

UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO



Escuela Superior de Apan

TESIS DE LICENCIATURA

Estudio de asociación amplia del genoma de trigos harineros (*Triticum aestivum*) nativos con sus rasgos de proteína en grano y sus propiedades viscoelásticas del gluten

Para obtener el título de

Licenciado en Ingeniería en Biociencias

P R E S E N T A

Jose Guillermo Alvarado Juarez

Directora:

Dra. Rosa Angela Pacheco Gil

Codirector:

Dr. Martin Peralta Gil

Apan, Hgo., México a 2 de mayo de 2025

Universidad Autónoma del Estado de Hidalgo

Escuela Superior de Apan

Título de la tesis

Estudio de asociación amplia del genoma de trigos harineros (*Triticum aestivum*)
nativos con sus rasgos de proteína en grano y sus propiedades viscoelásticas del
gluten

Tesis

Para obtener el título de la Licenciatura en Ingeniera en Biociencias

P R E S E N T A

Jose Guillermo Alvarado Juarez

Directora:

Dra. Rosa Angela Pacheco Gil

Codirector:

Dr. Martin Peralta Gil

Apan, Hgo., México a 2 de mayo de 2025

El presente trabajo se realizó en el Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT), bajo la dirección de Dra. Rosa Ángela Pacheco Gil y codirección de Dr. Martín Peralta Gil, apoyo que agradezco ampliamente.

Este trabajo fue financiado por la Secretaría de Agricultura y Desarrollo Rural (SADER), como parte de las iniciativas *Seeds of Discovery*, Modernización Sustentable de la Agricultura Tradicional (Mas-Agro) y hacer posible este trabajo.

DEDICATORIA

A mis padres, por su amor incondicional, apoyo constante y por estar siempre a mi lado durante todo este camino.

*A la **Dra. Carolina Paola Sansaloni**, la **Dra. Guadalupe Valdez Velázquez** y la **Dra. Rosa Ángela Pacheco Gil**, por su invaluable guía, conocimientos y por acompañarme en esta etapa de mi formación académica.*

Al Programa de Recursos Genéticos (GRP) y al Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT), así como a la Secretaría de Agricultura y Desarrollo Rural (SADER), como parte de las iniciativas Seeds of Discovery, Modernización Sustentable de la Agricultura Tradicional (Mas-Agro) y hacer posible este trabajo.

RESUMEN

Las especies silvestres no domesticadas, así como los parientes silvestres de los cultivos y las variedades locales, son consideradas esenciales para afrontar los desafíos contemporáneos en la agricultura. En este marco, mejorar la calidad del trigo se vuelve imperativo, centrándose particularmente en atributos cruciales como la proteína en el grano (GP) y las propiedades viscoelásticas del gluten (SDS_Sedim), que revisten gran importancia en la industria agroalimentaria. Este estudio tiene como foco la identificación genómica de regiones relacionadas con estas características, utilizando un análisis de asociación del genoma completo (GWAS). Se aplicaron tres enfoques analíticos: a) Modelo Lineal Mixto (MLM), b) Unificación de Probabilidad Circulante de Modelo Fijo y Aleatorio (FarmCPU), y c) Información Bayesiana y Desequilibrio de Ligamiento Anidado Iterativamente (BLINK), mediante el paquete GAPIT de R. Los datos, extraídos de 77,395 *loci* de polimorfismo de un solo nucleótido (SNP's) de 1,602 colectas de *T. aestivum* nativos de varios países y almacenados en el banco de germoplasma del CIMMYT, evidenciaron 28 marcadores asociados con GP y 15 con SDS_Sedim ($p < 0.01$), distribuidos en 20 cromosomas. Se identificaron 8 marcadores en los cromosomas 1A, 1B, 6A, 6B y 6D. Adicionalmente, se generaron haplotipos mediante desequilibrio de ligamiento con la herramienta PLINK, permitiendo la identificación de las mejores frecuencias en las que se detectaron haplotipos/marcadores asociados a variedades particulares. Este enfoque proporciona información contemporánea para desarrollar variedades de trigo elite futuras, explorando regiones genéticas o cromosómicas específicas e identificando germoplasma que conserve la diversidad alélica, para toma de decisiones futuras en programas de mejoramiento.

ABSTRACT

Wild, undomesticated species, along with crop wild relatives and local landraces, are regarded as essential resources for meeting contemporary challenges in agriculture. In this context, improving wheat quality becomes imperative, with a particular focus on key traits such as grain protein content (GP) and the viscoelastic properties of gluten (SDS_Sedim), both of which are of great importance to the agri-food industry. This study aimed to pinpoint genomic regions associated with these characteristics through a genome-wide association study (GWAS). Three analytical approaches were employed—(a) the Mixed Linear Model (MLM), (b) Fixed and Random Model Circulating Probability Unification (FarmCPU), and (c) Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK)—using the GAPIT package in R. Phenotypic and genotypic data were drawn from 77,395 single-nucleotide polymorphisms (SNPs) genotyped across 1,602 *Triticum aestivum* accessions originating from multiple countries and conserved in the CIMMYT germplasm bank. At a significance threshold of $p < 0.01$, GWAS revealed 28 markers associated with GP and 15 with SDS_Sedim, distributed over 20 chromosomes, including eight key markers located on chromosomes 1A, 1B, 6A, 6B, and 6D. Additionally, haplotypes were constructed via linkage disequilibrium analysis using PLINK, enabling the identification of the most frequent haplotype-marker combinations tied to specific varieties. This integrative approach delivers up-to-date genomic insights to guide the development of future elite wheat cultivars by targeting particular genetic or chromosomal regions and by selecting germplasm that maintains valuable allelic diversity for informed decision-making in breeding programs.

ÍNDICE

DEDICATORIA	I
RESUMEN	II
ABSTRACT	III
1. INTRODUCCIÓN	1
2. ANTECEDENTES	5
2.1. HISTORIA DEL TRIGO	5
2.2. GENERALIDADES DEL TRIGO.	6
2.3. TAXONOMÍA.....	6
2.4. IMPORTANCIA DEL TRIGO EN MÉXICO	7
2.5. CÁLIDA DE TRIGO.	7
2.6. MEJORAMIENTO TRADICIONAL DEL TRIGO.	9
2.7. MEJORA GENÉTICA EN LA CALIDAD DEL TRIGO MEDIANTE MARCADORES MOLECULARES.....	10
2.7.1. <i>Marcadores moleculares.</i>	10
2.7.2. <i>Polimorfismos de Nucleótido Único (SNP).</i>	11
2.7.3. <i>Polimorfismos de presencia-ausencia (PAV).</i>	12
2.8. SECUENCIACIÓN DE NUEVA GENERACIÓN (NGS).....	12
2.9. TECNOLOGÍA DARTSEQ.....	13
2.10. ESTUDIOS DE ASOCIACIÓN DEL GENOMA COMPLETO (GWAS).....	15
2.10.1. <i>Modelos estadísticos para GWAS.</i>	17
2.10.2. <i>Gráficos y medidas estadísticas para el análisis de GWAS.</i>	20
2.10.3. <i>Desequilibrio de ligamiento (LD).</i>	21
2.10.4. <i>Haplotipos.</i>	22
3. JUSTIFICACIÓN	23
4. HIPÓTESIS.....	24
5. OBJETIVOS	25
5.1. OBJETIVO GENERAL.	25
5.2. OBJETIVOS ESPECÍFICOS.....	25
• <i>Comprobar la calidad de la base de datos fenotípica y genotípica de las accesiones de trigos harineros nativos para garantizar la confiabilidad de los resultados en los análisis de asociación genómica.</i>	25
• <i>Realizar un análisis de asociación del genoma completo (GWAS) para detectar regiones asociadas con la calidad del grano en trigo harinero.</i>	25
• <i>Identificar y reportar haplotipos y SNP's significativos que puedan ser utilizados como marcadores moleculares en programas de mejoramiento genético.</i>	25
• <i>Explorar la posible función de las regiones genómicas detectadas mediante la búsqueda de genes en bases de datos y la comparación con estudios previos.</i>	25
6. METODOLOGÍA.....	26
6.1. CARACTERIZACIÓN GENÉTICA.....	26
6.2. ANÁLISIS FÍSICO Y QUÍMICO.	27
6.3. ANÁLISIS DE ASOCIACIÓN DEL GENOMA COMPLETO (GWAS).	27
6.4. HAPLOTIPOS.	29
6.5. SOFTWARE.	30

7. RESULTADOS Y DISCUSIÓN.....	32
7.1. ESTADÍSTICA DESCRIPTIVA Y ESTRUCTURA DE LA POBLACIÓN.....	32
7.2. ANÁLISIS DE ASOCIACIÓN.....	34
7.2.1. Identificación de Marcadores Significativos en Múltiples Modelos y Cromosomas.....	37
7.3. GRÁFICOS Q-Q.....	38
7.4. MAPEO PRECISO DE MARCADORES DE GWAS Y SUS BLOQUES DE ASOCIACIÓN.....	40
7.5. ANÁLISIS DE HAPLOTIPOS BASADO EN LD: IDENTIFICACIÓN Y DISTRIBUCIÓN CROMOSÓMICA .	44
8. CONCLUSIÓN.....	49
REFERENCIAS	50
LISTA DE FIGURAS.....	VI
LISTA DE TABLAS.....	VIII
GLOSARIO DE TÉRMINOS Y SIGLAS	IX
ANEXOS	X

1. Introducción

El trigo (*Triticum aestivum*) desempeña un papel crucial en la mayoría de los países a nivel mundial, destacándose como una fuente esencial de proteína vegetal. Su importancia se extiende a los ámbitos económicos y políticos debido a su estabilidad y capacidad para satisfacer la creciente demanda de alimentos. Según los datos de la Organización de las Naciones Unidas para la Alimentación y la Agricultura, FAO, (por sus siglas en inglés: Food and Agriculture Organization), aproximadamente el 50% de la superficie total de tierra cultivable a nivel mundial, que asciende a 219 millones de hectáreas en los últimos años, está ocupada por variedades de trigo (FAO, 2024). Aunque la producción anual mundial de trigo ha experimentado un aumento constante, alcanzando una proyección de 2,813 millones de toneladas para 2023/24, la demanda de trigo también ha crecido de manera sostenida (FAO, 2024). La mayoría de las variedades modernas de trigo se clasifican generalmente como "trigo común", cuya harina es altamente adecuada para la industria agroalimentaria (FAO, 2024).

En la actualidad, en el sector agroalimentario, múltiples variables desempeñan roles en la cadena de valor del trigo, desde el productor hasta el consumidor. Para el agricultor, un cultivo de trigo de alta calidad puede significar el uso reducido de recursos, la obtención del mayor rendimiento de granos y la posibilidad de venderlos al precio más elevado del mercado. De esta forma, comprende el concepto de calidad del trigo y cómo incorporarlo en el proceso de mejora. No obstante, el molinero evaluará las variedades en función del desempeño de los granos para la producción de harina, considerando también los requisitos energéticos necesarios. Por último, el panadero seleccionará y discriminará estos materiales según su idoneidad para cada producto horneado específico (Guzmán *et al.*, 2022a).

Sin embargo, para que esta cadena de valor funcione de manera efectiva, es esencial contar con el trigo nativo (*T. aestivum*), ya que cada una de ellas presenta características únicas (Kumar *et al.*, 2022). Uno de los aspectos importantes en la evaluación de la calidad del trigo es el contenido de proteína en el grano (GP), ya que determina tanto su valor nutricional como sus propiedades funcionales en la panificación. Entre estas proteínas destacan las gliadinas y gluteninas, las cuales constituyen aproximadamente el 80% del gluten y son responsables de la viscoelasticidad de la masa, afectando directamente la calidad de los productos horneados (Wrigley *et al.*, 2006; Shewry, 2019).

Otro parámetro fundamental es el ensayo de sedimentación con dodecil sulfato de sodio (SDS_Sedim), ampliamente utilizada para estimar la calidad del gluten en el trigo. Este análisis mide la capacidad de las proteínas de almacenamiento para formar una estructura reticulada en presencia de agentes reductores, lo que permite predecir la fuerza y elasticidad de la masa (Wang *et al.*, 2015). La SDS-Sedimentación es un indicador clave en la selección de variedades de trigo con buena calidad panadera, ya que correlaciona positivamente con el contenido de gluteninas de alto peso molecular (HMWG), las cuales son determinantes en la fuerza de la harina y su capacidad para retener gas en la fermentación (Barak *et al.*, 2013).

Dado su importancia, es fundamental considerar la relación de estas proteínas con los cromosomas 1 y 6 (A, B y D), ya que estos contienen los genes que codifican para las principales proteínas del gluten. Las gluteninas de bajo (LMWG) y alto peso molecular (HMWG), así como las gliadinas ω/γ , están localizadas en los cromosomas 1A, 1B y 1D, mientras que las gliadinas $\alpha/\beta/\gamma$ están presentes en los cromosomas 6A, 6B y 6D (Huo *et al.*, 2019; Sansaloni *et al.*, 2020; Guzmán *et al.*, 2022a). Esta asociación genética es clave para comprender la influencia de la variabilidad genómica en la calidad del trigo, y su estudio mediante herramientas genómicas avanzadas resulta indispensable para el mejoramiento genético.

La utilización eficaz del germoplasma en el proceso de mejoramiento radica en la necesidad de identificar las semillas que albergan variantes genéticas favorables y en abordar el problema del arrastre de ligamiento asociado con la introducción de variantes genéticas deseadas en el germoplasma (Sansaloni *et al.*, 2020). De acuerdo a lo anterior y gracias a los recientes avances en genómica y tecnologías moleculares, ha sido posible simplificar la caracterización de la diversidad genética en vastas colecciones de germoplasma, facilitando el uso de estos recursos genéticos en plantas para mejorar líneas de trigo de élite (Chu *et al.*, 2020).

Estos avances han permitido estimar con mayor precisión la diversidad molecular de los recursos genéticos en comparación con las líneas de élite (He *et al.*, 2019). También han proporcionado estrategias más eficientes, como la edición génica, para abordar problemas como el arrastre de ligamiento, un obstáculo clave en los programas de mejora genética (Sansaloni *et al.*, 2020). Un ejemplo de esto es el genotipado de alto rendimiento, que se plantea como una herramienta valiosa para la identificación de variantes genéticas (SNP y SilicoDArT basados en DArTseq) (Sansaloni *et al.*, 2011), Lo que ha permitido abrir nuevas oportunidades para el desarrollo de variedades de trigo más resilientes y productivas.

Los marcadores se emplean para detectar variaciones genéticas naturales a lo largo del tiempo y eventos históricos de recombinación cromosómica que tienen lugar en múltiples generaciones de poblaciones naturales (Pu *et al.*, 2022). Esos atributos específicos están modulados por diversas regiones genómicas que presentan efectos genéticos de magnitud reducida. La capacidad de los estudios de asociación del genoma completo o GWAS (por sus siglas en inglés, Genome-Wide Association Studies) identifica las correlaciones entre variaciones fenotípicas y cambios en los nucleótidos del ADN. Estos estudios utilizan una muestra diversa de población, la cual resulta fundamental en la comprensión de cómo ciertos rasgos específicos son afectados genéticamente (Pu *et al.*, 2022). Enfocándose en las

asociaciones entre los polimorfismos de un nucleótido simple o SNP (Single-Nucleotide Polymorphisms, en inglés).

Tomado en cuenta lo anterior, se utilizó el análisis de haplotipos, que consiste en la combinación coherente de marcadores de sitios polimórficos relevantes dentro de una región cromosómica específica. Este enfoque para explorar haplotipos permite aprovechar los eventos de recombinación antiguos para identificar los *loci* genéticos de los rasgos subyacentes con una resolución relativamente alta (Basile *et al.*, 2019). De esta manera, se busca identificar las mejores variedades que presenten las secuencias más favorables de estos haplotipos, proporcionando así una base más sólida para la selección de variedades con características agronómicas complejas junto con los genes posiblemente responsables a ellas en el estudio.

2. Antecedentes

2.1. Historia del trigo

El origen del cultivo del trigo es desconocido, se cree que surgió a partir de una planta silvestre en algún lugar del este de Asia, en una zona conocida como la región del creciente fértil, ubicada al norte de Irak y Siria, y al sur de Turquía. Esta región es conocida como la cuna de la civilización debido a sus suelos fértiles y porque existen indicios que muestran que el trigo comenzó a ser utilizado por los primeros asentamientos humanos en las primeras civilizaciones.

Los primeros trigos utilizados por el hombre provenían de una especie silvestre, doble haploide llamada trigos einkorn (*Triticum urartu*). Estos trigos tenían espiguillas delgadas y desarticuladas, con granos encerrados por las brácteas y aristas muy prominentes, se estima que esto ocurrió hace unos 30,000 años antes de Cristo. Con el tiempo, se cruzaron con otra especie de zacate silvestre llamada *Aegilops speltoides*, que contenía el genoma *BB*. El resultado fue un híbrido tetraploide con el genoma *AABB*, conocido como trigos emmer (*Triticum dicoccum*).

Estos fueron los primeros trigos cultivados, conocidos como trigos Emmer cultivados, y su domesticación probablemente ocurrió hace unos 10,000 años atrás. Posteriormente, los trigos emmer evolucionaron y se transformaron en una nueva especie, *Triticum durum* o trigo cristalino, en la que los granos estaban desnudos y las espiguillas eran articuladas. Se conocieron como los primeros trigos duros, que también eran tetraploides con el genoma *AABB*. Se estima que ocurrió aproximadamente hace unos 8,000 años debido a la evolución y selección. En el mismo período, se presume que los trigos emmer se cruzaron nuevamente con otra especie de zacate silvestre diploide llamada *Triticum tauschii*, que tenía el genoma *DD*. Resultando en una nueva especie de trigos hexaploides (*AABBDD*) llamados trigos espelta (*Triticum spelta*) esto ocurrió hace unos 9,000 años. Los trigos espelta evolucionaron nuevamente y dieron lugar a los primeros trigos harineros modernos o *Triticum aestivum*, con granos descubiertos (Miramontes *et al.*, 2023).

2.2. Generalidades del trigo.

La palabra "trigo" tiene su origen en el término latino "*Triticum*", el cual significa 'quebrado', 'triturado' o 'trillado', haciendo referencia a la acción necesaria para separar el grano de trigo de la capa externa que lo rodea. "Trigo" (*Triticum spp*) es el vocablo utilizado para describir el conjunto de cereales pertenecientes al género *Triticum*, ya sean cultivados o de origen silvestre. Este término se emplea tanto para referirse a la planta en sí como a sus semillas comestibles, similar a cómo se utiliza el nombre de otros cereales (Garcia, 2017).

El trigo es una planta que produce un grupo de frutos modificados que se unen a su única semilla en una espiga final y pueden encontrarse tanto en estado silvestre como cultivado. Los egipcios fueron los pioneros en descubrir la fermentación del trigo y utilizarlo en la preparación de alimentos. En la actualidad, el trigo es el segundo cereal más relevante en la dieta de los mexicanos, y en nuestro país se cultivan dos variedades de trigo grano: el cristalino y el panificable (SADER, 2022).

2.3. Taxonomía.

El trigo es clasificado dentro del orden Cyperales, perteneciente a la familia de las Gramíneas (Poaceae), subfamilia Poideae, y a la tribu *Triticeae*. En esta tribu, se reconocen 37 géneros, entre ellos, el género *Triticum*, que incluye las tres especies cultivadas de trigo (Miramontes *et al.*, 2023).

Una de estas especies es el *T. aestivum*, que, como se mencionó previamente, es una especie hexaploide (6X) con 42 cromosomas en sus células somáticas y presenta un genoma *AABBDD*. Esta especie es responsable del origen de los trigos harineros y también incluye la subespecie *T. aestivum* subesp. *compactum* (*T. compactum*), conocida como trigo club en Estados Unidos, que es la segunda especie cultivada de trigo (Miramontes *et al.*, 2023).

2.4. Importancia del trigo en México.

En México, el trigo constituye la tercera fuente de nutrientes de bajo costo en la alimentación de los mexicanos, siguiendo al maíz (*Zea mays spp*) y al frijol (*Phaseolus vulgaris spp*). Esto es especialmente relevante para las comunidades con recursos limitados, tanto en zonas urbanas como rurales. Además, el proceso de cultivo, transformación y consumo del trigo contribuye significativamente a la economía, generando una notable cantidad de empleos en diversos sectores y etapas de la cadena de producción del Sistema Producto Trigo. En el país, se cultiva *T. aestivum*, utilizado en la elaboración de pan, tortillas, galletas y repostería, así como *T. durum*, empleado en la producción de pastas alimenticias como espaguetis y sopas secas, entre otros (Peña *et al.*, 2008).

La producción de trigo tiene una significativa relevancia en el suministro interno del país. La importancia de este cultivo radica en su uso tanto para la alimentación humana como animal, así como en la extensión de terreno destinada a su siembra. Durante el año 2018, se cultivaron aproximadamente 560 mil hectáreas, de las cuales el 83.5% correspondió a terrenos de riego. Entre los estados de mayor importancia en términos de siembra se encuentran, en orden decreciente, Sonora, Guanajuato, Baja California Norte, Sinaloa y Michoacán. Estos estados en conjunto contribuyeron con el 89.1% del suministro a nivel nacional, el cual alcanzó una cifra de 3.3 millones de toneladas (Villaseñor *et al.*, 2020).

2.5. Cálida de trigo.

Hoy en día, definir el concepto de calidad resulta sumamente complejo, debido a la necesidad crucial de precisar las exigencias de la industria de procesamiento y comercialización. La cual busca aumentar la producción a bajo costo y maximizar sus beneficios. El proceso involucra varias etapas: molienda, procesamiento, uso final y calidad nutricional. Es esencial que la calidad de uso final de una variedad de trigo se adapte específicamente a la producción de alimentos que satisfagan las preferencias de los consumidores (Guzmán *et al.*, 2022a).

Los componentes del grano de trigo tiene una función crucial, junto con sus propiedades fisicoquímicas, para definir la calidad del grano de trigo. Entre los principales componentes del grano, se destacan: proteínas (7-18%), lípidos (1.5-2%) y carbohidratos (60-75%). Siendo la proteína, la más importante, influyendo de manera notable en tres características del grano; las propiedades viscoelásticas del gluten, algunas propiedades del almidón (gelatinización, empastado y gelificación) y la dureza o textura del grano. Propiedades directamente relacionadas con la molienda, la calidad del gluten en la harina, el procesamiento y la calidad final del producto (Guzmán *et al.*, 2022a).

En este sentido, el contenido de proteína en grano (GP) y el ensayo de sedimentación con dodecil sulfato de sodio (SDS_Sedim) son dos de los parámetros más utilizados para evaluar la calidad del trigo harinero. El GP está estrechamente relacionado con la cantidad de gluten en la harina, mientras que SDS_Sedim permite medir la calidad funcional de las proteínas del grano, en especial su capacidad para formar redes viscoelásticas en la masa (Wang *et al.*, 2015). Ambos parámetros han sido ampliamente utilizados en programas de mejoramiento genético para seleccionar variedades con características óptimas para la industria agroalimentaria.

Dado que las proteínas de almacenamiento del trigo, como las gliadinas y gluteninas, son las principales responsables de la calidad panadera, es fundamental considerar su relación con la composición genética del grano. Estas proteínas, que representan aproximadamente el 80% del gluten, están codificadas por genes ubicados en los cromosomas 1 y 6 (A, B y D). Las gluteninas de alto (HMWG) y bajo peso molecular (LMWG), junto con las gliadinas ω/γ , se encuentran en los cromosomas 1A, 1B y 1D, mientras que las gliadinas $\alpha/\beta/\gamma$ están presentes en los cromosomas 6A, 6B y 6D (Huo *et al.*, 2019; Sansaloni *et al.*, 2020; Guzmán *et al.*, 2022a).

Es fundamental que el trigo exhiba las características deseadas por los consumidores, todo ello mientras se minimizan los costos y se maximizan los beneficios, garantizando una producción exitosa después de la cosecha (Guzmán *et al.*,2022a).

2.6. Mejoramiento tradicional del trigo.

El mejoramiento tradicional del trigo basado en características físicas o fenotipos, implica seleccionar y cruzar plantas que presentan características deseadas observables, tales como: la altura de la planta, el tamaño y la calidad de los granos, la resistencia a enfermedades, y la adaptación a condiciones ambientales específicas. De esta forma, el mejoramiento tradicional se fundamenta en la observación de las características físicas de las plantas y en la selección de aquellas que muestran los rasgos más favorables para la producción y calidad del trigo (Khush, 2001).

El proceso comienza con la evaluación de las variedades de trigo disponibles, identificando las que exhiben fenotipos superiores en términos de rendimiento, resistencia a plagas y enfermedades, y adaptabilidad climática. Posteriormente, se realizan cruces entre estas variedades para combinar las características deseadas. Las plantas resultantes se evalúan a lo largo de varias generaciones para identificar aquellas que mantengan y mejoren los rasgos buscados (Gale & Youssefian, 1985).

A lo largo del tiempo, se realizan ciclos repetidos de selección y cruzamiento, con el objetivo de desarrollar nuevas variedades que no solo posean un alto rendimiento, sino también una mayor resistencia y calidad. Aunque este método es efectivo, es un proceso largo y dependiente de las condiciones ambientales, ya que los fenotipos están influenciados tanto por la genética de la planta como por el entorno en el que crecen (Evans, 1997).

2.7. Mejora genética en la calidad del trigo mediante marcadores moleculares.

Debido al incremento de la población, las necesidades sociales exigen una demanda en la mejora de la calidad del trigo, tanto para su uso en panificación como en otros productos derivados. Mientras que, la capacidad de identificar y seleccionar variedades con características superiores ha sido un desafío constante. Aunque, la selección tradicional basada en fenotipos observables ha sido fundamental, el mejoramiento de semillas se vio enormemente favorecido con las nuevas tecnologías, generadoras de datos a gran escala. En este sentido, el uso de marcadores moleculares denominado selección asistida por marcadores (MAS, del inglés Marker-Assisted Selection) ha mejorado drásticamente la eficiencia de estos procesos (Hassaneian & Abdi, 2023). Estos marcadores moleculares permiten asociar características de calidad del trigo, como el contenido de proteínas y la elasticidad del gluten, con regiones específicas del genoma (Khoshro *et al.*, 2022).

Gracias a estos avances, los mejoradores han acelerado la identificación de líneas de trigo con mejor calidad, lo que no solo mejora el rendimiento industrial del grano, sino también su valor nutritivo (Chegdali *et al.*, 2020). Por tanto, los marcadores moleculares juegan un papel crucial en la optimización de las cualidades de trigo, haciendo que el proceso de selección genética sea más preciso y eficaz.

2.7.1. Marcadores moleculares.

Los marcadores moleculares son biomoléculas esenciales para identificar rasgos genéticos específicos en diversos organismos. En el caso de la agricultura, su importancia se destaca especialmente en la mejora genética de plantas. Estos marcadores incluyen proteínas, como antígenos e isoenzimas, y fragmentos de ADN que se localizan en posiciones específicas en los cromosomas, conocidos como *loci*. La cercanía de estos segmentos en un cromosoma favorece su herencia conjunta a lo largo de las generaciones, facilitando así la asociación de ciertas características fenotípicas con genes específicos (Amiteye, 2021). Entre estos marcadores, el polimorfismo de nucleótido simple, SNP, y la variación en la

presencia-ausencia (PAV, por sus siglas en inglés) han transformado la genética molecular. Particularmente son valorados por su facilidad y costo en la detección, y no están sujetos a influencias ambientales ni a otros factores genéticos. Convirtiéndolos en herramientas poderosas para los investigadores y mejoradores de cultivos, facilitando la identificación de rasgos favorables y mejorando así la calidad y el rendimiento de las plantas cultivadas (Amiteye, 2021).

2.7.2. Polimorfismos de Nucleótido Único (SNP).

Los SNP, corresponden a cambios de una sola base en la secuencia de ADN y pueden manifestarse tanto dentro como fuera de los genes. Genes que contienen uno o varios SNP's pueden contribuir a dos o más variantes alélicas de ARNm. Las alteraciones en el ARNm pueden tener diversas funciones biológicas, ya que resultan en cambios en las estructuras primarias o superiores que interactúan con otros componentes celulares. Hasta ahora, en varios estudios se ha empleado un genoma de referencia único para analizar los SNP's entre diversas entidades. Sin embargo, la comprensión de los SNP's se enriquecería significativamente mediante un pangenoma, ya que proporcionaría información del genoma completo de cualquier especie (Amir, 2020).

En este sentido, en 2020, Chu y colaboradores evaluaron la idoneidad de los arreglos de SNP's, comparándola con la genotipificación por secuenciación para la genómica de bancos de germoplasma en trigo (Chu *et al.*, 2020). Ambos enfoques ofrecen ventajas particulares dependiendo del contexto de investigación (Chu *et al.*, 2020).

Posteriormente, Irshad y colaboradores identificaron un SNP en el gen TaSBEIII y desarrollaron un marcador KASP asociado con el peso del grano en trigo, lo cual representa un avance significativo en la mejora genética del cultivo (Irshad *et al.*, 2021).

2.7.3. Polimorfismos de presencia-ausencia (PAV).

Los polimorfismos denominados PAVs, son variaciones estructurales del genoma que se caracterizan por la existencia de secciones del ADN en ciertos individuos de una especie, mientras que están ausentes en otros. Este tipo de diferencias genómicas entre individuos dentro de cepas cultivadas, o domesticadas, se denomina comúnmente "variación de presencia-ausencia". Cuando una región PAV incluye un gen, esta variación puede alterar su función, ya que algunos organismos carecen de la porción del genoma que alberga dicho gen. Estudios a nivel de todo el genoma han documentado este fenómeno en diversas especies de plantas cultivadas (Takahashi-Kariyazono *et al.*, 2020).

2.8. Secuenciación de Nueva Generación (NGS).

La secuenciación de nueva generación (NGS, por sus siglas en inglés) es una tecnología de secuenciación que opera de manera altamente paralela, brindando un rendimiento, escalabilidad y velocidad extremadamente elevados. Este enfoque se emplea para determinar la secuencia de nucleótidos en genomas completos o en regiones específicas de ADN o ARN. La NGS ha revolucionado los campos de las ciencias biológicas, permitiendo a los laboratorios llevar a cabo una amplia gama de aplicaciones y analizar sistemas biológicos en niveles previamente inalcanzables.

