



UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE CIENCIAS VETERINARIAS

**TRABAJO DE TESIS PARA OPTAR AL GRADO DE DOCTOR
DE LA UNIVERSIDAD DE BUENOS AIRES**

**Identificación de variantes alélicas de genes candidatos
involucrados en producción de leche en un rodeo de bovinos
de raza Holando y cruza Holando x Jersey**

Área o disciplina: Genética

Bioq. María Agustina Raschia

Director: Dr. Mario A. Poli

Co-Director: Dr. Ariel F. Amadio

Institución: Instituto de Genética "Ewald A. Favret", CICVyA, INTA Castelar



Julio, 2015

*A mis hijos,
Irene y Joaquín*

AGRADECIMIENTOS.

A las instituciones gubernamentales, Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT) y al Instituto Nacional de Tecnología Agropecuaria (INTA), por haber sostenido el costo de mi formación en el marco de un doctorado.

Al Ing. Agr. Dr. Juan Carlos Salerno, director del Instituto de Genética “Ewald A. Favret” de INTA Castelar, por permitir la realización de la presente tesis.

A los Dres. Mario A. Poli y Ariel F. Amadio, por su confianza, su buena predisposición para transmitir sus conocimientos y por facilitarme las herramientas necesarias para la realización de este trabajo. Por su dedicación, su lectura crítica y sus valiosos aportes al presente manuscrito.

Al personal de Taperitas S. A. por su gentileza, cooperación y buena predisposición para facilitar la toma de muestras de los animales objeto de este estudio, indispensable para abordar la línea de investigación de este trabajo de tesis.

Al Dr. Daniel Maizon, por su buena voluntad para transmitir sus conocimientos y su colaboración en la estimación de valores de cría.

Al Dr. Ezequiel L. Nicolazzi, por su gran colaboración y sus valiosos aportes para la realización del análisis de asociación.

A Juan Pablo Nani, por su colaboración en la toma de muestras y sus comentarios y sugerencias frente a mis consultas.

A María José Beribe, por haber iniciado los estudios estadísticos de esta tesis y, sin proponérselo, alentarme a valerme por mí misma y adquirir independencia al encarar un campo ajeno a mi formación y totalmente nuevo para mí, como el de los modelos estadísticos.

A los evaluadores y la Comisión de Doctorado, por su evaluación criteriosa de la evolución de esta tesis.

A la secretaria de la Escuela de Graduados de la Facultad de Ciencias Veterinarias de la UBA, Rosana, por su excelente predisposición y su manera eficaz y expeditiva de actuar frente a mis solicitudes.

A Ana Rubio y Mara Saucedo, quienes cooperaron en la resolución de los inconvenientes que se presentaron con relación a mi beca.

A Euge, Marga, Irma, Lau y Maxi, por la amistad, la compañía, el apoyo, la solidaridad, los consejos, la ayuda prestada en numerosas tareas de laboratorio y tantos gratos momentos compartidos.

A mi familia y amigos, por su apoyo incondicional.

Especialmente a Matías, por comprenderme, apoyarme y acompañarme incondicionalmente a cada instante y en cada decisión tomada.

ÍNDICE.

ÍNDICE	4
ABREVIATURAS	9
LISTADO DE GENES	11
RESUMEN	15
ABSTRACT	16
INTRODUCCIÓN	17
1. <i>Producción lechera en Argentina. Razas de bovinos lecheros</i>	18
2. <i>Pago de la leche cruda</i>	20
3. <i>Polimorfismos del ADN</i>	20
4. <i>Marcadores moleculares</i>	21
5. <i>Loci para caracteres cuantitativos</i>	24
HIPÓTESIS Y OBJETIVOS	26
1. <i>Hipótesis de trabajo</i>	27
2. <i>Objetivos</i>	27
CAPÍTULO I	28
BÚSQUEDA, IDENTIFICACIÓN Y GENOTIPIFICACIÓN DE MARCADORES MOLECULARES EN GENES CANDIDATOS	28
INTRODUCCIÓN	30
1. <i>Glándula mamaria</i>	30
2. <i>Leche de vaca</i>	30
3. <i>Proteínas relacionadas con la composición y calidad de la leche</i>	30
4. <i>Genoma bovino</i>	31
5. <i>Genes candidatos</i>	32
6. <i>Metodologías de genotipificación de SNPs</i>	37
7. <i>Sistema de genotipificación SNPlex®</i>	37
OBJETIVOS ESPECÍFICOS	42
MATERIALES	43
1. <i>Animales</i>	43
2. <i>Extracción de ADN</i>	43

MÉTODOS.....	44
1. Selección de genes candidatos	44
2. Selección de SNPs a genotipificar	44
3. Genotipificación.....	45
4. Cálculo de factores descriptivos de los genotipos obtenidos.....	45
Equilibrio de Hardy-Weinberg.....	46
5. Determinación de fases de haplotipos e imputación de genotipos faltantes.....	46
RESULTADOS	47
1. Selección de genes candidatos	47
2. Selección de SNPs a genotipificar	49
3. Diseño de los oligonucleótidos utilizados por el sistema de genotipificación SNPlex®	55
4. Descripción de la población genotipificada.....	58
5. Genotipificación.....	61
6. Determinación de fases de haplotipos e imputación de genotipos faltantes.....	67
DISCUSIÓN.....	68
CAPÍTULO II.....	71
GENOTIPIFICACIÓN DE MEDIA DENSIDAD DE MARCADORES MOLECULARES DISTRIBUIDOS EN BLOQUES DE GENES CANDIDATOS.....	71
INTRODUCCIÓN.....	73
1. Genotipificación de baja, media y alta densidad.	73
2. Control de calidad de la genotipificación	73
3. Determinación de haplotipos y asignación de fases.....	78
4. Imputación de genotipos faltantes.....	78
OBJETIVOS ESPECÍFICOS.....	80
MATERIALES.....	81
1. Animales.....	81
2. ADN	81
MÉTODOS.....	82
1. Genotipificación.....	82
2. Cálculo de factores descriptivos de los SNPs evaluados.....	82
3. Control de calidad de la genotipificación	82
4. Comprobación de asignación de paternidades	83
5. Cálculo de factores descriptivos de los genotipos obtenidos.....	83
6. Determinación de haplotipos, asignación de fases e imputación de genotipos faltantes.....	85

RESULTADOS	86
1. Descripción de la población genotipificada	86
2. Descripción de los SNPs incluidos en el chip	87
3. Descripción de los bloques de genes candidatos	88
4. Control de calidad de la genotipificación	89
5. Comprobación de la asignación de paternidades	91
6. Cálculo de factores descriptivos de los genotipos obtenidos	92
7. Determinación de haplotipos, asignación de fases e imputación de genotipos faltantes	95
DISCUSIÓN	96
CAPÍTULO III	101
ANÁLISIS DE ASOCIACIÓN CON PRODUCCIÓN DE LECHE ACUMULADA A LOS 305 DÍAS	101
INTRODUCCIÓN	103
1. Curvas de lactancia del ganado bovino	103
2. Estimación de valores de cría. Modelo animal	105
3. Estudios de asociación	107
OBJETIVOS ESPECÍFICOS	112
MATERIALES Y MÉTODOS	113
1. Información fenotípica	113
2. Información genealógica	114
3. Estimación de valores de cría	114
4. Información genotípica	115
5. Análisis de asociación	115
6. Determinación del desequilibrio de ligamiento	117
RESULTADOS	118
1. Descripción de características fenotípicas	118
2. Estimación de la producción de leche a 305 días	118
3. Estimación de los valores de cría para producción de leche	118
4. Análisis de asociación	120
5. Determinación del desequilibrio de ligamiento y definición de bloques de SNPs en fuerte DL ...	123
DISCUSIÓN	126
CAPÍTULO IV	130
SECUENCIACIÓN DE REGIONES CANDIDATAS EN ANIMALES CON FENOTIPOS CONTRASTANTES	130
OBJETIVOS ESPECÍFICOS	132

MATERIALES Y MÉTODOS.....	133
1. Selección de animales con fenotipos extremos	133
2. Selección de regiones a secuenciar.....	133
3. Secuenciación	134
4. Comparación de las secuencias obtenidas	134
RESULTADOS	135
1. Selección de animales con fenotipos extremos	135
2. Selección de genes y regiones génicas a secuenciar.....	138
3. Secuenciación	143
DISCUSIÓN.....	152
CONSIDERACIONES FINALES	155
ANEXOS	160
ANEXO I	161
<i>Raza Holando</i>	<i>161</i>
<i>Raza Jersey.....</i>	<i>161</i>
<i>Cruza Holando x Jersey.....</i>	<i>162</i>
ANEXO II	163
1. Estructura de la glándula mamaria.....	163
2. Metabolismo durante la gestación y el inicio de la lactancia.....	164
3. Leche de vaca	165
4. Proteínas relacionadas con la composición y calidad de la leche.....	167
ANEXO III	177
<i>Metodologías de genotipificación de SNPs</i>	<i>177</i>
1. Técnicas de hibridación directa.....	177
2. Técnicas que involucran la generación y separación de productos específicos de alelo	177
ANEXO IV	179
<i>Protocolos utilizados</i>	<i>179</i>
1. Técnica de fenol-cloroformo para la extracción de ADN a partir de muestras de semen	179
2. Técnica de secuenciación de regiones seleccionadas de genes candidatos	179
ANEXO V	182
1. Estrategias utilizadas en la selección de genes candidatos.....	182
2. Selección de SNPs	192

ANEXO VI	195
<i>Equilibrio de Hardy-Weinberg</i>	195
ANEXO VII	196
<i>Modelos matemáticos utilizados para describir curvas de lactancia</i>	196
ANEXO VIII	199
<i>Aportes de efectos fijos al valor de cría estimado para producción de leche</i>	199
ANEXO IX	201
<i>Técnica de secuenciación de regiones seleccionadas en animales con fenotipos extremos</i>	201
BIBLIOGRAFÍA	203

ABREVIATURAS.

ACHA: Asociación de Criadores de Holando Argentino

ADN: Ácido desoxirribonucleico

AI: *Average Information*

AIL: Año de inicio de lactancia

ANOVA: *ANalysis Of VAriance*

ARNm: Ácido ribonucleico mensajero

ASO: *Allele-Specific Oligonucleotide*

BLUP: *Best Linear Unbiased Predictor*

BTA: *Bos taurus Autosome*

DL: desequilibrio de ligamiento

dNTP: Desoxinucleótido trifosfato

DE: desviación estándar

EBV: *Estimated Breeding Value*

EDTA: *Ethylenediaminetetraacetic acid*

EHW: Equilibrio de Hardy-Weinberg

EIL: Estación de inicio de lactancia

EM: *Expectation-maximization*

FAM: Frecuencia alélica mínima

FASTA: *FAMily based Score Test Approximation*

FDR: *False Discovery Rate*

GC: *Genomic Control*

GER: *Genomic Equivalent Region*

GWAS: *Genome-Wide Association Study*

h^2 : heredabilidad

IBD: *identity-by-descent*

IBS: *identity-by-state*

ICAR: *International Committee for Animal Recording*

IPA: *Ingenuity® Pathway Analysis*

kb: kilobases

LDL: *Low-density lipoprotein*

LEG: Polinomios de Legendre

LRT: *Likelihood Ratio Test*

LSO: *Locus- Specific Oligonucleotide*

miRNAs: microRNAs

NCBI: *National Center for Biotechnology Information*

NRM: *Numerator Relationship Matrix*

OLA: *Oligonucleotide Ligation Assay*

pb: pares de bases

PCA: *Principal Component Analysis*

PCR: *Polymerase Chain Reaction*

PH: Porcentaje de raza Holando

pH: potencial Hidrógeno

QTL: *Quantitative Trait Locus*

REML: *Restricted Maximum Likelihood*

RNA: *Ribonucleic acid*

RP: Registro Particular

RT: Retrotranscripción

SNP: *Single Nucleotide Polymorphism*

T_a: Temperatura de *annealing*

TE: Tris-EDTA

UDP: *Uridine diphosphate*

UMD: *University of Maryland*

UTR: *untranslated región*

LISTADO DE GENES.

5-OPase: 5-oxoprolinase

ABCG2/BCRP: breast cancer resistance protein - miembro 2 de la subfamilia G de los transportadores ATP binding cassette

ABT1: activator of basal transcription 1

AKR1C1: aldo-keto reductase family 1, member C1 (dihydrodiol dehydrogenase 1; 20-alpha (3-alpha)-hydroxysteroid dehydrogenase)-like

APP: amyloid precursor protein

ARFs: ADP-ribosylation factors

ARL4A: ADP-ribosylation factor-like 4A

BAIAP2: brain-specific angiogenesis inhibitor 1(BAI1)-associated protein 2

bGH: bovine growth hormone

BTN: butyrophilin

BTN1A1: butyrophilin subfamily 1 member A1

CACNG2: calcium channel, voltage-dependent, gamma subunit 2

CAMK2G: calcium/calmodulin-dependent protein kinase II gamma

CCDC90A: coiled-coil domain-containing protein 90A, mitochondrial

COL1A2: collagen, type I, alpha 2

COL22A1: collagen, type XXII, alpha 1

COL6A1: collagen, type VI, alpha 1

COL6A2: collagen, type VI, alpha 2

CSN1S1: α_{s1} -caseína

CSN1S2: α_{s2} -caseína

CSN2: β -caseína

CSN3: κ -caseína

CTBP2: C-terminal binding protein 2

CYHR1: cysteine/histidine-rich 1

DGAT1: diacylglycerol acyltransferase 1

DIP2A: disco-interacting protein 2 homolog A

EEF1D: eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein)

EGF: epidermal growth factor

EIF2C2: eukaryotic translation initiation factor 2C, 2

ETV1: ets variant 1

FANCC: Fanconi anemia, complementation group C

FASN: fatty acid synthase

FBLN5: fibulin 5

FOXA3: forhead box A3

FTCD: formiminotransferase cyclodeaminase

GALNT6: UDP-N-acetyl-alpha-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase 6 (GalNAc-T6)

GH: Growth Hormone

GHR: Growth Hormone Receptor

GML: glycosylphosphatidylinositol anchored molecule like protein

GRID2: glutamate receptor, ionotropic, delta 2

GUCY2C: guanylate cyclase 2C (heat stable enterotoxin receptor)

HIF1A: hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)

HMGN4: high mobility group nucleosome-binding domain-containing protein 4

Ig: Inmunoglobulina

IGF-1: Insulin-like growth factor 1

ITGB5: integrin beta 5

ITPR2: inositol 1,4,5-triphosphate receptor, type 2

KBTBD10: kelch repeat and BTB (POZ) domain containing 10

KCND3: potassium voltage-gated channel, Shal-related subfamily, member 3

KCNQ3: potassium voltage-gated channel, KQT-like subfamily, member 3

KHDRBS3: KH domain containing, RNA binding, signal transduction associated 3

LEC3: latrophilin 3

LEP: leptina

LSS: lanosterol synthase (2,3-oxidosqualene-lanosterol cyclase)

LTF: Lactoferrina o lactotransferrina

MAF1: MAF1 homolog (S. cerevisiae)

MAPK15: mitogen-activated protein kinase 15

MEOX2: mesenchyme homeobox 2

NFIB: nuclear factor I/B

NIPBL: Nipped-B homolog (Drosophila)

NRCAM: neuronal cell adhesion molecule

OLR1: oxidized LDL receptor 1

OPN/SPP1: osteopontina

PCBP3: poly(rC)-binding protein 3

PCNT: pericentrin

PDE9A: phosphodiesterase 9A

PKD2: polycystic kidney disease-2

PLCG1: phospholipase C, gamma 1

PPARGC1A: peroxisome proliferator activated receptor gamma, coactivator 1 alpha

PRL: prolactin

PRLR: prolactin receptor

PRMT2/HRMT1L1: protein arginine N-methyltransferase 2

PSAP: prosaposin

PTK2: protein tyrosine kinase 2

RAC2: ras-related C3 botulinum toxin substrate 2 (rho family, small GTP binding protein Rac2)

RANBP9: RAN binding protein 9

RHPN1: rhophilin, Rho GTPase binding protein 1

RICTOR: RPTOR independent companion of MTOR, complex 2

ROCK2: Rho-associated, coiled-coil containing protein kinase 2

S100B: S100 calcium binding protein B

SA: seroalbúmina

SCARB1: scavenger receptor class B, member 1

SCD1: stearyl-coenzyme A desaturase 1

SERPINA5: serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 5

SLC38A3: solute carrier family 38, member 3

SNX13: sorting nexin 13

SPATC1L: spermatogenesis and centriole associated 1-like

STAT5A: signal transducer and activator of transcription 5A

TG: thyroglobulin

TGF- α : transforming growth factor α

TGF- β : transforming growth factor β

UTMP: uterine milk proteins

*VPS28: vacuolar protein sorting 28 homolog (*S. cerevisiae*)*

YBEY: ybeY metallopeptidase (putative)

ZBTB20: zinc finger and BTB domain containing 20

ZC3H3: zinc finger CCCH-type containing 3

ZFH4: zinc finger homeobox 4

ZNF322A: zinc finger protein 322

α -LA: α -lactoalbúmina

β -LGB: β -lactoglobulina

RESUMEN.

