

**UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA
INSTITUTO DE INGENIERÍA**

MAESTRÍA Y DOCTORADO EN CIENCIAS E INGENIERÍA



**“Estudio de asociación de SNPs con características de conformación en ganado
Holstein de Baja California”**

TESIS PARA OBTENER EL GRADO DE:

Doctorado en Ciencias

PRESENTA

Mirvana Elizabeth González Macías

DIRECTOR

Dr. Rafael Villa Angulo

CODIRECTOR

Dr. Víctor Manuel González Vizcarra

Mexicali, B. C.

Febrero/2017

ÍNDICE

<i>Capítulo I</i>	1
<i>Introducción</i>	1
1.1 Biología Molecular.....	1
1.2 La célula, el ADN y los cromosomas.....	2
1.3 Genotipo y fenotipo.....	3
1.4 Mejoramiento genético de ganado bovino	4
1.5 Métodos de selección de ganado bovino	7
1.5.1 Selección tradicional y basada en marcadores.....	7
1.5.2 Selección genómica	9
1.6 Tecnologías de genotipificación de salida masiva	11
1.6.1 SNPchips	13
1.7 Imputación de genotipos	16
1.8 Bioestadística	18
1.8.1 Prueba X^2 de Pearson (Chi Square).	19
1.8.1.1 Prueba de Chi-Cuadrado de bondad de ajuste	19
1.8.1.2 Pruebas de Chi-Cuadrado de asociación e independencia.....	20
1.8.2 Intervalos de Confidencia	20
1.8.3 Significancia Estadística	21
1.8.4 El Valor-p	22
1.8.5 Regresión de mínimos cuadrados	22
1.8.6 Corrección de Bonferroni	23
1.9 Recursos Bionfornáticos	24
1.9.1 Lenguaje de programación Perl	24
1.9.2 Software estadístico R	25
1.9.3 FindHap V4	25
1.9.4 PLINK	26
1.9.5 Bases de Datos Genéticos y Genómicos Públicas	26
1.10 Métodos para estudios de asociación	27
1.10.1 Asociación Directa	28
1.10.2 Asociación Indirecta.....	28
1.11 Control de Calidad de los datos de los SNPs	28
1.11.1 Errores de genotipificación	28
1.11.2 Frecuencia del Alelo Menor (MAF).....	29
1.11.3 Equilibrio de Hardy-Weinberg (HWE)	29
<i>Capítulo II</i>	31

<i>Planteamiento del Problema</i>	<i>31</i>
2.1 Situación actual de la industria lechera en la región de Baja California.	31
2.2 Objetivo General	34
2.2.1 Objetivos específicos.....	35
<i>Capítulo III</i>	<i>36</i>
<i>Desarrollo y Resultados</i>	<i>36</i>
3.1 Descripción de la población	36
3.2 Muestreo para la recolección de sangre y medición de fenotipos	37
3.3 Extracción de ADN genómico de las muestras	39
3.4 Genotipificación con el SNPchip BovineLD	40
3.5 Imputación.....	40
3.6 Aplicación de algoritmos bioinformáticos para realizar estudios de asociación de SNPs con los 16 rasgos de conformación	41
<i>Capítulo IV</i>	<i>47</i>
<i>Discusión</i>	<i>47</i>
<i>Capítulo V.....</i>	<i>50</i>
<i>Conclusiones.....</i>	<i>50</i>
<i>Referencias</i>	<i>52</i>

LISTA DE FIGURAS

Figura 1. Mejoramiento genético en ganado bovino	5
Figura 2. Efecto epigenético	6
Figura 3. Marcadores SNP.	10
Figura 4 Selección Genómica.....	11
Figura 5. Ejemplo de una TGSM	12
Figura 6. Diagrama esquemático de un SNPchip.	13
Figura 7. Imputación fusionando dos conjuntos de genotipos	16
Figura 8. Imputación con paneles de referencia de haplotipos.....	17
Figura 9. Producción de leche en los últimos años en el estado de Baja California	32
Figura 10. Obtención de muestras de sangre de la vena coccígea y de la vena yugular	37
Figura 11. Extracción de ADN genómico	39
Figura 12. Análisis de ADN en gel agarosa	40
Figura 13. Gráfica de asociación de $-\log_{10}$ (P-values) de asociaciones de nucleótidos simples con (a) Profundidad corporal (b) Profundidad de la Ubre.....	45

LISTA DE TABLAS

Tabla 1. Cada organismo contiene diferente número de células	3
Tabla 2. Resumen de los SNPchips comercializados y el tipo de análisis que puede ser realizado con ellos.	15
Tabla 3. Resumen de las estadísticas del fenotipo.....	38
Tabla 4. Lista de SNPs significativos, utilizando la herramienta PLINK 1.07	43
Tabla 5. Lista de los QTLs que están asociados con los SNPs significativos.....	46

Agradecimientos

Gracias a Dios por estar siempre a mi lado, por ser mi fortaleza en los momentos de debilidad, por permitirme llegar a este momento. El señor es mi pastor nada me falta.

A mi familia, por todo el apoyo brindado a lo largo de mi vida. En especial a ti ma por enseñarme a ser una guerrera, te amo hasta el infinito y más allá.

A mis hermanas, hermanos y sobrinos por ser parte importante de mi vida y representar la unión familiar. A cada uno de ellos por ser mi fuerza motriz y razón de vivir. Los amo.

A mi director de tesis y amigo el Dr. Rafael Villa Angulo, por su confianza depositada en mí, por sus valiosas enseñanzas, por transmitirme ese entusiasmo y satisfacción de lograr las cosas.

Gracias Dr. Carlos Villa Angulo por todo su apoyo incondicional, por su tiempo, amistad y por los conocimientos que me trasmitió.

A los doctores del laboratorio de Biotecnología de la Universidad Politécnica de Madrid, por haberme recibido con los brazos abierto y adoptado como una madrileña. Majísimos!!

A mis sinodales por su apoyo, sus consejos y recomendaciones.

Agradezco el apoyo brindado por CONACYT, para la realización de mis estudios de Posgrado.

A todos y cada uno de mis maestros, compañeros y amigos:

GRACIAS POR SU PACIENCIA!!

Capítulo I

Introducción

1.1 Biología Molecular

La Biología Molecular es el campo de la ciencia en el que la Biología, la Informática y la Estadística se unen para formar una única disciplina cuyos objetivos son permitir el descubrimiento de nuevas visiones biológicas y crear una perspectiva global que unifique los criterios que forman parte de la Biología. Dentro del Proyecto Genoma Humano puede encontrarse la siguiente definición sobre la Biología Molecular: El estudio de la estructura, función y composición de las moléculas biológicamente importantes [1].

1.2 La célula, el ADN y los cromosomas

La célula es la unidad estructural y funcional básica de todos los organismos. Una célula es el elemento de menor tamaño que puede considerarse vivo. Las células animales y células vegetales difieren fundamentalmente. Dentro del núcleo de la célula se encuentra el Ácido Desoxirribonucleico (ADN), material genético de todos los organismos celulares y casi todos los virus. Es el tipo de molécula más compleja que se conoce. Su secuencia de nucleótidos contiene la información necesaria para poder replicar un ser vivo.

El ADN lleva la información necesaria para dirigir la síntesis de proteínas y la replicación. En casi todos los organismos celulares el ADN está organizado en forma de cromosomas, situados en el núcleo de la célula. Está formado por la unión de muchos desoxirribonucleótidos. La mayoría de las moléculas de ADN poseen dos cadenas antiparalelas (una 5'-3' y la otra 3'-5') unidas entre sí mediante las bases nitrogenadas, por medio de puentes de hidrógeno [1].

El ADN contiene cuatro nucleótidos llamados bases nitrogenadas; y estos son: adenina (A), guanina (G), citosina (C) y timina (T). La adenina enlaza con la timina; mediante dos puentes de hidrógeno, mientras que la citosina enlaza con la guanina; mediante tres puentes de hidrógeno. Al analizar cadenas de ADN dentro de la célula, podemos distinguir varios niveles de discretización. En su forma más general, el ADN dentro de la célula se organiza en cromosomas, los cuales son estructuras genéticas auto-replicantes. Podemos pensar en ellos como una secuencia lineal de nucleótidos. Cada cromosoma tiene un patrón de longitud y de bandas característico. El número de cromosomas en una célula es diferente de organismo a organismo, como se observa en la Tabla 1.

Tabla 1. Cada organismo contiene diferente número de células
Una célula haploide es aquella que contiene un solo juego de cromosomas o la mitad (n) y las células diploides ($2n$) son las células que tienen un número doble de cromosomas.

Organismo	Número de cromosomas Haploide (n)	Número de cromosomas Diploide ($2n$)
Home sapiens (Humano)	23	46
Chimpancé	24	48
Perro	39	78
Gato	19	38
Vaca	30	60
Caballo	32	64
Pollo	39	78
Rana	13	26
Carpa	52	104
Cerdo	19	38
Jirafa	31	62

1.3 Genotipo y fenotipo

Genotipo y fenotipo son términos muy comunes en el área de la genética y la genómica y son empleados para diferenciar la composición genética de un organismo y la forma en que esta composición se expresa. El genotipo representa todo el conjunto de genes que identifica a una especie. Es toda la información genética que se encuentre en el ADN.

Básicamente, el genotipo determina el tipo de rasgos que podrán observarse y medirse en el fenotipo. Por ejemplo, los rasgos genotípicos de los organismos determinan su susceptibilidad para desarrollar determinada enfermedad, o su capacidad para desarrollar determinadas características productivas, en el caso de especies pecuarias. De hecho, es la variación genotípica entre los cromosomas XX y XY la que determina el sexo del organismo. Es importante aclarar que aunque el genotipo determina ciertas características,

los factores ambientales también son importantes. Como contraste, el fenotipo está representado por los caracteres que se ven o se miden, como es el caso de los caracteres cuantitativos. A diferencia del genotipo, que es invariable, el fenotipo se puede ver modificado a lo largo de la vida de un animal. Por ejemplo, el fenotipo pueden ser rasgos físicos y/o fisiológicos como las enfermedades, la producción de leche en las diferentes lactaciones, proteína láctea y los porcentajes de grasas. No obstante, el fenotipo, respecto a caracteres cualitativos (el color de piel, la estatura, presencia o ausencia de cuernos) es invariable.

1.4 Mejoramiento genético de ganado bovino

El mejoramiento genético con fines de aumentar la productividad del ganado bovino comenzó a lograrse mucho antes de que el mecanismo hereditario fuera entendido. Siempre se ha buscado maximizar la probabilidad de que los padres transmitan sus características productivas a sus crías. Históricamente, todas las estrategias de mejora genética han estado basadas en la premisa de realizar la mejor selección de los animales a ser utilizados para reproducción [2,3].

- Haciendo cruzamiento de lo mejor con lo mejor hemos estado mejorando las especies por milenios



- Desde un inicio se a buscado maximizar la probabilidad de que los padres transmitan sus características productivas a sus crías

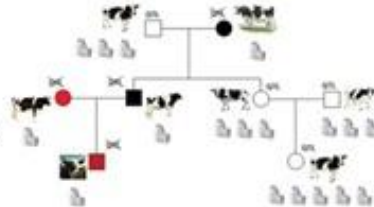


Figura 1. Mejoramiento genético en ganado bovino

Haciendo cruzamiento de los mejores sementales con las mejores vacas se ha logrado que los padres trasmitan sus características productivas a sus crías.

