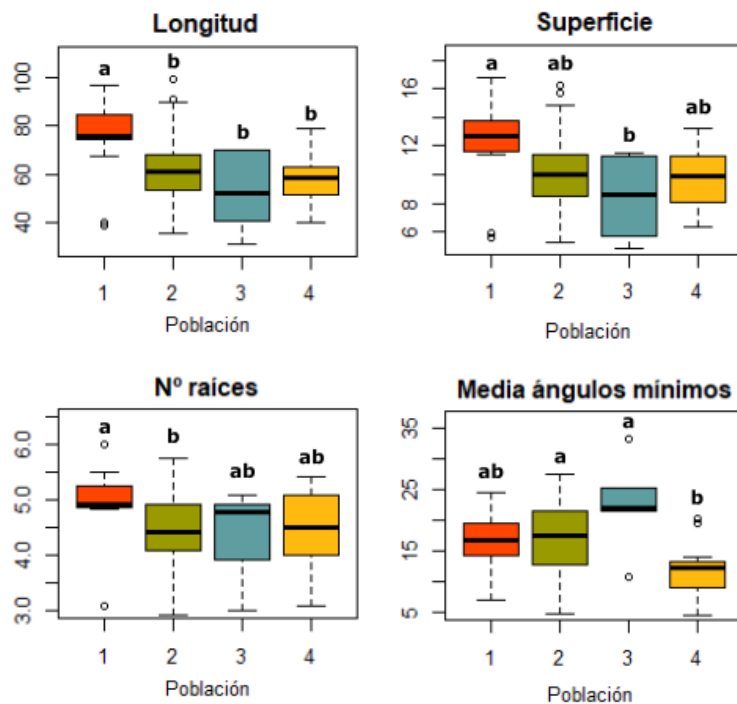


Figura 8. Diagramas de cajas y bigotes resultado de la comparación de poblaciones genéticas. La línea negra que atraviesa la caja indica el valor medio de los datos fenotípicos. Las letras situadas encima de los bigotes reflejan el grupo de significación al que el test de Tukey asigna cada población genética. $n_1 = 13$, $n_2 = 50$, $n_3 = 7$, $n_4 = 15$

3.2.4 PCA

Continuando con la exploración de los datos fenotípicos, se llevó a cabo un PCA en el que se reduce la complejidad presente en los datos a dos componentes principales (Figuras 9 y 10), las cuales consiguieron explicar gran parte de la variabilidad contenida en los datos fenotípicos (79,84%). En concreto, la primera componente explicó el 51,37% de la variabilidad y la segunda sintetizaba el 28,47% de la variabilidad.

Cuando las variedades se identificaron en función de la población genética (Figura 9) se observó que, mientras que en la segunda componente no se conseguía extraer ningún patrón apreciable, en la primera coordenada las variedades pertenecientes a la población 1 (salvo AGL-358 y AGL-372) mostraban una tendencia a presentar valores positivos. Existe demasiado grado de solapamiento como para que la primera componente principal permita diferenciar esta población de las demás, pero el hecho de que la componente principal 1 se viera influenciada en un 45,13% por los caracteres L, S y NR explicaría y corroboraría de forma visual los resultados obtenidos en el análisis de comparación poblacional: las variedades correspondientes a la población 1 presentan valores medios de los caracteres L, S y NR superiores.

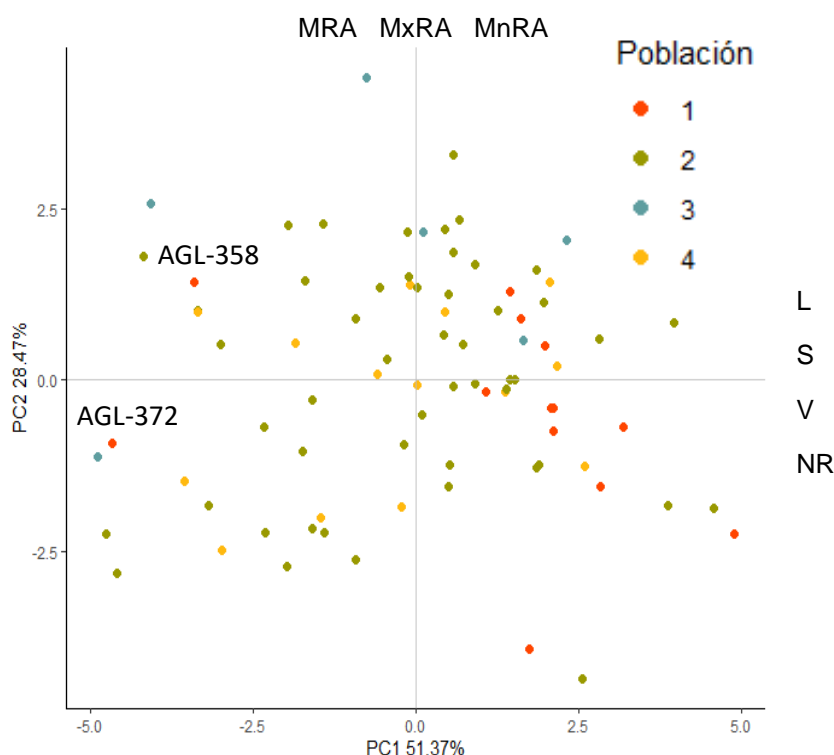


Figura 9. Representación del PCA en función de la población genética. Se representan los valores de cada variedad para las dos componentes principales. Los porcentajes indican la porción de la variabilidad total que explica cada componente. Se representan en los márgenes superior y derecho los caracteres con mayor influencia en cada una de las dos componentes principales. El color que presenta cada variedad hace referencia a la población genética a la que pertenece. Se destacan las dos variedades pertenecientes a la población 1 que no se agrupan con el resto de las variedades de la población.

$n_1 = 13$, $n_2 = 50$, $n_3 = 7$, $n_4 = 15$

El mismo procedimiento se empleó con el tipo de floración (Figura 10). De forma análoga a lo ocurrido al analizar la población genética, el PCA corroboró lo concluido en el ANOVA: no se observó ningún patrón visual en el PCA que permitiera diferenciar las variedades de invierno de las de primavera.

3.3 Análisis de asociación genómica

El análisis estadístico reveló que los datos fenotípicos mostraban una amplia variabilidad, propiedad necesaria para la realización de GWAS. Se empleó el software estadístico TASSEL, que permite la realización de dos tipos de modelos estadísticos de regresión de mínimos cuadrados: GLM y MLM (Zhang *et al.*, 2010). Para minimizar el número falsos positivos y asociaciones espurias (asociaciones estadísticamente significativas, pero biológicamente irrelevantes), ambos métodos llevan a cabo una corrección en base a la estructura genética poblacional (estimada mediante un PCoA (Figura 4B) de los datos genotípicos). Además, el MLM incorpora en el análisis una corrección en base a las relaciones genéticas entre las variables (estimadas

mediante una matriz de parentesco). Al tener en cuenta dos cofactores, el MLM constituye un método estadístico más restrictivo a la hora de encontrar asociaciones significativas.

