



UNIVERSIDAD NACIONAL DEL SUR

DEPARTAMENTO DE BIOLOGÍA, BIOQUÍMICA Y FARMACIA

**IDENTIFICACIÓN DE MARCADORES MOLECULARES
ASOCIADOS A CARACTERES RELACIONADOS CON EL
RENDIMIENTO DE UNA POBLACIÓN F7 (BI Quillen x
Buck Granate) DE TRIGO CANDEAL**

TESIS PARA OPTAR EL TÍTULO DE GRADO EN LICENCIATURA EN CIENCIAS
BIOLÓGICAS

ALUMNA: Rocio Laino

DIRECTOR: Dr. Pablo Federico Roncallo

Bahía Blanca, 26 de Agosto 2024

**IDENTIFICACIÓN DE MARCADORES MOLECULARES
ASOCIADOS A CARACTERES RELACIONADOS CON EL
RENDIMIENTO DE UNA POBLACIÓN F7 (BI Quillen x
Buck Granate) DE TRIGO CANDEAL**

TESIS PARA OPTAR AL GRADO EN LICENCIATURA EN CIENCIAS BIOLÓGICAS

Firmas



Laino Rocio

Alumna



Dr Pablo Federico Roncallo

Director

AGRADECIMIENTOS

- A mi director, el Dr. Pablo Federico Roncallo, por el tiempo, la dedicación, el compromiso y la ayuda durante la realización de este trabajo.
- Al Centro de Recursos Naturales de la Zona Semiárida (CERZOS-CONICET-UNS) por brindarme las instalaciones y medios necesarios para desarrollar la tesina.
- A la Universidad Nacional del Sur (UNS) y al Departamento de Biología, Bioquímica y Farmacia (BByF) por permitir que me forme profesionalmente en una universidad pública y de calidad.
- A mi docente consejera, la Dra. María Celina Zabaloy, por la orientación, paciencia y ayuda en todo momento.
- A mi familia, por la compañía, motivación, apoyo y amor incondicional. Gracias por todo. Este logro es tanto mío como de ustedes, vamos por más logros.
- A mis compañeros y amigos de la vida y de la carrera, con quienes fue posible llevar adelante esta etapa y quienes hicieron que este paso por la universidad sea una experiencia hermosa.
- Una mención especial para una gran compañera que ya no está con nosotros, pero que me acompañó desde los 9 años y fue un gran apoyo en todos los momentos, especialmente en los más difíciles. Gracias Toti, te extrañamos.

ÍNDICE

AGRADECIMIENTOS	3
RESUMEN	5
INTRODUCCIÓN	6
OBJETIVOS E HIPÓTESIS	12
Objetivo general.....	12
Objetivos específicos.....	12
Hipótesis.....	12
ACTIVIDADES Y METODOLOGÍA.....	13
Material vegetal y ensayos a campo	13
Evaluación de caracteres relacionados con el rendimiento	14
Genotipado	14
Análisis de resultados	15
RESULTADOS Y DISCUSIÓN	17
Resultados fenotípicos	
Histogramas.....	17
Medidas resumen	19
Inferencia basada en una muestra	20
Correlación	21
Regresión	21
Resultados genotípicos	
Manhattan plot	24
Estudio de mapeo por asociación en población RIL	26
Q-Q Plot	29
Evaluación del efecto del principal SNP identificado.....	31
Desequilibrio de ligamiento	33
Genes candidatos	36
Análisis de la diversidad genética.....	38
CONCLUSIÓN	40
BIBLIOGRAFÍA	41
ANEXOS.....	46

RESUMEN

El trigo candeal o trigo para fideos (*Triticum turgidum* L. spp. *durum*) (genoma AABB) es una especie tetraploide ($2n=2x=28$), miembro de la familia Poaceae. Las especies parentales más probables que contribuyeron con los genomas A y B respectivamente son *T. urartu* y *Aegilops speltoides*.

Esta especie se cultiva ampliamente en áreas relativamente secas del mundo. El uso principal de este cereal es la elaboración de pastas frescas y secas para la alimentación humana. En Argentina, el trigo candeal ocupa 1,3% de la superficie total de trigo y se localiza principalmente en el sudeste bonaerense (región de influencia de la UNS), aunque el área se encuentra en expansión hacia regiones como el NOA, NEA y la región Centro del país. En los últimos años, la demanda de trigo candeal para la producción de pasta ha aumentado significativamente a nivel mundial. Por lo tanto, garantizar el suministro para la industria de la pasta es una prioridad. Esto puede lograrse incrementando el rendimiento final o también mejorando los componentes del rendimiento.

El objetivo de este trabajo fue caracterizar y analizar la variabilidad genética molecular y fenotípica presente en la población F7 de líneas recombinantes endocriadas (RILs) de trigos candeales argentinos e identificar marcadores moleculares (MM) asociados a componentes del rendimiento. Utilizando la población F7 de RILs, obtenida de la cruce entre las variedades de trigo candeal BonINTA Quillen x Buck Granate, se llevó a cabo una evaluación a campo de caracteres relacionados con el rendimiento y el genotipado de la población RIL y sus parentales mediante marcadores DArTseq.

Para los caracteres analizados, se estimó un valor medio de 5 espigas por planta, 15 espiguillas por espiga, 41 granos por espiga y 3 granos por espiguilla. A su vez, todas las variables analizadas (NEP, NEE, NGE y NGe) mostraron presencia de segregación transgresiva. Los resultados del análisis de correlación mostraron una fuerte correlación positiva (y estadísticamente significativa) entre todas las variables estudiadas, a excepción de la fecha de espigazón, la cual presentó valores bajos de correlación con el resto de los caracteres. Por otra parte, los análisis de regresión lineal mostraron una asociación positiva (y estadísticamente significativa) entre todas las variables, a excepción de la fecha de espigazón vs. NEP, NGE y NGe. La evaluación fenotípica y de marcadores SNP sobre la población RIL se utilizó en un estudio de mapeo por asociación destacando la identificación de tres marcadores asociados a más de un carácter, estos fueron 1110438|F|0-14:C>T-14:C>T (asociado a los caracteres NEP, NEE y NGE), 1091747|F|0-52:G>A-52:G>A y 1140354|F|0-24:G>T-24:G>T (ambos asociados a los caracteres NGE y NGe). Por otro lado, se estimó la existencia de desequilibrio de ligamiento entre marcadores presentes en dos regiones independientes del cromosoma 2A, en la región comprendida entre 45851839 y 51847409 pb y la región comprendida entre 155070806 y 161696203 pb. Al analizar la diversidad genética dentro de la población se encontró que la misma está dividida en tres subgrupos, con un valor

mínimo de disimilaridad de 0.041 y un valor máximo de 0.838. El subgrupo 1 incluyó 68 RILs y los parentales, mientras que los subgrupos 2 y 3 incluyeron 55 y 5 RILs respectivamente.

INTRODUCCIÓN

Problemática y estado actual del conocimiento

Debido al aumento poblacional previsto para el 2050 y la mejora en la calidad de la dieta en los países en desarrollo y nuevos países emergentes, se espera un aumento en la demanda global de cereales (Fischer *et al.*, 2014). Dado que el área sembrada con trigo se encuentra estabilizada, e intentar ampliarla implicaría utilizar nuevas áreas con una alta susceptibilidad a la degradación (Andrade 2011), los principales esfuerzos para incrementar la producción mundial se han concentrado en aumentar el rendimiento potencial y real del cultivo (Fischer *et al.*, 2014; FAO 2018).

1. El trigo (*Triticum spp.*)

El trigo es una especie de la sub-tribu Triticinae dentro de la tribu Triticeae, perteneciente a la familia de gramíneas Poaceae. La sub-tribu Triticinae es de origen reciente, y contiene aproximadamente 35 géneros incluidos *Triticum*, *Aegilops*, *Thinopyrum*, *Dasypyrum*, *Lophopyrum* y *Secale*. Los trigos cultivados pertenecen a cuatro especies del género *Triticum*. Las especies poliploides de trigo son verdaderos alopoliploides, cuyo número cromosómico corresponde a un múltiplo del juego básico haploide (x) de siete cromosomas por gameto. Las especies de número cromosómico más bajo, como por ejemplo *T. monococcum* son normalmente diploides ($2n=2x=14$). Otras dos especies, como *T. turgidum* y *T. timopheevi*, son tetraploides ($2n=4x=28$). La especie cultivada de *Triticum* con número cromosómico más elevado, *T. aestivum*, es hexaploide ($2n=6x=42$).

1.1 *Triticum turgidum L. spp durum*

El trigo candeal o trigo para fideos (*T. turgidum L. spp. durum*) (genoma AABB) es un miembro de la familia Poaceae, tribu Triticeae, que evolucionó en Medio Oriente vía hibridación interespecífica y duplicación cromosómica, seguida de domesticación. Las especies parentales *T. urartu* y *Aegilops speltoides* contribuyeron con los genomas A y B respectivamente (Bozzini, 1988).

T. turgidum L. spp. durum se cosecha ampliamente en áreas relativamente secas del mundo, tales como la cuenca mediterránea, India, Rusia y otras repúblicas ex-soviéticas, las zonas de escasa pluviometría de las grandes llanuras de Norteamérica y Australia (Roncallo *et al.*, 2009).

El uso principal de este cereal es la alimentación humana, siendo una de las fuentes más abundantes de energía y proteínas. El gran valor nutritivo del candeal se debe a su alto contenido de proteínas (12 -

14%), aunque alcanza menores rendimientos que el trigo pan. El trigo candeal es utilizado principalmente para la elaboración de pastas secas (vermicelli, spaghetti y macarroni) y frescas. Existen otros productos como el cuscús, que se consume especialmente en el norte de África y el burgol, el frekeh y los chapatis que se elaboran con trigo entero y son muy populares en países de Medio Oriente, África, India y Pakistán (Patnaik & Khurana, 2001). En los países mediterráneos una gran parte de la cosecha de trigo candeal es dedicada a la preparación de pan (Quaglia, 1988; Blanco *et al.*, 2004) y de postres (Seghezze & Molfese, 1999). Sin embargo, el producto final por excelencia del candeal es la pasta en sus diferentes formas.

T. turgidum L. spp. durum posee características superiores para la elaboración de pasta de buena calidad. La calidad puede ser definida como la habilidad de los granos de trigo o sémola para satisfacer los requerimientos específicos de los usuarios. El contenido de proteína en el grano, la fuerza y elasticidad del gluten y el color son los principales factores involucrados en la calidad de la sémola de trigo candeal y estos redundan en mejores cualidades reológicas de la masa. El porcentaje de granos vítreos (vitreosidad) es también considerado por la industria fideera debido a su correlación positiva con el porcentaje de proteína, el rendimiento de sémola y la calidad de cocción. Mientras que la calidad del gluten, color de la sémola y cualidades reológicas de la masa están determinadas fundamentalmente por el genotipo, el contenido de proteína y la vitreosidad resultan más afectadas por el ambiente (Tróccoli *et al.*, 2000).

La importancia del trigo candeal a nivel mundial es sustancialmente mayor en algunos países de la Unión Europea (Italia, España, Francia y Grecia); de América (Canadá, Estados Unidos, México y Argentina); de África (Argelia, Marruecos y Túnez); Asia (India) y Australia; y otros como Rusia y Turquía (<http://www.fas.usda.gov>).

El trigo es el producto agrícola de mayor producción mundial correspondiendo el 95% al trigo pan. Casi la totalidad del 5% restante es trigo candeal, y ocupa una superficie de aproximadamente 14-16 millones de hectáreas (Tróccoli *et al.*, 2000). En la cosecha 2004/05 se llegó al récord de producción mundial (aproximadamente 33 millones de toneladas), que cayó bruscamente en la cosecha siguiente (26 millones), debido principalmente al desaliento de siembra por las reservas acumuladas del año anterior, y a la gran sequía que afectó a los principales países productores (Roncallo *et al.*, 2009).

En Argentina, el trigo candeal ocupa solo el 1,3% de la superficie total de trigo y se destina principalmente a la elaboración de pastas secas. El cultivo de candeal se localiza principalmente en el sudoeste bonaerense (región de influencia de una UNS), aunque el área se encuentra en expansión hacia regiones menos tradicionales como el NOA (Santiago del Estero, Tucumán, Salta), el NEA (Formosa) y la región Centro del país (San Luis, Córdoba, Santa Fe) (<http://datosestimaciones.magyp.gob.ar/>).

En el período 2019-2023, el rendimiento medio en Argentina de trigo candeal fue de 2,92 t/ha (<https://datosestimaciones.magyp.gob.ar/>). El menor rendimiento respecto al trigo pan es una clara

desventaja competitiva que desalienta su siembra. Es por ello que contribuir a descifrar la base genética del rendimiento en trigo candeal es un paso necesario para aumentar su producción. La tasa de ganancia genética anual alcanzada en trigo candeal ha sido de 0,81% año⁻¹ (1934-2015) equivalente a 26,94 kg ha⁻¹ año⁻¹.

2. Caracteres relacionados con el rendimiento

El rendimiento de grano es la expresión final de los múltiples y complejos procesos fisiológicos individuales de una planta que ha interactuado con el ambiente durante las temporadas de crecimiento del cultivo. La aparición de estreses ambientales durante el alargamiento del tallo, la época de floración y el llenado del grano es perjudicial para el trigo candeal, provocando importantes pérdidas de rendimiento y afectando negativamente la calidad del grano y las harinas. En los últimos años, la demanda de trigo duro para la producción de pasta ha aumentado significativamente a nivel mundial. Por lo tanto, garantizar el suministro para la industria de la pasta es una prioridad absoluta (Taranto *et al.*, 2023). Esto puede lograrse incrementando el rendimiento final modificando la morfología de la planta, fenología y los componentes del rendimiento, entre otros.

En trigo, el incremento en el rendimiento potencial alcanzado en los últimos años se logró aumentando el número de granos por unidad de área, estrechamente vinculado a los granos producidos por espiga, y particularmente incrementando el número de granos por espiguilla, sin una tendencia consistente en espigas por unidad de superficie (Lo Valvo *et al.*, 2018; Achilli *et al.*, 2022). El carácter número de granos por espiguilla se genera durante el periodo previo a la antesis del trigo. Este carácter se determina mediante la combinación de varios subcomponentes numéricos, incluidos el número de espigas por planta (NEP), y el número de granos por espiga (NGE) (Taranto *et al.*, 2023).

Otros caracteres cuyo incremento podría representar un aumento del rendimiento en trigo candeal (ya sea porque presentan una asociación positiva con el número de granos o porque benefician el desarrollo de la planta) son altura de planta, fecha de espigazón, entre otros.

En las zonas vulnerables a las tensiones ambientales y al cambio climático, es importante comprender los mecanismos genéticos, bioquímicos y fisiológicos que regulan el rendimiento de los cereales y desarrollar una nueva generación de variedades de trigo duro o candeal, más productivas incluso en condiciones ambientales extremas.