Las actuales y complejas interrogantes relacionadas con la genómica demandan una profundidad de información que excede las capacidades de las tecnologías tradicionales de secuenciación de ADN (Illumina, 2024).

Actualmente, la tecnología NGS ha transformado profundamente la naturaleza de las preguntas que los científicos pueden formular y responder. Las avanzadas opciones de preparación de muestras y análisis de datos han abierto el camino para una amplia gama de aplicaciones

En esencia, NGS ha ampliado significativamente las capacidades de investigación en una variedad de campos científicos (Illumina, 2024).

La tecnología de secuenciación de Illumina NGS utiliza un enfoque completamente distinto al método tradicional de terminación de cadena de Sanger. En lugar de eso, emplea la tecnología de secuenciación por síntesis, la cual sigue de cerca la incorporación de nucleótidos marcados durante la replicación de la cadena de ADN, de manera altamente paralela. Este tipo de secuenciación genera volúmenes masivos de datos y en comparación con la secuenciación tradicional de Sanger, es más económica y requiere menos tiempo. Los sistemas de secuenciación de Illumina pueden producir resultados que varían desde 300 kilobases hasta múltiples terabases en una sola ejecución, dependiendo del tipo de instrumento y su configuración específica (Illumina, 2024).

Rauf, y colaboradores, en 2020 señalaron que el trigo panadero tiene un genoma extremadamente grande, lo que dificulta su secuenciación con métodos convencionales (Rauf *et al.*, 2020). Sin embargo, destacaron que NGS ha permitido secuenciar rápidamente este genoma complejo, lo que facilita avances en estudios genómicos, como asociaciones genómicas y genética de poblaciones.

2.9. Tecnología DArTseq.

La tecnología DArTseq (del inglés Diversity Arrays Technology sequense) ha implementado la secuenciación de las representaciones genómicas en plataformas NGS. Esta tecnología tiene la capacidad de detectar dos tipos de polimorfismos: SilicoDArT (PAV's) y SNP's DArTseq. Y representa una plataforma de genotipado por secuenciación, altamente eficiente y permite descubrir decenas o incluso centenas de miles de polimorfismos (marcadores moleculares) de alta calidad en todo el genoma (Figura 1). El proceso comienza con la digestión del ADN utilizando enzimas de restricción específicas que fragmentan el ADN (Figura 1a), se realiza la amplificación de los fragmentos de interés mediante PCR para multiplicar selectivamente estas regiones (Figura 1b), los fragmentos amplificados se mezclan y cuantifican para la creación de una librería genética (Figura 1c), que posteriormente se secuencia utilizando tecnología NGS (Figura 1d), las secuencias

obtenidas se alinean contra un genoma de referencia para identificar variaciones genéticas (Figura 1e) y finalmente, se extraen datos sobre marcadores SNP's y PAVs a partir de los datos genómicos (Figura 1f). La generación de una gran cantidad de datos a través de esta tecnología es de gran importancia, ya que contribuye a la identificación de la diversidad genética en un cultivo y al entendimiento del control genético de las características evaluadas en una planta o cultivo de interés (Sansaloni *et al.*, 2011).

En 2020, un estudio realizado por Mahboubi y sus colaboradores mostraron la diversidad genética del genoma completo en genotipos de trigo de todo el mundo (Mahboubi *et al.*, 2020). En su análisis, utilizaron la plataforma DArTseq para evaluar dos tipos de marcadores: SNP's y SilicoDArT. Este enfoque permitió una evaluación exhaustiva de las variaciones genéticas y ayudó a delinear la estructura poblacional y el desequilibrio de ligamiento en las diferentes poblaciones de trigo. La combinación de estos métodos proporcionó una comprensión más profunda de la diversidad genética en trigo a nivel global.

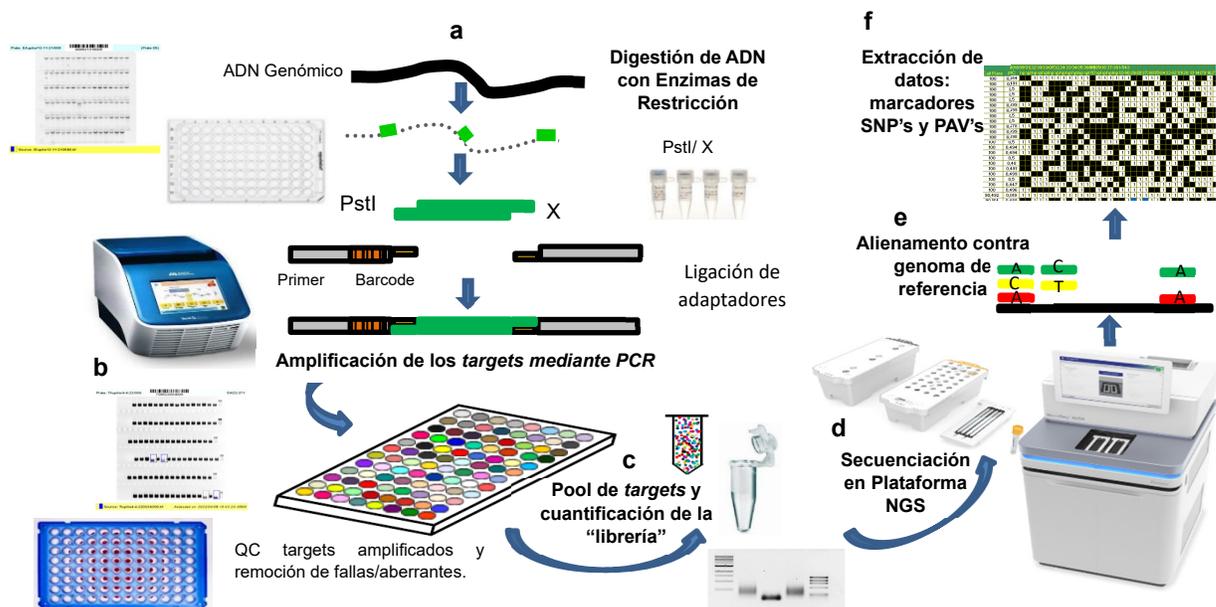


Figura 1. Genotipificación por Secuenciación: producción de datos primarios (perfiles genómicos de marcadores moleculares - DArT-Seq) a partir de muestras de ADN; **a)** Digestión de ADN con enzimas de restricción; **b)** Amplificación de los targets mediante PCR; **c)** Pool de targets y cuantificación de la librería; **d)** Secuenciación en plataforma NGS; **e)** Alineamiento contra genoma de referencia; **f)** Extracción de datos marcadores SNPs y PAVs (Tomada de Valdez, 2021).

2.10. Estudios de Asociación del Genoma Completo (GWAS).

Los GWAS (del inglés, Genome-Wide Association Studies) son ampliamente utilizados para descubrir relaciones entre variaciones comunes en la secuencia de ADN (Figura 2). Por lo general, los estudios involucran a miles de participantes y se enfocan en capturar la diversidad genética mediante SNP's en cada uno de los cromosomas (Figura 2c). Cada uno de estos SNP's, que pueden ser numerosos, se somete a análisis estadísticos para detectar posibles conexiones con un rasgo específico que se está investigando. Es importante tener en cuenta que los datos de GWAS deben ser interpretados considerando la probabilidad de obtener resultados falsos positivos debido al gran número de pruebas estadísticas realizadas. Además, se deben realizar ajustes para abordar posibles diferencias sistemáticas, especialmente en miles de factores genéticos, que pueden estar relacionadas con las diferencias en la ascendencia genética de los participantes en el estudio. Mediante el uso de técnicas adicionales, es posible combinar los resultados de múltiples GWAS con el fin de mejorar la capacidad para identificar SNP's con efectos muy pequeños en el rasgo en cuestión (Figura 2b-e). Aunque los GWAS no han revolucionado significativamente la capacidad de predecir rasgos comunes de enfermedades, sí han permitido descubrir miles de nuevas asociaciones entre las variaciones en la secuencia de ADN y los rasgos que influyen (Bush, 2019).

La importancia de este tipo de estudios ha permitido el análisis de GWAS de trigo, para evaluar germoplasma de pre-mejoramiento bajo estrés por sequía terminal (Shokat *et al.*, 2020). En este trabajo, analizaron 339 líneas en condiciones de riego y sequía, y se encontró que la sequía redujo significativamente el rendimiento del grano y otros rasgos importantes. El GWAS identificó asociaciones significativas en los cromosomas 4A, 2D, 3B, 6B y 3A para rasgos como el rendimiento, la longitud de la espiga y el número de granos por espiga. También identificaron veintiséis marcadores (SNP's) asociados con estos rasgos, los cuales podrían ser útiles para la selección asistida por marcadores en programas de mejoramiento de trigo (Shokat *et al.*, 2020).

En 2021, Gómez-Espejo y colaboradores analizaron la conexión entre los marcadores SNP y los patrones de crecimiento en las diferentes colecciones de *T. aestivum* (Gómez-Espejo *et al.*, 2021). Con la finalidad de identificar regiones que potencialmente guardan una relación funcional con esos patrones, emplearon un enfoque basado en GWAS.

Leonova y colaboradores, en 2023 utilizaron un análisis de GWAS para identificar *loci* genéticos asociados con el contenido de gluten en trigo panificable (Leonova *et al.*, 2023). El estudio reveló 17 loci significativos en varios cromosomas que explican hasta el 21% de la variación fenotípica. Encontraron que el haplotipo TT/AA/GG en el cromosoma 2A estaba asociado con un mayor contenido de gluten en comparación con otros haplotipos. Estos resultados sugieren que los *loci* identificados podrían ser útiles para mejorar la calidad del trigo panificable (Leonova *et al.*, 2023).

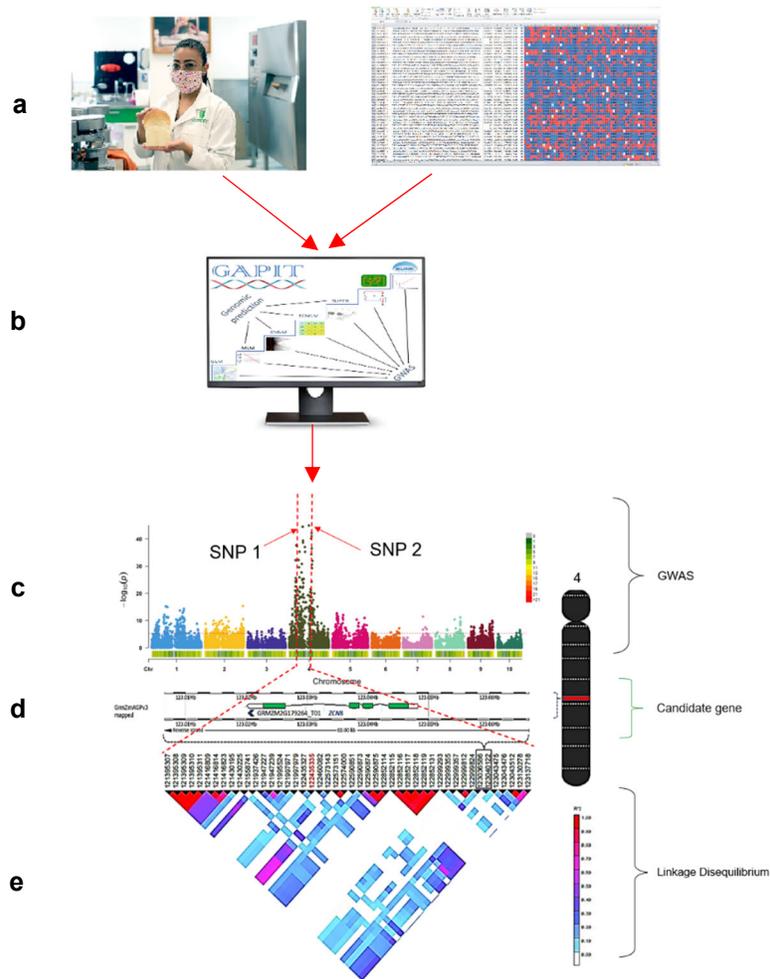


Figura 2. Proceso de análisis genómico utilizando GWAS; **a)** Recolección de los datos de la caracterización fisicoquímica y genotípica de las muestras (Rico & Dahm, 2022); **b)** Procesamiento de los datos mediante el software GAPIT para realizar estudios de asociación genómica (Lipka *et al.*, 2012); **c)** Identificación de SNPs significativos mediante un gráfico Manhattan, resaltando las variantes asociadas a los rasgos de interés, en cada cromosoma; **d)** Localización y análisis de genes candidatos cercanos a los SNPs identificados, en cada cromosoma; **e)** Evaluación del desequilibrio de ligamiento (LD) en la región genómica asociada para determinar la relación genética entre los SNPs detectados (Karn, 2019).

2.10.1. Modelos estadísticos para GWAS.

Los modelos estadísticos para GWAS son enfoques matemáticos utilizados para identificar asociaciones entre variantes genéticas y rasgos o enfermedades en un conjunto de datos genómicos. Estos modelos son esenciales para realizar análisis genéticos y comprender cómo las variaciones en el genoma pueden estar relacionadas con fenotipos específicos (Gorostiza, 2017). Algunos de los modelos

estadísticos comunes utilizados en GWAS son los descritos en las secciones siguientes.

2.10.1.1. Modelo Lineal Mixto (MLM).

El Modelo Lineal Mixto (MLM) abarca efectos tanto fijos como aleatorios. La inclusión de individuos como efectos aleatorios otorga al MLM la capacidad de incorporar información relativa a las relaciones existentes entre los individuos. Esta información acerca de las relaciones se transmite mediante la matriz de parentesco (K), la cual se emplea en el MLM como matriz que describe la varianza-covarianza entre los individuos. Cuando se utiliza una matriz de parentesco basada en marcadores genéticos (K) en combinación con la estructura de la población (a menudo denominada matriz "Q", que se puede obtener mediante métodos como STRUCTURE o un análisis de componentes principales), el enfoque "Q K" mejora la capacidad de detección estadística en comparación con el uso exclusivo de "Q" (Yu *et al.*, 2006).

El modelo busca representar la distribución de la varianza genética estimada para descubrimientos previamente conocidos en relación con un rasgo. Se hace mediante una mezcla de distribuciones exponenciales, similar a una mezcla de distribuciones normales escaladas para los coeficientes de regresión. Los coeficientes de regresión, por su parte, reflejan el grado de asociación entre la variante genética y el rasgo, teniendo en cuenta estas variables controladas (Holland *et al.*, 2016).

2.10.1.2. Unificación de Probabilidad Circulante de modelo fijo y aleatorio (FarmCPU).

FarmCPU, (del inglés, Fixed and random model Circulating Probability Unification) es un método ampliamente utilizado en el estudio GWAS. Lo que distingue a FarmCPU es su capacidad para combinar las ventajas del MLM y la regresión por pasos, representando un modelo de efectos fijos, al abordar sus desventajas de manera iterativa. En lugar de emplear el parentesco en un MLM y para eliminar cualquier posible confusión entre el parentesco y los genes

relacionados con un rasgo de interés, FarmCPU sustituye el parentesco con un conjunto de marcadores asociados a los genes causales. Estos marcadores asociados se incorporan como efectos fijos en un modelo de efectos fijos para evaluar cada marcador de forma individual en todo el genoma. Para evitar el ajuste excesivo del modelo al evaluar los marcadores, el conjunto de marcadores asociados se optimiza utilizando un enfoque de máxima verosimilitud en un MLM que considera la estructura de varianza y covarianza definida por los marcadores asociados (Liu *et al.*,2016).

2.10.1.3. Clave Anidada Iterativa de Desequilibrio de Vinculación y Evaluación Bayesiana (BLINK).

El método BLINK fue diseñado para tener un alto poder estadístico y eficiencia computacional. Se inspiró en el método FarmCPU con dos cambios importantes para lograr los objetivos. Una es eliminar la suposición de que los genes causales están distribuidos uniformemente en todo el genoma como lo requiere FarmCPU. Como la suposición causa la inclusión de genes no causales o la falta de genes causales que están en el mismo contenedor con otros genes causales con una señal más fuerte. BLINK funciona directamente en marcadores en lugar de contenedores. Se excluyen los marcadores que están en Desequilibrio de Ligamiento (LD, por sus siglas en inglés) con el marcador más significativo. Para el segundo marcador restante, la exclusión se realiza de la misma manera que para el marcador más significativo, y así sucesivamente hasta que no se pueda excluir ningún marcador. El otro cambio es utilizar el Contenido de Información Bayesiano (BIC, por sus siglas en inglés) de un modelo de efectos fijos para aproximar la probabilidad máxima de un modelo de efectos aleatorios para seleccionar los marcadores asociados entre los marcadores que siguen siendo la exclusión basada en LD. Tanto los modelos de prueba de marcadores como los de selección de marcadores asociados como cofactores son modelos de efectos fijos, la complejidad del cálculo alcanza el máximo (Huang *et al.*, 2019).

2.10.2. Gráficos y medidas estadísticas para el análisis de GWAS.

En un GWAS, se emplean diversos métodos estadísticos para examinar los resultados y determinar la relación entre las variaciones genéticas y los rasgos o enfermedades bajo investigación. A continuación, se describen algunas de las medidas estadísticas más utilizadas en los GWAS.

2.10.2.1. Valor p (p -value).

En el análisis estadístico, el valor p (p -value) representa la probabilidad de que la relación entre dos eventos sea resultado del azar. Un valor de p bajo indica una probabilidad reducida de que la asociación entre dos elementos sea un producto del azar, lo que sugiere una correlación o conexión entre ellos. En estudios GWAS, el umbral del valor p puede variar dependiendo del análisis, pero comúnmente se encuentra por debajo de 0,00001 (1×10^{-5}) e incluso puede ser inferior a 0,00000005 (5×10^{-8}) (Ellinghaus *et al.*, 2020).

2.10.2.2. Razón de probabilidad (OR).

La razón de probabilidad (OR, por sus siglas en inglés: *odds ratio*) es una medida de efecto ampliamente empleada para transmitir los resultados de una investigación. Matemáticamente, el OR representa la relación entre dos *odds*, siendo el *odds* una manera alternativa de expresar la probabilidad de que ocurra un evento de interés o la presencia de una exposición. Por ejemplo, para el mejoramiento del trigo, los OR se usan para interpretar la magnitud del efecto de ciertos marcadores sobre rasgos cuantitativos, ayudando a identificar genes clave (Yoosefzadeh-Najafabadi *et al.*, 2022). Los OR pueden calcularse en diseños prospectivos, retrospectivos y transversales, y en determinadas condiciones pueden sustituir al Riesgo Relativo (Cerdeira *et al.*, 2013)

2.10.2.3. Gráficos de Manhattan.

Los gráficos de Manhattan son una herramienta visual fundamental en los estudios de GWAS, ya que permiten identificar las asociaciones entre variantes genéticas y rasgos fenotípicos específicos, como el rendimiento, la resistencia a enfermedades o la tolerancia a factores de estrés abiótico, como la sequía o el calor.

Estos gráficos organizan las posiciones cromosómicas en el eje “x” y muestran los valores de p ($-\log_{10}$) en el eje “y”, lo que permite observar qué marcadores genéticos tienen asociaciones estadísticamente significativas con los rasgos analizados. La disposición de los picos en el gráfico refleja la estructura de asociación genética en el genoma del trigo, donde las regiones con variantes genéticas correlacionadas, debido a la baja frecuencia de recombinación genética, se agrupan, creando elevaciones en el gráfico (Rahimi *et al.*, 2023).

2.10.2.4. Gráficos Q-Q.

Los gráficos Q-Q (Quantile-Quantile), son fundamentales para evaluar la distribución de los valores p observados frente a la distribución esperada bajo la hipótesis nula (es decir, la ausencia de asociación genética). En estos gráficos, si los puntos se alinean cerca de la línea diagonal, se interpreta como que no hay desviaciones significativas de la hipótesis nula, lo que sugiere que no existen excesos de asociaciones falsas positivas. Por otro lado, si se observan desviaciones notables, especialmente en las colas superiores del gráfico (donde se encuentran las asociaciones más significativas), esto podría indicar la presencia de *loci* genuinamente asociados con el rasgo en estudio. Los gráficos Q-Q son, por lo tanto, herramientas útiles para detectar asociaciones genéticas significativas entre marcadores moleculares (SNP's) y características fenotípicas específicas, como la resistencia a enfermedades, el rendimiento y la tolerancia a factores de estrés abiótico (Choudhury *et al.*, 2019; Hu *et al.*, 2021).

2.10.3. Desequilibrio de ligamiento (LD).

El desequilibrio de ligamiento (LD, por sus siglas en inglés) es fundamental para entender la estructura genética y la relación entre marcadores genéticos y características fenotípicas. El LD se refiere a la asociación no aleatoria de alelos en diferentes *loci* dentro de un genoma. En el contexto del mejoramiento genético del trigo, el análisis de LD permite identificar regiones del genoma que están vinculadas a rasgos específicos, como tolerancia a sequía, resistencia a enfermedades o productividad de grano (Laido *et al.*, 2014).

2.10.4. Haplotipos.

Se describe un haplotipo como "dos o más alelos de SNP que tienden a heredarse como una unidad" (Bernardo, 2010). Sin embargo, la construcción de haplotipos basada en datos empíricos de marcadores no es sencilla (Korte & Kurt, 2013). Se describen métodos fundamentales para definir haplotipos, que incluyen (i) utilizar la diversidad de haplotipos de un segmento cromosómico específico como base para la asignación de haplotipos o (ii) emplear enfoques que asignen haplotipos basándose en el LD entre marcadores que muestran poca o ninguna evidencia de recombinación y que comparten herencia conjunta en el mismo bloque cromosómico, normalmente medido (Qian *et al.*, 2017).

Los enfoques basados en el desequilibrio de ligamiento presentan ventajas porque (i) se centran directamente en la detección de recombinación en una población dada al identificar bloques de haplotipos, (ii) son fácilmente aplicables a datos diploides donde se desconoce la fase del haplotipo, y (iii) los coeficientes de LD son fáciles de visualizar. Existen múltiples factores que afectan el LD, como el tipo de propagación y la tasa de endogamia en una especie, el tamaño de la población analizada, el grado de estratificación y subdivisión de la población, la deriva genética, la frecuencia de recombinación, la tasa de mutación o la fuerza y tipo de selección en un fragmento cromosómico específico (Gupta *et al.*, 2005).

De acuerdo a la importancia de este tema, se han publicado estudios para generar haplotipos asociados a rasgos de interés agronómico o industrial. En un estudio de esta índole, analizaron bloques de haplotipos que revelaron la presencia de 1268 bloques de diversos tamaños distribuidos a lo largo del genoma, incluyendo regiones altamente conservadas, como el brazo del cromosoma 1BS, que alberga la translocación trigo/centeno 1BL/1RS (Luján *et al.*, 2019).

3. Justificación

El trigo es un cereal esencial por su alto valor nutricional, siendo crucial mejorar su calidad y productividad para satisfacer las necesidades alimentarias mundiales.

Por ello, este estudio propone utilizar tecnologías avanzadas como la genotipificación para identificar marcadores genéticos relacionados con la calidad del gluten en accesiones nativas de *T. aestivum*. Los resultados permitirán seleccionar e incorporar estas accesiones en programas de mejoramiento genético, contribuyendo al desarrollo de líneas élite.

4. Hipótesis

Con la metodología aplicada se podrá identificar SNP's y haplotipos asociados con la calidad del grano y del gluten en trigos harineros nativos, favoreciendo la selección de líneas élite para su aplicación en mejoramiento genético.

5. Objetivos

5.1. Objetivo general.

Identificar y reportar marcadores moleculares que mejoren la calidad de los granos de *T. aestivum* mediante la aplicación de GWAS, proporcionando información útil para la selección en programas de mejoramiento genético.

5.2. Objetivos específicos.

- Comprobar la calidad de la base de datos fenotípica y genotípica de las accesiones de trigos harineros nativos para garantizar la confiabilidad de los resultados en los análisis de asociación genómica.
- Realizar un análisis de asociación del genoma completo (GWAS) para detectar regiones asociadas con la calidad del grano en trigo harinero.
- Identificar y reportar haplotipos y SNP's significativos que puedan ser utilizados como marcadores moleculares en programas de mejoramiento genético.
- Explorar la posible función de las regiones genómicas detectadas mediante la búsqueda de genes en bases de datos y la comparación con estudios previos.

6. Metodología

Se utilizaron 1602 accesiones, de *T. aestivum* variedad nativa, del banco de germoplasma de trigo, del Centro Internacional del Mejoramiento de Maíz y Trigo (CIMMYT) ubicado en carretera México-Veracruz KM.45, el Batán, Texcoco México. Proveniente de los siguientes países: Afganistán (AFG), Suiza (CHE), China (CHN), India (IND), Irán (IRN), México (MEX), Nepal (NPL), Pakistán (PAK), Portugal (PRT) y Rusia (RUS). Todas las variedades nativas se genotiparon mediante la tecnología DArTseq en la plataforma de genotipificación de alto rendimiento del Servicio de Análisis Genético para la Agricultura (SAGA). Y se realizó un análisis físico y químico de la calidad de los granos de trigo en el Laboratorio de Química, Bioquímica y Calidad Industrial del Trigo. Ambos laboratorios situados en CIMMYT.

6.1. Caracterización genética.

Como resultado de la caracterización mediante la tecnología de DArTseq se obtuvieron un total de 241,119 marcadores de tipo SNP's. A estos marcadores se les realizaron diferentes filtros, con la finalidad de conservar marcadores con un alto contenido de información polimórfica. El primer filtro es la cantidad de datos obtenidos de buena calidad, el segundo fue reproducibilidad (representa la proporción de datos que coinciden entre la muestra original y su réplica) y el último fue heterocigosidad (mide el porcentaje de marcadores que son heterocigotos) (Tabla 1).

Tabla 1. Número de marcadores eliminados en cada filtro.

Marcadores iniciales.	241,119
Por cantidad de datos obtenidos de buena calidad (<0.4).	60,034
Por Reproducibilidad (<0.9).	6,857
Por Heterocigosidad (>0.2).	3,639
Marcadores totales eliminados	70,530
Marcadores finales.	170,589

Al finalizar la serie de filtros aplicados, se obtuvieron 170,589 marcadores finales, garantizando que estos marcadores eran los más informativos. La posición física de cada marcador se determinó utilizando información obtenida a través de *BLAST* (del inglés, Basic Local Alignment Search Tool). Conservando únicamente aquellos que podían ser ubicados de forma clara y sin ambigüedad en el genoma del trigo, alrededor del 45.4%.

Teniendo así una matriz de datos, con un conjunto de 1,602 genotipos caracterizadas por 77,395 SNP's representándolos en un mapa de haplotipos o *HapMap* (abreviación para "Haplotype Map"). Estos SNP's tienen una cobertura exhaustiva en los 21 cromosomas de *T. aestivum*.

El genoma de referencia utilizado para alinear los marcadores fue el IWGSC RefSeq v2.1 (Wheat Chinese Spring v2.1) (Zhu *et al.*, 2021).

6.2. Análisis físico y químico.

Para el análisis físico y químico de los granos de trigos nativos de las 1,602 accesiones, se consideró dos parámetros importantes para el estudio. En el primero, se evaluó el volumen de sedimentación (SDS_Sedim) utilizando el método de Dick y Quick, modificado por Peña y colaboradores en 1990. En este método se utiliza 1 g de harina y se basa en el uso de dodecil sulfato de sodio para evaluar la capacidad de hidratación y expansión de las proteínas del gluten en un entorno ligeramente ácido, lo que permite medir la fuerza general de la masa (Peña *et al.*, 1990). Un mayor volumen de sedimentación indica una mayor fuerza en la masa. El segundo parámetro evaluado, considera el porcentaje de proteína en el grano y la harina (GP), el cual fue determinado utilizando el analizador NIR Infralyzer 6500 (método 39-10, AACC, 1995).

6.3. Análisis de Asociación del Genoma Completo (GWAS).

El análisis de GWAS se llevó a cabo ajustando tres modelos estadísticos diferentes, con el propósito de minimizar la cantidad de falos positivos: El modelo

MLM que combina un modelo de estructura poblacional (Q) y una matriz de parentesco (K) (Yu *et al.*, 2006), el modelo FarmCPU (Liu *et al.*, 2016), y el modelo BLINK (Huang *et al.*, 2019). A continuación, se proporciona una descripción general de cada uno de los modelos mencionados.

Modelo MLM:

$$Y = SNP + Q [PCs] + Kinship + \varepsilon \quad (1)$$

En la fórmula 1, Y corresponde a una matriz que contiene la información fenotípica, SNP representa la matriz de marcadores genéticos, Q denota la estructura poblacional, y $Kinship$ es la matriz que describe las relaciones entre los individuos incluidos en el modelo. Es importante destacar que, en este contexto, tanto SNP como Q se consideran efectos fijos, mientras que la matriz de parentesco ($Kinship$) se incorpora como un efecto aleatorio en el modelo (Yu, 2006).

El modelo anterior se basaba en un enfoque de *locus* único, que no es óptimo para abordar rasgos complejos. Por lo tanto, optamos por utilizar FarmCPU y BLINK para realizar una comparación.

El funcionamiento de FarmCPU se basa en la combinación de modelos de efectos fijos y aleatorios. En el modelo de efectos fijos, se prueba un solo marcador a la vez, utilizando múltiples marcadores asociados como covariables para controlar los falsos positivos. Por otro lado, el modelo de efectos aleatorios se utiliza para evitar el sobreajuste del modelo al emplear múltiples marcadores asociados para calcular el parentesco. En cada iteración, se unifican los valores p asociados a cada marcador probado y asociado. El modelo FarmCPU se puede expresar de la siguiente manera:

$$Y = \underbrace{SNP + QTN1 + QTN2 + \dots + QTNn}_{\text{efectos fijos}} + \underbrace{Q [PCs] + Kinship}_{\text{efectos aleatorios}} + \varepsilon \quad (2)$$

En los modelos de efectos aleatorios (Fórmula 2), los QTN se escogen en base a su distribución uniforme en el genoma. Para mejorar la eficiencia

computacional, se sustituyó el modelo de efectos aleatorios por uno de efectos fijos utilizando criterios de información bayesianos. Este innovador método se denomina BLINK, ya que los QTN son seleccionados tomando en cuenta la información de LD (Huang *et al.*, 2019).