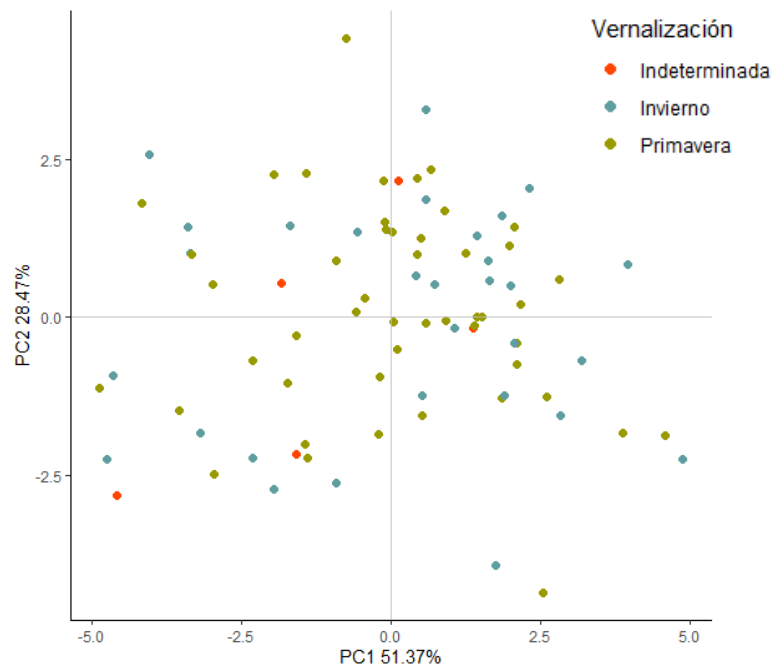


Figura 10. Representación del PCA en función del tipo de floración. Se representan los valores de cada variedad para las dos componentes principales. Los porcentajes indican la porción de la variabilidad total que explica cada componente. El color que presenta cada accesión hace referencia al tipo de variedad al que pertenece en función de si presenta necesidad de vernalización para florecer (variedad invernal) o no (variedad primaveral). Las accesiones que mostraron distintos comportamientos a lo largo de los sucesivos ensayos anuales de vernalización se representan como variedades de necesidad de vernalización indeterminada.

$n_{PRIMAVERA} = 48$, $n_{INVIERNO} = 32$, $n_{INDETERMINADA} = 5$

Estudios de asociación previos en trigo han empleado tanto modelos GLM, utilizando como único cofactor la estructura poblacional (Giraldo *et al.*, 2016), como modelos MLM, que incluyen, además, una matriz de parentesco (Roselló *et al.*, 2019). Ambas aproximaciones se empearon en este estudio, obteniéndose resultados muy similares, probablemente debido al bajo grado de parentesco en la colección de variedades analizadas.

Independientemente de la aproximación empleada, es necesaria la aplicación de una corrección sobre los p-values obtenidos para establecer un nivel de significación apropiado y evitar la aparición de falsos positivos. Esto es debido a que un GWAS conlleva la realización de un test de comparaciones múltiples en el que el número de test independientes realizados depende del desequilibrio de ligamiento (propiedad de los alelos de loci diferentes de no segregarse de forma independiente). Debido a lo complejo de su estimación, se ha usado un enfoque más pragmático en el que se considera que el máximo número de tests que se llevan a cabo es igual al número total de marcadores (4.713). La corrección por Bonferroni es el método de corrección más

estricto, pero más sencillo de aplicar. Consiste en dividir el nivel de significación original ($\alpha < 0,05$) entre el número de tests independientes realizados para obtener un nivel de significación global aplicable al conjunto de tests. Existen otros tipos de correcciones de la significancia estadística (FDR, test de variables al azar, etc.). En este estudio se han aplicado los métodos de corrección que contiene el paquete *multtest* (Pollard *et al.*, 2005): Bonferroni y modificaciones (Holm, Hochberg, SidakSS y SidakSD) y métodos basados en FDR (BH, BY, ABH y TSBH).

La aplicación de los distintos métodos de corrección de la significación estadística resultó en la ausencia de asociaciones estadísticamente significativas para todos los caracteres relacionados con la arquitectura radicular (Figura 11A).

