

Identificación de señales de selección reciente en el cromosoma 6 en ganado Holstein de Antioquia.

Juan David Rodríguez Mantilla

Universidad Tecnológica de Pereira, Facultad de Ciencias de la Salud, Programa de Medicina Veterinaria y Zootecnia, Tel: +57 316 285 47 17, e-mail: juandavirodriguez@utp.edu.co .

Grupo BIOGMA: Semillero de biodiversidad, genética y mejoramiento animal. Universidad Tecnológica De Pereira. Facultad de Ciencias de la salud .Pereira, Colombia.

Resumen

En Colombia el ganado Holstein, es una de las principales razas usadas para la producción lechera, por lo que identificar señales de selección reciente, puede ser de gran importancia para la producción de lácteos, de acuerdo con lo anterior el objetivo principal de este trabajo fue, identificar señales de selección reciente en el cromosoma 6 para ganado Holstein de Antioquia, mediante la metodología IHS. Para tal fin se utilizaron 145 animales para la extracción de DNA de sangre o semen mediante kit comercial, posteriormente se genotiparon los animales con el chip BovineLD (Illumina), que cuenta con un total de 6909 SNPs, se conservó sólo la información del cromosoma 6 que mapeaba en el ensamblaje bovino UMD 3.1 y se determinaron las señales de selección reciente mediante la metodología iHS. Además se realizó un modelo lineal generalizado para determinar el efecto de los marcadores con mayor señal de selección sobre la producción láctea, el porcentaje de proteína y de grasa. En total solo 2 marcadores focales para el cromosoma 6 fueron significativos ($p < 0.001$) para las señales de selección, en las cuales se encontraron algunas regiones importantes que estuvieron asociadas con genes como, CSN1S1, CSN2, CSN1S2, CSN3) que en general se encuentran relacionados

con funciones de estabilización de micelas de caseína para formación de proteína en la leche. Otro gen importante fue GnRHR (Receptor de gonadotropina), mediador en la liberación de la LH y FSH. Se encontraron frecuencias alélicas intermedias en los dos marcadores con señal de selección y uno de los marcadores presentó una asociación positiva con el porcentaje de proteína láctea, lo que coincide con un QTL previamente reportado.

Palabras claves: locus de rasgos cuantitativos, ganado lechero; producción de leche, polimorfismo de nucleótido simple.

Abstract

In Colombia, Holstein cattle are one of the main breeds used for dairy production, so identifying recent selection signals may be of great importance for dairy production. According to the above, the main objective of this work was, To identify recent selection signals on chromosome 6 for Holstein cattle of Antioquia, using the IHS methodology. For this purpose, 150 animals were used for the extraction of DNA by blood or semen with a commercial kit, the animals were then genotyped with the BovineLD chip (Illumina), which had a total of 6909 SNPs. Only the autosomal information that was mapped in the UMD cattle assembly 3.1. In addition, a generalized linear model was used to determine which production characteristics were associated. In total, only 2 focal markers for chromosome 6 were significant ($p < 0.001$) for the selection signals, in which some important regions were found that were associated with genes such as (CSN1S1, CSN2, CSN1S2, CSN3) that in general are related to stabilizing functions of casein micelles for protein formation in milk and genes such as GnRHR Receptor gonadotropin, mediator in the release of LH and FSH. Finally, the identification of selection signals using the iHS method allowed us to determine some regions of importance on the milk protein production of the Holstein cattle of Antioquia, some of these regions with previous reports of associated QTLs.

Key words: Quantitative trait locus, milk production, dairy cattle, Single-nucleotide polymorphism.

Introducción:

La industria láctea requiere ser cada vez más eficiente frente a las frecuentes amenazas que se vienen presentando en Colombia, principalmente debido a la firma de diferentes acuerdos de libre comercio y al aumento de los precios de los suplementos y agroquímicos necesarios para la producción, lo que hace necesario buscar alternativas con el fin de ser competitivo en un mercado cada vez más globalizado. Por lo anterior, es necesario buscar alternativas desde la genética que permitan seleccionar animales cada vez más productivos(1), por lo que identificar los genes de mayor importancia para la producción de leche se convierte en un factor primordial para las empresas de producción de lácteos(2)ⁱ, porque permite seleccionar animales con las mejores condiciones genéticas(3), y productivas para la industria lechera. Es importante resaltar que en el cromosoma 6 se han identificado muchos genes de importancia en la producción de leche, por lo que sería necesario buscar patrones de selección reciente sobre algunos de esos genes para identificar presiones de selección y genes de importancia para la industria lechera(4).

