

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

MAESTRÍA Y DOCTORADO EN CIENCIAS E INGENIERÍA



**“DETECCIÓN DE SEÑALES DE SELECCIÓN EN GANADO BOVINO
UTILIZANDO MARCADORES SNP DE ALTA DENSIDAD”**

**TESIS QUE PARA OBTENER EL GRADO DE:
MAESTRO EN CIENCIAS**

PRESENTA

JOSÉ LUIS BECERRA BUENROSTRO

DIRECTOR

DR. RAFAEL VILLA ANGULO

CODIRECTOR

DR. RICARDO SALOMÓN TORRES

MEXICALI, B.C.

MARZO 2020

AGRADECIMIENTOS

Desde lo más profundo de mi ser quiero agradecer el apoyo que he recibido de diversas fuentes a lo largo de esta etapa:

A mi querida familia, gracias por estar siempre a mi lado y por impulsarme a crecer como persona; por toda la paciencia, comprensión y amor recibido. Son ustedes con quien comparto este logro.

A mi director de tesis, el Dr. Rafael Villa, por brindarme la oportunidad de continuar con mis estudios de posgrado y de pertenecer a su grupo de investigación; por toda la guía, el apoyo y la paciencia durante esta etapa, así como los conocimientos y las innumerables lecciones de vida transmitidas.

A mi codirector de tesis, el Dr. Ricardo Salomón Torres, le agradezco su paciencia, orientación y aportes durante este proceso. A los miembros del sínodo y a los múltiples revisores de esta tesis, muchas gracias por su atención.

A todos los integrantes del cuerpo académico de Bioinformática y Física Aplicada: compañeros, amigos y hermanos. Gracias por todo el apoyo y convivencia, dentro y fuera del laboratorio. Fue un placer compartir esta experiencia con todos ustedes.

Al Instituto de Ingeniería y a la Universidad Autónoma de Baja California, por aceptarme en su programa de posgrado MyDCI y brindarme el espacio y los recursos necesarios en esta etapa; agradezco también el apoyo de todo el personal directivo, administrativo y académico que en ellos labora, sin el cual esto no hubiera sido posible.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por brindarme el apoyo económico a lo largo de mis estudios de posgrado y por su trascendental tarea de hacer de este país uno mejor.

ÍNDICE

CAPÍTULO 1	INTRODUCCIÓN.....	1
1.1	COMPOSICIÓN Y ESTRUCTURA DE LA INFORMACIÓN GENÉTICA	1
1.1.1	ÁCIDOS NUCLEICOS	1
1.1.2	ELEMENTOS GENÉTICOS.....	3
1.1.3	ALELOS	3
1.1.4	POLIMORFISMO DE NUCLEÓTIDO SIMPLE	5
1.1.5	SNPs COMO MARCADORES MOLECULARES	6
1.1.6	DESEQUILIBRIO DE LIGAMIENTO.....	7
1.1.7	GENOTIPO, FENOTIPO Y SU RELACIÓN CON EL AMBIENTE.....	8
1.1.8	LOCUS DE UN CARÁCTER CUANTITATIVO	8
1.2	BOVINOS	10
1.2.1	ORIGEN Y DEMOGRAFÍA HISTÓRICA	10
1.2.2	EL GENOMA BOVINO Y EL PROYECTO DE HAP MAP	12
1.2.3	TECNOLOGÍAS DE GENOTIPIFICACIÓN	13
1.2.4	SNP CHIPS	13
1.2.5	ANOTACION DEL GENOMA BOVINO UMD3.1.....	14
1.3	PLANTEAMIENTO DEL PROBLEMA	15
1.4	JUSTIFICACIÓN.....	15
1.5	OBJETIVOS.....	16
1.5.1	OBJETIVO GENERAL.....	16
1.5.2	OBJETIVOS ESPECÍFICOS	16
CAPÍTULO 2	MARCO TEÓRICO	17
2.1	TEORÍA DE LA EVOLUCIÓN	17
2.1.1	GENÉTICA DE POBLACIONES Y EVOLUCIÓN	17
2.1.2	TEOTRIA DE LA COALESCENCIA	17
2.1.3	SELECCIÓN NATURAL	18
2.1.4	TEORÍA NEUTRA DE LA EVOLUCIÓN.....	18
2.1.5	VARIACIÓN GENÉTICA Y SELECCIÓN	19
2.1.6	TIPOS DE SELECCIÓN	19

2.1.7	DERIVA GENÉTICA	20
2.1.8	EFEECTO HICHHIKING Y BARRIDOS DE SELECCIÓN	21
2.2	MÉTODOS PARA LA DETECCIÓN DE SEÑALES DE SELECCIÓN.....	22
2.2.1	REDUCCIÓN LOCAL DE DIVERSIDAD GENÉTICA	25
2.2.2	CAMBIOS EN EL ESPECTRO DE FRECUENCIA DE ALELOS	26
2.2.3	SEGMENTOS LARGOS DE HAPLOTIPOS.....	28
2.2.4	DIFERENCIACIÓN DE POBLACIONES	30
2.3	CONSIDERACIONES Y SOFTWARE PARA EL ANÁLISIS DE SS.....	31
2.4	LA EXPECTACIÓN DE NEUTRALIDAD.....	33
2.4.1	HIPÓTESIS NULA, ALTERNA Y VALORES P EN SS	33
2.5	HERRAMIENTAS BIOINFORMÁTICAS.....	34
2.5.1	BASES DE DATOS E INTERFASES	34
2.5.2	LENGUAJE DE PROGRAMACIÓN PERL.....	35
2.5.3	ENTORNO DE PROGRAMACIÓN R.....	35
CAPÍTULO 3 METODOLOGÍA.....		37
3.1	DESCRIPCIÓN DE LOS DATOS	37
3.1.2	ORIGEN DE LOS DATOS	37
3.1.3	FORMATO DE LOS DATOS.....	39
3.2	PREPROCESAMIENTO DE LA INFORMACIÓN	40
3.2.1	ESTRUCTURA DE DATOS.....	40
3.2.2	FILTROS DE CALIDAD.....	40
3.2.3	ARCHIVOS DE ENTRADA A LA PAQUETERÍA “rehh”.....	41
3.3	PROCESAMIENTO DE LA INFORMACIÓN	42
3.3.1	CÁLCULO DE LOS ESTADÍSTICOS DE iHS	42
3.3.2	ESTANDARIZACIÓN DE iHS	43
3.3.3	VALORES p	43
3.3.4	REGIONES GENÓMICAS SIGNIFICATIVAS	44
3.3.5	INFORMACIÓN BIOLÓGICA DE INTERÉS.....	44
CAPÍTULO 4 ANÁLISIS Y DISCUSIÓN DE RESULTADOS		46
4.1	RESOLUCIÓN DEL ESTUDIO.....	46
4.2	ESTADÍSTICOS DE iHS	47
4.3	REGIONES GENÓMICAS Y SEÑALES DE SELECCIÓN.....	54

4.4	CROMOSOMAS	55
4.5	ELEMENTOS GENÉTICOS DETECTADOS.....	58
4.5.1	GENES.....	58
4.5.2	RASGOS DE PRODUCCIÓN ASOCIADOS	59
CAPÍTULO 5 CONCLUSIONES Y TRABAJO FUTURO.....		68
5.1	CONCLUSIONES	68
5.2	TRABAJO FUTURO.....	70
REFERENCIAS.....		71
ANEXOS		74
ANEXO 1 COMPENDIO DE MARCADORES INICIALES		74
ANEXO 2 COMPENDIO DE MARCADORES RETENIDOS.....		76
ANEXO 3 DISTRIBUCIÓN DE VALORES P_{iHS}		78
ANEXO 4 NÚMERO DE REGIONES SIGNIFICATIVAS.....		88
ANEXO 5 REGIONES SIGNIFICATIVAS.....		90
ANEXO 6 ELEMENTOS GENÉTICOS DETECTADOS.....		95

LISTA DE FIGURAS

Figura 1 Estructura y composición del ADN.....	2
Figura 2 Estructura y composición del ARN.....	2
Figura 3 Cromosomas homocigotos y heterocigotos.....	4
Figura 4 Polimorfismo de Nucleótido Simple.	5
Figura 5 Desequilibrio de ligamiento en torno a una mutación ancestral.....	7
Figura 6 Cariotipo bovino.....	12
Figura 7 SNPchips para genotipificación en el genoma bovino.	14
Figura 8 Efecto de Hitchhiking.....	22
Figura 9 Descripción de la muestra.	38
Figura 10 Extracto del compendio de haplotipos codificados en “Alelos A/B”.....	39
Figura 11 Extracto del compendio de haplotipos codificados en bases nucleotídicas “ACGT”..	40
Figura 12 Extracto de los archivos de entrada a la paquetería de rehh.....	41
Figura 13 Extracto del objeto res.scan.....	42
Figura 14 Extracto del objeto res.ihs	43
Figura 15 Distribución de los valores de $ iHS $ en el genoma	53
Figura 16 Coincidencias en cromosomas con señales de selección	57
Figura 17 Proporción de elementos genéticos anotados en las regiones significativas	59
Figura 18 Distribución de rasgos de producción asociados según propósito de crianza	65

LISTA DE TABLAS

Tabla 1 Principales enfoques para la detección de señales de selección a partir de información genómica.....	24
Tabla 2 Softwares disponibles para el análisis de señales de selección	32
Tabla 3 Linajes y número especímenes que componen la muestra	37
Tabla 4 Distancias promedio entre marcadores adjuntos en cromosomas autosomales.....	46
Tabla 5 Proporción de Regiones significativas.....	54
Tabla 6 Distribución de regiones significativas.....	56
Tabla 7 Genes y rasgos asociados.....	60
Tabla 8 Referencia cruzada de genes.....	66

RESUMEN

Las señales de selección (SS) son definidas como regiones en el genoma que albergan variaciones funcionales importantes en secuencias de ADN y que han estado sujetas, ya sea a presión natural o artificial, dejando un patrón especial en dichas regiones. Existen diversos métodos para detectar SS, entre ellos los basados en la Homocigosidad Haplotípica Extendida, EHH (Extended Haplotype Homozygosity por sus siglas en inglés) definida como la probabilidad de que dos cromosomas elegidos al azar que portan el mismo alelo en un SNP particular (SNP focal), sean idénticos por descendencia para todo un intervalo (bloque), desde el SNP focal a una distancia dada en torno a él. Una extensión de EHH es el estadístico iHS (Integrated Haplotype Score, por sus siglas en inglés), que compara EHH entre los alelos ancestrales y derivados en una población.

El objetivo de este trabajo de tesis fue detectar la presencia de señales de selección positiva en 19 linajes bovinos, mediante el análisis de iHS en haplotipos de alta densidad. La información genética provenía de 507 especímenes de ganado bovino cuyos linajes pertenecen a las subespecies: *Bos taurus* y *Bos indicus*.

Después de aplicar los filtros de calidad, el número promedio de marcadores retenidos por cromosoma fue de 546,961 SNPs, lo que equivale al 73.62 % de los SNPs autosomales comprendidos en el BovineHD Genotyping BeadChip.

Para el cálculo de iHS se utilizó la paquetería rehh en el entorno de programación R, se promediaron los valores absolutos de iHS a lo largo de todo el genoma en regiones sin traslape de 500 Kb de longitud ($|iHS|$). Se generaron gráficas de Manhattan para observar la distribución de los valores de $|iHS|$ estandarizados y se establecieron como regiones significativas aquellas que superaron el valor extremo 2.33, que representa un intervalo de confianza del 98 % ($\alpha=0.02$). El

análisis reveló señales de selección positiva de intensidad variable en las 19 razas, con un valor promedio de regiones significativas por raza de 51, lo que corresponde al 1%.

Se analizaron las 5 regiones más significativas para inspeccionar la presencia de elementos anotados utilizando la base de datos *btaurus_gene_ensembl* (BioMart) y la base de datos *AnimalQTLdb* para rasgos de producción asociados. Numerosos elementos genéticos como genes, pseudogenes, transcritos de RNA y QTL fueron detectados, mayormente vinculados a procesos de producción, reproducción, metabolismo y sistema inmune, algunos de los cuales han sido reportados en estudios previos.

ABSTRACT

Selection signatures (SS) are defined as regions of the genome that harbor functionally important sequence variants and therefore are or have been under either natural or artificial selection, leaving special patterns of DNA behind. The SS are detected by different methods, such as the Extended Haplotype Homozygosity (EHH), defined as the probability that two randomly chosen chromosomes carrying the aim core haplotype are identical by descent for the entire interval between a certain locus and the core region. As an extension of the EHH, the Integrated Haplotype Score (iHS), compares EHH between ancestral and derived alleles within a population.

The aim of the thesis was to detect positive selection signatures in high density haplotypes of 19 bovine breeds using the iHS method. The genome information was obtained from 507 animals, belonging to the *Bos taurus* and *Bos indicus* species.

After quality control filters, the average number of markers remained per chromosome was 546, 961 SNPs, which represents the 73.62 % of the autosomal SNPs in the BovineHD Genotyping BeadChip.

The R package 'rehh' was used to compute the iHS values for each SNP across the genome. The genome was split into non-overlapping segments of 500 Kb and averaged, in each segment, the $|iHS|$ scores over the SNPs located in each window. In order to visualize the chromosomal distribution of $|iHS|$, Manhattan plots were created for all breeds and significant regions where those upper the 2.33 threshold value, which represents $\alpha=0.02$. The study revealed positive selection signatures in all breeds, with an average number of 51 regions, which represents the 1%.

The 5 most significant regions were inspected for genome annotations elements and quantitative traits using the *btaurus_gene_ensembl* and *AnimalQTLdb* databases. Several genetics elements

like genes, pseudogenes, RNA transcripts and QTL were detected, related to production, reproduction, metabolism and immune response, some of them previously reported in other studies.

CAPÍTULO 1 INTRODUCCIÓN

1.1 COMPOSICIÓN Y ESTRUCTURA DE LA INFORMACIÓN GENÉTICA

1.1.1 ÁCIDOS NUCLEICOS

Todas las especies de organismos vivos tienen un conjunto de características únicas heredadas que los hace ser diferentes de todas las demás formas de vida. Cada especie tiene su información biológica de desarrollo codificado en moléculas de Ácido Desoxirribonucleico (ADN) presentes en sus células. El ADN es un polímero lineal, no ramificado, cuyas subunidades monoméricas son cuatro nucleótidos químicamente distintos que se pueden unir en cualquier orden en cadenas de cientos, miles o, incluso, millones de unidades de longitud [1].

Cada nucleótido del polímero de ADN está formado por tres componentes:

- 2'-desoxirribosa, que es una pentosa (azúcar compuesta por cinco átomos de carbono).
- Una base nitrogenada que comprende dos grupos, las pirimidinas que incluye a la Citosina (C) y Timina (T), y las purinas que comprende a la Adenina (A) y Guanina (G).
- Un grupo fosfato (PO^4) [2].

EL ADN en las células se encuentra como una molécula de doble hebra con dos cadenas polinucleotídicas cuya secuencia de bases son complementarias. La complementariedad del ADN viene condicionada por el apareamiento específico de las bases púricas y pirimidínicas: Adenina con Timina y Guanina con Citosina. Las dos cadenas de la molécula de doble hebra adoptan una disposición antiparalela y un enrollamiento entre sí para formar una doble hélice como se muestra en la Figura 1 [3].

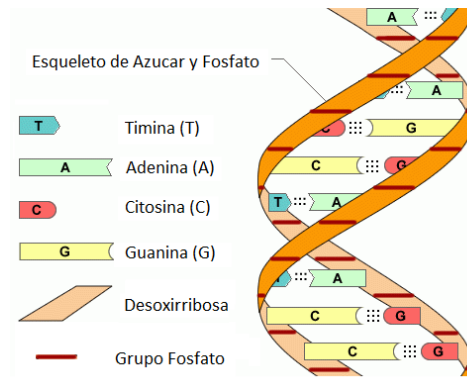


Figura 1 Estructura y composición del ADN. Bloques de construcción y ensamble helicoidal de la molécula de ADN.

El otro tipo de ácido nucleico es el Ácido Ribonucleico (ARN), un polinucleótido similar al ADN, pero con dos diferencias químicas importantes. Primero, el azúcar del nucleótido de ARN es ribosa y, segundo, el ARN contiene Uracilo en lugar de timina. Por lo tanto, los cuatro sustratos para la síntesis de ARN son: adenina, citosina, guanina y uracilo, como se muestra en la Figura 2 [4].

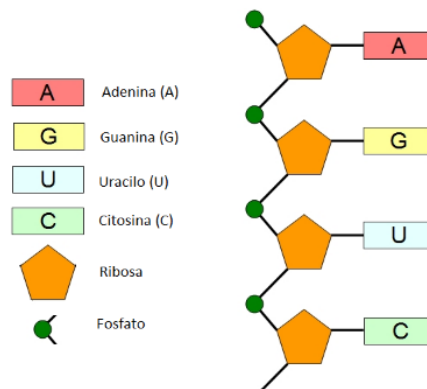


Figura 2 Estructura y composición del ARN. Bloques de construcción de una hebra sencilla de ARN.

Rara vez, las moléculas de ARN tienen más que unos pocos miles de nucleótidos de longitud y, aunque muchas forman pares de bases debido a interacciones a lo largo de la misma molécula, la mayoría son de una sola cadena y no de doble cadena [2].

Existen cuatro clases de moléculas de ARN: el ARN mensajero (ARNm), el cual está implicado directamente en la transferencia de información del ADN a la proteína; el ARN de transferencia (ARNt), que es una molécula más pequeña requerida para la síntesis de proteínas; el ARN ribosómico (ARNr), que es el principal componente de los ribosomas; finalmente, una cuarta clase de ARN consiste en pequeñas moléculas de ARN que participan en varios eventos metabólicos, incluyendo el procesamiento de ARN, mientras que muchas otras moléculas presentan actividad catalítica [5].

1.1.2 ELEMENTOS GENÉTICOS

La unidad física y funcional de la información genética es el gen. Todas las formas de vida contienen genes compuestos por ADN que físicamente están localizados en cromosomas u otras moléculas largas conocidas colectivamente como elementos genéticos [6]. Los genes actúan como instrucciones para fabricar las proteínas. El Proyecto del Genoma Humano ha estimado que los humanos contienen entre 20,000 y 25,000 genes, que varían su tamaño desde unos cuantos cientos de bases de ADN hasta más de 2 millones de bases.

Todas las personas contienen dos copias de cada gen, una copia heredada de cada padre. La mayoría de los genes son iguales en toda la población, pero una pequeña cantidad de genes, menor al 1 %, son ligeramente diferentes entre cada individuo; estas pequeñas diferencias hacen que cada persona presente propiedades físicas únicas [7].

1.1.3 ALELOS

Un locus es una posición fija en un cromosoma y debido a que los organismos diploides, presentan dos copias de cada cromosoma, existen dos valores para esta posición llamados alelos. Un individuo puede ser homocigoto si comparte el mismo valor nucleotídico en una posición en

particular, es decir, sus dos alelos son iguales o ser heterocigoto, si sus alelos presentan valores distintos como se muestra en la Figura 3.

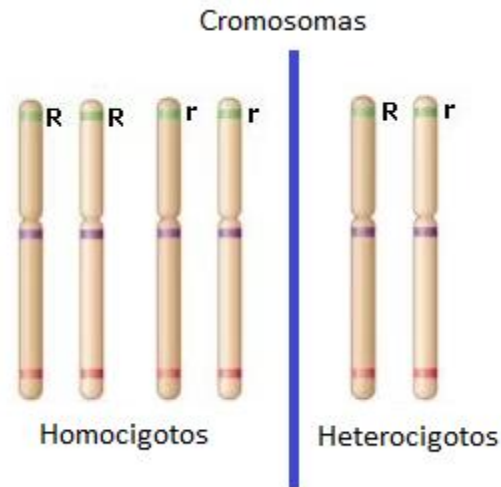


Figura 3 Cromosomas homocigotos y heterocigotos. Los pares de cromosomas de la izquierda presentan el mismo alelo R o r, es decir, son homocigotos; mientras que el par de la derecha presenta alelos diferentes, es decir, son heterocigotos.

Si en una posición en particular, tras analizar múltiples individuos de una población, no es posible encontrar variación en los valores alélicos, se dice que es una posición monomórfica o que presenta monomorfismo; si por el contrario, es posible encontrar valores distintos de alelos en al menos el 1% de la población, se dice que la posición es polimórfica o presenta polimorfismo.

Adicionalmente, un individuo puede ser homocigoto alelo menor, si el valor alélico corresponde al menos frecuente dentro de la población de estudio, comúnmente llamado MAF (Minor Allele Frequency por sus siglas en inglés); cuando presenta el valor alélico más frecuente, se dice que es homocigoto alelo mayor. Las variantes cuya frecuencia es menor al 1% pueden ser descritas como variantes raras.

1.1.4 POLIMORFISMO DE NUCLEÓTIDO SIMPLE

Desde el punto de vista molecular, los tres tipos principales de variación a nivel de ADN son cambios a nivel de nucleótidos, conocidos como Polimorfismo de Nucleótido Simple o SNP (Single Nucleotide Polymorphisms por sus siglas en inglés); inserciones o borrados en secuencias de ADN conocidos por su término en inglés “indels”; y los VNTR (Variable Number of Tandem Repeat por sus siglas en inglés).

Los SNPs son la forma más sencilla y más común de polimorfismo genético, ya que consisten en el cambio de un solo nucleótido en el contexto de una secuencia genética (ver Figura 4). Se considera que ellos determinan la mayor parte de la variabilidad genética entre los individuos, causando muchas de las diferencias fenotípicas observables en ellos.

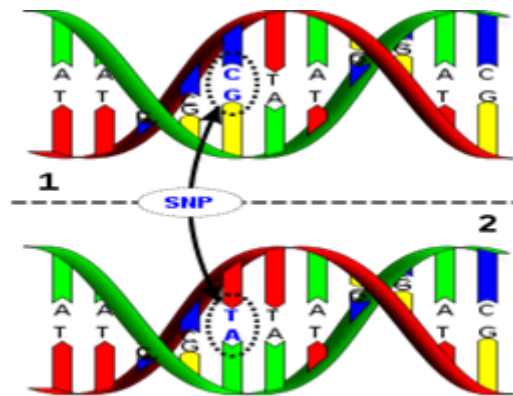


Figura 4 Polimorfismo de Nucleótido Simple. La posición señalada representa un polimorfismo entre las dos cadenas de ADN puesto que presentan valores distintos

Los SNPs se distribuyen de manera heterogénea por todo el genoma y se encuentran tanto en las regiones codificantes (exones), como no codificantes (intrones y regiones promotoras) de los genes, así como en las zonas del genoma donde no se conoce la presencia de genes o elementos con alguna función en particular.

Se han descrito varios millones de SNPs distribuidos por todo el genoma de humanos, animales y plantas. Para que una variación sea considerada como SNP debe ocurrir en al menos 1% de la población. Se estima que su frecuencia media, aunque con grandes variaciones según la especie, longitud de los cromosomas y regiones dentro de ellos, es de 1 SNP cada 500-1,000 nucleótidos [8].

1.1.5 SNPs COMO MARCADORES MOLECULARES

Un marcador genético es un polimorfismo del ADN que se puede detectar fácilmente mediante análisis fenotípico o molecular; puede hallarse dentro de un gen o en ADN con función desconocida, pero presenta la característica de ser heredable.

Con el advenimiento de las técnicas de biología molecular, aparecieron diversos métodos de detección de polimorfismo genético directamente a nivel de ADN: los marcadores genéticos moleculares basados en ADN, los cuales sirven de referencia para detectar la transmisión de un segmento de cromosoma de una generación a otra. El término marcador se utiliza en el sentido de marcador genético y muchas veces se le utiliza como sinónimo de marcador de locus, es decir, un sitio polimórfico que indica el genotipo del individuo que lo lleva, o el genotipo de uno o de varios loci ligados al marcador [9].

Como ventajas sobre otros marcadores moleculares, los SNPs son evolutivamente estables ya que no cambian mucho de una generación a otra, lo que facilita los estudios de poblaciones. También tienen elevada facilidad de análisis, porque admiten diferentes tipos de genotipaje. Existe un amplio abanico de técnicas para ser analizados, donde algunas de ellas pueden ser automatizadas y utilizadas en masa a gran escala [10].

1.1.6 DESEQUILIBRIO DE LIGAMIENTO

El desequilibrio de ligamiento LD (Linkage Disequilibrium por sus siglas en inglés) se refiere a la asociación no aleatoria entre alelos, es decir, cuando alelos en dos o más loci no se segregan de manera independiente. El LD cuantifica el desequilibrio como la diferencia entre la frecuencia observada de dos locus en un haplotipo y la frecuencia que se esperaría si los dos locus se segregaran de manera aleatoria. En la Figura 5 se muestra en color amarillo un cromosoma ancestral que presenta una mutación indicada por el triángulo rojo.

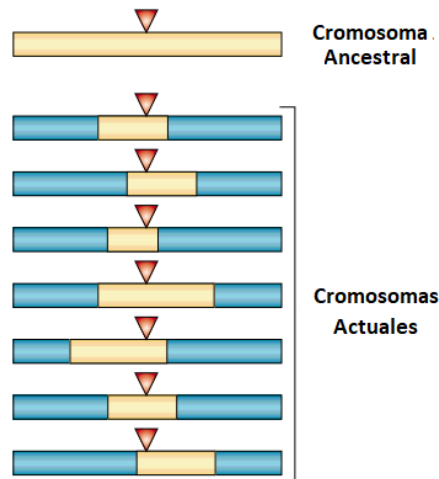


Figura 5 Desequilibrio de ligamiento en torno a una mutación ancestral

Con el paso del tiempo y debido al proceso de recombinación cromosómica (al generar nuevas generaciones de individuos), se introducen porciones nuevas de cromosomas que se muestran en color azul. Sin embargo, se puede observar un pequeño fragmento del cromosoma ancestral que sigue portando la mutación en los cromosomas actuales. Esto indica que los marcadores cercanos físicamente a la mutación tienden a permanecer asociados a ella a través del tiempo, formando lo que se conoce como un bloque de haplotipo, es decir, varios marcadores asociados en un mismo cromosoma.

1.1.7 GENOTIPO, FENOTIPO Y SU RELACIÓN CON EL AMBIENTE

Todo organismo tiene su material genético que contiene la información biológica necesaria para construir y mantener un ejemplo viviente de ese organismo, conocida colectivamente como genoma. La mayoría de los genomas, incluidos el genoma humano y los de todas las demás formas de vida celular, están compuestos por ADN, pero unos pocos virus tienen genomas de RNA.

El genotipo es la suma de los genes y combinación de alelos, que tiene un determinado individuo. Un alelo, en términos moleculares es la información genética codificada en el ADN que está presente en el núcleo, los plásmidos y mitocondrias; a nivel de especie se habla del genoma que incluye todos los genes y su potencial de expresión.

El fenotipo es el conjunto de caracteres que se manifiestan visiblemente a nivel del individuo o población, es decir, aquellas propiedades que podemos observar, medir, cuantificar, cosechar, etc., por ejemplo: altura, peso, color de ojos, rendimiento de producción, entre otros. Por su parte, el ambiente es el conjunto de variables externas que influyen sobre el desarrollo y las funciones de un organismo. Muchas veces el ambiente controla la expresión de los genes o las proteínas. El ambiente incluye todos los factores abióticos (parámetros químicos y físicos) y también factores bióticos de interacciones biológicas y ecológicas [11].

1.1.8 LOCUS DE UN CARÁCTER CUANTITATIVO

Con la generación de mapas de marcadores basados en polimorfismos de ADN, ha sido posible detectar loci particulares responsables de variación genética cuantificable, es decir, rasgos fenotípicos observables debido a influencia genética o ambiental [12]. Esta variación puede consistir en valores discretos o continuos, por ejemplo, altura, peso, presión arterial, porcentaje de grasa en leche, rendimiento de producción, etc. De manera general estos rasgos o características cuantificables son conocidos como QTL (Quantitative Trait Loci, por sus siglas en inglés).

El análisis para la detección de QTL es muy importante ya que la variación genética cuantitativa, es el substrato para la evolución de fenotipos en las poblaciones y para la reproducción selectiva de cultivos y ganados domésticos. También, ayuda a entender la susceptibilidad a enfermedades complejas, respuestas a terapias farmacológicas, así como ayudar a resolver preguntas evolutivas aún sin respuesta, acerca el mantenimiento de la variación genética de rasgos cuantitativos dentro de poblaciones y los mecanismos de divergencia entre poblaciones.

Para la identificación de genes que afectan rasgos cuantitativos, se requiere generar mapas de regiones donde uno o más alelos que afectan un rasgo, son segregados. Para este propósito existen dos enfoques: mapas de desequilibrio y mapas de asociación. Ambos se basan en el principio de que los QTLs pueden ser rastreados mediante su ligamiento genético a marcadores (SNPs, indels o microsatélites). Si un QTL está ligado a un marcador, entonces los individuos con diferentes valores de loci tendrán valores diferentes para el rasgo cuantitativo. En ambos casos se necesitan obtener mediciones del fenotipo y determinar valores de posiciones en los genotipos de los individuos en la población muestreada. Después se utilizan métodos estadísticos para determinar cuando existe diferencia entre los valores de los rasgos cuantificables en los individuos con diferentes valores en posiciones del genotipo [13].

Se han generado diversos proyectos colaborativos para generar mapas de genes que afectan rasgos económicos de gran importancia en diversas especies ganaderas. De estas colaboraciones también se están generando métodos estadísticos apropiados para el análisis de la información necesaria, como lo son datos fenotípicos, pedigrí e información de marcadores genéticos [14].

1.2 BOVINOS

La familia *Bovinae* comprende a los bovinos con joroba (*Bos indicus*) y sin joroba (*Bos taurus*), el yak (*Poephagus grunniens*), el gayal (*Bibos frontalis*) y el búfalo (*Bos bubalus bubalis*), entre otros. El número de especies de la familia *Bovinae* provee una amplia gama de contribuciones a la alimentación y la agricultura, representando cerca del 30% de la carne mundial y más del 87% de la producción mundial de leche. También tienen valor por su aporte en el área de la tracción, siendo utilizados para el transporte de mercancía, trabajo de la tierra de cultivo, y por su estiércol, utilizado como abono o fertilizante [15].

1.2.1 ORIGEN Y DEMOGRAFÍA HISTÓRICA

Se ha comprobado que los dos tipos principales de bovino domésticos, los cebúes (*Bos taurus indicus*) y los taurinos (*Bos taurus taurus*), descienden de los bueyes salvajes (*Bos primigenius*), los cuales se extinguieron en el año de 1627. Las dos poblaciones ancestrales de bueyes salvajes habrían divergido hace unos 280,000 años y fueron sujetos a eventos demográficos antes de ser domesticados por los humanos, como lo son cuellos de botella, mezclas y selección natural.

Evidencias moleculares en combinación con información histórica y arqueológica, indican que las especies salvajes fueron domesticadas en al menos dos episodios: hace unos 10,000 años los bovinos taurinos fueron capturados en la región conocida como la “Media Luna Creciente”, que comprende porciones de los países actuales de Israel, Jordania, Líbano, Chipre y Siria, así como Egipto, Turquía, Irak, Iran y Kuwait); aproximadamente 1,500 años después, el ganado cebú fue domesticado en Valle del Indo, que corresponde a porciones de Afganistán, Pakistán y el noreste de la India.

A partir de estos eventos trascendentales, ambos grupos de bovinos comenzaron a esparcirse a otras regiones siguiendo las migraciones humanas. El grupo de los taurinos se esparcieron lentamente por Asia y Europa, llegando hasta el Nuevo Mundo por las importaciones de los europeos, después del descubrimiento de América en el siglo XV. Los descendientes de éstos constituyen el grupo comúnmente llamado ganado criollo. Por su parte, el ganado cebú se expandió alrededor del globo acompañando migraciones humanas, pero se volvieron mayormente endémicos en regiones tropicales y subtropicales debido a su adaptabilidad a dichos ambientes. Este tipo de ganado fue probablemente introducido en el continente africano en el siglo VII, mediante la importación sementales de *B. Indicus* durante las invasiones árabes, mientras que en América Central y América del Sur en el siglo XX; después de 1950 su introducción al continente americano se intensificó.

Después de la domesticación, los ganaderos comenzaron a tomar control sobre el apareamiento de los individuos, de acuerdo con sus intereses en rasgos fenotípicos como son el tamaño, temperamento, producción de leche, entre otras. Esta estrategia de combinar a los mejores especímenes entre sí con la finalidad de reforzar los rasgos sobresalientes en ellos, puso en marcha una presión de selección muy fuerte provocando una serie de declives en el tamaño efectivo de la población, es decir, el número de individuos que aportan su información genética para generar las siguientes generaciones.

Cerca de 200 años atrás, los ganaderos designaron el concepto de especie y comenzaron a limitar el intercambio de germoplasma, con la finalidad de estandarizar los rasgos de morfología y desempeño de sus ganados, lo cual ha llevado al registro de cerca de 1,200 razas de ganado bovino; sin embargo, cerca de 200 especies se han extinguido y otras se encuentran en riesgo de desaparecer [16].