Las razas bovinas Holando y Jersey son dos de las razas productoras de leche más importantes a nivel mundial y las principales en nuestro país. El objetivo del trabajo de tesis fue realizar un análisis exploratorio que permitiera identificar variantes alélicas asociadas a la producción de leche en un rodeo comercial con vacas Holando y cruza Holando x Jersey. El material utilizado consistió en una base de datos de controles lecheros y registros de pedigrí de 1.859 bovinos pertenecientes a 25 familias de medias hermanas paternas de 16 tambos de la cuenca lechera central de la Argentina. A través de un abordaje con genes candidatos, todos los animales se genotipificaron para 56 SNPs de genes seleccionados por estar relacionados con características productivas y calidad de leche en bovinos mediante el sistema SNPlex®. Por otra parte, en 821 animales, se determinaron los genotipos de 12.460 SNPs incluidos en un microarreglo de mediana densidad y localizados en las regiones cromosómicas que contenían los genes candidatos seleccionados previamente. Con la información genotípica y los valores de cría (EBV) para producción de leche acumulada a 305 días durante la primera lactancia, estimados a través de un modelo lineal mixto, se realizó un análisis de asociación mediante la estrategia FASTA-GC del paquete GenABEL de R. Luego de realizar los respectivos controles de calidad de los genotipos, 10.227 SNPs se utilizaron en el análisis de asociación. Este análisis permitió la detección de 12 SNPs asociados al EBV para producción de leche en los BTAs 1, 20, 23 y 24. Por su cercanía a estos SNPs y por estar relacionados con la producción o composición de la leche, se secuenciaron regiones específicas de los genes *PRMT2*, *BTN1A1* y *S100B* en 34 vacas con EBVs extremos. Con las secuencias obtenidas se exploró la presencia de polimorfismos que pudieran contribuir a las diferencias observadas en los valores de EBV para producción de leche de la población bajo estudio. Para las regiones *PRMT2*, *BTN1A1_A* y *BTN1A1_B* se obtuvo el 7%, el 22% y el 24% de los SNPs reportados, respectivamente, mientras que en *S100B* no se halló ningún SNP. Nueve de los SNPs del gen *BTN1A1* generan mutaciones con cambio de sentido. Algunos de los SNPs hallados muestran diferencias apreciables de frecuencias alélicas en animales con diferentes fenotipos (bajo y alto EBV para producción de leche). Este análisis ha revelado tendencias que sería interesante estudiar en una población más numerosa, especialmente para el caso de SNPs en exones que dan lugar a un cambio en la estructura proteica que pueda afectar directamente la producción y las características composicionales de la leche. Los resultados obtenidos son de interés para la caracterización genotípica de los bovinos de Argentina de raza Holando y cruza Holando x Jersey por cuanto la población analizada es representativa de la genética utilizada actualmente en la actividad tambera y podría constituir parte de una población de referencia para la utilización de la metodología de Selección Genómica en mejoramiento de bovinos para leche del país.

ABSTRACT.

The Holstein and Jersey cattle breeds are two of the most important dairy breeds and the majors in our country. The aim of this study was to conduct an exploratory analysis that could identify allelic variants associated with milk production on a commercial herd with Holstein and Holstein x Jersey crossbred cows. The material used consisted in a database containing information about dairy production and pedigree records of 1,859 animals belonging to 25 half-sib families from 16 dairy farms located in the central dairy area of Argentina. Through a candidate gene approach, 1,859 animals were genotyped for 56 SNPs in selected genes, using the SNPlex® system. Those genes were selected for being related to productivity and milk quality in cattle. Moreover, in 821 animals, 12,460 SNPs genotypes included in a medium-density microarray and located in chromosomal regions containing the candidate genes selected previously, were determined. With the genotypic data and estimated breeding values (EBV) for first-lactation 305-days milk yield, estimated through a linear mixed model, an association analysis was performed using the FASTA-GC strategy of GenABEL package in R. After performing the respective genotypes quality controls, 10,227 SNPs were used in the association analysis. This analysis allowed the detection of 12 SNPs associated with EBVs for milk yield in BTAs 1, 20, 23 and 24. Due to its proximity to these SNPs and for being related to milk production or composition, specific regions of *PRMT2*, *BTN1A1* and *S100B* genes were sequenced in 34 cows with extreme EBVs. The obtained sequences were used to explore the presence of polymorphisms that could contribute to the observed differences in the studied population EBVs for milk yield. For *PRMT2*, *BTN1A1_A* and *BTN1A1_B* regions, the 7%, 22% and 24% of the SNPs previously reported were obtained, respectively, while in *S100B* none SNP was found. Nine of *BTN1A1* gene SNPs generate missense mutations. Some of the SNPs found showed significant differences in allele frequencies in animals with opposite phenotypes (lower and higher EBV values for milk yield). This analysis has revealed trends that would be interesting to study in a larger population, especially in the case of SNPs in coding regions that result in a change in the protein structure that may affect milk yield and compositional characteristics. The results obtained are of interest for the genotypic characterization of Argentinean Holstein and Holstein x Jersey cattle, because the studied population is representative of the genetic currently used in the dairy industry and could form part of a reference population to use the Genomic Selection methodology in dairy breeding cattle programs in the country.

INTRODUCCIÓN

Por más de 50 años las evaluaciones genéticas en ganado lechero han tomado información de características fenotípicas (como productividad y tipo) y de relaciones de parentesco para estimar el mérito genético de un reproductor. Sin embargo, en los últimos años se ha comenzado a utilizar información genotípica en forma complementaria a aquella proporcionada por los fenotipos y pedigrís. Actualmente la investigación en genética animal está orientada principalmente a la identificación de QTLs, genes y variantes alélicas que influyen características económicas que puedan ser útiles en programas de cría. En el ganado lechero el objetivo de la mayoría de las investigaciones es la búsqueda del trasfondo genético responsable de la variación en el rendimiento y composición de la leche. Gran cantidad de experimentos se han llevado a cabo con la finalidad de detectar *loci* para caracteres cuantitativos (QTLs). Khatkar y col. (2004) realizaron la primera revisión sobre la existencia de una cantidad importante de QTLs que afectaban caracteres en leche (cantidad y calidad) y sanidad (como por ejemplo, el *score* de células somáticas para el caso de infecciones intramamarias). Ese conocimiento brindó la oportunidad de incrementar el mejoramiento genético por selección directa sobre genes o regiones que afectaban los caracteres (Dekkers y Hospital, 2002).

El objetivo principal de este trabajo de tesis es la identificación de variantes alélicas asociadas a características productivas de la leche bovina. Si bien existen en el país controles lecheros llevados a cabo por diferentes instituciones, lo que genera información para la estimación del mérito genético de los reproductores, hasta el momento es escasa la información utilizada a nivel molecular.

Este trabajo permitiría incrementar la precisión en la estimación del mérito genético al integrar datos moleculares y fenotípicos y contribuiría a una mejor comprensión de los procesos moleculares involucrados en la producción lechera.

En la Introducción se mencionará el estado en que se encuentra la Argentina en relación a la producción lechera bovina y se hará referencia a la reglamentación del pago de la leche cruda en nuestro país. Posteriormente, se abordarán conceptos generales como el de *loci* para caracteres cuantitativos y polimorfismos del ácido desoxirribonucleico (ADN).

En el Capítulo I se hará una descripción del abordaje por genes candidatos para explicar la naturaleza de la variación genética responsable de un QTL y se expondrán los resultados correspondientes a la descripción genotípica de la población analizada. A continuación, en el Capítulo II, se describirá la metodología de barrido del genoma y se presentarán resultados obtenidos según este abordaje para los cromosomas seleccionados por presentar mayor influencia en las características de producción y composición de la leche bovina. Finalmente, con los genotipos obtenidos en los capítulos anteriores, se realizará un análisis de asociación con caracteres fenotípicos en el Capítulo III de este trabajo y, a través de la selección de individuos informativos, en el Capítulo IV se intentará hallar las mutaciones específicas implicadas en la generación de los efectos de los QTLs encontrados.

Por último se realizará una discusión general y se expondrán las conclusiones extraídas.

1. Producción lechera en Argentina. Razas de bovinos lecheros.

Según la Subsecretaría de Lechería del Ministerio de Agricultura, Ganadería y Pesca (http://www.minagri.gob.ar/site/subsecretaria_de_lecheria/lecheria/07_Estadisticas/index.php), en Argentina se produjeron 11.338 millones de litros de leche en el año 2012. Del total, el 76,1% se destinó a la elaboración de derivados lácteos sólidos (quesos, manteca, leche en polvo, crema, etc.), el 17,0% a la elaboración de leches líquidas (esterilizada, pasteurizada, chocolatada) y el 6,9% estuvo constituido por leche que no pasó por el circuito industrial formal. En el mismo año, se consumieron en promedio 216 litros por habitante, teniendo en cuenta tanto la leche fluida como los derivados lácteos.

En el transcurso del año 2012 se exportaron 429.108 toneladas de productos lácteos, principalmente a Venezuela, Brasil y Argelia, lo que representó una ganancia de aproximadamente 1.600 millones de dólares. Los derivados de mayor importancia para la exportación fueron leches en polvo entera y descremada, suero, leches modificadas, manteca y quesos de pasta blanda y semidura. Por otro lado, la importación de productos lácteos que tuvo lugar el mismo año fue de 26.211,51 toneladas (representando alrededor de 58,6 millones de dólares), siendo el suero el principal derivado lácteo importado (Fuente: http://www.minagri.gob.ar/site/subsecretaria_de_lecheria/lecheria/07_Estadisticas/index.php, Subsecretaría de Lechería, MAGyP).

Las principales cuencas lecheras están ubicadas en la "Pampa Húmeda", integrada por las provincias de Buenos Aires, Santa Fe, Córdoba, Entre Ríos y La Pampa. El 70% de la producción total de leche de nuestro país proviene de las cuencas de Santa Fe y Córdoba, el 22% de las cuencas de la provincia de Buenos Aires y el 5% de las cuencas entrerrianas y pampeanas.

En Argentina la principal raza bovina para leche es la Holando. La segunda en importancia es la raza Jersey y en los últimos años ha aumentado su utilización en cruzamientos para mejorar la calidad composicional de la leche. Ambas razas tienen características diferentes tanto en su fenotipo como a nivel productivo y en la composición de la leche que producen. En el Anexo I se describen las características generales, fenotípicas y productivas de los bovinos Holando, Jersey y de los cruzamientos entre ellos.

En la Figura 1 se exponen ejemplares de raza Holando, Jersey y cruza Holando x Jersey que formaron parte del material utilizado.



Figura 1. Razas Holando, Jersey y cruza Holando x Jersey.
Ejemplares de raza Holando (A), Jersey (B) y cruza H x J (C).

2. Pago de la leche cruda.

En Argentina, en el mes de octubre de 2011 entró en vigencia la resolución conjunta 739/2011 y 495/2011 sancionada por los Ministerios de Agricultura, Ganadería y Pesca y de Economía y Finanzas Públicas, que establece la creación de un sistema de pago de la leche cruda sobre la base de atributos de calidad composicional (grasa butirosa, proteínas y sólidos totales) e higiénico-sanitarios en un sistema de liquidación única, mensual, obligatoria y universal. Esta resolución constituye un punto de inflexión en la historia de la lechería argentina, ya que establece la obligatoriedad de pagar la leche cruda por calidad y componentes -proteína y grasa- en lugar de por volumen.

A nivel mundial existe una apreciable cantidad de información disponible gracias a los proyectos genoma de distintas especies, que permite hacer estudios comparativos en pos de describir e identificar los genes responsables de diferentes características fenotípicas, o al menos la relación con éstas. El sistema de producción de leche ofrece una excelente oportunidad técnica y económica para implementar programas de selección asistidos por marcadores (MAS, *Marker Assisted Selection*), debido a que, a nivel internacional y en distintas razas, ya se han identificado genes que afectan caracteres lecheros y han comenzado a publicarse los resultados de las evaluaciones genómicas (Khatkar y col., 2004; Druet y col., 2006; Guillaume y col., 2008; Harris y col., 2008; VanRaden y col., 2009). Incluso, muchas compañías están comercializando toros jóvenes en base a las estimaciones de su valor de cría genómico (Select Sires Inc., Plain City, Ohio, EE.UU.; ABS Global, Inc., Wisconsin, EE.UU.).

A continuación se exponen los conceptos de polimorfismos y marcadores moleculares y se hace una descripción de los principales marcadores genéticos a nivel de ADN utilizados en investigaciones de genética y mejoramiento animal. Luego se introduce el concepto de *loci* para caracteres cuantitativos (QTL, *Quantitative Trait Loci*) y se mencionan las posibilidades que ofrece su detección en el ganado bovino.

3. Polimorfismos del ADN.

Un polimorfismo genético tiene lugar cuando en una región determinada del ADN es posible identificar dos o más alelos alternativos dentro de una población. Estas regiones polimórficas pueden hallarse en secuencias codificantes (exones) o no codificantes (intrones, regiones promotoras, ADN intergénico) del ADN. En el caso de situarse dentro de un gen, cada uno de los alelos puede determinar una variante del producto génico. En otros casos, cuando se localizan en la región promotora de un gen, se puede ver alterada la regulación y el nivel de expresión del mismo. Existen varios tipos de polimorfismos genéticos que difieren entre sí en su estructura, distribución, estabilidad y forma de transmisión.

Los polimorfismos tienen una gran importancia biológica ya que determinan la variabilidad genética de los individuos, causando las diferencias fenotípicas de los mismos. El análisis de los polimorfismos del ADN permite establecer patrones genéticos característicos de cada individuo, que estarán determinados por las herencias materna y paterna.

4. Marcadores moleculares.

La búsqueda de variantes alélicas responsables de la variación fenotípica de un rasgo de interés se realiza frecuentemente a través de la utilización de marcadores moleculares informativos. Un marcador genético es una biomolécula que provee información acerca de la variación genética existente en un determinado *locus*, sin ser necesariamente responsable de la variabilidad a nivel fenotípico.

En los últimos años han aumentado las aplicaciones de la genética molecular en la evaluación de los caracteres hereditarios y la identificación animal. La utilización de marcadores moleculares asociados a determinadas características fenotípicas está adquiriendo un rol cada vez más importante en los programas de selección de reproductores. En el ganado vacuno en particular, la detección de marcadores moleculares relacionados con características productivas contribuirá por medio de esquemas de selección apropiados a mejorar la eficiencia de los animales y el rendimiento y calidad de la leche o carne, mientras que los marcadores asociados con enfermedades genéticas serán útiles para el control de dichas enfermedades en las poblaciones.