Por lo tanto, las empresas de producción lechera deben conseguir ser más eficientes para lograr enfrentar en mejores condiciones la agresiva competencia, esto se puede lograr desde la búsqueda de animales superiores genéticamente, como ha sido logrado por diferentes programas de mejoramiento genético en el mundo. Los programas de mejoramiento genético actuales buscan usar información molecular para mejorar las predicciones genéticas(5), por lo que el conocimiento de la ubicación de genes de importancia sobre la producción de leche se convierten un papel muy importante en la búsqueda de competitividad(6). Además, el conocimiento generado acerca de los genes de importancia puede generar un impulso en la industria lechera para obtener mejores resultados productivos y de calidad(7), asimismo permitirá entender los procesos de selección a los que ha sido sometido el ganado Holstein de Antioquia y finalmente posibilitará que las empresas productoras de leche generen muchos empleos.

Por otra parte el ganado Holstein ha sido seleccionado durante muchos años en diferentes lugares del mundo bajo diferentes criterios de selección(8). La identificación de regiones genómicas representa una excelente opción para identificar genes candidatos que responden a la selección genómica natural o artificial acordes con las condiciones productivas y de mercado de cada país (9),por lo que es posible utilizar bajo el criterio de identificación de señales de selección por medio de marcadores moleculares,(6) que faciliten este enfoque para identificar regiones asociadas a diferentes características de importancia para la industria lechera.

El mapeo y caracterización de genes generan información importante sobre la función y la posición de la densidad del genoma (10). Asimismo la búsqueda de estas señales de selección ya ha sido realizada en bovinos como indicadores de regiones segregantes, asociadas a locus de caracteres cuantitativos (QTL, por sus siglas en inglés, Quantitative trait locus). Las asociaciones a un marcador-QTL están vinculadas a características de la leche como, rendimiento de grasa y proteínas.(11)

Para la detección de señales de selección, Sabeti et al. (12), desarrollaron un estimador denominado homocigosidad haplotípica extendida (EHH) para buscar huellas de selección positiva en humanos. EHH constituye la probabilidad de que dos cromosomas escogidos aleatoriamente lleven el mismo alelo en un SNP focal y sean idénticos por descendencia en los marcadores que lo rodean. Voight et al. (13). Desarrollaron una prueba empírica basada en la integral del decaimiento observado de EHH, que fue definida como iHH y posteriormente complementada para definir la prueba iHS como un log-ratio de iHH, calculado en los alelos focales derivados y ancestrales. El enfoque planteado por Voight et al. (13)consiste en comparar el desequilibrio de ligamiento alrededor de un alelo seleccionado, con relación a un alelo no seleccionado que actúa como control con respecto al LD esperado en la región. La medida propuesta es conocida como la homocigosidad haplotípica extendida integrada estandarizada (iHS). Teniendo en cuenta lo anterior, el objetivo de este trabajo fue identificar señales de selección reciente para el cromosoma 6 en ganado Holstein de Antioquia y asociarlo a características de producción láctea.

Materiales y métodos:

Para la extracción de DNA y obtención de datos genotípicos, este trabajo utilizó la sangre del banco de suero de la Universidad Nacional con animales Holstein de Antioquia pertenecientes al programa de mejoramiento genético. El genotipado se realizó usando el BeadChip BovineLD (Illumina, San Diego CA) que tiene un total de 6909 SNPs con una distancia promedio entre marcadores contiguos de 383 kb. Se obtuvieron muestras de semen o sangre de 150 animales (37 toros y 113 vacas), evitando el maltrato de los animales. Las muestras de sangre fueron tomadas de la vena coccígea media, usando agujas número 18, con tubos BD vacutainer de 5ml y EDTA como anticoagulante (BD Vacutainer™). Una vez tomadas las muestras, fueron almacenadas a 4°C hasta su procesamiento. La extracción de DNA de los toros, se realizó a partir de semen en pajillas de 250 y 400 µl.

