



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
DOCTORADO EN CIENCIAS DE LA PRODUCCIÓN Y DE LA SALUD ANIMAL

ESTUDIO DE LA DIVERSIDAD GENÉTICA DE GANADO HOLSTEIN DETERMINADA CON BASE
EN INFORMACIÓN GENÓMICA EN MÉXICO

T E S I S

PARA OBTENER EL GRADO DE
DOCTOR EN CIENCIAS

P R E S E N T A:

ADRIANA GARCÍA RUIZ

TUTOR PRINCIPAL:

DR. FELIPE DE JESÚS RUIZ LÓPEZ
FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLÁN, UNAM

MIEMBROS DEL COMITÉ TUTOR:

DR. CURT PAUL VANTASSELL
DOCTORADO EN CIENCIAS DE LA PRODUCCIÓN Y DE LA SALUD ANIMAL

DR. HUGO H. MONTALDO VALDENEGRO
FACULTAD DE MEDICINA VETERINARIA Y ZOOTECNIA, UNAM

Cuautitlán Izcalli, Estado de México, Mex.

Mayo, 2015



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

DEDICATORIA

A mi mejor amigo y compañero que Dios me ha dado, Vicente Castañeda Rodríguez, por su apoyo incondicional brindado en todo momento.

A mis padres, Virginia Ruiz y Ángel García.

AGRADECIMIENTOS

A las siguientes instituciones, por su apoyo y aportación para el desarrollo del trabajo de tesis:

- Universidad Nacional Autónoma de México, en especial a la Facultad de Estudios Superiores Cuautitlán.
- Consejo Nacional de Ciencia y Tecnología (CONACYT).
- Centro Nacional de Investigación Disciplinaria en *Fisiología* y Mejoramiento Animal (Inifap).
- Asociación Holstein de México.
- Animal improvement program laboratory, USDA.

Al Dr. Felipe de Jesús Ruiz López, por la confianza depositada en mí para llevar a cabo el presente proyecto, por sus enseñanzas, paciencia, amistad y apoyo incondicional.

Al Dr. Curt Van Tassell por su paciencia, apoyo y enseñanza durante mi programa doctoral y sobre todo durante mi estancia en el ARS-USDA.

Al Dr. Hugo Montaldo, por sus consejos y enseñanzas durante todo el programa.

A mis compañeros y amigos.

A mi esposo y mi familia.

ABREVIATURAS

BS	Pardo Suizo
CAN	Canadá
Conv	Sistema de producción especializado
GDB	Base de datos genómica de Estados Unidos América
GWAS	Estudios de Asociación genómica
HO	Holstein
JE	Jersey
Kg	Kilogramos
LD	Desequilibrio de ligamiento
Lowi	Sistema de Producción familiar.
M1	Marcadores no identificados para el alelo 1
M2	Marcadores no identificados para el alelo 2
MEX	México
PCA	Análisis de Componentes Principales
QTL	Loci de características cuantitativas
SIAP	Servicio Mexicano de Información de Agroalimentaria y Pesquera
SNP	Polimorfismo de un sólo nucleótido
USA	Estados Unidos de América

CONTENIDO

	PÁGINA
Agradecimientos	II
Abreviaturas	III
Resumen	VII
Abstract	IX
Capítulo 1: Introducción	01
1.1. Origen del ganado y su evolución	02
1.2. Especialización de ganado productor de leche	04
1.2.1. Ganado Holstein	04
1.2.2. Ganado Jersey	05
1.2.3. Pardo Suizo	05
1.3. Ganado lechero en México y sus sistemas de producción	06
1.4. Avances tecnológicos y su aplicación en estudios genéticos	08
1.5. Usos de información genómica en estudios de diversidad genética	09
1.6. Referencias	12
Capítulo 2: Justificación	17
Capítulo 3: Objetivos	19
Capítulo 4: Genetic differentiation of Mexican Holstein cattle and its relationship with Canadian and U.S. Holstein	20
4.1. Summary	21
4.2. Introduction	20
4.3. Materials and methods	22
4.4. Results	23
4.5. Discussion	25

4.6.	Author Contributions	26
4.7.	Acknowledgments	27
4.8.	References	27
Capítulo 5: Huellas de Selección en Regiones Genómicas de Ganado Holstein de México, Canadá y Estados Unidos.								
5.1.	Resumen	30
5.2.	Introducción	31
5.3.	Materiales y Métodos	32
5.4.	Resultados	34
5.5.	Discusión	35
5.6.	Conclusiones e implicaciones	37
5.7.	Agradecimientos	38
5.8.	Abstract	39
5.9.	Literatura citada	40
5.10.	Cuadros y figuras	46
Capítulo 6: Effect of reference population size and available ancestor genotypes on imputation of Mexican Holstein genotypes								
6.1.	Interpretative summary	52
6.2.	Abstract	54
6.3.	Introduction	56

6.4.	Materials and methods	58
6.5.	Results and discussion	62
6.6.	Conclusions	68
6.7.	Acknowledgments	68
6.8.	References	69
Capítulo 7: Genome regions associated to milk production traits and somatic cell score in the Mexican Holstein population								
		79
7.1.	Abstract	80
7.2.	Introduction	80
7.3.	Materials and methods	80
7.4.	Results and discussion	81
7.5.	Conclusions	82
7.6.	Literature cited	82
Capítulo 8: Discusión general								
		83
8.1.	Referencias	88
Capítulo 9: Conclusiones								
		90

RESUMEN

Se estudió el grado de diversidad genética y la composición genómica de la población Holstein de México (caracterizadas por su sistema de producción en el que se encuentran), así como su relación con las poblaciones Holstein de Estados Unidos de América y Canadá con la finalidad de establecer las bases de un programa de mejoramiento genético y estudiar la posibilidad de realizar evaluaciones genómicas conjuntas de ganado Holstein de México de diferentes sistemas de producción e involucrando poblaciones de otros países.

El proyecto contempló el análisis de la diversidad genética desde diferentes puntos de vista: *a)* La estructura y composición genética de poblaciones en dos sistemas de producción en México; Especializado (CONV) y familiar (LOWI) y su grado de relación con otras poblaciones Holstein de las que han dependido genéticamente, como las de Estados Unidos (EUA) y Canadá (CAN). Estos estudios fueron realizados a través de análisis de mezcla, de componentes principales, del mapeo de desequilibrio de ligamiento y del mantenimiento de la fase entre las poblaciones, *b)* La persistencia de regiones del genoma que se han mantenido entre las poblaciones (Huellas de selección) y que se atribuyen a efectos similares del proceso de selección en cada una de las poblaciones y *c)* La asociación entre la variación que existe en las regiones del genoma de la población Holstein de México y la de otros países, con la producción de leche, grasa, proteína y el conteo celular somático.

Para realizar los estudios de asociación genómica y determinar las regiones del genoma que codifican para características de interés económico, fue necesario realizar estudios de imputación de genotipos, ya que los densidad de los marcadores de un solo polimorfismo (SNP) utilizados en los animales disponibles

eran diferentes y fue necesario realizar pruebas para determinar qué información era la más adecuada, para utilizarla en proyectos posteriores que incorporarían información genómica.

Los estudios sobre la estructura, composición y desequilibrio de ligamiento, permitieron concluir que las poblaciones de animales usadas en los dos sistemas de producción en México (CONV y LOWI), no son iguales genéticamente, aunque tienen ancestros comunes. La misma tendencia se pudo observar entre la población Holstein de México y la de EUA y CAN, aunque es bien sabido que la primera depende genética y comercialmente de las dos últimas.

La presencia de huellas de selección en las poblaciones CONV y LOWI de México y las de EUA y CAN demostró que existen amplias regiones del genoma que son iguales entre las poblaciones y que su persistencia podría ser producto de los sistemas de selección que se han desarrollado dentro de estas poblaciones.

Los estudios de asociación genómica de las características de producción, permitieron encontrar nuevas asociaciones de SNP en la población Holstein de México, principalmente para proteína, que no habían sido reportadas en otros estudios. Estos resultados son muestra de la diversidad genética que existe entre las poblaciones, incluso a nivel de la producción de componentes de la leche.

Con este proyecto, se puede concluir que las poblaciones lecheras de los sistemas de producción en México presentan diferencias genéticas y que éstas también son diferentes de las poblaciones Holstein de EUA y CAN; estas deberán ser consideradas en los programas de mejoramiento genético que se quieran realizar en las poblaciones, utilizando información genómica.

ABSTRACT

The degree of genetic diversity and genomic composition of the Mexican Holstein population (characterized by the production system in which they perform) and its relationship with the Holstein populations from the United States (USA) and Canada (CAN) was studied, in order to establish the basis of a genetic breeding program and evaluate the feasibility of performing joint genomic evaluations of the Mexican Holstein populations or with populations from other countries.

The project involved analysis of genetic diversity from different perspectives, a) The genetic structure and composition of two production systems in Mexico, the conventional (CONV) and the low income system (LOWI) and their degree of relationship with other Holstein populations as from the US and CAN populations. These studies were conducted through the admixture, principal component and linkage disequilibrium analyses and the persistence phase among populations; b) The persistence of common genomic regions (signatures of selection) conserved among populations which were attributed to common selection processes, applying a linkage disequilibrium analysis and c) The variation of genomic regions associated to economically important traits (particularly milk yield, fat, protein and somatic cell score) in the Mexican Holstein population through the use of genome wide association studies.

In order to perform the genome wide association studies proposed and to determine the genomic regions linked to economically important traits, it was required to impute genotypes to a common base, because different panel densities were available. Since it was important to determine which set of information was the most suitable for imputation, comparison of the procedure using different sources of information was performed.

The studies of genetic structure, composition and linkage disequilibrium, led to the conclusion that the two production systems in Mexico (CONV and LOWI) are not genetically identical although they share common ancestors. The same trend was observed between the Mexican Holstein population and the USA and CAN, although it is well known that the Mexican population depends genetically and commercially from the latter two populations.

The existence of signatures of selection in the Mexican and the USA and CAN populations showed that there are wide genomic regions common among populations, which demonstrates their dependency. The presence of these regions could be the result of the selection programs implemented in these populations since they were associated with regions where QTL for traits under selection have been reported to reside.

The genomic association studies of production traits, revealed new associations of SNP in the Mexican Holstein population, especially for protein production, , which have not been reported in other studies. These results showed the presence of genetic diversity at the synthesis of milk component level.

With this project, we can conclude that the production systems in Mexico have differences in their genetic makeup and these are also different from the USA and CAN Holstein populations. These differences need be considered, if future genomic breeding programs are to be implemented in these populations.

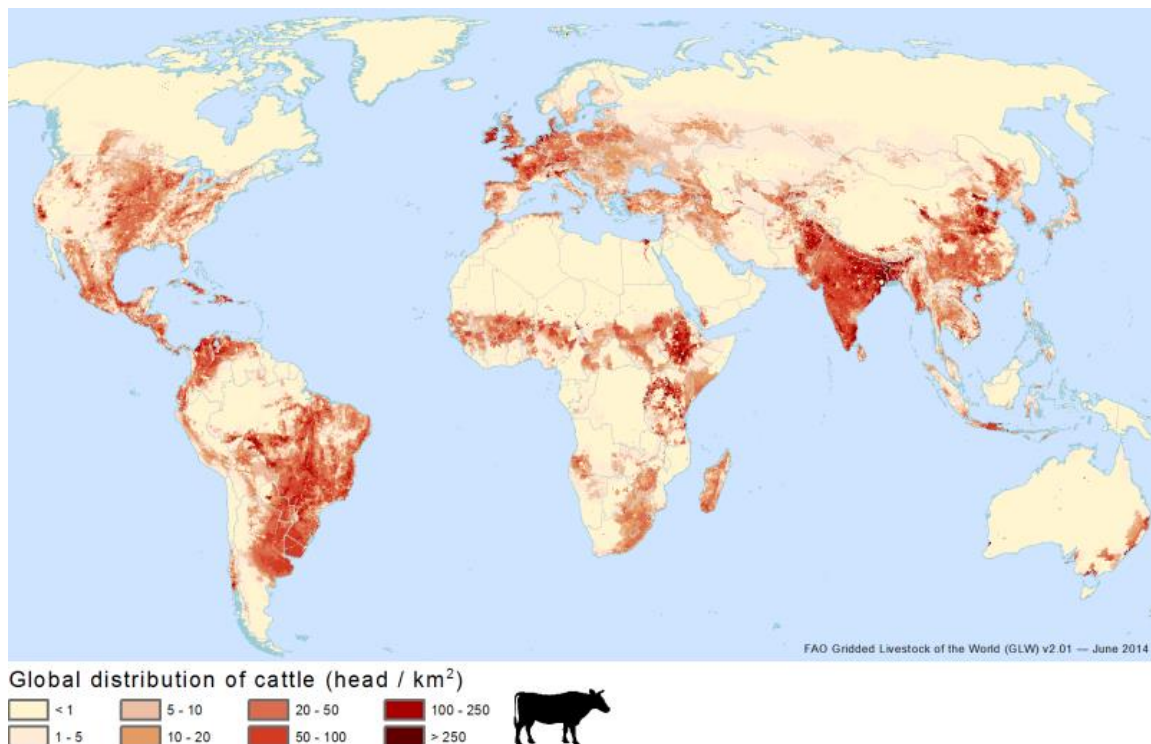
CAPÍTULO 1

INTRODUCCIÓN

Las investigaciones sobre el origen de los animales domésticos han experimentado un desarrollo considerable durante la última década, esto debido a la disponibilidad de herramientas bioinformáticas y moleculares.

Uno de los animales domésticos más importantes en la historia del ser humano es el ganado bovino (Lira, 2010). Esta especie ha sido usada para diferentes propósitos, incluyendo la producción de leche, carne, vestimenta (piel), trabajo de campo y espectáculos (Andersson y Georges, 2004).

Figura 1. Distribución de ganado Bovino a nivel mundial



Robinson et al., 2014

La organización de las Naciones Unidas para la Alimentación y Agricultura (FAO), determinó la presencia de 990 razas de ganado bovino, que constituyen el 22% del total de razas de mamíferos documentadas (FAO, 2007).

De acuerdo a estudios recientes, existen 1.43 billones de cabezas de ganado bovino distribuidas en todo el mundo (Robinson et al., 2014), figura 1.

1.1. Origen de Ganado Bovino y su evolución.

El ganado bovino se puede clasificar de manera general en dos grupos o subespecies que están bien clasificadas por sus rasgos fenotípicos: Con y sin la presencia de joroba (*Bos indicus* y *Bos Taurus* respectivamente) (Loftus et al., 1994; Badley et al., 1996). Estudios previos han demostrado que estas dos subespecies tuvieron un antecesor común y se separaron entre hace aproximadamente 1.7 a 2 millones de años (Troy et al., 2001; Hiendleder et al., 2008). Existen evidencias de que cada subespecie fue domesticada de manera independiente, el *Bos Taurus*, domesticada en el Cercano Oriente hace aproximadamente 8,000-10,000 años y el *Bos indicus* en el Sudoeste Asiático hace aproximadamente 6,000-10,000 años (Loftus et al., 1994). Otros autores consideran al *Bos indicus* como una variante del *Bos Taurus* (Payne, 1991; Lira, 2010).

Después de los eventos de domesticación de cada subespecie, se inició el proceso de selección genética realizada por los humanos y fue determinada en gran parte por la presencia de rasgos deseables como el fácil manejo y la capacidad de producción.

A partir de principios del siglo XIX, el desarrollo de las razas se fundamentó en la selección de fenotipos como la conformación, el color del pelaje y algunos aspectos productivos. El surgimiento de la civilización moderna fue acompañado por la adaptación, asimilación y mestizaje de los animales en cautiverio. El resultado de este proceso ha sido el desarrollo de razas con diferentes grados de especialidad que comprenden la producción de leche, la calidad de la carne, la tolerancia a climas y la

resistencia a enfermedades y plagas presentes en las regiones tropicales, entre otras características (The Bovine HapMap Consortium, 2009).

Durante los últimos 50 años, la cría de animales se ha basado en los progresos de la genética cuantitativa, que ha producido un notable progreso en la mejora de las características de producción de leche y carne (Anderson y Georges, 2004). Los procesos de selección (naturales y los implementados por los humanos) y los eventos demográficos han causado cambios en el genoma del ganado. Sus efectos combinados han creado una excepcional diversidad fenotípica y adaptación genética al medio ambiente local en todo el mundo a través de las razas vacunas modernas.

Estudios moleculares han evaluado la diversidad de las poblaciones de ganado bovino (Hayes et al., 2009; Flori et al., 2009) y han demostrado que la diversidad genética es producto de la especiación, la domesticación, la formación de la raza y el mejoramiento genético reciente (Xu et al., 2015). Hayes et al., (2009) generaron un mapa genómico para identificar las principales diferencias genéticas entre bovinos de razas productoras de leche y carne. Sus resultados demostraron que las diferencias en las frecuencias alélicas de bovinos lecheros y de carne se encuentran correlacionadas con los efectos de los marcadores genéticos asociados a las características de producción, lo que implica que la selección para producción (leche o carne) ha jugado un papel importante en la constitución del genoma bovino.

1.2. Especialización de ganado productor de leche.

En los últimos cincuenta años un número importante de razas bovinas se han especializado en producción de leche y gran parte del mejoramiento genético de éstas razas es producto de los procesos de selección y la adopción de tecnologías de la reproducción, como la inseminación artificial y el trasplante de embriones. Dentro de las razas especializadas más comunes de ganado lechero se encuentran la Ayrshire, Guernsey, Holstein, Jersey, Pardo Suizo y Shorthorn lechero. En México, las razas

especializadas en la producción de leche más comunes son la Holstein, Jersey y Pardo Suizo (SIAP-SAGARPA, 2013), por lo que a continuación se describirá de manera general la historia de formación de cada una de las razas lecheras que predominan en México.

1.2.1 Ganado Holstein

Los antecesores de la raza Holstein fueron las vacas negras de los bávaros y las blancas de los friesians, tribus que emigraron al oeste de Europa y que se asentaron en el delta del Rin hace aproximadamente 2000 años (French, 1968). La raza Holstein fue desarrollada en la parte norte de los Países Bajos (Holanda), en la provincia de Friesland y en provincias del norte de Alemania (Judkins-Keener, 1983). La raza se fue expandiendo a otros países europeos con un desarrollo rústico pero que le permitió en los últimos 300 años tener un valor importante en el mercado por sus características de producción y adaptación a los factores ecológicos de muchos países. Actualmente la raza Holstein tiene una distribución mundial (Farras, 1977).

No se cuenta con información bien documentada sobre la introducción de ganado Holstein en México, pero comunicaciones personales de la Asociación Holstein de México (Ing. Héctor de la Lanza, gerente técnico), mencionan que de acuerdo a investigaciones entre productores, se piensa que los primeros animales Holstein fueron importados a México a principios del siglo XIX y con el tiempo se fueron difundiendo en todo el territorio nacional, especialmente en el altiplano central y del norte.

1.2.2 Ganado Jersey

Raza originaria de la isla de Jersey en el Canal de la Mancha. Esta raza se fue desarrollando a partir del año 1700 adaptada a las necesidades de los habitantes de la isla y a las posibilidades forrajeras de un medio limitado. No existe información precisa de cuáles fueron las razas que la conformaron en su origen, pero existen teorías de que

el ganado negro pequeño de Bretaña y el colorado grande de Normandía fueron sus antecesores.

En 1743, los isleños, motivados por el interés de su ganado, decidieron preservar las características de la raza y prohibieron la introducción de otras razas bovinas a la isla. En 1784, se comenzó a exportar ejemplares a Inglaterra y las pruebas realizadas allí confirmaron su superioridad en el rendimiento de grasa, único componente valorado en esa época. La demanda de la vaca Jersey fue en aumento y durante el siglo XVIII muchos ejemplares llegaron a Sudáfrica, Austria, Tasmania y especialmente a Nueva Zelanda (Paulson et al., 2011).

1.2.3 Ganado Pardo Suizo

Esta raza es la más antigua de las razas lecheras, pues datos acerca de su existencia se remontan 2000 años antes de Cristo. Alcanzó su madurez en el siglo XIII en los valles de las montañas helvéticas y en el XIX se establecieron registros de producción sobre rendimiento en carne y leche.

Entre 1869 y 1880, se importó ganado a Estados Unidos que se utilizaron como base para el desarrollo de la raza en su evolución hacia la producción de leche y a partir de su selección, se creó el ganado Brown Swiss, la cual fue reconocida como raza pura lechera a finales del siglo XIX. Esta variedad estadounidense se conoce como Suiza Americana, y se caracteriza por su fácil adaptación, lo que ha impulsado su diseminación en diferentes altitudes que van del nivel del mar hasta por lo menos 3,800 metros (Paulson et al., 2011).

1.3. Ganado lechero en México y sus sistemas de producción.

El ganado lechero en México cuenta con una gran diversidad de razas y cruza, que se encuentran en diferentes sistemas de producción. El Servicio Mexicano de Información de Agroalimentaria y Pesquera (SIAP) informó que México cuenta con 2.3 millones de

cabezas de ganado lechero que se encuentran distribuidos en todo el país en diferentes sistemas de producción (sistema especializado, semi especializado, de doble propósito y familiar o de bajos recursos; Amendola, 2002). A continuación se mencionarán las principales características de cada sistema de producción, documentadas por SIAP-SAGARPA (2013).