1.2.2 EL GENOMA BOVINO Y EL PROYECTO DE HAP MAP

El genoma bovino contiene aproximadamente 22,000 genes, distribuido en 29 cromosomas autosomales y un cromosoma sexual [17], como se puede apreciar en el cariotipo de la Figura 6.

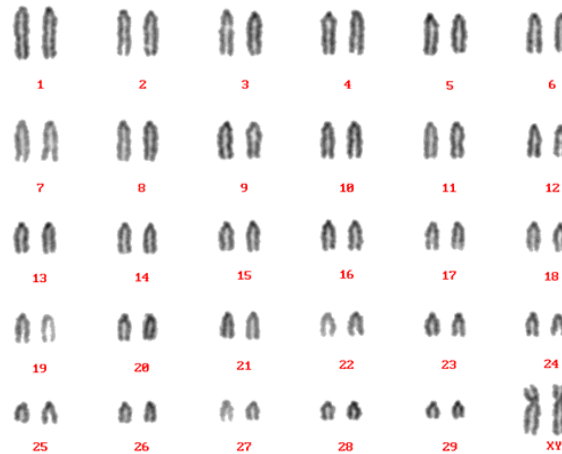


Figura 6 En el cariotipo bovino se puede apreciar la cantidad y el tamaño relativo de los cromosomas

En 2004 se inició el Proyecto Internacional para la generación del Mapa de Haplotipos del Ganado Bovino (Bovine HapMap Project) [18], para unificar esfuerzos en la secuenciación de todo el genoma bovino. La información generada condujo a la elaboración de un panel de genotipificación para bovinos de alta resolución con una capacidad de prueba de ~50,000 SNPs.

Son muchos los resultados generados por el Consorcio HapMap Bovino sobre esta especie, entre ellos están la estructura de los bloques de haplotipos, determinación del desequilibrio de ligamiento y estimación de la población efectiva, entre otras contribuciones [19], [20].

En años siguientes se desarrollaron metodologías diversas para realizar múltiples pruebas en los genomas disponibles para la detección de eventos de selección, a partir de distintos enfoques, de tal manera que los resultados de una prueba fueran contrastados con otras, resultando en la

identificación de señales de selección positiva asociadas a genes con implicaciones en reproducción, metabolismo y respuesta del sistema inmune [21]–[23].

1.2.3 TECNOLOGÍAS DE GENOTIPIFICACIÓN

La genotipificación es el proceso mediante el cual se determinan las variaciones genéticas que posee un individuo en particular. Este proceso puede ser llevado a cabo mediante una gran variedad de métodos, dependiendo de las variaciones de interés, la resolución del estudio, así como el presupuesto para tal propósito. Para investigar muchas variaciones a la vez, particularmente variaciones comunes, los chips de genotipificación son una alternativa eficiente y con buen grado de precisión [24].

1.2.4 SNP CHIPS

El descubrimiento y caracterización de miles de marcadores tipo SNP en el genoma bovino y otros genomas de especies pecuarias, trajo como consecuencia una reducción en los costos de genotipificación. En 2009 la casa comercial Illumina en Estados Unidos puso a la venta el SNPchip50 (Figura 7a), para capturar un conjunto de 55,609 polimorfismos a lo largo de los 30 cromosomas bovinos, teniendo tal grado de éxito, que pronto se convirtió en el estándar para realizar estudios genéticos de las razas de ganado pertenecientes a la evolución de *Bos taurus*.

En 2010 Illumina lanzó al mercado un SNPchip de alta densidad conocido como BovineHD Genotyping BeadChip (Figura 7b), el panel con mayor resolución para analizar variaciones genéticas de cualquier raza de ganado bovino ya sea con propósito de explotación lechero o cárnico.

Este chip interroga 777,962 SNPs que se distribuyen de manera uniforme en todo el genoma, permitiendo que sea implementado en estudios de señales de selección, la identificación de QTLs,

evaluación del mérito genético, estudios de desequilibrio de ligamiento, estudios de genética comparativa, entre otros.

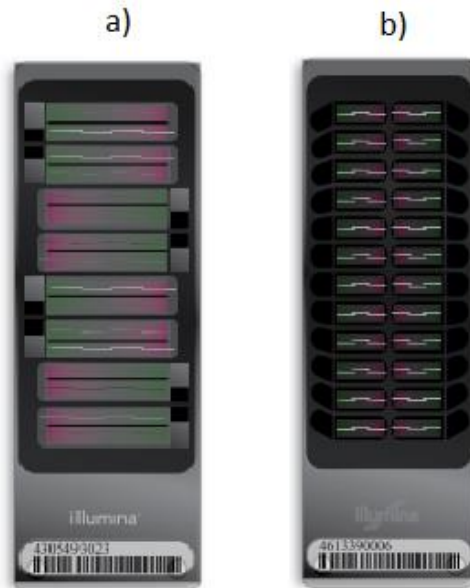


Figura 7 SNPchips para genotipificación en el genoma bovino. (a) Bovine SNPchip50; (b) BovineHD

1.2.5 ANOTACION DEL GENOMA BOVINO UMD3.1

En diciembre de 2009 se publicó el tercer ensamble del genoma bovino (*Bos taurus*) UMD 3.1, por parte del CBCB de la Universidad de Maryland (Center for Bioinformatics and Computational Biology por sus siglas en inglés). Las secuencias del genoma fueron generadas mediante una combinación de dos estrategias de secuenciación: secuenciación jerárquica shotgun (~11 millones de lecturas) y secuenciación de genoma completo shotgun (~24 millones de lecturas), mientras que el ensamblaje se realizó utilizando la herramienta Celera Assembler 5.2. La longitud total del ensamble UMD 3.1 es de 2.65 Gb y se identificaron un total de 19,981 genes codificantes de proteínas, 3,825 genes no codificantes, 797 pseudogenes y 26,740 transcritos de genes [25].

1.3 PLANTEAMIENTO DEL PROBLEMA

La genética de poblaciones ha ofrecido nuevos paradigmas para estudiar la relación entre la información genética y los fenotipos, los cuales se conocen como estudios de señales de selección y presentan las ventajas de requerir muestras de menor tamaño que los estudios tradicionales, así como su capacidad de revelar señales en genes que controlan rasgos que son difíciles, muy costosos o incluso imposibles de detectar en poblaciones grandes.

En años recientes se ha incrementado el interés por realizar este tipo de estudios en diversas especies de explotación económica, las cuales se encuentran bajo una fuerte presión de selección antropogénica, sin embargo, referente al ganado bovino, la mayoría de los estudios publicados utilizan genotipos de baja densidad, un número muy reducido de razas y propósitos de crianza limitados. Por tal motivo, en este trabajo de tesis se presenta un estudio de señales de selección en el genoma bovino empleando marcadores tipo SNP de alta densidad. La información utilizada proviene del Consorcio HapMap Bovino y comprende 19 linajes bovinos con propósitos de crianza diversos.

1.4 JUSTIFICACIÓN

Detectar eventos de selección en el genoma del ganado bovino es crucial para el entendimiento del pasado genético y su papel actual en cuestiones de adaptación, mejora genética, resistencia a enfermedades, calidad de producción, entre otros factores importantes en especies de explotación económica.

Los estudios de señales de selección representan un nicho de oportunidad para aportar conocimiento en la identificación, descripción y posible explicación biológica a variaciones genéticas que contribuyen a la diversidad de fenotipos de las especies, así como validar a nivel

genético los intensos y milenarios esfuerzos de domesticación y programas de mejora genética. De igual manera, los resultados de este estudio aportarán conocimiento nuevo al Consorcio HapMap Bovino y a diversas ramas de la ciencia afines.

1.5 OBJETIVOS

1.5.1 OBJETIVO GENERAL

Detectar señales de selección positiva en genomas bovinos mediante el enfoque de Homocigosidad Haplótica Extendida utilizando SNPs Chips de alta densidad y el entorno de programación R.

1.5.2 OBJETIVOS ESPECÍFICOS

Para el cumplimiento del objetivo general de este proyecto de tesis se planean los siguientes objetivos específicos:

- Desarrollar un algoritmo que permita la manipulación y preprocesamiento de los datos contenidos en los archivos con los haplotipos de alta densidad de los 19 linajes bovinos.
- Implementar el cálculo de *iHS* en el compendio de información, utilizando las funciones de la paquetería *rehh* de R Project.
- Identificar y reportar las regiones genómicas que correspondan a señales de selección positiva.
- Asociar la presencia de elementos genéticos anotados en las regiones genómicas de interés con las señales de selección positiva identificadas.

CAPÍTULO 2 MARCO TEÓRICO

2.1 TEORÍA DE LA EVOLUCIÓN

2.1.1 GENÉTICA DE POBLACIONES Y EVOLUCIÓN

La genética es el estudio de los rasgos biológicos hereditarios, incluyendo rasgos que son influenciados por el medio ambiente. Los genes representan los elementos de la herencia que son transmitidos de los padres a su descendencia mediante la reproducción.

El propósito de la genética de poblaciones es entender las fuerzas que producen y mantienen la variación genética en las especies. Esas fuerzas incluyen mutación, recombinación, selección natural, estructura de la población y la transmisión aleatoria de material genético de padres a hijos [1].

2.1.2 TEORÍA DE LA COALESCENCIA

La teoría de la coalescencia es una parte de la genética teórica de poblaciones, un campo que considera las variaciones genéticas en modelos matemáticos sustentados en matemáticas aplicadas, teoría de la probabilidad y estadística. Tiene sus raíces en la historia en la genética de poblaciones, y al mismo tiempo se orienta al análisis de datos genéticos modernos, especialmente datos de secuencias de ADN. La coalescencia describe la ascendencia genética de una muestra de secuencias bajo un modelo y realiza predicciones acerca de patrones de variación genética.

La ascendencia genética de una muestra o la genealogía de genes, es el conjunto de relaciones ancestrales a través de los miembros de una muestra. Esto incluye el tiempo con un ancestro común, el cual es medido en el presente con un enfoque retrospectivo [26].

2.1.3 SELECCIÓN NATURAL

Las dos ideas principales derivadas de la teoría de Darwin-Wallace sobre la evolución de las especies señalan que los organismos son producto de una historia evolutiva a partir de modificaciones de un ancestro común. Además, que el mecanismo principal de la evolución es la selección natural de estas variaciones hereditarias.

La selección se refiere a la reproducción diferencial de algunos fenotipos/genotipos, sobre otros bajo ciertas condiciones ambientales que prevalecen en un momento determinado. De esta manera, la selección natural es un mecanismo evolutivo que genera un cambio en las frecuencias relativas de los fenotipos/genotipos, de acuerdo con su adaptación dentro de la población.

El papel de la selección natural es ampliamente aceptado dentro de la comunidad científica, sobre todo en cuanto a caracteres morfológicos, es decir, a las variaciones fenotípicas. Sin embargo, a partir del desarrollo de nuevas técnicas para el estudio de las variaciones a nivel de ADN y proteínas, se encontró evidencia de que existía una gran variación genética contenida en éstas, misma que no podía ser justificada del todo por efectos de la selección natural. Estas ideas desembocan en la propuesta de la Teoría Neutra de la evolución, propuesta por Motoo Kimura a finales de los 60s.

2.1.4 TEORÍA NEUTRA DE LA EVOLUCIÓN

De acuerdo con la teoría neutral de la evolución molecular, la mayor parte de la variación al nivel molecular no es mantenida por la selección tipo darwiniano (selección adaptativa o positiva), sino

resultado de un balance entre la deriva genérica y las mutaciones neutras a casi neutras. Es decir, la mayor parte de la variación a nivel molecular ocurre de manera aleatoria y no presenta importancia adaptativa.

Debido a estas ideas, se ha generado un gran debate de si existen pruebas para demostrar la presencia de la selección positiva a nivel molecular; lo anterior llevó al desarrollo de numerosos métodos para su detección [27].

2.1.5 VARIACIÓN GENÉTICA Y SELECCIÓN

El fundamento clave de la evolución es el hecho de que la diversidad genética es heredable, de tal manera que existe un sustrato para los fenómenos de la selección natural y artificial. Para que la evolución se haga presente en una especie, debe existir una introducción continua de nuevas variaciones genéticas, las cuales se producen mediante diversos mecanismos en el genoma, destacando notablemente las mutaciones.

Las mutaciones son eventos raros y muchas de ellas tienen un efecto neutro en las secuencias (mutaciones neutras). Las variaciones genéticas también pueden ser materializadas como variaciones estructurales en cromosomas, por ejemplo: duplicaciones, borrados, inserciones, inversiones, translocaciones, entre otras. Por otra parte, los aspectos vinculados a la reproducción sexual de los individuos son los mecanismos principales que mantienen la diversidad genética dentro de las poblaciones [28].

2.1.6 TIPOS DE SELECCIÓN

Básicamente, la selección natural a nivel molecular puede ser considerada de dos tipos: selección positiva (que incluye a la selección balanceadora o diversificadora, y a la adaptativa darwiniana) y la selección negativa (purificadora).

La selección positiva adaptativa es el mecanismo evolutivo mediante el cual los nuevos mutantes poseen adecuaciones mayores que el promedio de la población, y las frecuencias de dichos mutantes se incrementan en las siguientes generaciones, de tal manera, que la selección positiva disminuye la variación genética al favorecer un alelo benéfico sobre otros [27].

En muchas situaciones, mutaciones que ocurren en secuencias funcionales importantes disminuyen la aptitud del individuo (mutaciones deletéreas); estas mutaciones pueden alterar procesos metabólicos u otras propiedades del individuo desembocando en una tasa de reproducción menos exitosa, de manera tal que, en una población, dichas mutaciones desfavorables terminan siendo removidas. Este tipo de proceso es conocido como selección de negativa o selección purificadora.

La selección purificadora produce una disminución en la diversidad genética, hasta que la población se estabiliza en una frecuencia de haplotipos mutantes en particular, mientras que la selección balanceadora se encarga de mantener el polimorfismo, es decir, de mantener frecuencias estables de dos o más tipos de fenotipos en la población [22].

2.1.7 DERIVA GENÉTICA

La deriva genética es un proceso por el cual se presentan cambios en la frecuencia alélica por aleatoriedad. De manera virtual, si todos los individuos en una población tienen exactamente la misma aptitud genética, de manera tal que la selección natural no tiene efecto alguno, las frecuencias alélicas cambian debido a un proceso aleatorio de generación de gametos. Esto se debe a que no todos los individuos de una población se reproducen o aportan información genética a la siguiente generación, de tal manera que, del total de gametos posibles, solo una fracción de estos

es utilizada. Los organismos homocigotos producirán un solo tipo de gametos, mientras que los organismos heterocigotos producirán dos variantes.

En poblaciones grandes este efecto es pequeño debido a que la fracción de gametos aportada a la siguiente generación será estadísticamente representativa. Sin embargo, en poblaciones pequeñas se pueden presentar cambios considerables en las frecuencias alélicas.

En la ausencia de nuevas mutaciones y factores como la selección que afecten la frecuencia alélica, los alelos sujetos a la deriva genética serán eventualmente fijados. La fijación de una variante en una población se alcanza cuando la frecuencia de un alelo es del 0% o 100% [29].

2.1.8 EFECTO HICHHIKING Y BARRIDOS DE SELECCIÓN

Las señales de selección son definidas como regiones en el genoma, que albergan variaciones funcionales importantes en secuencias de ADN y que han estado sujetas, ya sea a presión natural o artificial, dejando un patrón especial en dichas regiones. La extensión física de dichas señales en torno a la variante funcional es una consecuencia del efecto llamado “hitchhiking”.

En la Figura 8 se compara un fragmento del genoma de varios individuos de una población que presentan variabilidad entre ellos mismos. Considere que la mayor parte de la variación es neutral, lo que significa que no tiene un efecto detectable en la aptitud de los individuos.

Ahora considere que un alelo emerge como una mutación (en color rojo) en una copia de un cromosoma. En las siguientes generaciones, los haplotipos que portan el alelo bajo selección serán acompañados por alelos vecinos en desequilibrio de ligamiento, por el efecto de “hitchhiking” y se volverán muy comunes en la población. Este proceso es llamado barrido de selección.

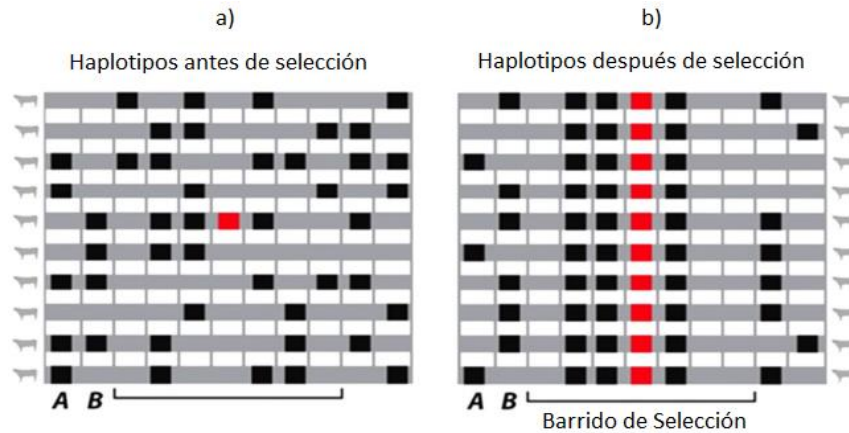


Figura 8 Efecto de Hitchhiking. Al comparar las imágenes (a) y (b) se observa la extensión física del fenómeno de hitchhiking en torno a la variante funcional.

Esto provoca que variaciones neutrales inicialmente a frecuencias intermedias incrementen su frecuencia casi hasta la fijación o disminuir hasta casi su extinción, de tal manera que las secuencias de ADN afectadas por los barridos de selección tendrán pocos alelos segregándose a frecuencia intermedia que lo esperado bajo equilibrio neutral.

Mientras que los alelos bajo el efecto de selección y los alelos bajo el efecto de hitchhiking incrementan su prevalencia en un barrido de selección, al mismo tiempo, los segmentos que abarcan dichos alelos irán disminuyendo lentamente de tamaño debido al proceso de recombinación. Esto genera una estructura específica de haplotipos segregándose en los barridos de selección. De esta manera, existen diferentes aspectos en la información genética que funcionarán como pautas para detectar barridos de selección [28].

2.2 MÉTODOS PARA LA DETECCIÓN DE SEÑALES DE SELECCIÓN

En años recientes se ha incrementado el interés por realizar estudios para detectar señales de selección SS (Selection Signatures por sus siglas en inglés), como consecuencia de la selección

natural o antropogénica, debido a que ofrecen una estrategia complementaria a los estudios tipo GWAS (Genome Wide Association Study por sus siglas en inglés), para el mapeado de variantes con impacto en rasgos de interés, contribuyendo a establecer relaciones entre fenotipos y funciones de genes.

Un estudio típico GWAS busca determinar el efecto que tienen las variaciones genéticas sobre un fenotipo en particular a partir del análisis de genotipos; por su parte, los estudios de señales de selección toman la dirección contraria, es decir, se inicia con la intención de determinar que mutaciones son responsables de algún fenotipo en particular en muestras que comparten un factor común como lo son la proximidad geográfica, factores ambientales o directamente el mismo fenotipo.

Una gran ventaja que presentan los estudios para detectar señales de selección, es que requieren muestras de menor tamaño que para estudios de GWAS, aproximadamente 10 veces más pequeñas; también los estudios de SS pueden revelar señales en genes que controlan rasgos que son difíciles, muy costosos o incluso imposibles de detectar en poblaciones grandes.

Sin embargo, su aplicación en datos de ganado bovino debe ser correctamente interpretada, debido a la existencia de eventos históricos demográficos como cuellos de botella, deriva genética, mezclas, así como el efecto fundador provenientes de los paneles de SNPs desarrollados [16].

Los diversos enfoques computacionales para la detección de SS pueden ser divididos en dos categorías: métodos que utilizan la divergencia de secuencias y diversidad de patrones entre especies, y métodos que consideran la variación genética de una población. Generalmente, las comparaciones entre especies se utilizan para identificar eventos más antiguos, mientras que los

métodos basados en una sola población revelan episodios más recientes de selección, como se indica en la Tabla 1.

La detección de una misma señal de selección de una región utilizando enfoques alternos puede constituir evidencia significativa para probar los efectos de selección. Sin embargo, el éxito en la detección de una prueba y el fracaso en otras no significa la negación del efecto de selección, puesto que las diversas metodologías tienen alcances de detección distintos [30].

Tabla 1 Principales enfoques para la detección de señales de selección a partir de información genómica.

Enfoque	Señales	Metodología	Modalidad de comparación	Selección detectable	Referencia temporal (años) ^a
Reducción local de la diversidad genética	Reducción local de la heterocigosidad en comparación con el resto del genoma	ZHp, ROH	Dentro de una población	Positiva	< 200,000
Cambios en el espectro de frecuencia de alelos	Incremento relativo en la frecuencia de alelos derivados	Tajima's D; Fu y Li D-test; Fay y Wu H-test	Dentro y entre poblaciones	Positiva y balanceadora	< 200,000
Homocigosidad haplotípica extendida	LD persistente e inusuales bloques de haplotipos extensos	LRH, EHH, iHS, XP-EHH, Rsb,	Dentro de una población	Positiva	< 30,000
Diferenciación entre poblaciones	Diferencias en las frecuencias de alelos entre poblaciones	F _{st}	Entre poblaciones	Positiva y balanceadora	< 80,000

^a La estimación de la referencia temporal en años está basada en información generacional de humanos, asumiendo intervalos de generación de 25 años.

2.2.1 REDUCCIÓN LOCAL DE DIVERSIDAD GENÉTICA

En cada generación las recombinaciones en los cromosomas rompen los bloques de haplotipos produciendo el efecto de equilibrio de ligamiento. Cuando un evento de selección incrementa la frecuencia de un alelo que otorga una ventaja en el individuo, los alelos vecinos quedan anclados a él y aumentan su frecuencia rápidamente, de tal manera que los eventos de recombinación no imposibilitan que se presente el fenómeno de desequilibrio de ligamiento, provocando que un segmento de cromosoma pierda su diversidad. De esta manera, las señales de selección positiva pueden ser detectadas al buscar regiones cromosómicas donde la heterocigosidad es mucho menor a la esperada bajo el supuesto de neutralidad.

Otro enfoque para la búsqueda de deficiencia en heterocigosidad en genotipos, es identificar islas de ROH (Runs Of Homozygosity por sus siglas en inglés), debido al efecto de idénticos por descendencia. Sin embargo, debido a que la recombinación es aleatoria, la distribución de ROH a través de una muestra se espera que sea muy heterogénea bajo la teoría de neutralidad, por lo que regiones puntuales del genoma con ROH significativos, pueden representar una señal de un barrido de selección. El tamaño de la isla muestra está correlacionado negativamente con el número de generaciones que han ocurrido desde el evento de selección, permitiendo fechar la edad del barrido de selección.

Enfocados en la identificación de ROH, Kim et. al en 2003 propusieron un nuevo estadístico llamado F_{RL} que compara la homocigosidad local entre una población bajo selección artificial y una población control. De manera general, para cada SNP, los animales son valorados con un 1 si el locus se encuentra dentro de un ROH o un 0 en caso contrario. Se calcula F_L como la proporción de animales con valores igual a 1 y finalmente se calcula F_{RL} como:

$$F_{RL} = \left(\frac{F_{L(\text{seleccionado})}}{F_{L(\text{control})}} \right)$$

Todos los valores son estandarizados para producir una distribución con media 0 y varianza 1. Valores extremos de F_{RL} representan cambios en la frecuencia alélica en la población bajo selección artificial en comparación del grupo control, reflejando un aumento en la homocigosidad debido a cruzamiento reciente, deriva genética o selección en el locus analizado [16].

2.2.2 CAMBIOS EN EL ESPECTRO DE FRECUENCIA DE ALELOS

Después de que se lleva a cabo un barrido de selección, nuevas mutaciones restauran lentamente la diversidad local con el curso de muchas generaciones. Debido a que las mutaciones son generalmente raras, pueden tomar muchas generaciones antes de que la frecuencia de dicha mutación aumente bajo el efecto de evolución neutral, por lo que la depresión local de heterocigosidad puede permanecer sin alteración por varias generaciones después de que la presión de selección haya actuado.

Las mutaciones nuevas o los llamados alelos derivados (variaciones que difieren con respecto a los valores originales o ancestrales), ocurren en baja frecuencia en comparación con los alelos ancestrales bajo el supuesto de neutralidad, pero cuando aparecen dentro de un barrido de selección quedan anclados y aumentan su frecuencia rápidamente en la población bajo selección. Diversos métodos para la detección han sido propuestos en esta categoría, incluyendo: Tajima's, Fay y Wu's, y ΔDAF .

La limitación para la implementación de estos métodos en datos de ganados bovinos es inferir los valores de los alelos ancestrales, puesto que las poblaciones ancestrales ya no existen. Generalmente son determinados comparando los genomas de ganado domesticado con genomas

de especies salvajes como búfalos (*Bubalus bubalis*) y yaks (*Bos grunniens*) entre otros, que se asume que derivan del ancestro común fundador de la familia de especies *Bovinae*.

Una alternativa para sobrellevar la limitación de no conocer los alelos ancestrales es utilizar diversas poblaciones genéticamente distintas para estimar las frecuencias de alelos promedio que pudieran representar el espectro en la población ancestral común. Por lo que, bajo este enfoque, en lugar de buscar alelos derivados o raros, se desea detectar cambios en el espectro de frecuencia de alelos de una población en comparación de los valores promedios calculados.

Para probar este enfoque, se puede calcular el CCL (Composite Log-likelihood por su acrónimo en inglés) para cada ventana de SNP como:

$$CCL = \sum_{i=1}^k \ln[1 - P_r(d < |p_i - \bar{p}_i| | \mu_i)]$$

donde para cada SNP i , d es cualquier valor aleatorio de la distribución teórica de frecuencias de alelos con media $\mu_i = \bar{p}_i$, y \bar{p}_i es el valor promedio de la frecuencia del alelo utilizado de referencia a través de las poblaciones; p_i es la frecuencia del alelo en la población investigada. La distribución teórica de las frecuencias de alelos puede ser modelada como a una distribución binomial o como una aproximación normal a binomial. Este tipo de análisis solo requiere un compendio de datos con múltiples razas en lugar de información de alelos ancestrales, con buenos resultados en el análisis de ganado bovino [16].

En un trabajo publicado en 2014 en la revista *Animal Breeding and Genetics*, se realizó una comparación de secuencias de genomas bovinos con secuencias de especies que comparten un ancestro en común, logrando generar los alelos ancestrales putativos para 14,339,107 SNPs bovinos, lo cual representa el 74% de los SNPs presentes en el BovineHD Genotyping BeadChip.

Dicha información se encuentra disponible en el archivo de libre acceso y descarga: Bovine_Ancestral_Allele en [31].

2.2.3 SEGMENTOS LARGOS DE HAPLOTIPOS

En 2002 *Sabeti et al.* definieron la Homocigosidad Haplotípica Extendida, EHH (Extended Haplotype Homozygosity por sus siglas en inglés) como la probabilidad de que dos cromosomas elegidos al azar que portan el mismo alelo en un SNP particular (SNP focal), sean idénticos por descendencia para todo un intervalo desde el SNP focal a una determinada distancia dada [32]. En este enfoque se propone identificar haplotipos cuya frecuencia se incrementa rápidamente en una población, impidiendo que los eventos de recombinación puedan interrumpir los haplotipos, generándose un prolongado desequilibrio de ligamiento persistente.

Si de manera general se consideran N cromosomas en una muestra y G haplotipos únicos, que se extienden desde un SNP focal hasta una distancia x en dirección ascendente o descendente, con cada uno de los grupos g presentando n_g observaciones, el valor de EHH para la muestra entera se calcula como:

$$EHH = \frac{\sum_{g=1}^G \binom{n_g}{2}}{\binom{N}{2}}$$

De tal manera que EHH detecta la transmisión de un segmento de haplotipo sin recombinarse. Si un alelo se encuentra bajo presión de selección, entonces el valor de EHH se espera que sea cercano a 1 en ambos sentidos entorno a dicho SNP [32].

Por definición, el valor de EHH para un SNP comienza en 1 y presenta una caída hasta el valor de 0 conforme se aumenta la distancia de inspección respecto al SNP inicial. En 2006 *Voight et al.*

propusieron una prueba empírica basada en la integral de la caída observada de EHH, la cual definieron como iHH .

$$iHH = \int_a^b EHH(x)dx$$

donde iHH representa el área bajo la curva evaluada en un segmento del cromosoma delimitada en el extremo superior por la posición a y extremo inferior b hasta que el valor de EHH decaiga hasta un valor arbitrario pequeño, por ejemplo 0.05.

Adicionalmente, propusieron la prueba estadística iHS (Integrated Haplotype Score, por sus siglas en inglés) como la razón log de iHH para aquellos haplotipos que portan el alelo ancestral iHH_A y el alelo derivado iHH_D :

$$iHS_i = \ln\left(\frac{iHH_{A,i}}{iHH_{D,i}}\right)$$

El iHS compara el EHH entre los alelos ancestrales y derivados en una población; estas mediciones se estandarizan usando la media y desviación estándar de los valores de todos los SNPs con frecuencias alélicas similares. Los valores extremos negativos han sido de particular importancia en estudios de genética humana, ya que ellos representan mutaciones recientes adquiridas cuya frecuencia ha aumentado rápidamente, sin embargo, un barrido de selección puede producir valores extremos positivos de iHS si los alelos ancestrales se anclan bajo el efecto de “hitchhike” a un sitio en particular.

En el contexto de los bovinos, la selección artificial y domesticación probablemente favorecen alelos benéficos sin importar si son ancestrales o derivados, por lo que los valores extremos, ya sean positivos o negativos, deben ser de interés. Esto implica que el valor absoluto del valor

estandarizado de iHS sea preferido sobre el signo, lo que conlleva a asumir una hipótesis de prueba de dos colas.

Una de las limitantes de este método es que, si un marcador presenta una fijación completa (o muy cercana de serlo) en la población analizada, este alelo no tendría una integral para ser evaluada (o tendría una integral muy cercana a cero), por lo que calcular la razón produciría un error matemático, arrojando valores infinitos positivos o negativos. Debido a lo anterior, se debe condicionar la existencia de MAF. Esta limitante provoca que iHS tenga un bajo poder para detectar barridos de selección muy recientes [16].

Tang et al. y *Sabeti et al.* en 2007 desarrollaron de manera independiente un procedimiento similar para comparar los perfiles de EHH entre poblaciones. Su orientación es la de detectar eventos de selección que han resultado en una fijación de alelos benéficos, ya sea de manera completa o cercana a serlo, comparando el comportamiento de EHH en datos de genotipos en poblaciones análogas [33].

$$XPEHH_i = Rsb_i = \ln \left(\frac{iHH_{pop1,i}}{iHH_{pop2,1}} \right)$$

donde $iHH_{pop1,i}$ es la integral de EHH en la primera población y $iHH_{pop1,2}$ es la integral de la segunda población. Los valores son estandarizados para producir una distribución estándar. Valores positivos indican barridos de selección en la población del numerador, mientras que valores negativos indican selección en la población del denominador [16].

2.2.4 DIFERENCIACIÓN DE POBLACIONES

Siguiendo los mismos principios que en Rsb , $XPEHH$ y F_{RL} donde la selección positiva actúa en poblaciones que comparten factores geográficos, ambientales o fenotipos comunes, es probable

que no compartan la misma presión de selección que poblaciones externas ubicadas en otros entornos. De esta manera los cambios en la frecuencia alélica ocasionados por la presión de selección en una población no serán observadas en otras poblaciones, generándose diferencias marcadas en la frecuencia alélica al ser comparadas.

El índice de fijación F_{ST} es el método estándar para detectar loci muy diferenciados entre poblaciones. En esencia, dado un valor promedio de la frecuencia de alelo \bar{p} a través de una subpoblación, F_{ST} es simplemente la razón de la variancia en la frecuencia alélica en diferentes subpoblaciones $\sigma_S^2 = \sum_{j=1}^k (p_j - \bar{p})^2$ y la varianza en la población total $\sigma_T^2 = \bar{p}(1 - \bar{p})$. En comparaciones pareadas, los cálculos se simplifican a:

$$F_{ST} = \frac{(p_1 - p_2)^2}{(p_1 + p_2)(2 - p_1 - p_2)}$$

Los valores pueden ser promediados a través de ventanas de SNP o a través de posiciones genómicas usando un estimado kernel local [16].

2.3 CONSIDERACIONES Y SOFTWARE PARA EL ANÁLISIS DE SS

El primer paso que se debe considerar antes de iniciar un análisis de SS después de haber elegido la metodología que satisface los objetivos de la investigación, es filtrar los datos de baja calidad, marcadores y muestras que no sean informativas o que puedan generar alteraciones en el análisis.