6.4. Haplotipos.

Los segmentos cromosómicos con haplotipos conservados, por lo general, abarcan varios genes y es probable que su actividad esté coordinada por diversos alelos reguladores. Lo cual, conduce a correlaciones entre características que probablemente se deben a la presencia de genes vinculados. Si en la población, los alelos de dos o más *loci* no están distribuidos al azar, se infiere que estos alelos están en LD.

En la fórmula 3 el alelo A del *locus* 1 tiene una frecuencia (P_A) y el alelo B del locus 2 tiene una frecuencia (P_B), en caso de que ambos alelos fueran independientes, se anticiparía que la frecuencia del haplotipo AB (P_{AB}) sería equivalente al producto de las frecuencias de los alelos ($P_{AB} = P_A * P_B$). Cualquier desviación significativa de la población en la frecuencia del haplotipo AB con respecto al producto $P_A * P_B$ indicaría la presencia de LD (Fórmula 3).

Si $P_{AB} \neq P_A * P_B$ confirma la existencia de LD, entonces, al evaluar las frecuencias de los cuatro posibles haplotipos ($P_{AB}, P_{A'B}, P_{AB'}, P_{A'B'}$), estas deberían ser iguales al producto de las frecuencias de los alelos que intervienen en los haplotipos. Se logra al sumar o restar a este producto de frecuencias el parámetro de LD denominado D .

$$\begin{aligned}
 P_{AB} &= P_A * P_B + D \\
 P_{A'B} &= P_{A'} * P_B - D \\
 P_{AB'} &= P_A * P_{B'} - D \\
 P_{A'B'} &= P_{A'} * P_{B'} + D
 \end{aligned}
 \tag{3}$$

En la fórmula 4, D se infiere de la siguiente manera:

$$D = P_{AB} * P_{A'B'} - P_{A'B} * P_{AB'} \quad (4)$$

El cual permite evaluar el nivel de LD entre dos *loci* bialélicos, mejor conocido como *desequilibrio gamético* de Lewontin (Fórmula 5), en el cual:

$$D' = \frac{D}{D_{max}} \quad (5)$$

Donde D representa el parámetro de LD estimado mediante la ecuación 4, y D_{max} es el menor de los productos entre $P_A * P_{B'}$ y $P_{B'} * P_A$ (Lewontin, 1988).

La evaluación del nivel de LD se realiza mediante una prueba de hipótesis de independencia entre los alelos de los marcadores, considerando que la ausencia de LD se interpreta como la independencia. El estadístico de prueba más común utilizado es el chi cuadrado (Weir, 1996).

Previo al análisis de haplotipos, se estableció empíricamente considerar una ventana de +-15 centimorgans (cM) tomando los marcadores aledaños a cada marcador significativo en el estudio, conforme a lo anterior, se generaron haplotipos para cada uno de ellos, respectivamente. El análisis se llevó a cabo utilizando la herramienta de análisis de datos conocida como PLINK v1.9, configurando al Software para formar bloques de haplotipos de alelos de mínimo 2 hasta máximo 10 (Chang, 2015). Posteriormente, se aplicó un filtro en donde se decidió seleccionar los haplotipos más frecuentes, arriba del 0.6 como un umbral arbitrario y se determinó con esto de manera precisa qué accesiones coincidían con cada uno de los haplotipos generados. Este enfoque permitió asociarlos al valor asignado con el análisis físico y químico que se tiene tanto para GP y para SDS_Sedim.

6.5. Software.

Para obtención y alineación, de los marcadores con el genoma de referencia, se utilizó el sistema de gestión de almacenamiento de datos (DArTDB). Y para

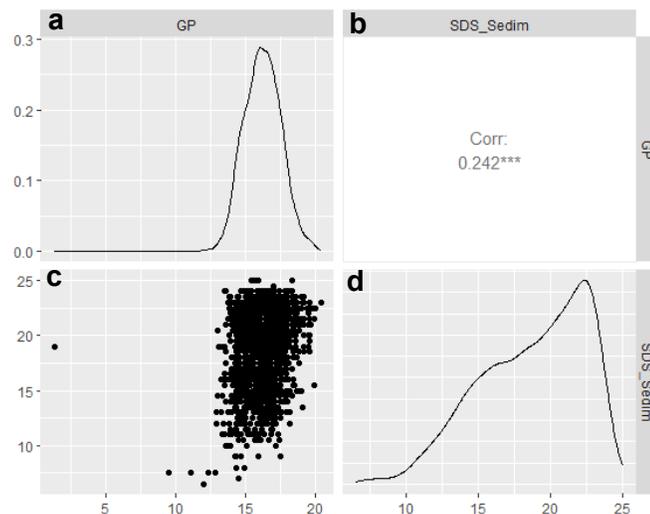
generar el formato HapMap a partir de dicho genoma, se utilizó DArTtoolbox. Ambos desarrollado por la empresa Diversity Arrays Technology (DArT).

Para realizar el análisis de asociación, se utilizó la Herramienta Integrada de Predicción y Asociación del Genoma v3.1.0 (GAPIT, del inglés Genome Association and Prediction Integrated Tool) (Lipka *et al.*, 2012). Para la generación de Haplotipos se utilizó la herramienta de análisis de datos a gran escala, PLINK v1.9 (Chang, 2015). Y para visualizar el correlograma, análisis de PCA y los haplotipos en gráficos de caja, se utilizó la paquetería de ggplot2 v3.4.2 (Wickham, 2016), todo esto desde R v4.1.2 (R Core, 2021).

7. Resultados y discusión

7.1. Estadística descriptiva y estructura de la población.

Para identificar patrones y valores de las muestras, proporcionadas por el laboratorio de Calidad de Trigo se realizó la estadística descriptiva de la distribución de los datos provenientes de los análisis físico y químico. Con la finalidad de, conocer su naturaleza y comportamiento (Figura 3). En la figura 3a y 3d se observan las distribuciones univariantes de cada variable mediante gráficos de densidad, mostrando que GP tiene una distribución asimétrica hacia la derecha. Esto significa que la mayoría de las muestras tienen valores concentrados de proteína, en un rango alto y alrededor de 15, mientras que la densidad (SDS_Sedim) presenta una distribución más amplia, con los valores se concentran entre 15 y 22.5. En la figura 3c, el diagrama de dispersión indica una leve tendencia positiva entre ambas variables, aunque con una notable dispersión de los datos. En la figura 3b, muestra la relación positiva de GP y SDS_Sedim, que por el coeficiente de correlación bajo se puede justificar la no existencia de multicolinealidad y por tanto realizar el GWAS para cada una de las variables.



La estructura de la población se realizó mediante el análisis de las primeras cinco componentes principales (ACP) (Figura 4). La figura 4a, muestra la relación entre CP1 (6.5%) y CP2 (4.5%), evidenciando una dispersión amplia en el eje “x” de CP1 y una agrupación central densa, este patrón sugiere que existen subpoblaciones que comparten cierta similitud genética, debido a factores geográficos o de selección local. La figura 4b, representa CP1 (6.5%) frente a CP3 (3.2%), mostrando una distribución central similar al primer gráfico, pero con algunas variaciones en la dispersión de los puntos, lo cual indica que PC3 capta aspectos adicionales de la variabilidad genética. Y la figura 4c, contrasta CP2 (4.5%) y CP3 (3.2%), este gráfico presenta una distribución uniforme con un núcleo central denso, sugiriendo que estas componentes reflejan variabilidad genética menos influenciada por CP1, lo que puede reflejar variaciones sutiles en la estructura genética que podrían estar relacionadas con diferencias climáticas, geográficas o de manejo en los orígenes de las accesiones.

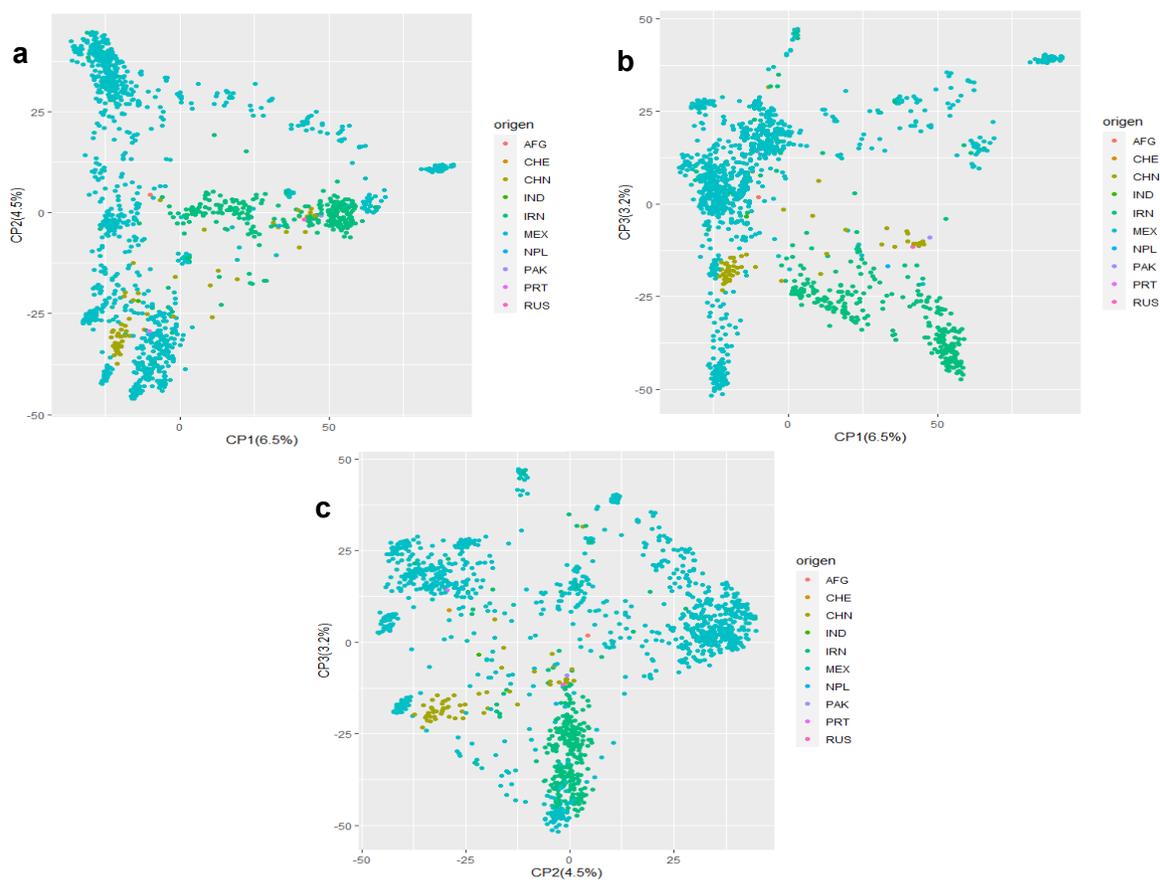


Figura 4. Gráficos ACP de 1.602 colectas de *T. aestivum* nativos de diferentes países a) CP1 vs CP2, b) CP1 vs CP3 y c) CP2 vs CP3.

7.2. Análisis de asociación.

Los marcadores significativos resultantes del GWAS en cada uno de los modelos estadísticos utilizados (MLM, FarmCPU y BLINK) se seleccionaron teniendo en cuenta su valor p ajustado por la tasa de falso descubrimiento (FDR) con un nivel de significancia de 0.01 lo que minimizó falsos positivos y abordó eficazmente los efectos multilocus, siguiendo un enfoque similar a investigaciones anteriores (Sandhu *et al.*, 2021).

Se localizaron un total de 43 marcadores significativos, 28 marcadores significativos para característica GP y 15 marcadores significativos para SDS_Sedim.

El modelo BLINK identificó un total de 21 marcadores, de los cuales 9 se asociaron a GP y 12 a SDS_Sedim (Figura 5). Para GP, estos marcadores se distribuyen en los cromosomas 1A, 2A, 3B, 3D, 4A, 4B, 5B, 6A y 6D (Figura 5a). Al contrastar estas asociaciones con estudios previos en *T. aestivum*, se observó que los cromosomas 2A y 6A ya habían sido reportados en relación con el contenido de proteína en grano (Sansaloni *et al.*, 2020). Específicamente, el marcador ubicado en el cromosoma 2A (a 737 cM) se encuentra próximo a otro descrito a 775 cM, mientras que el marcador en el cromosoma 6A (a 585 cM) se localiza cerca de uno registrado a 602 cM (Sansaloni *et al.*, 2020). Desafortunadamente, el artículo no especifica si dichos marcadores están directamente relacionados con la codificación de alguna proteína, sino que únicamente se reportan las asociaciones.

Por otra parte, para SDS_Sedim, los 12 marcadores identificados se detectaron en los cromosomas 1B, 2B, 2D, 3A, 4D, 5A, 5D, 6B y 7A, siendo los cromosomas 2D, 5A y 7A los que presentaron dos marcadores significativos cada uno (Figura 5b). Además, se han descrito asociaciones en *T. aestivum* en los cromosomas 2B y 7A; en el primero, la asociación corresponde al contenido de proteína de la harina, y en el segundo, a la sedimentación de la misma (Juliana *et al.*, 2019). En el cromosoma 2B (a 116 cM), el marcador se encuentra cercano a otro reportado a 83 cM, mientras que en el cromosoma 7A, el marcador ubicado a 25.8 cM se sitúa próximo a uno informado a 55 cM (Juliana *et al.*, 2019). Al igual

que en el caso de GP, estas asociaciones se describen sin aportar información adicional sobre una posible relación directa con la codificación de proteínas.

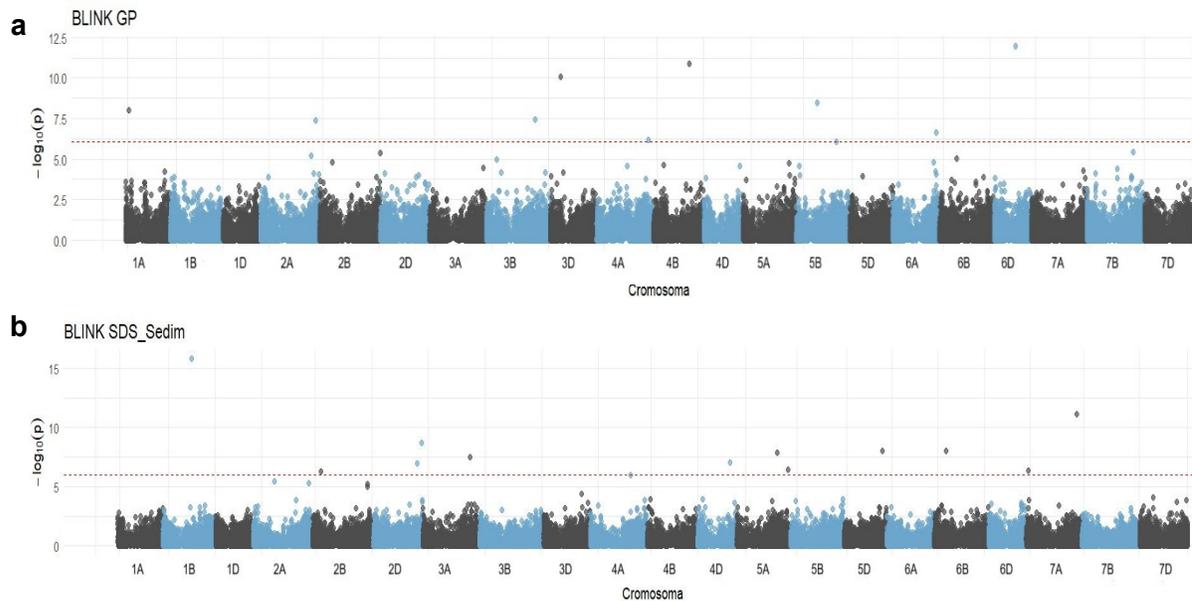


Figura 5. Gráficos de Manhattan del modelo BLINK que muestran la asociación de 77,395 SNP's en 1,602 colectas de *T. aestivum* nativos: **a)** Grafico manhattan para GP con 9 marcadores significativos; **b)** Grafico manhattan para SDS_Sedim con 12 marcadores significativos.

El modelo FarmCPU identificó un total de 17 marcadores, 12 asociados a GP y 5 a SDS_Sedim (Figura 6). En cuanto a GP, dichos marcadores se localizaron en los cromosomas 2B, 2D, 3D, 4B, 5A, 5B, 6A, 6B, 7B y 7D, destacándose que en los cromosomas 6A y 7D se detectaron dos marcadores significativos cada uno (Figura 6a). Además, se constataron asociaciones previamente reportadas en *T. aestivum*, en las que se relacionó el contenido de proteína de grano sin aportar información adicional en los cromosomas 2B, 5B y 6A (Kumar *et al.*, 2018; Juliana *et al.*, 2019). En el cromosoma 2B (a 174.22 cM), el marcador se encuentra próximo a uno descrito a 152.1 cM; en 5B (a 275.76 cM), está cercano a otro documentado a 256.49 cM. Por su parte, en 6A (a 32.59 cM) se identificaron dos marcadores reportados a 25.75 cM (Kumar *et al.*, 2018) y otro casi idéntico a 32 cM, el cual se localiza en una región de ADN asociada a un rasgo particular (QTL, del inglés Quantitative Trait Loci) Qcim.6A.4 descrito por Juliana y colaboradores (Juliana *et al.*, 2019) la cual integra genes conocidos y QTL informados en la región

significativa, entre los cuales destaca el Locus Gli-A2, asociado con la sedimentación de la harina (Manifiesto *et al.*, 1998; Huo *et al.*, 2018).

Los 5 marcadores identificados en SDS_Sedim, se ubicaron en los cromosomas 1B, 2B, 4A, 5A y 5D (Figura 6b). Se encontraron asociaciones previamente reportadas en *T. aestivum*, donde en el cromosoma 1B el marcador se asoció con el contenido de gluten y en el cromosoma 2B con el valor de sedimentación de la harina. En el cromosoma 1B (a 379.4 cM), el marcador está próximo a otro reportado a 343.26 cM, mientras que en 2B (a 744.46 cM) se halló uno cercano a 731.6 cM (Hao *et al.*, 2022). Es importante destacar que estas asociaciones se limitaron a las características mencionadas, sin aportar información adicional.

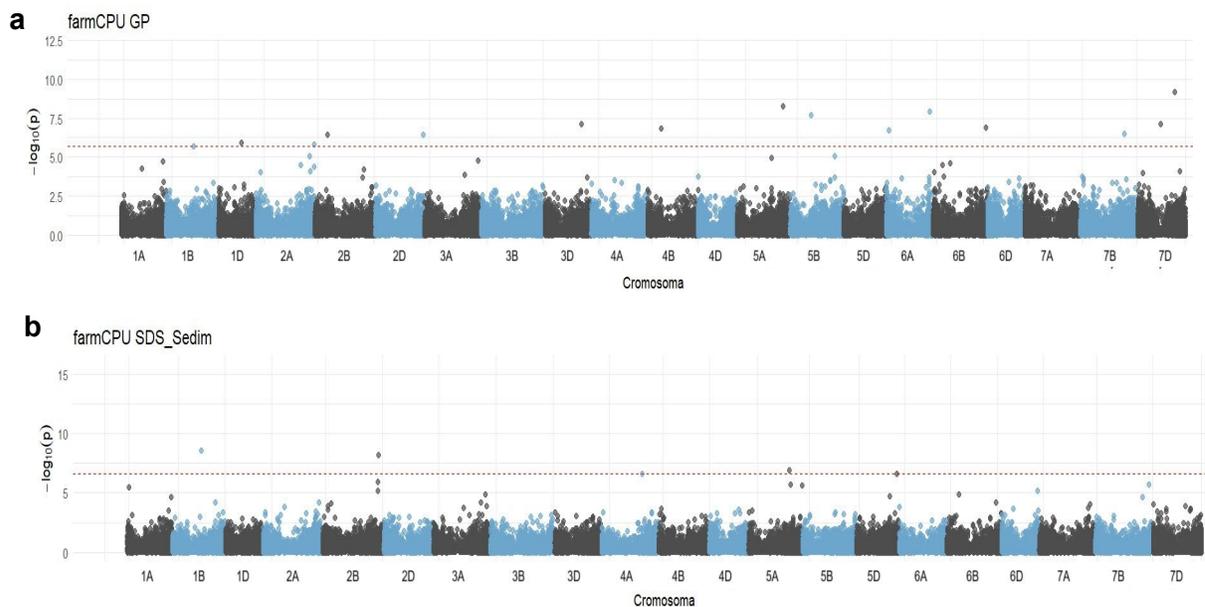


Figura 6. Gráficos de Manhattan del modelo FarmCPU que muestran la asociación de 77,395 SNP's en 1,602 colectas de *T. aestivum* nativos: **a)** Grafico manhattan para GP con 12 marcadores significativos; **b)** Grafico manhattan para SDS_Sedim con 5 marcadores significativos.

Finalmente, el modelo MLM identificó un total de 9 marcadores (Figura 7), ubicados en los cromosomas 2A, 2B, 2D, 3D, 5D y 7D para GP. Cabe destacar que en 5D y 7D se registraron dos y tres marcadores significativos, respectivamente. Otros estudios de asociación en *T. aestivum* para contenido de proteína en grano, también muestran SNP's en los cromosomas 2A, 2B y 5D (Lie *et al.*, 2023).

En el cromosoma 2A (a 71.32 cM) se localizó un marcador próximo a otro descrito a 89.1 cM; en 2B (a 170.83 cM) se detectó uno cercano a 168.1 cM, donde coincide el QTL QGpcqwert24, y en 5D (a 88.55 cM) se observó uno próximo a 83.4 cM (Lie *et al.*, 2023). Por otro lado, para SDS_Sedim no se obtuvo ningún marcador significativo mediante este modelo.

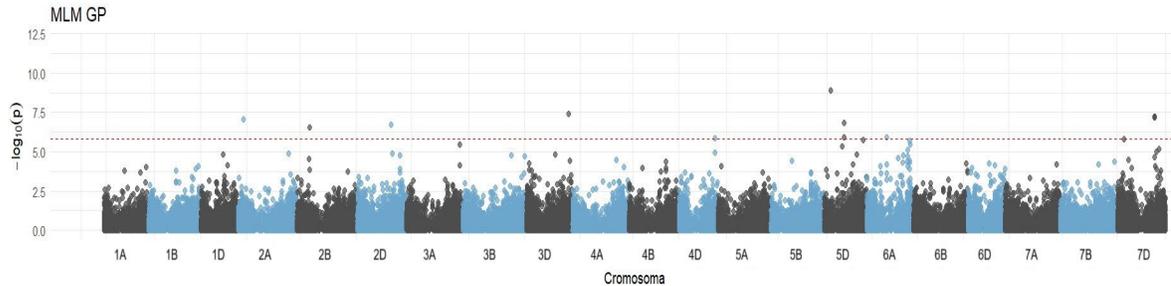


Figura 7. Gráfico de Manhattan del modelo MLM que muestran la asociación de 77,395 SNP's en 1,602 colectas de *T. aestivum* nativos: **a)** Gráfico Manhattan para GP con 9 marcadores significativos.

7.2.1. Identificación de Marcadores Significativos en Múltiples Modelos y Cromosomas

Los gráficos Manhattan circulares facilitaron la comparación entre modelos y permitieron evaluar la confiabilidad de los SNP's detectados.

Para GP, el modelo FarmCPU se destacó como el que mejor se ajustó, ya que identificó la mayor cantidad de asociaciones de SNP, detectando 12 marcadores en comparación con los 9 identificados por BLINK y los 9 (únicamente para GP) de MLM. Este desempeño superior se atribuye a su capacidad para equilibrar la sensibilidad en la detección de asociaciones y el control de falsos positivos, lo que resulta esencial para garantizar la robustez en el análisis (Kaler *et al.*, 2020).

En contraste, para SDS_Sedim el modelo BLINK demostró ser el más apropiado, ya que logró identificar 12 marcadores, frente a los 5 detectados por FarmCPU y la ausencia de detecciones por MLM. La eficiencia de BLINK se debe a su capacidad para maximizar la potencia estadística mediante la minimización de los efectos derivados de las relaciones genéticas entre individuos, lo que facilita la identificación de asociaciones significativas (Sandhu *et al.*, 2024).

Un hallazgo clave fue la identificación de marcadores significativos en al menos dos modelos, lo que refuerza su relevancia biológica. Para GP (Figura 8a), se encontraron SNP's en los cromosomas 5B (FarmCPU y BLINK) y 7D (MLM y FarmCPU). Para SDS_Sedim (Figura 8b), se identificaron en 1B y 5A (ambos en FarmCPU y BLINK). Estos resultados sugieren que las regiones identificadas podrían albergar genes candidatos de efecto mayor, constituyendo prospectos sólidos para futuros estudios funcionales y aplicaciones en programas de mejora genética (Sukumaran & Yu, 2014).

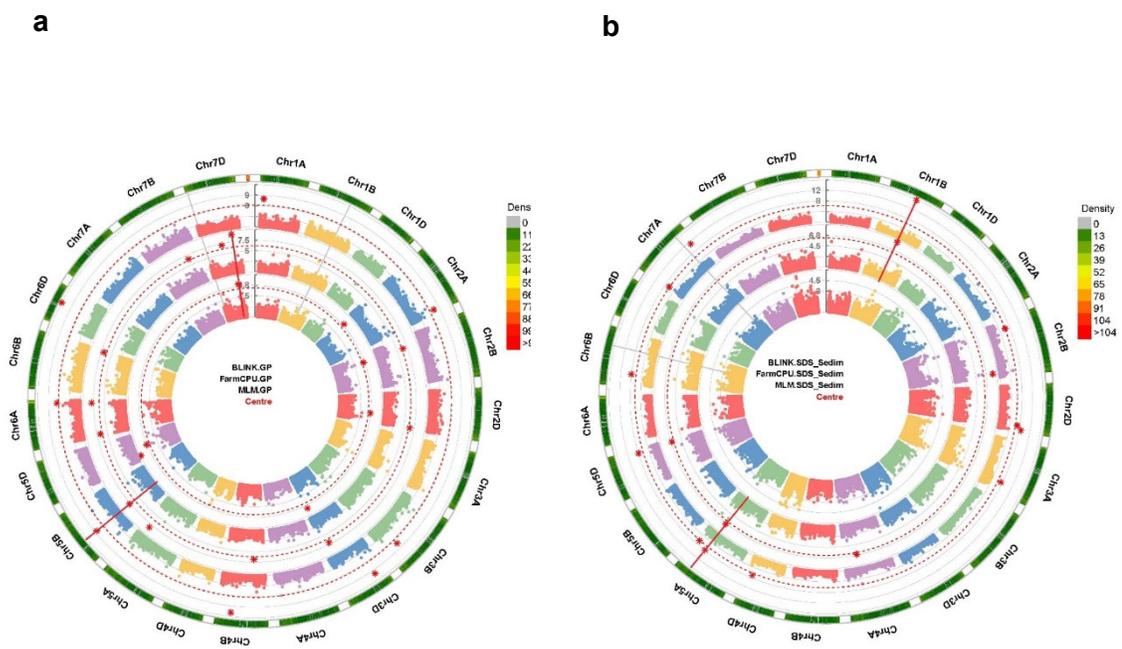


Figura 8. Gráficos de Manhattan circular de los tres modelos BLINK, FarmCPU y MLM. Se muestran la asociación de 77,395 SNP's en 1,602 colectas de *T. aestivum* nativos: **a)** Asociación de GP con un par de marcadores significativos que coincidieron en los modelos, para el caso del cromosoma 5B coincidieron en FarmCPU y BLINK y en el cromosoma 7D coincidieron en los modelos MLM y FarmCPU, **b)** Asociación de SDS_Sedim con un par de marcadores significativos que concordaron en los modelos de FarmCPU y BLINK, en los cromosomas 1B y 5A.

7.3. Gráficos Q-Q.

En esta sección, se presentan los gráficos Q-Q (Figura 9), que comparan las asociaciones observadas entre los marcadores para GP y SDS_Sedim con las esperadas tras considerar la estructura poblacional de los tres modelos empleados. En el caso de GP (Figura 9a), el modelo FarmCPU se destacó como el que mejor se ajustó, ya que en su gráfico las asociaciones se alinean con lo esperado por azar,

mostrando consistencia en todo el genoma y generando el mayor número de marcadores significativos. Por otro lado, para SDS_Sedim (Figura 9b), el modelo BLINK se reveló como el más adecuado, evidenciado por la capacidad de detectar asociaciones significativas reales con alta sensibilidad, especialmente en las colas superiores de la distribución observada. En contraste, MLM mostró un enfoque más conservador, minimizando los falsos positivos, mientras que FarmCPU presentó un desempeño intermedio. Estos hallazgos confirman que la elección del modelo debe ajustarse a la característica evaluada, optimizando la detección de asociaciones relevantes en cada caso (Sandhu et al., 2024).

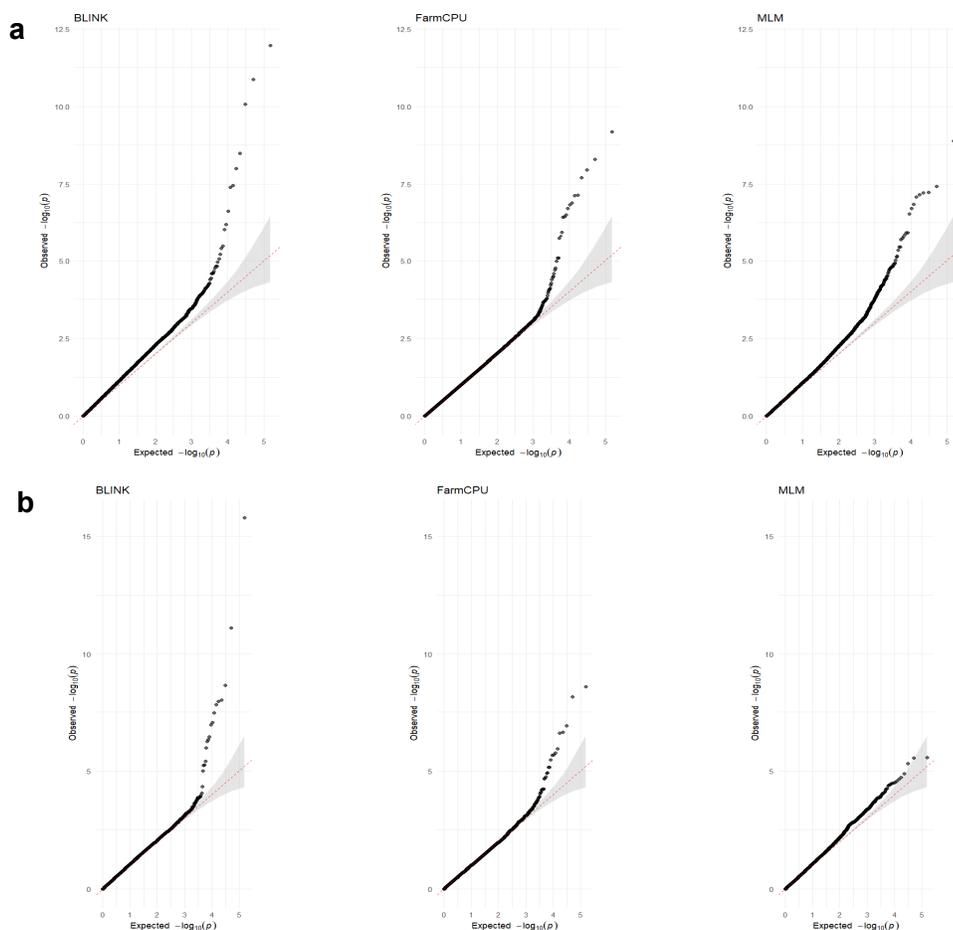


Figura 9. Los gráficos Q-Q ilustran las asociaciones observadas entre los marcadores obtenidos de los tres modelos de GWAS: **a)** Gráficos Q-Q para GP y **b)** Gráficos Q-Q para SDS_Sedim.