Sistema de producción especializado. Cuentan animales principalmente de la raza Holstein, Pardo Suizo y Jersey. Tienen tecnología altamente especializada y predomina el manejo de los animales en corrales (estabulado) realizando prácticas de medicina preventiva, reproducción y mejoramiento genético. Se desarrolla principalmente en el altiplano y las zonas áridas y semiáridas del norte del país, donde se ubican las cuencas lecheras de importancia, como es el caso de la comarca Lagunera (Coahuila y Durango), Los altos (Jalisco), Rincón de Romos (Aguascalientes), Delicias y Cuauhtémoc (Chihuahua), Zumpango y Jilotepec (Estado de México), Tizayuca (Hidalgo), Colón y Villa del Marqués (Querétaro) y Mexicali (Baja California), entre otras. La raza Holstein en el sistema especializado en México ha dependido genéticamente de Canadá y Estados Unidos, ya que por varias décadas se ha importado semen, novillas de reemplazo y embriones (Powell y Wiggans, 1991; Valencia et al., 1999).

Sistema de producción semiespecializado. Cuentan con animales de las razas Holstein y Pardo Suizo principalmente. Mantienen a los animales en semiestabulación, no cuentan con equipo especializado para la ordeña y la gran mayoría de establos carece de equipo propio para enfriamiento y conservación de la leche. Existe un control de producción y programas en reproducción pero no tan sofisticado como en el sistema especializado. Las principales entidades federativas vinculadas con este sistema son Baja California, Colima, Chihuahua, Hidalgo, Jalisco, México, Michoacán, Morelos, Puebla, Sinaloa, Sonora, Tlaxcala y Zacatecas.

Sistema de producción de doble propósito. El ganado de estas explotaciones se produce carne o leche, dependiendo de lo que exija el mercado. Se desarrolla principalmente en

las regiones tropicales de México. Utilizan las razas Cebuinas y sus cruzas con Pardo Suizo, Holstein y Simmental. Su alimentación se basa en el pastoreo y la ordeña se realiza generalmente de forma manual. Los estados que cuentan con el mayor número de vientres en producción bajo este sistema son: Chiapas, Veracruz, Jalisco, Guerrero, Guanajuato, Tabasco, Zacatecas, Nayarit, San Luis Potosí y Tamaulipas.

Sistema de producción familiar o de traspatio. En este sistema el ganado se explota en pequeñas superficies de terreno, principalmente en las viviendas (traspatio). Pueden tener a los animales en corrales o en pastoreo dependiendo de las condiciones de sus campos de cultivo. Tienen animales Holstein y Pardo Suizo, así como cruzas. Los productores no realizan prácticas reproductivas, de medicina preventiva o mejoramiento genético; se carece de registros de producción. Las instalaciones son rudimentarias y predomina el ordeño manual. En este sistema predominan los estados de Jalisco, México, Michoacán, Hidalgo, Sonora y en menor grado Aguascalientes, Baja California, Coahuila, Chihuahua, Durango y Nuevo León.

En el cuadro 1 se muestran las principales características de cada uno de los sistemas de producción lechera en México mencionados anteriormente.

Cuadro 1. Principales Características de los Sistemas de Producción

Características	Intensivo o especializado	Semi-especializado	Doble propósito	Pequeña escala o familiar
Distribución en el país	Centro y norte, zonas áridas y semi-áridas	Centro y norte	Sur, zonas tropicales	Altiplano central y en zonas templadas
Tamaño del hato	300-400	180-250	30-40	3-20
Razas	Holstein, Pardo Suizo y Jersey	Holstein y Pardo Suizo	Cebú y cruzas con Pardo Suizo, Holstein y Simmental	Holstein, Pardo Suizo, Criollo y sus cruzas
Producción de leche	20-27 litros por día (305 días de lactación)	18-20 litros por día (280-305 días de lactación)	3-9 litros por día (120-180 días de lactación)	16 litros por día (225-300 días de lactación)

1.4. Avances tecnológicos y su aplicación en estudios genéticos

El mejoramiento genético de ganado lechero en todo el mundo ha evolucionado en gran medida a lo largo de la segunda mitad del siglo XX. Existen aspectos clave que han contribuido a este desarrollo como los avances en Genética Cuantitativa y las posibilidades que han brindado el desarrollo de métodos estadísticos e informáticos derivados del avance progresivo en las capacidades de cálculo y almacenamiento de las computadoras (Tirados, 2001).

En la década de 1990, el potencial combinado de la clonación molecular y el análisis de ADN por microsatélites permitió desarrollar mapas completos de desequilibrio de ligamiento del genoma bovino (Bishop et al., 1994; Ihara et al., 2004). Estos mapas se han utilizado principalmente para identificar las ubicaciones cromosómicas de loci de rasgos cuantitativos (QTL) que son responsables de la variación en características de interés económico y que han sido ampliamente estudiados en ganado lechero (Heyen et al., 1999; Ashwell et al., 2004; Chen et al., 2006; Daetwyler et al., 2008). Durante la última década se han producido avances en genética molecular, incluyendo métodos para la determinación simultánea de cientos de miles de polimorfismos de un solo nucleótido (SNP) con alta precisión. Aunque cada SNP es

menos informativo que un marcador microsatélite individual, los SNPs en forma colectiva pueden monitorear la variación genética de manera efectiva a partir de numerosas muestras (The Bovine HapMap Consortium, 2009). Con la aplicación de éstas técnicas moleculares y mediante el uso de paneles de alta densidad de SNP ha sido posible identificar marcadores asociados a rasgos económicos importantes, lo que permitió predecir el valor genético de los animales basado en información genómica y que se le denominó selección genómica (Meuwissen et al., 2001).

La disponibilidad de información genómica ha estimulado el desarrollo de metodologías para realizar mapas genéticos de las poblaciones y estudiar el impacto de la selección sobre los niveles de diversidad, las desviaciones en las frecuencias de los marcadores de SNP, el aumento de desequilibrio de ligamiento en las poblaciones y la estructura de los haplotipos poblacionales (Xu et al., 2014).

1.5. Uso de información genómica en estudios de diversidad genética.

La historia de la evolución del ganado bovino está marcada por intensas presiones de selección que se llevaron a cabo durante la domesticación y la formación de cada una de las razas. La información genómica en el estudio de diversidad genética y evolución del ganado a través de la historia, ha proporcionado información de frecuencias alélicas en todo el genoma bovino, que permiten diferenciar desde el punto de vista genómico cada una de las razas e incluso poblaciones de la misma raza (The Bovine HapMap Consortium, 2009). Existen diferentes metodologías que se han desarrollado para estudiar las diferencias en las frecuencias alélicas de poblaciones y que permiten determinar el grado de diversidad genética.

El **análisis de mezcla** es una metodología establecida para estudiar la estructura de una población con respecto a otra(s) y calcular la deriva genética. Éstos análisis ayudan a definir el cambio de frecuencias alélicas debido a la naturaleza aleatoria de la recombinación (Winkler et al., 2010) o el origen de taxones parentales asociados a

grupos o razas genéticamente diferentes (Wang et al., 2005; Buerkle et al., 2008). Los **Análisis de Componentes Principales** (PCA), son una herramienta que ayudan interpretar mejor los análisis de mezcla, ya que a través de sus resultados, es posible construir mapas genéticos basados en la frecuencia de los alelos de diferentes grupos o poblaciones (Cavalli-Sforza et al., 1994; Silva et al., 2009).

El estudio del **Desequilibrio de Ligamiento y de persistencia de su fase** entre poblaciones, son herramientas que se ha utilizado para estudiar diferentes eventos de la población, como la selección, la deriva genética, la selección natural, la mutación y el origen de razas o líneas (Pritchard et al., 2001 y Waples et al., 2010), además de que en los análisis que incluyen a más de una población, permiten definir el grado de similitud entre las mismas.

Para determinar el efecto del proceso de selección natural o inducido (a través del establecimiento de programas de mejoramiento genético) entre una o más poblaciones, se ha recurrido a la identificación de **Huellas de Selección**, que son regiones en el genoma que han incrementado su presencia y se han establecido en poblaciones específicas debido a su importancia funcional (Pérez O'Brien et al., 2014). La identificación de huellas de selección en el genoma, permiten definir sucesos comunes y conocer los grados de similitud genética entre grupos o poblaciones.

El grado de variación genética de características de importancia económica en ganado lechero, ha llevado a la necesidad de detectar regiones del genoma que expliquen esa variación genética de características de importancia en los programas de mejoramiento genético de ganado (Pryce et al., 2010).

Los estudios de asociación genómica (GWAS) son una técnica reciente para la identificación de los genes causales o marcadores genéticos que explican la variación genética de características de interés (Zhang et al., 2012; Wang et al., 2005). Estos estudios han dado lugar a la posibilidad de ejecutar el mapeo de Loci de características cuantitativas (QTL) de manera precisa (Calus, 2009). Se han reportado diferencias entre

los marcadores genómicos asociados para las mismas características y razas en diferentes poblaciones, lo que podría atribuirse a una diferencia en la estructura de la población y por ende al grado de diversidad genética (Zhang et al. 2012).

REFERENCIAS

- Lira J. Revisión sobre la genética del origen del ganado vacuno y las aportaciones del ADN antiguo. *MUNIBE (Antropología-Arkeología)*, 2010. 61: 153-170.
- Anderson L y M Georges. Domestic-animal genomics: deciphering the genetics of complex traits. *Nat. Rev. Genet*, 2004. 5:202-212.
- FAO. *The State of the World's Animal Genetic Resources for Food and Agriculture*, 2007. Ed. RISCHKOWSKY, B. y PILLING, D. Rome
- Robinson TP, Wint GR, Conchedda W, Van Boeckel G, Ercoli TP, Palamara V, Cinardi E, D'Aietti G, Hay L y Gilbert S. Mapping the global distribution of livestock. *PLoS One*, 2014. 9(5), e96084. doi:10.1371/journal.pone.0096084
- Loftus RT, MacHugh DE, Bradley DG, Sharp PM and Cunningham P. Evidence for two independent domestications of cattle. *Proceedings of the National Academy of Sciences of the United States of America*, 1994. 91 (7):2757-2761.
- Bradley DG y Cunningham EP. Genetic aspects of domestication. *The Genetics of Cattle* 1998. pp. 15-32. CAB International, Oxon, UK.
- Troy CS, MacHugh DE, Bailey JF, Magee DA, Loftus RT, Cunningham P, Chamberlain AT, Sykes BC y Bradley DG. Genetic evidence for Near-Eastern origins of European cattle. *Nature*, 2001. 410:1088-1091.
- Hiendleder S, Lewalski H y Janke A. Complete mitochondrial genomes of *Bos taurus* and *Bos indicus* provide new insights into intra-species variation, taxonomy and domestication. *Cytogenet. Genome Res* 2008. 120:150-156.
- Payne WJA. Cattle genetic resources. In: *World Animal Series B7*, Amsterdam: Elsevier Science, 1991. 51-72.
- The Bovine HapMap Consortium. Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. *Science*, 2009. 324:528-532.

- Hayes BJ, Chamberlain AJ, Maceachern S, Savin K, McPartlan H, MacLeod I, Sethuraman L y Goddard ME. A genome map of divergent artificial selection between *Bos taurus* dairy cattle and *Bos taurus* beef cattle. *Animal Genetics*, 2009. 40(2), 176-84. doi:10.1111/j.1365-2052.2008.01815.
- Flori L, Fritz S, Jaffrézic F, Boussaha M, Gut I, Heath S, Foulley J y Gautier M. The genome response to artificial selection: a case study in dairy cattle. *PloS One*, 2009. 4(8), e6595. doi:10.1371/journal.pone.0006595
- Xu L, Bickhart DM, Cole JB, Schroeder SG, Song J, Van Tassell CP, Sonstegard T, Liu GE. Genomic Signatures Reveal New Evidences for Selection of Important Traits in Domestic Cattle. *Molecular Biology and Evolution*, 2015. 1-15. doi:10.1093/molbev/msu333
- SAGARPA. Situación Actual y perspectiva de la producción de leche en ganado Bovino en México. Consultado en Octubre de 2014, disponible en: www.sagarpa.gob.mx/ganadería/publicaciones/Situación-actual-y-perspectiva-de-la-produccion-de-leche-en-ganado-bovino-en-mexico
- SIAP-Sagarpa. Escenario Base 2009-2018, Proyecciones para el sector Agropecuario de México, Texas A&M University & University of Missouri, 2013.
- French MH. Razas Europeas de ganado bovino. Editorial FAO, Organización de Naciones Unidas para la Agricultura y la alimentación, 1968. Roma Italia.pp 99-100
- Judkins FH y Keener AH. La leche, su producción y procesos industriales. Editorial C.E.C.S.A. México 1983, DF p 71-80.
- Farras J. Cría lucrativa de la vaca lechera. 6ª ed. Editorial Sintés, S.A. Barcelona, España 1977. 67-90
- Paulson J, Salfer J, Newell S, Santi E, Seykora T, Litherland N, Endres M, Janni K, Reneau J, Broadwater N, Schwartz C, Rozeboom G y Hudson S. Learning About Dairy NCR593. 2011. 4-15.

- Amendola RD. A dairy system based on forages and grazing in temperate Mexico. PhD thesis, 2002. Wageningen University, The Netherlands.
- Powell RL y Wiggans GR. Animal model evaluations for Mexican Holsteins. *J Dairy Sci* 1991. 74:1420-1427.
- Valencia PM, Ruiz F, Montaldo HH, Keown JF y Van Vleck LD. Evaluación genética para la producción de leche en ganado Holstein en México. *Técnica Pecuaria Méx*, 1999. 37:1-7.
- Tirados SP. Livestock Genetic Improvement in the Second Half of the XX Century. *Arch Zootec*, 2001. 50: 517-546.
- Bishop MD, Kappes SM, Keele JW, Stone RT, Sunden S, Hawkins GA, Toldo SS, Fries R, Grosz MD Yoo J y Beattie CW. A genetic linkage map for cattle. *Genetics*, 1994. 136:619-639
- Ihara N, Takasuga A, Mizoshita K, Takeda H, Sugimoto M, Mizoguchi, Hirano T, Itoh T, Watanabe T, Reed K, Snelling W, Kappes S, Beattie C, Bennett G y Sugimoto Y. A Comprehensive Genetic Map of the Cattle Genome Based on 3802 Microsatellites. *Genome Research*, 2004. 14(10a), 1987-1998. doi:10.1101/gr.2741704
- Heyen DW, Weller JI, Ron M, Band M, Beever JE, Feldmesser E, Da Y, Wiggans GR, VanRaden PM y Lewin HA. A genome scan for QTL influencing milk production and health traits in dairy cattle *Physiol. Genomics*, 1999. 1:(3) 165-175.
- Ashwell MS, Heyen DW, Sonstegard TS, Van Tassell CP, Da Y, VanRaden PM y Lewin HA. Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle. *J Dairy Sci*, 2004. 87(2), 468-75. doi:10.3168/jds.S0022-0302(04)73186-0
- Chen HY, Zhang Q, Yin CC, Wang CK, Gong WJ y Mei G. Detection of quantitative trait loci affecting milk production traits on bovine chromosome 6 in a Chinese

- Holstein population by the daughter design. *J Dairy Sci*, 2006. 89(2), 782-90. doi:10.3168/jds.S0022-0302(06)72140-3
- Daetwyler HD, Schenkel FS, Sargolzaei M, Robinson JAB. A Genome Scan to Detect Quantitative Trait Loci for Economically Important Traits in Holstein Cattle Using Two Methods and a Dense Single Nucleotide Polymorphism Map. *J Dairy Sci*, 2008. 91:3225-3236.
- Meuwissen THE, Hayes BJ y Goddard ME. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics*, 2001. 157:1819-1829.
- Winkler AC, Nelson WG y Smith WM. Admixture Mapping Comes of Age. *Annual Review of Genomics and human Genetics*, 2010. Vol. 11: 65-89.
- Wang WYS, Barratt BJ, Clayton DG y Todd JA. Genome-wide association studies: theoretical and practical concerns. *Nature reviews. Genetics*, 2005. 6(2), 109-18.
- Buerkle CA y Lexer C. Admixture as the basis for genetic mapping. *Trends. Ecol. Evol*, 2008. 23:686-694. doi: 10.1016/j.tree.2008.07.008
- Cavalli-Sforza LL, Menozzi P y Piazza A. *The History and Geography of Human Genes*. Princeton University Press, Princeton, New Jersey, USA. 1994. 1059 p.
- Silva-Zolezzi I, Hidalgo-Miranda A, Estrada-Gil J, Fernandez-Lopez JC, Uribe-Figueroa L, Contreras A y Balam-Ortiz E. Analysis of genomic diversity in Mexican Mestizo populations to develop genomic medicine in Mexico. *PNAS*, 2009. 106: 8611-6. doi: 10.1073/pnas.0903045106
- Pritchard JK y Przeworski M. Linkage disequilibrium in humans: models and data. *American journal of human genetics*, 2001. 69(1), 1-14. doi:10.1086/321275.
- Waples RS y Do C. Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evolutionary Applications*, 2010. 3(3), 244-262.
- Pérez O'Brien AM, Utsunomiya YT, Mészáros G, Bickhart DM, Liu GE, Van Tassell CP, y Sölkner J. Assessing signatures of selection through variation in linkage

disequilibrium between taurine and indicine cattle. *Gen Sel, Evol: GSE*, 2014. 46, 19. doi:10.1186/1297-9686-46-19

Pryce JE, Bolormaa S, Chamberlain AJ, Bowman PJ, Savin K, Goddard ME y Hayes BJ. A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. *J Dairy Sci*, 2010. 93(7), 3331-45.

Zhang H, Wang Z, Wang S y Li H. Progress of genome wide association study in domestic animals. *J Anim Scie and Biotechnology*, 2012. 3(1), 26. doi:10.1186/2049-1891-3-26

Calus MPL. Genomic breeding value prediction: methods and procedures. *Animal?: An international journal of animal bioscience*, 2009. 4(2), 157-164.

CAPÍTULO 2

JUSTIFICACIÓN

La importancia de conocer la diversidad genética en animales domésticos, se orienta hacia satisfacer las necesidades actuales de producción en diversos entornos, establecer las estrategias a usar en los programas de selección de acuerdo a las condiciones y composición genética de las poblaciones y facilitar la adaptación rápida de los animales a los cambios de objetivos de los programas de mejoramiento genético.

Los animales de la raza Holstein son ampliamente usados en la producción láctea a nivel mundial y en México es la raza lechera más común, bajo diferentes sistemas de producción de leche.

Ante la necesidad de establecer un programa de mejoramiento genético de ganado Holstein en México incorporando información genómica, era indispensable primero evaluar el grado de diversidad genética de ganado Holstein en México que se encuentra en diferentes sistemas de producción y medir el grado de relación genética de éstos sistemas con otras poblaciones Holstein, como Estados Unidos y Canadá.

Debido a la amplitud del proyecto, fue necesario establecer cuatro subproyectos que estaban relacionados entre sí y que llevaron a cubrir los objetivos planteados en el proyecto original. Para explorar la diversidad genética fue necesario realizar mapas genómicos, estudiar la composición genética de los animales y evaluar el grado de relación genética medido como la asociación entre regiones del genoma a través de análisis de componentes principales, de análisis de mezcla y de estudios de desequilibrio de ligamiento (Capítulo 4), que indicarían diferencias en el material genético presente en las poblaciones estudiadas. Los resultados del primer subproyecto, nos llevaron a estudiar la presencia de huellas de selección comunes en la población Holstein de México y otras poblaciones (Capítulo 5), evaluar el impacto de incluir información genómica de otros animales en el proceso de imputación de genotipos de animales

Holstein mexicanos (Capítulo 6) y detectar regiones del genoma asociadas a características de interés económico y compararlas con las obtenidas en otras poblaciones (Capítulo 7).

Los resultados de cada subproyecto serán presentados en los siguientes cuatro capítulos en el formato en el que fueron publicados los resultados.

CAPÍTULO 3

3.1 OBJETIVO GENERAL

Evaluar el grado de diversidad genética y la composición genómica de las subpoblaciones mexicanas Holstein (caracterizadas por el sistema de producción en el que se encuentran) y su relación con otras poblaciones Holstein.

3.2 OBJETIVOS ESPECÍFICOS

- Estudiar la composición y diferenciación genética del ganado Holstein de México en dos sistemas de producción diferente (especializado y familiar) y su relación con el ganado Holstein de Estados Unidos y Canadá a través de estudios de análisis de mezcla, de componentes principales y de desequilibrio de ligamiento.
- Identificar huellas de selección en el genoma de ganado Holstein de México y de Estados Unidos y Canadá analizando el desequilibrio de ligamiento presente en las poblaciones.
- Evaluar el proceso de imputación de genotipos de los animales Holstein de México con la incorporación de información genómica de otras poblaciones disponibles (Estados Unidos y Canadá)
- Identificar SNP asociados a características de importancia económica en la población Holstein de México y compararlos con los reportados en otras poblaciones.

CAPÍTULO 4

Diferenciación genética de ganado Holstein de México y su relación con ganado Holstein de Canadá y Estados Unidos de América

Capítulo presentado con la publicación de un artículo científico publicado en la revista *Frontiers in Genetics*, en Febrero de 2015.



