



**FACULTAD DE CIENCIAS AGRARIAS
UNIVERSIDAD NACIONAL DE ROSARIO**

**INFLUENCIA DEL FONDO GENÉTICO SOBRE LA EXPRESIÓN DEL
CARÁCTER ALTA PROTEÍNA EN GRANO DE SOJA Y SU
CORRELACIÓN NEGATIVA CON EL RENDIMIENTO**

Licenciada en Biotecnología Mariana Gallo

TESIS PARA OPTAR AL TÍTULO DE MAGISTER EN GENÉTICA VEGETAL

DIRECTOR: Dr. José Rotundo

CO-DIRECTORES: MS. Ing. Agr. Martin Lisandro Oliva

Año 2018

**INFLUENCIA DEL FONDO GENÉTICO SOBRE LA EXPRESIÓN DEL CARÁCTER
ALTA PROTEÍNA EN GRANO DE SOJA Y SU CORRELACIÓN NEGATIVA CON EL
RENDIMIENTO**

Mariana Gallo

Licenciada en Biotecnología – Universidad Nacional de Rosario

Esta Tesis es presentada como parte de los requisitos para optar al grado académico de Magister en Genética Vegetal de la Facultad de Ciencias Agrarias de la Universidad Nacional de Rosario y no ha sido previamente presentada para la obtención de otro título en ésta u otra Universidad. La misma contiene los resultados obtenidos en investigaciones llevadas a cabo en las instalaciones del Departamento de Investigación Soja de Nidera Seeds Argentina S. A. U, Venado Tuerto, durante el período comprendido entre diciembre de 2014 y marzo de 2018, bajo la dirección del Dr. José Rotundo.

Nombre y firma del Maestrando: **Mariana Gallo**.....

Nombre y firma del Director: **José Rotundo**.... 

Nombre y firma del Consejero: **Martín Lisandro Oliva**.....

Defendida el día..... de.....de 20__

AGRADECIMIENTOS

Quiero agradecer a quienes hicieron posible la finalización de mis estudios de posgrado.

En primer lugar a mi hijo Juan Martin y a mi marido Gastón Codutti, por acompañarme y apoyarme en este desafío.

A mi director y co-director por su apoyo y guía.

A la empresa Nidera Seeds Argentina S. A. U., especialmente al equipo de Investigación Soja, por su colaboración, paciencia y apoyo estos 3 años.

A los profesionales y amigos que me incentivaron a cumplir este sueño Graciela Salas, Rodolfo Rossi, Cecilia Ghio, Sylvina Sartori, Marcos Mandrile, Esteban Martinez y Alejandra Peruzo.

A la Universidad Nacional de Rosario, y en particular a la Facultad de Ciencias Agrarias, por su contribución y el aporte de recursos humanos y materiales que forjaron mis conocimientos técnicos y mi desarrollo humano durante la maestría.

	Pág.
Índice General	i
Índice de figuras	iii
Índice de tablas	iv
Resumen	v
Abstract	vi
1. Introducción	1
1.1. Importancia de la proteína de soja a nivel mundial	1
1.2. Los niveles de proteína en las sojas de Argentina y rendimiento en grano	1
1.3. El mejoramiento genético de la soja	2
1.4. Los QTL como herramientas para el mejoramiento genético de soja direccionado a mejorar los niveles de proteína en soja	3
2. Hipótesis	4
3. Objetivos	5
3.1. Objetivos generales	5
3.2. Objetivos específicos	5
4. Materiales y métodos	6
4.1. Material vegetal	6
a. Selección de progenitores	6
b. Generación de las poblaciones	7
c. QTL alta proteína	9
4.2. Ensayos de líneas recombinantes	10
a. Ensayos comparativos de rendimiento	10
b. Análisis de proteína	10
4.3. Análisis estadístico	11
5. Resultados	12
5.1. Comportamiento para rendimiento de las poblaciones obtenidas	12
5.2. Comportamiento de la variable proteína sobre base seca de las poblaciones	13
5.3. Asociación entre las variables proteína y rendimiento en grano	14
5.4. Comportamiento de la variable proteína contra el estado del QTL dentro de cada población	16

5.5. Selección fenotípica contra selección asistida por marcadores moleculares	18
6. Discusión	23
7. Conclusiones	25
8. Bibliografía	26

Índice de Figuras	Pág.
Figura 1. Esquema de obtención de las líneas recombinantes	7
Figura 2. Gráficos de frecuencias de líneas contra nivel de proteína	14
Figura 3. Gráficos de dispersión de las líneas recombinantes de cada población de los datos de porcentaje de proteína en grano contra rendimiento en grano	15
Figura 4. Gráficos de dispersión de todos los datos de porcentaje de proteína en grano contra rendimiento en grano	16
Figura 5. Diagrama de cajas para la dispersión de los datos por población del porcentaje de proteína para QTL + contra QTL -	19
Figura 6. Gráfico de dispersión de los datos porcentaje de proteína contra rendimiento de los materiales selectos por marcadores moleculares	21
Figura 7. Gráfico de dispersión de la totalidad de los datos de porcentaje de proteína contra rendimiento de los materiales selectos por marcadores moleculares	22

Índice de Tablas	Pág.
Tabla 1. Progenitores y las características de selección	7
Tabla 2. Detalle de plantas F1 obtenidas de las poblaciones propuestas	8
Tabla 3. Detalle de las plantas F3 según estado de cigosidad del QTL	9
Tabla 4. Resumen de rendimiento de las poblaciones	12
Tabla 5. Análisis de medias de las poblaciones y los progenitores	13
Tabla 6. Análisis de regresión de las variables proteína y rendimiento	15
Tabla 7. Análisis de medias, mínimo y máximo del nivel de proteína de las poblaciones según las líneas recombinantes QTL + o QTL -	17
Tabla 8. Prueba de T para muestras independientes de las líneas QTL + contra QTL -	17
Tabla 9. Comparación del nivel de proteína de las poblaciones para el grupo de líneas QTL +	18
Tabla 10. Análisis de regresión de las variables proteína y rendimiento para líneas con QTL +	20
Tabla 11. Rendimiento en población total contra líneas con QTL +	24

RESUMEN

INFLUENCIA DEL FONDO GENÉTICO SOBRE LA EXPRESIÓN DEL CARÁCTER ALTA PROTEÍNA EN GRANO DE SOJA Y SU CORRELACIÓN NEGATIVA CON EL RENDIMIENTO.

La proteína de soja (*Glycine max* (L.) Merr) es una de las fuentes de proteína vegetal de alta calidad y bajo costo más utilizadas en el mundo. Una de las principales limitantes al desarrollo de variedades de alto porcentaje de proteína es su correlación negativa con el rendimiento. La asociación de un Locus de carácter cuantitativo (QTL por su nombre en inglés) de alta proteína del cromosoma 20 es una herramienta moderna para el mejoramiento en este carácter.

Los objetivos de esta tesis fueron evaluar la ganancia genética en proteína y su correlación con el rendimiento en poblaciones generadas a partir de diferentes variedades comerciales sin el QTL alta proteína cruzadas con líneas con el QTL alta proteína y comparar la ganancia en el porcentaje de proteína utilizando selección fenotípica contra selección asistida por marcadores moleculares.

Se generaron 4 poblaciones a partir de la cruce de dos genotipos con el QTL de alta proteína con dos materiales comerciales, uno de alta proteína y otro de baja proteína. Si bien los resultados del análisis rendimiento contra proteína fueron distintos en cada población (se observaron asociaciones tanto positivas como negativas), ninguna asociación fue significativa.

Cuando aplicamos selección fenotípica por nivel de proteína en grano sobre base seca, esta selección correlacionó en todos los casos con líneas con el QTL alta proteína presente. La información generada se utilizará para los mejoradores al momento de realizar mejoramiento orientado a aumentar los niveles de proteína, a partir de fuentes que contengan el QTL alta proteína asociado al cromosoma 20.

ABSTRACT

IMPACT OF GENETIC BACKGROUND OVER HIGH PROTEIN EXPRESSION ON SOYBEAN GRAIN AND NEGATIVE CORRELATION ON YIELD.

Soybean (*Glycine max* (L.) Merr) protein is a prominent plant-based protein source worldwide due to its high quality and relatively low cost. A major barrier to the development of high protein cultivars is the negative relationship between seed protein and seed yield. A protein quantitative trait locus (QTL) in chromosome 20 is a modern breeding tool to improve this character.

The objectives of this thesis were to evaluate the protein gain and its correlation with yield in populations developed from commercial cultivars and donor's cultivars; and to compare the percentages of protein gain by phenotypic selection versus molecular marker assisted selection.

Four populations were developed from two QTL donors and two commercial cultivars, one with high protein level and the other with low protein level. Even though the results obtained from the Yield versus Protein evaluation were different in each population, since some of them showed a positive correlation and others a negative one, none of them was significant.

When we used phenotypic selection by protein level in grain, the selection correlated with all materials containing QTL.

Information from this study will aid breeders in designing strategies to develop cultivars with increased seed protein levels from cultivars containing the high protein QTL associated with chromosome 20.

1.INTRODUCCIÓN

1.1. Importancia de la proteína de soja a nivel mundial.

La proteína de soja (*Glycine max* (L.) Merr) es una de las fuentes de proteína vegetal de alta calidad y bajo costo más utilizadas a nivel mundial para la cría de animales no rumiantes (Cromwell, 2012). El aceite de soja se obtiene a partir del proceso de extruzado y prensado del grano, de donde también se extrae posteriormente la harina. La harina de soja tiene los niveles más altos de proteína cruda entre las fuentes de proteína de origen vegetal. A nivel mundial, la proteína de soja cubre el 69 % de la proteína total utilizada en las dietas animales (Wilson, 2008; Cromwell, 2012). Los niveles de proteína en la harina deben estar entre 47-49 % de tenor proteico para su comercialización a nivel internacional. El precio de la harina es penalizado si esta no alcanza un nivel de proteína del 48 %, es por esto por lo que es de suma importancia que los fitomejoradores desarrollen cultivares con altos niveles de proteína (Brzostowski y Diers, 2017).

