



INSTITUTO POLITÉCNICO NACIONAL

CENTRO DE BIOTECNOLOGÍA GENÓMICA



**“ANÁLISIS DE GENOMA COMPLETO Y PREDICCIÓN GENÓMICA PARA
CARACTERES COMPLEJOS EN BOVINOS CHAROLAIS”**

TESIS

QUE PARA OBTENER EL TÍTULO DE

DOCTORADO EN CIENCIAS EN BIOTECNOLOGÍA

PRESENTA

M C. FRANCISCO JOEL JAHUEY MARTÍNEZ

REYNOSA, TAMPS.

JUNIO, 2018



INSTITUTO POLITÉCNICO NACIONAL

CENTRO DE BIOTECNOLOGÍA GENÓMICA



**“ANÁLISIS DE GENOMA COMPLETO Y PREDICCIÓN GENÓMICA PARA
CARACTERES COMPLEJOS EN BOVINOS CHAROLAIS”**

TESIS

QUE PARA OBTENER EL TÍTULO DE

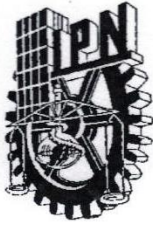
DOCTORADO EN CIENCIAS EN BIOTECNOLOGÍA

PRESENTA

M C. FRANCISCO JOEL JAHUEY MARTÍNEZ

REYNOSA, TAMPS.

JUNIO, 2018



INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

CARTA CESIÓN DE DERECHOS

En la Ciudad de Reynosa, Tamps. el día 02 del mes de Abril del año 2018, el que suscribe Francisco Joel Jahuey Martínez alumno del Programa de Doctorado en Ciencias en Biotecnología, con número de registro B140993, adscrito al Centro de Biotecnología Genómica, manifiesta que es el autor intelectual del presente trabajo de Tesis bajo la dirección del Dr. Gaspar Manuel Parra Bracamonte, Dr. Ana María Sifuentes Rincón y cede los derechos del trabajo titulado "Análisis de genoma completo y predicción genómica para caracteres complejos en bovinos Charolais.", al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o directores del trabajo. Este puede ser obtenido escribiendo a las siguientes direcciones Bld. del Maestro esq. con Elías Piña S/N Col. Narciso Mendoza, C.P. 88710 Cd. Reynosa, Tamaulipas, México Tels. 01-899 9243627, 9251656. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

Francisco Joel Jahuey Martínez



SIP-14-BIS

INSTITUTO POLITÉCNICO NACIONAL SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REVISIÓN DE TESIS

En la Ciudad de Reynosa, Tamps. siendo las 12:00 horas del día 02 del mes de Abril del 2018 se reunieron los miembros de la Comisión Revisora de la Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación de CBG para examinar la tesis titulada: Análisis de genoma completo y predicción genómica para caracteres complejos en bovinos Charolais.

Presentada por el alumno:

Jahuey	Martínez	Francisco Joel
Apellido paterno	Apellido materno	Nombre(s)
Con registro:		
B	1	4 0 9 9 3

aspirante de:

Doctorado en Ciencias en Biotecnología

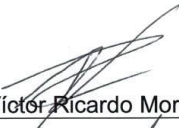
Después de intercambiar opiniones los miembros de la Comisión manifestaron **APROBAR LA TESIS**, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

LA COMISIÓN REVISORA

Directores de tesis


Dr. Gaspar Manuel Parra
Bracamonte


Dra. Ana María Sifuentes Rincón


Dr. Víctor Ricardo Moreno Medina


Dr. Carlos Armando García Pérez


Dr. Juan Manuel González Prieto

PRESIDENTE DEL COLEGIO DE PROFESORES


Dr. Mario Alberto Rodríguez Pérez



INSTITUTO POLITÉCNICO
NACIONAL
CENTRO DE BIOTECNOLOGIA
GENOMICA



AGRADECIMIENTOS

A Dios porque él es quien me permitió encontrar la gracia, sabiduría e inteligencia durante estos cuatro años de estudio. Definitivamente, sin su presencia este trabajo no hubiera sido posible. A él sea la gloria y la honra.

A mi Esposa Marisol quien llegó a mi vida en el momento preciso. Gracias por tu compañía, confianza y respeto.

Este trabajo es también dedicado a mis padres. Gracias por siempre encaminarme en el estudio.

Gracias a mi familia en Cristo, porque muchos de mis hermanos me animaron a seguir esforzándome siempre confiando en Dios.

Al personal del laboratorio de biotecnología animal, especialmente al Dr. Manuel y la Dra. Ana quienes dirigieron este trabajo. Gracias por haberme dado la oportunidad de trabajar con ustedes.

Al IPN, CONACYT y COFAA, instituciones que brindaron los recursos tanto académicos como económicos a un servidor.

Gracias a cada uno de los profesores que me asesoraron en el desarrollo de este trabajo.



INDICE DE CONTENIDO

AGRADECIMIENTOS.....	i
ÍNDICE DE CUADROS	iv
ÍNDICE DE FIGURAS	v
PREFACIO.....	vi
INTRODUCCIÓN GENERAL.....	1
LITERATURA CITADA.....	4
CAPITULO I: Estudio de asociación de genoma completo de características de crecimiento en ganado Charolais.....	7
RESUMEN.....	8
ABSTRACT	9
INTRODUCCIÓN.....	10
MATERIALES Y MÉTODOS.....	12
RESULTADOS	15
DISCUSIÓN.....	27
CONCLUSIÓN	34
REFERENCIAS	35
CAPITULO II: Identificación de señales de selección en ganado bovino Charolais.....	41
RESUMEN.....	42
ABSTRACT	43
INTRODUCCIÓN.....	44
MATERIALES Y MÉTODOS.....	46
RESULTADOS	49
DISCUSIÓN.....	54
CONCLUSIÓN	60
LITERATURA CITADA.....	61
CAPITULO III: Predicción genómica de características de peso corporal en ganado Charolais de México.....	66
RESUMEN.....	67
ABSTRACT	69
INTRODUCCIÓN.....	70
MATERIALES Y MÉTODOS.....	72



RESULTADOS Y DISCUSIÓN	77
CONCLUSIÓN	81
LITERATURA CITADA.....	83
CONCLUSIONES Y RECOMENDACIONES	86
APÉNDICES	87



ÍNDICE DE CUADROS

Cuadro 1.1. Estadísticos descriptivos para características de crecimiento en ganado Charolais.....	12
Cuadro 1.2 Parámetros y estadísticos de los polimorfismos (SNP) asociados a características de crecimiento de ganado Charolais.....	20
Cuadro 1.3 Genes y QTL previamente reportados en las regiones genómicas asociadas a características de crecimiento de ganado Charolais.....	23
Cuadro 1.4 QTL reportados en regiones genómicas asociadas a características de crecimiento de ganado Charolais.....	24
Cuadro 1.5 Genes cercanos a los polimorfismos (SNP) asociados a características de crecimiento de ganado Charolais.....	25
Cuadro 2.1 SNP estadísticamente significativos ($P < 0.00003$) en las regiones genómicas con señales de selección.....	52
Cuadro 2.2 Descripción de las regiones extendidas con señales de selección.....	52
Cuadro 2.3 Términos GO y MeSH estadísticamente significativos ($P < 0.0005$).....	53
Cuadro 3.1 Descripción estadística de los fenotipos ajustados y crudos de pesos corporales de ganado Charolais (kg).....	73
Cuadro 3.2 Estimaciones de los componentes de la varianza ¹ y heredabilidad (h^2) de pesos corporales (kg).....	74
Cuadro 3.3 Exactitud de predicción de varios métodos de predicción genómica ¹ vs BLUP.....	78
Cuadro 3.4 Estimaciones de los coeficientes de regresión y error cuadrático medio de BLUP y métodos de predicción genómica ¹	78



ÍNDICE DE FIGURAS

Figura 1.1. Gráfico de cuantiles de los resultados de los análisis de asociación de genoma completo para 5 características de crecimiento.....	19
Figura 1.2 Gráfico de Manhattan del $-\log_{10}$ de los valores de P del análisis de asociación de genoma completo para características de crecimiento de ganado Charolais mexicano.....	22
Figura 2.1 Gráficos de Manhattan de los valores integrated haplotype score (iHS) transformados en valores piHS.....	50
Figura 2.2 Señales de selección identificadas en ganado bovino Charolais.....	51
Figura 3.1 Decaimiento del DL (r^2) sobre varias distancias genómicas.....	77



PREFACIO

En ganado bovino, las características de importancia productiva son caracteres complejos, es decir, son influenciados por factores genéticos y ambientales, por lo que no siguen un patrón de herencia simple. En los sistemas de cría tradicional, el mejoramiento genético de estos caracteres se realiza seleccionando individuos en base a su mérito genético. Sin embargo, las nuevas tecnologías de genotipificación que utilizan miles de polimorfismos de un solo nucleótido (SNP) han traído nuevas oportunidades para el estudio y evaluación genética de animales domésticos a través del método denominado “genoma completo”. Este trabajo comprende el análisis genómico (mapeo genético y predicción) de características de importancia productiva en ganado bovino Charolais mediante el uso de un panel de 77 mil SNP. En el **capítulo I** se describe el mapeo genético de características de crecimiento (pesos corporales) mediante análisis de asociación de genoma completo (GWAS, por sus siglas en inglés). En este primer estudio se lograron identificar 18 marcadores genéticos asociados a las características estudiadas. Algunos SNP identificados mediante GWAS se localizaron cerca de genes funcionalmente relacionados con el crecimiento, entre ellos los genes TRAF6 (Factor 6 receptor asociado a TNF); CDH11 (Caderina 11, tipo 2, Obcaderina (osteoblasto)); KLF7 (Factor 7 tipo Kruppel (ubicuo)); MIR181A-1 (microRNA mir-181a-1); y adicionalmente PRCP (prolilcarboxipeptidasa [angiotensinasa C]), localizados en los cromosomas 15; 18; 2; 16; y 29, respectivamente. En el **capítulo II** se muestran los resultados de la identificación de señales de selección (GWSS, por sus siglas en inglés) mediante análisis de haplotipos. En este estudio se identificaron 10 SNP localizados en los cromosomas 1, 4, 6, 7 y 14. Particularmente, las señales de selección identificadas en los cromosomas 6 y 14 corresponden a regiones comúnmente reportadas en la literatura conteniendo *locus* (NCAPG-LCORL y PLAG1) asociados a una amplia variedad de características de importancia productiva, entre ellas el crecimiento. Tres nuevas regiones bajo selección se reportan en este estudio para ganado bovino Charolais, las cuales se localizan en 131.4-133.5 Mb, 51.8-60.3 Mb y 28.2-31.1 Mb de los cromosomas 1, 4 y 7, respectivamente. En estas regiones destacaron los genes CAV1 (caveolina 1), CAV2 (caveolina 1), MDFIC (proteína que contiene el dominio inhibidor MyoD), IFDR1 (regulador del desarrollo 1 relacionado a interferón) y ZNF608 (proteína dedo de cinc 608). Finalmente, en el **capítulo III** se describen los resultados del análisis de predicción de



características de crecimiento mediante los métodos GBLUP, BayesC y SSBR. Este último estudio mostró que la exactitud de predicción de los métodos genómicos no fue superior a la obtenida mediante BLUP (método basado en el pedigrí). En este estudio diferentes factores como la estructura de los datos, el tamaño de la población y la calidad de las variables de respuesta pudieron haber influido en el desempeño de los marcadores genéticos para predecir las características fenotípicas.



INTRODUCCIÓN GENERAL

Durante décadas se ha tratado de entender los mecanismos genéticos que influyen la variación de características económicamente importantes, tanto en bovinos como en otras especies de importancia doméstica, con el objetivo principal de acelerar el mejoramiento genético. Específicamente en ganado bovino para carne, los esfuerzos de investigación han sido enfocados hacia el estudio de características que son de especial interés para este sector, tales como las características productivas (pesos corporales, ganancias diarias de peso, eficiencia alimenticia), reproductivas (tasa de preñez,) y de calidad de la carne y la canal. Sin embargo, esta tarea ha sido difícil debido a que la mayoría de las características productivas son rasgos complejos, es decir, que dependen de la acción de muchos genes y efectos ambientales, y no pueden ser explicadas por herencia simple.

Recientemente, con la disponibilidad de tecnologías de genotipificación, que brindan la oportunidad de coleccionar información genómica de manera masiva, ha sido posible realizar estudios de “genoma completo” analizando con mayor detalle la arquitectura genética de características de interés. Estas tecnologías han permitido a los investigadores evaluar miles de marcadores moleculares, del tipo polimorfismos de un solo nucleótido (SNP) y de esta manera escanear el genoma de especies como el bovino. Entre las aplicaciones más importantes de las tecnologías de genotipificación masiva se encuentran los estudios de asociación (GWAS) y predicción genómica (PG). Ambas aplicaciones requieren del análisis de información fenotípica, genealógica y de marcadores genéticos para estimar el efecto de loci sobre una característica de interés. El fundamento en estos estudios radica en que el efecto de las mutaciones causales puede ser detectado indirectamente a través de marcadores genéticos en desequilibrio de ligamiento (DL). De esta manera, mediante GWAS se han tratado de identificar regiones genómicas y genes que influyen características de interés para posteriormente realizar la búsqueda de mutaciones causales en dichas regiones, mientras que en los estudios de PG se ha utilizado el efecto de los marcadores genéticos para realizar la predicción de valores genómicos (GEBV).

Por un lado, mediante estudios de GWAS se ha comenzado a descubrir los mecanismos genéticos de características complejas a través del descubrimiento de regiones genómicas,



genes y mutaciones implicadas en la variación fenotípica. Algunos de los avances y descubrimientos a partir de estudios de mapeo genético en especies de importancia doméstica han sido reportados por Fan *et al.* (2010), Zhang *et al.* (2012) y recientemente por Sharma *et al.* (2015). Específicamente, el efecto que tienen algunos genes sobre características de importancia productiva en ganado bovino se encuentran resumidos en Casas y Kehrli (2016). Entre los *loci* más reportados en la literatura para características de importancia en ganado de carne se encuentran los genes NCAPG-LCORL y PLAG1 los tanto bovinos, equinos, porcinos, caninos e incluso en humanos. Una extensa descripción de estos genes fue reportada por Takasuga (2015).

La literatura muestra muchos casos exitosos de mapeo genético en donde se han identificado regiones genómicas (QTL) con un efecto importante sobre diversas características. Por ejemplo, se han identificado QTL relacionados con la variación de características como el peso al nacimiento (Snelling *et al.* 2010), el peso de la canal (Lee *et al.* 2013), la estatura (Pryce *et al.* 2011) y la facilidad al parto (Pausch *et al.* 2011) en diversas poblaciones y razas de ganado bovino para carne. Incluso, estudios realizados en otras especies también han reportado la asociación de dichos QTL hacia características similares (Al-Mamum *et al.* 2015). Los GWAS también han sido de gran utilidad para identificar las mutaciones causales de enfermedades genéticas. Por ejemplo, en ganado Fleckvieh se logró identificar la mutación causal del síndrome tipo deficiencia de zinc el cual provoca una deficiencia en el sistema inmune, retraso en el crecimiento y alteraciones en la piel de los animales, por lo que después de este hallazgo, los productores ahora tienen una herramienta que les ayuda a tomar decisiones con respecto al manejo y la reproducción del ganado (Jung *et al.* 2014).

La PG es una metodología que consiste en aprovechar el desequilibrio de ligamiento entre miles de SNP y las mutaciones que influyen el fenotipo para cuantificar el efecto de todos los QTL en el genoma y predecir el GEBV de individuos en base al conocimiento de su genotipo (Meuwissen *et al.* 2001). Técnicamente, la PG inicia con la genotipificación de individuos que poseen registros fenotípicos o EBV para conformar poblaciones denominadas de “referencia”. Estas poblaciones son utilizadas para calcular mediante análisis de asociación el efecto de todos los loci (SNP) sobre la característica de interés y



para desarrollar la ecuación predictora de GEBV. Después, dicha ecuación es utilizada para estimar GEBV de los candidatos de selección. Mediante PG es posible estimar GEBV con mayor exactitud y seleccionar individuos a edades más tempranas, lo cual, teóricamente permite acelerar y precisar el mejoramiento genético de características de interés. En ganado bovino lechero, este método ha sido implementado exitosamente (García-Ruiz *et al.* 2016), sin embargo, su implementación en la mayoría de las razas bovinas para carne aún está en proceso de evaluación.

Se han reportado una gran cantidad de estudios de PG para una amplia variedad de características y razas bovinas (p.e. Saatchi *et al.* 2012, 2013; Gunia *et al.* 2014; Neves *et al.* 2014). En estos estudios se ha tratado de identificar la mejor estrategia para implementar la PG en ganado de carne a través de la evaluación de diversos factores que influyen en la exactitud de predicción de GEBV. Entre estos factores se encuentra la evaluación del tamaño y la estructura de la población de referencia, las relaciones genéticas entre los candidatos de selección y la de referencia, la densidad del panel de SNP, el método estadístico, así como la heredabilidad y arquitectura genética de la característica de interés.

La principal limitante para la aplicación de PG en ganado de carne es que las poblaciones de referencia aún no contienen una cantidad considerable de datos para estimar correctamente el efecto de los marcadores genéticos lo cual conlleva a producir estimaciones sesgadas de los valores genómicos. Esta carencia de datos, principalmente de sementales bien representados en la población, ha llevado a la necesidad de aplicar otras estrategias para desarrollar poblaciones de referencia ya sea genotipificando hembras, juntando poblaciones entre varios países o juntar eventualmente poblaciones de distintas razas. Por ejemplo, en diversos estudios se ha observado la colaboración de investigadores y organizaciones de diversos países (p. e. Estados Unidos, Australia y Canadá) para compartir datos, crear grandes poblaciones de referencia y realizar estudios encaminados al desarrollo de ecuaciones de predicción que puedan ser aplicadas a poblaciones uni y/o multirraciales (Miller 2010, Pollak *et al.* 2012). Sin embargo, un factor importante que debe considerarse es que los candidatos de selección de los cuales se desea conocer su GEBV deben estar genéticamente relacionados con la población de referencia. Por este motivo, los



expertos han sugerido que las poblaciones de referencia deben desarrollarse preferentemente en cada país.

A pesar de las dificultades técnicas, los estudios de GWAS y predicción genómica pueden ser capaces de detectar los QTL y sus efectos sobre características de interés siempre y cuando el estudio genómico esté diseñado adecuadamente, es decir, tomando en cuenta el tamaño y la diversidad de la población de estudio, la cantidad y calidad de los registros fenotípicos y genealógicos, la densidad de marcadores genéticos, entre otros aspectos. Por lo tanto, estas dos estrategias genómicas ofrecen nuevas oportunidades para acelerar el estudio de características de interés e implementar nuevos esquemas de selección animal.

LITERATURA CITADA

Al-Mamun H.A., Kwan P., Clark S.A., Ferdosi M.H., Tellam R., Gondro C. 2015. Genome-wide association study of body weight in Australian Merino sheep reveals an orthologous region on OAR6 to human and bovine genomic regions affecting height and weight. *Genet Sel Evol.* 47:66.

Casas E., M.E. Kehrli. 2016. A Review of Selected Genes with Known Effects on Performance and Health of Cattle. *Front Vet Sci.* 3:113.

Fan B., Du Z., Gorbach D. M., Rothschild M. F. 2010. Development and Application of High-density SNP Arrays in Genomic Studies of Domestic Animals. *Asian-Australasian Association of Animal Societies* 23 (7): 833-847.

García-Ruiz A., Cole J.B., VanRaden P.M., Wiggans G.R., Ruiz-López F.J., Van Tassell C.P. 2016. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proc. Natl. Acad. Sci. U S A.* 113(28):E3995-4004.

Gunia M., Saintilan R., Venot E., Hozé E., Fouilloux M.N., Phocas F. 2014. Genomic prediction in French Charolais beef cattle using high-density single nucleotide polymorphism markers. *Journal of Animal Science* 92: 3258-3269.



- Jung S., Pausch H., Langenmayer M.C., Schwarzenbacher H., Majzoub-Altweck M., Gollnick N.S., Fries R. 2014.** A nonsense mutation in PLD4 is associated with a zinc deficiency-like syndrome in Fleckvieh cattle. *BMC Genomics*. 15:623.
- Lee S.H., Choi B.H., Lim D., Gondro C., Cho Y.M., Dang C.G., Sharma A., Jang G.W., Lee K.T., Yoon D., Lee H.K., Yeon S.H., Yang B.S., Kang H.S., Hong S.K. 2013.** Genome-Wide Association Study Identifies Major Loci for Carcass Weight on BTA14 in Hanwoo (Korean Cattle). *PLoS One*. 8(10): e74677.
- Meuwissen T. H. E., Hayes B. J., Goddard M. E. 2001.** Prediction of total genetic value using genome-wide dense markers maps. *Genetics* 157 (4): 1819-1829.
- Miller S. 2010.** Genetic improvement of beef cattle through opportunities in genomics. *Revista Brasileira de Zootecnia*. 39: 247-255.
- Neves H.H.R., Carneiro R., O'Brien A.M.P., Utsunomiya Y.T., do Carmo A.S., Schenkel F.S., Sölkner J., McEwan J.C., Van Tassell C.P., Cole J.B., da Silva M.VGB., Queiroz S.A., Sonstegard T.S., Garcia J.F. 2014.** Accuracy of genomic predictions in *Bos indicus* (Nellore) cattle. *Genetics Selection Evolution* 46: 14.
- Pausch H., Flisikowski K., Jung S., Emmerling R., Edel C., Götz K., Fries R. 2011.** Genome-Wide Association Study Identifies Two Major Loci Affecting Calving Ease and Growth-Related Traits in Cattle. *Genetics*. 187(1): 289–297.
- Pollak E. J., Bennett G. L., Snelling W. M., Thallman R. M., Kuehn L. A. 2012.** Genomics and the global beef cattle industry. *Animal Production Science*. 52: 92-99.
- Pryce J.E., Hayes B.J., Bolormaa S., Goddard M.E. 2011.** Polymorphic Regions Affecting Human Height Also Control Stature in Cattle. *Genetics*. 187(3): 981–984.
- Saatchi M., Schnabel R.D., Rolf M.M., Taylor J.F., Garrick D.J. 2012.** Accuracy of direct genomic breeding values for nationally evaluated traits in US Limousin and Simmental beef cattle. *Genetics Selection Evolution* 44: 38.



Saatchi M., Ward J., Garrick D.J. 2013. Accuracies of direct genomic breeding values in Hereford beef cattle using national or international training populations. *Journal of Animal Science* 91: 1538-1551.

Sharma A, Lee JS, Dang CG, Sudrajad P, Kim HC, Yeon SH, Kang HS, Lee SH. 2015. stories and Challenges of Genome Wide Association Studies in Livestock - A Review. *Asian-Australas J Anim Sci.* 28(10):1371-1379.

Snelling W.M., Allan M.F., Keele J.W., Kuehn L.A., McDanel T., Smith T.P.L., Sonstegard T.S., Thallman R.M., Bennett G.L. 2010. Genome-wide association study of growth in crossbred beef cattle. *J. Anim. Sci.* 88, 837-848.

Takasuga A. 2016. PLAG1 and NCAPG-LCORL in livestock. *Anim Sci J.* 87:159-67.

Zhang H., Wang Z., Wang S., Li H. 2012. Progress of genome wide association study in domestic animals. *Journal of Animal Science and Biotechnology.* 3:26.



CAPITULO I: Estudio de asociación de genoma completo de características de crecimiento en ganado Charolais

Publicado en:

Jahuey-Martínez FJ, Parra-Bracamonte GM, Sifuentes-Rincón AM, Martínez-González JC, Gondro C, García-Pérez CA, López-Bustamante LA. Genomewide association analysis of growth traits in Charolais beef cattle. *Journal of Animal Science*. 2016. 94(11):4570-4582. doi: 10.2527/jas.2016-0359.



RESUMEN

Actualmente, los estudios de asociación de genoma completo (GWAS) son el método más utilizado para la identificación de regiones genómicas que influyen características de interés (QTL). Los QTL para características de crecimiento identificados en diversos estudios de GWAS en ganado bovino han sido la base para la búsqueda de genes y marcadores moleculares útiles para la selección asistida. El objetivo de este trabajo fue realizar análisis de GWAS para identificar SNP asociados a características de crecimiento de ganado Charolais y para definir los genes candidatos asociados a estas características. En este estudio, 855 animales fueron genotipificados con 76,883 mil polimorfismos de un solo nucleótido (SNP) del panel GeneSeek® Genomic Profiler Bovine HD™. Después del control de calidad, 68,337 SNP permanecieron en el análisis. Los datos fenotípicos solo incluyeron 823 registros de peso al nacimiento, 809 de peso al destete ajustado a 205 días y ganancia diaria al destete, 409 de peso al año ajustado a 365 días, ganancia diaria al año y ganancia diaria postdestete. El análisis de asociación se realizó aplicando el método de componentes principales mediante la función egscore del paquete GenABEL v1.8-0 en el entorno de R. Se identificaron 18 SNP asociados a las características de crecimiento ($P < 5 \times 10^{-5}$) localizados en las regiones de 68, 30, 72, 31, 12, 68, 95, 3, 42, 56, 24, 68, 34, 9, 80, 47 mega pares de bases de los autosomas 15, 4, 11, 18, 29, 6, 2, 1, 15, 21, 8, 7, 24, 9, 16 y 24, respectivamente. Los genes más importantes en estas regiones fueron TRAF6, CDH11, KLF7, MIR181A-1 y adicionalmente PRCP debido a que están relacionados con la sobrevivencia prenatal y postnatal, el desarrollo óseo, la adhesión celular y la regulación de la adipogénesis y el apetito. En conclusión este es el primer estudio de GWAS realizado en ganado bovino de México. Este estudio permitió detectar nuevos QTL asociados a características de crecimiento e identificar 5 genes candidatos que posiblemente están involucrados en la variación de las características analizadas. Estudios posteriores en estas regiones podrían ayudar a identificar marcadores moleculares útiles para la selección asistida y contribuir al conocimiento de las bases genéticas del crecimiento en bovinos.

Palabras clave: *Bos taurus*, locus de características cuantitativas, peso vivo, polimorfismo de una solo nucleótido



ABSTRACT

Currently, whole genome association studies (GWAS) are the most widely used method for the identification of genomic regions that influence characteristics of interest (QTL). The QTL for growth traits identified in several GWAS studies in cattle have been the basis for the search for genes and molecular markers useful for marker assisted selection. The objective of this study was to perform a genomewide association study (GWAS) for growth traits in Charolais beef cattle and to identify SNP markers and genes associated with these traits. Our study included 855 animals genotyped using 76,883 SNP from the GeneSeek Genomic Profiler Bovine HD panel. The examined phenotypic data included birth, weaning, and yearling weights as well as pre- and postweaning ADG. After quality control, 68,337 SNP and 823 animals were retained in the analysis. The association analysis was performed using the principal components method via the *egscore* function of the GenABEL version 1.8-0 package in the R environment. Eighteen SNP located in 13 BTA were associated with growth traits ($P < 5 \times 10^{-5}$). The most important genes in these genomic regions were TRAF6 (tumor necrosis factor receptor-associated factor 6), CDH11 (cadherin 11, type 2, OB-cadherin), KLF7 (Kruppel-like factor 7), MIR181A-1 (microRNA 181A-1), and PRCP (prolylcarboxypeptidase [angiotensinase C]), due to their relationships with perinatal and postnatal survival, bone growth, cell adhesion, regulation of adipogenesis, and appetite. In conclusion, this study is the first to describe a GWAS conducted in beef cattle in Mexico and represents a basis for further and future research. This study detected new QTL associated with growth traits and identified 5 positional and functional candidate genes that are potentially involved in variations of the analyzed traits. Future analyses of these regions could help to identify useful markers for marker-assisted selection and will contribute to the knowledge of the genetic basis of growth in cattle and be a foundation for genomic predictions in Mexican Charolais cattle.

Key words: *Bos taurus*, live weight, quantitative trait *loci*, single nucleotide polymorphism



INTRODUCCIÓN

El crecimiento animal es una característica de producción importante en la ganadería bovina y en México, esta característica constituye el principal objetivo de mejoramiento genético en la ganadería para carne. Los registros fenotípicos comúnmente utilizados para describir el crecimiento en bovinos incluyen el peso corporal y la ganancia diaria de peso desde el nacimiento hasta la edad adulta del animal. Estas características son excelentes indicadores para evaluar la eficiencia de las explotaciones dedicadas a la producción de carne. Por ejemplo, el peso al nacimiento es un indicador importante de la viabilidad animal, pero también es utilizado como un criterio de selección para mejorar la facilidad al parto. El peso al destete, por su parte, es un buen indicador de producción en los sistemas vaca-cría. El peso al año destaca debido a que es generalmente una característica correlacionada con el peso de la canal.

Actualmente, las tecnologías de genotipificación de alto rendimiento permiten evaluar miles de marcadores genéticos y realizar estudios de asociación, o GWAS, con el fin de identificar regiones en el genoma que influyen en características de interés productivo (QTL). En especies domésticas como el bovino, diversos estudios de GWAS han ayudado a definir QTL para características como el crecimiento, la eficiencia alimenticia y características de la calidad de la carne y la canal, principalmente (Zhang *et al.* 2012). Con respecto a las características de crecimiento, se ha demostrado que estos rasgos cuantitativos son controlados por múltiples QTL, la mayoría mostrando efectos muy pequeños (Snelling *et al.* 2010, Lu *et al.* 2013, Buzanskas *et al.* 2014). Sin embargo, una ventaja del análisis genético de características mediante GWAS es que ahora los QTL son mapeados a regiones cada vez más precisas y los genes en dichas regiones pueden ser investigados para localizar los nucleótidos que finalmente son los responsables de la variación de las características (también denominados QTN por sus siglas en inglés).

Al identificar los QTN, estos pueden ser utilizados como marcadores moleculares con fines de selección y manejo (Lu *et al.* 2013) o bien incorporados en evaluaciones genéticas y con esto ayudar a precisar la estimación de la varianza genética aditiva y los valores genéticos (Parra-Bracamonte *et al.* 2013) siempre y cuando estos marcadores expliquen una proporción significativa de la varianza fenotípica. Por lo tanto, es importante continuar con



el estudio de las características de crecimiento y descubrir nuevos QTL que puedan ayudar a explicar la variación de estas características. Además, aunque existen varios reportes de GWAS utilizando poblaciones de ganado Charolais (por ejemplo, Saatchi *et al.* 2014), hasta ahora en México no se han realizado estudios a una escala genómica utilizando poblaciones de ganado bovino locales. Por lo tanto, el objetivo de este trabajo fue realizar análisis de asociación de genoma completo para identificar SNP asociados a características de crecimiento de ganado Charolais mexicano y determinar los genes candidatos para la búsqueda de QTN.



MATERIALES Y MÉTODOS

Población y datos fenotípicos

El estudio incluyó 712 hembras y 143 machos Charolais registrados en la base de datos de la Asociación Charolais Charbray Herdbook de México. El ganado representó la progenie de 142 sementales y 593 vacas, provenía de hatos localizados en el noreste (86%) y noroeste (14%) de México y fue criado en sistemas de producción extensivos. Las características de crecimiento analizadas fueron el peso al nacimiento (PN), peso al destete ajustado a 205 días (PD_{aj205}), peso al año ajustado a 365 días (PA_{aj365}), ganancia diaria al destete ($GDP_{(N-205)}$), ganancia diaria al año ($GDP_{(N-365)}$) y ganancia diaria posdestete ($GDP_{(205-365)}$). En el cuadro 1.1 se muestran los estadísticos descriptivos para cada característica.

Cuadro 1.1. Estadísticos descriptivos para características de crecimiento en ganado Charolais.