7.4. Mapeo Preciso de Marcadores de GWAS y sus Bloques de Asociación.

La Tabla 2 y Tabla 3 contiene la información de los marcadores que resultaron significativos para el GWAS: la primera columna refiere el cromosoma al que pertenece el marcador, la siguiente representa el nombre que identifica al marcador. La tercera columna representa, el valor p resultante del GWAS, la cuarta columna muestra el modelo o modelos donde se obtuvo el marcador (SNP), la quinta columna, muestra la posición física en pares de bases del SNP y en la última columna muestra el bloque de asociación, estos bloques se construyeron con un rango de ± 1 cM partiendo de la posición física, teniendo una trazabilidad más amplia en la base de datos de EnsemblPlants (Martin *et al.*, 2023).

Es fundamental considerar la relevancia de los cromosomas 1 y 6 (A, B y D) en este estudio, ya que contienen genes clave para la calidad proteica del trigo. En la Tabla 2, se identificaron SNPs en los cromosomas 1A, 6A, 6B y 6D, mientras que en la Tabla 3, también se detectaron marcadores en 1B y 6B, lo que sugiere que estas regiones pueden ser de interés para la caracterización de alelos relacionados con el contenido de proteína en grano y la calidad del gluten.

Dado que estos cromosomas albergan genes que codifican las principales proteínas del gluten, como las gluteninas de bajo (LMWG) y alto peso molecular (HMWG) y las gliadinas (ω/γ en 1A, 1B y 1D; α/β en 6A, 6B y 6D) (Huo *et al.*, 2019; Sansaloni *et al.*, 2020; Guzmán *et al.*, 2022a), la presencia de SNPs en estas regiones resulta particularmente relevante. Sin embargo, la exploración en bases de datos genómicas como EnsemblPlants y UniProt no permitió establecer una relación directa entre estos marcadores y los genes de gliadinas y gluteninas.

Tabla 2. Resultados GWAS de marcadores significativos de los modelos BLINK, FarmCPU y MLM pertenecientes a GP en 1,602 colectas de *T. aestivum* nativos.

<i>Cromosoma</i>	<i>SNP</i>	<i>Valor ρ</i>	<i>Modelos significativos</i>	<i>Posición física (pb)</i>	<i>BA</i>
1A	994242 F 0-40:T>G-40:T>G	1.00E-08	BLINK	49000072	1
2A	1125593 F 0-28:C>T-28:C>T	8.29E-08	MLM	71323813	2
	999293 F 0-29:G>A-29:G>A	4.00E-08	BLINK	737416572	3
2B	5579663 F 0-66:A>C-66:A>C	2.96E-07	MLM	170832630	4
	2291704 F 0-68:C>T-68:C>T	3.53E-07	FarmCPU	174229727	5
2D	1230460 F 0-50:T>G-50:T>G	1.94E-07	MLM	449253906	6
	1071905 F 0-9:T>G-9:T>G	3.75E-07	FarmCPU	640830639	7
3B	1126067 F 0-49:T>C-49:T>C	3.53E-08	BLINK	648047722	8
	985642 F 0-34:T>C-34:T>C	3.74E-08	MLM	564045228	9
3D	996906 F 0-56:G>A-56:G>A	7.36E-08	FarmCPU	491540379	10
	1047222 F 0-30:C>T-30:C>T	8.62E-11	BLINK	142134747	11
4A	1203459 F 0-15:A>G-15:A>G	6.45E-07	BLINK	689693335	12
	1201010 F 0-9:C>A-9:C>A	1.48E-07	BLINK	183157442	13
4B	1088410 F 0-18:G>A-18:G>A	1.34E-11	FarmCPU	480447525	14
	999614 F 0-46:C>T-46:C>T	5.12E-09	FarmCPU	616159007	15
5B	100218022 F 0-40:C>G-40:C>G	1.16E-08	FarmCPU y BLINK	275768394	16
5D	2248887 F 0-13:T>G-13:T>G	1.32E-09	MLM	88551684	17
	2241741 F 0-37:T>G-37:T>G	1.43E-07	MLM	262914224	18
6A	2279397 F 0-55:C>G-55:C>G	1.93E-07	FarmCPU	32598078	19
	2254014 F 0-12:C>G-12:C>G	1.10E-08	FarmCPU	582647804	20
	1217805 F 0-29:C>T-29:C>T	2.37E-07	BLINK	585113505	21
6B	6027061 F 0-55:T>C-55:T>C	1.30E-07	FarmCPU	712735091	22
6D	1132926 F 0-13:T>G-13:T>G	1.10E-12	BLINK	281062455	23
7B	978206 F 0-9:C>G-9:C>G	3.12E-07	FarmCPU	587787084	24
	1124768 F 0-44:C>G-44:C>G	7.27E-08	FarmCPU	308754715	25
7D	977738 F 0-15:T>G-15:T>G	7.02E-08	MLM	500332533	
	977738 F 0-39:C>G-39:C>G	5.81E-08	MLM y FarmCPU	500332533	26
	977738 F 0-62:T>A-62:T>A	5.99E-08	MLM	500332533	

Nota. ρ = valor de p al 0.01 ajustado por tasa de falso descubrimiento (FDR), el símbolo E en los valores de p indica $\times 10^n$, pb = pares de bases y BA = Bloque de asociación.

Tabla 3. Resultados GWAS de marcadores significativos de los modelos BLINK y FarmCPU pertenecientes a SDS_Sedim en 1,602 colectas de *T. aestivum*. nativos.

<i>Cromosoma</i>	<i>SNP</i>	<i>Valor ρ</i>	<i>Modelos significativos</i>	<i>Posición física (pb)</i>	<i>BA</i>
1B	2275173 F 0-18:G>C-18:G>C	1.24E-09	FarmCPU y BLINK	379404953	1
2B	100274923 F 0-9:T>C-9:T>C	6.61E-09	FarmCPU	744463552	2
	2291631 F 0-7:C>T-7:C>T	5.68E-07	BLINK	113156951	3
2D	100248535 F 0-8:C>G-8:C>G	1.15E-07	BLINK	580054776	4
	1219375 F 0-17:C>G-17:C>G	2.17E-09	BLINK	641691858	5
3A	12419315 F 0-59:G>C-59:G>C	3.54E-08	BLINK	628184844	6
4A	1088015 F 0-9:T>G-9:T>G	2.58E-07	FarmCPU	533079975	7
4D	2242210 F 0-61:T>G-61:T>G	8.92E-08	BLINK	433162538	8
5A	29429694 F 0-64:A>G-64:A>G	6.98E-08	BLINK y FarmCPU	535572951	9
	1210571 F 0-7:G>A-7:G>A	3.61E-07	BLINK	682765400	10
5D	1159458 F 0-33:C>G-33:C>G	2.32E-07	FarmCPU	533815342	11
	1100180 F 0-60:G>T-60:G>T	1.02E-08	BLINK	510557316	12
6B	100220372 F 0-50:C>G-50:C>G	9.28E-09	BLINK	165040825	13
7A	1017632 F 0-57:A>G-57:A>G	4.82E-07	BLINK	25845577	14
	985710 F 0-31:G>T-31:G>T	7.88E-12	BLINK	677723033	15

Nota. ρ = valor de p al 0.01 ajustado por tasa de falso descubrimiento (FDR), el símbolo E en los valores de ρ indica $\times 10^n$, pb = pares de bases y BA = Bloque de asociación.

La base de datos de EnsemblPlant se utilizó para identificar los genes correspondientes a los bloques de asociación (Martin *et al.*, 2023). Y la función biológica, de las proteínas, se identificó utilizando la base de datos UniProt, la cual muestra los dominios funcionales de los genes identificados en EnsemblPlant (Tabla 1 y 2 de anexos) (Coudert *et al.*, 2023). Desafortunadamente, no se encontraron genes directamente relacionados con los rasgos evaluados y para algunos bloques de asociación no se identificó información. La falta de información sugiere que las asociaciones reportadas podrían no ser exclusivas para los rasgos estudiados. Esto podría atribuirse a que los datos provienen de *T. aestivum* nativos, cuya diversidad genética puede no estar completamente representada en la base de datos utilizada, así como al hecho de que el genoma de referencia empleado fue el más reciente (IWGSC RefSeq v2.1), lo que implica que aún existen regiones del genoma por caracterizar en profundidad.

En los bloques asociados con GP, se encontraron numerosos genes con funciones relacionadas con: transporte transmembranal, abarcando cadmio, calcio, zinc, xenobióticos, hierro férrico, fosfato, magnesio y azúcares. Además, exhiben actividades oxidorreductoras y participan en procesos catabólicos asociados al peróxido de hidrógeno. Estos genes también están vinculados a la regulación de la formación del cigoto y endospermo, al desarrollo de las flores, a respuestas defensivas contra otros organismos, a la aclimatación al frío y a respuestas a estímulos abióticos.

Por otro lado, en los bloques asociados con SDS_Sedim, se identificaron genes relacionados con el transporte transmembranal de calcio y zinc, la unión a iones metálicos, lípidos y carbohidratos. Además, están involucrados en el silenciamiento génico mediado por ácidos ribonucleicos no codificantes (ncARN), el desarrollo del xilema y las respuestas a otros organismos, así como al estímulo de factores abióticos. Para obtener información más detallada, se pueden consultar los Anexos 1 y Anexo 2.

7.5. Análisis de Haplotipos Basado en LD: Identificación y Distribución Cromosómica

El objetivo de este análisis fue identificar las accesiones con el mayor número de haplotipos en su genoma, proporcionando una visión más completa de la diversidad genética presente en las variedades utilizadas en el estudio. Esta información es clave para comprender la variabilidad genética en las poblaciones analizadas y su posible impacto en las características de calidad del trigo.

El análisis de haplotipos basado en LD, similar al de Al Bkhetan *et al.*, 2019, se inició seleccionando cada uno de los marcadores significativos identificados en el estudio, detallados en las Tablas 2 y 3 respectivamente. Posteriormente, se eligieron aquellos con frecuencias superiores al 0.6, teniendo haplotipos por cada uno de los cromosomas, que para GP, tenemos en los cromosomas 1A, 2A, 2B, 2D, 3B, 3D, 4A, 4B, 5A, 5B, 5D, 6A, 6B, 6D, 7B y 7D teniendo una cobertura en casi la totalidad de los cromosomas y para SDS_Sedim en los cromosomas 1B, 2B, 2D, 3A, 4A, 4D, 5A, 5D, 6B y 7A, con la identificación de estos haplotipos, se pudo ubicar con precisión las variedades que mostraban asociaciones con estos complejos de alelos contiguos.

Con toda la información recopilada durante el estudio, se logró crear una combinación de alelos favorables que se presentaron en gráficos de cajas. En estos gráficos, el eje "y" representa el valor de la caracterización fisicoquímica y el eje "x" representa los haplotipos creados. De esta manera, cada cromosoma fue descrito en un gráfico de cajas.

Debido a la metodología utilizada en la generación de los haplotipos, no contamos con gráficos para todos los cromosomas que son relevantes para este estudio, sin embargo, se han detectado 63 haplotipos vinculados a GP, de los cuales 9 están asociados al cromosoma 1A, 29 al cromosoma 6A, 12 al cromosoma 6B y 13 al cromosoma 6D.

En la Figura 10, se encuentra el cromosoma 1A (Figura 10a), donde podemos observar que el haplotipo “GCG” está por encima de la media estándar de GP en comparación de los demás, este haplotipo está compuesto por tres alelos, como se puede observar, sabiendo la posición física de inicio y fin, se pudo localizar e identificamos que está constituido por tres ID estables de genes: *TraesCS1A03G0157600*, *TraesCS1A03G0157800* y *TraesCS1A03G0157900* (consultar <https://plants.ensembl.org/biomart/martview/87103557fa6b24d53ddbf6eb3b981bdf>).

Este conjunto de alelos codifica una proteína que contiene el dominio RRM, es decir, unión a una molécula de ARN o una porción de la misma (como se describe en detalle en el Anexo 1, BA 1) (Coudert *et al.*, 2023). Además, es relevante destacar que este haplotipo está asociado a la variedad iraní CWI 55380 únicamente. Los haplotipos en el cromosoma 1A para GP, las accesiones CWI 59819 y CWI 61061 provenientes de México comparten 8 de los 9 haplotipos representados en la gráfica de caja (Figura 10a). Y no se pudo observar ninguna diferencia en el conjunto de los haplotipos correspondientes a los cromosomas 6A, 6B y 6D (Figura 10b, c y d).

En los gráficos de caja del cromosoma 6A de GP, la accesión CWI 67503, originaria de Irán, junto con las mexicanas CWI 69298 y CWI 65542, exhiben 26 de los 29 haplotipos identificados (Figura 10b). En el cromosoma 6B de GP, las accesiones CWI 63474, CWI 66087, CWI 69774 y CWI 69792 contienen los 12 haplotipos de la gráfica de caja (Figura 10c). Por último, el cromosoma 6D en GP, las accesiones CWI 62404, CWI 66619, CWI 60351, CWI 14643, CWI 73041, CWI 60451, CWI 60295, CWI 63908, CWI 49556, CWI 63413, CWI 60556 presentan los 13 haplotipos de la gráfica de caja (Figura 10d).

En el total de 63 haplotipos presentes en los gráficos de cajas del cromosoma 1A y el cromosoma 6 (A, B y D) para GP, la accesión CWI 60918 contiene 44 haplotipos, seguida por CWI 54234, CWI 65405 y BW 44438 con 36 haplotipos,

siendo estas las accesiones con la mayor cantidad de alelos beneficiosos para la característica de GP.

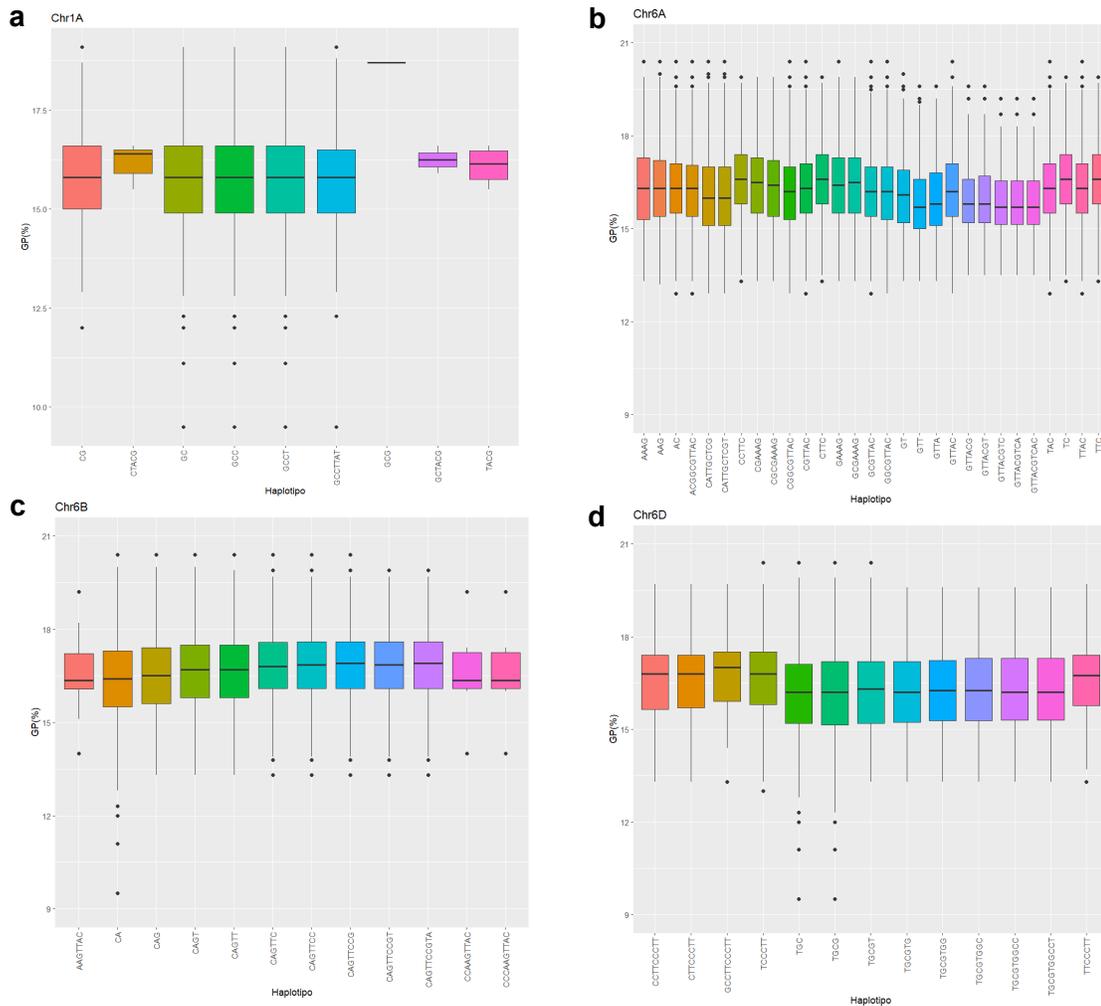


Figura 10. Los diagramas de caja indican los valores del análisis físico y químico de GP a los distintos grupos de haplotipos: **a)** Para GP en el cromosoma 1A, el haplotipo "GCG" esta significativamente alto a comparación de los demás; **b)** GP en el cromosoma 6A; **c)** GP en el cromosoma 6B; **d)** GP en el cromosoma 6D no hay diferencia relévale.

En cuanto a la característica, relacionada con la densidad de las proteínas, SDS_Sedim, también se realizaron gráficos de cajas para representar los haplotipos de los cromosomas relacionados con la síntesis de gliadinas y gluteninas (1B y 6B). Localizándose 20 haplotipos asociados, de los cuales 18 están vinculados al cromosoma 1B y únicamente 2 al cromosoma 6B (Figura 11).

En lo que respecta al cromosoma 1B, se observa que todos los haplotipos presentan similitudes dentro del rango de la media estándar, sin destacar ninguna diferencia significativa (Figura 11a). Del mismo modo, en el cromosoma 6B, no se observan diferencias notables que mencionar (Figura 11b).

En el cromosoma 1B para SDS_Sedim, 82 de 1,602 accesiones presentan la totalidad de los 18 haplotipos generados, como se muestra en la gráfica de caja (Figura 11a). Para el caso del cromosoma 6B en SDS_Sedim, 724 de 1602 accesiones, exhiben los 2 haplotipos de la gráfica de caja (Figura 11d).

Para SDS_Sedim, que solo incluye el cromosoma 1B y el 6B, de los 20 haplotipos presentes en el par de gráficos de cajas (Figura 11), las accesiones CWI 61283, CWI 54096, CWI 66075, CWI 65730, CWI 54178, CWI 63171, CWI 66890, CWI 61602, CWI 66068, CWI 60303, CWI 60272, CWI 49583, CWI 31372, CWI 62763, CWI 62792 y CWI 62943 poseen la totalidad de estos haplotipos, destacándose como las accesiones con todos los posibles alelos beneficiosos para SDS.

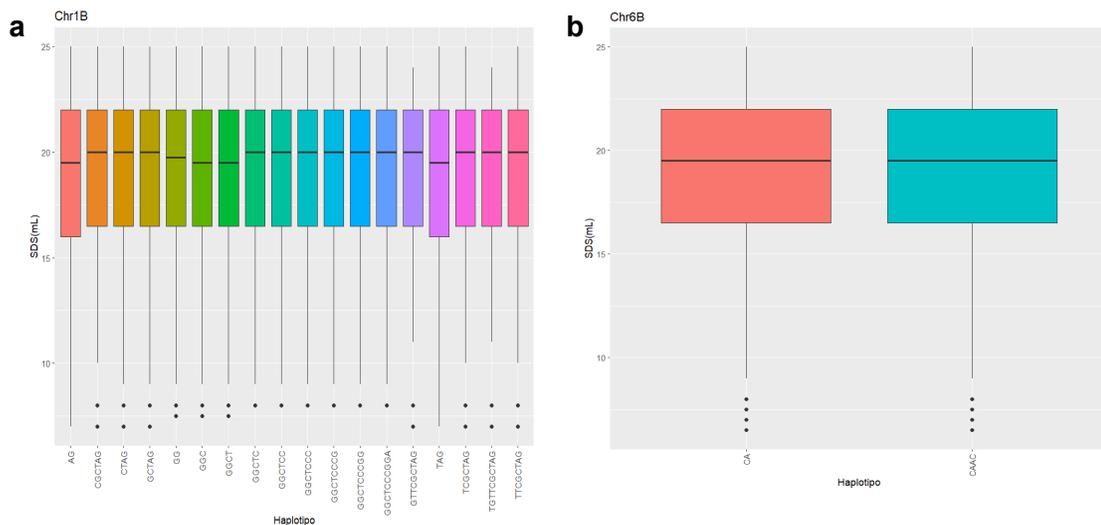


Figura 11. Los diagramas de caja indican los valores del análisis físico y químico de SDS_Sedim a los distintos grupos de haplotipos: **a)** SDS_Sedim en el cromosoma 1B y **b)** SDS_Sedim en el cromosoma 6B no hay diferencia relévale.

A pesar de que estos haplotipos están ubicados en cromosomas para la síntesis de gluteninas y gliadinas, no se ha identificado su vinculación con estos genes. Específicamente, en la búsqueda realizada en EnsemblPlant (Martin *et al.*, 2023) y UniProt (Coudert *et al.*, 2023) no se encontró la relación de los haplotipos con genes asociados a las proteínas del gluten, lo que sugiere que estos haplotipos no estén dentro de la región codificante de los genes de interés, sino en regiones intergénicas o reguladoras, afectando la expresión de genes clave sin estar directamente en su secuencia codificante (Guzmán *et al.*, 2022b). Además, aunque el IWGSC RefSeq v2.1 es una de las versiones más avanzadas del genoma del trigo, aún presenta limitaciones en la anotación de ciertos genes, especialmente en accesiones de trigo nativo, lo que podría explicar la falta de asociaciones directas en este estudio (Zhu *et al.*, 2021).

Finalmente, si se consideran conjuntamente los haplotipos de GP y SDS_Sedim, resultan 83 haplotipos, como se describe en las Figuras 10 y 11. Las accesiones CWI 53946 y CWI 14147 tienen 50 haplotipos, CWI 14346 tiene 51 haplotipos, BW 44438 tiene 52 haplotipos, CWI 54234 tiene 54 haplotipos y finalmente, CWI 60918 presenta un total de 62 haplotipos, lo que la hace la variedad con la mejor combinación de alelos para este estudio y es una de las accesiones que se puede recomendar para el programa de mejora genética en CIMMYT.

8. Conclusión

Este estudio constituyó el primer GWAS aplicado a accesiones nativas de *Triticum aestivum*, evaluando 1602 accesiones mediante 77 395 SNP's y empleando los modelos MLM, FarmCPU y BLINK en GAPIT. Se identificaron 43 marcadores significativos distribuidos en 21 cromosomas, de los cuales 28 se asociaron con el porcentaje de proteína en grano (GP) y 15 con la calidad del gluten (SDS_Sedim). Los datos del rasgo GP se ajustaron mejor al modelo FarmCPU, el cual identificó 12 marcadores con asociaciones robustas y un control eficiente de falsos positivos. Por otro lado, el modelo BLINK resultó más eficaz para SDS_Sedim, al detectar un mayor número de asociaciones significativas, incluyendo aquellas con efectos menores, manteniendo un adecuado control estadístico. Estos resultados metodológicos respaldan la utilidad de aplicar múltiples enfoques analíticos en el estudio de rasgos poligénicos mediante GWAS.

El análisis de haplotipos basado en desequilibrio de ligamiento reveló combinaciones alélicas favorables en los cromosomas 1A, 1B, 6A, 6B y 6D, señalando a la accesión CWI 60918 como la que acumula el mayor número de alelos beneficiosos para ambos rasgos. Aunque la exploración de genes en los bloques de asociación no evidenció vínculos directos con loci de gluteninas o gliadinas, probablemente debido a la anotación incompleta del genoma IWGSC RefSeq v2.1, estos hallazgos validan la solidez del enfoque multicéntrico. Desde una perspectiva práctica, los marcadores y genotipos priorizados ofrecen criterios precisos para la selección de líneas en los programas de mejoramiento del CIMMYT, optimizando el desarrollo de variedades de trigo con mayor valor nutricional y funcionalidad tecnológica. Asimismo, se sugiere validar funcionalmente los marcadores identificados e incorporar futuras versiones del genoma de referencia para mejorar la anotación y la detección de variantes regulatorias para mejoramiento genético llevados a cabo en el CIMMYT.

Referencias

1. Al Bkhetan, Z., Zobel, J., Kowalczyk, A., Verspoor, K., & Goudey, B. (2019). Exploring effective approaches for haplotype block phasing. *BMC Bioinformatics*, 20(1), 540. <https://doi.org/10.1186/s12859-019-3095-8>.
2. American Association of Cereal Chemists (AACC). (1995). *Approved methods of the AACC* (9th ed.). St. Paul, MN, USA: Author
3. Amir, R., Sani, Q.-u.-A., Maqsood, W., Munir, F., Fatima, N., Siddiqa, A., & Ahmad, J. (2020). Chapter 6 - Pan-genomics of plant pathogens and its applications. En D. Barh, S. Soares, S. Tiwari, & V. Azevedo (Eds.), *Pan-genomics: Applications, challenges, and future prospects* (pp. 121-145). Academic Press. <https://doi.org/10.1016/B978-0-12-817076-2.00006-8>.
4. Amiteye, S. (2021). Basic concepts and methodologies of DNA marker systems in plant molecular breeding. *Heliyon*, 7(10), e08093. <https://doi.org/10.1016/j.heliyon.2021.e08093>
5. Barak, S., Mudgil, D., & Khatkar, B. S. (2013). Relationship of gliadin and glutenin proteins with dough rheology, flour pasting and bread making performance of wheat varieties. *LWT - Food Science and Technology*, 51(1), 211-217. <https://doi.org/10.1016/j.lwt.2012.09.011>
6. Bazakos, C., Hanemian, M., Trontin, C., Jiménez-Gómez, J. M., & Loudet, O. (2017). New strategies and tools in quantitative genetics: How to go from the phenotype to the genotype. *Annual Review of Plant Biology*, 68, 435–455. <https://doi.org/10.1146/annurev-arplant-042916-040820>.
7. Bernardo, R. (2010). *Breeding for quantitative traits in plants* (Vol. 1, p. 369). Woodbury, MN: Stemma Press.
8. Bush, W. S. (2019). Genome-wide association studies. In S. Ranganathan, M. Gribskov, K. Nakai, & C. Schönbach (Eds.), *Encyclopedia of bioinformatics and computational biology* (pp. 235-241). Academic Press. <https://doi.org/10.1016/B978-0-12-809633-8.20232-X>.
9. Cerda, J., Vera, C., & Rada, G. (2013). Odds ratio: Aspectos teóricos y prácticos. *Revista Médica de Chile*, 141(10), 1329-1335. <https://doi.org/10.4067/S0034-98872013001000014>.