La reconstrucción de haplotipos y análisis de ligamiento para los progenitores fueron tomados teniendo en cuenta el menor parentesco posible entre grupos, pero en algunos casos se contó con unas pocas dupletas y tripletas (padre-madre-hija)(14). Después del proceso de edición, se reconstruyeron los haplotipos para cada cromosoma individualmente, teniendo en cuenta animales no relacionados, mediante el software Beagle 3.3. La reconstrucción de los haplotipos fue usada para los análisis posteriores(15). El ligamiento fue realizado paralelamente con Haploview v4.1 para estimar el desequilibrio de ligamiento mediante el estadístico r^2 .

La estimación de EHH e iHS evidencia selección positiva o negativa se realizó, basada en las frecuencias de haplotipos, de acuerdo con lo descrito por Voigh *et al.* (13). El iHS mide la expansión del LD local, teniendo en cuenta el haplotipo centrado en un SNP que lleva el alelo ancestral respecto al que lleva el derivado. El estadístico es aplicado a cada SNP individualmente, por lo que debe calcularse la integral del decaimiento de EHH para el alelo ancestral (iHH_A) y derivado (iHH_D). El iHS estandarizado de tal manera que tenga media 0 y varianza 1, de acuerdo con la siguiente ecuación:

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

Dónde: iHS es la medida de extensión del ligamiento, $\ln\left(\frac{iHH_A}{iHH_D}\right)$ corresponde a la medida de iHS no estandarizada a partir de un SNP específico; $E\left[\ln\left(\frac{iHH_A}{iHH_D}\right)\right]$ es la esperanza de iHS ; y $SD\left[\ln\left(\frac{iHH_A}{iHH_D}\right)\right]$, corresponde con la desviación estándar de iHS .

Los valores positivos o negativos más grandes, corresponden a haplotipos inusualmente largos llevando el alelo ancestral o derivado, respectivamente. Los puntajes iHS de los SNPs fueron transformados en $P_{iHS} = -\log[1 - 2|\Phi(iHS) - 0.5|]$, donde $\Phi(iHS)$ representa la distribución de la función acumulativa Gaussiana. Al asumir que los datos iHS se distribuyen normalmente (bajo neutralidad), P_{iHS} puede ser interpretado como $\log_{10}\left(\frac{1}{P}\right)$, donde P es el valor-P asociado a dos colas, para la hipótesis neutral en la que no hay selección. La proporción de valores de $iHS > 4$ fueron graficados teniendo en cuenta los 15 marcadores que rodearon cada SNP focal.

La estimación de EHH , iHS y P_{iHS} se realizó utilizando el paquete de R “rehh”, que usa una función para detectar señales de selección en datos de marcadores densos, usando la prueba basada en la homocigosidad haplotípica extendida (EHH) presentada anteriormente. El paquete permite la estimación de los estadísticos e incluso la posibilidad de realizar gráficos para visualizar e interpretar los resultados.

La identificación y anotación de regiones candidatas, se mapearon los SNPs de mayor efecto, con $P_{iHS} > 3$ ($p < 0.0001$). En los SNPs cercanos se determinó LD entre los marcadores contiguos y para la anotación se tuvo en cuenta que las ventanas asociadas a los QTLs no se solaparan. La anotación de los SNPs más importantes se realizó con base en el ensamblaje UMD 3.1 (*Bos taurus*) de NCBI y ENSEMBL, mediante la herramienta *Variant Effect Predictor* (VEP) (16). Las agrupaciones de genes y las figuras fueron realizadas de acuerdo a la información de ENSEMBL,

basadas en la construcción UMD 3.1 de VEP (16). Se recurrió a la ontología (*Gene ontology Consortium*) y se buscaron los reportes previos de QTLs señalados de acuerdo con la base de datos pública *Animal QTLdb* en la sección *cattleQTLdb* (17). La identificación y búsqueda de regiones candidatas se realizó con ventanas de 1 Mb de genoma, donde el SNP focal correspondió al centro (0.5 Mb para cada lado) de la ventana.