Un buen marcador molecular debe reunir una serie de condiciones que maximizan su utilidad, entre ellas, debe presentar un determinado grado de polimorfismo en la población y heredarse de forma estable. Además, la técnica utilizada para analizar el marcador debe ser rápida, práctica y reproducible en distintos laboratorios.

Aunque los primeros marcadores moleculares utilizados consistieron en variantes proteicas, los avances en las técnicas de biología molecular permitieron la introducción de marcadores basados en diferencias en la secuencia del ADN, siendo estos últimos los más difundidos en la actualidad.

4.A. Marcadores basados en variantes proteicas.

Los primeros marcadores desarrollados a finales de la década de 1970 se basaron en la identificación de isoenzimas y formas alternativas en que se presentaban las proteínas en general. Se utilizaron los sistemas de grupos sanguíneos, proteínas como las transferrinas y amilasas presentes en la sangre, y proteínas presentes en la leche, como las caseínas (α -, β - y κ -) y β -lactoglobulinas (Geldermann y col., 1985).

Si bien la variación a nivel proteico refleja sustituciones de aminoácidos causadas en última instancia por variaciones en la secuencia de nucleótidos del ADN, estos marcadores se clasifican como variantes proteicas debido a la metodología utilizada para su detección. Las variaciones en las enzimas y proteínas en general se diferenciaban a través de electroforesis nativas en gel, de acuerdo a las diferentes movilidades en presencia de un campo eléctrico que presentaban los polipéptidos según su tamaño y carga.

4.B. Marcadores basados en variantes del ADN.

Los marcadores a nivel del ADN tienen dos ventajas fundamentales respecto a los marcadores proteicos: en primer lugar, al evaluar directamente variaciones nucleotídicas

permiten detectar cambios en regiones intergénicas o no codificantes y, en segundo lugar, permiten cuantificar el número de mutaciones entre distintos alelos.

A continuación se presentan las principales características de los marcadores moleculares más comúnmente utilizados.

RFLPs (*Restriction Fragment Length Polymorphisms* o Polimorfismos en la longitud de los fragmentos de restricción)

El descubrimiento y aislamiento de endonucleasas de restricción permitió la generación de estos marcadores (Botstein y col., 1980) basados en las alteraciones en el tamaño de los fragmentos resultantes del corte del ADN con enzimas de restricción. Estas alteraciones son debidas a la presencia de inserciones, deleciones o mutaciones puntuales que resultan en la generación o pérdida de sitios de restricción, lo que trae como consecuencia una alteración en el patrón de corte del ADN con una enzima determinada.

Los RFLPs son generalmente bialélicos y su herencia es codominante. La detección del polimorfismo requiere disponer de una sonda de hibridización adecuada, así como también de un método de revelado específico, lo que muchas veces dificulta su utilización. Por este motivo, se ha desarrollado una variante denominada PCR-RFLP que permite un análisis rápido de RFLPs al utilizar la técnica de PCR en lugar de sondas para la detección del polimorfismo.

Estos marcadores han sido utilizados para la búsqueda de QTLs en ganado bovino (Rocha y col., 1992).

VNTRs (*Variable Number Tandem Repeats* o Repeticiones en tándem de número variable)

Este tipo de marcadores consiste en varias copias de una secuencia nucleotídica que se repiten en tándem, es decir, una a continuación de otra. El número de veces en que se repite dicha secuencia define los distintos alelos del marcador, que se heredan en forma codominante. Según el número de bases de la secuencia repetida y la técnica que se utiliza para su detección, los VNTRs se clasifican en mini o microsatélites.

- **Minisatélites.** Estos marcadores fueron descubiertos por Jeffreys y col. (1985) al notar que ciertas sondas de hibridización sobre secuencias repetitivas generaban un complejo patrón de bandas que contenían un polimorfismo heredable. La unidad de repetición en este marcador consta de entre 10 y 100 pb. Su análisis implica como primer paso, una digestión de ADN genómico con enzimas de restricción, seguida de una electroforesis en gel de agarosa y, finalmente, la hibridización con una sonda homóloga específica.

Estos marcadores son altamente polimórficos en la población, lo que hace posible su utilización en la identificación genética de individuos (a través de la “huella digital de ADN” o *ADN fingerprints*).

- **Microsatélites.** Consisten en secuencias de hasta 10 pb repetidas en tándem. Son altamente polimórficos en la población y se encuentran distribuidos abundantemente en el genoma, aunque se hallan con mayor frecuencia en regiones no codificantes que en exones. En

el genoma de mamíferos presentan una densidad promedio de 1 microsatélite cada 100.000 pb. Si bien cubren todo el genoma, no están uniformemente distribuidos a lo largo de los cromosomas ya que su número es menor en las regiones subteloméricas (Koreth y col., 1996).

El análisis de los microsatélites se basa en su amplificación mediante PCR y la detección por electroforesis de las diferencias en el tamaño de los productos obtenidos.

Una desventaja de estos marcadores radica en que son inestables meióticamente. Esta inestabilidad meiótica puede deberse a diferentes motivos: al proceso de deslizamiento (*slippage*) de la ADN polimerasa durante el proceso de replicación, consistente en una disociación transitoria de las hebras de ADN con posterior reasociación errónea (Levinson y Gutman, 1987; Tautz, 1989; Schlotterer y Tautz, 1992), o a un intercambio o recombinación desigual (Jeffreys y col., 1985). De todas formas, el elevado polimorfismo que presentan estos marcadores y la posibilidad de poder detectar ambos alelos los hace muy útiles para la identificación de individuos, ya que resulta muy poco probable que dos individuos elegidos al azar, si son analizados para una serie de marcadores, compartan todos sus alelos.

SNPs (*Single Nucleotide Polymorphisms* o Polimorfismos de nucleótido único o simple)

Un SNP es un cambio de nucleótido en la secuencia de ADN en una posición puntual, aunque en ciertas ocasiones también se considera SNP a un cambio de dos nucleótidos contiguos en una dada posición. El alelo (nucleótido) menos frecuente en dicha posición debe presentar una frecuencia de al menos el 1% en la población, de lo contrario, el polimorfismo no es considerado SNP sino “variante rara” (Brookes, 1999).

Los cambios de nucleótido en un *locus* pueden clasificarse como transiciones o transversiones, siendo las primeras, cambios entre dos purinas o entre dos pirimidinas, mientras que las segundas consisten en cambios entre una pirimidina y una purina. A pesar de existir más probabilidades de ocurrencia de transversiones que de transiciones, las transiciones son los cambios de nucleótidos que se observan con mayor frecuencia en los SNPs (Wang y col., 1998). Una explicación de este hecho radica en la alta tasa de desaminación espontánea de la 5-metilcitosina a timidina en los dinucleótidos CpG, lo que constituye una transición C por T, que se observa como G por A en la hebra de ADN complementaria (Cooper y col., 1989; Holliday y Grigg, 1993).

Estos marcadores son en general bialélicos y constituyen la forma más común de variación dentro del genoma de mamíferos, representando aproximadamente el 90% de los polimorfismos detectados (Collins y col., 1998).

En la Tabla 1 se exponen las principales ventajas y desventajas de los marcadores moleculares microsatélites y SNPs.

De los marcadores expuestos anteriormente, los más utilizados en los últimos años en estudios poblacionales han sido los SNPs. Varias razones justifican la popularidad de su uso como marcadores. La primera es su mayor prevalencia en el genoma en relación a la de los microsatélites, lo que suministra un gran número de marcadores cercanos o dentro del *locus*

de interés. En segundo lugar, puede esperarse que algunos de los SNPs localizados en genes afecten directamente las estructuras de las proteínas que codifican o el nivel de transcripción de los genes, considerándose de esta manera estas alteraciones genéticas como candidatos responsables de los fenotipos observados. Tercero, mientras que las secuencias repetitivas presentan alguna inestabilidad (es decir mutaciones que afectan el largo de la repetición), los SNPs son heredados de forma mucho más estable. Por último, existen numerosas técnicas de genotipificación de SNPs en masa, lo cual disminuye los costos, permite la automatización y posibilita la realización de estudios genéticos más confiables (Landegren y col., 1998).

Marcador	Ventajas	Desventajas
Microsatélites	<ul style="list-style-type: none"> • Muy informativos (gran cantidad de alelos, alta heterocigosidad) • Codominantes 	<ul style="list-style-type: none"> • Alta tasa de mutación • Dificultad para automatizar la técnica de detección • No suficientemente abundantes
SNPs	<ul style="list-style-type: none"> • Baja tasa de mutación • Muy abundantes • Disponibilidad de técnicas que permiten genotipificar gran cantidad de SNPs simultáneamente • Codominantes 	<ul style="list-style-type: none"> • Bajo contenido de información por ser generalmente bialélicos

Tabla 1. Comparación de marcadores moleculares microsatélites y SNPs. Principales ventajas y desventajas de cada uno (adaptación de Schlötterer, 2004).

5. *Loci* para caracteres cuantitativos.

La “Genética Mendeliana Clásica” estudia la relación entre el fenotipo y el genotipo. Las diferencias fenotípicas observadas entre individuos distintos se atribuyen al genotipo de un *locus* determinado. Sin embargo, la mayoría de los caracteres de interés económico en animales en lugar de presentar variaciones fenotípicas discretas exhiben una variación continua y una herencia multifactorial o compleja, hallándose influenciados por numerosos genes y factores ambientales. Las regiones cromosómicas que contienen uno o más genes que influyen un rasgo multifactorial se conocen como *Loci* para Caracteres Cuantitativos o QTLs.

Entre los rumiantes de interés agropecuario, el bovino tanto para leche como para carne, es la especie más estudiada en la búsqueda de QTLs. En el ganado lechero, la mayoría de los estudios tienen por objeto la búsqueda del trasfondo genético que explica la variación en la producción y composición de la leche. Sin embargo, el constante decremento en la fertilidad, así como la incidencia de enfermedades como la mastitis en vacas seleccionadas por su alta performance lechera, condujeron a un aumento en el interés por la identificación de genes que afectan caracteres no productivos como la eficiencia en la reproducción, la longevidad y la salud de los animales (Komisarek y Dorynek, 2009).

El primer intento de detección de QTLs en una población segregante existente (los estudios previos trabajaban con poblaciones experimentales generadas específicamente para la detección de QTLs) tuvo lugar en 1961, año en el que Neimann-Sorensen y Robertson

propusieron un diseño de medio-hermanos para la detección de QTLs en poblaciones comerciales de ganado lechero. Este estudio también fue pionero en el uso de grupos sanguíneos en lugar de marcadores morfológicos y en el uso de análisis estadísticos basados en un *test* χ^2 y en ANOVA (Weller, 2001).

Los estudios de ligamiento marcador-QTL requieren que los marcadores utilizados sean polimórficos en la población bajo estudio. Hasta 1980, los únicos *loci* con herencia mendeliana adecuados para el estudio de QTLs eran los grupos sanguíneos, por ser prevalentes en todas las poblaciones, generalmente multialélicos y no tener un efecto visible a nivel fenotípico sobre ningún carácter de interés. Posteriormente se utilizaron polimorfismos enzimáticos que se detectaban por la diferente velocidad de migración que presentaban en un campo eléctrico. En los últimos años, el descubrimiento de polimorfismos a nivel del ADN (acontecido hacia finales de la década de 1980) y la utilización de marcadores moleculares permitieron la detección de numerosos QTLs asociados a gran variedad de caracteres en el ganado bovino, así como a la determinación de las mutaciones implicadas en los mismos. El primer trabajo que reportó QTLs asociados a características productivas en bovinos lecheros fue realizado por Georges y col. (1995) utilizando microsatélites. Desde ese trabajo pionero han surgido numerosas publicaciones relacionadas con el tema (Kolbehdari y col., 2009; Olsen y col., 2007; Schnabel y col., 2005; Wang y col., 2012).

Los QTLs públicamente disponibles, así como la información sobre asociaciones entre genes y SNPs en animales domésticos se almacenan en la base de datos *Animal QTL database* (Hu y col., 2013; QTLdb; <http://www.animalgenome.org/QTLdb>). En bovinos, en particular, esta base de datos contiene 17.908 QTLs reportados, surgidos de 588 publicaciones y que representan 514 caracteres diferentes, siendo 1.043 los asociados al rendimiento o producción de leche (junio de 2015).

Una vez que un QTL que afecta un carácter particular ha sido detectado, éste puede ser utilizado en programas de selección. Hasta hace algunos años, la selección de animales destinados a producción se basaba únicamente en la información fenotípica de los mismos o de sus hijas. Sin embargo, dada la influencia de los factores medioambientales y de varios QTLs o genes sobre la expresión de ciertos caracteres, la información fenotípica resulta incompleta. El uso de técnicas moleculares provee información de los animales a nivel genético, posibilitando estudiar la segregación de segmentos particulares de ADN entre padres e hijos. Esto puede usarse, por ejemplo, para clasificar a las hijas de un toro en dos grupos, en base al alelo de un marcador molecular que han recibido de su padre. Las diferencias que muestran estos grupos con respecto a caracteres relevantes desde el punto de vista productivo sugieren que ese segmento de ADN (o QTL) está implicado en la expresión del rasgo evaluado (Schrooten, 2004). Soller y Beckmann (1982) fueron los primeros en visualizar el beneficio que brinda incluir en programas de cría a la selección por marcadores genéticos (MAS, *Marker Assisted Selection*) con respecto a la respuesta genética del ganado lechero.

HIPÓTESIS Y OBJETIVOS

1. Hipótesis de trabajo.

El desarrollo de diferentes tecnologías moleculares en las últimas dos décadas permitió la secuenciación completa del genoma de la mayoría de las especies de interés agropecuario. Prácticamente no existen limitaciones tecnológicas para la determinación de las variaciones a nivel del ADN en animales. Estas variaciones o polimorfismos pueden localizarse en regiones tanto codificantes como no codificantes, las primeras afectando en forma directa el producto proteico codificado y las segundas teniendo injerencia en los niveles de expresión, así como en el procesamiento de los ARNm. En ambos casos, los polimorfismos pueden ser responsables de una variación cuali/cuantitativa de determinados rasgos fenotípicos, en forma aislada o contribuyendo a complejas variaciones al implicar un conjunto de genes.

Numerosas experiencias a nivel mundial en bovinos para leche han permitido detectar, a diferentes niveles resolutivos, regiones cromosómicas, QTLs y variantes alélicas, que afectan características relacionadas con la producción y la composición de la leche. Sin embargo, actualmente es escasa la información en bovinos Holando y cruza Holando x Jersey de nuestro país. Es de esperar que debido a la conexión que existe entre los rodeos de Argentina y el resto del mundo, diversos QTLs y variantes alélicas implicados en la determinación de rasgos fenotípicos puedan ser compartidos por estos rodeos, aunque también pueden presentarse nuevas variantes, propias de cada rodeo. En ambas situaciones la estimación del efecto genético de esas variantes muy probablemente sea diferente debido a los efectos ambientales y al trasfondo genético de estos rodeos. La información surgida podría ser de utilidad para la estimación del mérito genético de los animales y en la aplicación de planes de mejoramiento destinados a incrementar la calidad y la producción de leche.

2. Objetivos.

2.A. General.

Identificar variantes alélicas asociadas a características productivas de la leche bovina.

2.B. Específicos.