3. Marcadores moleculares

La variación genética en individuos se puede presentar en distintos niveles. En uno de los extremos están las variantes genéticas manifestadas con efectos fácilmente distinguibles en el fenotipo (caracteres cualitativos). El otro extremo ocurre cuando una gran parte de la variación entre individuos no da lugar a clases fenotípicas naturalmente definidas (caracteres cuantitativos) por ejemplo, el número de granos por espiga. El supuesto biológico detrás de la variabilidad de los caracteres cuantitativos es que su

expresión está regulada por una gran cantidad de componentes genotípicos y sus interacciones con el ambiente. Para comprender las variaciones genéticas detrás de los caracteres cuantitativos se emplean técnicas como los marcadores moleculares, que permiten captar las diferencias genéticas de los individuos en el ADN, para luego, estudiar la expresión de estos genotipos en distintos ambientes (Paccapelo, 2015).

Las variaciones en las secuencias de ADN se pueden detectar utilizando una amplia variedad de técnicas (Lynch & Walsh, 1998) desde la secuenciación de ADN hasta caracterizaciones más precisas (más rápidas y menos costosas) logradas a través de distintos marcadores moleculares.

Uno de los marcadores moleculares más utilizados en la actualidad es el SNP (polimorfismo de nucleótido único). Estos marcadores identifican la variación de nucleótidos en un locus en el genoma. Dada la forma en que se disponen los pares de nucleótidos (A-T, C-G), en la práctica los SNPs son típicamente bialélicos, es decir, tienen dos alelos posibles A o T por un lado o bien, C o G por otro. Por lo tanto, los genotipos de un SNP pueden ser AA, AT/TA y TT o bien CC, CG/GC y GG.

Cuando los loci se heredan de manera independiente unos con otros, es decir, no existe asociación entre los loci, se dice que se produjo recombinación genética. Sin embargo, cuando dos o más loci están cercanos dentro de un cromosoma, estos pueden heredarse juntos y a este fenómeno de asociación se lo conoce como ligamiento. Los marcadores moleculares permiten no sólo describir el genoma de un individuo, sino también la construcción de mapas de ligamiento, con base en el estudio de la ocurrencia de recombinaciones. Estos mapas dan un ordenamiento a los marcadores y pueden ser utilizados para localizar regiones del cromosoma que contienen genes que controlan caracteres de interés (Paccapelo, 2015).

4. Genes que codifican para caracteres relacionados con el rendimiento

En las últimas dos décadas, se ha estudiado la base genética del rendimiento y los componentes del rendimiento en el trigo candeal gracias al mapeo de ligamiento basado en poblaciones biparentales y estudios de asociación de todo el genoma (GWAS), que utilizan marcadores moleculares basados en el ADN, como microsatélites o SSR (Repeticiones de secuencias simples), tecnologías de matriz de diversidad (DArT) y polimorfismos de nucleótido simple (SNP), entre otros.

Estos estudios llevaron a la identificación de cientos de loci para caracteres cuantitativos (QTL) en diferentes poblaciones de mapeo genético y con diferentes tipos de marcadores moleculares, pero con un impacto limitado en el rendimiento del grano.

Los rápidos avances en las tecnologías de secuenciación de próxima generación (NGS), junto con el ensamblaje mejorado del genoma del trigo y las plataformas de genotipado de alto rendimiento, así como la tecnología de edición del genoma, están facilitando la identificación de genes que controlan rasgos agronómicos críticos como el rendimiento y sus componentes. Esto significa que, para un rasgo genético complejo como el rendimiento del grano, controlado por múltiples loci/genes y fuertemente influenciado

por el entorno de crecimiento, se debe tener en cuenta un enfoque holístico más apropiado considerando todo el sistema de la planta de trigo, en lugar del estudio de los componentes individuales del rendimiento (Taranto *et al.*, 2023).

4.1 Estrategias de mapeo de QTL para el rendimiento y sus componentes en trigo candeal

La estrategia más poderosa y utilizada para descomponer los componentes del rendimiento de granos fue y sigue siendo el mapeo QTL, a través de enfoques de mapas de ligamiento. El mapeo de QTL se basa en la identificación de marcadores moleculares asociados con un rasgo fenotípico determinado en una población segregante, permitiendo así el posicionamiento de QTL dentro de mapas genéticos. La identificación de marcadores moleculares para QTL específicos puede ayudar a mejorar y acelerar los programas de mejoramiento.

4.2 Marcadores moleculares DArT

DArT es un método para el análisis del polimorfismo del ADN. Fue desarrollado por primera vez en arroz (Jaccoud *et al.*, 2001), aunque ha sido aplicado a más de 60 especies (www.diversityarrays.com). Este tipo de tecnología, para el análisis genético, ofrece un coste bajo, un alto rendimiento y un sistema robusto. Para obtener estos marcadores no es necesario tener conocimiento previo de la secuencia y, por lo tanto, ofrecen una alternativa interesante para el trigo. De hecho, esta tecnología ha sido aplicada con éxito en diferentes especies, con poca información genética como la avena (Tinker *et al.*, 2009), la banana (Risterucci *et al.*, 2009), el centeno (Bolibok-Bragoszewska *et al.*, 2009), entre otros. Por todo esto, las aplicaciones potenciales de los marcadores DArT incluyen: caracterización de germoplasma, mapeo genético, detección de genes, mejora asistida por marcadores y búsqueda de cambios de metilación en el genoma (Domínguez García, 2012).

Los marcadores moleculares DArT son marcadores bialélicos dominantes (presencia versus ausencia) o hemi dominantes (presentes en dosis doble, dosis simple o ausente). El tipo de clasificación del marcador depende tanto de los métodos estadísticos aplicados como del marcador específico (Kilian *et al.*, 2012).

4.3 Genotipificación por secuenciación, usando marcadores DArTseq

El genotipado por secuenciación usando marcadores DArTseq, por sus siglas en inglés, es un sistema altamente multiplexado que genera cientos de miles de marcadores de tipo SNPs (Polimorfismo de nucleótido simple) y PAVs (Variación de Presencia/Ausencia), que representan un perfil reducido del genoma y para ello se utilizan plataformas de secuenciación de última generación (NGS) (Ortega Jiménez, 2017). Para desarrollar la técnica DArTseq, se utilizan enzimas de restricción para reducir la complejidad del genoma y evitar la fracción repetitiva del mismo (Elshire *et al.*, 2011).

El genotipado con DArTseq se inicia definiendo un grupo de muestras de ADN representativas del germoplasma que se prevé analizar con la tecnología. Este grupo de muestras generalmente corresponde al acervo genético primario de una especie de cultivo, pero puede restringirse a los dos padres de una cruce o expandirse a acervos genéticos secundarios e incluso terciarios (Kilian *et al.*, 2012). Esta colección de ADNs se procesa utilizando enzimas de restricción y se selecciona una fracción definida de fragmentos genómicos, que se utiliza luego para construir una biblioteca de clones en *E. coli*. El proceso de clonación en bacterias cumple una función de “individualización” de las sondas obtenidas a partir de la colección resultante de fragmentos genómicos. Los insertos de los clones individuales se amplifican y se utilizan como sondas moleculares en matrices DArT (Kilian *et al.*, 2012).

La principal ventaja de DArTseq radica en la capacidad de obtener marcadores de una manera rentable y de alto rendimiento, así también como la capacidad de localizar físicamente al marcador en el genoma (Kilian *et al.*, 2012). Hasta el 90 % de los marcadores DArTseq se alinean de manera única con un genoma de referencia.

5. Líneas recombinantes endocriadas:

Para el desarrollo de líneas recombinantes endocriadas (RILs), el primer paso es la obtención de un híbrido inter o intraespecífico entre el parental donante y el parental recurrente. Luego, este híbrido se autofecunda (como alternativa, se cruzan dos plantas del híbrido) y las plantas así obtenidas continúan autofecundándose hasta obtener un conjunto de líneas homocigotas. Como una variante de este método, se pueden intercalar generaciones en las que se cruzan las plantas de la misma generación, finalizando con generaciones de autofecundación para generar homocigotas (Pollard, 2012). Finalmente, cada una de las plantas homocigotas diferirá de las demás en la combinación de alelos (aportados por los dos parentales) que contenga.

Las RILs, debido a su alto grado de homocigosis, constituyen un material muy estable (Bailey, 1971) y esto permite disponer de varias réplicas por experimento o evaluar caracteres en distintos ambientes. Las principales características de las líneas endocriadas son isogenicidad, homocigosis, individualidad, uniformidad fenotípica y estabilidad a largo plazo.

La uniformidad fenotípica de las RILs es un rasgo importante que nos permite afirmar que la variabilidad en los parámetros experimentales se debe exclusivamente a factores no genéticos (ambientales o metodológicos). Esta uniformidad nos permite lograr precisión estadística con muchas menos plantas que las necesarias al trabajar con grupos no consanguíneos. La uniformidad hace posible la comparación de resultados experimentales entre especies de diferentes ambientes y, además, debido a su relativa estabilidad genética, a lo largo del tiempo (Takuno *et al.*, 2012).

OBJETIVOS E HIPÓTESIS

Objetivo:

Caracterizar y analizar la variabilidad genética molecular y fenotípica presente en la población F7 de líneas recombinantes endocriadas (RILs) (BI Quillen x Buck Granate) de trigos candeales argentinos (*T. turgidum* L. spp. *durum*) e identificar marcadores moleculares asociados a componentes del rendimiento.

Objetivos específicos:

1. Evaluar componentes de rendimiento en espiga en las RILs sobre un ensayo a campo.
2. Analizar la diversidad genética de la población RIL utilizando técnica DArTseq
3. Utilizar herramientas bioinformáticas en la búsqueda de asociación entre los marcadores moleculares y caracteres relacionados con el rendimiento e identificar posibles genes candidatos.

Hipótesis:

Los fondos genéticos presentes en la población BI Quillen x Buck Granate poseen genes con efecto positivo que incrementan el rendimiento del cultivo, afectando a componentes cuantitativos del mismo tales como el número de espigas por planta, número de espiguillas por espiga, número de granos por espiga y número de granos por espiguilla.

ACTIVIDADES Y METODOLOGÍA

Material vegetal y ensayos a campo:

Se utilizó una población F7 de líneas recombinantes endocriadas (RILs) de la cruce entre las variedades de trigo candeal BI Quillen x Buck Granate. Se dispone de muestras tomadas sobre un ensayo a campo sembrado en el campo experimental del CERZOS (Bahía Blanca, Buenos Aires) el día 8 de agosto de 2023 en plantineras. La fecha de emergencia de las plantas fue el 14 de agosto del mismo año y las mismas fueron trasplantadas a campo.

En dicho ensayo se sembraron 128 RILs y los parentales (BI Quillen y Buck Granate), utilizando como unidad experimental parcelas de un surco de 2 m., espaciadas a 0,20 m., y 10 plantas por parcela. Las parcelas fueron aleatorizadas utilizando un diseño en alpha-lattice (Patterson y Williams, 1976) con dos repeticiones. Cada repetición incluyó 13 bloques incompletos de 10 parcelas cada uno y se ubicó físicamente sobre un tablón de 2 m. de ancho por 26 m. de largo, dentro de un campo experimental. Un bloque incompleto de cada réplica (R1-7 y R2-7, Figura 1) se trasplantó sobre un tercer tablón, debido a la organización del campo experimental.

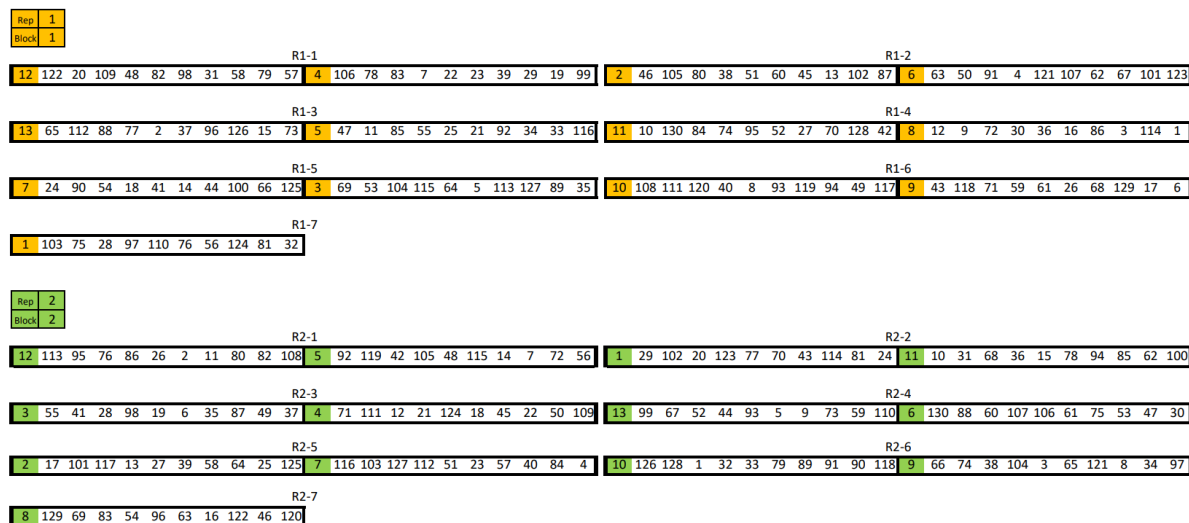


Figura 1. Diseño experimental según el diseño alpha-lattice (13x10) con dos repeticiones. En naranja y verde se indican el número de bloque incompleto de las réplicas 1 y 2 respectivamente.

Se aplicó riego complementario entre el 15 de agosto al 15 de diciembre y una fertilización a la siembra con 200 kg/ha de DAP y 300 kg/ha de Urea al macollaje. Se cosecharon tres espigas por parcela, tomando las espigas del macollo principal sobre tres plantas representativas de la parcela.

Evaluación de caracteres relacionados con el rendimiento:

Evaluaciones fenotípicas en campo:

- a) Altura de planta (cm) [ALT]: fue medida en el campo a la madurez como la altura en centímetros desde el suelo hasta la punta de las espigas, sobre tres tallos principales por parcela. Para la medición se utilizó una regla de 100 cm.
- b) Fecha de espigazón [FE]: medida como el número de días desde la siembra de los RILs y parentales (08/08/2023) hasta el momento en que el 50% de las espigas emergen de la vaina de la hoja bandera a la mitad de su longitud (Z 5.5) (Zadok *et al.*, 1974).
- c) Número de espigas/planta [NEP]: fue medido por conteo sobre las tres plantas representativas y sobre las cuales se tomó posteriormente la muestra de la espiga.

Evaluaciones fenotípicas a medir en espiga: estos caracteres fueron medidos sobre tres espigas por parcela, las cuales fueron previamente seleccionadas al azar y siendo representativas de la parcela, y marcadas (con un hilo distintivo) para luego ser posteriormente cosechadas.