Los inicios de los métodos tradicionales y modernos de selección datan de los 1700's, cuando el Ingles Robert Bakewell estableció las primeras estrategias para crianza de ganado bovino de carne [4]. Bakewell definió el ganado que consideró ideal (ejemplo: machos con buenos aplomos, bloques bien formados y de maduración rápida, entre otras características), y con esto propuso el objetivo de la crianza buscando obtener el ganado ideal. Su estrategia fue cruzar lo mejor con lo mejor sin importar la relación sanguínea entre los animales, cosa que no siempre lo condujo a los mejores resultados. Los métodos modernos de crianza están soportados en dos enseñanzas básicas del trabajo de Bakewell: 1) es de crucial importancia definir un objetivo o meta específica en un programa de crianza, y 2) las mediciones de mérito genético; tal como las diferencias esperadas de progenie, DEPs [5], cuentan en toda la información de rendimiento de los animales emparentados.

Asignar un valor cuantitativo a la genética de un animal no es un trabajo sencillo debido a que se sabe que muchos de los rasgos fisiológicos observables (fenotipos) del animal, de sus propiedades bioquímicas, de su morfología, de su desarrollo, comportamiento y producción, son epigenéticos. Esto es, son regulados por la interacción de los genes con el medio ambiente. Lo que genera; que según los factores de manejo, la raza, la región geográfica, la época del año y el estado de lactancia, entre otros; sumado a la naturaleza dominante de algunos genes que posee el animal, al efecto aditivo de genes que actúan de forma independiente y al efecto epistático de genes que actúan de forma interactiva para controlar los rasgos del animal, influya de forma determinante en el rendimiento productivo.

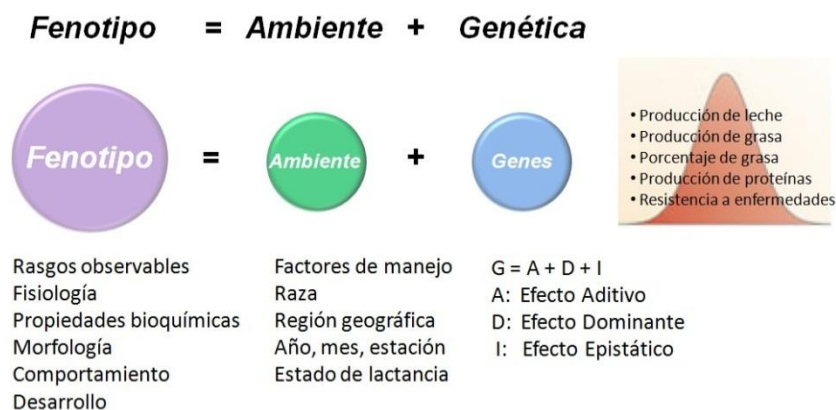


Figura 2. Efecto epigenético

La mayor parte de los rasgos de importancia económica en el ganado bovino son epigenéticos. La mayor parte de los fenotipos son regulados por la interacción de los genes con el medio ambiente.

A medida que se ha ido entendiendo el fenómeno epistático en el contexto productivo de los animales, se ha ido mejorando la técnica para asignarles un valor genético. De hecho, una fuente invaluable de información que ha contribuido al avance de

los programas de mejora fue la creación, hace más de 100 años, de las asociaciones de razas, con el objetivo de generar registros de manadas y recopilar de forma organizada información de pedigrí y de rendimiento del ganado. Esta información condujo eventualmente a experimentos efectivos de crianza selectiva, con el fin de incrementar la productividad, salud y conformación. Muchas razas de ganado bovino son ahora sujetas a selección de múltiple rasgos usando estimaciones del mérito genético, producido por métodos estadísticos y computacionales complejos [6,7]. Por ejemplo, el mérito genético para la entrega de leche en ganado lechero Holstein incrementa actualmente a una tasa de 1%, o 21 galones por año. A continuación se presenta una revisión de los métodos de selección de animales dentro de programas de mejora genética.

1.5 Métodos de selección de ganado bovino

Los métodos de selección de ganado dentro de los programas de mejora genética han evolucionado con la tecnología y la experiencia acumulada en los últimos 300 años. Los métodos difieren básicamente en la forma de obtener el valor genético de los animales. Estos podemos clasificarlos como sigue: método tradicional, método asistido por marcadores y métodos de selección genómica. A continuación se hace una descripción de cada uno de estos.

1.5.1 Selección tradicional y basada en marcadores

Con el método de selección tradicional, que utiliza información fenotípica y de pedigrí, se logra tener una efectividad del 30 – 40% en las predicciones. Esto se debe en gran parte a que, aun cuando logramos la mejora genética al ver un aumento en nuestra producción, el

conjunto de genes que han sido favorecidos (o desfavorecidos) en nuestro programa de selección sigue siendo una caja negra.

Un adelanto crucial en la integración del análisis genético en los esquemas de selección fue la inclusión de los marcadores moleculares, con los cuales se integra de forma específica el efecto de los genes en el rendimiento productivo del animal. Un marcador de ADN es una posición, o región, dentro de un cromosoma, físicamente identificable y cuya heredabilidad puede ser monitoreada. La función inicial y básica de los marcadores es identificar la expresión de genes relacionados con las características de importancia económica, y utilizar esta información para hacer una estimación más acertada de los valores estimados de crianza (EBVs).

Aun cuando la idea de usar marcadores de ADN para incrementar la tasa de ganancia genética en el ganado lechero, ha existido por décadas [8,9] , la adopción de la selección asistida por marcadores (MAS, “Marker Asisted Selection”, por sus siglas en ingles) en la industria lechera había estado limitada, hasta hace muy poco tiempo. Existían deferentes razones para esto; para muchos rasgos cuantitativos, de importancia económica, tales como los rasgos de producción y salud, en el ganado lechero, una gran cantidad de locus afectan a cada rasgo, donde cada locus captura una proporción limitada de la varianza genética total[10,11,12,13]. Consecuentemente, solo una ganancia relativamente pequeña era posible con el número limitado de marcadores que se tenían disponibles, y el costo de genotipificación de estos era alto. Adicionalmente, la complejidad para el cálculo de los valores genéticos de crianza incluyendo la información de los marcadores era una barrera más para la aplicación de la selección asistida por marcadores [14].

1.5.2 Selección genómica

La selección genómica (SG) es una forma de hacer selección asistida por marcadores genéticos (MAS), en la cual los marcadores están distribuidos a lo largo de todo el genoma, y todos los Locus de Características Cuantitativas (QTLs) están en desequilibrio de ligamiento con al menos un marcador [15]. La Selección Genómica está revolucionando a la industria del ganado. En esta, la toma de decisiones de selección de ganado para su reproducción está basada en el valor genómico de cría de los animales (GEBV, “Genomic Estimated Breeding Value”, por sus siglas en ingles).

Los GEBVs son calculados como la suma de los efectos de una cantidad grande de marcadores moleculares, distribuidos a lo largo de todo el genoma. Los genotipos y/o haplotipos de los marcadores involucrados capturan los efectos de todos los QTLs que contribuyen a la variación de un rasgo de valor económico [16,17,18].

La revolución de la selección genómica empezó con dos desarrollos claves. El primero fue la reciente secuenciación del genoma bovino[19], lo cual habilitó el descubrimiento de miles de marcadores de ADN, en forma de polimorfismo de nucleótido simple, SNPs[20]. El descubrimiento y caracterización de miles de marcadores SNPs en el genoma bovino y otros genomas de especies pecuarias trajo como consecuencia una reducción dramática en los costos de genotipificación.

El segundo desarrollo fue la demostración de que es posible tomar decisiones de selección con alta precisión cuando los valores genéticos de los animales son predichos usando una cantidad densa de marcadores [21].

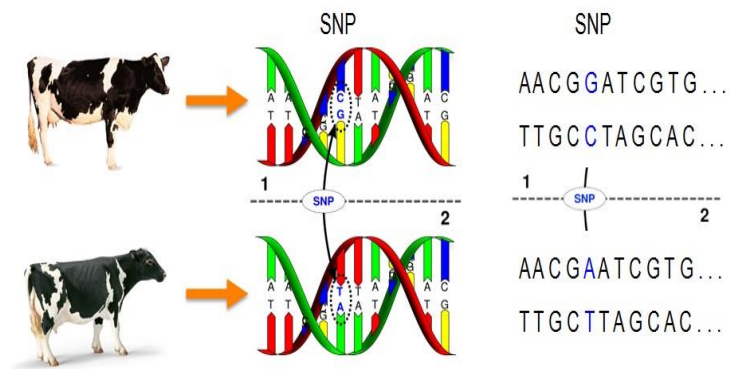


Figura 3. Marcadores SNP.

La forma más común de polimorfismos en los genomas son Polimorfismos de Nucleótido Simple.

La tecnología de selección genómica involucra tres pasos principales; 1) captura de las muestras y obtención del ADN genómico, 2) Aplicación de SNPchips para obtención de los genotipos y 3) Análisis bioinformático de la información para obtener los GEBVs. La captura de la muestras pueden ser hechas de esperma, sangre, saliva, segregación nasal o una pinchada en oreja, entre otras. Estas muestras deben ser transportadas al laboratorio para su procesamiento químico de extracción de ADN. El ADN debe ser ADN genómico completo y purificado. El ADN es amplificado y depositado en SNPchips. Se realiza una reacción de hibridización del ADN al SNPchip para posteriormente correr una reacción de extensión en la cual el valor alélico de cada posición polimórfica queda visible por elementos fluorescentes. Por medio de un escáner laser son medidos los valores de los alelos en los polimorfismos y se genera una base de datos con la información genética de cada uno de los animales de la muestra.



Figura 4 Selección Genómica

La selección genómica involucra tres pasos principales: 1) captura de la muestras y obtención de ADN genómico, 2) extracción de genotipos y 3) Análisis de la información.

Métodos bioinformáticos de análisis son aplicados a los datos para realizar estudios de asociación de genes con rasgos de importancia económica, con enfermedades y para obtener los GEBVs. Toda esta información resultante es integrada a los criterios establecidos para seleccionar sementales y vacas que serán utilizadas para generar las nuevas crías en el hato.

1.6 Tecnologías de genotipificación de salida masiva

En esta sección se describe con más detalle las tecnologías de genotipificación masiva que hacen uso de SNPchips para muestrear la genética de los animales.

Los polimorfismos de nucleótido simple son el tipo de variación genética más común en los genomas, millones de estos han sido detectados y caracterizados en distintos organismos, tales como los humanos y bovinos [20,22]. Existen métodos biomoleculares que pueden identificar de forma directa e independiente la información genotípica y haplotípica de los SNPs. Sin embargo, debido a que son de alto costo y consumen mucho tiempo, estos son

usados solo para cantidades pequeñas de muestras [23]. Para muestras a gran escala (cientos y miles de individuos), se han desarrollado las Tecnologías de Genotipificación de Salida Masiva (TGSM); con las cuales miles de SNPs son capturados de forma simultánea en un solo ensayo [24,25,26].