10. Chang, C. C. (2015, enero 25). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, 4(1). Obtenido de <http://pngu.mgh.harvard.edu/purcell/plink/>.
11. Cheg dali, Y., Ouabbou, H., Essamadi, A., Cervantes, F., Ibba, M. I., & Guzmán, C. (2020). Assessment of the glutenin subunits diversity in a durum wheat (*T. turgidum* ssp. *durum*) collection from Morocco. *Agronomy*, 10(7), 957. <https://doi.org/10.3390/agronomy10070957>
12. Chu, J., Zhao, Y., Beier, S., Schulthess, A. W., Stein, N., Philipp, N., Röder, M. S., & Reif, J. C. (2020). Suitability of single-nucleotide polymorphism arrays versus genotyping-by-sequencing for genebank genomics in wheat. *Frontiers in Plant Science*, 11, 42. <https://doi.org/10.3389/fpls.2020.00042>
13. Choudhury, S., Larkin, P., Xu, R., Hayden, M., Forrest, K., Meinke, H., Hu, H., Zhou, M., & Fan, Y. (2019). Genome-wide association study reveals novel QTL for barley yellow dwarf virus resistance in wheat. *BMC Genomics*, 20, 891. <https://doi.org/10.1186/s12864-019-6249-1>
14. Coudert, E., Gehant, S., de Castro, E., Pozzato, M., Baratin, D., Neto, T., Sigrist, C. J. A., Redaschi, N., Bridge, A., & UniProt Consortium. (2023). Annotation of biologically relevant ligands in UniProtKB using ChEBI. *Bioinformatics (Oxford, England)*, 39(1), btac793. <https://doi.org/10.1093/bioinformatics/btac793>.
15. Espitia Rangel, E., Martínez Cruz, E., Peña Bautista, R. J., Villaseñor Mir, H. E., & Huerta Espino, J. (2008). Polimorfismo de gluteninas de alto peso molecular y su relación con trigos harineros para temporal. *Agricultura Técnica en México*, 34(1), 57-67. Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias.
16. Evans, L. T. (1997). Adapting and improving crops: The endless task. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 352(1356), 901–906. <https://doi.org/10.1098/rstb.1997.0069>
17. FAO. (2024). *Cereal supply and demand brief*. Consultado el 10 de enero de 2024, de <https://www.fao.org/worldfoodsituation/csdb/en/>.
18. Foster, J. T., Bull, R. L., & Keim, P. (2020). Ricin forensics: Comparisons to microbial forensics. En B. Budowle, S. Schutzer, & S. Morse (Eds.), *Microbial forensics* (3ra ed., pp. 241-250). Academic Press. <https://doi.org/10.1016/B978-0-12-815379-6.00016-7>

19. Gale, M. D., & Youssefian, S. (1985). Dwarfing genes in wheat. En G. E. Russell (Ed.), *Progress in Plant Breeding* (Vol. 1, pp. 1-35). Butterworth-Heinemann. <https://doi.org/10.1016/B978-0-407-00780-2.50005-9>
20. Garcia Zavaleta, F. M. (2017). *Efecto de la sustitución parcial de harina de trigo (Triticum aestivum) por harina de tarwi (Lupinus mutabilis) sobre las características fisicoquímicas y aceptabilidad general de galletas tipos soda* (Tesis de licenciatura, Universidad Privada Antenor Orrego, Facultad de Ciencias Agrarias).
21. Gebhardt, C. (2007). Chapter 5. Molecular markers, maps and population genetics. En *Potato biology and biotechnology: Advances and perspectives* (pp. 77-86). <https://doi.org/10.1016/B978-044451018-1/50047-6>.
22. Gómez-Espejo, A. L., Sansaloni, C. P., Burgueño, J., Toledo, F. H., & Reyes-Valdés, M. H. (2021). Asociación de genoma completo para el hábito de crecimiento en trigo harinero (*Triticum aestivum* L.). *Ecosistemas y Recursos Agropecuarios*, 8(2), e2912. <https://doi.org/10.19136/era.a8n2.2912>.
23. Gupta, P. K., Rustgi, S., & Kulwal, P. L. (2005). Linkage disequilibrium and association studies in higher plants: Present status and future prospects. *Plant Molecular Biology*, 57(4), 461- 485. <https://doi.org/10.1007/s11103-005-0257-z>.
24. Guzmán, C., Itria Ibba, M., Álvarez, J. B., Sissons, M., & Morris, C. (2022a). Wheat quality. En M. P. Reynolds & H.-J. Braun (Eds.), *Wheat improvement: Food security in a changing climate* (pp. 177-194). Springer. <https://doi.org/10.1007/978-3-030-90673-3>.
25. Guzmán, C., Crossa, J., Mondal, S., Govindan, V., Huerta, J., Crespo-Herrera, L., Vargas, M., Singh, R. P., & Ibba, M. I. (2022b). Effects of glutenins (Glu-1 and Glu-3) allelic variation on dough properties and bread-making quality of CIMMYT bread wheat breeding lines. *Field Crops Research*, 284, 108585. <https://doi.org/10.1016/j.fcr.2022.108585>
26. Hao, S., Lou, H., Wang, H., Shi, J., Liu, D., Baogerile, ... Xin, M. (2022). Genome-Wide Association Study reveals the genetic basis of five quality traits in Chinese wheat. *Frontiers in Plant Science*, 13, 835306. <https://doi.org/10.3389/fpls.2022.835306>
27. Hassaneian Khoshro, H., & Abdi, H. (2023). Application of Marker-Assisted Selection in Wheat Quality Breeding Program. IntechOpen. doi: 10.5772/intechopen.113944.

28. He, F., Pasam, R., Shi, F., Kant, S., Keeble-Gagnere, G., Kay, P., Forrest, K., Fritz, A., Hucl, P., Wiebe, K., Knox, R., Cuthbert, R., Pozniak, C., Akhunova, A., Morrell, P. L., Davies, J. P., Webb, S. R., Spangenberg, G., Hayes, B., ... Akhunov, E. (2019). Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome. *Nature Genetics*, 51(6), 896–904. <https://doi.org/10.1038/s41588-019-0382-2>
29. Holland, D., Wang, Y., Thompson, W. K., Schork, A., Chen, C. H., Lo, M. T., Witoelar, A., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Enhancing Neuro Imaging Genetics through Meta Analysis Consortium, Werge, T., O'Donovan, M., Andreassen, O. A., & Dale, A. M. (2016). Estimating effect sizes and expected replication probabilities from GWAS summary statistics. *Frontiers in Genetics*, 7, 15. <https://doi.org/10.3389/fgene.2016.00015>.
30. Hu, P., Zheng, Q., Luo, Q., et al. (2021). Genome-wide association study of yield and related traits in common wheat under salt-stress conditions. *BMC Plant Biology*, 21, 27. <https://doi.org/10.1186/s12870-020-02799-1>
31. Huang, M., Liu, X., Zhou, Y., Summers, R. M., & Zhang, Z. (2019). BLINK: A package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience*, 8(2), giy154. <https://doi.org/10.1093/gigascience/giy154>.
32. Huo, N., Zhu, T., Zhang, S., Mohr, T., Luo, M. C., Lee, J. Y., Distelfeld, A., Altenbach, S., & Gu, Y. Q. (2019). Rapid evolution of α -gliadin gene family revealed by analyzing Gli-2 locus regions of wild emmer wheat. *Functional & Integrative Genomics*, 19(6), 993–1005. <https://doi.org/10.1007/s10142-019-00686-z>
33. Huo, N., Zhu, T., Altenbach, S., Dong, L., Wang, Y., Mohr, T., ... Gu, Y. Q. (2018). Dynamic evolution of α -gliadin prolamin gene family in homeologous genomes of hexaploid wheat. *Scientific Reports*, 8(1), 5181. <https://doi.org/10.1038/s41598-018-23570-5>
34. Illumina. (2024, 10 de enero). *Secuenciación de próxima generación (NGS)*. Illumina.com. Obtenido de <https://www.illumina.com/science/technology/next-generation-sequencing.html#:~:text=What%20is%20NGS%3F,regions%20of%20DNA%20or%20RNA>.

35. Irshad, A., Guo, H., Ur Rehman, S., Wang, X., Gu, J., Xiong, H., Xie, Y., Zhao, L., Zhao, S., Wang, C., & Liu, L. (2021). Identification of single nucleotide polymorphism in TaSBEIII and development of KASP marker associated with grain weight in wheat. *Frontiers in Genetics*, 12, 697294. <https://doi.org/10.3389/fgene.2021.697294>
36. Juliana, P., Poland, J., Huerta-Espino, J., Shrestha, S., Crossa, J., Crespo-Herrera, L., ... Singh, R. P. (2019). Improving grain yield, stress resilience and quality of bread wheat using large-scale genomics. *Nature Genetics*, 51(10), 1530–1539. <https://doi.org/10.1038/s41588-019-0496-6>
37. Kaler, A. S., Gillman, J. D., Beissinger, T., & Purcell, L. C. (2020). Comparing different statistical models and multiple testing corrections for association mapping in soybean and maize. *Frontiers in Plant Science*, 10, 1794. <https://doi.org/10.3389/fpls.2019.01794>
38. Karn, A. (2019). *Flow chart of candidate gene analysis post GWAS* [Imagen]. Avikarn. <https://avikarn.com/2019-07-22-GWAS/>
39. Khoshro, H. H., Bihamta, M. R., & Hassani, M. E. (2022). Relationship between allelic variation at the Glu-3 loci and qualitative traits in bread wheat. *Cereal Research Communications*, 50(3), 509-522. <https://doi.org/10.1007/s42976-021-00203-8>
40. Khush, G. (2001). Green revolution: the way forward. *Nature Reviews Genetics*, 2, 815–822. <https://doi.org/10.1038/35093585>
41. Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods*, 9, 29. <https://doi.org/10.1186/1746-4811-9-29>.
42. Krawczyk, B., & Kur, J. (2018). Molecular identification and genotyping of staphylococci: Genus, species, strains, clones, lineages, and interspecies exchanges. In V. Savini (Ed.), *Pet-to-man travelling staphylococci* (pp. 199-223). Academic Press. <https://doi.org/10.1016/B978-0-12-813547-1.00016-9>
43. Kumar, J., Saripalli, G., Gahlaut, V., Goel, N., Meher, P. K., Mishra, K. K., Mishra, P. C., Sehgal, D., Vikram, P., Sansaloni, C., Singh, S., Sharma, P. K., & Gupta, P. K. (2018). Genetics of Fe, Zn, β -carotene, GPC and yield traits in bread wheat (*Triticum aestivum* L.) using multi-locus and multi-traits GWAS. *Euphytica*, 214(11), 219. <https://doi.org/10.1007/s10681-018-2284-2>
44. Kumar, K., Jan, I., Saripalli, G., Sharma, P. K., Mir, R. R., Balyan, H. S., & Gupta, P. K. (2022). An update on resistance genes and their use in the development of

- leaf rust resistant cultivars in wheat. *Frontiers in Genetics*, 13, Article 816057. <https://doi.org/10.3389/fgene.2022.816057>
45. Laido, G., Marone, D., Russo, M. A., Colecchia, S. A., Mastrangelo, A. M., De Vita, P., & Papa, R. (2014). Linkage disequilibrium and genome-wide association mapping in tetraploid wheat (*Triticum turgidum* L.). *PLoS One*, 9(4), e95211. <https://doi.org/10.1371/journal.pone.0095211>
46. Leonova, I. N., Kiseleva, A. A., Berezhnaya, A. A., Orlovskaya, O. A., & Salina, E. A. (2023). Novel genetic loci from *Triticum timopheevii* associated with gluten content revealed by GWAS in wheat breeding lines. *International Journal of Molecular Sciences*, 24(17), 13304. <https://doi.org/10.3390/ijms241713304>
47. Lewontin, R. C. (1988). On measures of gametic disequilibrium. *Genetics*, 120(3), 849–852. <https://doi.org/10.1093/genetics/120.3.849>
48. Li, N., Miao, Y., Ma, J., Zhang, P., Chen, T., Liu, Y., ... Yang, D. (2023). Consensus genomic regions for grain quality traits in wheat revealed by Meta-QTL analysis and in silico transcriptome integration. *The Plant Genome*, 16(2), e20336. <https://doi.org/10.1002/tpg2.20336>
49. Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., Gore, M. A., Buckler, E. S., & Zhang, Z. (2012). GAPIT: Genome association and prediction integrated tool. *Bioinformatics*, 28(18), 2397–2399. <https://doi.org/10.1093/bioinformatics/bts444>.
50. Liu, X., Huang, M., Fan, B., Buckler, E. S., & Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genetics*, 12(2), e1005767. <https://doi.org/10.1371/journal.pgen.1005767>.
51. Luján Basile, S. M., Ramírez, I. A., Crescente, J. M., Conde, M. B., Demichelis, M., Abbate, P., Rogers, W. J., Pontaroli, A. C., Helguera, M., & Vanzetti, L. S. (2019). Haplotype block analysis of an Argentinean hexaploid wheat collection and GWAS for yield components and adaptation. *BMC Plant Biology*, 19(1), 553. <https://doi.org/10.1186/s12870-019-2015-4>.
52. Mahboubi, M., Mehrabi, R., Najj, A. M., & et al. (2020). Whole-genome diversity, population structure and linkage disequilibrium analysis of globally diverse wheat genotypes using genotyping-by-sequencing DArTseq platform. *3 Biotech*, 10(48). <https://doi.org/10.1007/s13205-019-2014-z>

53. Manifesto, M. M., Feingold, S., Hopp, H. E., Schlatter, A. R., & Dubcovsky, J. (1998). Molecular markers associated with differences in bread-making quality in a cross between bread wheat cultivars with the same high Mr glutenins. *Journal of Cereal Science*, 27(3), 217–227. [https://doi.org/10.1006/s0733-5210\(97\)91000-6](https://doi.org/10.1006/s0733-5210(97)91000-6)
54. Martin, F. J., Amode, M. R., Aneja, A., Austine-Orimoloye, O., Azov, A. G., Barnes, I., Becker, A., Bennett, R., Berry, A., Bhai, J., Bhurji, S. K., Bignell, A., Boddu, S., Branco Lins, P. R., Brooks, L., Budhanuru Ramaraju, S., Charkhchi, M., Cockburn, A., Da Rin Fiorretto, L., Davidson, C., Dodiya, K., Donaldson, S., El Houdaigui, B., El Naboulsi, T., Fatima, R., Garcia Giron, C., Genez, T., Ghattaoraya, G. S., Gonzalez Martinez, J., Guijarro, C., Hardy, M., Hollis, Z., Hourlier, T., Hunt, T., Kay, M., Kaykala, V., Le, T., Lemos, D., Marques-Coelho, D., Marugán, J. C., Merino, G. A., Mirabueno, L. P., Mushtaq, A., Hossain, S. N., Ogeh, D. N., Pandian Sakthivel, M., Parker, A., Perry, M., Piližota, I., Prosovskaia, I., Pérez-Silva, J. G., Abdul Salam, A. I., Saraiva-Agostinho, N., Schuilenburg, H., Sheppard, D., Sinha, S., Sipos, B., Stark, W., Steed, E., Sukumaran, R., Sumathipala, D., Suner, M. M., Surapaneni, L., Sutinen, K., Szpak, M., Tricomi, F. F., Urbina-Gómez, D., Veidenberg, A., Walsh, T. A., Walts, B., Wass, E., Willhoft, N., Allen, J., Alvarez-Jarreta, J., Chakiachvili, M., Flint, B., Giorgetti, S., Haggerty, L., Ilsley, G. R., Loveland, J. E., Moore, B., Mudge, J. M., Tate, J., Thybert, D., Trevanion, S. J., Winterbottom, A., Frankish, A., Hunt, S. E., Ruffier, M., Cunningham, F., Dyer, S., Finn, R. D., Howe, K. L., Harrison, P. W., Yates, A. D., & Fliceck, P. (2023). Ensembl 2023. *Nucleic Acids Research*, 51(D1), D933–D941. <https://doi.org/10.1093/nar/gkac958>.
55. Miguélez, P. A. (2021, 31 de mayo). *Marcadores genéticos*. Ampligen. Obtenido de <https://www.ampligen.es/adn-genetica/marcadores-geneticos/>.
56. Miramontes, J. A., Avila Salazar, J. M., Rivas Santoyo, F. J., & Martínez Heredía, D. (2023). *El cultivo del trigo*. Universidad de Sonora, División de Ciencias Biológicas y de la Salud, Departamento de Agricultura y Ganadería. Obtenido de <https://agricultura.unison.mx/memorias%20de%20maestros/EL%20CULTIVO%20DEL%20TRIGO.pdf>

57. Peña, R. J., Amaya, A., Rajaram, S., & Mujeeb-Kazi, A. (1990). Variation in quality characteristics associated with some spring 1B/1R translocation wheats. *Journal of Cereal Science*, 12(2), 105-112. [https://doi.org/10.1016/S0733-5210\(09\)80092-1](https://doi.org/10.1016/S0733-5210(09)80092-1)
58. Peña, R., Hernandez-Espinosa, N., Perez Herrera, P., Villaseñor Mir, H. E., Gomez Valdez, M. M., & Mendoza Lozano, M. A. (2008). *Calidad de la cosecha de trigo en México: Ciclo otoño-invierno 2006-07*. CONATRIGO, CONASIST. <http://hdl.handle.net/10883/1264>
59. Pu, Z., Ye, X., Li, Y., Shi, B., Guo, Z., Dai, S., Ma, J., Liu, Z., Jiang, Y., Li, W., Jiang, Q., Chen, G., Wei, Y., & Zheng, Y. (2022). Identification and validation of novel loci associated with wheat quality through a genome-wide association study. *Journal of Integrative Agriculture*, 21(11), 3131-3147. <https://doi.org/10.1016/j.jia.2022.08.085>
60. Qian, L., Hickey, L. T., Stahl, A., Werner, C. R., Hayes, B., Snowdon, R. J., & Voss-Fels, K. P. (2017). Exploring and harnessing haplotype diversity to improve yield stability in crops. *Frontiers in Plant Science*, 8, 1534. <https://doi.org/10.3389/fpls.2017.01534>
61. R Core Team. (2021). *R: A language and environment for statistical computing* (Version 4.1.2) [Software]. R Foundation for Statistical Computing. <https://cran.r-project.org/bin/windows/base/old/4.1.2/>
62. Rahimi, Y., Khahani, B., Jamali, A., Alipour, H., Bihamta, M. R., & Ingvarsson, P. K. (2023). Genome-wide association study to identify genomic loci associated with early vigor in bread wheat under simulated water deficit complemented with quantitative trait loci meta-analysis. *G3 Genes|Genomes|Genetics*, 13(2). <https://doi.org/10.1093/g3journal/jkac320>
63. Rauf, K., Rahman, R., Saeed, A., Ali, M., Noureen, F., Amir, R., & Gul, A. (2020). Next-generation sequencing in bread wheat. En M. Ozturk & A. Gul (Eds.), *Climate Change and Food Security with Emphasis on Wheat* (pp. 311-320). Academic Press. <https://doi.org/10.1016/B978-0-12-819527-7.00022-4>
64. Rico, S., & Dahm, M. (2021). *La ciencia detrás de la harina y la pasta perfectas* [Imagen]. Laboratorio de Calidad del Trigo del CIMMYT. <https://www.cimmyt.org/es/multimedia/la-ciencia-detras-de-la-harina-y-la-pasta-perfectas/>

65. Roncallo, P. F. (2010). *Mapeo de regiones genómicas asociadas a contenido de pigmentos carotenoides y color amarillo en trigo candeal* [Tesis de maestría, Universidad Nacional del Sur]. Repositorio Digital Institucional de la Biblioteca Central "Profesor Nicolás Matijevic" de la Universidad Nacional del Sur. https://repositoriodigital.uns.edu.ar/bitstream/handle/123456789/2093/ROCA_LLO-Introduccion-Hipotesis-Objetivos.pdf?sequence=1&isAllowed=y#:~:text=La%20especie%20cultivada%20de%20Triticum,2n%3D6x%3D42
66. Sandhu, K. S., Mihalyov, P. D., Lewien, M. J., Pumphrey, M. O., & Carter, A. H. (2021). Genomic selection and genome-wide association studies for grain protein content stability in a nested association mapping population of wheat. *Agronomy*, 11(2528). <https://doi.org/10.3390/agronomy11122528>
67. Sandhu, K., Burke, A., Merrick, L., Pumphrey, M., & Carter, A. (2024). Comparing performances of different statistical models and multiple threshold methods in a nested association mapping population of wheat. *Frontiers in Plant Science*, 15. <https://doi.org/10.3389/fpls.2024.1460353>
68. Sansaloni, C., Franco, J., Santos, B., Percival-Alwyn, L., Singh, S., Petrolí, C., Campos, J., Dreher, K., Payne, T., Marshall, D., Kilian, B., Milne, I., Raubach, S., Shaw, P., Stephen, G., Carling, J., Pierre, C. S., Burgueño, J., Crosa, J., Li, H., ... Pixley, K. (2020). Diversity analysis of 80,000 wheat accessions reveals consequences and opportunities of selection footprints. *Nature Communications*, 11(1), 4572. <https://doi.org/10.1038/s41467-020-18404-w>
69. Sansaloni, C., Petrolí, C., Jaccoud, D., Carling, J., Detering, F., Grattapaglia, D., & Kilian, A. (2011). Diversity Arrays Technology (DART) and next-generation sequencing combined: Genome-wide, high throughput, highly informative genotyping for molecular breeding of *Eucalyptus*. *BMC Proceedings*, 5(Suppl 7), P54. <https://doi.org/10.1186/1753-6561-5-S7-P54>.
70. Secretaría de Agricultura y Desarrollo Rural (SADER). (2022, 11 de octubre). *Cultivo de trigo en México: Todo lo que debes de saber*. Gobierno de México. <https://www.gob.mx/agricultura/articulos/cultivo-de-trigo-en-mexico-todo-lo-que-debes-de-saber#:~:text=El%20trigo%20es%20una%20planta,pueden%20ser%20silvestres%20o%20cultivadas>.

71. Sen, S. K. (2023, 12 de septiembre). *Restriction fragment length polymorphism (RFLP)*. Instituto Nacional de Investigación del Genoma Humano. <https://www.genome.gov/genetics-glossary/Restriction-Fragment-Length-Polymorphism>.
72. Severe Covid-19 GWAS Group, Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., Invernizzi, P., Fernández, J., Prati, D., Baselli, G., Asselta, R., Grimsrud, M. M., Milani, C., Aziz, F., Kässens, J., May, S., Wendorff, M., Wienbrandt, L., Uellendahl-Werth, F., Zheng, T., ... Karlsen, T. H. (2020). Genomewide association study of severe Covid-19 with respiratory failure. *The New England Journal of Medicine*, 383(16), 1522–1534. <https://doi.org/10.1056/NEJMoa2020283>.
73. Shokat, S., Sehgal, D., Liu, F., & Singh, S. (2020). GWAS Analysis of Wheat Pre-breeding Germplasm for Terminal Drought Stress Using Next Generation Sequencing Technology. Preprints 2020. <https://doi.org/10.20944/preprints202002.0272.v1>
74. Subijana, A. G. (2017). *Análisis de modelos estadísticos para estudios de asociación del genoma completo* (Trabajo Fin de Grado, Grado en Matemáticas). Universidad del País Vasco. Obtenido de https://addi.ehu.es/bitstream/10810/20111/1/TFG_GorostizaLopezDeSubijanaAnia.pdf.
75. Sukumaran, S., & Yu, J. (2014). Association mapping of genetic resources: Achievements and future perspectives. En *Advances in Agronomy* (pp. 207-235). https://doi.org/10.1007/978-94-007-7572-5_9
76. Shewry, P. (2019). What is gluten—Why is it special? *Frontiers in Nutrition*, 6. <https://doi.org/10.3389/fnut.2019.00101>.
77. Takahashi-Kariyazono, S., Sakai, K., & Terai, Y. (2020). Presence–absence polymorphisms of single-copy genes in the stony coral *Acropora digitifera*. *BMC Genomics*, 21(1), 158. <https://doi.org/10.1186/s12864-020-6566-4>.
78. Valdez Velázquez, G. (2021). *Genotipificación por Secuenciación: producción de datos primarios (perfiles genómicos de marcadores moleculares - DArT-Seq) a partir de muestras de ADN* [Ilustración]. Creación propia de Guadalupe Valdez Velázquez. Uso exclusivo de SAGA.

79. Villaseñor Mir, H. E., Huerta Espino, J., Hortelano Santa Rosa, R., Martínez Cruz, E., Rodríguez García, M. F., & Alvarado Padilla, J. I. (2020). Conatrigo F2015: Nueva variedad de trigo harinero para zonas de riego en México. *Revista Mexicana de Ciencias Agrícolas*, 11(8), 1993-1997. <https://doi.org/10.29312/remexca.v11i8.2616>.
80. Wang, P., Jin, Z., & Xu, X. (2015). Physicochemical alterations of wheat gluten proteins upon dough formation and frozen storage – A review from gluten, glutenin and gliadin perspectives. *Trends in Food Science & Technology*, 46(2, Part A), 189-198. <https://doi.org/10.1016/j.tifs.2015.10.005>
81. Wani, S. H., Gaikwad, K., Razzaq, A., Samantara, K., Kumar, M., & Govindan, V. (2022). Improving zinc and iron biofortification in wheat through genomics approaches. *Molecular Biology Reports*, 49(8), 8007–8023. <https://doi.org/10.1007/s11033-022-07326-z>
82. Weir, B. (1996). Genetic Data Analysis II. Sinauer Associates. Sunderland MA. USA, 321-327.
83. Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* (2nd ed., pp. 189-201). Use R! Series. Springer, New York, NY, USA. <https://doi.org/10.1007/978-3-319-24277-4>.
84. Wrigley, C., Bekes, F., & Bushuk, W. (2006). Gluten: A balance of gliadin and glutenin. En **C. Wrigley, F. Bekes, & W. Bushuk** (Eds.), *Chapter 1* (pp. 3-32). ISBN: 978-1-891127-51-9. <https://doi.org/10.1094/9781891127519.002>.
85. Yang, Y., Chai, Y., Zhang, X., Lu, S., Zhao, Z., Wei, D., Chen, L., & Hu, Y.-G. (2020). Multi-locus GWAS of quality traits in bread wheat: Mining more candidate genes and possible regulatory network. *Frontiers in Plant Science*, 11. <https://doi.org/10.3389/fpls.2020.01091>.
86. Yoosefzadeh-Najafabadi, M., Eskandari, M., Belzile, F., & Torkamaneh, D. (2022). Genome-Wide Association Study Statistical Models: A Review. In D. Torkamaneh & F. Belzile (Eds.), *Genome-Wide Association Studies. Methods in Molecular Biology* (Vol. 2481). Humana, New York, NY. https://doi.org/10.1007/978-1-0716-2237-7_4
87. Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D. M., Holland, J. B., Kresovich, S., & Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for

multiple levels of relatedness. *Nature Genetics*, 38(2), 203–208.
<https://doi.org/10.1038/ng1702>.

88. Zhu, T., Wang, L., Rimbart, H., Rodriguez, J. C., Deal, K. R., De Oliveira, R., Choulet, F., Keeble-Gagnère, G., Tibbits, J., Rogers, J., Eversole, K., Appels, R., Gu, Y. Q., Mascher, M., Dvorak, J., & Luo, M. C. (2021). Optical maps refine the bread wheat *Triticum aestivum* cv. Chinese Spring genome assembly. *The Plant Journal: For Cell and Molecular Biology*, 107(1), 303–314.
<https://doi.org/10.1111/tpj.15289>.

Lista de figuras

No. Figuras	Pagina
<p>Figura 1. Genotipificación por Secuenciación: producción de datos primarios (perfiles genómicos de marcadores moleculares - DArT-Seq) a partir de muestras de ADN: a) Digestión de ADN con enzimas de restricción: Fragmentación del ADN mediante el uso de enzimas de restricción específicas, b) Amplificación de los targets mediante PCR: Multiplicación selectiva de regiones de interés del ADN a través de la técnica de PCR, c) Pool de targets y cuantificación de la librería: Mezcla y cuantificación de los fragmentos amplificados para la creación de una librería genética. d) Secuenciación en plataforma NGS: Secuenciación de la librería mediante tecnología del secuenciador <i>NovaSeq 6000</i>. e) Alineamiento contra genoma de referencia: Comparación de las secuencias obtenidas con un genoma de referencia para identificar variaciones. f) Extracción de datos marcadores SNP's y PAVs: Identificación de polimorfismos de nucleótido simple (SNP's) y variaciones de presencia/ausencia (PAVs) en los datos genómicos (Valdez, 2021).</p>	14
<p>Figura 2. Diagrama del proceso de análisis genómico utilizando GWAS; a) Recolección de los datos de la caracterización fisicoquímica y genotípica de las muestras, b) Procesamiento de los datos mediante el software GAPIT para realizar estudios de asociación genómica, c) Identificación de SNP's significativos mediante un gráfico Manhattan, resaltando las variantes asociadas a los rasgos de interés, d) Localización y análisis de genes candidatos cercanos a los SNP's identificados, e) Evaluación del desequilibrio de ligamiento (LD) en la región genómica asociada para determinar la relación genética entre los SNP's detectados (Karn, 2019; Rico & Dahm, 2022).</p>	17
<p>Figura 3. Correlograma de GP y SDS_Sedim</p>	32
<p>Figura 4. Gráfico PCA de 1.602 colectas de <i>T. aestivum</i> nativos de diferentes países a) PC1 vs PC2, b) PC1 vs PC3 y c) PC2 vs PC3.</p>	33
<p>Figura 5. Gráficos de Manhattan del modelo BLINK que muestran la asociación de 77,395 marcadores SNP's en 1,602 colectas de <i>T. aestivum</i> nativos para cada característica evaluada.</p>	35

Figura 6. Gráficos de Manhattan del modelo FarmCPU que muestran la asociación de 77,395 marcadores SNP's en 1,602 colectas de <i>T. aestivum</i> nativos para cada característica evaluada.	36
Figura 7. Gráfico de Manhattan del modelo MLM que muestra la asociación de 77,395 marcadores SNP's en 1,602 colectas de <i>T. aestivum</i> nativos solo para GP.	37
Figura 8. Gráficos de Manhattan circular de los tres modelos BLINK, FarmCPU y MLM que muestran la asociación de 77,395 marcadores SNP's en 1,602 colectas de <i>T. aestivum</i> nativos a) asociación de GP con un par de SNP que coincidieron en los modelos, para el caso del cromosoma 5B (14) coincidieron en FarmCPU y BLINK y en el cromosoma 7D (21) coincidieron en los modelos MLM y FarmCPU, b) Asociación de SDS_Sedim con un par de SNP que concordaron en los modelos de FarmCPU y BLINK, en los cromosomas 1B (2) y 5A (13).	38
Figura 9. Gráficos Q-Q para GP de los tres modelos en el panel GWAS, b) Gráficos Q-Q para SDS_Sedim de los tres modelos en el panel GWAS	39
Figura 10. Los diagramas de caja indican los valores de la caracterización fisicoquímica correspondientes a los distintos grupos de haplotipos diferentes a) para GP en el cromosoma 1A y el haplotipo "GCG" esta significativamente alto a comparación de los demás y b) SDS en el cromosoma 1B	46
Figura 11. Los diagramas de caja indican los valores de la caracterización fisicoquímica correspondientes a los distintos grupos de haplotipos diferentes a) GP en el cromosoma 6A, b) GP en el cromosoma 6B, b) GP en el cromosoma 6D no hay diferencia relévate y d) SDS en el cromosoma 6B.	47

Lista de tablas

No. Tablas	Pagina
Tabla 1. Numero de marcadores eliminados por cada uno de los filtros.	26
Tabla 2. Resultados GWAS de marcadores significativos pertenecientes a GP en 1,602 colectas de T. aestivum nativos.	41
Tabla 3. Resultados GWAS de marcadores significativos pertenecientes a SDS_Sedim en 1,602 colectas de T. aestivum.	42

Glosario de términos y siglas

ADN	Ácido Desoxirribonucleico
ARN	Ácido Ribonucleico
ARNm	Ácido Ribonucleico Mensajero
ncARN	Ácido Ribonucleico no codificante
SNP	Polimorfismo de un Solo Nucleótido
PAP	Polimorfismos de Presencia-Ausencia
NGS	Secuenciación de Nueva Generación
SBS	Secuenciación por Síntesis
GP	Proteína en Grano
SDS	Sedimentación con Dodecilsulfato de Sodio
BLAST	Herramienta Básica de Búsqueda de Alineamiento Local
GWAS	Estudios de Asociación de Todo el Genoma
MLM	Modelo Lineal Mixto
BLINK	Clave Anidada Iterativa de Desequilibrio de Vinculación y Evaluación Bayesiana
FarmCPU	Unificación de Probabilidad Circulante de Modelo Fijo y Aleatorio
GAPIT	Herramienta Integrada de Predicción y Asociación del Genoma
LD	Desequilibrio de Ligamiento
FDR	Tasa de Falso Descubrimiento
OR	Odds Ratio
PCA	Análisis de Componentes Principales
cM	Centimorgan

Anexos

Anexo 1. Información bioinformática de bloques de asociación GP.