- d) Número de espiguillas por espiga [NEE]: fue medido como el conteo de espiguillas en cada espiga.
- e) Número de granos por espiga [NGE]: fue medido por el conteo de los granos luego de la trilla de cada espiga.
- f) Número de granos por espiguilla [NGe]: fue obtenido por cálculo utilizando la siguiente fórmula
$$NGe = NGE / NEE$$

Genotipado de la población RIL y sus parentales:

Sobre tejido fresco de cada RIL y sus progenitores se extrajo ADN genómico llevando a cabo un protocolo de extracción en pequeña escala de ADN de muy alta calidad descrito por Dreisigacker *et al.* (2016) y desarrollado en el Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT®, www.cimmyt.org).

Una vez obtenido el ADN genómico de cada RIL y sus progenitores, estos fueron caracterizados utilizando la técnica DArTseq (Kilian *et al.*, 2012) en colaboración con la Dra. Carolina Saint Pierre (CIMMYT, México).

Análisis de resultados:

Variables fenotípicas

La variabilidad de los caracteres relacionados con el rendimiento (medidos) y su relación con las variables ya disponibles (altura y fecha de espigazón) se evaluó utilizando el software InfoStat (<http://www.infostat.com.ar/>) versión estudiantil.

Se obtuvo el promedio de cada variable por parcela, en base a las tres espigas o plantas analizadas, y luego se analizaron los siguientes caracteres: fecha de espigazón, altura promedio, número de espigas por planta promedio (NEP promedio), número de espiguillas por espiga promedio (NEE promedio), número de granos por espiga promedio (NGE promedio), número de granos por espiguilla promedio (NGe promedio). Cada uno de ellos se evaluó teniendo en cuenta el genotipo del individuo, el bloque incompleto y la repetición en que este se hallaba.

La variabilidad existente dentro y entre los caracteres se evaluó utilizando diversas herramientas estadísticas (Libro de apuntes de Bioestadística, UNS).

Evaluación de los marcadores moleculares obtenidos

Los marcadores moleculares obtenidos en cada RIL y sus progenitores, a través de la técnica DArTseq, se analizaron utilizando la planilla de cálculo Excel, dentro del paquete OFFICE. Se realizó el filtrado de los datos de ADN, para eliminar aquellos que no cumplan con estándares de calidad necesarios para los análisis de asociación fenotipo-genotipo. Sobre los marcadores moleculares de tipo SNP, se eliminaron aquellos que fueran monomórficos (una única variante alélica), aquellos que tuvieran más de 10% de heterocigotos, más de 10% de datos perdidos o donde la frecuencia del menor alelo (MAF) fuera menor al 5%. Además, se eliminaron aquellos marcadores SNP que mostraron ser homocigotos en los padres.

Búsqueda de marcadores asociados mediante la metodología de mapeo de asociación

Para este análisis se utilizaron herramientas dentro del programa TASSEL v.5.2 (<https://www.maizogenetics.net/tassel>). La búsqueda de asociación entre marcadores moleculares (SNP) y los caracteres fenotípicos (NEE, NGE, NGe y NEP) se obtuvo mediante los modelos lineal general (GLM) y un modelo lineal mixto (MLM [PCA+K]) (Yu *et al.*, 2006). Se realizó un Análisis de Componentes Principales (PCA) para corregir desviaciones en los resultados debidas a la estructura de la población RIL y se estimó la matriz de parentesco (K) entre los genotipos. Ambos resultados se utilizaron en el análisis de mapeo de asociación (PCA+K) y finalmente se consideró este modelo para describir los resultados de asociación.

Búsqueda de genes candidatos en base a los marcadores SNP identificados

La búsqueda de genes candidatos se realizó en base a la posición física de los marcadores identificados y la informada para las anotaciones funcionales de genes sobre el genoma de trigo candeal (https://figshare.com/articles/dataset/Durum_Wheat_cv_Svevo_annotation/6984035).

Los genes candidatos para los marcadores SNP asociados a los caracteres fenotípicos, se seleccionaron de una distancia de 80 Kb alrededor del marcador. Se consideraron las principales asociaciones identificadas en los cromosomas 2A, 2B y 6A y 7B.

Análisis de la diversidad genética

Para el análisis de la diversidad genética se utilizó el software DARwin (<https://darwin.cirad.fr/>), el árbol filogenético con el que se analizó dicha diversidad se construyó utilizando el índice Simple Matching y el método de asociación Ward (varianza mínima). Posteriormente se editó el árbol filogenético obtenido con el software Fig Tree (<http://tree.bio.ed.ac.uk/software/figtree/>).

RESULTADOS Y DISCUSIÓN

1. Resultados Fenotípicos

a) Histogramas de frecuencias:

Para cada una de las variables analizadas se obtuvo un histograma de frecuencias (Figuras 2-5). Para estos análisis se utilizaron los datos promedio de cada una de las variables, considerando las dos repeticiones de cada genotipo.

Sobre cada uno de los histogramas se identificaron los genotipos parentales utilizando líneas verticales (líneas de corte), con la finalidad de analizar la presencia o ausencia de segregación transgresiva.

Es importante aclarar que, en genética, la segregación transgresiva es la formación de fenotipos extremos o genotipos transgresivos, observados en poblaciones híbridas segregantes, comparados con los fenotipos de las líneas progenitoras. La presencia de estos fenotipos transgresivos (extremos) puede ser positiva o negativa en términos de aptitud (Johansen & Latta, 2006).

Las líneas híbridas mostraron mayor variación genética y en la expresión de genes en sus progenitores. Como resultado, las líneas híbridas tendrán algunas características que son transgresivas (extremas) en la naturaleza en relación a sus progenitores (Velasco López, 2013).

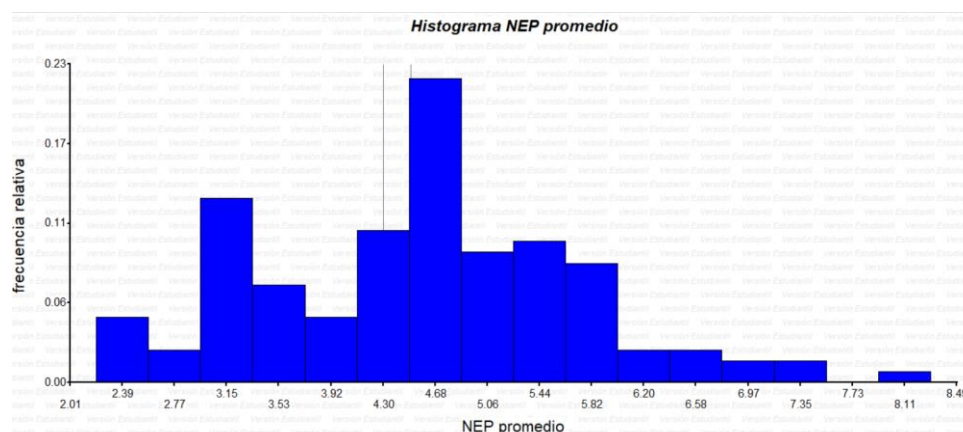


Figura 2. Histograma de frecuencias para el número de espigas por planta (NEP) utilizando valores promedio

La Figura 2 muestra el histograma de frecuencias para la variable número de espigas por planta (NEP), en la misma se observa la presencia de segregación transgresiva, la cual es evidente debido a la presencia de RILs cuyo número de espigas por planta supera o es menor al de los parentales (NEP del genotipo parental BonINTA Quillen: 4.3; NEP del genotipo parental Buck Granate: 4.5).

En el mismo se observa una distribución de normal, teniendo en cuenta el supuesto de normalidad.

Por otro lado, se observan valores atípicos (alejados de la mayoría de los valores y con menor frecuencia) los mismos se encuentran en el rango entre 7,95 a 8,29 espigas por planta. Estos corresponden con los genotipos 19, 28, 55, 62 y 87 los cuales estaban ubicados cercanos entre sí en la réplica dos, pero a una distancia considerable del riego por lo que podría tratarse de una distribución desigual de nutrientes en

el suelo y de genotipos que acumularon alelos favorables de distintos loci. Estos genotipos, bajo las condiciones ambientales de siembra, lograron desarrollar un número de granos por espiguilla cercanos a la media.

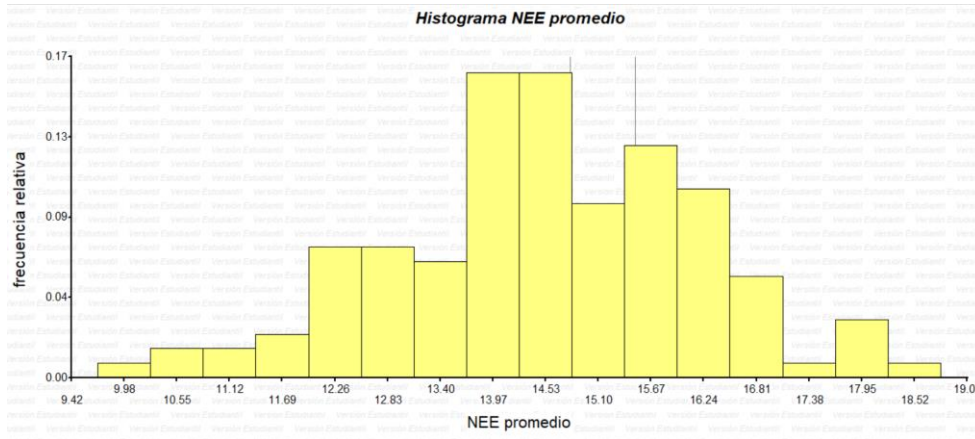


Figura 3. Histograma de frecuencias para el número de espiguillas por espiga (NEE) utilizando valores promedio

En la Figura 3 se presenta el histograma de frecuencias obtenido para el análisis del número de espiguillas por espiga (NEE) en el cual se observa una relativa simetría, por lo que se asume que esta variable sigue una distribución de tipo normal (gaussiana). Esto nos indica que, para esta variable, podemos conocer la probabilidad de encontrar un valor de la variable que sea igual o inferior a un cierto valor, conociendo la media, el desvío estándar y la varianza de un conjunto de datos (al sustituirlos en la función que propone el modelo)

Por otro lado, se observa que existe la presencia de segregación transgresiva ya que vemos un gran número de RILs con un NEE inferior y superior al de los parentales (NEE genotipo parental BonINTA Quillen: 14.8; NEE genotipo parental Buck Granate: 15.5).

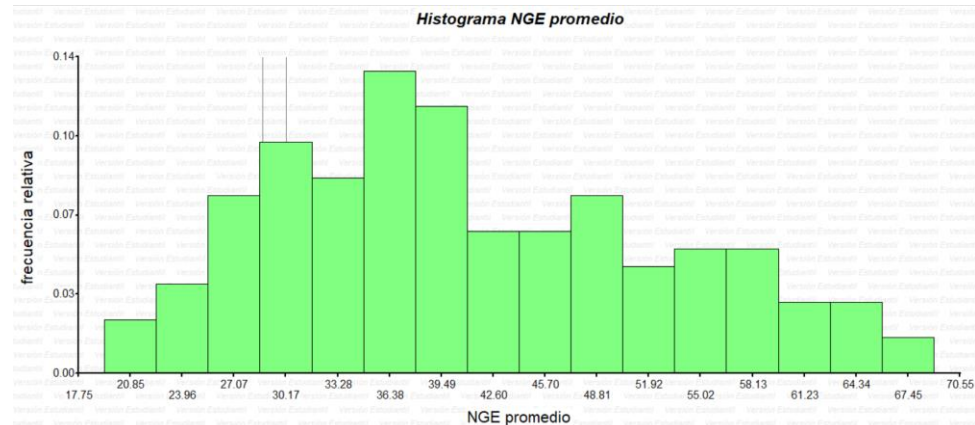


Figura 4. Histograma de frecuencias para el número de granos por espiga (NGE) utilizando valores promedio

La Figura 4 muestra el Histograma de frecuencias obtenido para la variable NGE promedio (número de granos por espiga promedio). En este histograma se observa una distribución de tipo bimodal (Moda 1: 30,17; Moda 2: 38,01) y podría tratarse de dos alelos que se encuentran afectando la variable analizada.

En el histograma también se puede observar la presencia de segregación transgresiva, la cual se vuelve evidente por la presencia de genotipos con un número de granos por espiga superior e inferior al de los individuos parentales (NGE del genotipo parental BonINTA Quillen: 28.8; NGE del genotipo parental Buck Granate: 30.2).

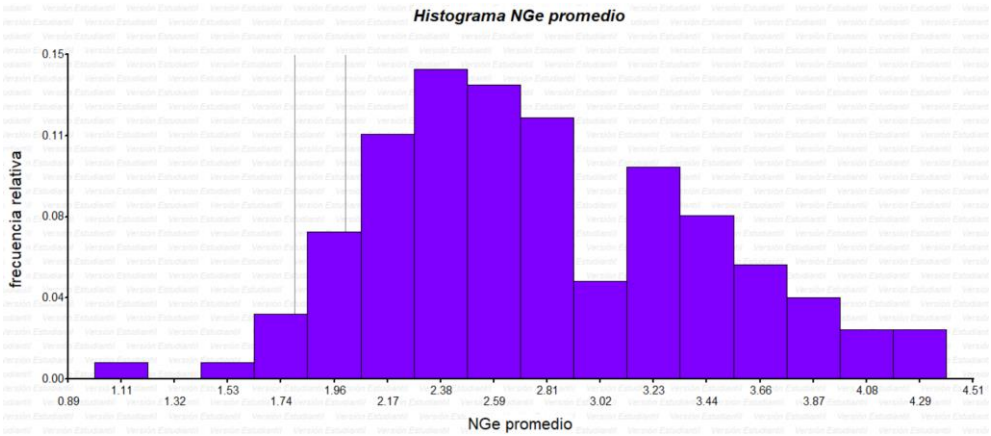


Figura 5. Histograma de frecuencias para el número de granos por espiguilla (NGe) utilizando valores promedio

En el histograma obtenido para el análisis del número de granos por espiguilla (Figura 5) se evidencia una distribución normal y la presencia segregación transgresiva ya que se observan RILs con un NGe inferior y superior al de los parentales (NGe genotipo parental BonINTA Quillen: 1.8; NGe genotipo parental Buck Granate: 2.00). Además, se observan valores atípicos tal como se mencionó para la Figura 2. Estos valores se encuentran en el rango entre 1,02 a 1,31 granos por espiguilla. La causa de estos valores podría deberse a diferencias en la disposición del riego, la variación de nutrientes en el suelo o a factores genéticos. Podría ser que estos genotipos, bajo las condiciones ambientales de siembra, no lograron desarrollar un número de granos por espiguilla cercanos a la media.

b) Medidas resumen:

En la tabla 1 se presentan los resultados obtenidos para la media, el desvío estándar (D.E), valores mínimos y máximos y la mediana para cada una de las variables analizadas.