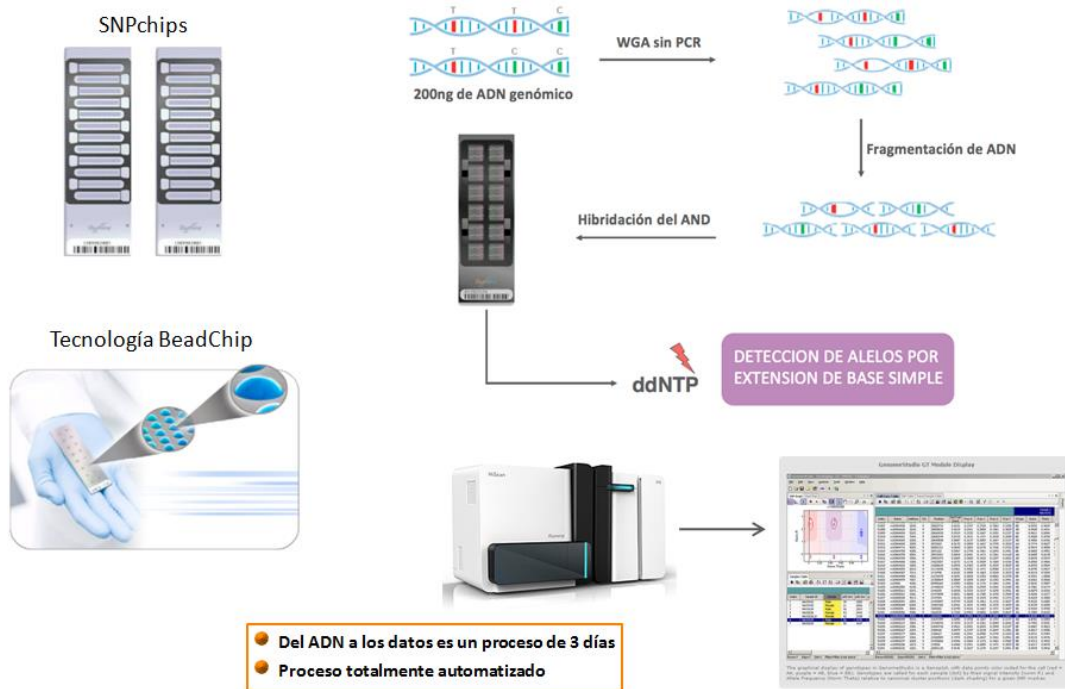


Figura 5. Ejemplo de una TGSM
Ensayo *Infinitum* de la empresa Illumina. Puede muestrear miles de SNPs en distintos individuos de forma simultánea.

Estas tecnologías están permitiendo la implementación de estudios de asociación de genoma completo a gran escala para detectar variaciones genéticas comunes, que tienen un rol significativo en la identificación de locus asociados con enfermedades, proteínas, biomarcadores, y respuestas farmacogenómicas [27]. Las TGSMs están basadas en el uso de SNPchips para la captura de los genotipos. A continuación se describe a más detalle los SNPchips.

1.6.1 SNPchips

Un SNPchip es un arreglo matricial en el cual cientos de miles de pruebas (oligos) son colocadas en un chip (laminilla) habilitando la medición de miles de SNPs de forma simultánea [28]. El valor de los alelos en la posición de un SNP es capturado hibridizando un oligo de ADN complementario en el sitio del SNP. La figura 8 muestra de forma esquemática la arquitectura de un SNPchip.

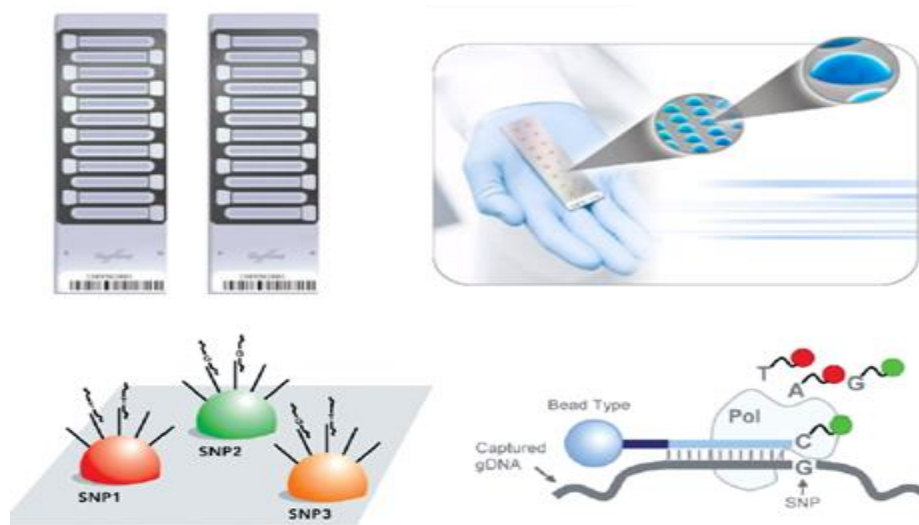


Figura 6. Diagrama esquemático de un SNPchip.
La estructura física y la reacción química general que se efectúa en el chip.

Actualmente, en el mercado existen diferentes versiones de SNPchips para genotipificar ganado bovino, entre las cuales encontramos los de la empresa Illumina, que capturan 777962, 54609 y 6912 SNPs a lo largo de los 30 cromosomas del genoma bovino. En especial el de 50000 SNPs (BovineSNP50) fue utilizado para caracterizar la estructura genética del genoma bovino [29], es actualmente el estándar para realizar Selección Genómica, debido a su costo y a que nos permite realizar estudios de asociación de genoma completo [30]. De la empresa Affimetrix: Axiom Bos 1 que captura 648875 SNPs.

El desarrollo de las TGSM y el aumento de la capacidad de los SNPchips han permitido la implementación de Selección Genómica dentro de los programas de mejora genética [13,15]. Con ésta, es posible obtener el valor genómico de cada animal con respecto a la característica de importancia que deseamos mejorar en nuestro hato (ej: producción de leche, tasa de preñez, peso al nacer, susceptibilidad a cierto padecimiento, etc.), y seleccionar ganado genéticamente óptimo para la producción, garantizando la transmisión genética adecuada de acuerdo a las características propias de medio ambiente y recursos de las regiones donde se sitúan los hatos lecheros. Los países desarrollados han adoptado rápidamente esta tecnología, y proyectan grandes avances en sus programas de mejora genética [15]. En la tabla 2 se observa un resumen de los SNPchips.

Tabla 2. Resumen de los SNPchips comercializados y el tipo de análisis que puede ser realizado con ellos.

SNPchip	Total de SNPs	Animales por chip	Propósito principal
BovineLD	6,909	24	Imputación, cálculo de Merito Neto
BovineLD V1.1	6,912	24	Imputación, cálculo de Merito Neto
BovineSNP50V.1	54,001	24	Selección Genómica, identificación de QTLs, evaluación de mérito genético, comparación genética
BovineSNP50V.2	54,609	24	Selección Genómica, identificación de QTLs, evaluación de mérito genético, comparación genética
BovineHD	777,962	8	Variación genética dentro de poblaciones, estudios de asociación GWAS, identificación de QTLs, evolución de mérito genético, mapeo entre razas, desequilibrio de ligamiento, comparación genética, caracterización de razas para estudios de biodiversidad
Affimetrix BOS 1	648,874	10	Merito genético, estudios de asociación GWAS, evaluación genómica, estudios de respuesta a fármacos, desequilibrio de ligamiento

Sería de gran impacto adoptar en la región de Baja California la tecnología moderna (Selección Genómica), para garantizar el progreso de la industria pecuaria, en general, y lechera en particular. Sin embargo, el factor más importante que limita la transferencia de esta tecnología es el factor económico; debido a que la situación actual de los ganaderos productores de leche es complicada.

1.7 Imputación de genotipos

La imputación de genotipos es el proceso de inferir o predecir genotipos que no han sido directamente capturados en un ensayo de genotipificado [40]. La imputación puede ser usada para diferentes propósitos. Dos de las aplicaciones más utilizadas son:

- 1.- Inferencia de genotipos faltantes en un panel, al fusionarlo con otro panel que ha sido genotipificado en un ensayo distinto.
- 2.- Inferencia de genotipos faltantes en un panel, utilizando un panel completo de referencia de haplotipos de mayor densidad; muestreado en individuos distintos de la misma raza.

En la figura 7 se muestra la fusión de información de diferentes fuentes para estudios combinados. La meta es asignar genotipos en dos muestras que han sido genotipificadas en dos ensayos distintos.

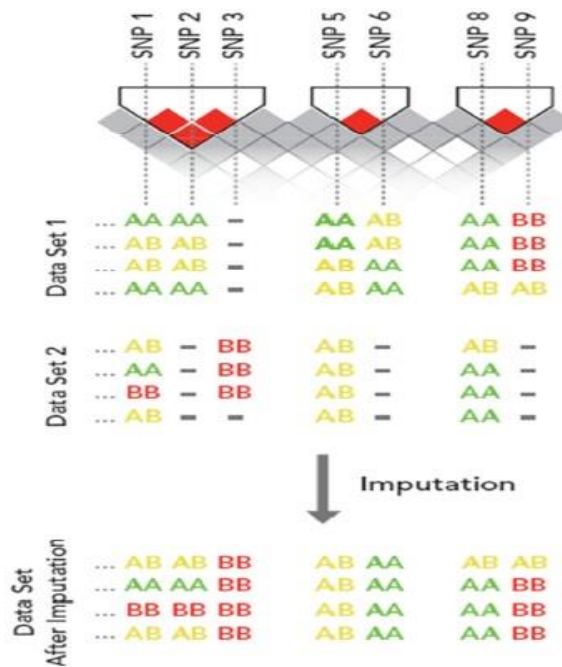


Figura 7. Imputación fusionando dos conjuntos de genotipos

La imputación provee un valor de probabilidad para cada uno de los tres posibles genotipos (AA, Aa y aa) y decide sobre el genotipo más probable.

Para la imputación basada en paneles de referencia de haplotipos, el razonamiento biológico es que los SNPs pueden ser modelados como mosaicos de haplotipos que contendrán valores iguales debido a que son idénticos por descendencia. Se infieren los haplotipos de los genotipos existentes en la muestra, y se busca la mejor alineación con el panel de referencia. Los genotipos faltantes son imputados, basados en los haplotipos con mejor alineación [41]. La figura 8 muestra un ejemplo de este tipo de imputación.

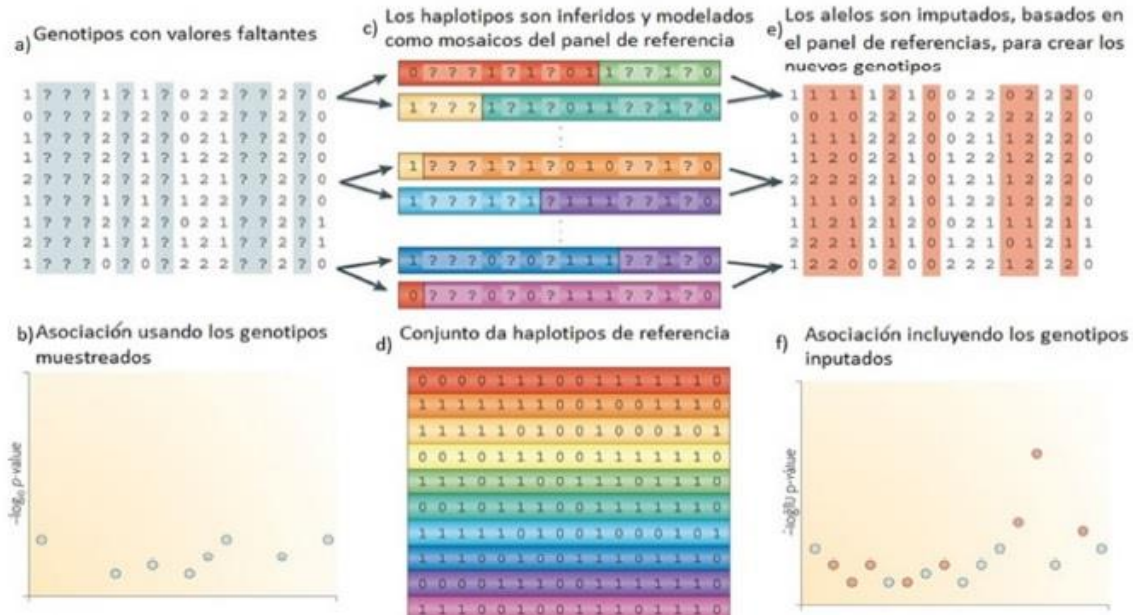


Figura 8. Imputación con paneles de referencia de haplotipos

Una definición formal del problema de imputación se expresa como sigue:

Asumiendo que se cuenta con los datos de L SNPs autosomales bi-alelicos, y que los dos alelos en cada SNP han sido codificados con 0 y 1:

- Sea H un conjunto de haplotipos en L SNPs, de un conjunto de individuos de cierta raza.
- Sea G el conjunto de genotipos de los L SNPs en K individuos de la misma raza, distintos a los individuos del conjunto anterior, donde
- $G_i = (G_{i1}, \dots, G_{iL})$ denota los genotipos del i_{th} individuo.
- Los genotipos individuales son ya sea observados, $G_{ik} \in \text{existente}$, o no observados, $G_{ik} = ? = \text{faltante}$.
- El objetivo es predecir, utilizando el grupo completo de haplotipos del primer conjunto, los genotipos de aquellos SNPs que no han sido genotipificados en el segundo conjunto.