BA	ID estable del gen EnsemblPlants	ID	Dominio	UniProt	Descripción del dominio
1	TraesCS1A03G0156300	A0A3B5XUP6	DUF4005	-	Este dominio se encuentra hacia el extremo C de varias proteínas que contienen el dominio IQ de plantas. Estas proteínas pueden estar involucradas en interacciones cooperativas con calmodulinas o proteínas similares a la calmodulina, y pueden asociarse con ácidos nucleicos y regular la expresión génica a nivel transcripcional o postranscripcional.
	TraesCS1A03G0156700	A0A3B5XUQ4	-	-	-
		A0A3B5XUY0	RRM	Unión de ARN	Unión a una molécula de ARN o una porción de la misma
	A0A3B5XV03				
	TraesCS1A03G0157100	A0A3B5XVZ3			
	TraesCS1A03G0157600	A0A3B5XUQ1			
	TraesCS1A03G0157800	A0A3B5XUQ9			
TraesCS1A03G0157900	A0A3B5XUY2				
TraesCS1A03G0157900	A0A3B5XV08				
TraesCS1A03G0158400	A0A3B5XVZ7	DUF295	-	Este dominio se encuentra en proteínas vegetales, muchas de las cuales también contienen un dominio F-box. Este dominio se puede encontrar en KIB3 de Arabidopsis. KIB3 es una ubiquitina ligasa F-box E3 que promueve la degradación de la quinasa BIN2 similar a GSK3 mientras bloquea el acceso a su sustrato.	
2	TraesCS2A03G0234900	-	-	-	-
	TraesCS2A03G0235400	A0A3B6AS22	CARM1	metilación	-
		A0A3B6ASM3			-
	TraesCS2A03G0235500	A0A3B6ATV5	-	señalización mediada por calcio	-
	TraesCS2A03G0235600	W5BKZ9	NIP7	biogénesis de subunidades grandes ribosómicas	-
	TraesCS2A03G0235700	A0A3B6AS27	-	fotosíntesis	-
		W5C572	-		-
	TraesCS2A03G0235900	A0A3B6ASM9	-	regulación del desarrollo de las flores	-
TraesCS2A03G0236000	A0A3B6ATW0	-	-	-	
TraesCS2A03G0236100	A0A3B6AT09	PRORP	-	-	
3	TraesCS2A03G1170200	A0A3B6B6D1	DUF6598	-	Este supuesto dominio se encuentra en una o más copias en varias proteínas vegetales. Se desconoce la función del dominio.
	TraesCS2A03G1175400	A0A3B6B712	P450	Actividad oxidoreductasa que actúa sobre donantes emparejados, con incorporación o reducción de oxígeno molecular.	-
	TraesCS2A03G1173700	A0A3B6B706	-	actividad endopeptidasa tipo serina	-
	TraesCS2A03G1171800	A0A3B6B7R3	-	Unión de ATP	-
	TraesCS2A03G1175500	A0A3B6B5X5	-	sacarosa: actividad simportadora de protones	-
	TraesCS2A03G1176600	A0A3B6B6G4	-	Actividad del transportador transmembrana de iones de zinc.	-
TraesCS2A03G1172300	A0A3B6B6R7	RPM1	respuesta de defensa	-	

BA = Bloque de asociación, ID = Identificación.

	TraesCS2A03G1171700	A0A3B6B6E2	-	Unión de ATP	-
4	TraesCS2B03G0448600	A0A3B6C4T3	DYW	unión de iones de zinc	-
	TraesCS2B03G0448900	A0A3B6C2I9	-	unión al ADN	-
	TraesCS2B03G0449100	A0A3B6C296	P450	unión de iones de hierro	-
	TraesCS2B03G0449300	A0A3B6C269	LEA-2	respuesta de defensa a otro organismo	Se expresan diferentes tipos de proteínas LEA en diferentes etapas de la embriogénesis tardía en embriones de semillas de plantas superiores y en condiciones de estrés por deshidratación.
	TraesCS2B03G0449400	A0A3B6C278		-	
	TraesCS2B03G0449500	A0A3B6C4T8	J	-	desempeña muchas funciones diversas en la célula, incluido el plegamiento de proteínas nascentes, la translocación de polipéptidos a través de las membranas de los orgánulos, la coordinación de respuestas al estrés y la selección de proteínas para su degradación.
	TraesCS2B03G0449600	A0A3B6C2J3	-	Encuadernación de trampa	-
	TraesCS2B03G0449700	A0A3B6C2A1	SAP	unión de ácido nucleico	El motivo SAP se encuentra en una variedad de proteínas nucleares involucradas en la transcripción, la reparación del ADN, el procesamiento del ARN o la degradación apoptótica de la cromatina.
	TraesCS2B03G0449800	A0A3B6C273	C-terminal	unión de ARNm	el dominio C-terminal de la proteína de la subunidad ribosómica grande bL12 (anteriormente conocida como L7/L12). bL12 está presente en cada subunidad 50S en cuatro copias organizadas como dos dímeros. El complejo proteico L8 que consta de dos dímeros de bL12 y L10 en los ribosomas de Escherichia coli se ensambla en la región conservada del ARNr 23 S denominada dominio asociado a GTPasa.
	TraesCS2B03G0450100	A0A3B6C283	Tim44	unión de chaperonas de plegamiento de proteínas	Tim44 es un componente esencial de la maquinaria que media la translocación de proteínas codificadas en el núcleo a través de la membrana interna mitocondrial. Se cree que Tim44 se une a los fosfolípidos de la membrana interna mitocondrial mediante interacciones electrostáticas y mediante la penetración en la región del grupo de la cabeza polar
A0A3B6C4U3		transporte de proteínas intracelulares			
5	TraesCS2B03G0456800	A0A3B6C2N0	serpina	actividad inhibidora de la endopeptidasa tipo serina	La estructura es un pliegue multidominio que contiene un haz de hélices y un sándwich beta.
	TraesCS2B03G0456900	A0A2X0SLI3	-	constituyente estructural del ribosoma	-
	TraesCS2B03G0457100	A0A3B6C2B4	proteína quinasa	transducción de señales intracelulares	Las proteínas quinasas eucariotas son enzimas que pertenecen a una familia muy extensa de proteínas que comparten un núcleo catalítico conservado común a las proteínas quinasas de serina/treonina y tirosina
		A0A3B6C2C5		actividad de la proteína serina/treonina quinasa	
		A0A3B6C4Y3			
	TraesCS2B03G0457300	A0A3B6C2N4		unión de iones de zinc	-
	TraesCS2B03G0457500	A0A3B6C2E6	enoil reductasa (ER)	Actividad de alcohol deshidrogenasa, dependiente de zinc.	-
	TraesCS2B03G0457600	A0A3B6C2B8	c oxidasa	-	-
	TraesCS2B03G0457700	-	-	-	-
	TraesCS2B03G0458300	A0A3B6C4Y8	-	actividad helicasa	-
6	TraesCS2D03G0800300	A0A3B6DF45	-	respuesta de defensa	-
	TraesCS2D03G0796000	A0A3B6DG02	-	actividad de la proteína serina/treonina quinasa	-
		A0A2X0SK24	-		-
	TraesCS2D03G0797300	A0A3B6DG18	-	actividad del transportador transmembrana	-
		A0A3B6DGP3	-		-
		A0A3B6DJ95	-		-

	TraesCS2D03G0796700	W5BGP8	-	actividad del transportador transmembrana de piruvato	-
	TraesCS2D03G0796600	-	-	-	-
7	TraesCS2D03G1271900	A0A2X0SEE6	-	-	-
	TraesCS2D03G1265000	E0WC53	-	-	-
	TraesCS2D03G1270200	A0A3B6DPX4	proteína quinasa	actividad de la proteína serina/treonina quinasa	Las proteínas quinasas eucariotas son enzimas que pertenecen a una familia muy extensa de proteínas que comparten un núcleo catalítico conservado común a las proteínas quinasas de serina/treonina y tirosina
	TraesCS2D03G1264400	A0A1D6RNX2	-	unión de iones metálicos	-
	TraesCS2D03G1264300	A0A1D6RNX3	-	actividad de mevalonato quinasa	-
	TraesCS2D03G1268800	A0A3B6DRV1	SART-1	Empalme cis de ARNm, a través de espliceosoma	-
	TraesCS2D03G1266400	A0A3B6DNH7	DUF674	-	-
	TraesCS2D03G1264800	A0A3B6DSQ2	-	Actividad 1,3-beta-D-glucano sintasa	-
		A0A2X0S216	-		-
		TraesCS2D03G1264500	A0A2X0S2B1	-	actividad de la proteína quinasa
8	TraesCS3B03G0999800	A0A3B6FWT7	C-terminal Hs1pro-1	unión del hemo	-
	TraesCS3B03G0998600	A0A3B6FSX6	-	-	-
	TraesCS3B03G0997500	A0A077RXG2	-	Unión de ADN específica de secuencia de la región reguladora cis de la ARN polimerasa II	-
	TraesCS3B03G0997900	-	-	-	-
	TraesCS3B03G0997600	A0A341NXR6	-	-	-
	TraesCS3B03G0998800	A0A3B6FW86	L35A	constituente estructural del ribosoma	-
9	TraesCS3B03G0866900	A0A077S0H0	LEA-2	-	Se expresan diferentes tipos de proteínas LEA en diferentes etapas de la embriogénesis tardía en embriones de semillas de plantas superiores y en condiciones de estrés por deshidratación.
	TraesCS3B03G0869100	-	-	-	-
10	TraesCS3D03G0831500	A0A3B6H2I3	P450	Actividad oxidoreductasa, que actúa sobre donantes emparejados, con incorporación o reducción de oxígeno molecular.	-
	TraesCS3D03G0832500	A0A3B6GZF2	-	-	-
		A0A3B6GXQ1	-	-	-
	TraesCS3D03G0831800	A0A3B6H2I8	-	actividad amidofosforribosiltransferasa	-
	TraesCS3D03G0831000	A0A3B6H232	-	actividad del transportador transmembrana	-
	TraesCS3D03G0832200	A0A3B6GYW0	-	Actividad del transportador transmembrana de iones de magnesio.	-
	TraesCS3D03G0830500	A0A3B6H2I0	-	Actividad exonucleasa de ARN 5'-3'	-
		A0A3B6H227	-		-
	A0A3B6GZE2	-	-		
	A0A3B6GYV2	-	-		

11	TraesCS3D03G0358000	A0A3B6GTW2	-	unión al ADN	-	
	TraesCS3D03G0358100	A0A3B6GUC4	-	-	-	
		A0A3B6GQS1	-	-	-	
	TraesCS3D03G0358500	A0A3B6GRE7	RWP-RK		Actividad del factor de transcripción de unión al ADN.	-
		A0A3B6GPC7				-
		A0A3B6GTW7				-
	TraesCS3D03G0359700	A0A3B6GUC9	-		actividad de la amilopectina maltohidrolasa	-
		A0A3B6GQS6	-			-
TraesCS3D03G0360000	A0A3B6GRF0	-		-	-	
TraesCS3D03G0361400	A0A3B6GPD1	-		-	-	
12	TraesCS4A03G1028900	A0A3B6I156	-	Actividad del ARNt 2'-fosfotransferasa	-	
	TraesCS4A03G1027300	A0A3B6I1K7	N-terminal Rx	-	Esta entrada representa el dominio N-terminal que se encuentra en muchas proteínas de resistencia de las plantas. Se ha predicho que este dominio será una bobina enrollada, sin embargo, la estructura muestra que adopta un pliegue de cuatro haces helicoidales.	
	TraesCS4A03G1031000	-	-	-	-	
	TraesCS4A03G1029100	U5HTD8	S20	constituente estructural del ribosoma	-	
	TraesCS4A03G1029200	A0A3B6I5I5	fosfatasa del ácido fosfatídico tipo 2/haloperoxidasa	actividad fosfatidato fosfatasa	Las enzimas PAP2 tienen una estructura central que consta de un haz de 5 hélices, donde el comienzo de la tercera hélice se une al cofactor. Las enzimas PAP2 catalizan la desfosforilación del fosfatidato, produciendo diacilglicerol y fosfato inorgánico. En las células eucariotas, la actividad de la PAP tiene un papel central en la síntesis de fosfolípidos y triacilglicerol a través de su producto diacilglicerol, y también genera y/o degrada moléculas de señalización de lípidos relacionadas con el fosfatidato.	
	TraesCS4A03G1026600	A0A3B6I4K8	-	-	-	
	TraesCS4A03G1026400	A0A3B6I360	-	-	-	
	TraesCS4A03G1026200	A0A0E3IHU1	-	Actividad del transportador transmembrana de fosfato inorgánico.	-	
TraesCS4A03G1026500	A0A3B6I151	JmjC	Actividad desmetilasa de la histona H3K4me/H3K4me2/H3K4me3	-		
TraesCS4A03G1030900	A0A3B6I161	-	Actividad hidrolasa, que actúa sobre enlaces éster.	-		
13	TraesCS4B03G0317800	-	-	-	-	
	TraesCS4B03G0317900	A0A3B6IR92	-	Actividad oxidorreductasa, que actúa sobre donantes emparejados, con la oxidación de un par de donantes dando como resultado la reducción del oxígeno molecular a dos moléculas de agua.	-	
	TraesCS4B03G0318100	A0A3B6IS17	N-terminal RIX1	-	Rix1 es una partícula nucleoplásmica involucrada en el procesamiento de ARNr/ensamblaje de ribosomas. Se asocia con otras dos proteínas, Ipi1 e Ipi3, para formar el complejo RIX1 que permite que Rea1 (la ATPasa AAA) se asocie con la subunidad ribosómica 60S. Más de 170 factores de ensamblaje participan en la construcción y maduración de los ribosomas de levadura y, una vez que estos factores han	
A0A3B6IPK1		-				
A0A3B6IPK1	-					

					completado su función, deben liberarse de los prerribosomas. Rea1 induce la liberación del complejo proteico de ensamblaje de forma mecánica.
	TraesCS4B03G0318300	A0A3B6IN20	-		-
		A0A3B6IR97	-	unión de cadena ligera de clatrina	-
		A0A3B6IS22	-		-
	TraesCS4B03G0318500	A0A3B6IPK8	EF	unión de iones calcio	Muchas proteínas de unión a calcio pertenecen a la misma familia evolutiva y comparten un tipo de dominio de unión a calcio conocido como EF-mano. Este tipo de dominio consta de un bucle de doce residuos flanqueado en ambos lados por un dominio helicoidal α de doce residuos. En un bucle manual EF, el ion calcio está coordinado en una configuración bipiramidal pentagonal. Los seis residuos implicados en la unión están en las posiciones 1, 3, 5, 7, 9 y 12; estos residuos se indican con X, Y, Z, -Y, -X y -Z. El Glu o Asp invariante en la posición 12 proporciona dos oxígenos para ligar Ca (ligando bidentado). La unidad estructural/funcional básica de las proteínas de la mano EF suele ser un par de motivos de la mano EF que juntos forman un dominio estable de haz de cuatro hélices. El emparejamiento de EF-hand permite la cooperatividad en la unión de iones Ca ²⁺ .
14	TraesCS4B03G0629400	A0A3B6IT40	N-terminal de ciclina	Actividad reguladora de la proteína serina/treonina quinasa dependiente de ciclina	Las ciclinas regulan las quinasas dependientes de ciclinas (CDK). Suiza: P22674 es una uracilo-ADN glicosilasa relacionada con otras ciclinas. Las ciclinas contienen dos dominios de pliegue todo alfa similar, de los cuales esta familia corresponde con el dominio N-terminal.
	TraesCS4B03G0628100	A0A3B6IUVO	proteína quinasa	Unión de ATP	-
	TraesCS4B03G0627500	A0A3B6IT43	C2H2	regulación negativa de la vía de señalización activada por ácido abscísico	Los dominios de 'dedos de zinc' son estructuras proteicas de unión a ácidos nucleicos identificadas por primera vez en el factor de transcripción TFIIIA de Xenopus.
	TraesCS4B03G0629000	A0A3B6IUUV5	fosfatasa relacionada con polifosfato de inositol	actividad fosfatidilinositol-3,4,5-trifosfato 5-fosfatasa	Este dominio se encuentra en diversas proteínas homólogas al inositol monofosfatasa. Estas proteínas son fosfatasas sensibles a Li ⁺ dependientes de Mg ²⁺ que catalizan una variedad de reacciones.
	TraesCS4B03G0627600	A0A3B6IRH3	OBERÓN 3	regulación de la expresión genética	-
	TraesCS4B03G0628300	A0A3B6IVJ7	serpina	actividad inhibidora de la endopeptidasa tipo serina	Las serpinas (inhibidores de la SERina proteínasa) pertenecen a la familia I4 de inhibidores de MEROPS, ID de clan. La mayoría de los miembros de la familia de las serpinas son de hecho inhibidores de la serina proteasa, pero varios tienen funciones de inhibición cruzada adicionales e inhiben a los miembros de la familia de la cisteína proteasa, como las caspasas y las catepsinas.
	TraesCS4B03G0628700	A0A3B6IT55	-	unión de iones metálicos	-
		A0A3B6IRI4	-	unión de hierro férrico	-
15	TraesCS5A03G1016800	A0A3B6KQS2	-	actividad de la ATP sintasa transportadora de protones, mecanismo de rotación	-
	TraesCS5A03G1015700	A0A3B6KRG9	similar a la prolamina	Regulación de la doble fecundación formando cigoto y endospermo.	Prolamin_like (en el que se han fusionado DUF784 y DUF1278) se expresa en el saco embrionario de la planta y está regulado por el factor de transcripción Myb98. El análisis computacional ha revelado que los miembros son homólogos de la superfamilia de prolaminas vegetales (familia LTP, inhibidor de proteasa, almacenamiento de semillas,

					Pfam:PF00234). A diferencia de los miembros típicos de prolamina que tienen ocho residuos Cys conservados que forman cuatro pares de enlaces disulfuro, este dominio contiene sólo seis residuos Cys conservados que pueden formar tres pares de enlaces disulfuro. El dominio puede tener una función potencial en la transferencia o protección de lípidos durante el desarrollo y la reproducción del saco embrionario de la planta.
	TraesCS5A03G1013300	A0A3B6KP19	-	unión al ADN	-
	TraesCS5A03G1013700	W5EUY8	-	Actividad del receptor SNAP	-
	TraesCS5A03G1014600	A0A3B6KQR6	BTB	respuesta al estímulo abiótico	-
		A0A3B6KQR6			-
	TraesCS5A03G1016200	A0A3B6KR12	similar a la prolamina	Regulación de la doble fecundación formando cigoto y endospermo.	Prolamin_like (en el que se han fusionado DUF784 y DUF1278) se expresa en el saco embrionario de la planta y está regulado por el factor de transcripción Myb98. El análisis computacional ha revelado que los miembros son homólogos de la superfamilia de prolaminas vegetales (familia LTP, inhibidor de proteasa, almacenamiento de semillas, Pfam:PF00234). A diferencia de los miembros típicos de prolamina que tienen ocho residuos Cys conservados que forman cuatro pares de enlaces disulfuro, este dominio contiene sólo seis residuos Cys conservados que pueden formar tres pares de enlaces disulfuro. El dominio puede tener una función potencial en la transferencia o protección de lípidos durante el desarrollo y la reproducción del saco embrionario de la planta.
	TraesCS5A03G1015300	A0A3B6KQR7	-	unión de iones de zinc	-
	TraesCS5A03G1018000	A0A3B6KP46	YTH	Unión de ARN que contiene N6-metiladenosina	-
		A0A3B6KNE4			-
16	TraesCS5B03G0398400	A0A3B6LK26	C-terminal Tubby	-	-
	TraesCS5B03G0398700	A0A3B6LJN6		-	-
	TraesCS5B03G0398800	A0A3B6LJT9	-	actividad adenililulfatasa	-
	TraesCS5B03G0398900	A0A3B6LI50	-	-	-
	TraesCS5B03G0399100	A0A3B6LJQ0	-	-	-
	TraesCS5B03G0399400	A0A3B6LK30	proteína quinasa	actividad de la proteína serina/treonina quinasa	Las proteínas quinasa eucariotas son enzimas que pertenecen a una familia muy extensa de proteínas que comparten un núcleo catalítico conservado común a las proteínas quinasa de serina/treonina y tirosina
		A0A3B6LJP1		Actividad de la proteína serina/treonina quinasa dependiente de ciclina	
	TraesCS5B03G0399700	A0A3B6LJU5	-	-	-
	TraesCS5B03G0400000	A0A3B6LI57	LEA-2	-	Se expresan diferentes tipos de proteínas LEA en diferentes etapas de la embriogénesis tardía en embriones de semillas de plantas superiores y en condiciones de estrés por deshidratación.
	TraesCS5B03G0400100	A0A3B6LJQ5	-	-	-
17	TraesCS5D03G0201700	A0A3B6MMM04	-	Actividad GDP-Man:Man3GlcNAc2-PP-Dol alfa-1,2-manosiltransferasa	-

	TraesCS5D03G0201800	A0A3B6MLP4	AP2/ERF	Actividad del factor de transcripción de unión al ADN.	El etileno es una hormona vegetal endógena que influye en muchos aspectos del crecimiento y desarrollo de las plantas. Se descubrió que algunos genes relacionados con la defensa inducidos por el etileno contienen un elemento regulador cis conocido como elemento sensible al etileno (ERE). El análisis de secuencia en varias regiones ERE identificó un motivo corto rico en nucleótidos G/C, la caja GCC, esencial para la respuesta al etileno. Este breve motivo es reconocido por una familia de factores de transcripción, los factores de unión ERE (ERF).
	TraesCS5D03G0202100	A0A3B6MKG1	-	actividad del canal catiónico activado por calcio	-
A0A3B6MN39		-	-		
A0A3B6MLU2		CSC1/OSCA1	-		
	TraesCS5D03G0202200	A0A3B6MM09	amidasa	-	Las enzimas de firma amidasa (AS) son un gran grupo de enzimas hidrolíticas que contienen un tramo conservado de aproximadamente 130 aminoácidos conocido como secuencia AS. Están muy extendidos y se encuentran tanto en procariotas como en eucariotas. Las enzimas AS catalizan la hidrólisis de los enlaces amida (CO-NH ₂), aunque la familia ha divergido ampliamente con respecto a la especificidad y función del sustrato. No obstante, estas enzimas mantienen una estructura central $\alpha/\beta/\alpha$, donde las topologías de las mitades N y C-terminales son similares. Las enzimas AS se caracterizan por tener una región C-terminal altamente conservada, rica en residuos de serina y glicina, pero desprovista de residuos de ácido aspártico e histidina, por lo que difieren de la serina hidrolasas clásicas. Estas enzimas poseen una tríada catalítica Ser-Ser-Lys única y altamente conservada que se utiliza para la hidrólisis de amidas, aunque el mecanismo catalítico para la formación del intermedio acil-enzima puede diferir entre enzimas.
		A0A3B6MLP8	-	-	-
	TraesCS5D03G0202500	A0A3B6MKG5	-	-	-
	TraesCS5D03G0203100	A0A3B6MLU6	PIN	-	-
		A0A3B6MN44		Unión de ARNr de subunidad ribosomal pequeña	-
18	TraesCS5D03G0404900	A0A3B6MPY1	-	Actividad del transportador transmembrana del ion cadmio.	-
		A0A3B6MR61	-	Actividad del transportador transmembrana de iones metálicos.	-
		A0A3B6MQ40	-	-	-
	TraesCS5D03G0405000	A0A3B6MPS7	-	organización del cloroplasto	-
	TraesCS5D03G0405100	A0A3B6MNK2	-	Actividad oxidorreductasa, actuando sobre el grupo CH-OH de los donantes, NAD o NADP como aceptor.	-
		A0A3B6MPY5	-	actividad fosfoglicerato deshidrogenasa	-
		A0A3B6MR68	-	Actividad oxidorreductasa, actuando sobre el grupo CH-OH de los donantes, NAD o NADP como aceptor.	-

	TraesCS5D03G0405500	A0A3B6MQ45	F-box	Proceso catabólico proteico dependiente de ubiquitina proteasomal dependiente de SCF	Este dominio tiene aproximadamente 50 aminoácidos de longitud y generalmente se encuentra en la mitad N-terminal de una variedad de proteínas. Dos motivos que se encuentran comúnmente asociados con el dominio F-box son las repeticiones ricas en leucina (LRR; Pfam:PF00560 y Pfam:PF07723) y la repetición WD (Pfam:PF00400). El dominio F-box tiene un papel en la mediación de las interacciones proteína-proteína en una variedad de contextos, como la poliubiquitinación, el alargamiento de la transcripción, la unión del centrómero y la represión traduccional.
	TraesCS5D03G0405600	A0A3B6MPT0	proteína quinasa	Actividad de la proteína serina/treonina quinasa del receptor transmembrana	Las proteínas quinasas eucariotas son enzimas que pertenecen a una familia muy extensa de proteínas que comparten un núcleo catalítico conservado común a las proteínas quinasas de serina/treonina y tirosina
	TraesCS5D03G0405900	A0A3B6MNL0	-	unión al ADN monocatenario	-
19	TraesCS6A03G0120900	A0A3B6NKS4	F-box	-	Este dominio tiene aproximadamente 50 aminoácidos de longitud y generalmente se encuentra en la mitad N-terminal de una variedad de proteínas. Dos motivos que se encuentran comúnmente asociados con el dominio F-box son las repeticiones ricas en leucina (LRR; Pfam:PF00560 y Pfam:PF07723) y la repetición WD (Pfam:PF00400). El dominio F-box tiene un papel en la mediación de las interacciones proteína-proteína en una variedad de contextos, como la poliubiquitinación, el alargamiento de la transcripción, la unión del centrómero y la represión traduccional.
	TraesCS6A03G0121000	A0A3B6NIG5		-	-
	TraesCS6A03G0121100	A0A3B6NJY4	proteína quinasa	unión de polisacárido	Las proteínas quinasas eucariotas son enzimas que pertenecen a una familia muy extensa de proteínas que comparten un núcleo catalítico conservado común a las proteínas quinasas de serina/treonina y tirosina
	TraesCS6A03G0121500	A0A3B6NKZ5	NB-ARC	Unión ADP	Este es el dominio NB-ARC, un nuevo motivo de señalización que se encuentra en bacterias y eucariotas, compartido por productos genéticos de resistencia de plantas y reguladores de la muerte celular en animales. Este dominio se ha caracterizado estructuralmente en la proteína humana factor 1 activador de proteasa apoptótica (Apaf-1). Contiene el pliegue α - β de tres capas y la posterior región helicoidal α corta característica de la superfamilia del dominio ATPasa AAA+. Si bien se cree que este dominio se une e hidroliza ATP, sólo se ha verificado experimentalmente la unión de ADP. Se propone que la unión e hidrólisis del ATP por este dominio induce cambios conformacionales en la proteína general, lo que conduce a la formación del apoptosoma.
	TraesCS6A03G0121600	A0A3B6NK78	N-terminal Rx	-	Esta entrada representa el dominio N-terminal que se encuentra en muchas proteínas de resistencia de las plantas. Se ha predicho que este dominio será una bobina enrollada, sin embargo, la estructura muestra que adopta un pliegue de cuatro haces helicoidales.
	TraesCS6A03G0121700	A0A3B6NKS9	-	-	-
	TraesCS6A03G0121800	A0A3B6NIG9	Bowman-Birk	-	-
	TraesCS6A03G0122000	A0A077RYQ1	-	constituyente estructural de la cromatina	-
	TraesCS6A03G0122100	A0A3B6NJY8	-	ubiquitinación de proteínas	-

	TraesCS6A03G0122300	A0A077RYQ1	-	actividad de heterodimerización de proteínas	-
20	TraesCS6A03G0892600	A0A3B6NW97	DUF676	-	Esta familia de proteínas son probablemente enzimas de tipo serina esterasa con un pliegue alfa/beta hidrolasa.
	TraesCS6A03G0893300	A0A3B6NWA1	(MFS)	actividad del transportador transmembrana de azúcar	Entre los diferentes transportadores, las dos familias más grandes que se encuentran de manera ubicua en todas las clasificaciones de organismos son la superfamilia de transportadores primarios del casete de unión a ATP (ABC) y la superfamilia de facilitadores principales (MFS). Los transportadores MFS son portadores secundarios de un solo polipéptido capaces sólo de transportar pequeños solutos en respuesta a gradientes de iones quimiosmóticos. Funcionan como uniportadores, simportadores o antiportadores. Además, su especificidad de soluto también es diversa. Las proteínas MFS contienen 12 regiones transmembrana (con algunas variaciones).
	TraesCS6A03G0890300	A0A3B6NVW5	SCP	-	-
	TraesCS6A03G0892900	A0A3B6NVD3	SCP	-	-
	TraesCS6A03G0893600	A0A3B6NTM5	Bowman-Birk	-	-
	TraesCS6A03G0891500	A0A3B6NVX5	RING	-	Varias proteínas eucariotas y virales contienen un dominio conservado rico en cisteína de 40 a 60 residuos (llamado dedo de zinc C3HC4 o dedo 'RING') que une dos átomos de zinc. Hay dos variantes diferentes, el tipo C3HC4 y el tipo C3H2C3, que está claramente relacionado a pesar del diferente patrón de cisteína/histidina. Este último tipo a veces se denomina "dedo RING-H2".
	TraesCS6A03G0888800	A0A3B6NVA8	citocromo c	actividad del transportador transmembrana hemo	Esta familia está formada por varias proteínas implicadas en el ensamblaje del citocromo c de mitocondrias y bacterias; CycK de Rhizobium, CcmC de E. coli y Paracoccus denitrificans y orf240 de las mitocondrias del trigo. Los miembros de esta familia son probablemente proteínas de membrana integrales con seis hélices transmembrana previstas. Se ha propuesto que los miembros de esta familia comprendan un componente de membrana de un complejo transportador ABC (casete de unión de ATP). También se propone que este transportador sea necesario para el transporte de algún componente necesario para el ensamblaje del citocromo c. Un miembro CycK contiene un supuesto motivo de unión al hemo, orf240 también contiene un supuesto motivo de unión al hemo y es un transportador ABC propuesto con hemo de tipo c como sustrato propuesto. Sin embargo, parece poco probable que todos los miembros de esta familia transporten hemo ni apocitocromos tipo C porque CcmC en el supuesto transportador CcmABC no transporta ninguno de los dos. CcmF forma un módulo de trabajo con CcmH y CcmI, CcmFHI, y es poco probable que se una directamente al hemo.
	TraesCS6A03G0890000	A0A3B6NTJ8 A0A3B6NV72 A0A3B6NVW1	proteína quinasa	actividad de la proteína serina/treonina quinasa actividad de la proteína quinasa	Las proteínas quinasas eucariotas son enzimas que pertenecen a una familia muy extensa de proteínas que comparten un núcleo catalítico conservado común a las proteínas quinasas de serina/treonina y tirosina

