



UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA

**INSTITUTO DE INGENIERÍA
MAESTRÍA Y DOCTORADO EN CIENCIAS E INGENIERÍA**

***“ESTUDIO DE VARIACIÓN ALÉLICA ENTRE LOS
MARCADORES DE PATERNIDAD DE LA RAZA HEREFORD
MEXICANA Y LOS DEL HAPMAP BOVINO”***

T E S I S

QUE PARA OBTENER EL GRADO DE:

DOCTOR EN CIENCIAS

PRESENTA:

GUADALUPE NELSON AGUILAR PALMA

DIRECTOR DE TESIS:

DR. RAFAEL VILLA ANGULO

CODIRECTOR:

DR. JOEL DOMÍNGUEZ VIVEROS

Mexicali, B. C.

Junio del 2018

DEDICATORIA

El esfuerzo y sacrificios requeridos para completar este propósito de mi vida se los dedico a mi esposa e hijas, a mis compañer@s de trabajo y estudio; pero sobre todo a mis alumn@s.

Predicar con las palabras y el ejemplo de forma coherente, resultan para mí las virtudes excelsas de un Profesional, de un verdadero Maestro y de un Líder.

Por lo tanto, admitiendo anticipadamente como un error de mi parte cualesquier omisión inoportuna, enlisto algunas de las personas quienes en la práctica me demostraron su dominio de la coherencia, a ell@s también les dedico mi trabajo; Eddy Nelson, John Hazelwood, Juan López, Alfredo Villagrán, Juan Ángel Ortega, Felipe Rodríguez, Javier Martínez, Rafael Villa, Carlos Villa, Rogelio Ramos, Néstor Santillán,,, y much@s más.

A tod@s l@s que se tomen el tiempo necesario para leer y entender el trabajo aquí documentado, sean y siéntanse merecedores de ser incluidos en esta dedicatoria, gracias por su curiosidad y deseo de entendimiento.

AGRADECIMIENTOS

Acatando como regla básica las siguientes definiciones, que de la palabra "Institución" presenta la Real Academia de la Lengua Española; a) Organismo que desempeña una función de interés público, especialmente benéfico o docente y b) Alguien con la facultad, de ser y tener en cualquier agrupación humana el prestigio debido a la antigüedad o a poseer todos los caracteres representativos de aquella.

Doy gracias a todas las Instituciones que me ayudaron y permitieron concluir el Proyecto de Investigación aquí descrito:

A Dios

A mi familia

A mis amig@s

A mis Mentores y Tutores

A mi Universidad Autónoma de Chihuahua

A mi Universidad Autónoma de Baja California

A mi Laboratorio de Bioinformática y Biofotónica

A mi Instituto de Ingeniería en la UABC y el MYDCI

Al PRODEP

Al CONACYT por el PNPC

Al CONARGEN y la SAGARPA

Al CA En Consolidación UACH-CA-07

RESUMEN

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA

INSTITUTO DE INGENIERÍA

MAESTRÍA Y DOCTORADO EN CIENCIAS E INGENIERÍA

LABORATORIO DE BIOFOTÓNICA Y BIOINFORMÁTICA

Nuevas más precisas y menos costosas técnicas de extracción y genotipado han sido desarrolladas en los últimos años permitiendo el incremento en los usos y aplicaciones de los Polimorfismos de Nucleótido Simple (SNP) en múltiples áreas, sustituyendo en algunos casos Micro satélites o Repeticiones de Tándem Cortas (STR), sin disminuir la confianza o precisión de los resultados de las pruebas. La prueba de paternidad usando marcadores genéticos ha probado ser un excelente y veraz medio para demostrar la paternidad. El ganado para producción de carne ha estado en las manos de la humanidad por miles de años, en ese tiempo ha soportado el método de apareamiento selectivo bajo los criterios que satisfacen al hombre. Sin embargo las duras condiciones actuales como el calentamiento global, cambio climático, cambios culturales globalizados, sobre población y otros riesgos relativos están dificultando la situación para los productores de carne, incrementando los requisitos de calidad de los mercados internacionales (alimentos orgánicos y libres de hormonas), costos extras, incertidumbre en los recursos naturales que impide la planeación efectiva aumentando los riesgos de pérdida. La selección Genómica, las pruebas de paternidad y trazabilidad (origen) pronto serán las técnicas regulares para administrar los ranchos de ganado de carne y los negocios dependen de eso. Para apoyar a los granjeros productores a tener éxito, los SNPchips son la herramienta correcta por su confiabilidad, bajo costo y capacidad de genotipado. Simultáneamente los científicos del campo de la bioinformática desarrollan algoritmos para tomar decisiones de forma simple y asertiva. ISAG e ICAR certificaron un panel de 206 marcadores SNP's para pruebas de paternidad de bovinos a nivel internacional en cualesquier raza. La asociación de criadores Hereford de México colaboró para genotipar 1317 animales para la prueba de paternidad. La información fue analizada para verificar y certificar registros de pedigrí, pero además fue aprovechada para responder si era posible utilizar menos marcadores en la prueba, reduciendo costos pero manteniendo el nivel de confianza. Dos fuentes de información fueron estudiadas México y un panel internacional del HapMap Consortium. Filtros de control de calidad se desplegaron para asegurar completés, consistencia y coherencia. Se desarrolló la reconstrucción de los haplotipos para el panel Mexicano y se verificó usando los 2218 tríos obtenidos de la muestra original de ganado. La variabilidad alélica fue evaluada en ambos paneles usando MAF, H_o , H_e , HW y LD. El poder de exclusión acumulado total de la prueba de paternidad (Q) se calculó para un conjunto de 80 marcadores SNP's, en tres posibles escenarios; ambos padres, el padre, la madre. Los resultados demuestran que 63, 74 y 72 de los marcadores seleccionados pueden aportar valores de Q equivalentes en confiabilidad al conjunto de 206 del panel original.

ABSTRACT

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA
INSTITUTO DE INGENIERÍA
MAESTRÍA Y DOCTORADO EN CIENCIAS E INGENIERÍA
LABORATORIO DE BIOFOTÓNICA Y BIOINFORMÁTICA

New more precise and less expensive DNA extraction and genotyping techniques have been developed in the last years allowing to increase the uses and applications of Single Nucleotide Polymorphisms (SNP) in multiple areas, substituting in some cases Microsatellites or Short Tandem Repetitions (STRs), without diminishing confidence or precision at the test results. Paternity testing using genetic markers has proved to be an excellent and trustable way to demonstrate parenthood. Beef cattle has been at human hand for thousands of years, in that time it has endure the mating selection method to satisfy criteria defined by man. However actual harsh conditions like global warming, climate change, globalized cultural changes, over human population and other relative risks are struggling around beef producers, increasing quality requirements from international markets (organic and certified food free from hormones), costs, uncertainty in nature resources impeding better planning and increasing loss risks. Genome selection, paternity testing and identification will be soon the regular techniques to manage beef cattle ranches and the business depend on it. To allow farmers to succeed SNPchips are the right tool because their high reliability, low cost, genotyping capacities. Simultaneously bioinformatics field scientists developed algorithms to let decision making be simple and asertive . ISAG and ICAR certified a 206 SNP's panel for bovine parental testing in any breed. Hereford Mexican Breeders Asociation collaborate to genotype 1317 animals to be paternity tested. The information was used to certified pedigree records but also analyzed to find out if less SNP's can do the same test, reducing cost but keeping reliability. Two main information sources were used Mexico and International Hap Map Consortium panels. Quality control filters were used to assure completeness, consistency and coherency. Haplotype reconstruction was performed for the Mexico panel, and verified with 218 trios obtained from the original sample. Allelic variability in both panels was evaluated using MAF, Ho, He, HW and LD. The Q cumulative exclusion power was calculated for a set of 80 markers, selected and discriminated on Ho difference bases. To find out how many markers can do the test without diminishing reliability, as many as 80 variants for the three possible tests were made. Both parents, one parent, the other parent relatively. Results demonstrate that 63, 74, and 72 of the selected markers can get Q values equally reliable as the 206 set.

ÍNDICE

DEDICATORIA	i
AGRADECIMIENTOS	ii
RESUMEN	iii
ABSTRACT	iv
CAPÍTULO I. INTRODUCCIÓN	3
1.1 Situación actual de la ganadería de engorda en México	3
1.2 El mejoramiento Genético	8
1.2.1 Método tradicional de mejora genética de ganado	8
1.2.2 Método de mejora genética asistida por marcadores	11
1.2.3 Método de mejora genética por Selección Genómica	14
1.3 Parámetros de caracterización de la estructura genética	18
1.3.1 Análisis de Proporciones polimórficas.....	19
1.3.2 Análisis de Equilibrio de Hardy-Weinberg.....	20
1.3.3 Estudios de Heterosigocidad	21
1.3.4 Desequilibrio por ligamiento	23
1.3.5 Coeficiente de endogamia.....	24
1.3.6 Índice de fijación.....	25
1.3.7 Estudio del Índice global de fijación	25
1.3.8 Caracterización del desequilibrio por ligamento	26
1.3.9 Estudio de la tasa de mutaciones	29
1.3.10 Estimación de la población efectiva, estimación en retrospectiva.	30
1.3.11 Generación de mapas de haplotipos.....	32
1.4 La importancia de las pruebas de paternidad en ganado bovino	33
1.5 Pruebas de paternidad utilizando marcadores genéticos	33
CAPÍTULO II. PLANTEAMIENTO DEL PROBLEMA Y OBJETIVOS	35
2.1 Planteamiento del Problema	35
2.2 Objetivo general.....	35
2.3 Objetivos específicos.....	36

CAPÍTULO III. DESCRIPCIÓN DE LOS DATOS	37
3.1 Datos de ganado Hereford	37
3.1.1 Obtención de las muestras sanguíneas	37
3.1.2 Organización y depuración de los datos.....	38
3.2 Datos del consorcio HapMap bovino	39
CAPÍTULO IV. ANÁLISIS Y RESULTADOS	42
4.1 Estimaciones para la reconstrucción de Haplotipos Hereford México (PHASE).....	42
4.2 Estimadores de Variabilidad Alélica como indicadores de Variabilidad Genética (HAPLOVIEW)	46
4.3 Resultados para frecuencias MAF, H_e , H_o y HW	47
4.4 Resultados del Desequilibrio por Ligamiento	50
4.5 Selección de Marcadores significativos para Paternidad México.....	51
CAPÍTULO V. DISCUSIÓN DE RESULTADOS	52
CAPÍTULO VI. CONCLUSIONES Y TRABAJO FUTURO	54
BIBLIOGRAFÍA.....	55
ANEXOS	57

CAPÍTULO I. INTRODUCCIÓN

1.1 Situación actual de la ganadería de engorda en México

Históricamente la domesticación de especies de interés para el Hombre implicó una presión evolutiva paralela a la selección natural ya encontrada en el estado salvaje, la "selección artificial" o mejora de las especies siguiendo criterios y patrones de fenotipos *ad hoc* para satisfacer las preferencias del criador ha moldeado la genética de múltiples especies, algunas para servicio, otras para consumo e incluso ambos propósitos simultáneamente. El ganado bovino es parte del fenómeno desde hace aproximadamente 6000 años que inició el proceso de su domesticación. El *bos-primigenius* es el ancestro original de las vacas y toros actuales, unos de sus descendientes el *bos-primigenius-taurus* predominaron en la zona euroasiática hasta que fueron cazados por el hombre y llevados a la extinción. Actualmente las razas bovinas se clasifican generalmente en tres grupos; *Bos taurus* (ganado Africano Europeo) *Bos Indicus* (ganado Asiático) y las mezclas entre ambos. Estas últimas originadas por el hombre, desde que los medios de transporte humanos facilitaron la movilidad del ganado en todo el orbe mundial.

La producción de leche y carne son los usos más importantes del ganado bovino para el consumo humano en la historia reciente, además de la alimentación de otras especies co-dependientes del hombre como los perros, gatos, etcétera. La industria de la piel también se ha favorecido de las propiedades admirables de los bovinos, incluso en el área del entretenimiento existen tradiciones como "Las corridas de Toros" , el "Rodeo" y todas sus mezclas o variedades.

En esta Tesis el propósito central se focaliza en el uso del Ganado de Engorda Mexicano de la raza Hereford (*Bos Taurus*) para la producción de carne con estándares internacionales y su posible exportación a mercados extranjeros, o como alternativa para el consumo nacional en los segmentos del mercado que prefieren pagar un precio adicional por cortes de primera clase.

El **Gráfico 1.1.a** presenta el consumo de carne de bovino en volumen total, por países o regiones, sin embargo el consumo per cápita debe ser también considerado, sobre

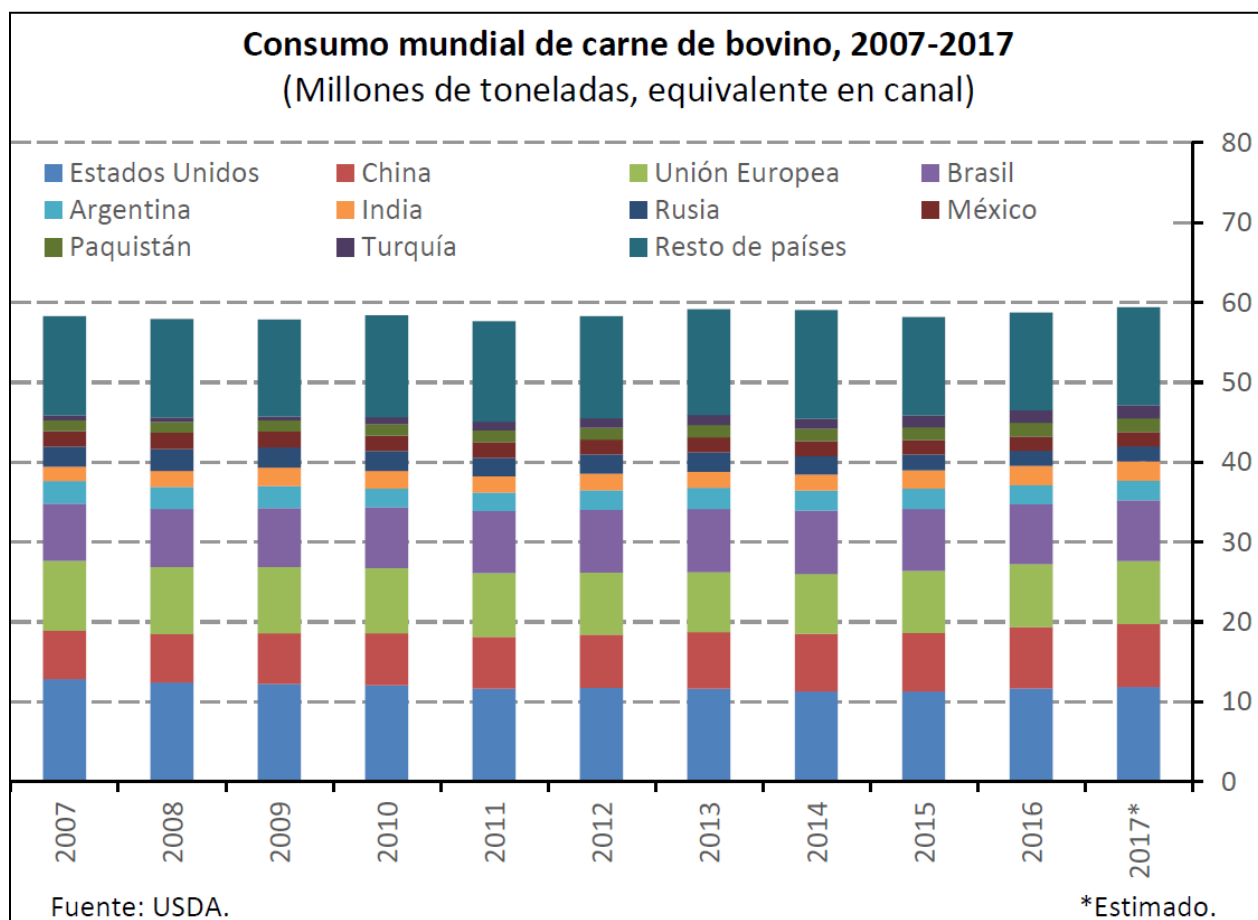


Gráfico 1.1.a. Consumo Mundial en toneladas por país o región en canal.

todo por la importancia de los cambios culturales globales y regionales, que últimamente han impactado a la baja el nivel de disposición de los consumidores de carnes rojas, en la cantidad a ingerir, pero también han afectado positivamente en un aumento en la disposición por pagar

mayores precios por la aseguranza de la calidad de los cortes a consumir. La calidad del corte, implica la trazabilidad (rastreadibilidad), desde su origen hasta el consumidor final.