Característica	N	Media	DS	Min	Max	CV
PN, kg	855	40.488	6.712	20	62	0.165
PD_{aj205} , kg	840	227.554	39.571	103.7	377.8	0.173
PA_{aj365} , kg	418	356.693	58.137	211.2	582.2	0.162
$GDP_{(N-205)}$, kg	840	0.912	0.186	0.291	1.603	0.204
$GDP_{(N-365)}$, kg	418	0.866	0.157	0.218	1.750	0.352
$GDP_{(205-365)}$, kg	418	0.807	0.284	0.499	1.469	0.181

CV= coeficiente de variación, DS= desviación estándar, $GDP_{(N-205)}$ = ganancia diaria al destete, $GDP_{(N-365)}$ = ganancia diaria al año, $GDP_{(205-365)}$ = ganancia diaria posdestete, kg= kilogramos, Min= valor mínimo, Max= valor máximo, N= número de registros, PA_{aj365} = peso al año ajustado a 365 días, PD_{aj205} = peso al destete ajustado a 205 días, PN= peso al nacimiento.

Genotipificación y control de calidad

Un total de 855 muestras fueron genotipificadas con el panel GeneSeek® Genomic Profiler Bovine HDTM (GGPHD 77K) el cual contiene 76,883 SNP. Previo al análisis de asociación, la calidad de los datos genotípicos fue revisada utilizando el programa SNPQC (Gondro *et al.* 2014). Los genotipos fueron considerados adecuados si presentaban un valor de GenCall (GC) mayor a 0.50, por lo que todos los SNP con valores menores fueron descartados (n=2,629). También fueron eliminados los SNP que resultaron monomórficos (n=363), que presentaron tasa de llamado menor al 90% (n=5,368), frecuencia del alelo menor <0.01 (n=1,162) y desviación al equilibrio de Hardy-Weinberg de acuerdo a la prueba exacta de Fisher y un valor de $P > 1 \times 10^{-15}$ (n=723). Además, los SNP que no se conocían sus



coordenadas en la secuencia del genoma bovino UMD v3.1 (Zimin *et al.* 2009) ($n=99$) y SNP localizados en cromosomas no autosómicos ($n=2,112$) fueron descartados. También fueron eliminadas las muestras que presentaron tasa de llamado menor al 80% ($n=14$) o niveles de heterocigosidad (HE) mayor a 3 desviaciones estándar (DS) ($n=26$) tomando en cuenta que la media y DS de la HE observada fue de 0.385 y 0.018, respectivamente. Finalmente, se calculó la correlación (Pearson) entre cada par de muestras de acuerdo a su genotipo obteniendo una r promedio de 0.57 y valores mínimos y máximos de 0.037 y 0.86, respectivamente. Se consideró un valor máximo de $r = 0.98$ con el fin de detectar muestras posiblemente duplicadas ($n=0$). Al final, solo 68,337 SNP y 823 muestras pasaron los criterios del control de calidad y permanecieron para los análisis posteriores. Todos estos procedimientos, así como los análisis posteriores fueron realizados en el entorno de R. Los códigos utilizados se encuentran en el apartado de apéndices.

Identificación de la estructura de la población y análisis de asociación

Se esperaba encontrar estructura en la población de estudio debido a que el ganado provenía de dos regiones distintas del norte de México. Lo anterior era apoyado por el hecho de que los hatos evaluados tienen objetivos de selección distintos (por ejemplo para la característica de PN) y como consecuencia también difieren en el tipo de material genético diseminado. Para investigar la estructuración de la población, primero se calculó la matriz de relaciones genómicas (MRG) utilizando la información de los genotipos de acuerdo a VanRaden (2008) y después se realizó una descomposición en valores singulares y un análisis de componentes principales (CP). El análisis preliminar indicó que el 28.6% de la varianza de los datos era explicada por los dos primeros CP. Por lo tanto, se decidió realizar un análisis de asociación de genoma completo mediante el método de componentes principales (CP) propuesto por Price *et al.* (2006). Para esto, se utilizó la función *egscore* del paquete GenABEL v1.8-0 (Aulchenko *et al.* 2007). Esta función permite ajustar los fenotipos de acuerdo a los efectos fijos y los dos CP en un modelo lineal. Para el análisis del PN, el modelo incluyó el grupo contemporáneo (GC) y el efecto lineal y cuadrático de la edad de la madre. Los GCs fueron definidos por el año (A) y época de nacimiento del animal (E), sexo (S) y hato (H). Las épocas de nacimiento fueron divididas en enero-marzo,



abril-junio, julio-septiembre y octubre-diciembre. El modelo estadístico para las otras características solo incluyó el GC. Por otro lado, los genotipos fueron ajustados obteniendo los residuales de la regresión de los genotipos en los CP. Finalmente, la asociación entre los genotipos y fenotipos fue evaluada mediante correlación y los valores de P fueron obtenidos calculando el cuadrado de esta correlación por $(N-K-1)$, donde N fue el número de individuos genotipificados y K fue el número de CP. Los SNP con valores de $P < 5 \times 10^{-5}$ fueron considerados significativos y asociados a las características de crecimiento. La proporción de la varianza fenotípica explicada por los SNP fue estimada mediante $VAR(\%) = (Chi.Idf)/N \times 100$, donde *Chi.Idf* fue el estadístico de chi cuadrada para un grado de libertad y N fue el número de individuos utilizados para al análisis de cada marcador.

Análisis de regiones genómicas con SNP significativos

Los genes más cercanos a los marcadores significativos así como aquellos localizados dentro de una distancia de 250 kb a ambos lados del SNP fueron identificados. La lista de genes fue obtenida utilizando la función *snp2gene.LD* del paquete *Postgwas* (Hiersche *et al.* 2014). La distancia entre los SNP y los genes fue calculada como la diferencia entre la posición del marcador y el inicio o final del gen de acuerdo con las coordenadas en el ensamblaje del genoma bovino UMD v3.1 (Zimin *et al.* 2009). La función de los genes fue investigada en la base de datos de Uniprot (www.uniprot.org). Las anotaciones en humanos o ratón fueron utilizadas cuando no existía información de los genes en bovinos. Los genes fueron considerados candidatos funcionales y posicionales siempre y cuando su función biológica tuviera relación con la característica de estudio y apoyado por evidencia experimental en la literatura. Finalmente, se determinó si los SNP significativos coincidían con QTL previamente asociados a características de crecimiento depositados en la base de datos QTLdb (Hu *et al.* 2013). Las posiciones de los SNP en la secuencia del genoma Btau4.6 fueron utilizadas para comparar con la base de datos QTLdb. Lo anterior debido a que muchos de los QTL previamente reportados no contaban con posiciones bien definidas en el ensamblaje del genoma bovino UMD v3.1.



RESULTADOS

Un total 68,337 SNP del panel GGPHD fueron evaluados por asociación a características de crecimiento de ganado Charolais. De acuerdo con el mapa físico de los marcadores, en promedio 2,356 SNP fueron evaluados en cada BTA, siendo los BTA 1 y 25 los que presentaron el mayor y menor número de SNP, respectivamente. La distancia promedio entre SNP continuos fue 36,298 pares de bases (pb), y la distancia mínima (75 pb) y máxima (3,085,904 pb) entre SNP fue encontrada en los BTA 2 y 6, respectivamente.

El efecto de la estructura de la población fue removido de los análisis realizados en este estudio, lo cual puede ser observado en los gráficos de cuantiles para GWAS (figura 1.1). De acuerdo con el valor de significancia establecido ($P = 5 \times 10^{-5}$), 14 SNP estuvieron asociados con las características de crecimiento (cuadro 1.2). Estos marcadores estuvieron distribuidos en los BTA 1, 2, 4, 6, 7, 8, 9, 11, 15, 16, 18, 21, 24 y 29. La figura 1.2 muestra los gráficos de Manhattan donde se grafica el $-\log_{10}$ de los valores de P para cada uno de los GWAS. En el cuadro 1.3 se muestran los genes y QTL presentes en cada una de las regiones genómicas asociadas a las características de crecimiento. Una descripción más detallada incluyendo el número de identificador y la localización exacta de cada uno los genes y QTL es proporcionada en los cuadros 1.4 y 1.5.

Peso al nacimiento (PN)

El único SNP asociado con el PN fue rs42512403, el cual se localiza en 68.0 mega pares de bases (Mb) del BTA 15. Este SNP presentó un efecto de sustitución alélica de -0.160 kg y explicó el 2.6% de la varianza fenotípica del PN. Los genes encontrados en esta región fueron C15H11orf74 (Marco de lectura abierto del cromosoma 15 humano C11orf74), RAG2 y RAG1 (Genes activadores de la recombinación 2 y 1, respectivamente) y TRAF6 (Factor 6 asociado al receptor de TNF).



Peso al destete (PD_{aj205}) y ganancia diaria al destete ($GDP_{(N-205)}$)

Dos SNP estuvieron asociados con el PD_{aj205} y la $GDP_{(N-205)}$, uno de ellos fue rs43378829. Este SNP se localiza en 30.8 Mb del BTA 4 y presentó un efecto de sustitución alélica de -0.753 kg sobre el PD_{aj205} y -0.003 kg sobre la $GDP_{(N-205)}$. Este SNP fue encontrado en el gen CDCA7L (Proteína tipo 7 asociada al ciclo de división celular). Otros genes también localizados en esta región fueron DNAH11 (Dineína axonemal de cadena pesada 11) y RAPGEF5 (Factor 5 de intercambio de nucleótidos guanina Rap). El segundo SNP significativamente asociado con ambas características fue encontrado en 72.4 Mb del BTA 11. El SNP rs109462000 presentó un efecto de sustitución alélica de 1.293 y 0.005 kg, respectivamente. El análisis de esta región en el ensamblaje del genoma bovino UMD v3.1 mostró la presencia de 26 genes abarcando desde IFT172 (Proteína de transporte intraflagelar 172) hasta DPYSL5 (Dihidropiridiminasa tipo 5).

Otros dos SNP asociados únicamente con el PD_{aj205} fueron rs109028958 y rs42339359. El primero de ellos se localiza en la región de 31.9 Mb del BTA 18 y presentó un efecto de sustitución alélica de 0.915 kg, mientras que el SNP rs42339359, localizado en 12.3 Mb del BTA 29, presentó un efecto sustitución alélica de -1.695 kg. El gen más cercano al SNP rs109028958 fue encontrado a 873.5 kb río arriba y fue el gen CDH11 (Caderina 11, tipo 2, OB-caderina (osteoblasto)). En BTA 29, el SNP rs42339359 está ubicado cerca de los genes DLG2 (Homólogo 2 de discos largos), CCDC90B (Proteína 90B que contiene el dominio espiral-espiral), ANKRD42 (Proteína 92 que contiene el dominio de repeticiones de anquirina) y PCF11 (Subunidad del factor de poliadenilación y escisión PCF11). La varianza fenotípica promedio explicada por cada marcador, tanto en el PD_{aj205} y la $GDP_{(N-205)}$, fue de 2.5%.

Peso al año (PA_{aj365}) y ganancia diaria al año ($GDP_{(N-365)}$)

Nueve SNP estuvieron asociados con el PA_{aj365} , de los cuales 7 también estuvieron asociados significativamente con la $GDP_{(N-365)}$. Tres de estos marcadores, incluyendo rs133575890, rs110875592 y rs135029160, estuvieron localizados en 68.0 Mb del BTA 6, región donde también se localiza el gen CORIN (Corin Serina Peptidasa). El efecto



promedio de sustitución alélica asociado a estos marcadores fue de 1.468 y 0.003 kg sobre el PA_{aj365} y la GDP_(N-365), respectivamente. Otros genes también localizados en esta región genómica fueron ATP10D (ATPasa, clase V, tipo 10D) y NFXL1 (Factor de transcripción nuclear, tipo 1 de unión a la caja X). El SNP rs109300035 fue otro marcador también localizado en BTA 6 pero en 68.4 Mb. Este SNP fue ubicado dentro del gen TXK (Tirosina Kinasa TXK) y presentó un efecto de sustitución alélica de -1.411 y -0.003 kg, respectivamente. Los genes NIPAL1 (Proteína 1 que contiene el dominio tipo NIPA), CNGA1 (Canal alfa 1 cerrado por nucleótidos cíclicos), TEC (Proteína tirosina cinasa Tec), al igual que NFXL1 y CORIN, también estuvieron localizados en esta región genómica del BTA6.

Los otros tres marcadores asociados significativamente con el PA_{aj365} y la GDP_(N-365) fueron rs110545556, rs109532225 y rs43538101 ubicados en 95.8, 42.9 y 24.5 Mb en los BTA 2, 15 y 8, respectivamente. El efecto promedio de sustitución alélica asociado a estos tres marcadores fue de 2.091 y 0.005 kg, respectivamente. En BTA 2, el SNP rs110545556 se localiza dentro del gen KLF7 (factor 7 tipo Kruppel (ubicuo)) y cerca de los genes MIR2355 (MicroRNA mir-2355) y CPO (Carboxipeptidasa O). En BTA 15, el SNP rs109532225 está ubicado cerca de los genes ADM (Adrenomedulina), RNF141 (Proteína 141 ring finger), LYVE1 (Receptor 1 hialuronano endotelial del vaso linfático) y MRVII (Homólogo 1 del sitio de integración del retrovirus murino). En BTA 8, el SNP rs43538101 se encuentra dentro del gen SLC24A2 (Acarreador de solutos, familia 24 (intercambiador de sodio/potasio/calcio), miembro 2).

Otros dos marcadores únicamente asociados con el PA_{aj365} fueron rs134262111 y rs109419103, los cuales están ubicados en 3.5 y 56.6 Mb del BTA 1 y 21, respectivamente. El primero de ellos, presentó un efecto de sustitución alélica de 5.767 kg, mientras que el SNP rs109419103 mostró un efecto sustitución alélica de 2.6 kg. Los genes identificados en BTA 1 fueron MIR2284I (microRNA mir-2284i) y MIR2284X (microRNA mir-2284x), mientras que en BTA 21, los genes identificados fueron GPR68 (Receptor 68 acoplado a proteínas G), CCDC88C (Proteína 88C que contiene el dominio espiral-espiral) y SMEK1 (Homólogo 1 del gen SMEK, supresor de mek1 (Dictyostelium)).



Adicionalmente, el SNP rs135567729 también estuvo asociado con la GDP_(N-365). Este marcador se encuentra en la región de 68.1 Mb del BTA 7 y presentó un efecto de sustitución alélica de 0.005 kg. Los genes presentes en esta región genómica fueron LARP1 (Miembro 1, familia dominio La de Ribonucleoproteína), CNOT8 (Complejo de transcripción CCR4-NOT, subunidad 8), GEMIN5 (Proteína 5 asociada a gem (organelo nuclear)) y MRPL22 (Proteína ribosomal mitocondrial L22). La varianza fenotípica promedio explicada por los 9 SNP antes mencionados fue de 4.6% tanto del PA_{aj365} como de la GDP_(N-365).

Ganancia diaria posdestete (GDP₍₂₀₅₋₃₆₅₎)

Finalmente, 4 marcadores estuvieron asociados con la GDP₍₂₀₅₋₃₆₅₎. En BTA 24, los SNP rs41566285 y rs41585202 presentaron un efecto de sustitución alélica de -0.008 y 0.006 kg, respectivamente. Estos 4 marcadores explicaron en promedio el 4.8% de la varianza fenotípica de la GDP₍₂₀₅₋₃₆₅₎. El SNP rs41566285 está ubicado cerca del gen RBBP8 (Proteína 8 de unión al retinoblastoma), mientras que los genes identificados cerca del SNP rs41585202 fueron KATNAL2 (Subunidad 2 tipo A catanina p60), HDHD2 (Proteína 2 que contiene el dominio tipo haloácido-deshalogenasa), IER3IP1 (Proteína 1 interactuante 3 de respuesta temprana inmediata) y PIAS2 (Proteína 2 inhibidora de STAT activado). Las otras dos asociaciones significativas fueron encontradas en 9.3 y 80.1 Mb en los BTA 9 y 16, respectivamente. En BTA 9, el SNP rs41569509 presentó un efecto de sustitución alélica de -0.007 kg, mientras que en BTA 16, el efecto de sustitución alélica del SNP rs42219323 fue de 0.016 kg. Los genes más cercanos a estos marcadores fueron COL9A1 (Proteína 1 alfa tipo IX, colageno) MIR181A-1 (microRNA mir-181a-1), respectivamente. La varianza fenotípica promedio explicada por estos 4 SNP fue de 4.8%.

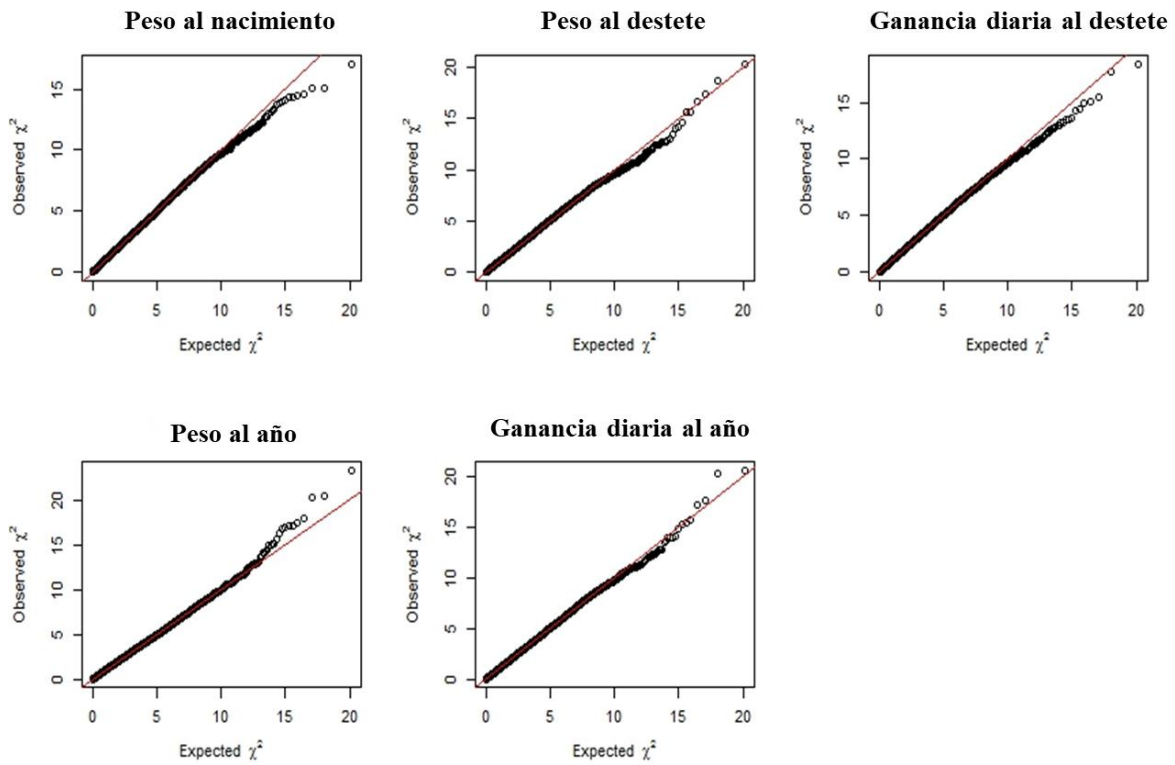


Figura 1.1. Gráfico de cuantiles de los resultados de los análisis de asociación de genoma completo para 5 características de crecimiento.



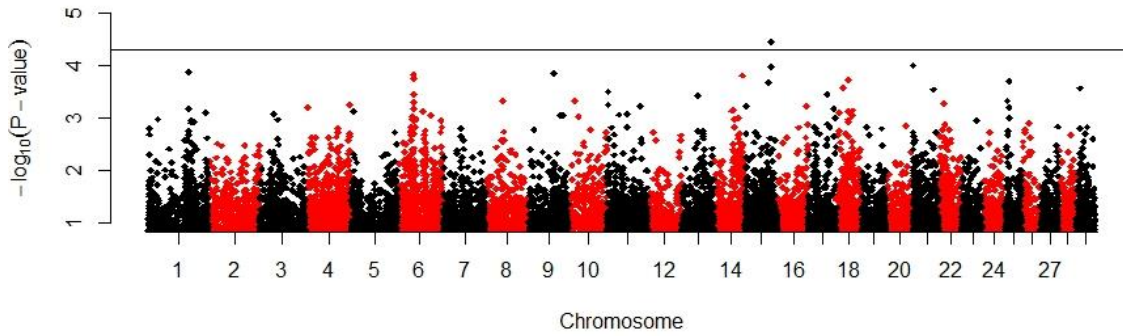
Cuadro 1.2 Parámetros y estadísticos de los polimorfismos (SNP) asociados a características de crecimiento de ganado Charolais.

Característica/SNP ID	BTA	UMD3.1 (pb)	Btau4.6 (pb)	Alelos	MAF	β	ES	% Var	P
PN									
rs42512403	15	68021790	66969167	G/A	0.44	-0.160	0.034	2.6	3.59×10^{-5}
PD _{aj205}									
rs43378829	4	30834682	31656994	C/A	0.46	-0.753	0.156	2.8	6.48×10^{-6}
rs109462000	11	72437674	74486805	A/G	0.26	1.293	0.280	2.6	1.59×10^{-5}
rs109028958	18	31925240	30754012	G/A	0.36	0.915	0.205	2.4	3.05×10^{-5}
rs42339359	29	12381867	12854197	A/G	0.16	-1.695	0.389	2.3	4.42×10^{-5}
PA _{aj365}									
rs133575890	6	68061364	68907863	G/A	0.40	1.643	0.339	5.7	1.42×10^{-6}
rs110545556	2	95861435	99648226	A/G	0.29	2.131	0.469	5.0	6.24×10^{-6}
rs110875592	6	68059441	68909786	A/C	0.43	1.445	0.319	5.0	6.57×10^{-6}
rs134262111	1	3515994	3339804	A/G	0.06	5.767	1.356	4.4	2.33×10^{-5}
rs109532225	15	42922452	41199707	G/A	0.22	2.536	0.605	4.4	3.08×10^{-5}
rs109419103	21	56601433	56469386	G/A	0.21	2.600	0.624	4.2	3.40×10^{-5}
rs43538101	8	24510080	26019429	C/A	0.37	1.608	0.386	4.2	3.48×10^{-5}
rs109300035	6	68410488	69261925	G/A	0.42	-1.411	0.341	4.1	3.80×10^{-5}
rs135029160	6	68027231	68871374	A/G	0.45	1.317	0.319	4.1	4.14×10^{-5}
GDP ^(N-205)									
rs43378829	4	30834682	31656994	C/A	0.46	-0.003	0.000	2.6	1.76×10^{-5}
rs109462000	11	72437674	74486805	A/G	0.26	0.005	0.001	2.5	2.61×10^{-5}
GDP ^(N-365)									
rs133575890	6	68061364	68907863	G/A	0.40	0.004	0.000	6.1	5.06×10^{-7}
rs110875592	6	68059441	68909786	A/C	0.43	0.004	0.000	5.5	1.90×10^{-6}
rs110545556	2	95861435	99648226	A/G	0.29	0.005	0.001	5.4	2.76×10^{-6}
rs109300035	6	68410488	69261925	G/A	0.42	-0.003	0.000	4.1	3.64×10^{-5}
rs109532225	15	42922452	41199707	G/A	0.22	0.006	0.001	4.2	4.42×10^{-5}
rs43538101	8	24510080	26019429	C/A	0.37	0.004	0.001	4.0	4.49×10^{-5}
rs135029160	6	68027231	68871374	A/G	0.45	0.003	0.000	4.1	4.49×10^{-5}
rs135567729	7	68102276	66868984	A/G	0.26	0.005	0.001	4.0	4.97×10^{-5}
GDP ⁽²⁰⁵⁻³⁶⁵⁾									
rs41566285	24	34013385	34397620	A/G	0.34	-0.008	0.001	5.2	5.85×10^{-6}
rs41569509	9	9372653	8968100	A/C	0.41	-0.007	0.001	5.4	6.69×10^{-6}
rs42219323	16	80103005	76347639	C/A	0.12	0.016	0.001	4.5	2.80×10^{-5}
rs41585202	24	47181172	48010063	A/G	0.40	0.006	0.001	4.3	3.55×10^{-5}

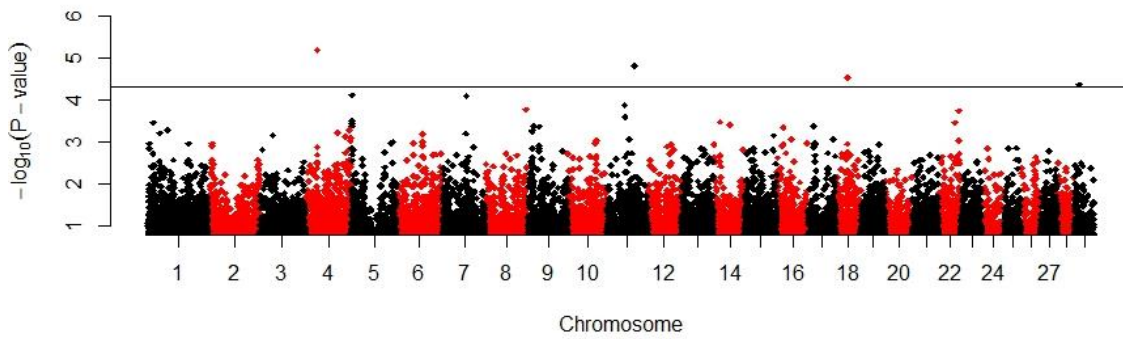
BTA= autosoma bovino, ES= error estándar, GDP^(N-205)= ganancia diaria al destete, GDP^(N-365)= ganancia diaria al año, GDP⁽²⁰⁵⁻³⁶⁵⁾= ganancia diaria posdestete, SNP ID= identificar del polimorfismo de un solo nucleótido, P= valor de probabilidad, pb= pares de bases, MAF= frecuencia del alelo menor, PA_{aj365}= peso al año ajustado a 365 días, PD_{aj205}= peso al destete ajustado a 205 días, PN= peso al nacimiento, % VAR= porcentaje de la varianza fenotípica explicada por cada SNP, β = efecto de sustitución alélica.



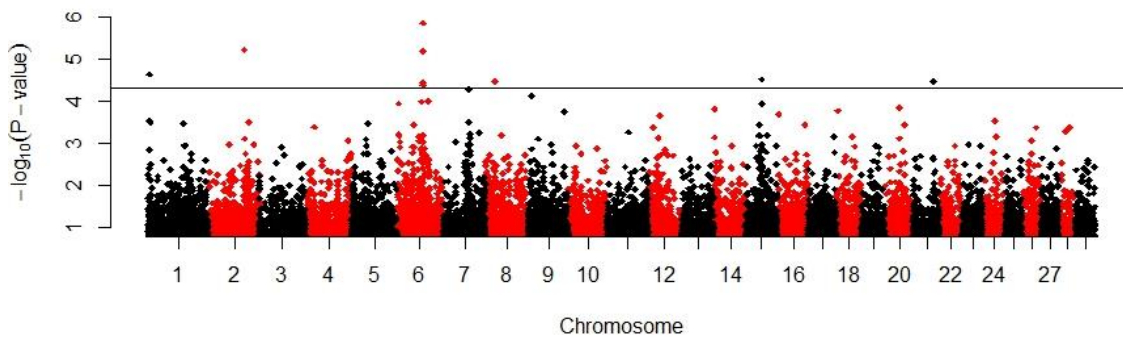
Peso al nacimiento (PN)



Peso al destete (PD_{aj205})

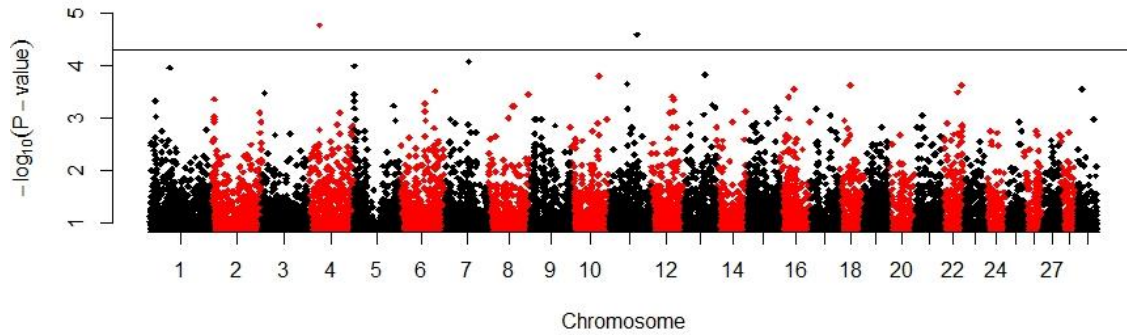


Peso al año (PD_{aj365})





Ganancia diaria al destete (GDP_(N-205))



Ganancia diaria al año (GDP_(N-365))

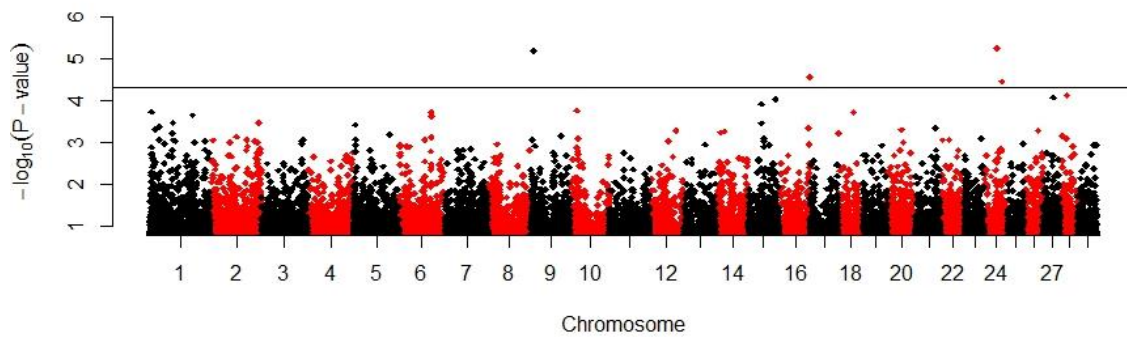


Figura 1.2 Gráfico de Manhattan del $-\log_{10}$ de los valores de P del análisis de asociación de genoma completo para características de crecimiento de ganado Charolais mexicano. La línea horizontal indica el nivel de significancia de $P < 5 \times 10^{-5}$.



Cuadro 1.3 Genes y QTL previamente reportados en las regiones genómicas asociadas a características de crecimiento de ganado Charolais.