Tabla 1. Medidas resumen (número de parcelas, media, desvío estándar, mínimos, máximos y mediana) obtenidos con InfoStat

Variable	n	Media	D.E	Mín.	Máx.	Mediana
NEP promedio	257	4,54	1,71	1,00	10,00	4,30
NEE promedio	253	14,49	2,10	7,70	20,00	14,70

NGE promedio	253	40,86	15,10	7,00	82,70	41,00
NGe promedio	252	2,77	0,86	0,80	4,80	2,70

n= número de parcelas; D.E= desvío estándar; Mín= valor mínimo; Máx= valor máximo

c) Inferencia basada en una muestra:

En la Tabla 2 se observan los resultados para el análisis de Inferencia basada en una muestra, en el mismo se obtuvieron resultados para la Prueba de Shapiro-Wilks, D.E, media y coeficiente de variación (C.V). Los datos utilizados para el análisis fueron obtenidos a partir del cálculo del promedio (para cada una de las variables) de las dos repeticiones para cada genotipo.

Tabla 2. Datos obtenidos a partir del análisis de inferencia basada en una muestra con InfoStat

Variable	n	Media	D.E	C.V.	p (Unilateral D)
NEP promedio	130	4,55	1,16	25,49	0,1726
NEE promedio	130	14,50	1,70	11,72	0,7981
NGE promedio	130	41,08	11,54	28,09	<0,0001
NGe promedio	130	2,77	0,65	23,46	0,0746

C.V.=coeficiente de variación $[(\sigma/\mu)*100]$

Como se muestra en la tabla, a partir de los promedios calculados y con el uso del software InfoStat, se obtuvo una media de, aproximadamente, 5 espigas por planta, 15 espiguillas por espiga, 41 granos por espiga y 3 granos por espiguilla.

Los valores de D.E para las variables NEP promedio, NEE promedio y NGe promedio resultaron bajos, por lo que los datos obtenidos para estas variables en la población estarían cercanos a la media. Esto no ocurre con la variable NGE promedio, cuyo valor de D.E es alto, lo que indica dispersión en los datos y variabilidad en las observaciones (estas se alejan al valor de la media obtenido). Esto puede respaldarse con los resultados de C.V., para todas las variables, excepto NGE promedio, estos valores resultaron alejados al 30% (lo que significa que la media aritmética es representativa del conjunto de datos), mientras que, para la variable antes mencionada, el C.V. resulto cercano al 30% indicando que el promedio no es representativo del conjunto de datos.

Estos resultados se respaldan con la prueba de Shpiro-Wilk, para la cual se obtuvo un valor p . Esta prueba indica que la variable NGE promedio no sigue una distribución normal ($p<0.05$, rechazo H_0).

d) Correlación entre variables:

Los datos utilizados para el análisis de correlación se obtuvieron promediando las dos repeticiones (parcelas) por genotipo y para cada una de las variables. En la Tabla 3 se muestran los resultados del análisis del Correlación de Pearson considerando las variables estudiadas.

Tabla 3. Análisis de Correlación de Pearson: Coeficientes/Probabilidades

	Espigazón	Altura promedio	NEP promedio	NEE promedio	NGE promedio	NGe promedio
Espigazón	1,00	0,0027	0,52	3,1 E-08	0,24	0,27
Altura promedio	0,26	1,00	0,00	0,00	0,00	2,8 E-07
NEP promedio	0,06	0,78	1,00	0,00	0,00	7,9 E-10
NEE promedio	0,46	0,68	0,62	1,00	0,00	3,1 E-05
NGE promedio	0,10	0,60	0,63	0,65	1,00	0,00
NGe promedio	-0,10	0,43	0,51	0,36	0,92	1,00

En el triángulo inferior se muestran los coeficientes de correlación mientras que en el triángulo superior se presentan los valores *p* de significancia estadística

Las relaciones significativas entre variables encontradas fueron positivas. Los resultados del análisis de correlación exponen una fuerte asociación entre las variables Altura promedio y NEP promedio, lo cual refleja que, ante una mayor altura de planta, mayor será el número de espigas por planta. Otras variables que mostraron una fuerte asociación fueron NGE promedio y NGe promedio, esta última asociación indicaría que, a mayor número de granos por espiga, mayor es el número de granos por espiguilla. Por otro lado, los resultados demostraron que la variable que menor coeficiente de correlación tuvo con el resto de las variables fue la Espigazón promedio, teniendo incluso, una correlación negativa con la variable NGe, aunque no significativa.

e) Análisis de Regresión:

En las siguientes tablas y en el Anexo A se presentan los análisis de regresión lineal simple para cada una de las variables estudiadas. Los datos utilizados para estos análisis se obtuvieron promediando las dos repeticiones de cada genotipo y para cada una de las variables.

Tabla 4. Análisis de la regresión lineal entre la Espigazón y el resto de las variables estudiadas

Espigazón vs.	N	P	F	R²	R² ajustado
Altura	130	0,0027	9,34	0,07 (7%)	0,06 (6%)
NEP	130	0,5212	0,41	0,0032 (0,32%)	0,00 (0%)
NEE	130	<0,0001	34,70	0,21 (21%)	0,21 (21%)
NGE	130	0,2382	1,40	0,01 (1%)	0,0031 (0,31%)
NGe	130	0,2734	1,21	0,01 (1%)	0,0016 (0,16%)

Valor crítico (VC)= 2,731

Los resultados obtenidos en la Tabla 4 demuestran que existe relación positiva y estadísticamente significativa entre la Espigazón y las variables Altura ($p= 0,027$) y NEE ($p= <0,0001$) pero no hay evidencias estadísticas de que exista tal relación entre la Espigazón y el resto de las variables analizadas (debido a que la probabilidad p de la Espigazón con dichas variables es mayor que 0,005).

La Prueba F refuerza los resultados previamente mencionados ya que, las únicas variables cuyo valor de F superan al valor crítico de tabla son Altura y NEE (lo cual nos permite rechazar la H_0). En relación con los valores de R^2 , los resultados muestran que el modelo de regresión lineal explica bajos porcentajes para todas las variables analizadas.

Tabla 5. Análisis de la regresión lineal entre la Altura y el resto de las variables estudiadas

Altura vs.	N	P	F	R²	R² ajustado
NEP	130	<0,0001	201,21	0,61 (61%)	0,61 (61%)
NEE	130	<0,0001	107,81	0,46 (46%)	0,45 (45%)
NGE	130	<0,0001	72,54	0,36 (36%)	0,36 (36%)
NGe	130	<0,0001	29,44	0,19 (19%)	0,18 (18%)

En la Tabla 5 se puede evidenciar que todas las variables analizadas (NEP, NEE, NGE, NGe) presentan una relación positiva y estadísticamente significativa con el carácter Altura (ya que $p<0,0001$), esto sería un indicador de que, a mayor altura de planta, mayor número de espigas por planta, número de espiguillas por espiga y mayor número de granos por espiga y espiguilla. Estos resultados son reforzados por el valor del estadístico F (el cual supera en todas las variables al valor crítico de tabla).

En relación con el coeficiente de determinación, se puede observar que la variable con mayor porcentaje es NEP, para esta variable la variación en altura explica un 61% de la variabilidad total (más de la mitad de la variabilidad total). Mientras que, otros factores explican la variabilidad restante.

Tabla 6. Análisis de la regresión lineal entre NEP y el resto de las variables estudiadas

NEP vs.	N	P	F	R ²	R ² ajustado
NEE	130	<0,0001	80,00	0,38 (38%)	0,38 (38%)
NGE	130	<0,0001	86,12	0,40 (40%)	0,40 (40%)
NGe	130	<0,0001	44,14	0,26 (26%)	0,25 (25%)

En la tabla 6 se observa que las variables NEE, NGE y NGe presentan una relación positiva y estadísticamente significativa con la variable número de espigas por planta al analizar los valores de probabilidad p (vemos que todos son menores a 0,05). Estos resultados son respaldados por los valores de la prueba F (todos superan al valor crítico de tabla) lo cual indica que, a medida que aumenta el número de espigas por planta, aumentan con este el NEE, NGE y NGe.

En este análisis, se puede observar que la variable con mayor porcentaje de R^2 es el número de granos por espiga (NGE). Para esta variable, la variación en número de espigas por planta explica un 40% de la variabilidad total. Mientras que, el resto de las variables estudiadas (NEE y NGe) explican la variabilidad restante.

Tabla 7. Análisis de la regresión lineal entre el NEE y el resto de las variables estudiadas

NEE vs	N	P	F	R ²	R ² ajustado
NGE	130	<0,0001	95,17	0,43 (43%)	0,42 (42%)
Nge	130	<0,0001	18,68	0,13 (13%)	0,12 (12%)

La Tabla 7 muestra el análisis de correlación lineal entre NEE y las variables NGE y NGe. A partir de estos resultados se puede observar una relación entre NEE y el resto de las variables analizadas ($p<0,0001$) lo cual indica que, a mayor número de espiguillas por espigas, el número de granos por espiga y espiguilla será mayor. Estos resultados están reforzados por los valores de la Prueba F (estadístico $F>VC$).

En relación con el R^2 , la relación entre variables con mayor porcentaje de la varianza explicada se observa para NEE-NGE.

Tabla 8. Análisis de la regresión lineal entre el NGE y el NGe

NGE vs.	N	P	F	R ²	R ² ajustado
Nge	130	<0,0001	758,18	0,86 (86%)	0,86 (86%)

Al analizar los valores de probabilidad p y del estadístico F de la Tabla 8 observamos que existe relación lineal y altamente significativa entre las variables NGE y NGe ($p<0,0001$) ($F>VC$). La variable Número

de granos por espiguilla (NGe) fue medida por el cálculo (NGE/NEE), de esta variable se espera una correlación-regresión alta, siendo el $R^2=86\%$, dado que se calcula con el carácter NGE. El alto valor de R^2 indica que gran parte de la variabilidad es explicada por el modelo y, solo un 14% de esta variabilidad explica otros factores.

2. Resultados del análisis genotípico

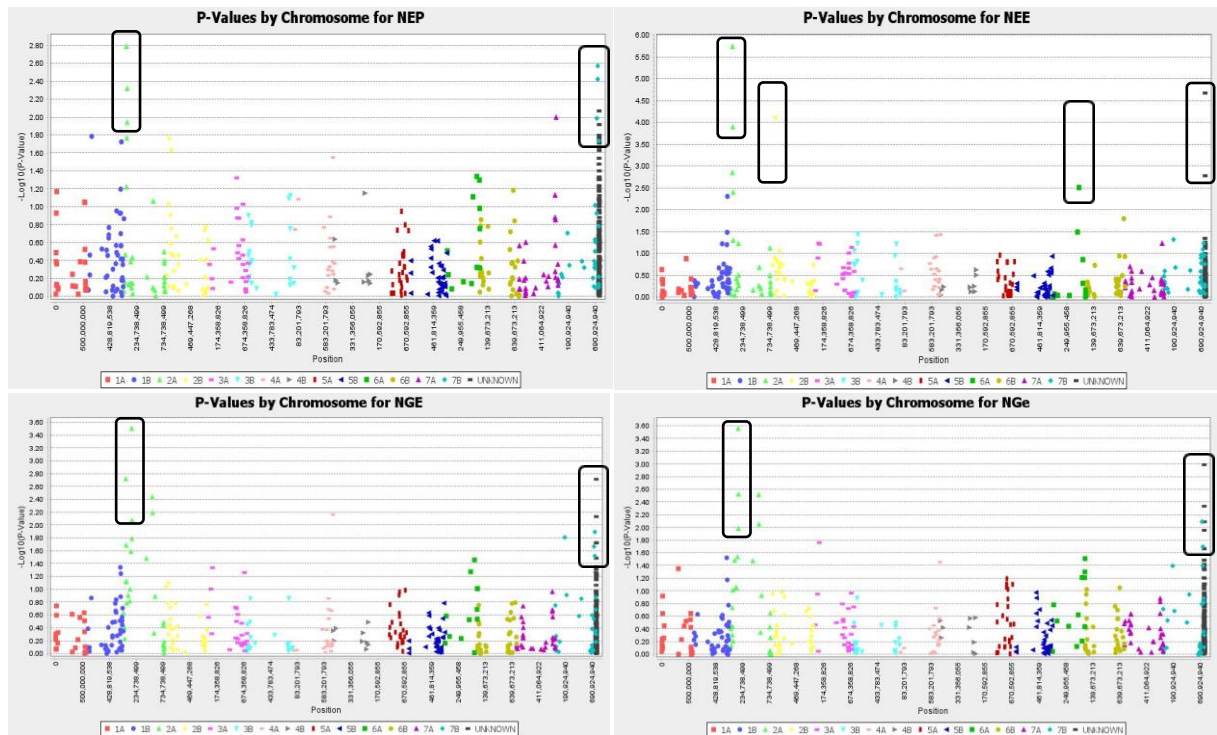
El análisis de la información molecular se llevó a cabo para las 128 RILs y los parentales, los mismos se resumen en las siguientes secciones.

Sobre las RILs y los genotipos parentales se amplificaron un total de 18.439 marcadores correspondientes a polimorfismos de nucleótido simple (SNP). Luego de aplicar los filtros descriptos en la metodología, se retuvieron 812 marcadores.

a. Gráficos Manhattan plot

Un diagrama de Manhattan es un tipo específico de diagrama de dispersión ampliamente utilizado en genómica para estudiar los resultados del GWAS (*Genome Wide Association Study*). Cada punto representa una variante genética. El eje X muestra su posición en un cromosoma, el eje Y indica cuánto está asociado con un rasgo.

Si una variante muestra un valor de asociación que supera el umbral de significación marcado en el estudio (LOD=3 o 2,5), está será juzgada como “presuntamente relacionada con el rasgo” (*Manhattan plot in R: a review*, s.f.). (https://r-graph-gallery.com/101_Manhattan_plot.html).



Figuras 6 A, B, C y D. Diagramas de Manhattan. En estos gráficos se muestra el valor de asociación de cada SNP con el rasgo estudiado (A= NEP, B=NEE, C=NGE, D=NGE). La escala indica el valor LOD ($= -\log(p)$), cuando el valor de asociación obtenido en un SNP supera el valor umbral ($p=3$ o 2.5), se determina que el SNP podría estar asociado con ese rasgo (Uffelmann *et al.*, 2021)

En las Figuras 6 A, B, C y D se observa que, para los caracteres número de granos por espiguilla (NGe) y número de granos por espiga (NGE), el cromosoma 2A (representado como triángulos verdes) y un cromosoma desconocido (*unknown*, representado con líneas negras) presentan variantes con un valor de asociación que supera el valor umbral por lo que se determina que dichos SNPs podrían estar asociados con los rasgos mencionados.