Algunos criterios como el Equilibrio de Hardy-Weinberg (EHW) y MAF comúnmente contemplados en estudios GWAS, deben ser aplicados con precaución. La eliminación de marcadores con desviaciones extremas al EHW puede comprometer la resolución del análisis de SS, puesto que en estos estudios valores con comportamiento atípicos son buscados. De igual

manera, eliminar MAF pudiera eliminar por completo señales que están por alcanzar una fijación completa.

Otro aspecto que se debe considerar es eliminar muestras de animales con relaciones de parentesco cercanas para reducir el número de señales que pudiesen resultar en falsos positivos. El siguiente punto es generar haplotipos de buena calidad a partir de los genotipos. La Tabla 2 presenta una serie de software que contempla las consideraciones previas y llevan a cabo diversos análisis de SS [16].

Tabla 2 Software disponibles para el análisis de señales de selección

Software	Análisis	Referencia
PLINK v1.07 y v1.90beta	Control de calidad, procesamiento de información, revisión de parentesco, ROH y FST	Purcell et al., 2007
SNP & Variation Suite v7.6.8	Control de calidad, procesamiento de información, revisión de parentesco, ROH y FST	Golden Helix, Bozeman, MT, USA
GCTA v1.24.2	Control de calidad, procesamiento de información, revisión de parentesco	Yang et al., 2011
fastPHASE v1.2	Inferencia e imputación de genotipos faltantes	Scheet and Stephens, 2006
Beagle v3.3 y v4.0	Inferencia e imputación de genotipos faltantes	Browning and Browning, 2008
SHAPEIT2	Inferencia e imputación de genotipos faltantes	O'Connell et al., 2014
Sweep	Métodos basados en EHH	Sabeti et al., 2002
Selscan	Métodos basados en EHH	Szpiech and Hernández, 2014
Rehh	Métodos basados en EHH	Gautier and Vitalis, 2012
FLK	Cálculo de los estadísticos de FLK	Bonhomme et al., 2010; Fariello et al., 2013
SweepFinder	Cálculo de los estadísticos de CLR	Nielsen et al., 2005
cosi y cosi2	Simulaciones de coalescencia	Schaffner et al., 2005; Shlyakhter et al., 2014
MSMS	Simulaciones de coalescencia	Ewing and Hermisson, 2010

2.4 LA EXPECTACIÓN DE NEUTRALIDAD

2.4.1 HIPÓTESIS NULA, ALTERNA Y VALORES P EN SS

De manera histórica la selección positiva es estudiada indirectamente cuando un grupo de genes parece desviarse del modelo neutral, es decir, al no encajar en las predicciones neutrales se supone de forma indirecta que estas desviaciones son producto de la selección que las está modificando.

Para poder detectar la participación de la selección en una región del genoma o en un gen en particular, lo primero que se tiene que probar es el rechazo de la hipótesis nula de que la selección no ha actuado en las secuencias en cuestión. Los postulados de la teoría de Kimura nos permiten establecer qué patrones de evolución de las secuencias se deben esperar si existen desviaciones del modelo neutro originadas por la mutación y la deriva genética sin la participación de la selección.

A pesar de los distintos enfoques para los análisis de SS, todos los métodos intentan probar la misma hipótesis nula y alterna, formuladas a partir de teoría de neutralidad, es decir, determinar si un locus en particular se encuentra bajo algún tipo de selección:

H_0 : la selección no ha actuado en el locus

H_1 : la selección ha actuado en el locus

Para los valores de iHS, la probabilidad de que el SNP i con valor de iHS x_i sea neutral, dado que los alelos de referencia (ancestrales) y los alelos alternos son de interés, puede ser aproximada mediante el valor p de dos colas: [16]

$$\begin{aligned} Pr(|iHS| > |x_i| \mid neutral) &= 1 - |\Phi(x_i) - \Phi(-x_i)| \\ &= 1 - 2|\Phi(x_i) - 0.5| \end{aligned}$$

2.5 HERRAMIENTAS BIOINFORMÁTICAS

2.5.1 BASES DE DATOS E INTERFASES

Una base de datos es un “almacén” que permite guardar grandes cantidades de información de forma organizada para que posteriormente puede ser consultada y utilizada fácilmente [34]. En años recientes enormes cantidades de datos biológicos como secuencias de genes, genomas, proteínas, etc., han sido almacenados y puesto a disposición de usuarios en estos almacenes electrónicos públicos, abarcando una gran cantidad de organismos modelo como: *Homo sapiens*, *Gallus gallus*, *Drosophila melanogaster*, *Bos taurus*, *Caenorhabditis elegans*, etc. Existe gran cantidad y diversidad de bases de datos, entre ellas NCBI, KEGG, UCSC, UniProt, Prosite, etc.

La gran cantidad y diversidad de información almacenada en bases de datos, ha generado la necesidad de crear rutas de acceso para llevar a cabo el análisis bioinformático de dichos recursos. Atendiendo esta necesidad, el paquete de programas biomaRt provee una interface a una colección de bases de datos en constante crecimiento, que permite seleccionar y extraer grandes cantidades de información sin la necesidad de escribir búsquedas complejas. Algunos ejemplos de bases de datos de BioMart son: Ensembl, COSMIC, Uniprot, HGNC, Gramene, Wormbase, entre otras [35].

Ensembl es una herramienta para la búsqueda en genomas de vertebrados con aplicaciones a genómica comparativa, evolución, variaciones de secuencia y regulación transcripcional. Ensembl puede realizar la anotación de genes, alineamientos de secuencias múltiples, predecir funciones regulatorias y almacenar información de enfermedades. Como parte de las herramientas de Ensembl se incluye BLAST, BLAT, BioMart y VEP para todas las especies registradas [36].

La base de datos *AnimalQTLdb* (Animal Quantitative Trait Loci Database por su acrónimo en inglés) almacena información sobre QTLs de ganados y datos de asociación curados extraídos de

bases de datos públicas. Particularmente *AnimalQTLdb* está diseñada para facilitar a los usuarios el proceso de comparar, confirmar y localizar la mejor ubicación de genes responsables de caracteres cuantitativos de gran importancia en la producción ganadera. Esta base de datos presenta una serie de herramientas en línea para realizar las búsquedas deseadas o en su defecto, descargar el material de interés en el ordenador. La última actualización contiene 98,090 QTLs/asociaciones de 773 publicaciones, que comprenden 563 caracteres diferentes.

2.5.2 LENGUAJE DE PROGRAMACIÓN PERL

Perl es un lenguaje de programación de propósito general, originalmente desarrollado para la manipulación de texto y actualmente es utilizado para una amplia gama de tareas, incluyendo administración de sistemas, desarrollo web, programación en red, etc. El lenguaje Perl adopta lo mejor de otros lenguajes como C, awk, sed, sh, BASIC, entre otros. En el campo de la bioinformática es de especial apreciación debido a su poder y naturaleza para la manipulación de cadenas, por ejemplo, el uso de expresiones regulares, así como la gran cantidad de módulos desarrollados que permiten, entre otras cosas, conectarse a bases de datos y extraer información de manera sencilla [37].

2.5.3 ENTORNO DE PROGRAMACIÓN R

R es un conjunto integrado de programas para manipulación de datos, cálculo y gráficos. El término “entorno” lo caracteriza como un sistema completamente diseñado y coherente, útil para el desarrollo de nuevos métodos de análisis iterativo de datos. Entre otras características dispone de almacenamiento y manipulación efectiva de datos, operadores para cálculo sobre variables indexadas, en particular matrices, entre otras. Adicionalmente existe un gran número de paquetes para R que han sido desarrollados para aplicarse en el campo de la bioinformática, por ejemplo, la búsqueda de genes en bases de datos, análisis de microarreglos de ADN, etc. [38].

En 2012 fue publicada y puesta a disposición de los investigadores un paquete en R denominado “rehh”, el cual permite realizar cálculos en datos genéticos, tomando como modelos de operación las aportaciones realizadas por *Sabeti et al.*, *Voight et al.*, y *Tang et al.* centrado en el cálculo de EHH y variables relacionadas, permitiendo la generación asistida de gráficos, diagramas y otras herramientas de cálculo de libre costo [39].

CAPÍTULO 3 METODOLOGÍA

3.1 DESCRIPCIÓN DE LOS DATOS

3.1.2 ORIGEN DE LOS DATOS

La presente tesis, parte del análisis de un compendio de haplotipos que representan la información genética de 507 especímenes de ganado bovino, los cuales abarcan 19 razas distintas pertenecientes a las subespecies del género *Bos*: *Bos taurus* y *Bos indicus*, como se muestra en la Tabla 3.

Tabla 3 Linajes y número especímenes que componen la muestra

Linaje	Abreviatura	Número de especímenes	Subespecie
Angus	ANG	27	<i>Bos taurus</i>
Brown Swiss	BSW	24	<i>Bos taurus</i>
Charolais	CHL	22	<i>Bos taurus</i>
Guernsey	GNS	21	<i>Bos taurus</i>
Hereford	HFD	27	<i>Bos taurus</i>
Holstein	HOL	59	<i>Bos taurus</i>
Jersey	JER	32	<i>Bos taurus</i>
Limousin	LMS	40	<i>Bos taurus</i>
Norwegian Red	NRC	17	<i>Bos taurus</i>
N'Dama	NDA	24	<i>Bos taurus</i>
Piedmontese	PMT	24	<i>Bos taurus</i>
Red Angus	RGU	11	<i>Bos taurus</i>
Romagnola	RMG	23	<i>Bos taurus</i>
Brahman	BRM	25	<i>Bos indicus</i>
Gir	GIR	30	<i>Bos indicus</i>
Nelore	NEL	35	<i>Bos indicus</i>
Beefmaster	BMA	24	<i>Bos taurus</i> x <i>Bos indicus</i>
Sheko	SHK	18	<i>Bos taurus</i> x <i>Bos indicus</i>
Santa Gertrudis	SGT	24	<i>Bos taurus</i> x <i>Bos indicus</i>

Los datos analizados proceden del Consorcio HapMap Bovino y corresponden a genotipos de alta densidad provenientes de la genotipificación con el BovineHD Genotyping BeadChip, el cual interroga un total de 777,962 SNPs a lo largo de todo el genoma bovino, de los cuales 742,910 se encuentran en los cromosomas autosomales, es decir, en los cromosomas no sexuales.

La fuente del material genético fue muestras de sangre o semen criopreservado de animales seleccionados de tal manera que no estuvieran emparentados en al menos de 4-5 generaciones ancestrales. Todos los linajes de la muestra pertenecen a las subespecies *Bos taurus* y *Bos indicus* como se detalla en la Figura 9.



Figura 9 Descripción de la muestra. Clasificación de las razas según su origen y descripción del propósito de crianza.

Los genotipos fueron filtrados para eliminar los errores de tecnología de “no llamado” (“not call” en inglés) ocurridos durante la genotipificación de las muestras. Posterior a dicha depuración, el valor promedio de SNPs retenidos en las 19 razas fue de 735,202 SNPs, lo que representa un 94.5% de la información inicial (ver Anexo 1). Finalmente, los haplotipos analizados en el presente estudio provienen de la inferencia mediante el algoritmo Beagle 3.3.2 [40].

3.1.3 FORMATO DE LOS DATOS

Originalmente el compendio de haplotipos se encontraba dividido en 551 archivos de texto, lo que representa los 29 cromosomas autosomales del genoma bovino, para cada una de las 19 razas de manera individual. La información de estos archivos estaba codificada en el formato de “Alelos A/B” [41], como se aprecia en el extracto de información en la Figura 10. En el primer renglón se encuentra el identificador (id) de los individuos de la muestra, cuyos primeros tres caracteres señalan la raza en cuestión; a partir del segundo renglón se encuentran los marcadores (M) iniciando con el identificador numérico del SNP y posteriormente los valores de los alelos.

```

I id ANG000001 ANG000001 ANG000002
M 16947 B B B B B B B A B A B B B
M 36337 A A A B A A A A A B A A A
M 67130 A A A A A A A A A A A A
M 78655 A A A B A A A A A A A A
M 83412 B B B B B B B A B A B B B
M 89725 A A A A A A A B A B A A A
M 100260 B B B A B B B A B A B B B
M 111645 B B B A B B B B B B B B B
M 120183 B B A B B A B B B B A B A
M 127885 A A A A A A A A A A A A
M 135098 B B B A B B B B B B B B B

```

Figura 10 Extracto del compendio de haplotipos codificados en “Alelos A/B”

3.2 PREPROCESAMIENTO DE LA INFORMACIÓN

3.2.1 ESTRUCTURA DE DATOS

Debido al alto grado de segregación que la información presentaba, se optó por disponer de ella en una estructura de datos tipo *hash en PERL*, dado que se requiere asociar diversos campos de información a un mismo elemento en particular.

Se determinó el valor de la base nucleotídica que le correspondía a cada alelo basados en la anotación del genoma bovino UMD 3.1 (como se observa en la Figura 11), y se asoció la identidad de los alelos ancestrales y derivados haciendo uso de los valores putativos publicados por Rocha *et al.* en [31].

```
ANG000001  ANG000001  ANG000002
16947  CCCCCCACACCCCCCCCCCCCCACCAC(
36337  AAAGAAAAAGAAAAAGAAAAAGGAAAI
67130  AAAAAAAAAAAAAAAAAAAAAAAAAAAAI
78655  AAAGAAAAAGAAAAAGAAAAAGGAAAI
83412  GGGGGGAGAGGGGGGGGGGGGAGGAG(
89725  TTTTTTCTCTTTTTTTTTTTTCTTCT(
100260  CCCTCCCTCTCCCCCCTCCCTTCT(
111645  CCCACCCCCCCCCCCCCACCCCACCC(
120183  GGTGGTGGGTGTTTTTGGGGTGGTG(
127885  TTTTTTTTTTTTTTTTTTTTTTTTTT(
135098  GGGAGGGGGGGGGGGGGAGGGGGAGG(
```

Figura 11 Extracto del compendio de haplotipos codificados en bases nucleotídicas “ACGT”.

3.2.2 FILTROS DE CALIDAD

Al compendio de haplotipos preprocesados se aplicó las siguientes medidas como filtros de calidad: se eliminaron los SNPs monomórficos, marcadores con frecuencias de alelo menor inferior a 0.05 (MAF <0.05) y aquellos marcadores cuya información sobre alelos ancestrales y derivados era desconocía (ver Anexo 2).

3.2.3 ARCHIVOS DE ENTRADA A LA PAQUETERÍA “rehh”

A partir de la información almacenada en el *hash* se generaron los archivos de entrada a la paquetería de análisis *rehh* de R Project, siguiendo las especificaciones particulares de formato. A continuación, se detallan los archivos de texto generados:

3.2.3.1 ARCHIVOS DE HAPLOTIPOS

Se crearon siguiendo el formato estándar de haplotipos [42], donde cada renglón representa un haplotipo; el primer elemento es el ID del haplotipo y los SNPs de los haplotipos se encuentran separados en columnas (de manera alternada para las dos cadenas de secuencias). Los alelos se codificaron de la siguiente manera: 0 para alelos faltantes, 1 para alelos ancestrales y 2 para alelos derivados (ver Figura 12 a).

3.2.3.2 ARCHIVOS CON INFORMACIÓN DE SNPS

Estos archivos brindan información específica de cada uno de los SNP en la muestra. Al inicio de cada renglón se encuentra el nombre del SNP, seguido del cromosoma de origen, su posición en el cromosoma, el alelo ancestral y el alelo derivado, codificados como 1 y 2 respectivamente. Los SNPs se acomodaron en el mismo orden de aparición que en los archivos de haplotipos (ver Figura 12 b).

a)								b)					
ANG000001	1	1	1	1	2	2	2	1	BovineHD0100000005	1	16947	1	2
ANG000002	1	1	1	1	2	2	2	1	BovineHD0100000015	1	36337	1	2
ANG000003	1	1	1	1	2	2	2	1	BovineHD0100000024	1	67130	1	2
ANG000004	1	2	1	2	2	2	1	2	BovineHD0100000026	1	78655	1	2
ANG000005	1	1	1	1	2	2	2	1	BovineHD0100000027	1	83412	1	2
ANG000006	1	1	1	1	2	2	2	1	BovineHD0100046367	1	89725	1	2
ANG000007	1	1	1	1	2	2	2	1	BovineHD0100000030	1	100260	1	2
ANG000008	2	1	1	1	1	1	1	1	BovineHD0100000033	1	111645	1	2
ANG000009	1	1	1	1	2	2	2	1	BovineHD0100000035	1	120183	1	2
ANG000010	2	2	1	1	1	1	1	1	BovineHD0100000037	1	127885	1	2

Figura 12 Extracto de los archivos de entrada a la paquetería de *rehh*: a) archivo de haplotipos, b) archivo con información de SNPs

3.3 PROCESAMIENTO DE LA INFORMACIÓN

3.3.1 CÁLCULO DE LOS ESTADÍSTICOS DE iHS

La siguiente secuencia de instrucciones se programó y anidó en una estructura de control tipo *for* para calcular el estadístico iHS en las razas bovinas utilizando la paquetería *rehh* en R.

Los archivos de entrada tipo texto (haplotipos e información de SNPs) se convirtieron en objetos de clase tipo “*haplohh*” mediante la función *data2haplohh*, utilizando los siguientes argumentos: *min_perc_genos.hap*=100, *min_perc_genos.snp*=100, *min_maf*=0.05. Mediante la función *scan_hh*, se calculó el IHH para los alelos ancestrales y derivados, así como el IES para todos los SNPs del objeto *haplohh*, empleando sus parámetros predeterminados: *limehh*=*limehhs*=0.05 y *limhaplo*=2.

Como resultado se obtiene un objeto tipo *res.scan*, que es una matriz de dimensiones [SNPs declarados x 6], donde las columnas indican: el número de cromosoma, la posición del SNP, la frecuencia del alelo ancestral, el valor de iHH para el alelo ancestral (iHHa), el valor de iHH para el alelo derivado (iHHd) y finalmente el valor de IES. Para cada linaje bovino se generó un archivo en el que se concatenó las matrices de los 29 cromosomas en una sola matriz resultante como se muestra en la Figura 13.

	CHR	POSITION	FREQ_a	IHHa	IHHd	IES
BovineHD0100000005	1	16947	0.85185185	281389.085	422472.24	284069.928
BovineHD0100000015	1	36337	0.81481481	327174.752	412162	323696.893
BovineHD0100000026	1	78655	0.88888889	257848.269	1973339.62	265908.835
BovineHD0100000027	1	83412	0.14814815	601500.907	282400.31	289205.109
BovineHD0100046367	1	89725	0.14814815	601500.907	282400.31	289205.109
BovineHD0100000030	1	100260	0.25925926	240335.919	402885.142	386195.512
BovineHD0100000033	1	111645	0.88888889	252655.581	1973339.62	260771.982
BovineHD0100000035	1	120183	0.59259259	226547.025	1016248.01	454829.146

Figura 13 Extracto del objeto *res.scan*. La concatenación de los 29 objetos (uno por cada cromosoma) es la matriz resultante de la función *scan_hh* para cada linaje

3.3.2 ESTANDARIZACIÓN DE iHS

Para cada linaje se calculó los valores de iHS estandarizados utilizando la matriz de estadísticos iHH mediante la función *ihh2ihs* con el argumento *freqbin=0*, la cual establece que cada frecuencia observada sea tratada como una marca de clase, al momento de estandarizar los valores de iHH. Esta asignación se recomienda en estudios donde la muestra cuente con un gran número de SNPs y pocos haplotipos como se indica en [42]. Como resultado se obtuvo un objeto tipo *res.ihs*, que es una lista con dos elementos: una matriz de iHS de los SNPs con sus respectivos valores p_{iHS} , asumiendo que los valores de *iHS* se distribuyen de acuerdo con la hipótesis de neutralidad (Figura 14 a), así como una matriz donde se presentan los parámetros estadísticos para datos agrupados basados en las frecuencias observadas (Figura 14 b).

	CHR	POSITION	iHS	Pvalue
BovineHD0100000005	1	16947	0.385069482	0.154786616
BovineHD0100000015	1	36337	0.543783386	0.231664943
BovineHD0100000026	1	78655	-0.738822867	0.337228418
BovineHD0100000027	1	83412	0.045384866	0.016012833
BovineHD0100046367	1	89725	0.045384866	0.016012833
BovineHD0100000030	1	100260	-0.657374754	0.291630113
BovineHD0100000033	1	111645	-0.751650509	0.344610611
BovineHD0100000035	1	120183	-0.934781447	0.456054887

	Freq Class	Size	Mean iHH	SD iHH
1	0.074074074	10797	-4.388088401	1.185624395
2	0.111111111	12418	0.559971499	1.666509064
3	0.148148148	14036	0.689160036	1.47499397
4	0.185185185	14788	0.674981724	1.363346521
5	0.222222222	15593	0.521997359	1.318717311
6	0.259259259	16510	0.348722363	1.316351544
7	0.296296296	16780	0.265011763	1.296124147
8	0.333333333	18055	0.185562274	1.29474639

Figura 14 Extracto del objeto *res.ihs* donde se muestra en a) los valores estandarizados de iHS y en b) la información sobre la estandarización

3.3.3 VALORES p

A partir de los valores de probabilidad p_{iHS} , donde $p_{iHS} = -\log[1 - 2|\Phi(iHS) - 0.5|]$, y $\Phi(x)$ representa la función de distribución acumulativa gaussiana de los valores de iHS, se calcularon los valores p como en [23], debido a que el p_{iHS} es el valor p de la prueba estadística de dos colas asociado a la hipótesis de neutralidad, es decir, la hipótesis nula de “no selección” para dicho

estadístico de iHS , el valor p_{iHS} se puede interpretar como $\log_{10}\left(\frac{1}{p}\right)$. Se generaron gráficas de Manhattan para observar la distribución de los valores de p_{iHS} a lo largo del genoma utilizando la función *ihplot* (ver Anexo 3).

3.3.4 REGIONES GENÓMICAS SIGNIFICATIVAS

Para establecer regiones de búsqueda donde interrogar señales de selección se calcularon los valores absolutos de los estadísticos de prueba iHS , se promediaron a lo largo de todo el genoma en regiones sin traslape de 500 Kb de longitud como en [43], [44] y se identificaron como $|iHS|$. Se generaron gráficas de Manhattan para observar la distribución de los valores de $|iHS|$ estandarizados en todo el genoma, y se establecieron como regiones significativas aquellas que superaron el valor extremo 2.33, que representa un intervalo de confianza del 98 % ($\alpha=0.02$), similar como en [23], [42], [43].

3.3.5 INFORMACIÓN BIOLÓGICA DE INTERÉS

Debido a la magnitud de regiones significativas detectadas en las 19 razas bovinas, para inspeccionar la presencia de información biológica de interés se limitó a utilizar las cinco regiones con las señales de selección más intensas.

3.3.5.1 GENES

Para inspeccionar la presencia de genes se desarrolló un algoritmo iterativo en R que utiliza el paquete estadístico *biomaRt* como interfase entre la terminal y la base de datos *btaurus_gene_ensembl* disponible en el sitio BioMart. Los parámetros de entrada que se utilizaron para realizar la búsqueda fueron: el número de cromosoma, el inicio y fin de la región (ambos en Kb). Como resultado de cada iteración se generó un archivo con la información necesaria para identificar los genes.

3.3.5.2 QTL

Para investigar la presencia de QTL se descargó y procesó el compendio de QTL reportados y almacenados en la base de datos *AnimalQTLdb*. Mediante un algoritmo en R y de manera similar en la búsqueda de genes se acotaron las regiones de búsqueda y se generaron una serie de archivos donde se identifican y describen los QTL detectados.

CAPÍTULO 4 ANÁLISIS Y DISCUSIÓN DE RESULTADOS

4.1 RESOLUCIÓN DEL ESTUDIO

Después de aplicar los filtros de calidad, el número promedio de marcadores retenidos por cromosoma fue de 546,961 SNPs, lo que equivale al 73.62% de los SNPs autosomales comprendidos en el BovineHD Genotyping BeadChip (ver Anexo 2), abarcando 2,512.08 Mbp del genoma bovino (UMD3.1) con una distancia media de 4.59 Kb entre marcadores adjuntos (ver Tabla 4).

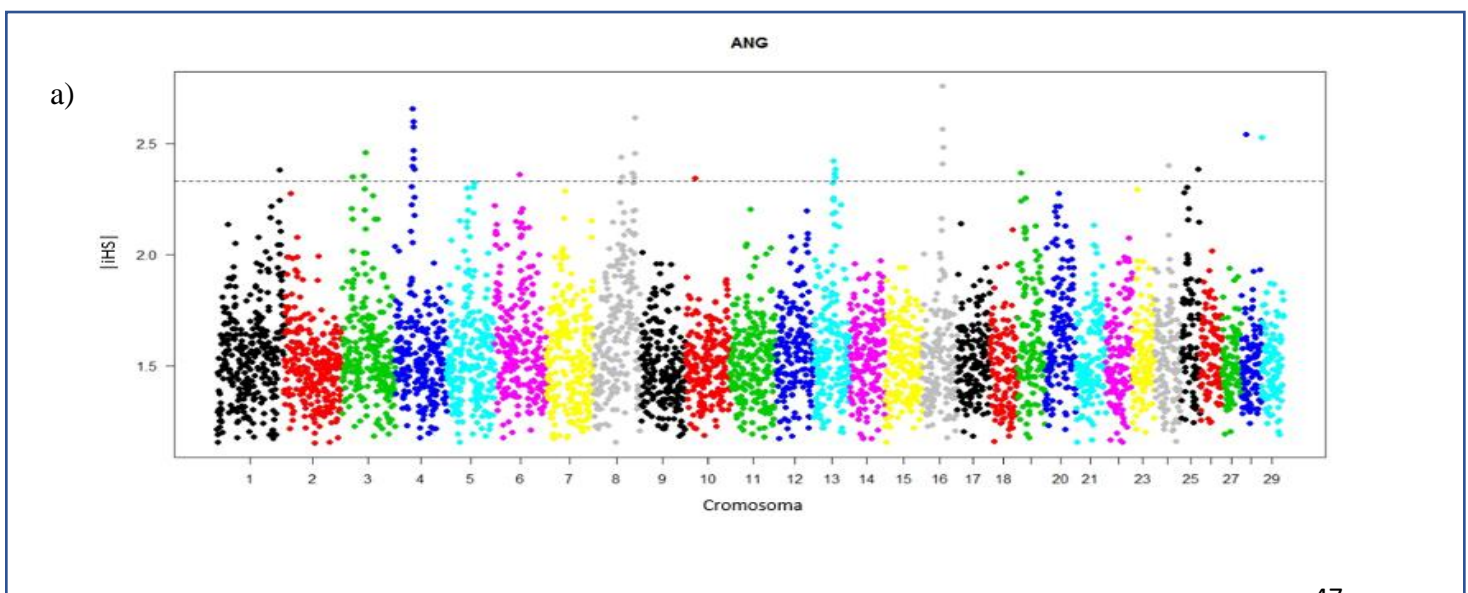
Tabla 4 Distancias promedio entre marcadores adjuntos en cromosomas autosomales

Cromosoma	Número de SNPs promedio	Longitud del cromosoma (Mb)	Distancia promedio (Kb)
1	34,563	158.34	4.58
2	29,405	137.06	4.66
3	26,371	121.43	4.60
4	25,924	120.83	4.66
5	25,114	121.19	4.83
6	26,764	119.46	4.46
7	24,552	112.64	4.59
8	22,242	113.38	5.10
9	23,454	105.71	4.51
10	23,148	104.31	4.51
11	24,051	107.31	4.46
12	19,471	91.16	4.68
13	15,863	84.24	5.31
14	16,692	84.65	5.07
15	18,636	85.3	4.58
16	18,163	81.72	4.50
17	17,111	75.16	4.39
18	15,148	66	4.36
19	14,388	64.06	4.45

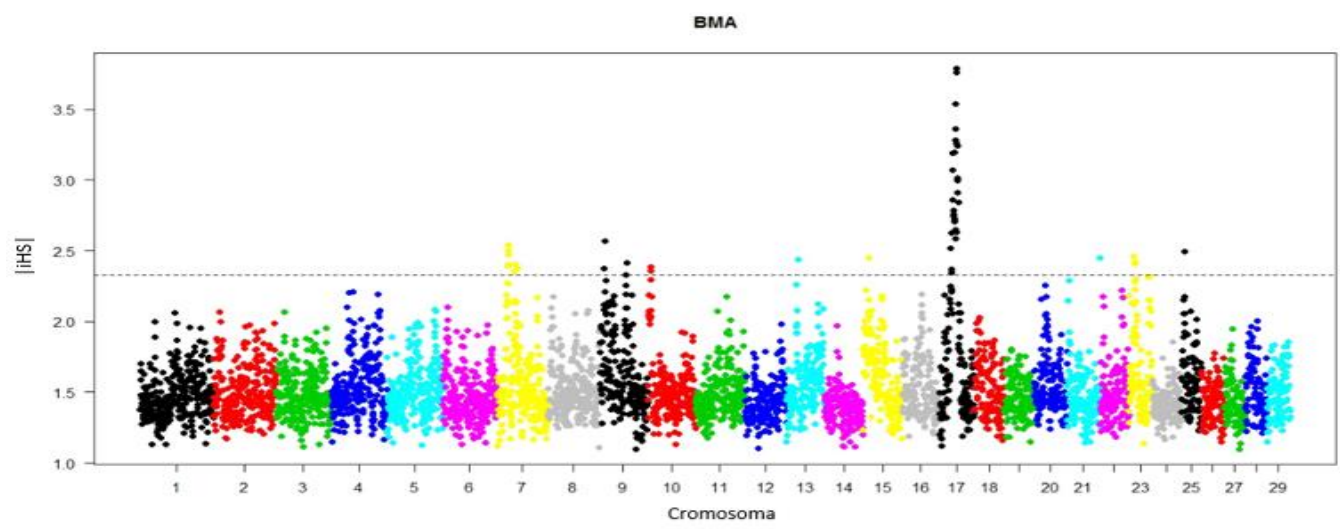
20	16,506	72.04	4.36
21	15,856	71.6	4.52
22	14,022	61.44	4.38
23	11,933	52.53	4.40
24	14,043	62.71	4.47
25	10,028	42.9	4.28
26	11,785	51.68	4.39
27	10,297	45.41	4.41
28	10,303	46.31	4.49
29	11,129	51.51	4.63
Total	546,961	2,512.08	4.59

4.2 ESTADÍSTICOS DE *iHS*

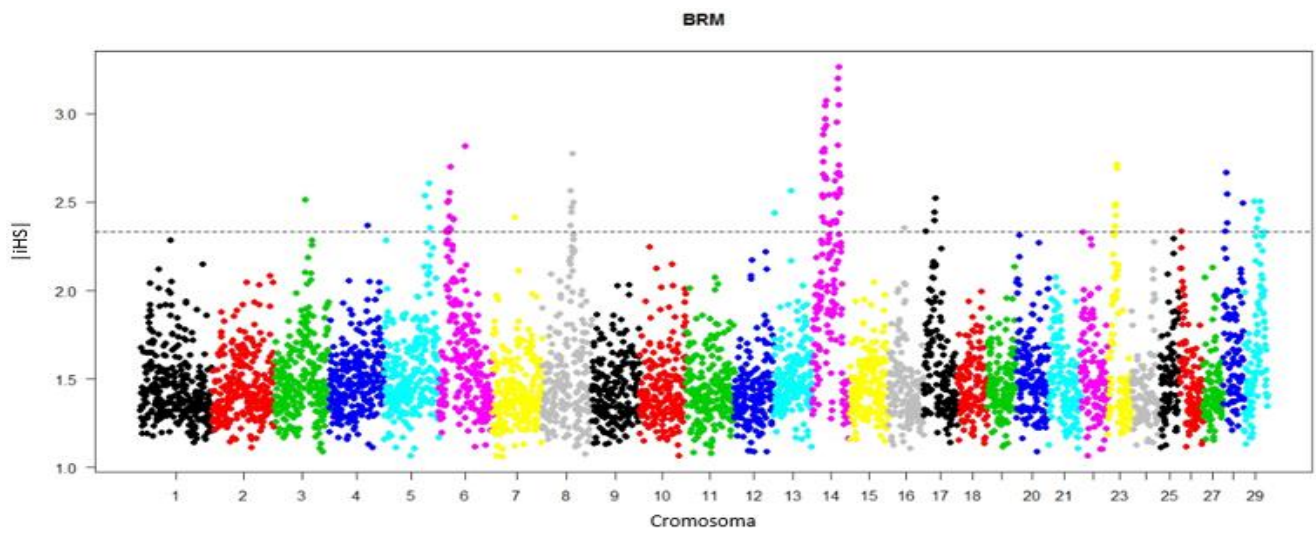
En la Figura 15 se presentan las gráficas con la distribución de los valores de $|iHS|$ a lo largo del genoma para cada uno de los linajes bovinos. En el eje horizontal se tienen los cromosomas autosomales, cada punto representa una región de 500 Kb, y en el eje vertical se presenta el promedio del valor absoluto de los estadísticos de *iHS* estandarizados. La línea punteada horizontal indica el umbral de significancia considerando un α de 0.02, es decir, aquellos valores extremos de $|iHS| > 2.33$ [23], [43], [44]. Con dicho nivel de exigencia estadística se detectaron regiones significativas en diversos cromosomas en todos los linajes bovinos.