BTA_Mb_SNP ID	Genes	QTL	Referencia del QTL
15_68.0_rs42512403	C15H11orf74, RAG2, RAG1, TRAF6	FP; GDP	Sahana <i>et al.</i> (2011); Peters <i>et al.</i> (2012)
4_30.8_rs43378829	DNAH11, CDCA7L , RAPGEF5	AML, MARM; FP; AA, AM, MARM, PD, PDLM, P365; MARM	Mizoshita <i>et al.</i> (2004); Thomassen <i>et al.</i> (2008); McClure <i>et al.</i> (2010); Yokouchi <i>et al.</i> (2009); McClure <i>et al.</i> (2010)
11_72.4_rs109462000	IFT172, KRTCAP3, NRBP1, PPM1G, ZNF513, SNX17, EIF2B4, GTF3C2, MPV17, UCN, TRIM54, DNAJC5G, SLC30A3, CAD, SLC5A6, TCF23, PRR30, PREB, ABHD1, CGREF1, KHK, EMILIN1, AGBL5, TMEM214, MAPRE3, DPYSL5	MARM	McClure <i>et al.</i> (2010)
18_31.9_rs109028958	CDH11	MF; CRA; PD	Seidensspinner <i>et al.</i> (2009); Sherman <i>et al.</i> (2009); McClure <i>et al.</i> (2010)
29_12.3_rs42339359	DLG2, CCDC90B, ANKRD42, PCF11	PC; PC, TER; CRA; AM, AML, EG12a, PC, PDLM	Kim <i>et al.</i> (2003); Casas <i>et al.</i> (2000, 2003); Nkrumah <i>et al.</i> (2007); McClure <i>et al.</i> (2010)
6_68.0_rs133575890, rs110875592, rs135029160	ATP10D, CORIN , NFXL1	-	-
6_68.4_rs109300035	CORIN, NFXL1, CNGA1, NIPAL1, TXK , TEC	-	-
2_95.8_rs110545556	CPO, KLF7 , MIR2355	-	-
1_3.5_rs134262111	MIR2284I, MIR2284X	-	-
15_42.9_rs109532225	MRVII, LYVE1, RNF141, ADM	AM, FP, PD, PDLM	McClure <i>et al.</i> (2010)
21_56.6_rs109419103	GPR68, CCDC88C, SMEK1	-	-
8_24.5_rs43538101	SLC24A2	MF; AA, AM, PN, PC	Kühn <i>et al.</i> (2003); McClure <i>et al.</i> (2010)
7_68.1_rs135567729	LARP1, CNOT8, GEMIN5, MRPL22	AM, MARM, PN	McClure <i>et al.</i> (2010)
24_34.0_rs41566285	RBBP8	FP, PD, P365	McClure <i>et al.</i> (2010)
9_9.3_rs41569509	COL9A1	MARM; CRA; FP, MARM, PA, PD	Imai <i>et al.</i> (2007); Sherman <i>et al.</i> (2009); McClure <i>et al.</i> (2010)
16_80.1_rs42219323	MIR181A-1	-	-
24_47.1_rs41585202	PIAS2, KATNAL2, HDHD2, IER3IP1	TCA; MARM	Sherman <i>et al.</i> (2009); McClure <i>et al.</i> (2010)

AA= altura al año, AM= altura a la madurez, AML= área del músculo *longissimus*, BTA= autosoma bovino, CRA= consumo residual de alimento, EG12a= espesor de grasa en la 12ª costilla, FP= facilidad al parto, GDP= ganancia diaria de peso, MARM= marmoleo, MF= muerte fetal, Mb= mega pares de bases, pb= pares de bases, PC= peso de la canal, PD= peso al destete, PDLM= peso al destete-leche materno, PN= peso al nacimiento, P365= peso al año, QTL= *locus* de característica cuantitativa, QTL ID= identificador del QTL, SNP ID= identificador del polimorfismo de un solo nucleótido, TCA= tasa de conversión alimenticia, TER= puntaje de ternera.



Cuadro 1.4 QTL reportados en regiones genómicas asociadas a características de crecimiento de ganado Charolais.

BTA_Mb_SNP ID	QTL	QTL ID	Inicio (pb)	Final (pb)	Referencia
15_68.0_rs42512403	GDP	22798	66162109	75437676	Peters <i>et al.</i> (2012)
	FP	15192-4	66823500	67865000	Sahana <i>et al.</i> (2011)
4_30.8_rs43378829	MARM	10013	8553221	50937899	Yokouchi <i>et al.</i> (2009)
	AA	10709	10994551	32226979	McClure <i>et al.</i> (2010)
	MARM	10707	10994551	32226979	McClure <i>et al.</i> (2010)
	AM	10712	10994551	32226979	McClure <i>et al.</i> (2010)
	PD	10708	10994551	35086352	McClure <i>et al.</i> (2010)
	PDLM	10711	19135469	32226979	McClure <i>et al.</i> (2010)
	P365	10710	19135469	32226979	McClure <i>et al.</i> (2010)
	FP	4655	28268762	35086352	Thomasen <i>et al.</i> (2008)
11_72.4_rs109462000	MARM	1729	28268762	38039020	Mizoshita <i>et al.</i> (2004)
	AML	1728	28268762	38039020	Mizoshita <i>et al.</i> (2004)
	MARM	10909	62118964	82382031	McClure <i>et al.</i> (2010)
18_31.9_rs109028958	MF	11362	2437515	33011652	Seidenspinner <i>et al.</i> (2009)
	CRA	5294	17918086	34957653	Sherman <i>et al.</i> (2009)
	CRA	5293	17918086	35720845	Sherman <i>et al.</i> (2009)
	PD	11063	22926481	36500992	McClure <i>et al.</i> (2010)
29_12.3_rs42339359	PC	1316	2819208	27838509	Kim <i>et al.</i> (2003)
	PC	1344	5310955	36108874	Casas <i>et al.</i> (2003)
	AM	11296	6513888	26822859	McClure <i>et al.</i> (2010)
	TER	1373	7471912	37390549	Casas <i>et al.</i> (2000)
	CRA	4403-4	7760040	28758488	Nkrumah <i>et al.</i> (2007)
	EG12a	11291	11626483	17915367	McClure <i>et al.</i> (2010)
	AM	11290	11626483	17915367	McClure <i>et al.</i> (2010)
	AML	11289	11626483	17915367	McClure <i>et al.</i> (2010)
	AM	11293	11626483	26822859	McClure <i>et al.</i> (2010)
	PC	11292	11626483	26822859	McClure <i>et al.</i> (2010)
	PDLM	11294	11626483	26822859	McClure <i>et al.</i> (2010)
15_42.9_rs109532225	AM	11001	29452266	50231208	McClure <i>et al.</i> (2010)
	FP	11000	36441038	51210066	McClure <i>et al.</i> (2010)
	PD	11002	40819540	51210066	McClure <i>et al.</i> (2010)
	PDLM	11003	40819540	51210066	McClure <i>et al.</i> (2010)
8_24.5_rs43538101	PN	10825	15196699	29543811	McClure <i>et al.</i> (2010)
	PC	10826	15196699	29543811	McClure <i>et al.</i> (2010)
	AM	10828	24157132	29543811	McClure <i>et al.</i> (2010)
	AA	10830	24157132	29543811	McClure <i>et al.</i> (2010)
	MF	11443	24157132	48094361	Kühn <i>et al.</i> (2003)
7_68.1_rs135567729	PN	10810	54839121	67327176	McClure <i>et al.</i> (2010)
	MARM	10812	64165448	69026040	McClure <i>et al.</i> (2010)
	AM	10813	64165448	69026040	McClure <i>et al.</i> (2010)
24_34.0_rs41566285	P365	11197	30004563	41590595	McClure <i>et al.</i> (2010)
	FP	11196	30004563	41590595	McClure <i>et al.</i> (2010)
	PD	11198	30004563	41590595	McClure <i>et al.</i> (2010)
9_9.3_rs41569509	CRA	5279	5803533	12440486	Sherman <i>et al.</i> (2009)
	PD	10852	8888721	18194890	McClure <i>et al.</i> (2010)
	PA	10851	8888721	18194890	McClure <i>et al.</i> (2010)
	MARM	10850	8888721	18194890	McClure <i>et al.</i> (2010)
	FP	10853	8888721	18194890	McClure <i>et al.</i> (2010)
24_47.1_rs41585202	MARM	4507	8888721	25471420	Imai <i>et al.</i> (2007)
	TCA	5308	25435547	51266517	Sherman <i>et al.</i> (2009)
	MARM	11200	41590595	50955296	McClure <i>et al.</i> (2010)

AA= altura al año, AM= altura a la madurez, AML= área del músculo *longissimus*, BTA= autosoma bovino, CRA= consume residual de alimento, EG12a= espesor de grasa en la 12ª costilla, FP= facilidad al parto, GDP= ganancia diaria de peso, MARM= marmoleo, MF= muerte fetal, Mb= mega pares de bases, pb= pares de bases, PC= peso de la canal, PD= peso al destete, PDLM= peso al destete-leche materno, PN= peso al nacimiento, P365= peso al año, QTL= *locus* de característica cuantitativa, QTL ID= identificador del QTL, SNP ID= identificador del polimorfismo de un solo nucleótido, TCA= tasa de conversión alimenticia, TER= puntaje de ternera.



Cuadro 1.5 Genes cercanos a los polimorfismos (SNP) asociados a características de crecimiento de ganado Charolais.

BTA_SNP ID	Gen	Gen ID	Distancia (kb)	Descripción
15_rs42512403	TRAF6	539124	U 232.3	Factor 6 receptor asociado a TNF
	RAG1	506302	U 191.4	Gen 1 activador de la recombinación
	RAG2	782387	U 172.2	Gen 2 activador de la recombinación
4_rs43378829	C15H11orf74	614192	U 110.6	Marco de lectura abierto del cromosoma 15 humano C11orf74
	DNAH11	497208	U 35.0	Dineína axonemal de cadena pesada I1
	CDCA7L	514631	Dentro	Proteína tipo 7 asociada al ciclo de división celular
11_rs109462000	RAPGEF5	519566	D 182.3	Factor (GEF) 5 de intercambio de nucleótidos guanine Rap
	IFT172	100848219	U 214.7	Proteína 172 de transporte intraflagelar
	KRTCAP3	508550	U 212.9	Proteína 3 asociada al queratinocito
	NRBP1	532919	U 200.3	Proteína tipo 1 de unión al receptor nuclear
	PPM1G	286880	U 164.5	Proteína fosfatasa 1G, dependiente de Mg2+/Mn2+
	ZNF513	100138621	U 160.5	Proteína 513 de anillo de zinc
	SNX17	529972	U 154.7	Nexina de clasificación 17
	EIF2B4	521926	U 149.9	Factor 2B iniciador de la traducción eucariótica, subunidad 4 delta, 67kDa
	GTF3C2	782752	U 125.6	Factor de transcripción general IIIC, polipéptido 2, beta 110kDa
	MPV17	505763	U 113.1	Proteína mitocondrial de membrana interna MpV17
18_rs109028958 29_rs42339359	UCN	518336	U 110.7	Urocortina
	TRIM54	535320	U 89.1	Proteína 54 que contiene el dominio tripartito
	DNAJC5G	616608	U 84.1	Homólogo de DnaJ (Hsp40), subfamilia C, miembro 5 gamma
	SLC30A3	512803	U 64.4	Acarreador de solutos, familia 30 (transportador de zinc), miembro 3
	CAD	504261	U 32.9	Carbamil Fosfato sintasa 2, aspartato transcarbamilasa y dihidroorotasa
	SLC5A6	516021	U 12.2	Acarreador de solutos, familia 5 (cotransportador de yoduro de sodio/multivitamina)
	TCF23	616841	D 32.6	Factor de transcripción 23
	PRR30	782932	D 53.1	Proteína 30 rica en prolina
	PREB	525256	D 55.5	Proteína de unión al elemento regulador de prolactina
	ABHD1	510774	D 59.6	Proteína 1 que contiene el dominio abhidrolasa
	CGREF1	507586	D 83.8	Proteína que contiene el dominio mano EF con el regulador del crecimiento celular
	KHK	614868	D 88.0	Cetohexocinasa (fructocinasa)
	EMILIN1	540451	D 101.6	Interfacedor 1 de microfibrillas de elastina
	AGBL5	538585	D 118.1	Proteína tipo 5 de unión a ATP/GTP
	TMEM214	514683	D 144.6	Proteína 214 transmembranal
6_rs133575890 rs110875592 rs135029160	MAPRE3	528839	D 160.4	Proteína asociada a los microtubulos, familia RP/EB, miembro 3
	DPYSL5	100126171	D 233.7	Dihidropiridiminas tipo 5
	CDH11	785475	U 873.5	Caderina 11, tipo 2, OB-caderina (osteoblasto)
	DLG2	505383	U 57.7	Homólogo 2 de discos largos
	CCDC90B	506046	D 78.3	Proteína 90B que contiene el dominio espiral-espiral
6_rs109300035	ANKRD42	522004	D 110.9	Proteína 92 que contiene el dominio de repeticiones de anquirina
	LOC510604	510604	D 173.9	Subunidad del factor de poliadenilación y escisión PCF11
	ATP10D	536495	U 138.1	ATPasa, clase V, tipo 10D
	CORIN	527107	Dentro	Corin serina peptidase
	NFXL1	784953	D 179.9	Factor de transcripción nuclear tipo 1 de unión a la caja X
	CORIN	527107	U 233.8	Corin serina peptidase
	NFXL1	784953	U 110.8	Factor de transcripción nuclear tipo 1 de unión a la caja X
	CNGA1	281700	U 56.2	Canal alfa 1 cerrado por nucleótidos cíclicos
	NIPAL1	539757	U 18.9	Proteína 1 que contiene el dominio tipo NIPA
	TXK	504782	Dentro	Tirosina cinasa TXK
2_rs110545556	TEC	504733	D 58.4	Proteína tirosina cinasa tec
	CPO	508767	U 160.5	Carboxipeptidasa O
	KLF7	537747	Dentro	Factor 7 tipo Kruppel (ubiquo)
1_rs134262111	MIR2355	100313333	D 26.7	MicroRNA mir-2355
	MIR2284I	100313114	D 64.2	microRNA mir-2284i
	MIR2284X	100526413	D 91.7	microRNA mir-2284x
15_rs109532225	MRVII	281918	U 247.8	Homólogo 1 del sitio de integración al retrovirus murino
	LYVE1	404179	U 229.9	Receptor 1 hialuronano endotelial del vaso linfático
	RNF141	539455	U 168.8	Proteína 141 con el dominio dedo anular
	ADM	280713	U 9.1	Adrenomedulina
	SBF2	510498	D 48.5	Factor 2 de unión a SET
21_rs109419103	GPR68	281799	U 10.1	Receptor 68 acoplado a proteínas G
	CCDC88C	515039	D 28.3	Proteína 88C que contiene el dominio espiral-espiral



8_rs43538101	SMEK1 SLC24A2	407220 525618	D 199.4 Dentro	Homologo 1 del gen SMEK, supresor de mek1 (Dictyostelium) Acarreador de solutos, familia 24 (intercambiador de sodio/potasio/calcio), miembro 2
7_rs135567729	LARP1 CNOT8 GEMIN5 MRPL22	505690 533443 533444 614639	U 94.7 U 44.8 U 0.5 D 3.7	Miembro 1, familia dominio La de Ribonucleoproteína Complejo de transcripción CCR4-NOT, subunidad 8 Proteína 5 asociada a gem (organelo nuclear)
24_rs41566285	RBBP8	512977	U 65.1	Proteína ribosomal mitocondrial L22
9_rs41569509	COL9A1	282195	D 144.7	Proteína 8 de union al retinoblastoma
16_rs42219323	MIR181A-1	100313401	U 417.0	Proteína 1 alfa tipo IX, colágeno
24_rs41585202	PIAS2 KATNAL2 HDHD2 IER3IP1	533403 514354 505403 768079	U 183.6 D 4.2 D 64.9 D 123.7	microRNA mir-181a-1 Proteína inhibidora de STAT activado, 2 Subunidad 2 tipo A catanina p60 Proteína 2 que contiene el dominio tipo haloácido-deshalogenasa Proteína 1 interactuante 3 de respuesta temprana inmediata

BTA= autosoma bovino, D= río abajo, Gen ID= identificador, kb= kilopares de bases, SNP ID= identificador del polimorfismo de un solo nucleótido, U= río arriba.



DISCUSIÓN

Peso al nacimiento (PN)

El SNP rs42512403 fue el único marcador asociado con el PN y estuvo localizado en 68.0 Mb del BTA15. De los 4 genes identificados en esta región, el más importante fue TRAF6, localizado a 232.3 kilo pares de bases (kb) río arriba del SNP rs42512403. Este gen pertenece a la familia de los factores asociados al receptor del factor de necrosis tumoral o TNFR, y entre sus funciones destacan la formación y diferenciación de osteoclastos funcionales, así como la proliferación, apoptosis e invasión de células del osteosarcoma (Meng *et al.* 2012). Diversos estudios han destacado su importancia para la sobrevivencia pre y postnatal en ratones. Lomaga *et al.* (1999) observaron que al inducir una mutación en el gen TRAF6, los ratones homocigotos para la mutación fueron fenotípicamente normales al nacimiento, pero murieron prematuramente y presentaron osteopetrosis, incluso los pocos animales sobrevivientes mostraron 20% a 30% menos masa corporal y menor longitud. Esta información sugiere que TRAF6 podría participar en las primeras etapas del desarrollo en bovinos y por lo tanto influir en el peso al nacimiento. Además, en esta región se han reportado QTL asociados con la facilidad al parto de ganado Holstein (Sahana *et al.* 2011) y con la ganancia diaria de peso de ganado Brahman (Peters *et al.* 2012).

Peso al destete (PD_{aj205}) y ganancia diaria al destete ($GDP_{(N-205)}$)

Dos de los 4 SNP asociados con el PD_{aj205} (rs43378829 y rs109462000) también estuvieron asociados con la $GDP_{(N-205)}$ y mostraron efectos en el mismo sentido. En BTA 4, el SNP rs43378829 fue encontrado en el gen CDCA7L. Otros genes localizados en este cromosoma fueron DNAH11 y RAPGEF5. Aunque no se encontró evidencia sobre la influencia de estos genes en procesos relacionados con el crecimiento, diversos QTL previamente asociados con características de crecimiento han sido encontrados en esta región. Por ejemplo, McClure *et al.* (2010) encontraron QTL asociados con el peso al destete, peso al año, altura, entre otras características, mientras que Thomasen *et al.* (2008) identificaron un QTL asociado con la facilidad al parto. Otros QTL también reportados en



esta región están asociados con el marmoleo y el área del músculo *longissimus* (Mizoshita *et al.* 2004, Yokouchi *et al.* 2009, McClure *et al.* 2010).

Otro de los marcadores asociados tanto con el PD_{aj205} y la GDP_(N-205) fue encontrado en una región densa en genes del BTA 11. De los 26 genes encontrados en esta región, solo TCF23 (Factor de transcripción 23) y CGREF1 (Proteína que contiene el dominio EF con el regulador de crecimiento celular) presentaron funciones interesantes. De acuerdo con la literatura consultada, TCF23 podría participar en la inhibición de la miogénesis, es decir en el proceso de determinación y formación de las células del tejido muscular. Con respecto a CGREF1, este gen se encarga de la adhesión celular de una manera dependiente de calcio y es capaz de inhibir el crecimiento en diversas líneas celulares. Por otro lado, CGREF1 ha sido identificado como uno de los genes diferencialmente expresados en muestras de carne de alta y baja calidad (Bernard *et al.* 2007). En su estudio Bernard *et al.* (2007) analizaron muestras de *Longissimus thoracis* de ganado Charolais y evaluaron los perfiles de expresión de 5,418 genes. CGREF1 apareció entre los 23 genes sobreexpresados cuando las muestras de carne presentaban mayor terneza, jugosidad y sabor. Coincidentemente, el único QTL reportado en esta región ha sido asociado con el score o puntaje de marmoleo (McClure *et al.* 2010).

Los dos SNP que únicamente estuvieron asociados con el PD_{aj205} fueron rs109028958 y rs42339359 localizados en BTA 18 y 29, respectivamente. En el único gen identificado en BTA 18, fue CDH11. Este gen codifica un tipo de glicoproteínas dependientes de calcio que pertenecen a la familia de las caderinas las cuales participan en los procesos de adhesión celular y han sido relacionadas con la osteogénesis (Marie *et al.* 2014). Estudios realizados en ratones han sugerido la importancia de las caderinas, principalmente CDH2 y CDH11, en el desarrollo óseo postnatal y en el mantenimiento de la masa ósea (Di Benedetto *et al.* 2010, Farber *et al.* 2011). En esta región, McClure *et al.* (2010) también encontraron un QTL asociado con el peso al destete, mientras que Sherman *et al.* (2009) y Seidenspinner *et al.* (2009) reportaron QTL asociados con el consumo residual de alimento y muerte fetal, respectivamente.

En BTA 29, ninguno de los genes identificados cerca del SNP rs42339359 (DLG2, CCDC90B, ANKRD42, PCF11) presentó información funcional relacionada con el



crecimiento. Sin embargo, a 438.2 kb río abajo del SNP rs42339359 se encuentra el gen PRCP (Prolinocarboxipeptidasa (angiotensinasa C)) cuya función está relacionada con la regulación del apetito. Este gen codifica una enzima carboxipeptidasa que es responsable de la degradación e inactivación de la hormona estimulante de α -melanocitos o MSH- α , la cual es un regulador importante del metabolismo de la energía (Wallingford *et al.* 2009, Diano 2011). Experimentos in vivo realizados por Wallingford *et al.* (2009) demostraron que al inhibir su actividad enzimática ocurre una disminución en la ingesta de alimento y el peso en ratones. Jeong *et al.* (2012) reportaron resultados similares y observaron una reducción de la ganancia de peso corporal y de grasa, esteatosis hepática y una mejora del metabolismo de la glucosa cuando los ratones fueron expuestos a una dieta alta en grasa. Por esta razón, PRCP ha sido propuesto como un marcador genético para la regulación de la ganancia de peso y la obesidad (Shariat-Madar *et al.* 2010). Los QTL previamente reportados en esta región han sido asociados a características como el peso de la canal (Casas *et al.* 2003, Kim *et al.* 2003, McClure *et al.* 2010), consumo residual de alimento (Nkrumah *et al.* 2007), entre otras características (McClure *et al.* 2010). Es probable que la asociación encontrada en el BTA 29 pueda deberse a la acción del gen PCRCP y su posible influencia en la ganancia de peso y el crecimiento postnatal en bovinos.

Peso al año (PA_{aj365}) y ganancia diaria al año ($GDP_{(N-365)}$)

Siete de los 9 SNP asociados con el PA_{aj365} también estuvieron asociados con la $GDP_{(N-365)}$. En BTA 6, los SNP rs133575890, rs110875592 y rs135029160 estuvieron ubicados en regiones intrónicas del gen CORIN. Este gen codifica una endopeptidasa tipo serina y participa en el procesamiento de la hormona de péptidos natriuréticos y en la regulación de la presión sanguínea. Sin embargo, el gen CORIN ha sido relacionado únicamente con el color de la capa en bovinos de acuerdo con un estudio realizado en una población de ganado Nelore-Angus (Hanna *et al.* 2014). Los otros genes identificados en esta región (ATP10D y NFXL1) tampoco mostraron alguna relación funcional con características de crecimiento. A diferencia de los SNP encontrados en el gen CORIN, el SNP rs109300035 localizado en 68.4 Mb presentó un efecto negativo sobre el PA_{aj365} y la $GDP_{(N-365)}$. Este SNP fue ubicado dentro del gen TXK y cerca de los genes NIPAL1, CNGA1, TEC,



NFXL1 y CORIN. Sin embargo, nuevamente ninguno de estos genes estuvo funcionalmente asociado a procesos relacionados con el crecimiento. De acuerdo con la búsqueda realizada en la base de datos QTLdb tampoco se encontraron reportes de QTL en las dos regiones identificadas del BTA 6.

Los otros marcadores asociados con el PA_{aj365} y la GDP_(N-365) fueron encontrados en los BTA 2, 15 y 8. En BTA 2, el SNP rs110545556 estuvo localizado en el gen KLF7 y cerca de los genes MIR2355 y CPO. KLF7 es un factor de transcripción perteneciente a la familia de los factores tipo Kruppel. Como otros miembros de esta familia, KLF7 es considerado un regulador importante del metabolismo de la energía. La sobreexpresión de este gen inhibe la diferenciación de los preadipocitos (Wu y Wang 2013) y por lo tanto, tiene un efecto negativo en la adipogénesis, es decir en el proceso del desarrollo de adipocitos y el metabolismo de las grasas. Zhang *et al.* (2013) asociaron los niveles de expresión del RNAm de KLF7 con el contenido de grasa abdominal en gallinas seleccionadas divergentemente y efectivamente observaron que la sobreexpresión de KLF7 inhibió la diferenciación de los preadipocitos pero también incrementó la proliferación de los mismos. Ma *et al.* (2011) identificaron 2 SNP localizados en el intrón 2 (T42025C y A42075G), los cuales fueron asociados a características de crecimiento de ganado Nanyang. Específicamente, el SNP A42075G estuvo asociado con la ganancia diaria de peso a los 6 meses de edad y el peso corporal a los 12 meses de edad.

En BTA 15, los genes identificados cerca del SNP rs109532225 fueron ADM, RNF141, LYVE1 y MRVII. Un estudio reciente reportado por Widmann *et al.* (2013), indicó la posible influencia de MRVII en el crecimiento divergente en bovinos. Es importante mencionar que el estudio de Widmann *et al.* (2013) estuvo basado en la construcción de redes de interacción génica a partir de resultados de GWAS en características de crecimiento. De acuerdo con sus resultados, Widmann *et al.* (2013) encontraron a MRVII como uno de los genes con mayor número de interacciones con el gen NCAPG (Complejo condensina I no SMC, subunidad G). Este último, localizado en la región de 38.57 a 38.1 Mb del BTA 6, ha sido uno de los genes más citados en literatura que posiblemente influyen el crecimiento en bovinos (por ejemplo, Snelling *et al.* 2010, Saatchi *et al.* 2014). En el modelo propuesto por Widmann *et al.* (2013), MRVII y NCAPG podrían influir el



crecimiento corporal a través de sus efectos en la vía del óxido nítrico-arginina. En BTA 8, el SNP rs43538101 fue localizado dentro del gen SLC24A2. Aunque no se encontró información relevante acerca de la función de este gen, en esta región genómica se encontraron QTL previamente asociados con el peso al nacimiento y de la canal, altura (McClure *et al.* 2010) y muerte fetal (Kühn *et al.* 2003).

Los dos SNP que únicamente estuvieron asociados con el PA_{aj365} fueron rs134262111 y rs109419103 localizados en BTA 1 y 21, respectivamente. En BTA 1, el SNP rs134262111 se localiza cerca de los microRNAs MIR2284I y MIR2284X. Los microRNAs son considerados los reguladores generales de la expresión génica ya que se sabe que estas moléculas controlan muchos procesos biológicos, incluyendo la diferenciación celular, el desarrollo animal, el metabolismo, entre otros, a través de la regulación postranscripcional de los factores de transcripción y así como de otros genes. Recientes hallazgos han indicado que los microRNAs juegan un papel importante en la regulación de la adipogénesis (Peng *et al.* 2014), pero hasta ahora solo algunos microRNAs han sido identificados como diferencialmente expresados en el tejido adiposo de ganado para carne (Jin *et al.* 2010). Por otro lado, los genes identificados en BTA 21 fueron GPR68, CCDC88C y SMEK1. Aunque se conoce la función de estos genes ninguno de ellos mostró alguna relación directa con el crecimiento. Tampoco se encontraron QTL previamente asociados a características de crecimiento en estos cromosomas.

En 68.1 Mb del BTA 7, el SNP rs135567729 estuvo asociado con la GDP_(N-365). En esta región se encuentran los genes LARP1, CNOT8, GEMIN5 y MRPL22. De estos genes, LARP1 fue el más informativo ya que entre sus funciones se encuentra la regulación de la proliferación y el crecimiento celular, pero también participa en la mitosis y en la migración y supervivencia de las células (Burrows *et al.* 2010). Este gen codifica una proteína de unión al RNA la cual promueve la traducción de ciertas clases específicas de RNAm después del complejo TORC1m. En esta región genómica, McClure *et al.*, 2010 reportaron algunos QTL asociados con el peso al nacimiento, altura y marmoleo.



Ganancia diaria postdestete ($GDP_{(205-365)}$)

Dos de los 4 marcadores asociados con la $GDP_{(205-365)}$ estuvieron ubicados en las regiones de 34.0 y 47.1 Mb del BTA 24. En 34.0 Mb, el gen RBBP8 fue encontrado muy cerca del SNP rs41566285 (65.1 kb río arriba). Aunque no está claro como RBBP8 pudiera relacionarse con el crecimiento, este gen ha sido reportado como uno de los *locus* asociados con la estatura en humanos (Hirschhorn y Lettre, 2009). Por otro lado, en la región de 47.1 Mb cerca del SNP rs41585202 fueron identificados los genes KATNAL2, HDHD2, IER3IP1 y PIAS2, sin embargo no se encontraron reportes acerca de la influencia de estos genes en el crecimiento. Los únicos QTL encontrados en estas regiones genómicas han sido asociados con la facilidad al parto, el peso al destete y el peso al año (McClure *et al.* 2010) pero también con la conversión alimenticia (Sherman *et al.* 2009).

Los otros marcadores asociados significativamente con la $GDP_{(205-365)}$ fueron encontrados en las regiones de 9.3 y 80.1 Mb en los BTA 9 y 16, respectivamente. En BTA 9, el gen COL9A1 fue identificado a 144.6 kb río abajo del SNP rs41569509. Algunos SNP (ss86352089 y ss86352090) localizados en COL9A1 han sido asociados con el área del músculo *longissimus dorsi* en cerdos (Fan *et al.* 2009). Sin embargo, no está claro como COL9A1 podría influir en el crecimiento. En esta región genómica se encontraron QTL previamente asociados con el consumo residual de alimento (Sherman *et al.* 2009), peso corporal (McClure *et al.* 2010), entre otras características (Imai *et al.* 2007).

En BTA 16, el único gen identificado cerca del SNP rs42219323 fue MIR181A-1, el cual se localiza a 417.0 kb río arriba del marcador. Como se mencionó anteriormente, los microRNAs son de gran importancia en la regulación génica de distintos procesos biológicos, incluyendo la adipogénesis. De acuerdo con Li *et al.* (2013), miR-181a participa en la adipogénesis a través de la regulación del Factor de necrosis tumoral- α (TNF- α). En breve, TNF- α es una citoquina multifuncional que inhibe la diferenciación de adipocitos a través de la regulación negativa de los principales factores de transcripción de la adipogénesis (como C/EBP α y PPAR γ). En su estudio, Li *et al.* (2013) sobreexpresaron miR-181a en adipocitos porcinos observando que la expresión de TNF- α (blanco potencial de miR-181a) era inhibida, así también ocurrió una acumulación de lípidos y un incremento en la cantidad de triglicéridos. De esta manera se demostró que el incremento en los niveles



de expresión de miR-181a favorecía la adipogénesis. Li *et al.* (2013) también observaron la sobreexpresión de algunos genes asociados con la síntesis de ácidos grasos como PDE3B (Fosfodiesterasa 3B), LPL (Lipoproteína lipasa), PPAR γ (Receptor γ activado por el proliferador), GLUT1 (Transportador de glucosa 1), GLUT4 (Transportador de glucosa 1), Adiponectina y FASN (Sintasa de ácidos grasos) así como HSL (Lipasa sensible a hormonas) y ATGL (Lipasa de triglicéridos adiposos). En esta región genómica no se encontraron QTL previamente asociados a características de crecimiento.

Frecuentemente, otros estudios de GWAS han reportado la asociación de regiones genómicas en los BTA 6 y 14 hacia características de crecimiento (Snelling *et al.* 2010, Saatchi *et al.* 2014). Aunque una de estas regiones genómicas, localizada entre 43.0 y 45.0 Mb del BTA 6, mostró una ligera señal de asociación (al menos con el PN), no fue lo suficientemente significativa para alcanzar el umbral de P establecido. La discrepancia entre nuestros resultados y otros estudios de GWAS puede ser explicado por las diferencias entre las poblaciones de ganado estudiadas (raza, tamaño y diseño de la población), los métodos y modelos estadísticos aplicados, la densidad de marcadores, entre otros factores. Sin embargo, este estudio identificó algunas regiones genómicas que anteriormente no habían sido asociadas con el crecimiento, por ejemplo en las QTL localizados en los BTA 1, 2, 6, 9 y 21. Algunos de nuestros resultados también coincidieron con regiones genómicas previamente asociadas a características de crecimiento reportadas en la literatura. Sin embargo, no podemos asegurar que se trate del mismo *locus* en cuestión debido a que los QTL aquí citados abarcan regiones de 14.8 Mb en promedio.

La varianza fenotípica explicada por los SNP aquí identificados fue muy pequeña (3.9% en promedio). Es esperado que en estudios de asociación de características de crecimiento la mayoría de los SNP identificados solo expliquen una pequeña proporción de la varianza fenotípica debido a que este tipo de características son controladas por múltiples genes, los cuales de manera individual, influyen ligeramente el fenotipo. Sin embargo, al considerar el conjunto de SNP significativamente asociados con cada característica es posible explicar una mayor proporción de la varianza fenotípica. Por ejemplo, los 4 SNP asociados con el PD_{aj205} logran explicar hasta el 10.1% de la varianza del PD_{aj205}. Valores de varianza fenotípica mayores fueron encontrados para los SNP asociados con el PA_{aj365}, GDP₍₂₀₅₋₃₆₅₎ y



GDP_(N=365). Sin embargo, es importante notar que para estas características se contó con un menor número de observaciones, lo cual pudo influir en la sobreestimación de la varianza explicada por los SNP.