Por otro lado, para la variable número de espiguillas por espiga (NEE) vemos una mayor cantidad de cromosomas cuyas variantes superan el valor umbral, estos son; el cromosoma 2A, el 2B (representado como rombos amarillos), el 6A (representado como cuadrados verdes) y un cromosoma desconocido (*unknown*). Esto nos indica que dichos SNPs estarían presuntamente relacionados con el rasgo nombrado.

Para el carácter NEP se encontraron dos cromosomas con SNPs asociados (los cuales presentan un valor de asociación que supera el umbral $p=2.5$), estos cromosomas son el 2A y el 7B (representado como rombos celestes).

b. Estudio de mapeo por asociación en la población RIL

Se determinó la asociación entre los caracteres fenotípicos medidos en espiga y la matriz molecular de marcadores SNP obtenidos mediante la técnica DArTseq. Se identificaron un total de 12 marcadores moleculares SNP asociados a uno o más caracteres fenotípicos. Para cada uno de los caracteres estudiados, la estructura de datos quedó mejor representada por el modelo de asociación MLM (K+PCA). Este modelo consideró cinco componentes del análisis de componentes principales utilizando datos moleculares como covariable y, además, incorporó la correlación de a pares entre las líneas mediante la matriz de parentesco (*Kinship*).

En la tabla 9 se presentan los marcadores moleculares asociados con las variables cuantitativas estudiadas. También se presenta su localización cromosómica y la posición (en pares de bases [pb]) en el genoma. En particular, tres marcadores presentaron asociación con más de un carácter, mientras que los otros nueve marcadores restantes solo presentaron asociación con un solo carácter.

Las características de estos tres marcadores fueron las siguientes:

- **1110438|F|0-14:C>T-14:C>T**: Localizado en el cromosoma 2A (posición 46429124 pb) presentó asociación con los caracteres número de espigas por planta (NEP), número de espiguillas por espiga (NEE) y número de granos por espiga (NGE).
- **1091747|F|0-52:G>A-52:G>A**: Localizado en el cromosoma 2A (posición 155070906 pb) presentó asociación con los caracteres número de granos por espiga (NGE) y número de granos por espiguilla (NGe).
- **1140354|F|0-24:G>T-24:G>T**: Localizado en un cromosoma desconocido, no anclado físicamente, (*UNKNOWN*) presentó asociación con los caracteres número de granos por espiga (NGE) y número de granos por espiguilla (NGe), al igual que el marcador molecular previamente mencionado.

Tabla 9. Información sobre los marcadores moleculares identificados (MM), localización cromosómica, posición física en el genoma y asociación con los caracteres cuantitativos estudiados en una población de 128 RILs y sus respectivos parentales.

Carácter	MM	Cromosoma	Posición (pb)	F	P	LOD	Add effect	R ² (R ² %)
NEP	1110438 F 0-14:C>T-14:C>T	2A	46429124	6,81	0,00162	2,79	-4,54E-01	0,115 (11,5%)
NEE	1110438 F 0-14:C>T-14:C>T	2A	46429124	14,93	1,81E-06	5,74	-1,18E+00	0,262 (26,23%)
NEE	32940342 F 0-13:A>G-13:A>G	2A	45851839	6,97	0,0014	2,85	0,83089	0,132 (13,29%)
NEE	4004331 F 0-18:T>C-18:T>C	2A	51874409	9,72	1,27E-04	3,89	0,94883	0,168 (16,88%)
NGE	1110438 F 0-14:C>T-14:C>T	2A	46429124	6,62	0,00192	2,7	-4,79E+00	0,114 (11,43%)
NGE	1091747 F 0-52:G>A-52:G>A	2A	155070906	8,66	3,13E-04	3,50	6,77324	0,141 (14,13%)
NGe	1091747 F 0-52:G>A-52:G>A	2A	155070906	8,81	2,76E-04	3,5	0,36771	0,138 (13,88%)
NGe	1215845 F 0-38:A>G-38:A>G	2A	161696203	6,11	0,00298	2,5	-3,00E-01	0,099 (9,97%)
NGe	994673 F 0-35:G>C-35:G>C	2A	536736281	6,10	0,00305	2,5	0,27175	0,104 (10,46%)

NEE	12779268 F 0-25:A>C-25:A>C	2B	74498489	10,30	7,72E-05	4,11	0,91869	0,176 (17,63%)
NEE	1714673 F 0-64:C>T-64:C>T	6A	483681213	6,06	0,0031	2,5	-4,29E-01	0,101 (10,14%)
NEP	997146 F 0-23:C>G-23:C>G	7B	736481446	6,22	0,00269	2,57	0,37215	0,100 (10,01%)
NEE	1095899 F 0-19:T>A-19:T>A	UNKNOWN	-	11,85	2,13E-05	4,67	1,00712	0,200 (20,04%)
NEE	1128705 F 0-11:A>G-11:A>G	UNKNOWN	-	6,76	0,00167	2,77	0,79997	0,115 (11,55%)
NGe	1140354 F 0-24:G>T-24:G>T	UNKNOWN	-	7,31	0,00103	2,98	-3,28E-01	0,124 (12,43%)
NGE	1140354 F 0-24:G>T-24:G>T	UNKNOWN	-	6,60	0,00194	2,71	-5,69E+00	0,115 (11,5%)

Número de espigas por planta (NEP), número de espiguillas por espiga (NEE), número de granos por espiga (NGE) y número de granos por espiguilla (NGe)

La hipótesis nula (H_0) que se prueba con el modelo plantea que no existe asociación entre marcadores moleculares y caracteres relacionados con el rendimiento (cuantitativos). Para evaluar la hipótesis de trabajo se verifica el valor de p asociado con F y los valores críticos de la Tabla de Valores de la Distribución de Fisher. Si el valor del estadístico F supera al valor crítico, los resultados son estadísticamente significativos y puedo rechazar la H_0 . En cambio, si el valor de F es menor al del valor crítico, se entiende que los resultados no son estadísticamente significativos y no puedo rechazar la hipótesis nula.

En la Tabla 9 se puede observar el valor del estadístico F para cada una de los marcadores moleculares y el carácter cuantitativo asociado. Teniendo en cuenta que $N=170$ (por ende, $VC=3,895$), rechazamos la H_0 para cada uno de los marcadores estudiados y confirmamos que existe asociación entre marcador y carácter estudiado.

De los valores presentados en la tabla previamente mencionada, el valor de significancia de la prueba (p) nos indica que la relación entre marcadores moleculares y variables estudiadas existe y es estadísticamente significativa ya que, para todos los marcadores estudiados, el valor de p fue menor a 0,05 ($p<0,05$).

El Logaritmo de las probabilidades (LOD) es una estimación estadística de la probabilidad de que se hereden juntos dos loci genéticos (debido a la cercanía física, loci ligados) o un locus y un carácter (*Diccionario de genética del NCI*, s.f.).

Para este trabajo, el LOD representa una estimación de la probabilidad de que el marcador molecular identificado (o gen ligado) este asociado al carácter cuantitativo estudiado y que, por ende, sean heredados juntos. El LOD (logaritmo de las probabilidades), se calculó mediante el programa TASSEL como $-\log_{10}(\text{valor } p)$ y, una vez obtenidos los valores de LOD para cada marcador, se seleccionaron aquellos marcadores cuyo valor de LOD fuera superior a 2,5 (valor umbral).

El marcador 1110438|F|0-14:C>T-14:C>T, asociado a NEE, NEP y NGE mostró el mayor valor de LOD y se localizó en posición 46429124 pb sobre el cromosoma 2A con un valor de LOD=5.74. El SNP presentó alelos C/T. El marcador 1714673|F|0-64:C>T-64:C>T y otros tres SNP, presentaron valores de LOD muy cercanos a 2,5.

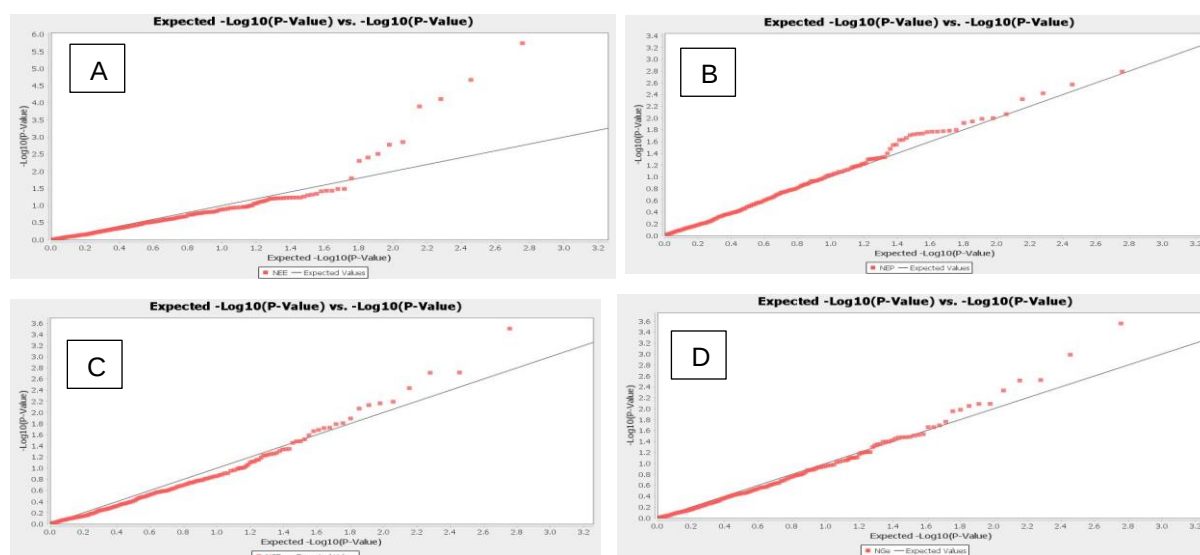
c. Gráfico cuantil-cuantil (Q-Q Plot)

Un gráfico cuantil-cuantil (Q-Q Plot) es un tipo de gráfico que podemos usar para determinar si un conjunto de datos proviene o no de una distribución teórica. Muchas pruebas estadísticas suponen que un conjunto de datos sigue una distribución normal y, a menudo, se utiliza un gráfico Q-Q Plot para evaluar si se cumple (o no) dicha hipótesis de normalidad. Un gráfico cuantil-cuantil proporciona una manera fácil de verificar visualmente si un conjunto de datos sigue una distribución normal. Si la

distribución teórica constituye una buena aproximación de la distribución empírica, cabría esperar que los cuantiles (puntos) de los datos estén muy próximos a los de la distribución teórica y, por ende, los puntos del gráfico se dispondrán muy próximos a la recta ("Q-Q Plot Normal. Los puntos de posición gráfica", 2007).

En este trabajo, los gráficos Q-Q plot comparan los valores observados del $-\log_{10}$ (valor- p) para cada marcador molecular, con los valores esperados del $-\log_{10}$ (valor- p) bajo la hipótesis nula de no asociación entre marcadores moleculares y los caracteres en estudio (Ruiz *et al.*, 2021). Cuanto más próximos se encuentran los puntos a la recta, mejor será el ajuste de datos, es decir, se observa una buena asociación entre variable y marcadores.

Los gráficos fueron elaborados a partir del análisis de la matriz MLM, utilizando el software TASSEL.



Figuras 7 A, B, C y D. Gráficos cuartil-cuartil para cada una de las variables estudiadas (A=NEE, B=NEP, C=NGE, D=NGe).

A partir de los gráficos obtenidos se observa que las variables altura, número de espigas por planta (NEP), número de granos por espiga (NGE) y número de granos por espiguilla (NGe) presentan la nube de puntos dispuesta muy próxima a la recta, por lo que se asume que existe una buena asociación entre marcadores y variables. En el caso de la variable número de espiguillas por espiga (NEE) se observa un buen ajuste (una buena asociación entre variables) para valores bajos de $-\log_{10}$ (valor- p), pero a medida que estos incrementan, el gráfico no muestra una buena asociación entre los valores esperados y observados.

d. Evaluación del efecto del principal SNP identificado

Considerando que el valor de LOD para el marcador *1110438|F|0-14:C>T-14:C>T* fue el más alto, se evaluó su efecto sobre los caracteres medidos utilizando un análisis de la varianza (ANOVA) y luego del ANOVA se utilizó una prueba post hoc de Tukey.

Se plantearon hipótesis para la utilización de una prueba de comparación de medias utilizando el test de ANOVA para las variables NEP, NEE, NGE, NGe, diseñando como una relación causal y se enuncian de la siguiente forma:

- Ha: hay diferencias en el promedio de las medias obtenidas para cada uno de los alelos.
- H₀: no hay diferencias en el promedio de las medias para cada uno de los alelos.
- Prueba estadística: el valor de significancia de la prueba es de $\alpha=0.05$ (5%), si es mayor no se rechaza la hipótesis nula, si es menor se rechaza la hipótesis nula.

Los valores medios de cada componente del rendimiento y para cada alelo del marcador *1110438|F|0-14:C>T-14:C>T* se presentan utilizando gráficos de barra (Figuras 8 a 11). Las medias se estimaron considerando el total de valores de la población RIL para cada alelo. La tabla (Tabla 10) y gráficos correspondientes con cada uno de los test de Tukey se elaboraron con el software INFOSTAT.