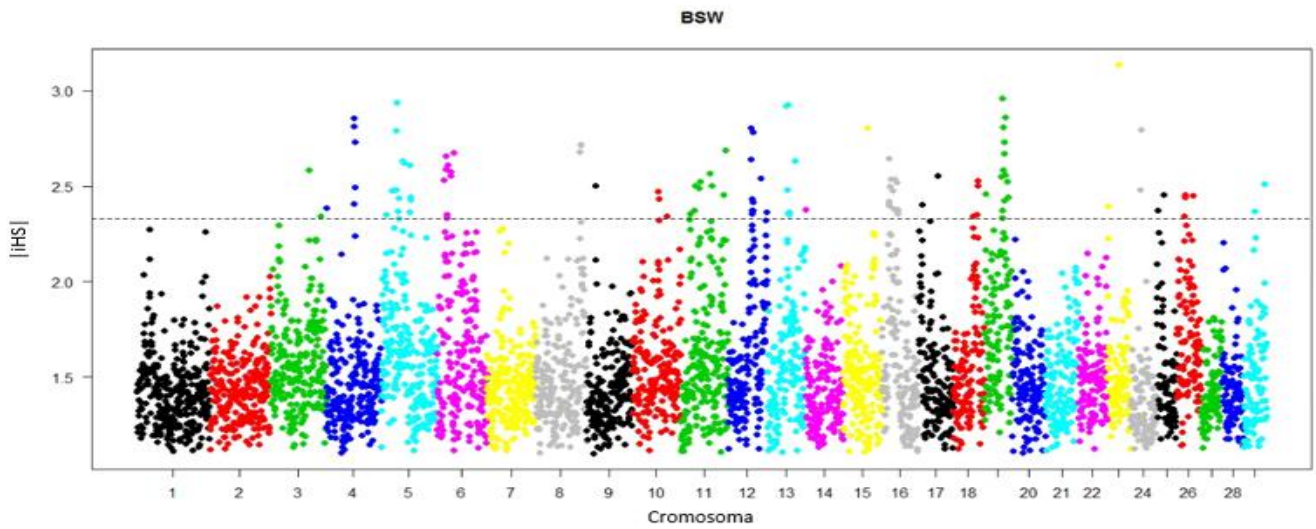
b)



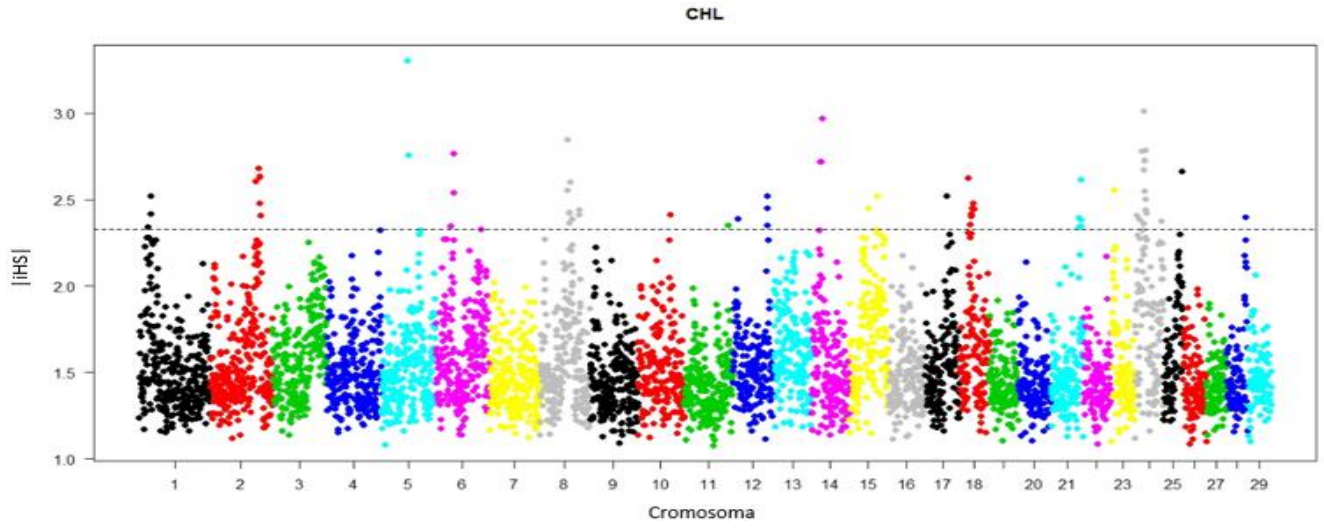
c)



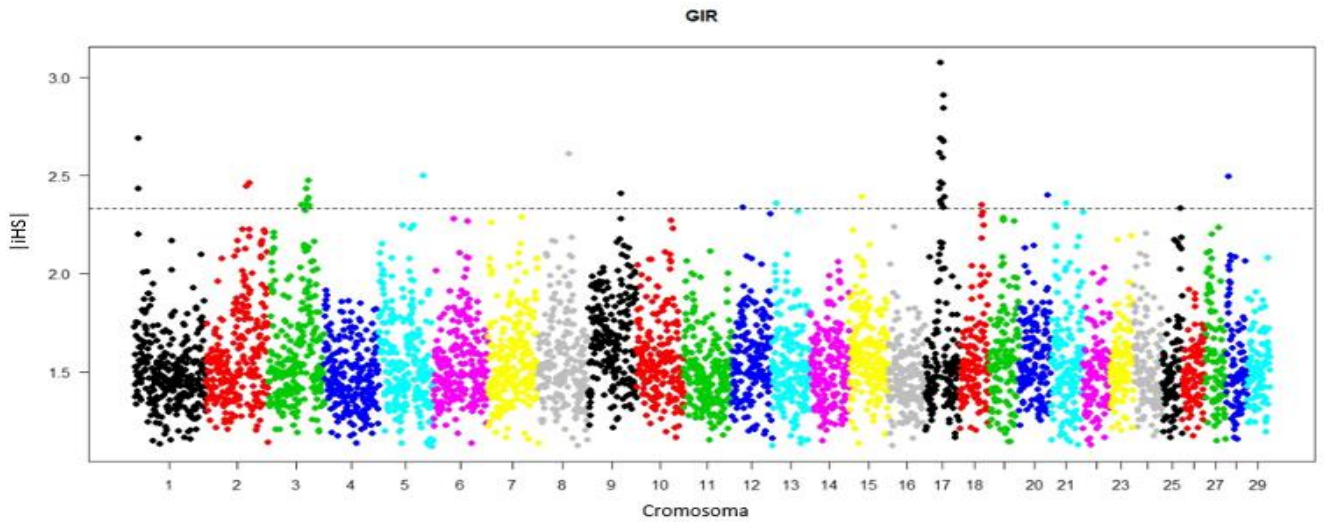
d)



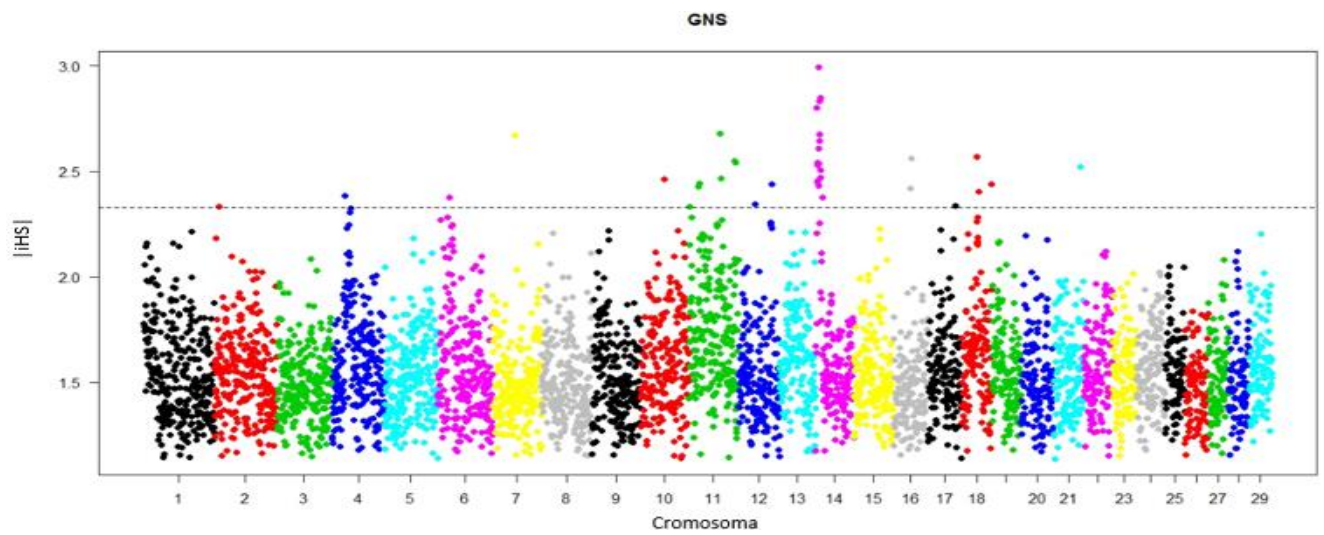
e)



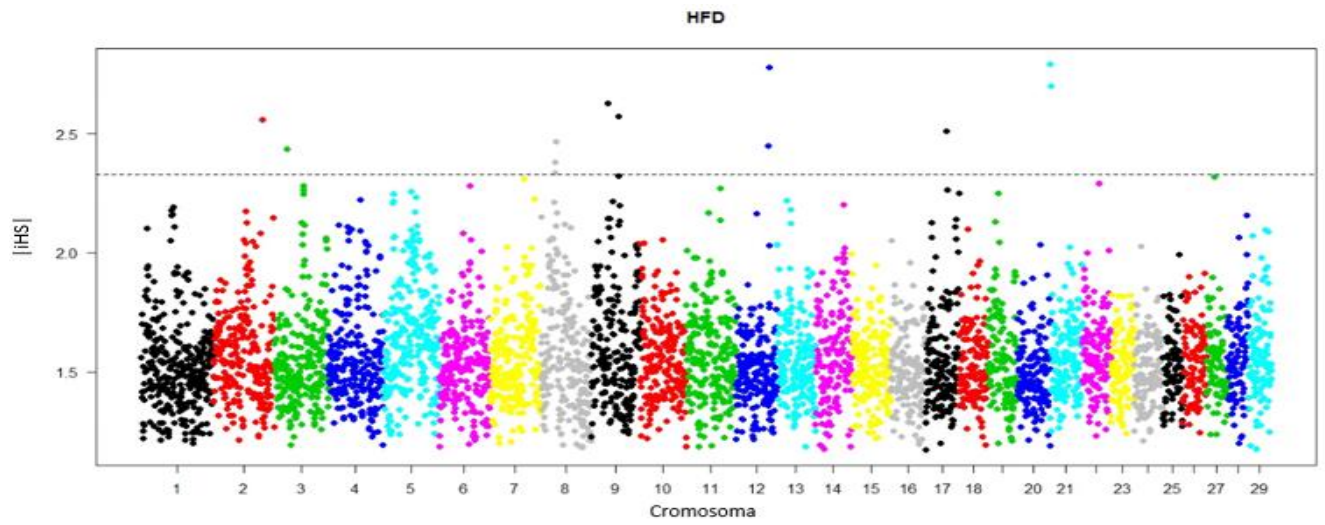
f)



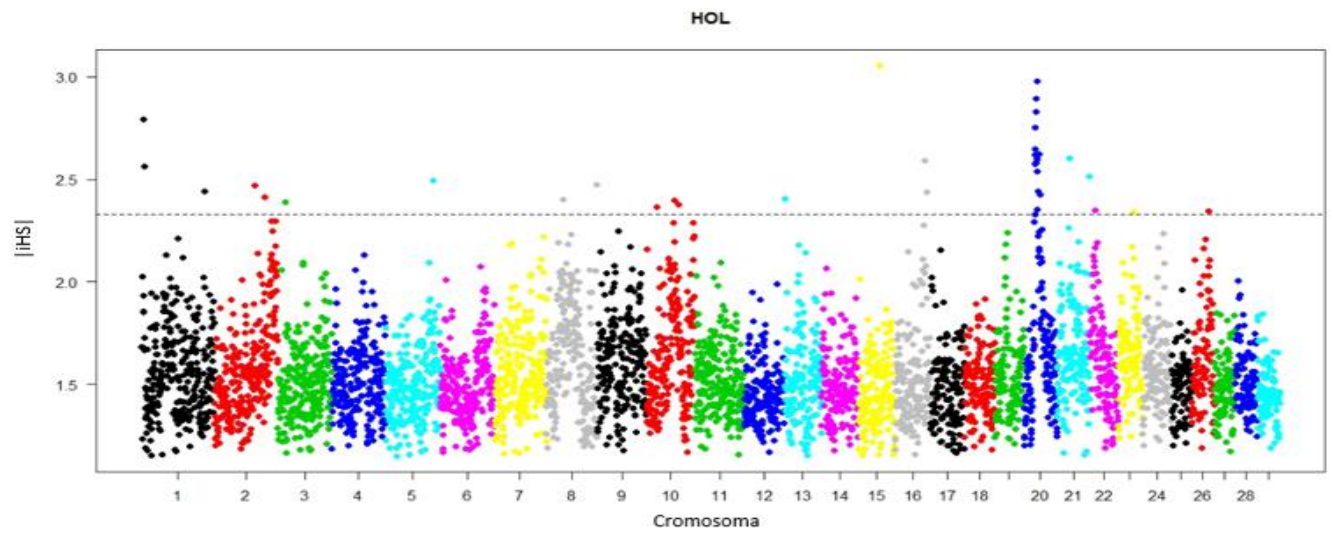
g)



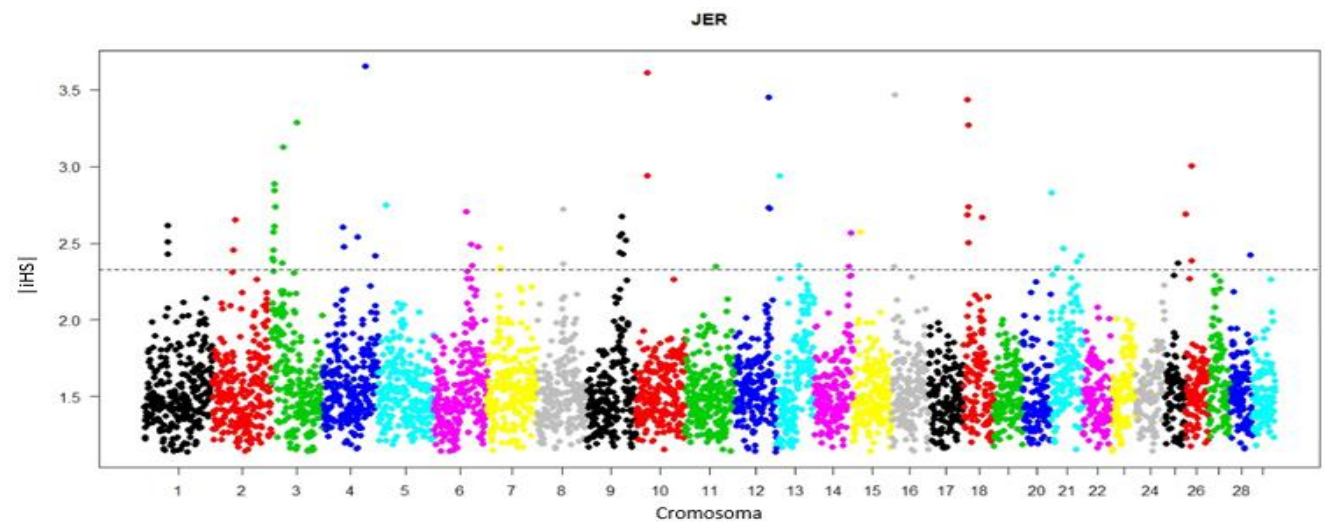
h)



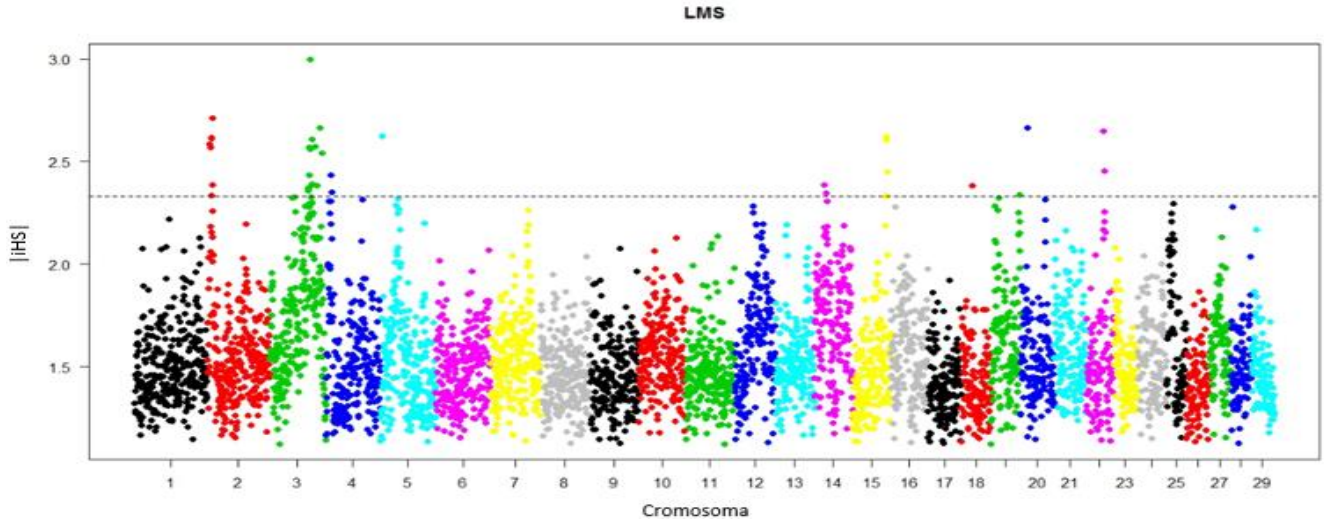
i)



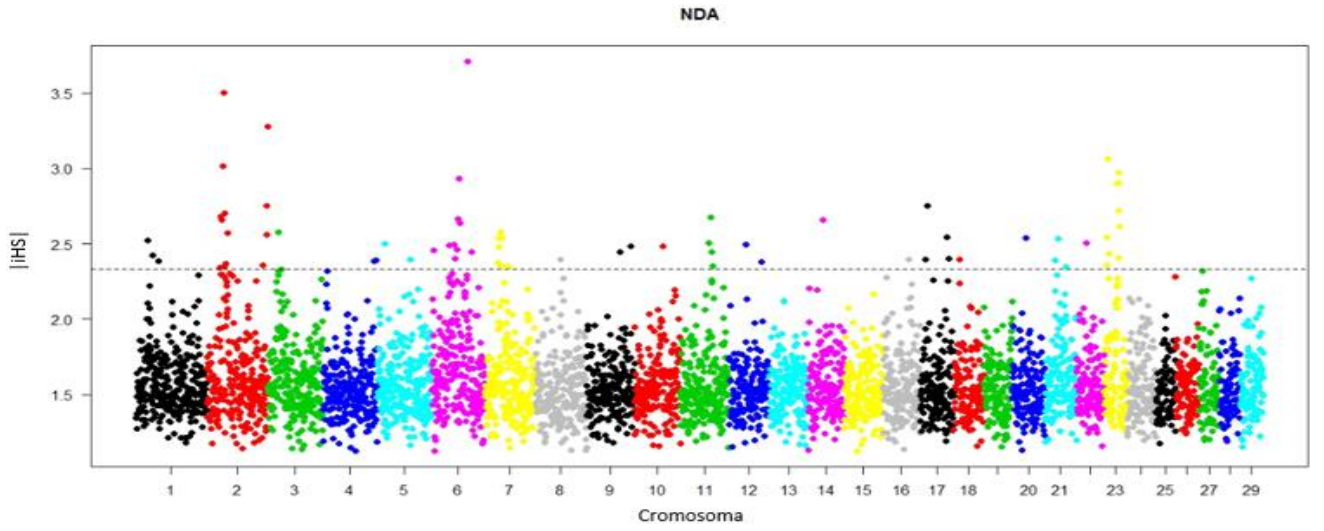
j)



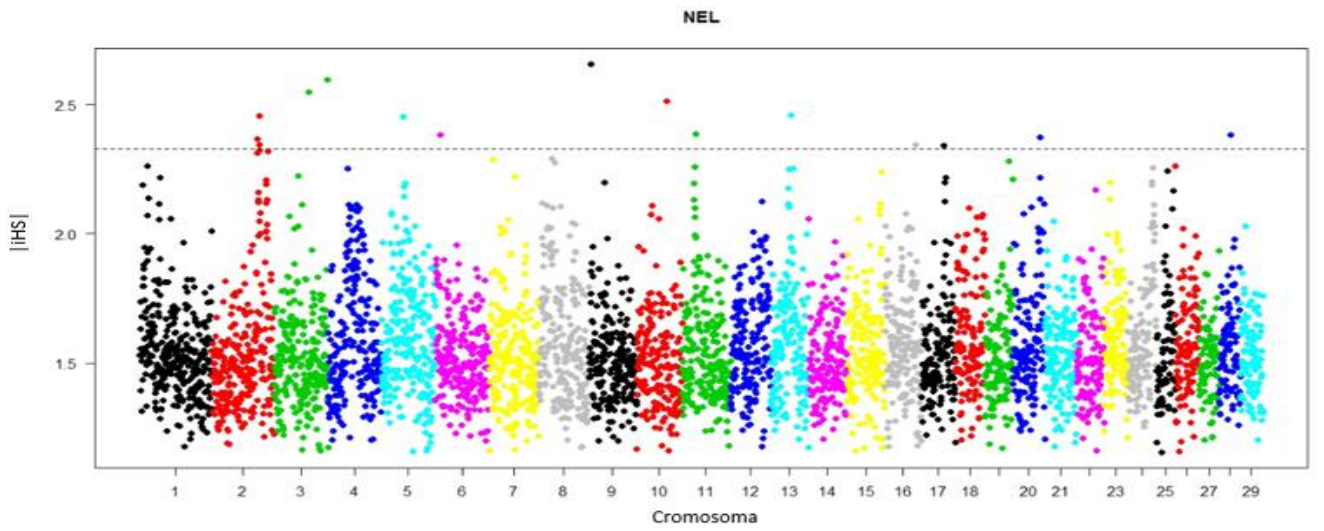
k)



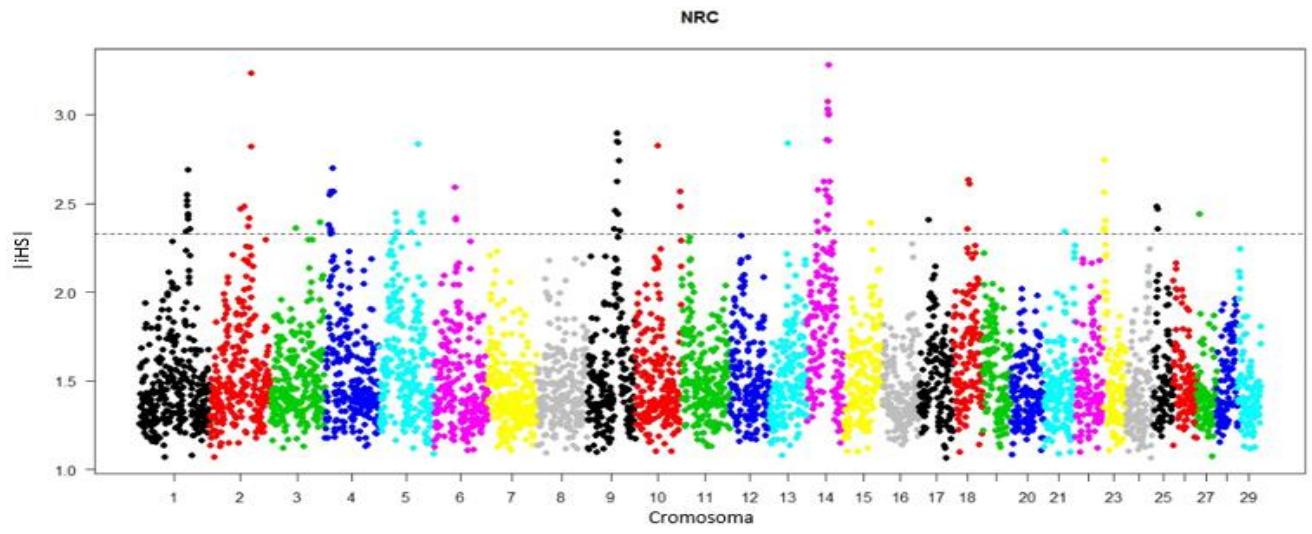
l)



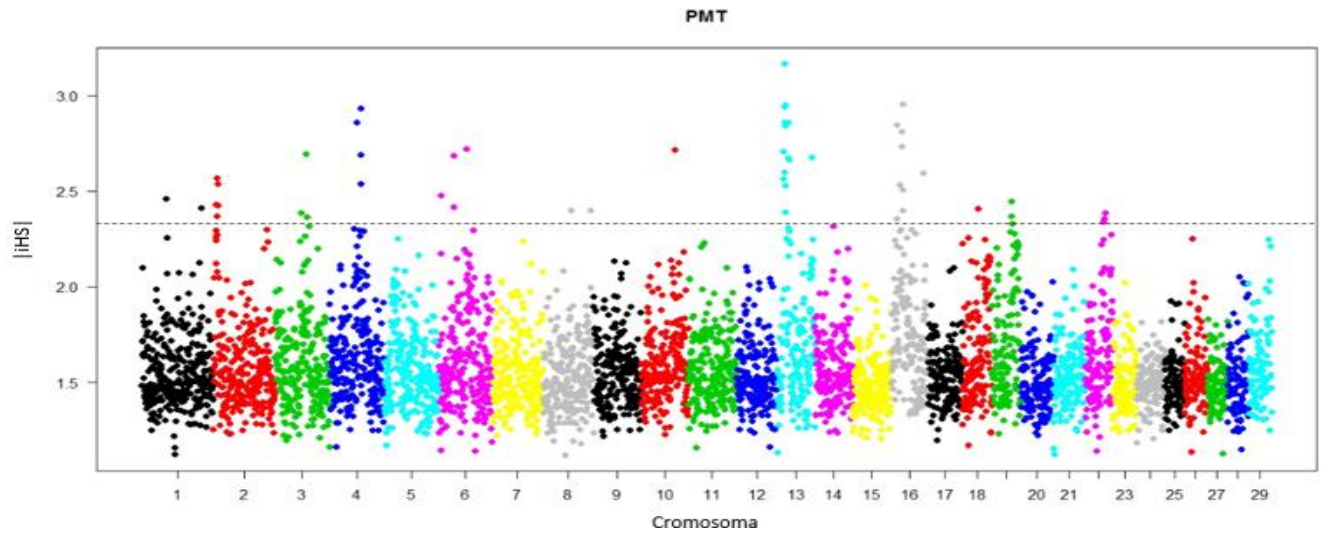
m)



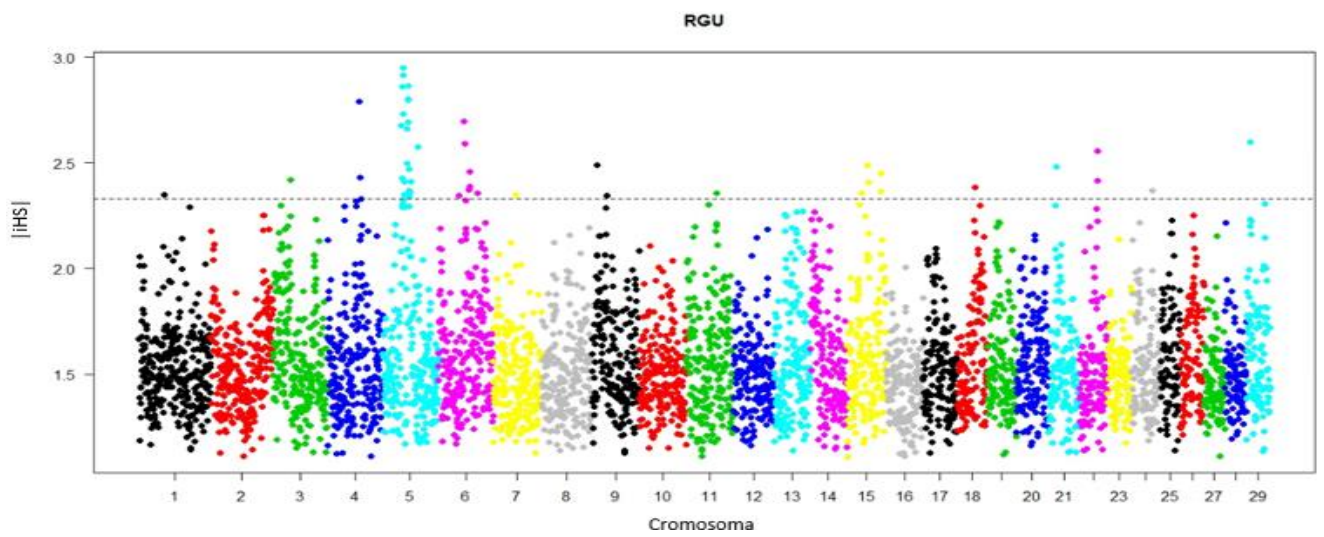
n)



ñ)



o)



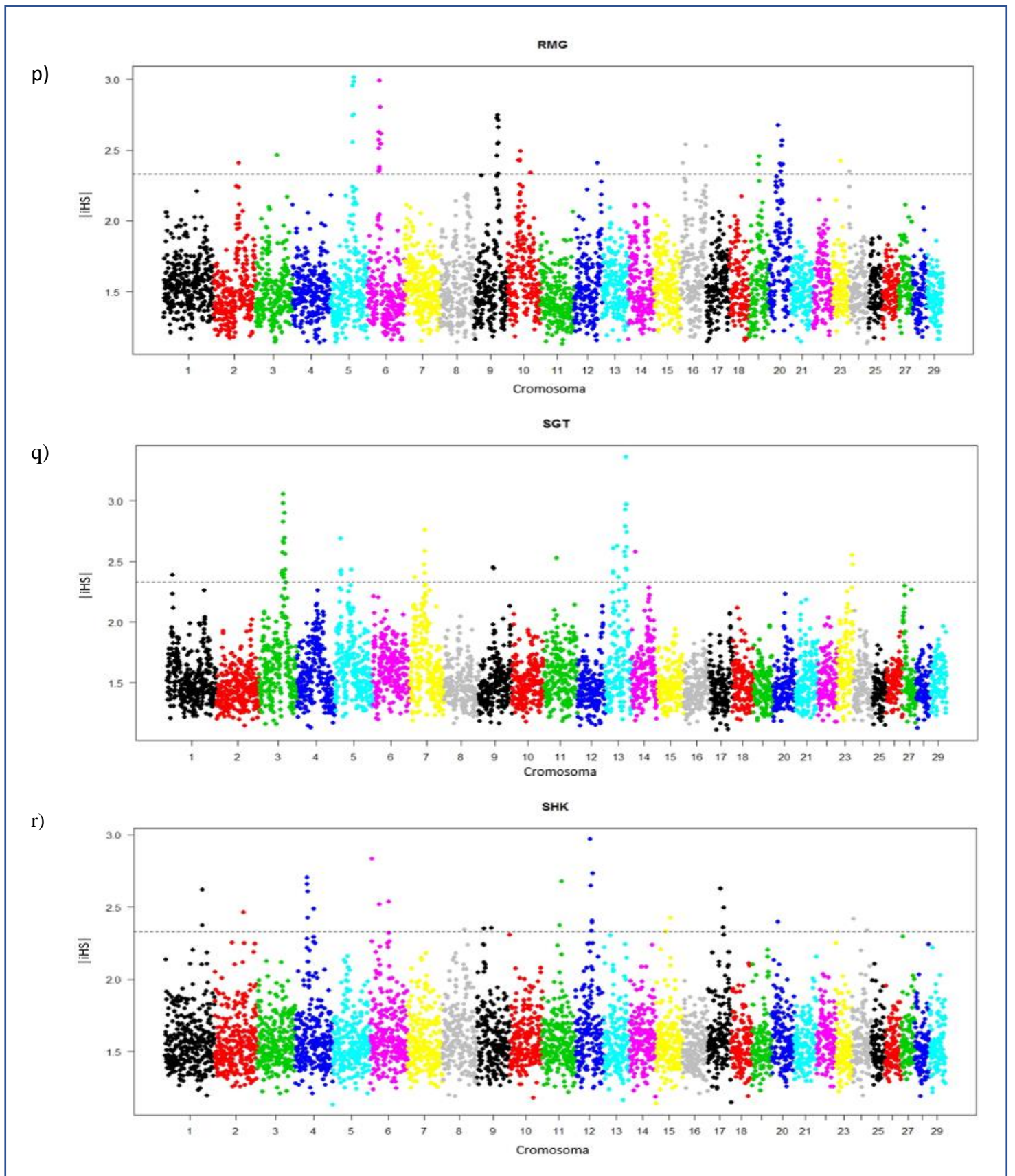


Figura 15 Distribución de los valores de $|iHS|$ en el genoma

En las gráficas se pueden observar claramente el patrón adecuado para detectar una señal de selección, de acuerdo con el fundamento de análisis de *iHS*, es decir, cúmulos de marcadores que forman picos claramente marcados de gran altura, en conjunción de exceder un umbral de significancia.