CONCLUSIÓN

El análisis de asociación de genoma completo basado en componentes principales permitió identificar 18 SNP asociados a las características de crecimiento de ganado Charolais. Ocho de estos SNP fueron encontrados en regiones intrónicas de los genes CDCA7L, CORIN, TXK, KLF7 y SLC24A2. En este estudio no se tomó en cuenta la corrección por Bonferroni para explicar los múltiples análisis estadísticos debido a que cuando esta prueba fue aplicada ninguno de los marcadores resultó significativo (datos no mostrados). Aunque estas asociaciones pueden considerarse moderadamente significativas (en comparación a otros estudios de GWAS), existe evidencia de que algunos de los genes están funcionalmente relacionados con el crecimiento a través de procesos que involucran la sobrevivencia prenatal y postnatal, el desarrollo óseo, la adhesión celular y la regulación de la adipogénesis y el apetito. De esta manera, pudimos definir cinco genes candidatos con potencial de asociación hacia rasgos de crecimiento de ganado Charolais, entre los que se encuentran: TRAF6, CDH11, KLF7, MIR181A-1 y adicionalmente PRCP. Estudios posteriores en estas regiones genómicas, podrían conducir a la identificación de marcadores moleculares útiles para la selección asistida y a su vez a entender las bases genéticas de las características de crecimiento en bovinos.



REFERENCIAS

Aulchenko Y.S., Ripke S., Isaacs A., Van Duijn C.M. 2007. GenABEL: an R library for genome-wide association analysis. *Bioinformatics*. 23, 1294–1296.

Bernard C., Cassar-Malek I., Le Cunff M., Dubroeuq H., Renand G., Hocquette J.F. 2007. New indicators of beef sensory quality revealed by expression of specific genes. *J. Agric. Food Chem.* 55, 5229-5237.

Burrows C., Abd Latip N., Lam S.J., Carpenter L., Sawicka K., Tzolovsky G., Gabra H., Bushell M., Glover D.M., Willis A.E., Blagden S.P. 2010. The RNA binding protein Larp1 regulates cell division, apoptosis and cell migration. *Nucleic Acids Res.* 38, 5542-5553

Buzanskas M. E., Gross D. A., Ventura R. V., Schenkel F. S., Sargolzaei m., Meirelles S. L. C., Mokry F. B. Higa R. B., Madadu M. A, Barbosa da Silva M. V. G., Niciura S. C. M., Torres Júnior R. A. A., Alencar M. M., Regitano L. C. A., Munari D. P. 2014. Genome-Wide Association for Growth Traits in Canchim Beef Cattle. *PLoS ONE*. 9, e94802.

Casas E., Shackelford S.D., Keele J.W., Koohmaraie M., Smith T.P., Stone R.T. 2003. Detection of quantitative trait *loci* for growth and carcass composition in cattle. *J. Anim. Sci.* 81, 2976-2983.

Casas E., Shackelford S.D., Keele J.W., Stone R.T., Kappes S.M., Koohmaraie M. 2000. Quantitative trait *loci* affecting growth and carcass composition of cattle segregating alternate forms of myostatin. *J. Anim. Sci.* 78, 560-569.

Di Benedetto A., Watkins M., Grimston S., Salazar V., Donsante C., Mbalaviele G., Radice G.L., Civitelli R. 2010. N-cadherin and cadherin 11 modulate postnatal bone growth and osteoblast differentiation by distinct mechanisms. *J. Cell. Sci.* 123, 2640-2648.

Diano S. 2011. New aspects of melanocortin signaling: a role for PRCP in alpha-MSH degradation. *Front. Neuroendocrinol.* 32, 70-83.



Fan B., Onteru S., Nikkila M., Stalder K., Rothschild M. 2009. The COL9A1 gene is associated with *longissimus dorsi* muscle area in the pig. *Anim. Genet.* 40:788.

Farber C.R., Kelly S.A., Baruch E., Yu D., Hua K., Nehrenberg D.L., de Villena F.P., Buus R.J., Garland T. Jr., Pomp D. 2011. Identification of quantitative trait *loci* influencing skeletal architecture in mice: emergence of *Cdh11* as a primary candidate gene regulating femoral morphology. *J. Bone Miner. Res.* 26, 2174-2183.

Gondro C., Porto-Neto L.R., Lee S.H. 2014. SNPQC - an R pipeline for quality control of Illumina SNP genotyping array data. *Anim. Genet.* 45, 758-761.

Hanna L.L., Sanders J.O., Riley D.G., Abbey C.A., Gill C.A. 2014. Identification of a major *locus* interacting with *MC1R* and modifying black coat color in an F₂ Nellore-Angus population. *Genet. Sel. Evol.* 24, 46-54.

Hiersche M., Rühle F., Stoll M. 2013. Postgwas: Advanced GWAS Interpretation in R. *PLoS ONE.* 8, e71775.

Hirschhorn J.N., Lettre G. 2009. Progress in genome-wide association studies of human height. *Horm. Res.* 71, 5-13.

Hu Z.L., Park C.A., Wu X.L., Reecy J.M. 2013. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Res.* 41, 871–879.

Huang D.W., Sherman B.T., Lempicki R.A. 2009. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1-13.

Imai K., Matsughige T., Watanabe T., Sugimoto Y., Ihara N. 2007. Mapping of a quantitative trait *locus* for beef marbling on bovine chromosome 9 in purebred Japanese black cattle. *Anim Biotechnol.* 18, 75-80.

Jeong J.K., Szabo G., Raso G.M., Meli R., Diano S. 2012. Deletion of prolyl carboxypeptidase attenuates the metabolic effects of diet-induced obesity. *Am. J. Physiol. Endocrinol. Metab.* 302: E1502–E1510.



Kim J.J., Farnir F., Savell J., Taylor J.F. 2003. Detection of quantitative trait *loci* for growth and beef carcass fatness traits in a cross between *Bos taurus* (Angus) and *Bos indicus* (Brahman) cattle. *J. Anim. Sci.* 81, 1933-1942.

Kühn Ch., Bennewitz J., Reinsch N., Xu N., Thomsen H., Looft C., Brockmann G.A., Schwerin M., Weimann C., Hiendleder S., Erhardt G., Medjugorac I., Förster M., Brenig B., Reinhardt F., Reents R., Russ I., Averdunk G., Blümel J., Kalm E. 2003. Quantitative trait *loci* mapping of functional traits in the German Holstein cattle population. *J. Dairy Sci.* 86, 360-368.

Li H., Chen X., Guan L., Qi Q., Shu G., Jiang Q., Yuan L., Xi Q., Zhang Y. 2013. MiRNA-181a Regulates Adipogenesis by Targeting Tumor Necrosis Factor- α (TNF- α) in the Porcine Model. *PLoS ONE.* 8, e71568.

Lomaga M.A., Yeh W.C., Sarosi I., Duncan G.S., Furlonger C., Ho A., Morony S., Capparelli C., Van G., Kaufman S., Heiden A.V.D, Itie A., Wakeham A., Khoo W., Sasaki T., Cao Z., Penninger J.F., Paige C.J., Lacey D.L., Dunstan C.R., Boyle W.J., Goeddel D.V., Mak T.W. 1999. TRAF6 deficiency results in osteopetrosis and defective interleukin-1, CD40, and LPS signaling. *Genes. Dev.* 13, 1015–1024.

Lu D., Miller S., Sargolzaei M., Kelly M., Vander Voort G., Caldwell T., Wang Z., Plastow G., Moore S. 2013. Genome-wide association analyses for growth and feed efficiency traits in beef cattle. *J. Anim. Sci.* 92, 3612-3633.

Ma L., Qu Y.J., Huai Y.T., Li Z.J., Wang J., Lan X.Y., Zhang C.L., Wang J.Q., Chen H. 2011. Polymorphisms identification and associations of KLF7 gene with cattle growth traits. *Livest. Sci.* 135, 1-7.

Marie P.J., Häy E., Modrowski D., Revollo L., Mbalaviele G., Civitelli R. 2014. Cadherin-mediated cell-cell adhesion and signaling in the skeleton. *Calcif. Tissue Int.* 94, 46-54.

McClure M.C., Morsci N.S., Schnabel R.D., Kim J.W., Yao P., Rolf M.M., McKay S.D., Gregg S.J., Chapple R.H., Northcutt S.L., Taylor J.F. 2010. A genome scan for



quantitative trait *loci* influencing carcass, post-natal growth and reproductive traits in commercial Angus cattle. *Anim. Genet.* 41, 597–607.

Meng Q., Zheng M., Liu H., Song C., Zhang W., Yan J., Qin L., Liu X. 2012. TRAF6 regulates proliferation, apoptosis, and invasion of osteosarcoma cell. *Mol. Cell Biochem.* 371, 177-186.

Mizoshita K., Watanabe T., Hayashi H., Kubota C., Yamakuchi H., Todoroki J., Sugimoto Y. 2004. Quantitative trait *loci* analysis for growth and carcass traits in a half-sibfamily of purebred Japanese Black (Wagyu) cattle. *J. Anim. Sci.* 82, 3415-320.

Nkrumah J.D., Sherman E.L., Li C., Marques E., Crews D.H. Jr., Bartusiak R., Murdoch B., Wang Z., Basarab J.A., Moore S.S. 2007. Primary genome scan to identify putative quantitative trait *loci* for feedlot growth rate, feed intake, and feed efficiency of beef cattle. *J. Anim. Sci.* 85, 3170-3181.

UniProt. Disponible en: <http://www.uniprot.org/>. Consultado el 10 de Marzo de 2015.

Parra-Bracamonte G.P., López-Villalobos N., Sifuentes-Rincón A.M., Morris S., López-Bustamante L.A., Meza-García L.A. 2013. Single and composite influence of growth-related candidate gene polymorphisms on additive genetic variation of birth weight in charolais beef cattle. *Trop. Anim. Health Prod.* 46, 509-512.

Peng Y., Yu S., Li H., Xiang H., Peng J., Jiang S. 2014. MicroRNAs: emerging roles in adipogenesis and obesity. *Cell Signal.* 26, 1888-1896.

Peters S.O., Kizilkaya K., Garrick D.J., Fernando R.L., Reecy J.M., Weaber R.L., Silver G.A., Thomas M.G. 2012. Bayesian genome-wide association analysis of growth and yearling ultrasound measures of carcass traits in Brangus heifers. *J. Anim. Sci.* 90, 3398-409.

Price A.L., Patterson N.J., Plenge R.M., Weinblatt M.E. Shadick N.A., Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 38, 904-909.



- Saatchi M, Schnabel R.D., Taylor J.F., Garrick D.J. 2014.** Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. *BMC Genom.* 15, 442.
- Sahana G., Guldbbrandtsen B., Lund M.S. 2011.** Genome-wide association study for calving traits in Danish and Swedish Holstein cattle. *J. Dairy Sci.* 94, 479-86.
- Seidenspinner T., Bennewitz J., Reinhardt F., Thaller G. 2009.** Need for sharp phenotypes in QTL detection for calving traits in dairy cattle. *J. Anim. Breed. Genet.* 126, 455-462.
- Shariat-Madar B., Kolte D., Verlangieri A., Shariat-Madar Z. 2010.** Prolylcarboxypeptidase (PRCP) as a new target for obesity treatment. *Diabetes Metab. Syndr. Obes.* 3, 67-78.
- Sherman E.L., Nkrumah J.D., Li C., Bartusiak R., Murdoch B., Moore S.S. 2009.** Fine mapping quantitative trait *loci* for feed intake and feed efficiency in beef cattle. *J. Anim. Sci.* 87, 37-45.
- Snelling W.M., Allan M.F., Keele J.W., Kuehn L.A., McDanel T., Smith T.P.L., Sonstegard T.S., Thallman R.M., Bennett G.L. 2010.** Genome-wide association study of growth in crossbred beef cattle. *J. Anim. Sci.* 88, 837-848.
- Thomasen J.R., Guldbbrandtsen B., Sorensen P., Thomsen B., Lund M.S. 2008.** Quantitative trait *loci* affecting calving traits in Danish Holstein cattle. *J. Dairy Sci.* 91, 2098-2105.
- VanRaden PM. 2008.** Efficient methods to compute genomic predictions. *J Dairy Sci.* 91, 4414-4423.
- Wallingford N., Perroud B., Gao Q., Coppola A., Gyengesi E., Liu Z.W., Gao X.B., Diamant A., Haus K.A., Shariat-Madar Z., Mahdi F., Wardlaw S.L., Schmaier A.H., Warden C.H., Diano S. 2009.** Prolylcarboxypeptidase regulates food intake by inactivating alpha-MSH in rodents. *J. Clin. Invest.* 119, 2291–2303.



Widmann P., Reverter A., Fortes M.R., Weikard R., Suhre K., Hammon H., Albrecht E., Kuehn C. 2013. A systems biology approach using metabolomic data reveals genes and pathways interacting to modulate divergent growth in cattle. *BMC Genomics*.14, 798.

Wu Z., Wang S. 2013. Role of kruppel-like transcription factors in adipogenesis. *Dev. Biol.* 373, 235-243.

Yokouchi K., Mizoguchi Y., Watanabe T., Iwamoto E., Sugimoto Y., Takasuga A. 2009. Identification of a 3.7-Mb region for a marbling QTL on bovine chromosome 4 by identical-by-descent and association analysis. *Anim. Genet.* 40, 945-951.

Zhang H., Wang Z., Wang S., Li H. 2012. Progress of genome wide association study in domestic animals. *J. Anim. Sci. Biotech.* 3, 26.

Zhang Z., Wang H., Sun Y., Li H., Wang N. 2013. Klf7 modulates the differentiation and proliferation of chicken preadipocyte. *Acta Biochim. Biophys. Sin.* 45: 280–288.

Zimin A.V., Delcher A.L., Florea L., Kelley D.R., Schatz M.C., Puiu D., Hanrahan F., Pertea G., Van Tassel C.V., Sonstegard T.S., Marçais G., Roberts M., Subramanian P., Yprke J.A., Salzberg S.L. 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* 10, R42.



CAPITULO II: Identificación de señales de selección en ganado bovino Charolais



RESUMEN

En México, el ganado Charolais es una raza bovina importante para la producción de carne y es seleccionada principalmente para características de pesos corporales. Un estudio previo utilizando poblaciones de ganado Charolais de México ya ha reportado *locus* de características cuantitativas (QTL) asociados a características de crecimiento. Otra estrategia para el mapeo de regiones genómicas importantes es mediante la identificación de señales de selección las cuales son *loci* que muestran patrones de desequilibrio de ligamiento (DL) y homocigosidad extendida. Mediante esta estrategia, estudios realizados en varias razas y poblaciones de ganado bovino han logrado identificar 283 regiones genómicas importantes las cuales contienen QTL para diferentes características productivas. Por lo tanto, el objetivo de este estudio fue identificar señales de selección en ganado Charolais genotipificado GeneSeek Genomic Profiler Bovine HD con el panel de 77 K polimorfismos de un solo nucleótido (SNP). Un total de 71,871 SNP y 819 muestras que pasaron el control de calidad fueron utilizados para el análisis. La identificación de señales de selección se realizó mediante el método integrated haplotype score (iHS) implementado en el paquete rehh de R. Diez SNP localizados en los BTA 1, 4, 6, 7 y 4 fueron estadísticamente significativos ($P < 0.00003$). Las regiones candidatas fueron definidas mediante la identificación de marcadores en DL con SNP significativos las cuales fueron 132.4-132.5 Mb de BTA 1; 52.8-53.4, 55.3-57.2 y 57.6-59.3 Mb de BTA 4; 38.9-40.6 Mb de BTA 6; 29.2-30.1 Mb de BTA 7; y 27.6-29.5 Mb de BTA 14. Estas regiones genómicas coinciden con señales de selección previamente reportadas en otras razas de ganado bovino. Interesantemente, los genes y QTL localizados en estas regiones estuvieron asociados al crecimiento y otras características de importancia productiva. Las regiones genómicas identificadas en este estudio indican que la selección de características productivas en la población de ganado Charolais está conservando haplotipos en varios cromosomas reflejando el origen poligénico de estas características. Estas regiones genómicas pudieran servir como base para estudios de asociación de haplotipos y para la identificación de genes causales relacionados a características de crecimiento.

Palabras clave: Charolais, Integrated haplotype score, Señales de selección, Polimorfismo de un solo nucleótido



ABSTRACT

Charolais cattle are an important beef breed in Mexico and are mainly selected for body weight traits. A previously study using Charolais populations from Mexico reported quantitative trait *locus* (QTL) associated for growth traits. Another strategy for mapping important genomic regions is by the identification of signatures of selection which are *loci* showing extended linkage disequilibrium (LD) and homozygosity patterns. By this approach, 263 important genomic regions containing QTL for productive traits have been identified in various cattle breeds and populations. The objective of this study was to identify signatures of selection in Charolais cattle genotyped with the GeneSeek Genomic Profiler Bovine HD panel of 77 K single nucleotide polymorphism (SNP). A total 71,871 SNP and 819 samples that passed a quality control were used for the analysis. The identification of signatures of selection was carried out by the integrated haplotype score (iHS) methodology implemented in the rehh R package. Ten SNP located on BTA 1, 4, 6, 7 y 14 were statistically significant ($P < 0.00003$). By identifying markers in LD with significant SNP the defined candidate regions mapped to the regions of 132.4-132.5 Mb on BTA 1; 52.8-53.4, 55.3-57.2, and 57.6-59.3 Mb on BTA 4; 38.9-40.6 Mb on BTA 6; 29.2-30.1 Mb on BTA 7; and 27.6-29.5 Mb on BTA 14. These candidate regions match with signatures of selection reported in other cattle breeds. Interestingly, genes and QTL located on these regions were associated to growth and other important productive traits. The genomic regions identified in this study indicates that selection for growth traits on Charolais population is conserving haplotypes on various chromosomes reflecting the polygenic background of these traits. These genomic regions could serve as basis for haplotype association studies and for the identification of causal genes related to growth traits.

Keywords: Charolais, Integrated haplotype score, Signatures of selection, Single nucleotide polymorphism



INTRODUCCIÓN

En México, el ganado bovino Charolais fue introducido en 1930 y desde entonces ha sido manejado con fines de producción de carne y seleccionado principalmente para características de peso corporal (Parra-Bracamonte *et al.* 2016). Con el objetivo de incrementar el peso corporal, pero sin dar lugar a la presencia de dificultades al parto, algunos productores seleccionan animales que presentan bajo peso al nacimiento pero alto peso al posdestete. Otras características, como la talla y circunferencia escrotal, también son consideradas para la selección animal. Adicionalmente, la raza Charolais también presenta el fenotipo de doble músculo, sin embargo, esta condición no es aceptada por los productores en México y como consecuencia seleccionan en contra a este fenotipo.

El efecto de la selección no natural (antropocéntricas) sobre las características productivas en las poblaciones de ganado Charolais podría haber incrementado la frecuencia de algunos alelos en genes relacionados a las características seleccionadas (de crecimiento) así como también en regiones asociadas debido al desequilibrio de ligamiento (LD). De esta manera, podrían haber surgido algunas regiones genómicas presentando patrones de DL y homocigosidad extendida (Utsunomiya *et al.* 2015). Recientemente, estudios genómicos utilizando paneles de alta densidad de polimorfismos de un solo nucleótido (SNP) se han enfocado en la identificación de este tipo de señales de selección, así como de *locus* de características cuantitativas (QTL). Mediante estudios de señales de selección de genoma completo (GWSS), se han reportado 263 regiones genómicas bajo selección en diferentes razas de ganado bovino (Randhawa *et al.* 2016). Algunas de estas regiones en el genoma bovino contienen genes con efectos mayores sobre el color de la capa, cornamenta e hipertrofia muscular y genes que influyen características poligénicas como la estatura, adaptación, eficiencia alimenticia, inmunidad, comportamiento, reproducción y producción en ganado cárnico y lechero.

En México, las poblaciones de ganado Charolais han sido investigadas para la identificación de QTL asociados a características de crecimiento mediante estudios de asociación de genoma completo, GWAS (Jahuey-Martínez *et al.* 2016). Sin embargo, la identificación de QTL mediante la metodología de GWAS es sensible a algunos factores como el tamaño de la población, la variable de respuesta, heredabilidad de la característica,



modelo estadístico, entre otros. Afortunadamente, la sola identificación de patrones extendidos de DL podría ayudar al mapeo genético de regiones que influyen características productivas. Por lo tanto, el objetivo de este estudio fue identificar señales de selección en ganado Charolais utilizando haplotipos obtenido con un arreglo de SNP de alta densidad 77 K (GGPHD 77 K).



MATERIALES Y MÉTODOS

Población y datos de estudio

En este estudio se utilizó la información genotípica de 855 animales Charolais. Estos animales fueron previamente genotipificados con el panel GGPHD 77 K (Jahuey-Martínez *et al.* 2016). Se realizó un control de calidad de los datos genotípicos mediante el paquete GenABEL (Aulchenko *et al.* 2007) de R. El filtrado de los datos consistió en eliminar SNP monomórficos, no autosómicos o sin posición definida en el ensamblaje del genoma bovino UMD versión 3.1, así como SNP con frecuencias del alelo menor (MAF) < 0.05 y tasa de llamado < 0.90. También se eliminaron las muestras que presentaron tasa de llamado < 0.80. Después del control de calidad, los datos genotípicos fueron faseados y los genotipos faltantes fueron imputados utilizando el programa SHAPEIT versión 2 (O'Connell *et al.* 2014). Para este procedimiento no se utilizó la información del pedigrí, por lo que los animales fueron considerados como no relacionados.

Identificación de señales de selección

El método integrated haplotype score (iHS) (Voight *et al.* 2006) fue utilizado para identificar señales de selección en haplotipos de ganado Charolais. Como una variante del método extended haplotype homozygosity (EHH) (Sabeti *et al.* 2002), iHS permite identificar eventos recientes de selección positiva en donde las frecuencias de un haplotipo se han incrementado recientemente pero los alelos ancestrales todavía están presentes. El estadístico integrated haplotype (iHS) fue calculado mediante:

$$iHS = \frac{\ln\left(\frac{iHH_A}{iHH_D}\right) - E\left[\ln\left(\frac{iHH_A}{iHH_D}\right)\right]}{SD\left[\ln\left(\frac{iHH_A}{iHH_D}\right)\right]},$$

donde iHH_A e iHH_D representan el estadístico integrated EHH para los alelos ancestrales (A) y derivados (D); E y SD son la desviación estándar y esperada de iHH_A/iHH_D , respectivamente. Después de la estandarización ($iHS \sim N(0,1)$), los valores de iHS fueron transformados a valores π_{iHS} mediante la fórmula $\pi_{iHS} = -\log_{10}[1 - 2|\Phi(iHS) - 0.5|]$ en donde $\Phi(iHS)$ es una función gaussiana de la distribución acumulada de iHS. Los valores de π_{iHS} pueden ser interpretados como valores de P (en una escala de $-\log_{10}$) asociados con la hipótesis nula, por lo que en este estudio valores de π_{iHS} mayores o iguales a 4.5 fueron



considerados estadísticamente significativos ($P < 0.00003$) e indicadores de regiones genómicas con señales de selección. Todos los análisis estadísticos para la identificación de señales de selección fueron realizados utilizando la librería *rehh* (Gautier *et al.* 2017) de R. Las regiones significativas fueron investigadas mediante un procedimiento de agrupación el cual consistió en la identificación de grupos de marcadores en DL ($r^2 > 0.2$) incluyendo para esto solo los SNP significativos y SNP que presentaran valores de $P < 0.01$. Este procedimiento de agrupación fue realizado alrededor de cada SNP significativo considerando una ventana de 1 Mb.

Todos los grupos de marcadores identificados, fueron considerados como regiones candidatas con señales de selección. Los genes y QTL localizados alrededor de 2 Mb de estas regiones, fueron identificados a partir de la anotación del genoma bovino UMD versión 3.1 y de la base de datos QTLdb (Hu *et al.* 2016), respectivamente. Las anotaciones funcionales de los genes fueron investigadas a partir de literatura y de la base de datos de Uniprot. Además, otras regiones genómicas presentando menores valores de significancia fueron también consideradas para realizar una caracterización funcional de la lista de genes.

Caracterización funcional de regiones candidatas

Para investigar las implicaciones biológicas de los genes localizados en las regiones candidatas, se realizó un análisis de sobrerrepresentación (ORA) utilizando los resultados completos de la identificación de señales de selección. Este método consiste en la identificación de grupos funcionales de genes sobrerrepresentados por genes de interés o genes significativos. Los grupos funcionales de genes fueron obtenidos a partir de las anotaciones de bovino de Gene Ontology (GO) y Medical Subject Headings (MeSH) de las librerías de R *org.Bt.eg.db* (versión 3.4.0) and *MeSH.Bta.eg.db* (versión 1.7.0). Solo las categorías de procesos biológicos (PB) de GO así como las de anatomía (A) y procesos y fenómenos (G) de MeSH fueron consideradas para el ORA. Para definir la lista de genes para el ORA, se realizó un procedimiento de anotación génica de todo el panel de SNP, el cual consistió en asignar cada SNP a todos los genes localizados parcial o completamente en una región de 0.75 Mb alrededor de cada *loci*. Una vez realizada la asignación de SNP, el valor de significancia de cada SNP fue utilizado para calcular el valor global de significancia de



cada gen. Estos valores de significancia fueron calculados mediante el método GATES (Li *et al.* 2011) el cual considera tanto el número como el DL (r^2) de los SNP asignados a cada gen. Solo los genes que estuvieron representados por al menos 3 SNP fueron considerados en los análisis posteriores. La lista completa de genes anotados por el panel de SNP fue utilizada como universo de genes en el ORA. Sin embargo, una lista de genes de interés diferente fue utilizada en ORA de GO y MeSH debido a que la base de datos GO contiene un menor número de anotaciones funcionales en comparación a la base de datos MeSH. En el ORA basado en GO, la lista de genes de interés contenía genes con valores de $P < 0.04$ mientras que en el ORA basado en MeSH se utilizó una lista de genes con valores de $P < 0.01$. Para evaluar si un grupo funcional de genes estaba sobrerrepresentado por genes de interés se utilizó una prueba hipergeométrica de la siguiente manera:

$$P = 1 - \sum_{i=0}^{g-1} \frac{\binom{S}{i} \binom{N-S}{k-i}}{\binom{N}{k}},$$

donde S es el número total de genes de interés, N es el universo total de genes y k es el número de genes el grupo funcional evaluado. Los grupos funcionales con valores de $P < 0.0005$ fueron considerados como estadísticamente sobrerrepresentados por genes de interés. Los ORA fueron realizados utilizando las librerías clusterProfiler (Yu *et al.* 2012) y meshr (Morota *et al.* 2015). Todos los códigos utilizados en este estudio se encuentran en el apartado de apéndices.



RESULTADOS

En total, 71,871 SNP y 819 individuos que pasaron el control de calidad fueron utilizados en este estudio. A partir de estos, se pudo calcular el iHS para 62,318 marcadores SNP. El método iHS proporciona una medida del decaimiento de la homocigosidad haplotípica alrededor de *loci* mediante la comparación del EHH entre el alelo ancestral (considerado como control) y el derivado (seleccionado) dentro de una población. En términos de interpretación, valores extremos de iHS positivos o negativos indican una excesiva homocigosidad haplotípica sobre los alelos ancestrales y derivados, respectivamente. Las regiones genómicas presentando picos de iHS en una o ambas direcciones son de gran interés ya que, en especies de importancia doméstica, reflejan la selección artificial relacionada a características productivas. En la figura 2.1 se muestra la distribución genómica de los valores de iHS y piHS en cada uno de los autosomas bovinos (BTA).

Diez SNP localizados en los BTA 4, 1, 6, 7 y 14 fueron estadísticamente significativos (cuadro 2.1). Mediante gráficos regionales (figura 2.2) se muestran las anotaciones génicas y QTL en cada una de las señales de selección. Solo los SNP rs43391925 y rs43394545 de BTA estuvieron localizados en regiones intrónicas mientras que el resto de los marcadores estuvieron localizados en regiones intergénicas. El análisis de agrupamiento alrededor de los SNP significativos indicó las regiones candidatas de 132.4-132.5 Mb de BTA 1; 52.8-53.4, 55.3-57.2 y 57.6-59.3 Mb de BTA 4; 38.9-40.6 Mb de BTA 6; 29.2-30.1 Mb de BTA 7; y 27.6-29.5 Mb de BTA 14 (cuadro 2.2). El número total de SNP en cada una de las regiones candidatas fue de 7, 19, 10, 10, 17, 13 y 18, respectivamente. Un total de 88 genes estuvieron localizados en las regiones candidatas.

Un total de 22,530 genes fueron anotados por el panel de SNP pero un número reducido fue utilizado para ORA. El número total de genes factibles, es decir, genes con anotaciones funcionales en las bases de datos de GO, MeSH y KEGG fue de 4,611 (255 de interés), 16,654 (296 de interés), 8444 (364 de interés), respectivamente. Utilizando estos genes en ORA se identificaron 21 PB, 1 A y 6 G términos estadísticamente significativos ($P < 0.0005$). La lista completa de genes de interés en cada término identificado se muestra en el cuadro 2.3.

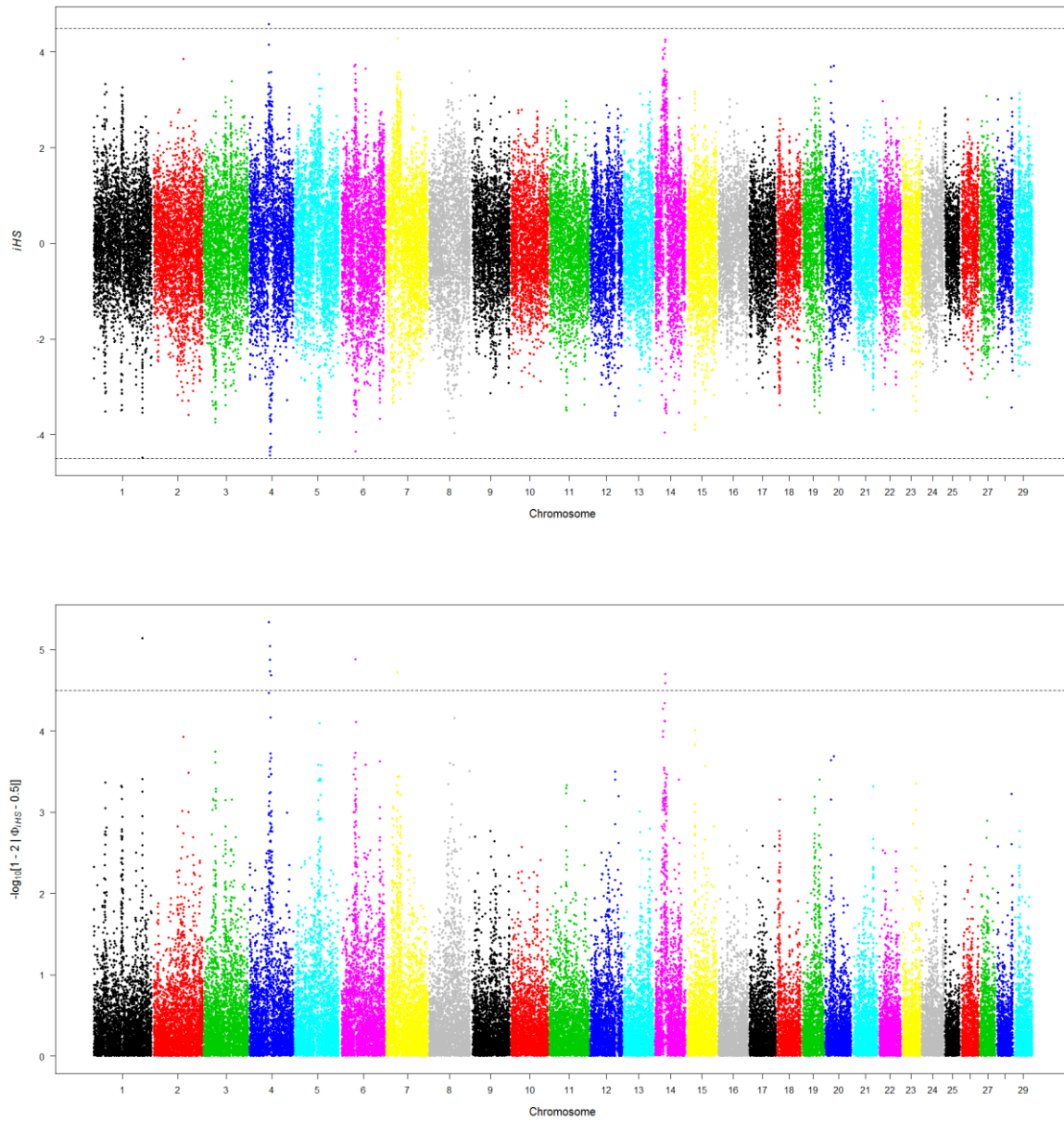


Figura 2.1 Gráficos de Manhattan de los valores integrated haplotype score (iHS) transformados en valores piHS. Los SNP con valores piHS estadísticamente significativos ($P < 0.00003$) se encuentran arriba de la línea punteada.