1.2. Los niveles de proteína en las sojas argentinas y rendimiento de grano.

Argentina se posiciona como el primer exportador de dicho subproducto del proceso de industrialización de la soja, debido a que posee un importante polo aceitero y un reducido consumo interno (BCR, 2012). La comercialización de la harina de soja tiene estándares de porcentaje de proteína mínimo para su venta. La caída de los valores ocasiona esfuerzos por parte del complejo de aceitero nacional para lograr una harina de soja que cumpla con los estándares de comercialización internacionales, situados en un 47-49 % de tenor proteico. Los principales países importadores de nuestras harinas de soja requieren harinas con 47 % de proteína, valor que se logra a partir de una molienda de granos con 38 % de proteína a 11 % de humedad. En la zona núcleo sojera de la República Argentina, los contenidos promedio de proteína en grano, hace dos décadas, oscilan entre el 38 y el 42 % sobre base seca (s. b. s.) y los de aceite entre el 20 y el 24 % s. b. s., ubicándose el grueso de la producción en valores de alrededor del 39 % de proteína y del 23 % de aceite sobre base seca (Cuniberti *et al.*, 2004; Cuniberti *et al.*, 2011).

En nuestro país se viene observando una caída en el contenido proteico de la soja que llegó, en las campañas 2012/13 y 2013/14, a los valores más bajos de los últimos 17 años, no sólo en proteína sino también en aceite (Cuniberti *et al.*, 2014). En muchos casos, esta baja concentración de proteína impidió alcanzar los estándares de harinas High Pro. Esta categoría posee mayor nivel

de proteína y menor nivel fibra, por lo que es adecuada para la alimentación de cerdos y aves. Se obtiene de un descascarado previo del poroto de soja, que al eliminar la cáscara favorece la concentración del contenido proteico y de granos con un nivel de proteína superior a 38%. Lo que provocó la necesidad de renegociar algunos contratos de venta de harina. Desde hace unos años, Argentina no comercializa harina con valores del 47 % de proteína. Inclusive, viene usando como estrategia la compra de grano en otros países, como Paraguay, para hacer mezclas y lograr mejores niveles de proteína. También se ha venido utilizando como estrategia el secado del producto para aumentar los niveles de proteína, implicando un mayor gasto energético y su consecuente aumento de los costos de producción. Debido a esta depresión en los niveles proteicos, se han definido nuevos estándares nacionales para la categoría High Pro. La nueva base de comercialización de harina de producción argentina es 46,5 % base de proteína. Porcentajes de proteína menores a la base, son sujetos a penalizaciones en la comercialización.

Esta tendencia en la baja de los niveles de proteína no solo afecta a Argentina, este mismo patrón se viene dando en Brasil y Estados Unidos (Pípolo et al., 2016), en el país vecino en la cosecha 2014/2015 los valores de proteína sobre base seca estuvieron en torno al 36%, según el análisis llevado a cabo por el Embrapa Soja, a través de todo el territorio. Este patrón se viene observado desde los noventas en dicho país y muestra las mismas dificultades comerciales que en Argentina.

1.3. El mejoramiento genético de la soja

En los últimos años, el mejoramiento genético en el país se ha enfocado en la demanda del productor, que se concentra principalmente en el rendimiento. Una de las consecuencias de la priorización de este carácter fue la reducción del porcentaje de proteína (De Felipe *et al.* 2016). Dado que hasta la actualidad no ha habido incentivos económicos directos para mejorar la calidad, este carácter no ha sido tenido demasiado en cuenta en el desarrollo de cultivares. El productor calcula su margen en base al precio del grano y su margen bruto se basa en el rendimiento por hectárea.

Una de las principales limitantes al desarrollo de variedades de alto porcentaje de proteína es su correlación negativa con el rendimiento. (Hanson, 1991; Chung *et al.*, 2003; Cober y Voldeng, 2000). Sin embargo, el lanzamiento al mercado de algunas variedades de soja con altos niveles de proteínas y buenos rendimientos (Bosaz *et al.*, 2017; Herrero *et al.*, 2017) sugiere que

existe la posibilidad de desarrollar variedades con estas características mediante la combinación apropiada de genética, ambientes e interacción genotipo por ambiente, a través del uso de las herramientas de mejoramiento correctas (Panthee *et al.*, 2006).

1.4. Los QTL como herramientas para el mejoramiento genético de soja direccionado a mejorar los niveles de proteína en soja.

La proteína en soja se hereda como una característica cuantitativa y hay Locus de carácter cuantitativo (*QTL*) asociados a nivel de proteína en grano en todos los cromosomas (SoyBase, 2018). A la fecha, se encuentran reportados 254 marcadores asociados al nivel de proteína en grano en la base de datos genéticos de la soja SoyBase (SoyBase, 2018).

En 1992, Diers *et al.* identificó dos QTL asociados a alta proteína en una población derivada de la PI468916, uno en el cromosoma 20 y otro en el cromosoma 15. Se han publicado investigaciones donde se evalúan QTL que mapean en el cromosoma 15 (Brzostowski *et al.*, 2017) y se confirma que la presencia de este QTL aumentó el nivel de proteína en dos fondos genéticos distintos y no afectó el rendimiento de los materiales. Mientras que en el caso del cromosoma 20, el QTL está altamente asociado a alta proteína (Warrington *et al.*, 2011), se han reportado efectos contrastantes de este QTL en el rinde de soja. Brzostowski *et al.* (2017) concluye sobre un efecto negativo, mientras que Harris (2001) y Yates (2006) encontraron efectos positivos. El QTL evaluado en esta tesis está ligado al cromosoma 20.

Es importante contar con QTL asociados a nivel de proteína, pero a su vez es necesario evaluar los QTL a través de múltiples fondos genéticos y ambientes antes de incorporarlos a los programas de mejoramiento debido a la compleja relación entre los componentes de la semilla y el rendimiento (Rotundo *et al.*, 2009; Rotundo *et al.*, 2011). A la fecha hay mayor disponibilidad de herramientas moleculares para incorporar en los programas de mejoramiento a menores costos y más industrializadas.

2. HIPÓTESIS

1.- La correlación entre el nivel proteína en grano y el rendimiento es dependiente de la fuente de alta proteína y del fondo genético de la variedad comercial.

2.- El QTL asociado al cromosoma 20 explica un porcentaje elevado de la variación del porcentaje de proteína y es un QTL de fácil selección molecular.

3. OBJETIVOS

3.1 Objetivos generales

1.- Evaluar la ganancia en el nivel de proteína y su correlación con el rendimiento, en líneas recombinantes generadas de la combinación de líneas de alta proteína con el QTL de alta proteína del cromosoma 20 con variedades comerciales sin el QTL.

2.- Demostrar que el QTL asociado a alta proteína del cromosoma 20 explica un porcentaje elevado de la variación en proteína y comparar la selección fenotípica contra selección asistida por marcadores moleculares.

3.1 Objetivos específicos

1.- Generar 4 set de líneas recombinantes, desde el cruzamiento hasta la generación F5, a partir de la combinación de 4 progenitores.

2.- Caracterizar molecularmente según el QTL de alta proteína todas las plantas de cada población en la generación F3.

3.- Llevar a cabo un ensayo comparativo de rendimiento entre las líneas recombinantes caracterizadas molecularmente y sus progenitores.

4.- Analizar los niveles de proteína del grano de las líneas participantes en el ensayo comparativo de rendimiento.

5.- Evaluar estadísticamente la correlación entre los niveles de proteína y el rendimiento de los datos obtenidos en el ensayo comparativo de rendimiento.

6.- Llevar a cabo la selección fenotípica del 20% de las líneas de cada set de líneas recombinantes, según el carácter nivel de proteína en grano.

7.- Llevar a cabo la selección de las líneas recombinantes con el QTL alta proteína presente en los 4 set de líneas.

8.- Comparar estadísticamente los resultados de la selección fenotípica con la selección a partir del QTL alta proteína.

4. MATERIALES Y MÉTODOS

4.1. Material vegetal

a. Selección de progenitores

Los progenitores comerciales utilizados para crear las poblaciones se seleccionaron a partir de los datos generados por la RECSO (Red Nacional de Evaluación de Cultivares de soja) en la sección Calidad de Grano de la campaña 2012-2013, cuyo análisis corresponde a datos desde la campaña 2009/2010 hasta 2012/2013, a datos de Cuniberti *et al.*, 2011 y a datos propios de la empresa y su elección se basó en los factores tales como estabilidad y rendimiento, baja proteína y alta proteína. (Tabla 1).

Se seleccionaron progenitores de grupos de madurez corto para lograr ciclos de cultivo cortos y cumplir los plazos de evaluación (Tabla 1).

Los genotipos de alta proteína con el QTL asociado al cromosoma 20 (Spro 12-61 y Spro 12-17), fueron obtenidos por el departamento de biotecnología de Nidera Seeds Argentina S.A.U, así como los marcadores asociados al QTL alta proteína. Los genotipos se obtuvieron por retrocruza entre NA 4613RG y Danbaekkong (PI619083), utilizando la selección asistida por marcadores moleculares (Warrington et al., 2011). NA 4613RG es un cultivar comercial que tiene estándares relativamente altos de proteína, según datos obtenidos por Cuniberti, et al. (2014), además de ser una variedad altamente cultivada y de estabilidad reconocida. Danbaekkong (PI619083) fue el progenitor utilizando para aumentar los niveles de proteína, ya que se caracteriza por tener niveles muy altos de proteína, pero con características agronómicas poco comerciales. Las líneas en BC3F3 fueron evaluadas a campo en un ensayo comparativo de rendimiento en la campaña 2013-2014, por el equipo de trabajo del departamento de Biotecnología de Nidera Seeds S.A.U. A partir de los ensayos mencionados, se seleccionaron dos genotipos alta proteína (Spro 12-61 y Spro 12-17) que se caracterizaron por presentar altos contenidos de proteína (44,9 % sobre base seca (s.b.s.) y 41,8 % s.b.s.), en comparación con el progenitor NA 4613RG, con 37,8 % (s.b.s.). El rendimiento de las líneas de alta proteína selectas fue de 2,3 t/ha y 3,3 t/ha mientras que NA 4613RG rindió 3,8 t/ha.