- Desarrollar una metodología de selección de genes candidatos relacionados con características productivas y atributos de calidad de leche en bovinos lecheros. Identificar marcadores moleculares en los genes candidatos seleccionados.
- Determinar el genotipo de los marcadores moleculares seleccionados en genes candidatos así como el de otros marcadores moleculares no necesariamente relacionados con el fenotipo en estudio, en un rodeo comercial de razas Holando y cruza Holando x Jersey y caracterizarlos en función de parámetros poblacionales.
- Integrar la información molecular con información genealógica y fenotípica de modo de establecer asociaciones entre las mismas.

CAPÍTULO I

**BÚSQUEDA, IDENTIFICACIÓN Y GENOTIPIFICACIÓN DE
MARCADORES MOLECULARES EN GENES CANDIDATOS**

En el primer capítulo de este trabajo de tesis se describe el criterio seguido y la metodología empleada para realizar la búsqueda, identificación y genotipificación de marcadores moleculares en genes candidatos.

En primer lugar se hace una breve introducción sobre la estructura y fisiología de la glándula mamaria, los genes que codifican las principales proteínas de la leche y aquellos que participan en los procesos de producción y secreción de la leche bovina. A continuación se expone el concepto de genes candidatos, se mencionan distintas estrategias utilizadas para la búsqueda de los mismos y se describen las alteraciones que pueden tener lugar a nivel de las proteínas como consecuencia de mutaciones ocurridas en estos genes. Finalmente, se mencionan las distintas metodologías existentes para realizar la genotipificación de polimorfismos de nucleótido simple.

La introducción se continúa con una descripción de los materiales y métodos utilizados para realizar este trabajo, los objetivos específicos propuestos para esta primera parte y los resultados obtenidos. El capítulo finaliza con una discusión en la que se exponen las conclusiones extraídas.

INTRODUCCIÓN.

1. Glándula mamaria.

La glándula mamaria bovina o ubre es una glándula exocrina, ya que la leche es sintetizada en células especializadas agrupadas en alvéolos y luego excretada fuera del cuerpo por medio de un sistema de conductos.

Una descripción de la estructura y fisiología de la glándula mamaria bovina, así como del tipo de metabolismo que tiene lugar en las vacas durante la preñez y el inicio de la lactancia puede encontrarse en el Anexo II.

2. Leche de vaca.

La leche es un líquido de color blanco opaco de composición y estructura compleja. La materia grasa se encuentra en emulsión, las proteínas constituyen una suspensión, mientras que los restantes componentes (lactosa, otras sustancias nitrogenadas, minerales, etc.) están disueltos. La composición de la leche varía considerablemente con la raza de la vaca, el estado de lactancia, el alimento ingerido, la época del año y muchos otros factores. La Tabla I.1 indica los promedios de la composición química de la leche de vaca según Taverna y col. (2001).

Componentes	Promedios generales (g/Kg leche)
Agua	881,5
Materia seca	119,5
Lactosa	46,1
Grasa	35,1
Proteína total	31,7
Cenizas	6,6
Minerales (Calcio, Fósforo, Magnesio, Potasio, Sodio, Cloro)	5,8

Tabla I.1. Composición química promedio de la leche bovina.

3. Proteínas relacionadas con la composición y calidad de la leche.

Entre las proteínas relacionadas directa o indirectamente con la composición y calidad de la leche bovina pueden encontrarse proteínas con función estructural, nutritiva, enzimas pertenecientes a rutas metabólicas, proteínas involucradas en vías de transducción de señales, etc. A continuación se enumerarán y describirán las principales características de diferentes proteínas que componen la leche bovina o que participan en los procesos de su producción y secreción.

3.A. Proteínas que componen la leche bovina.

Las proteínas que componen la leche bovina pueden dividirse en dos grandes grupos: la fracción caseínica y las proteínas del suero de leche. Farell y col. (2004) hicieron una extensa revisión sobre ellas. La inducción de los genes que codifican estas proteínas está bajo el control

de un complejo proceso multihormonal que involucra esencialmente a la prolactina, glucocorticoides, insulina, hormona de crecimiento y otros factores lactogénicos. A continuación se mencionarán las proteínas que componen cada uno de estos grupos y en el Anexo II de esta tesis se describirán sus características principales.

3.A.1. Caseínas.

Las caseínas de la leche del género *Bos* se definieron originalmente como aquellas fosfoproteínas capaces de precipitar de la leche cruda descremada por acidificación a pH 4,6 a 20°C (Farrell y col., 2004). Según su estructura primaria, las caseínas pueden ser clasificadas en cuatro familias: α_{s1} -, α_{s2} -, β - y κ -caseínas. Todas ellas son codificadas por genes ubicados en el cromosoma 6 del genoma bovino.

3.A.2. Proteínas del suero.

La denominación “proteínas del suero de leche” hace referencia al conjunto de proteínas de la leche que permanecen solubles luego de la precipitación de las caseínas a pH 4,6 y 20°C (Farrell y col., 2004). Entre ellas se encuentran la β -lactoglobulina, la α -lactoalbúmina, la seroalbúmina, las inmunoglobulinas, la lactoferrina y la fracción proteosa-peptona. Asimismo, permanecen solubles luego de la precipitación de caseínas, y por lo tanto forman parte de este grupo, pequeños fragmentos que surgen de la proteólisis de caseínas, así como también proteínas que constituyen la membrana de los glóbulos de grasa.

3.B. Proteínas que participan en los procesos de producción, secreción y eyección de la leche bovina.

Los procesos de producción, secreción y eyección de la leche bovina están controlados de forma endócrina. La producción y secreción de leche son los procesos que comprenden la síntesis de los componentes de la leche por las células alveolares y su excreción hacia el lumen y están regulados por las hormonas prolactina y hormona de crecimiento de la glándula pituitaria anterior (Svennersten-Sjaunja y Olsson, 2005). La eyección de la leche consiste en el pasaje de la leche secretada y mantenida en los alvéolos hacia grandes ductos y cisternas a partir de las cuales se hace accesible a las crías por succión. Este proceso es dependiente de la liberación de la hormona oxitocina por la glándula pituitaria posterior (Cowie, 1972).

Una descripción más detallada de la endocrinología de la producción y de la secreción de leche puede encontrarse en el Anexo II.

4. Genoma bovino.

Las razas bovinas actuales pertenecen a las razas vacunas originales *Bos indicus*, de origen asiático, o *Bos taurus*, de origen europeo. El *Bos taurus* es característico de zonas templadas continentales, mientras que el *Bos indicus* lo es de países tropicales y presenta más resistencia al calor y a ciertas enfermedades y parásitos que los descendientes del *Bos taurus*. Las diferencias existentes entre los distintos tipos son tanto fenotípicas como genotípicas. La especie *Bos indicus* incluye los bovinos con giba pertenecientes al grupo conocido como cebuínos, mientras que los bovinos *Bos taurus* no presentan giba. A nivel genético, una

diferencia entre las dos especies se encuentra en el cromosoma Y: en *Bos indicus* este cromosoma es acrocéntrico mientras que en *Bos taurus* es submetacéntrico. Las principales razas productoras de leche (Holando, Jersey, Ayrshire y Pardo Suizo) pertenecen al grupo *Bos taurus* (Fuente: Enciclopedia Alfabética del Campo).

El genoma bovino consiste en 29 pares de autosomas (BTAs) y un par de cromosomas sexuales. El primer borrador de la secuencia de su genoma fue generado a partir de un individuo de la raza Hereford, mientras que para la búsqueda de SNPs se utilizaron cinco razas diferentes *Bos taurus*: Holando, Angus, Jersey, Limousin y Norwegian Red, y una *Bos indicus*: Brahman (The Bovine HapMap Consortium, 2009). Existen dos ensamblados del genoma bovino: el Btau4.0 y el UMD3.1. El primero fue dado a conocer por el Centro de Secuenciación del Genoma Humano del *Baylor College of Medicine* (Houston, Texas, EE.UU.) en octubre del año 2007, mientras que el último fue realizado por el Centro de Bioinformática y Biología Computacional de la Universidad de Maryland (EE.UU.) y está disponible desde diciembre del año 2009. En este trabajo se hará referencia al ensamblado UMD3.1 por haber ubicado en cromosomas mayor cantidad de pares de bases del genoma bovino que el Btau4.0. De la información obtenida tanto de los mapas físicos y genéticos como de la secuenciación, se estima que el total del genoma bovino consta de 2,87 Gpb (2.870 millones de pares de bases) y un mínimo de 22.000 genes (The Bovine Genome Sequencing and Analysis Consortium, 2009).

5. Genes candidatos.

Los rasgos fenotípicos complejos, como la producción y composición de la leche bovina, involucran varias vías biológicas, por lo que implican numerosos genes, con mayor o menor contribución a la generación del fenotipo.

Se entiende como **genes candidatos** aquellos genes codificantes de proteínas cuyas funciones sugieran que pueden estar involucradas en la expresión de un fenotipo o enfermedad de interés (Georges, 2012). En este trabajo de tesis en particular, se consideran genes candidatos aquellos que potencialmente pueden causar alteraciones en la producción o las características composicionales de la leche bovina.

Los estudios de asociación que utilizan genes candidatos asumen que éstos están involucrados en un aspecto fisiológico o vía metabólica del rasgo en estudio y que su variación genética es responsable, al menos en parte, de la variación fenotípica observada. Por lo tanto, la búsqueda en el genoma se hace de manera orientada y limitada. El objetivo principal es encontrar una correlación estadística entre variantes genéticas específicas de genes candidatos y el rasgo de interés. Los distintos pasos seguidos al utilizar la vía de los genes candidatos pueden resumirse de la siguiente manera:

- 1- En primer lugar se generan hipótesis e identifican genes candidatos que se relacionan directa o indirectamente con el rasgo complejo, para lo que es necesario investigar cuáles son los tejidos, órganos, células, proteínas y genes involucrados de alguna manera en las distintas vías metabólicas implicadas en la generación del fenotipo;
- 2- Posteriormente, se identifican variantes nucleotídicas dentro de los genes candidatos o cercanas a ellos que puedan causar un cambio a nivel de la estructura de la proteína

o de la regulación de la expresión génica, o bien que puedan hallarse en desequilibrio de ligamiento con las variantes causantes de dichas alteraciones;

- 3- Finalmente, se genotifican las variantes seleccionadas en la población y se emplean métodos estadísticos de modo de establecer la existencia o no de una asociación entre dichas variantes y el fenotipo estudiado.

Los estudios de asociación que utilizan genes candidatos presentan ciertas limitaciones (Tabor y col., 2002; Georges, 2012):

- Están limitados al estudio de genes con funciones conocidas, los que representan una pequeña fracción del genoma. Muchos críticos argumentan que la capacidad de detección de las variantes responsables del fenotipo es limitada en relación a estudios que utilizan marcadores distribuidos a lo largo de todo el genoma, independientemente de su función o ubicación en el contexto de genes específicos.
- La probabilidad de encontrar el alelo responsable de las variaciones observadas a nivel fenotípico es baja. En la mayoría de los casos, las asociaciones que se detectan son indirectas, es decir, debidas al desequilibrio de ligamiento existente entre los polimorfismos analizados y las verdaderas variantes responsables del fenotipo. Por este mismo motivo, la ausencia de una asociación significativa entre genotipo y fenotipo no excluye la posibilidad de que el gen candidato presente una variante responsable de la generación del fenotipo estudiado (que no esté en desequilibrio de ligamiento con ninguno de los polimorfismos analizados).
- En varias ocasiones, una asociación estadísticamente significativa entre un gen candidato y un fenotipo determinado no ha podido ser reproducida en diferentes estudios de asociación. Esto puede deberse a que muchos estudios con genes candidatos no tienen en cuenta la estratificación poblacional, lo que puede ser responsable de asociaciones falsamente positivas.

Sin embargo, estos estudios son considerados muy útiles como primer paso en la exploración de las vías genéticamente determinadas y potencialmente responsables de variantes fenotípicas de rasgos complejos. La detección de una asociación significativa para una variante genética lleva a hacer un estudio más profundo del gen correspondiente y esa u otras variantes dentro del mismo en una población independiente. Asimismo, puede estudiarse la relación entre esas mismas variantes génicas y otras características fenotípicas que impliquen las mismas vías biológicas o similares. Además, la detección de asociaciones positivas permite formular hipótesis acerca de los mecanismos moleculares y bioquímicos responsables de los cambios a nivel fenotípico asociados a las variantes alélicas.

5.A. Estrategias de selección de genes candidatos.

La selección de genes candidatos relacionados al fenotipo en estudio requiere conocer los procesos metabólicos y productos génicos implicados en la generación del rasgo. Es necesario definir los criterios de búsqueda a seguir y realizar una evaluación cuidadosa de los genes con mayor probabilidad de ser responsables del fenotipo.

Los criterios de selección de genes candidatos definidos por Ogorevc y col. (2009) resumen las estrategias de búsqueda mencionadas a continuación:

- Genes codificantes de proteínas relacionadas al fenotipo estudiado que presenten diferentes variantes genéticas;
- Genes con variantes alélicas que hayan sido reportadas en bibliografía como asociadas al rasgo en estudio;
- Genes con perfiles de expresión asociados a la variación fenotípica del rasgo analizado;
- Genes localizados en QTLs asociados al fenotipo en estudio;
- Genes regulados post-transcripcionalmente por miRNAs expresados en el/los órganos relacionados al rasgo estudiado;
- Genes cuya expresión alterada en animales *knock out* o transgénicos haya resultado en fenotipos específicos asociados al rasgo en estudio;

Una vez seleccionados los genes candidatos a estudiar, se deben elegir las variantes alélicas que se analizarán. Es importante determinar si las variantes a evaluar pueden ser responsables de alteraciones en la estructura o función de los productos génicos, para lo cual se debe tener en cuenta el tipo de mutación, su ubicación dentro del gen y su frecuencia poblacional.

5.B. Selección de variantes alélicas de genes candidatos.

5.B.1. Tipos de mutaciones.

Una mutación es un cambio en la secuencia del ADN. Las mutaciones tienen lugar a lo largo de todo el genoma, aunque su frecuencia de aparición es mayor en regiones intergénicas o no codificantes ya que la presión de selección en estas regiones es menor que en las regiones codificantes.

Los tipos de mutaciones que pueden encontrarse en los genes son las siguientes:

- **Mutaciones dinámicas:** son repeticiones en *tándem* de una secuencia en general corta de nucleótidos. Son responsables de algunas enfermedades humanas como la Enfermedad de Huntington y el Síndrome del X-frágil. Estas enfermedades suelen presentar el fenómeno de anticipación, es decir, su manifestación es más grave y de inicio más temprano al pasar de una generación a la siguiente. Este fenómeno puede explicarse de acuerdo a la inestabilidad meiótica de los microsatélites (Thompson & Thompson, 2008).
- **Indels:** están constituidos por inserciones o deleciones de uno o más nucleótidos en la secuencia génica. Si el número de nucleótidos contiguos insertados o delecionados es múltiplo de tres y la mutación tiene lugar en un exón, se obtiene como producto una proteína con uno o más aminoácidos extras o ausentes, lo que puede o no tener consecuencias a nivel fenotípico. Por otro lado, si el número de nucleótidos implicados en el indel no es múltiplo de

tres, cambia el marco de lectura de la región codificante de la proteína, lo que implica consecuencias generalmente muy negativas sobre la estructura y función de la misma.

▪ **Mutaciones puntuales:** son cambios de un nucleótido por otro en la secuencia de ADN. Sus consecuencias a nivel fenotípico presentan un amplio rango de variación, según el tipo de mutación puntual presente. Su clasificación se basa en su ubicación dentro del gen y su efecto funcional y se indica en la Tabla I.2.