Cuadro 2.1 SNP estadísticamente significativos ($P < 0.00003$) en las regiones genómicas con señales de selección.

SNP ^a	BTA ^b	Posición (pb) ^c	iHS ^d	piHS ^e
rs135183459	4	53,163,338	4.581854	5.336419
rs109384003	1	132,555,764	-4.48526	5.137713
rs41656231	4	56,329,704	-4.43745	5.040802
rs43458270	6	39,837,065	-4.35478	4.875487
rs43391925	4	56,479,030	-4.3521	4.870186
rs43394545	4	56,409,650	-4.28078	4.729919
rs110685244	7	29,519,033	4.275014	4.718682
rs41579042	14	28,575,850	4.261253	4.691897
rs108981994	4	58,434,323	-4.25655	4.68277
rs109933597	14	28,671,618	4.203729	4.580783

^aSNP: nombre de referencia del SNP

^bBTA: autosoma de *Bos taurus*

^cpb: pares de bases

^diHS: integrated haplotype score

^epiHS: integrated haplotype score transformado

Cuadro 2.2 Descripción de las regiones extendidas con señales de selección.

BTA ^a	Inicio	Final	Tamaño (kb) ^b	Genes
4	51,849,283	60,346,408	8497.125	MET, CAV1, CAV2, TES, TFEC, ENSBTAG00000045017, MDFIC, ENSBTAG00000043905, FOXP2, ENSBTAG00000047618, PPP1R3A, SMIM30, GPR85, LSMEM1, IFRD1, ZNF277, BMT2, TMEM168, STIP1, DOCK4, ENSBTAG00000044727, IMMP2L, ENSBTAG00000045592, LRRN3, ENSBTAG00000020620, SNORA20, DNAJB9, THAP5, ENSBTAG00000046491, GPR141
1	131,456,625	133,572,701	2116.076	PIK3CB, FAIM, CEP70, ESYT3, MRAS, NME9, ARMC8, DBR1, A4GNT, DZIP1L, CLDN18, ENSBTAG00000019535, ENSBTAG00000043591, IL20RB, ENSBTAG00000046781, NCK1, SLC35G2, ENSBTAG00000048305
6	37,939,012	41,612,700	3673.688	ABCG2, PKD2, SPP1, MEPE, IBSP, ENSBTAG00000045804, LAP-3, MED28, FAM184B, DCAF16, NCAPG, LCORL, SLIT2, bta-mir-218-1
7	28,221,112	31,179,226	2958.114	MARCH3, LMNB1, ENSBTAG00000047241, TEX43, PHAX, ALDH7A1, GRAMD3, bta-mir-2458, ENSBTAG00000043765, ENSBTAG00000046696, ZNF608, ENSBTAG00000047546
14	26,604,507	30,575,157	3970.650	TOX, CA8, RAB2A, CHD7, ENSBTAG00000045005, ENSBTAG00000043978, ASPH, ENSBTAG00000044415, ENSBTAG00000014045, ENSBTAG00000045031, GGH, TTPA, YTHDF3, ENSBTAG00000046964

^aBTA: autosoma *Bos taurus*

^bkb: kilo pares de bases



Cuadro 2.3 Términos GO y MeSH estadísticamente significativos ($P < 0.0005$).

ID	Proporción de genes	P	Descripción del término: genes informativos
GO:0032272	6/23	0.00123	Regulación negativa de la polimerización de proteínas: CAPZA2, SNCA, TUBB4A, DMTN, ARFGEF1, TBCD
GO:0051952	4/10	0.00147	Regulación del transporte de aminas: SNCA, DRD2, CHGA, SNCG
GO:0017015	5/18	0.00235	Regulación de la vía de señalización del receptor del factor de crecimiento transformante beta: CAV1, CAV2, PIN1, FLCN, ITGA3
GO:1903844	5/18	0.00235	Regulación de la respuesta celular a los estímulos del factor de crecimiento transformante beta: CAV1, CAV2, PIN1, FLCN, ITGA3
GO:0051494	7/36	0.00305	Regulación negativa de la organización del citoesqueleto: TMEFF2, CAPZA2, SNCA, TUBB4A, DMTN, ARFGEF1, TBCD
GO:0015837	4/12	0.00318	Transporte de aminas: SNCA, DRD2, CHGA, SNCG
GO:0048732	10/67	0.00339	Desarrollo de glándulas: WNT2, CAV1, PKD2, PRKCSH, TGFBR1, NR5A1, SNAI2, CRH, DRD2, GH1
GO:0071559	7/37	0.00359	Respuesta al factor de crecimiento beta: WNT2, CAV1, CAV2, PIN1, TGFBR1, FLCN, ITGA3
GO:0071560	7/37	0.00359	Respuesta celular a los estímulos del factor de crecimiento transformante beta: WNT2, CAV1, CAV2, PIN1, TGFBR1, FLCN, ITGA3
GO:0043271	5/20	0.00388	Regulación negativa del transporte de iones: CAV1, IGF1, SNCA, PKD2, DRD2
GO:0030198	7/38	0.00420	Organización de la matriz extracelular: IBSP, DMP1, TGFBR1, LOXL2, COL1A1, FBLN5, EXOC8
GO:1902904	7/38	0.00420	Regulación negativa de la organización fibrilar: TMEFF2, CAPZA2, SNCA, TUBB4A, DMTN, ARFGEF1, TBCD
GO:0007179	6/29	0.00436	Ruta de señalización del receptor del factor de crecimiento transformante beta: CAV1, CAV2, PIN1, TGFBR1, FLCN, ITGA3
GO:0034766	4/13	0.00439	Regulación negativa del transporte transmembranal de iones: CAV1, IGF1, PKD2, DRD2
GO:0090263	4/13	0.00439	Regulación positiva de la ruta de señalización canónica Wnt: WNT2, CAV1, PIN1, COL1A1
GO:1901983	4/13	0.00439	Regulación de la acetilación de proteínas: SNCA, SNAI2, FLCN, GH1
GO:1904063	4/13	0.00439	Regulación negativa del transporte transmembranal de cationes: CAV1, IGF1, PKD2, DRD2
GO:2001021	4/13	0.00439	Regulación negativa de la respuesta a estímulos de daño al ADN: PARPBP, CDKN2D, SNAI2, PGAP2
GO:0007043	5/21	0.00487	Ensamblaje de unión célula-célula: CLDN1, CAV1, SNAI2, FLCN, TBCD
GO:0043062	7/39	0.00488	Organización de la estructura extracelular: IBSP, DMP1, TGFBR1, LOXL2, COL1A1, FBLN5, EXOC8
D040641	9/76	0.00001	Locus de características cuantitativas: OSTN, STAT1, NHLH2, CDC10, IGF-1, FAM13A1, ABCG2, PKD2, OPN, LAP-3, NCAPG, XKR4, PLAG1, RDHE2, GH
D012741	4/16	0.00015	Maduración sexual: IGF-1, NCAPG, IL1A, IL1B
D015430	4/21	0.00046	Ganancia de peso: IGF-1, OPN, NCAPG, PLAG1
D020641	15/328	0.00082	Polimorfismo de un solo nucleótido: OSTN, STAT1, CRYGC, ATP1A1, GSTM1, CFTR, WNT2, CDC10, MYBPC1, PMCH, IGF-1, FAM13A1, NAP1L5, ABCG2, PKD2, OPN, LAP-3, NCAPG, AGF, SIX6, IL1B, XKR4, PLAG1, RDHE2, CRF, DRD2, HSBP1, GH, Gpr14
D001693	3/13	0.00139	Transporte biológico, Activo: CLDN1, ABCG2, SLC39A14, CHGA
D005247	3/17	0.00314	Comportamiento alimenticio: IGF-1, NCAPG, PLAG1, GH
D001842	4/28	0.00143	Hueso y huesos: OPN, IBSP, NCAPG, TRAP, PLAG1, CHAD, COL1A1



DISCUSIÓN

Las regiones candidatas identificadas en este estudio coinciden con los resultados obtenidos en otros GWSS. En el meta análisis de señales de selección en ganado bovino, Randhawa *et al.* (2016) definió las regiones genómicas detectadas por 64 GWSS. Las regiones definidas por Randhawa *et al.* (2016) que coinciden con nuestros resultados fueron las siguientes: 130.8-135.0 Mb de BTA 1; 3 regiones localizadas entre 43.6-57.9 Mb de BTA 4; 7 regiones localizadas entre 37.9-40.4 Mb de BTA 6; 28.3-32.0 Mb de BTA 7; y 6 regiones localizadas entre 20.5-40.22 Mb de BTA 14.

En BTA 1, la región extendida de 131.4-133.5 Mb contiene 18 genes, de los cuales 16 fueron genes que codifican proteínas, 1 pseudogen y 1 snRNA. Ninguno de los genes identificados se localizó dentro de la región candidata. Los genes más cercanos al SNP rs109384003 fueron ENSBTAG00000019535 (pseudogen) y CLDN18 (Claudin 18) localizados 104.7 kb y 364.0 kb río abajo, respectivamente. Cerca de la región candidata, se han reportado señales de selección para ganado Jersey (Kim *et al.* 2015 y Porto-Neto *et al.* 2013). Además, también se han reportado QTL para características como promedio de ganancia diaria, índice al nacimiento, peso corporal del nacimiento al sacrificio, ganancia de peso corporal, facilidad al parto, tasa de preñez, producción de carne magra, estatura, muerte fetal y fuerza. Por ejemplo, Snelling *et al.* (2010) reportaron QTL empezando de la posición 131,806,615 pb hasta 149,852,860 pb asociados a pesos corporales (del nacimiento al año de edad) en una población de cruza. Sorbolini *et al.* (2017) también reportaron QTL en 133.0 Mb asociados con peso corporal al año y al sacrificio en ganado Marchigiana. Mokry *et al.* (2013) reportaron un QTL en 132,385,787 pb que explicó el 9.4% del grosor de grasa dorsal en ganado Canchim y propusieron los genes CLDN18 y DZIP1L (Zinc finger protein DZIP1L) como candidatos para este QTL. Interesantemente, CLDN18 y DZIP1L fueron dos de los genes más cercanos a la región candidata de selección en el presente estudio. De acuerdo con estos reportes, la región candidata de BTA 1 podría estar relacionada a la selección de características de crecimiento en la población de ganado Charolais estudiada. Sin embargo, en el estudio previo de GWAS no se encontró ningún QTL asociado a características de crecimiento en esta región genómica (Jahuey-Martinez *et al.* 2016).



Otros genes interesantes en BTA 1 fueron PIK3CB (Phosphatidylinositol 4, 5-bisphosphate 3-kinase catalytic subunit beta isoform) y MRAS (Ras-related protein M-Ras). Utilizando datos de secuencias genómicas, Taye *et al.* (2017) encontraron algunas regiones bajo selección en ganado Ankole incluyendo las posiciones de 131.5-131.55 y 131.73-131.78 Mb y propusieron los genes PIK3CB y MRAS como genes candidatos. Las anotaciones de Uniprot mostraron que PIK3CB participa en el metabolismo de los fosfolípidos, específicamente en la ruta de biosíntesis del fosfato fosfatidilinositol y MRAS puede servir como un importante transductor de señales para nuevos estímulos en el control de la proliferación celular. Otros genes localizados en la región candidata de BTA 1 con anotaciones importantes en Uniprot fueron ESYT3 (Extended synaptotagmin-3) y SLC35G2 (Solute carrier family 35 member G2) los cuales podrían participar en el transporte de lípidos celulares y en la proliferación celular, respectivamente.

En BTA 4, la región extendida de 51.8-60.3 Mb abarcó las tres regiones candidatas y contiene 30 genes de los cuales 24 fueron genes que codifican proteínas, 1 pseudogen, 1 miRNA, 1 snoRNA y 3 snRNA. Varios reportes indican la presencia de señales de selección en esta región. Por ejemplo, se han identificado señales de selección para ganado Simental y Holstein (por ejemplo, Rothammer *et al.* 2013, Bomba *et al.* 2015). En contraste, solo muy pocos QTL (n=7) han sido reportados cerca de estas regiones candidatas. La comparación de la región extendida con la base de datos QTLdb permitió identificar QTL asociados con consumo residual de alimento, promedio de consumo diario de alimento, promedio de ganancia diaria, profundidad corporal, estructura, intervalo entre partos, grosor de grada en la 12^a costilla, área pélvica, ancho de la grupa, estatura y fuerza.

En la región candidata de 52.8-53.4 Mb de BTA 4, el gen más cercano al SNP más significativo rs135183459 fue ENSBTAG00000045017 (snRNA) localizado 173.8 kb río abajo. También en esta región, los genes TES (Testin) y TFEC (Transcription factor EC), localizados río abajo, también fueron importantes ya que han sido propuestos como genes candidatos para características de grosor de grasa dorsal. Mokry *et al.* (2013) reportaron que estos 2 genes se encontraron cerca de un QTL localizado en 52,535,108 bp identificado en ganado Canchim. Las anotaciones de Uniprot mostraron que TES es una proteína de ensamblaje que podría participar en la adhesión celular y como supresor de tumores



mientras que TFEC es un regulador transcripcional que actúa como un supresor o activador reconociendo o uniéndose a promotores de una manera secuencia específica. Otros genes relevantes localizados río abajo fueron CAV1 (Caveolin-1) y CAV2 (Caveolin-2) los cuales aparecieron en términos significativos identificados mediante ORA relacionados a la ruta de señalización del receptor del factor de crecimiento transformante beta (GO:0017015, GO:1903844, GO:0071559, GO:0071560, GO:0007179), desarrollo de glándulas (GO:0048732) y otros PB (GO:0043271, GO:0034766, GO:1904063, GO:0007043). Por otro lado, el gen localizado río arriba MDFIC (MyoD family inhibitor domain-containing protein) junto con ENSBTAG00000043905 (snRNA) fueron reportados como candidatos en una región identificada bajo selección en ganado Maasai (Mslaya *et al.* 2017). Además, el gen MDFIC también fue detectado bajo selección en un estudio utilizando poblaciones de cruza de ganado porcino diferenciadas por presentar altos o bajos pesos al nacimiento (Zhang *et al.* 2014). La anotación en Uniprot indica que MDFIC actúa como un activador o represor transcripcional.

En la región candidata de 55.3-57.2 Mb de BTA 4, los SNP rs43394545 y rs43391925 estuvieron localizados dentro de los genes BMT2 (Probable methyltransferase BMT2 homolog) y TMEM168 (Transmembrane protein 168), respectivamente. Sin embargo, ninguno de estos genes presentaba anotaciones biológicas en la literatura revisada. También dentro de esta región, el gen IFRD1 (Interferon-related developmental regulator 1) resultó de mayor interés ya que se ha propuesto que participa la expresión del fenotipo de doble músculo, la cual es una condición también presente en la raza Charolais, pero indeseable en la población de estudio. Al estudiar el doble músculo en ganado Piedmontese mediante una red de coexpresión codiferencial, Hudson *et al.* (2012) encontraron que el fenotipo de doble músculo se debe al efecto combinado de los genes MSTN, IFRD1 y MYOD1 (Myoblast determination protein 1). Hudson *et al.* (2012) demostraron que estos tres genes forman una subred génica con niveles similarmente altos y coordinados de coexpresión diferencial. La anotación mediante Uniprot indica que IFRD1 podría participar en la regulación de la actividad génica en las rutas proliferativas y/o diferenciadoras inducidas por el factor de crecimiento nervioso. Por otro lado, junto con los genes localizados río arriba DOCK4 (Dedicator of cytokinesis protein 4) y LRRN3 (Leucine-rich repeat neuronal protein 3), IFRD1 también ha sido propuesto como candidato para la resistencia a parásitos en ovejas



(Atlija *et al.* 2016). Otros genes interesantes fueron STIP1 (Stress-induced-phosphoprotein 1) y PPP1R3A (Serine/threonine-protein phosphatase PP1-gamma catalytic subunit) los cuales estuvieron localizados río abajo de la región candidata. STIP1 ha sido sugerido como gen candidato para terneza de la carne en ganado Charolais (Ramayo-Caldas *et al.* 2016) mientras que PPP1R3A ha sido identificado como diferencialmente expresado en músculo e hígado de hembras y machos de ganado bovino (Seo *et al.* 2016). Las anotaciones de Uniprot indican que PPP1R3A se asocia con más de 200 proteínas regulatorias para formar holoenzimas altamente específicas las cuales desfosforilizan cientos de blancos biológicos.

En la región candidata de 57.6-59.3 Mb de BTA 4, los genes más cercanos al SNP rs108981994 fueron IMMP2L (Mitochondrial inner membrane protease subunit 2) y ENSBTAG00000020620 (pseudogene) localizados a 350.7 kb río abajo y 289.9 kb río arriba, respectivamente. Sin embargo, no se encontraron anotaciones biológicamente relevantes relacionadas a estos genes.

En BTA 6, la región extendida de 37.9-41.6 Mb contiene 14 genes de los cuales 12 son genes que codifican proteínas y 2 miRNA. En la región candidata de 38.9-40.6 Mb, LCORL (Ligand-dependent nuclear receptor corepressor-like protein) fue el único gen encontrado y estuvo localizado a 844.9 kb río abajo del SNP rs43458270. Un gran número de señales selección y QTL han sido reportados en esta región genómica. Siete señales de selección localizadas entre 37.9-40.4 Mb han sido identificadas en ganado Piedmontese, Romagnola, Charolais, Holstein, Braunvieh, Pardo Suizo y Jersey de acuerdo con reportes, por ejemplo, Porto-Neto *et al.* (2013), Kemper *et al.* (2014), Bomba *et al.* (2015), entre otros. A lo largo de la región extendida se encuentran muchos QTL (n=132) asociados principalmente a promedio de ingesta diaria de alimento, promedio de ganancia diaria, peso corporal del nacimiento al año de edad, ganancia de peso corporal, peso de la canal, tamaño de la cría, facilidad al parto, conformación, grosor de grasa en la 12^a costilla, porcentaje de grasa intramuscular, área del músculo *longissimus*, área pélvica, estatura, entre otras características.

En BTA 7, la región extendida de 28.2-31.1 Mb contiene 12 genes de los cuales 9 son genes que codifican proteínas, 1 miRNA y 2 snRNA. En la región candidata de 29.2-30.1 Mb, los genes más cercanos al SNP rs110685244 fueron ENSBTAG00000043765



(snRNA) y ENSBTAG00000046696 localizados a 319.7 kb río abajo y 129.2 kb río arriba, respectivamente. Cerca de la región candidata, se han reportado señales de selección identificadas en ganado Holstein (Qanbari *et al.* 2010, Druet *et al.* 2013, Lee *et al.* 2014 y Bomba *et al.* 2015). Los QTL (n=10) reportados en la región extendida han sido asociados con promedio de ganancia diaria, capacidad corporal, peso corporal, facilidad al parto, tasa de preñez de hijas, longitud de gestación, producción de carne magra, área pélvica y muerte fetal. Mediante GWAS, Serão *et al.* (2013) descubrieron un QTL en 30,506,622 pb asociado con promedio de ganancia diaria en cruza de ganado bovino y encontraron ZNF608 (Zinc finger protein 608) era el gen más cercano a este QTL. Curiosamente, este gen ha sido comúnmente asociado con el índice de masa corporal en humanos (por ejemplo, Ntalla *et al.* 2013) y se encontró a 1125.6 kb río arriba del SNP rs110685244. Río debajo de la región candidata, Buzanskas *et al.* (2014) reportaron un QTL en 28,515,652 pb asociado con el peso al año en ganado Canchim. Para este QTL, Buzanskas *et al.* (2014) propusieron los genes MARCH3 (E3 ubiquitin-protein ligase MARCH3), LMNB1 (Lamin-B1), PHAX (Phosphorylated adapter RNA export protein), ALDH7A1 (Alpha-aminoaliphatic semialdehyde dehydrogenase), GRAMD3 (GRAM domain-containing protein 2B) y bta-mir-2458 (miRNA) como genes candidatos posicionales. Sin embargo, no se encontró información relevante de estos genes.

En BTA 14, la región extendida de 26.6-30.5 Mb contiene 14 genes de los cuales 9 son genes que codifican proteínas, 1 pseudogen, 2 miRNA, 1 rRNA y 1 snoRNA. En la región candidata de 27.6-29.5 Mb, los genes más cercanos al SNP rs41579042 fueron ENSBTAG00000043978 y ASPH (Aspartyl/asparaginyl beta-hydroxylase) localizados a 36.6 kb río abajo y 125.0 kb río arriba, respectivamente. Al igual que en BTA 6, un gran número de señales de selección y QTL han sido reportados en esta región de BTA 14. Seis señales de selección localizada entre 20.5-40.2 Mb han sido identificadas en ganado Limousin, Nellore, Holstein, Charolais, Braunvieh, Belga Azul en reportes de, por ejemplo, Druet *et al.* (2013), Kemper *et al.* (2014), entre otros. Los QTL reportados en esta región (n=44) han sido asociados a características como edad a la pubertad, promedio de ganancia diaria, peso corporal del nacimiento al año, ganancia de peso corporal, facilidad al parto, peso de la canal, intervalo al primer estro después del parto, producción de carne magra,



área del musculo *longissimus*, consumo residual de alimento, circunferencia escrotal, muerte fetal, entre muchas otras características.

Específicamente para ganado Charolais, estudios previos de identificación de señales de selección han indicado las regiones de 37.9-39.3 Mb de BTA 6 (Gibbs *et al.* 2009, Kemper *et al.* 2014, Xu *et al.* 2014, Zhao *et al.* 2015) y 22.2-31.7 Mb de BTA 14 (Druet *et al.* 2013, Gibbs *et al.* 2009, Kemper *et al.* 2014, Xu *et al.* 2014, Zhao *et al.* 2015). En particular, las regiones genómicas de BTA 6 y BTA 14 han sido comúnmente reportadas en la literatura como regiones que contienen QTL con efectos pleiotrópicos para características importantes como producción (crecimiento y conversión alimenticia), reproducción (fertilidad) y características exteriores (conformación). En estas regiones, la mayoría de los reportes han atribuido los efectos detectados a los *loci* comprendiendo los genes NCAPG (Condensin complex subunit 3) - LCORL y PLAG1 (Zinc finger protein PLAG1) - CHCHD7 (Coiled-coil-helix-coiled-coil-helix domain-containing protein 7). De acuerdo con estudios de mapeo fino y validación (por ejemplo, Fink *et al.* 2017), estos *loci* son los genes causales en los BTA 6 y 14, respectivamente. En nuestros resultados, LCORL fue el gen más cercano al SNP rs43458270 en la región candidata de BTA 6. Por otro lado, los genes PLAG1-CHCHD7 no aparecieron en la región candidata de BTA 14 pero estuvieron localizados muy cerca de la señal de selección como se puede observar en el gráfico regional (figura 2.2). De hecho, dos SNP (rs134249915 y rs42403970) localizados en 22,517,710 y 27,155,254 bp pudieran ser indicadores de estos genes. Sin embargo, es importante destacar que estos dos SNP, aunque no fueron significativos, aparecieron entre el 20 SNP con los menores valores de P. Además, el análisis funcional reveló que los genes NCAPG y PLAG1 aparecieron en 6 de los 7 términos significativos en el ORA. Estos genes junto con otros interesantes estuvieron sobrerrepresentados en términos relacionados con *loci* de características cuantitativas (D040641), maduración sexual (D012741), ganancia de peso (D015430), polimorfismo de un solo nucleótido (D020641), comportamiento alimenticio (D005247) y hueso y huesos (D001842). Información funcional acerca de estos genes se encuentra bien documentada en Takasuga (2016).



CONCLUSIÓN

Dos de los *loci* identificados en este estudio se encuentran localizados en regiones genómicas de BTA 6 y 14 las cuales han sido previamente reportadas y validadas conteniendo señales de selección y QTL en varias razas bovinas incluyendo Charolais. Más importante, tres nuevas regiones candidatas (en BTA 1, 4 y 7) de señales de selección para la raza Charolais se reportan este estudio. De acuerdo al ORA y los reportes de la base de datos QTLdb, los productos génicos localizados en estas regiones candidatas (CAV1, CAV2, MDFIC, BMT2, TMEM18, IFDR1, ZNF608, NCAPG, LCORL, PLAG1) podrían actuar de manera conjunta influyendo mecanismos biológicos importantes que impactan en las características productivas.



LITERATURA CITADA

Atlija M., Arranz J.J., Martinez-Valladares M., Gutiérrez-Gil B. 2016. Detection and replication of QTL underlying resistance to gastrointestinal nematodes in adult sheep using the ovine 50K SNP array. *Genet Sel Evol.* 48:4.

Aulchenko Y.S., Ripke S., Isaacs A., van Duijn C.M. 2007. GenABEL: an R library for genome-wide association analysis. *Bioinformatics.* 23:1294-96.

Bomba L., Nicolazzi E., Milanesi M., Negrini R., Mancini G., Biscarini F., Stella A., Valentini A., Ajmone-Marsan P. 2015. Relative extended haplotype homozygosity signals across breeds reveal dairy and beef specific signatures of selection. *Genet Sel Evol.* 47(1):25.

Buzanskas M.E., Grossi D.A., Ventura R.V., Schenkel F.S., Sargolzaei M., Meirelles S.L., Mokry F.B., Higa R.H., Mudadu M.A., da Silva M.V., Niciura S.C., Torres R.A. Jr., Alencar M.M., Regitano L.C., Munari D.P.1. 2014. Genome-wide association for growth traits in Canchim beef cattle. *PLoS One.* 9:e94802.

Druet T., Pérez-Pardal L., Charlier C., Gautier M. 2013. Identification of large selective sweeps associated with major genes in cattle. *Anim Genet.* 44:758-62.

Fink T., Tiplady K., Lopdell T., Johnson T., Snell R.G., Spelman R.J., Davis S.R., Littlejohn M.D. 2017. Functional confirmation of PLAG1 as the candidate causative gene underlying major pleiotropic effects on body weight and milk characteristics. *Sci Rep.* 7:44793.

Gautier M., Klassmann A., Vitalis R. 2017. rehh 2.0: a reimplementaion of the R package rehh to detect positive selection from haplotype structure. *Mol. Ecol. Resour.* 17:78-90.

Gibbs RA, Taylor JF, Van Tassell CP, Barendse W, Eversole KA, Gill CA, et al. 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science.* 324:528-32.

Hu Z., Carissa A., Reecy P.J.M. 2016. Developmental progress and current status of the Animal QTLdb. *Nucleic Acids Research.* 44: D827-33.



- Hudson N.J., Dalrymple B.P., Reverter A. 2012.** Beyond differential expression: the quest for causal mutations and effector molecules. *BMC Genomics*. 13:356.
- Jahuey-Martínez F.J., Parra-Bracamonte G.M., Sifuentes-Rincón A.M., Martínez-González J.C., Gondro C., García-Pérez C.A., López-Bustamante L.A. 2016.** Genomewide association analysis of growth traits in Charolais beef cattle. *J. Anim. Sci.* 94(11):4570-4582.
- Kemper K., Saxton S., Bolormaa S., Hayes B., Goddard M. 2014.** Selection for complex traits leaves little or no classic signatures of selection. *BMC Genomics*. 15:246.
- Kim E.S., Sonstegard T., Rothschild M. 2015.** Recent artificial selection in U.S. Jersey cattle impacts autozygosity levels of specific genomic regions. *BMC Genomics*. 16:302.
- Lee H-J, Kim J, Lee T, Son JK, Yoon H-B, Baek K-S., Jeong J.Y., Cho Y-M., Lee K-T., Yang B-C., Lim H-J., Cho K., Kim T-H., Kwon E.G., Nam J., Kwak W., Cho S., Kim H. 2014.** Deciphering the genetic blueprint behind Holstein milk proteins and production. *Genome Biol Evol.* 6:1366-74.
- Li M.X., Gui H.S., Kwan J.S., Sham P.C. 2011.** GATES: a rapid and powerful gene-based association test using extended Simes procedure. *Am J Hum Genet.* 88:283-93.
- Mokry F.B., Higa R.H., de Alvarenga Mudadu M., Oliveira de Lima A., Meirelles S.L., Barbosa da Silva M.V., Cardoso F.F., de Oliveira M.M., Urbinati I., Niciura M.S.C., Tullio R.R., de Alencar M.M., Regitano C.A.L. 2013.** Genome-wide association study for backfat thickness in Canchim beef cattle using Random Forest approach. *BMC Genet.*14:47.
- Morota G., Peñagaricano F., Petersen J.L., Ciobanu D.C., Tsuyuzaki K., Nikaido I. 2015.** An application of MeSH enrichment analysis in livestock. *Anim Genet.* 46:381-7.
- Msalya G., Kim E.S., Laisser E.L., Kipanyula M.J., Karimuribo E.D., Kusiluka L.J., Chenyambuga S.W., Rothschild M.F.2017.** Determination of Genetic Structure and Signatures of Selection in Three Strains of Tanzania Shorthorn Zebu, Boran and Friesian Cattle by Genome-Wide SNP Analyses. *PLoS One.* 12:e0171088.



- Ntalla I., Panoutsopoulou K., Vlachou P., Southam L., Rayner N.W., Zeggini E., Dedoussis G.V. 2013.** Replication of established common genetic variants for adult BMI and childhood obesity in Greek adolescents: the TEENAGE study. *Ann Hum Genet.* 77:268-74.
- O'Connell J., Gurdasani D., Delaneau O., Pirastu N., Ulivi S., Cocca M., Traglia M., Huang J., Huffman J.E., Rudan I., McQuillan R., Fraser R.M., Campbell H., Polasek O., Asiki G., Ekoru K., Hayward C., Wright A.F., Vitart V., Navarro P., Zagury J.F., Wilson J.F., Toniolo D., Gasparini P., Soranzo N., Sandhu M.S., Marchini J. 2014.** A General Approach for Haplotype Phasing across the Full Spectrum of Relatedness. *PLoS Genet.* 10: e1004234.
- Parra-Bracamonte G.M., Lopez-Villalobos N., Morris S.T., Sifuentes-Rincón A.M., Lopez-Bustamante L.A. 2016.** Genetic trends for live weight traits reflect breeding strategies in registered Charolais Farms in Mexico. *Trop Anim Health Prod.* 48:1729-38.
- Porto-Neto L., Sonstegard T., Liu G., Bickhart D., Da Silva M., Machado M., Utsunomiya Y.T., Garcia J.F., Gondro C., Van Tassell C.P. 2013.** Genomic divergence of zebu and taurine cattle identified through high-density SNP genotyping. *BMC Genomics.* 14(1):876.
- Qanbari S., Pimentel E.C.G., Tetens J., Thaller G., Lichtner P., Sharifi A.R., Simianer H. 2010.** A genome-wide scan for signatures of recent selection in Holstein cattle. *Anim Genet.* 41:377-89.
- Ramayo-Caldas Y., Renand G., Ballester M., Saintilan R., Rocha D. 2016.** Multi-breed and multi-trait co-association analysis of meat tenderness and other meat quality traits in three French beef cattle breeds. *Genet Sel Evol.* 48:37.
- Randhawa I.A.S., Khatkar M.S., Thomson P.C., Raadsma H.W. 2016.** A Meta-Assembly of Selection Signatures in Cattle. *PLoS One.* 11:e0153013.
- Rothammer S., Seichter D., Forster M., Medugorac I. 2013.** A genome-wide scan for signatures of differential artificial selection in ten cattle breeds. *BMC Genomics.* 14(1):908.