Tabla 1: Progenitores y las características de selección.

PROGENITORES	GRUPO DE MADUREZ	HÁBITO DE CRECIMIENTO	CARACTERÍSTICA	PORCENTAJE DE PROTEÍNA s.b.s.
Spro 12-61	3,7	Indeterminado	alta proteína	44,9
Spro 12-17	3,7	Determinado	alta proteína	41,8
DM 3810	3,8	Indeterminado	Comercial baja proteína	35,2
A 3302RG	3,3	Indeterminado	Comercial alta proteína	38,6

Tabla 1. Progenitores: Listado de variedades utilizadas para generar las poblaciones, Grupo de Madurez: corresponde al tiempo que dura el ciclo del cultivo, desde emergencia a madurez (R8), Hábito de crecimiento: tipos de desarrollo del tallo principal de la planta de soja, Determinado, detiene su crecimiento cuando comienza a florecer e Indeterminado, sigue elongando el tallo principal luego del inicio de floración, Característica: cualidad por la que se seleccionó el material para este trabajo, Porcentaje de proteína: nivel de proteína del grano.

b. Generación de líneas recombinantes:

Las poblaciones se generaron en el departamento de investigación Soja de Nidera Seeds Argentina S.A.U. Figura (1)

Figura 1: Esquema de obtención de las líneas recombinantes

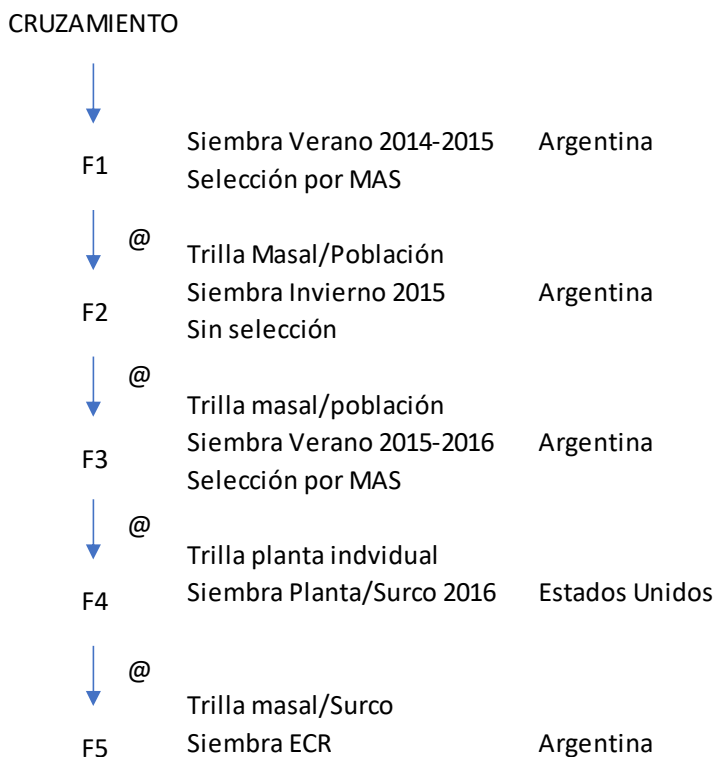


Figura 1. A la izquierda los distintos ciclos de crecimiento y autofecundación (@). Al medio las actividades llevadas a cabo en esa generación y a la derecha el país donde se llevó a cabo el ciclo de crecimiento.

Los cruzamientos se llevaron a cabo en invierno del 2014. Utilizándose en todos los casos los genotipos con el QTL de alta proteína como macho y los materiales comerciales como hembra. Además, los genotipos comerciales se cruzaron por ambos genotipos de alta proteína.

Las semillas de cada cruzamiento fueron sembradas en Venado Tuerto durante la campaña agrícola 2014-2015, donde se verificó la cruce por medio de marcadores moleculares y se seleccionaron plantas con el QTL de alta proteína en estado heterocigota. De estas cruza, se obtuvieron 16 plantas F1 (Tabla 2). Las plantas F1 obtenidas se trillan en masal por población.

Tabla 2: Detalle de plantas F1 obtenidas de las poblaciones propuestas.

POBLACION	PLANTAS F1
DM 3810*Spro 12-61	1
DM 3810*Spro 12-17	8
A 3302RG*Spro 12-61	3
A 3302RG*Spro 12-17	4

Tabla 2. Población: nombre asignado a cada cruzamiento efectuado. Planta F1: Plantas logradas de cada población, verificadas por marcadores moleculares.

A continuación, se sembró toda la semilla F2 en contra estación (invierno de 2015) en Bella Vista, Corrientes, para realizar un segundo ciclo de autofecundación. La cosecha se realizó en masal de todas las plantas F2-semilla F3 de cada población.

En la estación de crecimiento 2015-2016, se sembraron las semillas F3 de cada población y en el estado V3-V5 se llevó a cabo un muestreo de ADN en papel nucleico para caracterizar las plantas según el estado de cigosidad del QTL de alta proteína. Se analizaron 882 plantas F3 que se trillaron individualmente, distribuidas entre heterocigotas, homocigotas alta proteína (QTL +) y homocigotas tipo silvestre (QTL-) para el QTL alta proteína del cromosoma 20. En la (Tabla 3) se detalla la distribución en las distintas poblaciones.

Cada planta F4 fue sembrada individualmente en surcos de 3 metros de longitud en contra estación (invierno 2016) en Arkansas, Estados Unidos. En este ciclo de crecimiento, se caracterizó cada surco por su color de flor, grupo de madurez (según testigos de referencia colocados en el mismo ensayo) y hábito de crecimiento. Se cosecharon las semillas F5 de cada surco por separado en masal.

En la generación F5, se llevaron a cabo los ensayos comparativos de rendimiento en el ciclo de crecimiento 2016-2017 en Argentina.

Tabla 3: Detalle de plantas F3 según estado de cigosidad del QTL

POBLACION	CIGOSIDAD QTL alta proteína	PLANTAS F3	TOTAL PLANTAS
DM 3810*SPRO 12-61	HETEROCIGOTA	13	48
	HOMOCIGOTA ALTA PROTEINA	24	
	HOMOCIGOTA TIPO SILVESTRE	11	
DM 3810*SPRO 12-17	HETEROCIGOTA	67	251
	HOMOCIGOTA ALTA PROTEINA	86	
	HOMOCIGOTA TIPO SILVESTRE	96	
	SIN DATO	2	
A 3302RG*SPRO 12-61	HETEROCIGOTA	27	141
	HOMOCIGOTA ALTA PROTEINA	41	
	HOMOCIGOTA TIPO SILVESTRE	72	
	SIN DATO	1	
A 3302RG*SPRO 12-17	HETEROCIGOTA	132	442
	HOMOCIGOTA ALTA PROTEINA	157	
	HOMOCIGOTA TIPO SILVESTRE	153	

Tabla 3. Resultado del análisis por marcadores moleculares de las 4 poblaciones en las plantas F3.

c. QTL alta proteína:

El análisis de genotipado de los individuos pertenecientes a las poblaciones desarrolladas en el presente trabajo se realizó en una plataforma de genotipado SNPLINE LGC Genomics. Para el desarrollo del ensayo KASP NAGMPR01 utilizado en la presente tesis, se envió la secuencia de nucleótidos del SNP correspondiente a grupo de ligamiento “I” (Chr20), previamente identificados en el departamento de Biotecnología, a la empresa LGC. Los cebadores alelo específicos fueron diseñados de forma que la variante asociada al carácter de alto contenido de proteína en semilla (HPC) se marcó con el fluoróforo HEX (Hexachlorofluorescein) y el alelo nativo con el fluoróforo FAM (Fluorescein amidite). La reacción de PCR se realizó en formato de placa de 384 pocillos en un volumen final de 3 µl: 10-20 ng/µl de ADN seco, 0.042 µl de mix de ensayo, 1.46 µl de KASP Mix y H₂O csp. 3 µl. La reacción de amplificación fue realizada en un equipo Hydrocycler 16, utilizando el siguiente protocolo: 95 °C por 15 min, seguido por diez ciclos de touchdown (95 °C for 20s/ touchdown a 65 °C inicialmente y

disminuyendo 1 °C por ciclo, 25 s), seguido por 30 ciclos de hibridación y extensión (95 °C por 10 s; 57 °C por 60 s). El producto de amplificación obtenido para cada reacción fue analizado en un equipo Pherastar.

4.2. Ensayos de líneas recombinantes.

a. Ensayos comparativos de rendimiento:

Se realizó un ensayo en Bloques Completos al Azar con 2 repeticiones (unidad experimental: 2 surcos a 0.52 m por 4 metros de largo) con todas las líneas homocigotas para el QTL y homocigotas del tipo silvestre de las 4 poblaciones en estudio, más DM 3810, Spro 12-17, Spro 12-61 y A 3302RG estas como testigos.

Durante el ciclo del cultivo, se llevó a cabo un muestreo de ADN a los fines de realizar un control por marcadores moleculares de todos los materiales, lo que permitió definir aquellos que estaban sin datos en el análisis original y verificar los datos originales de las plantas F3.

Se llevo a cabo la siembra de las 649 parcelas (correspondientes a las líneas de cada población más los testigos), la misma se llevó a cabo el día 1 de diciembre del 2016 en el campo experimental de Nidera Seeds Argentina S.A.U., El Recuerdo, ubicado en Venado Tuerto (-62.060855°;-33.729288°). El ensayo, se realizó sobre rastrojo de maíz, se mantuvo libre de malezas e insectos durante todo el ciclo de crecimiento. El registro de lluvias fue de 1137 mm durante el ciclo del cultivo.