Genetic differentiation of Mexican Holstein cattle and its relationship with Canadian and U.S. Holsteins

Adriana García-Ruiz¹, Felipe de J. Ruiz-López^{1,2*}, Curtis P. Van Tassell³, Hugo H. Montaldo¹ and Heather J. Huson^{3,4}

¹ Facultad de Estudios Superiores Cuautitlán, Universidad Nacional Autónoma de México, Ajuchitlán, Mexico

² Centro Nacional de Investigación Disciplinaria en Fisiología y Mejoramiento Animal, Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias – Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación, Ajuchitlán, Mexico

³ Animal Genomics and Improvement Laboratory, Agricultural Research Service, United State Department of Agriculture, Beltsville, MD, USA

⁴ Department of Animal Science, Cornell University, Ithaca, NY, USA

Edited by:

Johann Sölkner, BOKU – University of Natural Resources and Life Sciences Vienna, Austria

Reviewed by:

Ikhide G. Imumorin, Cornell University, USA
Bertrand Servin, Institut National de la Recherche Agronomique, France

*Correspondence:

Felipe de J. Ruiz-López, Centro Nacional de Investigación Disciplinaria en Fisiología y Mejoramiento Animal, Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias – Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación, Km. 1 Carretera a Colón, Ajuchitlán, Querétaro 76280, Mexico
e-mail: ruiz.felipe@nifap.gob.mx

The Mexican Holstein (HO) industry has imported Canadian and US (CAN + USA) HO germplasm for use in two different production systems, the conventional (Conv) and the low income (Lowi) system. The objective of this work was to study the genetic composition and differentiation of the Mexican HO cattle, considering the production system in which they perform and their relationship with the Canadian and US HO populations. The analysis included information from 149, 303, and 173 unrelated or with unknown pedigree HO animals from the Conv, Lowi, and CAN + USA populations, respectively. Canadian and US Jersey (JE) and Brown Swiss (BS) genotypes (162 and 86, respectively) were used to determine if Mexican HOs were hybridized with either of these breeds. After quality control filtering, a total of 6,617 out of 6,836 single nucleotide polymorphism markers were used. To describe the genetic diversity across the populations, principal component (PC), admixture composition, and linkage disequilibrium (LD; r^2) analyses were performed. Through the PC analysis, HO × JE and HO × BS crossbreeding was detected in the Lowi system. The Conv system appeared to be in between Lowi and CAN + USA populations. Admixture analysis differentiated between the genetic composition of the Conv and Lowi systems, and five ancestry groups associated to sire's country of origin were identified. The minimum distance between markers to estimate a useful LD was found to be 54.5 kb for the Mexican HO populations. At this average distance, the persistence of phase across autosomes of Conv and Lowi systems was 0.94, for Conv and CAN + USA was 0.92 and for the Lowi and CAN + USA was 0.91. Results supported the flow of germplasm among populations being Conv a source for Lowi, and dependent on migration from CAN + USA. Mexican HO cattle in Conv and Lowi populations share common ancestry with CAN + USA but have different genetic signatures.

Keywords: genetic differentiation, Holstein, admixture, linkage disequilibrium

INTRODUCTION

Dairy farms in Mexico are extremely heterogeneous (e.g., different herd sizes, feeding systems, reproductive management, etc.). Conventional (Conv) dairy farms have an average herd size of 230 head and are highly mechanized and milk yield is relatively high. Cows on these farms are typically grouped in pens and rations usually include high proportions of concentrates. Low income (Lowi) systems vary with region with sizes ranging from 3 to 30 cows, with animals usually spending part of the day grazing, but may also be housed in pens (Amendola, 2002). Additionally, Lowi farms rely heavily on the use of unpaid family labor and the typical herd is smaller than Conv herds. The Mexican Holstein (HO) population has depended on US and Canada genetics for some years (Powell and Wiggans, 1991). Recently, germplasm from European populations have been introduced, however, these populations also tend to be highly influenced by US and Canadian bulls (Norman and Powell, 1999). This information was expected in the Conv Mexican

systems where pedigree information was available, but little was known about the genetic influence on the Lowi system because of incomplete or unavailable pedigree information.

Before single nucleotide polymorphism (SNP) data became available, genetic diversity and relatedness was studied through the analysis of pedigree information which, unfortunately, is not always complete. This incomplete data renders results inaccurate or limits the performance of studies of individuals or populations (Boichard et al., 1997). Currently, with the availability of SNP data, it is possible to estimate breed or population compositions without the previous knowledge of ancestry information (Sölkner et al., 2010; Frkonia et al., 2012). Different analyses based on genomic information have been used to study genetic diversity of populations. In this study, principal component analysis (PCA) was performed to describe the breed or geographic allele variation, whereas admixture analysis was used to describe the structure of populations. Persistence phase of linkage disequilibrium (LD)

analysis allowed the characterization of the degree of agreement of LD across distances between populations.

Principal component analysis was first used in human populations to generate maps summarizing the allele frequency of different geographic areas (Cavalli-Sforza et al., 1994; Silva-Zolezzi et al., 2009) and more recently for understanding relationship among cattle breeds (Bovine HapMap Consortium, 2009; Lewis et al., 2011). Currently, PCA is used to control spurious genome wide association in populations structured with individuals of different geographic areas or even when geography does not explain the genetic background (Novembre and Stephens, 2008). In this study, PCA was used to identify differences among HO cattle originating from CAN + USA and two Mexican production systems.

Admixture is defined as the mixing of genomes of divergent parental origins (Buerkle and Lexer, 2008), which implies the presence of multiple genetically distinct groups or breeds in a population (Wang et al., 2005). Admixture can be studied at an individual (Tang et al., 2005) or population level (Buerkle and Lexer, 2008; Frkonia et al., 2012). These haplotype blocks vary in size because of the random nature of recombination, but become progressively shorter by further recombination with increasing generations (Winkler et al., 2010).

Linkage disequilibrium is defined as a non-random association of alleles at different loci (Du et al., 2007; Waples and England, 2011) because the recombination rate differs from that expected if the loci segregated independently. LD is common between alleles at neighboring loci that tend to be inherited together and associated in a segregating population (Du et al., 2007), but can also be associated with selection (Bulmer, 1971). Characterization of LD is used to assess whether two or more populations can be jointly analyzed in genomic studies, because markers in LD in one population may not be in LD in another population (de Roos et al., 2008), and to make meaningful inferences in populations other than the reference population will depend on the persistence of LD phase between the two populations (Dekkers and Hospital, 2002). The LD level in a population is also used for determining the required marker map resolution to be used in a genomic selection program and testing associations based on QTL scans (McKay et al., 2007). The LD of sufficiently large degree that allows the QTL scan is known as the useful LD (Lu et al., 2012).

To establish genomic evaluations in Mexico, it is important to determine whether multiple HO subpopulations exist. This information determines whether it is necessary to stratify the population by production system and if foreign genomic information, especially from Canada and US, would improve accuracy of predicted breeding values. Thus, the objective of this work was to study the genetic composition and differentiation of the Mexican HO population and the relationship between Mexican cattle with those from Canada and the US. The primary source of stratification of the Mexican population considered was the production system in which those cows performed.

MATERIALS AND METHODS

ANIMALS, BREEDS AND GENOTYPES

A total of 625 HO and HO like unrelated cows and sires, born on or after 2005 and genotyped with the Illumina BovineSNP50 or

BovineLD Bead Chips were used in this analysis. A total of 149 and 303 animals assigned to the Conv and Lowi systems, respectively, and 173 were from the Canadian and US HO populations. In addition, 162 Jersey (JE) and 86 Brown Swiss (BS) sires from Canada and US were included to help determine if crossbred cows were present in the populations. The animals from the Conv system originated from 16 herds in 6 states of Mexico (Aguascalientes, Guanajuato, Estado de México, Querétaro, San Luis Potosí, and Zacatecas) while the animals of the Lowi system were from 21 herds in 4 states (Estado de México, Jalisco, Puebla, and Tlaxcala). From the 6,836 common SNP markers in both the Illumina BovineSNP50 and the BovineLD Bead Chips a total of 6,617 SNP were included in the analysis after quality control. Markers with a minor allele frequency less than 2% and call rate less than 90% were excluded. Individuals with a call rate less than 90% were also excluded. **Table 1** shows the number and frequency of markers per chromosome included in the analysis.

Table 1 | Number and frequency of single nucleotide polymorphism (SNP) per chromosome included in the analysis.

Chromosome	Number of SNP by chromosome	SNP percentage by chromosome
1	398	6.01
2	349	5.27
3	310	4.68
4	307	4.64
5	304	4.59
6	310	4.68
7	284	4.29
8	299	4.52
9	272	4.11
10	269	4.07
11	285	4.31
12	225	3.40
13	218	3.29
14	226	3.42
15	223	3.37
16	206	3.11
17	195	2.95
18	176	2.66
19	182	2.75
20	204	3.08
21	186	2.81
22	169	2.55
23	152	2.30
24	177	2.67
25	142	2.15
26	145	2.19
27	139	2.10
28	128	1.93
29	137	2.07

POPULATION COMPOSITION

The PCA proposed by Price et al. (2006) was used in this study because it models ancestry differences along continuous axes of variation. Genotypes of BS, HO, and JE animals were used. Sire country of origin was considered to explain variation among HO populations and for Mexican cattle; the production system was used as classification variable resulting in three distinct HO populations (Lowi, Conv, and CAN + USA).

GENETIC STRUCTURE OF MEXICAN HO

The ADMIXTURE package (Alexander et al., 2009) was used for this purpose because it implements a fast model-based estimation that assumes that individuals come from an admixed population with contributions from K ancestral populations. Each K population contributes a fraction q_{ik} for each individual i . Only Mexican HO animals were used for this analysis. Once the K ancestral populations were determined, the most common sire's country of origin of each ancestral population was identified based on pedigree information.

LINKAGE DISEQUILIBRIUM DIFFERENCES AMONG HO POPULATIONS

The LD, measured as r^2 for alleles at two loci was calculated as:

$$r^2 = \frac{D_{ij}^2}{p_1 p_2 q_1 q_2}$$

Where D is the difference between the observed and the expected frequency of two loci, based on population allele frequencies and assuming random assortment and can be estimated directly from the allele frequencies (Waples and England, 2011), $p_1 p_2 q_1$ and q_2 are the observed frequencies of alleles 1, 2 respectively (Hill and Robertson, 1968). The value, r^2 , is considered the most robust measure of LD. Persistence phase of LD was calculated as the Pearson correlation coefficient between the root of r^2 between populations for the same pair of SNP (Badke et al., 2012).

Quality control, PCA and LD analysis were performed with SVS Golden Helix software (SNP and Variation Suite Manual v7, 2013; Golden Helix, Inc.), persistence phase of LD was calculated using SAS 9.2 (SAS Institute, 2009).

RESULTS

POPULATION COMPOSITION

In this study, the first three of 625 components of PCA explained 13% of the observed variation. Population differentiation is observed between the Mexican Lowi and Conv subgroups and the Canadian and US HO cattle (Figure 1), despite many common ancestors across these groups. The Conv system seems to be intermediate between the Lowi and CAN + USA groups. Neither Mexican system demonstrated a clustering of animals by country of origin of the sire. PCs for all North American HO cattle along with the JE and BS cattle are shown in Figure 2, where the individuals were color coded by breed and population of origin. Crossbred individuals derived from HO and JE or BS were represented by points located between those pure breeds. Crossbreeding was rare in the Conv herds but was much more common in the Lowi systems with a higher proportion of crossing to JE influenced animals than BS.

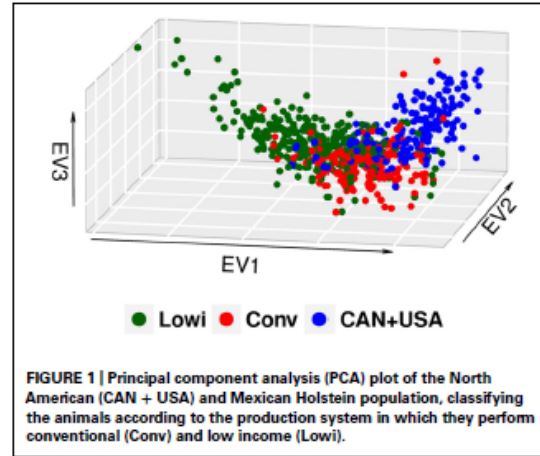


FIGURE 1 | Principal component analysis (PCA) plot of the North American (CAN + USA) and Mexican Holstein population, classifying the animals according to the production system in which they perform conventional (Conv) and low income (Lowi).

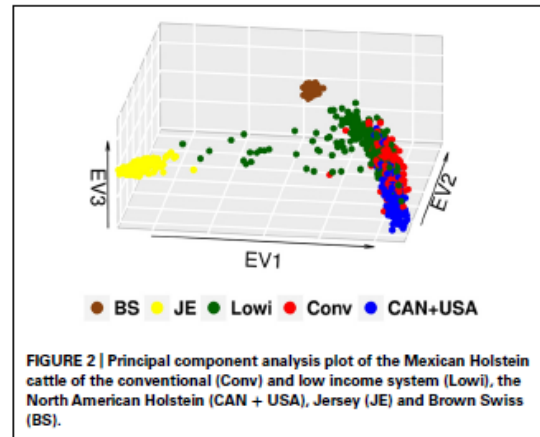
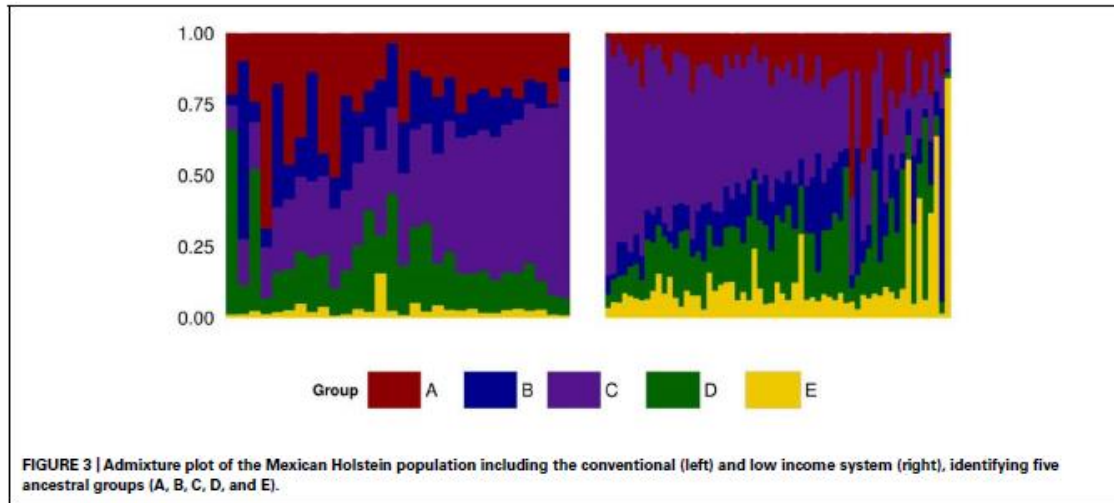


FIGURE 2 | Principal component analysis plot of the Mexican Holstein cattle of the conventional (Conv) and low income system (Lowi), the North American Holstein (CAN + USA), Jersey (JE) and Brown Swiss (BS).

GENETIC STRUCTURE OF MEXICAN HO

To identify the genetic structure of the Mexican HO systems and its dependency on the Canadian and US populations, an admixture analysis was performed first for the entire Mexican population and then for each sub-population. For both analyses, the value that best explained the stratification of the population was $K = 6$. Pedigree records of individual Mexican HO identified their sire's country of origin as Canadian, US, or Mexican, hence providing an indirect measure of the genetic contribution of these countries to Mexican HO. Indeed, the ADMIXTURE strata identified at $K = 6$ (Figure 3) correlated to the sire's country of origin. Two of the ADMIXTURE strata were directly related to sires originating from Canada and were subsequently combined into group A with the purpose of investigating country of origin influence. Group B consisted of sires most commonly from the United States (USA). The third strata, group C, consisted of sires from both Canada and US. No unique lineages or discernable characteristics were associated with group C. Another stratum was associated to sires registered in the Mexican herd book and assigned as group

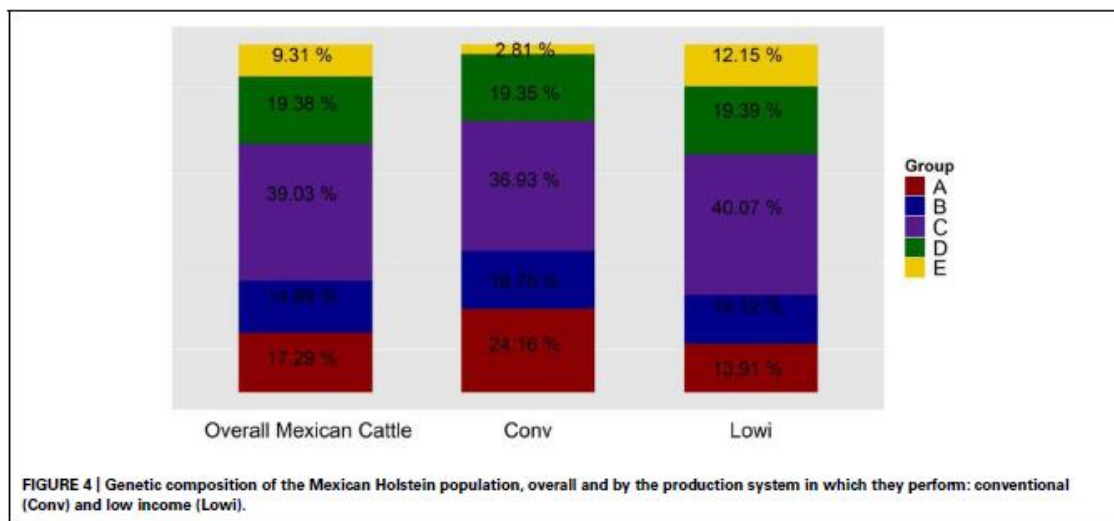


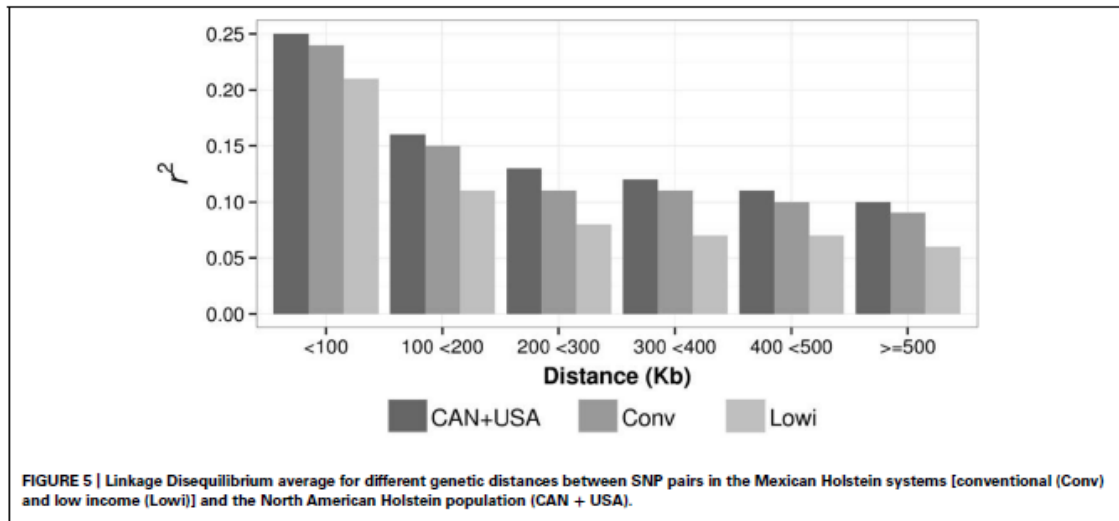
D. The last stratum included crossbred animals as evidenced by their genetic similarity to BS and JE breeds in the PC analysis and were assigned to group E. Average ancestry contributions, both overall and by production system, can be observed in **Figure 4**. Population structure showed differences in the genetic composition of the Mexican HO production systems where group C had the largest contribution to the Conv and Lowi populations at 37 and 40%, respectively. In the Conv population, group A (24%), D (19%), B (17%) and E (2.8%) followed while for Lowi animals, group C was followed by D (19%), B (14%), A (14%) and the E (13%). Overall, the contributions of the US, Canada/US, and the Mexican sires were relatively similar among the production systems. The primary differences were within the Canadian

and crossbred lineages. The Conv population had approximately 1.7 fold greater contribution from Canadian lineage (group A; Conv-24.16% vs. Lowi-13.91%) and the Lowi population had ~4.3 fold increase in crossbred lineages (group E; Conv-2.81% vs. Lowi-12.51%).