- Sabeti P.C., Reich D.E., Higgins J.M., Levine H.Z.P., Richter D.J., Schaffner S.F., Gabriel S.B., Platko J.V., Patterson N.J., McDonald G.J., Ackerman H.C., Campbell S.J., Altshuler D., Cooper R., Kwiatkowski D., Ward R., Lander E.S. 2002.** Detecting recent positive selection in the human genome from haplotype structure. *Nature*. 419:832-7.
- Seo M., Caetano-Anolles K., Rodriguez-Zas S., Ka S., Jeong J.Y., Park S., Kim M.J., Nho W.G., Cho S., Kim H., Lee H.J. 2016.** Comprehensive identification of sexually dimorphic genes in diverse cattle tissues using RNA-seq. *BMC Genomics*. 17:81.
- Serão N.V., González-Peña D., Beaver J.E., Bollero G.A., Southey B.R., Faulkner D.B., Rodriguez-Zas S.L. 2013.** Bivariate genome-wide association analysis of the growth and intake components of feed efficiency. *PLoS One*. 8:e78530.
- Snelling W.M., Allan M.F., Keele J.W., Kuehn L.A., McDanel T., Smith T.P.L., Sonstegard T.S., Thallman R.M., Bennett G.L. 2010.** Genome-wide association study of growth in crossbred beef cattle. *J. Anim. Sci.* 88, 837-848.
- Sorbolini S., Bongiorno S., Cellesi M., Gaspa G., Dimauro C., Valentini A., Macciotta N.P. 2017.** Genome wide association study on beef production traits in Marchigiana cattle breed. *J Anim Breed Genet*. 134:43-48.
- Takasuga A. 2016.** PLAG1 and NCAPG-LCORL in livestock. *Anim Sci J*. 87:159-67.
- Taye M, Kim J, Yoon SH, Lee W, Hanotte O, Dessie T, et al. 2017.** Whole genome scan reveals the genetic signature of African Ankole cattle breed and potential for higher quality beef. *BMC Genet*. 18:11.
- Utsunomiya Y.T., Pérez O'Brien A.M., Sonstegard T.S., Sölkner J., Garcia J.F. 2015.** Genomic data as the “hitchhiker's guide” to cattle adaptation: tracking the milestones of past selection in the bovine genome. *Front Genet*. 6:36.
- Voight B.F., Kudaravalli S., Wen X., Pritchard J.K. 2006.** A map of recent positive selection in the human genome. *PLoS Biol*. 4(3):e72.
- Xu L., Bickhart D.M., Cole J.B., Schroeder S.G., Song J., Tassell C.P., Sonstegard T.S., Liu G.E. 2014.** Genomic signatures reveal new evidences for selection of important traits in domestic cattle. *Mol Biol Evol*. 32:711-25.



Yu G., Wang L., Han Y., He Q. 2012. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS*. 16: 284-287.

Zhang L., Zhou X., Michal J.J., Ding B., Li R., Jiang Z. 2014. Genome wide screening of candidate genes for improving piglet birth weight using high and low estimated breeding value populations. *Int J Biol Sci*. 10:236-44.

Zhao F., McParland S., Kearney F., Du L., Berry D. 2015. Detection of selection signatures in dairy and beef cattle using high-density genomic information. *Genet Sel Evol*. 47:49.



**CAPITULO III: Predicción genómica de características de peso corporal en ganado
Charolais de México**



RESUMEN

Comúnmente, la selección y mejoramiento genético de características de importancia productiva se realiza mediante el uso de los valores genéticos de los animales (EBV), los cuales son una estimación del mérito productivo de un animal para una o varias características, y su estimación exacta es necesaria para el logro del mejoramiento genético. Mediante la metodología tradicional utilizada para la estimación de EBV, denominada BLUP (por sus siglas en inglés best linear unbiased prediction), se requiere de un gran número de registros fenotípicos y genealógicos para lograr estimaciones con alta exactitud (>0.7). El uso de las herramientas genómicas ha logrado la evolución en las estimaciones de EBV, dando lugar a la denominada predicción genómica (PG). Esta metodología se fundamenta en la posibilidad de detectar indirectamente el efecto de los genes que influyen una característica de interés a través de la tipificación de miles de polimorfismos de un solo nucleótido (SNP) en desequilibrio de ligamiento (DL) y utilizar dichos efectos para incrementar la exactitud en la estimación de EBV ahora denominados GEBV. El objetivo de este estudio fue evaluar la exactitud de predicción de GEBV para peso al nacimiento (PN) y peso al destete (PD) de ganado Charolais de registro de México. La población de estudio consistió de 823 animales (683 hembras y 140 machos) genotipificados con el panel de 77 mil SNP. Registros de otros 5,575 animales no genotipificados también fueron utilizados en este estudio para estimar EBV mediante BLUP y para ajustar los fenotipos por efectos ambientales como grupo contemporáneo (GC) y edad de la madre. Los fenotipos ajustados fueron utilizados como variables de respuesta en todos los análisis posteriores. La predicción genómica (PG) se realizó mediante los métodos Genomic-based best linear unbiased prediction (GBLUP), Bayes C (BC) y Single-step Bayesian regression (SSBR). La exactitud de predicción también fue comparada mediante grupos de validación (GV) formados aleatoriamente y mediante GC. Los resultados mostraron que los tres métodos de PG proporcionaron exactitudes de predicción similares entre los GV pero no proporcionaron exactitudes de predicción que BLUP. La exactitud de predicción de GBLUP y BLUP fue de 0.35 y 0.37 para PN mientras que para PD fue de 0.30 y 0.41, respectivamente. Estos resultados muestran que la PG no es factible bajo los escenarios evaluados. De acuerdo con la literatura, mayores exactitudes de PG podrían obtenerse si se



incrementa el número de animales genotipificados (principalmente sementales) y mediante el uso de EBV desregresados como variables de respuesta.

Palabras clave: Charolais, peso vivo, polimorfismo de un solo nucleótido, predicción genómica



ABSTRACT

Commonly, the selection and genetic improvement of important productive traits is done through the use of the genetic values of the animals (EBV), which are an estimate of the productive merit of an animal for one or several characteristics, and its exact estimation is necessary for the achievement of genetic improvement. Using the traditional methodology used to estimate EBV, called BLUP (best linear unbiased prediction), a large number of phenotypic and genealogical records are required to obtain estimates with high accuracy (> 0.7). The use of genomic tools has achieved evolution in EBV estimates, giving rise to the so-called genomic prediction (PG). This methodology is based on the possibility of indirectly detecting the effect of genes that influence a characteristic of interest through the genotyping of thousands of single nucleotide polymorphisms (SNP) in linkage disequilibrium (DL) and using these effects to increase the accuracy in the estimation of EBV now called GEBV. The aim of this study was to evaluate the accuracy of estimation of GEBV for BW at birth (BWT) and weaning (WWT) of registered Charolais cattle from Mexico. The populations study included 823 animals (683 females and 140 males) genotyped with an array of 77 thousand SNP. Records from other 5,575 non-genotyped animals were also used in this study to estimate EBV by BLUP and to adjust the phenotypes by environmental effects like contemporary groups (CG) and age of dam. Adjusted phenotypes were used as variable response in all analysis. Genomic prediction (GP) was done by testing the methods Genomic-based best linear unbiased prediction (GBLUP), Bayes C (BC), and single-step Bayesian regression (SSBR). Accuracy of estimation was also compared by testing random and CG validation sets (VS). Our results showed that the three GP methods provided similar accuracy of estimation across VS but they did not produce higher accuracy than BLUP. The prediction accuracy of GBLUP versus BLUP was 0.35 and 0.37 for BWT while 0.30 and 0.41 for WWT. These results show that GP is not feasible under the tested scenarios. According to the literature, higher accuracies of GP methods could be obtained by increasing the number of genotyped animals (mainly sires) and by using deregressed EBV as variable response.

Key words: genomic prediction, single nucleotide polymorphism, live weight, Charolais



INTRODUCCIÓN

Mediante la predicción genómica (PG) ha sido posible la estimación de valores genómicos (GEBV) en muchas especies domésticas. A diferencia del método de evaluación genética tradicional denominado pedigrí BLUP (PBLUP), el método de PG utiliza la información de marcadores genéticos en lugar de las relaciones del pedigrí para estimar GEBV. Para la implementación ideal de PG se requiere de una población de referencia (PR) conformada por sementales ampliamente probados donde el efecto de todos los genes será estimado para una característica de interés mediante un gran número de polimorfismos de un solo nucleótido (SNP), por lo que la población de predicción necesita estar genéticamente relacionada a la PR. Asumiendo que los marcadores genéticos están en desequilibrio de ligamiento (DL) con los genes causales, los GEBV de los candidatos de selección pueden ser estimados como una función del genotipo individual multiplicado por los efectos previamente estimados. Mediante PG, la selección de animales puede realizarse de manera más exacta a edades tempranas, provocando una reducción en el intervalo generacional y un incremento en las ganancias genéticas por año (Meuwissen *et al.* 2001).

Las ventajas de la PG pueden observarse en razas de ganado lechero. Por ejemplo, en ganado Holstein de EU se ha logrado reducir a menos de la mitad el intervalo generacional y cuadruplicar la ganancia genética para algunas características esto como consecuencia de la implementación de la PG. (García Ruíz *et al.* 2016). La exitosa implementación de la PG en razas lecheras ha sido posible gracias a las PR idealmente estructuradas caracterizadas por sementales con un gran número de progenie, y también debido al adecuado registro fenotípico y genealógico implementado por los productores.

Los estudios en ganado de carne también han demostrado que la PG proporciona mayor exactitud de predicción al estimar GEBV en comparación con el método tradicional PBLUP incluso cuando las PR son limitadas en cuanto al número de animales y no están igualmente estructuradas como en ganado lechero. Sin embargo, las exactitudes de PG en ganado de carne no son tan altas como las obtenidas en ganado lechero. Por lo tanto, los estudios de PG en poblaciones de ganado de carne se han enfocado en la exploración para encontrar la mejor estrategia para implementar la PG investigando, por ejemplo, diferentes modelos estadísticos de predicción (Lee *et al.* 2017), densidades de marcadores (Gunia *et*



al. 2014), relaciones entre la PR y la población de predicción (Saatchi *et al.* 2012), predicciones uni o multirraciales (Neves *et al.* 2014), entre otros aspectos que impactan la exactitud de predicción.

Como uno de los mayores productores de ganado de carne a nivel mundial (USDA, 2017), México podría incrementar su producción cárnica a través de la selección para mejorar las características productivas del ganado. Sin embargo, el mejoramiento genético de características productivas en algunas razas bovinas para carne (incluida la Charolais) ha sido limitado (Parra-Bracamonte *et al.* 2016). Por lo tanto, el objetivo de este estudio fue investigar el impacto de la información genómica en la exactitud de predicción de GEBV de dos características de peso corporal de ganado mexicano Charolais de registro.



MATERIALES Y MÉTODOS

Población y datos de estudio

La población de estudio consistió de 823 bovinos Charolais (683 hembras y 140 machos) que previamente habían sido genotipificados con el panel GeneSeek Genomic Profiler Bovine HD de 77 mil SNP (GGPHD, Neogen Corp., Lincoln, NE) (Jahuey-Martínez *et al.* 2016). Esta población pertenece a 3 hatos del norte de México y representa la progenie de 138 Sementales y 577 Madres. Los registros fenotípicos de peso al nacimiento (PN) y peso al destete (PD) de la población genotipificada, así como los registros fenotípicos de 5,575 animales no genotipificados, fueron proporcionados por la Asociación Charolais Charbray Herd Book de México. Estos animales y sus ancestros conformaron un pedigrí de 6 generaciones que incluyó 7,256 animales.

Control de calidad de los datos genotípicos

Los datos genotípicos fueron previamente evaluados para descartar SNP y muestras cuya calidad fuera deficiente (Jahuey-Martínez *et al.* 2016). Los parámetros de calidad para los SNP fueron: GenCall > 50%, polimorfismo, tasa de llamado > 90%, frecuencia del alelo menor > 0.01, no desviación al equilibrio de Hardy-Weinberg (valor de $P > 1 \times 10^{-15}$), localización autosómica y con posición definida en el ensamblaje del genoma bovino UMD versión 3.1. Los parámetros de calidad para las muestras fueron: tasa de llamado > 80%, heterocigocidad < a 3 desviaciones estandar y correlación genotípica < 0.98. Un total de 68,337 SNP cumplieron con los parámetros de calidad y fueron utilizados en los análisis de predicción genómica. Los genotipos faltantes fueron reemplazados por el valor promedio observado para cada SNP. El grado de DL fue investigado calculando el estadístico r^2 entre cada par de SNP separados por diferentes distancias genómicas.

Variables de respuesta

Fenotipos ajustados (y^*) por efectos ambientales fueron calculados para ser utilizados como variables de respuesta en los análisis de predicción genómica. Los y^* fueron



obtenidos como los residuales de un modelo lineal que, en el caso de PN, incluyó los efectos fijos de grupo contemporáneo (GC) y edad de la madre, mientras que en PD solo incluyó el efecto fijo del GC. Los GC fueron formados como la interacción del sexo, hato, año y época de nacimiento del animal. En el cuadro 3.1 se muestran los estadísticos descriptivos de las características fenotípicas y las variables de respuesta utilizadas en este estudio. Los fenotipos ajustados fueron utilizados como variables de respuesta en lugar de EBV o EBV desregresados debido a que la mayoría de los animales en este estudio fueron hembras y en consecuencia tuvieron exactitudes bajas de EBV.

Cuadro 3.1 Descripción estadística de los fenotipos ajustados y crudos de pesos corporales de ganado Charolais (kg).

Variable ¹	n	Media	DS	Mediana	Min ²	Max ³	ES
PN	6,195	40.51	7.39	40	15	66	0.09
PD	5,727	228.64	41.83	226.9	101.47	396.59	0.55
PN*	6,195	0	5.87	0	-28.9	21.74	0.07
PD*	5,727	0	30.64	-0.81	-141.82	159.64	0.4

¹ PN, PD, and * son peso al nacimiento, peso al destete y fenotipos ajustados por efectos ambientales, respectivamente.

² min=valor mínimo

³ max= valor máximo

Predicción de EBV basada en el pedigrí (BLUP)

El análisis de BLUP fue realizado utilizando los datos fenotípicos de animales tanto genotipificados como no genotipificados. El modelo general utilizado en PBLUP fue el siguiente:

$$y = I\mu + Zu + e,$$

donde y es el vector de fenotipos corregidos (y^*), I es un vector de 1s, μ es el intercepto, Z es la matriz de diseño que asigna los registros a los efectos genético aditivos, u es el vector de efectos genético aditivos y e es un vector aleatorio de residuales. En este modelo se asume que $u \sim N(0, A\sigma_g^2)$, donde A es la matriz de relaciones del numerador derivada del pedigrí y σ_g^2 es la varianza genética aditiva. Se asumió que los residuales se distribuían idéntica e independientemente con media cero y varianza σ_e^2 . Los componentes de la varianza y las estimaciones de h^2 fueron calculados a partir de los datos fenotípicos crudos



y utilizando el modelo anterior más los efectos fijos antes mencionados (cuadro 3.2). Estos análisis fueron realizados utilizando REML en el paquete breedR de R (2017).

Cuadro 3.2 Estimaciones de los componentes de la varianza¹ y heredabilidad (h^2) de pesos corporales (kg).

Variable ²	σ^2_g	σ^2_e	h^2
PN	13.213	25.254	0.343
PD	232.78	620.92	0.229

¹ Los componentes de varianza son la varianza genética aditiva (σ^2_g) y la varianza ambiental (σ^2_e).

² PN y PD son el peso al nacimiento y al destete.

Métodos de predicción genómica

Los métodos Genomic-based best linear unbiased prediction (GBLUP) y Bayes C (BC) fueron evaluados utilizando únicamente los datos de la población genotipificada. El modelo de GBLUP fue similar al utilizado en BLUP pero una matriz de relaciones genómicas (G) estimadas de los marcadores genéticos fue utilizada en lugar de la matriz A. La matriz G fue estandarizada para tener un valor diagonal de aproximadamente uno seguido de una descomposición de autovectores. En el modelo BC, Zu fue remplazado por Wg donde W es la matriz de genotipos codificada como 0, 1 y 2 representando el número de alelos B para cada marcador y para cada animal y g es el vector efectos de los marcadores. Los métodos GBLUP y BC fueron evaluados en el paquete BGLR (Pérez y de los Campos, 2014) de R utilizando los parámetros por defecto para cada modelo corriendo 50,000 iteraciones incluyendo un periodo de descarte (burnin) de 5,000 muestras. En BGLR, el supuesto a priori para los efectos de los SNP en el modelo BC es una mezcla de una distribución con mayor masa de probabilidad en el entorno del cero y una distribución normal (Habier *et al.* 2011).

Adicionalmente, con el objetivo de combinar la información de animales genotipificados y no genotipificados en un modelo de predicción genómica, se evaluó el método Single-step Bayesian Regression (SSBR) (Fernando *et al.* 2014). Como una clase del método Single step GBLUP (Legarra *et al.* 2009), SSBR permite incorporar la matriz de relaciones genómicas derivada de los SNP en la matriz de relaciones genéticas derivada del pedigrí y de esta manera, la matriz de relaciones combinada es utilizada para calcular GEBVs para

todos los animales en el pedigrí (tanto genotipificados como no genotipificados). En el modelo SSBR, los fenotipos ajustados fueron modelados como en Lee *et al.* (2017): $y = X\beta + ZM\alpha + Z_n\epsilon + e$, donde y es el vector de fenotipos ajustados para los animales genotipificados y no genotipificados, $X = \begin{bmatrix} 1 & -Z_n A_{ng} A_{gg}^{-1} 1 \\ 1 & -Z_g 1 \end{bmatrix}$, $\beta = \begin{bmatrix} \mu \\ \mu_g \end{bmatrix}$, μ es la media general, μ_g representa la diferencia en los efectos genético aditivos entre los animales genotipificados y no genotipificados, Z es la matriz de diseño, $M = \begin{bmatrix} \widehat{M}_n \\ M_g \end{bmatrix}$, donde M_g es la matriz de covariables SNP para los animales genotipificados y $\widehat{M}_n = A_{ng} A_{gg}^{-1} M_g$, representa las covariables SNP imputadas para los animales no genotipificados que son derivados de sus parientes genotipificados, ϵ es el residual de imputación, Z_n y Z_g son las matrices de diseño asignando registros a los efectos genético aditivos de los animales no genotipificados y genotipificados, A_{ng} es la partición de A que corresponde a las relaciones entre los animales genotipificados y no genotipificados y A_{gg} es la partición de A que corresponde a los animales genotipificados. Los supuestos para μ , μ_g , e y ϵ fueron los mismos a los descritos en Lee *et al.* (2017). Para este método se utilizó el modelo BC para asumir que una proporción de 0.995 de los SNP no tenían efecto sobre la característica bajo estudio. El método SSBR fue evaluado utilizando 50,000 iteraciones en el paquete JWAS (2017) de Julia. Los códigos utilizados para cada modelo de PG se muestran en el apartado de apéndices.

Validación cruzada

Dos escenarios de poblaciones de validación (PV) basados en la elección de animales elegidos aleatoriamente (PV_AL) y elegidos de acuerdo a GC (PV_GC) fueron evaluados en este estudio. Las PV_AL fueron creadas seleccionando 82-83 animales de la población genotipificada hasta que se obtuvieran 10 PV_AL. Por otro lado, las PV_GC fueron creadas seleccionando todos los animales pertenecientes a GC que previamente habían sido clasificados por su tamaño (número de animales). Para clasificar los GC, se definieron tres clases de tamaños: GC pequeños, medianos y grandes los cuales incluían 1-6, 7-14 y 15-49 animales, respectivamente. Como una breve descripción, la clase más abundante en nuestra



base de datos fue la clase pequeña con 105 GC y 270 animales, seguida por la clase mediana con 26 GC y 259 animales y finalmente la clase grande con 13 GC y 294 animales. Al final, se seleccionaron 10, 2 y 1 GC de las clases pequeña, mediana y grande, respectivamente y creamos 12 PV_GC las cuales contenían en promedio 69 animales.

La exactitud de predicción de cada método fue calculada como la correlación entre los GEBV y los fenotipos ajustados de los animales en el conjunto de validación dividido por la raíz cuadrada de la h^2 de la característica (Meuwissen *et al.* 2013). También se calcularon el coeficiente de regresión (CR) del fenotipo ajustado sobre el GEBV así como el error cuadrático medio (ECM).

RESULTADOS Y DISCUSIÓN

En este estudio se analizaron los datos de ganado Charolais con el objetivo de evaluar el impacto de la información genómica en la exactitud de la estimación de GEBV para características de peso corporal. El fundamento de la PG se basa en que el efecto de todos los genes sobre una característica puede ser estimado de manera aproximada en una PR mediante un gran número de marcadores SNP en DL con los genes causales. Como primer paso, en este estudio se caracterizó el DL a partir de los marcadores SNP. Los valores de r^2 promedio entre pares de SNP en distancias cromosómicas de 0-25 kb, 25-50 kb, 50-100 kb, 100-500 kb, 0.5-1 Mb, 1-5 Mb y 5-10 Mb fueron 0.36, 0.27, 0.19, 0.10, 0.07, 0.04 y 0.02, respectivamente. La figura 3.1 muestra el decaimiento del DL sobre varias distancias genómicas. El DL observado en este estudio fue similar a lo reportado en una población de ganado Charolais de Canadá (Lu *et al.* 2012) así como en otras razas bovinas (Zhu *et al.* 2013).

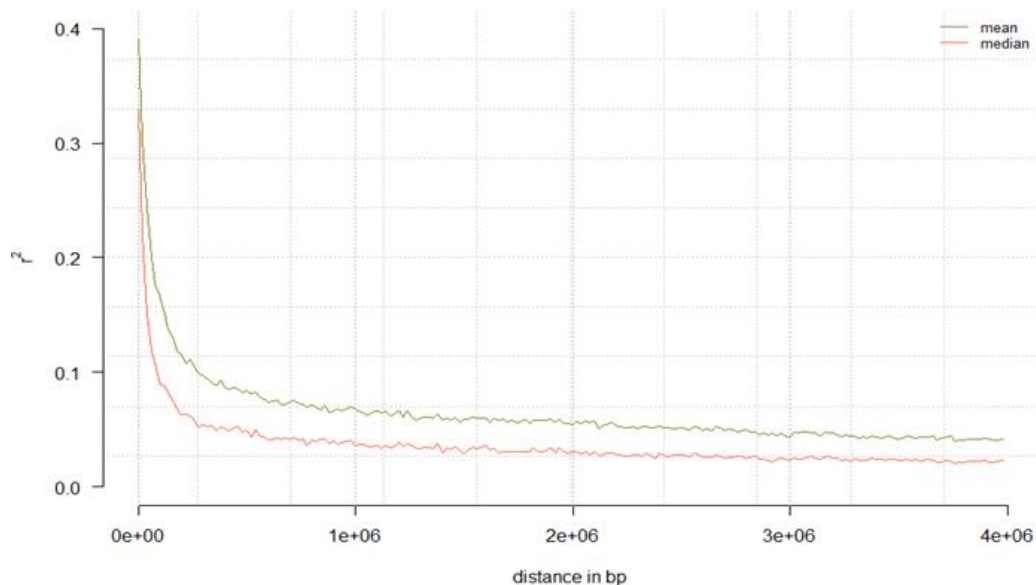


Figura 3.1 Decaimiento del DL (r^2) sobre varias distancias genómicas.

Las exactitudes de predicción obtenidas en este estudio fueron similares entre los modelos de PG y las clases de PV (Cuadro 3.3). Con el fin de comparar los resultados, las exactitudes de predicción obtenidas mediante GBLUP utilizando PV-GC serán utilizadas durante la discusión. Contrario a muchos estudios, las exactitudes de predicción obtenidas mediante los modelos de PG no fueron mayores a las exactitudes de predicción obtenidas



mediante BLUP. Por ejemplo, la exactitud de predicción obtenida mediante GBLUP fue de 0.35 y 0.30 mientras que la exactitud de predicción mediante BLUP fue de 0.37 y 0.41 para PN y PD respectivamente.

El coeficiente de regresión lineal más alto de la variable de respuesta sobre el GEBV se observó en los rasgos de PN y PD mediante el modelo BC, mientras que el modelo SSBR mostró el valor más pequeño (0,67) para PD (cuadro 3.4). El error cuadrático medio fue muy similar en todos los modelos.

Cuadro 3.3 Exactitud de predicción de varios métodos de predicción genómica¹ vs BLUP.

Variable ²	PV	GBLUP	BC	SSBR	BLUP
PN*	AL	0.312	0.311	0.379	0.381
	GC	0.356	0.353	0.371	0.372
PD*	AL	0.290	0.295	0.345	0.393
	CG	0.307	0.310	0.370	0.414

¹ Los métodos de predicción genómica son Genomic-based best linear unbiased prediction (GBLUP), Bayes C (BC) y Single-step Bayesian regression (SSBR).

² PN, PD y * son peso al nacimiento, peso al destete y fenotipos ajustados por efectos ambientales, respectivamente.

³ PV= tipo de población de validación.

Cuadro 3.4 Estimaciones de los coeficientes de regresión y error cuadrático medio de BLUP y métodos de predicción genómica¹.

Variable ⁴	PV ⁵	CR ²				ECM ³			
		GBLUP	BC	SSBR	BLUP	GBLUP	BC	SSBR	BLUP
PN*	AL	0.82	0.84	0.81	0.75	28.06	27.97	27.79	27.68
PN*	GC	0.85	0.89	0.76	0.71	26.87	26.88	26.90	26.82
PD*	AL	0.80	0.84	0.67	0.72	681.31	679.18	102.49	675.41
PD*	GC	0.84	0.87	0.78	0.81	674.77	673.62	99.90	669.34

¹ Los métodos de predicción genómica son Genomic-based best linear unbiased prediction (GBLUP), Bayes C (BC) y Single-step Bayesian regression (SSBR).

² RC= coeficiente de regresión.

³ MSE= error cuadrático medio.

⁴ PN, PD y * son peso al nacimiento, peso al destete y fenotipos ajustados por efectos ambientales, respectivamente.

⁵ VS= tipo de población de validación.

Se ha demostrado que los métodos de predicción de GEBV basados en el método de single-step, los cuales toman en cuenta la información de animales no genotipificados, predicen con mayor exactitud que los modelos bayesianos y GBLUP (por ejemplo Lee *et al.* 2017). Por esta razón, en este estudio también se evaluó el método basado en single-step denominado SSBR esperando obtener mayores exactitudes de predicción. Sin embargo, la inclusión de datos fenotípicos y de pedigrí de animales no genotipificados no tuvo ningún



efecto sobre la exactitud de predicción. La exactitud de predicción obtenida mediante SSBR fue de 0.37 tanto para PN como para PD (cuadro 3.3).

Estudios en otras razas bovinas para carne han reportado mayores exactitudes de predicción y demuestran que mediante PG se pueden obtener estimaciones más exactas del mérito genético en comparación a BLUP. Las exactitudes promedio de GEBV para PN y PD en las razas Angus, Hereford, Charolais, Limousin y Simmental fueron 0.70 y 0.53 (Saatchi *et al.* 2011), 0.42 y 0.21 (Saatchi *et al.* 2013), 0.42 y 0.45 (Gunia *et al.* 2014), 0.58 y 0.58 (Saatchi *et al.* 2012), 0.65 y 0.52 (Saatchi *et al.* 2012), respectivamente. Es importante destacar que en estos estudios se utilizaron mayor número de animales (principalmente sementales) en los análisis de predicción. Además, los GEBV fueron estimados a partir de EBV desregresados, en lugar de los fenotipos observados.

La exactitud de predicción en estudio pudo haber sido influida por factores como el tipo de variable de respuesta y el tamaño de la población los cuales son de los principales factores que impactan la PG. Los fenotipos corregidos como variables de respuesta no son la mejor opción para utilizar en PG debido a que estos fenotipos aún permanecen influenciados por factores no genéticos. En este estudio, las dos características de peso corporal solo fueron corregidas por los efectos de GC y edad de la madre. Otros factores como el efecto permanente de la madre podrían ser modelados en el modelo de ajuste para obtener fenotipos más corregidos. Sin embargo, los datos utilizados no contenían suficiente información para considerar este efecto. Además, en el modelo de ajuste el número de observaciones por nivel en cada factor ambiental fue bajo como fue el caso del efecto de GC. Considerando únicamente los datos de la población genotipificada, se observaron más de 100 niveles de GC con menos de seis observaciones.

Otro factor limitante en este estudio fue que los animales genotipificados fueron en su mayoría hembras, seleccionadas con base a la participación de los productores y a la disponibilidad de datos fenotípicos, en lugar de realizar una selección de sementales con alta exactitud de EBV. El uso de PR representadas por sementales altamente probados sería una mejor opción para investigar el impacto de la información genómica en los modelos de PG. En México, el número total de sementales de la raza Charolais con 10 o más progenie ha ascendido a 3,879 de acuerdo con la última evaluación genética de esta raza (Sumario de



sementales Charolais y Charbray, 2017). En el futuro, la construcción de una adecuada PR para un programa de selección genómica de mayor magnitud debería necesariamente considerar la identificación de sementales altamente probados y la inclusión de un mayor número de productores.

A pesar de que se han realizado muchos estudios localmente demostrando la importancia del registro de pedigrís, el impacto de utilizar EBV como herramienta de selección (Parra-Bracamonte *et al.* 2016) y el potencial del uso de marcadores genéticos para mejorar características productivas, los resultados de estas investigaciones aún no han sido puestos en práctica. Desafortunadamente, la cultura de cría de ganado en México todavía sigue siendo dominada por algunas tradiciones como son la selección de animales en base a aspectos fenotípicos observables y esquemas de reproducción que no toman en cuenta la importancia genética de los animales. Por lo tanto, se requiere de la implementación de esquemas de reproducción y técnicas de selección efectivas que utilicen o no información genómica. Para esto, una tarea de los investigadores sería implementar cursos de capacitación que introduzcan a los productores en los conceptos generales de la selección animal y proporcionar programas estadísticos útiles para el registro, manejo y simulación de procedimientos donde los productores puedan observar fácilmente estadísticas de sus hatos e indicadores acerca de posibles cambios de acuerdo con ciertas condiciones de selección y manejo. De esta manera, los productores podrían reconocer la importancia del correcto registro de pedigrís e información fenotípica o incluso la importancia de la información genómica.