Los productores y las asociaciones de criadores, enfrentan además de las condiciones adversas relacionadas con los mercados y la economía globalizada, el fenómeno del cambio climático, el cual amenaza el flujo de los recursos naturales necesarios para la sustentabilidad de su industria. Mantener los niveles de producción mientras simultáneamente se elevan los estándares de calidad y los requisitos para cumplirlos, logrando salir a los mercados confiando en que su producto no puede ser rechazado o verse involucrado en situaciones indeseables de afectación a la salud de los consumidores, representa un reto mayúsculo y requiere de todas las herramientas científicas posibles y disponibles. El **Gráfico 1.1.b** muestra la producción a nivel

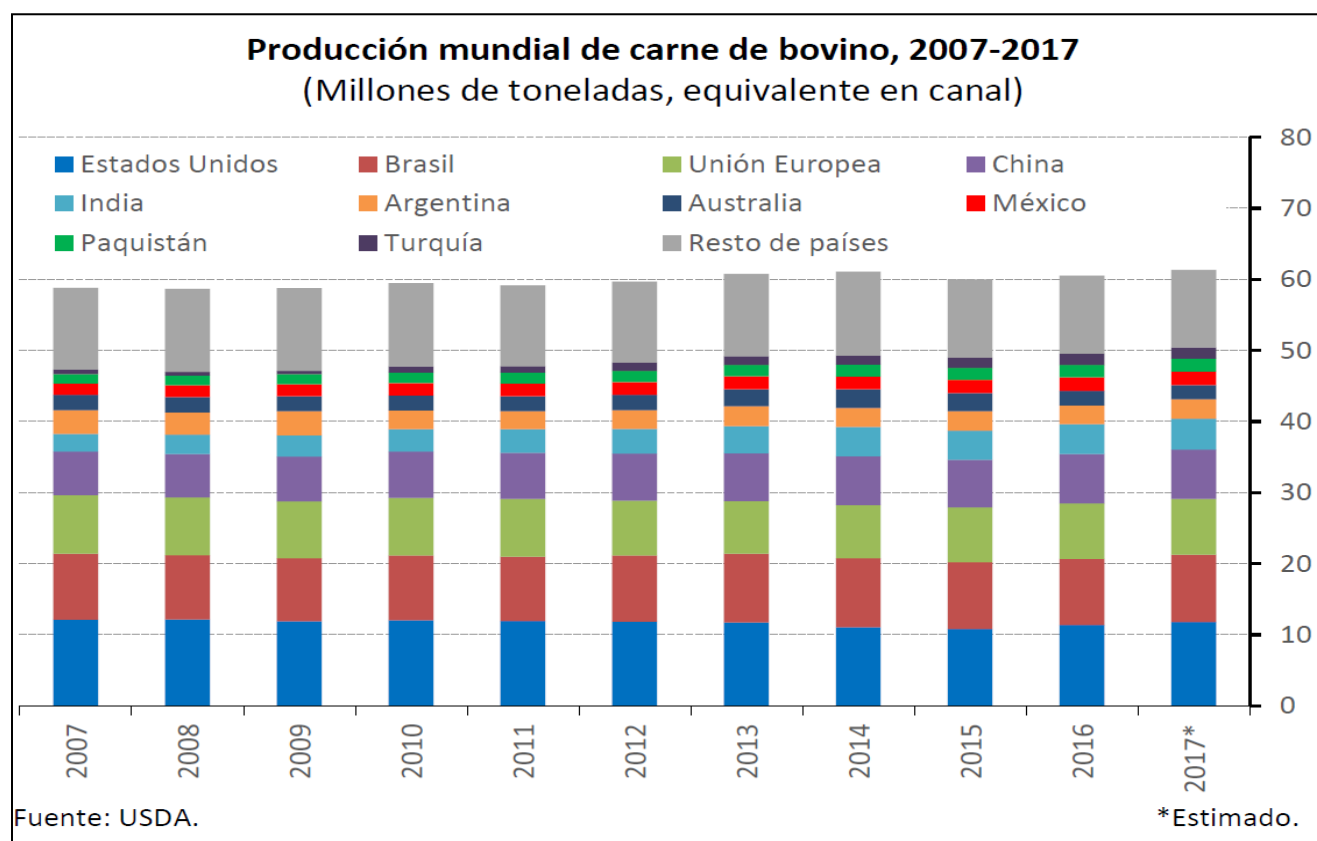


Gráfico 1.1.b. Producción Mundial en toneladas por país o región, en canal.

mundial, México aun sin contar con la extensión territorial de los principales países productores de carne, aparece en el panorama mundial en el grupo de los más relevantes.

En la última década, la producción nacional de carne de bovino ha presentado un continuo crecimiento, con excepción de 2013. Así, entre 2007 y 2016, creció a una tasa media anual de 1.6 por ciento, para ubicarse en 1.88 millones de toneladas de carne en canal. [1]

Para 2017, de acuerdo con el Departamento de Agricultura de Estados Unidos (USDA), la producción de carne de res se ubicó en un máximo histórico de 1.91 millones de toneladas, es decir, registró un crecimiento anual de 1.6 por ciento según lo muestra el **Gráfico 1.1.c**.

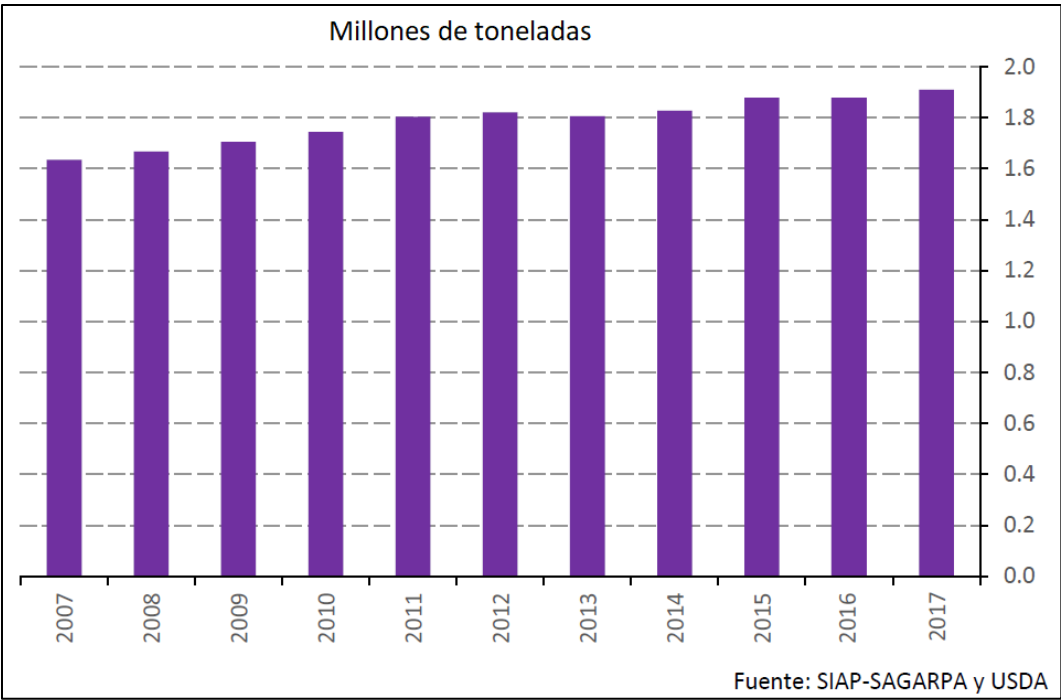


Gráfico 1.1.c Producción de carne de bovino en México, 2007-2017

En 2016, el 62.5 por ciento de la producción nacional de carne de bovino se concentró en diez entidades: Veracruz (13.4 por ciento), Jalisco (11.5 por ciento), Chiapas (6.1 por ciento), San Luis Potosí (5.5 por ciento), Sinaloa (4.9 por ciento), Baja California (4.8 por ciento),

Durango (4.5 por ciento), Michoacán (4.1 por ciento), Chihuahua (4.0 por ciento) y Sonora (3.7 por ciento). El **Gráfico 1.1.d** presenta la relación de producción por estados.

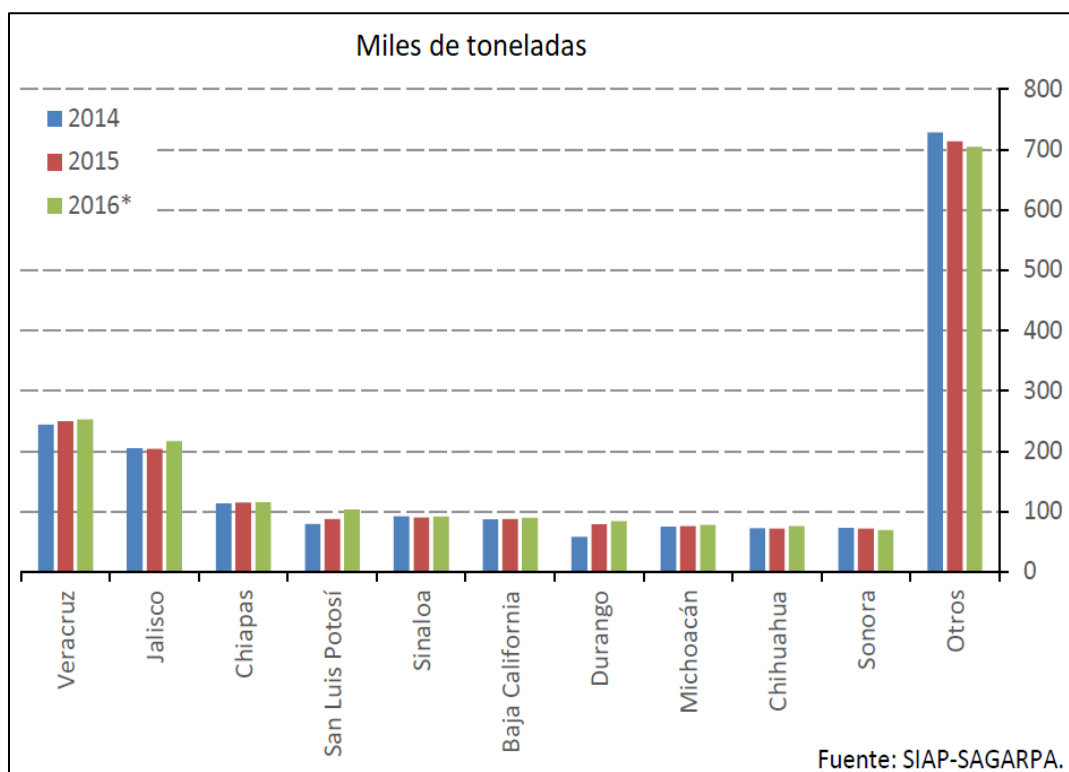


Gráfico 1.1.d. Principales estados productores de carne de bovino, 2014-2016.

La amplia disponibilidad de granos forrajeros, así como de pastizales, combinado con los elevados precios de la carne de bovino en el país, incentivan a que los ganaderos mantengan por más tiempo las reses en engorda. Así, por tercer año consecutivo, el incremento en la producción de carne sería resultado del mayor peso al sacrificio, que compensaría la disminución del número de cabezas sacrificadas.

Se estima que en 2018 continúe la recuperación del hato ganadero nacional, y que igual que en años previos, contribuya a lo anterior la existencia de incentivos financieros y los programas de mejoramiento genético, a través de apoyos gubernamentales. [2]

1.2 El mejoramiento Genético

1.2.1 Método tradicional de mejora genética de ganado

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. Esto se ha venido realizando de forma sistemática haciendo cruzamiento de lo mejor con lo mejor **Figura 1.2.1.a.**

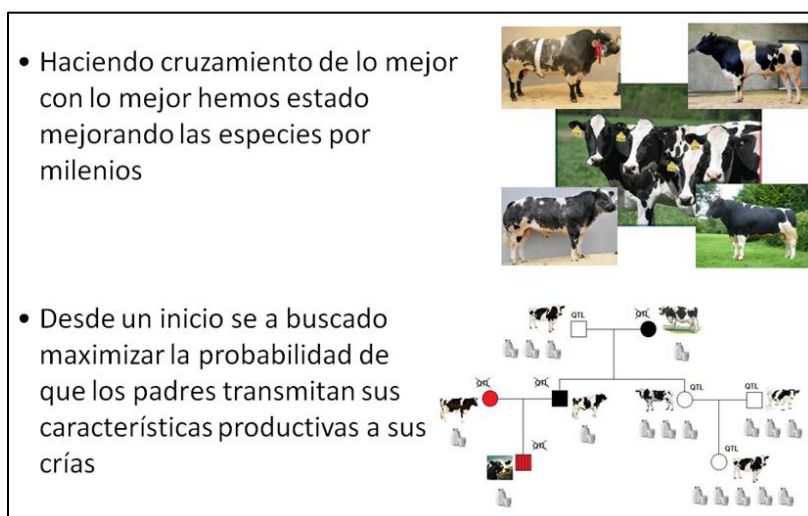


Figura 1.2.1.a. Haciendo cruzamiento de los mejores sementales con las mejores vacas se ha logrado que los padres trasmitan sus características productivas a sus crías.

La recopilación organizada de pedigrís y rendimiento productivo del ganado, a través de la formación de registros de manadas y asociaciones de razas, se inició hace más de 100 años. 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 [3,4]. 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.

Durante los últimos 50 años, los programas de mejoramiento genético en ganado lechero han sido muy exitosos en la mejora de la productividad [5]. El modelo tradicional de crianza selectiva, basada en principios sofisticados de selección cuantitativa, ha rendido frutos satisfactorios, aún cuando no sabemos con exactitud cuáles son los genes que han incrementado o decrementado su frecuencia en las poblaciones de ganado. Generalmente se describe el método de selección como una “caja negra” debido a que no se sabe exactamente cuales genes han sido afectados. La implementación de programas de pruebas efectivas de progenie y de sementales élitos bien validados se ha convertido en el método que mejores resultados arroja en programas de mejora genética. Varios avances han hecho que éste método básico haya mejorado de una forma sustancial; la adopción de comparaciones con animales contemporáneos y posteriormente del modelo animal para realizar la evaluación genética son ejemplo de ello. En décadas recientes, los productores y genetistas han reconocido la importancia de la crianza en la mejora de la salud, reproducción y longevidad, por lo que rasgos de salud, reproducción y vida productiva se han convertido en una parte muy importante para la selección de ganado. En un programa de selección tradicional, el valor genético de crianza de un animal (EBV, “Estimated Breeding Value”, por sus siglas en inglés) es estimado para cada rasgo de importancia económica. Métodos estadísticos avanzados evalúan de forma simultánea a sementales y vacas, integrando una gran gama de información fenotípica y de pedigrí, tal como lo muestra la **Figura 1.2.1.b** y son:

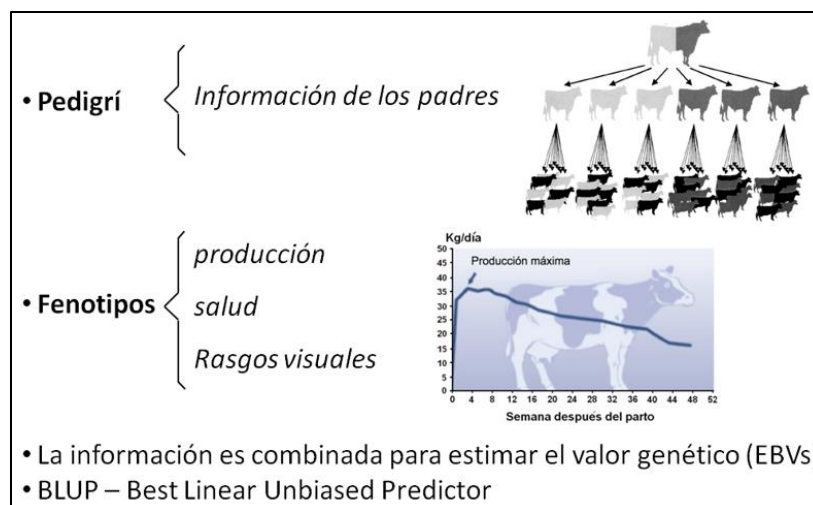


Figura 1.2.1.b. El método tradicional de selección está basado en el mérito genético, calculado de información fenotípica y de pedigrí.

datos sobre los padres y/o hijos y hermanos, distribución de las hijas en el hato, longitudes de lactancia, número de records de lactancia por hija y diferencia genéticas entre miembros del hato, entre otros. Estos métodos son ejecutados en computadoras de alta velocidad para estimar de forma precisa el mérito genético de un animal (Ej. BLUP [6] es el método más utilizado actualmente para evaluar sementales). En este proceso deben ser identificadas todas las posibles fuentes de variación ambiental, y cada animal debe ser evaluado en relación a otros animales criados bajo un manejo igual o similar. De esta forma, el mérito genético para un animal específico es calculado como la desviación de su valor genético de crianza estimado con respecto a la media en su hato, o en hatos con condiciones de manejo similar [7].

El EBV representa la contribución genética en el record del animal. Por ejemplo; un EBV = +2000 lbs. para una vaquilla, nos indica que bajo las mismas condiciones de manejo, cuando esta vaquilla sea madura, se espera que produzca 2000 lbs. más que una vaquilla con un EBV = 0. La mitad del EBV de un animal es utilizado para predecir su habilidad de transmitir su valor

genético a sus crías (PTA, “Predicted Transmmiting Ability”, por sus siglas en ingles). Por ejemplo; un PTA = +1000 lbs. para la vaquilla del ejemplo anterior, nos indica que bajo las mismas condiciones de manejo, se espera que una hija madura de esta vaca, produzca 1000 lbs. más que una vaca con PTA = 0. En el caso de animales jóvenes sin record, se obtiene el promedio de la habilidad transmisora de sus padres (PA, “Parent Average”, por sus siglas en ingles). Esto es:

$$PA = \frac{1}{2} PTA \text{ del padre} + \frac{1}{2} PTA \text{ de la madre}$$

La certeza de la estimación del mérito genético se calcula una vez teniendo el rendimiento del animal, o sus crías, según el caso. Se utiliza el record del animal y su información de pedigrí. Con esto se mide la precisión de PTA, que es el cuadrado del grado de correlación de los valores de rendimiento real y los estimados de PTA. Se puede estimar el valor de certeza para animales jóvenes sin record utilizando la certeza de sus padres:

$$Certeza = (certeza \text{ padre} + certeza \text{ madre})/4$$

Este es un valor estimado que utiliza la información se los padres pero no garantiza que las crías hereden los rasgos genéticos deseados de sus padres.