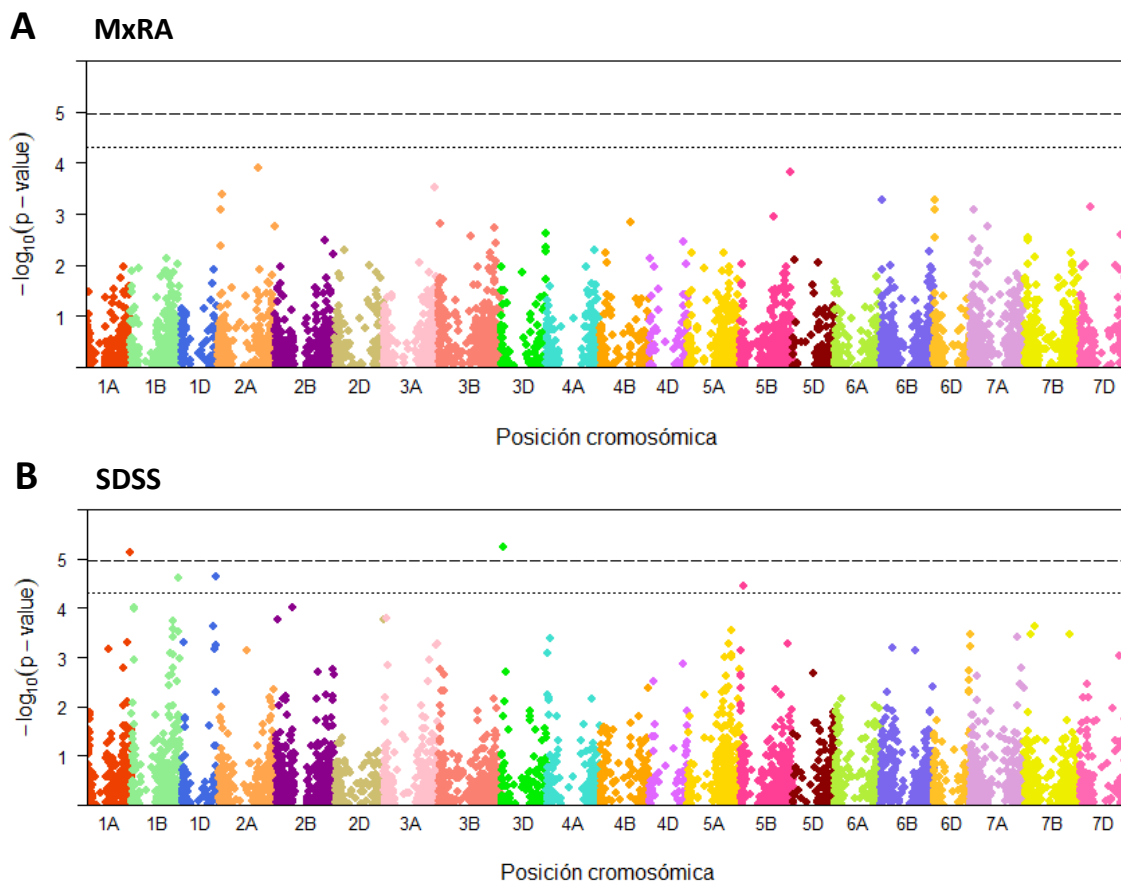


Figura 11. Manhattan plots de los estudios de asociación genómica (GWAS) para los caracteres MxRA (A) y SDSS (B). Las líneas discontinuas superior e inferior representan el valor de significancia estadística umbral en base al método de Bonferroni y FDR (False Discovery Rate), respectivamente.

Con el fin de validar la metodología, se realizó el mismo estudio de asociación para la fuerza del gluten (SDSS), un carácter altamente correlacionado con el contenido proteico. Es un carácter cuantitativo controlado por genes de efecto mayor: los genes que codifican las gluteninas, las principales proteínas del endospermo (Morris *et al.*, 2007). Estos genes se localizan principalmente en el cromosoma 1 de los tres genomioms que porta *T. aestivum* (AABBDD). Por

tanto, se espera encontrar marcadores asociados en las regiones próximas a dichos genes. El método Bonferroni y sus modificaciones identificaron dos marcadores asociados significativamente al carácter SDSS en los cromosomas 1A y 3D, respectivamente (Figura 8B). Los métodos basados en FDR detectaron tres marcadores adicionales asociados significativamente al carácter SDSS en los cromosomas 1B, 1D y 5B. Estos resultados ponen de manifiesto que los métodos basados en FDR aplican algoritmos menos restrictivos en la identificación de asociaciones estadísticamente significativas (Diz *et al.*, 2011) y que los falsos positivos (asociaciones detectadas en los cromosomas 3D y 5B) constituyen un grave problema.

Los resultados obtenidos para el carácter SDSS revelan que en la colección de 85 variedades empleadas en este trabajo es posible detectar asociaciones, al menos para caracteres en los que hay genes con claramente efectos mayores. Aun así, el número de asociaciones detectadas es muy bajo en comparación con un estudio que emplea el set completo de 189 variedades, donde se detectan más de 20 asociaciones para ese carácter (Pascual *et al.*, 2021)

El principal problema de los estudios GWAS es la aparición de falsos positivos, una forma de validar los resultados es mediante la representación de gráficos QQPlots (significancia estadística obtenida en el análisis vs. aquella que se obtendría si se considera la H_0 : no hay asociaciones). Puesto que la mayoría de los fenotipos están controlados por unos pocos genes, lo que esperamos es que existan pocas asociaciones, es decir, que en los QQplots difieran poco los valores observados de los esperados por azar. es decir, que se ajusten a la diagonal, especialmente en los primeros tramos. Cuando el QQPlot de todos los caracteres incluidos en el análisis, se observa que los caracteres de arquitectura radicular cumplen este principio, mientras que el SDSS se desvía de lo esperado, especialmente en el tramo final (Figura 12). Esto está en consonancia con lo descrito en el apartado anterior: para los caracteres de arquitectura radicular no se detectan asociaciones y para el SDSS sí se detectan, pero pocas.

El poder de GWAS para identificar una verdadera asociación entre un SNP y un rasgo depende de la varianza fenotípica dentro de la población explicada por el SNP, y esto está relacionado con el tamaño poblacional y con el número de marcadores analizados (Korte & Farlow, 2013). El número de marcadores utilizado en este trabajo se considera lo suficientemente elevado para una especie de genoma grande y complejo como es el trigo. Otros estudios de GWAS en trigo han utilizado un número de marcadores inferior (Giraldo *et al.*, 2016; Rosello *et al.*, 2019)

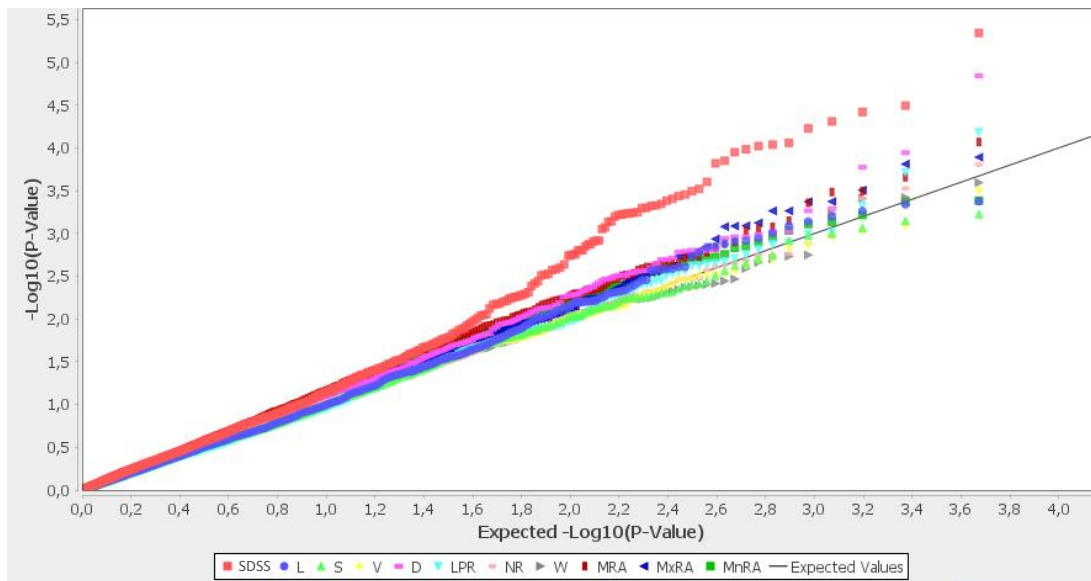


Figura 12. QQPlots de los caracteres analizados. Las curvas de colores representan el $-\log_{10}$ de los p -valores obtenidos en el estudio de asociación para cada carácter individual en función de los p -valores que se obtendrían si no existen asociaciones. La línea negra corresponde a la diagonal.