Los cúmulos más prominentes se encuentran en: BMA (cromosoma 17), JER (cromosoma 18), NDA (cromosomas 23) y PMT (cromosoma 13). En el resto de los linajes se aprecian múltiples cúmulos de altura variable.

4.3 REGIONES GENÓMICAS Y SEÑALES DE SELECCIÓN

En la Tabla 5 se muestra la cantidad de regiones de 500 Kb utilizadas para promediar los valores de $|iHS|$ en cada una de las razas, así como la proporción de regiones de interés según los criterios de significancia estadística.

Tabla 5 Proporción de Regiones significativas

Raza	Número de regiones	Regiones significativas	% Regiones significativas
ANG	4,868	33	0.68
BMA	4,990	49	0.98
BRM	4,964	93	1.87
BSW	4,760	111	2.33
CHL	4,982	65	1.30
GIR	4,909	37	0.75
GNS	4,800	37	0.77
HFD	4,849	12	0.25
HOL	4,829	36	0.75
JER	4,731	65	1.37
LMS	4,892	35	0.72

NDA	4,933	72	1.46
NEL	4,875	15	0.31
NRC	4,953	84	1.70
PMT	5,011	51	1.02
RGU	4,867	46	0.95
RMG	4,913	47	0.96
SGT	4,998	50	1.00
SHK	5,008	30	0.60

El valor promedio de regiones por raza fue de 4,902, siendo 5,011 el valor más alto para la raza PMT y 4,731 el valor inferior en la raza JER. Esta variación se debe a que, en algunas regiones, el valor promediado de $|iHS|$ fue de cero, es decir, no se contaba con algún marcador en dicha región y por ende fueron descartadas.

El valor promedio de regiones significativas por raza fue de 51, con un valor máximo de 111 regiones en la raza BSW y un valor mínimo de 12 en HFD. En términos de porcentaje el valor promedio de regiones significativas es del 1%, con un valor máximo en BSW de 2.33 % y un valor mínimo en HFD de 0.25 %. El conteo y las regiones significativas por cromosoma de cada uno de los linajes se encuentran en los Anexos 4 y 5.

4.4 CROMOSOMAS

Para analizar de manera cuantitativa la distribución de las regiones significativas graficadas en la Figura 15, en la Tabla 6 se presentan los cromosomas (Cr) con el mayor número de regiones significativas (# R). En la tabla se omiten los cromosomas donde se presenta un empate con la cantidad mínima de 1 región significativa por cromosoma o un empate entre 5 o más cromosomas con la misma cantidad de regiones.

Tabla 6 Distribución de regiones significativas

ANG		BMA		BRM		BSW		CHL		GIR		GNS	
Cr	# R	Cr	# R	Cr	# R	Cr	# R	Cr	# R	Cr	# R	Cr	# R
4	7	17	29	14	40	5	14	24	14	17	15	14	16
8	7	7	8	6	13	19	14	8	9	3	7	11	7
13	4	9	3	23	7	16	13	18	7	1	2	18	3
16	4	23	3	8	6	12	10	2	6	2	2	12	2
3	3	10	2	29	6	6 11	9	12 21	5	-	-	16	2
HFD		HOL		JER		LMS		NDA		NEL		NRC	
Cr	# R	Cr	# R	Cr	# R	Cr	# R	Cr	# R	Cr	# R	Cr	# R
8	3	20	15	3	11	3	13	2	13	2	3	14	19
9	2	1	3	9	6	2	7	6	10	3	2	1	9
12	2	10	3	18	6	15	5	23	10	-	-	9	9
21	2	2	2	4	5	4	2	7	8	-	-	4	8
-	-	8 16 21		21	5	14 22		11 17	4	-	-	5	8
PMT		RGU		RMG		SGT		SHK		/			
Cr	# R	Cr	# R	Cr	# R	Cr	# R	Cr	# R				
13	15	5	20	6	11	13	17	12	6				
16	9	6	7	9	8	3	16	4	5				
2	5	15	5	20	7	5	5	6	3				
4	4	4	2	5	6	7	5	17	3				
6	4	9 22		10	5	9 23	2	-	-				

El valor máximo de regiones significativas se presenta en el cromosoma 14 de la raza BRM con un total de 40 regiones; en segunda posición se tiene el cromosoma 17 de la raza BMA con 29 regiones; y en tercer lugar se tiene el cromosoma 5 de la raza RGU con 20 regiones significativas.

Para analizar las coincidencias entre los cromosomas con la mayor cantidad de regiones significativas, se distribuyeron los linajes bovinos en tres grupos, atendiendo su propósito de crianza como sigue:

- Producción de Leche: Guernsey (GNS), Jersey (JER), Brown Swiss (BSW), Holstein (HOL) y Norwegian Red (NRC).
- Producción de Carne: Angus (ANG), Charolais (CHL), Hereford (HFD), Limousin (LMS), Piedmontese (PMT), Red Angus (RGU), Romagnola (RMG), Brahman (BRM), Santa Gertrudis (SGT).
- Multipropósito: Beefmaster (BMA), N'Dama (DNA), Nelore (NEL), Gir (GIR), Sheko (SHK).

Adicionalmente, los cromosomas en consideración son aquellos que presentan al menos dos regiones significativas detectadas. Los resultados se presentan en la Figura 16.

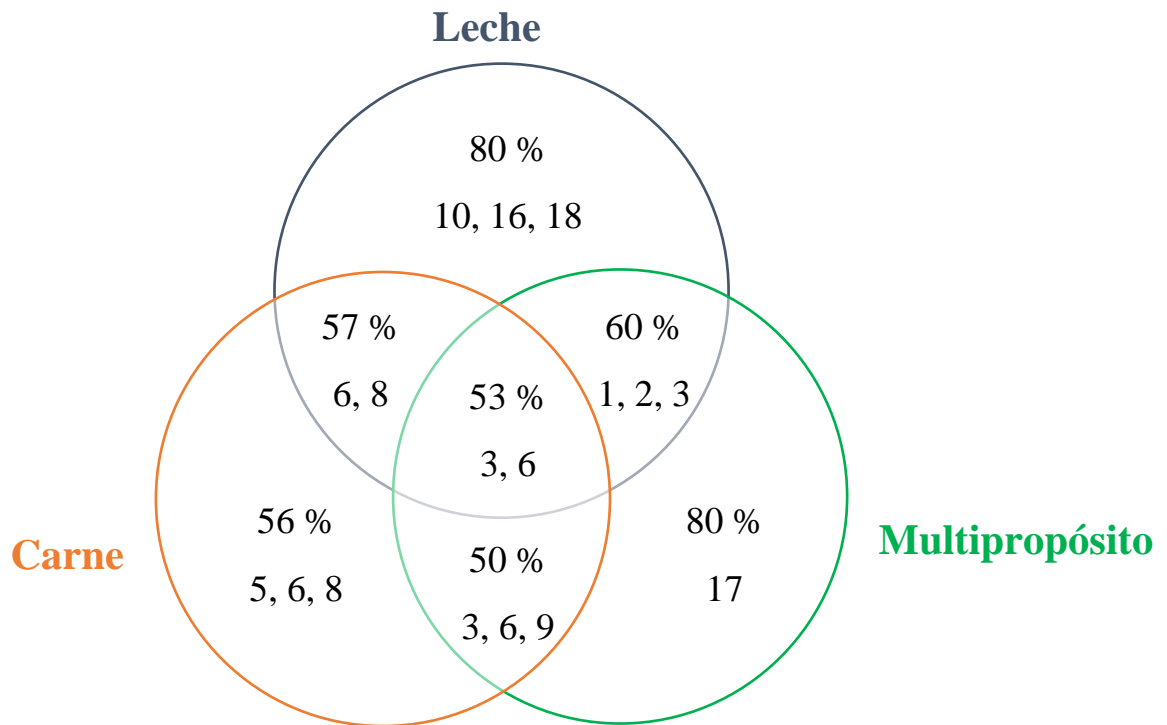


Figura 16 Coincidencias en cromosomas con señales de selección

Se puede observar que el 80% de los linajes destinados a la producción de leche tienen una coincidencia de al menos dos regiones significativas en los cromosomas 10, 16 y 18. De igual modo, en el grupo multipropósito se tiene una coincidencia del 80% en el cromosoma 17. Por su parte, en el ganado de carne existe una coincidencia del 56% en los cromosomas 5, 6 y 8.

Respecto a comparaciones entre grupos, el porcentaje de coincidencia mayor se obtuvo en la comparación Leche-Multipropósito, con un valor de 60% en los cromosomas 1, 2 y 3; el valor menor se obtuvo en la comparación Carne-Multipropósito con 50% en los cromosomas 3, 6 y 9. Finalmente al comparar los cromosomas de los tres grupos, se tiene una coincidencia del 53% respecto a los cromosomas 3 y 6.

4.5 ELEMENTOS GENÉTICOS DETECTADOS

4.5.1 GENES

El análisis de las regiones genómicas comprendidas por las señales de selección en la base de datos *btaurus_gene_ensembl* de BioMart, reveló la presencia de gran cantidad de elementos genéticos anotados, principalmente genes formalmente reconocidos, así como pseudogenes y transcritos de ARN (ver Anexo 6).

Para visualizar esta información, se calculó una proporción de la siguiente manera: se dividió el total de elementos anotados entre la cantidad de regiones significativas en cada raza y se graficaron los valores, como se muestra en la Figura 17.

La gráfica generada revela que existe una distribución similar en la proporción de elementos anotados entre los linajes bovinos, con un valor promedio de 3.23; el valor superior se presentó en la raza BSW con 4.99 y el valor inferior en la raza BMA con una proporción de 1.53.

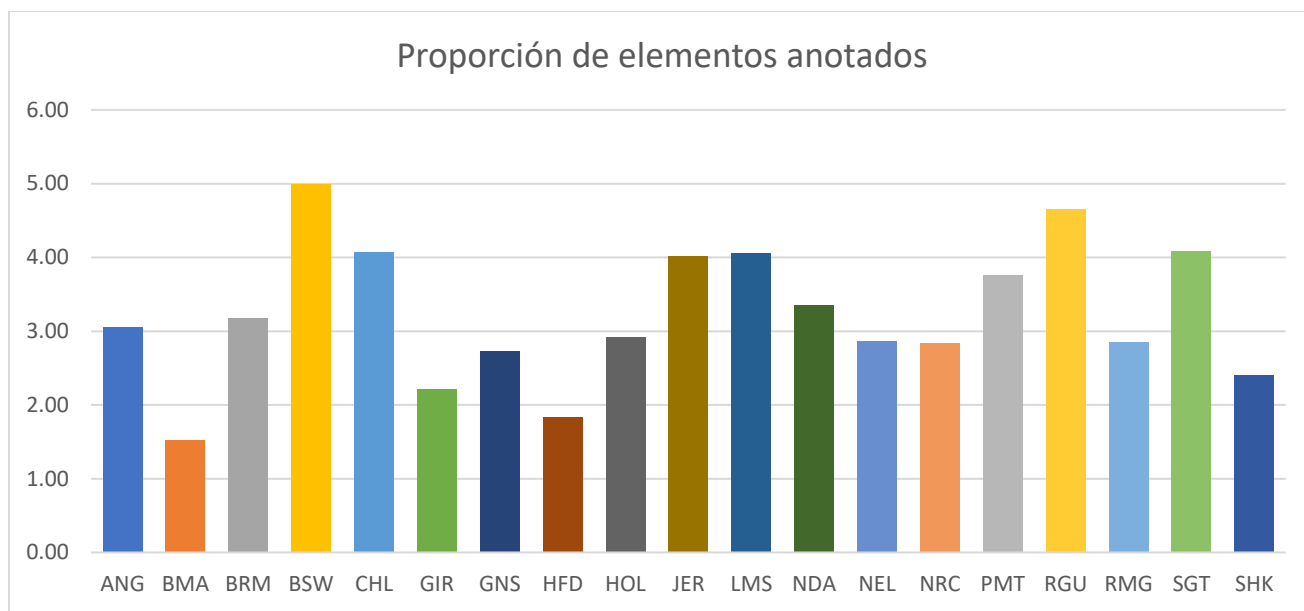


Figura 17 Proporción de elementos genéticos anotados en las regiones significativas

De igual manera, se hace notar que la raza BSW, quien obtuvo el mayor porcentaje de regiones significativas (2.33 %), es quien presenta la mayor cantidad de elementos genéticos anotados.

4.5.2 RASGOS DE PRODUCCIÓN ASOCIADOS

Los genes anotados para las cinco regiones más significativas de cada linaje se presentan en la Tabla 7, junto con los resultados complementarios del análisis en la base de datos *AnimalQTLdb*, utilizada para detectar la presencia de rasgos de producción asociados o QTLs.

En la columna “Tipo de rasgo/QTL” aparece la siguiente nomenclatura para la clasificación de la información.

AR: Asociación a Reproducción

AS: Asociación a Salud

AP: Asociación a Producción

ACC: Asociación a Cuerpo y Carne

ACE: Asociación a Características Externas

APL: Asociación a Producción de Leche

QTL-R: QTL Reproducción

QTL-L: QTL Leche

QTL-P: QTL Producción

QTL-CC: QTL Cuerpo y Carne

QTL-CE: QTL Características Externas

Tabla 7 Genes y rasgos asociados

Raza	Región (Kb)	iHS promedio	Genes	Num. rasgos/QTL	Tipo de rasgo/ QTL
ANG	16:48500-49000	2.757			
	4:41000-41500	2.655	GNAI1	1	AP
	8:97000-97500	2.615	FSD1L, FKTN, TMEM38B, TAL2		
	4:42000-42500	2.597			
	4:41500-42000	2.572		1	AP
BMA	17:36500-37000	3.789	TRPC3		
	17:37000-37500	3.759	SNORA72		
	17:34000-34500	3.537			
	17:34500-35000	3.362	SPRY1, SPATA5		
	17:35000-35500	3.285	SPATA5, SNORA11, NUDT6, U4, BBS12, IL21		
BRM	14:60000-60500	3.265	OXR1		
	14:57000-57500	3.198	ENY2, NUDCD1, TRHR		
	14:57500-58000	3.138	TRHR, TMEM74		
	14:32000-32500	3.072	DNAJC5B, TRIM55, CRH	1	AR
	14:60500-61000	3.051	ZFPM2		
BSW	23:25500-26000	3.138	BoLA DR-ALPHA, BTNL2, BLA-DQB		
	19:44000-44500	2.962	DHX8, ETV4, MEOX1, SOST, DUSP3, C17orf105, MPP3, CD300LG, MPP2, SNORD112, PPY, PYY, PYY2, NAGS, TMEM101		
	5:35000-35500	2.940	ANO6, DBX2		
	13:45000-45500	2.924	PITRM1		
	13:42000-42500	2.923	U6, THBD, CD93, NXT1, GZF1, NAPB, CSTL1, CST11, MGC133636		
CHL	5:59500-60000	3.306	OR10A7		
	24:19000-19500	3.014		2	AP
	14:25500-26000	2.970	IMPAD1	4	AP, QTL-P

	8:59500-60000	2.849	FAM205C, PHF24, U6, DNAJB5, VCP, FANCG, PIGO, STOML2, FAM214B, UNC13B		
	24:22500-23000	2.787	DTNA, U6		
GIR	17:32500-33000	3.077	FAT4		
	17:39500-40000	2.910	ULBP3		
	17:38000-38500	2.844			
	1:7000-7500	2.693			
	17:33000-33500	2.692	ANKRD50		
GNS	14:8000-8500	2.993	ZFAT		
	14:11000-11500	2.848	ADCY8, ASAP1		
	14:10500-11000	2.831	ADCY8		
	14:4500-5000	2.800	TRAPPC9, KCNK9		
	11:67000-67500	2.677	ARHGAP25, BMP10, GKN2, GKN1, ANTXR1		
HFD	21:1000-1500	2.792	U6		
	12:74500-75000	2.777			
	21:1500-2000	2.698	U6, OR5D13		
	9:36500-37000	2.626	HS3ST5, HDAC2	1	QTL-CC
	9:60000-60500	2.573		2	AP
HOL	15:52000-52500	3.056	STIM1, RHOG, PGAP2, NUP98, CHRNA10, bta-mir-2317, RNF121, IL18BP, NUMA1, LRRC51	4	APL
	20:29500-30000	2.978	HCN1	58	AP, APL, QTL-L, ACC
	20:28500-29000	2.895	PARP8, SNORD28, EMB	14	APL, ACC
	20:27000-27500	2.829		7	AP, QTL-R
	1:2500-3000	2.794	MIS18A, HUNK	21	AP, AS, QTL-R, AR, ACE, APL
JER	4:94000-94500	3.654	AHCYL2, STRIP2, NRF1, bta-mir-182, bta-mir-96, bta-mir-183, UBE2H		
	10:24000-24500	3.609			
	16:6500-7000	3.465	KCNT2	3	APL
	12:71500-72000	3.452			
	18:8500-9000	3.435	PLCG2, SNORA70, SDR42E1, HSD17B2, U6, MPHOSPH6	1	QTL-CE
LMS	3:87000-87500	2.998	FGGY, U6		
	2:10000-10500	2.712			

	3:108500-109000	2.666	UTP11, FHL3, SF3A3, INPP5B, MTF1, C1orf122, YRDC, MANEAL, EPHA10, CDCA8, C1orf109, RSPO1, GNL2, DNALI1, SNIP1, MEAF6, ZC3H12A, POU3F1, U6		
	20:14000-14500	2.666	ADAMTS6, U4, CWC27	4	AP
	22:38500-39000	2.650	CADPS	1	QTL-CE
NDA	6:81500-82000	3.712	TECRL		
	2:38000-38500	3.505	U12, UPP2, U6		
	2:136500-137000	3.282	EPHA2, U6, FAM131C, HSPB7, C2H1orf64		
	23:8000-8500	3.065	GRM4, HMGA1, NUDT3, SNORA31, PACSIN1		
	2:36000-36500	3.017	RBMS1, ITGB6, PLA2R1		
NEL	9:4500-5000	2.656			
	3:120500-121000	2.595	KIF1A, AGXT, C2orf54, CROCC2, SNED1, MTERF4, PASK, PPP1R7, HDLBP, SEPT2		
	3:76500-77000	2.550			
	10:69000-69500	2.514	TMEM260, Y_RNA, OTX2		
	13:45000-45500	2.460	PITRM1		
NRC	14:48500-49000	3.281	EXT1, MED30		
	2:93000-93500	3.232	U6		
	14:47500-48000	3.077	SAMD12		
	14:48000-48500	3.033	SAMD12, bta-mir-2489, EXT1		
	14:50000-50500	3.008			
PMT	13:17000-17500	3.167	PRKCQ, PFKFB3, RBM17		
	16:25000-25500	2.955	U6atac, HLX		
	13:18500-19000	2.949	CREM, CUL2, PARD3		
	13:15500-16000	2.941	GATA3		
	4:69500-70000	2.932	SKAP2, 7SK		
RGU	5:44000-44500	2.950	BEST3, LRRC10, CCT2, FRS2, SNORA25, 5S_rRNA, YEATS4, U5, U6	1	AP

	5:45000-45500	2.915	CPM, MDM2, SLC35E3, NUP107, SNORA38	1	AP
	5:55500-56000	2.863	ATP23, CTDSP2, bta-mir-26a-2, AVIL		
	5:43000-43500	2.859	PTPRR, PTPRB, KCNMB4, CNOT2, SNORA18	1	AP
	5:54500-55000	2.800	LRIG3		
RMG	5:75000-75500	3.017	MYH9, U6, TXN2, FOXRED2, EIF3D, CACNG2		
	6:37500-38000	2.995	HERC3, NAP1L5, PYURF, HERC5, HERC6, PPM1K, ABCG2, U6		
	5:75500-76000	2.987	PVALB, NCF4, CSF2RB, TEX33, TST, MPST, KCTD17, Tmprss6, IL2RB, C1QTNF6		
	5:72000-72500	2.959	LARGE1		
	6:39500-40000	2.805			
SGT	13:66000-66500	3.362	EPB41L1, AAR2, DLGAP4, MYL9, TGIF2, TGIF2-C20orf24, SLA2, NDRG3, TLDC2, SAMHD1		
	3:75500-76000	3.060	LRRC7		
	3:76000-76500	2.981			
	13:68500-69000	2.973			
	13:67000-67500	2.971	SRC, BLCAP, NNAT, CTNBL1, VSTM2L		
SHK	12:46000-46500	2.971			
	6:5000-5500	2.836			
	12:55000-55500	2.733			
	4:38500-39000	2.706	CACNA2D1		
	11:62000-62500	2.679	UGP2, VPS54, PELI1		

La tabla anterior refleja varias cuestiones de interés, por ejemplo, en las 95 regiones con las señales de selección más intensas, se presentan varios genes y elementos de producción asociados. De manera numérica se tiene que el 76.8 % de ellas comprenden al menos un gen; el 20 % presenta rasgos de producción asociados o QTLs; y finalmente, el 14.7% presenta ambos elementos.

La región 108,500-109,000 Kb del cromosoma 3 de la raza LMS es la que presentó la mayor concentración de genes detectados con un total de 17, mientras que la región 29500-30000 Kb del cromosoma 20 de la raza HOL presentó el mayor número de asociaciones y QTLs con un total de 58 elementos, abarcando las categorías de: asociación a producción, a producción de leche, a cuerpo y carne y QTL Leche.

Cabe mencionar que durante el análisis en la base de datos *AnimalQTLdb*, se verificó que el rasgo de producción (QTL) detectado estuviera registrado en el mismo linaje bovino en estudio, es decir, que efectivamente ese rasgo hubiese sido reportado para el mismo tipo de ganado. De igual manera, se corroboró que cada elemento detectado presentara un identificador único (QTL_ID). Respetando los criterios anteriores, fue posible identificar rasgos de producción asociados en al menos 8 linajes bovinos de los 19 manejados en este estudio.

Atendiendo la agrupación de razas bovinas según el propósito de crianza, se destaca dos principales categorías, como se aprecia en la Figura #18.

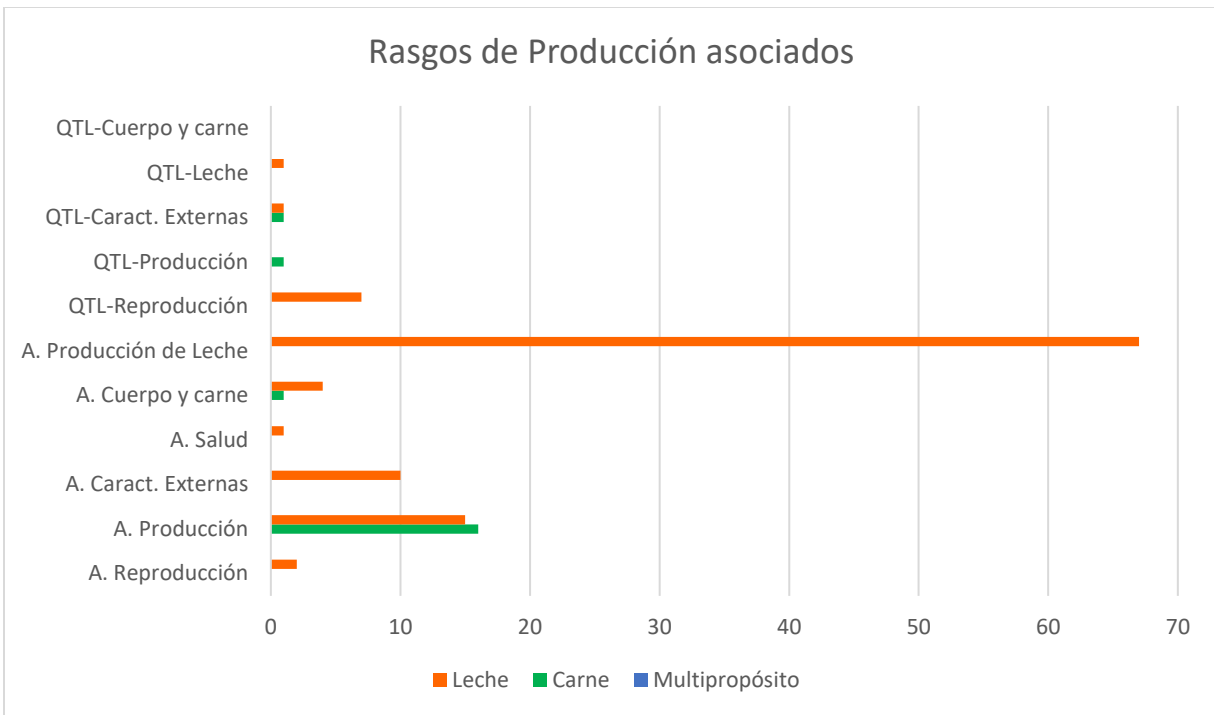


Figura 18 Distribución de rasgos de producción asociados según propósito de crianza

En el grupo “Producción de leche” se obtuvo el conteo mayor de asociaciones con 67 en la categoría “Asociación a Producción de Leche”, seguida de la categoría “Asociación a Producción” con un conteo de 15; lo anterior refleja por completo los rasgos distintivos deseados según el propósito de crianza.

De manera similar, en el grupo “Producción de carne” la categoría predominante fue “Asociación a Producción” con 16 conteos, seguidas por las categorías “Asociación a cuerpo y carne”, “QTL_características externas” y “QTL_Producción” con un valor de 1. Por otra parte, no fue posible detectar ningún elemento de asociación para los linajes que componen el grupo “Multipropósito”.

Finalmente, una breve revisión en la literatura permitió vincular varios genes detectados en este estudio con genes y/o familias de genes reportados previamente. En la Tabla 8 se muestran una referencia cruzada con la información en cuestión.

Tabla 8 Referencia cruzada de genes

Autor	Gen de referencia	Gen detectado	Raza	Función
Zhao, 2015	FAM159A	FAM131C	NDA	
S. Qanbari, 2011	ADAMTS6	ADAMTS6	LMS	
Zhao, 2014, 2015	ANTXR2	ANTXR1	GNS	Codifica una proteína reconocida como receptor de las toxinas de ántrax.
S. Qanbari, 2010	BMP4	BMP10	GNS	Vinculado al desarrollo cardiovascular.
Zhao, 2015	CACNA2D2	CACNA2D1	SHK	Asociado a rasgos de producción y mastitis clínica.
M. Hossein, 2012	CACNB3	CACNG2	RMG	
Zhao, 2015	CD109, CD63	CD93	BSW	Relacionado a rutas del sistema inmune.
Zhao, 2015	DNAJC14	DNAJC5B	BRM	Relacionado al correcto plegamiento de proteínas.
S. Qanbari, 2010	HSD11B1	HSD17B2	JER	Asociado al metabolismo de estrógeno.
Shameek, 2006	IL4	IL21	BMA	Vinculado a respuestas innatas y adaptativas del sistema inmune.
Zhao, 2015	LRRC41	LRRC7	SGT	Vinculado al sistema inmune innato.
Holly Ramey, 2013	OR5D18	OR5D13	HFD	Receptor olfativo.
Gautier, 2011	PPP1R8	PPP1R7	NEL	Asociado a funciones reproductivas (desarrollo embrionario).
M. Hossein, 2012	PPP2CA	PPP1R7	NEL	
Zhao, 2015	PPP1R42	PPP1R7	NEL	Asociado a funciones reproductivas (desarrollo embrionario).
Gautier, 2011	SAMD12	SAMD12	NRC	Involucrado en señalización celular durante respuestas inmune.
Holly Ramey, 2013	SAMD11	SAMD12	NRC	Involucrado en señalización celular durante respuestas inmune.

Holly Ramey, 2013	SAMHD1	SAMHD1	SGT	Involucrado en señalización celular durante respuestas inmune.
Gautier, 2011	SLC7A5	SLC35E3	RGU	Involucrado en pigmentación de piel (por exposición a luz UV).
Shameek, 2006	SLC24A5	SLC35E3	RGU	Involucrado en pigmentación de piel (por exposición a luz UV).
Zhao, 2015	SLC5A9	SLC35E3	RGU	Involucrado en pigmentación de piel (por exposición a luz UV).
Zhao, 2014, 2015	SPATA18	SPATA5	BMA	Asociado a espermatogénesis
S. Qanbari, 2010, 2011	SPATA6	SPATA5	BMA	Asociado a espermatogénesis
Zhao, 2015	TAL1	TAL2	ANG	Asociado a la activación de las células T en procesos de leucemia.
M. Hossein, 2012	TMEM8B	TMEM38B	ANG	
Zhao, 2015	ZC3H10	ZC3H12A	LMS	Relacionado al proceso de apoptosis celular.

De la tabla anterior se destaca que los genes: SAM12, SAMHD1 y ADAMTS6, han sido reportados previamente en la literatura como sujetos a presión de selección en estudios similares al presente. El resto de los genes fueron seleccionados por afinidad a la familia del gen reportada.

La anotación de los genes SAM12 y SAMHD1 indica que ambos participan en procesos de señalización celular durante respuestas inmunes. Analizando las funciones biológicas de los otros genes señalados, se destacan dos principales aspectos de participación: sistema inmune y reproducción. Por ejemplo, se tiene una triple referencia para la familia del gen detectado PPP1R7, asociado a funciones reproductivas y desarrollo embrionario; la familia del gen SLC35E3 involucrada a pigmentación de piel por exposición a luz UV; y una doble referencia para la familia del gen SPATA5 asociada al proceso de espermatogénesis.

CAPÍTULO 5 CONCLUSIONES Y TRABAJO FUTURO

5.1 CONCLUSIONES

El análisis de la Homocigosidad Haplotípica Extendida, a través de su estadístico de prueba iHS, reveló un total de 968 señales de selección positiva de intensidad variable en regiones específicas del genoma bovino en las 19 razas, siendo la principal aportación de este trabajo de investigación, esta colección de regiones genómicas obtenidas tras una inspección del genoma completo con SNPs de alta densidad, las cuales se sumarán al panel público del Consorcio HapMap Bovino.

En las múltiples etapas del procesamiento de la información se generaron diversos algoritmos en lenguaje Perl, donde la estructura de datos tipo *hash* fue la más utilizada para almacenar y manipular los datos. Por su parte, el ambiente de programación en R facilitó el cálculo de los valores de iHS, generación de gráficas, así como la búsqueda y extracción de información en bases de datos.

Tras los diversos filtros de calidad aplicados a los datos, la resolución del estudio no se vio comprometida, pues se lograron retener 546, 961 SNPs, lo que equivale al 73.62 % de los SNPs autosomales del BovineHD Genotyping BeadChip.

El umbral de exigencia utilizado para determinar un estadístico de iHS como significativo, permitió identificar señales de selección positiva en todas las razas, concluyéndose que el valor promedio de regiones bajo presión de selección en el genoma bovino es del 1 %.

El análisis de las señales de selección utilizando la clasificación de los linajes según el propósito de crianza, reveló una alta coincidencia entre los cromosomas dentro de los grupos, lo que sugiere que existe una afinidad entre el propósito de explotación y los cromosomas con regiones bajo presión de selección, siendo los cromosomas 3 y 6 comunes en los tres grupos.

El estudio también revela que las regiones señaladas bajo presión de selección más significativas son ricas en elementos genéticos como genes, pseudogenes y transcritos de RNA, teniendo una distribución uniforme en las 19 razas, con un valor promedio de 3 elementos genéticos por región.

Fue posible detectar rasgos de producción asociados y QTLs en al menos 8 linajes bovinos, destacándose la presencia de 79 elementos en el cromosoma 20 de la raza Holstein, distribuidos en 3 regiones muy cercanas entre sí; dichos elementos están vinculados a la producción de leche, propósito de crianza de dicha raza.

Nuevamente, si se agrupan las razas según el propósito de crianza en lechero, cárnico y multipropósito, se observa una relación directa entre éstos y las categorías de los rasgos de producción asociados.

El hecho de que en una región significativa en particular no se detectara la presencia de un algún gen o rasgo de producción asociado, no indica que en dicha región no exista una presión de selección justificable, puesto que las bases de datos y anotaciones de los genomas son compendios de información en constante crecimiento y continuos procesos de curación, abriendo la posibilidad de que en un futuro algún elemento sea anotado.

Finalmente, este estudio detectó varios genes y familias de genes que han sido reportados en estudios previos, donde el común denominador son los genes involucrados con la producción, reproducción, metabolismo y sistema inmune.