LINKAGE DISEQUILIBRIUM DIFFERENCES AMONG HO POPULATIONS

Mean LD, calculated as r^2 for different distances (at intervals of 100 Kbp) between SNP were calculated for the Conv, Lowi, and CAN + USA populations (**Figure 5**). At all distances, average r^2 was highest for CAN + USA animals, intermediate for individuals from the Mexican Conv farms, and smallest for cattle representing the Lowi systems. The differences between the CAN + USA and





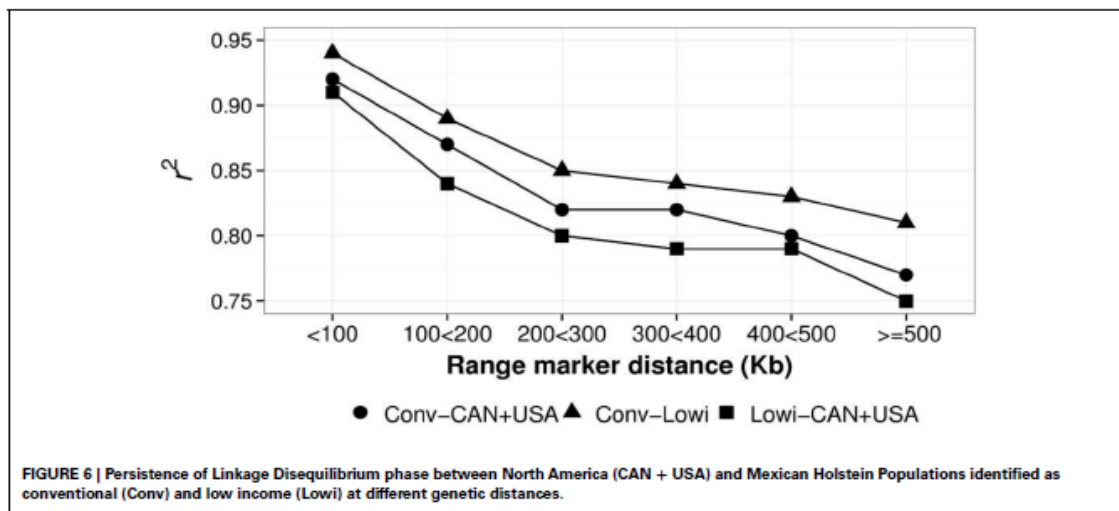
Conv populations r^2 were quite small (~ 0.01) while differences between Conv and Lowi were larger and consistent, ranging from 0.03 to 0.04. The persistence of LD phase between Conv, Lowi, and CAN + USA populations was calculated at the same interval distances as was LD (Figure 6). As expected, the persistence of LD phase decreased when the distance between markers was increased. At all intervals, the highest correlations were between Conv and Lowi populations, with the lowest correlation being between Lowi and CAN + USA. At a distance of <100 Kb (with an average of 54.5 kb), the correlations ranged from 0.91 to 0.94 and for a distance > 500 kb correlations varied from 0.75 to 0.81 between the Conv-Lowi and Lowi-CAN + USA, respectively.

DISCUSSION

The results obtained in this study, provide important details to be considered for future genetic research on the Mexican HO cattle and provided the opportunity to measure the genetic relationship of the Mexican animals with the HO populations of Canada and the US.

The Conv system seems to be intermediate in most aspects between the Lowi and CAN + USA, and this population may act as a conduit for germplasm flow from Canada and the US to Mexico.

When the PCA in the HO breed was explored, a difference in the genetic structure was observed between the Mexican production



systems and the CAN + USA population, although population overlap suggests that all three groups share common genetic material. In this case the Conv system seems to be a link between the Lowi system and CAN + USA population because breeders of the Conv system provide genetic material (heifers and semen) to the Lowi one and the Conv system has depended genetically from US and Canada for many years (Powell and Wiggans, 1991). The Lowi also obtains genetic material directly from CAN + USA, but at a lower proportion than the Conv one. No specific tendency to group animals for geographic area was found in the Mexican populations, but as in other studies (Novembre and Stephens, 2008), these analyses give us an idea of the genetic background of the population.

The inclusion of JE and BS resulted in individual clusters for both breeds as expected, but more importantly, they helped define the differences between the Lowi and the Conv populations. Results suggest that breeders of the Lowi system occasionally used genetic material of other breeds.

Admixture analysis showed that five different populations, linked to origin country of the sires, comprise the ancestral background of the Mexican HO populations. The two systems showed variation in the average proportion of genetic similarity to the different ancestral populations. The results show a substantial influence of the North American HOs on the Mexican population, and agree with previous results in the Conv system reported by Powell and Wiggans (1991). JE and BS breed influence is visible in both Mexican production systems, but stands out in the Lowi where the use of these dairy breeds seems to be more common. The MEX population was linked through pedigree information to registered Mexican HO bulls related to North American animals. This block also included individuals with unknown sires presumably with similar origins to those of the Mexican sires.

In line with previous studies, decay of LD was observed in this study when the distance between markers was increased (Figure 5; Du et al., 2007; de Roos et al., 2008; Sargolzaei et al., 2008; Badke et al., 2012). In general, average r^2 estimates for the populations in this study were in the range of those reported in other HO populations (McKay et al., 2007; de Roos et al., 2008; Sargolzaei et al., 2008; Qanbari et al., 2010; Zhou et al., 2013), although r^2 averages for the CAN + USA are slightly higher than those reported for the same population in a group of animals born after 1990 at similar distances between markers (Sargolzaei et al., 2008), except when the distance is less than 100 Kb. Note that the LD of the CAN + USA were higher than the Mexican systems at all compared genetic distances and those of the CONV are slightly higher than those of Lowi. The lowest r^2 values among the populations were for the Lowi. These lower values may be the result of breeders in the Lowi system introducing other breeds through crossbreeding and incorporating gene migration and genetic drift. This practice may explain the reduction in LD (Lynch and Walsh, 1998). The r^2 averages for the Lowi decreased rapidly when the distance increased from <100 to 100 and <200 Kb then decreased slowly at distances >200 Kb.

A practical application of LD is determining the number of markers necessary to perform genome wide associations studies

(Gautier et al., 2007; McKay et al., 2007; de Roos et al., 2008) at a useful LD ($r^2 \geq 0.20$; de Roos et al., 2008). The useful LD of 54.5 kb found in this study, suggests that at least 53,000 markers should be used to perform genomic analysis within this population. Similar numbers of markers were suggested for other cattle populations (McKay et al., 2007; de Roos et al., 2008).

Because the persistence phase or correlation of r among populations show the genetic relationship between them (Badke et al., 2012), it was used as a measure among the HO populations included in this analysis. Results shown in Figure 6 confirm the results of PCA, because the higher persistence phase was reported between the Conv and Lowi, followed by the one between Conv and CAN + USA ending with confirmation of the lowest relationship between Lowi and CAN + USA. As it was also reported in other studies (de Roos et al., 2008; Badke et al., 2012) the correlation of r among all populations decreased rapidly when the distance between markers increased. The difference between persistence phases among the Conv, Lowi, and CAN + USA ranged from 0.01 to 0.04, lower than that present in other breeds like Angus, Charolais, and JE (de Roos et al., 2008; Lu et al., 2012) or species like pigs (Badke et al., 2012). At distances <100 Kb, the persistence phase between the Conv and Lowi, Conv and CAN + USA, and CAN + USA and Lowi were lower than that reported between Chinese and Nordic HO cattle (0.97; Zhou et al., 2013) and at all measured intervals, similar values were found between Dutch black and white and Dutch red and white HO Friesian bulls and lower values were reported for Australian bulls and New Zealand Friesian cows (de Roos et al., 2008).

Results showed that the US and Canadian and the Mexican HO cattle of the Conv and Lowi have different genetic structures although these populations share much common ancestry. The main difference between the Mexican HO systems is the result of crossbreeding with other breeds, especially in the Lowi system. If joint genomic studies are to be performed between these populations, stratification of populations is recommended. Joint genetic improvement programs of HO animals across North America, i.e., including Mexico, may be established as these populations share genetic material. The useful LD founded in this populations, will determine the minimum number of SNP markers need if joint genomic studies are to be performed.

The considerable similarity between the Conv subgroup with US and Canadian populations means that integration of these groups would be straightforward and should be considered.

AUTHOR CONTRIBUTIONS

The authors have made the following declarations about their contributions: Conceived and designed the experiments Adriana García-Ruiz, Felipe de J. Ruiz-López, Curtis P. Van Tassell, and Hugo H. Montaldo. Performed experiments and analyze data: Adriana García-Ruiz Data acquisition and interpretation: Adriana García-Ruiz, Felipe de J. Ruiz-López, Curtis P. Van Tassell, Hugo H. Montaldo, and Heather J. Huson. Wrote the paper: Adriana García-Ruiz. All authors approve the manuscript final version.

ACKNOWLEDGMENTS

We would like to thank the Mexican HO Association for providing samples of the animals included in the analysis, the Council on

Dairy Cattle Breeding (CDCB) for providing material included in the analysis and Dr. George Wiggins, for his contribution and the edition of databases.

This study was supported by CONACYT, CONARGEN and the research projects: Study of Genetic Diversity of Mexican HO Cattle based on Genomic Information (SIGI: 1523542158) and Incorporation of Genomic information in the Genetic Evaluation Process of Mexican Dairy Cattle (SIGI: 1056821832).

REFERENCES

- Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast Model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109
- Amendola, R. D. (2002). *A Dairy System Based on Forages, and Grazing in Temperate Mexico*. Ph.D. thesis, Wageningen University, Wageningen.
- Badke, Y. M., Bates, R. O., Ernst, C. W., Schwab, C., and Steibel, J. P. (2012). Estimation of linkage disequilibrium in four US pig breeds. *BMC Genomics* 13:24. doi: 10.1186/1471-2164-13-24
- Boichard, D., Maigner, L., and Verrier, É. (1997). The value of using probabilities of gene origin to measure genetic variability in a population. *Genet. Sel. Evol.* 29, 5–23. doi: 10.1051/gse:19970101
- Bovine HapMap Consortium. (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324, 528–532. doi: 10.1126/science.1167936
- Buerkle, C. A., and Lexer, C. (2008). Admixture as the basis for genetic mapping. *Trends Ecol. Evol.* 23, 686–694. doi: 10.1016/j.tree.2008.07.008
- Bulmer, M. G. (1971). The effect of selection on genetic variability. *Am. Nat.* 105, 201–211. doi: 10.1086/282718
- Cavalli-Sforza, L. L., Menozzi, P., and Piazza, A. (1994). *The History, and Geography of Human Genes*. Princeton, NJ: Princeton University Press. doi: 10.1086/282718
- Dekkers, J. C., and Hospital, F. (2002). The use of molecular genetics in the improvement of agricultural populations. *Nat. Rev. Genet.* 3, 22–32. doi: 10.1038/nrg701
- de Roos, A. P. W., Hayes, B. J., Spelman, R. J., and Goddard, M. E. (2008). Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics* 179, 1503–1512. doi: 10.1534/genetics.107.084301
- Du, F. X., Clutter, A. C., and Lohuis, M. M. (2007). Characterizing linkage disequilibrium in pig populations. *Int. J. Biol. Sci.* 3, 166–178. doi: 10.7150/ijbs.3.166
- Frkonia, A., Gredler, B., Schnyder, U., Curik, I., and Sölkner, J. (2012). Prediction of breed composition in an admixed cattle population. *Anim. Genet.* 43, 696–703. doi: 10.1111/j.1365-2052.2012.02345.x
- Gautier, M., Faraut, T., Moazami-Goudarzi, K., Navratil, V., Foglio, M., Grohs, C., et al. (2007). Genetic and haplotypic structure in 14 European and African cattle breeds. *Genetics* 177, 1059–1070. doi: 10.1534/genetics.107.075804
- Hill, W. G., and Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* 38, 226–231. doi: 10.1111/j.1365-2052.2012.02345.x
- Lewis, J., Abas, Z., Dadousis, C., Lykidis, D., Paschou, P., and Drineas, P. (2011). Tracing cattle breeds with principal components analysis ancestry informative SNPs. *PLoS ONE* 6:e18007. doi: 10.1371/journal.pone.0018007
- Lu, D., Sargolzaei, M., Kelly, M., Li, C., Vander Voort, G., Wang, Z., et al. (2012). Linkage disequilibrium in Angus, Charolais, and Crossbred beef cattle. *Front. Genet.* 3:152. doi: 10.3389/fgene.2012.00152
- Lynch, M., and Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer Associates.
- McKay, S. D., Schnabel, R. D., Murdoch, B. M., Matukumalli, L. K., Aerts, J., Coppeters, W., et al. (2007). Whole genome linkage disequilibrium maps in cattle. *BMC Genet.* 8:74. doi: 10.1186/1471-2156-8-74
- Norman, H. D., and Powell, R. L. (1999). Dairy cows of high genetic merit for yields of milk, fat and protein. *Asian Australas. J. Anim. Sci.* 12, 1316–1323. doi: 10.1186/1471-2156-8-74
- Novembre, J., and Stephens, M. (2008). Interpreting principal component analyses of spatial population genetic variation. *Nat. Genet.* 40, 646–649. doi: 10.1038/ng.139
- Powell, R. L., and Wiggins, G. R. (1991). Animal model evaluations for Mexican Holsteins. *J. Dairy Sci.* 74, 1420–1427. doi: 10.3168/jds.S0022-0302(91)78298-2
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909. doi: 10.1038/ng1847
- Qanbari, S., Pimentel, E. C. G., Tetens, J., Thaller, G., Lichtner, P., Sharifi, A. R., et al. (2010). A genome-wide scan for signatures of recent selection in Holstein cattle. *Anim. Genet.* 41, 377–389. doi: 10.1111/j.1365-2052.2009.02016.x
- Sargolzaei, M., Schenkel, F. S., Jansen, G. B., and Schaeffer, L. R. (2008). Extent of linkage disequilibrium in Holstein cattle in North America. *J. Dairy Sci.* 91, 2106–2117. doi: 10.3168/jds.2007-0553
- SAS Institute. (2009). *SAS OnlineDoc 9.2. SAS Institute Inc., Cary, NC (Online)*. Available at: http://support.sas.com/documentation/cdl_main/index.html
- Silva-Zolezzi, I., Hidalgo-Miranda, A., Estrada-Gil, J., Fernandez-Lopez, J. C., Uribe-Figueroa, L., Contreras, A., et al. (2009). Analysis of genomic diversity in Mexican Mestizo populations to develop genomic medicine in Mexico. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8611–8616. doi: 10.1073/pnas.0903045106
- SNP and Variation Suite Manual v7. (2013). *Golden Helix, Inc., Bozeman, MT, (Online)*. Available at: www.goldenhelix.com
- Sölkner, J., Frkonia, A., Raadsma, H. W., Jonas, E., Thaller, G., Gootwine, E., et al. (2010). Estimation of individual levels of admixture in crossbred populations from SNP chip data: examples with sheep and cattle populations. *Interbull Bull.* 42, 62–66.
- Tang, H., Peng, J., Wang, P., and Risch, N. J. (2005). Estimation of individual admixture: analytical and study design considerations. *Genet. Epidemiol.* 28, 289–301. doi: 10.1002/gepi.20064
- Wang, W. Y., Barrat, B. J., Clayton D. G., and Todd, J. A. (2005). Genome-wide association studies: theoretical and practical concerns. *Nat. Rev. Genet.* 6, 109–118. doi: 10.1038/nrg1522
- Waples, R. S., and England, P. R. (2011). Estimating contemporary effective population size on the basis of linkage disequilibrium in the face of migration. *Genetics* 189, 633–644. doi: 10.1534/genetics.111.132233
- Winkler, A. C., Nelson, W. G., and Smith, W. M. (2010). Admixture mapping comes of age. *Annu. Rev. Genomics Hum. Genet.* 11, 65–89. doi: 10.1146/annurev-genom-082509-141523
- Zhou, L., Ding, X., Zhang, Q., Wang, Y., Lund, M. S., and Su, G. (2013). Consistency of linkage disequilibrium between Chinese and Nordic Holsteins and genomic prediction for Chinese Holsteins using a joint reference population. *Genet. Sel. Evol.* 45:7. doi: 10.1186/1297-9686-45-7

Conflict of Interest Statement: The Review Editor Ikhid G. Immorin declares that, despite being affiliated with the same institute as the author Heather J. Huson, the review process was handled objectively. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 17 August 2014; accepted: 09 January 2015; published online: 09 February 2015.

Citation: García-Ruiz A, Ruiz-López FJ, Van Tassell CP, Montaldo HH and Huson HJ (2015) Genetic differentiation of Mexican Holstein cattle and its relationship with Canadian and U.S. Holsteins. *Front. Genet.* 6:7. doi: 10.3389/fgene.2015.00007

This article was submitted to *Livestock Genomics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2015 García-Ruiz, Ruiz-López, Van Tassell, Montaldo and Huson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

CAPÍTULO 5

Huellas de Selección en Regiones Genómicas de Ganado Holstein de México, Canadá y Estados Unidos

Este capítulo será abordado a través de un artículo científico sometido en la Revista Mexicana de Ciencias Pecuarias en Noviembre de 2014, documento se encuentra en el formato en el que fue presentado.

**Huellas de Selección en Regiones Genómicas de Ganado Holstein de México,
Canadá y Estados Unidos**

**Signatures of Selection on Genomic regions of Holstein Cattle in Mexico, Canada
and the United States of America.**

Adriana García R.^{1,2}, Felipe de J. Ruiz L.^{1,2}, Curtis P. VanTassell³ y Hugo H. Montaldo⁴

¹ Facultad de Estudios Superiores Cuautitlán, Universidad Nacional Autónoma de México.

² Centro Nacional de Investigación Disciplinaria en Fisiología y Mejoramiento Animal, Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, SAGARPA.

³ Bovine Functional Genomics Laboratory, USDA Beltsville MD.

⁴ Departamento de Genética y Bioestadística, Facultad de Medicina Veterinaria y Zootecnia, Universidad Nacional Autónoma de México.

Autor de correspondencia: Felipe de J Ruiz López.

Email: ruiz.felipe@inifap.gob.mx

Dirección: Km.1 de la Carretera a Colón, poblado de Ajuchitlán, Municipio de Colón, Querétaro C.P. 76280

Teléfono: (419) 292 0036

Fax: (419) 292 0033

RESUMEN

La presencia de huellas de selección en el genoma, se puede detectar por la frecuencia de bloques en desequilibrio de ligamiento (LD) en ciertas poblaciones. El objetivo de este estudio fue localizar los bloques del genoma (>150 Kb) que se encuentran en LD en la población Holstein de México, Canadá y Estados Unidos, y determinar si estas regiones del genoma se encuentran asociadas a características de interés económico en éstas y otras poblaciones. En el presente estudio se incluyeron 322 animales Holstein no emparentados, de los cuales 149 fueron del sistema especializado en México y 173 de Estados Unidos y Canadá. Después de un control de calidad, se incluyeron un total de 6,847 marcadores de un solo polimorfismo (SNP). El LD entre los marcadores SNP se calculó a través de r^2 , utilizando el software SVS Golden Helix y los bloques del genoma en LD más grandes, fueron referenciados en base de datos de ganado bovino (QTLdb). Los bloques en LD mayores a 150 kb en las poblaciones fueron localizados en los cromosomas 3, 7, 8, 9, 12, 14, 16 y 20. Siete de las ocho regiones en LD han sido previamente referenciadas en estudios de asociación o identificación de QTL en ganado Holstein. En las regiones anteriores se incluyen rasgos de salud, producción y composición de la leche y fertilidad. Las regiones en LD del cromosoma 7 y 12 no contaron con referencias previas. Las regiones en LD asociadas a características de interés económico, podrían explicar los antecedentes de selección genética en ambas poblaciones.

PALABRAS CLAVE: Desequilibrio de ligamiento, Ganado Lechero.

INTRODUCCIÓN

La selección natural y la inducida a través de los programas de mejoramiento genético son aspectos que determinan la persistencia o eliminación de material genético en las poblaciones. La selección continua y positiva hacia un conjunto de genes en una población específica puede resultar en la formación de regiones en LD (1,2) y éstas se pueden mantener a través de varias generaciones en la población.

Las huellas de selección son regiones en el genoma que han incrementado su presencia y se han establecido en poblaciones específicas debido a su importancia funcional (3). Existen diversas metodologías propuestas para determinar la presencia de huellas de selección en el genoma (4, 5, 6, 7). Una forma de detectar éstas regiones, es a través del estudio de la persistencia de la regiones en desequilibrio de ligamiento (7, 8), que también es utilizado para analizar otros eventos en las poblaciones, como la selección, deriva génica, mutación u origen genético de poblaciones (9, 10). El desequilibrio de ligamiento (LD) se define como la asociación no aleatoria de alelos en diferentes loci que al encontrarse cercanos en el genoma son heredados juntos en una población segregante (11, 12), lo que genera falta de independencia entre dos o más loci. Cuando dos o más loci se encuentran en LD en una población, no necesariamente se encuentran en la misma condición en otras poblaciones (13).

Diversas metodologías se han propuesto para calcular el LD entre marcadores genéticos (14, 15, 16). Se considera que la r^2 como la medida más robusta para evaluar el LD (10, 12).

La disponibilidad de paneles densos de marcadores genéticos ha generado una gama amplia de publicaciones científicas que hacen referencia a regiones del genoma que se encuentran asociadas a características de interés económico o de salud en diferentes poblaciones. En el caso de los animales domésticos, miles de QTL han sido reportados (<http://www.animalgenome.org/QTLdb/>) en diferentes especies, poblaciones o razas (17). Esta información puede ser utilizada libremente para conocer referencias previas de regiones del genoma de acuerdo a la posición de los marcadores.