		A0A3B6NW83	PB1	Unión de ATP	<p> dominio Phox y Bem1p, presentes en muchas proteínas de señalización citoplasmáticas eucariotas. El dominio adopta un pliegue de agarre beta, similar al que se encuentra en los dominios de unión a Ras y ubiquitina. Un motivo, denominado OPR, PC y AID, representa la región más conservada de la mayoría de los dominios PB1 y es necesario para la función del dominio PB1. Esta función es la formación de heterodímeros del dominio PB1, aunque no todos los pares de dominios PB1 se asocian.</p>
21	TraesCS6A03G0897200	A0A3B6NTP3	Dof	-	-
	TraesCS6A03G0898900	A0A3B6NTQ2	similar a ciclina	Actividad reguladora de la proteína serina/treonina quinasa dependiente de ciclina	Un dominio helicoidal presente en ciclinas y TFIIB (dos veces) y Retinoblastoma (una vez). Un dominio de reconocimiento de proteínas que funciona en el ciclo celular y el control de la transcripción.
	TraesCS6A03G0897300	A0A3B6NVB6	-	-	-
	TraesCS6A03G0899300	A0A3B6NWD2	LEA-2	respuesta de defensa a otro organismo	Se expresan diferentes tipos de proteínas LEA en diferentes etapas de la embriogénesis tardía en embriones de semillas de plantas superiores y en condiciones de estrés por deshidratación.
	TraesCS6A03G0896900	A0A3B6NW01	DUF1618	-	-
	TraesCS6A03G0898600	A0A3B6NWC7	-	-	-
	TraesCS6A03G0898800	A0A3B6NW11	-	unión de iones de zinc	-
	TraesCS6A03G0899800	A0A3B6NVD1	-	aclimatación al frío	-
	TraesCS6A03G0900600	A0A3B6NW21	-	actividad del transportador transmembrana xenobiótico	-
	TraesCS6A03G0901900	A0A3B6NTR7	-	actividad inhibidora de pectinesterasa	-
22	TraesCS6B03G1218400	A0A3B6PV35	PGG	-	El dominio PGG lleva el nombre del motivo de secuencia altamente conservado que se encuentra al comienzo del dominio. La función no se conoce.
	TraesCS6B03G1217400	A0A3B6PV55	DNL	unión de chaperonas de plegamiento de proteínas	El dedo de zinc tipo DNL es una molécula en forma de L. Los dos motivos CXXC están ubicados al final de la L y están intercalados por láminas β antiparalelas de dos cadenas. Dos hélices α cortas constituyen otra pata de la L. La cara exterior (convexa) de la L tiene un gran surco ácido, que está revestido con cinco residuos ácidos, mientras que la cara interior (cóncava) de la L tiene dos residuos cargados positivamente, junto a los motivos CXXC.
	TraesCS6B03G1211800	A0A3B6PSE8	NAC	unión al ADN	El dominio NAC (para petunia NAM y Arabidopsis ATAF1, ATAF2 y CUC2) es un módulo N-terminal de ~ 160 aminoácidos, que se encuentra en proteínas de la familia NAC de reguladores transcripcionales específicos de plantas. Las proteínas NAC participan en los procesos de desarrollo, incluida la formación del meristemo apical del brote, los órganos florales y los brotes laterales, así como en el control y la defensa hormonal de las plantas. El dominio NAC está acompañado por diversos dominios de activación transcripcional C-terminal. Se ha demostrado que el dominio NAC es un dominio de unión al ADN (DBD) y un dominio de dimerización.
	TraesCS6B03G1213600	A0A3B6PV11	HMA	unión de iones metálicos	-
	TraesCS6B03G1216400	A0A3B6PSG0	OTU	-	actividad deubiquitinasa de tipo cisteína
A0A3B6PT17		-			
A0A3B6PTT4		-			
A0A3B6PV25		-			

		A0A3B6PV51		-	
	TraesCS6B03G1211400	A0A3B6PV06	-	vía de señalización activada por auxinas	-
		A0A3B6PV32	-		-
23	TraesCS6B03G0556000	A0A3B6PJ78	ciclofilina PPlasa	Unión de ciclosporina A.	La ciclofilina es la principal proteína de unión de alta afinidad en los vertebrados para el fármaco inmunosupresor ciclosporina A (CSA). Exhibe una actividad peptidil-prolil cis-trans isomerasa (EC 5.2.1.8) (PPlasa o rotamasa). PPlasa es una enzima que acelera el plegamiento de proteínas catalizando la isomerización cis-trans de enlaces peptídicos imídicos de prolina en oligopéptidos. Es probable que la CSA medie algunos de sus efectos mediante una acción inhibidora de la PPlasa.
		A0A3B6PJ83			
		A0A3B6PKE3			
		A0A3B6PKF0			
		A0A3B6PKY0			
		A0A3B6PM95			
		A0A3B6PMA0			
		A0A3B6PMC9			
		A0A3B6PMD4			
	TraesCS6B03G0556100	A0A3B6PKY4	U-box	actividad de la proteína ligasa de ubiquitina	-
24	TraesCS7B03G0880500	A0A3B6SK11	transportador ABC	Actividad del transportador transmembrana de fosfato inorgánico.	Los transportadores ABC pertenecen a la superfamilia de casetes de unión a ATP (ABC) que utiliza la hidrólisis de ATP para energizar diversos sistemas biológicos. Los transportadores ABC están constituidos mínimamente por dos regiones conservadas: un casete de unión a ATP (ABC) altamente conservado y un dominio transmembrana menos conservado (TMD). Estas regiones se pueden encontrar en la misma proteína o en dos diferentes. La mayoría de los transportadores ABC funcionan como un dímero y, por lo tanto, están constituidos por cuatro dominios, dos módulos ABC y dos TMD.
	TraesCS7B03G0880900	A0A3B6SL61	MATH	ubiquitinación de proteínas	Las meprinas son metaloendopeptidasas específicas de tejido de mamíferos de la familia de la astacina implicadas en procesos de desarrollo, normales y patológicos mediante la hidrólisis de una variedad de proteínas. Varios factores de crecimiento, citoquinas y proteínas de la matriz extracelular son sustratos de las meprinas. Están compuestos por cinco dominios estructurales: un dominio de endopeptidasa N-terminal, un dominio MAM, un dominio MATH, un dominio similar a EGF y una región transmembrana C-terminal. La meprina A y B forman homotetrámeros unidos a la membrana, mientras que los homooligómeros de la meprina A se secretan. Un sitio proteolítico adyacente al dominio MATH, sólo presente en meprina A, permite la liberación de la proteína de la membrana.
	TraesCS7B03G0880300	A0A3B6SHR3	RING	-	-
	TraesCS7B03G0881800	A0A3B6SK16	-	proceso catabólico de peróxido de hidrógeno	-
	TraesCS7B03G0877900	A0A3B6SR83	-	actividad hidrolasa, hidrolizando compuestos O-glicosilo	-

	TraesCS7B03G0881100	A0A3B6SHR7	neprosina	-	Las plantas de jarra son insectívoras y secretan un líquido digestivo en la jarra. Este líquido contiene una mezcla de enzimas que incluyen peptidasas. Uno de ellos es la neprosina, obtenida de la planta carnívora <i>Nepenthes ventrata</i> . Esta peptidasa es de tipo catalítico desconocido y no se ve afectada por los inhibidores de peptidasas estándar. Inusualmente, la actividad se dirige hacia los enlaces proliolo, pero a diferencia de la mayoría de las peptidasas que se escinden después de la prolina, no hay restricción en la longitud de la secuencia o la posición del residuo de prolina. La peptidasa se secreta y se presume que posee un péptido de activación N-terminal. El dominio neprosina corresponde a la peptidasa madura. No se sabe si otras proteínas con este dominio son peptidasas.
	TraesCS7B03G0879000	A0A3B6SPU8	-	actividad de inositol tetrakisfosfato quinasa	-
	TraesCS7B03G0877800	A0A3B6SKH7	-	actividad hidrolasa, hidrolizando compuestos O-glicosilo	-
	TraesCS7B03G0879300	A0A3B6SL58	-	-	-
	TraesCS7B03G0879600	-	-	-	-
25	TraesCS7D03G0656000	A0A3B6TRD4	-	actividad de glucosa-1-fosfato adenililtransferasa	-
		A0A3B6TJU4	-	-	-
	TraesCS7D03G0656300	A0A3B6TF02	YTH	Unión de ARN que contiene N6-metiladenosina	-
		A0A3B6TF06		regulación de la estabilidad del ARNm	-
		A0A3B6TJU6		-	
		A0A3B6TJV0		-	
		A0A3B6TRD9		-	
		A0A3B6TRZ0	-	-	
	TraesCS7D03G0656400	A0A3B6TJV1	uS7	constituyente estructural del ribosoma	Esta familia contiene la proteína ribosómica S7 de procariotas y S5 de eucariotas.
	TraesCS7D03G0656800	A0A3B6TRZ7	-	-	-
26	TraesCS7D03G0902200	A0A3B6TNQ9	-	-	-
		A0A3B6TUT2	-	-	-
		A0A3B6TVB2	-	-	-
	TraesCS7D03G0902600	A0A3B6TI96	-	actividad carboxi-liasa	-
	TraesCS7D03G0902700	A0A3B6TN54	-	-	-
	TraesCS7D03G0903400	A0A3B6TVB7	-	-	-
	TraesCS7D03G0904000	A0A3B6TUT6	GRF	-	-
	TraesCS7D03G0904400	A0A3B6TNR3	-	transporte mediado por vesículas del retículo endoplásmico al Golgi	-
	TraesCS7D03G0904500	A0A3B6TIA0	-	constituyente estructural del ribosoma	-
	TraesCS7D03G0905600	A0A3B6TNR8	-	actividad de la proteína quinasa	-

Anexo 2. Información bioinformática de bloques de asociación a SDS.

BA	ID estable del gen EnsemblPlants	ID	Dominio	UniProt	Descripción del dominio
1	TraesCS1B03G0601000	A0A3B5YWR2	-	Actividad 3-hidroxiisobutil-CoA hidrolasa	-
	TraesCS1B03G0601300	A0A3B5YY49	AT	surco menor de unión al ADN rico en adenina-timina	-
		A0A3B5YX47			-
	TraesCS1B03G0601900	A0A3B5YXX0	PfkB	actividad quinasa	Esta familia incluye una variedad de carbohidratos y pirimidina quinasa.
	TraesCS1B03G0602000	A0A1D5SHP6	AMMECR1	-	El síndrome de delección de genes contiguos se caracteriza por síndrome de Alport (A), retraso mental (M), hipoplasia de la parte media de la cara (M) y eliptocitosis (E), así como hipoplasia generalizada y anomalías cardíacas. Está causada por una delección en Xq22.3, que comprende varios genes, incluido el gen 1 de la región cromosómica AMME (AMMECR1), que codifica una proteína con una ubicación nuclear y una función actualmente desconocida. La región C-terminal de AMMECR1 (del residuo 122 al 333) está bien conservada y aparecen homólogos en especies que van desde bacterias y arqueas hasta eucariotas. El alto nivel de conservación del dominio AMMECR1 apunta a una función celular básica, potencialmente en la maquinaria de transcripción, replicación, reparación o traducción.
	TraesCS1B03G0602200	A0A3B5YWR7	-	-	-
	TraesCS1B03G0602900	A0A3B5Y53	-	-	-
	TraesCS1B03G0603200	A0A3B5YX52	-	actividad inhibidora de la proteína fosfatasa	-
2	TraesCS2B03G1355200	A0A3B6CEV4	-	actividad endopeptidasa tipo cisteína	-
		A0A3B6CH67	-		-
	TraesCS2B03G1353600	A0A3B6CH63	-	constituyente estructural del ribosoma	-
	TraesCS2B03G1353400	A0A3B6CEU2	-	Actividad galactosilgalactosilxilosilproteína 3-beta-glucuronosiltransferasa	-
	TraesCS2B03G1356300	A0A3B6CER4	proteína quinasa	actividad de la proteína quinasa	Las proteínas quinasa eucariotas son enzimas que pertenecen a una familia muy extensa de proteínas que comparten un núcleo catalítico conservado común a las proteínas quinasa de serina/treonina y tirosina. Hay varias regiones conservadas en el dominio catalítico de las proteínas quinasa. Hemos seleccionado dos de estas regiones para crear patrones característicos. La primera región, que se encuentra en el extremo N-terminal del dominio catalítico, es un tramo de residuos rico en glicina en las proximidades de un residuo de lisina, que se ha demostrado que está involucrado en la unión de ATP. La segunda región, que se encuentra en la parte central del dominio catalítico, contiene un residuo de ácido aspártico conservado que es importante para la actividad catalítica de la enzima; Hemos derivado dos patrones distintivos para esa región: uno específico para serina/treonina quinasa y el otro para tirosina quinasa. También desarrollamos

					un perfil que se basa en la alineación en y cubre todo el dominio catalítico.	
	TraesCS2B03G1354300	A0A3B6CEU7	granulinas	-	Las granulinas metazoarias son una familia de péptidos ricos en cisteína de aproximadamente 6 Kd que pueden tener múltiples actividades biológicas. Una proteína precursora (conocida como acrogranina) codifica potencialmente siete formas diferentes de granulina (grnA a grnG) que probablemente se liberan mediante procesamiento proteolítico postraduccional. Las granulinas están relacionadas evolutivamente con PMP-D1, un péptido extraído de la pars intercerebralis de las langostas migratorias.	
	TraesCS2B03G1353700	A0A3B6CEV0	proteína quinasa	unión de polisacárido	Las proteínas quinasas eucariotas son enzimas que pertenecen a una familia muy extensa de proteínas que comparten un núcleo catalítico conservado común a las proteínas quinasas de serina/treonina y tirosina. Hay varias regiones conservadas en el dominio catalítico de las proteínas quinasas. Hemos seleccionado dos de estas regiones para crear patrones característicos. La primera región, que se encuentra en el extremo N-terminal del dominio catalítico, es un tramo de residuos rico en glicina en las proximidades de un residuo de lisina, que se ha demostrado que está involucrado en la unión de ATP. La segunda región, que se encuentra en la parte central del dominio catalítico, contiene un residuo de ácido aspártico conservado que es importante para la actividad catalítica de la enzima; Hemos derivado dos patrones distintivos para esa región: uno específico para serina/treonina quinasas y el otro para tirosina quinasas. También desarrollamos un perfil que se basa en la alineación en y cubre todo el dominio catalítico.	
	TraesCS2B03G1353900	-	-	-	-	
3	TraesCS2B03G0332400	-	-	-	-	
	TraesCS2B03G0332500	-	-	-	-	
	TraesCS2B03G0333500	A0A3B6C0L3	proteína quinasa	actividad de la proteína quinasa	actividad de la proteína serina/treonina quinasa	Las proteínas quinasas eucariotas son enzimas que pertenecen a una familia muy extensa de proteínas que comparten un núcleo catalítico conservado común a las proteínas quinasas de serina/treonina y tirosina. Hay varias regiones conservadas en el dominio catalítico de las proteínas quinasas. Hemos seleccionado dos de estas regiones para crear patrones característicos. La primera región, que se encuentra en el extremo N-terminal del dominio catalítico, es un tramo de residuos rico en glicina en las proximidades de un residuo de lisina, que se ha demostrado que está involucrado en la unión de ATP. La segunda región, que se encuentra en la parte central del dominio catalítico, contiene un residuo de ácido aspártico
		A0A3B6C0P2				
		A0A3B6C0R0				
		A0A3B6C0Y2				
A0A3B6C398						

					conservado que es importante para la actividad catalítica de la enzima ; Hemos derivado dos patrones distintivos para esa región: uno específico para serina/treonina quinasa y el otro para tirosina quinasa. También desarrollamos un perfil que se basa en la alineación en y cubre todo el dominio catalítico.
	TraesCS2B03G0333600	A0A3B6C0Y7	lipasa fúngica	-	Esta entrada representa un dominio con un pliegue de α/β hidrolasa que se encuentra en un grupo de proteínas de eucariotas y bacterias, incluida la feruloil esterasa A de <i>Aspergillus niger</i> , Triacilglicerol lipasa OBL1 de <i>Arabidopsis thaliana</i> , diacilglicerol lipasa-alfa humana, Toxina TseL de <i>Vibrio cholerae</i> . Es similar al encontrado en las lipasas fúngicas.
	TraesCS2B03G0333900	A0A3B6C0R4	C3H1	unión de iones metálicos	-
	TraesCS2B03G0334600	A0A3B6C0L8			-
4	TraesCS2D03G1056200	A0A3B6DJZ0	C-terminal ataxina-2	unión de ARNm	-
	TraesCS2D03G1059400	A0A1D5UQA8	-	actividad del transportador transmembrana	-
	TraesCS2D03G1059500	A0A3B6DKE3	-	desarrollo del xilema	-
	TraesCS2D03G1057800	A0A3B6DJZ5	GRF con dedos de zinc	-	-
	TraesCS2D03G1061100	A0A3B6DKF0	APO	Unión de ARN	La ACUMULACIÓN DEL FOTOSISTEMA UNO1 (APO1) es necesaria para la acumulación del fotosistema I (PSI) y otras proteínas del cloroplasto codificadas por plastidios [4FE-4S]. APO1 pertenece a una familia de genes con cuatro grupos definidos (APO1 a APO4) que sólo se encuentran en los genomas nucleares de las plantas vasculares. Los miembros de la familia de genes APO están presentes tanto en monocotiledóneas como en dicotiledóneas y se predice que estarán localizados en cloroplastos y mitocondrias. APO1 a APO4 muestran muchas menos similitudes en el extremo N que en el resto de las proteínas, lo que indica diferentes localizaciones y/o funciones. Cada uno de los grupos, APO1 a APO4, muestra características distintas y conservadas, lo que sugiere que todos son importantes para la célula vegetal y poseen funciones relacionadas en diferentes compartimentos celulares.
	TraesCS2D03G1061400	-	-	-	-
	TraesCS2D03G1059300	A0A3B6DJ91	-	Actividad del factor de transcripción de unión al ADN, específico de la ARN polimerasa II	-
	TraesCS2D03G1059100	A0A3B6DK02	AAA+ ATPasa	Unión ADP	-
	TraesCS2D03G1056100	A0A3B6DJ76	RRM	Unión de ARN	Muchas proteínas eucariotas que se sabe o se supone que se unen a ARN monocatenario contienen una o más copias de un supuesto dominio de unión a ARN de aproximadamente 90 aminoácidos. Este dominio se conoce como
		A0A3B6DKW4		unión de ARNm	

					motivo de reconocimiento de ARN (RRM).
5	TraesCS2D03G1271900	A0A2X0SEE6	-	-	-
	TraesCS2D03G1276000	A0A1D5UGY5	-	actividad de la proteína quinasa	-
	TraesCS2D03G1270200	A0A3B6DPX4	proteína quinasa	actividad de la proteína serina/treonina quinasa	Las proteínas quinatas eucariotas son enzimas que pertenecen a una familia muy extensa de proteínas que comparten un núcleo catalítico conservado común a las proteínas quinatas de serina/treonina y tirosina. Hay varias regiones conservadas en el dominio catalítico de las proteínas quinatas. Hemos seleccionado dos de estas regiones para crear patrones característicos. La primera región, que se encuentra en el extremo N-terminal del dominio catalítico, es un tramo de residuos rico en glicina en las proximidades de un residuo de lisina, que se ha demostrado que está involucrado en la unión de ATP. La segunda región, que se encuentra en la parte central del dominio catalítico, contiene un residuo de ácido aspártico conservado que es importante para la actividad catalítica de la enzima; Hemos derivado dos patrones distintivos para esa región: uno específico para serina/treonina quinatas y el otro para tirosina quinatas. También desarrollamos un perfil que se basa en la alineación en y cubre todo el dominio catalítico.
	TraesCS2D03G1274000	-	-	-	-
	TraesCS2D03G1273800	A0A1D5UPG4	-	unión de fosfatidilinositol-4,5-bisfosfato	-
	TraesCS2D03G1268800	A0A3B6DRV1	-	-	-
	TraesCS2D03G1274200	A0A1D5UJ71	-	-	-
	TraesCS2D03G1268500	A0A3B6DPW4	-	-	-
	TraesCS2D03G1270400	A0A3B6DNB2	Nudix hidrolasa	Actividad bis(5'-adenosil)-pentafofatasa	La superfamilia Nudix está muy extendida entre eucariotas, bacterias, arqueas y virus y se compone principalmente de pirofosfohidrolasas que actúan sobre sustratos de estructura general NUCleósido Difosfato unido a otro resto, X (NDP-X), para producir NMP más PX. Dichos sustratos incluyen (d)NTP (derivados canónicos y oxidados), azúcares y alcoholes de nucleótidos, polifosfatos de dinucleósidos (NpnN), coenzimas de dinucleótidos y ARN protegidos. Sin embargo, también se ha descrito la actividad de la fosfohidrolasa, incluida la actividad hacia los propios NDP, y sustratos no nucleotídicos como los polifosfatos de difosfoinositol (DIP), el 5-fosforribosil 1-pirofosfato (PRPP), el pirofosfato de tiamina (TPP) y el trifosfato de dihidroneopterina (DHNTTP). Algunos miembros de la superfamilia, como Escherichia coli mutT, tienen la capacidad de degradar nucleótidos oxidados potencialmente mutagénicos, mientras que otros controlan los niveles de intermediarios metabólicos y compuestos de

					señalización. En procariontes y eucariotes simples, el número de genes Nudix varía de 0 a más de 30, lo que refleja la complejidad metabólica y la adaptabilidad del organismo. Las hidrolasas Nudix son típicamente proteínas pequeñas, las más grandes tienen dominios adicionales con funciones interactivas u otras funciones catalíticas.
	TraesCS2D03G1271300	A0A1D5ULI4	-	-	-
6	TraesCS3A03G0891600	A0A3B6EMQ5	-	Unión de ATP	-
	TraesCS3A03G0892600	A0A3B6EQX0	parche G	unión de ácido nucleico	El dominio del parche G es un dominio de aproximadamente 48 aminoácidos, que se encuentra en una sola copia en varias proteínas asociadas a ARN y en poliproteínas retrovirales tipo D. Está muy extendido entre los eucariotes, pero está ausente en arqueas y bacterias. El dominio del parche G recibe su nombre de su característica más notable, la presencia de seis residuos de glicina altamente conservados. La posición que sigue a la primera glicina conservada está ocupada casi invariablemente por un residuo aromático, y varias otras posiciones están ocupadas predominantemente por residuos hidrofóbicos o pequeños. Varios grupos de proteínas que contienen parche G se conservan en animales, plantas y hongos. En algunas de estas proteínas, el parche G es el único dominio reconocible, pero en la mayoría de ellas se combina con otros dominios, que incluyen dominios de unión a ARN bien definidos, como el RRM, dsRBD, SURP y R3H. Se ha sugerido que el dominio del parche G tiene una función específica en el procesamiento del ARN y, en particular, que podría ser un dominio de unión al ARN no detectado previamente que media en un tipo distinto de interacción ARN-proteína.
	TraesCS3A03G0892300	A0A3B6ELE9	NAC	Actividad del factor de transcripción de unión al ADN.	El dominio NAC (para petunia NAM y Arabidopsis ATAF1, ATAF2 y CUC2) es un módulo N-terminal de ~ 160 aminoácidos, que se encuentra en proteínas de la familia NAC de reguladores transcripcionales específicos de plantas. Las proteínas NAC participan en los procesos de desarrollo, incluida la formación del meristemo apical del brote, los órganos florales y los brotes laterales, así como en el control y la defensa hormonal de las plantas. El dominio NAC está acompañado por diversos dominios de activación transcripcional C-terminal. Se ha demostrado que el dominio NAC es un dominio de unión al ADN (DBD) y un dominio de dimerización.
		A0A3B6ELU5			
		A0A3B6EMR0	-	transcripción unión a región reguladora cis	-

		A0A3B6EQW5	NAC	Actividad del factor de transcripción de unión al ADN.	El dominio NAC (para petunia NAM y Arabidopsis ATAF1, ATAF2 y CUC2) es un módulo N-terminal de ~ 160 aminoácidos, que se encuentra en proteínas de la familia NAC de reguladores transcripcionales específicos de plantas. Las proteínas NAC participan en los procesos de desarrollo, incluida la formación del meristemo apical del brote, los órganos florales y los brotes laterales, así como en el control y la defensa hormonal de las plantas. El dominio NAC está acompañado por diversos dominios de activación transcripcional C-terminal. Se ha demostrado que el dominio NAC es un dominio de unión al ADN (DBD) y un dominio de dimerización.
	TraesCS3A03G0895400	A0A1D5WMA7	-	transporte mediado por vesículas del retículo endoplásmico al Golgi	-
	TraesCS3A03G0895600	A0A3B6ELW0	F-box	-	Identificada por primera vez en la ciclina-F como un motivo de interacción proteína-proteína, la caja F es un dominio conservado que está presente en numerosas proteínas con una estructura bipartita. A través de la F-box, estas proteínas están unidas a la proteína Skp1 y al núcleo de los complejos SCF (Skp1-cullin-F-box proteína ligasa). Los complejos SCF constituyen una nueva clase de ligasas E3. Funcionan en combinación con la enzima E2 Cdc34 para ubiquitinar ciclinas G1, inhibidores de Cdk y muchas otras proteínas, para marcarlas para su degradación. La unión de los sustratos específicos mediante complejos SCF está mediada por motivos divergentes de interacción proteína-proteína presentes en las proteínas F-box, como repeticiones WD40, repeticiones ricas en leucina o repeticiones ANK.
	TraesCS3A03G0892000	A0A1D5VCR6	-	Actividad de fructosa 1.6-bifosfato 1-fosfatasa	-
	TraesCS3A03G0894200	A0A3B6EMS2	GRF con dedos de zinc	-	-
7	TraesCS4A03G0598900	A0A3B6HV99	-	actividad sacarosa-fosfato sintasa	-
		A0A3B6HVV5	-		-
	TraesCS4A03G0599000	A0A3B6HUP1	-	actividad de la glicerofosfodiéster fosfodiesterasa	-
		A0A3B6HYA3	-		-
	TraesCS4A03G0599600	A0A3B6HZ99	VQ	-	Este breve motivo se encuentra en una variedad de proteínas vegetales. Estas proteínas varían mucho en longitud y en su mayoría están compuestas por regiones de baja complejidad. Todos conservan un motivo corto FXhVQChTG, donde X es cualquier aminoácido y h es un aminoácido hidrofóbico. La función de este motivo es incierta; sin embargo, se ha descubierto que una proteína de esta familia se une al factor suizo SigA sigma: Q9LDH1. Parecería plausible que este motivo sea necesario para esta actividad y que toda esta familia pueda estar involucrada en la modulación de los factores sigma

					de los plástidos (Bateman A, obs. personal).
	TraesCS4A03G0599700	A0A3B6HVA4	-	actividad del transportador transmembrana	-
	TraesCS4A03G0599800	A0A3B6HWW0	PHD con dedos de zinc	unión de iones metálicos	El dedo del homeodominio vegetal (PHD) es un motivo similar a un dedo de zinc C4HC3 que se encuentra en proteínas nucleares que se cree que están involucradas en la epigenética y la regulación transcripcional mediada por la cromatina. El dedo PHD une dos iones de zinc mediante el llamado motivo de "correa cruzada" y, por lo tanto, está estructuralmente relacionado con el dedo RING y el dedo FYVE. Aún no se sabe si los dedos de PHD tienen una función molecular común. Varios informes sugieren que puede funcionar como un dominio de interacción proteína-proteína y recientemente se demostró que el dedo PHD de p300 puede cooperar con el dominio BROMO adyacente en la unión de nucleosomas in vitro. Otros informes que sugieren que el dedo PHD es una ubiquitina ligasa han sido refutados ya que estos dominios eran dedos RING identificados erróneamente como dedos PHD
	TraesCS4A03G0600100	A0A3B6HUP5	TF-B3	unión al ADN	El dominio B3 es una región de residuos de aproximadamente 120 aminoácidos definida originalmente como la tercera región básica del maíz VP1, que comparte el bloque contiguo más grande de identidad de secuencia con su ortólogo, Arabidopsis ABI3. El dominio B3, que también ha sido llamado dominio REM, se encuentra sólo en especies de plantas superiores. Es característico de varios factores de transcripción y se ha demostrado que posee una actividad de unión al ADN altamente cooperativa y específica de secuencia. El dominio B3 se puede encontrar en una o más copias y en asociación con otros dominios, como el dominio de unión al ADN AP2, un dominio ácido o el dominio Aux/IAA-ARF.
	TraesCS4A03G0600600	A0A3B6HYA8	alfa/beta hidrolasa plegado 3	actividad hidrolasa	El pliegue de la α/β hidrolasa es común a varias enzimas hidrolíticas de origen filogenético y función catalítica muy diferentes. El núcleo de cada enzima es una lámina α/β (en lugar de un barril), que contiene 8 hebras conectadas por hélices. Se cree que las enzimas se separaron de un ancestro común, preservando la disposición de los residuos catalíticos. Todos presentan una tríada catalítica, cuyos elementos se apoyan en bucles, que son los rasgos estructurales mejor conservados del pliegue. La esterasa (EST) de Pseudomonas putida es un miembro de la superfamilia de enzimas de la α/β hidrolasa.
	TraesCS4A03G0600900	A0A3B6HVA8	-	silenciamiento génico mediado por ncRNA regulador	-
		A0A3B6HWW4	-		-
		A0A3B6HZA4	-		-
8	TraesCS4D03G0625400	A0A3B6JNW6	-	-	-

TraesCS4D03G0625500	A0A341PHP9	GRF	unión de iones de zinc	El GRF-ZF comprende una lámina β antiparalela de tres hebras (β 1- β 3) que se pliega en una estructura similar a una garra en forma de media luna. Un ion $Zn(2+)$ unido único desempeña un papel estructural central en este dominio y está coordinado con la geometría tetraédrica mediante un motivo de secuencia "CHCC". La identidad de estos ligandos $Zn(2+)$ se conserva para la mayoría de las proteínas que contienen GRF-ZF; sin embargo, un subconjunto de proteínas GRF-ZF [p. ej., Top3 α (Topoisomerasa 3 α)] sustituye la His de este motivo por un residuo Cys (coordinación CCCC). Los dos primeros ligandos de $Zn(2+)$ se encuentran en un bucle que precede a β 1, mientras que la segunda mitad del motivo se asigna al bucle de conexión β 2- β 3.
TraesCS4D03G0624600	-	-	-	-
TraesCS4D03G0623600	A0A3B6JLM5	-	actividad serina-tRNA ligasa	-
TraesCS4D03G0624000	A0A3B6JN15	-	Actividad de ADN-3-metiladenina glicosilasa	-
TraesCS4D03G0623800	A0A3B6JM76	PPPDE	actividad deubiquitinasa	-
TraesCS4D03G0623500	A0A3B6JNV8	WW	-	El dominio WW (también conocido como rsp5 o WWP) se descubrió originalmente como una región corta conservada en varias proteínas no relacionadas, entre ellas la distrofina, el gen responsable de la distrofia muscular de Duchenne. El dominio, que abarca unos 35 residuos, se repite hasta 4 veces en algunas proteínas. Se ha demostrado que se une a proteínas con motivos de prolina particulares, [AP]-PP-[AP]-Y, y por lo tanto se parece un poco a los dominios SH3. Parece contener cadenas β agrupadas alrededor de cuatro posiciones aromáticas conservadas; generalmente Trp. El nombre WW o WWP deriva de la presencia de estos Trp así como de la de un Pc
TraesCS4D03G0625600	A0A3B6JM83	F-box	-	Identificada por primera vez en la ciclina-F como un motivo de interacción proteína-proteína, la caja F es un dominio conservado que está presente en numerosas proteínas con una estructura bipartita. A través de la F-box, estas proteínas están unidas a la proteína Skp1 y al núcleo de los complejos SCF (Skp1-cullin-F-box proteína ligasa). Los complejos SCF constituyen una nueva clase de ligasas E3. Funcionan en combinación con la enzima E2 Cdc34 para ubiquitinar ciclinas G1, inhibidores de Cdk y muchas otras proteínas, para marcarlas para su degradación. La unión de los sustratos específicos mediante complejos SCF está mediada por motivos divergentes de interacción proteína-proteína presentes en las proteínas F-box, como repeticiones WD40, repeticiones ricas en leucina o repeticiones ANK.