1.8 Bioestadística

Los actuales desarrollos en la biotecnología han llevado a un crecimiento exponencial en los estudios sobre el análisis del ADN en organismos, sus funciones, interacciones entre sus biomoléculas y sus rutas metabólicas. Con estas nuevas tecnologías, también se ha incrementado la información en las bases de datos, mismas que demandan de nuevas herramientas para almacenar, analizar e interpretar los grandes volúmenes de información. Debido a que la información es compleja y cruda, se requiere de la adecuación a través de la normalización, estandarización, escalamiento u otros medios, en la información para poder aplicar los métodos estadísticos, y al final obtener datos que sean fáciles de interpretar.

En los siguientes apartados veremos la descripción de algunos métodos aplicados en la Bioestadística.

1.8.1 Prueba X^2 de Pearson (Chi Square).

Es una prueba de hipótesis que compara la distribución observada de los datos con una distribución esperada de los datos.

Existen varios tipos de pruebas de Chi-Cuadrado como se describe a continuación

1.8.1.1 Prueba de Chi-Cuadrado de bondad de ajuste

Este análisis se utiliza para probar qué tan bien una muestra de datos categóricos se ajusta a una distribución.

Se entiende por bondad de ajuste a la asimilación de los datos observados de una variable a una función matemática previamente establecida y reconocida. A través de esta es posible entonces predecir el comportamiento de la variable en estudio [31].

En el caso de que las diferencias sean grandes entre O (frecuencias observadas) y E (frecuencias esperadas) el valor de la X cuadrada X^2 será mayor y, en ese caso se rechaza hipótesis nula. Es difícil suponer que las diferencias entre O y E son iguales o idénticas.

En resumen, si el valor de X^2 es mayor de lo que el azar anticiparía, se puede descartar la hipótesis nula.

$$X^2 = \sum_{i=1}^r \frac{(O_i - E_i)^2}{E_i}$$

1.8.1.2 Pruebas de Chi-Cuadrado de asociación e independencia

Los cálculos para estas pruebas son iguales, pero la pregunta que se está tratando de contestar puede ser diferente.

- **Prueba de asociación:** utiliza una prueba de asociación para determinar si una variable está asociada a otra variable.
- **Prueba de independencia:** utiliza una prueba de independencia para determinar si el valor observado de una variable depende del valor observado de otra variable.

$$X^2 = \sum_{i=1}^k \frac{(n_i - E_i)^2}{E_i}$$

1.8.2 Intervalos de Confidencia

El cálculo de intervalos de confidencia para la estimación de parámetros son técnicas que nos permiten hacer declaraciones sobre qué valores podemos esperar para un parámetro.

El intervalo calculado dependerá de:

- Lo estimado en la muestra (porcentaje, media,...). El intervalo de confianza esta formado por valores ligeramente menores y mayores que la aproximación ofrecida por la muestra.
- El tamaño muestral. Cuantos más datos hayan participado en el cálculo, más pequeño esperamos que sea la diferencia entre el valor estimado y el valor real desconocido.

- La probabilidad (nivel de confianza) con la que el método dará una respuesta correcta. Niveles de confianza habituales para los intervalos de confianza son el 95% y el 99%.

Por consiguiente, la siguiente expresión proporciona un intervalo de confianza para la media de la población a un nivel de confianza $1 - \alpha$:

$$\left(\bar{x} - t_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n-1}}, \bar{x} + t_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n-1}} \right)$$

En lo sucesivo no tendremos más que sustituir los valores muestrales, en la expresión anterior, para obtener las correspondientes estimaciones [31].

1.8.3 Significancia Estadística

En estadística, un resultado se denomina **estadísticamente significativo** cuando no es probable que haya sido debido al azar. Una "diferencia estadísticamente significativa" significa que hay evidencias estadísticas de que existe una diferencia entre las variables estudiadas. El nivel de significancia es un concepto estadístico asociado a la verificación de una hipótesis. En pocas palabras, se define como la probabilidad de tomar la decisión de rechazar una hipótesis nula (H_0) cuando ésta es verdadera (decisión conocida como Error tipo I, o "falso positivo"). La decisión se toma a menudo utilizando el valor P (o p-valor): si el valor P es inferior al nivel de significación, entonces la hipótesis nula es rechazada. Cuanto menor sea el valor P, más significativo será el resultado[31].

1.8.4 El Valor-p

El *Valor-P* es un valor de probabilidad, que oscila entre 0 y 1, por lo que los valores muy próximos a 1, no rechazan la hipótesis nula y sus valores muy aproximados a 0, rechazan la hipótesis nula.

- Un valor bajo de p (por ejemplo, $p=0.0002$) prueba que la diferencia encontrada se debe a la intervención, si es un estudio experimental, o a la exposición en uno de los casos y controles.
- Un valor bajo de p demuestra relación causa-efecto entre una exposición y padecimiento.
- Un valor bajo de p establece la fuerza de la asociación.
- Un valor bajo de p demuestra la validez del estudio.
- Un valor de p alto (por ejemplo, $p>0.05$) significa que no hay asociación entre la exposición A y el resultado B o que no hay diferencia entre exposición y no exposición, etc.

1.8.5 Regresión de mínimos cuadrados

En general, los Modelos de Regresión son métodos matemáticos para modelar la relación estocástica cuantitativa entre una variable de interés y una (o un conjunto) de variables explicativas [31]. Estos modelos pueden ser expresados como sigue:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + \varepsilon_i$$

Dónde:

Y_i : variable de interés, dependiente o regresando,

$X_{1i}, X_{2i}, \dots, X_{pi}$: variables explicativas, independientes o regresores,

β_0 : intersección o termino constante,

$\beta_1, \beta_2, \dots, \beta_p$: parámetros, miden la influencia que las variables explicativas tienen sobre el regresando,

p : número de parámetros independientes a tomar en cuenta,

ε : error de observación debido a variables no controladas,

$i: 1, 2, \dots, n$ número de observaciones de las variables.

Con estos modelos es posible estudiar relaciones lineales entre múltiples variables y el efecto que estas tienen sobre la variable dependiente. Los β_i se estiman siguiendo el criterio de mínimos cuadrados:

$$\min_{\beta \in n} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} - \dots - \beta_p X_{pi})^2$$

y los estimadores de mínimos cuadrados son obtenidos a partir de la ecuación:

$$\hat{\beta} = (X^T X)^{-1} X^{-T} Y$$

1.8.6 Corrección de Bonferroni

La corrección de Bonferroni es una herramienta para ajustar los resultados estadísticos en comparaciones múltiples. Por ejemplo, si probamos una hipótesis al nivel $p=0.05$, entonces tenemos un 5% de probabilidad de cometer un error de tipo "I", es decir, de rechazar falsamente la hipótesis nula. Pero si ponemos a prueba más hipótesis, la posibilidad de cometer un error de tipo "I" en cualquiera de ellas aumenta [31].

La corrección de Bonferroni establece como significativos valores de $\alpha/Nsnps$

Donde:

$$\alpha=0.05,$$

N_{snps} = es la cantidad de SNPs utilizados en el estudio.

1.9 Recursos Bionfórmáticos

Una nueva disciplina denominada Bioinformática que agrupa conceptos tanto de biología como de informática, matemática o de la física y que permite cubrir la necesidad de manejo de enormes y complejas cantidades de información provenientes, tanto de la secuenciación de las macromoléculas (ADN, proteínas y glúcidos) como de las técnicas de análisis masivo del comportamiento de genes y proteínas. El Proyecto del Genoma Humano (HapMap Project) ha constituido ser uno de los retos científico-tecnológicos más grandes que ha enfrentado la comunidad científica, la primera secuencia completa y estable fue declarada en el 2003.

A partir de ese momento grupo de investigadores han desarrollado esplendidos lugares en línea donde se encuentra una gran diversidad de bases de datos, herramientas sofisticadas de análisis, y distintos tipos de software, los cuales están disponibles, con el fin de acelerar el avance a favor de la ciencia. A continuación se presenta una breve descripción de algunas herramientas bioinformáticas utilizadas y las bases de datos consultadas.

1.9.1 Lenguaje de programación Perl

Perl (Practical Extraction y Report Language, por sus siglas en inglés), es un lenguaje de programación de alto nivel. Es el lenguaje más utilizado en análisis de información biológica, es de libre acceso, eso quiere decir que es gratuito. Un script genérico de Perl puede ejecutarse en cualquier plataforma en la que tengamos un intérprete disponible. Perl se deriva del lenguaje C de programación, así como el sed, awk y el Shell de Unix, además

de otras herramientas y lenguajes. Las facilidades para la manipulación de procesos, archivos, y texto hace que este lenguaje este particularmente bien situado en las tareas donde involucran el rápido desarrollo de programas, desarrollo de utilerías para el sistema operativo, herramientas de software, manejo de base de datos, programación de gráficas, redes y el World Wide Web. Se puede visitar www.perl.org para mayor información sobre este lenguaje [32].

1.9.2 Software estadístico R

R es uno de los software estadísticos más flexibles, potentes y profesionales que existen actualmente. Este permite realizar análisis estadísticos de todo tipo, desde lo más elemental, hasta lo más avanzado y es el más usado en la comunidad científica. Cuenta, además, con la ventaja de ser gratuito y de descarga e instalación sencilla. Este programa está disponible en la página web: <http://www.r-project.org> y consta de una aplicación central y de librerías de multitud de temas que se pueden instalar según la necesidad [33].

1.9.3 FindHap V4

La imputación de genotipos se realizó con el algoritmo FindHapV4, descrito por Vanraden et. al [42]. El algoritmo parte del hecho de que genotipos que no se conocen pueden ser elucidados (imputados) partiendo de genotipos observados en los mismos loci (o loci cercanos), de individuos de la misma familia, utilizando haplotipado de pedigrí, o comparando patrones en los alelos (sin usar el pedigrí) usando haplotipado de la población. De forma general, el algoritmo genera una lista de haplotipos, partiendo de los genotipos, y en un proceso iterativo los haplotipos son refinados tomando en cuenta el pedigrí o patrones

de alelos. Los genotipos son comparados con haplotipos refinados para decidir el valor a imputar.

Para este trabajo, obtuvimos los genotipos reales de nuestra población (47 animales), utilizando el SNPchip BovienLD el cual captura 6912 SNPs. Y, utilizando un panel de referencia de ganado Holstein del proyecto Bovine HapMap Consortium (http://bovinegenome.org/?q=hapmap_funding) hicimos la imputación a 46523 SNPs del chip BovineSNP50 versión 2.