Tabla 10. Información sobre el valor de N, media de alelos (CC/TT), valor *p*, CV, y coeficiente de determinación (*R*²) para la prueba post hoc de Tukey

Carácter	N	Media CC	Media TT	Valor <i>p</i>	CV	<i>R</i> ²
NEP	108	4,26	4,88	0.0077	25,62	0,07
NEE	108	13,67	15,58	<0,0001	10,22	0,29
NGE	108	39,32	43,33	0,0819	28,49	0,03
NGe	108	2,74	2,80	0,6599	24,32	0,0018

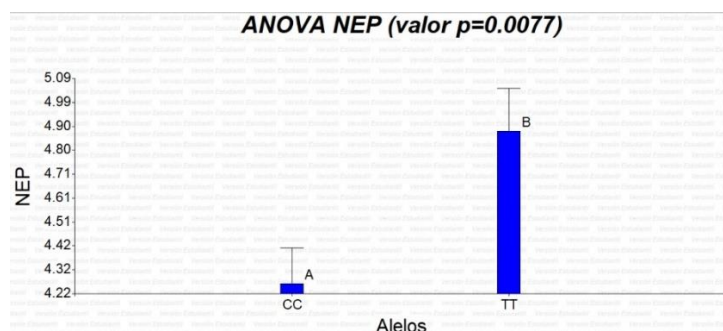


Figura 8. Gráfico de barras para el carácter número de espigas por planta (NEP) utilizando test post hoc de Tukey

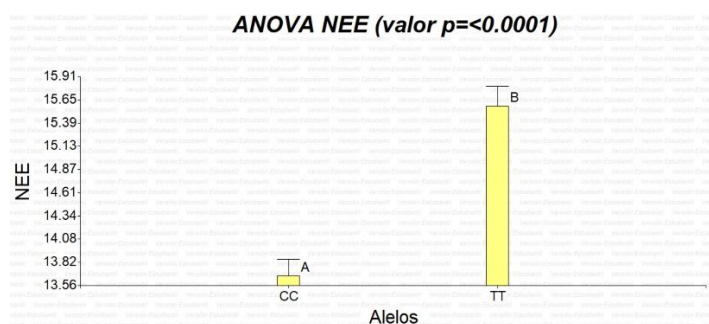


Figura 9. Gráfico de barras para el carácter número de espiguillas por espiga (NEE) utilizando test post hoc de Tukey

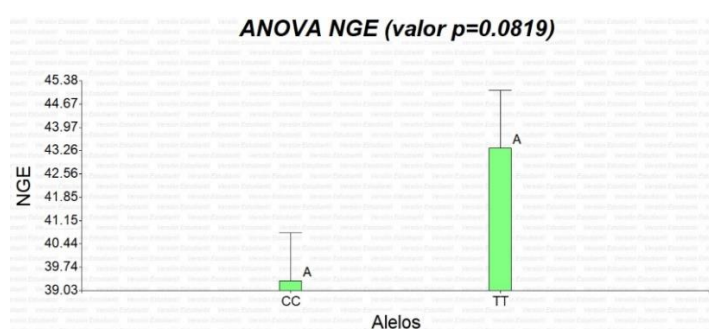


Figura 10. Gráfico de barras para el carácter número de granos por espiga (NGE) utilizando test post hoc de Tukey

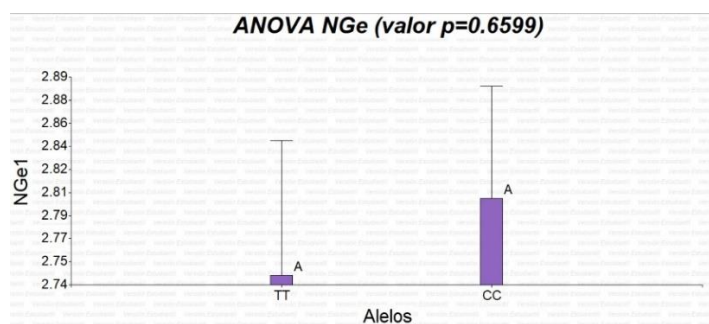


Figura 11. Gráfico de barras para el carácter número de granos por espiguilla (NGe) utilizando el test post hoc de Tukey

Los resultados obtenidos para la prueba de ANOVA y, utilizando como prueba post hoc el test de Tukey, mostraron que las variables número de espigas por planta (NEP) y número de espiguillas por espiga (NEE) presentan un valor de significancia (p) menor a 0,05, lo cual indica que existen diferencias altamente significativas entre las medias para ambos alelos (rechazo H_0). Mientras que, para las variables número de granos por espiga (NGE) y número de granos por espiguilla (NGe) no se hallaron diferencias estadísticamente significativas entre las medias ($p > 0,05$).

El coeficiente de determinación R^2 es una medida estadística que se utiliza para evaluar la proporción de la varianza en una variable dependiente que es explicada por una o más variables independientes en un modelo de regresión. En el contexto de los marcadores moleculares, el R^2 se utiliza para evaluar la fuerza y significancia de la asociación entre marcadores genéticos y características fenotípicas.

En los estudios de asociación del genoma completo (GWAS) el coeficiente de determinación se utiliza para medir la fuerza de la asociación entre un marcador genético (en este caso, SNPs) y un rasgo fenotípico (fecha de espigazón, número de espiguillas por espiga, etc.).

Los valores de R^2 oscilan entre 0 y 1 o expresado en porcentaje, estos valores oscilan entre 0 y 100%. Un valor de R^2 cercano a 1 (o 100%) indica que la proporción significativa de la variabilidad del rasgo puede ser explicada por el marcador genético.

Para tres de las variables analizadas, no se mostraron valores altos de R^2 (menores al 10%), lo cual indica que, para los caracteres cuantitativos NEP, NGE y NGe los marcadores moleculares utilizados no lograrían explicar una proporción significativa de la variabilidad de los rasgos mencionados. Si bien, para el carácter NEE el marcador 1110438|F|0-14:C>T-14:C>T explicó el 29% de la variación en el carácter, este valor corresponde a lo que se considera como un QTL mayor. Se encontraron diferencias significativas entre alelos solo para las variables NEP y NEE por lo que, si bien es necesario validar el efecto del marcador en otras poblaciones, este podría resultar útil en selección asistida por marcadores (MAS).

e. Desequilibrio de ligamiento (Linkage disequilibrium o LD)

El desequilibrio de ligamiento (LD, por sus siglas en inglés) se refiere a la asociación que existe entre alelos de genes que se encuentran ubicados muy cercanamente en un mismo segmento cromosómico o sobre diferentes cromosomas (Suazo S *et al.*, 2005). El LD describe la no aleatoriedad con la que los alelos en diferentes loci se asocian entre sí en una población. En otras palabras, el desequilibrio de ligamiento mide la tendencia de ciertos alelos a ser heredados juntos más frecuentemente de lo que se esperaría por azar.

El desequilibrio de ligamiento puede medirse utilizando diferentes métodos y métricas que cuantifican la no aleatoriedad en la asociación de alelos en diferentes loci, como son el valor de p (p -value), el coeficiente de determinación (r^2), el coeficiente de desequilibrio de ligamiento (D) y su medida estandarizada (D'). El r^2 se utiliza para cuantificar el desequilibrio de ligamiento entre dos loci, un valor de r^2 cercano a 1 indica un alto LD (lo que significa presencia de loci ligados físicamente o co-heredados aunque sobre distintos cromosomas), y contribuye a identificar regiones del genoma donde los alelos están altamente correlacionados. El coeficiente D mide la desviación de las frecuencias observadas de haplotipos con respecto a las frecuencias esperadas si los loci fueran independientes.

D' es una versión normalizada del coeficiente D que varía entre -1 y 1, facilitando la comparación entre diferentes pares de loci.

Valores de r^2 y D' cercanos a 1 indican un fuerte desequilibrio de ligamiento (sugiriendo que los loci pueden estar físicamente cercanos en el cromosoma y/o son heredados juntos con mayor frecuencia), mientras que valores iguales o cercanos a 0 indican que no hay LD o, en otras palabras, que existe independencia entre loci. En el caso de una población segregante, un r^2 o D' cercano a uno puede indicar desequilibrio de ligamiento debido a la proximidad física de dos loci, pero no es exclusivo en todos los casos; puede ocurrir también que dos genes seleccionados juntos se encuentren en distintos cromosomas y también puedan encontrarse en desequilibrio de ligamiento.

Tabla 11. Análisis de la presencia de desequilibrio de ligamiento en MM utilizando valor de p (*p-value*), coeficiente de determinación (r^2), coeficiente de desequilibrio de ligamiento (D)

Locus 1	MM	Posición	Locus 2	MM	Posición	r^2	D'	p-Value
2A		45851839	-		-	-	-	-
2A	1110438 F 0- 14:C>T- 14:C>T	46429124	2A	32940342 F 0- 13:A>G- 13:A>G	45851839	1	1	7,48E-28
2A	4004331 F 0- 18:T>C- 18:T>C	51874409	2A	32940342 F 0- 13:A>G- 13:A>G	45851839	0,96	1	1,63E-27
2A	4004331 F 0- 18:T>C- 18:T>C	51874409	2A	1110438 F 0-14:C>T- 14:C>T	46429124	0,9166682 4	0,957 4279	9,66E-25
2A	1091747 F 0- 52:G>A- 52:G>A	155070906	2A	32940342 F 0- 13:A>G- 13:A>G	45851839	0,0768716 3	0,283 04347	0,00829
2A	1091747 F 0- 52:G>A- 52:G>A	155070906	2A	1110438 F 0-14:C>T- 14:C>T	46429124	0,0457875 5	0,238 09524	0,06129 NS
2A	1091747 F 0- 52:G>A- 52:G>A	155070906	2A	4004331 F 0-18:T>C- 18:T>C	51874409	0,0792144	0,312 05675	0,00796
2A	1215845 F 0-	161696203	2A	32940342 F 0-	45851839	0,0724002 9	0,269 073	0,00907

	38:A>G- 38:A>G			13:A>G- 13:A>G				
2A	1215845 F 0- 38:A>G- 38:A>G	161696203	2A	1110438 F 0-14:C>T- 14:C>T	46429124	0,0386743 7	0,227 47636	0,06561NS
2A	1215845 F 0- 38:A>G- 38:A>G	161696203	2A	4004331 F 0-18:T>C- 18:T>C	51874409	0,0706785 5	0,282 19062	0,00959
2A	1215845 F 0- 38:A>G- 38:A>G	161696203	2A	1091747 F 0-52:G>A- 52:G>A	155070906	1	1	1,26E-30
2A	994673 F 0- 35:G>C- 35:G>C	536736281	2A	32940342 F 0- 13:A>G- 13:A>G	45851839	0,0407936 9	0,206 37898	0,06144 NS
2A	994673 F 0- 35:G>C- 35:G>C	536736281	2A	1110438 F 0-14:C>T- 14:C>T	46429124	0,0113666 4	0,118 94737	0,39781 NS
2A	994673 F 0- 35:G>C- 35:G>C	536736281	2A	4004331 F 0-18:T>C- 18:T>C	51874409	0,0199917 7	0,144 38502	0,21857 NS
2A	994673 F 0- 35:G>C- 35:G>C	536736281	2A	1215845 F 0-38:A>G- 38:A>G	161696203	0,4463795 4	0,737 5643	5,70E-12
2B		74498489	-		-	-	-	-
6A		483681213	-		-	-	-	-
7B		736481446	-		-	-	-	-

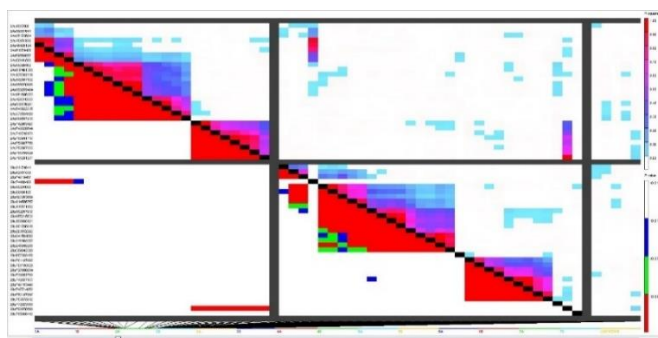


Figura 12. Diagrama de disequilibrio de ligamiento (LD plot) expandido a la región del cromosoma 2A

En la Tabla 11 se observan los valores de p , r^2 y D' para los marcadores seleccionados (aquellos con valores de LOD altos).

A partir de los resultados obtenidos se puede deducir que existe desequilibrio de ligamiento en dos regiones independientes del genoma comprendidas entre los 45851839 y 51874409 pb y los 155070806 y 161696203 pb, presente en el cromosoma 2A. Entre ambas regiones no se encontró LD.

Para el caso de los MM presentes en los cromosomas 2B, 6A y 7B, no se encontró desequilibrio de ligamiento con otros marcadores asociados por lo que se asume que presentan posición única en el genoma y no muestran desequilibrio de ligamiento.

En el trabajo de Roncallo (2021) se indica que un alto LD entre loci puede deberse a ligamiento físico, deriva genética, cuellos de botella o selección. Sin embargo, algunos cromosomas, como el 2A, muestran un número extendido y creciente de SNP en alto o completo desequilibrio de ligamiento en función de la distancia, probablemente debido a una presión de selección diferencial ejercida sobre estos cromosomas para mantener loci agronómicamente ventajosos o epistáticos durante el proceso de mejoramiento.

El estudio del desequilibrio de ligamiento puede ayudar a entender el efecto de las presiones de selección ejercidas sobre el germoplasma, ocurrida durante el proceso de mejora genética. Este desequilibrio de ligamiento es probablemente causado por un ligamiento físico (Roncallo *et al.*, 2021). Tanto la capacidad de capturar asociaciones significativas entre loci polimórficos y la varianza fenotípica como la utilidad de las estrategias de mapeo de asociaciones dependen de la extensión del LD a lo largo del genoma. La extensión del LD como función de la distancia genética es indicativa de la profundidad de la resolución, así como de la densidad de marcadores necesarios para obtener resultados confiables en estudios de mapeo de asociaciones (Roncallo *et al.*, 2019).

f. Genes candidatos

La búsqueda de genes candidatos se realizó en base a la posición física de los marcadores identificados y la informada para las anotaciones funcionales de genes sobre el genoma de trigo candeal. Se consideraron las principales asociaciones identificadas en los cromosomas 2A, 2B, 6A y 7B.

En la Tabla 12 del Anexo B se observan las posiciones en el genoma de trigo candeal de los marcadores SNP asociados a uno o más caracteres cuantitativos y la posición en el genoma de los posibles genes candidatos para dichos marcadores en los cromosomas previamente nombrados.

El marcador 32940342|F|0-13:A>G-13:A>G asociado al carácter NEE, localizado en el cromosoma 2A (posición 45851839 pb) presenta gran cercanía con el gen TRITD2Av1G023480 (la distancia mínima entre el marcador molecular y el gen mencionado fue de 57405 pb). Dicho gen se encuentra localizado

en el cromosoma 2A (posición en el genoma entre 45791690 y 45794434 pb). Será necesario validar experimentalmente el gen TRITD2Av1G023480 como responsable de la variación en el carácter cuantitativo número de espiguillas por espiga.

Por otro lado, en función de lo observado en la tabla, vemos que el marcador molecular 4004331|F|0-18:T>C-18:T>C, asociado al carácter NEE, localizado en el cromosoma 2A (posición 51874409 pb) presenta una gran proximidad con tres genes ubicados en este cromosoma, estos genes son TRITD2Av1G026050 (posición en el genoma entre 51811235 y 51813781 pb), TRITD2Av1G026060 (posición en el genoma entre 51822211 y 51823989 pb) y TRITD2Av1G026070 (posición en el genoma entre 51826660 y 51829915 pb). Estos posibles candidatos, en la región comprendida entre 51811235 y 51829915 pb, deberán estudiarse con mayor detalle para determinar cuál de ellos esta potencialmente afectando al carácter número de espiguillas por espiga.

El marcador 1091747|F|0-52:G>A-52:G>A asociado a los caracteres NGE y NGe, localizado en el cromosoma 2A (posición 155070906 pb) resulto ser el marcador con menor distancia (10329 pb) a un gen candidato. El gen TRITD2Av1G066860 (Phototropic-responsive NPH3) localizado en 2A entre 155081235 y 155083808 pb, resulta interesante para continuar la etapa de validación como responsable de la variación en el número de granos por espiga y espiguilla.