Análisis estadístico

El análisis estadístico se realizó utilizando un modelo lineal generalizado para cada una de las variables dependientes Rendimiento, porcentaje de grasa, porcentaje de proteína). El modelo general utilizado para este análisis fue el siguiente:

$$Y_{ij} = \mu + G_i + A_j + e_{ij}$$

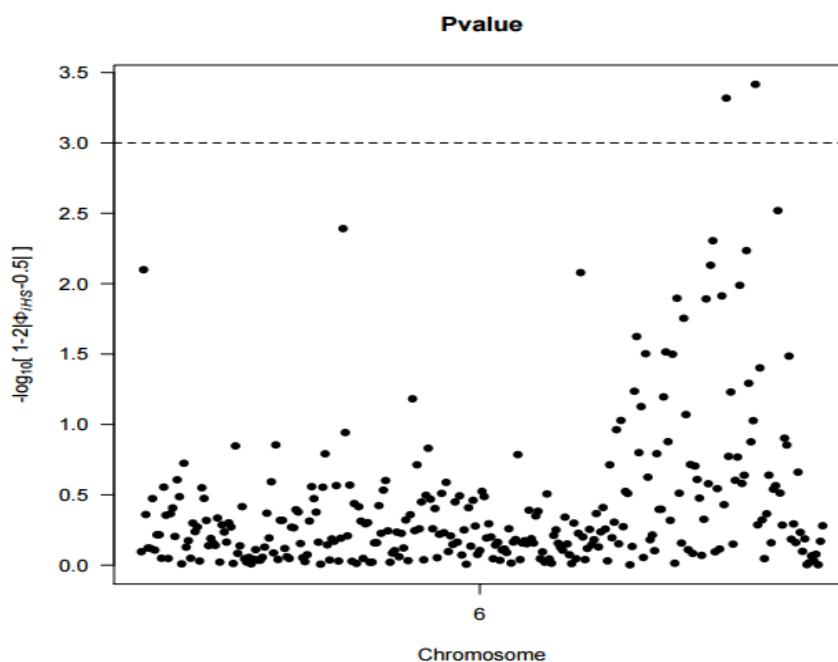
Donde, Y_{ij} es la variable respuesta, (que corresponde al valor de cría para la producción de leche, el porcentaje de grasa y el porcentaje de proteína.), μ es la media global, G_i el efecto del i -ésimo genotipo para el gen Hapmap26296-BTC-051940, A_j el efecto del j -ésimo genotipo para el gen ARS-BFGL-NGS-118182 y e_{ij} el error estadístico.

Resultados y discusión

Una vez hecha la edición de los datos para el cromosoma 6, se tuvieron en cuenta 306 SNPs (polimorfismo de nucleótido simple), y se buscaron los marcadores en la base de datos GenBank y en ENSEMBL. Adicionalmente, se buscó la función teniendo en cuenta la ontología reportada en la base de datos SWIS-PROT. En todos los casos se trabajó con la versión UMD3.1 del ensamblaje del genoma bovino.

Cuando se tuvo en cuenta un umbral para el valor absoluto $P > (p < 0.0001)$ sólo dos marcadores superaron el umbral y corresponden a regiones intergenicas en el cromosoma 6, pero cercana a genes de importancia en la producción de leche (Fig 1).

Figura 1: Grafico de la magnitud de *iHS* para el cromosoma (Bta) 6.



Es importante resaltar que en el cromosoma 6 presentó un pico significativo, que aunque no fue el más grande correspondió a un SNP (rs110527224) muy cercano a unos de genes de las caseínas (CSN1S1, CSN2, CSN1S2, CSN3) y de la GnRH. . Además se encontró otro SNP (rs109408260) que también se encontraba cercano a los genes mencionados anteriormente, evidenciando un QTL conocido asociado con la producción de proteína de la leche (18). Las frecuencias genotípicas en los dos marcadores fueron similares con mayor proporción de animales heterocigóticos (tabla 1), con respecto a las frecuencias alélicas los dos marcadores tuvieron frecuencias cercanas a 0.5 (Tabla 1.1).

Tabla 1: frecuencias genotípicas para los polimorfismos de nucleótido simple rs 109408260 y rs110527224 del cromosoma 6 bovino.