1.9.4 PLINK

Es un conjunto de herramientas de código abierto para realizar análisis de asociación de genoma completo, diseñado para analizar cantidades grandes de marcadores y fenotipos [34].

1.9.5 Bases de Datos Genéticos y Genómicos Públicas

Las bases de datos se derivaron a partir del proyecto de secuenciación del genoma humano (HapMap). Existen tres servidores internacionales que albergan los datos del análisis del genoma humano: EMBL-Bank del EBI europeo, DDBJ (DNA Data Bank of Japan) en el CIB/NIG y GenBank en el NCBI de USA. Estas tres plataformas están conectadas en alianza para asegurar la disponibilidad de las secuencias al público en general y ninguna revista científica o análisis de diagnóstico molecular puede describir resultados de una secuencia de nucleótidos, proteínas o su interpretación sin hacer referencia a los depósitos de una de estas tres principales bases de datos.

En este caso en particular, utilizamos como referencia la base de datos del Genoma Bovino [19], en la cual se tiene acceso a distintas herramientas para análisis de secuencias de ADN en distintas anotaciones UMD 3.1 por ejemplo.

Existen otras bases de datos de servicio gratuito, entre las cuales se encuentra la Enciclopedia de Genes y Genomas Kyoto (KEGG, Kyoto Encyclopedia of Genes and Genomes) [35]. Esta base de datos alberga información sobre las funciones y utilidades de alto nivel de sistemas biológicos, en el mismo portal nos permite tener acceso a rutas biológicas (pathways), genes, genomas, fármacos y compuestos relacionados con dichas rutas. El portal de DAVID [36], mismo que realiza una consulta sobre el enriquecimiento funcional de los genes, utilizando la Ontología de Genes (GO) [37].

1.10 Métodos para estudios de asociación

Los estudios de asociación de polimorfismos de nucleótido simple (SNP) involucran métodos de análisis que buscan establecer la relación estadística con características de importancia económica en especies domesticas es la estrategia más utilizada para la caracterización de marcadores genéticos relacionados con enfermedades, con rasgos de importancia económica y con programas de mejora genética.

Ahora, dado que es poco probable que un único método analítico sea adecuado en todas las situaciones, es importante evaluar empíricamente las fortalezas y debilidades de una variedad de métodos computacionales.

1.10.1 Asociación Directa

Cuando un SNP que influencia al sistema biológico que regula la expresión del fenotipo es directamente genotipificado en el estudio, se encuentra que esta estadísticamente asociado con el fenotipo o rasgo [38].

1.10.2 Asociación Indirecta

Cuando un SNP que influencia al sistema biológico que regula la expresión del fenotipo no es directamente genotipificado en el estudio, pero en cambio, un SNP tag que se encuentra en alto desequilibrio de ligamiento con éste, es genotipificado; se encuentra estadísticamente asociado con el fenotipo o rasgo [38].

1.11 Control de Calidad de los datos de los SNPs

Al obtener los datos crudos, después del proceso de genotipificación es necesario aplicar filtros de control de calidad con el propósito de trabajar con datos homogéneos. Algunos de los filtros que pueden ser aplicados son los siguientes:

1.11.1 Errores de genotipificación

En el contexto de las aplicaciones que involucra este trabajo de tesis, un error de genotipificación se refiere a un error de la tecnología, ya sea en el manejo del ADN, en el depósito del mismo en el SNPchip o en el sistema de lectura de la información, que finaliza en una posición con falta de información. Este tipo de error es comúnmente llamado “error de no-call”. Por convención, un individuo o un SNP que presenta más del 2% de valores faltantes, es eliminado [20].

1.11.2 Frecuencia del Alelo Menor (MAF)

La frecuencia del alelo menor, abreviada MAF (Minor Allele Frequency, por sus siglas en Ingles), se define como la frecuencia del alelo menos común en un determinado locus, en una población. Para considerar una variación de nucleótido simple como un SNP, la variante (alelo menor) debe estar presente en al menos el 5% de la población, Aquellos alelos con $MAF \leq 0.05$ son considerados raros, y son candidatos a ser mutaciones [39]. Por convención, todos aquellos SNPs con un $MAF \leq 0.05$ son eliminados, y por ende no utilizados en los estudios de asociación.

1.11.3 Equilibrio de Hardy-Weinberg (HWE)

El equilibrio de Hardy-Weinberg hace referencia a la independencia de alelos en un locus entre los dos cromosomas homólogos. En concreto establece la relación entre las frecuencias alélicas y genotípicas en una población lo suficientemente grande, sin selección, mutación o migración. Si consideramos un locus con dos alelos “A” y “a” con frecuencias $p_A = p$ y $p_a = 1 - p = q$, si estos dos alelos que porta un individuo se heredan independientemente, el número de copias del alelo A sigue una distribución binomial $\text{Bin}(2, p)$. Por lo tanto las probabilidades de los genotipos AA, Aa y aa serán p^2 , $2pq$ y q^2 respectivamente. La probabilidad de que un alelo aparezca en un homólogo es independiente de que alelo esté presente en el segundo homólogo.

En un estudio de asociación es necesario verificar que los SNPs genotipificados se encuentran en equilibrio HWE, ya que desviaciones de este equilibrio pueden ser indicativas de errores de genotipificación o bien de la existencia de factores que podrían afectar a las frecuencias alélicas (por ejemplo, migración reciente) diferentes del rasgo en

estudio. Si no se examina puede dar lugar a asociaciones espurias. La comprobación del HWE se realiza sobre la población control, ya que desviaciones del HWE en casos pueden ser indicativos de una asociación .

$$p^2 + 2pq + q^2 = (p + q)^2 = 1$$

Por convención, todos los SNPs que no están en equilibrio de Hardy-Weinberg son eliminados de un estudio de asociación.

Capítulo II

Planteamiento del Problema

2.1 Situación actual de la industria lechera en la región de Baja California.

En México, la industria del ganado bovino representa, en términos monetarios, el 47.3% de la producción nacional pecuaria [43]. De este porcentaje, el 43.7% proviene del ganado utilizado para la producción de leche, siendo la tercera actividad pecuaria más importante, por encima de la producción de cerdo y huevo. Sin embargo, a pesar de que en México contamos con un importante número de cabezas de ganado bovino dedicado a la producción de leche, nos situamos entre las naciones con menor productividad [44]. Países como Canadá y Argentina que poseen una cantidad mucho menor de cabezas de ganado lechero, han reportado una producción mayor en los últimos años. En general, los factores

que han limitado los esfuerzos por incrementar la producción son de índole económico, de infraestructura, genéticos y ambientales; los cuales han provocado que la producción sea insuficiente, por lo que se ha tenido que importar una gran cantidad de leche, principalmente de países como Estados Unidos de América, Nueva Zelanda y Uruguay [43].

En la región de Baja California, la ganadería de leche ha presentado, históricamente problemas de producción, principalmente en las zonas áridas, tales como los municipios de Mexicali y San Felipe. Y, a pesar del esfuerzo que han realizado los ganaderos de la región por mantenerla como una industria necesaria para el desarrollo de la región, la producción ha sufrido una disminución paulatina y constante en los últimos años [45]. La figura 7 muestra como la producción ha venido disminuyendo en los últimos 11 años.

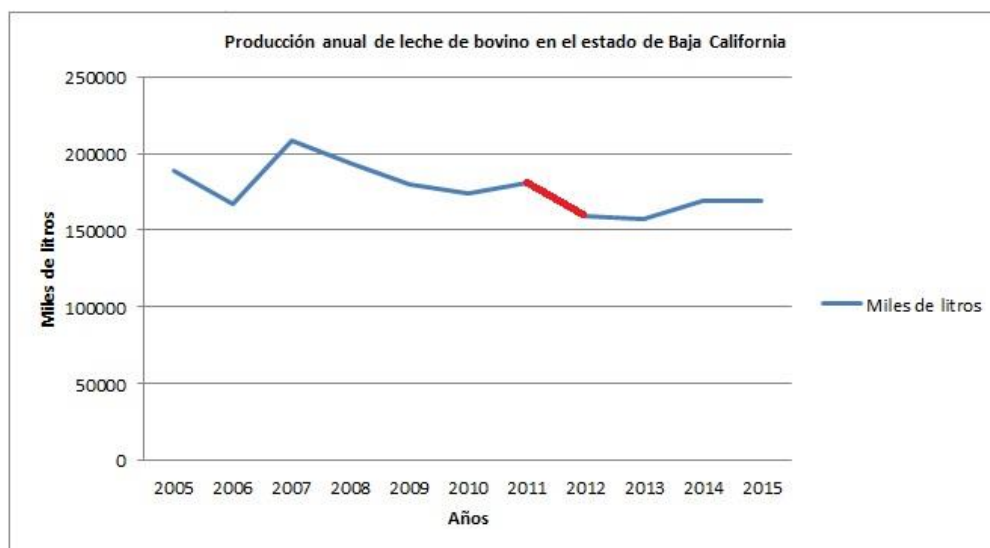


Figura 9. Producción de leche en los últimos años en el estado de Baja California (la información fue obtenida de la página de internet de la SAGARPA, en su sección de Servicios de Información Agroalimentaria Pesquera, diciembre de 2015).

En el año 2011, en adición a la disminución de la productividad lechera manifestada en los últimos años, se sumó un alto desabasto de forraje para alimentación del ganado, generando que las crisis en el sector lechero se agudizara. Esto provocó la pérdida de más de 10,000 cabezas de ganado lechero, que fueron vendidas para alimentar otras [46]. En un esfuerzo por frenar la agudización de la crisis en el sector lechero, miembros de la Unión Ganadera del Estado y directivos de la Comisión de Agricultura, Ganadería, Asuntos Pecuarios y Pesca del Congreso llevaron a cabo varias mesas de trabajo, acordando buscar y aplicar medidas para rescatar la industria lechera de Baja California y evitar que la crisis actual haga desaparecer esta industria en el estado [47].

Es de crucial importancia el establecimiento de programas integrales a ser implementados en los hatos lecheros del estado, para mejorar la producción ganadera. En especial la mejora genética del ganado, debe ser enfocada, con programas de reproducción, a las características productivas de interés económico. Para realizar mejora genética, será necesario realizar estudios de asociación de genes con rasgos productivos y de conformación en el ganado de Baja California. Por lo que, para formular este trabajo de tesis partimos de la importancia de las características de conformación del ganado lechero Holstein en las regiones áridas del estado, las cuales son tan importantes como otros rasgos (e.j. de producción y salud). Los rasgos de conformación, tales como la estatura, profundidad y tamaño del cuerpo, ángulo y ancho de la grupa, ubres y condiciones de piernas y patas, entre otras, están directamente asociados con la salud de los animales, por lo que son valorados y utilizados como indicadores de selección para la facilidad de reproducción y longevidad. Finalmente, el encontrar los genes asociados a características de conformación ayudará a generar las bases genéticas de distintos padecimientos y de

producción, en particular en zonas áridas y semiáridas, tales como el noroeste de México. Sin embargo, el abordar esta estrategia de solución impone dos retos que deben resolverse; por un lado el aspecto económico, dado que la condición de los ganaderos es limitada, y tendrá que encontrarse una solución eficaz y económica; y por otro lado la disponibilidad y uso de la tecnología genómica actual, que no se tiene en la región; por lo que la solución deberá implementarse de forma gradual, estudiando un rasgo de interés a la vez.