Con anterioridad a la cosecha, se verificó la longitud de las parcelas y se corrigió el largo en los casos que fue necesario. Los 4 metros de parcela se cosecharon a mano y a máquina. En todas las parcelas se llevó adelante el pesaje y la medición del porcentaje de humedad. Además, se realizó el cálculo de rendimiento en toneladas sobre hectárea (t/ha), corregido al 13 % de humedad.

b. Análisis de proteína:

Se analizaron los niveles de proteína de todas las parcelas del ensayo, ambas repeticiones por separado. Para medir los niveles de proteína sobre base seca se utilizó el equipo NIR (espectroscopia de infrarrojo cercano), perteneciente a la Facultad de Ciencias Agrarias de Zavalla, puesto a punto con Foss Argentina (De Felipe et al., 2016). Las 1298 muestras, de 200 centímetros cúbicos de semilla cada una, se analizaron por NIR.

4.3. Análisis estadístico:

El análisis de datos se realizó mediante el módulo de modelos lineales y mixtos utilizando el software estadístico Infostat (Di Rienzo J.A; *et al.*, versión 2017), donde el modelo empleado para las variables proteína y rendimiento fue:

$$Y_{ijk} = \mu + \text{Repetición}_i + \text{Población}_j + \text{Líneas}_{k(j)} + e_{ijk}.$$
; donde Repetición y Población fueron tomados como efectos fijos, en tanto que Líneas dentro de Población fue tomada como efecto aleatorio. Entiéndase por Población a las poblaciones y sus progenitores. (Spro 12-61, Spro 12-17, DM 3810, A 3302RG, DM 3810*Spro 12-61, DM 3810*Spro 12-17, A 3302RG*Spro 12-61 y A 3302RG*Spro 12-17)

A modo de poder conocer y describir cómo fue la variabilidad de los datos es que se estimaron los parámetros de curtosis y asimetría. La curtosis de los datos de cada población se midió promediando la cuarta potencia de la diferencia entre cada elemento del conjunto y la media, dividido entre la desviación típica elevado también a la cuarta potencia. Las curvas se pueden clasificar en tres grupos según el signo de su curtosis, es decir, según la forma de la distribución:

- Leptocúrtica: la Curtosis > 0. Los datos están muy concentrados en la media, siendo una curva muy apuntada.
- Mesocúrtica: la Curtosis = 0. Distribución normal.
- Platicúrtica: la Curtosis < 0. Muy poca concentración de datos en la media, presentando una forma muy achatada.

El coeficiente de asimetría de Fisher CAF evalúa la proximidad de los datos a su media $\bar{x}$. Cuanto mayor sea la suma $\sum (x_i - \bar{x})^3$, mayor será el coeficiente. Si CAF < 0: la distribución tiene una asimetría negativa y se alarga a valores menores que la media.

- Si CAF = 0: la distribución es simétrica.
- Si CAF > 0: la distribución tiene una asimetría positiva y se alarga a valores mayores que la media.

5. RESULTADOS

5.1. Comportamiento para rendimiento de las poblaciones obtenidas.

Para el componente rendimiento en grano se detectaron diferencias significativas entre las distintas poblaciones (p-valor: 0,0001). En la (Tabla 4) se detallan el rendimiento promedio y otros parámetros poblacionales y de los progenitores que robustecen el análisis.

Tabla 4: Resumen de rendimiento de las poblaciones.

Resumen de rendimiento (t/ha)						
Poblaciones	n	Media	(*)	D.E.	Mín.	Máx.
Spro12-17	12	4,54	a	0,65	3,43	5,61
A3302RG	5	4,50	a b	0,52	4,04	5,34
DM 3810*SPRO 12-17	318	4,45	a b	0,55	2,70	5,66
DM3810	25	4,37	a b	0,56	3,43	5,36
A 3302RG*SPRO 12-61	198	4,08	b	0,60	2,56	5,72
A 3302RG*SPRO 12-17	575	4,06	b	0,55	2,65	5,71
DM 3810*SPRO 12-61	58	3,80	c	0,58	2,64	5,09
Spro12-61	11	3,69	c	0,65	2,73	4,52

Tabla 4: n: número de parcelas evaluados, Media: media del rendimiento en t/ha de toda la población, (*) Valores seguidos de igual letra no difieren entre sí significativamente, con un valor de Diferencia Mínima Significativa (DMS) de 0,405 t/ha. D.E.: desvío estándar del rendimiento, Mín.: valor mínimo de rendimiento de la población, Máx.: Valor máximo de rendimiento de la población. En negrita los progenitores.

Para el caso de las dos poblaciones con la hembra A 3302RG (A 3302RG*SPRO 12-61 y A 3302RG*SPRO 12-17), se observaron rendimientos mínimos semejantes a los progenitores alta proteína con el QTL y rendimientos máximos superiores a los 3 progenitores, mostrando ambas, casi la misma dispersión de datos solo que en el caso de la población A 3302RG*SPRO 12-61 con una asimetría negativa, siendo la asimetría positiva para el caso de la población A 3302RG*SPRO 12-17. Es importante aclarar que la asimetría nos indica si la mayor concentración de datos se encuentra hacia mayores valores (asimetría negativa) o hacia los menores valores (asimetría positiva).

Para el caso de las poblaciones con la hembra DM 3810 (DM 3810*SPRO 12-61 y DM 3810*SPRO 12-17), el rendimiento mínimo de la población DM 3810*SPRO 12-17 fue ligeramente superiores a los de las otras tres poblaciones, igualmente fue menor que el de los progenitores. El rendimiento máximo fue superior que el de los progenitores en la población DM 3810*SPRO 12-17 y menores a los de los progenitores en la DM 3810*SPRO 12-61,

correlacionando con el dato del menor rendimiento observado en progenitor Spro 12-61, genotipo alta proteína en esta población. En el caso de la población DM 3810*SPRO 12-17, la asimetría fue negativa, mientras que en la población DM 3810*SPRO 12-61 la asimetría fue positiva.

5.2. Comportamiento de la variable proteína sobre base seca de las poblaciones.

Para el componente proteína en grano se detectaron diferencias significativas entre las distintas poblaciones (p-valor: 0,0001), ya que se observaron divergencias entre las poblaciones generadas, como así también entre estas y sus progenitores. En la Tabla 5 se detallan los niveles promedio de la variable y otros parámetros poblacionales que robustecen el análisis.

Tabla 5: Análisis de medias de proteína de las poblaciones y los progenitores.

Proteína (% sbs)						
Poblaciones	n	Media	(*)	D.E.	Mín.	Máx.
A 3302RG*SPRO 12-17	573	39,41	a	1,13	36,64	42,58
Spro12-61	10	39,29	a	1,30	35,94	40,42
A 3302RG*SPRO 12-61	195	39,19	a b	1,17	37,16	41,88
Spro12-17	12	38,73	a b	1,22	35,86	40,18
DM 3810*SPRO 12-61	56	38,56	a b	1,33	35,63	40,37
DM 3810*SPRO 12-17	310	37,47	c	1,17	34,93	40,19
A 3302RG	6	36,84	c d	1,48	35,10	38,71
DM 3810	25	36,00	d	0,84	34,94	38,47

Tabla 5: n: número de muestras analizadas, Media: media de porcentaje de proteína en grano sobre base seca de todas las líneas de la población, (*) Valores seguidos de igual letra no difieren entre sí significativamente, con un valor de DMS de 0,957. D.E.: desvío estándar del porcentaje de proteína, Mín.: valor mínimo de proteína en grano sobre base seca de la población, Máx.: Valor máximo de proteína en grano sobre base seca de la población. El nivel de proteína es tomado en porcentaje sobre base seca. En negrita los progenitores.

La población A 3302RG*SPRO 12-17 fue la que mostró el mayor nivel de proteína promedio y registró el valor máximo entre todas las poblaciones (42,58 % s.b.s.) e inclusive superó a todos los progenitores. Si bien el valor de proteína de A 3302RG fue inferior al esperado (Cuniberti *et al.*, 2011), el correspondiente a dos poblaciones con esta línea como hembra mostró los mayores índices de proteína entre las 4 poblaciones.

En todas las poblaciones, se pudieron detectar líneas con niveles del 40 % de proteína dentro de sus valores máximos.

La curtosis de las 4 poblaciones fue semejante entre ellas (-0,79, -0,88, -0,89 y -1,07) (Figura 2).

Figura 2: Gráficos de frecuencia de líneas contra nivel de proteína.

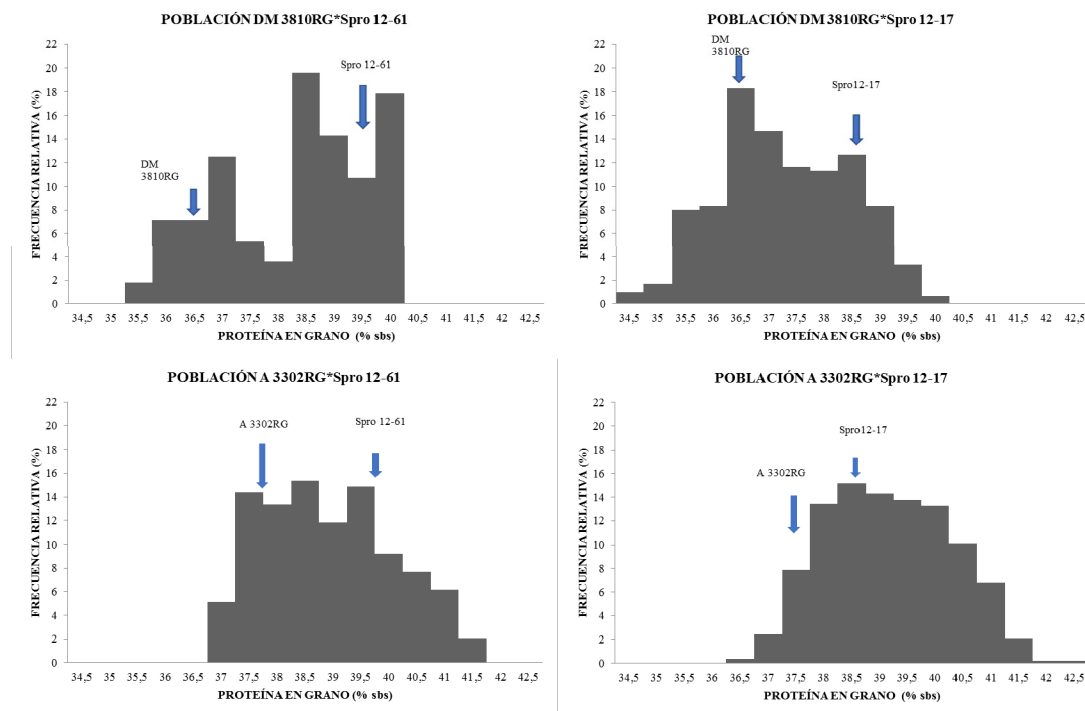


Figura 2. Gráficos de frecuencias, donde se indican las ubicaciones relativas que ocuparon los progenitores en la distribución de niveles de proteína de cada población y donde se refleja el grado de curtosis de los datos.