1.2.2 Método de mejora genética asistida por 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, aún 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. Se sabe que un buen balance entre el manejo adecuado de la crianza y la selección minuciosa de animales con características fenotípicas mayores en los

rasgos que deseamos, ha conducido a lograr progreso, pero de igual forma se sabe que muchos de los rasgos observables (fenotipos) de fisiología del animal, de sus propiedades bioquímicas, de su morfología, de desarrollo, comportamiento y producción, son epigenéticos. Esto es, son regulados por la interacción de los genes con el medio ambiente. Lo que hace 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. Tal como se presenta en la **Figura 1.2.2.a.**

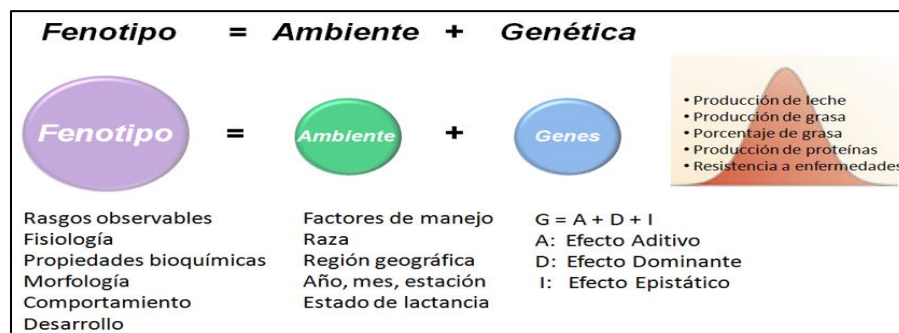


Figura 1.2.2.a. La mayor parte de los rasgos de importancia económica en el ganado bovino son epigenéticos. La mayor parte de los fenotipos son regulados por la interacción de los genes con el medio ambiente.

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 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 (ej. [8, 9]), la adopción de la selección asistida por marcadores (MAS, “Marker Assisted Selection”, por sus siglas en inglés) en la industria lechera había estado limitada, hasta hace muy poco tiempo. Existían diferentes 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 (ej. [10-13]). 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. En adición, la complejidad para el cálculo de los valores genéticos de cría incluyendo la información de los marcadores era una barrera más para la aplicación de la selección asistida por marcadores [14].

En el año 2001, investigadores de Holanda y Australia demostraron a través de simulaciones por computadora, que usando marcadores de ADN distribuidos a lo largo de todo el genoma permitía capturar de forma simultánea los efectos de todos los genes que afectan y regulan las características de importancia económica, aumentando con esto de forma substancial la precisión de la selección asistida por marcadores [15]. Este trabajo inicial sobre los beneficios de utilizar información de todo el genoma sirvió para establecer las necesidades tanto de información como de tecnología para lograr los avances demostrados en las simulaciones. Estos requerimientos incluían: 1) contar con poblaciones de animales con mediciones fenotípicas precisas y sus muestras de ADN genómico, 2) marcadores polimórficos de ADN posicionados en todo el genoma, 3) tecnología de laboratorio para realizar análisis de

una alta densidad de marcadores y 3) algoritmos y métodos para el análisis de datos genotípicos y fenotípicos.

1.2.3 Método de mejora genética por Selección Genómica

La revolución de la Selección Genómica empezó con dos desarrollos claves. El primero fue la reciente secuenciación del genoma bovino [16], lo cual habilitó el descubrimiento de miles de marcadores de ADN, en forma de polimorfismo de nucleótido simple, SNPs [17]. 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. La **Figura 1.2.3.a** describe la forma más común de los SNP's.

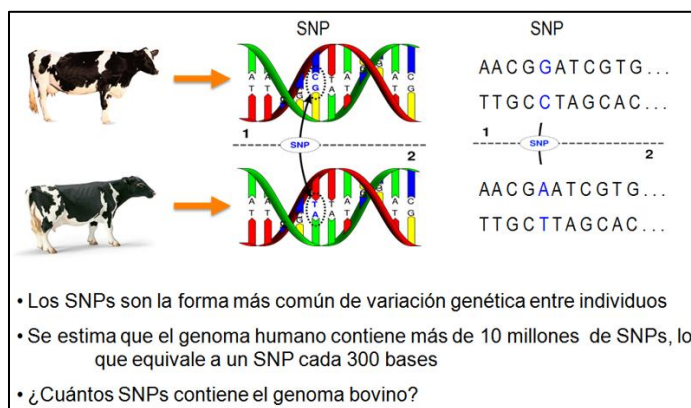


Figura 1.2.3.a. La forma más común de polimorfismos en los genomas son los SNP. Un SNP es una posición en el genoma, en la cual, al menos el 5% de los individuos de una población tienen un alelo distinto.

Se Realizó un estudio de descubrimiento de polimorfismos de nucleótido simple en la especie bovina, obteniendo como resultado más de 6 millones de SNPs putativos. Inicialmente de validaron 54,609 SNPs y se realizaron los primeros estudios de caracterización de la estructura genética de la especie [18, 19]. Se muestrearon 565 animales de 19 razas distintas, ubicadas alrededor de mundo, **Figura 1.2.3.b**.

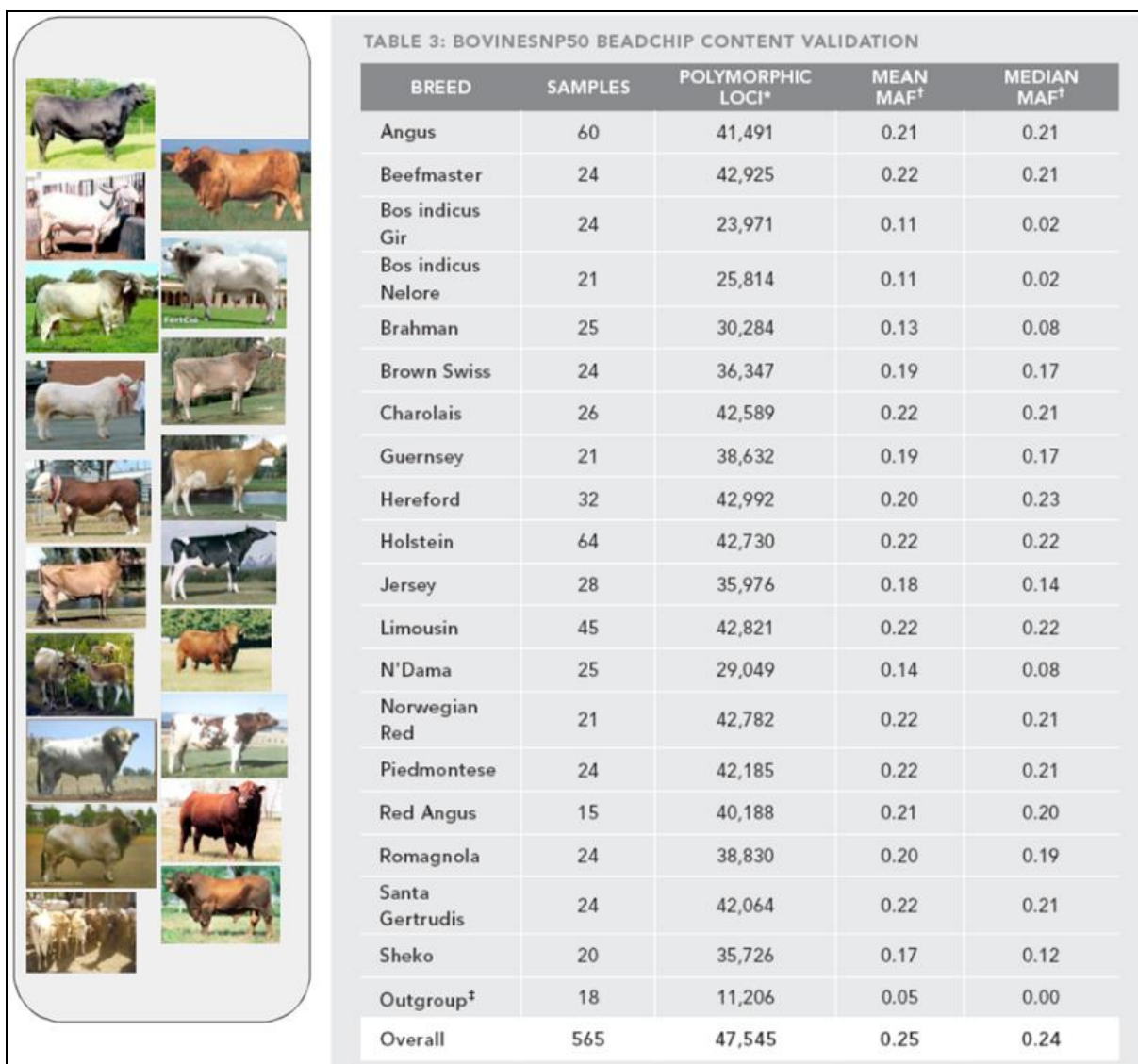


Figura 1.2.3.b. Se tomaron muestras de 565 animales pertenecientes a 21 razas de ganado bovino y se realizaron los primeros estudios de estructura genética de genoma completo (imagen tomada de Illumina.com).

Este estudio de validación de SNP's detonó el desarrollo de la tecnología de genotipado de salida masiva para hacer captura de polimorfismos en poblaciones grandes de animales. El SNPchip50, en solo unos meses se convirtió en el estándar para realizar estudios genéticos de las razas de ganado pertenecientes a la evolución *Bos Taurus*, dando inicio a la revolución de la *Selección Genómica*. Este nuevo desarrollo tecnológico **Figura 1.2.3.c** y **Figura 1.2.3.d**, está

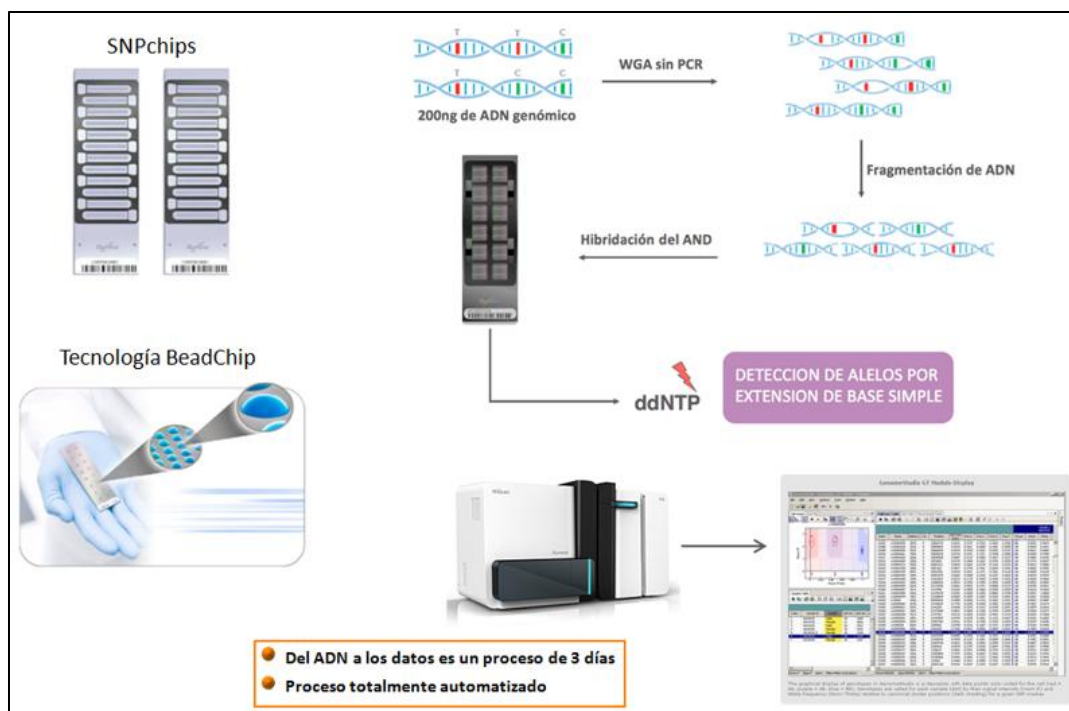


Figura 1.2.3.c. Proceso de genotipificación de salida masiva de la empresa Illumina. La empresa Illumina en USA, a mediados del 2009 puso a la venta el SNPchip50 para capturar el conjunto de los 55,609 polimorfismos validados en poblaciones grandes de ganado.

- Herramienta estándar para realizar genotipado en el genoma-completo para investigación y aplicaciones en la industria
- 54,609 SNPs distribuidos en los 30 cromosomas del genoma bovino
- 60,800 Beads

Parámetro	Rendimiento*	Especificación del producto
Tasa de lectura	99.50 %	> 99 %
Inconsistencia Mendeliana	0.04 %	> 0.1 %
Reproducibilidad	100 %	< 99 %
Frecuencia de alelo menor	0.19 media / 0.2 mediana	x
Locus polimórficos	37,758 media / 40,188 mediana	x

* Basado en 565 animales de 19 razas de ganado bovino, 46 tríos, y dos replicas

BovineSNP50 BeadChip

Figura 1.2.3.d. Primer SNPchip comercial para bovinos: SNP50 (Illumina, 2009)

revolucionando a la industria del ganado. La Selección Genómica se refiere a la toma de decisiones de selección 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 Locus de Características Cuantitativas (QTL, “Quantitative Trait Loci” por sus siglas en inglés) que contribuyen a la variación de un rasgo de valor económico. El efecto de los QTLs, inferidos ya sea de los haplotipos o genotipos de marcadores de tipo polimorfismo de nucleótido simple (SNP), es primeramente estimado en una población grande de referencia con información fenotípica. En las generaciones subsecuentes solo es necesaria la información de los marcadores para calcular los GEBVs.

La tecnología de selección genómica involucra tres pasos principales; 1) captura de las muestras y obtención del ADN genómico, 2) Aplicación de los SNPchips para obtención de los genotipos y 3) Análisis bioinformático de la información para obtener los GEBVs. La captura de las muestras pueden ser hechas de esperma, sangre, saliva, segregación nasal o una pinchada en la 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 los SNPchip. 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 láser 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 1.2.3.e.**

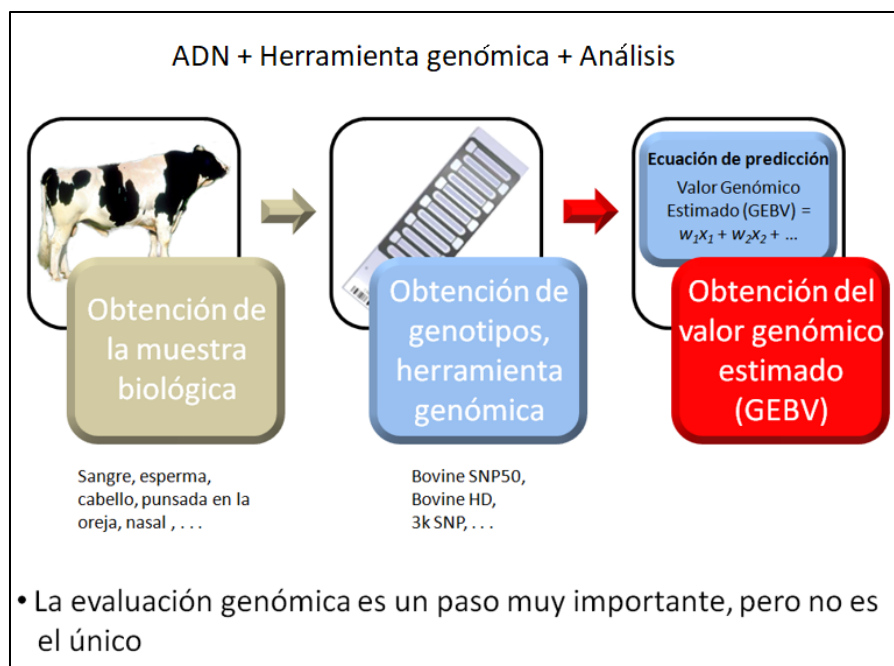


Figura 1.2.3.e. La selección genómica involucra tres pasos principales: 1) captura de la muestras y obtención de ADN genómico, 2) extracción de genotipos y 3) Análisis de la información.

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

1.3 Parámetros de caracterización de la estructura genética

Se entiende por estructura genética de una población a la cantidad y distribución de la variación genética presente en la misma. El conjunto de genes de una población recibe el nombre de pool génico, y su estructura genética se expresa en términos de frecuencias de los alelos (frecuencias génicas) presentes en la población y, consecuentemente, genotípicas. La

variabilidad está determinada por el número y la frecuencia de los distintos alelos del conjunto de genes de la población.

Para definir la estructura genética, o genómica, de una especie o una población, es necesario muestrear grupos de individuos y hacer genética y/o genómica comparativa. Los estudios más comunes de estructura genómica de poblaciones son:

1. Análisis de proporciones polimórficas.
2. Estudios de Equilibrio de Hardy-Weinberg.
3. Estudios de Heterosigosidad.
4. Estudios de endogamia.
5. Estudio del Índice de fijación.
6. Estudio del índice global de fijación.
7. Caracterización del desequilibrio por ligamento.
8. Estudio de la tasa de mutaciones.
9. Estimación de la población efectiva, estimación en retrospectiva.
10. Generación de mapas de haplotipos.
11. Otras.

1.3.1 Análisis de Proporciones polimórficas

Se refiere a clasificar los SNPs por la frecuencia de su alelo menor (MAF). Esto nos permite visualizar las proporciones de marcadores polimórficos en una población, e igualmente permite ver la cobertura de marcadores informativos en un ensayo. La **Figura 1.3.1.a** muestra las proporciones polimórficas de distintas razas.