Variante	Ubicación	Consecuencia funcional	Frecuencia	Riesgo sobre el fenotipo
Sin sentido	Secuencia codificante	Terminación prematura de la traducción	Muy baja	Muy alto
No sinónima (de cambio de sentido)/ no conservativa	Secuencia codificante	Cambio de un aminoácido por otro de propiedades diferentes	Baja	Moderado a muy alto
No sinónima (de cambio de sentido)/ conservativa	Secuencia codificante	Cambio de un aminoácido por otro de propiedades similares	Baja	Bajo a muy alto
Sinónima	Secuencia codificante	Puede alterar el <i>splicing</i> del ARNm	Media	Bajo a alto
En la región reguladora	Promotor, 5'UTR, 3'UTR	Puede afectar el nivel, la ubicación o el momento de la expresión génica	Baja media ^a	Bajo a alto
En los sitios de unión intrón/exón	Sitio de <i>splicing</i> del ARNm	Puede alterar el <i>splicing</i> y ensamblado del ARNm	Baja	Bajo a alto
Intrónica	Intrón	Puede afectar la expresión o la estabilidad del ARNm	Media	Muy bajo

Tabla I.2. Tipos de mutaciones puntuales.

Distintos tipos de mutaciones puntuales clasificadas en función de su posición en el genoma, así como la consecuencia que pueden tener a nivel funcional, su frecuencia de aparición y el riesgo que implican sobre el fenotipo.

Adaptación de Tabor y col., 2002.

De las mutaciones expuestas anteriormente, las que presentan mayor frecuencia son las mutaciones puntuales, constituyendo polimorfismos (SNPs) si tienen una frecuencia alélica mínima en la población superior al 1%.

En este trabajo de tesis se utilizarán SNPs en lugar de otros marcadores moleculares debido a las razones siguientes (Landegren y col., 1998):

-su abundancia y distribución en el genoma posibilita la existencia de un gran número de marcadores cercanos o dentro del locus de interés,

-los SNPs localizados en genes pueden ser responsables de los fenotipos observados al afectar directamente el nivel de transcripción génica o la estructura de las proteínas codificadas por los mismos,

-presentan, al contrario que las secuencias nucleotídicas repetidas en *tándem*, estabilidad meiótica,

-la existencia de numerosas técnicas de genotipificación de SNPs en masa, disminuye los costos y permite la automatización.

Si bien los microsatélites son más informativos que los SNPs por presentar gran cantidad de alelos en la población en lugar de los dos alelos que generalmente presentan los SNPs, si se analiza un mayor número de SNPs que el que se analizaría de microsatélites, se obtienen resultados comparables en estudios de mapeo/ligamiento e identificación de individuos.

5.B.2. Selección de SNPs de genes candidatos.

Dado que en las bases de datos públicas existe gran cantidad de SNPs reportados en la mayoría de los genes, es necesario establecer un orden de prioridades al hacer la selección de SNPs de genes candidatos que se genotificarán. Es deseable estudiar sólo aquellos polimorfismos que afecten la función o la expresión de una proteína, ya que son los más probablemente responsables de desarrollar un determinado fenotipo.

Para establecer el orden de prioridades, debe utilizarse la información acerca del tipo y ubicación de mutaciones en el ADN. Muchas variaciones en el ADN implican una alteración de la función proteica, como por ejemplo mutaciones no sinónimas que cambian un aminoácido por otro en la proteína, o las mutaciones sin sentido que resultan en la aparición de un codón de terminación prematuro. Estos tipos de polimorfismos es más probable que sean responsables de la generación de fenotipos particulares, por lo que se les debe asignar una alta prioridad en el momento de seleccionar los SNPs de genes candidatos a genotipificar.

Los polimorfismos localizados en secuencias promotoras o reguladoras de la expresión génica también es probable que impliquen consecuencias a nivel de la función proteica, sobre todo si su ubicación precisa tiene lugar en secuencias de ADN altamente conservadas entre promotores de genes relacionados o en individuos de diferentes especies. Por lo tanto, también se debe asignar una alta prioridad para genotipificar este tipo de variantes génicas.

Los polimorfismos que tienen lugar en regiones codificantes pero que no implican un cambio de aminoácidos, o los que se dan fuera de las regiones codificantes de los genes, aunque no resultan en un cambio de la secuencia primaria de la proteína pueden alterar la estabilidad o el *splicing* del ARNm. Por lo tanto, si bien no se les asigna una prioridad tan alta como a los polimorfismos expuestos en los párrafos precedentes, debe asignárseles mayor prioridad que a los polimorfismos localizados en intrones (Tabor y col., 2002).

Además de considerar las consecuencias que pueden tener los polimorfismos a nivel de la función proteica, es importante tener en cuenta la frecuencia que presentan en la población a ser analizada. La asociación de aquellos SNPs que presenten una frecuencia alélica mínima muy baja será detectada por estudios de asociación que utilizan genes candidatos sólo si son responsables de una gran parte de la variación fenotípica del rasgo estudiado (Tabor y col., 2002).

Finalmente, una última consideración a tener en cuenta al seleccionar SNPs para ser incluidos en estudios de asociación consiste en evaluar la existencia de desequilibrio de ligamiento en el gen candidato en la población a estudiar. Los SNPs que se seleccionan específicamente para capturar la variación que tiene lugar en sitios cercanos del genoma se denominan “*tag SNPs*” (Bush y Moore, 2012). Si varios SNPs de un gen están en desequilibrio de ligamiento, es posible inferir a partir del genotipo de un único SNP del grupo (“*tag SNP*”), el genotipo de los restantes. Dado que los patrones de desequilibrio de ligamiento son específicos de la población estudiada, también lo son los *tag SNPs*.

6. Metodologías de genotipificación de SNPs.

Existen varias técnicas para genotipificar SNPs. Algunas de ellas son técnicas de hibridación directa basadas en la utilización de oligonucleótidos específicos de alelo, que no implican etapas de separación de reactivos o fases. Entre ellas se encuentran las metodologías que involucran la utilización de sondas TaqMan® o las que utilizan *Molecular Beacons*. Mientras que otras involucran la generación y separación de productos específicos de alelo, como las técnicas de digestión con enzimas de restricción, el estudio de la conformación del ADN de hebra simple, la PCR alelo-específica, la reacción de ligación de oligonucleótidos y la pirosecuenciación. Una descripción de cada una de estas técnicas o metodologías puede hallarse en el Anexo III.

En la genotipificación de SNPs por el abordaje de genes candidatos de este trabajo de tesis se utilizó un sistema comercial que consiste en una reacción de ligación de oligonucleótidos, seguida de una PCR y posterior separación de los productos obtenidos por electroforesis capilar. La descripción de cada una de las etapas que implica el empleo de este sistema, así como las prestaciones del mismo se exponen a continuación.

7. Sistema de genotipificación SNPlex®.

De los métodos de genotipificación disponibles y accesibles económicamente, este sistema comercial, desarrollado por la compañía Applied Biosystems (California, Foster City, EE.UU.), fue el método de elección teniendo en consideración el número de muestras a genotipificar, la posibilidad de analizar simultáneamente 48 marcadores SNPs bialélicos por muestra y el equipamiento disponible en el laboratorio.

7.A. Técnica SNPlex®.

La técnica comprende siete pasos que pueden ser llevados a cabo en dos días (Figura I.1). El fundamento de cada uno de ellos se expone a continuación (Tobler y col., 2005; De La Vega y col., 2005):

Reacción de ligación de oligonucleótidos (OLA, *Oligonucleotide Ligation Assay*).

En este primer paso tiene lugar la hibridación de oligonucleótidos específicos de alelo (ASO) y oligonucleótidos específicos de *locus* (LSO) al ADN a analizar. Luego se produce la ligación de las sondas ASO y LSO. Esta reacción requiere que la complementariedad de bases entre las sondas ASO y la secuencia de ADN a analizar sea perfecta en el sitio del SNP. En

simultáneo, se ligan adaptadores universales a los extremos de los ASO y LSO. Estos adaptadores contienen secuencias de unión a *primers* de PCR universales, así como también secuencias complementarias a las sondas ASO y LSO. En las sondas ASO se distinguen dos regiones: una región de secuencia complementaria al ADN a analizar (“GER”, *Genomic Equivalent Region*) ubicada hacia el extremo 3’, y una secuencia única denominada ZipCode hacia el extremo 5’. Consecuentemente, en virtud de la secuencia ZipCode unida a productos de ligación específicos, la reacción OLA distingue la información genotípica de cada SNP.

En la reacción OLA se utilizan 97 adaptadores (“*linkers*”) universales: dos *linkers* diferentes que se unirán a cada uno de los ASO correspondientes a cada SNP a genotipificar, más un *linker* común capaz de unirse a todos los LSO. Cada uno de estos adaptadores es sintetizado con un espaciador interno (“*spacer*”) que previene la digestión completa del oligonucleótido durante el paso de purificación con exonucleasas (paso 2). Los adaptadores que se unen a las sondas ASO forman una estructura de horquilla cuyo extremo 3’ es complementario a la secuencia ZipCode del ASO correspondiente. Asimismo, estos adaptadores también contienen una secuencia idéntica a la del *primer* directo que se usará en la reacción de PCR que tendrá lugar en el paso 3. Por otro lado, el adaptador que se une a las sondas LSO también forma una estructura de horquilla y contiene una secuencia complementaria al *primer* reverso que se utilizará en la reacción de PCR.

Tanto los adaptadores como las sondas ASO y LSO son activados por fosforilación previamente a la reacción de ligación sobre el ADN genómico a analizar. El sistema requiere la utilización de ADN genómico fraccionado térmicamente de forma que posea un tamaño de entre 2 y 7 kb.

Purificación de los productos de la reacción OLA por digestión con exonucleasas.

Este segundo paso consiste en la remoción de las sondas ASO y LSO y los adaptadores no ligados o parcialmente ligados, así como del ADN genómico en exceso mediante digestión enzimática. Las enzimas que catalizan la reacción son la exonucleasa-I y la λ -exonucleasa. Este paso asegura la eficiencia de la reacción de PCR que tiene lugar a continuación.

Reacción en cadena de la polimerasa (PCR).

Consiste en la amplificación simultánea de los productos de ligación purificados y diluidos utilizando un único par de *primers* de PCR común a todos los *loci* a analizar. Uno de los *primers* está biotinilado.

Captura de los productos de PCR biotinilados en placas cubiertas con estreptavidina.

Este paso consiste en una incubación que favorece la unión de los amplicones de hebra simple biotinilados a la estreptavidina que recubre los pocillos de una microplaca provista por el fabricante, seguida de la remoción de las hebras no biotiniladas.

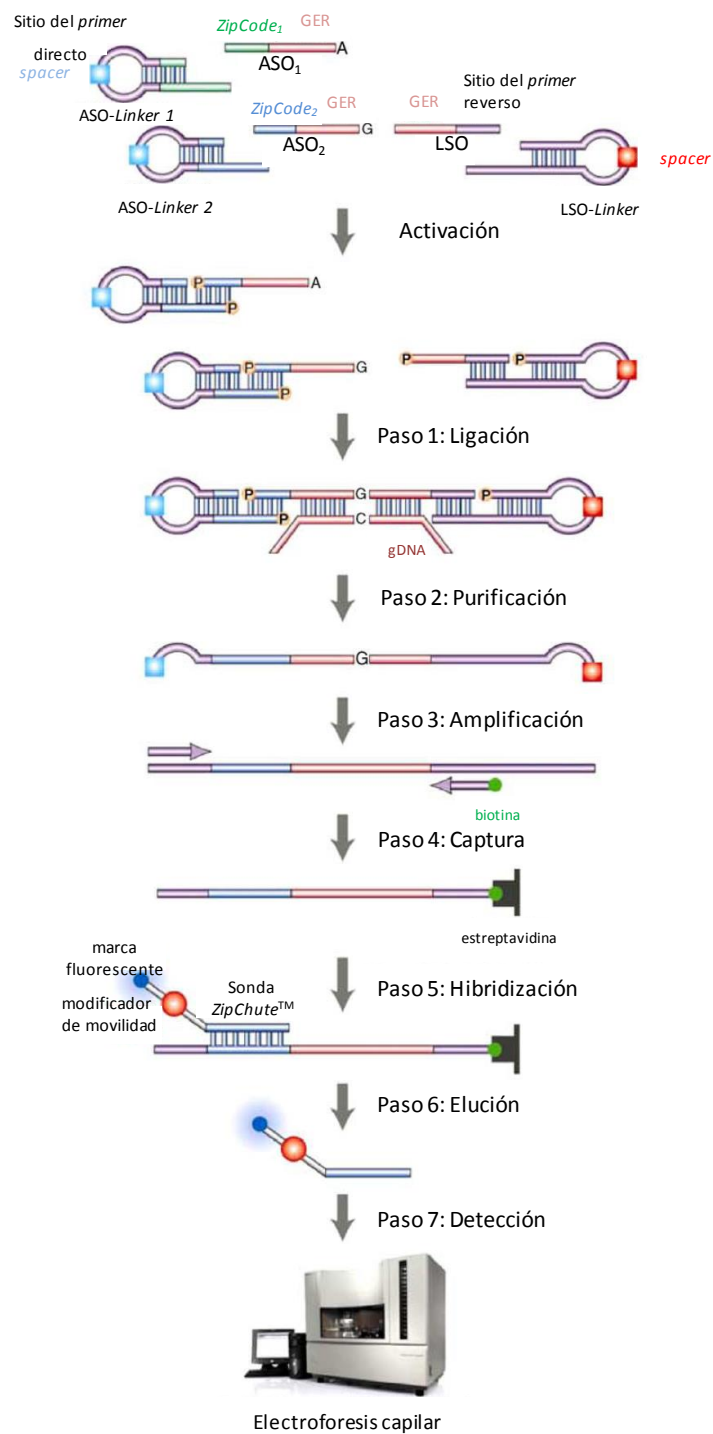


Figura I.1. Sistema de Genotipificación SNPlex®

Hibridación de sondas ZipChute a los productos de PCR capturados.

Una mezcla de sondas universales denominadas ZipChute se agrega a cada pocillo con amplicones capturados con el objetivo de decodificar la información genotípica. Cada sonda ZipChute contiene una secuencia complementaria a las secuencias únicas ZipCode

correspondientes a cada sonda ASO, a su vez contenidas en los productos de PCR unidos a la placa. Las ZipChute asimismo contienen una estructura que determina una movilidad electroforética específica y única para cada una de ellas, y un fluoróforo para su detección.

Cada *kit* SNPlex® provee al usuario una mezcla de 102 sondas ZipChute universales, 96 de las cuales corresponden a los 96 posibles alelos de los 48 SNPs susceptibles de ser analizados en paralelo por el *kit*, y las 6 restantes consisten en controles internos. Para la marcación de las ZipChute se utilizan dos fluoróforos, estando cada par de sondas correspondientes a cada alelo de un mismo SNP marcado con el mismo fluoróforo (los alelos se diferencian, por lo tanto, en base a la movilidad electroforética de las ZipChute que los representen).

Elución de las ZipChute hibridizadas.

Después de la hibridización del paso anterior y de un intenso lavado de las sondas en exceso o unidas inespecíficamente, las sondas ZipChute específicamente unidas a los productos de PCR son eluidas en un *buffer* adecuado.

Electroforesis capilar.

El último paso del proceso consiste en la separación y detección de sondas ZipChute por electroforesis capilar.

Finalmente, la asignación de alelos en base al perfil electroforético obtenido se realiza mediante el programa GeneMapper®. Este programa detecta los picos asociados a cada muestra de ADN y determina su tamaño en base a su movilidad relativa a un estándar que se agrega en la corrida electroforética. Además, analiza los picos que se obtienen a partir de un marcador alélico que contiene todas las sondas ZipChute posibles y se utiliza con el fin de corregir variaciones de tamaño entre diferentes corridas electroforéticas. Luego de la detección y la determinación del tamaño de cada pico, el programa evalúa la intensidad de la señal de los picos correspondientes a cada sonda ZipChute para asignar los genotipos. También permite la visualización de los genotipos obtenidos para un SNP particular en las muestras de ADN analizadas al representar gráficamente los ADNs homocigotas y heterocigotas en ese locus como nubes de puntos de diferentes colores (Figura 1.2) y calcula distintos parámetros con los que evalúa la performance de la técnica y la calidad de la genotipificación.