2.2 Objetivo General

En este trabajo de tesis se propone implementar un estudio inicial para encontrar los genes asociados con los rasgos de conformación en ganado lechero Holstein de Baja California. Estos rasgos son de suma importancia en regiones desérticas porque influye directamente en la facilidad de parto, en la salud, en aspectos funcionales y de longevidad [48, 49]. Igualmente se propone utilizar la tecnología de Genotipificación de salida masiva, y en especial el panel BeadChip BovineLD de la empresa Illumina, el cual muestrea 6,912 marcadores SNP a lo largo de todo el genoma bovino, debido a que es el panel más económico actualmente en el mercado.

De lo anterior se ha establecido el siguiente objetivo general de este trabajo:

“Evaluar el uso del panel de baja densidad BovineLD, que muestrea 6,912 SNPs, más 39,611 SNPs imputados a lo largo del genoma bovino, para encontrar genes asociados con 16 rasgos de importancia económica en ganado lechero Holstein de Baja California, y habilitar la implementación de programas de mejora genética de los rasgos que sean significativos, de una forma económica y eficaz”.

2.2.1 Objetivos específicos

1. Definición de la población de estudio. (Modelo experimental de producción lechera del Instituto de Investigaciones en Ciencias Veterinarias de la UABC),
2. Aplicación de muestreo para recolección de sangre y medición del fenotipo,
3. Extracción de ADN genómico de las muestras,
4. Genotipificación con el panel BeadChip BovineLD,
5. Utilizar un panel de referencia de ganado Holstein del proyecto Bovine HapMap Consortium (http://bovinegenome.org/?q=hapmap_funding) del chip BovineSNP50 versión 2, para realizar imputación,
6. Aplicar algoritmos bioinformáticos para realizar estudios de asociación de SNPs en los 16 rasgos de conformación,
7. Definición de los genes asociados con los rasgos que sean significativos,
8. Validación de los hallazgos por asociación con bases de datos Genéticos y Genómicos públicas.

Capítulo III

Desarrollo y Resultados

3.1 Descripción de la población

El hato de ganado lechero experimental del Instituto de Investigaciones en Ciencias Veterinarias de la Universidad Autónoma de Baja California, en Mexicali, Baja California, México; localizado a 3.5 km a lo largo de la carretera San Felipe de Mexicali, Baja California 32°24'27"N y 115°23'03"O. La altitud de la finca es de 8 m, alcanzando una temperatura máxima de 50 °C durante el verano, una temperatura mínima de -5 °C [50].

Un inventario físico de ganado de raza Holstein registradas en la Asociación Holstein de México, al 22 de abril 2015, constituido de la siguiente manera: a) 45 vacas en producción; b) 12 becerras; c) 23 vacas en gestación y d) 4 becerros. Todas las vacas nacieron después de inseminación artificial; todos eran clínicamente sanas y libres de

brucelosis y tuberculosis. Este hato es considerado como un hato joven con promedio general de edad de 4 años y 3 meses.

3.2 Muestreo para la recolección de sangre y medición de fenotipos

Se recolectaron muestras de sangre, en la vacas utilizando el método de inmovilización de la cola para la obtención de muestras de sangre de la vena coccígea, en los becerros de la yugular. La sangre se recolectó en tubos vacutainer (Vacutainer, Hemogar, USA) con anticoagulante EDTA, como se observa en la figura 10. Las muestras fueron etiquetadas con el número de muestra y número de arete del animal. Las cuales se depositaron en un refrigerador de -20 centígrados, en el Laboratorio de Biología Molecular del Instituto de Investigación en Ciencias Veterinarias.



Figura 10. Obtención de muestras de sangre de la vena coccígea y de la vena yugular

Se muestrearon 16 rasgos de interés económico, utilizando los rangos de medición establecidos en el documento de características descriptivas lineales de la Asociación Holstein, USA (Holstein Association USA). Los rasgos son los siguientes: Estatura, Pecho, Profundidad corporal, Angulosidad, Angulo de la grupa, Anchura de la grupa, Vista

posterior de las patas, Vista lateral de las patas, Angulo podal, Inserción anterior, Colocación de pezones anteriores, Longitud de pezones, Profundidad de la ubre, Altura inserción posterior, Ligamento suspensor medio y Colocación de pezones posteriores. La tabla 3 muestra la estadística resultante del fenotipo utilizado en este estudio.

Tabla 3. Resumen de las estadísticas del fenotipo

Fenotipo	N	Min	Max	Media	Sd	IC (95%)
Estatura (cm)	46	135	159	146.36	4.85	(144.95,147.75)
Pecho (cm)	46	30	36	34.62	2.45	(33.91,35.33)
Profundidad Corporal	46	6	9	7.53	1.51	(7.09,7.97)
Angulosidad	46	3	9	7.6	1.64	(7.12,8.07)
Angulo de la Grupa	46	3	9	7.6	1.64	(3.26,4.60)
Anchura de la Grupa (cm)	46	26	30	28.44	1.28	(28.047,28.81)
Vista posterior de las patas	46	1	8	5.84	2.11	(5.23,6.45)
Vista lateral de las patas	46	1	3	1.86	0.50	(1.72,2.01)
Angulo Podal	46	1	9	4.91	2.31	(4.24,5.57)
Inserción Anterior	46	1	3	2.13	0.62	(1.95,2.31)
Colocación Pezones Anteriores	46	1	8	4.48	1.53	(4.04,4.93)
Longitud de Pezones	46	3	9	6.04	1.93	(5.48,6.60)
Profundidad de la Ubre	46	1	8	5.6	2.73	(4.81,6.38)
Altura Inserción Posterior	46	0	6	0.26	1.25	(-0.09,0.62)
Ligamento Suspensor Medio	46	1	9	5.44	2.98	(4.58,6.30)
Colocación de Pezones	46	1	9	6.42	2.84	(5.59,7.24)
Posteriores						

N= Número de observaciones; Min= Mínimo, Max= Máximo, Media=Media, SD= Desviación estándar IC=Intervalos de confianza.

3.3 Extracción de ADN genómico de las muestras

La extracción y la purificación de ADN se realizaron con el kit de QIAGEN (QIAamp DNA Blood, QIAGEN, Germany). El cual comprende 4 etapas:

- Lisis de las células presentes en la muestra de sangre
- Unión del ADN genómico del lisado celular a la membrana de una columna de centrifugación QIAamp
- Lavado de la membrana
- Elución del ADN genómico de la membrana



Figura 11. Extracción de ADN genómico
Todas las muestras de ADN fueron analizadas por espectroscopia y electroforesis en gel de agarosa.

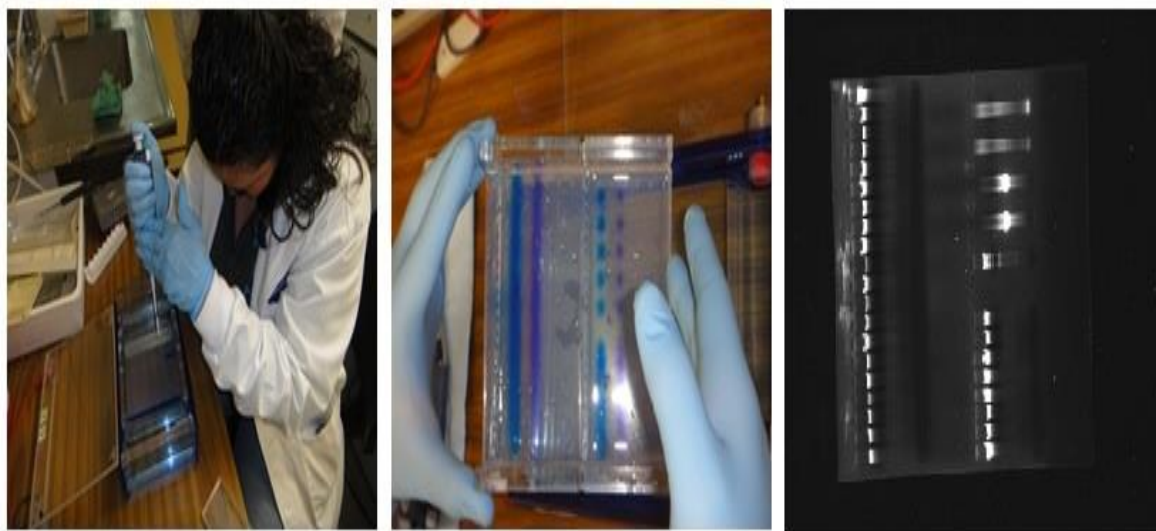


Figura 12. Análisis de ADN en gel agarosa

3.4 Genotipificación con el SNPchip BovineLD

Los chips permiten la genotipificación simultánea de miles de SNPs y su análisis aporta información adicional de caracteres económicamente importantes para la producción bovina relacionados con la producción lechera, reproducción, el crecimiento y la calidad de la carne y de la canal. La genotipificación fue realizada con el SNPchip BovineLD Illumina muestreando un total de 6,912 SNPs a lo largo de todo el genoma bovino; utilizando la secuencia del genoma bovino UMD 3.1 [35]. En total se genotipificaron 47 vacas elegidas entre las que estaban en producción y/o en gestación.

3.5 Imputación

La imputación de genotipos se realizó con el algoritmo FindHapV4, descrito por [42]. El algoritmo parte del hecho de que genotipos que no se conocen pueden ser elucidados (imputados) partiendo de genotipos observados en los mismos loci (o loci cercanos), de individuos de la misma familia, utilizando haplotipado de pedigrí, o comparando patrones

en los alelos (sin usar el pedigrí) usando haplotipado de la población. De forma general, el algoritmo genera una lista de haplotipos, partiendo de los genotipos, y en un proceso iterativo los haplotipos son refinados tomando en cuenta el pedigrí o patrones de alelos. Los genotipos son comparados con haplotipos refinados para decidir el valor a imputar.

Para este trabajo, obtuvimos los genotipos reales de nuestra población (47 animales), utilizando el SNPchip BovienLD, y, utilizando un panel de referencia de ganado Holstein del proyecto Bovine HapMap Consortium (http://bovinegenome.org/?q=hapmap_funding) hicimos la imputación a 46,523 SNPs del chip BovineSNP50 versión 2.

Con el propósito de trabajar con un conjunto de datos homogéneo minimizando errores de genotipificación e inconsistencias, a través del filtrado de control de calidad se eliminaron SNPs con errores de genotificación mayores al 0.05%, SNPs que no cumplieron con el principio de Hardy Weinberg ($<1.0E-3$), SNPs con frecuencia de alelo menor ($MAF < 0.05\%$) e individuos con un porcentaje mayor a 2% de valores faltantes, Finalmente, 46 muestras y un total de 22,151 SNPs a lo largo de los 29 cromosomas autosomales, que representa el 47.61% de la información inicial, fueron examinados en este estudio.

3.6 Aplicación de algoritmos bioinformáticos para realizar estudios de asociación de SNPs con los 16 rasgos de conformación

La asociación se realizó con dos herramientas distintas, con el objetivo de verificar la consistencia en los resultados. Primeramente se ejecutó una regresión por mínimos cuadrados utilizando la herramienta PLINK 1.07, en modo de línea de comando con los siguientes argumentos: `plink -bfile inputFile -assoc -pheno qt.phe -out outputFile`; donde *inputFileName* es el archivo con los genotipos y el pedigrí en formato

linkage, *assoc* le indica al algoritmo que realice la asociación con regresión cuadrática, *qt.phe* es un archivo con el vector de fenotipos, y *outputfileName* es el archivo de salida. Y, posteriormente se realizó la asociación utilizando Modelos Lineales Generalizados, con el software R. Para este último se ejecutó un procedimiento de dos pasos: en el primero se ejecutó una asociación simple para cada SNP; seleccionando aquellos que resultaron con un *valor p* moderado (≤ 0.1); en el segundo se ejecutó una asociación múltiple con los SNPs seleccionados, cuyos *valores p* resultantes fueron corregidos con el método de Bonferroni [52], antes de declarar la asociación.