5.3. Asociación entre las variables proteína y rendimiento en grano

Se realizaron regresiones con el conjunto total de datos y luego por población donde se observaron tendencias positivas y negativas muy leves (valores “b”, Tabla 6) con niveles de correlaciones despreciables (valores “r”, Tabla 6), excepto para el conjunto de datos total. En él, la correlación resultó significativa (significancia < 0,0001) pero con muy baja correlación ($r = -0,15$), lo cual demuestra, para este set de datos, una asociación pobre entre ambas variables. Se realizó un gráfico de dispersión Rendimiento contra Proteína en grano (Figura 3), donde se puede observar el comportamiento de cada población y el total de los datos (Figura 4).

Tabla 6: Análisis de regresión de las variables proteína y rendimiento.

Poblaciones	B	r	significancia
TOTAL	-0,06	-0,15	< 0,0001
DM 3810*SPRO 12-61	-0,03	-0,06	0,65
DM 3810*SPRO 12-17	0,02	0,05	0,38
A 3302RG*SPRO 12-61	-0,06	-0,11	0,13
A 3302RG*SPRO 12-17	0,03	0,06	0,16

Tabla 6. b: pendiente de la regresión de los datos, tipo de correlación de las variables, r: ajuste de los datos a la regresión, TOTAL: análisis llevado a cabo con la totalidad de los datos de todas las poblaciones.

Se puede observar la pendiente casi nula para las regresiones de las poblaciones DM 3810*SPRO 12-61, DM 3810*SPRO 12-17 y A 3302RG*SPRO 12-17, mientras que la población A 3302RG*SPRO 12-61 mostró una pendiente muy semejante a la Total, negativa y de mayor magnitud. Mostrando que casi no habría correlación negativa entre el rendimiento y el porcentaje de proteína para estas poblaciones en este ambiente.

Figura 3: Gráfico de dispersión de las líneas recombinantes de cada población de los datos de porcentaje de proteína en grano contra rendimiento en grano.

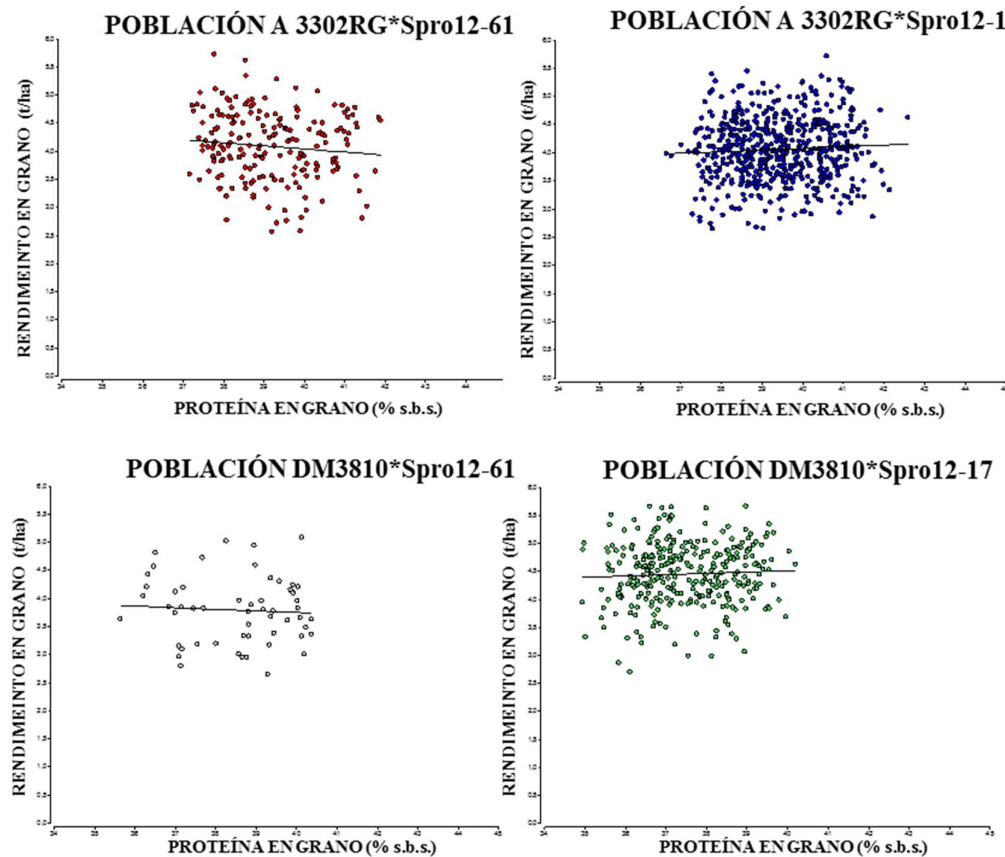


Figura 3. Análisis de dispersión rendimiento (t/ha) contra porcentaje de proteína sobre base seca, de las distintas poblaciones. No incluye a los progenitores.

Figura 4: Gráfico de dispersión de todos los datos de porcentaje de proteína en grano contra rendimiento en grano.

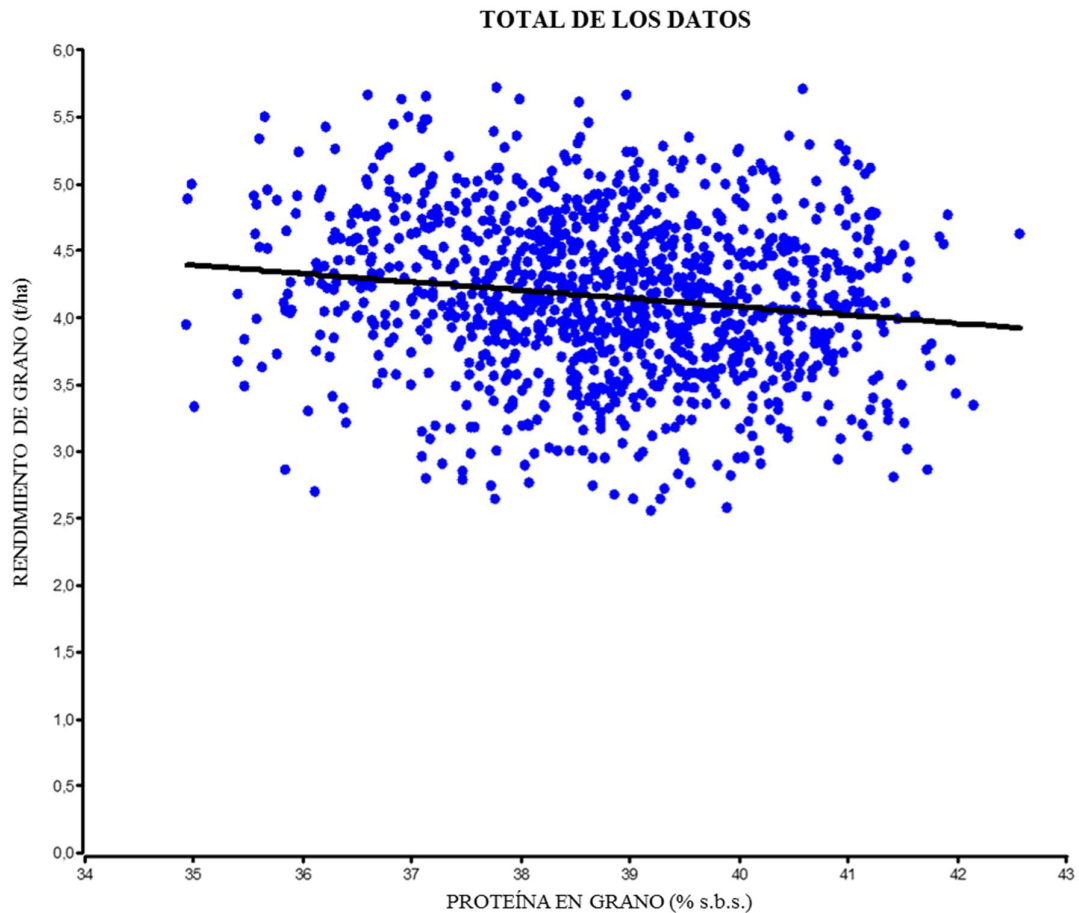


Figura 4. Análisis de dispersión rendimiento (t/ha) contra porcentaje de proteína sobre base seca, de las distintas poblaciones como así también del conjunto total de datos. No incluye a los progenitores.

5.4. Comportamiento de la variable proteína contra el estado del QTL dentro de cada población.

Se llevó a cabo el análisis de las poblaciones segregando las variedades que fueron caracterizadas con el QTL homocigota (QTL +) y el QTL homocigota del tipo silvestre (QTL -). Cuando se analizaron por separado los materiales con el QTL +, las cuatro poblaciones mostraron mayores valores promedio que aquellas con el QTL - (Tabla 7).

Tabla 7: Análisis de medias, mínimo y máximo del nivel de proteína de las poblaciones según las líneas recombinantes con el QTL + o QTL -.