5.2 TRABAJO FUTURO

Como trabajo futuro se plantea extender el análisis a la totalidad de las regiones con señales de selección detectadas, inspeccionando la presencia de otros elementos genéticos de interés además de genes y QTLs, por ejemplo, variaciones estructurales tipo CNV.

También se propone profundizar el análisis de los genes detectados de tal manera que se puedan asociar a rutas metabólicas específicas, lo cual ayudaría a entender la presencia y el efecto de la presión de selección de acuerdo con el propósito de crianza.

REFERENCIAS

- [1] E. W. Hartl, Daniel L. Jones, *Genetics: analyses of genes and genomes.*, 6th ed. Jones and Barlett Publishers, 2005.
- [2] B. Terry, *Genomas*, 3rd ed. Panamericana, Buenos Aires: Médica, 2008.
- [3] “Estructura del ADN | webioQenfermeria.” [Online]. Available: <https://webioqenfermeria.wordpress.com/estructura-del-adn/>. [Accessed: 01-Aug-2017].
- [4] “RNA - science biology.” [Online]. Available: <http://taescience.weebly.com/rna.html>. [Accessed: 01-Aug-2017].
- [5] R. Horton, L. Moran, G. Scrimgeour, M. Perry, and D. Rawn, *Principios de Bioquímica*, 4th ed. Pearson Education, Inc., 2008.
- [6] C. Michael, Madigan, Jhon, Martinko, Paul, Dunlap, David, *Biología de los Microorganismos*, 12th ed. Pearson Education, Inc., 2009.
- [7] “What is a gene? - Genetics Home Reference.” [Online]. Available: <https://ghr.nlm.nih.gov/primer/basics/gene>. [Accessed: 03-Aug-2017].
- [8] “Polimorfismo de nucleótido simple - EcuRed.” [Online]. Available: https://www.ecured.cu/Polimorfismo_de_nucleótido_simple. [Accessed: 03-Aug-2017].
- [9] Q. T. Loci and L. Qtl, “Marcadores genéticos,” pp. 1–42.
- [10] A. T. Favier, “Conceptos básicos y metodologías innovadoras de mejoramiento genético de maíz.”
- [11] A. Tiessen Favier, *Fundamentos de mejoramiento genético vegetal fundamentos de mejoramiento genético vegetal*. 2012.
- [12] C. S. Haley and S. A. Knott, “A simple regression method for mapping quantitative trait loci in line crosses using flanking markers.” *Heredity (Edinb.)*, vol. 69, no. 4, pp. 315–24, 1992.
- [13] T. F. C. Mackay, “Q&A: Genetic analysis of quantitative traits,” *J. Biol.*, vol. 8, p. 23, 2009.
- [14] Q. Zhang *et al.*, “Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree.” *Genetics*, vol. 149, no. 4, pp. 1959–1973, 1998.
- [15] “Bovino - EcuRed.” [Online]. Available: <https://www.ecured.cu/Bovino>. [Accessed: 04-Aug-2017].
- [16] Y. T. Utsunomiya, A. M. P. Pérez O’Brien, T. S. Sonstegard, J. Sölkner, and J. F. Garcia, “Genomic data as the ‘hitchhiker’s guide’ to cattle adaptation: Tracking the milestones of past selection in the bovine genome,” *Front. Genet.*, vol. 5, no. FEB, pp. 1–13, 2015.
- [17] T. J. Ortega and P. L. García, “El genoma bovino, métodos y resultados de su análisis,” *Rev. MVZ Cordoba*, vol. 16, no. 1, pp. 2410–2424, 2011.

- [18] C. G. Elsik *et al.*, “Bovine Genome Database: new tools for gleaning function from the *Bos taurus* genome.,” *Nucleic Acids Res.*, vol. 44, no. D1, pp. D834–9, Jan. 2016.
- [19] R. Villa-Angulo, L. K. Matukumalli, C. A. Gill, J. Choi, C. P. Van Tassell, and J. J. Grefenstette, “High-resolution haplotype block structure in the cattle genome.,” *BMC Genet.*, vol. 10, p. 19, 2009.
- [20] The Bovine HapMap Consortium, “Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds,” *Science (80-.)*, vol. 324, no. April, pp. 528–532, 2009.
- [21] S. R. Grossman *et al.*, “A composite of multiple signals distinguishes causal variants in regions of positive selection.,” *Science (80-.)*, vol. 327, no. 5967, pp. 883–6, 2010.
- [22] T. K. Oleksyk, M. W. Smith, and S. J. O. Brien, “Genome-wide scans for footprints of natural selection,” pp. 185–205, 2010.
- [23] M. Gautier and M. Naves, “Footprints of selection in the ancestral admixture of a New World Creole cattle breed,” *Mol. Ecol.*, 2011.
- [24] “What is the difference between genotyping and sequencing? – 23andMe Customer Care.” [Online]. Available: <https://customercare.23andme.com/hc/en-us/articles/202904600-What-is-the-difference-between-genotyping-and-sequencing->. [Accessed: 06-Aug-2017].
- [25] “Bos taurus - Ensembl genome browser 89.” [Online]. Available: http://www.ensembl.org/Bos_taurus/Info/Annotation. [Accessed: 15-Jul-2017].
- [26] W. Jhon, *Coalescent Theory: An introduction.*, 12th ed. United States of America: Ben Roberts, 2009.
- [27] a Castillo, “La Selección Natural a Nivel Molecular,” *Ecol. Mol.*, pp. 11–48, 2007.
- [28] S. Qanbari and H. Simianer, “Mapping signatures of positive selection in the genome of livestock,” *Livest. Sci.*, vol. 166, no. 1, pp. 133–143, 2014.
- [29] T. Strachan and A. Read, *Human Molecular Genetics*, 4th ed. Garland Science, 2011.
- [30] T. K. Oleksyk, M. W. Smith, and S. J. O’Brien, “Genome-wide scans for footprints of natural selection.,” *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 365, no. 1537, pp. 185–205, 2010.
- [31] D. Rocha, C. Billerey, F. Samson, D. Boichard, and M. Boussaha, “Identification of the putative ancestral allele of bovine single-nucleotide polymorphisms,” *J. Anim. Breed. Genet.*, vol. 131, no. 6, pp. 483–486, 2014.
- [32] P. C. Sabeti *et al.*, “Detecting recent positive selection in the human genome from haplotype structure,” *Nature*, vol. 419, no. October, pp. 832–837, 2002.
- [33] K. Tang, K. R. Thornton, and M. Stoneking, “A new approach for using genome scans to detect recent positive selection in the human genome,” *PLoS Biol.*, vol. 5, no. 7, pp. 1587–1602, 2007.
- [34] “Base de Datos.” [Online]. Available:

- <http://docencia.fca.unam.mx/CAACS/index.php/areas/base-de-datos>. [Accessed: 02-Aug-2017].
- [35] “Bioconductor - biomaRt.” [Online]. Available: <https://bioconductor.org/packages/3.5/bioc/html/biomaRt.html>. [Accessed: 01-Aug-2017].
- [36] “Ensembl genome browser 89.” [Online]. Available: <http://www.ensembl.org/index.html>. [Accessed: 01-Aug-2017].
- [37] “Perl Introduction.” [Online]. Available: https://www.tutorialspoint.com/perl/perl_introduction.htm. [Accessed: 02-Aug-2017].
- [38] M. Dowle, “Introduction to the data . table package in R,” vol. 2014, pp. 1–9, 2014.
- [39] M. Gautier and R. Vitalis, “Rehh An R package to detect footprints of selection in genome-wide SNP data from haplotype structure,” *Bioinformatics*, vol. 28, no. 8, pp. 1176–1177, 2012.
- [40] S. R. Browning and B. L. Browning, “Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering,” *Am. J. Hum. Genet.*, vol. 81, no. 5, pp. 1084–1097, 2007.
- [41] T. Note, “‘ TOP / BOT ’ Strand and ‘ A / B ’ Allele.”
- [42] M. Gautier and R. Vitalis, “Tutorial for the R package rehh,” pp. 1–33, 2012.
- [43] S. Qanbari *et al.*, “Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle.,” *BMC Genomics*, vol. 12, no. 1, p. 318, 2011.
- [44] F. Zhao, S. McParland, F. Kearney, L. Du, and D. P. Berry, “Detection of selection signatures in dairy and beef cattle using high-density genomic information,” *Genet. Sel. Evol.*, vol. 47, no. 1, p. 49, 2015.

ANEXOS

ANEXO 1 COMPENDIO DE MARCADORES INICIALES

En la siguiente tabla se muestra el compendio de marcadores utilizados como datos de partida para el análisis. Se detalla el número de marcadores en cada uno de los cromosomas para cada linaje.

Cromosoma	Razas									
	ANG	BMA	BRM	BSW	CHL	GIR	GNS	HFD	HOL	JER
1	46,491	46,493	46,492	46,490	46,493	46,489	46,492	46,493	46,492	46,490
2	40,056	40,056	40,056	40,054	40,056	40,043	40,053	40,056	40,056	40,054
3	35,574	35,575	35,574	35,573	35,575	35,570	35,574	35,575	35,573	35,574
4	34,975	34,976	34,973	34,974	34,976	34,965	34,973	34,976	34,975	34,975
5	34,838	34,840	34,839	34,830	34,840	34,825	34,829	34,840	34,839	34,838
6	35,515	35,515	35,514	35,510	35,514	35,513	35,513	35,516	35,515	35,511
7	33,166	33,166	33,163	33,166	33,166	33,163	33,164	33,166	33,166	33,159
8	33,525	33,526	33,523	33,524	33,525	33,519	33,525	33,526	33,525	33,525
9	31,056	31,056	31,055	31,055	31,056	31,053	31,055	31,056	31,056	31,056
10	30,448	30,449	30,449	30,448	30,449	30,442	30,448	30,449	30,448	30,449
11	32,011	32,014	32,011	32,012	32,014	32,007	32,011	32,014	32,012	32,012
12	26,121	26,122	26,119	26,118	26,121	26,116	26,120	26,123	26,122	26,122
13	23,592	23,592	23,592	23,592	23,591	23,588	23,590	23,592	23,592	23,591
14	24,777	24,778	24,776	24,774	24,778	24,768	24,776	24,778	24,776	24,775
15	24,751	24,751	24,751	24,750	24,751	24,744	24,750	24,751	24,750	24,749
16	24,177	24,177	24,176	24,177	24,177	24,175	24,175	24,177	24,177	24,176
17	22,263	22,265	22,264	22,263	22,264	22,262	22,263	22,264	22,263	22,263
18	19,383	19,383	19,383	19,383	19,384	19,379	19,383	19,385	19,385	19,385
19	18,907	18,908	18,906	18,908	18,908	18,906	18,907	18,908	18,907	18,907
20	21,489	21,489	21,489	21,488	21,489	21,487	21,489	21,489	21,489	21,489
21	21,172	21,173	21,173	21,173	21,173	21,169	21,173	21,173	21,173	21,172
22	18,032	18,033	18,030	18,032	18,033	18,029	18,031	18,033	18,032	18,032
23	15,213	15,214	15,213	15,213	15,213	15,212	15,214	15,214	15,214	15,213
24	18,619	18,619	18,619	18,618	18,618	18,618	18,618	18,619	18,619	18,617
25	12,931	12,931	12,931	12,931	12,931	12,930	12,931	12,931	12,931	12,931
26	15,240	15,241	15,238	15,240	15,240	15,238	15,239	15,241	15,240	15,239
27	13,151	13,151	13,151	13,151	13,151	13,149	13,151	13,151	13,151	13,149
28	13,036	13,037	13,037	13,037	13,037	13,036	13,035	13,037	13,037	13,036
29	14,710	14,710	14,708	14,708	14,710	14,708	14,709	14,710	14,709	14,707
Total SNPs	735,219	735,240	735,205	735,192	735,233	735,103	735,191	735,243	735,224	735,196

Cromosoma	Razas								
	LMS	NDA	NEL	NRC	PMT	RGU	RMG	SGT	SHK
1	46,493	46,484	46,487	46,493	46,492	46,491	46,492	46,492	46,487
2	40,056	40,049	40,047	40,056	40,055	40,055	40,056	40,056	40,055
3	35,575	35,572	35,573	35,574	35,575	35,572	35,574	35,575	35,574
4	34,975	34,971	34,968	34,975	34,976	34,975	34,974	34,976	34,975
5	34,840	34,838	34,828	34,837	34,839	34,838	34,832	34,840	34,840
6	35,515	35,510	35,512	35,514	35,514	35,512	35,514	35,514	35,514
7	33,166	33,163	33,164	33,165	33,166	33,166	33,166	33,166	33,166
8	33,526	33,521	33,518	33,525	33,525	33,523	33,525	33,525	33,524
9	31,056	31,051	31,052	31,056	31,055	31,056	31,056	31,056	31,055
10	30,449	30,446	30,442	30,448	30,449	30,447	30,448	30,449	30,449
11	32,014	32,012	32,008	32,011	32,014	32,011	32,014	32,013	32,012
12	26,121	26,120	26,115	26,120	26,123	26,119	26,120	26,121	26,120
13	23,592	23,588	23,588	23,591	23,591	23,592	23,592	23,592	23,591
14	24,777	24,771	24,772	24,774	24,777	24,775	24,776	24,777	24,775
15	24,750	24,744	24,745	24,750	24,750	24,747	24,751	24,751	24,749
16	24,177	24,176	24,175	24,177	24,177	24,177	24,176	24,177	24,177
17	22,265	22,262	22,264	22,264	22,265	22,264	22,264	22,265	22,265
18	19,385	19,383	19,381	19,385	19,385	19,385	19,384	19,385	19,383
19	18,907	18,905	18,906	18,906	18,907	18,907	18,908	18,908	18,907
20	21,489	21,486	21,483	21,488	21,489	21,489	21,489	21,489	21,488
21	21,173	21,167	21,171	21,173	21,173	21,172	21,173	21,172	21,172
22	18,033	18,029	18,026	18,032	18,033	18,032	18,033	18,033	18,030
23	15,214	15,212	15,213	15,214	15,214	15,213	15,213	15,213	15,212
24	18,619	18,617	18,616	18,618	18,618	18,618	18,619	18,619	18,619
25	12,931	12,931	12,930	12,931	12,931	12,931	12,931	12,931	12,931
26	15,240	15,237	15,241	15,239	15,241	15,240	15,240	15,241	15,239
27	13,151	13,149	13,149	13,151	13,151	13,151	13,150	13,151	13,149
28	13,036	13,034	13,035	13,036	13,037	13,036	13,037	13,037	13,035
29	14,710	14,709	14,707	14,709	14,710	14,710	14,710	14,710	14,709
Total SNPs	735,235	735,137	735,116	735,212	735,232	735,204	735,217	735,234	735,202

ANEXO 2 COMPENDIO DE MARCADORES RETENIDOS

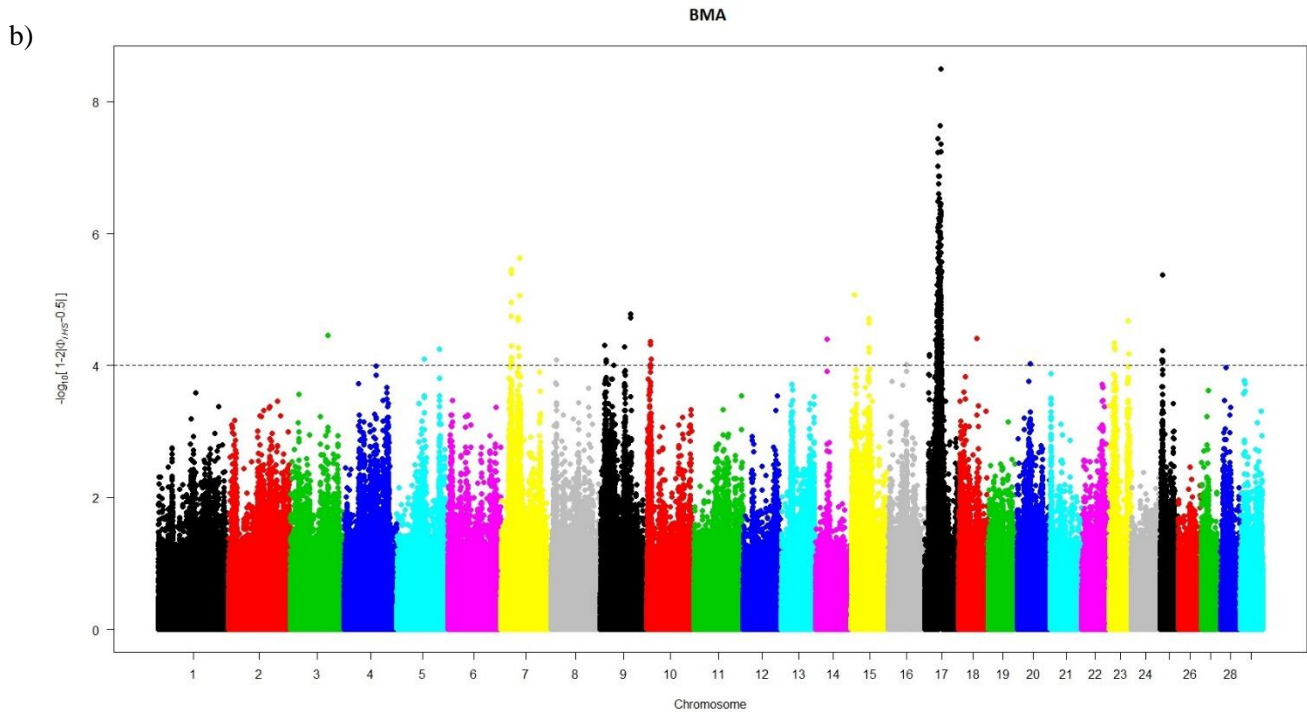
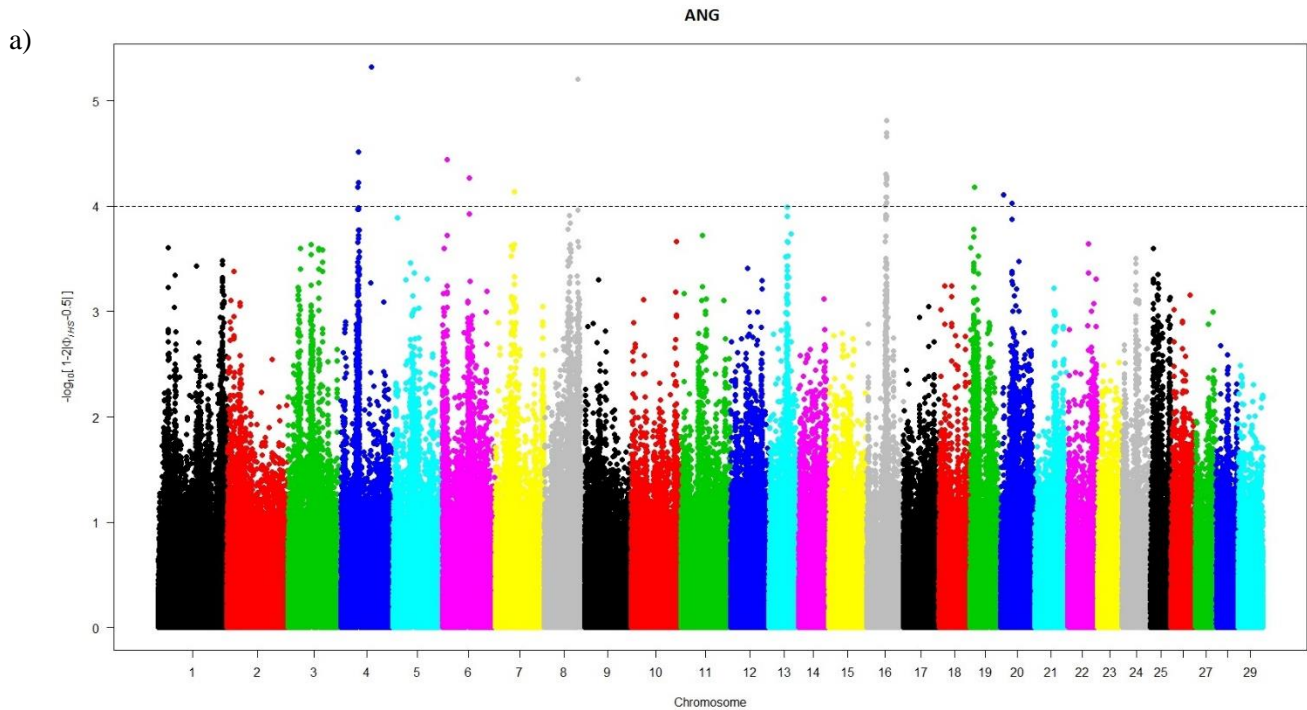
Marcadores retenidos después de aplicar los filtros de calidad requeridos por el estudio: eliminación de SNPs monomórficos, marcadores con frecuencias de alelo menor inferior a 0.05 y aquellos marcadores cuya información sobre alelos ancestrales y derivados era desconocía.

Raza	ANG	BMA	BRM	BSW	CHL	GIR	GNS	HFD	HOL	JER
1	34,661	42,333	34,435	32,939	36,952	27,477	32,093	37,208	35,927	30,312
2	29,991	36,723	28,587	27,824	31,454	22,802	28,047	31,029	29,686	25,612
3	26,552	32,169	24,854	25,614	28,649	20,701	24,865	28,951	27,683	22,801
4	26,574	31,783	23,945	25,235	27,704	19,836	24,344	28,043	27,438	24,084
5	24,284	30,341	22,911	22,916	27,234	18,810	24,699	27,681	26,868	23,870
6	27,010	32,374	25,482	22,436	29,158	23,233	25,627	28,424	28,298	23,752
7	24,592	29,930	23,103	23,943	26,779	19,598	23,254	26,490	24,881	22,209
8	20,452	29,937	23,709	20,006	23,162	20,696	19,368	21,355	21,219	18,931
9	23,391	28,475	23,390	22,698	25,351	18,950	21,867	24,608	24,469	21,808
10	24,880	28,102	20,401	23,017	25,396	16,420	21,589	25,737	24,405	21,848
11	24,845	29,108	21,573	22,261	26,441	17,645	21,643	27,089	26,113	23,596
12	19,477	23,503	17,626	18,808	20,533	14,677	18,450	21,235	20,688	18,800
13	15,399	20,779	17,283	14,435	16,035	13,850	13,794	16,677	15,685	13,508
14	15,405	21,983	19,226	14,950	17,122	15,333	15,365	17,035	16,238	14,689
15	18,876	22,213	18,072	17,936	20,521	13,684	17,719	20,384	19,985	17,359
16	18,539	21,932	17,281	16,349	19,598	15,623	17,017	19,596	18,574	17,140
17	17,893	20,418	16,715	16,473	18,331	12,674	15,892	18,628	18,070	16,003
18	15,497	17,688	14,653	14,824	16,118	11,537	15,043	16,150	16,383	14,082
19	15,335	17,195	12,262	13,891	16,400	9,922	13,572	16,042	15,520	14,092
20	17,320	20,007	14,280	16,138	18,536	12,136	15,788	18,578	17,446	14,349
21	16,397	19,002	15,644	15,339	17,374	11,857	15,031	17,052	16,618	14,657
22	14,960	16,484	11,672	13,522	15,605	9,957	13,569	15,834	15,078	13,615
23	12,699	13,696	11,472	11,912	12,883	9,219	11,816	13,076	12,528	10,993
24	14,283	17,353	14,060	12,947	15,117	10,857	13,175	14,834	14,589	12,937
25	10,590	11,450	8,883	9,825	11,046	7,289	9,901	11,124	10,971	9,677
26	12,042	13,628	11,502	11,558	12,538	8,793	11,335	13,168	12,421	11,195
27	10,628	11,922	9,706	10,020	11,216	7,619	10,015	11,535	11,026	9,783
28	10,859	11,677	8,882	10,343	11,404	7,674	10,186	11,399	11,110	10,027
29	11,927	13,155	9,788	10,450	12,385	7,835	10,613	12,536	12,063	10,169
SNPs raza	555,358	665,360	521,397	518,609	591,042	426,704	515,677	591,498	571,980	501,898

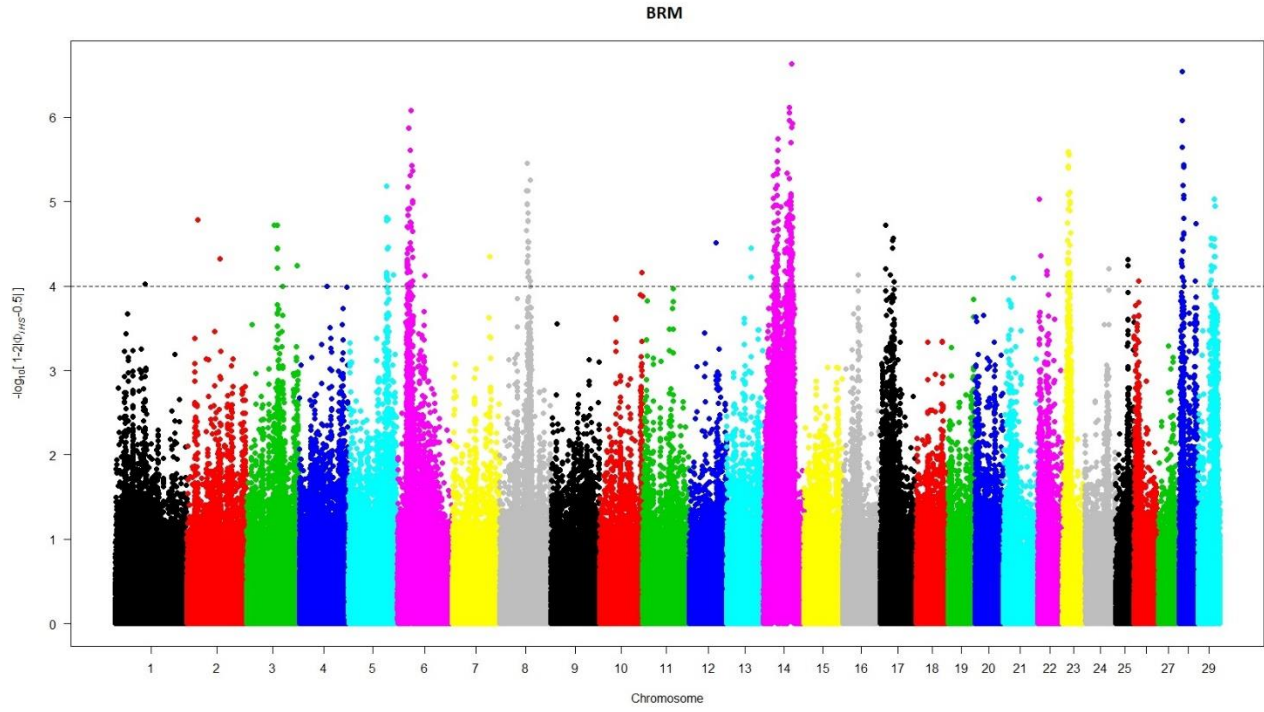
Raza	LMS	NDA	NEL	NRC	PMT	RGU	RMG	SGT	SHK
1	36,220	27,322	28,128	36,432	37,093	32,392	35,779	41,372	37,627
2	30,131	22,934	23,701	30,779	31,086	28,436	31,767	35,992	32,107
3	28,550	21,391	20,751	27,969	28,441	24,791	27,378	30,771	28,169
4	28,059	20,797	19,118	27,159	28,236	24,704	27,144	30,889	27,468
5	27,420	19,959	19,149	25,980	27,725	24,181	26,224	29,638	27,281
6	28,301	21,078	23,173	28,336	28,939	25,747	27,179	30,996	28,973
7	25,167	19,557	19,493	25,663	26,068	23,772	26,119	29,303	26,572
8	21,592	16,135	21,793	21,523	22,395	19,609	23,242	30,068	27,410
9	24,538	18,327	20,509	24,098	24,828	22,049	23,700	27,507	25,057
10	24,591	18,360	17,561	23,938	25,372	23,067	23,658	27,509	23,961
11	26,131	19,269	18,246	25,828	26,595	23,043	25,120	27,837	24,592
12	20,657	15,770	15,018	20,936	21,068	18,652	20,593	23,004	20,449
13	16,493	12,195	13,817	15,788	16,781	14,056	16,685	19,261	18,882
14	16,387	12,262	16,225	15,724	17,159	14,474	16,569	20,769	20,224
15	20,031	14,445	14,540	19,625	20,507	17,794	19,181	21,852	19,361
16	19,109	14,543	14,323	19,479	19,686	17,395	17,419	22,009	19,486
17	18,349	14,064	13,584	18,268	18,650	16,808	16,928	19,764	17,596
18	16,576	12,161	11,534	15,828	16,644	14,546	15,929	17,107	15,505
19	15,702	10,679	10,351	15,474	15,985	14,269	15,616	16,502	14,560
20	18,002	13,276	12,451	17,931	18,353	16,333	16,411	19,512	16,764
21	16,471	12,587	12,286	16,240	17,147	15,186	16,632	18,753	16,999
22	15,357	11,184	10,326	15,038	15,496	14,280	14,402	15,939	14,101
23	12,885	9,989	9,488	12,350	12,956	11,315	11,886	13,312	12,243
24	14,753	11,007	10,964	14,738	15,077	13,726	14,649	16,831	14,918
25	10,850	7,641	7,101	10,995	11,214	9,745	10,655	11,497	10,071
26	12,487	9,348	9,298	12,472	12,878	11,045	12,525	13,585	12,106
27	10,983	8,225	7,926	10,896	11,207	10,075	10,667	11,707	10,479
28	10,898	8,560	7,706	10,951	11,217	10,099	10,654	11,564	10,547
29	11,955	8,714	8,352	11,992	12,336	11,082	11,947	12,528	11,624
SNPs raza	578,645	431,779	436,912	572,430	591,139	522,671	566,658	647,378	585,132

ANEXO 3 DISTRIBUCIÓN DE VALORES P_{iHS}

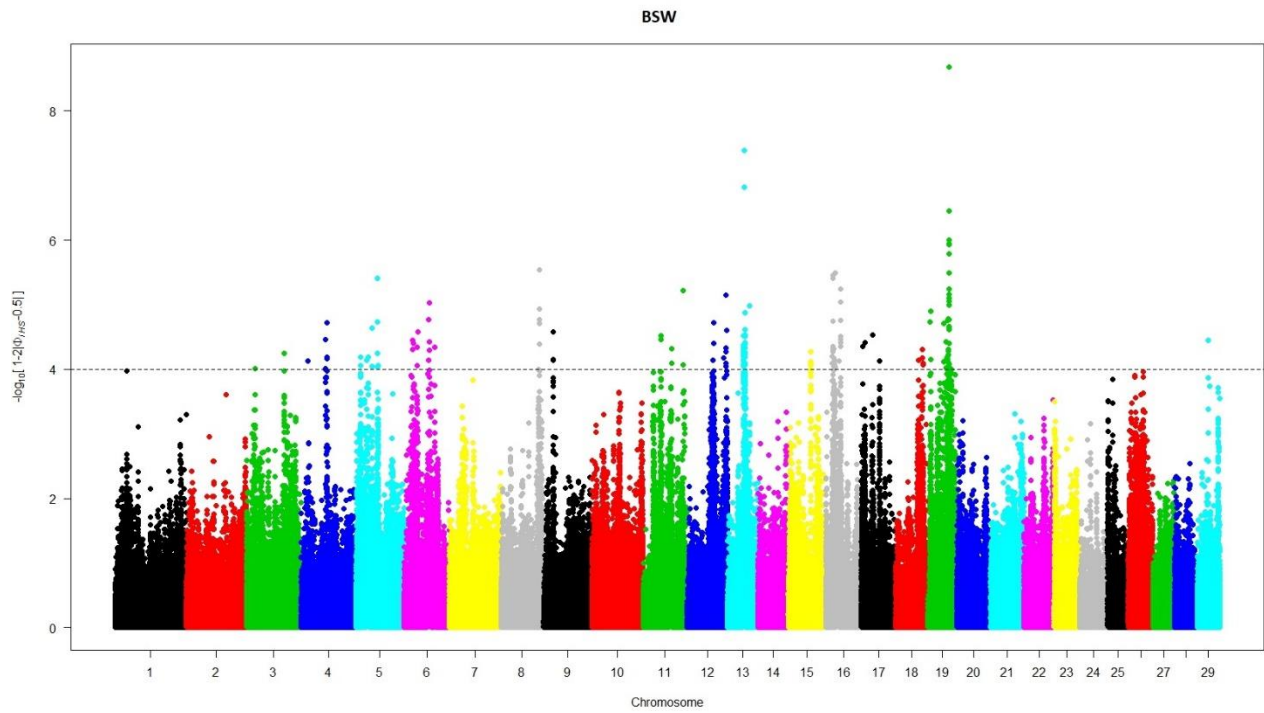
En la siguiente figura se muestran las gráficas de Manhattan con la distribución de los valores de p_{iHS} a lo largo del genoma.