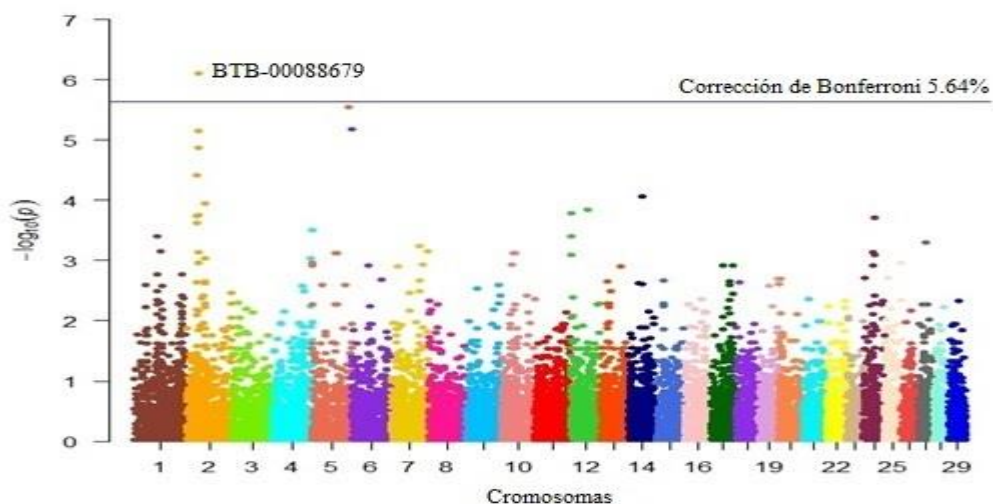
De la asociación con PLINK para cada fenotipo, y después de la aplicación de la corrección de Bonferroni (*valor p* corregido $\leq 2.25723\text{E-}06$), 8 SNPs resultaron significativos, en dos fenotipos. La tabla 4 muestra los resultados de la asociación.

Tabla 4. Lista de SNPs significativos, utilizando la herramienta PLINK 1.07

Fenotipo	Nombre SNP	dbSNP- Identificador	Chr	Posición (bp)	Gen más cercano	Distancia (bp)	-Log10(P)
Profundidad Corporal	BTB-00088679	rs43293152	2	34194876	GCA	Dentro	6.11
Profundidad de la Ubre	BTA-22344-no-rs	rs41606710	10	5598788	SFXN1	Dentro	5.90
	ARS-BFGL-BAC-12251	rs109798047	10	5767015	LOC781028	414,403	6.83
	ARS-BFGL-NGS-4488	rs110012582	10	6476252	ANKRD31	Dentro	5.90
	ARS-BFGL-NGS-18543	rs109007992	10	6521978	ANKRD31	Dentro	5.90
	ARS-BFGL-NGS-8362	rs108984322	10	7943252	LOC100296562	35,024	7.06
	ARS-BFGL-NGS-97105	rs110898362	10	8255705	LOC107131388	15,532	7.06
	ARS-BFGL-NGS-21248	rs109191831	10	8553040	WDR41	Dentro	7.06

En el rasgo de profundidad corporal, el SNP BTB-00088679 ubicado en el cromosoma 2, resultó significativo con un valor $p = 7.74E-07$. Su anotación fue verificada en NCBI (<http://www.ncbi.nlm.nih.gov/>). Este está dentro del gen GCA que codifica proteínas. Como podemos ver en la tabla 2, para el rasgo de profundidad de la ubre siete SNPs resultaron significativos, todos ubicados en el cromosoma 10 (BTA10), he identificados como BTA-22344-no-rs, ARS-BFGL-BAC-12251, ARS-BFGL-NGS-4488, ARS-BFGL-NGS-18543, ARS-BFGL-NGS-8362, ARS-BFGL-NGS-97105 y ARS-BFGL-NGS-21248 (Figura 13). No se detectaron SNPs significativos para los otros rasgos.

(a) Profundidad Corporal



(b) Profundidad de la Ubre

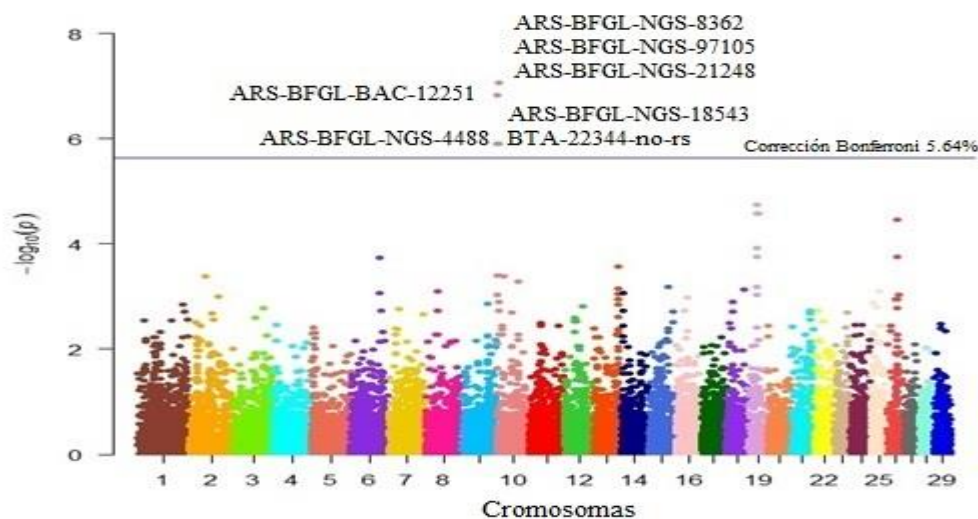


Figura 13. Gráfica de asociación de $-\log_{10}(P\text{-values})$ de asociaciones de nucleótidos simples con (a) Profundidad corporal (b) Profundidad de la Ubre. La línea horizontal representa el límite de la corrección de Bonferroni ($-\log p > 5.64$), que es equivalente a $P < 0.05$.

Todos SNPs están asociados con Locus de Características Cuantitativas (QTL), se realizó una búsqueda de QTL genómicos en la base de datos del genoma bovino (http://bovinegenome.org/bovineqtl_v2/findQTL.html). Los siete SNPs resultaron asociados a cinco QTLs relacionados con puntaje de marmolado, la producción de leche, rendimiento de proteínas, porcentaje de proteínas y rendimiento de grasa. La tabla 5 muestra el concentrado de los QTLs asociados.

Tabla 5. Lista de los QTLs que están asociados con los SNPs significativos

Número	Rasgo	Cromosoma	Inicio de Posición	Fin de Posición
180	Puntaje de marmolado	BTA10	2225595	14961855
267	Producción de Leche	BTA10	8292052	13636885
430	Producción de Leche	BTA10	2225595	14270735
431	Rendimiento de Grasa	BTA10	2225595	14270735
432	Rendimiento de Proteínas	BTA10	2225595	14270735
433	Porcentaje de Proteínas	BTA10	2225595	14270735

Para verificar la consistencia del resultado, se realizó de forma alternativa el análisis de asociación con el software R 3.2.3, implementando el siguiente procedimiento: 1) se ejecutó una regresión lineal simple a cada SNP, seleccionando todos los SNPs que tuvieran un $valor\ p \leq 0.1$; 2) se ejecutó una regresión lineal múltiple utilizando el conjunto de SNPs seleccionado en el paso anterior, y se declaró como significativos a todos aquellos con $-\log_{10}(valor\ p) > 5.64$, después de una corrección de Bonferroni. En este análisis resultaron significativos los mismos SNPs obtenidos significativos con la herramienta PLINK 1.07.

Capítulo IV

Discusión

Como resultado del análisis de la asociación de un conjunto de SNPs muestreados e imputados con 16 caracteres de conformación en una población de vacas Holstein mexicanas, se encontraron 8 SNPs significativos: uno asociado a la profundidad corporal y siete asociados a la profundidad de la ubre. Se utilizó la imputación de 6,912 a 46,523 SNPs como una estrategia de optimización de bajo costo, y aunque no se evalúa la exactitud de imputación debido a la muestra de baja densidad de datos (6,912 SNPs), los resultados indicaron buena asociación dado que se han reportado exactitud de imputación muy significativa, utilizando la misma herramienta de imputación (FindHap). Como ejemplo: Boichard [52] y Pimentel [53], reportaron una exactitud del 96.97% en Brown-Swiss y 97% en varias razas, cuando imputaron de 6k a 50k, respectivamente.

El SNP BTB-00088679 en el cromosoma 2 se encontró asociado a la profundidad corporal. Este SNP se encuentra en el gen GCA que codifica la proteína Grancalcin. Esta es una proteína de unión al calcio que desempeña un papel en la adhesión de los neutrófilos a la fibronectina, y puede desempeñar un papel en la formación de adherencias focales [UniProtKB-P28676: <http://www.uniprot.org/uniprot/P28676>]. No se encontró ningún trabajo previo de asociación entre el gen GCA con el rasgo de profundidad corporal, pero en un estudio de [54] reportó una correlación moderada de rendimiento de leche, rendimiento de grasa y rendimiento de proteína con estatura y rasgos de la profundidad del cuerpo. Específicamente, informaron que las vacas con mayor profundidad, más anchas y altas están genéticamente predispuestas a una mayor producción de leche.

En el caso del rasgo de profundidad de la ubre, resultaron asociados 7 SNPs en el cromosoma 10; donde cuatro SNPs se encuentran dentro de genes, mientras que el resto se encuentran a una distancia media de 155kb cerca de los genes anotados (véase la tabla 4). La función de estos genes varía: el gen SFXN1 codifica la proteína Sideroflexin-1 cuya evidencia experimental indica que puede estar implicado en el transporte de un componente requerido para la utilización del hierro dentro y fuera de las mitocondrias. El gen LOC781028, es un pseudogen de cadena ligera de clatrina. El gen ANKRD31 codifica la proteína putativa 31, que contiene el dominio de repetición de ankyrina. El gen LOC100296562 codifica la proteína 58 que contiene un dominio de bobina helicoidal. El gen LOC107131388 es un gen de codificación de una proteína tipo fosfodiesterasa 8B tipo 3, 5'-cíclico de alta intensidad de cAMP y sensible a IBMX. Y el gen WDR41 es una proteína que contiene la repetición de WD 41. Todos los genes caen

dentro de 5 QTLs asociados a Puntaje de Marmolado, Rendimiento de Leche, Rendimiento de Grasa, Rendimiento de Proteína y Porcentaje de Proteína.

Por último, el haber encontrado genes asociados a los rasgos de conformación puede ser un indicador de la presión de selección que ha tenido la raza Holstein para adaptarse a las condiciones climáticas extremas de las zonas del noroeste de México. Este resultado apoya la posibilidad de utilizar la talla y la profundidad corporal como rasgos que pueden indicar el mérito genético de las vacas para la producción de leche, grasa y proteínas.