Genotipos		rs 109408260	rs110527224
aa	0	57 (35,7%)	48 (33,1%)
Aa	1	62 (43,3%)	57 (39,3%)
AA	2	30 (21%)	40 (27,6%)
Total		143	145

Tabla 1.1: Frecuencias alélicas para los polimorfismos de nucleótido simple rs 109408260 y rs110527224 del cromosoma 6 bovino.

Alelos	rs 109408260	rs110527224
a	57,30%	52,80%
A	42,70%	47,20%

El equilibrio de Hardy-Weinberg se obtuvo mediante el software R y plink v1.07. En el cual se encontró que los dos marcadores están en equilibrio, es decir que no se han sometido a procesos fuertes de migración, selección, mutación o deriva génica.

Los marcadores seleccionados por su señal de selección fuerte estuvieron asociados a algunos de genes con importancia metabólica evidente sobre la producción de leche, como es el caso de los CSN1, GNRH, TECLR, como se puede observar en la tabla 2.

Tabla 2: Descripción de los SNP con mayor señal de selección reciente y su ubicación en el ensamblaje del genoma bovino UMD 3.1 para el cromosoma 6.

Código rs	Cromosoma/posición	Gen/consecuencia	Función
rs109408260	6:82388706	EPHA5	Receptor del metabolismo celular.
		TECRL	(Trans-2,3-enoylCoA) reductasa interviene en procesos de biosíntesis de ácidos grasos.
		GnRHR	Receptor de gonadotropina, mediador en la liberación de la LH y FSH.

rs110527224	6:88592295	SLC4A4	Gen de la familia co-transportadora de bicarbonato de sodio.
		GC	Proteínas de unión a la vitamina D, almacenamiento y transporte de vitamina D, además interfiere en los mecanismos de inflamación.
		CSN-1,2-3	Estabiliza la formación de micelas, para la formación de leche y queso.

El marcador rs109408260 tuvo un efecto significativo sobre la producción de proteína en porcentaje, pero no tuvo ningún efecto sobre la producción de leche en cantidad, ni el porcentaje de grasa en leche, esto se debe a que en la región del marcador, hay genes importantes para la producción láctea, como las caseínas (CSN1S1, CSN2, CSN1S2, CSN3) de carácter importante en la industria lechera para la elaboración de yogur y queso entre otros.

Al analizar cada genotipo para el marcador rs109408260 se encontró un comportamiento aditivo (19), solo para el valor de cría relacionado con la producción de proteína, esto puede observarse en la tabla 3, donde los individuos AA presentaban el menor valor genético, mientras que los individuos aa el mayor.

Tabla 3: Comportamiento aditivo del marcador rs 109408260 relacionado con el porcentaje de proteína.

Genotipos	VCE% proteína
aa	0,02
Aa	0,01
AA	-0,02

Por lo anterior, el efecto de sustituir un alelo *A* por un alelo *a* es de 0,02% más en la producción de proteína en promedio. Es decir que seleccionar individuos con el genotipo *aa* permitirá obtener animales que contribuyan a mejorar la cantidad de proteína en leche del ganado Holstein de Antioquia y que pueda ser utilizado para elaborar derivados lácteos y para apoyar programas de selección asistida por marcadores moleculares.

Con respecto al segundo marcador rs110527224, no se encontró efecto sobre ninguna de las características, a pesar de tener señales de selección reciente, lo que hace pensar que está asociado a una característica diferente a las que se evaluaron en este trabajo, y finalmente se realizó una búsqueda en la base de datos Cattle QTL, y no se encontraron asociaciones previas para los mismos marcadores, pero si en regiones cercanas, donde se sugiere que existe un QTL asociado a la producción de proteína y caseínas en la leche, el cual se reporta alrededor de la posición 87,2Mpb, todo esto coincide con los hallazgos de este trabajo.

Conclusiones y recomendaciones:

Mediante la determinación de la Homosigocidad Haplotípica Extendida (IHS), se encontró que existen procesos de selección, específicamente para el cromosoma 6 en el ganado Holstein de Antioquia, se identificaron SNPs encontradas en regiones cercanas a genes importantes para la producción de leche como las caseínas, importantes para la formación de proteína en la leche, y del gen *GnHR* importante

para la reproducción, se evidencio además que uno de los marcadores con más pico de selección, tuvo un efecto positivo en la producción de proteína en porcentaje y asimismo se encontró asociado a un QTL previamente reportado.