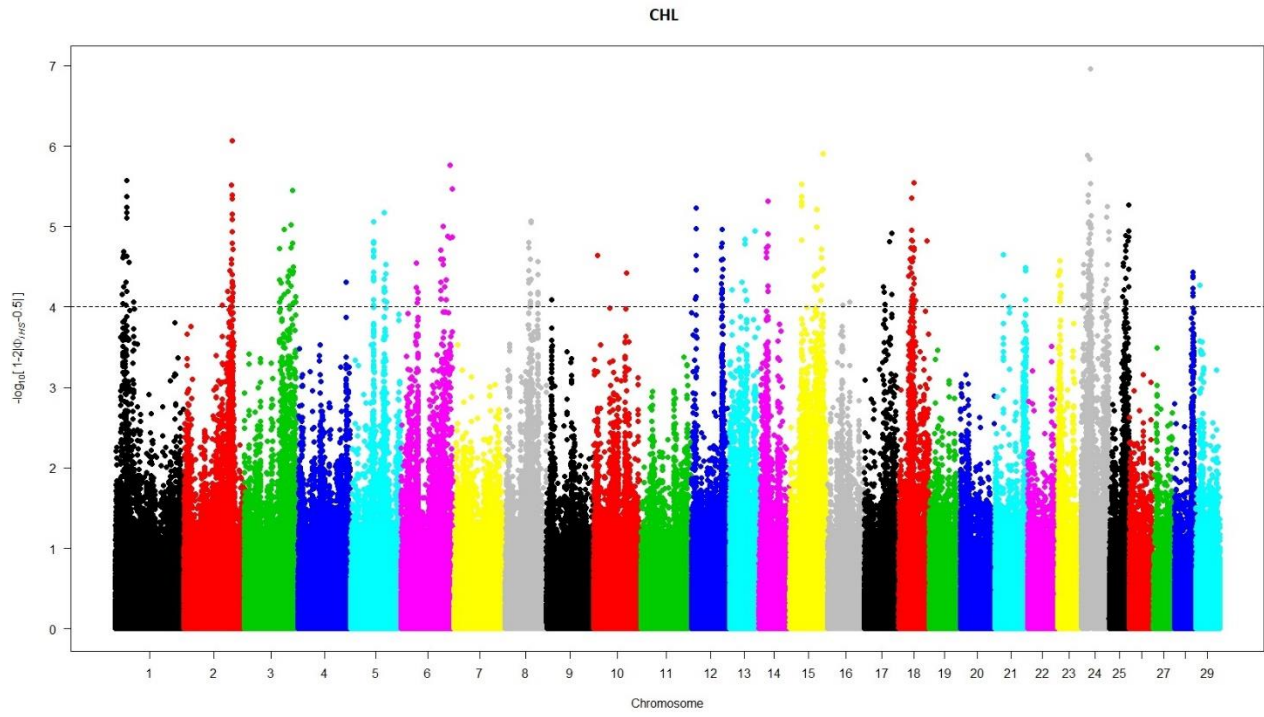
c)



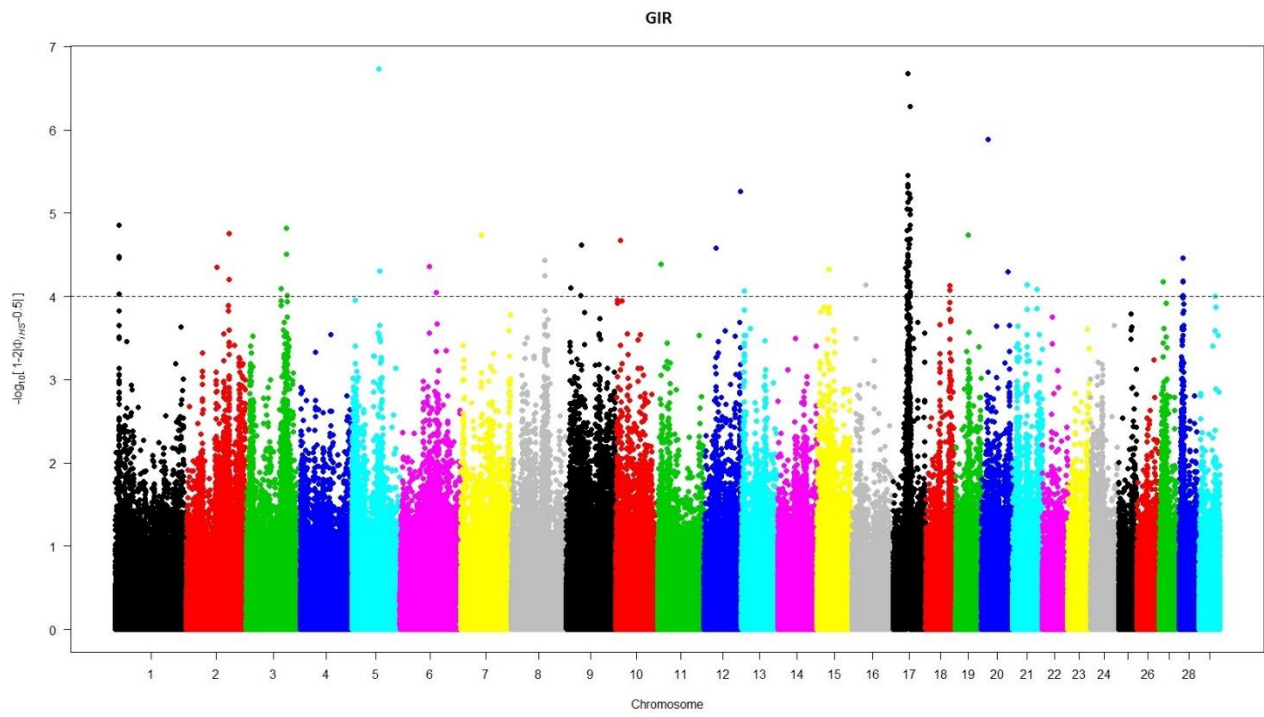
d)



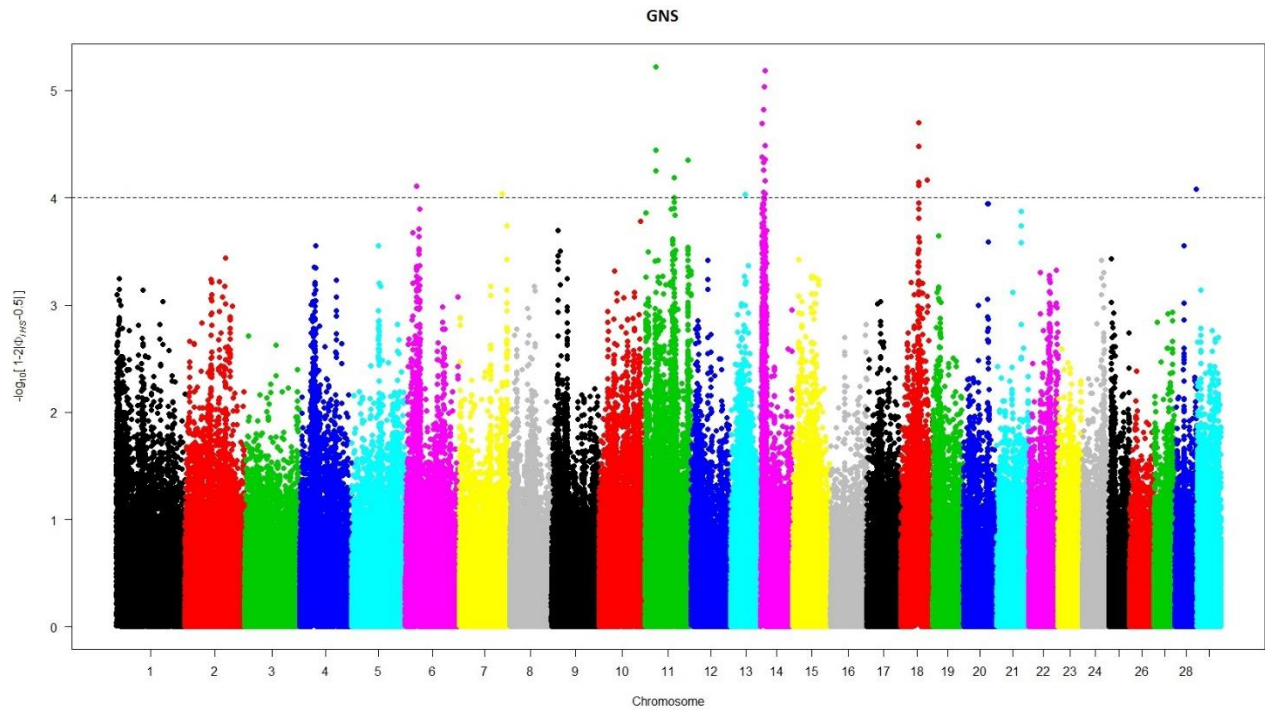
e)



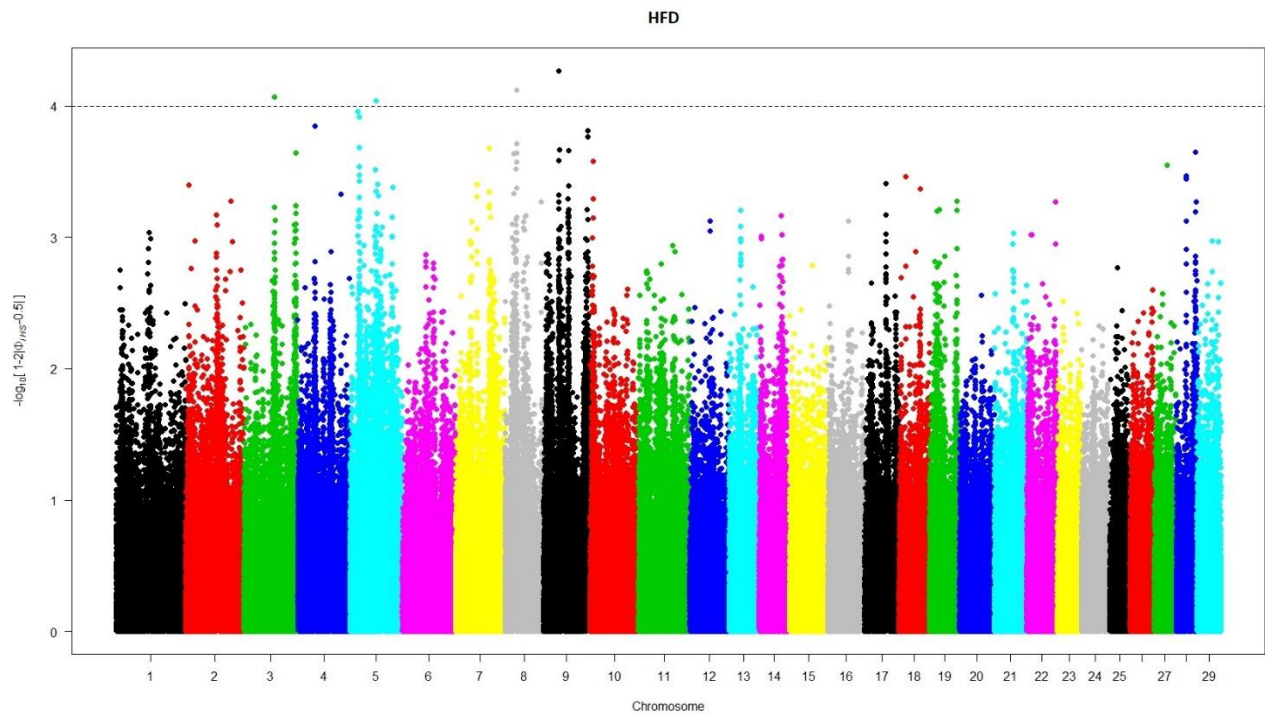
f)



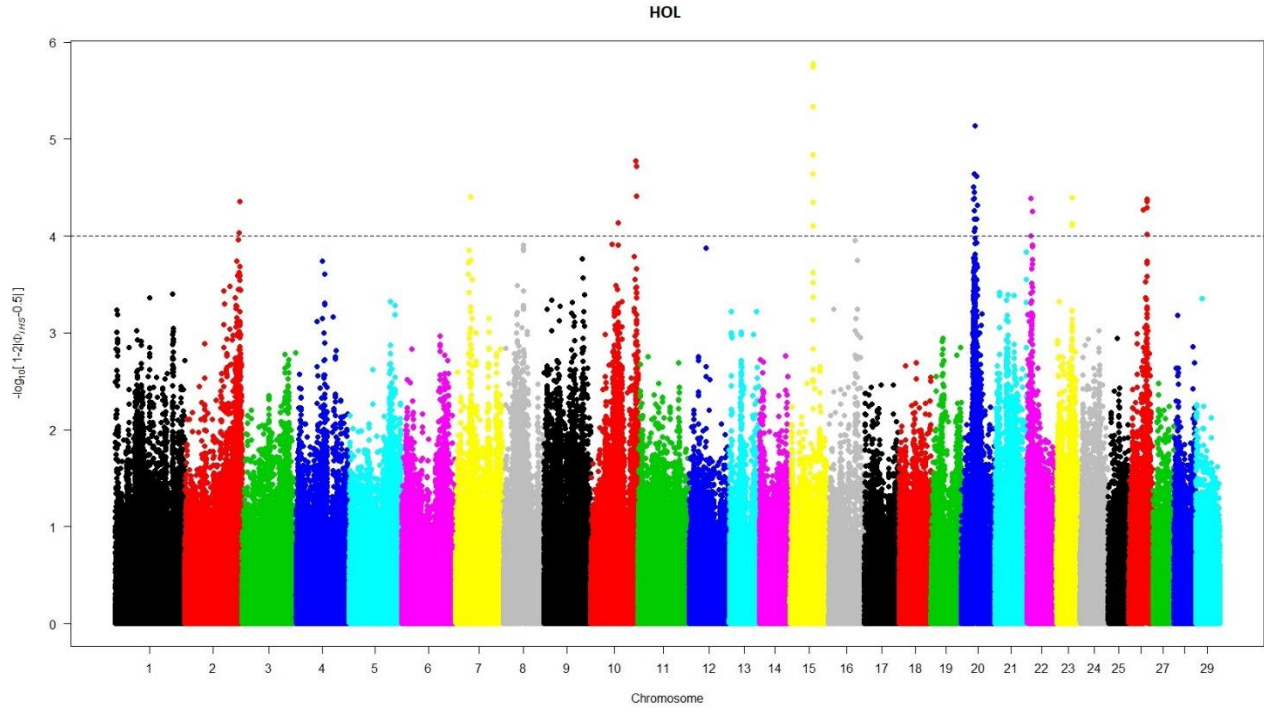
g)



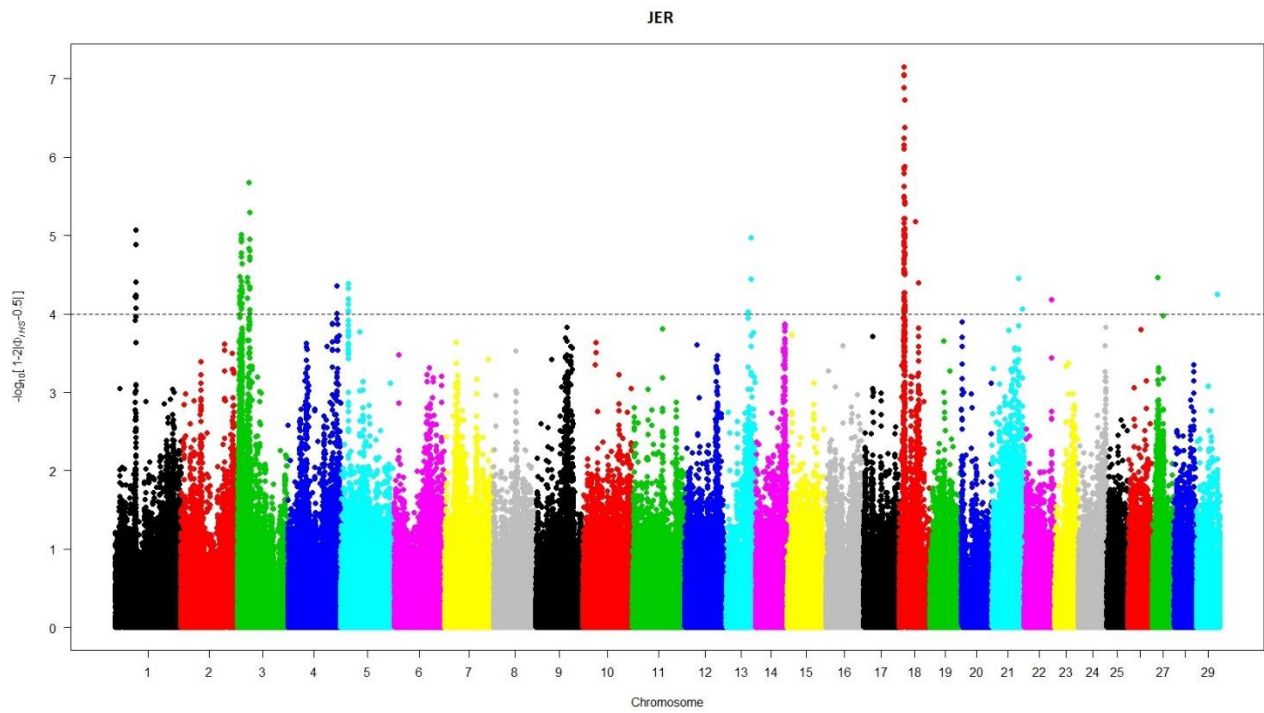
h)



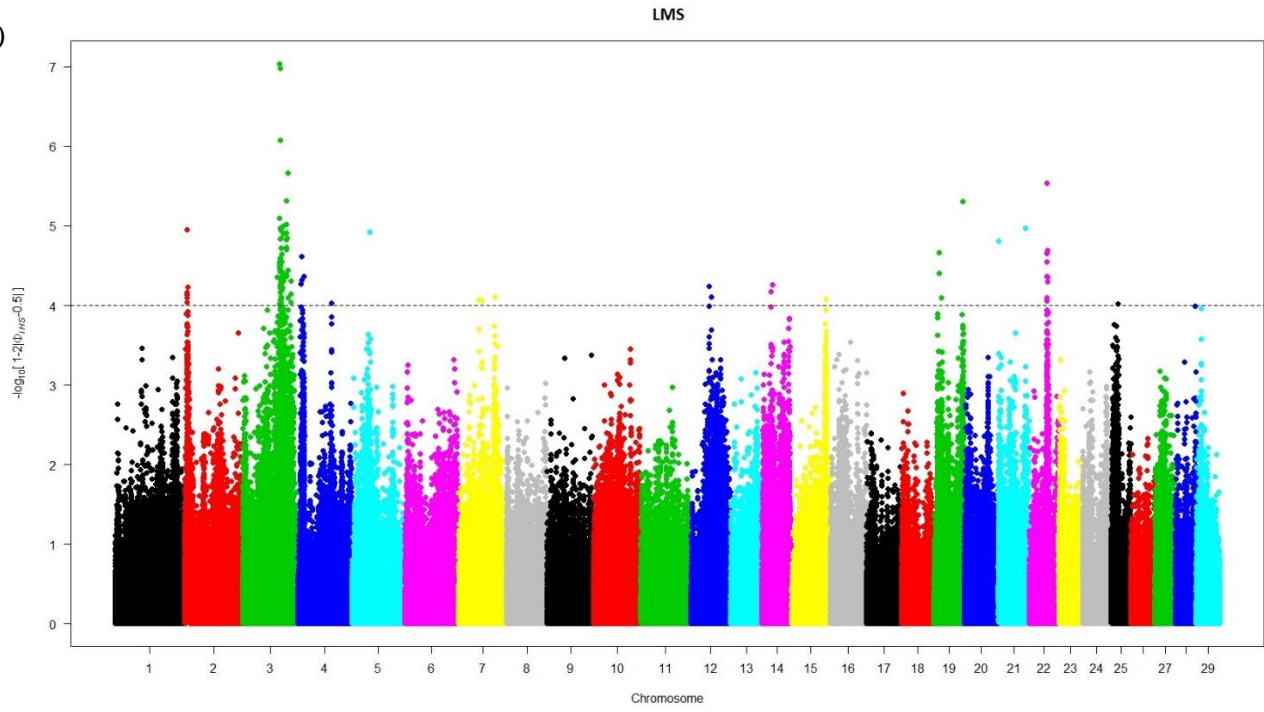
i)



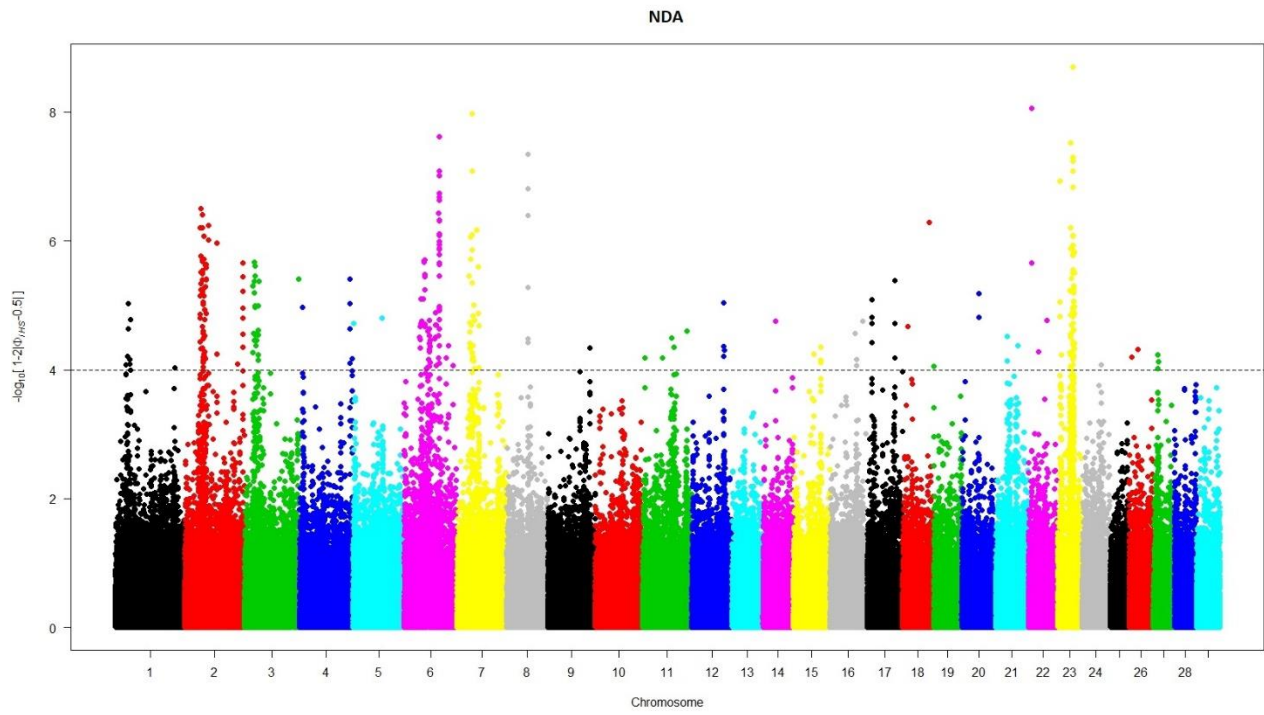
j)



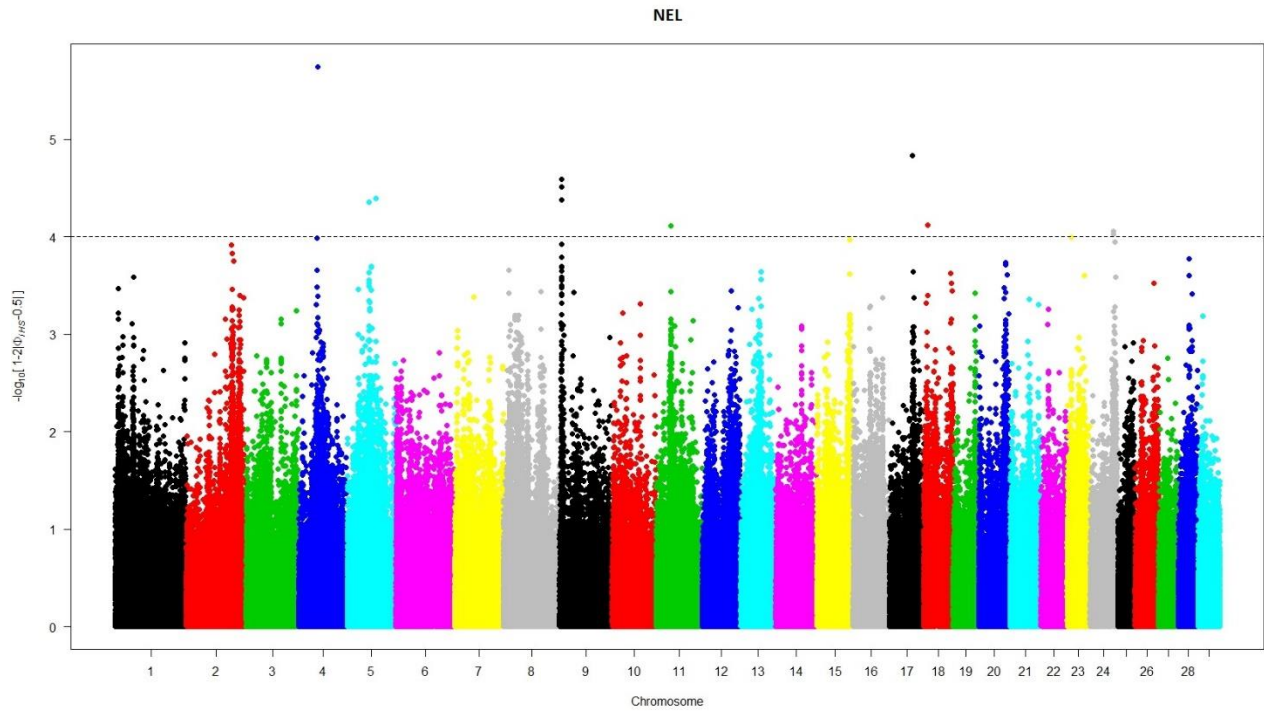
k)



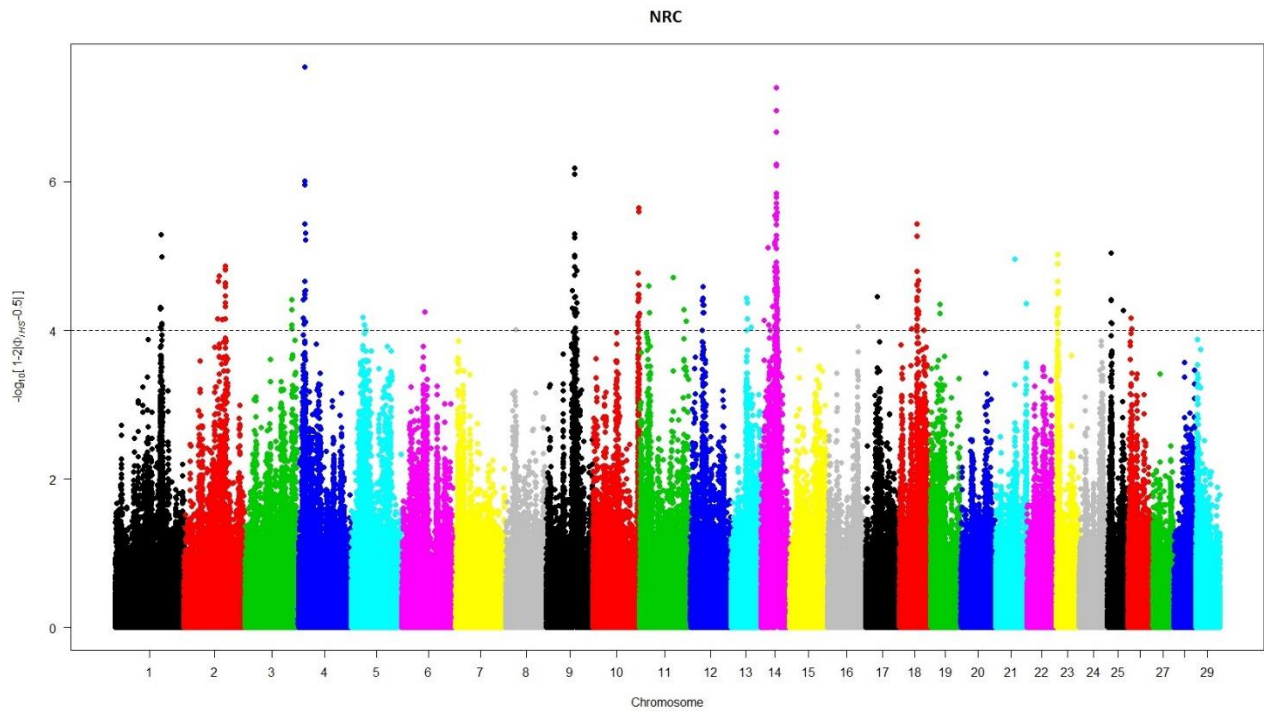
l)



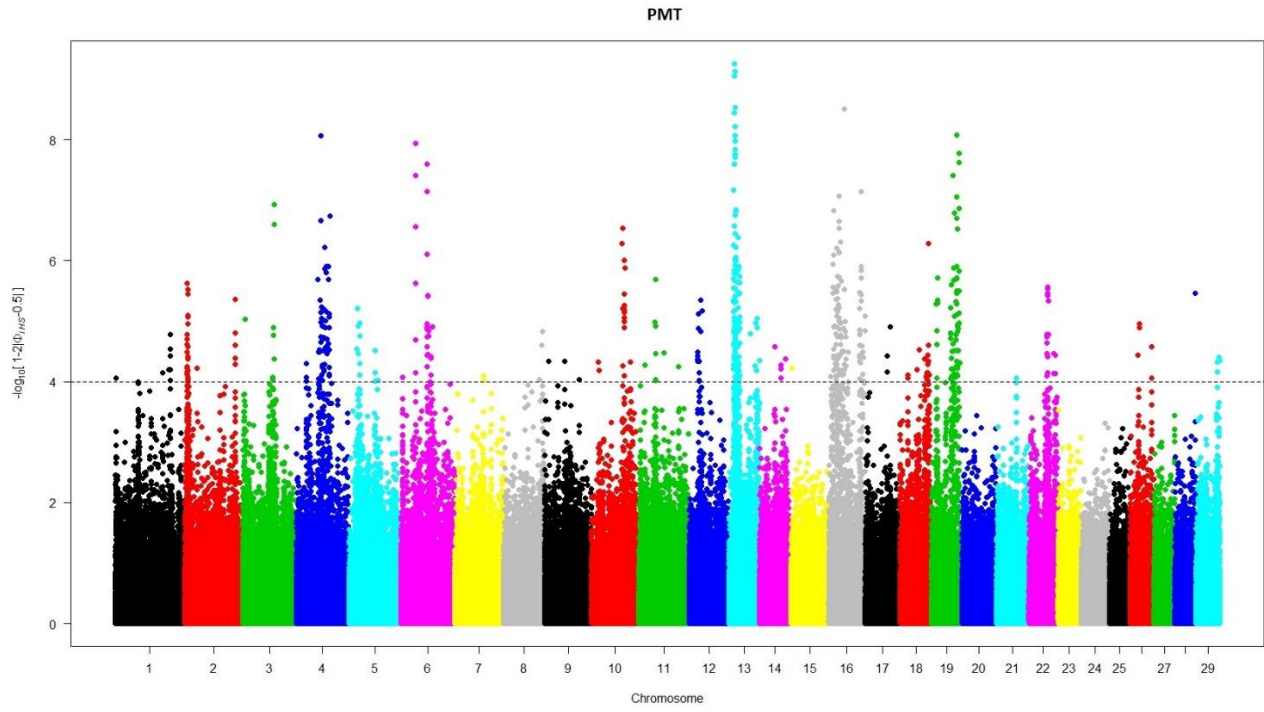
m)