Capítulo V

Conclusiones

En este estudio se presenta un análisis de la asociación de un conjunto de SNPs muestreados e imputados con 16 caracteres de conformación en una población de vacas Holstein de una zona desértica del noroeste de México. La imputación de 6k a 50k SNPs se realizó como estrategia de optimización de bajo costo. Los resultados muestran 8 SNPs asociados a dos rasgos de conformación. El rasgo de profundidad de la ubre resultó con siete SNPs asociados en el cromosoma 10, relacionados con Puntaje de Marmoleo, Rendimiento de Leche, Rendimiento de Grasa, Rendimiento de Proteína y Porcentaje de Proteína. El rasgo de profundidad corporal resultó con un SNP asociado en el cromosoma 2, pero no se encontró relación con QTL.

Finalmente, se observa que debido a la presión de selección que ha tenido la raza Holstein para adaptarse a las condiciones climáticas extremas de las zonas del noroeste de Mexico, puede ser la causa de haber encontrado genes asociados a los rasgos de conformación. Este resultado puede apoyar la posibilidad de usar la talla y la profundidad corporal como rasgos que pueden indicar el mérito genético de las vacas para la producción de leche, grasa y proteínas.

Referencias

- [1] P. C. D. T. Madigan Michael, M. Martinko John, V. Dunlap Paul, *Biología de los microorganismos*, 12a Edición. Madrid.
- [2] M. C. Andrabi SMH, "Selection of Dairy Cow Bulls for Artificial Insemination.," *Int. J. Agric. Biol.* 9 4, vol. 9, p. 4, 2007.
- [3] R. J. Syrstad O, "Prospects and strategies for genetic improvement of the dairy potential of tropical cattle by selection," *Trop Anim Heal. Prod* 30, pp. 257–268., 1998.
- [4] S. Newman, "Genetics Research at Fort Keogh: A Historical Perspective. USDA." 2008.
- [5] SP. Greiner, "Understanding Expected Progeny Differences (EPDs)," *Virginia Coop. Ext.*, vol. Tech. pp. , p. 2009, 2009.
- [6] G. K. Robinson, "That BLUP Is a Good Thing : The Estimation of Random Effects," vol. 6, no. 1, pp. 15–32, 2007.
- [7] C. J. Wilcox, D. W. Webb, and M. A. Delorenza, "Genetic Improvement of Dairy Cattle 1," 2003.
- [8] C. Smith, "Improvement of metric traits through specific genetic loci.," *Anim. Prod.*, vol. 9, p. 1967, 1967.
- [9] B. J. Soller M, "Genetic polymorphism in variety identification and genetic improvement," *Theor. Appl. Genet.*, vol. 9, p. 67, 1983.

- [10] A. Shrimptom, AE; Robertson, “The isolation of polygenic factors controlling bristle score in *Drosophila Melanoganster*,” *Genetics*, vol. 148, p. 14, 1988.
- [11] B. Hayes and M. E. Goddard, “The distribution of the effects of genes affecting quantitative traits in livestock,” vol. 33, pp. 209–229, 2001.
- [12] S. Sanna, A. U. Jackson, R. Nagaraja, C. J. Willer, L. L. Bonnycastle, H. Shen, N. Timpson, G. Lettre, G. Usala, P. S. Chines, H. M. Stringham, M. Dei, S. Lai, G. Albai, L. Crisponi, S. Naitza, K. F. Doheny, W. Elizabeth, Y. Ben-shlomo, S. Ebrahim, D. A. Lawlor, N. Richard, J. N. Hirschhorn, A. R. Shuldiner, D. Schlessinger, S. Francis, G. D. Smith, E. Boerwinkle, A. Cao, M. Boehnke, G. R. Abecasis, and K. L. Mohlke, “Common variants in the GDF5-UQCC region are associated with variation in human height,” vol. 40, no. 2, pp. 198–203, 2010.
- [13] PM. VanRaden, CP. Van tassell, GR. Wiggans, TS. Sonstegard, RD. Schnabel, JF. Taylor, FS. Schenkel, “Invited review : Reliability of genomic predictions for North American Holstein bulls,” *J. Dairy Sci.*, vol. 92, no. 1, pp. 16–24, 2009.
- [14] B. J. Hayes and M. E. Goddard, “Technical note: Prediction of breeding values using marker-derived relationship matrices,” pp. 2089–2092, 2014.
- [15] G. M. Hayes BJ, Bowman PJ, Chamberlain AJ, “Genomic selection in dairy cattle: Progress and challenges,” no. iii, p. 3168, 2009.
- [16] T. G. Seidenspinner T, Bennewitz J, Reinhardt F, “Need for sharp phenotypes in QTL detection for calving traits in dairy cattle,” *J Anim Breed Genet*, vol. 126, pp. 455–462, 2009.

- [17] F. A. Abdel-Azim G, “Superiority of QTL - Assisted Selection in Dairy Cattle Breeding Schemes,” *J Dairy Sci* 85 1869-1880., p. 2002, 2002.
- [18] D. W. Heyen, J. I. Weller, M. Ron, M. Band, J. E. Beever, E. Feldmesser, E. Feldmesser, Y. Da, G. R. Wiggans, and H. A. A. Lewin, “A genome scan for QTL influencing milk production and health traits in dairy cattle,” pp. 165–175, 1999.
- [19] M. D. Elsik CG, Tellam RL, Worley KC, Gibbs RA, “The genome sequence of taurine cattle: a window to ruminant biology and evolution,” *Science* (80-.), vol. 324, no. 5926, pp. 522–528, 2010.
- [20] CP. Van Tassell, TP. Smith, LK, Matukumalli, JF, Taylor, RD. Schnabel, “SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries,” *Nat Methods*, vol. 5, pp. 247–252, 2008.
- [21] G. M. Meuwissen THE, Hayes BJ, “Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps,” *Genetics*, vol. 157, p. 11, 2001.
- [22] S. L. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, “A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms,” *Nature*, vol. 409, pp. 928–933, 2001.
- [23] N. DA Crawford DC, “Definition and clinical importance of haplotypes,” *Annu Rev Med*, vol. 56, pp. 303–320, 2005.
- [24] B. J. Appleby N, Edwards D, “New technologies for ultra-high throughput genotyping in plants Methods Mol Biol 513: 19-39.,” *Methods Mol Biol*, vol. 513, pp. 19–39, 2009.

- [25] D. AM, “High throughput genotyping technologies Brief Funct Genomic Proteomic 1: 139-150.” *Br. Funct Genomic Proteomic*, vol. 1, pp. 139–150, 2002.
- [26] B. MD, “High throughput genotyping technologies for pharmacogenomics,” *Am J Pharmacogenomics*, vol. 1, pp. 295–302, 2001.
- [27] K. FJ, Steemers, Gunderson, “Whole genome genotyping technologies on the BeadArray platform,” *Biotechnol J*, vol. 2, pp. 41–49, 2007.
- [28] H. S. Rapley R, “Molecular Analysis and Genome Discovery,” *Chichester, Ed. John Wiley Sons Ltd.*, p. 2004, 2004.
- [29] E. K. Gibbs RA, Taylor JF, Van Tassell CP, Barendse W, “Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds,” *Science (80-.)*, vol. 324, pp. 528–532, 2009.
- [30] I. Inc, “BovineSNP50 Genotyping BeadChip,” no. Table 3, 2010.
- [31] B. Rosner, *Fundamentals of biostatistics*, Seventh ed. 2011.
- [32] J. E. Stajich, “The Bioperl Toolkit: Perl Modules for the Life Sciences,” *Genome Res.*, vol. 12, pp. 1611–1618, 2002.
- [33] T. R Development Core, “R: A Language and Environment for Statistical Computing. Vienna, Austria : the R Foundation for Statistical Computing. ISBN: 3-900051-07-0.” *PhD Propos.*, vol. 1, p. 900051, 2011.
- [34] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham, “PLINK: A tool set

- for whole-genome association and population-based linkage analyses,” *Am. J. Hum. Genet.*, vol. 81, no. 3, pp. 559–575, 2007.
- [35] M. Kanehisa, S. Goto, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, “Data, information, knowledge and principle: Back to metabolism in KEGG,” *Nucleic Acids Res.*, vol. 42, no. D1, pp. 199–205, 2014.
- [36] H. Colella, S., Yau, C., Taylor, J. M., Mirza, G., Butler, “QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data,” *Nucleic Acids Res*, vol. 35.
- [37] Seringhaus, Michael., Mark Gerstein, “Gene Ontology (GO),” *Investig. Sci.*, vol. 390, pp. 73–81, 2009.
- [38] C. E. Gang, Zheng., Yaning, Yang., Xiaofeng, Zhu., Robert, “Analysis of Genetic Association Studies,” *Springer*, p. 2011, 2011.
- [39] M. Durbin, *All About High-Frequency Trading*. 2010.
- [40] Marchini J., Howie B, Myers S, McVean G, Donnelly P “ A new multipoint method for genome-wide association studies by imputation of genotypes,” *Nat Genet.*, vol.39, pp. 906-9713, 2011.
- [41] Marchini J., Howie B, “ Genotype imputation for genome-wide association studies ,” *Nat. Rev-Genet.*, vol.11, pp. 499-511, 2010.
- [42] PM. Vanraden, JRO. Connell, GR. Wiggans, KA. Weigel and Improvement A, “ Combining different marker densities in genomic evaluation,” pp. 1-4, 2010.

- [43] CANILEC (2009) Bovino y sus derivados. Mexico DF.
- [44] USDA (2011) Dairy: World Markets and Trade.
- [45] SIAP (2015) Leche de bovino, produccion, precio y valor. In: SAGARPA, editor. Servicio de informacion agroalimentaria y pesquera. Mexico DF..
- [46] Vargas E (2011) Pierde B.C. 10 mil cabezas de ganado lechero. Noticias Ensenadanet. Ensenada.
- [47] Garcia RG (2011) En crisis la industria lechera de BC. La Voz de la Frontera. Mexicali.
- [48] Avendaño-Reyes L, Álvarez-Valenzuela FD, Correa-Calderón A, Algándar-Sandoval A, Rodríguez-González E, Pérez-Velázquez R, Macías-Cruz U, Díaz-Molina R, Robinson PH and Fadel JG (2010) Comparison of three cooling management systems to reduce heat stress in lactating Holstein cows during hot and dry ambient conditions. *Livest Sci* 132:48–52. doi: 10.1016/j.livsci.2010.04.020.
- [49] Avendaño-Reyes L, Alvarez-Valenzuela FD, Correa-Calderón A, Saucedo-Quintero JS, Robinson PH and Fadel JG (2006) Effect of cooling Holstein cows during the dry period on postpartum performance under heat stress conditions. *Livest Sci* 105:198–206. doi: 10.1016/j.livsci.2006.06.009.
- [50] García E (2004) Modificaciones al Sistema de Clasificación Climática de Köppen (para adaptarlo a las condiciones de la República Mexicana). Instituto de Geografía, 6a ed. Universidad Nacional Autónoma de México, D.F. México.
- [51] Nakagawa S and Sciences P (2004) A farewell to Bonferroni: the problems of low

statistical power and publication bias. 15:1044–1045. doi: 10.1093/beheco/arh107

- [52] Boichard D, Chung H, Dasonneville R, David X, Eggen A, Fritz S, Gietzen KJ, Hayes BJ, Lawley CT, Sonstegard TS et al. (2012) Design of a bovine low-density snp array optimized for imputation. PLoS One 7:1–10. doi: 10.1371/journal.pone.0034130.
- [53] Pimentel ECG, Edel C, Emmerling R and Götz K-U (2015) How imputation errors bias genomic predictions. J Dairy Sci 98:4131–8. doi: 10.3168/jds.2014-9170.
- [54] Haas Y., Janss L.L.G., Kadarmideen H.N. (2007): Genetic and phenotypic parameters for conformation and yield traits in three Swiss dairy cattle breeds. Journal of Animal Breeding and Genetics, 124, 12–19