El objetivo del presente estudio fue identificar las regiones del genoma que están en LD en las poblaciones Holstein de México y de Estados Unidos y Canadá y determinar si estas regiones del genoma se encuentran asociadas a características de interés económico que se encuentran bajo selección genética en diferentes poblaciones de ganado lechero.

MATERIALES Y MÉTODOS

En el presente estudio se incluyeron genotipos de 322 animales de la raza Holstein sin parentesco conocido entre ellos, para que, donde fuera posible, disminuir la probabilidad de formar bloques que provinieran de alelos idénticos por descendencia. De los animales incluidos 149 provinieron de la población Holstein de México y 173 de Estados Unidos de América y Canadá. Para el análisis se utilizó información de 6,842 marcadores de un solo polimorfismo (SNP por sus siglas en inglés) contenidos en el Bovine LD v1.1 Bead Chip de Illumina. Los marcadores con frecuencias del alelo menor (FAM) menor al 2% y tasa de asignación menor del 90% fueron excluidos del análisis; así como los individuos con tasa de asignación menor al 90%.

Después del control de calidad, en el estudio se incluyeron 6,647 marcadores de SNP de 322 animales.

Posteriormente se calculó el LD para cada par de marcadores. LD se calculó de la siguiente manera:

$$r^2 = \frac{D_{ij}^2}{p_1 p_2 q_1 q_2}$$

Donde p_1, p_2, q_1 y q_2 son las frecuencias observadas de los alelos 1 y 2 respectivamente y D_{ij}^2 es el cuadrado de D_{ij} que se calcula como:

$$D_{ij} = p(A_i B_j) - p(A_i) p(B_j)$$

Donde $p(A_i B_j)$ es la frecuencia del haplotipo $A_i B_j$ y $p(A_i)$ y $p(B_j)$ son las frecuencias de los alelos A_i y B_j respectivamente. Para la formación de haplotipos, se siguió la metodología descrita por Gabriel, et al (18). El uso de r^2 para calcular LD, tiene la ventaja de eliminar el posible sesgo generado por las frecuencias iniciales de los alelos (14). La r^2 se calculó utilizando el software SVS Golden Helix (19) y se detectaron los bloques del genoma en LD con mayor longitud y con valor de $r^2 \geq 0.5$ entre marcadores, seleccionando regiones del genoma considerando como grandes a las que incluyeran 150 Kb o más.

Para buscar asociaciones previas o QTLs reportados en las regiones que se encontraron en LD entre las poblaciones estudiadas, se obtuvo la base de datos que contiene información de QTLs o marcadores genéticos asociados a diferentes características de interés económico (dbQTL) en otras poblaciones de ganado lechero (UMD3.1 *Bos taurus* genome). Esta esta bases de datos destaca la información del cromosoma, posición del marcador o QTL en pares de bases, longitud de la región

genómica y la(s) característica(s) asociada(s) en cada región e información de la publicación. Posteriormente se buscaron las referencias de cada una de las regiones grandes en LD encontradas en este estudio, para lo cual fue necesario hacer referencia a la posición inicial y final del conjunto de marcadores en LD.

RESULTADOS

Se encontraron 8 regiones grandes del genoma en LD con una longitud mayor a 150Kb. El rango de longitud de las regiones grandes del genoma varió de 188 a 1358 Kb. Los bloques del genoma en LD fueron localizados en los cromosomas 3, 7, 8, 9, 12, 14, 16 y 20, con tamaños de 640, 1358, 241, 188, 221, 321, 611 y 317 Kb respectivamente. En el cuadro 1 se presenta la lista de marcadores que se encuentran en LD en cada cromosoma y el valor de r^2 entre cada par de marcadores. El valor de r^2 entre los marcadores varía de 0.57 a 0.99 entre los diferentes marcadores. Como ejemplo, en la Figura 1 se muestra la región en LD del cromosoma 20, que incluye en el eje horizontal la lista de SNP que se encuentran dentro del bloque en LD y en el vertical el valor de r^2 , el cual se presenta a través de una gama de colores de rojo a blanco y de blanco a azul, el rojo más intenso representando una r^2 de 1 y su tonalidad va disminuyendo hasta tomar un color blanco, que representa una r^2 de 0.5. A partir del color blanco, el valor de r^2 disminuye aún más hasta llegar a 0, el cual está representado, por un color azul intenso. En el eje vertical a la derecha de la figura se presenta la escala de r^2 a través de los colores mencionados anteriormente. El valor de r^2 entre dos marcadores (clasificados como marcador inicial y final) está representado

por el color que se presenta en la intersección de las diagonales de cada uno de los marcadores.

Las regiones que se encontraron en LD se buscaron en la base de datos bovina de QTL's y se encontró que siete de las ocho regiones en LD han sido previamente referenciadas en estudios de asociación o identificación de QTL en ganado Holstein. Las regiones en LD encontradas han sido asociadas a diversas características como composición de la leche (grasa o proteína en porcentaje y producción), producción de leche, rasgos de salud, conformación y fertilidad. En el cuadro 2 se muestra el listado de asociaciones genéticas o QTL que se encuentran en las regiones de LD identificadas en el presente estudio. De todas las en LD, las que se encuentran en el cromosoma 7 y 12 no tuvieron referencias de QTL o asociaciones con características de interés económico.

DISCUSIÓN

La población Holstein de México ha dependido de material genético de la población de Estados Unidos y Canadá por varias décadas, lo que podría influir en la persistencia de huellas de selección en el genoma (a través de la presencia de bloques en LD) entre ambas poblaciones. Además, la selección de características de interés económico en ambas poblaciones son similares, lo que podría explicar los antecedentes de selección genética en las ambas poblaciones. Cabe mencionar que en la mayoría de estas regiones en LD en las poblaciones Holstein de México y de Estados Unidos y Canadá, están detectados QTLs asociados a las características de importancia económica. En la región en LD del cromosoma 3, se ha encontrado la presencia de regiones

genómicas o QTL asociados a producción de leche (20), grasa (21) y proteína (22). En el cromosoma 8, QTLs asociados a muerte fetal, distocia (23), mastitis clínica y conteo celular somático (24). En el cromosoma 9, QTLs asociados a características de fertilidad, como número de inseminaciones por concepción, tasa de no retorno (25) e intervalo entre la primera y la última inseminación (26). En la región en LD del cromosoma 14, se han detectado la presencia del gen DGAT1, que afecta la producción de grasa (27, 28, 29, 30), leche (27, 30) y proteína (30, 31), tasa de no retorno, anchura del anca, y características de conformación asociadas al sistema de estructura y capacidad y patas y pezuñas (30). En el cromosoma 16, QTL's que afectan facilidad de parto (32), desplazamiento de abomaso (33) así como producción de leche (34, 35) y proteína (34). En el cromosoma 20, QTL's asociados con producción de proteína (36, 37), de tamaño de la cría (38), conteo celular somático (39) y susceptibilidad a *Mycobacterium paratuberculosis* (40).

Los bloques grandes del genoma que se encontraron en LD, hacen referencia a regiones del genoma que se encuentran asociadas a características de interés económico, para las cuales se han implementado programas de selección en ambas poblaciones, especialmente en Estados Unidos y Canadá, donde el número de características incorporadas en los programas de mejoramiento genético es mayor. En la población Holstein de México, su persistencia podría atribuirse a la dependencia de material genético que esta población ha tenido de las de Estados Unidos y Canadá.

La permanencia de grandes bloques en LD, puede ser también un factor importante en los programas de selección genómica, sobre todo cuando se requieren conjuntar animales genotipados con chips de diferente densidad de y realizar las imputaciones de

genotipos. De ésta manera, se podría aumentar el número de marcadores en la población de referencia, lo que a su vez favorece las confiabilidades y predicciones genómicas (41).

La conservación de las regiones de LD detectadas en las poblaciones de estudio podría incrementar la tasa de mejoramiento genético en la población Holstein de México, sobre todo de las características que aún no son incorporadas en los programas de selección en México y se puede realizar un mejoramiento para múltiples características. Además, se podría realizar un programa de selección basado en la presencia de haplotipos que se encuentren en LD con características de interés económico.

CONCLUSIONES E IMPLICACIONES

A pesar de la relativamente baja densidad de los paneles de marcadores usados para el presente estudio, se encontraron 8 grandes bloques en LD, de los cuales seis de ellos se encuentran asociados a características de interés económico en diversas poblaciones de ganado Holstein. Las características a las cuales se hacen referencia en las regiones en LD son objeto de selección, lo que podría implicar que esas regiones del genoma han estado bajo selección genética en ambas poblaciones y son huellas de selección generadas por los programas de mejoramiento genético utilizados. La conservación de regiones en LD entre las poblaciones de estudio podría mejorar el proceso de selección genética, mejorando la precisión de las evaluaciones y por lo tanto las tasas de mejoramiento genético.

AGRADECIMIENTOS

El presente estudio ha sido parcialmente financiado el proyecto INIFAP Num. 1523542158, “Estudio de la diversidad genética de ganado Holstein determinada con base en información genómica en México” y el proyecto INIFAP Num. 1056821832, “Incorporación de información genómica a los procesos de evaluación genética del ganado productor de leche”.

Los autores agradecen el apoyo a CONACYT, la asociación Holstein de México al Animal Improvement Programs Laboratory AIPL-USDA, Beltsville MD y al Council on Dairy Cattle Breeding (*CDCB*), por proporcionar información genómica usada en el presente estudio.

ABSTRACT

Signatures of selection on genomic material could be detected by the presence of Linkage Disequilibrium (LD) blocks in specific populations. The objective of this study was to find genomic blocks in LD (greater than 150 kb) in the Mexican and in the Canadian and U.S. Holstein populations and to investigate if those regions are related to previous genome annotations. The analysis included information of 322 unrelated Holstein animals, 149 from the Mexican conventional system and 173 from the U.S. and Canadian populations. After a genome quality control analysis, a total of 6,847 Single Nucleotide Polymorphism (SNP) were included in the analysis. The LD between the SNP markers was calculated through the r^2 , using the SVS Golden Helix software, and the largest blocks were referenced in the cattle QTL database (QTLdb). The largest block markers (>150 Kb) in LD in the populations were located in chromosomes 3, 7, 8, 9, 12, 14, 16 and 20. Seven of the eight LD regions have been related to association studies or QTL identifications in Holstein cattle. The references are related to health traits, to milk yield and composition, and fertility traits. LD region of chromosome 7 and 12 has not references for association of QTL in Holstein cattle. Traits referenced in the LD regions, could explain the genetic selection background in the studied populations.

KEYWORDS: Linkage disequilibrium, Dairy cattle

LITERATURA CITADA

1. Stephan W, Song YS, Langley CH: The hitchhiking effect on linkage disequilibrium between linked neutral loci. *Genetics* 2006, 172:2647–2663.
2. Biswas S, Akey JM: Genomic insights into positive selection. *Trends Genet* 2006, 22:437–446.
3. Pérez O'Brien AM, Utsunomiya YT, Mészáros G, Bickhart DM, Liu GE, Van Tassell CP, et al. Assessing signatures of selection through variation in linkage disequilibrium between taurine and indicine cattle. *Genet Sel Evol* 2014; 46: 19. doi:10.1186/1297-9686-46-19
4. Rubin CJ, Zody MC, Eriksson J, Meadows JRS, Sherwood E et al. Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* 2010; 464: 587–593.
5. Oleksyk TK, Zhao K, De La Vega FM, Gilbert DA, O'Brien SJ et al. 2008. Identifying Selected Regions from Heterozygosity and Divergence Using a Light-Coverage Genomic Dataset from Two Human Populations. *PLoS ONE* (3): e1712. doi: 10.1371.
6. Grossman SR, Shylakhter I, Karlsson EK, Byrne EH, Morales S et al. A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science* 2010; 327: 883–886.
7. Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ et al. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 2002; 419: 832–837.

8. Tang K, Thornton KR, Stoneking M. A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biology* 2007; 5: e171.
9. Pritchard JK, Przeworski M. Linkage disequilibrium in humans: models and data. *Amer J of Hum Genet* 2001; 69(1): 1–14.
10. Waples RS, Do C. Linkage disequilibrium estimates of contemporary N_e using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evol Appl* 2010; 3(3): 244–262.
11. Waples RS, England PR. Estimating contemporary effective population size on the basis of linkage disequilibrium in the face of migration. *Genetics* 2011; 189(2): 633–644.
12. Du FX, Clutter AC, Lohuis MM. Characterizing linkage disequilibrium in pig populations. *Inter J of Biol Scie* 2007; 3(3): 166–178.
13. de Roos APW, Hayes BJ, Spelman RJ, Goddard ME. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics* 2008; 179(3):1503-1512.
14. Hill WG, Robertson A. Linkage Disequilibrium in Finite Populations. *Theor Appl Genet J* 1968; 38: 226–231.
15. Hedrick PW. Gametic disequilibrium measures: proceed with caution. *Genetics* 1987; 117: 331-341.
16. Zhao H, Nettleton D, Soller M, Dekkers JCM. Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between markers and QTL. *Genet Res* 2005; 86(1): 77-87.

17. Zhang H, Wang Z, Wang S, Li H. Progress of genome wide association study in domestic animals. *J of Anim Scie Biotech* 2012; 3(1): 26.
18. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, et al. The structure of haplotype blocks in the human genome. *Science* 2002; 296: 2225–2229
19. SNP & Variation Suite Manual v7. Copyright © 2013 Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com.
20. Daetwyler HD, Schenkel FS, Sargolzaei M, Robinson JA. A genome scan to detect quantitative trait loci for economically important traits in Holstein cattle using two methods and a dense single nucleotide polymorphism map. *J of Dairy Scie* 2008; 91(8):3225-3236.
21. Liu Y, Jansen GB, Lin CY. Quantitative trait loci mapping for dairy cattle production traits using a maximum likelihood method. *J Dairy Scie*, 2004; 87(2):491-500.
22. Pimentel EC, Bauersachs S, Tietze M, Simianer H, Tetens J, Thaller G, Reinhardt F, Wolf E, König S. Exploration of relationships between production and fertility traits in dairy cattle via association studies of SNPs within candidate genes derived by expression profiling. *Anim Genet*, 2011; 42(3):251-62.
23. Kühn CH, Bennewitz J, Reinsch N, Xu N, Thomsen H, Looft C, et al. Quantitative trait loci mapping of functional traits in the German Holstein cattle population. *J of Dairy Scie* 2003; 86(1): 360-368.
24. Klungland H, Sabry A, Heringstad B, Olsen HG, Gomez-Raya L, Vage DI, et al. Quantitative trait loci affecting clinical mastitis and somatic cell count in dairy cattle. *Mamm Genome* 2001; 12(11): 837-842.

25. Holmberg M, Andersson-Eklund L. Quantitative trait loci affecting fertility and calving traits in Swedish dairy cattle. *J of Dairy Scie* 2006; 89(9):3664-3671.
26. Hoglund JK, Guldbrandtsen B, Su G, Thomsen B, Lund MS. Genome scan detects quantitative trait loci affecting female fertility traits in Danish and Swedish Holstein cattle. *J of Dairy Scie* 2009; 92(5): 2136-2143.
27. Boichard DB, Rohs CG, Bourgeois F, Cerqueira F, Faugeras R, Neau A, et al. Detection of genes influencing economic traits in three French dairy cattle breeds. *Genet Sel Evol* 2003; (35): 77–101.
28. Bennewitz JB, Einsch NR, Rohs CG, Rockmann GAB, Chwerin MS, Eimann, CW. Combined analysis of data from two granddaughter designs?: A simple strategy for QTL confirmation and increasing experimental power in dairy cattle. *Genet. Sel. Evol.* 2003;(35): 319–338.
29. Bennewitz J, Reinsch N, Paul S, Looft C, Kaupe B, Weimann C, et al. The DGAT1 K232A mutation is not solely responsible for the milk production quantitative trait locus on the bovine chromosome 14. *J of Dairy Scie.* 2004; 87(2):431-42.
30. Kaupe B, Brandt H, Prinzenberg EM, Erhardt G. Joint analysis of the influence of CYP11B1 and DGAT1 genetic variation on milk production, somatic cell score, conformation, reproduction, and productive lifespan in German Holstein cattle. *J of Anim Scie* 2007; 85(1): 11–21.
31. Casas E, Shackelford SD, Keele JW, Koohmaraie M, Smith TPL, Stone RT. Detection of Quantitative Trait Loci for Growth and Carcass Composition in Cattle. *J Anim Sci*, 2003; 81: 2976-2983.

32. Schnabel RD, Sonstegard TS, Taylor JF, Ashwell MS. Whole-genome scan to detect QTL for milk production, conformation, fertility and functional traits in two US Holstein families. *Anim Genet* 2005; 36(5): 408-16.
33. Momke S, Scholz H, Doll K, Rehage J, Distl O. Mapping quantitative trait Loci for left-sided displacement of the abomasum in German Holstein dairy cows. *J Dairy Scie*, 2008; 91 (11): 4383-92.
34. Daetwyler HD, Schenkel FS, Sargolzaei M, Robinson JA. A genome scan to detect quantitative trait loci for economically important traits in Holstein cattle using two methods and a dense single nucleotide polymorphism map. *J of Dairy Scie* 2008; 91(8):3225-3236.
35. Chamberlain AJ, Hayes BJ, Savin K, Bolormaa S, McPartlan HC, Bowman PJ, et al. Validation of single nucleotide polymorphisms associated with milk production traits in dairy cattle. *J Dairy Scie*, 2012; 95 (2): 864-75.
36. Olsen HG, Gomez-Raya L, Våge DI, Olsaker I, Klungland H, Svendsen M, et al. A genome scan for quantitative trait loci affecting milk production in Norwegian dairy cattle. *J of Dairy Scie* 2002; 85(11): 3124–3130.
37. Lund MS, Sorensen P, Madsen P, Jaffrezic F. Detection and modelling of time-dependent QTL in animal populations. *Gen Sel Evol*, 2008; 40 (2): 177-94.
38. Thomasen JR, Guldbrandtsen B, Sorensen P, Thomsen B, Lund MS. Quantitative trait loci affecting calving traits in Danish Holstein cattle. *J of Dairy Scie* 2008; 91(5): 2098-2105. doi: 10.3168/jds.2007-0602.

39. Rodriguez-Zas SL, Southey BR, Heyen DW, Lewin HA. Interval and composite interval mapping of somatic cell score, yield, and components of milk in dairy cattle. *J of Dairy Scie* 2002; 85(11): 3081-3091.
40. Kirkpatrick BW, Shi X, Shook GE, Collins MT. Whole-Genome association analysis of susceptibility to paratuberculosis in holstein cattle. *Anim Genet*, 2011; 42(2): 149-160.
41. Zhou L, Ding X, Zhang Q, Wang Y, Lund MS, Su G. Consistency of linkage disequilibrium between Chinese and Nordic Holsteins and genomic prediction for Chinese Holsteins using a joint reference population. *Genet Sel Evol* 2013; 45(1):7.

CUADROS Y GRÁFICAS

Cuadro 1. Lista de marcadores que se encuentran en desequilibrio de ligamiento en cada cromosoma en los animales Holstein de México y de Estados Unidos y Canadá.