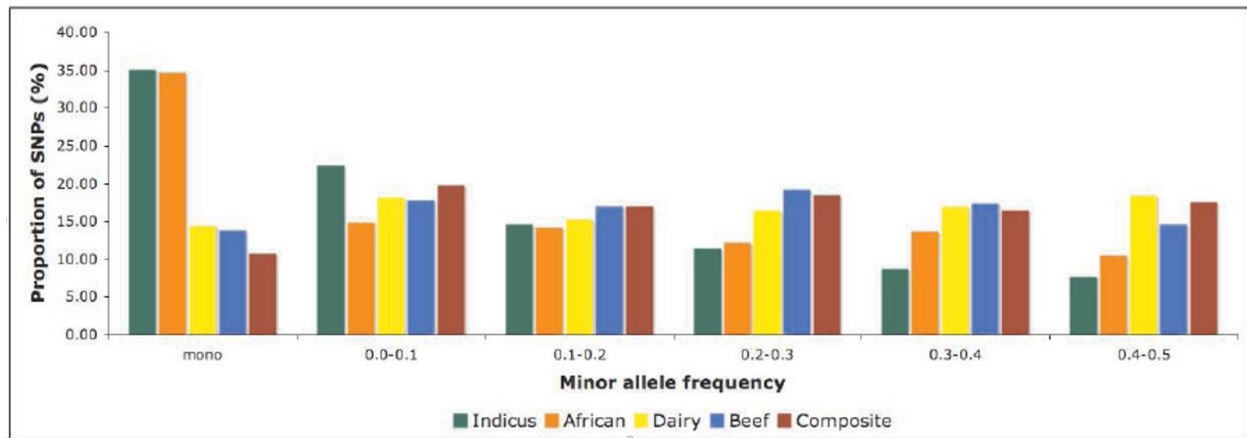


Figura 1.3.1.a. Proporciones polimórficas de 5 grupos de razas, incluyendo razas de ganado lechero, de carne, y las evoluciones *Bos Taurus*, *Bos Indicus* y razas híbridas.

Los valores de las frecuencias del alelo menor son un indicador de la cantidad de variabilidad genética en los seres vivos, en el caso de los organismos diploides solo hay dos posibles alelos en cada locus y aquel que es más frecuente en la población muestreada se le denomina Alelo Mayor (q); a su complemento Alelo Menor (p).

1.3.2 Análisis de Equilibrio de Hardy-Weinberg.

La ley del equilibrio de Hardy-Weinberg establece que cuando una población se encuentra en "equilibrio" la relación matemática de los alelos Menor y Mayor se puede representar por:

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

Al cumplirse la relación se considera un buen indicador de la "salud" genética de tal población. Y para verificar su cumplimiento, es necesario realizar unas pruebas de desviación de las frecuencias observadas de los genotipos, con respecto a las frecuencias esperadas de Hardy – Weinberg. Estadísticamente, esta corresponde una prueba *prueba χ^2 de Pearson*, para independencia de proporciones. Esto debido a que podemos obtener como proporción la frecuencia de cada alelo para cada SNP. La prueba es como sigue:

Prueba χ^2 de Pearson para independencia de proporciones:

La hipótesis que deseamos probar es: $H_0: p_1 = p_2$ vs : $H_1: p_1 \neq p_2$.

Donde p_1 = frecuencia observada, y p_2 = frecuencias esperada; en el SNPs.

1. Calculamos la estadística:

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

donde: χ^2 = estadística acumulativa de Pearson.

O_i = frecuencia observada.

E_i = frecuencia esperada (Hardy-Weinberg).

n = número de genotipos distintos = 3 (AA, Aa y aa).

2. Para un nivel de significancia α :

Rechazar H_0 si $\chi^2 > \chi^2_{df, 1-\alpha}$

Aceptar H_0 si $\chi^2 \leq \chi^2_{df, 1-\alpha}$

donde: df = número de posibles genotipos – número de alelos

3. Usar esta prueba sólo si ninguno de los valores esperados es menor de 5. De lo contrario utilice la prueba exacta de Fisher.

1.3.3 Estudios de Heterosigocidad

La presencia de alelos iguales en el mismo locus de un SNP dentro de un animal diploide se denomina Homocigoto, puede ser Homocigoto Mayor cuando el alelo más frecuente en el locus es el presente, y Menor cuando el menos frecuente aparece. En otra variante el animal

Heterocigoto es la condición donde ambos alelos están presentes en el SNP. H_0 es la llamada Heterosigocidad Observada, y es la frecuencia de los animales de la muestra con esa condición en ese locus o SNP. H_e es la Heterosigocidad Esperada, o predicha, y se puede aproximar su estimación, extrayendo de la ecuación de HW el término donde se encuentran los conteos del alelo mayor y del alelo menor:

$$H_e = 2pq = 2p(1 - p)$$

recordando también que la suma de las frecuencias relativas tiene la propiedad:

$$p + q = 1$$

de donde:

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

desarrollando la potencia del binomio:

$$HW = p^2 + 2pq + q^2 = 1$$

para detectar algún posible "desequilibrio" en la relación de Heterosigocidad, Wigginton J.E. *et al* (2005) desarrollaron un procedimiento matemático para realizar la prueba exacta de HW de tal modo que la probabilidad P_{value} se aproxime al valor real para que la desviación del "equilibrio natural" de HW pudiera explicarse como algo debido a la aleatoriedad o variabilidad normal de la distribución de frecuencias de los alelos en el locus del marcador SNP.

Una medida básica para cuantificar las diferencias en la estructura genética entre individuos, grupos y sub grupos está basada en medir el grado de heterosigocidad a nivel individuo, sub grupo y población total. La batería de cálculos para hacer estos estudios es llamada Estadísticas- F , de Wright; las cuales son un conjunto de índices que describen el nivel

de heterosigiosidad en una población. Específicamente, las estadísticas- F miden el grado de reducción en la homosisidad comparado con los valores de Hardy-Weinberg esperados. Las estadísticas- F son generalmente usadas para cuantificar el nivel de diferenciación genética entre subpoblaciones.

Su formulación general se basa en la siguiente definición:

Definición de Hudson, Slatkin y Maddison (1992)

$$F = \frac{\prod_{entre} - \prod_{dentro}}{X_{entre}}$$

Donde:

$\prod_{entre}$ y $\prod_{dentro}$ representan el número promedio de posiciones con diferentes alelos, entre dos individuos seleccionados de diferente $\prod_{entre}$ o $\prod_{dentro}$ de la misma población. Dada la definición anterior, estimamos de las mediciones jerárquicas de heterosigiosidad como sigue:

H_I = Heterosigiosidad promedio observada por individuo dentro de las subpoblaciones.

H_S = Heterosigiosidad esperada entre subpoblaciones con apareamiento aleatorio = $2p_iq_i$.

H_T = Heterosigiosidad esperada en la población total con apareamiento aleatorio = $2pq$.

1.3.4 Desequilibrio por ligamiento

La asociación no aleatoria entre los alelos de dos diferentes loci dentro de una población es conocida como "Desequilibrio por Ligamiento" (LD), considerando la situación utópica de

una muestra de individuos no relacionados consanguíneamente. Existen varias formas para cuantificar la asociación entre pares de locus, por ejemplo D y su versión máxima normalizada D', en HAPLOVIEW se obtiene la correlación de Pearson, dadas sus propiedades respecto del tamaño de la muestra y frecuencias relativas con valores muy extremos. Para calcular el valor se usó la formula:

$$r^2 = \frac{(p_{11} - p_1 q_1)^2}{(p_1 q_1 p_2 q_2)}$$

donde p_1 y p_2 son las frecuencias relativas de los alelos menor y mayor del SNP₁ en el locus #1 y q_1 y q_2 son las frecuencias relativas de los alelos menor y mayor del SNP₂ en el locus #2, además p_{11} es la frecuencia relativa observada a través de la población muestreada, el conteo de los animales que presentan ambos alelos menores de ambos SNP's.

1.3.5 Coeficiente de endogamia

De las defunciones de Slatkin y Maddison, de H_I , H_S y H_T obtenemos el coeficiente de endogamia, cuya expresión es la siguiente:

F_{IS} = Coeficiente de endogamia (Inbreeding coefficient): es la reducción promedio de H_0 en un individuo debido al apareamiento no aleatorio dentro de una subpoblación.

$$F_{IS} = \frac{H_S - H_I}{H_S}$$

F_{IS} es una medida del grado de endogamia (consanguinidad) dentro de una sub población. Esta varía desde -1 (todos los individuos son heterocigotos) a +1 (no heterocigotos observados).

1.3.6 Índice de fijación

El índice de fijación es definido como sigue:

F_{ST} = Índice de fijación (fixation index): es la reducción promedio de H_0 en una subpoblación (en relación a la población total) debido a deriva genética entre subpoblaciones:

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

F_{ST} es una medida del grado de diferenciación genética entre subpoblaciones. Varía desde 0 (no existe diferenciación) a 1 (existe total diferenciación, lo que significa que las subpoblaciones se han fijado para diferente alelos).

Wright, 1978, siguiere la siguiente guía cualitativa para la interpretación de F_{ST}

1. El rango de 0.0 a 0.05 indica una diferenciación genética pequeña.
2. El rango de 0.05 a 0.15 indica una diferenciación genética moderada.
3. El rango 0.15 a 0.25 indica una diferenciación genética grande.
4. Valores de F_{ST} mayores a 0.25 indica una diferenciación genética muy grande.

1.3.7 Estudio del Índice global de fijación

El índice global de fijación es la reducción promedio de H_0 de un individuo en relación a la población global. Su expresión de cálculo es la siguiente:

F_{IT} = Índice global de fijación (overall fixation index):

$$F_{IT} = \frac{H_T - H_I}{H_T}$$

F_{IT} combina las contribuciones entre apareamiento no aleatorio dentro de subgrupos (F_{IS}) y efectos de deriva aleatoria en todos los subgrupos (F_{ST}).

La relación entre las tres estadísticas F es la siguiente:

$$(1 - F_{IT}) = (1 - F_{IS})(1 - F_{ST})$$

1.3.8 Caracterización del desequilibrio por ligamento

El desequilibrio por ligamento (LD) es un término utilizado para estudiar la asociación no aleatoria entre dos o más locus, no necesariamente en el mismo cromosoma. LD es generalmente causado por la interacción entre genes, por ligado genético, por la tasa de recombinación, por la deriva genética aleatoria, por el apareamiento no aleatorio y por la estructura poblacional.

Caracterizar el desequilibrio por ligamento de una muestra es importante porque si un polimorfismo que está relacionado directamente con la causa de un fenotipo no es genotipificado en el ensayo, mantenemos cierto grado de probabilidad de detectar su efecto a través del desequilibrio por ligamento de los marcadores que sí han sido genotipificados. El desequilibrio por ligamento es un fenómeno no cuantitativo, por lo que no existe una escala natural para medirlo. Se han propuesto diversas formas de cuantificarlo, dentro de las cuales; las más usadas son los coeficientes D , D' y r^2 que están basadas en los haplotipos de pares de marcadores.

Coefficiente de desequilibrio por ligamiento, D

Sean A y B dos locus segregando dos alelos cada uno; A_1 y A_2 con frecuencias p_1 y p_2 en A , y con frecuencias q_1 y q_2 en B .

	SNP1	SNP2
Indv 1	A ₁	A ₂
Indv 2	A ₂	A ₁
Indv 3	A ₁	A ₁
Indv 4	A ₁	A ₁
Indv 5	A ₁	A ₂
Indv 6	A ₁	A ₂
Indv 7	A ₂	A ₂
Indv 8	A ₁	A ₂
Indv 9	A ₂	A ₂
Indiv 10	A ₂	A ₂
	$p_1 = 6/10$	$q_1 = 3/10$
	$p_2 = 4/10$	$q_2 = 7/10$

Si los SNPs segregaran alelos de forma aleatoria e independiente, entonces esperaríamos que:

$$p_{11} = p_1 q_1 = (6/10)(3/10) = 0.18$$

$$p_{12} = p_1 q_2 = (6/10)(7/10) = 0.42$$

$$p_{21} = p_2 q_1 = (4/10)(3/10) = 0.12$$

$$p_{22} = p_2 q_2 = (4/10)(7/10) = 0.28$$

Las proporciones observada en la muestra son:

$$p_{11} = 2/10 = 0.2$$

$$p_{12} = 4/10 = 0.4$$

$$p_{21} = 1/10 = 0.1$$

$$p_{22} = 3/10 = 0.3$$

Si las proporciones fueran iguales, diríamos que la muestra está en equilibrio.

Podemos expresar la desviación de los valores observados con respecto a los valores esperados de la siguiente forma:

	A_1	A_2	$Total$
A_1	$p_{11} = p_1 q_1 + D$	$p_{12} = p_1 q_2 - D$	p_1
A_2	$p_{21} = p_2 q_1 - D$	$p_{22} = p_2 q_2 + D$	p_2
$Total$	q_1	q_2	1

De la relación anterior tenemos que: $D = p_{11} - p_1 q_1$. Donde, D es dependiente de las frecuencias de los alelos en los SNPs.

Coeficiente de Desequilibrio por Ligamiento, D'

Podemos calcular una versión de LD, que es no dependiente de las frecuencias de los alelos, como sigue:

$$D' = \begin{cases} \frac{D}{\min(p_1 q_1, p_2 q_2)} , & \text{si } D < 0 \\ \frac{D}{\min(p_1 q_2, p_2 q_1)} , & \text{si } D > 0 \end{cases}$$

Es sensitivo a las recombinaciones entre los locus, aun cuando el número de recombinaciones sea muy pequeño. Tiene una caída exponencial con respecto al tiempo. Puede tener valores grandes, indicando alto desequilibrio por ligamento, aun cuando uno de los alelos tenga una frecuencia demasiado pequeña, lo cual lo convierte en un alelo de poco interés.

Desequilibrio por ligamiento por coeficiente de correlación cuadrada, r^2

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

Es la medida preferida por los genetistas. Es independiente de las frecuencias de los alelos. Su rango de valores es entre 0 y 1. $r^2 = 1$ implica que los SNP's proveen exactamente la

misma información. Existe un desequilibrio máximo. $r^2 = 0$ cuando los SNP's están en perfecto equilibrio.

1.3.9 Estudio de la tasa de mutaciones

La tasa de mutaciones corresponde a la frecuencia de aparición de nuevas mutaciones en un gen u organismo, a través del tiempo [20]. La tasa de mutaciones no es contante y no está limitada a un tipo específico de mutación. Esta puede ser medida utilizando una prueba llamada Prueba de Fluctuación, también conocida como Experimento de Luria-Delbrück [21].

Como se muestra en la **Figura 1.3.9.a**, El experimento de Luria-Delbrück realiza dos posibles pruebas: a) si la mutación es inducida por el medio, en cuyo caso, aproximadamente el mismo número de mutantes se esperaría que aparecieran en cada plato de crecimiento, y b) si la mutación aparece de forma espontánea durante la división celular, entonces cada plato de crecimiento tendrá una número muy distinto de mutantes.

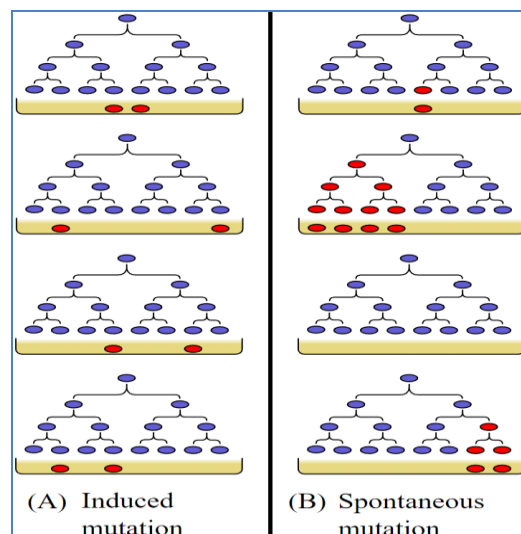


Figura 1.3.9.a. Experimento de Luria-Delbrück para medir la tasa de mutaciones.

1.3.10 Estimación de la población efectiva, estimación en retrospectiva.

La población efectiva (N) dentro de una población general es la cantidad de individuos que contribuyen genéticamente en la formación de la siguiente generación. En los vertebrados esto es, los individuos que son cruzados sexualmente, y cuyos hijos forman la siguiente generación. Una definición más técnica, en términos genéticos, de población efectiva es:

“El número de individuos con valor positivo de endogénia en una población idealizada, que mostraría la misma cantidad de dispersión en las frecuencias alélicas, bajo efectos de deriva genética, o la misma cantidad de endogénia que se observa en la población en estudio”.

Cálculo de la población efectiva:

Una forma de estimar la población efectiva es utilizando una aproximación al valor esperado del desequilibrio por ligamento (r^2) en una población. Esto es:

$$E(r^2) = \frac{1}{4Nc + 1}$$

donde:

- N = población efectiva $1/(2c)$ generaciones en el pasado,
- $E(r^2)$ = promedio de los valores de r^2 para todos los pares de SNPs en un rango determinado,
- c = mediana del rango, en *centiMorgans*.

Procedimiento para calcular la población efectiva para generaciones específicas en el pasado:

1. Se selecciona el número de generaciones en el pasado, para la cual se desea calcular la población efectiva.
2. Se estima la mediana c , en Morgans, del rango.

3. Se obtiene el promedio de r^2 para todos los pares de SNPs en el rango considerado.
4. Se estima N utilizando la formula:

$$E(r^2) = \frac{1}{4Nc + 1}$$

De esto que:

$$N = \frac{1}{4c} \left[\frac{1}{E(r^2)} - 1 \right]$$

La **Figura 1.3.10.a** muestra la población efectiva a través de una serie de generaciones pasadas. En la gráfica se observa como la población ha ido disminuyendo con el tiempo, y donde han sucedido eventos de exterminio y/o reducción de la población efectiva de distintas razas de ganado bovino.