Poblaciones		Media	D.E.	Mín.	Máx.
A 3302RG*SPRO 12-17	QTL +	40,27	0,74	38,19	42,58
A 3302RG*SPRO 12-17	QTL -	38,52	0,69	36,64	41,92
A 3302RG*SPRO 12-61	QTL +	40,42	0,7	38,80	41,88
A 3302RG*SPRO 12-61	QTL -	38,48	0,7	37,16	40,29
DM 3810*SPRO 12-17	QTL +	38,45	0,76	35,87	40,19
DM 3810*SPRO 12-17	QTL -	36,57	0,63	34,93	38,06
DM 3810*SPRO 12-61	QTL +	39,33	0,68	37,70	40,37
DM 3810*SPRO 12-61	QTL -	37,06	0,95	35,63	40,37

Tabla 7. Análisis de la media de cada población agrupada por QTL + y QTL -Media: media del porcentaje de proteína sobre base seca, D. E. desvío estándar; Mín.: valor mínimo de porcentaje de proteína sobre base seca; Máx.: valor máximo de porcentaje de proteína sobre base seca. No incluye a los progenitores.

Cuando se evaluaron dentro de cada población, las líneas con QTL + contra QTL - siempre mostraron una diferencia significativa en cada población (Tabla 8).

Tabla 8: Prueba de T para muestras independientes de las líneas QTL + contra QTL -.

Poblaciones	Grupo 1	Grupo 2	n(1)	n(2)	Media (1)	Media (2)	Media(1)-Media (2)	p-valor
A 3302RG*SPRO 12-17	QTL +	QTL -	290	279	40,27	38,52	1,75	<0,0001
A 3302RG*SPRO 12-61	QTL +	QTL -	71	124	40,42	38,48	1,94	<0,0001
DM 3810*SPRO 12-17	QTL +	QTL -	148	162	38,45	36,57	1,87	<0,0001
DM 3810*SPRO 12-61	QTL +	QTL -	37	19	39,33	37,06	2,27	<0,0001

Tabla 8. Prueba de T para muestras independiente entre las líneas QTL +: Grupo 1 y QTL -: Grupo 2. n: número de líneas de cada grupo (1) o (2) de cada población. Media del porcentaje de proteína sobre base seca de cada población para el subgrupo (1) y (2). Media (1)-Media (2): media del grupo (1) menos la media del grupo (2) para cada población. p-valor: valor de probabilidades asociado a la prueba T de cada población. No incluye a los progenitores.

En los 4 casos analizados, las líneas con QTL + presentaron mayores niveles promedio de proteína, y los mínimos fueron entre un punto y dos mayores a las líneas con el QTL -. Solo en la población DM 3810*SPRO 12-61, las líneas con QTL+ y -, presentaron valores máximos iguales, mientras que el valor mínimo de QTL + se mantuvo superior.

No se detecta asociación aparente entre los genotipos alta proteína y los niveles de proteína cuando se realizan comparaciones a través de las hembras utilizadas (Tabla 9). Pero si se observa una diferencia significativa entre los niveles de proteína cuando analizamos las poblaciones a partir de las hembras utilizadas (Tabla 9). Claramente se observa que la hembra A3302RG (selecta por su alto nivel de proteína) aporta más a este carácter, diferenciándose significativamente de las poblaciones con la hembra DM 3810 (comercial de bajo contenido en proteína).

Tabla 9: Comparación del nivel de proteína de las poblaciones para el grupo de líneas QTL +

Población	(*)	Medias	n	E.E.
A3302RG*Spro12-61	a	40,42	71	0,09
A3302RG*Spro12-17	a	40,27	290	0,04
DM3810*Spro12-61	b	39,33	37	0,12
DM3810*Spro12-17	c	38,45	148	0,06

Tabla 9: n: número de muestras analizadas, Media: media de porcentaje de proteína en grano sobre base seca de las líneas caracterizadas con el QTL + de cada población, (*) Valores seguidos de igual letra no difieren entre sí significativamente, con un valor de DMS de 0,23121 E.E.: error estándar del porcentaje de proteína

5.5. Selección fenotípica contra selección asistida por marcadores moleculares.

La selección fenotípica se llevó a cabo de la siguiente manera: a partir de los datos del porcentaje de proteína del grano obtenidos, se realizó la selección de líneas dentro de cada población. Para la selección se eligió el 20 % de las líneas con los mayores valores de porcentaje de proteína.

La selección asistida por marcadores moleculares se llevó a cabo seleccionando, dentro de cada población, las líneas con el QTL presente en homocigosis.

Se compararon ambos listados y se observó que en la selección fenotípica todas las líneas tenían presente el QTL alta proteína. Salvo un solo caso en la población A 3302RG*SPRO 12-17, al comparar la media en el carácter proteína para ambos tipos de selecciones de las 4 poblaciones esta fue la misma.

En los diagramas de cajas (Figura 5) se puede observar la separación de los grupos QTL+ contra QTL- para las cuatro poblaciones.

Figura 5: Diagrama de cajas para la dispersión de los datos por población del porcentaje de proteína para QTL + contra QTL-.

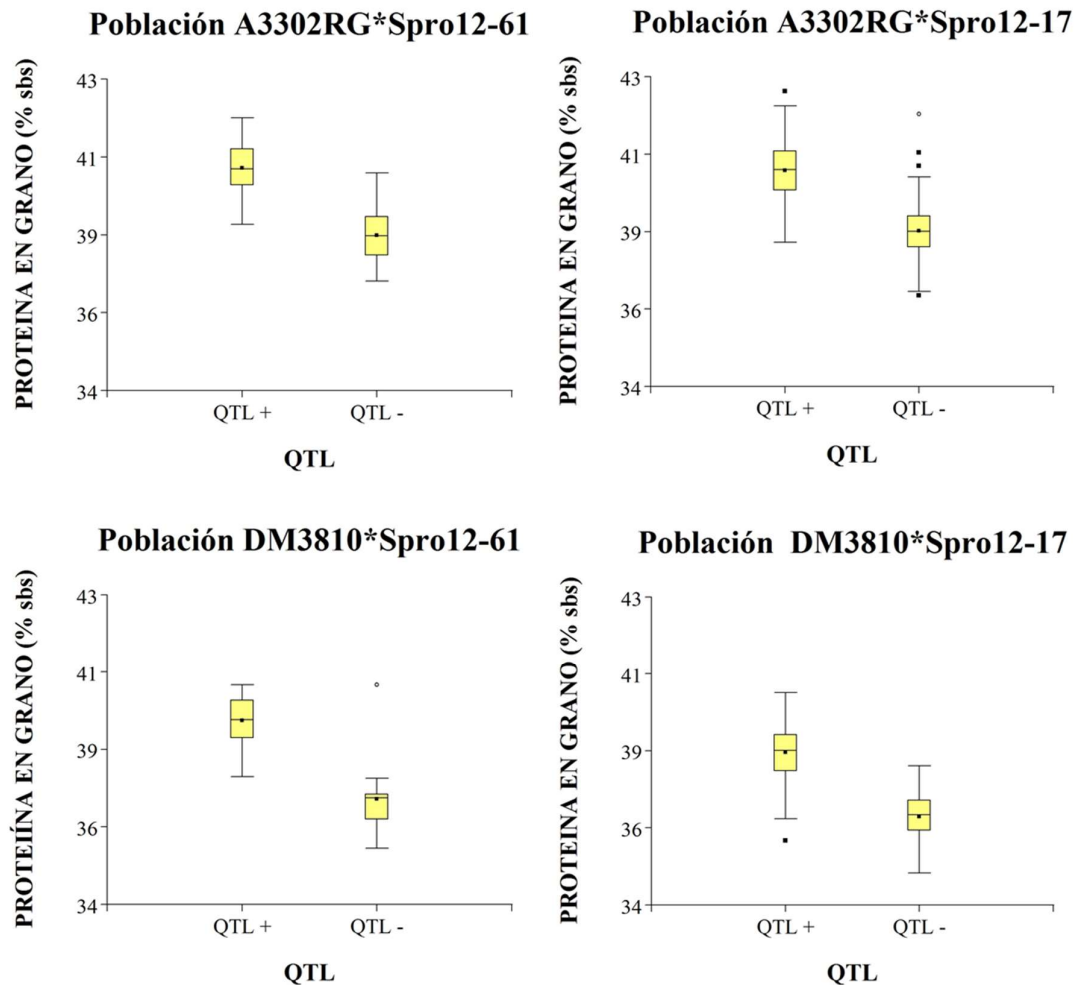


Figura 5. Análisis de dispersión de porcentaje de proteína sobre base seca de cada población segregando las líneas entre aquellas con el QTL + contra QTL -.

Se realizó el análisis de asociación entre las variables proteína y rendimiento en grano (Tabla 10, Figura 6) solo con aquellas líneas con el QTL + de cada población y no se observó una correlación importante ($r: 0,1$), pese a que la pendiente fue negativa ($-0,06$) para el análisis con la totalidad de las líneas QTL +. En general, se observa que todas las pendientes (b) son positivas para cada población, pero ésta pendiente no es significativa estadísticamente. Cuando comparamos estos datos con los obtenidos en la Tabla 6, se puede observar un aumento de la correlación positiva para el caso de las poblaciones. Mientras que la correlación cuando se

analizan la totalidad de los datos sigue siendo negativa y del mismo valor. En todos los casos, dicha correlación es no significativa.

Tabla 10: Análisis de regresión de las variables proteína y rendimiento para líneas con QTL +

Poblaciones	b	r	Significancia
A 3302RG*SPRO 12-17	0,04	0,05	0,4067
A 3302RG*SPRO 12-61	0,15	0,20	0,1165
A 3302RG*SPRO 12-61	0,1	0,14	0,0619
A 3302RG*SPRO 12-61	0,08	0,10	0,5958
TOTAL	-0,06	0,10	0,0066

Tabla 10. Análisis llevado a cabo solo con las líneas con el QTL + de cada población. b: pendiente de la regresión de los datos, tipo de correlación de las variables, r: ajuste de los datos a la regresión, TOTAL: incluye todos los datos de líneas QTL + de todas las poblaciones en estudio.