Cromosoma	Marcador inicial	Marcador final	r^2
3	ARS-BFGL-NGS-111118	BTB-02075948	0.59
3	ARS-BFGL-NGS-112952	Hapmap46132-BTA-29718	0.70
3	ARS-BFGL-NGS-112952	INRA-417	0.70
3	ARS-BFGL-NGS-43853	ARS-BFGL-NGS-112952	0.82
3	ARS-BFGL-NGS-43853	Hapmap46132-BTA-29718	0.57
3	ARS-BFGL-NGS-43853	INRA-417	0.57
3	Hapmap46132-BTA-29718	INRA-417	0.99
7	Hapmap41358-BTA-79117	ARS-USMARC-Parent-DQ650635-rs29012174	0.61
7	Hapmap41358-BTA-79117	Hapmap35191-BES11_Contig367_1030	0.63
7	ARS-USMARC-Parent-DQ650635-rs29012174	Hapmap35191-BES11_Contig367_1030	0.94
7	ARS-USMARC-Parent-DQ650635-rs29012174	Hapmap51384-BTA-108597	0.67
8	BTB-01258879	ARS-BFGL-NGS-37680	0.60
8	BTB-01258879	Hapmap49049-BTA-119736	0.69
8	ARS-BFGL-NGS-37680	ARS-BFGL-NGS-16472	0.66
8	ARS-BFGL-NGS-37680	Hapmap49049-BTA-119736	0.87
8	Hapmap49049-BTA-119736	ARS-BFGL-NGS-16472	0.60
9	Hapmap49335-BTA-83363	Hapmap48549-BTA-99373	0.94
9	Hapmap49335-BTA-83363	BTA-94256-no-rs	0.94
9	Hapmap49335-BTA-83363	Hapmap44492-BTA-88469	0.96
9	BTA-94256-no-rs	Hapmap48549-BTA-99373	1.00
9	BTA-94256-no-rs	Hapmap44492-BTA-88469	0.98
9	Hapmap44492-BTA-88469	Hapmap48549-BTA-99373	0.98
12	ARS-BFGL-NGS-30815	ARS-BFGL-NGS-112364	0.99
12	ARS-BFGL-NGS-30815	Hapmap40318-BTA-122815	0.99
12	ARS-BFGL-NGS-30815	Hapmap43221-BTA-27481	1.00
12	ARS-BFGL-NGS-112364	Hapmap40318-BTA-122815	1.00
12	ARS-BFGL-NGS-112364	Hapmap43221-BTA-27481	0.99
12	Hapmap43221-BTA-27481	Hapmap40318-BTA-122815	0.99
14	Hapmap30383-BTC-005848	ARS-BFGL-NGS-4939	0.59
14	Hapmap30383-BTC-005848	ARS-BFGL-NGS-57820	0.64

14	BTA-34956-no-rs	ARS-BFGL-NGS-34135	0.66
14	BTA-34956-no-rs	ARS-BFGL-NGS-94706	0.64
14	ARS-BFGL-NGS-57820	ARS-BFGL-NGS-4939	0.91
14	ARS-BFGL-NGS-34135	ARS-BFGL-NGS-94706	0.94
16	ARS-BFGL-NGS-53805	ARS-BFGL-NGS-13513	0.55
16	ARS-BFGL-NGS-53805	ARS-BFGL-NGS-22302	0.71
16	ARS-BFGL-NGS-53805	ARS-BFGL-NGS-37851	0.71
16	ARS-BFGL-NGS-22302	ARS-BFGL-NGS-13513	0.76
16	ARS-BFGL-NGS-22302	ARS-BFGL-NGS-37851	1.00
16	ARS-BFGL-NGS-37851	ARS-BFGL-NGS-13513	0.76
20	ARS-BFGL-NGS-23012	ARS-BFGL-NGS-112221	0.92
20	ARS-BFGL-NGS-23012	ARS-BFGL-NGS-113227	0.75
20	ARS-BFGL-NGS-23012	ARS-BFGL-NGS-114380	0.79
20	ARS-BFGL-NGS-114380	ARS-BFGL-NGS-112221	0.86
20	ARS-BFGL-NGS-114380	ARS-BFGL-NGS-113227	0.71
20	ARS-BFGL-NGS-112221	ARS-BFGL-NGS-113227	0.81
20	ARS-BFGL-NGS-113227	ARS-BFGL-NGS-118166	0.57

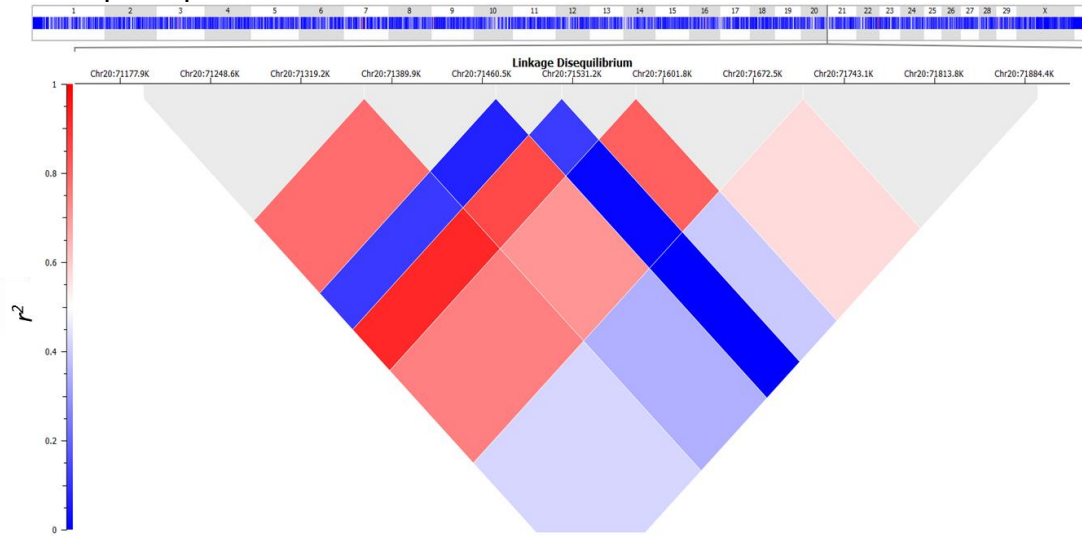
- 1 Cuadro 2. Posición y referencias de las asociaciones genéticas encontradas en las regiones en desequilibrio de
 2 ligamiento en la población de estudio.

QTL-ID	CARACTERÍSTICA ASOCIADA	GEN CANDIDATO	CROMOSOMA	POSICIÓN INICIAL*	POSICIÓN FINAL *	TIPO DE ANÁLISIS	REFERENCIA
6155	Producción láctea		3	37375696	37567489	Asociación genómica	20
10382	Porcentaje de grasa		3	45563451	45755244	QTL	21
16221	Porcentaje de proteína	HFM1 (HFM1, ATP-dependent DNA helicase homolog (S. cerevisiae))	3	45755244	45947036	Asociación genómica	22
11442	Distocia		8	44393492	89588554	QTL	23
11443	Muerte fetal		8	44393492	89588554	QTL	23
2497	Mastitis clínica	Prnp (prion protein)	8	41285255	80980500	QTL	24
2498	Conteo Celular Somático	Prnp (prion protein)	8	41285255	80980500	QTL	24
3565	Inseminaciones por concepción		9	0	75409523	QTL	25
3566	Tasa de no retorno		9	2148415	9159784	QTL	25
5006	Intervalo entre primera y última inseminación		9	0	75409523	QTL	26
3523	Producción de leche		14	1641277	12858984	QTL	27
1333	Porcentaje de grasa	DGAT1	14	1641277	12892474	Asociación genómica	27, 28, 29, 30
3407	Producción de grasa	DGAT1	14	1641277	12892474	Asociación genómica	27, 28, 29, 31
3522	Producción de leche		14	1641277	12892474	Asociación genómica	30

3524	Tasa de no retorno	14	1641277	12892474	Asociación genómica	30
3525	Anchura del anca	14	1641277	12892474	Asociación genómica	30
3526	Sistema de estructura corporal, patas y pezuñas	14	1641277	12892474	Asociación genómica	30
3520	Porcentaje de proteína	14	1641277	12892474	Asociación genómica	30, 31
3521	Producción de proteína	14	1641277	12892474	Asociación genómica	30, 31
3527	Producción láctea	14	1641277	12093363	QTL	31
1700	Facilidad de parto	16	39070184	62523468	QTL	32
5123	Desplazamiento de abomaso	16	25756620	55832211	QTL	33
6219	Producción de leche (Valor genético)	16	33660604	66993830	Asociación genómica	34
6220	Producción de proteína (Valor genético)	16	33660604	66993830	Asociación genómica	34
20167	Producción de leche	16	30257623	55009939	Asociación genómica	35
4669	Tamaño de cria	20	61960897	67895566	QTL	30
2722	Producción de proteína	20	35800031	67895566	QTL	36
10103	Producción de proteína	20	277899	71951703	QTL	37
10188	Conteo Celular Somático	20	277899	71951703	QTL	39
14870	Susceptibilidad a M. paratuberculosis	20	67670883	67843716	Asociación genómica	40

3 *La posición inicial y final se encuentra en especificada en pares de bases en el cromosoma.

Figura 1. Mapa de disequilibrio de ligamiento de un conjunto de marcadores genéticos ubicados en el cromosoma 20 en los animales Holstein de México y Estados Unidos y Canadá. El valor de r^2 entre dos marcadores está representado por el color que se presenta en la intersección de las diagonales de cada uno de los marcadores. A la derecha de la gráfica se presenta la gama de colores de rojo a azul que representan la escala de los valores de r^2 .



CAPÍTULO 6

Efecto del tamaño de la población de referencia y los genotipos de ancestros disponibles sobre el proceso de imputación de genotipos Holstein Mexicanos

El presente capítulo será abordado con la presentación de un artículo científico que fue aceptado para su publicación en Febrero de 2015 en la revista: Journal of Dairy Science. El documento se encuentra en el formato en el que fue sometido.

Effect of reference population size and available ancestor genotypes on imputation of Mexican Holstein genotypes. *By García-Ruiz et al., page 000.* The effects of reference population size and the availability of information from genotyped ancestors on the accuracy of imputation of genetic markers were investigated for Mexican Holstein cattle. A larger reference population and the availability of genotyped ancestors improved imputation; animals with genotyped parents in a large reference population had higher imputation accuracy than those with no or few genotyped relatives in a small reference population. For small local populations, including genotypes from other related populations can be an important tool for improving accuracy of imputation.

Effect of reference population size and available ancestor genotypes on imputation of Mexican Holstein genotypes¹

A. García-Ruiz,* F. J. Ruiz-Lopez,*^{†2} G. R. Wiggans,‡ C. P. Van Tassell‡, and H. H. Montaldo*

* Facultad de Estudios Superiores Cuautitlán, Facultad de Medicina Veterinaria y Zootecnia, Universidad Nacional Autónoma de México, México, DF 04510, México

†Centro Nacional de Investigación en Fisiología y Mejoramiento Animal, Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, Ajuchitlán, Querétaro 76280, México

‡Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD, 20705-2350

Corresponding author: Felipe de Jesús Ruiz-López
CENID Fisiología y Mejoramiento Animal
Carretera Colón-Querétaro Km 1, Ajuchitlán, Colón
Querétaro.
ruiz.felipe@inifap.gob.mx

Received November 21, 2014.

Accepted February 02, 2015.

¹The use of trade, firm, or corporation names in this publication is for the information and convenience of the reader. Such use does not constitute an official endorsement or approval by the US Department of Agriculture or the Agricultural Research Service of any product or service to the exclusion of others that may be suitable.

²Corresponding author: ruiz.felipe@inifap.gob.mx

ABSTRACT

The effects of reference population size and the availability of information from genotyped ancestors on the accuracy of imputation of single nucleotide polymorphisms (SNP) were investigated for Mexican Holstein cattle. Three scenarios for reference population size were examined: (1) a local population of 2,011 genotyped Mexican Holsteins, (2) animals in scenario 1 plus 866 Holsteins in the US genotype database (GDB) with genotyped Mexican daughters, and (3) animals in scenario 1 and all US GDB Holsteins (338,073). Genotypes from 4 different chip densities (2 low density, 1 mid-density, and 1 high density) were imputed using findhap (version 3) to the 45,195 markers on the mid-density chip. Imputation success was determined by comparing the numbers of SNP with 1 or 2 alleles missing and the numbers of differently predicted SNP (conflicts) among the 3 scenarios. Imputation accuracy improved as chip density and numbers of genotyped ancestors increased, and the percentage of SNP with 1 missing allele was greater than that for 2 missing alleles for all scenarios. The largest numbers of conflicts were found between scenarios 1 and 3. The inclusion of information from direct ancestors (dam or sire) with US GDB genotypes in the imputation of Mexican Holstein genotypes increased imputation accuracy by 1 percentage point for low-density genotypes and 0.5 percentage points for high-density genotypes, which was about half the gain found with information from all US GDB Holsteins. A larger

reference population and the availability of genotyped ancestors improved imputation; animals with genotyped parents in a large reference population had higher imputation accuracy than those with no or few genotyped relatives in a small reference population. For small local populations, including genotypes from other related populations can aid in improving imputation accuracy.

Keywords: imputation, Mexican Holstein, ancestor genotype, reference population

INTRODUCTION

Genomic selection is a recent technology that has enhanced genetic improvement in dairy cattle, but its success relies on numerous factors. The most important are methodology for calculating marker effects (Hayes et al., 2009; Daetwyler et al., 2010; VanRaden et al., 2010), marker panel density (Solberg et al., 2008; Hayes et al., 2009), trait heritability (Hayes et al., 2009), and number of genotyped animals in the reference population (Hayes et al., 2009). To improve reliability of genomic predictions, increasing the number of genotyped animals is more important than using higher density panels (VanRaden et al., 2010). In developing countries such as Mexico, genotyping is still expensive, and high-density marker panels generally are more expensive than low-density panels. Using low-density panels is an alternative that may lead to more genotyped animals.

Because genomic technology has evolved rapidly, the number and sets of markers to be used for genomic prediction have changed over time (Hayes et al., 2009). To combine information based on different markers or marker densities, missing markers for animals genotyped with lower density panels are imputed from genotype information of relatives or from haplotypes of animals genotyped with higher density panels (Druet et al., 2010; VanRaden et al., 2010). Druet et al. (2010) showed that imputation efficiency is higher when the size of the reference population, marker density, and expected proportion of the genome inherited from

the reference population are increased. Bouwman et al. (2014) reported that the number of genotyped relatives has a direct effect on imputation accuracy for animals without genotypes, and Khatkar et al. (2012) demonstrated the importance of considering genotyped relatives when imputing low-density genotypes, particularly when the sire's genotype was not included in the reference genotypes. Although different imputation methods with high accuracy have been implemented (Browning and Browning, 2011; VanRaden et al., 2011; Hickey et al., 2012), the choice of the optimal method depends on population structure (Johnston et al., 2011).

Using imputation as part of genomic selection reduces genotyping costs and increases both the size of the reference population and the number of markers for which effects are estimated, which increases the reliability of genetic predictions and consequently the expected genetic improvement. Reducing genotyping cost also makes the technology more accessible to breeders (Berry and Kearney, 2011; VanRaden et al., 2011). For dairy cattle, reliability of genomic predictions using imputation varies according to trait and according to population size and structure. Reliability improvements of approximately 2 percentage points have been reported in simulation studies that included a set of 500,000 SNP per imputed genotype compared with a 50,000-marker subset (VanRaden et al., 2011).

Canadian and U.S. dairy cattle genetics are widely used around the world. Using genotypes from those animals to impute local genotypes would increase the size of the local reference population considerably and could increase the accuracy of imputing genotypes, thus improving genomic evaluations in many countries. Such an improvement should be possible for Mexico, because Mexican Holstein breeders have depended heavily on US and Canadian genetics. The primary objective of this study was to determine the effect on imputation of Mexican Holstein genotypes from increasing the size of the Mexican Holstein reference population by adding Holsteins from the US genotype database. The effect of relatedness of animals in the reference population on imputation also was evaluated.

MATERIALS AND METHODS

Scenarios

Three imputation scenarios were defined based on source and number of genotyped animals. For scenario 1, only genotypes of a local population of 2,011 genotyped Mexican Holsteins were included. For scenario 2, genotypes of animals in scenario 1 plus genotypes of 866 Holsteins in the US genotype database (**GDB**) with genotyped Mexican daughters were included. For scenario 3, genotypes of

animals in scenario 1 and all US GDB Holsteins available at the time of the study (338,073) were included.

Data

Genotypes. The genotyped Mexican Holstein population (scenario 1) included 1,971 cows and 40 sires. For cows, 183 were genotyped with the Illumina BovineLD BeadChip v1.1 (**6K**; Illumina, 2013), 277 with the GeneSeek Genomic Profiler BeadChip v1 (**9K**; Neogen Corporation, 2013a), 686 with the *Illumina BovineSNP50 BeadChip v2* (**50K**; Illumina, 2011), and 825 with the GeneSeek Genomic Profiler HD BeadChip (**77K**; Neogen Corporation, 2013b). All Mexican sires had 50K genotypes. Of the US GDB Holsteins, 839 bulls and 47 cows had genotyped daughters in Mexico (scenario 2). All US GDB dams of genotyped Mexican daughters had 50K genotypes; the US GDB sires included 533 US, 270 Canadian, and 22 European bulls with 50K genotypes and 10 US and 4 Canadian bulls with 77K genotypes. For scenario 3, 338,073 US GDB genotypes were included.

Pedigrees. Two different pedigree files were used in the analysis: 27,625 animals for scenarios 1 and 2; 938,662 animals for scenario 3.

Imputation

Missing genotypes were predicted by combining population and pedigree haplotypes with findhap (version 3) software (VanRaden, 2014). The imputation goal for all scenarios was to fill in any missing genotypes for the 45,195 SNP from the 50K chip that were then being used in US genomic evaluations. Of those 45,195 SNP, 6,842 were included in 6K genotypes, 8,196 in 9K genotypes, and 28,048 in 77K genotypes, and a different number of SNP were imputed for each chip density: 38,353 for 6K genotypes, 36,999 for 9K genotypes, and 17,147 for 77K genotypes. For imputation, genotypes were first coded as 0 = BB, 1 = AB, 2 = AA, or 5 = both alleles unknown. Then SNP genotypes for each scenario were imputed, and the results were compared using SAS (version 9.3; SAS Institute Inc., Cary NC).

A SNP genotype is not always able to be determined through imputation. If only 1 parental allele could be determined (i.e., B and unknown allele or A and unknown allele), then the SNP genotype was designated as M1. If neither parental contribution could be determined (i.e., both alleles missing), the SNP genotype was designated as M2. Squared correlations (R^2) between estimated and true SNP genotypes (VanRaden et al., 2011), percentages of missing alleles (M1 and M2), and differences in imputed SNP (conflicts) between scenarios were used to determine the success of genomic imputation (imputation accuracy). Khatkar et al. (2012) reported that increasing the number of reference animals has a positive

effect on imputation. Therefore, scenario 3 was expected to be the most accurate for imputation, and scenarios 1 and 2 were compared with scenario 3. Conflict frequency was calculated as the percentage of 50K SNP that were conflicts between scenarios.

Ancestor Genotypes

Genotyped animals were grouped by category of genotyped ancestors: animals without parental genotypes; animals with dam or sire genotyped; and animals with dam, sire, and a grandsire genotyped. Animals with other combinations of genotyped ancestors were not considered because of small group sizes. Numbers of M1 and M2 SNP genotypes and SNP conflicts were compared using SAS within chip density by ancestor group. Only scenarios 2 and 3 were analyzed, because scenario 1 had too few genotyped ancestors to form groups.

Comparison of Imputed and True Genotypes

To determine the accuracy of imputed genotypes from low-density chips (6K and 9K), 10 groups were formed by random sampling of the 686 cows without replacement; each group contained 10 cows with 50K genotypes. Group size was limited to 10 cows to avoid excessive reduction in imputation accuracy from removing many high-density genotypes that are critical for accurate SNP

imputation; a group size of 10 cows was considered large enough that a single animal could not have undue influence. For each animal, only 6K or 9K SNP were included in the 50K genotype, and the genotype was filled through imputation and compared with the original 50K genotype. The same animal groups were used for both chip densities. The total numbers of missing SNP (**MM**), which was calculated as $M1 + M2$, and conflicts were used to compare true and imputed genotypes.

RESULTS AND DISCUSSION

Imputation Accuracy

Imputation accuracy was improved by increasing genotype chip density and the number of reference animals (Figure 1). Imputation R^2 using only local genotypes (scenario 1) were 96, 96, 99, and 99% when imputing from 6K, 9K, 50K, and 77K genotypes, respectively. When information from US GDB Holsteins with genotyped Mexican daughters was added to the Mexican genotypes (scenario 2), imputation R^2 increased by almost 1 percentage point for 6K and 9K genotypes and by half a percentage point for 77K genotypes. When all US GDB Holstein data were included (scenario 3) and compared with the results for scenario 1, an increase of almost 3 percentage points was observed for 6K and 9K genotypes and 1 percentage point for 77K genotypes. As expected, no increase in imputation R^2 was found for 50K genotypes in any scenario because of the small number of SNP that actually were

imputed. These results were consistent with those reported in other studies using the same (Wiggans et al., 2012) or different (Zhang and Druet, 2010; Johnston et al., 2011; Khatkar et al., 2012) imputation methods. All studies reported higher imputation accuracy with increased number of reference animals.

Missing Alleles

Percentages of 50K alleles that were missing decreased as more animals were added to the reference population for both M1 and M2 SNP genotypes (Table 1) regardless of chip density used for genotyping. Numbers of animals with missing alleles in ≥ 1 SNP genotype followed that pattern only for 6K, 9K, and 77K M2 SNP genotypes. As expected, the corresponding M1 and M2 percentages by scenario were lower for 50K genotypes than for SNP genotypes from other chip densities, especially for M1 SNP. These results can be explained partially by the imputation process, which first divides each chromosome into segments, determines the haplotypes, and then matches genotypes with the haplotype list (VanRaden et al., 2011). When the original genotype has more markers, haplotype matching is more precise and imputation accuracy tends to be higher. In addition, the proportion of SNP to be imputed decreases as chip density increases.

Percentage of SNP genotypes designated as M1 was larger than that for M2 for all scenarios regardless of chip density except for animals with 50K genotypes in

scenario 1. The M1 percentage was largest for 9K genotypes followed by 6K and 77K genotypes for scenarios 1 and 2. For scenario 3, the M1 percentage was largest for 6K genotypes followed by 9K and 77K genotypes. The M2 percentage generally decreased as number of SNP to be imputed increased for all scenarios.

Conflicts

The number of detected conflicts increased as the number of genotyped animals or the number of imputed SNP (Table 2) increased. Conflict frequency never exceeded 3.8%, possibly because the 6K chip was the lowest density chip studied. Conflicts have a greater effect on evaluation reliability than do missing markers because missing markers are replaced by population allele frequencies during the calculation of genomic evaluations (Wiggans et al., 2011).