La falta de asociaciones estadísticamente significativas en este trabajo es más probable que se deba principalmente al tamaño poblacional. Aunque no hay un tamaño mínimo establecido, los estudios GWAS con menos de 100 individuos no son recomendables, aunque el éxito del análisis depende también de la variabilidad existente en la colección y de la base genética de los caracteres analizados. La mayoría de los caracteres de interés, como los relacionados con la arquitectura radicular, son poligénicos, controlados por varios genes de efecto pequeño; en estos casos, aumentar el tamaño de la muestra aumentará el poder de recuperar asociaciones significativas (Korte & Farlow, 2013).

3.4 Análisis de asociación de marcadores individuales

En un trabajo anterior de asociación genómica realizado con 139 variedades locales españolas de trigo duro se obtuvieron 50 marcadores asociados significativamente a los mismos 10 caracteres de arquitectura radicular analizados en este trabajo (Pascual *et al.*, 2021). Aprovechando el elevado grado de parecido genómico entre las dos principales especies de trigo, se analizó si alguno de los marcadores asociados identificados en ese estudio en trigo duro se encontraba en el set de 4.713 marcadores analizados en este trabajo y se hallaron 5 marcadores comunes. Estos 5 marcadores candidatos fueron sometidos a un análisis de asociación individual que incluyó un *Single Marker Analysis* (SMA) y una comparación de las frecuencias alélicas de las variedades de fenotipos extremos. Este análisis no es adecuado cuando se analizan el elevado número de marcadores necesarios para una apropiada cobertura

del genoma, pero es útil para la validación de un reducido grupo de marcadores como paso posterior al GWAS.

El SMA consiste en la realización de una prueba estadística de comparación de medias (t-test o ANOVA, en función del número de alelos) para comprobar la existencia de diferencias en los valores medios fenotípicos de los grupos de variedades que presentan cada uno de los dos alelos. Así, se pudo relacionar la presencia de un alelo a la mayor o menor expresión de un carácter cuantitativo. Los resultados del SMA dieron un resultado positivo para el marcador 1204133, ya que la presencia del alelo “G” del marcador se encontraba asociada a valores superiores del carácter MxRA (Tabla 1).

Tabla 1. Resultado del Single Marker Analysis (SMA) de los marcadores candidatos comunes con un estudio GWAS en trigo duro. El resultado del t-test se representa mediante “ns” (no significativo) o ** (p-value < 0,01

Marcador	Carácter	Resultado	Alelo 1	Media alelo 1	Media alelo 2	Variedades alelo 1	Variedades alelo 2
1203632 G/A	LPR	ns	A	21,14	19,98	9	71
1054888 T/G	MRA	ns	T	31,26	30,52	43	38
1245259 C/T	D	ns	T	0,052	0,051	15	65
2294505 A/T	MxRA	ns	A	43,94	43,77	52	27
1204133 G/C	MxRA	**	G	45,65	41,76	41	35

Posteriormente, se seleccionaron las 8 variedades (10% del total) que presentaban un fenotipo extremo para los caracteres analizados y se compararon las alélicas de los 5 marcadores presentes en los dos grupos de variedades. Para considerar asociado un marcador, sus frecuencias deben ser significativamente distintas en los dos grupos. El marcador “1204133|G/C” mostró un p-value < 0,05 y unas frecuencias de 0,75 (C) y 0,25 (G) en el grupo de variedades de menor expresión del carácter cuantitativo MxRA (Tabla 2). Estas frecuencias se invertían en el grupo complementario. Por tanto, se concluyó que dicho marcador estaba asociado al carácter MxRA. En el estudio GWAS realizado, este marcador exhibió un p-value de 0,0236 y no fue detectado tras aplicar la corrección de la significancia estadística, lo que pone de manifiesto, de nuevo, la necesidad de incrementar el tamaño muestral de la colección a analizar.

Tabla 2. Resultado de la comparación de las frecuencias alélicas entre el 10% de variedades de fenotipo extremo en los marcadores candidatos compartidos con un estudio realizado en trigo duro.

Marcador	Carácter	Resultado	Frecuencias mínimas		Frecuencias máximas	
			Alelo 1	Alelo 2	Alelo 1	Alelo 2
1203632 G/A	LPR	ns	0,125	0,875	0,125	0,875
1054888 T/G	MRA	ns	0,625	0,375	0,5	0,5
1245259 C/T	D	ns	0,875	0,125	0,625	0,375
2294505 A/T	MxRA	ns	0,875	0,125	0,875	0,125
1204133 G/C	MxRA	**	0,75	0,25	0,25	0,75

Finalmente, se procedió a la evaluación de la asociación de marcadores relacionados con la vernalización (*Vrn-A1* y *Vrn-B1*) y el estrés hídrico (enzimas claves del metabolismo vegetal), dos fenómenos estrechamente relacionados con el desarrollo radicular. Con respecto a los marcadores de vernalización, no se obtuvieron resultados significativos, lo que apoya los resultados obtenidos en el análisis del tipo de floración descritos en 3.2.3. No obstante, el gen *Vrn-A1*, que codifica para un factor de transcripción, se ha asociado en otros estudios con la longitud de la raíz seminal en trigo (Voss-Fels *et al.*, 2018).

Debido a la naturaleza multialélica del set de marcadores relativos al estrés hídrico, en este caso se llevó a cabo un SMA en el que la comparación de las medias de las variedades que presentan cada alelo se realizó mediante un ANOVA. Se encontraron diferencias estadísticamente significativas para los alelos de la enzima PEPC (*PhosphoEnolPyruvate Carboxylase*) en los caracteres MRA y MxRA (p-values = 0,00617 y 0,00669) (Apéndice C).

La PEPC carboxilasa es una enzima implicada en la biosíntesis de metabolitos secundarios que se ha relacionado con la tolerancia al stress hídrico en cereales (Quin *et al.*, 2016; González *et al.*, 2003). En trigo, se ha visto que su expresión es elevada en la raíz, y que se induce específicamente por la falta de agua (González *et al.*, 2003). La existencia de variabilidad alélica para esta enzima en nuestra colección podría estar relacionado con distintos niveles de expresión de esta enzima que confieran distinta adaptabilidad al stress hídrico y que eso esté relacionado con la arquitectura radicular. Sería interesante profundizar en este tema, analizando cuál es la influencia del polimorfismo detectado en la expresión génica, y ampliando el estudio de variabilidad a una colección más amplia de variedades, y a la colección de trigos duro.