Figura 6: Gráfico de dispersión de los datos porcentaje de proteína contra rendimiento de los materiales selectos por marcadores moleculares.

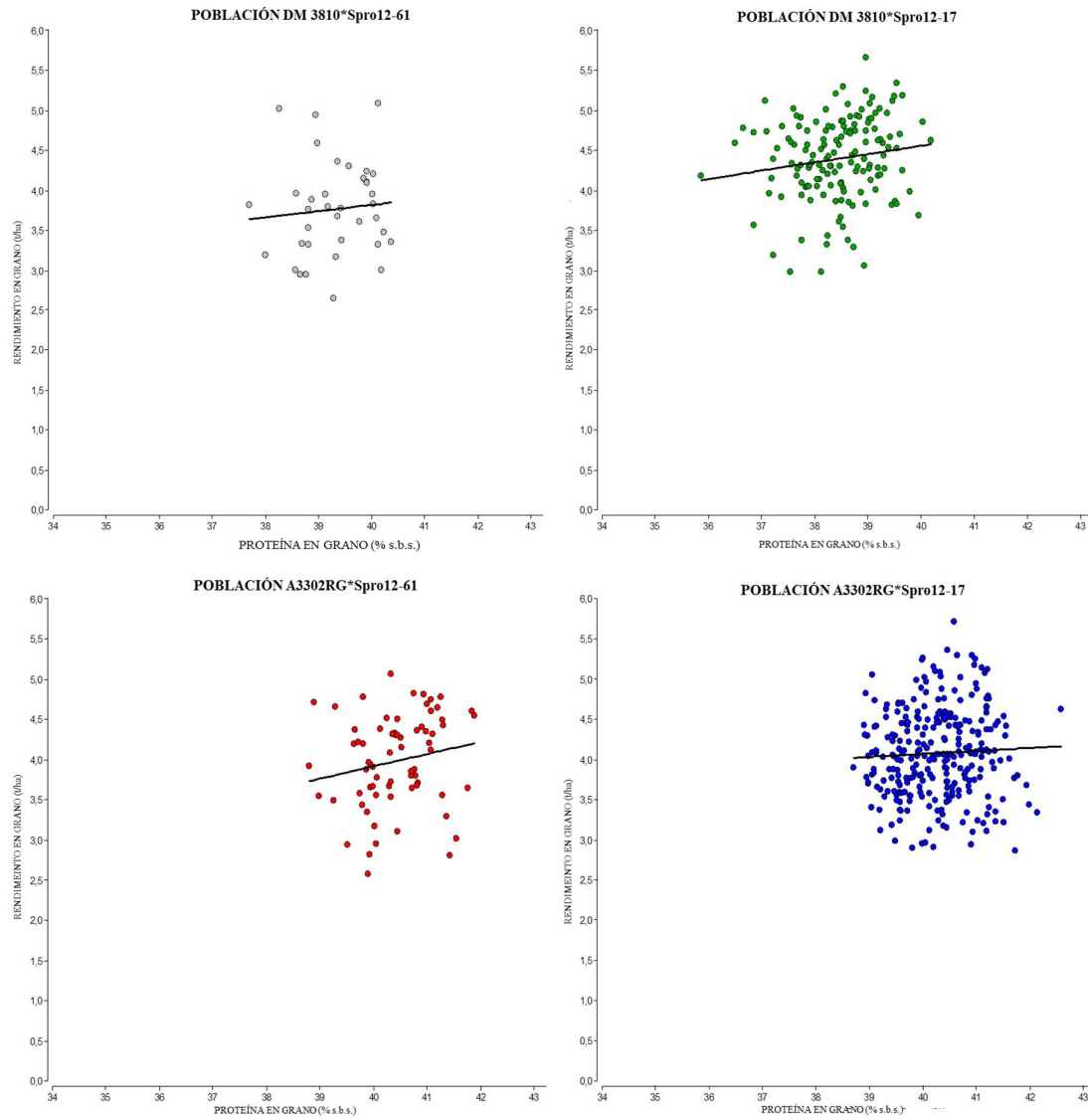


Figura 6. Análisis de dispersión rendimiento (t/ha) contra porcentaje de proteína sobre base seca de cada población para las líneas con QTL +. Sin incluir a los progenitores. En negro las líneas de tendencia de cada set de datos.

Figura 7: Gráfico de dispersión de la totalidad de los datos de porcentaje de proteína contra rendimiento de los materiales selectos por marcadores moleculares.

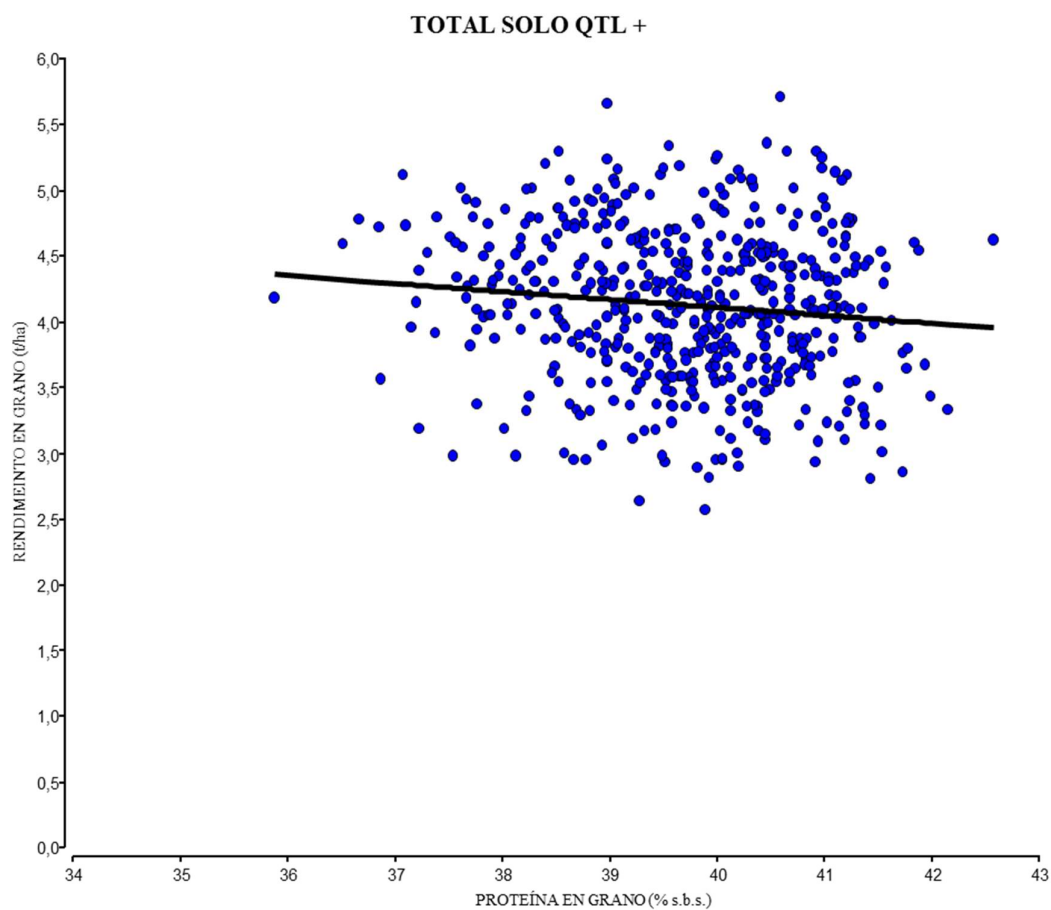


Figura 7. Análisis de dispersión rendimiento (t/ha) contra porcentaje de proteína sobre base seca de las líneas con QTL +. Sin incluir a los progenitores. En negro línea de tendencia de los datos Regresión TOTAL: $b:-0.13$ $r: 0.1$. significativa

6. DISCUSIÓN

Durante los últimos 20 años, los niveles de proteína en soja en Argentina han disminuido debido a que el porcentaje de proteína en grano no es bonificado *per se*. Esta tendencia ocasionó la pérdida de oportunidades de negocio para el país. El carácter carece de importancia en el mejoramiento tradicional debido a una supuesta correlación negativa con el rendimiento (Hanson, 1991; Chung *et al.*, 2003; Cober and Voldeng, 2000) y a la baja relevancia que se le otorga a nivel comercial. En consecuencia, al no ser un factor de demanda importante por parte de los productores, los semilleros no abordan el asunto. Desde hace un tiempo, ACSOJA (Asociación de la Cadena de la Soja), viene poniendo en consideración y buscando soluciones a esta pérdida de nivel de proteína en Argentina (participación por parte de PROSOJA de los talleres llevados a cabo en los últimos años). El advenimiento de una mayor oferta de herramientas moleculares que simplifican el proceso de selección en etapas tempranas abre la posibilidad de evaluar y, potencialmente, incluir el carácter proteína en los programas de introgresión en mejoramiento comercial.

Cuando se evaluó rendimiento contra porcentaje de proteína en grano en la totalidad de los datos generados de las cuatro poblaciones, la regresión mostró una correlación negativa significativa con un nivel muy bajo de correlación (Figura 4). Sin embargo, cuando se analizó cada población por separado, las regresiones fueron positivas y negativas según la población, pero no fueron significativas (Tabla 6). Cuando se analizaron los materiales con el QTL + por separado, las correlaciones pasaron a ser todas positivas salvo la que considera la totalidad de los datos (TOTAL), que se mantuvo negativa (Tabla 10 y Figura 7). Se deberían realizar ensayos en una mayor cantidad de ambientes para confirmar este comportamiento para este QTL. Esto concuerda con lo analizado por Brzostowski *et al.* (2017), donde encuentra una correlación negativa significativa entre el nivel de proteína en grano y el rendimiento cuando se analiza en el mismo QTL. Nuestro resultado no se correlaciona con los de Brzostowski y Diers (2017) cuando analizaron un QTL del cromosoma 20 asociado a alta proteína.