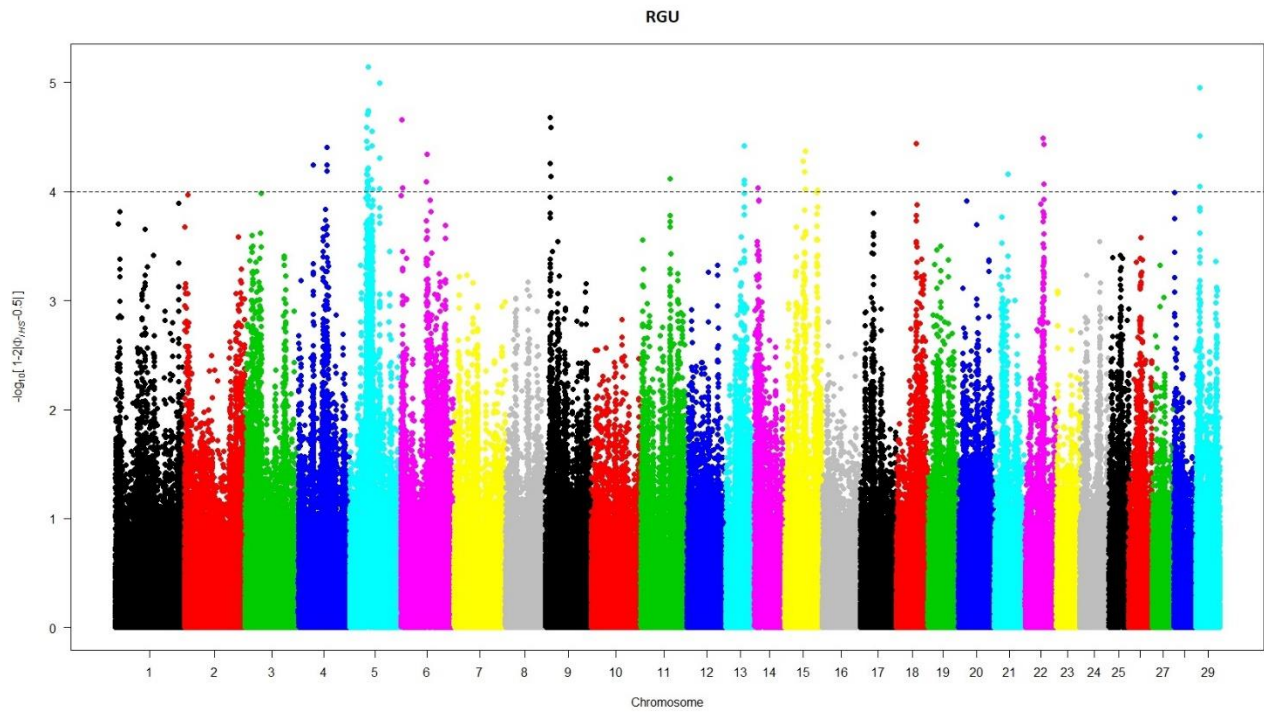
n)



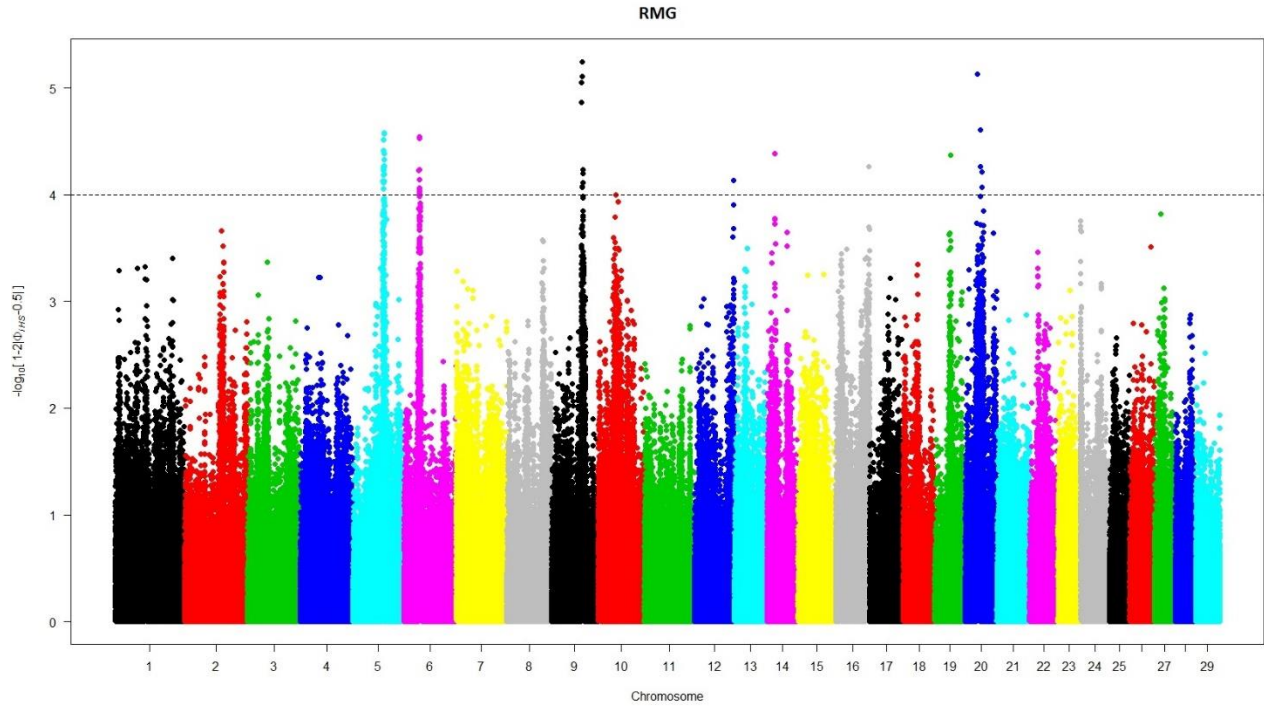
n)



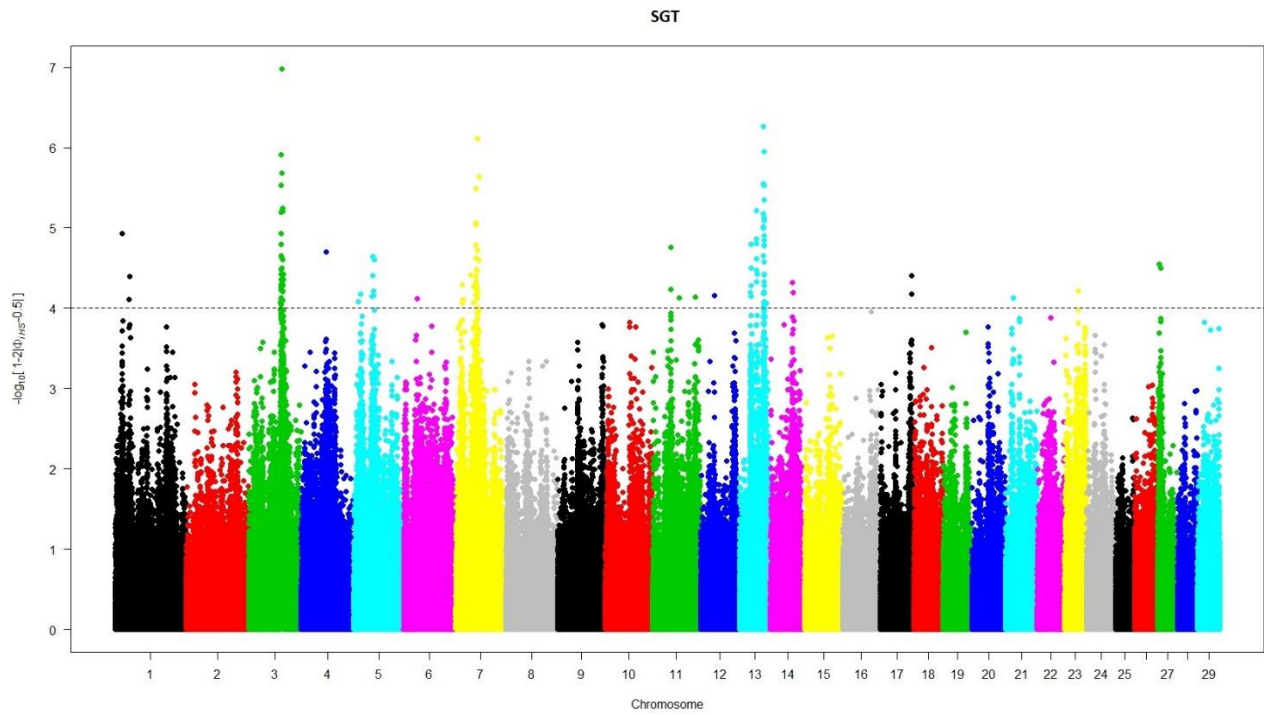
o)



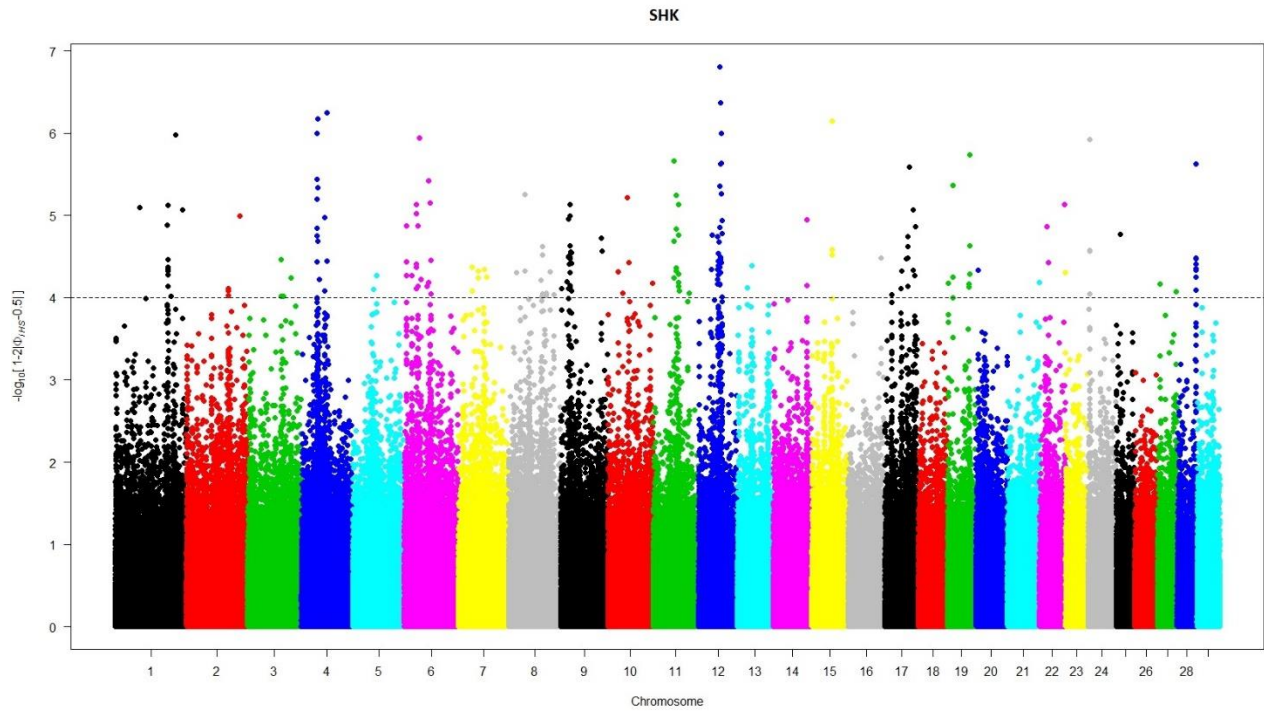
p)



q)



r)



ANEXO 4 NÚMERO DE REGIONES SIGNIFICATIVAS

En la siguiente tabla se muestran el número de regiones significativas detectadas en cada cromosoma de los 19 linajes bovinos.

	ANG	BMA	BRM	BSW	CHL	GIR	GNS	HFD	HOL	JER
Cr	# R	# R	# R	# R	# R	# R	# R	# R	# R	# R
1	1				3	2			3	3
2					6	2	1	1	2	2
3	3		1	2		7		1	1	11
4	7		1	6			1			5
5			4	14	2	1			1	1
6	1		13	9	3		1			4
7		8	1				1			2
8	7		6	3	9	1		3	2	2
9		3		1		1		2		6
10	1	2		3	1		1		3	2
11				9	1		7			1
12				10	5	1	2	2		3
13	4	1	2	7		1			1	2
14			40	1	3		16			2
15		1		1	2	1			1	1
16	4		1	13			2		2	2
17		29	4	2	1	15	1	1		
18				4	7	1	3			6
19	1			14						
20						1			15	
21		1			5	1	1	2	2	5
22			1						1	
23		3	6	2	1				1	
24	1			2	14					
25	1	1		2	1	1				1
26			1	4					1	3
27										
28	1		5		1	1				1
29	1		7	2						

	LMS	NDA	NEL	NRC	PMT	RGU	RMG	SGT	SHK
Cr	# R	# R	# R	# R	# R	# R	# R	# R	# R
1		3		9	2	1		1	2
2	7	13	3	6	5		1		1
3	13	2	2	2	3	1	1	16	
4	2	2		8	4	2			5
5	1	2	1	8		20	6	5	
6		10	1	3	4	7	11		3
7		8				1		5	
8		1			2				1
9		2	1	9		2	8	2	2
10		1	1	3	1		5		
11		4	1			1		1	2
12		2					1		6
13			1	1	15			17	
14	2	1		19				1	
15	5			2		5			2
16		1	1		9		3		
17		4	1	1					3
18	1	1		3	1	1			
19	1				2		2		
20	1	1	1				7		1
21		3		1		1			
22	2	1			3	2			
23		10		5			1	2	
24						1	1		2
25				3					
26									
27				1					
28			1						
29						1			

ANEXO 5 REGIONES SIGNIFICATIVAS

En la siguiente tabla se describen a detalle las regiones significativas detectadas en el estudio. El primer número corresponde al cromosoma y los siguientes valores identifican la posición exacta de la región de 500 Kb.

Raza	Regiones
ANG	ANG,16:48500-49000,4:41000-41500,8:97000-97500,4:42000-42500,4:41500-42000,16:48000-48500,28:12000-12500,29:0-500,16:49000-49500,4:42500-43000,3:54000-54500,8:96500-97000,8:64000-64500,4:43000-43500,13:46500-47000,16:47000-47500,24:32500-33000,4:40500-41000,25:39500-40000,13:52000-52500,4:44000-44500,1:149000-149500,19:8500-9000,8:91000-91500,13:48500-49000,8:92000-92500,6:58500-59000,3:49500-50000,8:66500-67000,3:23000-23500,13:51500-52000,8:93500-94000,10:23500-24000
BMA	17:36500-37000,17:37000-37500,17:34000-34500,17:34500-35000,17:35000-35500,17:37500-38000,17:38000-38500,17:32500-33000,17:33000-33500,17:28000-28500,17:28500-29000,17:39000-39500,17:38500-39000,17:39500-40000,17:27500-28000,17:40000-40500,17:30500-31000,17:31000-31500,17:30000-30500,17:29500-30000,17:31500-32000,17:32000-32500,17:35500-36000,17:36000-36500,17:27000-27500,17:33500-34000,9:12000-12500,7:26000-26500,17:23000-23500,7:27000-27500,25:6000-6500,7:26500-27000,23:12000-12500,21:71500-72000,15:13500-14000,13:25500-26000,23:14500-15000,9:59500-60000,23:13500-14000,7:25000-25500,7:41500-42000,7:24500-25000,10:7000-7500,9:11500-12000,7:46500-47000,17:26500-27000,10:5500-6000,7:38500-39000,17:25000-25500
BRM	14:60000-60500,14:57000-57500,14:57500-58000,14:32000-32500,14:60500-61000,14:29000-29500,14:30000-30500,14:56500-57000,14:31000-31500,14:27000-27500,14:26000-26500,14:58500-59000,6:59500-60000,14:27500-28000,14:28000-28500,14:24000-24500,8:68000-68500,14:25500-26000,23:19000-19500,14:61000-61500,6:27000-27500,23:19500-20000,28:7000-7500,14:59500-60000,14:56000-56500,14:26500-27000,14:62000-62500,14:29500-30000,14:30500-31000,14:31500-32000,14:52000-52500,5:99500-100000,14:62500-63000,13:38000-38500,8:62000-62500,6:25000-25500,14:61500-62000,14:51500-52000,28:8500-9000,14:23000-23500,14:38500-39000,5:90500-91000,14:25000-25500,14:55000-55500,17:26500-27000,3:67500-68000,6:21500-22000,29:35000-35500,29:23000-23500,8:70000-70500,6:20000-20500,28:42500-43000,23:18000-18500,23:14500-15000,5:99000-99500,8:65500-66000,29:36000-36500,29:37500-38000,8:66000-66500,17:23000-23500,13:500-1000,14:63000-63500,23:17500-18000,6:23500-24000,14:22500-23000,7:52000-52500,29:29000-29500,6:33000-33500,6:32500-33000,17:24500-25000,14:54000-54500,14:46500-47000,14:28500-29000,28:8000-8500,14:54500-55000,14:38000-38500,14:41000-41500,4:83500-84000,8:64000-64500,23:15000-15500,29:26500-27000,6:25500-26000,16:35500-36000,5:101500-102000,6:26500-27000,6:17500-18000,28:5000-5500,26:4500-5000,17:4500-5000,29:44000-44500,6:22500-23000,22:4500-5000,6:19000-19500

BSW	<p>23:25500-26000,19:44000-44500,5:35000-35500,13:45000-45500,13:42000-42500,19:48500-49000,4:64500-65000,4:62500-63000,19:45500-46000,15:51500-52000,12:55500-56000,24:26000-26500,5:34000-34500,12:59000-59500,19:48000-48500,4:65000-65500,8:100000-100500,8:100500-101000,11:100500-101000,8:99000-99500,6:39000-39500,19:46500-47000,6:20500-21000,16:17000-17500,12:55000-55500,13:60500-61000,5:48500-49000,5:50000-50500,6:26500-27000,5:64000-64500,6:21000-21500,19:45000-45500,3:84500-85000,6:32500-33000,11:64500-65000,19:49500-50000,6:32000-32500,17:41000-41500,19:40500-41000,12:75500-76000,16:28000-28500,16:20000-20500,6:18000-18500,18:53500-54000,11:43500-44000,19:54000-54500,16:34000-34500,29:46500-47000,9:22000-22500,11:33000-33500,18:52500-53000,11:67500-68000,16:19000-19500,4:65500-66000,11:42000-42500,24:25500-26000,13:43500-44000,5:33500-34000,5:25000-25500,16:25000-25500,10:55000-55500,19:8000-8500,26:17500-18000,25:16000-16500,11:95000-95500,26:34000-34500,26:18500-19000,5:65500-66000,19:54500-55000,5:38500-39000,12:57000-57500,10:57500-58000,5:64500-65000,19:47500-48000,16:16500-17000,12:58500-59000,16:15500-16000,4:64000-64500,17:7500-8000,16:16000-16500,23:3000-3500,4:0-500,16:26000-26500,14:1000-1500,16:34500-35000,12:59500-60000,11:32000-32500,19:43500-44000,25:2000-2500,12:56500-57000,16:35500-36000,5:38000-38500,29:23500-24000,5:63500-64000,13:47000-47500,12:89000000-89500000,13:46000000-46500000,16:35000000-35500000,12:56000000-56500000,11:23000000-23500000,13:46500000-47000000,6:22500000-23000000,5:11000000-11500000,18:50500000-51000000,26:17000000-17500000,3:11000000-110500000,18:43000000-43500000,10:75000000-75500000,19:43000000-43500000,6:24000000-24500000,5:39500000-40000000</p>
CHL	<p>5:59500-60000,24:19000-19500,14:25500-26000,8:59500-60000,24:22500-23000,24:15000-15500,6:40000-40500,5:60500-61000,24:21000-21500,24:21500-22000,14:20500-21000,14:23000-23500,2:106500-107000,24:18500-19000,25:41000-41500,2:107000-107500,2:108000-108500,18:19000-19500,21:66500-67000,2:99000-99500,8:67500-68000,8:60000-60500,23:6500-7000,24:22000-22500,6:39500-40000,1:25500-26000,15:61000-61500,17:47500-48000,12:78000-78500,24:24000-24500,2:107500-108000,18:30000-30500,12:77500-78000,15:40000-40500,18:27500-28000,18:31500-32000,24:23000-23500,8:87500-88000,8:64000-64500,24:26500-27000,8:63500-64000,18:27000-27500,1:24500-25000,8:87000-87500,10:72500-73000,24:5000-5500,2:109500-110000,18:26000-26500,28:42000-42500,21:61500-62000,12:10000-10500,8:70000-70500,24:11000-11500,21:66000-66500,24:57500-58000,8:64500-65000,24:19500-20000,18:23000-23500,12:76500-77000,11:96500-97000,12:77000-77500,21:67000-67500,6:34000-34500,21:62500-63000,1:17500-18000</p>
GIR	<p>17:32500-33000,17:39500-40000,17:38000-38500,1:7000-7500,17:33000-33500,17:37000-37500,17:38500-39000,17:31000-31500,8:68000-68500,17:36500-37000,5:99500-100000,28:5500-6000,3:86500-87000,17:32000-32500,2:94500-95000,17:37500-38000,2:89500-90000,3:83500-84000,17:29500-30000,1:8000-8500,9:71000-71500,20:65000-65500,17:40000-40500,15:26000-26500,3:86000-86500,3:84000-84500,17:31500-32000,13:8000-8500,21:36000-36500,17:34000-34500,18:49000-49500,3:78000-78500,3:71000-71500,3:88500-89000,12:24000-24500,17:39000-39500,25:37500-38000</p>
GNS	<p>14:8000-8500,14:11000-11500,14:10500-11000,14:4500-5000,11:67000-67500,14:9500-10000,7:52000-52500,14:10000-10500,14:7500-8000,18:32500-33000,16:45500-46000,11:99500-100000,11:100000-100500,14:5000-5500,14:7000-7500,14:6000-6500,21:64500-65000,14:11500-12000,14:12000-12500,11:69500-70000,10:51500-52000,14:6500-7000,14:5500-6000,11:22000-22500,12:73000-73500,18:65500-66000,14:8500-9000,11:21000-21500,16:41000-41500,18:36000-36500,4:28000-</p>

	28500,14:15500-16000,6:26000-26500,12:37000-37500,17:61000-61500,2:9000-9500,11:1500-2000
HFD	21:1000-1500,12:74500-75000,21:1500-2000,9:36500-37000,9:60000-60500,2:111500-112000,17:47000-47500,8:32000-32500,12:71500-72000,3:29000-29500,8:29500-30000,8:30000-30500
HOL	15:52000-52500,20:29500-30000,20:28500-29000,20:27000-27500,1:2500-3000,20:26000-26500,20:25500-26000,20:33000-33500,20:26500-27000,21:29500-30000,20:29000-29500,16:68000-68500,20:27500-28000,20:28000-28500,20:25000-25500,1:4500-5000,20:30500-31000,21:70000-70500,5:105500-106000,8:108500-109000,2:87500-88000,20:31000-31500,1:137000-137500,16:74500-75000,20:35000-35500,2:108000-108500,13:0-500,8:35000-35500,10:62000-62500,3:17500-18000,10:70000-70500,10:24000-24500,20:30000-30500,22:11000-11500,26:40500-41000,23:34000-34500
JER	4:94000-94500,10:24000-24500,16:6500-7000,12:71500-72000,18:8500-9000,3:58500-59000,18:9000-9500,3:26000-26500,26:15500-16000,13:10500-11000,10:24500-25000,3:7500-8000,3:7000-7500,21:1000-1500,5:17000-17500,18:9500-10000,3:8000-8500,12:73500-74000,12:76000-76500,8:58500-59000,6:74000-74500,26:2000-2500,18:8000-8500,9:75500-76000,18:38000-38500,2:49000-49500,1:53000-53500,3:6500-7000,4:44000-44500,15:17000-17500,3:4000-4500,14:80500-81000,9:76000-76500,9:71500-72000,4:77500-78000,9:83500-84000,1:52000-52500,18:10000-10500,6:83000-83500,6:100000-100500,4:46000-46500,21:29000-29500,7:33000-33500,2:44500-45000,3:5000-5500,9:70500-71000,9:76500-77000,1:51500-52000,28:42000-42500,4:114500-115000,21:66500-67000,3:3500-4000,26:15000-15500,3:5500-6000,21:59500-60000,25:27000-27500,3:22500-23000,8:58000-58500,13:53000-53500,6:86500-87000,11:65500-66000,16:4000-4500,14:77000-77500,7:32500-33000,21:17000-17500
LMS	3:87000-87500,2:10000-10500,3:108500-109000,20:14000-14500,22:38500-39000,5:4500-5000,15:75000-75500,2:9500-10000,2:9000-9500,3:91000-91500,15:74500-75000,2:5500-6000,3:98500-99000,2:6500-7000,3:84500-85000,3:86500-87000,3:113500-114000,22:39000-39500,15:76000-76500,4:10500-11000,3:85000-85500,3:91500-92000,2:10500-11000,14:23000-23500,3:89500-90000,18:24500-25000,3:101500-102000,3:83000-83500,3:82500-83000,4:12500-13000,14:26500-27000,19:61500-62000,2:8500-9000,15:74000-74500,15:75500-76000
NDA	6:81500-82000,2:38000-38500,2:136500-137000,23:8000-8500,2:36000-36500,23:33500-34000,6:61000-61500,23:32000-32500,23:28500-29000,2:133000-133500,17:15000-15500,23:32500-33000,2:39500-40000,2:31500-32000,11:68000-68500,6:60500-61000,2:32500-33000,14:36500-37000,6:64000-64500,23:34000-34500,3:22000-22500,7:35500-36000,2:46000-46500,2:133500-134000,17:58000-58500,23:7500-8000,7:37000-37500,20:30500-31000,7:30500-31000,21:29500-30000,1:26500-27000,11:62500-63000,22:20500-21000,5:17500-18000,12:37000-37500,6:50500-51000,6:39500-40000,10:66500-67000,9:96500-97000,7:31000-31500,6:57500-58000,6:3500-4000,6:89000-89500,11:69500-70000,9:74500-75000,23:10500-11000,1:36000-36500,23:33000-33500,6:54000-54500,17:63500-64000,8:55000-55500,18:12000-12500,17:11000-11500,5:72000-72500,16:56500-57000,21:23500-24000,4:117500-118000,4:114500-115000,1:50000-50500,12:72000-72500,7:30000-30500,2:42000-42500,23:6500-7000,2:125500-126000,7:36000-36500,2:39000-39500,7:50500-51000,11:70500-71000,21:48000-48500,7:52500-53000,2:29000-29500,3:29000-29500,
NEL	9:4500-5000,3:120500-121000,3:76500-77000,10:69000-69500,13:45000-45500,2:104000-104500,5:49500-50000,11:29500-30000,28:26000-26500,6:11500-12000,20:62000-62500,2:100000-100500,2:103500-104000,16:69000-69500,17:50000-50500

NRC	14:48500-49000,2:93000-93500,14:47500-48000,14:48000-48500,14:50000-50500,14:49000-49500,9:66500-67000,14:44000-44500,14:49500-50000,9:65500-66000,9:68000-68500,13:42000-42500,5:86500-87000,10:51000-51500,2:92500-93000,23:3500-4000,9:70500-71000,4:17000-17500,1:109000-109500,18:33500-34000,9:66000-66500,14:50500-51000,14:39000-39500,18:36000-36500,6:47500-48000,14:26000-26500,14:43500-44000,10:101500-102000,4:17500-18000,4:13000-13500,23:4000-4500,4:11000-11500,1:107000-107500,14:44500-45000,14:51000-51500,1:107500-108000,14:51500-52000,1:106500-107000,2:78500-79000,10:102000-102500,25:7000-7500,25:8000-8500,2:68500-69000,9:62500-63000,5:36500-37000,5:94000-94500,1:108500-109000,9:67500-68000,27:5500-6000,14:47000-47500,1:108000-108500,5:92500-93000,6:48500-49000,2:88500-89000,1:109500-110000,6:48000-48500,17:21500-22000,23:6000-6500,14:23500-24000,5:39000-39500,5:97000-97500,3:110000-110500,15:59000-59500,15:58000-58500,4:7500-8000,2:86500-87000,3:55500-56000,23:3000-3500,14:40500-41000,25:9000-9500,4:11500-12000,18:32500-33000,9:60000-60500,1:112500-113000,14:46000-46500,23:4500-5000,9:71500-72000,1:105500-106000,4:12000-12500,14:25500-26000,21:46000-46500,5:37500-38000,5:71500-72000,4:13500-14000,
PMT	13:17000-17500,16:25000-25500,13:18500-19000,13:15500-16000,4:69500-70000,13:16000-16500,13:25500-26000,4:60500-61000,16:12500-13000,13:18000-18500,16:23000-23500,16:23500-24000,6:61000-61500,10:75500-76000,13:13500-14000,13:14500-15000,3:68500-69000,4:69000-69500,6:34000-34500,13:78000-78500,13:26000-26500,13:28000-28500,13:16500-17000,16:72000-72500,2:8000-8500,13:15000-15500,2:9500-10000,4:68000-68500,16:19000-19500,13:19500-20000,16:25500-26000,6:5000-5500,1:54000-54500,19:42500-43000,2:6500-7000,2:10000-10500,6:33500-34000,1:130500-131000,18:36000-36500,16:24500-25000,8:105500-106000,8:62000-62500,13:19000-19500,22:44000-44500,3:57000-57500,19:43000-43500,2:9000-9500,3:70500-71000,16:12000-12500,22:41000-41500,22:40000-40500
RGU	5:44000-44500,5:45000-45500,5:55500-56000,5:43000-43500,5:54500-55000,5:55000-55500,4:69500-70000,5:43500-44000,6:57000-57500,5:54000-54500,5:40500-41000,5:52000-52500,29:9000-9500,6:58500-59000,5:75500-76000,22:38500-39000,5:52500-53000,9:11500-12000,15:44000-44500,21:19500-20000,5:57500-58000,6:68000-68500,15:75000-75500,4:71500-72000,5:41000-41500,3:38000-38500,22:39000-39500,5:60500-61000,5:45500-46000,15:47500-48000,6:69000-69500,18:38000-38500,6:67500-68000,24:45500-46000,5:57000-57500,15:74000-74500,5:58000-58500,6:85500-86000,15:31500-32000,11:70000-70500,7:53500-54000,1:54500-55000,5:46000-46500,6:45500-46000,9:33500-34000,5:56500-57000,
RMG	5:75000-75500,6:37500-38000,5:75500-76000,5:72000-72500,6:39500-40000,5:76000-76500,9:74000-74500,5:71500-72000,9:70500-71000,9:75000-75500,20:29500-30000,9:74500-75000,6:36500-37000,6:42000-42500,6:36000-36500,20:42500-43000,5:72500-73000,9:75500-76000,6:41500-42000,6:39000-39500,9:73000-73500,16:16000-16500,20:40500-41000,16:79500-80000,6:35000-35500,10:39500-40000,3:72000-72500,9:72000-72500,19:33000-33500,10:38500-39000,10:34500-35000,23:26000-26500,10:37500-38000,16:6500-7000,12:74500-75000,2:84000-84500,20:36500-37000,19:30500-31000,20:43500-44000,20:39500-40000,6:37000-37500,6:38500-39000,6:35500-36000,24:1000-1500,20:37500-38000,10:72500-73000,9:76000-76500

SGT	13:66000-66500,3:75500-76000,3:76000-76500,13:68500-69000,13:67000-67500,13:65000-65500,3:79500-80000,3:75000-75500,13:64500-65000,13:65500-66000,7:48500-49000,13:68000-68500,3:81000-81500,5:17000-17500,3:76500-77000,3:77500-78000,13:39500-40000,13:69500-70000,13:25500-26000,7:50000-50500,14:14500-15000,13:62500-63000,3:74000-74500,3:83000-83500,23:48500-49000,13:64000-64500,11:38500-39000,7:47500-48000,23:49000-49500,9:45500-46000,13:66500-67000,9:46500-47000,3:81500-82000,5:51500-52000,13:69000-69500,5:19500-20000,3:78500-79000,3:74500-75000,13:26000-26500,5:18000-18500,7:49500-50000,3:70500-71000,13:25000-25500,5:19000-19500,3:79000-79500,1:17000-17500,3:73500-74000,13:43500-44000,7:17500-18000,3:78000-78500,
SHK	12:46000-46500,6:5000-5500,12:55000-55500,4:38500-39000,11:62000-62500,4:37500-38000,12:47500-48000,17:39000-39500,1:119000-119500,4:39000-39500,6:60000-60500,6:28000-28500,17:51000-51500,4:59000-59500,2:93000-93500,15:47000-47500,4:40000-40500,24:4500-5000,12:53500-54000,20:19500-20000,12:51500-52000,11:55500-56000,1:118500-119000,17:48000-48500,9:46500-47000,9:21500-22000,8:70500-71000,24:49000-49500,12:50500-51000,15:30500-31000,

ANEXO 6 ELEMENTOS GENÉTICOS DETECTADOS

En la siguiente tabla se muestra el número de elementos anotados que se detectaron en las regiones significativas, principalmente genes formalmente reconocidos, así como pseudogenes y transcritos de ARN.

Raza	Regiones significativas	Elementos anotados	Promedio por región
ANG	33	101	3.06
BMA	49	75	1.53
BRM	93	296	3.18
BSW	111	554	4.99
CHL	65	265	4.08
GIR	37	82	2.22
GNS	37	101	2.73
HFD	12	22	1.83
HOL	36	105	2.92
JER	65	261	4.02
LMS	35	142	4.06
NDA	72	242	3.36
NEL	15	43	2.87
NRC	84	238	2.83
PMT	51	192	3.76
RGU	46	214	4.65
RMG	47	134	2.85
SGT	50	204	4.08
SHK	30	72	2.40