7.B. Diseño del *kit*.

El sistema SNPlex® impone ciertas condiciones a los marcadores para poder ser incluidos en el diseño del *kit*. Las condiciones que deben reunir las secuencias conteniendo los SNPs se mencionan a continuación:

- no deben ser redundantes en el genoma, es decir, debe evitarse el diseño de sondas en regiones repetitivas o duplicadas ya que esto conduciría a una baja especificidad en la unión de los oligonucleótidos a su secuencia complementaria;
- no deben contener otros SNPs distintos al SNP *target*;
- no deben contener dominios incompatibles con componentes del sistema de genotipificación;

- deben evitarse las interacciones entre sondas y las ligaciones espurias.

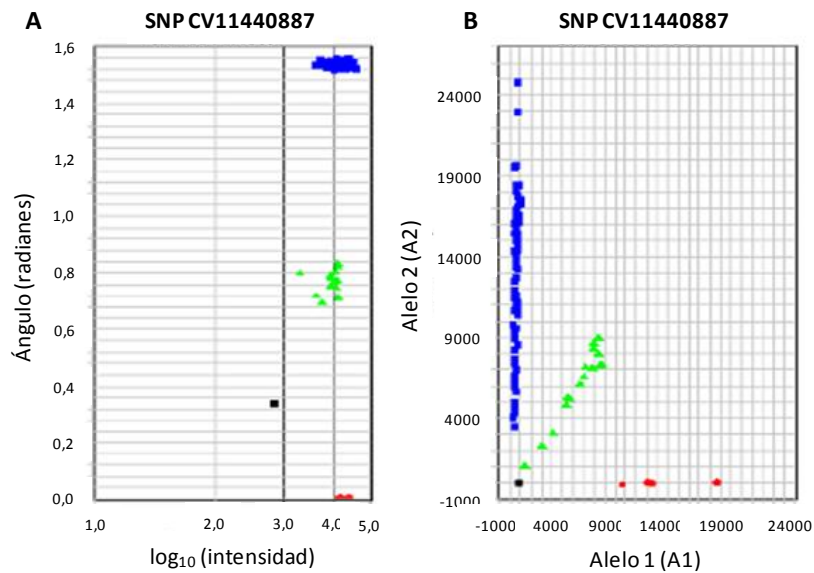


Figura I.2. Visualización de genotipos mediante el programa GeneMapper®.

A. Gráfico polar en el que la intensidad de los picos se grafica en el eje x y la relación entre las alturas de los dos picos en el eje y. **B.** Gráfico cartesiano en el que la intensidad de los dos picos se grafica en los ejes x e y. Las nubes de puntos azules representan las muestras de ADN homocigotas para el alelo A2, las nubes rojas las muestras homocigotas para el alelo A1, y las nubes verdes las muestras heterocigotas.

OBJETIVOS ESPECÍFICOS.

- Aplicar una metodología de selección de genes candidatos relacionados con características productivas y atributos de calidad de leche en bovinos lecheros.
- Identificar marcadores moleculares en los genes candidatos seleccionados.
- Determinar el genotipo de los marcadores moleculares seleccionados en un rodeo comercial de razas Holando y cruza Holando x Jersey y caracterizarlos en función de parámetros poblacionales (frecuencias alélicas, genotípicas y haplotípicas).

MATERIALES.

1. Animales.

Se trabajó con un rodeo comercial de 1.859 bovinos de razas Holando, Jersey y cruza Holando x Jersey pertenecientes a 25 familias de medias hermanas paternas, de 16 tambos de la empresa Las Taperitas S.A., ubicados en la región centro de la provincia de Santa Fe.

La población estudiada consistió en 29 toros (24 de raza Holando y 5 de raza Jersey) y 1.830 vacas Holando y distintos grados de cruce Holando x Jersey.

Todas las vacas poseen registros genealógicos completos y datos fenotípicos para producción (Kg) y composición (Kg y % de grasa y Kg y % de proteína en la leche) de leche (una descripción de los datos fenotípicos puede hallarse en el Capítulo III).

2. Extracción de ADN.

Vacas. Se extrajo ADN a partir de muestras de sangre fresca anticoaguladas con EDTA, mediante *kit* comercial (AxyPrep Blood Genomic DNA Miniprep Kit, Axygen Biosciences; EE.UU.) según el protocolo provisto por el fabricante. Se trabajó con sangre extraída de la vena yugular de los animales sin anestesiar, utilizándose una alícuota de la que se toma en los procedimientos de rutina destinados a realizar controles sanitarios en el establecimiento. Esto se adecúa a las pautas establecidas en el “Manual de cuidado y uso de animales de experimentación” aprobado por el INTA.

Toros. Se extrajo ADN a partir de muestras de semen mediante la técnica de fenol-cloroformo (el protocolo detallado puede consultarse en el Anexo IV). Luego de la extracción y precipitación, el ADN fue resuspendido en *buffer* TE (Tris-EDTA).

Independientemente del método de extracción utilizado, la concentración y la calidad del ADN obtenido se determinaron a través de la lectura de su absorbancia a 260 nm y la relación de absorbancias 260nm/280nm y 260nm/230nm respectivamente, utilizando un espectrofotómetro NanoDrop 1000 Thermo Scientific (Wilmington, Delaware, EE.UU.).

MÉTODOS.

1. Selección de genes candidatos.

La selección de genes candidatos se realizó siguiendo las seis estrategias de búsqueda bibliográfica mencionadas en la introducción de este capítulo (en la sección 5.A.), consultando en las bases de datos de acceso público a través de la página *web* de PubMed (www.ncbi.nlm.nih.gov/pubmed) las publicaciones más relevantes relacionadas con la producción y composición de la leche en el ganado bovino. Asimismo, en la página *web* www.animalgenome.org/QTldb se consultó la cantidad y distribución por cromosoma de los QTLs relacionados con producción y composición de leche reportados hasta el momento en el ganado bovino (Hu y col., 2007; Hu y col., 2013).

2. Selección de SNPs a genotipificar.

La selección de SNPs a incluir en el ensayo de genotipificación se llevó a cabo siguiendo tres abordajes de búsqueda, que se mencionan a continuación.

- Búsqueda en la base de datos dbSNP del NCBI (www.ncbi.nlm.nih.gov/snp), en el Proyecto Genoma Bovino (www.bovinegenome.org) y en bibliografía de los SNPs reportados para cada gen candidato seleccionado;
- Secuenciación de genes candidatos (regiones seleccionadas de modo que contengan los SNPs descritos en literatura) en la población de toros de Las Taperitas S.A. y comparación de las secuencias obtenidas a fin de determinar la presencia de polimorfismos previamente reportados, o nuevos polimorfismos en la población analizada;
- Búsqueda bibliográfica de SNPs asociados al fenotipo en estudio, localizados en regiones intergénicas o en genes predichos *in silico*.

La secuenciación de regiones seleccionadas de genes candidatos en la población de toros en estudio se realizó mediante la utilización de la química BigDye® Terminator Cycle Sequencing Kit de Applied Biosystems (Foster City, California, EE.UU.). La descripción de la técnica así como la lista de los *primers* utilizados en las reacciones de PCR y secuenciación pueden consultarse en el Anexo IV. Los productos de la reacción de secuenciación se resolvieron utilizando un secuenciador de capilares automático ABI 3130xl (Applied Biosystems, California, EE.UU.). Las secuencias obtenidas se procesaron utilizando el programa PreGap4 (Staden Package; Staden y col., 1998) con el fin de eliminar regiones de baja calidad. Posteriormente se ensamblaron mediante el programa Gap4 (Staden Package; Staden y col., 1998), obteniendo finalmente una secuencia consenso corregida para cada combinación de región génica y ADN analizada.

Una vez seleccionados los marcadores a analizar, mediante la utilización de herramientas de visualización de secuencias "*genome browser*", se ubicaron los SNPs en el contexto génico y se buscaron regiones del genoma bovino de aproximadamente 200 pares de bases que comprendían dichos SNPs. En estas regiones se basó el diseño de los oligonucleótidos específicos de alelo que emplea el sistema de genotipificación utilizado.

3. Genotipificación.

La evaluación de los genotipos de toros y crías se realizó mediante la utilización del sistema de genotipificación “SNPlex Genotyping System” de Applied Biosystems (Foster City, California, EE.UU.) (Tobler y col., 2005), que permite determinar simultáneamente el genotipo de hasta 48 SNPs bialélicos por muestra de ADN. Dada la cantidad de SNPs seleccionados para el análisis, fue necesario diseñar dos *sets* o *kits* de oligonucleótidos específicos de alelo a utilizar por el SNPlex Genotyping System®. En el diseño del primer *kit* fueron incluidos 39 SNPs pertenecientes a 22 genes candidatos relacionados con la producción y la composición de la leche en bovinos, mientras que en el segundo se incluyeron 23 marcadores relacionados con el mismo fenotipo.

Como material de partida se utilizaron entre 40 y 100 ng de ADN genómico de cada uno de los animales estudiados (1.856 en el primer *kit* y 581 en el segundo). El ADN fue fragmentado térmicamente de acuerdo a las siguientes condiciones: 1 min a 4°C, seguido de 3 min a 99°C, con el fin de obtener segmentos de entre 2 y 7 Kb para su genotipificación. La secuencia de reacciones que implica la genotipificación mediante el sistema seleccionado se llevó a cabo según protocolo provisto por el fabricante. Las reacciones que implicaron el uso de cicladores térmicos fueron realizadas en cicladores Eppendorf Mastercycler y Mastercycler gradient. El análisis de los resultados obtenidos y la asignación de genotipos se realizó mediante la aplicación del *software* GeneMapper® v4.0.

4. Cálculo de factores descriptivos de los genotipos obtenidos.

Una vez asignados los genotipos, se estimaron las frecuencias genotípicas y alélicas para cada SNP en la población analizada. Dichas estimaciones se realizaron discriminando los animales por raza, de modo de establecer el aporte de cada raza a la variabilidad alélica de la población en su conjunto. Asimismo, se evaluó la calidad de los genotipos obtenidos a través de la determinación de la proporción de SNPs con genotipos asignados (tasa de genotipificación), los conflictos de parentesco y la reproducibilidad en la asignación de genotipos en muestras repetidas del mismo animal. Se evaluó además si la población genotipificada se encontraba o no en equilibrio de Hardy-Weinberg (EHW).

Tasa de genotipificación (*Call rate*).

La tasa de genotipificación es una medida del éxito de la genotipificación y se calculó tanto para los SNPs como para las muestras de ADN analizadas.

La **tasa de genotipificación para un SNP** determinado consiste en el cociente entre la cantidad de genotipos asignados para el SNP sobre el número total de muestras ensayadas. Aquellos SNPs que presenten una tasa de genotipificación baja deben ser descartados puesto que, al sugerir que puede existir una falla en el sistema de genotipificación para ese SNP, los genotipos asignados no constituyen resultados confiables. El valor umbral que determina si un SNP será incluido o no en el análisis posterior depende del sistema de genotipificación utilizado, pudiendo variar entre 0,80 y 0,95 (Jiang y col., 2010; Schopen y col., 2011; Strucken y col., 2011). En nuestro caso, se estableció un valor mínimo de 0,5 para la tasa de

genotipificación de los SNPs para retenerlos en el análisis. Este criterio se adoptó de modo de descartar la mínima cantidad posible de SNPs, teniendo en cuenta que cualquier error en los genotipos asignados por el sistema SNPlex® podrá ser detectado a través de la información provista por la genotipificación con una densidad media de marcadores (explicada con detalle en el Capítulo II de este trabajo de tesis). Aquellos SNPs que no alcanzaron el umbral mencionado previamente fueron descartados y no considerados para ningún cálculo posterior.

La **tasa de genotipificación para una muestra** determinada consiste en el cociente entre el número de genotipos asignados para la muestra sobre todos los SNPs evaluados.

Porcentaje de éxito en las reacciones de genotipificación.

El éxito en las reacciones de genotipificación es una medida de la performance de la química del sistema utilizado, y se define como el cociente entre el total de genotipos asignados y el número de *loci* evaluados, multiplicado por 100, siendo el número de genotipos asignados el número de resultados obtenidos considerando los SNPs y las muestras analizadas que presentaron una tasa de genotipificación $\geq 0,5$ en cada *kit*, y el número de *loci* evaluados aquellos con una tasa de genotipificación $\geq 0,5$ para cada uno de los *kits*.

Reproducibilidad en la asignación de genotipos.

La reproducibilidad en las reacciones de genotipificación se mide como la concordancia entre los genotipos obtenidos al genotipificar en reacciones y corridas diferentes dos alícuotas de la misma muestra de ADN.

Errores mendelianos.

Los errores mendelianos hacen referencia a todos aquellos genotipos asignados que generen inconsistencias genotípicas en el núcleo familiar padre-madre-hijo o bien padre-hijo. La determinación del número y *locus* de ocurrencia de errores mendelianos se realizó mediante el programa PLINK (Purcell y col., 2007).

Equilibrio de Hardy-Weinberg.

La comprobación de la desviación del EHW se llevó a cabo mediante la prueba χ^2 , utilizando las frecuencias genotípicas observadas (obtenidas de los datos) y las frecuencias genotípicas esperadas según el principio de Hardy-Weinberg. En este *test* de hipótesis χ^2 , la hipótesis nula fue que la población tiene proporciones de Hardy-Weinberg y la alternativa que no las tiene.

5. Determinación de fases de haplotipos e imputación de genotipos faltantes.

La determinación de fases haplotípicas fue realizada por medio del programa BEAGLE (Browning y Browning, 2007; Browning y Browning, 2009). En el mismo proceso, fueron imputados los genotipos no asignados y aquellos que presentaban inconsistencias mendelianas. Esto se llevó a cabo al reunir y procesar simultáneamente los genotipos obtenidos a partir del sistema SNPlex® y los obtenidos mediante el *chip* Bovine SNP50 v2 BeadChip (Illumina Inc., San Diego, California).

RESULTADOS.

1. Selección de genes candidatos.

La búsqueda de genes candidatos se realizó siguiendo las seis estrategias mencionadas en la sección 5.A. de la introducción de este capítulo. A continuación se enumeran los genes seleccionados según los distintos abordajes de búsqueda bibliográfica. Un resumen de las razones que justifican su elección en el que se exponen las principales características funcionales de las proteínas codificadas por los genes seleccionados, así como los trabajos científicos consultados, se realiza en el Anexo V.

1.A. Genes codificantes de proteínas relacionadas al fenotipo estudiado que presenten diferentes variantes genéticas.

Las proteínas seleccionadas en esta categoría fueron las caseínas (α_{s1} -, α_{s2} -, β - y κ -caseínas) y la β -lactoglobulina.

1.B. Genes con variantes alélicas que hayan sido reportadas en bibliografía como asociadas al rasgo en estudio.

Numerosos estudios han reportado asociaciones significativas entre polimorfismos en los genes codificantes de la hormona de crecimiento (*GH*, *Growth Hormone*), el receptor de la hormona de crecimiento (*GHR*, *Growth Hormone Receptor*), el receptor 1 de LDL oxidada (*OLR1*, *Oxidized LDL Receptor 1*), la proteína de resistencia al cáncer de mama (*BCRP* o *ABCG2*), la sintasa de ácidos grasos (*FASN*, *Fatty Acid Synthase*), la estearoil-Coenzima A desaturasa 1 (*SCD1*, *Stearoyl-Coenzyme A Desaturase 1*), la α -lactoalbúmina (α -*LA*) y la leptina (*LEP*) bovinas, con características relacionadas a la producción y composición de la leche.