Estos resultados resultan concordantes con en el trabajo de Achilli (2023) en el cual se lograron identificar SNPs distribuidos en el cromosoma 2A entre las posiciones 86.161.841 y 157.827.531, con valores máximos de LOD en 149414703 pb y 154732736 pb, asociados a tres caracteres, al número de granos por espiga, número de granos por espiguilla y número de granos por espiguilla fértil. Los resultados de esta tesis de grado evidencian que esta región en 2A podría afectar el número de granos por espiga a través de su efecto sobre el número de granos por espiguilla. Para esta región, se informan en la bibliografía un total de 20 QTLs. Sin embargo, solo tres de estos QTLs estuvieron asociados con el rendimiento o sus componentes, uno de ellos fue informado por Milner *et al.* (2015) asociado al rendimiento y los otros dos fueron informados por Roncallo *et al.* (2017), uno asociado al número de granos por espiguilla y el otro al número de granos por espiguilla fértil (Achilli *et al.*, 2023).

Por otro lado, el marcador SNP 12779268|F|0-25:A>C-25:A>C, asociado al carácter NEE y localizado en el cromosoma 2B (posición 74498489 pb) se ubica próximo a dos genes en el mismo cromosoma. Estos genes son TRITD2Bv1G031560 (posición en el genoma entre 74514953 y 74520155 pb), TRITD2Bv1G031540 (posición en el genoma entre 74452070 y 74452965 pb). Deberá entonces, continuarse la validación experimental para determinar si estos genes candidatos al carácter número de espiguillas por espiga.

De acuerdo con el rango establecido (80 kb) no fue posible encontrar genes candidatos para los marcadores identificados sobre los cromosomas 6A y 7B.

g. Análisis de la diversidad genética utilizando árbol filogenético

Un árbol filogenético o cladograma es una estructura en la cual las especies están organizadas en ramas que las unen de acuerdo con su relación de parentesco y distancia genética. La distancia genética representa el número de mutaciones entre especies desde su divergencia.

La estructura de la población se estudió sobre las 128 RILs y sus parentales (BI Quillen, Buck Granate). El árbol filogenético (Figura 13) se construyó utilizando un total de 347 marcadores SNP, en base a los cuales se obtuvo una matriz de disimilaridad mediante el cálculo del índice Simple Matching. Finalmente, se utilizó el método de agrupamiento Ward (varianza mínima) en la obtención del árbol. Los marcadores previamente nombrados también fueron útiles en la construcción de un gráfico de análisis discriminante de las componentes principales (PCA) (Figura 14).

La población estuvo principalmente dividida en 3 subgrupos con un valor mínimo de disimilaridad de 0.041 y un valor máximo de 0.838 (Figura 13). El subgrupo 1 (indicada con color verde) incluyó 68 RILs y los parentales (en total 70 líneas) mientras que los subgrupos 2 y 3 (representadas con color amarillo y violeta) incluyeron 55 y 5 RILs respectivamente.

Con el número y tipo de marcadores utilizados, los parentales (BIQ y BG) no mostraron ser genéticamente diferentes entre sí, lo cual podría deberse a la metodología utilizada. A pesar de no poder diferenciar genéticamente a los parentales, se observa una gran diversidad genética entre las líneas derivadas de la población, producto de la cruce BI Quillen x Buck Granate.

Estos resultados indican la existencia de 3 subgrupos con moderada a alta diferenciación, ligeramente menor a lo reportado en el trabajo de Roncallo (2021) quien analizando 111 accesiones de trigo candeal sudamericano y utilizando el mismo método de agrupamiento (Ward) encontró 5 subpoblaciones y ligeramente mayor a lo reportado en el trabajo de Ren (2013), quien analizando 150 accesiones pudo dividir la población de trigo candeal analizada en dos subpoblaciones genéticamente distintas.

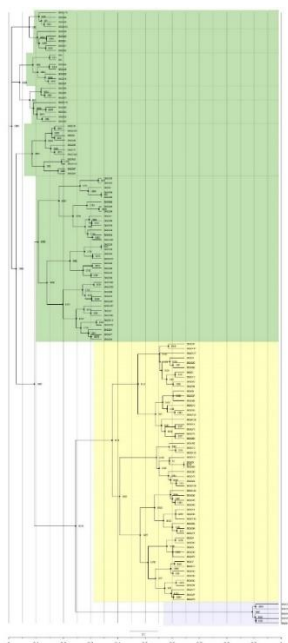


Figura 13. Árbol filogenético utilizado para el análisis de la estructura de la población obtenido en base a 347 marcadores SNP, implementando el índice Simple Matching y el método de agrupamiento Ward

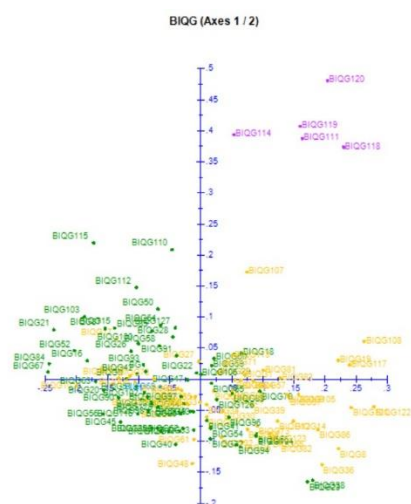


Figura 14. Análisis discriminante de las componentes principales (ACP) para las 128 RILs y sus parentales utilizando 347 marcadores SNP

CONCLUSIÓN

Este trabajo aporta valiosa información sobre la población de trigo candeal estudiada y confirma la presencia de genes con efecto positivo que incrementan el rendimiento del cultivo, afectando a componentes cuantitativos del mismo como son el número de espigas por planta, número de espiguillas por espiga y número de granos por espiga y espiguilla.

El estudio de solo los caracteres fenotípicos no permite conocer en profundidad la variabilidad genética presente en la población en estudio. Por ello, la implementación de marcadores moleculares (como lo son los marcadores SNPs) es fundamental para diferenciar RILs en las cuales sus componentes del rendimiento presenten una mejora en el rendimiento final del cultivo. Además, el uso de SNP permite caracterizar la población y conocer la diversidad genética presente.

Este trabajo, permitió postular genes asociados a caracteres relacionados con el rendimiento y sustentan las bases para futuras investigaciones. Los resultados presentados podrían tener fuerte implicancia en estudios genéticos en el cultivo de trigo, en el proceso de mejora y en la selección de líneas segregantes. Además, pueden contribuir en la conservación de la diversidad genética de la especie, ya que estudiar genes asociados a caracteres cuantitativos puede ayudar a entender cómo las poblaciones responden a los cambios ambientales y cómo se puede mantener la variabilidad genética necesaria para la adaptación y supervivencia a largo plazo.

BIBLIOGRAFÍA

- Achilli A. L. (2023). Análisis de la estructura del genoma y mapeo por asociación para rendimiento y sus componentes en trigo candeal. [Tesis de doctorado, Universidad Nacional del Sur]. RID-UNS.
- Achilli, A.L., Roncallo, P.F., Echenique, V. (2022). Genetic Gains in Grain Yield and Agronomic Traits of Argentinian Durum Wheat from 1934 to 2015. *Agronomy*, 12, 2151. <https://doi.org/10.3390/agronomy12092151>
- Andrade F. (2011). La tecnología y la producción agrícola. El pasado y los actuales desafíos. Editorial El Vikingo. Balcarce, Provincia de Buenos Aires, Argentina. 2da Edición.
- Bailey DW. (1971). Recombinant inbred strains. *Transplantation* 11:325-327
- Blanco, A., R. Simeone, A. Cenci, A. Gadaleta, O.A.Tanzarella, E. Porceddu, S. Salvi, R. Tuberosa, G. Figliuolo, P. Spagnoletti, M.S. Roder & V. Korzun. (2004). Extension of the Messapia x dicoccoides linkage map of *Triticum turgidum* (L.) *Theil Cellular and Molecular Biology Letters*, 9(3), 529-541.
- Bolibok-Bragoszewska H., Heller-Uszynska K., Wenzl P., Uszynski G., Kilian A., Rakoczy Trojanowska M. (2009). DArT markers for the rye genome genetic diversity and mapping. *BMC Genomics*, 10, 578. <https://doi.org/10.1186/1471-2164-10-578>
- Bozzini, A. (1988). Origin, distribution and production of durum wheat in the world. En: *Durum Chemistry and Technology*. Cap 1. Edit. Fabiani, G. Lintas, C. pp.1-16.
- CIMMYT. (2006). Protocolos de laboratorio: Laboratorio de Genética Molecular Aplicada del CIMMYT. Tercera edición. México, D.F.: CIMMYT.
- Diccionario de genética del NCI. (s.f.). Comprehensive Cancer Information - NCI. <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-genetica/def/puntuacion-del-logaritmo-de-las-probabilidades>
- Domínguez García, M. d. C. (2012). *Marcadores moleculares de ADN: análisis de la variabilidad, relaciones genéticas y mapeo en olivo (Olea europaea L.)* [Tesis doctoral inédita, Instituto de Investigación y Formación Agraria, Pesquera y Alimentaria (IFAPA)]. Centro Alameda del Obispo de Córdoba y en el Departamento de Genética de la Universidad de Córdoba.

- Dreisigacker, S. Sehgal D., Reyes Jaimez A.E, Luna Garrido B., Muñoz Zavala S., Núñez Ríos C., Mollins J., Mall S. (Eds.) (2016). *CIMMYT Wheat Molecular Genetics: Laboratory Protocols and Applications to Wheat Breeding*. Mexico, D.F.: CIMMYT.
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE*, 6(5),10.
- Fischer R.A., Byerlee D. and Edmeades G.O. (2014). *Crop yields and global food security: will yield increase continue to feed the world?*. ACIAR Monograph No. 158. Australian Centre for International Agricultural Research: Canberra. xxii + 634 pp.
- Jaccoud, D., Peng, K., Feinsein D., Kilian A. (2001). Diversity arrays: a solid state technology for sequence information independent genotyping. *Nuclei. Aci. Rese*, 29 (4), e25. <https://doi.org/10.1093/nar/29.4.e25>
- Johansen-Morris, A.D. & R.G. Latta. (2006). Fitness consequences of hybridization between ecotypes of *Avena barbata*: hybrid breakdown, hybrid vigor, and transgressive segregation. *Evolution; international journal of organic evolution*, 60(8), 1585–1595.
- Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., Caig, V., Heller-Uszynska, K., Jaccoud, D., Hopper, C., Aschenbrenner-Kilian, M., Evers, M., Peng, K., Cayla, C., Hok, P., & Uszynski, G. (2012). Diversity arrays technology: a generic genome profiling technology on open platforms. *Methods in molecular biology* (Clifton, N.J.), 888, 67–89. https://doi.org/10.1007/978-1-61779-870-2_5
- Lo Valvo, P., Miralles, D., Serrago, R. (2018). Genetic progress in Argentine bread wheat varieties released between 1918 and 2011: Changes in physiological and numerical yield components. *Field Crops Res.* 221: 314-321. <http://dx.doi.org/10.1016/j.fcr.2017.08.014>
- Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits* (Vol. 1, pp. 535-557). Sunderland, MA: Sinauer.
- Milner, S.G., Maccaferri, M., Huang, B.E., Mantovani, P., Massi, A., Frascaroli, E., Tuberosa, R., Salvi, S. (2015) A multiparental cross population for mapping QTL for agronomic traits in durum wheat (*Triticum turgidum* ssp. *durum*). *Plant Biotech. J.* 14(2), 735-748. <https://doi.org/10.1111/pbi.12424>

- Ortega Jiménez, N. G. (2017). Relación entre la distancia genética y geográfica en maíces Tuxupeño y Cónico norteño en México [Tesis de maestría inédita, Universidad Autónoma Chapingo]. Repositorio Chapingo. <https://repositorio.chapingo.edu.mx/handle/123456789/1269>
- Paccapelo, M. V. (2015). Modelos de selección genómica para caracteres cuantitativos basados en marcadores moleculares aplicados al mejoramiento de maíz [Tesis de maestría, Universidad Nacional de Córdoba]. Repositorio Digital de la Universidad Nacional de Córdoba. <http://hdl.handle.net/11086/2355>
- Patnaik, Debasis & Khurana, Paramjit. (2001). Wheat biotechnology: A minireview. *Electronic Journal of Biotechnology* (ISSN: 0717-3458) Vol 4 Num 2. 4. 10.2225/vol4-issue2-fulltext-4.
- Patterson, H. D., & Williams, E. R. (1976). A new class of resolvable incomplete block designs. *Biometrika*, 63(1), pp. 83-92.
- Pollard D. A. (2012). Design and construction of recombinant inbred lines. *Methods in molecular biology* (Clifton, N.J.), 871, 31–39. https://doi.org/10.1007/978-1-61779-785-9_3
- Quaglia, G.B. (1988). Other durum wheat products. *Durum Chemistry and Technology*, Fabriani, G. and Lintas, C. (eds). AACCC, St. Paul, Minnesota, pp. 263-282.
- Ren, J., Sun, D., Chen, L., You, F., Wang, J., Peng, Y., Nevo, E., Sun, D., Luo, M.-C., & Peng, J. (2013). Genetic Diversity Revealed by Single Nucleotide Polymorphism Markers in a Worldwide Germplasm Collection of Durum Wheat. *International Journal of Molecular Sciences*, 14(4), 7061–7088. <https://doi.org/10.3390/ijms14047061>
- Risterucci A.M., Hippolyte I., Perrier X., Xia L., Caig V., Evers E., Huttner E., Kilian A., Glaszmann A.C. (2009). Development and assessment of Diversity Arrays Technology for high-throughput DNA analyses in Musa. *Theor. Appl. Genet.*, 119, 1093-1103.
- Roncallo PF, Akkiraju PC, Cervigni GL, Echenique V (2017) QTL mapping and analysis of epistatic interactions for grain yield and yield-related traits in *Triticum turgidum* L. var. durum. *Euphytica* 213, 277. <https://doi.org/10.1007/s10681-017-2058-2>
- Roncallo P.F., Beaufort V., Larsen A.O., Dreisigacker S., Echenique V. (2019) Genetic diversity and linkage disequilibrium using SNP (KASP) and AFLP markers in a worldwide durum wheat (*Triticum turgidum* L. var durum) collection. *PLoS ONE*, 14(6): e0218562. <https://doi.org/10.1371/journal.pone.0218562>