Ancestor Genotypes

For animals grouped by availability of genotyped ancestors, Table 3 shows differences between scenarios 2 and 3 for M1, M2, and conflict numbers by genotyping chip density and ancestor genotype group. Imputation problems for scenario 2 were more numerous with fewer genotyped ancestors, a pattern also reported by Johnston et al. (2011), who concluded that the number of correctly imputed genotypes increases as relatedness between genotyped ancestors and

animals to be imputed increases. Other studies on genotype imputation from low to high density (e.g., Carvalheiro et al., 2014; Ma et al., 2013) reported only marginal benefits on imputation accuracy as the relatedness between imputed and reference animals increased. The greater benefit in this study may be caused partially by the imputation method used. In a comparison of 5 methods for imputing Swedish and Finnish Red Cattle marker genotypes, Ma et al. (2013) reported that IMPUTE2 (Howie et al., 2009) and BEAGLE (Browning and Browning, 2009) had the greatest imputation accuracy; however the effect of relatedness on percentage of correctly imputed alleles was largest for findhap (version 2; VanRaden, 2014) and other imputation methods that used pedigree information. Carvalheiro et al. (2014) used BEAGLE as well as FImpute (Sargolzaei et al., 2014) to study Nelore cattle; although they found higher imputation accuracies with FImpute than with Beagle, they concluded that the optimal imputation method is unique to the genotyping chip and the population studied. When Khatkar et al. (2012) examined the effect of including sire genotype in the imputation process, a slightly lower allelic error rate was found when sire was included (2.61% with sire vs. 3.34% without sire). Similar findings were observed in this study, although the effect of including a parent was less for 50K and 77K genotypes than for 6K and 9K genotypes (Table 3).

In addition to reference population size and relatedness of genotypes, chip density was also a factor that determined imputation success between scenarios 2 and 3. Within genotyped ancestor group, animals with 50K genotypes had the fewest imputation problems, and differences between scenarios 2 and 3 decreased slightly with more ancestors with 50K genotypes (Table 3). Because imputation improvement only pertains to SNP that are not called during genotyping and the 50K chip contains all SNP, the number of correctly filled SNP increased slightly when genotypes for immediate family members were available.

The largest differences between scenarios 2 and 3 for imputation problems (17.2% of SNP) were found for animals with 6K genotypes and no parental genotypes (Table 3). However, comparison with the other genotyped ancestor groups was not possible because only animals with no parental genotypes had 6K genotypes. For animals with 9K genotypes, those with only sire or dam genotyped had more imputation problems than those with a grandsire also genotyped (14.8 vs. 3.3% of SNP). Animals with 77K genotypes had imputation problems for 7.8% of SNP for scenario 2 compared with scenario 3 when no parental genotypes were available, which decreased to 6.7% when a dam or sire genotype was available and to 1.6% when genotypes were available for dam, sire, and a grandsire. When the reference population is small, related genotyped animals with high density genotypes should be included to increase imputation success.

Comparison of Imputed and True Genotypes

Imputed genotypes were validated only for scenarios 1 and 2, because only extremely small differences were found between scenarios 2 and 3 and computational demands of the imputation process would have been large for scenario 3. Within scenario, more MM were found for 9K genotypes than for 6K genotypes (Table 4), whereas the number of conflicts was slightly higher for 6K genotypes. The percentage of markers with 1 or 2 missing alleles (Table 4) was higher for 9K genotypes (7.2% for scenario 1 and 5.8% for scenario 2) than for 6K genotypes (6.6 and 5.3%, respectively). However, 6K genotypes had a higher conflict rate (2.8% for scenario 1 and 2.3% for scenario 2) than did 9K genotypes (2.6 and 2.1%, respectively). Similar error rates (2.7%) were reported by Zhang and Druet (2010) when a panel of 6,000 markers was imputed to 45,836 markers using a reference group of 2,000 animals. Weigel et al. (2010) reported error rates of 3.6 to 5.8% for imputation from 8,680 markers using a reference group of 2,542 Jerseys, which was slightly higher than the 9K conflict rates in this study. Although imputation for 9K genotypes is expected to be slightly more accurate than for 6K genotypes (0.3 percentage points for all scenarios; Figure 1), the 6K and 9K genotyping chips are no longer available commercially.

CONCLUSIONS

Imputation efficiency in Mexican Holstein cattle based on M1, M2, and conflicts was affected by the size of the reference population used for imputation, the SNP density of the chip used for the original genotype, and the availability of ancestor genotypes. The inclusion of information from direct ancestors with US GDB genotypes in the imputation of Mexican Holstein genotypes increased imputation accuracy by 1 percentage point for low-density (6K and 9K) genotypes and 0.5 percentage points for high-density (77K) genotypes, which was about half the gain found with information from all US GDB Holsteins. Numbers of MM markers and imputation conflicts decreased when all US GDB genotypes were included. A larger reference population and the availability of genotyped ancestors improved imputation; animals with genotyped parents in a large reference population had higher imputation accuracy than those with no or few genotyped relatives in a small reference population. For small local populations, including genotypes from other related populations can be an important tool for improving accuracy of imputation.

ACKNOWLEDGMENTS

This study was supported by CONACYT, CONARGEN, and research project 1056821832 (Incorporation of Genomic information in the Genetic Evaluation
García-Ruiz A.

Process of Mexican Dairy Cattle). We thank Asociación Holstein de México (Querétaro, México) and the Council on Dairy Cattle Breeding (Reynoldsburg, OH) for providing data. The authors thank S. M. Hubbard, (Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD) for technical manuscript review.

REFERENCES

- Berry, D. P., and J. F. Kearney. 2011. Imputation of genotypes from low- to high-density genotyping platforms and implications for genomic selection. *Animal* 5:1162–1169. <http://doi.org/10.1017/S1751731111000309>.
- Bouwman, A. C., J. M. Hickey, M. P. L. Calus, and R. F. Veerkamp. 2014. Imputation of non-genotyped individuals based on genotyped relatives: Assessing the imputation accuracy of a real case scenario in dairy cattle. *Genet. Sel. Evol.* 46:6. <http://doi.org/10.1186/1297-9686-46-6>.
- Browning, B. L., and S. R. Browning. 2009. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* 84:210–223.
- Browning, B. L., and S. R. Browning. 2011. A fast, powerful method for detecting identity by descent. *Am. J. Hum. Genet.* 88:173–182. <http://doi.org/10.1016/j.ajhg.2011.01.010>.

- Carvalho, R., S. A. Boison, H. H. R. Neves, M. Sargolzaei, F. S. Schenkel, Y. T. Utsunomiya, A. M. Pérez O'Brien, J. Sölkner, J. C. McEwan, C. P. Van Tassell, T. S. Sonstegard, and J. F. Garcia. 2014. Accuracy of genotype imputation in Nelore cattle. *Genet. Sel. Evol.* 46:69. <http://doi.org/10.1186/s12711-014-0069-1>.
- Daetwyler, H. D., R. Pong-Wong, B. Villanueva, and J. A. Woolliams. 2010. The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 185:1021–1031. <http://doi.org/10.1534/genetics.110.116855>.
- Druet, T., C. Schrooten, and A. P. W. de Roos. 2010. Imputation of genotypes from different single nucleotide polymorphism panels in dairy cattle. *J. Dairy Sci.* 93:5443–5454. <http://doi.org/10.3168/jds.2010-3255>.
- Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.* 92:433–443. <http://doi.org/10.3168/jds.2008-1646>.
- Hickey, J. M., B. P. Kinghorn, B. Tier, J. H. J. van der Werf, and M. A. Cleveland. 2012. A phasing and imputation method for pedigreed populations that results in a single-stage genomic evaluation. *Genet. Sel. Evol.* 44:9. <http://doi.org/10.1186/1297-9686-44-9>.
- Howe, B. N., P. Donnelly, and J. Marchini. 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 5:e1000529.

Illumina Inc. 2011. BovineSNP50 Genotyping BeadChip. Accessed Dec. 30, 2014.

http://res.illumina.com/documents/products/datasheets/datasheet_bovine_sn_p50.pdf.

Illumina Inc. 2013. BovineLD v1.1 Genotyping BeadChip. Accessed Dec. 30, 2014.

http://res.illumina.com/documents/products/datasheets/datasheet_bovineld.pdf.

Johnston, J., G. Kistemaker, and P. G. Sullivan. 2011. Comparison of different imputation methods. *Interbull Bull.* 44:25–33.

Khatkar, M. S., G. Moser, B. J. Hayes, and H. W. Raadsma. 2012. Strategies and utility of imputed SNP genotypes for genomic analysis in dairy cattle. *BMC Genomics* 13:538. <http://doi.org/10.1186/1471-2164-13-538>.

Ma, P., R. F. Brøndum, Q. Zhang, M. S. Lund, and G. Su. 2013. Comparison of different methods for imputing genome-wide marker genotypes in Swedish and Finnish Red Cattle. *J. Dairy Sci.* 96:4666–4677. <http://doi.org/10.3168/jds.2012-6316>.

Neogen Corporation. 2013a. GeneSeek Genomic Profiler for dairy cattle. Accessed Dec. 30, 2014. http://www.neogen.com/Agrigenomics/pdf/Slicks/GGP-LD_Dairy.pdf.

- Neogen Corporation. 2013b. GeneSeek Genomic Profiler HD for dairy cattle. Accessed Dec. 30, 2014. http://www.neogen.com/Agrigenomics/pdf/Slicks/GGP_HD_Dairy.pdf.
- Sargolzaei, M., J. P. Chesnais, and F. S. Schenkel. 2011. FImpute—An efficient imputation algorithm for dairy cattle populations. *J. Dairy Sci.* 94(E-Suppl. 1):421. (Abstr.)
- Solberg, T. R., A. K. Sonesson, J. A. Woolliams, and T. H. E. Meuwissen. 2008. Genomic selection using different marker types and densities. *J. Anim. Sci.* 86:2447–2454. <http://doi.org/10.2527/jas.2007-0010>.
- VanRaden, P. 2014. findhap.f90, Find haplotypes and impute genotypes using multiple chip sets. Accessed Dec. 30, 2014. <http://aipl.arsusda.gov/software/findhap/>.
- VanRaden, P. M., J. R. O’Connell, G. R. Wiggans, and K. A. Weigel. 2010. Combining different marker densities in genomic evaluation. *Interbull Bull.* 42:113–117.
- VanRaden, P. M., J. R. O’Connell, G. R. Wiggans, and K. A. Weigel. 2011. Genomic evaluations with many more genotypes. *Genet. Sel. Evol.* 43:10. <http://doi.org/10.1186/1297-9686-43-10>.
- Weigel, K. A, G. de los Campos, A. I. Vazquez, G. J. M. Rosa, D. Gianola, and C. P. Van Tassell. 2010. Accuracy of direct genomic values derived from imputed

single nucleotide polymorphism genotypes in Jersey cattle. *J. Dairy Sci.*

93:5423–5435. <http://doi.org/10.3168/jds.2010-3149>.

Wiggans, G. R., T. A. Cooper, P. M. VanRaden, K. M. Olson, and M. E. Tooker. 2012.

Use of the Illumina Bovine3K BeadChip in dairy genomic evaluation. *J. Dairy Sci.*

95:1552–1558. <http://doi.org/10.3168/jds.2011-4985>.

Wiggans, G. R., P. M. VanRaden, and T. A. Cooper. 2011. The genomic evaluation

system in the United States: Past, present, future. *J. Dairy Sci.* 94:3202–3211.

<http://doi.org/10.3168/jds.2010-3866>.

Zhang, Z., and T. Druet. 2010. Marker imputation with low-density marker panels in

Dutch Holstein cattle. *J. Dairy Sci.* 93:5487–5494.

<http://doi.org/10.3168/jds.2010-3501>.

Table 1. Percentages of 45,195 (50K) genotype alleles that were missing¹ and numbers of animals with missing alleles by genotyping chip density and reference population scenario

Genotyping chip density ²	Scenario ³	Animals (no.)	Missing 50K alleles (%)		Animals with missing alleles at ≥ 1 SNP genotype (no.)	
			M1	M2	M1	M2
6K	1	183	4.75	0.92	183	183
	2	183	3.16	0.05	183	180
	3	183	0.38	0.01	183	26
9K	1	277	6.03	0.94	277	277
	2	277	3.58	0.04	277	217
	3	277	0.22	0.02	277	25
50K	1	726	0.17	0.86	726	726
	2	1,598	0.21	0.03	896	802
	3	1,598	0.01	0.01	690	196
77K	1	825	4.36	1.13	820	820
	2	839	2.23	0.07	835	739
	3	839	0.13	0.02	815	153

¹M1 = 1 allele missing, and M2 = both alleles missing.

²6K = 6,842 markers, 9K = 8,196 markers, 50K = 45,195 markers, and 77K = 28,048 markers.

³Scenario 1 = 2,011 Mexican genotyped Holsteins, scenario 2 = 2,011 Mexican genotyped Holsteins plus 886 related Holsteins in the US genotype database, and scenario 3 = 2,011 Mexican genotyped Holsteins and 338,073 Holsteins in the US genotype database.

Table 2. Numbers of imputed alleles for scenarios 1 and 2 that conflict with those imputed for scenario 3¹ and conflict frequencies by genotyping chip density

Genotyping chip density ²	Conflicts (no.)		Conflict frequency ³ (%)	
	Scenario 1	Scenario 2	Scenario 1	Scenario 2
6K	1,718	325	3.80	0.72
9K	1,705	506	3.77	1.12
50K	28	16	0.06	0.04
77K	746	689	1.65	1.52

¹Scenario 1 = 2,011 Mexican genotyped Holsteins, scenario 2 = 2,011 Mexican genotyped Holsteins plus 886 related Holsteins in the US genotype database, and scenario 3 = 2,011 Mexican genotyped Holsteins and 338,073 Holsteins in the US genotype database.

²6K = 6,842 markers, 9K = 8,196 markers, 50K = 45,195 markers, and 77K = 28,048 markers.

³Conflict frequency = 100(number of conflicts/45,195).

Table 3. Differences in the numbers and percentages of missing alleles¹ and imputed allele conflicts for scenario 2 compared with scenario 3² by availability of genotyped ancestors and genotyping chip density

	Genotyping		Imputation difference							
	g chip density	Animals (no.)	M1		M2		Conflicts		All	
			No.	%	No.	%	No.	%	No.	%
Neither dam nor sire	6K	18	2,897	6.4	40	0.1	4,827	10.7	7,764	17.2
	50K	12	159	0.4	20	<0.1	168	0.4	347	0.8
	77K	33	1,627	3.6	62	0.1	1,837	4.1	3,526	7.8
Dam or sire	9K	23	2,896	6.4	20	<0.1	3,762	8.3	6,678	14.8
	50K	299	124	0.3	10	<0.1	136	0.3	270	0.6
	77K	156	1,444	3.2	34	0.1	1,560	3.5	3,038	6.8
Dam, sire, and a grandsire	9K	20	584	1.3	6	<0.1	898	2.0	1,488	3.3
	50K	36	85	0.2	8	<0.1	99	0.2	192	0.4
	77K	47	583	1.3	11	<0.1	705	1.6	1,299	2.9

¹M1 = 1 allele missing, and M2 = both alleles missing.

²Scenario 2 = 2,011 Mexican genotyped Holsteins plus 886 related Holsteins in the US genotype database, and scenario 3 = 2,011 Mexican genotyped Holsteins and 338,073 Holsteins in the US genotype database.

³6K = 6,842 markers, 9K = 8,196 markers, 50K = 45,195 markers, and 77K = 28,048 markers

Table 4. Numbers and percentages of missing markers (MM) and conflicts after SNP validation by genotyping chip density and reference population scenario

Genotyping chip density ¹	Scenario ²	MM		Conflicts	
		No.	%	No.	%
6K	1	2,965	6.6	1,277	2.8
	2	2,398	5.3	1,022	2.3
9K	1	3,252	7.2	1,181	2.6
	2	2,641	5.8	936	2.1

¹Scenario 1 = 2,011 Mexican genotyped Holsteins and scenario 2 = 2,011 Mexican genotyped Holsteins plus 886 related Holsteins in the US genotype database.

²6K = 6,842 markers and 9K = 8,196 markers.

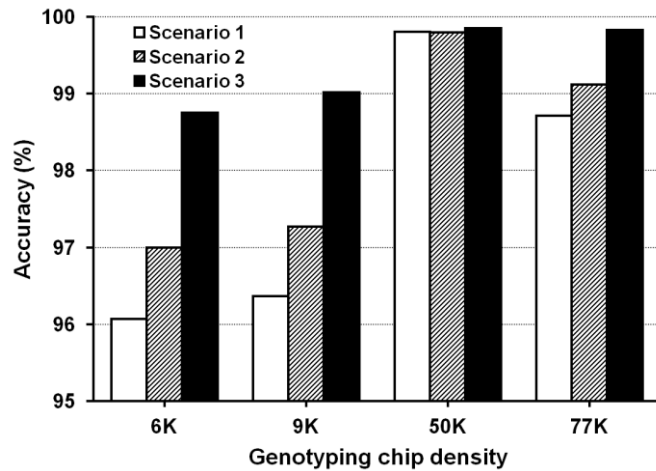


Figure 1. Accuracy of imputation to 45,195 (50K) markers for genotypes based on 4 genotyping chip densities (6K = 6,842 markers, 9K = 8,196 markers, 50K = 45,195 markers, and 77K = 28,048 markers) and 3 reference population scenarios (1 = 2,011 Mexican genotyped Holsteins, 2 = 2,011 Mexican genotyped Holsteins plus 886 related Holsteins in the US genotype database, and 3 = 2,011 Mexican genotyped Holsteins and 338,073 Holsteins in the US genotype database); accuracy was measured as the squared correlation between estimated and true genotypes and expressed as a percentage.

CAPÍTULO 7

Regiones del genoma asociadas a características de producción láctea y conteo celular somático en la población Holstein de México

Éste capítulo fue presentado como resumen en el mes de Agosto de 2014 en el 10th World Congress on Genetics Applied to Livestock Production, Vancouver Canadá. El documento se encuentra en el formato en el que fue publicado en las memorias del congreso.

Genome Regions Associated to Milk Production Traits and Somatic Cell Score in the Mexican Holstein Population

A. García-Ruiz*, F.J. Ruiz*†, C.P. Van Tassell‡, Hugo H. Montaldo*

*National Autonomous University of Mexico, México, D.F., México, †National Institute of Forestry, Agricultural and Livestock Research, Ajuchitlán, Querétaro, México, ‡Bovine Functional Genomics Laboratory, Agricultural Research Service, USDA, Beltsville, Maryland, USA

ABSTRACT: The aim of this study was to identify the genomic markers associated to milk production traits and somatic cell scores of Mexican Holstein cattle. A total of 1,975, 1,602, and 1,595 animals were included in a genome wide association studies (GWAS) for milk yield, milk components, and somatic cell score, respectively. The genotypes included 45,195 SNP. The analyses were performed with a mixed model and regression analysis implemented in GenABEL (R-software). For all traits, a total of 251 markers had statistically significant association ($p < 0.001$) and 69.72% of the SNP were located in genomic regions earlier reported to traits studied in other populations. BTA13 contains 40 SNP with statistically significant association that have not been reported in other studies. These GWAS results confirm the evidence on the well-known genes and identify a suite of new associations in the Mexican Holstein population.

Keywords:

GWAS

Somatic cell scores

Milk production

Introduction

A better understanding of genetic improvement of livestock production, has led to the necessity of detecting genomic regions that explain the genetic variation of economical important traits in livestock animals (Pryce et al. (2010)). Conventional quantitative trait locus (QTL) mapping has been used to detect QTL related to important traits; nevertheless this tool was limited because of the low precision of the QTL position (Mai et al. (2010)). The availability of the genome wide scan through the single nucleotide polymorphism (SNP) markers and the reducing cost of genome sequencing had allowed the detection of numerous genetic regions associated to important traits and the identification of the best animals according to their genomic information (Daetwyler et al. (2010)). The genome wide association study (GWAS) is a recent technique for the identification of causal genes and surveys common genetic variation of traits of interest (Zhang et al. (2012); Wang et al. (2005)). These studies have triggered the opportunity of performing QTL mapping more accurate (Calus (2009); Wang et al. (2005)), including causal genes with modest effects (Hirschhorn et al. (2005)).

In dairy cattle, the majority of GWAS used a bull reference population (Bolormaa et al. (2010); Mai et al. (2010); Pryce et al. (2010)) due to the high reliability of their phenotypes, compared with that of cows. But in some populations the availability of biological samples of bulls

for extracting DNA is limited and the inclusion of female's phenotypes is necessary. Previous studies show a greater number of detected regions in the genome when males and females are included in the reference population (Calus et al. (2013)). For farm animals, including dairy cattle, numerous genomic regions are reported frequently explaining variance of economically important traits. GWAS studies have included traits for different breeds within countries, but many inconsistencies have been reported. Differences among the same trait and breed in different countries could be attributed to reference population size, panel densities or population structure (Zhang et al. (2012)). The aim of this study was to identify genomic markers associated to milk production traits and somatic cell scores in the Mexican Holstein population and compare them with those reported in other populations.

Materials and Methods

Phenotypes. Breeding values of Holstein males and females estimated by INIFAP in Mexico (Ruiz et al. (2013)) were deregressed according to the methodology described by Garrick et al. (2009) and were used as phenotypes in GWAS.

For milk production, bulls, sires of bulls, dams of bulls, cows, sires of cows and dams of cows with low reliability (<60 , <60 , <50 , <50 , <60 , <50 , respectively) were excluded from the analysis. Reliabilities of milk components and somatic cell scores were slightly lower than milk production. For fat and protein production (Kilograms and percentage), the limits for reliability was decreased 5 units and 10 units for somatic cell score compared to milk production limits.