3.5 Análisis de marcadores asociados

Se consideró que la región de asociación era la que delimitada por los marcadores adyacentes al marcador asociado. En el caso de “1204133|G/C”, esta región abarcaba 4,64 M y contenía 51 genes anotados como de elevada calidad en las bases de datos. Entre los tipos de proteínas más repetidas se encontraban factores de transcripción, proteínas de unión a metales, quinasas y proteínas de membrana. La elevada variabilidad funcional complicó la búsqueda de genes candidatos para el control parcial del carácter MxRA. Apareamientos de las secuencias codificantes de los genes candidatos contra la base de datos de proteínas de *Arabidopsis thaliana* del NCBI vía BLASTp y exploraciones en bases de datos como Uniprot revelaron que uno de los genes candidatos (TraesCS7B02G299100) codificaba para la proteína GNOM, un factor intercambiador de guaninas que actúa activando proteínas ARF (*Auxin Response Factor*) mediante el intercambio de GDP por GTP libre. Este factor involucrado en el tráfico de vesículas endosomales determinando la localización de los transportadores de auxinas PIN (Steinmann *et al.*, 1999). Además, se ha descrito su relación con el hidrotropismo de las raíces en estado de plántula, lo que puede explicar que este gen esté asociado al carácter MxRA.

En vista de algunos resultados prometedores en el GWAS realizado cercanos a la significación (Figura 11A), también se quiso investigar el marcador “2244360|C/T”, localizado en el cromosoma 6D, donde se ha descrito la existencia de QTLs asociados al carácter MxRA (Roselló *et al.*, 2019). La región de asociación abarcaba 0,4 Mb y contenía 18 genes, muchos de ellos factores de transcripción. Uno de los genes candidatos detectado (TraesCS6D02G023100) codificaba para AHP2, una histidina fosfotransferasa implicada en la cascada de señalización de las citoquininas. Además, se ha descrito que sus niveles de expresión son elevados en raíces y se inducen por estrés salino, térmico e hídrico (Miyata *et al.*, 1998).

En conjunto, estos dos resultados reflejan la enorme importancia de la regulación hormonal en el control de la arquitectura radicular. Mientras que las auxinas actúan como promotores del crecimiento radicular, las citoquininas ejercen de inhibidores (Bielach *et al.*, 2012). Dado el papel antagonista de este par de hormonas en el desarrollo de la raíz, se teoriza que el polimorfismo del marcador 1204133|G/C estaría relacionado con una mejora de la funcionalidad de GNOM y el de 2244360|C/T afectaría también positivamente al desempeño de AHP2, ya que su inducción en condiciones de estrés resultaría en un menor desarrollo radicular. Con todo, ambos genes se postulan como buenos candidatos para futuros análisis.

CAPÍTULO 4. CONCLUSIONES

- I. La colección de 85 variedades locales españolas de trigo blando analizada en este trabajo presenta una variabilidad fenotípica superior a la descrita en estudios previos e idónea para la realización de estudios de asociación.
- II. El análisis PCoA en base a datos genotípicos de marcadores moleculares corroboró la existencia de una baja estructura poblacional adecuada para llevar a cabo estudios GWAS.
- III. De acuerdo con el análisis de la varianza (ANOVA), el factor poblacional influye de forma significativa en cuatro de los diez caracteres analizados, donde la población 1 exhibió valores superiores. A su vez, el factor tipo de floración no mostró una influencia significativa en ninguno de los caracteres
- IV. El estudio GWAS reveló que podrían existir asociaciones de interés pero que para ser detectadas significativamente se debería aumentar el tamaño de la colección analizada.
- V. Se ha detectado una asociación entre la variabilidad alélica para la enzima PEPC, implicada en la respuesta al estrés hídrico, y los caracteres MRA y MxRA
- VI. Se han encontrado dos marcadores asociados al carácter MxRA, con genes relacionados con el equilibrio hormonal en la raíz, lo que los convierte en candidatos a controlar la morfología de las raíces en trigo. Por tanto, presentan un potencial interés para el desarrollo de marcadores moleculares de aplicación en programas de mejora.