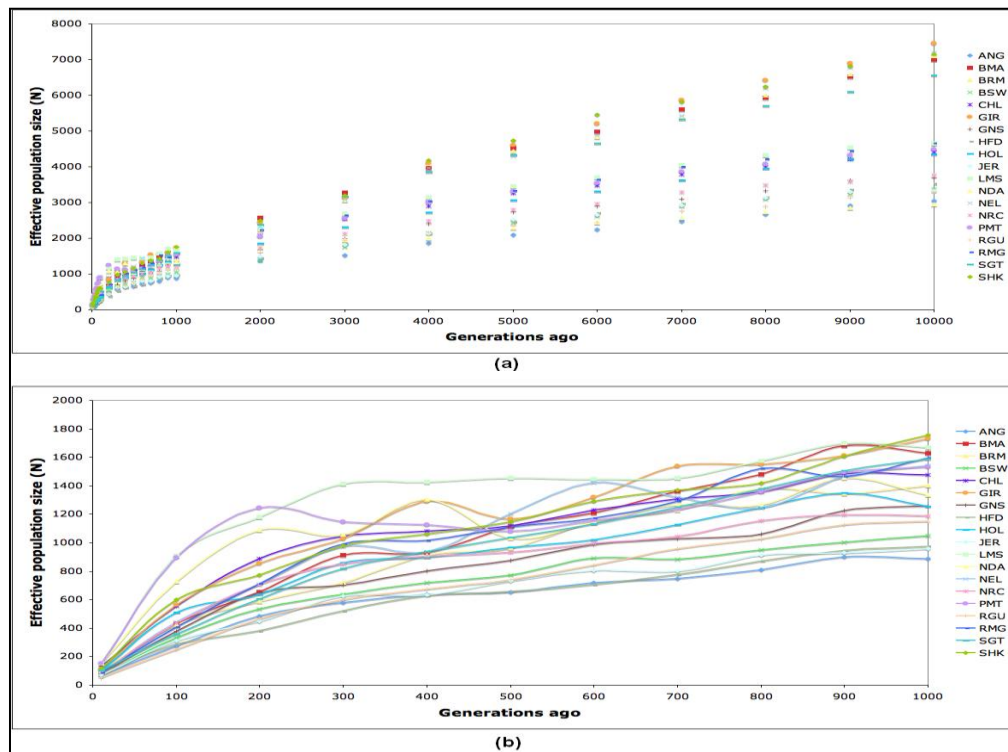


Figura 1.3.10.a. Estimación de la población efectiva para las 19 razas de ganado bovino del HapMap.

1.3.11 Generación de mapas de haplotipos

Generar un mapa de haplotipos del genoma nos permite encontrar los valores exactos de alelos tanto en locus individuales como en grupos de éstos. Igualmente nos permite definir bloques que son heredados de forma conjunta. Esto nos ofrece la máxima resolución en estudios de asociación. Los bloques de haplotipos nos muestran de forma muy concreta la estructura genética y nos permiten estudiar la evolución de las especies.

Para generar bloques de haplotipos es necesario descifrar los haplotipos de cada individuo en la muestra. Una vez elucidados los haplotipos de los individuos de la muestra, se aplica alguna definición de bloque de haplotipos para obtener un mapa de sus tamaños y posiciones en los cromosomas. A continuación se describe una definición de bloque de haplotipos:

Bloques de haplotipos basados en el coeficiente de correlación, r^2 . [19]

Este algoritmo inicia un bloque buscando el par de SNPs adyacentes que contenga el valor de r^2 mas alto (no menor a α). Extiende el bloque hacia un SNP adyacente; si y solo si: el valor promedio de r^2 entre el SNP adyacente y todos los miembros actuales del bloque es mayor a β , y todos los valores individuales de r^2 , del SNP adyacente y los SNP's miembros del bloque, son mayores a γ .

Siempre; $\alpha > \beta > \gamma$. Por default, $\alpha = 0.4$, $\beta = 0.3$ y $\gamma = 0.1$

Los SNP's adyacentes, candidatos a ser agregados a un bloque son seleccionados de forma alternada; uno a la derecha del bloque actual y el siguiente a la izquierda, o viceversa.

Después de identificar el primer bloque, un nuevo par de SNP's adyacentes con el valor de r^2 más alto (no menor a α) es utilizado para iniciar el siguiente bloque. Ocasionalmente un SNP puede ser asignado a un nuevo bloque adyacente, en este caso debemos asignar el SNP al bloque previamente existente.

Nota: Los valores de $\alpha = 0.4$, $\beta = 0.3$ y $\gamma = 0.1$ fueron asignados debido a que mostraron tener mejor consistencia definiendo patrones de bloques en poblaciones de prueba con la misma región geográfica.

1.4 La importancia de las pruebas de paternidad en ganado bovino

La calidad de la información en el registro del pedigrí es importante debido a su utilización en el diseño de los modelos matemáticos y estadísticos usados para estimar los EPD's en las evaluaciones genéticas y las imprecisiones del registro de pedigrí tienen un efecto negativo en la calidad de tales predicciones.

Se ha vuelto necesario verificar la consistencia y congruencia de los registros de las ganaderías, se han buscado posibles estrategias para atender el problema, sin embargo la Técnica moderna más precisa y económica vino a ser la genotípica, con las nuevas tecnologías de captura, purificación, amplificación y lectura del DNA mediante el uso de SNPChips, permite actualmente a muy bajo costo certificar la paternidad y los registros de pedigrí.

1.5 Pruebas de paternidad utilizando marcadores genéticos

En situaciones donde solamente se encuentran padres putativos el uso de SNP's en la prueba de paternidad resulta muy atractivo, la exclusión de los supuestos padres en un trío que

pertenece a una población en condiciones favorables puede simplificarse bastante ante el hecho de hallar únicamente dos alelos en cada locus.

Según la información de los haplotipos o genotipos disponibles se pueden completar dos tipos de pruebas en la exclusión utilizando los alelos y su frecuencias relativas [22]:

Para un padre:

$$P = 1 - 4 \sum_{i=1}^n p_i^2 + 2 \left(\sum_{i=1}^n p_i^2 \right)^2 + 4 \sum_{i=1}^n p_i^3 - 3 \sum_{i=1}^n p_i^4$$

Para dos padres:

$$P = 1 + 4 \sum_{i=1}^n p_i^4 - 4 \sum_{i=1}^n p_i^5 - 3 \sum_{i=1}^n p_i^6 - 8 \left(\sum_{i=1}^n p_i^2 \right)^2 + 8 \left(\sum_{i=1}^n p_i^2 \right) \left(\sum_{i=1}^n p_i^3 \right) + 2 \left(\sum_{i=1}^n p_i^3 \right)^2$$

Las probabilidades se calculan para cada locus probado, donde p_i es la frecuencia relativa del alelo i , n el numero de alelos (usualmente 2 por SNP para organismo diploides). El poder total de exclusión se calculó combinando todas las probabilidades P de cada unos de los locus, con la siguiente fórmula:

$$Q = 1 - (1 - P_1)(1 - P_2)(1 - P_3) \dots (1 - P_k),$$

donde k es el número de locus usado[23].

CAPÍTULO II. PLANTEAMIENTO DEL PROBLEMA Y OBJETIVOS

2.1 Planteamiento del Problema

La calidad de los registros de pedigrí en las ganaderías implica un trabajo arduo y metódico, la cantidad de factores que pueden alterar la exactitud del mismo es grande y cambia según el sistema de producción (estabulado, extensivo o combinado), por lo que se ha tenido que recurrir a pruebas genéticas para identificar con mayor precisión y exactitud los padres y ancestros de cada animal. Otra vertiente es la trazabilidad o rastreabilidad de los productos del ganado bovino, es decir identificar correctamente el origen de un corte de carne, o un vaso de leche calificada como orgánica, asignando así la responsabilidad al productor cuando haya un incumplimiento en los estándares especificados para su producto. El estudio del genoma, el varioma y el transcriptoma de los bovinos está incidiendo en la identificación de marcadores SNP's significativos para múltiples propósitos; como un mínimo costo, mayor precisión y exactitud en las predicciones, convirtiéndose en un elemento clave para la selección de empadramientos, en el proceso de mejora genética mediante selección genómica.

2.2 Objetivo general

Encontrar una combinación de marcadores SNP's que mejore significativamente el proceso de verificación de Pedigrís, reduciendo la cantidad de marcadores, simplificando los cálculos, y reduciendo costos del procesamiento físico de los laboratorios bio-moleculares e informáticos

2.3 Objetivos específicos

- I. Estudiar la variabilidad de los SNP's propuestos por ISAG e ICAR en los paneles de pruebas de paternidad, analizando su posible estructura y patrones de variabilidad alélica.
- II. Comparar la población Hereford Mexicana con los datos del Bovine HapMap Project para evaluar las diferencias, distancias y posibles estructuras.
- III. Discriminar aquellos SNP's que aportan una cantidad menor de información relevante para el éxito del proceso. Reduciendo la cantidad de SNP's necesarios, identificando aquellos exclusivos o indispensables para la población Hereford México.

CAPÍTULO III. DESCRIPCIÓN DE LOS DATOS

3.1 Datos de ganado Hereford

La ganadería Hereford de México aportó 1317 animales registrados provenientes de 4 regiones del país y 18 unidades de producción relevantes de acuerdo a la producción anual reportada en el año 2015. La selección de cada elemento muestral fue aleatoria dentro de cada unidad de producción y el trabajo de campo para la obtención del fluido sanguíneo se desarrolló durante el verano del 2016, mediante un equipo de Técnicos entrenado para tal propósito y contando con el apoyo del personal de cada una de las UPR's (Unidad de Producción Representativa) muestreadas, se detectó un total de 218 tríos. Los animales en el estudio fueron segregados y manejados siguiendo las directrices de investigación en seres vivos planteadas en las Normas Mexicanas para el estudio de especies de interés zootécnico NOM-003.ZOO-1994, NOM-062-ZOO-1999 además la CONARGEN (Consejo Nacional de los Recursos Genético Pecuarios) y la SAGARPA (Secretaría de Agricultura Ganadería Desarrollo Rural Pesca y Alimentación) contribuyeron en el diseño y la supervisión durante todas las etapas del trabajo de campo para el adecuado cumplimiento de las disposiciones legales pertinentes.

3.1.1 Obtención de las muestras sanguíneas

La sangre utilizada para la extracción y Genotipificación de los marcadores polimórficos de nucleótido simple (SNP's) se obtuvo de una de las orejas localizando a simple vista alguno de los vasos sanguíneos mediante punción con agujas esterilizadas nuevas y el fluido depositado por gravedad en tarjetas Whatman® FTA las cuales fueron posteriormente introducidas en el sobre metalizado provisto en el kit, mismo que fue usado para la preservación y transporte del

ADN desde México a GeneSeek® en Lincoln, Nebraska, Estados Unidos de Norte América. Los laboratorios de esta empresa han sido certificados para cumplir con los procedimientos estándar aprobados por la ISAG (International Society for Animal Genetics) y el ICAR (International Committee for Animal Recording). Ahí se extrajo, depuró y amplificó el ADN, se identificaron y validaron los alelos de los marcadores presentes en cada animal y pertinentes a cada uno de los 206 SNP's del panel de paternidad para razas bovinas, también aprobado por los organismos internacionales mencionados.

3.1.2 Organización y depuración de los datos

Una vez recibidos los resultados del proceso de lectura de los marcadores SNP's en GeneSeek®, de acuerdo al orden de los lotes (grupos de animales) muestreados, se ensambló una Base de Datos en forma de genotipos con 206 locus de marcadores, cada uno con su par de alelos y distribuidos a través de los 29 cromosomas autosomales de los 1317 animales. Asimismo se acopló una segunda Base de Datos en forma de haplotipos de alta densidad (~735 kbp) obtenida directamente del Bovine Hapmap Consortium y construida a partir de 27 sementales de corte internacional. Para hacer posible la confrontación entre ambas Bases, se aplicó un código escrito especialmente para tal propósito utilizando el lenguaje de programación PERL (Practical Extraction & Reporting Language), mediante el cual se extrajeron de la Base de alta densidad (HapMap) los marcadores de Paternidad correspondientes al panel de ganado Hereford México y se ordenaron coincidiendo con la estructura de datos del archivo preparado con la información recibida de GeneSeek®.

Ambas Bases de Datos fueron filtradas usando idénticos criterios de control de calidad, pero de forma independiente en cada archivo, en la **Tabla 3.1.2.a**, se presentan los valores discriminantes en cada criterio, tanto para marcadores SNP's como para animales.

Tabla 3.1.2.a. Criterios y valores de exclusión para Control de Calidad.

CRITERIO	SNP's	ANIMAL
Errores Mendelianos	Alelos en locus >2 por columna	Alelos en locus <2 ó fuera de protocolo
No informativo	Monomórfico (no informativo)	Monomórfico (no informativo)
Datos faltantes (No Call Rate)	> 5% Por columna	> 5% Por renglón
Frecuencia de Alelo Menor	< 0.05	No Aplica
Hardy-Weinberg (Prueba exacta de Fisher)	$P_{(valor)} \leq 0.15$	No Aplica

Después de aplicar estos criterios de calidad a la muestra genotipada de la población Hereford Mexicana la Base de Datos filtrada incluyó un total de 171 marcadores SNP's y únicamente un animal fue excluido, totalizando 1316 individuos válidos en la muestra. El promedio de lecturas alélicas obtenidas (Gen Call Rate) por marcador fue de $97.85\% \pm 0.603\%$ y por animal $98.3\% \pm 0.031\%$.

3.2 Datos del consorcio HapMap bovino

Los datos crudos del Bovine HapMap Consortium se integraron desde 29 archivos diferentes, uno para cada cromosoma autosomal, conteniendo en promedio 25,350 bp de alelos por cromosoma y un total de 735,158 bp, en forma de haplotipos. Se lograron identificar

160 marcadores SNP's (coincidentes) como materia prima para el estudio y en la **Tabla 3.2.1.a** se presenta un resumen por cromosoma, indicando la convergencia de los controles de calidad en ambas fuentes.

Tabla 3.2.1.a Cantidad de marcadores SNP's Iniciales y Finales.

CROMOSOMA	GENOTIPOS GENESEEK®		HAPLOTIPOS HAPMAP	
	CRUDOS	FILTRADOS	CRUDOS	FILTRADOS
1	11	10	46,492	10
*2	10	9	40,047	8
3	8	7	35,573	7
4	8	6	34,968	6
5	9	7	34,828	7
*6	9	8	35,512	6
7	10	8	33,164	8
8	6	4	33,518	4
9	5	4	31,052	4
10	9	8	30,442	8
11	11	10	32,008	10
*12	5	5	26,120	4
13	6	6	23,591	6
14	5	3	24,774	3
15	5	4	24,750	4
16	8	8	24,177	8
*17	8	8	22,263	5
*18	7	6	19,385	5
19	9	7	18,906	7
20	6	5	21,488	5
*21	7	7	21,173	6
22	8	6	18,032	6
23	4	3	15,214	3
24	4	3	18,618	3
25	5	3	12,931	3
*26	4	4	15,241	2
27	6	3	13,149	3
28	5	5	13,035	5
29	4	4	14,707	4
TOTALES	202	171	735,158	160

*Al menos un SNP no coincidente con el HapMap.

Los haplotipos de marcadores SNP's del HapMap coincidentes y posteriormente filtrados para control de calidad arrojaron un número de lecturas alélicas obtenidas por locus (Gen Call Rate) del 100% idénticamente para cada uno de los 27 animales muestreados dentro del panel internacional seleccionado por el Consorcio.

En 7 de los 29 cromosomas se presentó al menos una omisión en la coincidencia de marcadores con las lecturas obtenidas de GeneSeek®, un total de 11 SNP's del HapMap se omitieron en los análisis posteriores.

CAPÍTULO IV. ANÁLISIS Y RESULTADOS

Para los cálculos necesarios en la estimación de los parámetros de variabilidad alélica se utilizaron los "Haplotipos" dadas sus propiedades de heredabilidad y que además a su vez facilitan la trazabilidad de los bloques de DNA recombinados durante la meiosis; en el caso del material genotipado en GeneSeek® proveniente de animales Mexicanos se tuvo que agregar una etapa adicional intermedia antes de procesar la información en el análisis numérico, mientras que en el caso del HapMap se recibieron los alelos ordenados y agrupados ya en forma de haplotipos por lo tanto no se requirió de preprocesamiento alguno.

4.1 Estimaciones para la reconstrucción de Haplotipos Hereford México (PHASE)

La reconstrucción de Haplotipos de animales genotipados de la ganadería Hereford Mexicana se basó en el trabajo de Stephens *et al.*(2001), Stephens *and* Scheet (2005) con su programa computacional PHASE 2.1.1 en sistema operativo LINUX el cual implementa un algoritmo basado en estadística Bayesiana que toma en cuenta la caída del desequilibrio por ligamiento al aumentar las distancias entre los marcadores involucrados así como la posible tasa de recombinación encubierta por el fenómeno, sin embargo el efecto es mayor cuando la densidad de marcadores es muy grande y las distancias entre ellos muy pequeñas, no es el caso aquí, solamente tenemos 171 SNP's y las distancias entre ellos dentro de cada cromosoma son demasiado grandes. La información de los tríos de animales identificados (218) no se incluyó en el proceso de reconstrucción, para posteriormente utilizar la información como validación de la calidad de las estimaciones del algoritmo de PHASE.

Para alimentar el programa se prepararon con un código propio en PERL 29 archivos con extensión ".inp" para declarar los datos de entrada. Como resultados del proceso computacional PHASE 2.1.1 exporta los haplotipos estimados en un archivo por cromosoma extension ".out_pairs" y dentro del mismo a cada animal se le puede reportar una de dos diferentes condiciones; primero, en caso de unicidad del haplotipo sugerido hay un resultado único por lo tanto la asignación es automática, pero en segunda instancia, otra situación se presenta cuando hay más de un haplotipo posible, entonces se enlistan todos los haplotipos candidatos cada uno con su probabilidad estimada pero discriminando aquellos con $P_{(H)} \leq 0.001$.

El **Gráfico 4.1.a** presenta la cantidad de haplotipos posibles según las combinaciones de alelos usados en cada cromosoma, el promedio del conteo fue de 159 ± 47 .