El registro de características productivas en ganado de carne es de gran importancia en términos de realizar PG. La PG no puede ser utilizada si no existen los datos fenotípicos en la PR y no habría ninguna ventaja de genotipificar animales si tampoco existe un objetivo de cría así como los correspondientes registros fenotípicos. Para la raza Charolais, solo ocho características productivas (pesos corporales, talla, circunferencia escrotal, permanencia productiva) son consideradas para evaluación genética. Por lo tanto, solo para estas características sería posible estimar GEBV si se implementara la PG en esta raza. Desafortunadamente, muchas características importantes que impactan en la producción, por ejemplo eficiencia alimenticia, temperamento, resistencia a parásitos o a diversos



climas, no son registradas en las razas de ganado bovino para carne en México y es precisamente en este tipo de características donde la PG puede ser de mayor beneficio ya que son características de baja heredabilidad y difíciles, o costosas de medir.

El costo de genotipificación seguirá siendo un factor determinante en la implementación de la PG debido a la alta inversión económica que representa para los productores en México. Actualmente, en México no existen compañías que proporcionen servicios de genotipificación de especies de importancia doméstica, y como consecuencia se debe acudir a servicios de compañías en el extranjero, siendo el costo una limitante importante. Experiencias y estrategias de otros proyectos genómicos podrían ser consideradas en futuras aplicaciones de PG para ayudar a minimizar el costo de genotipificación y maximizar la exactitud de predicción. Por ejemplo, la imputación de genotipos de animales no genotipificados podría realizarse para incrementar el tamaño de la PR (Pimmentel *et al.* 2013). Además, el genotipo de sementales probados por progenie puede ser predicho de manera exacta mediante la estructuración de una base de datos que incluya el genotipo de familias de medios hermanos y mediante el uso de programas de imputación como *hspase* (Ferdosi *et al.* 2014). Siguiendo esta estrategia, fue posible predecir el genotipo de 66 sementales no genotipificados a partir de la información genotípica de 636 animales genotipificados. Se observó que el genotipo del semental puede ser imputado de manera exacta hasta 66 mil SNP teniendo al menos 15 crías genotipificadas por semental (datos no mostrados). Otras estrategias como la imputación de baja a alta densidad de SNP también podrían implementarse (Huang *et al.* 2012).

CONCLUSIÓN

En este estudio se demostró que es posible la implementación de métodos de PG para características de peso corporal en ganado Charolais pero sin ningún incremento en la exactitud de predicción del mérito genético para estas características. Este resultado es debido principalmente a lo reducido de la PR y el uso de fenotipos ajustados en lugar de EBV desregresados de sementales con un gran número de progenie. Se espera que en un futuro se pueda tener acceso a la información de EBV y exactitudes recientes para evaluar



EBV desregresados como variables de respuesta y considerar los genotipos imputados de
sementales en los modelos de PG.



LITERATURA CITADA

breedR: Statistical Methods for Forest Genetic Resources Analysis. R package version 0.11-5. <https://github.com/famuvie/breedR> (Accessed 20 February 2017)

Ferdosi M.H., Kinghorn B.P., van der Werf J.H., Lee S.H., Gondro C. 2014. hspase: an R package for pedigree reconstruction, detection of recombination events, phasing and imputation of half-sib family groups. *BMC Bioinformatics*. 15:172.

Fernando R.L., Dekkers J.C.M., Garrick D.J. 2014. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. *Genet. Sel. Evol.* 46:50.

García-Ruiz A., Cole J.B., VanRaden P.M., Wiggans G.R., Ruiz-López F.J., Van Tassell C.P. 2016. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proc. Natl. Acad. Sci. U S A*. 113(28):E3995-4004.

Gunia M., Saintilan R., Venot E., Hozé C., Fouilloux M.N., Phocas F. 2014. Genomic prediction in French Charolais beef cattle using high-density single nucleotide polymorphism markers. *J. Anim. Sci.* 92(8):3258-3269.

Habier D., Fernando R.L., Kizilkaya K., Garrick D.J. 2011. Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics*. 12:186.

Huang Y., Hickey J.M., Cleveland M.A., Maltecca C. 2012. Assessment of alternative genotyping strategies to maximize imputation accuracy at minimal cost. *Genet. Sel. Evol.* 44:25.

Jahuey-Martínez F.J., Parra-Bracamonte G.M., Sifuentes-Rincón A.M., Martínez-González J.C., Gondro C., García-Pérez C.A., López-Bustamante L.A. 2016. Genomewide association analysis of growth traits in Charolais beef cattle. *J. Anim. Sci.* 94(11):4570-4582.

JWAS (Julia implementation of whole-genome analyses software using univariate and multivariate Bayesian mixed effects model). <http://QTL.rocks> (Accessed 20 February 2017)



Lee J., Cheng H., Garrick D., Golden B., Dekkers J., Park K., Lee D., Fernando R. 2017. Comparison of alternative approaches to single-trait genomic prediction using genotyped and non-genotyped Hanwoo beef cattle. *Genet. Sel. Evol.* 49(1):2.

Legarra A., Aguilar I., Misztal I. 2009. A relationship matrix including full pedigree and genomic information. *J. Dairy. Sci.* 92:4656-4663.

Saatchi M., McClure M.C., McKay S.D., Rolf M.M., Kim J., Decker J.E., Taxis T.M., Chapple R.H., Ramey H.R., Northcutt S.L., Bauck S., Woodward B., Dekkers J.C.M., Fernando R.L., Schnabel R.D., Garrick D.J., Taylor J.F. 2011. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. *Genet. Sel. Evol.* 43(1): 40.

Meuwissen T., Hayes B., Goddard M. 2013. Accelerating improvement of livestock with genomic selection. *Annu. Rev. Anim. Biosci.* 1:221-37.

Meuwissen T.H., Hayes B.J., Goddard M.E.. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics.* 157(4):1819-1829.

Parra-Bracamonte G.M., Lopez-Villalobos N., Morris S.T., Sifuentes-Rincón A.M., Lopez-Bustamante L.A. 2016. Genetic trends for live weight traits reflect breeding strategies in registered Charolais Farms in Mexico. *Trop. Anim. Health Prod.* 48(8):1729-1738.

Pérez P., de los Campos G. 2014. Genome-Wide Regression and Prediction with the BGLR Statistical Package. *Genetics.* 198(2): 483-495.

Pimentel E.C., Wensch-Dorendorf M., König S., Swalve H.H. 2013. Enlarging a training set for genomic selection by imputation of un-genotyped animals in populations of varying genetic architecture. *Genet. Sel. Evol.* 45:12.

Saatchi M., Ward J., Garrick D.J. 2013. Accuracies of direct genomic breeding values in Hereford beef cattle using national or international training populations. *J. Anim. Sci.* 91(4):1538-1551.

Saatchi M., McClure M.C., McKay S.D., Rolf M.M., Kim J., Decker J.E., Taxis T.M., Chapple R.H., Ramey H.R., Northcutt S.L., Bauck S., Woodward B., Dekkers J.C.M.,



Fernando R.L., Schnabel R.D., Garrick D.J., Taylor J.F. 2011. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. *Genet. Sel. Evol.* 43(1): 40.

Saatchi M., Schnabel R.D., Rolf M.M., Taylor J.F., Garrick D.J. 2012. Accuracy of direct genomic breeding values for nationally evaluated traits in US Limousin and Simmental beef cattle. *Genet. Sel. Evol.* 44(1): 38.

Sumario de sementales Charolais y Charbray 2017. <http://charolais.org.mx/> (Accessed 15 august 2017)

USDA. 2017. Livestock and Poultry: World Markets and Trade. https://apps.fas.usda.gov/psdonline/circulars/livestock_poultry.pdf (Accessed 30 July 2017)



CONCLUSIONES Y RECOMENDACIONES

El mapeo genético realizado en este trabajo mediante análisis de asociación (GWAS) e identificación de señales de selección permitió definir regiones genómicas y genes (TRAF6, CDH11, KLF7, MIR181A-1, PRCP, LCOLL, CAV1, CAV2, MDFIC, IFRD1, ZNF608) funcionalmente asociados con características de crecimiento en ganado bovino. Futuros estudios de exploración y validación serán necesarios para lograr la identificación de las variaciones genéticas que influyen directamente en características de crecimiento. Por lo tanto, este trabajo abre las puertas para el estudio de regiones específicas del genoma para la búsqueda de mutaciones que influyan características de importancia productiva en ganado Charolais. Las regiones genómicas propuestas en este trabajo podrían ser exploradas primeramente mediante análisis *in silico* (computacional) y posteriormente mediante una estrategia de genes candidatos.

Con respecto al uso de la información genómica para la evaluación genética de la raza Charolais, el presente trabajo demuestra que la población utilizada no cumple con las características adecuadas tanto en número como en estructura y, por lo tanto, fue imposible observar las ventajas de la predicción genómica en la población de estudio. A medida que sea posible la construcción de poblaciones de referencia en México, los datos generados en este trabajo podrán ser utilizados para evaluar otros escenarios de predicción evaluando nuevas variables de respuesta, diferentes modelos estadísticos, de validación, etc., que ayuden a identificar la mejor estrategia para el uso de la información genómica en la predicción de características de importancia productiva. Para una futura construcción de una población de referencia, se recomienda realizar previamente un estudio extenso de los datos productivos y genealógicos de la raza Charolais para lograr la identificación de los sementales más representados y con mejor calidad de datos en las poblaciones de México. Actualmente, la asociación de ganado Charolais en México cuenta con un inventario de casi cuatro mil sementales representados en sus datos, por lo que esta población de sementales pudiera ser útil para comenzar un proyecto nacional de selección genómica. Por supuesto, la definición clara y precisa de los objetivos de selección y el registro de nuevas características fenotípicas serán de suma importancia lo cual ayudaría a expandir y explotar el potencial de la raza Charolais para la producción (y posiblemente calidad) de carne.



APÉNDICES

Códigos correspondientes al capítulo I

leer archivos de genotipos y fenotipos en GenABEL

```
library(GenABEL)
```

```
genabel_feno <- read.table("genabel_feno.txt", header=T, sep="\t", na.strings="NA",  
dec=".", strip.white=TRUE)
```

```
convert.snp.illumina(infile="genabel_genotipos.txt", out="genabelgenotipos855.raw", strand="+")
```

```
genabelgrowth855 <- load.gwaa.data(phe="genabel_feno.txt", gen="genabelgenotipos855.raw", force=TRUE,  
sort=TRUE, id="id")
```

realizar control de calidad

```
qc1 <- check.marker(genabelgrowth855, callrate = 0.90, perid.call=0.80, extr.call = 0.1, extr.perid.call = 0.1,  
het.fdr = 0.01, ibs.threshold = 0.95, ibs.mrk = 7000, ibs.exclude="both", maf= 0.05, p.level = 0, odds = 1000,  
, redundant = "no", minconcordance = 2.0, qoption = "bh95", imphetasmising = TRUE,  
XXY.call=0.8, intermediateXF = c(0.5, 0.5))
```

```
summary(qc1)
```

```
genabelgrowth855 <- genabelgrowth855[qc1$idok, qc1$snpok]
```

calcular relaciones genómicas

```
genabelgrowth855.gkin <- ibs(genabelgrowth855[, autosomal(genabelgrowth855)], weight="freq")
```

realizar escalamiento multidimensional

```
genabelgrowth855.dist <- as.dist(0.5-genabelgrowth855.gkin)
```

```
genabelgrowth855.mds <- cmdscale(genabelgrowth855.dist)
```

```
plot(genabelgrowth855.mds)
```

realizar agrupamiento kmeans

```
km <- kmeans(genabelgrowth855.mds, centers=2, nstart=1000)
```

realizar análisis de asociación mediante el método egscore de GenABEL

```
diag(genabelgrowth855.gkin) <- hom(genabelgrowth855[, autosomal(genabelgrowth855)])$Var
```



```
# análisis de peso al nacimiento (residuales). Para analizar otros fenotipos solo se cambia el nombre de la variable (p.e. reemplazar PN_res por PD_res)
```

```
GWAS_PN <- egsscore(PN_res, genabelgrowth855, kinship=genabelgrowth855.gkin)
```

```
# generar gráfico de manhattan
```

```
plot(GWAS_PN, df="Pc1df", col=c("black", "darkgray"), cex=0.8, pch=20, ylim=c(1,6), main="")
```

```
abline(h=-log10(0.00005), col="black", lty=1, lwd=1.5)
```

```
##### Anotación de GWAS
```

```
library(postgwas)
```

```
library(org.Bt.eg.db)
```

```
# reemplazar sobre cada una de las listas de "biomartConfigs"
```

```
for(i in seq_along(biomartConfigs)){
```

```
  biomartConfigs[[i]]$snp$host = "www.ensembl.org"
```

```
  biomartConfigs[[i]]$snp$mart<-"ENSEMBL_MART_SNP"
```

```
  biomartConfigs[[i]]$gene$host<-"www.ensembl.org"
```

```
  biomartConfigs[[i]]$gene$mart<-"ENSEMBL_MART_ENSEMBL"
```

```
  biomartConfigs[[i]]$path$host<-"www.ensembl.org"
```

```
  biomartConfigs[[i]]$proteincomplex$host<-"www.ensembl.org" }
```

```
# elegir columnas SNP, Pc1df y CHR del objeto GWAS_PN y crear un dataframe
```

```
GWAS_PN_DF<-data.frame(SNP=snpnames(GWAS_PN), P= GWAS_PN[, "Pc1df"], BP= GWAS_PN[, "Position"], CHR=chromosome(GWAS_PN))
```

```
# utilizar el dataframe anterior y lo convierte a formato PLINK
```

```
PLINK_PN<- postgwas:::readGWASdatasets(GWAS_PN_DF)
```

```
# del dataframe seleccionar solo los SNP de interés en base al P-value
```

```
SNP_PN<- removeNeighborSnps(PLINK_PN[PLINK_PN$P<= 5*10^-5, ], maxdist=100000)
```

```
# extraer los genes más cercanos a cada SNP de interés (requiere conexión a internet)
```

```
GENC_PN<- snp2gene.prox(SNP_PN, level=0, use.buffer=TRUE, biomart.config=biomartConfigs$btaurus)
```



```
# extraer los genes localizados dentro de una ventana de 1Mb alrededor de cada SNP de
# interés y calcula valores de desequilibrio de ligamiento (requiere conexión a internet)
```

```
GENLD_PN<- snp2gene.LD(SNP_PN, use.buffer=T, biomart.config=biomartConfigs$btaurus,
ld.win=1000000, by.genename=T, gts.source=gtdata(genabelgrowth855))
```

```
##### crear archivos PED y MAP de PLINK manualmente desde GenABEL
```

```
# este método fue tomado del siguiente link: http://rstudio-pubs-static.s3.amazonaws.com/10819\_cd03428731b74392bcabd16cef8bc485.html
```

```
# Archivo Ped
```

```
ped<-data.frame(FID=1:nids(genabelgrowth855), IID=idnames(genabelgrowth855), PAT=paste("S_",
phdata(genabelgrowth855)[,"Padres_NoPriv_Reg"], sep=""), MAT=paste("S_",
phdata(genabelgrowth855)[,"Madres_NoPriv_Reg"], sep=""), SEX=phdata(genabelgrowth855)[,"sex"],
PHE=rep(0, nids(genabelgrowth855)))
```

```
ped$PAT[ped$PAT=="S_NA"]<-NA
```

```
ped$SEX[ped$SEX==0]<-2
```

```
# exportar archivos de genotipos por cromosoma
```

```
for(i in 1:29){
```

```
chr<-data.frame(as.character(genabelgrowth855[,genabelgrowth855@gtdata@chromosome==i]))
```

```
chr <- lapply(chr,
```

```
function(x) {
```

```
my.df <- data.frame(one = sapply(strsplit(as.character(x), split="/"), "[", 1),
```

```
two = sapply(strsplit(as.character(x), split="/"), "[", 2)))
```

```
chr <- cbind(as.data.frame(chr))
```

```
chr<-cbind(ped,chr)
```

```
write.table(chr, paste("chr_", i, ".ped", sep=""), sep="\t", row.names=F, col.names=F, quote=F, na="0");
rm(chr)}
```

```
# crear archivo MAP de cada cromosoma
```

```
for(i in 1:29){
```

```
map<-data.frame(chr=rep(i, nsnps(genabelgrowth855[,genabelgrowth855@gtdata@chromosome==i])),
```

```
snps=names(map(genabelgrowth855[,genabelgrowth855@gtdata@chromosome==i])),
```

```
gdist=rep(0, nsnps(genabelgrowth855[,genabelgrowth855@gtdata@chromosome==i])),
```



```
pos=map(genabelgrowth855[,genabelgrowth855@gtdata@chromosome==i])  
write.table(map, paste("chr_", i, ".map", sep=""), sep="\t", row.names=F, col.names=F, quote=F); rm(map)}
```



Códigos correspondientes al capítulo II

Faseo de genotipos en el programa SHAPEIT2 (Linux). La población de estudio fue considerada como no relacionada

Faseo del cromosoma 1. Repetir para cada autosoma

```
shapeit --input-ped chr_1.ped chr_1.map \  
-O chr_1.phased \  
--effective-size 300 \  
--noped \  
--thread 10
```

convertir datos faseados de SHAPEIT2 a archivos de entrada de la librería GHap de R

```
tail -n +3 chr_29.phased.sample | cut -d' ' -f1,2 > chr_29.samples  
cut -d' ' -f1-5 chr_29.phased.haps > chr_29.markers  
cut -d' ' -f1-5 --complement chr_29.phased.haps > chr_29.phase
```

Análisis de identificación de señales de selección

```
library(rehh)
```

leer archivos "map" y cambiar el orden de las columnas 1 y 2 de acuerdo con el archivo de entrada del paquete rehh de R

```
for(i in 1:29){  
  snp_info<-read.table(paste("chr_", i, ".markers", sep=""))  
  snp_info<-snp_info[,c(2, 1, 3, 4, 5)]  
  write.table(snp_info, paste("chr_", i, ".markers" ,sep=""), sep="\t", row.names=F, col.names=F, quote=F)  
  rm(snp_info)}
```

crear el map file de todos los cromosomas

```
library(readbulk)  
snp_info<- read_bulk(extension=".markers", fun=read.table, stringsAsFactor=FALSE)  
str(snp_info)  
snp_info<- snp_info[order(snp_info$V2, decreasing=F),]
```



```
snp_info<- snp_info[,1:5]
write.table(snp_info, "snp_info", sep="\t", row.names=F, col.names=F, quote=F)

# transformar los genotipos 0 y 1 a alelos A,G,C,T de acuerdo con las columnas a0 y a1
(columnas 4 y 5 de DF info)

# leer los archivos de salida ".phased_haps" de SHAPEIT2 y reescribirlos con el nuevo
formato

for(i in 1:29){

chr<-read.table(paste("chr_", i, ".phased.haps", sep=""), header=F, sep=" ")
info<-chr[,1:5]
info$V4<-as.character(info$V4)
info$V5<-as.character(info$V5)
chr<-chr[,6:ncol(chr)]
chr <- apply(chr, 2, function(x) {x[x==0] <- info$V4[x==0]; x})
chr <- apply(chr, 2, function(x) {x[x==1] <- info$V5[x==1]; x})
write.table(chr, paste("chr_", i, ".phased.haps", sep=""), sep="\t", row.names=F, col.names=F, quote=F)}

#####  Análisis mediante el método integrated haplotype score dentro de una población

for(i in 1:29){

datos<-data2haplohh(hap_file=paste("chr_", i, ".phased.haps", sep=""), map_file=paste("chr_", i, ".markers",
sep=""), haplotype.in.columns=TRUE, recode.allele=TRUE, chr.name=i)

res<-scan_hh(datos)

if(i==1){wg.res<-res}else{wg.res<-rbind(wg.res, res)}

wg.ihs<-ihh2ihs(wg.res)

head(wg.ihs$iHS)

head(wg.ihs$frequency.class)

# gráfico de Manhattan

ihplot(wg.ihs, plot.pval=TRUE, ylim.scan=4.5, main="")

# los valores de P ya estan transformados a log10

head(wg.ihs$iHS[order(wg.ihs$iHS`-log10(p-value)`, decreasing=T),], n=10)
```



```
# generar gráficos de bifurcación alrededor de un SNP (p.e. SNP rs132673479)

chr<-data2haplohh(hap_file="chr_23.phased.haps", map_file="chr_23.markers",
haplotype.in.columns=TRUE, recode.allele=TRUE, chr.name=23)

layout(matrix(1:2,2,1))

which(grepl("rs132673479", chr@snp.name))

bifurcation.diagram(chr, mrk_foc=3684, all_foc=1, nmrk_l=20, nmrk_r=20, main="Bifurcation diagram
(rs29024564 SNP on BTA1): Ancestral Allele")

bifurcation.diagram(chr, mrk_foc=3684, all_foc=2, nmrk_l=20, nmrk_r=20, main="Bifurcation diagram
(rs29024564 SNP on BTA1): Derived Allele")
```

Anotación de las señales de selección

```
library(postgwas)
library(org.Bt.eg.db)
```

seleccionar solo SNP que tienen P-value

```
GWIS_DF<-wg.ihs$iHS[!is.na(wg.ihs$iHS$`-log10(p-value)`),]
```

seleccionar columnas y convertir los valores de P

```
GWIS_DF<-data.frame(SNP=rownames(GWIS_DF), P=1*10^-GWIS_DF[,4],
BP=GWIS_DF[, "POSITION"], CHR=GWIS_DF[, "CHR"])
```

```
GWIS_DF<- postgwas:::readGWASdatasets(GWIS_DF)
```

```
save(GWIS_DF, file = "GWIS_DF.RData")
```

```
load("GWIS_DF.RData")
```

crear un dataframe eligiendo los SNP de interés

```
SNP_IS<- removeNeighborSnps(GWIS_DF[GWIS_DF$P<= 5*10^-6, ], maxdist=10)
```

```
save(SNP_IS, file = "SNP_IS.RData")
```

cargar en el sitio de trabajo donde se encuentran depositados los genotipos (análisis en GenABEL)

```
load("SNP_IS.RData")
```

extraer los genes más cercanos a cada SNP de interés (requiere conexión a internet)



```
GENC_IS<- snp2gene.prox(SNP_IS,level=0, use.buffer=TRUE, biomart.config=biomartConfigs$btaurus)

# extraer los genes localizados dentro de una ventana de 1Mb alrededor de cada SNP de
interés y calcula valores de desequilibrio de ligamiento (requiere conexión a internet)

GENLD_IS<- snp2gene.LD(SNP_IS, use.buffer=T, biomart.config=biomartConfigs$btaurus,
ld.win=2000000, by.genename=T, gts.source=gtdata(genabelgrowth855))

# cargar nuevamente los resultados del análisis de señales de selección

load("res_SS.RData")

GWIS_DF<-wg.ihs$iHS[!is.na(wg.ihs$iHS)`-log10(p-value)`],]
rownames(GWIS_DF)<-GWIS_DF$SNP
names(GWIS_DF)[c(4,2)]<-c("Chromosome", "P1df")
GWIS_DF$SNP<-as.character(GWIS_DF$SNP)
GWIS_DF$Chromosome<-as.character(GWIS_DF$Chromosome)
head(GWIS_DF[order(GWIS_DF$P1df, decreasing=F),], n=10)

# identificar grupos de SNP en DL con SNP significativos y guardar resultado en dataframe
SS_clumps<-vector(mode="list", length=length(SS_sign$chr))
names(SS_clumps)<-SS_sign$chr
for(i in SS_sign$chr){
  borrar<-clump.markers2(data=genabelgrowth855, gwas.result=GWIS_DF, chr=i, r2=0.2, bp.dist=1000000,
p1=1*10^-4.5)
  SS_clumps[names(SS_clumps)==i]<-list(borrar)
  rm(borrar)}

SS_clumps_DF<-data.frame()
for(i in 1:length(SS_clumps)){
  gr<-1:length(SS_clumps[[i]])
  for(j in gr){
    borrar<-data.frame(chr=head(SS_clumps[[i]][j]$chr, n=1), SNP=names(SS_clumps[[i]][j]),
Start=head(SS_clumps[[i]][j]$coord, n=1),End=tail(SS_clumps[[i]][j]$coord, n=1),
nSNP=length(SS_clumps[[i]][j]$coord), SNP_clump=min(SS_clumps[[i]][j]$pval))
    SS_clumps_DF <- rbind(SS_clumps_DF, borrar)
  }
  rm(borrar)}}}
```



```
# definir región extendida en cada agrupamiento

SS_sign<-data.frame(chr=rep(NA, 5), Start=rep(NA, 5), End=rep(NA, 5), size=rep(NA, 5), nSNPs=rep(NA,
5))

SS_sign$chr<-c(4,1,6,7,14)
for(i in c(4,1,6,7,14)){
DF<-SS_clumps_DF[SS_clumps_DF$chr==i,]
SS_sign$Start[SS_sign$chr==i]<-min(DF$Start)-1000000
SS_sign$End[SS_sign$chr==i]<-max(DF$End)+1000000
SS_sign$size[SS_sign$chr==i]<-((max(DF$End)+1000000)-(min(DF$Start)-1000000))/1000}

##### Comparación de las señales de selección con QTL de bovino

# descargar archivo ".gff" de QTLdb de bovinos

# en EXCEL eliminar filas de comentarios y asignar valores de 0 y 1 pb a QTL con
posiciones faltantes. Tener cuidado con las posiciones que si tienen el inicio o el final.

# realizar ordenaciones en base al inicio y al final de cada QTL para identificar regiones
que si tienen el inicio o el final.

# en el editor de CRIMSON eliminar los caracteres de comillas "

# analisis del archivo ".gff" obtenido de QTLdb bovino

library(rtracklayer)
QTLdb<-(import("QTL_UMD_3.1_rel_34_2.gff"))

# crear un archivo de texto y nombrarlo "QTLdb_T.txt". Este archivo debe contener las
columnas Trait_type, Trait_subtype, Trait, Abbrev, Phenotype que corresponden al tipo,
subtipo, característica, abreviación y traducción de cada una de las características
encontradas en la QTLdb de bovino.

# lee los nombres y abreviaturas de QTL

QTLdb_T<-read.table("QTLdb_T.txt", sep="\t", header=T, na.strings="NA", quote="")

# cambiar todas las columnas a caracteres

QTLdb_T[sapply(QTLdb_T, is.factor)]<-lapply(QTLdb_T[sapply(QTLdb_T, is.factor)], as.character)

# funcion para eliminar espacios en blanco iniciales y finales

trim <- function( x ) {gsub("^[[:space:]]+|[[:space:]]+$", "", x)}
```



```
# elimina el primer y último espacio en cada string
QTLdb_T$Trait<-trim(QTLdb_T$Trait)
QTLdb_T$Trait_type<-trim(QTLdb_T$Trait_type)
QTLdb_T$Trait_subtype<-trim(QTLdb_T$Trait_subtype)

# convertir primera letra a mayúsculas
simpleCap <- function(x) {paste(toupper(substring(x, 1, 1)), substring(x, 2), sep="", collapse=" ")}
QTLdb_T$Trait<-sapply(QTLdb_T$Trait, simpleCap)

# asignar nuevos nombres, tipos y subtipos a la QTLdb de bovino
QTLdb@elementMetadata$Trait2<-as.character(sapply(QTLdb@elementMetadata$Trait, paste, collapse=","))
QTLdb@elementMetadata$Trait2<-trim(QTLdb@elementMetadata$Trait2)

# convertir primera letra a mayusculas
QTLdb@elementMetadata$Trait2<-sapply(QTLdb@elementMetadata$Trait2, simpleCap)

# checar cuales características no están en la tabla total de características (QTLdb_T)
unique(QTLdb@elementMetadata$Trait2)[!(unique(QTLdb@elementMetadata$Trait2) %in%
QTLdb_T$Trait)]

# asignar tipos, subtipos y traducción a QTLdb
QTLdb@elementMetadata$Trait_type<-QTLdb_T$Trait_type[match(QTLdb@elementMetadata$Trait2,
QTLdb_T$Trait, nomatch=NA_integer_)]
QTLdb@elementMetadata$Trait_subtype<-
QTLdb_T$Trait_subtype[match(QTLdb@elementMetadata$Trait2,QTLdb_T$Trait,nomatch=NA_integer_)]
QTLdb@elementMetadata$Phenotype<-
QTLdb_T$Phenotype[match(QTLdb@elementMetadata$Trait2,QTLdb_T$Trait,nomatch=NA_integer_)]

# convertir QTLdb a dataframe
QTLdb<-as.data.frame(QTLdb)

# identificar QTL que coinciden con las señales de selección
traits<-c("Conformation","Anatomy","Fatness","Feed conversion","Growth","Fertility","General")
SS_QTLs<-data.frame()
for(i in SS_sign$chr){
SS<-SS_sign[SS_sign$chr==i,]
```



```
QTL<-QTLdb[QTLdb$Chr==i & QTLdb$Trait_subtype %in% traits,]
QTL<-QTL[!is.na(QTL$Start),]
match_QTL<-QTL[(QTL$Start >= SS$Start & QTL$End <= SS$End) | (QTL$Start <= SS$Start &
(QTL$End >= SS$Start & QTL$End <= SS$End)) | ((QTL$Start >= SS$Start & QTL$Start <= SS$End) &
QTL$End >= SS$End) | (SS$Start >= QTL$Start & SS$End <= QTL$End),]
SS_QTLs<-rbind(SS_QTLs, match_QTL)
rm(SS,QTL,match_QTL)}
SS_QTLs<-SS_QTLs[rowSums(is.na(SS_QTLs))!=ncol(SS_QTLs),]
```