CAPÍTULO 5. BIBLIOGRAFÍA

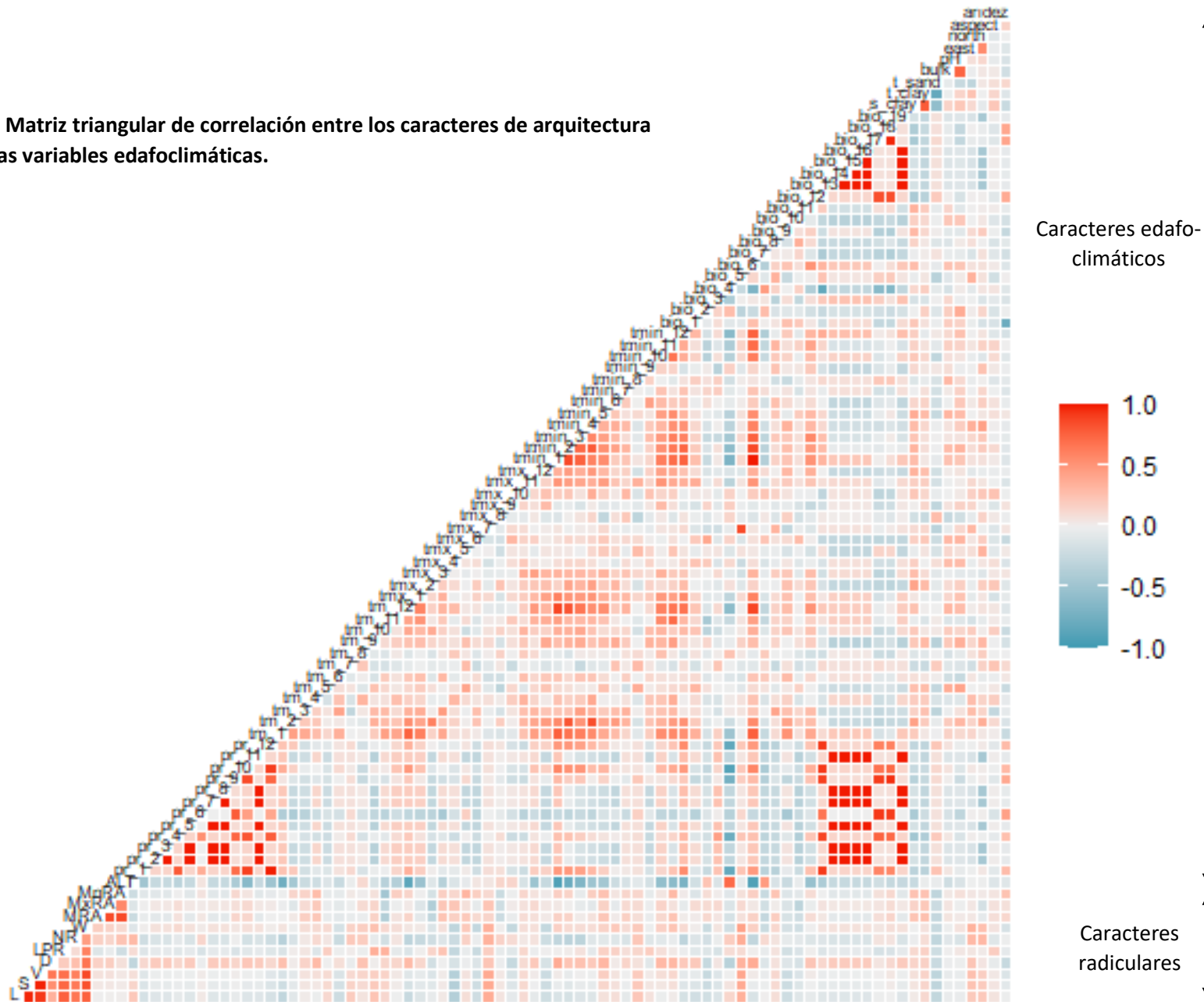
- Abdi, H. (2007). The Bonferonni and Šidák Corrections for Multiple Comparisons. *Encyclopedia of Measurement and Statistics*, 3, 103–107.
- Alahmad, S., El Hassouni, K., Bassi, F. M., Dinglasan, E., Youssef, C., Quarry, G., Aksoy, A., Mazzucotelli, E., Juhász, A., Able, J. A., Christopher, J., Voss-Fels, K. P., & Hickey, L. T. (2019). A major root architecture QTL responding to water limitation in durum wheat. *Frontiers in Plant Science*, 10(April), 1–18. <https://doi.org/10.3389/fpls.2019.00436>
- Altschul, S. F., Gish, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215, 403–410. <https://pubmed.ncbi.nlm.nih.gov/2231712/>
- Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., Pozniak, C. J., Choulet, F., Distelfeld, A., Poland, J., Ronen, G., Barad, O., Baruch, K., Keeble-Gagnère, G., Mascher, M., Ben-Zvi, G., Josselin, A. A., Himmelbach, A., Balfourier, F., ... Wang, L. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, 361(6403). <https://doi.org/10.1126/science.aar7191>
- Bibi, S., Dahot, M. U., Khan, I. A., Khatri, A., & Naqvi, M. H. (2009). Study of genetic diversity in wheat (*Triticum aestivum* L.) using random amplified polymorphic DNA (RAPD) markers (*TRITICUM AESTIVUM* L.) USING RANDOM AMPLIFIED. *Pakistan Journal of Botany*. 41(3), 1023-1027.
- Bielach, A., Duclercq, J., Marhavý, P., Benková, E., Marhavy, P., & Benkova, E. (2012). Genetic approach towards the identification of auxin – cytokinin crosstalk components involved in root development Genetic approach towards the identification of auxin – cytokinin crosstalk components involved in root development. *Philosophical transactions of the royal society*. 367, 1469-1478. <https://doi.org/10.1098/rstb.2011.0233>
- Boudiar, R., González, J. M., Mekhlouf, A., Casas, A. M., & Igartua, E. (2020). Durum wheat seminal root traits within modern and landrace germplasm in Algeria. *Agronomy*, 10(5). <https://doi.org/10.3390/agronomy10050713>
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). *TASSEL : software for association mapping of complex traits in diverse samples*. 23(19), 2633–2635. <https://doi.org/10.1093/bioinformatics/btm308>
- Clark, L. J., Price, A. H., Steele, K. A., & Whalley, W. R. (2008). Evidence from near-isogenic lines that root penetration increases with root diameter and bending stiffness in rice. *Functional Plant Biology*, 35, 1163–1171.
- Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., & De Hoon, M. J. L. (2009). Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11), 1422–1423. <https://doi.org/10.1093/bioinformatics/btp163>
- Demir, P., Onde, S., & Severcan, F. (2015). Spectrochimica Acta Part A : Molecular and Biomolecular Spectroscopy Phylogeny of cultivated and wild wheat species using ATR – FTIR spectroscopy q. *SPECTROCHIMICA ACTA PART A: MOLECULAR AND BIOMOLECULAR SPECTROSCOPY*, 135, 757–763. <https://doi.org/10.1016/j.saa.2014.07.025>
- Diz, A. P., Carvajal-Rodríguez, A., & Skibinski, D. O. F. (2011). Multiple hypothesis testing in proteomics: A strategy for experimental work. *Molecular and Cellular Proteomics*, 10(3). <https://doi.org/10.1074/mcp.M110.004374>
- Dubcovsky, J., & Dvorak, J. (2007). Genome Plasticity a Key Factor in the Success of Polyploid Wheat Under Domestication. *Science*, 316, 1862–1866. <https://doi.org/10.1126/science.1143986>
- Ehret, G. B. (2010). Genome-wide association studies: Contribution of genomics to understanding blood pressure and essential hypertension. *Current Hypertension Reports*, 12(1), 17–25. <https://doi.org/10.1007/s11906-009-0086-6>
- El-Beltagy, A., & Madkour, M. (2012). Impact of climate change on arid lands agriculture. *Agriculture and Food Security*, 1, 1–12. <https://doi.org/10.1186/2048-7010-1-3>
- Food and Agriculture Organization of the United Nations (FAO). (2019). *FAOSTAT database*. <http://www.fao.org/faostat/en/#data/QC>
- Giraldo, P., Royo, C., González, M., Carrillo, J. M., & Ruiz, M. (2016). Genetic diversity and association mapping for agromorphological and grain quality traits of a structured collection of durum wheat landraces including subsp. durum, turgidum and diccocon. *PLoS ONE*, 11(11), 1–24.