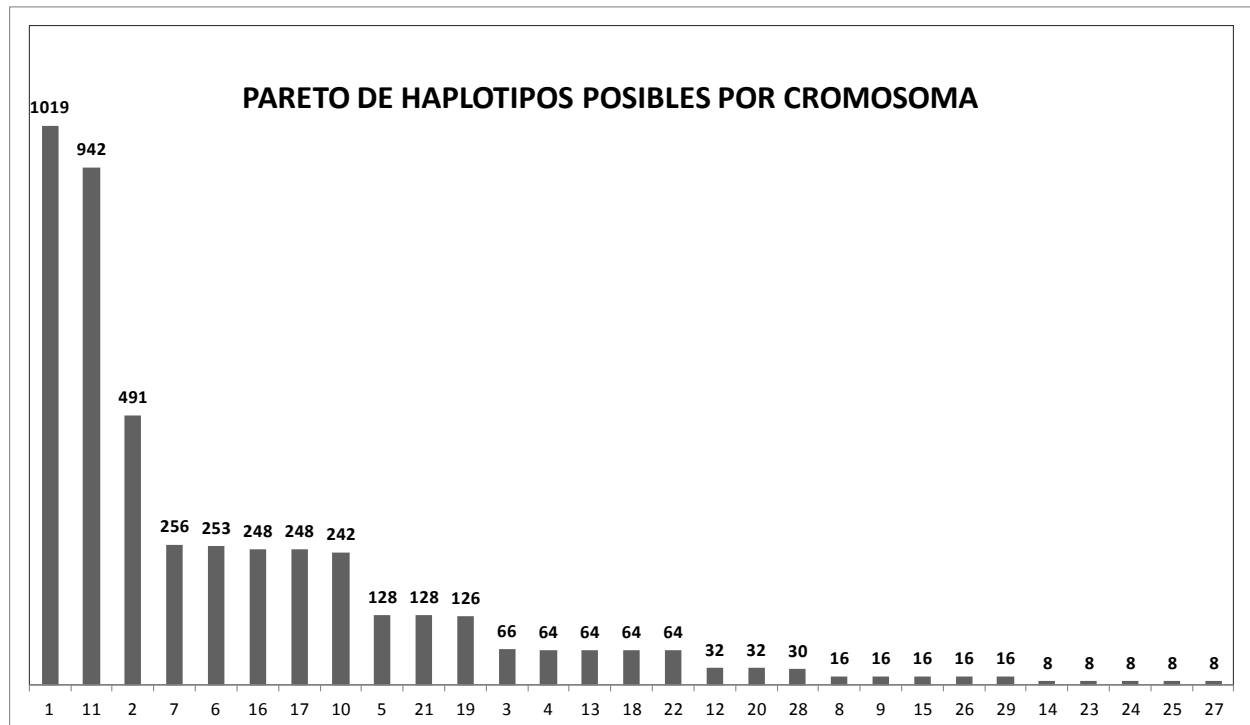


Gráfico 4.1.a. Contribución de cada cromosoma a las frecuencias de posibles haplotipos.

Un código en PERL se diseñó para extraer del archivo con extensión ".out_pairs" los haplotipos correspondientes a cada animal, ya sean caso único o de lo contrario identificando el haplotipo con el par de cadenas más probable. La lista depurada de cada cromosoma se formateó de modo que pudiera ser usada como entrada de datos para el programa HAPLOVIEW con una extensión ".ped".

El cromosoma 1 y 11 son los más abundantes en SNP's muestreados y por lo tanto, dado que la cantidad de posibles combinaciones de alelos en los haplotipos es geométrica, resultan los primeros en el orden de contribución al total, sin embargo hay una diferencia considerable ($\Delta=1019-942=77$) en su conteo. El **Gráfico 4.1.b** enfatiza la relación geométrica y el alto valor del determinante R^2 abona en la demostración, sin embargo ya se logra percibir una señal de diferencia en el segmento de los cromosomas 3,5,19,21 todos con 7 SNP's pero cada uno con diferentes cantidades de posibles haplotipos.

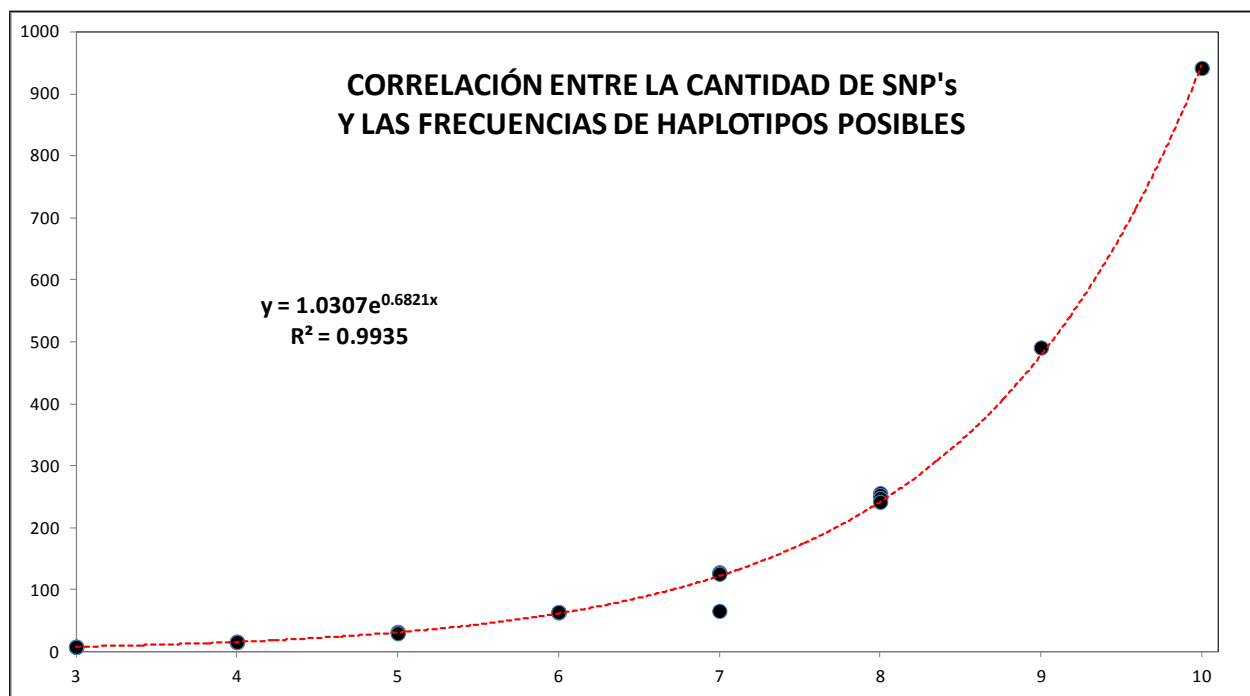


Gráfico 4.1.b. Correlación de SNP's muestreados por cromosoma y la cantidad de posibles haplotipos.

Las frecuencias haplotípicas por cromosoma después de la selección (optima) para la reconstrucción más probable se pueden apreciar en el **Gráfico 4.1.c** nótese que algunos cromosomas no siguen el mismo orden que el de las frecuencias posibles y es más evidente el caso de los cromosomas ya mencionados.

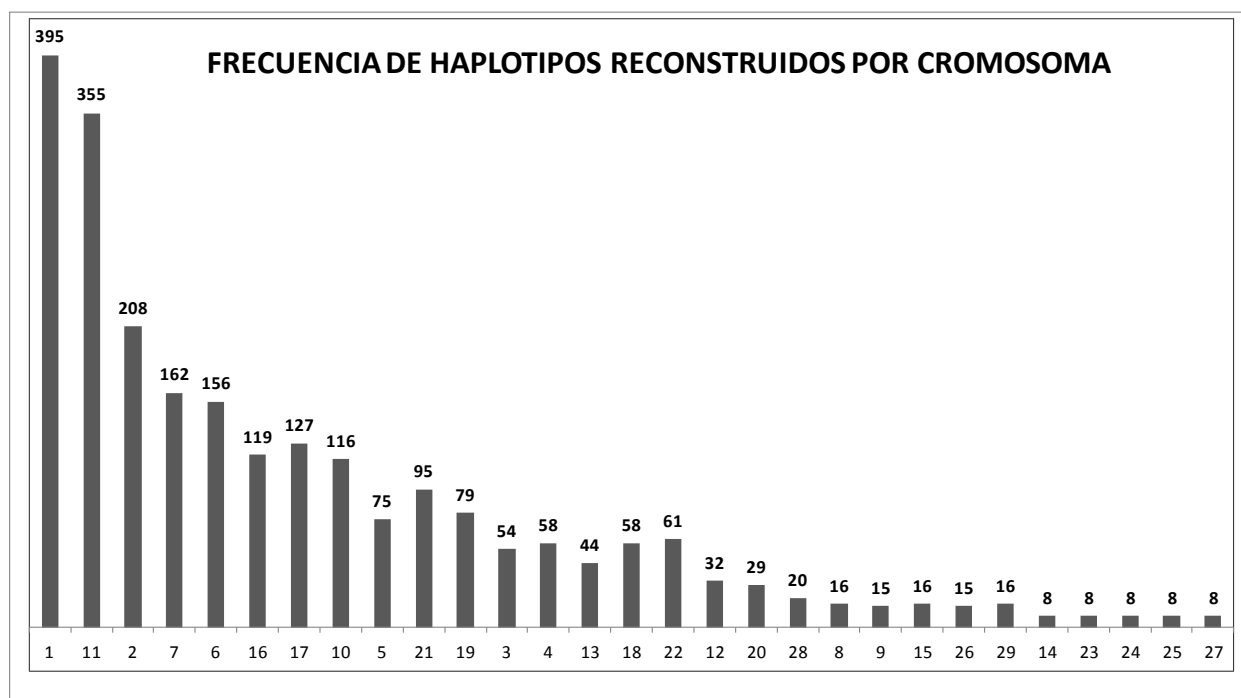


Gráfico 4.1.c. Frecuencias de haplotipos optimizados por cromosoma para la mejor reconstrucción posible.

Para la validación de calidad del proceso de reconstrucción de haplotipos, se contrastaron los haplotipos predichos para cada descendiente en cada trío, con los haplotipos de sus padres putativos, registrando las no coincidencias para cada locus, sin penalizar los casos donde ambos alelos fueron predichos erróneamente, es decir , ambos Padres con Alelo Mayor vs Hijo con ambos alelos menores y viceversa. De un total de 67,846 posiciones comparadas, se detectaron 5,671 diferencias, lo que arroja una tasa de 8.36% de posibles fallas en la reconstrucción de los haplotipos, sin embargo usar la estimación de haplotipos resulta la mejor opción para el análisis del Ganado Mexicano en este estudio.

4.2 Estimadores de Variabilidad Alélica como indicadores de Variabilidad Genética (HAPLOVIEW)

Ambas bases de datos conteniendo los haplotipos reconstruidos fueron procesadas por separado y de forma independiente, utilizando el Interface Gráfico de Usuario del Programa HAPLOVIEW 4.2 desarrollado por Barret, J.C. *et al* (2005). Se analizaron cada uno de los 58 (58/2=29) archivos por separado y los resultados se concentraron en un archivo tipo ".csv" y así facilitar el desarrollo del análisis final. En la primera etapa del proceso computacional HAPLOVIEW desempeñó un nuevo filtrado de calidad para cada marcador, en la **Tabla 4.2.a** se presentan las etiquetas de columna en la salida de resultados primarios. Es relevante mencionar que los límites de este segundo filtro no son los mismos de la primera etapa, cuando las lecturas mostraban genotipos, en HAPLOVIEW se dan de entrada haplotipos reconstruidos mediante el algoritmo de PHASE .

Tabla 4.2.a. Descripción de las etiquetas de columna en los resultados de HAPLOVIEW

ETIQUETA HAPLOVIEW	DESCRIPCIÓN
#	Número del Marcador
Name	Identificación del Marcador especificada en el Archivo _entrada.info
Position	Posición del Marcador especificada en el Archivo_entrada.info (bp)
ObsHET	Heterosigocidad observada
PredHET	Heterosigocidad predicha (esperada)
HWpval	Equilibrio de Hardy-Weinberg (P_{value})=Hipotesis Nula de equilibrio
%Geno	Porcentaje de genotipos no-faltantes
FamTrio	Cantidad de trios genotipados
MendErr	Número de errores hereditarios Mendelianos (=0 para animales sin pedigree)
MAF	Frecuencia del alelo menor ($p=1-q$)
Alleles	Identificación de la base correspondiente a cada alelo (A, T, C, G)
Rating	Identificador de la discriminación del marcador por no cumplir algun(s) criterio

4.3 Resultados para frecuencias MAF, H_e , H_o y HW

Aún y cuando los cromosomas 1 y 11 presentaron la mayor frecuencia de haplotipos entre los 29 cromosomas, en la población Hereford México el MAF correspondiente los sitúa en la 4a y 17ava posición respectivamente, el **Gráfico 4.3.a** enseña el Pareto de la contribución promedio al MAF por cromosoma.

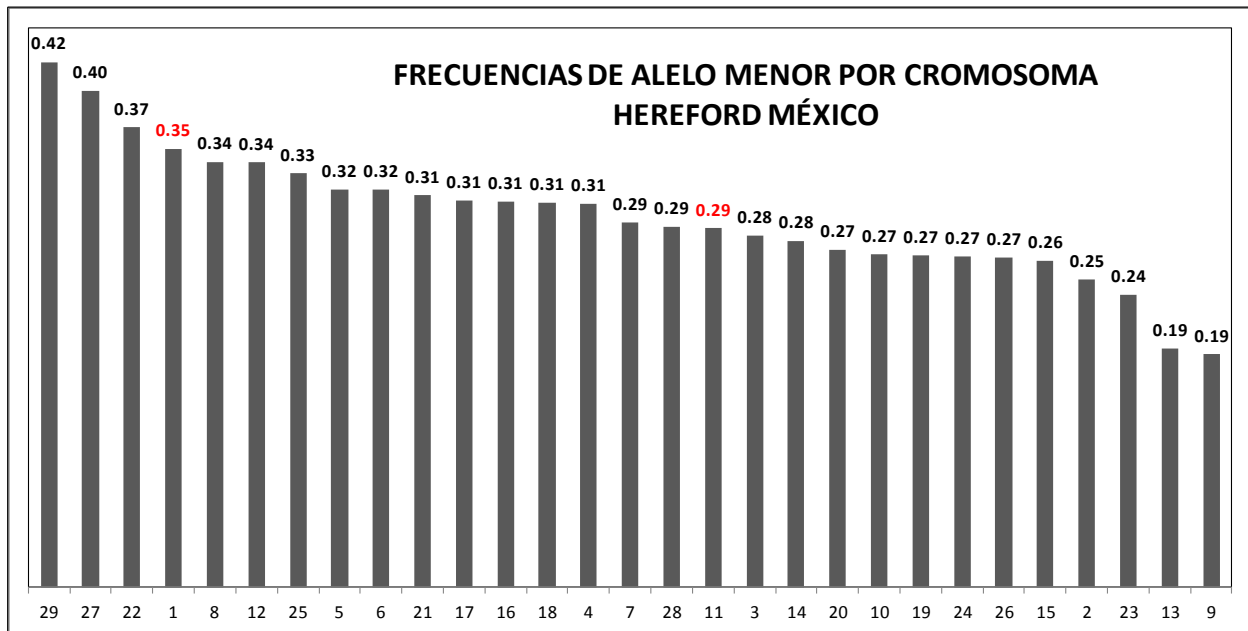


Gráfico 4.3.a. MAF para el Hereford Mexicano.

Agrupando la información del MAF por marcador, de forma independiente a su cromosoma, los valores indican una diferencia importante entre las poblaciones México y HapMap. En el **Gráfico 4.3.b** se construyó un comparativo agrupando por clases de menor a mayor, y se agregó una línea de regresión lineal simple enfatizando la tendencia en ambas muestras. El cruce de las líneas en el centro de la distribución aporta evidencia en la hipótesis de una posible diferencia en la variabilidad alélica entre las poblaciones.

En el **Gráfico 4.3.c** los MAF por SNP del ganado Hereford HapMap, los cromosomas 1

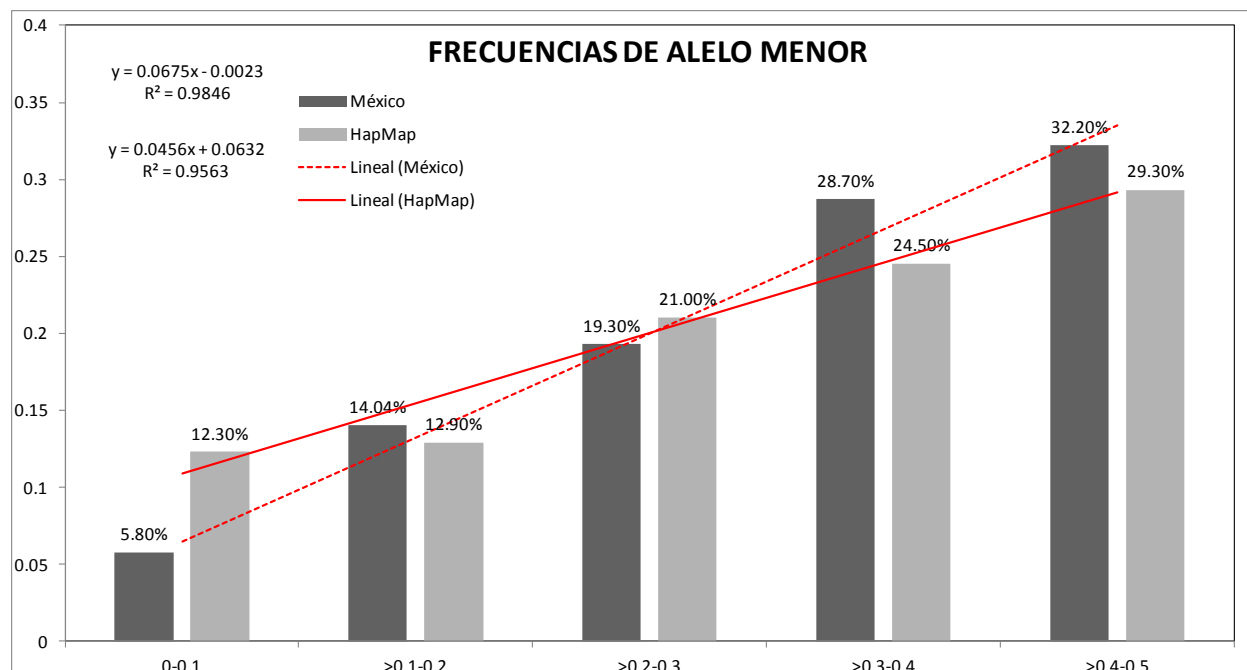


Gráfico 4.3.b. MAF comparativo agrupado para ambos orígenes de información.

y 11 tienen el mismo orden de aparición que en Hereford México, pero los valores no están ordenados de Mayor a menor, por lo tanto su ubicación real es la 2a y 4a posición , la contribución del cromosoma 11 sugiere una mayor importancia en la variabilidad alélica del ganado HapMap.