- Roncallo P.F., Garbus I., Picca A., Echenique V., Carrera D.A., Cervigni G.L., Miranda R. (2009) Análisis de las bases genéticas del color en trigo candeal.1 *Rev.Fac.Agron.* Vol 108 (1): 9- 23.
- Roncallo, P. F., Larsen, A. O., Achilli, A. L., Pierre, C. S., Gallo, C. A., Dreisigacker, S., & Echenique, V. (2021). Linkage disequilibrium patterns, population structure and diversity analysis in a worldwide durum wheat collection including Argentinian genotypes. *BMC Genomics*, 22(1). <https://doi.org/10.1186/s12864-021-07519-z>
- Ruiz, M., Rossi, E. A., Bonamico, N. C., & Balzarini, M. G. (2021). Multi-trait models for genomic regions associated with mal de Río Cuarto and bacterial disease in maize. *Journal of Basic and Applied Genetics*, 32(Issue 1), 25–33. <https://doi.org/10.35407/bag.2021.32.01.03>
- Sakuma, S., Golan, G., Guo, Z., Ogawa, T., Tagiri, A., Sugimoto, K., Bernhardt, N., Brassac, J., Mascher, M., Hensel, G., Ohnishi, S., Jinno, H., Yamashita, Y., Ayalon, I., Peleg, Z., Schnurbusch, T., & Komatsuda, T. (2019). Unleashing floret fertility in wheat through the mutation of a homeobox gene. *Proceedings of the National Academy of Sciences*, 116(11), 5182–5187. <https://doi.org/10.1073/pnas.1815465116>
- Seghezzo, M.L. & E. Molfese. (1999). Trigo Candeal. Criterios para la evaluación de la calidad. Publicación Miscelaneas N° 2. Seghezzo, M.L. y Molfese E. (eds). Chacra Experimental Integrada Barrow, INTA-Ministerio de Asuntos Agrarios de la Provincia de Buenos Aires
- Suazo S, J., Santos M, L., Silva A, V., Jara S, L., Palomino Z, H., & Blanco C, R. (2005). Estudio de asociación por desequilibrio de ligamiento entre los genes TGFA, RARA, y BCL3 y fisura labiopalatina no sindrómica (FLPNS) en la población chilena. *Revista médica de Chile*, 133(9). <https://doi.org/10.4067/s0034-98872005000900008>
- Takuno S, Terauchi R, Innan H (2012) The Power of QTL Mapping with RILs. *PLoS ONE*, 7(10): e46545. <https://doi.org/10.1371/journal.pone.0046545>
- Taranto, F., Esposito, S., De Vita, P. (2023) Genomics for Yield and Yield Components in Durum Wheat. *Plants* 12, 2571. <https://doi.org/10.3390/plants12132571>
- Tinker N.A., Kilian A., Wight C.P., Heller-Uszynska K., Wenzl P., Rines H.W., Bjornstad A., Howarth C.J., Jannink J.L., Anderson J.M., Rossonagel B.G., Stuthman D.D., Sorrells M., Jackson E.W., Tuvevsson S., Kolb F.L., Olson O., Federizzi C., Carson M.L., Ohm H.W., Molnar S.J., Scoles G.J., Eckstein P.E., Bonman J.M., Ceplitis A., Langdon T., (2009). New DArT markers for oat provide enhanced map coverage and global germplasm characterization. *BMC genomics*.

doi:10.1186/1471-2164-10-39.

- Troccoli, A., Borrelli, G.M., De Vita P., Fares C., Di Fonzo, N. (2000). Durum wheat quality: a multidisciplinary concept. *Journal of Cereal Science* 32: 99–113.
- Uffelmann, E., Huang, Q.Q., Munung, N.S., De Vries J., Pkada Y., Martin A.R., Martin H.C., Lappalainen T., Pstuma D. (2021). Genome-wide association studies. *Nat Rev Methods Primers* 1, 59. <https://doi.org/10.1038/s43586-021-00056-9>
- Velasco López, J. L. (2013). *Segregación transgresiva en familias F2:7 de triticale forrajero (X Triticosecale Wittmack) en diferentes ambientes del Norte de México* [Tesis de maestría, Universidad Autónoma Agraria Antonio Narro]. Repositorio UAAAN. <http://repositorio.uaaan.mx:8080/xmlui/handle/123456789/7069>
- Yu, J.M., Pressoir, G., Briggs, W.H., Bi, I.V., 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. *Nat. Genet.* 38: 203-208.
- Zadoks, J.C., Chang, T.T., Konzak, C.F. (1974) A decimal code for the growth stage of cereals. *Weed Res.* 14:415-421.

Anexo A

Figura A. Análisis de regresión entre la Espigazón y las demás variables estudiadas

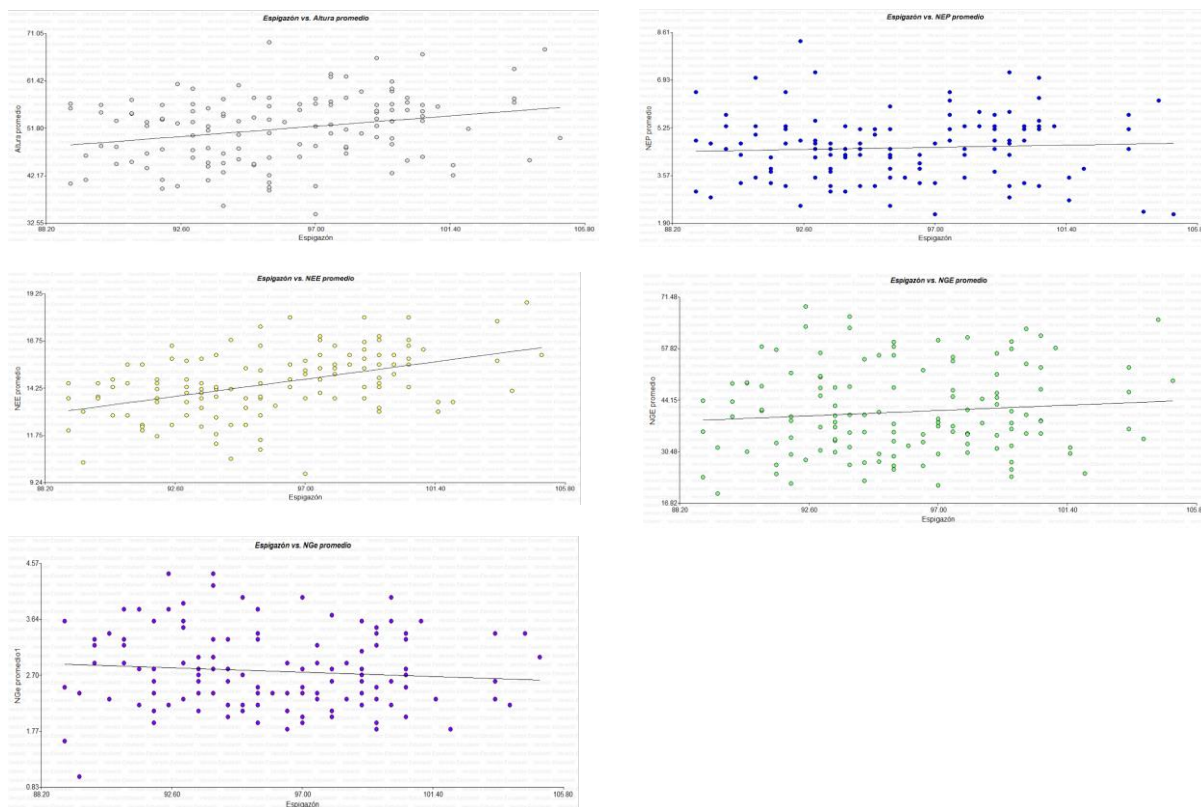


Figura B. Análisis de regresión entre la Altura promedio y el resto de las variables estudiadas

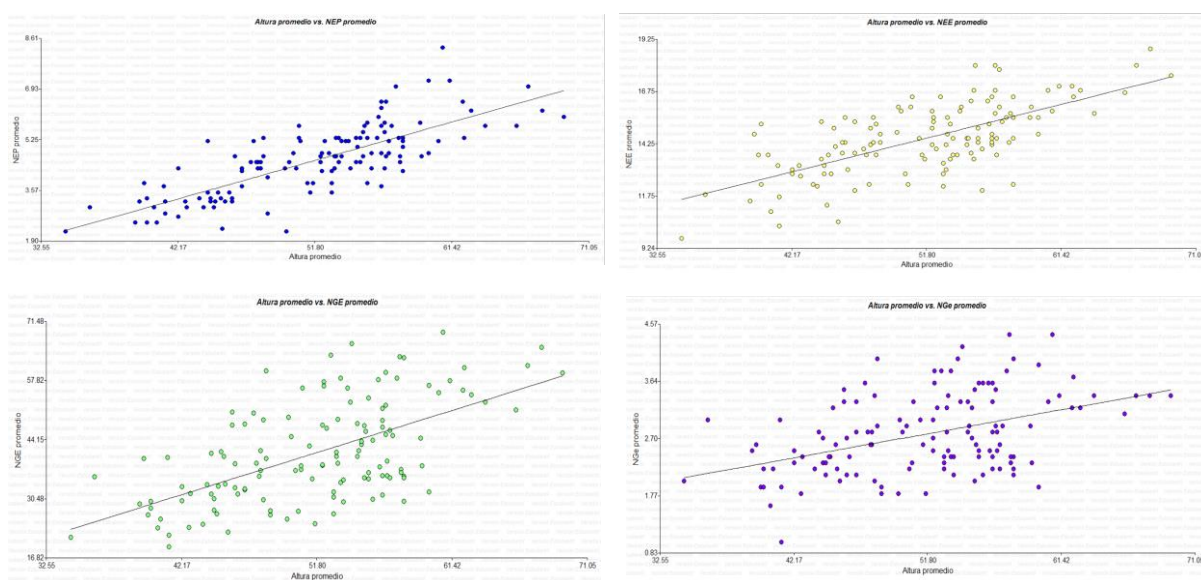


Figura C. Análisis de regresión entre NEP promedio y el resto de las variables estudiadas

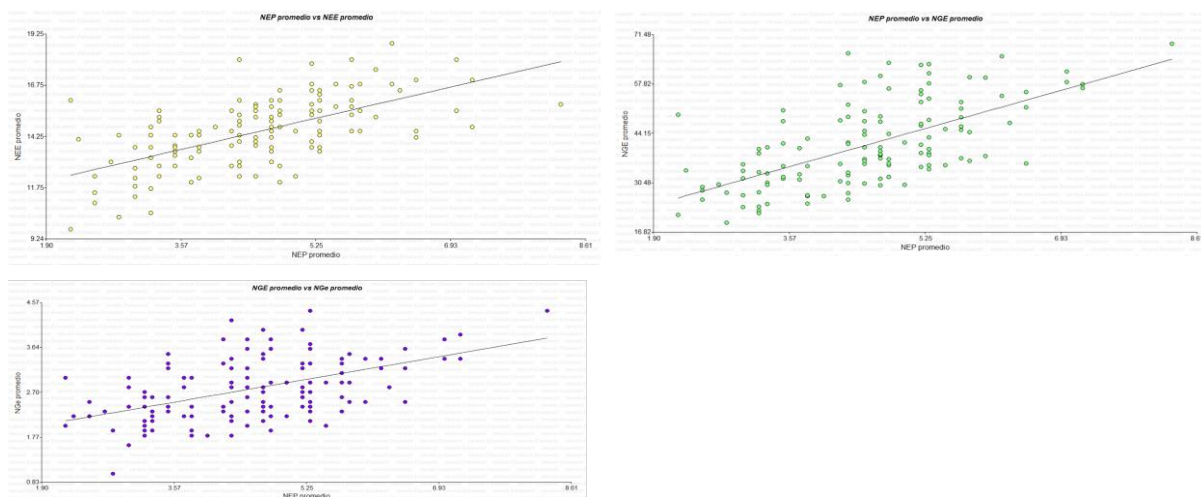


Figura D. Análisis de regresión entre NEE promedio y el resto de las variables estudiadas

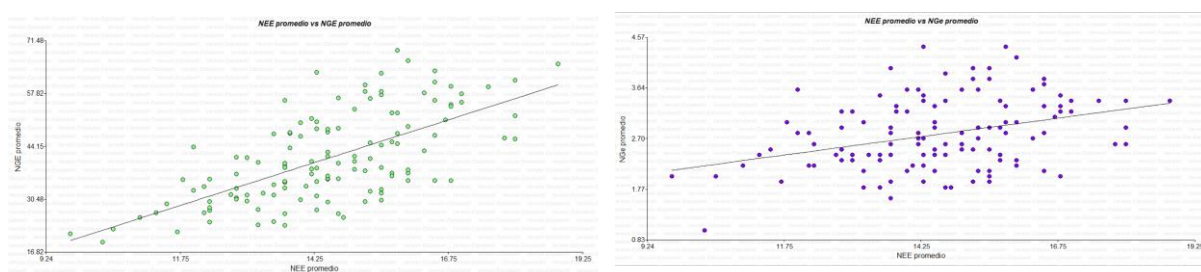
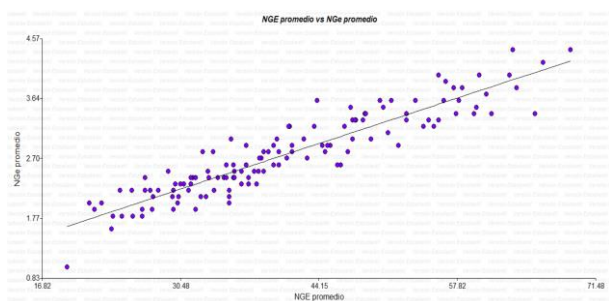


Figura E. Análisis de regresión entre NGE promedio y el resto de las variables estudiadas



Anexo B

Tabla 12. Posición de marcadores SNP asociados a los caracteres bajo estudio y los posibles genes candidatos (identificación del gen, cromosoma en el que se encuentra, posición en el genoma en pb y descripción funcional del mismo)

MM	Chr	Posición MM	ID del gen candidato	Chr*	pb inicio ID	pb de fin ID	Descripción funcional	Distancia mínima al gen candidato
32940342 F 0-13:A>G-13:A>G	2A	45851839	TRITD2Av1G023480	2A	45791690	45794434	NBS-LRR disease resistance protein%2C putative%2C expressed	57405
1110438 F 0-14:C>T-14:C>T	2A	46429124	No identificado					
4004331 F 0-18:T>C-18:T>C	2A	51874409	TRITD2Av1G026050	2A	51811235	51813781	Short chain dehydrogenase/reductase	60628
			TRITD2Av1G026060	2A	51822211	51823989	Subtilisin-like protease	50420
			TRITD2Av1G026070	2A	51826660	51829915	Serine/threonine-protein kinase	44494
1091747 F 0-52:G>A-52:G>A	2A	155070906	TRITD2Av1G066860	2A	155081235	155083808	Phototropic-responsive NPH3	10329
1215845 F 0-38:A>G-38:A>G	2A	161696203	No identificado					
12779268 F 0-25:A>C-25:A>C	2B	74498489	TRITD2Bv1G031560	2B	74514953	74520155	Dipeptidyl-peptidase%2C putative	16464
			TRITD2Bv1G031540	2B	74452070	74452965	Myb/SANT-like DNA-binding domain protein	45524
1714673 F 0-64:C>T-64:C>T	6A	483681213	No identificado					
997146 F 0-23:C>G-23:C>G	7B	736481446	No identificado					