<https://doi.org/10.1371/journal.pone.0166577>

- Goh, L., & Yap, V. B. (2009). *Effects of normalization on quantitative traits in association test. BMC Bioinformatics*, 10, 415. <https://doi.org/10.1186/1471-2105-10-415>
- 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. *Methods in Molecular Biology*, 888, 67–89. https://doi.org/10.1007/978-1-61779-870-2_5
- Klopfenstein, D. V., Zhang, L., Pedersen, B. S., Ramírez, F., Vesztröcy, A. W., Naldi, A., Mungall, C. J., Yunes, J. M., Botvinnik, O., Weigel, M., Dampier, W., Dessimoz, C., Flick, P., & Tang, H. (2018). GOATOOLS: A Python library for Gene Ontology analyses. *Scientific Reports*, 8(1), 1–17. <https://doi.org/10.1038/s41598-018-28948-z>
- Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS : a review Self-fertilisation makes Arabidopsis particularly well suited to GWAS. *Plant Methods*, 9(1), 29.
- Kuijken, R. C. P., Van Eeuwijk, F. A., Marcelis, L. F. M., & Bouwmeester, H. J. (2015). Root phenotyping: From component trait in the lab to breeding. *Journal of Experimental Botany*, 66(18), 5389–5401. <https://doi.org/10.1093/jxb/erv239>
- Leng, P. fei, Lübberstedt, T., & XU, M. liang. (2017). Genomics-assisted breeding – A revolutionary strategy for crop improvement. *Journal of Integrative Agriculture*, 16(12), 2674–2685. [https://doi.org/10.1016/S2095-3119\(17\)61813-6](https://doi.org/10.1016/S2095-3119(17)61813-6)
- Levene, H. (1960). Robust tests for equality of variances. In I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, & H. B. Mann (Eds.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling* (Stanford U, pp. 278–292).
- Li, T., Ma, J., Zou, Y., Chen, G., Ding, P., Zhang, H., Yang, C., Mu, Y., Tang, H., Liu, Y., Jiang, Q., Chen, G., Qi, P., Wei, Y., Zheng, Y., & Lan, X. (2020). Quantitative trait loci for seeding root traits and the relationships between root and agronomic traits in common wheat. *Genome*, 63(1), 27–36. <https://doi.org/10.1139/gen-2019-0116>
- Lobet, G., Draye, X., & Périlleux, C. (2013). An online database for plant image analysis software tools. *Plant Methods*, 9(1). <https://doi.org/10.1186/1746-4811-9-38>
- Lopes, M. S., El-Basyoni, I., Baenziger, P. S., Singh, S., Royo, C., Ozbek, K., Aktas, H., Ozer, E., Ozdemir, F., Manickavelu, A., Ban, T., & Vikram, P. (2015). Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change. *Journal of Experimental Botany*, 66(12), 3477–3486. <https://doi.org/10.1093/jxb/erv122>
- López-Fernandez, M., Pascual, L., Faci, I., Fernández, M., Ruiz, M., Benavente, E., & Giraldo, P. (2021). Exploring the End-Use Quality Potential of a Collection of Spanish Bread Wheat Landraces. *Plants*, 10(4), 620. <https://doi.org/https://doi.org/10.3390/plants10040620>
- Manschadi, A. M., Christopher, J., Devoil, P., & Hammer, G. L. (2006). The role of root architectural traits in adaptation of wheat to water-limited environments. *Functional Plant Biology*, 33(9), 823–837. <https://doi.org/10.1071/FP06055>
- Marcussen, T., Sandve, S. R., Heier, L., Spannagl, M., Pfeifer, M., Jakobsen, K. S., Wulff, B. B., Steuernagel, B., Mayer, K. F., & Olsen, O. A. (2014). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome Ancient hybridizations among the ancestral genomes of bread wheat. *Science*, 345(6194). <https://doi.org/10.1126/science.1250092>
- Mérida-García, R., Gálvez, S., Paux, E., Dorado, G., Pascual, L., Giraldo, P., & Hernandez, P. (2020). High Resolution Melting and Insertion Site-Based Polymorphism Markers for Wheat Variability Analysis and Candidate Genes Selection at Drought and Heat MQTL Loci. *Agronomy*, 10(9), 1294. <https://doi.org/10.3390/agronomy10091294>
- Miyata, S., Urao, T., Y, K. Y., & Shinozaki, K. (1998). Characterization of genes for two-component phosphorelay mediators with a single Hpt domain in *Arabidopsis thaliana*. *FEBS Letters*, 437(1–2), 11–14.
- Morris, C. F., Paszczyńska, B., Bettge, A. D., & King, G. E. (2007). A critical examination of the sodium dodecyl sulfate (SDS) sedimentation test for wheat meals. *Journal of the Science of Food and Agriculture*, 87, 607–615. <https://doi.org/https://doi.org/10.1002/jsfa.2740>
- Nagel, K. A., Kastenholz, B., Jahnke, S., & van Dusschoten, D. (2009). Temperature responses of roots : impact on growth , root system architecture and implications for phenotyping. *Functiona*, 36, 947–959.

- Pascual, L., González, J. M., García-Abadillo, J., Loarce, Y., Mérida, R., Hernández, P., Giraldo, P., & Ruiz, M. (2021). GWAs in Durum wheat identified new genomic regions related with root system architecture. *Comunicación Presentada Al XLII Congreso de La SEG, 14-18 Junio de 2021*.
- Pascual, L., Ruiz, M., López-Fernández, M., Pérez-Penã, H., Benavente, E., Vázquez, J. F., Sansaloni, C., & Giraldo, P. (2020). Genomic analysis of Spanish wheat landraces reveals their variability and potential for breeding. *BMC Genomics*, 21(1). <https://doi.org/10.1186/s12864-020-6536-x>
- Pollard, K. S., Dudoit, S., & van der Laan, M. J. (2005). *Multiple Testing Procedures: R multtest Package and Applications to Genomics, in Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (pp. 251–272). Springer (Statistics for Biology and Health Series).
- R Core Team. (2021). *R: A Language and Environment for Statistical Computing* (R version 4.1.0). R Foundation for Statistical Computing. <https://www.r-project.org/>
- Rasheed, A., Mujeeb-Kazi, A., Ogonnaya, F. C., He, Z., & Rajaram, S. (2018). Wheat genetic resources in the post-genomics era: Promise and challenges. *Annals of Botany*, 121(4), 603–616. <https://doi.org/10.1093/aob/mcx148>
- Roselló, M., Royo, C., Sanchez-García, M., & Soriano, J. M. (2019). Genetic Dissection of the Seminal Root System Architecture in Mediterranean Durum Wheat Landraces by Genome-Wide Association Study. *Agronomy*, 9, 364.
- Rufo, R., Salvi, S., Royo, C., & Soriano, J. M. (2020). Exploring the genetic architecture of root-related traits in Mediterranean bread wheat landraces by genome-wide association analysis. *Agronomy*, 10(5). <https://doi.org/10.3390/agronomy10050613>
- Ruiz, M., Giraldo, P., & González, J. M. (2018). Phenotypic variation in root architecture traits and their relationship with eco-geographical and agronomic features in a core collection of tetraploid wheat landraces (*Triticum turgidum* L.). *Euphytica*, 214(3). <https://doi.org/10.1007/s10681-018-2133-3>
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3–4), 591–611. <https://doi.org/10.1093/biomet/52.3-4.591>
- Skovmand, B., Reynolds, M. P., & Delacy, I. H. (2001). Mining wheat germplasm collections for yield enhancing traits. In *Wheat in a Global Environment* (pp. 761–771). Springer.
- Steinmann, T., Geldner, N., Grebe, M., Mangold, S., Jackson, C. L., Paris, S., Gälweiler, L., Palme, K., & Jürgens, G. (1999). Coordinated polar localization of auxin efflux carrier PIN1 by GNOM ARF GEF. *Science*, 286(5438), 316–318. <https://doi.org/10.1126/science.286.5438.316>
- Tukey, J. W. (1949). Comparing Individual Means in the Analysis of Variance. *International Biometric Society*, 5(2), 99–114. <https://www.jstor.org/stable/3001913>
- Voss-Fels, K. P., Robinson, H., Mudge, S. R., Richard, C., Newman, S., Wittkop, B., Trevaskis, B., Snowdon, R. J., & Hickey, L. T. (2018). VERNALIZATION1 Modulates Root System Architecture in Wheat and Barley. *Molecular Plant*, 11(1), 226–229. <https://doi.org/10.1016/j.molp.2017.10.005>
- Wickland, D. P., Battu, G., Hudson, K. A., Diers, B. W., & Hudson, M. E. (2017). A comparison of genotyping-by-sequencing analysis methods on low-coverage crop datasets shows advantages of a new workflow, GB-eaSy. *BMC Bioinformatics*, 18(1), 58. <https://doi.org/10.1186/s12859-017-2000-6>
- Zargar, S. M., Raatz, B., Sonah, H., Muslimanazir, Bhat, J. A., Dar, Z. A., Agrawal, G. K., & Rakwal, R. (2015). Recent advances in molecular marker techniques: Insight into QTL mapping, GWAS and genomic selection in plants. *Journal of Crop Science and Biotechnology*, 18(5), 293–308. <https://doi.org/10.1007/s12892-015-0037-5>
- Zhang, Z., Ersoz, E., Lai, C., Todhunter, R. J., Tiwari, H. K., Gore, M. A., Bradbury, P. J., Yu, J., Arnett, D. K., Ordoñas, J. M., & Buckler, E. S. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nature Publishing Group*, 42(4), 355–360. <https://doi.org/10.1038/ng.546>
- Zhu, J., Ingram, P. A., Benfey, P. N., & Elich, T. (2011). From lab to field, new approaches to phenotyping root system architecture. *Current Opinion in Plant Biology*, 14(3), 310–317. <https://doi.org/10.1016/j.pbi.2011.03.020>