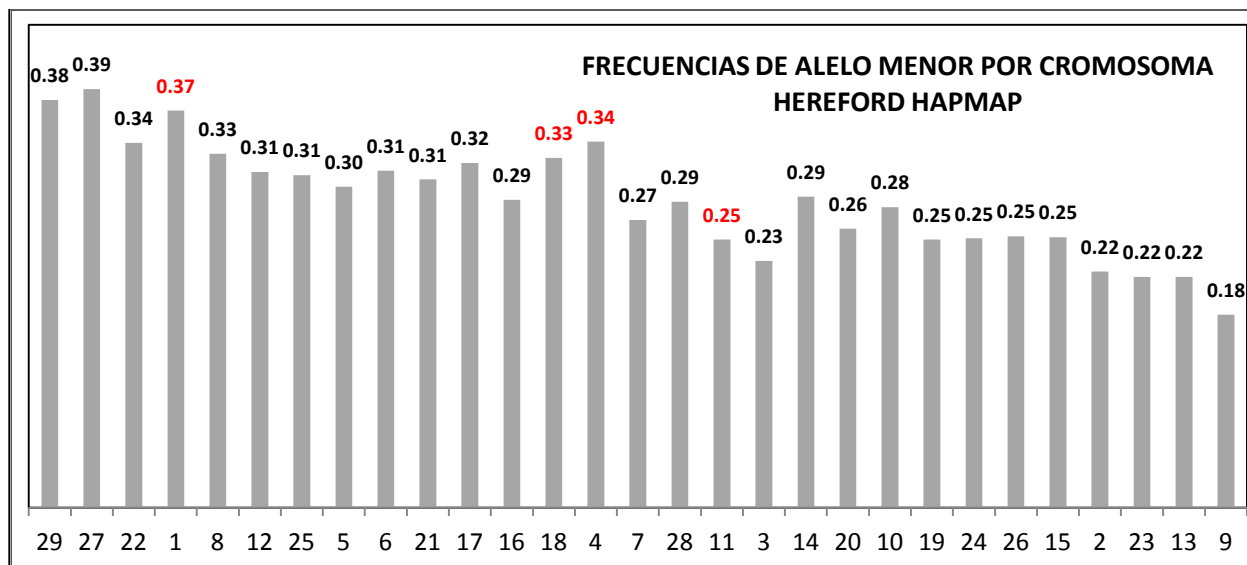


Gráfico 4.3.c. MAF para el Hereford del HapMap.

Debido a que la H_e depende directamente de la frecuencia relativa del alelo menor, dado que participa en su cálculo de estimación, la H_o resulta una mejor indicadora de la cantidad de variabilidad de alelos en la población muestreada. En el **Gráfico 4.3.d** se presenta la contribución de cada marcador a la H_o resaltando que 36% de los marcadores del panel Hereford México tiene H_o por arriba de 0.46 , casi la mitad de la muestra de animales de la población tienen una alta H_o .

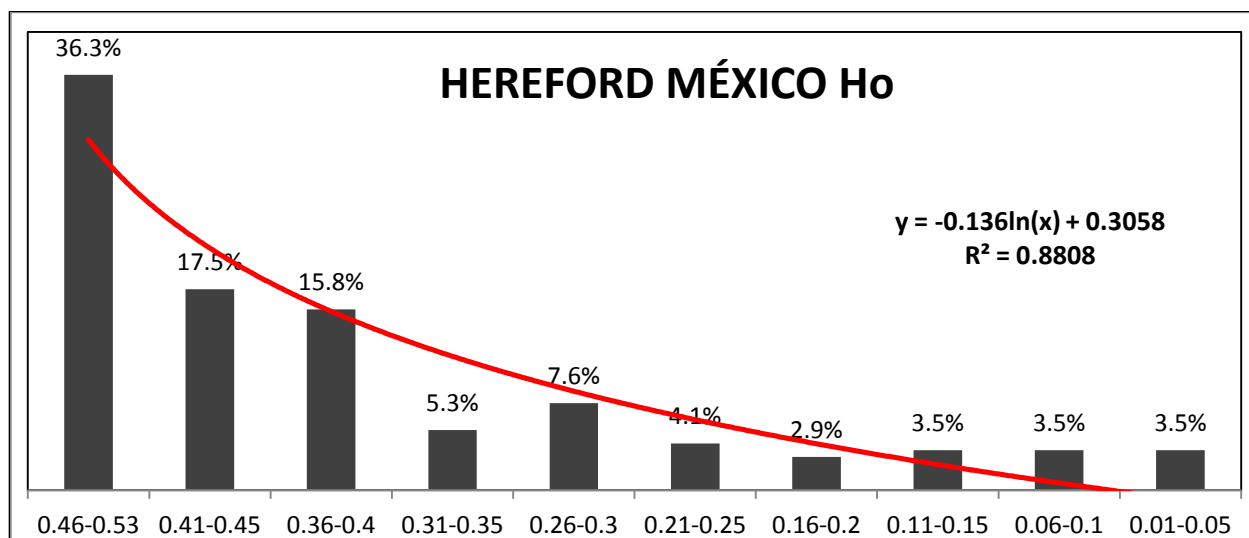


Gráfico 4.3.d. Heterosigocidad Observada en Hereford México.

En cambio en el panel internacional muestreado por el HapMap Consortium se aprecia una diferencia alta de 8.48% en la frecuencia de SNP's con H_o entre 0.41 a 0.45, tal como lo muestra el **Gráfico 4.3.f** , además un total de 55.21% de los SNP's tienen H_o entre 0.45 hasta 0.53, sin embargo la pendiente de la línea de regresión logarítmica es muy similar en ambos casos (-0.136, -0.131), coincidiendo la tendencia a lo largo de los 29 cromosomas en ambas poblaciones.

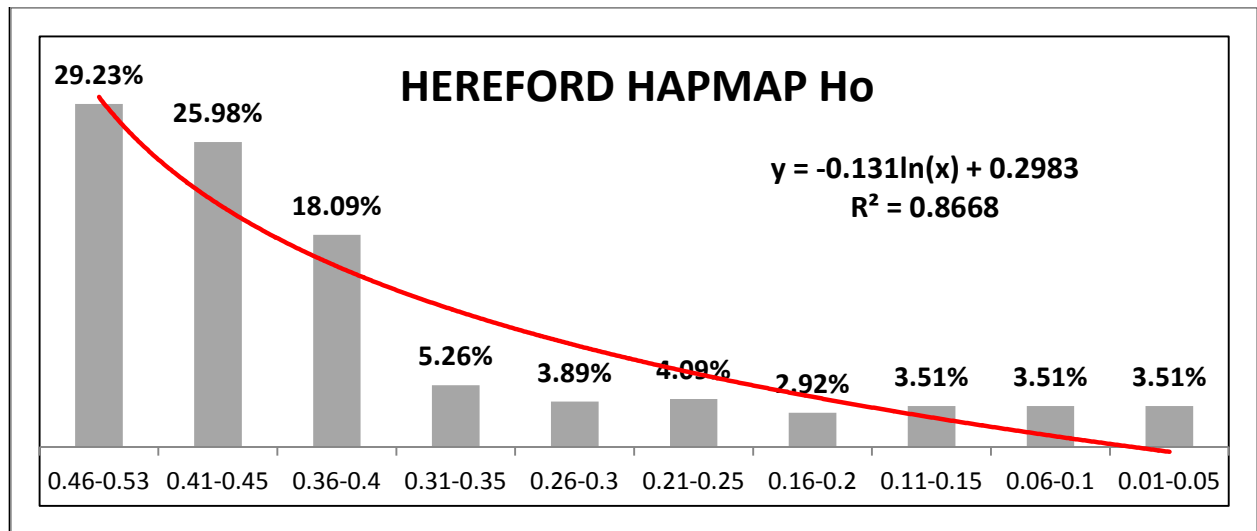


Gráfico 4.3.e. Heterosigicidad Observada en Hereford HapMap.

4.4 Resultados del Desequilibrio por Ligamiento

Únicamente se consideró el LD dentro de cada cromosoma, y se identificaron aquellos pares de marcadores con los valores más altos de asociación r^2 , sin embargo, la condición de las distancias entre ellos dificultó su poder de discriminación, únicamente 6 pares de marcadores mostraron una leve presencia del fenómeno, en el **Gráfico 4.4.a** se puede observar la caída del LD aún a distancias tan grandes, pero los valores de r^2 son irrelevantes para el caso.

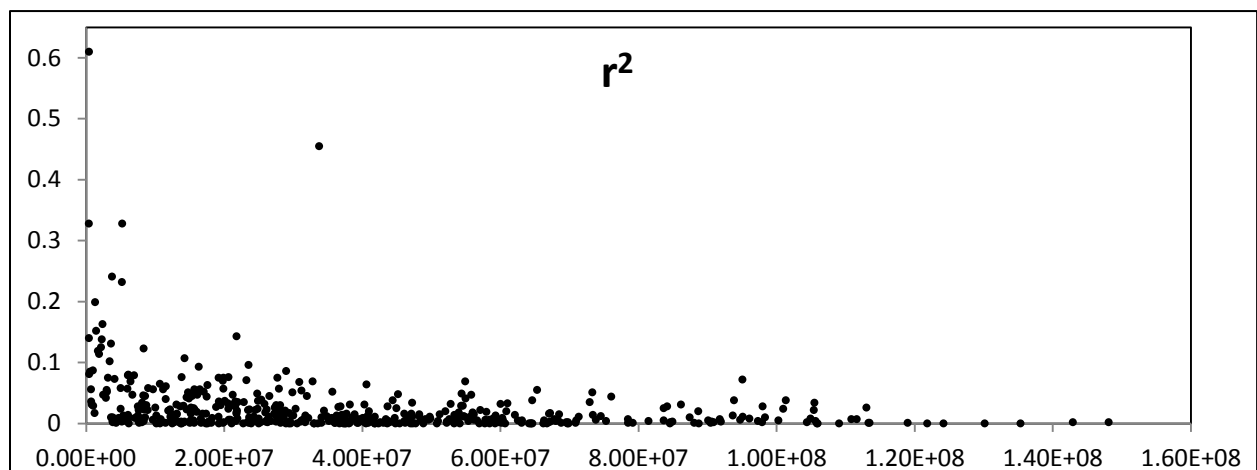


Gráfico 4.4.a. Gráfico de caída de LD por distancia en Bases Par (bp)

En cuanto a los resultados de D' como estimador de LD, en el **Gráfico 4.4.b** puede apreciarse mejor la función de caída y su ecuación de regresión logarítmica.

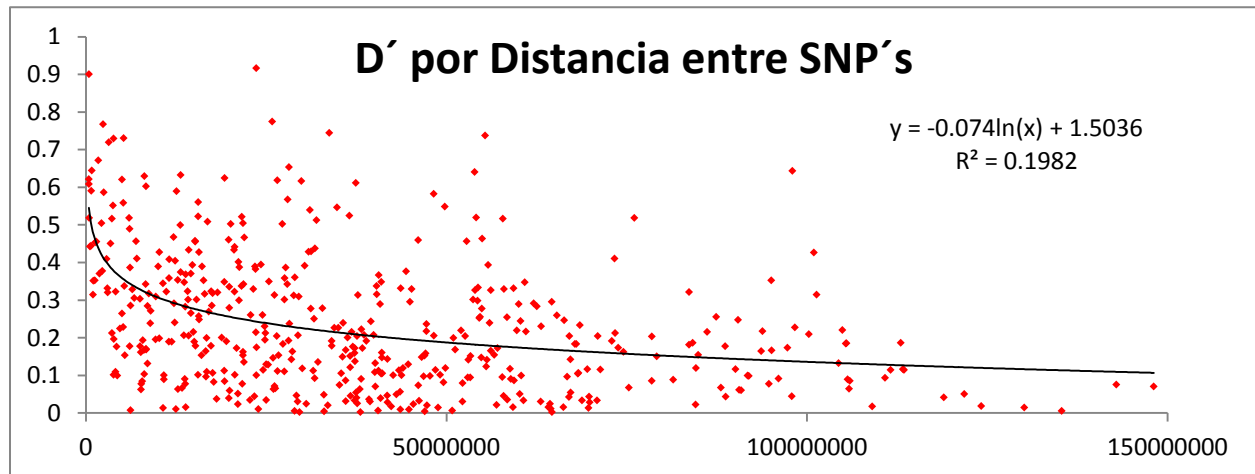


Gráfico 4.4.b. Caída de LD por la distancia en Bases Par (bp)

4.5 Selección de Marcadores significativos para Paternidad México

Basándose en la H_o como la mejor fuente de discriminación para la informatividad de los SNP's se seleccionaron desde el Pareto de H_o y sin importar el cromosoma de origen, 80 marcadores quedaron como los candidatos viables para iniciar el proceso de estimación del Poder Acumulado de Exclusión Q . El **Gráfico 4.5.a** muestra los valores de H_o .

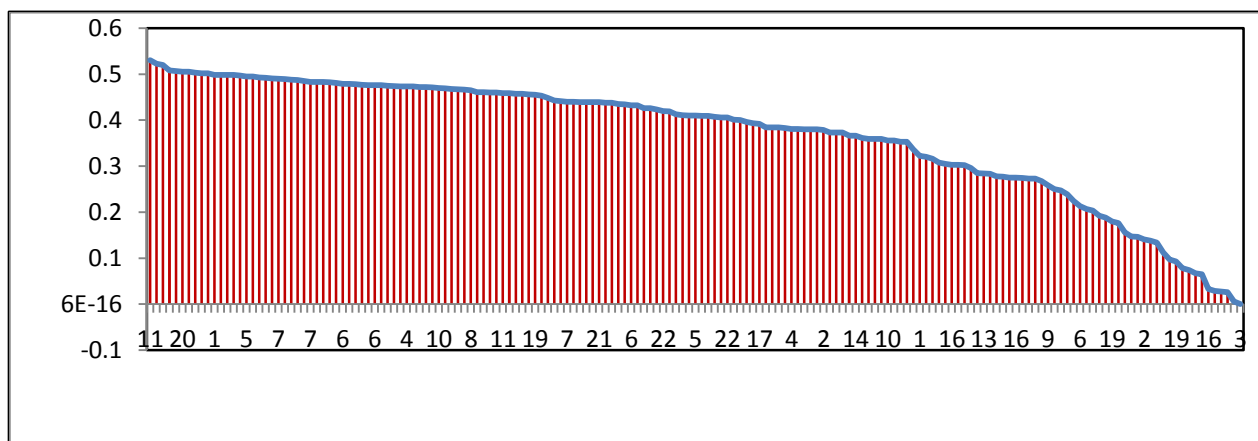


Gráfico 4.5.a Heterosigocidad Observada en el Ganado Hereford México.

CAPÍTULO V. DISCUSIÓN DE RESULTADOS

Tal como se puede apreciar en el **Gráfico 5.1.a** el Poder de Exclusión Acumulado (Q) de la prueba de paternidad, aumenta al adicionar la cantidad de SNP's en el algoritmo de computo, Fernández, M.E. *et al* (2013)[22], demostraron las ventajas del uso de SNP's en ganado Angus Argentino y sus resultados sugieren en ese caso un mínimo de 24 para igualar los resultados encontrados con el set de STR's (Short Tandem Repetitions) recomendado por el ISAG.