Checar las anotaciones genómicas disponibles en UCSC

```
library(rtracklayer)
```

lista de genomas en UCSC

```
ucscGenomes()
```

```
library(biomaRt)
```

conectarse a ENSEMBL

```
ensembl<- useEnsembl(biomart="ensembl")
```

lista de genomas en ENSEMBL

```
head(listDatasets(ensembl))
```

Descargar la anotación del genoma de interés

extraer anotación utilizando BiomaRt

```
library(biomaRt)
```

conectarse a ENSEMBL

```
ensembl<- useEnsembl(biomart="ensembl")
```

lista de especies en ENSEMBL

```
head(listDatasets(ensembl))
```

información disponible en ENSEMBL

```
head(listAttributes(ensembl), 75)
```

seleccionar la especie de interes

```
ensembl_Bt<- useEnsembl(biomart="ensembl", dataset="btaurus_gene_ensembl")
```



```
Bt_UMD_3_1<- getBM(attributes=c("entrezgene", "hgnc_symbol", "chromosome_name", "start_position",
"end_position", "ensembl_gene_id", "ensembl_transcript_id", "ensembl_peptide_id", "description", "strand"),
mart=ensembl_Bt)

Bt_UMD_3_1<-Bt_UMD_3_1[Bt_UMD_3_1$chromosome %in% c(1:29, "X"),]

# asignar nombres a columnas
names(Bt_UMD_3_1)[1:5]<-c("geneid", "genename", "chr", "start", "end")

# llenar espacios vacios con "NA"
is.na(Bt_UMD_3_1) <- Bt_UMD_3_1=="

# función para contar el numero de NA por filas
count_NA_byrows<-function(datos){sum(is.na(datos))}

# crear columna con el número de NAs por filas
Bt_UMD_3_1$num_NA<-apply(Bt_UMD_3_1,1,count_NA_byrows)

# ordenar en base al num de NAs en cada fila
Bt_UMD_3_1<-Bt_UMD_3_1[order(Bt_UMD_3_1$num_NA, decreasing=F),]

# eliminar genes con posiciones repetidas.
Bt_UMD_3_1<-Bt_UMD_3_1[!duplicated(Bt_UMD_3_1[,c("start","end","chr")]),]
Bt_UMD_3_1$num_NA<-NULL

library(GenomicFeatures)

# extraer anotación de Biomart

# crear objeto TxDb
BT_GF1<- makeTxDbFromBiomart(dataset="btaurus_gene_ensembl")

# convertir TxDb a lista de dataframe
BT_GF1_list<-as.list(BT_GF1)

# unir "transcripts" y "genes"
BT_GF1_DF<- cbind(BT_GF1_list$transcripts,BT_GF1_list$genes)
BT_GF1_DF<- BT_GF1_DF[BT_GF1_DF$tx_chrom %in% c(1:29,"X"),]

# eliminar genes repetidos
BT_GF1_DF<-BT_GF1_DF[!duplicated(BT_GF1_DF[,c("tx_start","tx_end"),]),]
BT_GF1_DF<-BT_GF1_DF[!duplicated(BT_GF1_DF[,c("gene_id"),]),]
```



```
# identificar genes que coinciden con las señales de selección
```

```
SS_genes1<-data.frame()
for(i in SS_sign$chr){
SS<-SS_sign[SS_sign$chr==i,]
QTL<-Bt_UMD_3_1[Bt_UMD_3_1$chr==i,]
QTL<-QTL[!is.na(QTL$start),]
match_QTL<-QTL[(QTL$start >= SS$start & QTL$end <= SS$end) | (QTL$start <= SS$start & (QTL$end
>= SS$start & QTL$end <= SS$end)) | ((QTL$start >= SS$start & QTL$start <= SS$end) & QTL$end >=
SS$end) | (SS$start >= QTL$start & SS$end <= QTL$end),]
SS_genes1<-rbind(SS_genes1,match_QTL)
rm(SS,QTL,match_QTL)}
```

```
SS_genes2<-data.frame()
for(i in SS_sign$chr){
SS<-SS_sign[SS_sign$chr==i,]
QTL<-BT_GF1_DF[BT_GF1_DF$tx_chrom==i,]
QTL<-QTL[!is.na(QTL$tx_start),]
match_QTL<-QTL[(QTL$tx_start >= SS$start & QTL$tx_end <= SS$end) | (QTL$tx_start <= SS$start &
(QTL$tx_end >= SS$start & QTL$tx_end <= SS$end)) | ((QTL$tx_start >= SS$start & QTL$tx_start <=
SS$end) & QTL$tx_end >= SS$end) | (SS$start >= QTL$tx_start & SS$end <= QTL$tx_end),]
SS_genes2<-rbind(SS_genes2,match_QTL)
rm(SS,QTL,match_QTL)}
SS_genes2<-SS_genes2[!duplicated(SS_genes2$gene_id),]
```

```
# unir nombres y entrez a genes
```

```
SS_genes2<-cbind(SS_genes2, SS_genes1[match(SS_genes2$gene_id,
SS_genes1$ensembl_gene_id,nomatch=NA_integer_), c("geneid", "genename")],
eg[match(SS_genes2$gene_id, eg$ENSEMBL, nomatch=NA_integer_),])
```

```
##### asignar genes a SNP utilizando el procedimiento de la función "snp2gene" de
postgwas. Obetener genes para cada SNP de todo el panel (dentro de una ventana + el más
cercano)
```

```
DFann<-BT_GF1_DF[BT_GF1_DF$chr %in% c(1:29),] # Anotacion del genoma
```

```
DFsnp<-GWIS_DF # Panel de SNP con sus posiciones
```



```
nchr=29

ld_win=1500000 # ventana en pb

snp2geneWIN<-vector(mode="list", length=nchr)

for(i in 1:29){
  CHRgen<-DFann[DFann$chr == i,] # selecciona genes en cromosoma i
  CHRsnp<-DFsnp[DFsnp$CHR == i,] # selecciona SNP en cromosoma i
  # extraer genes para cada SNP
  for(g in seq_along(CHRsnp$SNP)) # por cada SNP
    nearest_bp <- min(abs(CHRgen$start - CHRsnp[g,"BP"]), abs(CHRgen$end - CHRsnp[g,"BP"]))
    wind<-c((CHRsnp[g,"BP"] - ld_win/2),(CHRsnp[g,"BP"] + ld_win/2))

  snp2geneWIN[[i]][[g]]<-unique(rbind(
    # genes dentro del rango "ld_win" alrededor del SNP
    CHRgen[(CHRgen$start >= wind[1] & CHRgen$end <= wind[2]) | (CHRgen$start <= wind[1] &
    (CHRgen$end >= wind[1] & CHRgen$end <= wind[2])) | ((CHRgen$start >= wind[1] & CHRgen$start <=
    wind[2]) & CHRgen$end >= wind[2]) | (wind[1] >= CHRgen$start & wind[2] <= CHRgen$end),,
    # gen cercano a SNP
    CHRgen[abs(CHRgen$start - CHRsnp[g,"BP"]) == nearest_bp | abs(CHRgen$end - CHRsnp[g,"BP"]) ==
    nearest_bp,] ))
  snp2geneWIN[[i]][[g]]$SNP<-CHRsnp[g,"SNP"]
  snp2geneWIN[[i]][[g]]$gene_loc<-NA
  snp2geneWIN[[i]][[g]]$bp_dist<-NA
  # Asignar localizacion y distancia del gen respecto al SNP
  for(j in 1:nrow(snp2geneWIN[[i]][[g]])){
    # SNP dentro de genes
    if (snp2geneWIN[[i]][[g]][j, "start"] <= CHRsnp[g, "BP"] & snp2geneWIN[[i]][[g]][j, "end"] >= CHRsnp[g,
    "BP"])
      {snp2geneWIN[[i]][[g]][j, "gene_loc"] <- "within"
      snp2geneWIN[[i]][[g]][j, "bp_dist"] <- 0}
    # SNP rio arriba de genes
    else if (snp2geneWIN[[i]][[g]][j, "start"] <= CHRsnp[g, "BP"] & snp2geneWIN[[i]][[g]][j, "end"] <=
    CHRsnp[g, "BP"])
```



```
{snp2geneWIN[[i]][[g]][j, "gene_loc"]<-"down"
snp2geneWIN[[i]][[g]][j, "bp_dist"]<-CHRsnp[g, "BP"] - snp2geneWIN[[i]][[g]][j, "end"]}
# SNP rio abajo de genes
else if (snp2geneWIN[[i]][[g]][j, "start"] >= CHRsnp[g, "BP"] & snp2geneWIN[[i]][[g]][j, "end"] >=
CHRsnp[g, "BP"])
{snp2geneWIN[[i]][[g]][j, "gene_loc"]<-"up"
snp2geneWIN[[i]][[g]][j, "bp_dist"]<-snp2geneWIN[[i]][[g]][j, "start"] - CHRsnp[g, "BP"]}
else {snp2geneWIN[[i]][[g]][j, "gene_loc"]<- NA
snp2geneWIN[[i]][[g]][j, "bp_dist"] <- NA}
}}

# genes para el SNP 1
snp2geneWIN[[1]][[1]]

# unir dataframe
snp2geneWIN_DF<-snp2geneWIN

# unir listas de DF
for(i in 1:29){snp2geneWIN_DF[[i]]<-rbindlist(snp2geneWIN_DF[[i]])}

# unir listas de DF
snp2geneWIN_DF<-rbindlist(snp2geneWIN_DF)
snp2geneWIN_DF<-data.frame(snp2geneWIN_DF)
snp2geneWIN_DF<-snp2geneWIN_DF[snp2geneWIN_DF$bp_dist<=1000000,]

# leer los valores de P del análisis de señales de selección
load("GWIS_DF.RData")

# seleccionar columnas y convertir los valores de P
GWIS_DF<-data.frame(SNP=rownames(GWIS_DF), P=1*10^-GWIS_DF[,4],
BP=GWIS_DF[, "POSITION"], CHR=GWIS_DF[, "CHR"])

library(postgwas)
GWIS_DF<- postgwas:::readGWASdatasets(GWIS_DF)
GWIS_DF$SNP<-as.character(GWIS_DF$SNP)
```




```
GWIS_DF$CHR<-as.character(GWIS_DF$CHR)
```

```
# convertir identificadores de genes
```

```
library(clusterProfiler)
```

```
x<-BT_GF1_DF$gene_id
```

```
eg<-bitr(x, fromType="ENSEMBL", toType=c("ENTREZID", "UNIPROT", "ALIAS", "GENENAME"),  
OrgDb="org.Bt.eg.db")
```

```
eg<-eg[!duplicated(eg$ENTREZID),]
```

```
# unir nombres y entrez a genes
```

```
BT_GF1_DF<-cbind(BT_GF1_DF, Bt_UMD_3_1[match(BT_GF1_DF$gene_id,  
Bt_UMD_3_1$ensembl_gene_id, nomatch=NA_integer_), c("geneid", "genename")],  
eg[match(BT_GF1_DF$gene_id, eg$ENSEMBL, nomatch=NA_integer_),])
```

```
BT_GF1_DF$chr<-as.character(BT_GF1_DF$chr)
```

```
names(BT_GF1_DF)[c(4,6,7)]<-c("chr", "start", "end")
```

```
# seleccionar SNP que tienen valores de P (del objeto de GenABEL)
```

```
genabelgrowth855<-genabelgrowth855[,GWIS_DF$SNP]
```

```
# calcular calcular valores de P mediante método GATES
```

```
# vector de valores de P de SNP
```

```
pvalm<-GWIS_DF$P
```

```
names(pvalm)<-GWIS_DF$SNP
```

```
# crear una lista de dataframe con la información de los SNP en cada gen por cada  
cromosoma
```

```
nchr=29
```

```
SNPinf<-vector(mode="list", length=nchr)
```

```
# DF de anotación de genes a SNP
```

```
DFgen<-snp2geneWIN_DF
```

```
for(i in 1:29){
```

```
CHRi<-DFgen[DFgen$chr == i,]
```



```
gene_id<-CHRi$gene_id[!duplicated(CHRi$gene_id)]
SNPinf[[i]]<-vector(mode="list", length=length(gene_id))
names(SNPinf[[i]])<-gene_id
for(g in 1:length(gene_id)){
  SNPg<-CHRi[CHRi$gene_id == gene_id[g], c("SNP")]
  SNPinf[[i]][[g]]<-SNPg}
rm(CHRi, gene_id, SNPg, DFgen)

# mostrar los SNP en el gen 1 del cromosoma 1
SNPinf[[1]][[1]]

# mostrar los nombres de cada dataframe (cada gen, ENTREZ ID)
head(names(SNPinf[[1]]))
# total de genes
sum(sapply(SNPinf,length))

# función para asignar el triángulo inferior de las matrices de desequilibrio de ligamiento
simM <- function(m) {
  m[lower.tri(m)] <- t(m)[lower.tri(m)]
  diag(m) <- 1
  m}

##### calcular valores de P para cada gen mediante el método GATES
library(aSPU)
Gps<-NULL;
for(i in 1:29){ # por cada cromosoma
# calcular matriz de desequilibrio de ligamiento
LDm <- r2fast(genabelgrowth855[,genabelgrowth855@gtdata@chromosome==i])
LDm<-simM(LDm)
```



```
for(g in 1:length(SNPinf[[i]])){ # por cada gen de CHR
snps<-SNPinf[[i]][[g]]
newP <-pvalm[snps]
ldsub <- LDm[snps, snps]
if(length(snps) > 1) {
opvec = order(newP)
ldmat <- ldsub[opvec, opvec]
gatesp <- GATES2(ldmat, sort(newP))[1]
Gps <- rbind(Gps, c(length(newP), gatesp))
} else if (length(snps) == 1) {
gatesp <- newP
Gps <- rbind(Gps, c(length(newP), gatesp))}}
rm(i, g, snps, newP, ldsub, opvec, ldmat, gatesp, LDm)

##### genes para análisis de enriquecimiento
genEnrich<-snp2geneWIN_DF[!duplicated(snp2geneWIN_DF$gene_id), 1:16]
# Unir los valores de P calculatos mediante GATES2
genEnrich<-cbind(genEnrich, Gps)

# análisis de sobrerrepresentación de Gene Ontology utilizando la librería clusterprofiler
library(clusterProfiler)
over_CIP <- enrichGO(gene=genEnrich$gene_id[genEnrich$pvalue<=0.04], universe=genEnrich$gene_id,
OrgDb="org.Bt.eg.db", keytype='ENSEMBL', ont="BP", pAdjustMethod="none", pvalueCutoff=0.005,
qvalueCutoff=1, readable=TRUE)
dim(over_CIP@result)
head(over_CIP,20)
barplot(over_CIP, drop=TRUE, showCategory=10)
dotplot(over_CIP)
enrichMap(over_CIP, n=20)
cnetplot(over_CIP, categorySize="pvalue")

# análisis de sobrerrepresentación de MESH utilizando la librería meshr
```



```
library(meshr)
```

```
library(MeSH.Bta.eg.db)
```

```
meshParams <- new("MeSHHyperGParams", geneIds =  
genEnrich$ENTREZID[!is.na(genEnrich$ENTREZID) & genEnrich$pvalue<=0.01], universeGeneIds =  
genEnrich$ENTREZID[!is.na(genEnrich$ENTREZID)], annotation = "MeSH.Bta.eg.db", category = "G",  
database = "gene2pubmed", pvalueCutoff = 0.005, pAdjust = "none")
```

```
str(meshParams)
```

```
meshR <- meshHyperGTest(meshParams)
```

```
category(meshParams) <- "A"
```



Códigos correspondientes al capítulo III

Leer datos fenotípicos curados

```
Hatos_data_6398<-read.table("Hatos_data_6398.txt", header=T, sep="\t",na.strings="NA", dec=".",
strip.white=TRUE)
```

Evaluar y corregir pedigrí

```
library(breedR)
# renombrar columnas
colnames(Hatos_data_6398)[42:44]<-c("self", "dad", "mum")
# evaluar pedigrí
check_pedigree(Hatos_data_6398[, c('self', 'dad', 'mum')])
pedCH_6398<- build_pedigree(c("self", "dad", "mum"),data= Hatos_data_6398)
check_pedigree(pedCH_6398)
# guardar pedigrí
pedCH_DF_7460<-data.frame(id=pedCH_6398@label)
pedCH_DF_7460$sire<-pedCH_6398@sire
pedCH_DF_7460$dam<-pedCH_6398@dam
# reemplazar NAs por 0
pedCH_DF_7460[is.na(pedCH_DF_7460)] <- 0
library(pedigree)
pedCH_DF_7460$generation<-countGen(pedCH_DF_7460)
```

BLUP de peso al nacimiento y al destete en toda la población fenotipificada

```
mod_PNac_6398 <- remlf90(fixed= PNac ~ as.factor(AEHS)+ as.factor(EM_y), genetic= list(model=
'add_animal', pedigree= Hatos_data_6398[, 42:44], id= 'self'), data= Hatos_data_6398)
mod_P205_6398_1 <- remlf90(fixed= P205 ~ as.factor(AEHS), genetic= list(model= 'add_animal', pedigree=
Hatos_data_6398[, 42:44], id= 'self'), data= Hatos_data_6398)
```

heredabilidad y componentes de la varianza

```
summary(mod_PNac_6398)
summary(mod_P205_6398_2)
mod_PNac_6398$var
```



```
mod_P205_6398_1$var

# extraer residuales

residuales<-data.frame(PN_res=residuals(mod_PNac_res_1), PD_res=residuals(mod_P205_res_1),
PD_res2=residuals(mod_P205_res_2))

# asignar residuales

Hatos_data_6398<-cbind(Hatos_data_6398, round(residuales, digits=3))

library(psych)

describe(Hatos_data_6398 [c("PNac", "P205", "P365", "PN_res", "PD_res", "PD_res2")])

rm(residuales)

# crear datos 7460

# ordenar fenotipos con respecto al pedigrí 7460

Hatos_data_7460<-Hatos_data_6398[match(pedCH_DF_7460$self,
Hatos_data_6398$ID_cron_Anim,nomatch=NA_integer_),]

# unir datos 7460 a ped

pedCH_DF_7460<-cbind(pedCH_DF_7460,Hatos_data_7460[,54:56])

##### Obtener datos fenotípicos de 823 animales (población genotipificada)

# crear datos 823

Hatos_data_823<- Hatos_data_6398[match(paste(buscar_823$NoPrivado, buscar_823$MadreRegistro,
buscar_823$PNac,buscar_823$IdSexo, buscar_823$PadreRegistro,
sep='_'),paste(Hatos_data_6398$NoPrivado, Hatos_data_6398$MadreRegistro, Hatos_data_6398$PNac,
Hatos_data_6398$IdSexo, Hatos_data_6398$PadreRegistro, sep='_'), nomatch=NA_integer_),]

Hatos_data_823<-cbind(Hatos_data_823, pedCH_DF_7460[match(Hatos_data_823$self,
pedCH_DF_7460$self, nomatch=NA_integer_), c(8:13, 17:22)])

# recodificar pedigrí

library(breedR)

# renombrar columnas id, sire, dam
```



```
colnames(Hatos_data_823)[42:44]<-c("self", "dad", "mum")
# evaluar pedigrí
check_pedigree(Hatos_data_823[,c('self', 'dad', 'mum')])
pedCH_823<- build_pedigree(c("self","dad","mum"),data= Hatos_data_823)
pedCH_DF_823<-data.frame(id=pedCH_823@label)
pedCH_DF_823$sire<-pedCH_823@sire
pedCH_DF_823$dam<-pedCH_823@dam
pedCH_DF_823[is.na(pedCH_DF_823)] <- 0
library(pedigree)
pedCH_DF_823$generation<-countGen(pedCH_DF_823)

##### BLUP de peso al nacimiento y al destete en la población genotipificada
mod_PNac_823 <- remlf90(fixed= PNac ~ as.factor(AEHS)+ as.factor(EM_y), genetic= list(model=
'add_animal', pedigree= Hatos_data_823[, 42:44], id= 'self'), data= Hatos_data_823)
mod_P205_823_1 <- remlf90(fixed= P205 ~ as.factor(AEHS), genetic= list(model=
'add_animal', pedigree= Hatos_data_823[, 42:44], id= 'self'), data= Hatos_data_823)

# heredabilidad y componentes de la varianza
summary(mod_PNac_823)
summary(mod_P205_823_2)
mod_PNac_823$var
mod_P205_823_1$var

##### Validacion cruzada (aleatoria) mediante PBLUP_REML
library(breedR)
# matriz de incidencia
Za_823 <- model.matrix(mod_PNac_823)$genetic
nRep<-10
# matriz de correlaciones
CV_REML_PN_cor_823<-matrix(nrow=nRep, ncol=6, NA)
# matriz de coeficientes de regresión
```



```
CV_REML_PN_lm_coeff_823<-matrix(nrow=nRep, ncol=6, NA)
# matriz de error cuadrático medio
CV_REML_PN_mse_823<-matrix(nrow=nRep, ncol=6,NA)
# matriz de EBV
CV_REML_PN_EBV_PBLUP_823<-matrix(nrow=nRep,ncol=823, NA)
variable<-c("PN_res","PD_res","PD_res2","PN_dr","PD_dr","PD_dr2")
colnames(CV_REML_PN_cor_823)<-variable
colnames(CV_REML_PN_lm_coeff_823)<-variable
colnames(CV_REML_PN_mse_823)<-variable

# utilizar residuales de peso al nacimiento
P_REML_823<-Hatos_data_823[,c("PN_res","self","dad","mum")]
colnames(P_REML_823)[1]<- "P"
rownames(P_REML_823)<-pedCH_DF_823_acom$id
for(i in seq_along(PV_AL_PN)) {
# determinar filas con NA
tst<-as.numeric(rownames(P_REML_823)) %in% PV_AL_PN[[i]]
yNA_REML<-P_REML_823; yNA_REML$P[tst]<-NA
CV_REML_823 <- remlf90(fixed= P ~ 1, genetic= list(model= 'add_animal', pedigree= yNA_REML[, 2:4],
id= 'self'), data= yNA_REML)
# EBV de todo el pedigrí
EBV_1247 <- with(ranef(CV_REML_823), cbind(value=genetic, 's.e.'=attr(genetic, 'se')))
# EBV de los ids con fenotipos. Multiplica matriz de incidencia por EBV del pedigrí
EBV_823<- Za_823 %*% EBV_1247
# calcular correlacion entre fenotipos y EBV
CV_REML_PN_cor_823[i, 1]<-cor(P_REML_823$P[tst], EBV_823[, 1][tst])
CV_REML_PN_EBV_PBLUP_823[i,byrow=TRUE]<-EBV_823[,1]
# regresión de los fenotipos en los EBV
lm1 <- lm(P_REML_823$P[tst] ~ as.numeric (EBV_823[,1][tst]))
CV_REML_PN_lm_coeff_823[i,1]<- lm1$coefficients[2]
```




```
CV_REML_PN_mse_823[i, 1]<- mean((P_REML_823$P[tst] - as.numeric(EBV_823[, 1][tst]))^2);  
rm(CV_REML_823);rm(EBV_1247);rm(EBV_823); rm(lm1)}
```

Validación cruzada mediante predicción genómica

```
library(BGLR)
```

crear matriz de genotipos

```
geno_823_012_snpbyind<-  
read.table("geno_823_012_snpbyind.txt",sep="\t",header=F,na.strings="NA",dec=".",strip.white=TRUE)
```

```
map<-read.table("snp_map.txt",sep="\t",header=T)
```

```
geno_823_012_indbysnp<-t(geno_823_012_snpbyind)
```

```
dim(geno_823_012_indbysnp)
```

```
colnames(geno_823_012_indbysnp) = geno_823_012_indbysnp[1, ]
```

```
geno_823_012_indbysnp = geno_823_012_indbysnp[-1, ]
```

```
geno_823_012_indbysnp[1:5,1:5]
```

función para sustituir los NAs por el genotipo promedio en la población

```
replace_NAbymean=function(geno_823_012_indbysnp){  
  geno_823_012_indbysnp[is.na(geno_823_012_indbysnp)]=mean(geno_823_012_indbysnp,na.rm=T)
```

```
  geno_823_012_indbysnp}
```

```
  geno_823_012_indbysnp_new=apply(geno_823_012_indbysnp, 2, replace_NAbymean)
```

```
  geno=round (geno_823_012_indbysnp_new,0) # redondea los valores promedio
```

```
  colnames(geno)<-map$SNP_name
```

```
  rownames(geno) <- pedCH_DF_823_acom$id
```

```
  rm(geno_823_012_snpbyind, geno_823_012_indbysnp, geno_823_012_indbysnp_new)
```

crear matriz A del pedigrí de la población genotificada

```
library(breedR)
```

```
A_all<-as.matrix(getA(get_pedigree(mod_PNac_6398)))
```

```
A_all<-A_all[match(IDs_all_and_823$ID_corregido_7460, rownames(A_all), nomatch=NA_integer_),  
  match(IDs_all_and_823$ID_corregido_7460, colnames(A_all), nomatch=NA_integer_)]
```

```
A_all<-A_all/mean(diag(A_all))
```

```
colnames(A_all)<-IDs_all_and_823$ID_corregido_823
```

```
rownames(A_all)<-IDs_all_and_823$ID_corregido_823
```



```
# calcular matriz de relaciones genómicas
X<-scale(geno,center=TRUE,scale=TRUE)
G<-tcrossprod(X)/ncol(X)

# descompone G en autovectores.
EVD<-eigen(G)

# vector de IDs y fenotipos (residuales)
PN<-data.frame(PN=Hatos_data_823$PN_res)
rownames(PN)<-pedCH_DF_823_acom$Id

# predictor lineal
ETA_PBLUP_PN<-list(list(K=A_all, model="RKHS"))
ETA_GBLUP_PN<-list(list(V=EVD$vectors,d=EVD$values, model="RKHS"))
ETA_BC_PN<-list(list(X=geno,model="BayesC"))

# matrices de información
set.seed(123)
nRep<-10; nIter<-50000; burnIn<-5000
CV_PN_cor<-matrix(nrow=nRep, ncol=6, NA)
CV_PN_cor_s<-matrix(nrow=nRep, ncol=6, NA)
CV_PN_lm_coeff<-matrix(nrow=nRep, ncol=6, NA)
CV_PN_mse<-matrix(nrow=nRep, ncol=6, NA)
CV_PN_EBV_PBLUP<-matrix(nrow=nRep, ncol=823, NA)
CV_PN_EBV_GBLUP<-matrix(nrow=nRep, ncol=823, NA)
CV_PN_EBV_BC<-matrix(nrow=nRep,ncol=823, NA)
metodo<-c("PBLUP", "GBLUP", "BC")
colnames(CV_PN_cor)<-metodo
colnames(CV_PN_cor_s)<-metodo
```



```
colnames(CV_PN_lm_coeff)<-metodo
colnames(CV_PN_mse)<-metodo

#### validación cruzada
for(i in seq_along(PV_AL_PN)) {
  tst<-as.numeric(rownames(PN)) %in% PV_AL_PN[[i]] # elige IDs
  yNA<-PN$PN; yNA[tst]<-NA
  # PBLUP
  CV<-BGLR(y=yNA,ETA=ETA_PBLUP_PN, nIter=nIter, burnIn=burnIn)
  CV_PN_cor[i,1]<-cor(PN$PN[tst],CV$yHat[tst]) # pearson
  CV_PN_cor_s[i,1]<-cor(PN$PN[tst],CV$yHat[tst],method="spearman") # spearman
  lm1 <- lm(PN$PN[tst]~as.numeric(CV$yHat[tst])) # Regresión
  CV_PN_lm_coeff[i,1]<- lm1$coefficients[2] # coeficientes de regresión
  CV_PN_mse[i,1] <- mean((PN$PN[tst]-as.numeric(CV$yHat[tst]))^2) # MSE
  CV_PN_EBV_PBLUP[i,byrow=TRUE]<-predict(CV);rm(CV);rm(lm1) # GEBV
  # GBLUP
  CV<-BGLR(y=yNA,ETA=ETA_GBLUP_PN, nIter=nIter, burnIn=burnIn)
  CV_PN_cor[i,2]<-cor(PN$PN[tst],CV$yHat[tst])
  CV_PN_cor_s[i,2]<-cor(PN$PN[tst],CV$yHat[tst],method="spearman")
  lm1 <- lm(PN$PN[tst]~as.numeric(CV$yHat[tst]))
  CV_PN_lm_coeff[i,2]<- lm1$coefficients[2]
  CV_PN_mse[i,2] <- mean((PN$PN[tst]-as.numeric(CV$yHat[tst]))^2)
  CV_PN_EBV_GBLUP[i,byrow=TRUE]<-predict(CV);rm(CV);rm(lm1)
  # BayesC
  CV<-BGLR(y=yNA,ETA=ETA_BC_PN, nIter=nIter, burnIn=burnIn)
  CV_PN_cor[i,5]<-cor(PN$PN[tst],CV$yHat[tst])
  CV_PN_cor_s[i,5]<-cor(PN$PN[tst],CV$yHat[tst],method="spearman")
  lm1 <- lm(PN$PN[tst]~as.numeric(CV$yHat[tst]))
  CV_PN_lm_coeff[i,5]<- lm1$coefficients[2]
  CV_PN_mse[i,5] <- mean((PN$PN[tst]-as.numeric(CV$yHat[tst]))^2)
  CV_PN_EBV_BC[i,byrow=TRUE]<-predict(CV);rm(CV);rm(lm1)
```



```
}
```

```
##### exportar archivo de pedigrí para SSBR en julia
```

```
write.table(pedCH_DF_7460[,1:3], "pedCH_DF_7460.txt", quote = FALSE, sep="\t", row.names=F,  
col.names=F)
```

```
# exportar archivos de identificadores y fenotipos para SSBRjulia (p.e. validacion cruzada  
aleatoria de peso al nacimiento)
```

```
for(i in seq_along(PV_AL_PN_SSBR)){
```

```
write.table(pedCH_DF_7460[complete.cases(pedCH_DF_7460$PN_res) & !pedCH_DF_7460$Sid %in%  
PV_AL_PN_SSBR[[i]], c(1,14),
```

```
file=paste("PV_AL_PN_SSBR_p_", names(PV_AL_PN_SSBR[i]), ".txt", sep=""), quote=FALSE, sep="\t",  
row.names=F, col.names=F)
```

```
write.table(data.frame(id=pedCH_DF_7460[complete.cases(pedCH_DF_7460$PN_res) &  
!pedCH_DF_7460$Sid %in% PV_AL_PN_SSBR[[i]],c(1)),
```

```
MU="mu"),file=paste("PV_AL_PN_SSBR_f_", names(PV_AL_PN_SSBR[i]), ".txt", sep=""),  
quote=FALSE, sep="\t", row.names=F, col.names=T)}
```

```
# crear matriz genotipos para SSBR julia
```

```
M_CH<-as.matrix(read.table("geno_823_012_snpbyind.txt", sep="\t", header=F, na.strings="NA", dec=".",  
strip.white=TRUE))
```

```
M_CH<-M_CH[,-1]
```

```
rownames(head(M_CH))
```

```
row.names(M_CH)<-NULL
```

```
M_CH<-t(M_CH)
```

```
row.names(M_CH)<-NULL
```

```
replace_NAbymean=function(M_CH){
```

```
M_CH[is.na(M_CH)]=mean(M_CH, na.rm=T)
```

```
M_CH}
```

```
M_CH<-apply(M_CH,2,replace_NAbymean)
```

```
M_CH<-round (M_CH, 0)
```

```
# exportar genotipos para SSBR julia
```



```
write.table(M_CH,"M_CH.txt", quote = FALSE, sep="\t", row.names =
as.character(IDs_all_and_823$ID_corregido_7460), col.names = F)

rm(M_CH)
```

```
##### Analisis SSBR en julia y Windows
```

```
using JWAS: Datasets, SSBR, misc
```

```
# asignar archivos (población de validación 1 en base a grupos contemporáneos)
```

```
pedfile = "pedCH_DF_7460.txt"
```

```
genofile = "M_CH.txt"
```

```
phenofile = "PV_GC_PD_SSBR_p_fold1.txt"
```

```
fixedfile = "PV_GC_PD_SSBR_f_fold1.txt"
```

```
# asignar parámetros del modelo (p.e. para peso al destete)
```

```
#input=InputParameters()
```

```
input=InputParameters(314, "BayesC", 50000,1000, 0.995, 15.0, 15.0, true, false, false, 4.0, 4.0, 4.0, false)
```

```
input.method = "BayesC"
```

```
input.varGenotypic = 377.63
```

```
input.varResidual = 656.69
```

```
input.probFixed = 0.995
```

```
MCMCinfo(input)
```

```
# correr el modelo SSBR y exportar GEBV
```

```
fold=runSSBR(input, pedigree=pedfile, genotype=genofile, phenotype=phenofile, fixedfile=fixedfile);
```

```
using DataFrames
```

```
writetable("EBV_GC_PD_SSBR_fold1.txt", fold.ebv)
```

```
##### Calcular correlación de GEBV de SSBR vs fenotipos (residuales)
```

```
# leer los nombres de los archivos de EBV de SSBR
```

```
EBV_julia<-list.files(pattern="EBV")
```

```
# acomodar los archivos de EBV de SSBR
```



```
EBV_julia<-EBV_julia[order(nchar(EBV_julia), EBV_julia)]
# matriz de correlación de EBV vs fenotipos
CV_SSBR_AL_cor<-matrix(nrow=10,ncol=2,NA)
# matriz de coeficientes de regresión
CV_SSBR_AL_lm_coeff<-matrix(nrow=10,ncol=2,NA)

for(i in seq_along(EBV_julia)){
# leer el archivo de GEBV SSBR
SSBR_GEBV<-read.table(EBV_julia[i], header=T, sep=",", na.strings="NA", dec=".", strip.white=TRUE)
# acomodar GEBV de acuerdo con el pedigrí
SSBR_GEBV<-SSBR_GEBV[match(pedCH_DF_7460r$id, SSBR_GEBV$ID,nomatch=NA_integer_),]
CV_SSBR_AL_cor[i,1]<- cor(SSBR_GEBV[SSBR_GEBV$ID %in% PV_AL_PN_SSBR[[i]],
"EBV"],pedCH_DF_7460r[pedCH_DF_7460r$id %in% PV_AL_PN_SSBR[[i]], "PN_res"])
lm1 <- lm(pedCH_DF_7460r[pedCH_DF_7460r$id %in% PV_AL_PN_SSBR[[i]],"PN_res"] ~
SSBR_GEBV[SSBR_GEBV$ID %in% PV_AL_PN_SSBR[[i]], "EBV"])
CV_SSBR_AL_lm_coeff[i,1]<- lm1$coefficients[2]}
rm(SSBR_GEBV,lm1)
```