	TraesCS4D03G0626400	A0A3B6JNX6	MFS	actividad del transportador transmembrana de carbohidratos	Entre los diferentes transportadores, las dos familias más grandes que se encuentran de manera ubicua en todas las clasificaciones de organismos son la superfamilia de transportadores primarios del casete de unión a ATP (ABC) (ver < PDOC00185 >) y la superfamilia de facilitadores principales (MFS). Los transportadores MFS son portadores secundarios de un solo polipéptido capaces sólo de transportar pequeños solutos en respuesta a gradientes de iones quimiosmóticos. Funcionan como uniportadores, simportadores o antiportadores. Además, su especificidad de soluto también es diversa. Las proteínas MFS contienen 12 regiones transmembrana (con algunas variaciones).
	TraesCS4D03G0625700	A0A3B6JKR2	GRF con dedos de zinc	-	-
9	TraesCS5A03G0780700	A0A3B6KMB6	Clp R	Actividad de hidrólisis de ATP	Las chaperonas moleculares reconocen proteínas desplegadas o mal plegadas uniéndose a parches de superficie hidrofóbica que normalmente no están expuestos en las proteínas nativas. Los miembros de la familia de chaperonas Clp/Hsp100 están presentes en eubacterias y dentro de los orgánulos de todos los eucariotas, promoviendo la desagregación y el desmontaje de complejos proteicos y participando en la degradación de proteínas dependiente de energía. Las subfamilias ClpA, ClpB y ClpC de las ATPasas Clp/Hsp100 contienen un dominio N-terminal conservado de ~150 aminoácidos, que a su vez consta de dos repeticiones de ~75 residuos. Aunque el dominio de repetición Clp (R) contiene dos repeticiones de secuencia aproximadas, se comporta como una única unidad plegada cooperativamente. Se cree que el dominio Clp R proporciona un medio para regular la especificidad y ampliar el conjunto de sustratos disponibles para los complejos de proteasa o chaperona Clp/Hsp100. Estas funciones pueden verse favorecidas mediante la unión de una proteína adaptadora. Las proteínas adaptadoras se unen al dominio Clp R, modulan la especificidad objetivo del complejo Clp/Hsp100 hacia un sustrato particular de interés y también pueden regular la actividad del complejo.
		A0A3B6KN04			
	TraesCS5A03G0778700	A0A3B6KMA6	-	Unión de ADN específica de secuencia de región reguladora cis	-
	TraesCS5A03G0783700	A0A3B6KKN3	-	actividad pectina acetiltransferasa	-
	TraesCS5A03G0784200	A0A3B6KN19	-	actividad celulasa	-

	TraesCS5A03G0780200	A0A3B6KME8	fosfoesterasa similar a la calcineurina	Actividad hidrolasa, que actúa sobre enlaces éster.	Este dominio se encuentra en una amplia gama de fosfoesterasas, incluyendo bis(5'-nucleosil)-tetrafosfatasa (apaH), nucleotidasas, esfingomieliina fosfodiesterasas y 2'-3' AMPc fosfodiesterasas, así como nucleasas tales como SbcD bacteriana o Mre11 de arqueas/levaduras. Las regiones más conservadas en este dominio se centran alrededor de los residuos quelantes de metales.
	TraesCS5A03G0781200	A0A3B6KJS4	-	Actividad de hidrólisis de ATP	-
	TraesCS5A03G0778100	A0A3B6KKL3	-	-	-
	TraesCS5A03G0784300	A0A3B6KMH3	N-terminal de ciclina	Actividad reguladora de la proteína serina/treonina quinasa dependiente de ciclina	-
	TraesCS5A03G0780000	A0A3B6KMZ9	fosfoesterasa similar a la calcineurina	Actividad hidrolasa, que actúa sobre enlaces éster.	Este dominio se encuentra en una amplia gama de fosfoesterasas, incluyendo bis(5'-nucleosil)-tetrafosfatasa (apaH), nucleotidasas, esfingomieliina fosfodiesterasas y 2'-3' AMPc fosfodiesterasas, así como nucleasas tales como SbcD bacteriana o Mre11 de arqueas/levaduras. Las regiones más conservadas en este dominio se centran alrededor de los residuos quelantes de metales.
10	TraesCS5A03G1212300	A0A3B6KU84	similar a la ubiquitina	-	La ubiquitina es una proteína de setenta y seis residuos de aminoácidos, que se encuentra en todas las células eucariotas y cuya secuencia está extremadamente bien conservada desde los protozoos hasta los vertebrados. Desempeña un papel clave en una variedad de procesos celulares, como la degradación selectiva de proteínas celulares dependiente de ATP, el mantenimiento de la estructura de la cromatina, la regulación de la expresión genética, la respuesta al estrés y la biogénesis de los ribosomas.
	TraesCS5A03G1210200	A0A3B6KTJ6	epimerasa/deshidratasa dependiente de NAD	Actividad oxidoreductasa, actuando sobre el grupo CH-OH de los donantes, NAD o NADP como aceptor.	Este dominio se encuentra en proteínas que utilizan NAD como cofactor y utilizan sustratos de azúcar de nucleótidos para una variedad de reacciones químicas. Una de las proteínas mejor estudiadas es la UDP-galactosa 4-epimerasa, que cataliza la conversión de UDP-galactosa en UDP-glucosa durante el metabolismo de la galactosa.
	TraesCS5A03G1214100	A0A3B6KTV1	-	-	-
	TraesCS5A03G1216200	A0A3B6KTW5	-	actividad del transportador transmembrana	-
	TraesCS5A03G1209500	A0A3B6KTJ1	TF-B3	Actividad del factor de transcripción de unión al ADN.	El dominio B3 es una región de residuos de aproximadamente 120 aminoácidos definida originalmente como la tercera región básica del maíz VP1, que comparte el bloque contiguo más grande de identidad de secuencia con su ortólogo, Arabidopsis ABI3. El dominio B3, que también ha sido llamado dominio REM, se encuentra sólo en especies de plantas superiores. Es característico de varios factores de transcripción y se ha demostrado que posee una actividad de unión al ADN altamente cooperativa y

					específica de secuencia. El dominio B3 se puede encontrar en una o más copias y en asociación con otros dominios, como el dominio de unión al ADN AP2, un dominio ácido o el dominio Aux/IAA-ARF.
	TraesCS5A03G1213100	A0A3B6KTK8	similar a la ubiquitina	-	La ubiquitina es una proteína de setenta y seis residuos de aminoácidos, que se encuentra en todas las células eucariotas y cuya secuencia está extremadamente bien conservada desde los protozoos hasta los vertebrados. Desempeña un papel clave en una variedad de procesos celulares, como la degradación selectiva de proteínas celulares dependiente de ATP, el mantenimiento de la estructura de la cromatina, la regulación de la expresión genética, la respuesta al estrés y la biogénesis de los ribosomas.
	TraesCS5A03G1213500	A0A3B6KTU7	-	Actividad de quercetina 7-O-glucosiltransferasa	-
	TraesCS5A03G1215600	A0A3B6KRC8	MADS-box	Unión de ADN específica de secuencia de la región reguladora cis de la ARN polimerasa II	-
	TraesCS5A03G1212400	A0A3B6KTU1	-	-	-
	TraesCS5A03G1213000	A0A3B6KRB1	-	actividad del canal de agua	-
11	TraesCS5D03G1095400	A0A3B6N1P5	pequeño metiltransferasa	Actividad metiltransferasa dependiente de S-adenosilmetionina	Este dominio se encuentra en la metiltransferasa C de la subunidad pequeña del ARN ribosomal y en otras metiltransferasas.
	TraesCS5D03G1101100	A0A3B6MZZ4	GRAM	-	El dominio GRAM se encuentra en glucosiltransferasas, miotubularinas y otras supuestas proteínas asociadas a membrana. Normalmente tiene una longitud de unos 70 aminoácidos. Se cree que es un dominio de señalización de unión a proteínas o de lípidos intracelular, que tiene una función importante en los procesos asociados a la membrana. La estructura del dominio GRAM es similar a la que se encuentra en los dominios PH. Las mutaciones en el dominio GRAM de las miotubularinas causan una enfermedad muscular, lo que sugiere que el dominio es esencial para el funcionamiento completo de la enzima. Las proteínas relacionadas con la miotubularina son una gran subfamilia de proteínas tirosina fosfatasa (PTP) que desfosforilan los lípidos de inositol fosforilados en D3.
		A0A3B6N1C1		-	
	TraesCS5D03G1097400	A0A3B6MZX8	proteína quinasa	actividad de la proteína tirosina quinasa	Las proteínas quinasa eucariotas son enzimas que pertenecen a una familia muy extensa de proteínas que comparten un núcleo catalítico conservado común a las proteínas quinasa de serina/treonina y tirosina. Hay varias regiones conservadas en el dominio catalítico de las proteínas quinasa. Hemos seleccionado dos de estas regiones para crear patrones característicos. La primera región, que se encuentra en el extremo N-terminal del dominio catalítico, es un tramo de residuos rico en glicina en las proximidades de un residuo de lisina, que se ha demostrado
		A0A3B6N118			

					que está involucrado en la unión de ATP. La segunda región, que se encuentra en la parte central del dominio catalítico, contiene un residuo de ácido aspártico conservado que es importante para la actividad catalítica de la enzima; Hemos derivado dos patrones distintivos para esa región: uno específico para serina/treonina quinasa y el otro para tirosina quinasa. También desarrollamos un perfil que se basa en la alineación en y cubre todo el dominio catalítico.
	TraesCS5D03G1100700	A0A3B6N1S9	-	actividad de la proteína tirosina/serina/treonina fosfatasa	-
	TraesCS5D03G1096600	A0A3B6N1Q7	N-terminal similar a NusG	-	-
	TraesCS5D03G1094700	A0A3B6N179	-	actividad pectina acetilesterasa	-
	TraesCS5D03G1100600	A0A3B6N2M9	-	transporte mediado por vesículas del retículo endoplásmico al Golgi	-
	TraesCS5D03G1094200	A0A3B6N171	NB-ARC	Unión ADP	Este es el dominio NB-ARC, un nuevo motivo de señalización que se encuentra en bacterias y eucariotas, compartido por productos genéticos de resistencia de plantas y reguladores de la muerte celular en animales. Este dominio se ha caracterizado estructuralmente en la proteína humana factor 1 activador de proteasa apoptótica (Apaf-1). Contiene el pliegue α - β de tres capas y la posterior región helicoidal α corta característica de la superfamilia del dominio ATPasa AAA+. Si bien se cree que este dominio se une e hidroliza ATP, sólo se ha verificado experimentalmente la unión de ADP. Se propone que la unión e hidrólisis del ATP por este dominio induce cambios conformacionales en la proteína general, lo que conduce a la formación del apoptosoma
12	TraesCS5D03G1025000	A0A3B6MZQ3	-	constituyente estructural del ribosoma	-
	TraesCS5D03G1025300	A0A3B6N1C6	-	unión de lípidos	-
	TraesCS5D03G1024800	A0A3B6N0G1	serpina	actividad inhibidora de la endopeptidasa tipo serina	Las serpinas comparten una estructura central altamente conservada que es fundamental para su funcionamiento como inhibidores de serina proteasa. Las serpinas inhibidoras comprenden varias hélices alfa y cadenas beta junto con un bucle central reactivo externo (RCL) que contiene el sitio activo reconocido por el enzima objetivo. Las serpinas forman complejos covalentes con proteasas diana. Su mecanismo de inhibición de la proteasa se conoce como "atrapamiento" irreversible en el que un rápido cambio conformacional atrapa la proteasa afín en un complejo covalente.
	TraesCS5D03G1024200	-	-	-	-
	TraesCS5D03G1023100	A0A3B6N1B5	BTB	respuesta al estímulo abiótico	-
	TraesCS5D03G1026800	A0A3B6MYR5	NPH3	-	-

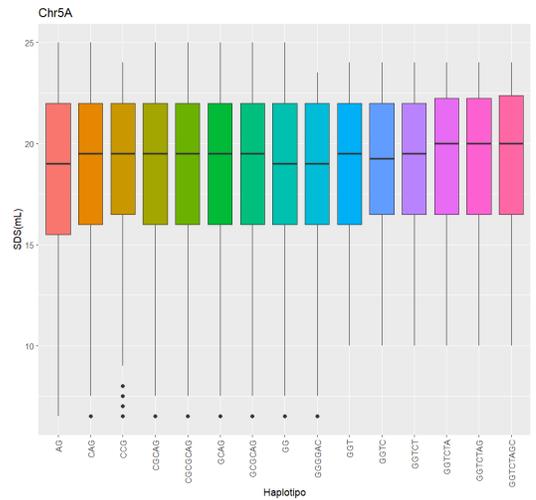
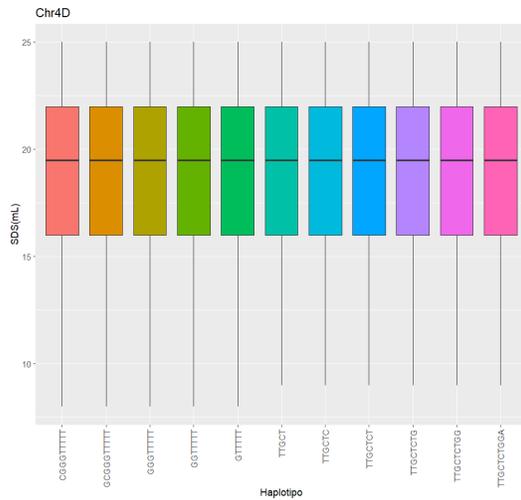
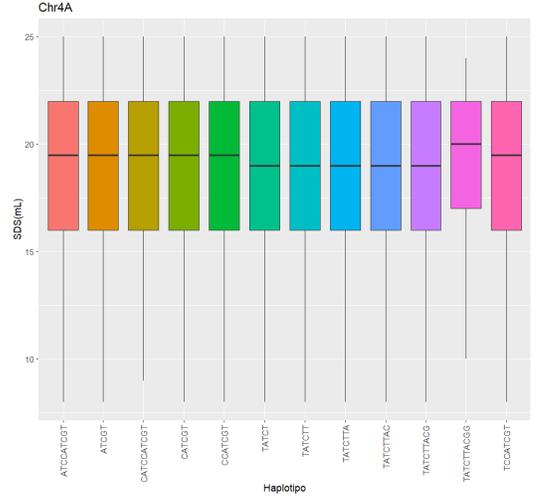
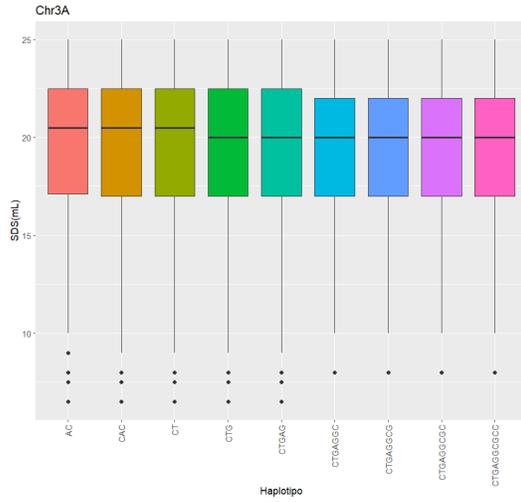
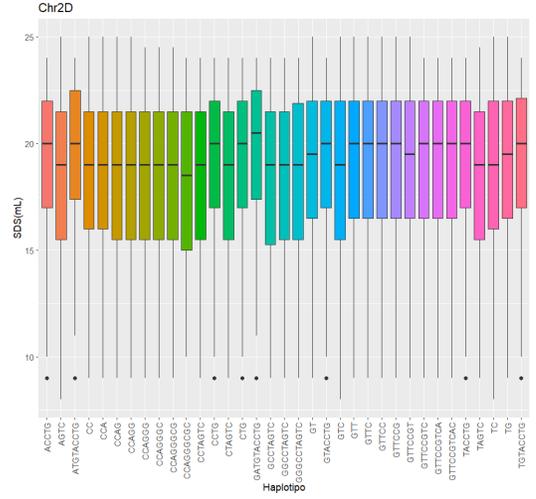
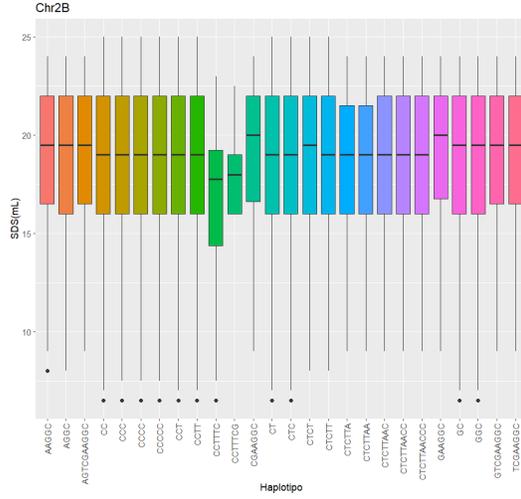
		A0A3B6MZZ9		-	La familia NRL (para NPH3/RPT2-Like) está formada por moléculas de señalización específicas de plantas superiores. Varias regiones de secuencia y conservación estructural predicha definen a los miembros de la familia NRL, siendo tres dominios los más notables: un dominio BTB N-terminal, un dominio NPH3 ubicado centralmente y un dominio en espiral C-terminal. La función del dominio NPH3 aún no se conoce.
		A0A3B6N1D5		-	
	TraesCS5D03G1027400	A0A3B6N0I4	-	Actividad UDP-glucosiltransferasa	-
	TraesCS5D03G1023900	A0A3B6N1B8	-	actividad del factor de iniciación de traducción	-
13	TraesCS6B03G0390600	A0A3B6PKF2	N-terminal	-	-
	TraesCS6B03G0391000	A0A3B6PKC3		-	-
	TraesCS6B03G0391100	A0A3B6PIV5	-	-	-
	TraesCS6B03G0391200	A0A3B6PIM9	-	-	-
	TraesCS6B03G0391400	A0A3B6PH81	N-terminal	-	Este dominio se encuentra a menudo en el extremo N de repeticiones ricas en leucina en tándem, principalmente en proteínas vegetales. Las repeticiones ricas en leucina (LRR) constan de 2 a 45 motivos de 20 a 30 aminoácidos de longitud que generalmente se pliegan en forma de arco o de herradura. Los LRR se encuentran en proteínas que van desde virus hasta eucariotas y parecen proporcionar un marco estructural para la formación de interacciones proteína-proteína. Las proteínas que contienen LRR incluyen receptores de tirosina quinasa, moléculas de adhesión celular, factores de virulencia y glicoproteínas de unión a la matriz extracelular, y participan en una variedad de procesos biológicos, incluida la transducción de señales, la adhesión celular, la reparación del ADN, la recombinación, la transcripción y el procesamiento del ARN., resistencia a enfermedades, apoptosis y respuesta inmune.
	TraesCS6B03G0391500	A0A3B6PKF4	-	-	-
	TraesCS6B03G0392000	A0A3B6PKC9	N-terminal	-	Este dominio se encuentra a menudo en el extremo N de repeticiones ricas en leucina en tándem, principalmente en proteínas vegetales. Las repeticiones ricas en leucina (LRR) constan de 2 a 45 motivos de 20 a 30 aminoácidos de longitud que generalmente se pliegan en forma de arco o de herradura. Los LRR se encuentran en proteínas que van desde virus hasta eucariotas y parecen proporcionar un marco estructural para la formación de interacciones proteína-proteína. Las proteínas que contienen LRR incluyen receptores de tirosina quinasa, moléculas de adhesión celular, factores de virulencia y glicoproteínas de unión a la matriz extracelular, y participan en una variedad de procesos biológicos, incluida la transducción de señales, la adhesión celular, la reparación del ADN, la recombinación, la
	TraesCS6B03G0392100	A0A3B6PIW1		-	

					transcripción y el procesamiento del ARN, resistencia a enfermedades, apoptosis y respuesta inmune.
	TraesCS6B03G0392200	A0A3B6PIN4	NB-ARC	Unión ADP	Este es el dominio NB-ARC, un nuevo motivo de señalización que se encuentra en bacterias y eucariotas, compartido por productos genéticos de resistencia de plantas y reguladores de la muerte celular en animales. Este dominio se ha caracterizado estructuralmente en la proteína humana factor 1 activador de proteasa apoptótica (Apaf-1). Contiene el pliegue α - β de tres capas y la posterior región helicoidal α corta característica de la superfamilia del dominio ATPasa AAA+. Si bien se cree que este dominio se une e hidroliza ATP, sólo se ha verificado experimentalmente la unión de ADP. Se propone que la unión e hidrólisis del ATP por este dominio induce cambios conformacionales en la proteína general, lo que conduce a la formación del apoptosoma
	TraesCS6B03G0393000	A0A3B6PH86	N-terminal	-	Este dominio se encuentra a menudo en el extremo N de repeticiones ricas en leucina en tándem, principalmente en proteínas vegetales. Las repeticiones ricas en leucina (LRR) constan de 2 a 45 motivos de 20 a 30 aminoácidos de longitud que generalmente se pliegan en forma de arco o de herradura. Los LRR se encuentran en proteínas que van desde virus hasta eucariotas y parecen proporcionar un marco estructural para la formación de interacciones proteína-proteína. Las proteínas que contienen LRR incluyen receptores de tirosina quinasa, moléculas de adhesión celular, factores de virulencia y glicoproteínas de unión a la matriz extracelular, y participan en una variedad de procesos biológicos, incluida la transducción de señales, la adhesión celular, la reparación del ADN, la recombinación, la transcripción y el procesamiento del ARN, resistencia a enfermedades, apoptosis y respuesta inmune.
14	TraesCS7A03G0110300	A0A3B6RBG4	LEA-2	respuesta de defensa a otro organismo	-
	TraesCS7A03G0110400	A0A3B6RBK8	-	-	-
	TraesCS7A03G0110800	A0A3B6R845	-	-	-
	TraesCS7A03G0111000	A0A3B6R7N2	DUF6598	-	Este dominio se encuentra en el terminal C de varias proteínas vegetales, incluida la proteína inducida por jasmonato de 60 kDa y en algunas secuencias de hongos. Se desconoce la función del dominio.
	TraesCS7A03G0111700	A0A3B6RBG7 A0A3B6RBG9	START	unión de lípidos	- -
	TraesCS7A03G0111800	A0A3B6RBL2	-	modificación del ARNm mitocondrial	-
	TraesCS7A03G0112400	A0A3B6R850	Clp R	Actividad de hidrólisis de ATP	-
	TraesCS7A03G0112700	-	-	-	-

	TraesCS7A03G0113000	A0A3B6RBH2	mano EF	unión de iones calcio	Muchas proteínas de unión a calcio pertenecen a la misma familia evolutiva y comparten un tipo de dominio de unión a calcio conocido como EF-mano. Este tipo de dominio consta de un bucle de doce residuos flanqueado en ambos lados por un dominio helicoidal α de doce residuos. En un bucle manual EF, el ion calcio está coordinado en una configuración bipiramidal pentagonal. Los seis residuos implicados en la unión están en las posiciones 1, 3, 5, 7, 9 y 12; estos residuos se indican con X, Y, Z, -Y, -X y -Z. El Glu o Asp invariante en la posición 12 proporciona dos oxígenos para ligar Ca (ligando bidentado). La unidad estructural/funcional básica de las proteínas de la mano EF suele ser un par de motivos de la mano EF que juntos forman un dominio estable de haz de cuatro hélices. El emparejamiento de EF-hand permite la cooperatividad en la unión de iones Ca^{2+}
15	TraesCS7A03G1163700	A0A3B6RLZ6	peptidasa A1	actividad endopeptidasa de tipo aspártico	Las aspartil proteasas (AP), también conocidas como proteasas ácidas, (EC 3.4.23.-) son una familia ampliamente distribuida de enzimas proteolíticas que se sabe que existen en vertebrados, hongos, plantas y retrovirus. y algunos virus vegetales. Los AP utilizan una díada Asp para hidrolizar los enlaces peptídicos.
	TraesCS7A03G1158100	A0A3B6RLX7 A0A3B6RQR0	transportador ABC	Actividad del transportador tipo ABC.	Los transportadores ABC participan en la exportación o importación de una amplia variedad de sustratos que van desde pequeños iones hasta macromoléculas. La función principal de los sistemas de importación ABC es proporcionar nutrientes esenciales a las bacterias. Se encuentran únicamente en procariotas y sus cuatro dominios constitutivos suelen estar codificados por polipéptidos independientes (dos proteínas ABC y dos proteínas TMD). Los importadores procarióticos requieren proteínas de unión extracitoplasmáticas adicionales (una o más por sistema) para funcionar. Por el contrario, los sistemas de exportación participan en la extrusión de sustancias nocivas, la exportación de toxinas extracelulares y el ataque a componentes de membrana. Se encuentran en todos los organismos vivos y, en general, el TMD está fusionado con el módulo ABC en una variedad de combinaciones. Algunos exportadores eucariotas codifican los cuatro dominios en la misma cadena polipeptídica.
	TraesCS7A03G1161900	A0A3B6RQS9	F-box	-	-
	TraesCS7A03G1161000	A0A3B6RQV7	-	unión de glucosa	-
	TraesCS7A03G1161600	A0A3B6RQW1	-	actividad de la histona H2AQ104 metiltransferasa	-
	TraesCS7A03G1161800	A0A3B6RLZ2 A0A3B6RMS5	HMA	Actividad del transportador transmembrana de cationes	Para evitar la unión y reactividad no específica y/o indeseable de los iones metálicos con componentes

		A0A3B6RQV9		monoatómicos acoplados a ATPasa	celulares, los organismos han desarrollado sistemas específicos de metales para el tráfico de proteínas. Aunque los sistemas difieren, aquellos que manejan iones de metales blandos como Hg(2+), Cu(+), Zn(2+), etc., utilizan proteínas asociadas a metales pesados (HMA) y dominios de ~70 aminoácidos con un motivo GMXCXC conservado en un pliegue estructural $\beta\alpha\beta\alpha\beta$. Si bien las cisteínas conservadas definen un sitio de unión de metal común en estas proteínas, se deben utilizar otras características estructurales para crear iones metálicos, proteínas asociadas y especificidades contextuales.
	TraesCS7A03G1165400	A0A3B6RQU0	F-box	-	Identificada por primera vez en la ciclina-F como un motivo de interacción proteína-proteína, la caja F es un dominio conservado que está presente en numerosas proteínas con una estructura bipartita. A través de la F-box, estas proteínas están unidas a la proteína Skp1 y al núcleo de los complejos SCF (Skp1-cullin-F-box proteína ligasa). Los complejos SCF constituyen una nueva clase de ligasas. Funcionan en combinación con la enzima E2 Cdc34 para ubiquitinar ciclinas G1, inhibidores de Cdk y muchas otras proteínas, para marcarlas para su degradación. La unión de los sustratos específicos mediante complejos SCF está mediada por motivos divergentes de interacción proteína-proteína presentes en las proteínas F-box, como repeticiones WD40, repeticiones ricas en leucina o repeticiones ANK.

Anexo 4. Diagramas de caja indican los valores del análisis físico y químico correspondientes a los distintos grupos de haplotipos de los cromosomas restantes para SDS.



Anexo 4. (continuación).