1.C. Genes con perfiles de expresión asociados a la variación fenotípica del rasgo analizado.

Dentro de esta categoría se seleccionaron genes que se expresan en determinados momentos de la preñez o la lactancia y codifican proteínas que contribuyen al desarrollo de estructuras fisiológicas características de estos procesos o al mantenimiento de los mismos, como son las proteínas de leche uterina (*UTMP*, *Uterine Milk Proteins*) y la prolactina (*PRL*).

1.D. Genes localizados en QTLs asociados al fenotipo en estudio.

Desde el primer barrido del genoma bovino completo para la búsqueda de QTLs en bovinos lecheros realizado en 1995 (Georges y col., 1995), se han hecho gran cantidad de estudios similares que condujeron a la detección de QTLs relacionados a características productivas y sanitarias, entre otras. En bovinos lecheros se han determinado numerosas localizaciones para QTLs que afectan caracteres productivos como por ejemplo el rendimiento de leche, producción y porcentaje de grasa y producción y porcentaje de proteína en la leche (Khatkar y col., 2004).

En la *web* existen bases de datos de acceso público que pueden consultarse para averiguar la localización de los QTLs relacionados a algún rasgo de interés en un determinado organismo (www.animalgenome.org/QTLdb/). Actualmente (junio de 2015) se han reportado

7.579 QTLs asociados a leche en bovinos; de ellos 1.998 corresponden a la composición proteica de la leche, 4.470 a la composición grasa de la leche, 1.043 al rendimiento o producción de leche, 50 a otras características composicionales y 18 a características del procesamiento de la leche. El objetivo último del mapeo de QTLs consiste en identificar los genes y las variantes alélicas responsables de la variación de un fenotipo en particular.

Rincón y col. (2009) desarrollaron un análisis de asociación con microsatélites y uno de ligamiento con marcadores SNP sobre muestras independientes con el objetivo de aumentar la resolución del mapeo de un QTL previamente reportado en una región del BTA 4. Los principales efectos sobre caracteres productivos fueron observados en cinco SNPs localizados en la región del QTL analizado. Dichos SNPs se ubican dentro o en la proximidad de los siguientes genes candidatos: *ARL4A* (codificante de la proteína *ADP-ribosylation factor-like*), *XR_027435.1* (gen similar al codificante de la proteína ribosomal S4), *ETV1* (*ets variant 1*), *SNX13* (*Sorting nexin 13*), *MEOX2* (*Mesenchyme homeobox 2*).

Si bien en el BTA 4 se han reportado gran cantidad de QTLs relacionados con la producción y composición de la leche, el BTA 6 es el que alberga el mayor número de QTLs asociados a estos caracteres. Varios grupos de investigación han propuesto genes candidatos para explicar los efectos de los QTLs hallados en el cromosoma 6 bovino en base a la relación entre su rol fisiológico y el fenotipo en estudio. Algunos de estos genes candidatos son el coactivador 1- α del receptor γ activado por proliferadores peroxisómicos (*PPARGC1A*, *Peroxisome Proliferator Activated Receptor Gamma*, *Coactivator 1 Alpha*), el codificante de la osteopontina (*OPN*, también denominado *SPP1*) y el que codifica la poliquistina 2 (*PKD2*).

Por otro lado, en el año 2002 Grisart y col. identificaron un candidato posicional en un intervalo de 3 cM en el extremo telomérico del BTA 14, donde había sido reportado previamente un QTL con gran efecto sobre la producción y composición de la leche bovina, en particular sobre el contenido de grasa en la leche. El gen candidato hallado fue el *DGAT1* (*AcylCoA:diacylglycerol acyltransferase 1*).

En otros estudios, Blott y col. (2003) y Viitala y col. (2006), en el intento de explicar el efecto del QTL del BTA 20 relacionado a la composición y el rendimiento de leche, hallaron dos fuertes candidatos con polimorfismos significativamente asociados al fenotipo mencionado: el receptor de la hormona de crecimiento (*GHR*) y el receptor de prolactina (*PRLR*).

1.E. Genes regulados post-transcripcionalmente por miRNAs expresados en el/los órganos relacionados al rasgo estudiado.

Se ha reportado que al transfectar un miRNA perfectamente complementario a una secuencia conservada del gen de la lactoferrina humano y de otros organismos en una línea celular del epitelio mamario de ratón se obtiene un bloqueo de la estimulación del gen y la represión de su expresión proteica. Por este motivo dentro de esta categoría se seleccionó al gen de la lactoferrina (*LTF*) como gen candidato.

1.F. Genes cuya expresión alterada en animales *knock out* o transgénicos haya resultado en fenotipos específicos asociados al rasgo en estudio.

Liu y col. (1997), a través de la delección del gen *STAT5A* demostraron que la proteína codificada por el mismo está implicada en el desarrollo de la glándula mamaria adulta y en la lactogénesis. Por tal motivo el gen transductor de señales y activador de la transcripción 5A (*STAT5A*, *Signal Transducer and Activator of Transcription 5A*) es un gen candidato a ser estudiado en este trabajo de tesis.

La Tabla I.3 resume los genes candidatos seleccionados a partir de cada una de las estrategias de búsqueda seguidas.

Estrategia de búsqueda	Genes candidatos
A	<i>CSN1S1</i> , <i>CSN1S2</i> , <i>CSN2</i> , <i>CSN3</i> y β -LGB
B	<i>GH</i> , <i>GHR</i> , <i>OLR1</i> , <i>ABCG2</i> , <i>FASN</i> , <i>SCD1</i> , α -lactoalbúmina y <i>LEP</i>
C	<i>UTMP</i> y <i>PRL</i>
D	<i>ARL4A</i> , GenBank: XR_027435.1 (gen predicho), <i>ETV1</i> , <i>SNX13</i> , <i>MEOX2</i> , <i>PPARGC1A</i> , <i>OPN</i> , <i>PKD2</i> , <i>DGAT1</i> y <i>PRLR</i>
E	<i>LTF</i>
F	<i>STAT5A</i>

Tabla I.3. Genes candidatos seleccionados a partir de distintas estrategias de búsqueda.

A- Genes codificantes de proteínas relacionadas al fenotipo estudiado que presenten diferentes variantes genéticas. B- Genes con variantes alélicas que hayan sido reportadas en bibliografía como asociadas al rasgo en estudio. C- Genes con perfiles de expresión asociados a la variación fenotípica del rasgo analizado. D- Genes localizados en QTLs asociados al fenotipo en estudio. E- Genes regulados post-transcripcionalmente por miRNAs expresados en el/los órganos relacionados al rasgo estudiado. F- Genes cuya expresión alterada en animales *knock out* o transgénicos haya resultado en fenotipos específicos asociados al rasgo en estudio.

Abreviaturas: *CSN1S1*: α_{s1} -caseína, *CSN1S2*: α_{s2} -caseína, *CSN2*: β -caseína, *CSN3*: κ -caseína, β -LGB: β -lactoglobulina, *GH*: Hormona de crecimiento, *GHR*: Receptor de hormona de crecimiento, *OLR1*: Receptor 1 de LDL oxidada, *ABCG2*: miembro 2 de la subfamilia G de la familia de transportadores ATP-binding cassette, *FASN*: Sintasa de ácidos grasos, *SCD1*: Estearoil-Coenzima A desaturasa 1, *LEP*: Leptina, *UTMP*: Proteínas de leche uterina, *PRL*: Prolactina, *ARL4A*: ADP-ribosylation factor-like 4A, *ETV1*: *ets variant 1*, *SNX13*: *Sorting nexin 13*, *MEOX2*: *Mesenchyme homeobox 2*, *PPARGC1A*: Coactivador 1- α del receptor γ activado por proliferadores peroxisómicos, *OPN*: Osteopontina, *PKD2*: Poliquistina 2, *DGAT1*: *AcylCoA:diacylglycerol acyltransferase 1*, *PRLR*: Receptor de prolactina, *LTF*: Lactoferrina, *STAT5A*: Transductor de señales y activador de la transcripción 5A.

Una vez definidos los genes candidatos a estudiar, se procedió a seleccionar los marcadores SNP a genotipificar dentro de los mismos.

2. Selección de SNPs a genotipificar.

La selección de SNPs a incluir en el ensayo de genotipificación se llevó a cabo siguiendo los tres abordajes de búsqueda que se mencionan en la sección de Métodos de este capítulo. A continuación se detallan los SNPs hallados según cada una de estas estrategias.

2.A. Búsqueda en bases de datos de acceso público (www.ncbi.nlm.nih.gov/snp, www.bovinegenome.org) y en bibliografía de los SNPs reportados para cada gen candidato seleccionado.

Los SNPs seleccionados para cada uno de los genes candidatos se presentan en la Tabla I.4, en la que figura su código *rs* en la base de datos del NCBI, su ubicación en el genoma bovino según el ensamblado UMD 3.1 y, en los casos en que hubiera, la asociación con caracteres relacionados a producción y composición de la leche bovina reportada en bibliografía.

Los SNPs rs135164815 y rs136247583 se ubican uno a continuación del otro en el genoma bovino, por lo que se analizarán en forma conjunta como un polimorfismo de nucleótido simple consistente en una variación de dos nucleótidos (GT/AC). A este SNP se lo denominará rs135164815-136247583.

2.B. Secuenciación de genes candidatos en la población de toros y detección de nuevos polimorfismos en la población analizada.

Las secuencias obtenidas por secuenciación de regiones seleccionadas dentro de seis genes candidatos en los toros de la población analizada permitieron la asignación de alelos en forma correcta. A continuación se presenta, a modo de ejemplo, una porción de 64 pb de la secuencia obtenida para el gen candidato *PRL* en el toro 112 (Figura I.3).

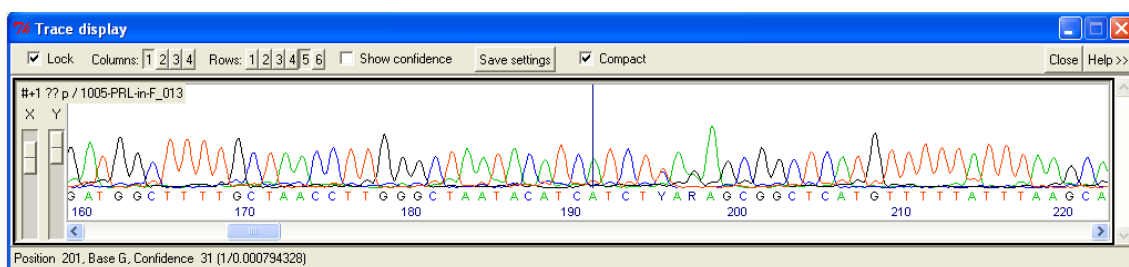


Figura I.3. Visualización de secuencias en el programa Trev de Staden Package.

Picos obtenidos para una porción de 64 pb de la secuencia obtenida para el gen candidato *PRL* en el toro de RP 112.

Gen candidato	SNP	Variación nucleotídica	Ubicación (BTA: pb)	Asociaciones reportadas
<i>ARL4A</i>	rs43375517	C/G, 5' UTR	4:20296999	Estos SNPs explicarían los efectos del QTL del BTA4 asociado a caracteres productivos (Rincón y col., 2009).
<i>XR_027435.1</i>	rs43376185	C/T	4:20708087	
<i>ETV1</i>	rs42213673	C/T, intrón	4:22111819	
<i>SNX13</i>	rs41595314	A/T	4:26425057	
<i>LEP</i>	rs29004488 p.Arg25Cys	C/T, exón	4:93262056	Asociado al rendimiento de leche y de proteína en la leche (Buchanan y col., 2003; Giblin y col., 2010).
<i>α-lactoalbúmina</i>	g.1966G>A	G/A, 5'UTR	5:31347875	Asociado a características productivas en bovinos lecheros (Bleck y Bremel, 1993).
<i>OLR1</i>	g.8232C>A p.Cys223Ala	C/A, 3'UTR	5:100254823	
<i>ABCG2</i>	rs43702337 p.Tyr581Ser	A/C, exón	6:38027010	Asociado al rendimiento de leche y porcentajes de grasa y proteína en la leche (Cohen-Zinder y col., 2005; Olsen y col., 2007).
<i>OPN</i>	OPN3907	T9-T10 Indel, 5'UTR	6:38128806	Asociado al rendimiento de leche y porcentajes de grasa y proteína en la leche (Schnabel y col., 2005; Olsen y col., 2007).
	g.8514C>T	C/T, intrón	6:38122665	Asociado a porcentajes de grasa y proteína en la leche (Leonard y col., 2005).
<i>PPARGC1A</i>	rs109579682 c.1892+19T>C	A/G, intrón	6:44875251	Asociado al contenido de grasa de la leche (Weikard y col., 2005).
<i>CSN1S1</i>	CSN1_AB	T/A, exón	6:87146017	CSN1_AB es el SNP responsable de la generación de la variante A, por delección en el ARNm de los nucleótidos que codifican los residuos 14-26. Sustitución en +6 del exón 4.
	p.Ala53ThrP	G/A, exón	6:87148464	SNP CSN1_BD determinante de la variante D.
	p.Gln59Lys	C/A, exón	6:87149981	SNP CSN1_BE determinante de la variante E.
	p.SerP66Leu	C/T, exón	6:87150003	SNP CSN1_BF determinante de la variante F.

Tabla I.4. SNPs en genes candidatos seleccionados reportados en bibliografía.

La denominación del SNP figura con el código rs, la mutación a nivel de ADN genómico, la mutación a nivel de proteína o una denominación arbitraria.

Gen candidato	SNP	Variación nucleotídica	Ubicación (BTA: pb)	Asociaciones reportadas
<i>CSN1S1</i>	rs43703010 p.Glu192Gly	A/G, exón	6:87157262	El SNP <i>CSN1_BCF</i> es responsable del cambio Glu192Gly, presente en las variantes C y E (solo este SNP determina la variante C, si se agrega Gln59Lys, entonces es E).
<i>CSN2</i>	rs43703011 p.Pro67His	C/A, exón	6:87181619	SNP responsable de la generación de las variantes A2 y A1.
	rs43703013 p.Ser122Arg	G/C, exón	6:87181453	El cambio Ser122Arg, es el responsable de la diferenciación de alelos A y B.
<i>CSN3</i>	rs43706475 p.Ser104Ala	T/G, exón	6:87390479	SNP <i>CSN3_AI</i> determinante de la variante I.
	rs43703015 p.Thr136Ile	C/T, exón	6:87390576	Este SNP <i>CSN3_AB</i> es uno de los SNP determinantes de la variante B.
	rs43703017 p.Ser155Gly	A/G, exón	6:87390632	SNP <i>CSN3_AE</i> determinante de la variante E.
<i>LGB</i>	rs41255683	G/C, promotor	11:103301455	
	rs41255685	C/T, 5'UTR	11:103301690	
	rs43691049	C/T, intrón	11:103303343	
	rs110066229 p.Asp64Gly	A/G, exón	11:103303475	Este SNP permite la diferenciación de variantes A y B.
	rs109625649 p.Ala118Val	C/T, exón	11:103304757	Este SNP permite la diferenciación de variantes A y B.
<i>DGAT1</i>	p.Lys232Ala	ApA/GpC, exón	14:1802265- 14:1802266	Asociado al contenido de grasa en la leche (Grisart y col., 2002).
<i>STAT5A</i>	g.9501G>A	G/A, intrón	19:43047829	Asociado a la producción de leche y contenido de grasa y proteína en la leche (Brym y col., 2004).

Tabla I.4. SNPs en genes candidatos seleccionados reportados en bibliografía (continuación).