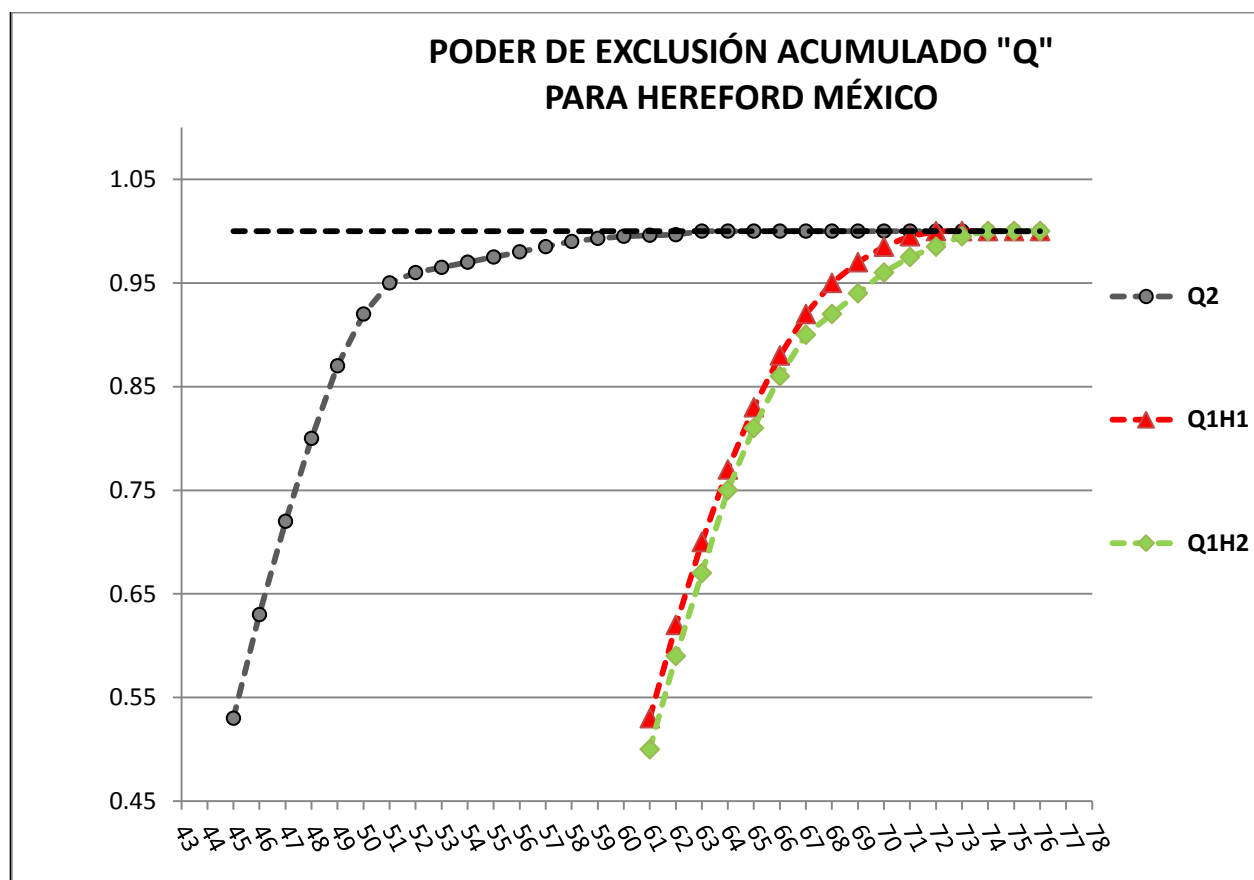


Gráfico 5.1.a Poder de Exclusión de la Prueba de Paternidad vs cantidad SNP's

En el estudio desarrollado por Werner, F.A.O. et al [23] se encontraron 37 SNP's como una base fiable en un Panel de Paternidad para razas de bovinos lecheras Europeas. Schütz, E. et al [24] trabajaron con la raza Holstein de Alemania y demostraron una reducción confiable de 176 a 78 SNP's. Sin embargo las características de los sistemas de producción y de crianza varían significativamente de los usados para la producción de Carne, Van-Eennaam A.L. et al [25] comprobaron paneles de 36 y 40 SNP's en Ranchos comerciales grandes en Estados Unidos de Norte América, incluso utilizaron la información de los marcadores para la evaluación genética del ganado (EBV's y EPD's).

Los resultados encontrados en el panel de paternidad aplicado a la raza Hereford Mexicana, demuestran que con un mínimo de 63 SNP's seleccionados con el procedimiento aquí descrito, es posible replicar resultados equivalentes al panel de 206 SNP's adquirido para el estudio, cuando se tiene información de ambos padres, en el caso de solamente contar con información de uno de ellos la potencia de la prueba requiere de mayor cantidad de marcadores para alcanzar un nivel máximo 72 para el Haplotipo 1 y 74 para el Haplotipo 2, de forma general 75 marcadores son los recomendables.

Considerando que los marcadores SNP's son los seleccionados por ISAG e ICAR para los estudios de paternidad de cualesquier bovino a nivel mundial, se asume que son altamente informativos y que algunos de ellos aplican mejor a razas *Bos Taurus* o *Bos Indicus*, por eso se podría explicar fácilmente la cantidad de información redundante en el panel internacional.

CAPÍTULO VI. CONCLUSIONES Y TRABAJO FUTURO

Dada la estructura de variabilidad alélica de la muestra se ha demostrado que con 63, 72 y 74 marcadores SNP's; con ambos padres putativos, con un padre putativo y con una madre putativa, respectivamente, es factible y viable realizar pruebas de paternidad en ganado Hereford Mexicano, sin demerito alguno en las probabilidades de exclusión o los Niveles de Confianza de la prueba. Permitiendo así sugerir un panel con 75 marcadores SNP's suficientes para cumplir con ese propósito, equivalentes al panel internacional general de 206 SNP's.

En un futuro se pretende continuar con otras razas de México, integrando una red de investigadores para la zona Noroeste del país, donde actualmente se usa ganado Angus, Brangus, Limousin, Charolais y Criollo.

BIBLIOGRAFÍA

- [1] FIRA (2017). Carne de bovino 2017. Panorama agroalimentario. Dirección de investigación y evaluación económica y sectorial.
- [2] USDA-FAS (2016). México, Livestock and Products Annual. October, 2016.
- [3] Wilcox, C.J., D.W. Webb, and M.A. DeLorenza, *Genetic Improvement of Dairy Cattle*, EDIS, Editor. 1992, University of Florida IFAS extension: Gainesville. p. 11.
- [4] Robinson, G.K., *That BLUP is a Good Thing: The Estimation of Random Effects*. Statistical.
- [5] Rogers, G.W., et al. *Four ways genomic selection will change dairy cattle genetic improvement in the near future*. Progressive Dairyman [cited 2010 10/14/2010]; Available from: http://www.progressivedairy.com/pd/features/2008/0808/0808_rogers.html.
- [6] Henderson, C.R., *Best linear unbiased estimation and prediction under a selection model*. Biometrics, 1975. **31**(2): p. 423-47.
- [7] Long, T.E., *Estimating Genetic Merit*, in *Swine Genetics*, NSIF, Editor. 2003, Purdu University. p. 4.
- [8] Smith, C., *Improvement of metric traits through specific genetic loci*. Animal Production, 1967. **9**: p. 10.
- [9] Soller, M. and J.S. Beckmann, *Genetic polymorphism in variety identification and genetic improvement*. Theoretical and Applied Genetics, 1983. **67**(1): p. 9.
- [10] Shrimpton, A.E. and A. Robertson, *The isolation of polygenic factors controlling bristle score in Drosophila Melanogaster*. Genetics, 1988. **148**: p. 14.
- [11] Hayes, B. and M.E. Goddard, *The distribution of the effects of genes affecting quantitative traits in livestock*. Genet Sel Evol, 2001. **33**(3): p. 209-29.
- [12] Sanna, S., et al., *Common variants in the GDF5-UQCC region are associated with variation in human height*. Nat Genet, 2008. **40**(2): p. 198-203.
- [13] VanRaden, P.M., et al., *Invited review: reliability of genomic predictions for North American Holstein bulls*. J Dairy Sci, 2009. **92**(1): p. 16-24.
- [14] Hayes, B.J. and M.E. Goddard, *Technical note: prediction of breeding values using marker-derived relationship matrices*. J Anim Sci, 2008. **86**(9): p. 2089-92.
- [15] Meuwissen, T.H., B.J. Hayes, and M.E. Goddard, *Prediction of total genetic value using genome-wide dense marker maps*. Genetics, 2001. **157**(4): p. 1819-29.
- [16] Elsik, C.G., et al., *The genome sequence of taurine cattle: a window to ruminant biology and evolution*. Science, 2009. **324**(5926): p. 522-8.
- [17] Van Tassell, C.P., et al., *SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries*. Nat Methods, 2008. **5**(3): p. 247-52.
- [18] Gibbs, R.A., et al., *Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds*. Science, 2009. **324**(5926): p. 528-3
- [19] Villa-Angulo, R., et al., *High-resolution haplotype block structure in the cattle genome*. BMC Genet, 2009. **10**: p. 19.
- [20] Wikipedia. Mutation rate. Consultado del día 31 de mayo de 2018 de: https://en.wikipedia.org/wiki/Mutation_rate.
- [21] Wikipedia. Luria-Delbruck Experiment. Consultado el día 31 de mayo de 2018 de: https://en.wikipedia.org/wiki/Luria%E2%80%93Delbr%C3%BCck_experiment

- [22] Fernandez, M. *et al* (2013). *Comparison of the effectiveness of microsatellites and SNP panels for genetic identification, traceability and assessment of parentage in an inbred Angus herd*. Genetics and molecular biology. **36**. p 185-91.
- [23] Werner, F.A.O. *et al* (2014) *Detection and characterization of SNPs useful for identity control and parentage testing in major European dairy breeds* Animal Genetics, **35**, p 44–49.
- [24] Schütz E, Brenig B. (2015) *Analytical and statistical consideration on the use of the ISAG-ICAR-SNP bovine panel for parentage control, using the Illumina BeadChip technology: example on the German Holstein population*. Genetics, Selection, Evolution : GSE.; **47**(1):3.
- [25] Van Enneman E.L. *et al* (2014) *DNA-based paternity analysis and genetic evaluation in a large, commercial cattle ranch setting* J. Anim. Sci. **85** p 3159–3169 doi:10.2527/jas.2007-0284 p 3159-3169.

ANEXOS

Chrom	r²	D'	MAF	Ho	He	HWpval
1	0.02371	0.15038	0.35250	0.39240	0.43110	0.15442
2	0.01681	0.22175	0.24733	0.34633	0.35433	0.40183
3	0.03427	0.21913	0.28243	0.33700	0.34900	0.50520
4	0.03273	0.32707	0.30833	0.40150	0.41017	0.53030
5	0.04224	0.28476	0.31957	0.38700	0.39700	0.37363
6	0.01536	0.17418	0.31938	0.37613	0.38588	0.23795
7	0.01993	0.22196	0.29313	0.37838	0.38825	0.39909
8	0.02783	0.24750	0.34175	0.38350	0.40375	0.31243
9	0.02050	0.28883	0.18750	0.25800	0.26975	0.39245
10	0.01307	0.23143	0.26725	0.33838	0.35113	0.31334
11	0.01907	0.21631	0.28840	0.37700	0.38060	0.32965
12	0.04330	0.28310	0.34140	0.42720	0.44140	0.29968
13	0.01760	0.30487	0.19200	0.24867	0.28200	0.27795
14	0.05667	0.32800	0.27833	0.38100	0.39033	0.28800
15	0.01900	0.34150	0.26250	0.32875	0.35925	0.23635
16	0.01993	0.24475	0.31013	0.36250	0.37525	0.41663
17	0.03529	0.19132	0.31100	0.40338	0.41150	0.28843
18	0.03427	0.28207	0.30917	0.38317	0.40400	0.36555
19	0.02319	0.33186	0.26700	0.33829	0.34243	0.31469
20	0.01550	0.26120	0.27160	0.34780	0.36420	0.42450
21	0.02948	0.22205	0.31486	0.41657	0.42271	0.40777
22	0.02553	0.20627	0.36950	0.41900	0.44917	0.20845
23	0.02200	0.35000	0.23533	0.28000	0.29033	0.43467
24	0.01700	0.26967	0.26600	0.34900	0.35600	0.27813
25	0.02733	0.26500	0.33267	0.43367	0.42467	0.38320
26	0.02700	0.22300	0.26500	0.35675	0.36850	0.30800
27	0.02400	0.18467	0.39867	0.45067	0.47467	0.18787
28	0.01940	0.27390	0.29000	0.33160	0.35780	0.35604
29	0.08533	0.31367	0.42225	0.44625	0.46975	0.28973
Genome	0.02482	0.23555	0.29805	0.36875	0.38340	0.33611
AVGxChr	0.02784	0.25725	0.29810	0.36827	0.38293	0.33503

SNP	CROMOSOMA	POSICION	SNP	CROMOSOMA	POSICION
ARS-USMARC-569	17	30360942	AY943841	1	138583183
AY761135	1	127426647	BAC-19454	16	13695211
AY776154	2	26997623	BAC-27364	1	8437530
AY841151	2	45832887	BAC-35552	2	12823369
AY842473	3	49703647	BES10_Contig566_926	24	6408329
AY842474	3	51976646	BTA-100621	22	25598780
AY842475	4	20181749	BTA-103548	20	30928803
AY842475	4	20181749	BTA-107561	6	29556343
AY844963	5	98102349	BTA-109898	21	27927781
AY844963	5	98102349	BTA-112834	4	72987997
AY849380	6	90562665	BTA-11701	11	61989432
AY849381	6	23562312	BTA-122724	19	4670556
AY850194	8	59996431	BTA-122725	13	83437845
AY851162	11	46411100	BTA-136127	21	62028408
AY851163	11	103047474	BTA-163266	27	10764825
AY853302	13	47397987	BTA-30857	12	80058952
AY853303	13	75383374	BTA37062	15	51528617
AY856094	29	9160939	BTA-40051	16	72931875
AY857620	16	67607778	BTA-40319	16	19882683
AY858890	17	29936157	BTA-41809	17	67216420
AY860426	17	56512519	BTA-47447	2	41938704
AY863214	18	46647177	BTA-59790	25	27912102
AY914316	18	48812014	BTA-63031	27	41049981
AY916666	19	44799390	BTA-73768	5	67031023
AY919868	20	46066109	BTA-80741	7	13279972
AY929334	23	7219975	BTA-83292	22	54103673
AY937242	23	27306795	BTA-92021	11	9385008
AY939849	24	56415794	BTA-92999	11	36713471
AY941204	25	14683151	BTA-97415	6	44461834
AY942198	25	40851475	BTB-00188171	4	57787437
DQ786757	2	111155237	BTB-00394801	9	59383405
UA-IFASA-6532	22	21770149	BTB-00420215	10	44920804

SNP	CROMOSOMA	POSICION	SNP	CROMOSOMA	POSICION
BTB-00818821	21	40408803	DQ786758	7	18454636
BTB-01057979	1	114085424	DQ786759	7	94259472
BTB-01285245	1	63277846	DQ786760	10	14673344
BTB-01371672	15	50566485	DQ786761	10	44103665
BTB-01416427	19	52911677	DQ786762	10	81572252
BTB-01478115	4	42103377	DQ786763	12	11824653
BTB-01902778	6	114165533	DQ786764	12	25668974
BTB1980499	12	61345454	DQ789028	6	46936182
DQ381152	17	17616950	DQ832700	12	80629629
DQ381153	1	3249057	DQ837643	11	66341589
DQ404149	1	99314925	DQ837644	8	88974063
DQ404150	1	59409838	DQ837645	11	24553007
DQ404151	1	151349514	DQ837646	11	1703612
DQ404152	2	5306838	DQ839235	3	116448759
DQ404153	29	44756502	DQ846688	5	119261609
DQ422949	2	65069037	DQ846688	5	119261609
DQ422950	3	3129167	DQ846690	14	10171919
DQ435443	3	58040470	DQ846691	14	48380429
DQ451555	1	29524658	DQ846691	14	48380429
DQ468384	5	113137320	DQ846692	14	80082923
DQ470475	5	7651053	DQ846692	14	80082923
DQ485413	4	94176209	DQ846693	16	9855276
DQ489377	3	98188384	DQ846693	16	9855276
DQ500958	5	27825118	DQ846694	16	33521338
DQ647186	4	17200594	DQ846695	16	80747557
DQ647187	3	21146877	DQ866817	15	38078775
DQ647188	4	118165814	DQ866817	15	38078775
DQ647189	5	63273386	DQ866818	15	79187295
DQ647190	6	13897068	DQ888309	7	8272794
DQ650635	7	55116289	DQ888309	7	8272794
DQ650636	8	28799249	DQ888310	17	887216
DQ674265	8	106174871	DQ888311	19	8505317

SNP	CROMOSOMA	POSICION	SNP	CROMOSOMA	POSICION
DQ888312	19	36437188	EF093510	22	46211533
DQ888313	20	17837675	EF093511	21	26620013
DQ916057	7	81591587	EF093512	27	15141319
DQ916058	8	1554706	EF141102	27	37513923
DQ916058	8	1554706	EF150946	26	13229219
DQ916059	18	23426214	EF164803	19	55174260
DQ984825	10	98230479	Hapmap59876	2	13853729
DQ984826	14	27751888	NGS-10035	26	39952154
DQ984827	10	55611885	NGS-101456	7	82825197
DQ984828	20	676757	NGS-102169	27	12930436
DQ990832	22	11038205	NGS-106015	6	61199572
DQ990833	24	15447771	NGS-111053	5	72264603
DQ990834	26	8221270	NGS-111076	21	21220448
DQ995976	21	3088886	NGS-111114	10	84446642
DQ995977	24	1854953	NGS-112094	13	69187742
EF026084	19	15345312	NGS-112325	12	79643103
EF026085	21	65198296	NGS-11383	17	12891997
EF026086	28	35331560	NGS-114006	25	20114319
EF026087	13	1982209	NGS-117319	16	12238539
EF028073	18	1839733	NGS-117322	11	77884744
EF034080	29	28647816	NGS-118319	17	9279323
EF034081	13	25606469	NGS-118340	18	26844474
EF034082	22	56526462	NGS-119662	6	118292328
EF034083	25	3126438	NGS-14740	9	14436985
EF034084	27	21480570	NGS-15506	5	88334676
EF034085	28	5913226	NGS-15731	7	96936145
EF034086	26	38233337	NGS-24419	22	11756783
EF034087	28	16097749	NGS-26517	21	52139183
EF042090	15	21207529	NGS-27577	7	26066004
EF042091	28	44261945	NGS-31640	9	52430564
EF089234	23	50884052	NGS-31807	10	61684704
EF093509	22	22573121	NGS-36513	19	17236707

SNP	CROMOSOMA	POSICION
NGS-38423	3	12709854
NGS-38620	18	64382705
NGS-39978	2	5757355
NGS-42283	11	95543411
NGS-42505	14	62478242
NGS-55943	15	64628706
NGS-57711	19	40692429
NGS-67146	2	118773633
NGS-70946	10	14574453
NGS-72471	7	103375625
NGS-76191	11	62732451
NGS-76330	28	33449126
NGS-86662	1	40206017
NGS-93119	9	28149131
NGS-96125	20	58449212
rs29010990	8	103137122
rs29012198	4	110776624
rs29012632	9	83320792
rs29018046	2	13853729
rs29023153	18	6776532
rs29023471	7	81668727
SCAFFOLD86180_8791	23	41700829
SCAFFOLD90561_9460	20	2289049
UA-IFASA-5034	29	29052623
UA-IFASA-6532	22	21770149