Apéndice A. Matriz triangular de correlación entre los caracteres de arquitectura radicular y las variables edafoclimáticas.



Apéndice B. Leyenda de abreviaturas de la matriz triangular de correlación entre los caracteres de arquitectura radicular y las variables edafoclimáticas

Componentes ecogeográficas	Descripción	Abreviatura	Unidades
Bioclimáticas	Precipitaciones mensuales (Enero – Diciembre)	prec_1 a 12	mm
	Temperaturas medias mensuales (Enero – Diciembre)	tmean_1 a 12	°C
	Temperaturas mínimas mensuales (Enero – Diciembre)	tmin_1 a 12	°C
	Temperaturas máximas mensuales (Enero – Diciembre)	tmax_1 a 12	°C
	Temperatura media anual	bio_1	°C
	Rango medio de temperatura diaria (media mensual (max temp – min temp))	bio_2	°C
	Isotermalidad (bio_2/bio_7) (* 100)	bio_3	-
	Temperatura estacional (desviación estándar * 100)	bio_4	-
	Temperatura máxima del mes más cálido	bio_5	°C
	Temperatura mínima del mes más frío	bio_6	°C
	Rango de temperatura anual (BIO5-BIO6)	bio_7	°C
	Temperatura media del trimestre más lluvioso	bio_8	°C
	Temperatura media del trimestre más seco	bio_9	°C
	Temperatura media del trimestre más cálido	bio_10	°C
	Temperatura media del trimestre más frío	bio_11	°C
	Precipitación anual	bio_12	mm
	Precipitación del mes más lluvioso	bio_13	mm
	Precipitación del mes más seco	bio_14	mm
	Geofísicas	Precipitación estacional (coeficiente de variación)	bio_15
Precipitación del trimestre más lluvioso		bio_16	mm
Precipitación del trimestre más seco		bio_17	mm
Precipitación del trimestre más cálido		bio_18	mm
Precipitación del trimestre más frío		bio_19	mm
Norticidad (1 si mira al norte, - 1 si mira al sur)		North	°
Esticidad (1 si mira al este, - 1 si mira al oeste)		East	°
Edáficas	Altitud	Metros sobre el mar	m
	Aspecto (orientación de la superficie del terreno)	Aspect	°
	Fracción superior arcillosa	t_clay	% peso
	Fracción superior arenosa	t_sand	% peso
	Fracción inferior arcillosa	s_clay	% peso
pH de la fracción superior	pH	-log(H ⁺)	
Densidad másica de la fracción superior	bulk	% peso	

Apéndice C. Resultado del SMA para los marcadores relacionados con estrés hídrico y tipo de floración. Se destacan en rojo las asociaciones significativas.
 RBCL = RuBisCO subunidad grande; RBCS = RuBisCO subunidad pequeña; PEPC = fosfoenolpiruvato carboxilasa; CAT = catalasa; RLI = inhibidor de RNasa L; GS = glutamina sintetasa; DHN = dehidrina

Carácter	RBCL	RBCS	PEPC	CAT	RLI	GS1	GS2	DHN11	DHN16	Marcador 1 VrnA1	Marcador 2 VrnA1	Marcador 3 VrnA1	VrnB1
L	0.125	0.904	0.420	0.235	0.459	0.411	0.312	0.543	0.502	0.716	0.583	0.304	0.468
S	0.186	0.964	0.587	0.303	0.731	0.324	0.327	0.567	0.503	0.669	0.600	0.269	0.445
V	0.273	0.986	0.713	0.379	0.910	0.271	0.345	0.479	0.553	0.637	0.623	0.266	0.445
D	0.960	0.802	0.704	0.859	0.597	0.771	0.662	0.087	0.777	0.766	0.841	0.393	0.455
LPR	0.767	0.545	0.296	0.264	0.830	0.526	0.143	0.434	0.726	0.241	0.903	0.171	0.885
NR	0.065	0.843	0.533	0.440	0.691	0.438	0.853	0.432	0.436	0.151	0.263	0.735	0.160
W	0.932	0.748	0.902	0.304	0.932	0.909	0.886	0.280	0.434	0.927	0.825	0.327	0.485
MRA	0.971	0.666	0.006	0.566	0.841	0.751	0.794	0.447	0.189	0.834	0.862	0.936	0.991
MxRA	0.613	0.662	0.007	0.986	0.958	0.865	0.681	0.276	0.153	0.838	0.345	0.377	0.402
MnRA	0.924	0.517	0.055	0.336	0.689	0.547	0.498	0.639	0.138	0.526	0.482	0.415	0.414