A total of 1,975 animals (453 sires and 1,522 cows) were included in the GWAS for milk production, 1,602 (387 sires and 1,215 cows) for milk components and 1,595 (411 sires and 1,184 cows) for somatic cell scores. Table 1 contains general statistics of phenotypes for all traits.

Genotypes. Genomic information of 2,018 Mexican animals and 886 ancestors from North America were used for imputing a total of 45,195 single nucleotide polymorphism (SNP), using the Find Hap software Version 2 (VanRaden et al. (2011)). After the imputation process and deletion of low accuracy phenotypes, the quality control analysis was performed excluding markers with minor allele frequency less than 2.5%, call rate less than 90%, or failure of Hardy Weinberg equilibrium ($P < 0.15$). Individuals with a call rate across loci less than 90% were also excluded. A total of 34,856; 34,776 and 34,505 SNP

markers were used in the GWAS for milk production, milk component traits and somatic cell score, respectively.

Model. GWAS were performed using a mixed model and regression analysis (GRAMMAR). This method, obtains the residuals adjusted for family effects first (constructing the relationship matrix with SNP information) and subsequently analyzes the association between the residuals and genetic polymorphisms using rapid least-squares methods (Aulchenko et al. (2007)). To account for relatedness, the covariance matrix was estimated from the genomic kinship matrix. This methodology was implemented in the GenABEL package of R (Aulchenko et al. (2007b)). The same package was used for the quality control analysis.

Results and Discussion

To reference the location of SNP with estimated effects statistically different than zero, these were searched by base pair position in the cattle QTL data base (<http://www.animalgenome.org/QTLdb/cattle.html>, January, 2013) and compared with our results. Previous studies reported QTL, genes or significant SNP in the same positions that our findings and others had not been identified. SNP markers with a $p < 0.001$ were assumed as significant. Genomic regions which included more than 3 significant SNP closely located and were not previous referenced in Holstein cattle will be discussed.

A total of 251 SNP markers had statistically significant association in this study. Twenty eight for milk yield, 30 and 83 for milk fat yield and percentage, respectively, 35 and 44 for milk protein yield and percentage and 31 for somatic cell score. From all the statistically significant association SNP, 70% were previously reported in the same phenotype in this study. A total of 40 SNP with significant association were newly identified on chromosome (BTA) 13 in this study.

Milk yield. Four SNP on BTA2 had statistically significant association (BTB-00088621, BTB-00969752, ARS-BFGL-BAC-30703 and ARS-BFGL-NGS-25349) were related to milk production in our study, and those markers were not previously associated with this trait in Holsteins, but QTL related to milk yield in Ayrshire cattle (Viitala et al. (2003)) were reported in the same genomic region.

Milk fat yield. Four SNP on BTA15 (ARS-BFGL-NGS-117777, ARS-BFGL-NGS-32123, Hapmap52074-ss46527004, BTB-00605195) had statistically significant association for milk fat yield in this study that had not previously found associated with fat yield in dairy cattle.

Milk protein yield. This trait presented the largest number of genomic regions across 12 BTA with low to moderate association level. Nine regions were previously identified. Novel SNP associations were identified in protein yield for Mexican Holsteins. Those associations were located in BTA4 (ARS-BFGL-NGS-102687, ARS-

BFGL-NGS-113152), BTA13 (Hapmap42348-BTA-88290, Hapmap51209-BTA-32563, BTA-117294-no-rs, ARS-BFGL-NGS-31560, ARS-BFGL-NGS-38064, ARS-BFGL-NGS-21302, ARS-BFGL-NGS-32055) and BTA15 (Hapmap52074-ss46527004, ARS-BFGL-NGS-31679, ARS-BFGL-NGS-25439). In the same region of BTA13, were detected QTL (Plante et al. (2001)) affecting fat yield, but no associations were identified with milk traits on BTA15.

Milk Fat percentage. The GWAS results for this trait showed fewer genomic regions associated with fat percentage than with milk production, but with moderate to large effect. SNP on BTA9 (ARS-BFGL-NGS-75693 and Hapmap45687-BTA-83338) had not been associated with fat percentage previously, although they had been linked with fat yield (Wiener et al. (2000)), and some SNP on BTA14 (Hapmap30381-BTC-005750, Hapmap30383-BTC-005848, BTA-34956-no-rs and ARS-BFGL-NGS-57820) have not previously been associated with fat percentage, although they are in the neighborhood of DGAT1 (Kaupe et al. (2007)). This result might indicate a wider DGAT1 region or linkage disequilibrium between these markers and that mutation.

Milk protein percentage. SNP with estimated effects that were statistically different than zero for protein percentage in this study were located in 9 different BTA and 5 of these BTA were earlier reported in other Holstein populations. Nineteen SNP on BTA13 (ARS-BFGL-NGS-23363, ARS-BFGL-NGS-65199, ARS-BFGL-NGS-115482, ARS-BFGL-NGS-40188, ARS-BFGL-NGS-13061, UA-IFASA-4272, ARS-BFGL-NGS-41237, ARS-BFGL-NGS-80072, ARS-BFGL-BAC-15734, BTB-01414766, BTA-92697-no-rs, ARS-BFGL-NGS-14463, ARS-BFGL-NGS-75174, ARS-BFGL-NGS-25461, ARS-BFGL-NGS-31462, ARS-BFGL-NGS-116624, BTB-00529466, ARS-BFGL-NGS-104967 and ARS-BFGL-NGS-14592) with large impact on protein percentage in this study had not been reported previously. SNP affecting milk fat yield with estimated effects that were statistically different than zero in this study have been reported within the genomic region with statistically significant association reported in other studies (Plante et al. (2001)).

Somatic cell score. SNP markers had a low to moderate effect for this trait. Genome regions without previous references were located in BTA13 (Hapmap52431-ss46526792, BTB-01721718, ARS-BFGL-BAC-11276, ARS-BFGL-NGS-39930 and ARS-BFGL-NGS-103355), and BTA26 (ARS-BFGL-NGS-12828, ARS-USMARC-Parent-EF034086-no-rs and ARS-BFGL-NGS-114564). While QTL associated with different traits have been reported in the BTA26 region reported here, no previous references were found for the SNP in the region where the statistically significant association markers of BTA13 are located. For example, studies identified milk yield QTL (Bennowitz et al. (2003)), milk fat yield QTL (Plante et al. (2001)) and milk fat percentage QTL in Ayrshire cattle (Viitala et al. (2003)).

Conclusion

The majority of the SNP with statistically significant association of this study were located in genomic regions earlier related to the studied traits. Others were near reported QTL or genes or were located in regions previously associated with other production traits. Two groups of SNP with estimated effects that were statistically different than zero were located on BTA13 and had moderate to large effects on milk protein yield and percentage. The region includes 7 and 19 SNP for each trait, respectively. These GWAS results identify a suite of novel associations that are not reported previously as well as confirm the evidence on the well-known genes.

Literature Cited

- Aulchenko, Y. S., Koning, D. De, and Haley, C. (2007). *Genetics*, 585: 577–585.
- Aulchenko, Y. S., Ripke, S., Isaacs, A., and Duijn, C. M. Van. (2007b). *Bioinformatics* 23: 1294–1296.
- Bennewitz, J., Reinsch, N., Grohs, C. et al. (2003). *Gen. Sel. Evol.*, 35, 319–338.
- Bolormaa S, Pryce J. E., Hayes B.J. et al., (2010). *J. Dairy Sci.*, 93:3818–3833.
- Calus, M. P. L. (2009). *Animal: Int J. Anim. Biosci.* 4:157–164.
- Calus, M. P. L., de Haas, Y., and Veerkamp, R. F. (2013). *J. Dairy Sci.* 96: 6703–15.
- Daetwyler, H.D., R. Pong-Wong, B. Villanueva, et al. (2010). *Genetics* 185:1021–1031.
- Garrick, D. J., Taylor, J. F. and Fernando, R. L. (2009). *Gen. Sel. Evol.* 41:55.
- Hirschhorn, J. N. and Daly M.J. (2006). *Nat. Rev.* 6: 95–108.
- Kaube, B., Brandt, H., Prinzenberg, E. M. et al. (2007). *J. Anim. Sci.*, 85(1), 11–21.
- Mai, M. D., Sahana, G., Christiansen, F. B., et al. (2010). *J. Anim. Sci.*, 88: 3522–3528.
- Plante, Y., Gibson, J. P., Nadesalingam, J. et al. (2001). *J. Dairy Sci.* 84(6):1516–24.
- Pryce, J. E., Bolormaa, S., Chamberlain, J. et al. (2010). *J. Dairy Sci.*, 93(7), 3331–45.
- Ruiz, F. L., Toledo, H., Duran, A.M., (2013). ¿Qué toro?, INIFAP, 44.
- VanRaden, P. M., O'Connell, J. R., Wiggans, G. R. et al. (2011). *Gen. Sel. Evol.* 43: 10.
- Viitala, S. M., Schulman, N. F., de Koning, D. J. et al. (2003). *J. Dairy Sci.*, 86(5):1828–36.
- Wang, W. Y., Barratt, B. J., Clayton, D. G. et al. (2005). *Nat. Rev. Genetics* 6: 109–18.
- Zhang, H., Wang, Z., Wang, S. et al. (2012). *J. Anim. Sci. and Biotechnology* 3:26

Table 1. Mean and standard deviation of phenotypes for Milk production, milk components and somatic cell score included in the analysis.

Trait Breeding Values	Mean	Standard deviation
Milk production (kg)	1178.21	886.88
Fat production (kg)	20.40	27.69
Protein Production (kg)	23.28	24.47
Fat test (%)	-0.107	0.191
Protein test (%)	-0.041	0.100
Somatic cell score	-0.189	0.388

CAPÍTULO 8

DISCUSIÓN GENERAL

Existen numerosos beneficios de conocer la diversidad genética, tanto en plantas (Zhang et al., 2011) como en animales (Groeneveld et al., 2010), ya que a través de ella es posible establecer programas de conservación de especies en peligro de extinción (Glowatzki-Mullis et al., 2008; Tejas et al., 2012), o en el caso de especies domésticas, planear programas de producción en diversos entornos, establecer estrategias en los programas de mejoramiento genético y programas de selección de animales de acuerdo a las condiciones ambientales y composición genética de las poblaciones y facilitar la adaptación rápida de los animales cuando se requieran cambios en los objetivos de los programas de mejoramiento (Notter, 1999; Edea et al., 2015).

El ganado Holstein es una de las razas lecheras más usadas a nivel mundial (Farras, 1977) y las estrategias usadas en los programas de mejoramiento genético han sido muy eficaces en incrementar rápidamente las tasas de mejoramiento genético y lograr los objetivos planteados por la industria (Zenger et al., 2007). Estas ganancias han sido facilitadas en gran medida por la intensidad de selección de sementales y el uso extendido de técnicas reproductivas, como la inseminación artificial y trasplante de embriones (Flori et al., 2009). Sin embargo, esta práctica ha limitado potencialmente el nivel de diversidad genética (Zenger et al., 2007), sobre todo el ganado productor de leche, que es una industria altamente especializada.

Con los avances en las tecnologías moleculares, particularmente con disponibilidad de marcadores de SNP, ha sido posible realizar estudios de diversidad genética en las poblaciones y en regiones específicas del genoma, debido

a que éstos son abundantes y estables genéticamente (Vignal et al., 2002; Engelsma et al., 2010).

Se han realizado numerosos estudios sobre diversidad genética en ganado Holstein en diferentes poblaciones usando marcadores de SNP (Engelsma et al., 2010; Zhang et al., 2011; Xu et al., 2015). En México éste proyecto es el primero que realiza estudios conducentes a determinar la diversidad genética de poblaciones lecheras, por lo que servirá como referencia futuros estudios genéticos de ganado Holstein

Es bien conocido que la producción de leche en México ha dependido comercialmente de EUA y Canadá, y en particular México ha importado material genético por varias décadas en forma de semen, embriones y vaquillas de estos países. Por esta razón, se hubiera podido considerar que los animales de las poblaciones de EUA, Canadá y México eran genéticamente similares, pero los resultados obtenidos en este proyecto, han permitido detectar diferencias genéticas en el ganado Holstein de dos sistemas de producción en México (CONV y LOWI) y entre éstos y las poblaciones de EUA y CAN (Capítulo 4).

La principal diferencia encontrada entre los dos sistemas de producción en México (CONV y LOWI) fue la detección de material genético de otras razas, especialmente en el sistema LOWI, lo que podría indicar que los propietarios de los animales del sistema LOWI detectaron en algún momento la necesidad de cruzar a los animales Holstein con otras razas de acuerdo a sus objetivos de producción. Para determinar el impacto de usar cruza en el sistema LOWI, sería necesario desarrollar investigaciones adicionales para comprobar si los cruzamientos favorecen la adaptación o producción para las características de interés bajo las condiciones de ese sistema de producción.

Actualmente no se ha documentado el flujo comercial de material genético en México (semen, vaquillas, etc.), pero con los resultados obtenidos en el capítulo 4, se dio una posible explicación sobre la distribución y uso de material genético entre los dos sistemas de producción en México y las poblaciones de USA y CAN, ya que se encontró que el sistema CONV se encuentra entre el sistema LOWI y la población CAN + USA, probablemente porque los criadores del sistema CONV proporcionan material genético (semen y novillas) al sistema LOWI y los animales del sistema CONV se encuentran más relacionados genéticamente a las poblaciones CAN y USA.

Posteriormente fue de interés detectar las regiones del genoma mantenidas a través de varias generaciones (huellas de selección) y corroborar que son producto de los objetivos de los programas de mejoramiento genético (Capítulo 5). En este estudio se observó que las huellas de selección detectadas, estuvieron relacionadas con características que no han sido incorporadas en los programas de mejoramiento genético en México (reproductivas) lo que sugirió que por causa del flujo comercial del material genético y de la selección para el mejoramiento de características reproductivas en las poblaciones de USA y CAN, los grupos mexicanos han modificado estas características sin haberlo considerado en sus programas de selección.

La necesidad de realizar imputación de genotipos de animales de la población Holstein de México (Capítulo 6), permitió observar la importancia que tiene el tamaño de la población de referencia en el proceso de imputación de animales genotipados con los chips de baja densidad, ya que a mayor tamaño de la población de referencia, se obtiene mayor confiabilidad en el proceso de imputación. Ésta tendencia había sido previamente reportada (Druet et al., 2010), pero no se sabía si éste patrón se mantendría en poblaciones que no mantienen el

mismo nivel de desequilibrio de ligamiento, como fue el caso de las poblaciones incorporadas en el presente estudio.

La asociación de marcadores de SNP con características de interés económico tiene como objetivo identificar loci de características cuantitativas (McCarthy et al., 2008). Sin embargo, se han encontrado diferencias en las asociaciones entre SNP y características reportadas en diferentes poblaciones, aunque se hable de la misma característica y raza de estudio. Estas diferencias podrían explicarse por el grado de diversidad genética que existe entre las poblaciones (Zhang et al., 2012). En la población Holstein de México, se encontraron asociaciones de SNP que no habían sido previamente reportadas en otras poblaciones, resaltando los asociados con producción de proteína (Capítulo 7) en el cromosoma 13. Las nuevas asociaciones encontradas podrían ser objeto de futuros estudios genéticos y funcionales para explicar sus efectos e implicaciones con la expresión de genotipos. Las diferencias genéticas entre poblaciones pueden tener diferentes orígenes, que podrían venir desde el proceso de domesticación hasta la formación de las razas y los programas recientes de mejoramiento genético (Xu et al., 2015).

Los conocimientos obtenidos a través del presente estudio tienen aplicaciones prácticas en futuros estudios genómicos y en los programas de mejoramiento genético. Por ejemplo, si se requiere realizar análisis de asociación incluyendo información de varias poblaciones (CONV, USA o CAN), será necesario estratificar las poblaciones, para corregir las variaciones ocasionadas por la estructura poblacional.

Para cuestiones prácticas en los programas de mejoramiento genómico, se deberán desarrollar y usar ecuaciones de predicción genómica para cada población

de manera de que los valores genéticos que se obtengan sean un indicador confiable para la selección de los animales.

REFERENCIAS

- Zhang Y, Wang XF, Li ZK, Zhang GY y Ma ZY. Assessing genetic diversity of cotton cultivars using genomic and newly developed expressed sequence tag-derived microsatellite markers. *Genet Mol Res*, 2011. 10(3):1462-1470.
- Groeneveld LF, Lenstra JA, Eding H, Toro MA, Scherf B, Pilling D, Negrini R, Finlay EK, Jianlin H, Groeneveld E, Weigend S y The GLOBALDIV Consortium. Genetic diversity in farm animals; a review. *Anim Genet*, 2010. 41 (1): 6-31. doi:10.1111/j.1365-2052.2010.02038.x
- Glowatzki-Mullis ML, Muntwyler J, Bäümle E y Gaillard C. Genetic diversity measures of Swiss goat breeds as decision-making support for conservation policy. *Small Ruminant Res*, 2008. 74, 202-11.
- Tejas MS, Patel J, Chandrakant DB, Aakash D, Uday D, Shivnandan SP, Dharamshibhai NR, Jitendra VS y Chaitanya GJ. Evaluation of genetic diversity and population structure of west-central Indian cattle breeds, 2012. 44(4): 442-445. DOI: 10.1111/age.12013
- Notter DR. The importance of genetic diversity in livestock populations of the future. *J Anim Sci*, 1999. 77(1):61-9
- Edea Z, Bhuiyan MS, Dessie T, Rothschild MF, Dadi H, Kim KS. Genome-wide genetic diversity, population structure and admixture analysis in African and Asian cattle breeds. *Animal*, 2015. 9(2):218-26. doi: 10.1017/S1751731114002560
- Farras J. Cría lucrativa de la vaca lechera. 6ª ed. Editorial Sintes, S.A. Barcelona, España 1977. 67-90
- Zenger KR, Khatkar MS, Cavanagh JA, Hawken RJ, Raadsma HW. Genome-wide genetic diversity of Holstein Friesian cattle reveals new insights into

- Australian and global population variability, including impact of selection. *Anim Genet*, 2007. 38(1):7-14.
- Flori L, Fritz S, Jaffrézic F, Boussaha M, Gut I, Heath S, Foulley J y Gautier M. The genome response to artificial selection: a case study in dairy cattle. *PLoS One*, 2009. 4(8), e6595. doi:10.1371/journal.pone.0006595
- Vignal A, Milan D, Sancristobal M y Eggen A. A review on SNP and other types of molecular markers and their use in animal genetics. *Genet Sel Evol*, 2002. 34:275-305.
- Engelsma KA, Veerkamp RF, Calus MPL y Windig JJ. Differences in genetic diversity in Holstein cattle with high and low genetic merit. 9th World Congress on Genetics Applied to Livestock Production, 2010.
- Xu L, Bickhart DM, Cole JB, Schroeder SG, Song J, Van Tassell CP, Sonstegard T, Liu GE. Genomic Signatures Reveal New Evidences for Selection of Important Traits in Domestic Cattle. *Molecular Biology and Evolution*, 2015. 1-15. doi:10.1093/molbev/msu333
- Druet T, Schrooten C, y de Roos APW. Imputation of genotypes from different single nucleotide polymorphism panels in dairy cattle. *J. Dairy Sci*, 2010. 93:5443-5454. <http://doi.org/10.3168/jds.2010-3255>.
- McCarthy, Mark I, Hirschhorn, Joel N. Genome-wide association studies: potential next steps on a genetic journey. *Human Molecular Genetics*, 2008. 17(2):156,165.

CAPÍTULO 9

CONCLUSIONES

Con los estudios realizados en el presente proyecto, se llegó a las siguientes conclusiones:

- Las poblaciones en los sistemas de producción especializado (CONV) y familiar (LOWI) de ganado Holstein en México y las poblaciones de Estados Unidos y Canadá tienen diferentes estructuras genéticas, aunque comparten ancestros comunes y regiones mantenidas del genoma asociadas a características de interés económico.
- La principal diferencia entre los sistemas de producción CONV y LOWI en México es debida a que en éste último se realizaron cruzamientos de ganado Holstein con otras razas, por ejemplo: Jersey y Pardo Suizo.
- Para realizar futuros estudios de asociación del genoma y predicciones genómicas entre las poblaciones, será necesario considerar el nivel de desequilibrio de ligamiento y usar por lo menos 54,000 marcadores de SNP.
- Aunque se observaron diferencias en la estructura genética de las poblaciones Holstein de México y de Estados Unidos y Canadá, la información genómica disponible de éstas dos últimas mejora el proceso de imputación de genotipos en el ganado Holstein de México.
- El tamaño de la población de referencia, el número de genotipos disponibles y su densidad (número de SNP) fueron factores que determinaron la eficiencia del proceso de imputación de la población Holstein de México.

- En los estudios de asociación del genoma de la población Holstein de México, se encontró una región del cromosoma 13 asociada a producción de proteína láctea que no había sido previamente reportada en otros estudios de ganado lechero a nivel mundial.
- Debido a las diferencias genéticas encontradas entre las poblaciones incluidas en este estudio, será necesario desarrollar evaluaciones genómicas dentro de cada población o realizar estratificaciones en caso de participar en evaluaciones genómicas conjuntas.