Se recomienda para futuras investigaciones contar con una muestra de animales más grande, para que los resultados sean mejores y estén más cercanos a la realidad. Por otro lado, también se recomienda realizar el mapeo para la totalidad del genoma bovino, ya que nos permitirá encontrar una gran cantidad de marcadores, los cuales podrían estar asociados a características importantes de la producción lechera. Finalmente, hacer una revisión de los marcadores que no mostraron un pico alto de selección pero que podrían estar relacionados con características de productivas.

Bibliografía

1. Druet T, Pérez-Pardal L, Charlier C, Gautier M. Identification of large selective sweeps associated with major genes in cattle. *Anim Genet.* 2013;44(6):758–62.
2. Qanbari S, Pimentel ECG, Tetens J, Thaller G, Lichtner P, Sharifi AR, et al. A genome-wide scan for signatures of recent selection in Holstein cattle. *Anim Genet.* 2010;41(4):377–89.
3. de Simoni Gouveia JJ, da Silva MVGB, Paiva SR, de Oliveira SMP. Identification of selection signatures in livestock species. Vol. 37, *Genetics and Molecular Biology.* 2014. p. 330–42.
4. Hayes BJ, Lien S, Nilsen H, Olsen HG, Berg P, Maceachern S, et al. The origin of selection signatures on bovine chromosome 6. *Anim Genet.* 2008;39(2):105–11.
5. Qanbari S, Pausch H, Jansen S, Somel M, Strom TM, Fries R, et al. Classic Selective Sweeps Revealed by Massive Sequencing in Cattle. *PLoS Genet.* 2014;10(2).
6. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics.* 2001;157(4):1819–29.
7. Riquet J, Coppieters W, Cambisano N, Arranz JJ, Berzi P, Davis SK, et al.

- Fine-mapping of quantitative trait loci by identity by descent in outbred populations: application to milk production in dairy cattle. *Proc Natl Acad Sci U S A*. 1999;96(16):9252–7.
8. Kemper KE, Saxton SJ, Bolormaa S, Hayes BJ, Goddard ME. Selection for complex traits leaves little or no classic signatures of selection. *BMC Genomics* [Internet]. 2014;15(1):246. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3986643&tool=pmcentrez&rendertype=abstract>
 9. Gautier M, Vitalis R. Rehh An R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics*. 2012;28(8):1176–7.
 10. Olsen HG, Lien S, Gautier M, Nilsen H, Roseth A, Berg PR, et al. Mapping of a milk production quantitative trait locus to a 420-kb region on bovine chromosome 6. *Genetics*. 2005;169(1):275–83.
 11. Spelman RJ, Coppieters W, Karim L, Van Arendonk J a M, Bovenhuis H. Quantitative trait loci analysis for five milk production traits on chromosome six in the Dutch Holstein-Friesian population. *Genetics*. 1996;144(Table 1):1799–808.
 12. Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, et al. Detecting recent positive selection in the human genome from haplotype structure. *Nature*. 2002;419(6909):832–7.
 13. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol*. 2006;4(3):e72.
 14. Pérez-Enciso M. Genomic relationships computed from either next-generation sequence or array SNP data. *J Anim Breed Genet*. 2014;131(2):85–96.
 15. Qanbari S, Gianola D, Hayes B, Schenkel F, Miller S, Moore S, et al. Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. *BMC Genomics*. 2011;12(1):318.
 16. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics*. 2010;26(16):2069–70.
 17. Hu ZL, Park CA, Wu XL, Reecy JM. Animal QTLdb: An improved database tool

for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Res.* 2013;41(D1).

18. Ramey HR, Decker JE, McKay SD, Rolf MM, Schnabel RD, Taylor JF. Detection of selective sweeps in cattle using genome-wide SNP data. *BMC Genomics.* 2013;14(1):382.
 19. Su G, Christensen OF, Ostersen T, Henryon M, Lund MS. Estimating Additive and Non-Additive Genetic Variances and Predicting Genetic Merits Using Genome-Wide Dense Single Nucleotide Polymorphism Markers. *PLoS One.* 2012;7(9).
-