A su vez, se puede observar, en concordancia con Bosaz *et al.*, (2018), (Figura 7) que hay un límite de rendimiento para la alta proteína. Si se desea generar materiales con alta proteína y alto rinde (mayor a 4,5 t/ha), existe una barrera que podría estar asociada a la captura de nitrógeno.

Cuando se analizaron los materiales con el QTL de cada población (Tabla 7 y Tabla 10), se observó claramente un mayor nivel promedio de proteína en las dos poblaciones que tienen

como hembra a A3302RG, por lo que se observa un aporte significativo de este fondo genético al contenido en proteína. En cuanto a rendimiento, el menor valor se observó en la población DM 3810*SPRO 12-61. Si consideramos el rendimiento promedio de cada población respecto al rendimiento promedio de las líneas de la población con el QTL +, la media poblacional se mantuvo casi igual, en algún caso subió y en otros bajó (Tabla 11).

Tabla 11: Rendimiento en población total contra líneas con QTL+.

RENDIMIENTO (Promedio t/ha)		
Poblaciones	Total	QTL +
A 3302RG*SPRO 12-17	4,06	4,08 SUBE
A 3302RG*SPRO 12-61	4,08	4,02 BAJA
DM 3810*SPRO 12-17	4,45	4,48 SUBE
DM 3810*SPRO 12-61	3,79	3,70 BAJA

Tabla 11 Total: promedio de rendimiento de todas las líneas de la población. QTL+: promedio de rendimiento en t/ha de las líneas con el QTL +.

Se observa una pequeña tendencia con respecto a los genotipos alta proteína. El genotipo de mayor rendimiento (Spro 12-17) aportó a este carácter también y se lograron mejores rendimientos con respecto a la media cuando se analizaron los materiales con QTL +. El rendimiento bajó, sin embargo, para las cruzas con el genotipo de menor rendimiento (Spro 12-61). Se debería hacer una evaluación en más ambientes y más fondos genéticos para obtener mayor consistencia de datos. Con la información obtenida, si tuviera que utilizar alguno de estos materiales para un proceso de mejoramiento, utilizaría lo que fenotípicamente se encuentre en mejores condiciones de la población A 3302RG*SPRO 12-17. Esto lleva a pensar que, para optimizar procesos de mejoramiento, se pueden generar poblaciones con progenitores que tengan este QTL alta proteína (con el objetivo de aumentar el nivel) y en F2 o F4-5 seleccionar por marcadores moleculares aquellas líneas con el QTL homocigota y avanzar con poblaciones más reducidas. Esto permitiría asegurar un piso de proteína, sin perder oportunidades de mejorar el rendimiento. En la Figura 5 se puede observar que seleccionando para la presencia del QTL quedan selectas las líneas con el mayor rango de nivel de proteína de las poblaciones generadas. Solo quedan solapados algunos casos que se pueden considerar valores atípicos.

7. CONCLUSIONES

Con respecto a la primera hipótesis planteada, “la correlación entre el nivel proteína en grano y el rendimiento es dependiente de la fuente de alta proteína y del fondo genético de la variedad receptora”, en las poblaciones evaluadas, el carácter alta proteína en grano se correlacionó con los niveles propios de la variedad comercial, utilizándose dos materiales alta proteína con el QTL asociado a alta proteína. Se detectó una correlación negativa y significativa con el rendimiento a nivel de todas las líneas analizadas en su conjunto. Sin embargo, la correlación no fue significativa para cada población por sí misma. Aun cuando se analizaron solo aquellos materiales de cada población con el QTL +, se pudo observar que en todas las poblaciones la correlación es positiva (no significativa) y con mayores valores absolutos, y la TOTAL se mantuvo con el mismo valor de correlación, y siguió siendo significativa. En consecuencia, para este material de estudio se confirma la hipótesis planteada.

En cuanto a la segunda hipótesis “El QTL asociado al cromosoma 20 explica un porcentaje elevado de la variación del porcentaje de proteína y es un QTL de fácil selección”, en las cuatro poblaciones analizadas, la propuesta de selección del 20 % de líneas con contenido superior en porcentaje de proteína en grano tuvo una alta correlación con la presencia del QTL alta proteína. Cuando comparamos las medias de las líneas con el QTL + con la media de las líneas con el QTL – para el porcentaje de proteína en grano, vemos claramente que el QTL alta proteína aporta mucho al carácter en cuestión. Por lo que vemos factible poder hacer una selección por marcadores moleculares para este QTL en el cromosoma 20 en poblaciones generadas a partir de progenitores alta proteína con este QTL. Por lo tanto, se confirma la hipótesis planteada.

A partir de estas conclusiones, se dejan planteados los siguientes interrogantes para seguir analizando: ¿los materiales con el QTL presente, mantendrán su comportamiento de alto rinde en otros ambientes? ¿Este QTL mostrará el mismo comportamiento en otros fondos genéticos comerciales de Argentina? ¿Habrà algún efecto negativo pleiotrópico que no se haya visto en este análisis?

Se han seleccionado líneas con el QTL presente y de alto rendimiento para incorporar en el programa de mejoramiento de la empresa Nidera Seeds Argentina S.A.U..

8. BIBLIOGRAFÍA

- Bosaz, L., Gerde, J. A., Borrás, L., Cipriotti, P., Gallo, S., Rotundo, J. L. (2018). Caracterización regional y manejo para aumentar la concentración de proteína en soja. AACREA, 448, 48-XXX.
- Brzostowski, L. F., & Diers, B. W. (2017). Agronomic Evaluation of a High Protein Allele from PI407788A on Chromosome 15 across Two Soybean Backgrounds. *Crop Science*, 57(6), 2972-2978.
- Brzostowski, L. F., Pruski, T. I., Specht, J. E., & Diers, B. W. (2017). Impact of seed protein alleles from three soybean sources on seed composition and agronomic traits. *Theoretical and Applied Genetics*, 130(11), 2315-2326.
- Chung, J., Babka, H. L., Graef, G. L., Staswick, P. E., Lee, D. J., Cregan, P. B., ... & Specht, J. E. (2003). The seed protein, oil, and yield QTL on soybean linkage group I. *Crop science*, 43(3), 1053-1067.
- Cober, E. R., & D Voldeng, H. (2000). Developing high-protein, high-yield soybean populations and lines. *Crop Science*, 40(1), 39-42.
- Cromwell, G. L., 2012. Soybean meal - An exceptional protein source. Soybean Meal InfoCenter, Ankeny, IA
- Cuniberti, M. B., Rossi, R. L., Herrero, R., & Ferrari, B. (2004). Calidad industrial de la soja argentina. *Revista Aposgran.*, 17(87), 45-52.
- Cuniberti, M., Herrero, R., & Conde, B. (2014). Cultivares de soja superiores en proteína Interacción Genotipo x Ambiente Fertilización.
- Cuniberti, M., Herrero, R., & Masiero, B. (2011, September). Evolución del contenido de proteína y de aceite en la región sojera argentina. In *Mercosojá 2011 y 5 Congreso de la soja del Mercosur, I Foro de la Soja Asia-Mercosur* (Vol. 14).
- de Felipe, M., Gerde, J. A., & Rotundo, J. L. (2016). Soybean genetic gain in maturity groups III to V in Argentina from 1980 to 2015. *Crop Science*, 56(6), 3066-3077.
- Diers, B. W., Keim, P., Fehr, W. R., & Shoemaker, R. C. (1992). RFLP analysis of soybean seed protein and oil content. *Theoretical and Applied Genetics*, 83(5), 608-612.
- Di Rienzo, J. A., Macchiavelli, R., & Casanoves, F. (2017). Modelos lineales generalizados mixtos aplicaciones en InfoStat.
- Hanson, W. D. (1991). Seed protein content and delivery of assimilates to soybean seed embryos. *Crop science*, 31(6), 1600-1604.

- Harris, D. K. (2001). Genetic improvement of soybean seed traits and resistance to bud blight and root-knot nematodes (Doctoral dissertation, uga).
- Herrero, R., Cuniberti, M., Fuentes, F., Conde, M. Belén, Fernández, D. & Giménez, F. (2017). Calidad industrial de cultivares de soja de los grupos de madurez II-III corto al VII L-VIII, en Argentina. Campaña 16/17, Línea 9.
- Panthee, D. R., Pantalone, V. R., West, D. R., Saxton, A. M., & Sams, C. E. (2005). Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. *Crop Science*, 45(5), 2015-2022.
- Pípolo, A. E., & Mandarino, J. M. G. (2016) Os teores de proteína da soja e a qualidade para a indústria. Embrapa Soja-Artigo em periódico indexado (ALICE).
- Poeta, F., Borrás, L., & Rotundo, J. L. (2016). Variation in seed protein concentration and seed size affects soybean crop growth and development. *Crop Science*, 56(6), 3196-3208.
- Rotundo, J. L., Borrás, L., & Westgate, M. E. (2011). Linking assimilate supply and seed developmental processes that determine soybean seed composition. *European journal of agronomy*, 35(3), 184-191.
- Rotundo, J. L., Borrás, L., Westgate, M. E., & Orf, J. H. (2009). Relationship between assimilate supply per seed during seed filling and soybean seed composition. *Field crops research*, 112(1), 90-96.
- Warrington, C. V., Abdel-Haleem, H., Hyten, D. L., Cregan, P. B., Orf, J. H., Killam, A. S., & Boerma, H. R. (2015). QTL for seed protein and amino acids in the Benning× Danbaekkong soybean population. *Theoretical and Applied Genetics*, 128(5), 839-850.
- Wilson, R. F. (2008). Soybean: market driven research needs. In *Genetics and genomics of soybean* (pp. 3-15). Springer, New York, NY.
- www.meteoblue.com
- www.soybase.org
- Yates, J. L. (2006). Use of diverse germplasm to improve peanut root-knot nematode resistance and seed protein content in soybean (Doctoral dissertation, uga).