Gen candidato	SNP	Variación nucleotídica	Ubicación (BTA: pb)	Asociaciones reportadas
<i>GH</i>	rs41923484 p.Leu153Val	G/C, exón	19:48768916	
<i>FASN</i>	g.16024G>A p.Thr1953Ala	G/A, exón	19:51400139	Asociado al contenido de grasa en la leche (Schennink y col., 2009).
	g.17924G>A	G/A, exón	19:51402032	Asociado al porcentaje de grasa en la leche (Schennink y col., 2009).
<i>GHR</i>	GHR_140 p.Phe279Tyr	T/A, exón	20:31909478	Asociado al rendimiento y composición de la leche (Blott y col., 2003) y porcentaje de grasa y de proteína en la leche (Viitala y col., 2006).
	rs42551798	T/G, intrón	20:39090309	
<i>PRLR</i>	rs42551796	A/T, intrón	20:39090450	
	rs135164815 p.Ser18Asn	G/A, exón	20:39115344	Asociados a la producción de leche y de grasa y proteína en la leche (Viitala y col., 2006).
	rs136247583	T/C, exón	20:39115345	
<i>UTMP</i>	g.1296G>A	G/A, exón	21:59667572	Asociado con la vida productiva (Khatib y col., 2005).
<i>LTF</i>	rs41256920	C/A, 5'UTR	22:53521978	
	rs43706485	C/G, 5'UTR	22:53522038	
<i>PRL</i>	PRL_in1_188	C/T, intrón	23:35110974	
	g.8398G>A	G/A	23:35106206	Posee un efecto sobre la composición de la grasa de la leche (Schennink y col., 2009).
<i>SCD1</i>	rs41255693 p.Ala293Val	C/T, exón	26:21144708	Asociado a la composición de ácidos grasos en la leche (Conte y col., 2010; Mele y col., 2007, Moiola y col., 2007).

Tabla I.4. SNPs en genes candidatos seleccionados reportados en bibliografía (continuación).

La secuenciación realizada permitió detectar nueve polimorfismos nuevos, no incluidos en las bases de datos de SNPs ni reportados en la bibliografía consultada. Los SNPs detectados, el gen al que pertenecen y la variación nucleotídica implicada figuran en la Tabla I.5.

Gen candidato	SNP	Variación nucleotídica	Ubicación (BTA: pb)
<i>ABCG2</i>	ABCG2_190	G/A	6:38027001
<i>OPN</i>	SPP1_11740	A/C	6:38120968
<i>PPARGC1A</i>	PPARGC1A_230	G/A, exón	6:44875315
	PPARGC1A_337	C/T, intrón	6:44875421
<i>LGB</i>	LGBprom_37	G/A, promotor	11:103301242
	LGB_in3ex4_134	C/G	11:103304617
<i>GH</i>	GH_51	C/T	19:48769040
	GH_335	A/C	19:48768766
<i>PRL</i>	PRL_in1_39	C/T, intrón	23:35111122

Tabla I.5. Nuevos SNPs detectados por secuenciación.

2.C. Búsqueda bibliográfica de SNPs asociados al fenotipo en estudio, localizados en regiones intergénicas o en genes predichos *in silico*.

La tercera estrategia de selección de SNPs a incluir en el ensayo de genotipificación se basó en la búsqueda bibliográfica de SNPs asociados a la producción y composición de la leche bovina, sin pertenecer necesariamente a genes directa, indirecta o potencialmente relacionados a alguna vía metabólica implicada en la generación del rasgo en estudio. Para llevar a cabo esta búsqueda se consultaron publicaciones en las que se realizó una genotipificación de alta densidad sobre un elevado número de bovinos lecheros, con posteriores análisis de asociación a caracteres de producción (Kolbehdari y col., 2009; Jiang y col., 2010). En el Anexo V (punto 2) se exponen todos los SNPs que formaron parte de la preselección efectuada, mientras que a continuación se presentan los SNPs que resultaron seleccionados para ser considerados en el diseño de los *kits* de genotipificación, todos ellos con asociaciones significativas reportadas con cada uno de los caracteres testeados. Se indican, asimismo, el cromosoma y el gen (entre paréntesis) al que pertenece cada SNP.

- Producción de leche.

BTA 1: rs41631818 (*MGC128242*)

BTA 5: rs41655901 (*GALNT6*)

BTA 11: rs41569023 (*ROCK2*)

BTA 14: rs41256919 (*MAF1*), rs41628862 (*BIG1*), rs41587081 (*ZFHX4*)

BTA 21: rs41644615 (*SERPINA5*)

- Contenido de proteína en la leche.

BTA 1: rs41586699 (*PDE9A*), rs41631818 (*MGC128242*)

BTA 11: rs41569023 (*ROCK2*)

BTA 14: rs41628862 (*BIG1*)

BTA 21: rs41644615 (*SERPINA5*)

- Contenido de grasa en la leche.

BTA 1: rs41637121 (*APP*), rs41631818 (*MGC128242*),

BTA 5: rs41652648 (*ITPR2*), rs41592942 (*GUCY2C*)

BTA 14: rs41579049 (*5-OPase*)

BTA 24: rs41645253 (*MGC139244*)

- Porcentaje de proteína en la leche.

BTA 5: rs29014633 (*CACNG2*)

BTA 6: rs29018333 (*GRID2*), rs41622323 (*PKD2*)

BTA 20: rs29018751 (*NIPBL*)

- Porcentaje de grasa en la leche.

BTA 6: rs29018853 (*LEC3*)

BTA 14: rs41256919 (*MAF1*), rs41624797 (*PTK2*), rs55617160 (*NIBP*), rs41579063 (*BIG1*), rs41587081 (*ZFHX4*)

3. Diseño de los oligonucleótidos utilizados por el sistema de genotipificación SNPlex®.

Una vez seleccionados los SNPs relacionados directa, indirecta o potencialmente con caracteres productivos y composicionales de la leche bovina, y por lo tanto interesantes para ser incluidos en el diseño del *kit* del sistema de genotipificación a utilizar, los mismos fueron ubicados manualmente en el contexto génico mediante el programa Artemis (Rutherford y col., 2000) con el fin de seleccionar una región flanqueante del SNP de 200 pb. Estas regiones se enviaron al fabricante del *kit* y en ellas se basó el diseño de los oligonucleótidos específicos de alelo que emplea el sistema de genotipificación utilizado.

A modo de ejemplo, en la Figura I.4 se visualiza, mediante el programa Artemis, una región de alrededor de 140 pb que contiene el SNP rs41255693 (g.10329C>T) del gen *SCD1* que fue enviada al fabricante del *kit*.

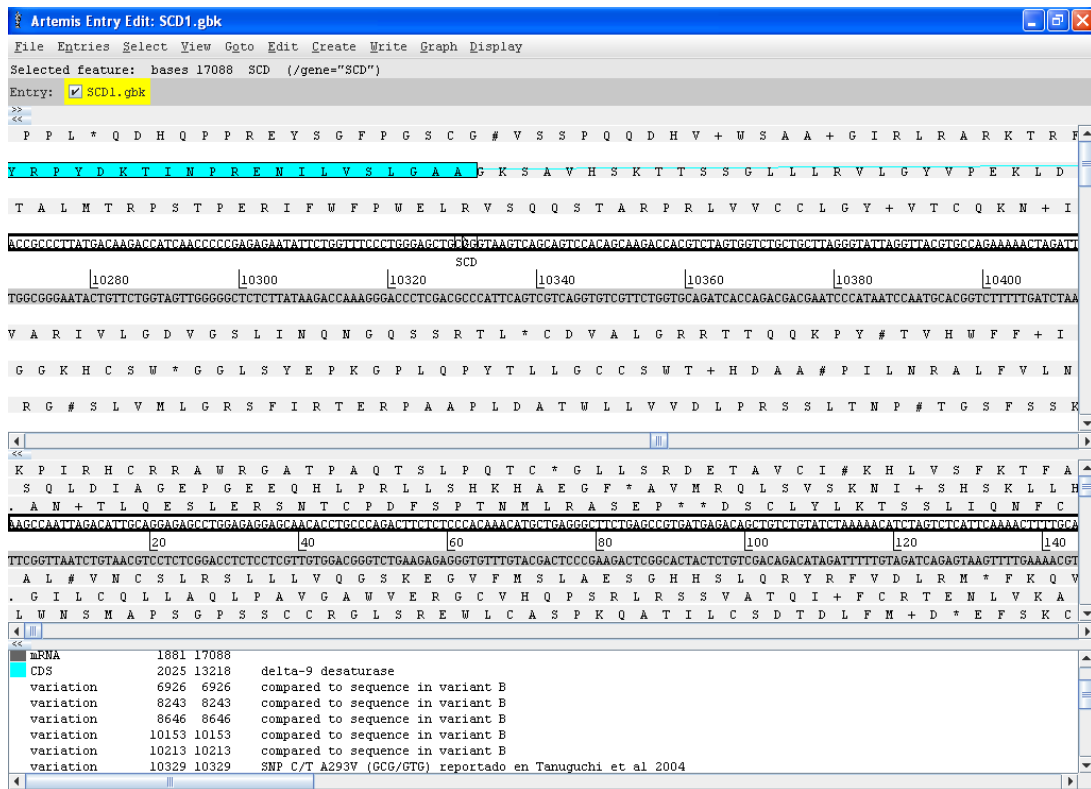


Figura I.4. Visualización de secuencias en el programa Artemis.
Región flanqueante al SNP rs41255693 (g.10329C>T) del gen *SCD1*.

Los SNPs que reunieron las condiciones que impone el sistema SNPlex® y resultaron compatibles con el ensayo permitieron el diseño de dos kits. En las Tablas I.6 y I.7 se muestran los SNPs incluidos en cada uno de ellos.

SNP	Gen	Ubicación (BTA: pb)
rs43375517	<i>ARL4A</i>	4:20296999
rs42213673	<i>ETV1</i>	4:22111819
rs41595314	<i>SNX13</i>	4:26425057
rs29004488	<i>LEP</i>	4:93262056
OLR1_g.8232C>A	<i>OLR1</i>	5:100254823
rs43702337	<i>ABCG2</i>	6:38027010
SPP1_11740	<i>OPN</i>	6:38120968
OPN_g.8514C>T	<i>OPN</i>	6:38122665
OPN3907	<i>OPN</i>	6:38128806
rs109579682	<i>PPARGC1A</i>	6:44875251
PPARGC1A_230	<i>PPARGC1A</i>	6:44875315
PPARGC1A_337	<i>PPARGC1A</i>	6:44875421
CSN1_AB	<i>CSN1S1</i>	6:87146017
CSN1_p.Ala53ThrP	<i>CSN1S1</i>	6:87148464
CSN1_p.SerP66Leu	<i>CSN1S1</i>	6:87150003
rs43703010	<i>CSN1S1</i>	6:87157262

Tabla I.6. SNPs incluidos en el primer kit SNPlex®.

SNPs incluidos en el diseño del primer kit SNPlex®, gen al que pertenecen y ubicación en el genoma bovino.

SNP	Gen	Ubicación (BTA: pb)
rs43703013	<i>CSN2</i>	6:87181453
rs43703011	<i>CSN2</i>	6:87181619
rs43706475	<i>CSN3</i>	6:87390479
rs43703015	<i>CSN3</i>	6:87390576
rs43703017	<i>CSN3</i>	6:87390632
LGBprom_37	<i>LGB</i>	11:103301242
rs41255685	<i>LGB</i>	11:103301690
rs43691049	<i>LGB</i>	11:103303343
rs109625649	<i>LGB</i>	11:103304757
DGAT1_p.Lys232Ala	<i>DGAT1</i>	14:1802265 - 14:1802266
STAT5_g.9501G>A	<i>STAT5</i>	19:43047829
GH_335	<i>GH</i>	19:48768766
GH_51	<i>GH</i>	19:48769040
FASN_g.16024G>A	<i>FASN</i>	19:51400139
FASN_g.17924G>A	<i>FASN</i>	19:51402032
GHR_140	<i>GHR</i>	20:31909478
rs135164815	<i>PRLR</i>	20:39115344
UTMP_g.1296G>A	<i>UTMP</i>	21:59667572
rs41256920	<i>LTF</i>	22:53521978
rs43706485	<i>LTF</i>	22:53522038
PRL_g.8398G>A	<i>PRL</i>	23:35106206
PRL_in1_188	<i>PRL</i>	23:35110974
rs41255693	<i>SCD1</i>	26:21144708

Tabla I.6. SNPs incluidos en el primer *kit* SNPlex® (continuación).

SNP	Gen	Ubicación (BTA: pb)
rs41637121	<i>APP</i>	1:9852144
rs41631818	<i>MGC128242</i>	1:57192123
rs41586699	<i>PDE9A</i>	1:144598855
rs41655901	<i>GALNT6</i>	5:28604763
rs29014633	<i>CACNG2</i>	5:75370225
rs41652648	<i>ITPR2</i>	5:83830316
rs41592942	<i>GUCY2C</i>	5:95666368
rs29018333	<i>LOC536367</i>	6:33396376
rs41622323	<i>PKD2</i>	6:38063313
rs29018853	<i>LEC3</i>	6:79203343
rs41569023	<i>ROCK2</i>	11:86480524
rs41256919	<i>MAF1</i>	14:1923292
rs41579049	<i>5-OPase</i>	14:1958592
rs41624797	<i>PTK2</i>	14:3956956
rs55617160	<i>NIBP</i>	14:4468478
rs41579063	<i>BIG1</i>	14:33364431
rs41628862	<i>BIG1</i>	14:33408326
rs41587081	<i>ZFH4</i>	14:42175361
rs42575474	<i>ZNF215</i>	15:46392886
GH_299G-C	<i>GH</i>	19:48768916
rs29018751	<i>NIPBL</i>	20:37297072
rs41644615	<i>SERPINA5</i>	21:59787612
rs41645253	<i>MGC139244</i>	24:22099387

Tabla I.7. SNPs incluidos en el segundo *kit* SNPlex®.

SNPs incluidos en el diseño del segundo *kit* SNPlex®, gen al que pertenecen y ubicación en el genoma bovino.

Un esquema de la cobertura cromosómica alcanzada con los SNPs seleccionados para ambos *kits* se muestra en la Figura I.5.



Figura I.5. Cobertura cromosómica alcanzada con los SNPs analizados.

En **A** se expone la distribución cromosómica (en escala aproximada) de los SNPs del primer (azul) y segundo (rojo) *kits* SNPlex® en los cromosomas que albergan genes candidatos, y en **B** (ver hoja siguiente) los nombres y genes a los que pertenecen cada uno de los SNPs analizados según su ubicación relativa.

4. Descripción de la población genotipificada.

El ensayo de genotipificación fue realizado sobre una población de 1.859 bovinos de razas Holando, Jersey y cruza Holando x Jersey pertenecientes a 25 familias de medias hermanas paternas, de 16 tambos de la cuenca lechera central de la provincia de Santa Fe.

La población estudiada consistió en 29 toros (24 de raza Holando y 5 de raza Jersey, de los cuales 21 y 4 respectivamente contaban con al menos una hija genotipificada) y 1.830 vacas. El promedio de hijas por toro fue de 73, presentando un rango de variación de 1 a 193. A modo descriptivo de la población de vacas, a continuación se grafica el número de vacas analizadas que son hijas de cada toro (Figura I.6), el porcentaje en que están representados en la población la raza Holando y los distintos grados de cruce con Jersey (Figura I.7), la distribución

de vacas según el año de nacimiento (Figura I. 8) y su distribución en base al tambo de pertenencia, denominado “tambo de extracción” (Figura I.9).

B

Ubicación relativa	SNP	Gen	Ubicación relativa	SNP	Gen		
BTA 1	1	rs41637121	APP	1	rs41569023	ROCK2	
	2	rs41631818	MGC128242	2	LGBprom_37	LGB	
	3	rs41586699	PDE9A	BTA 11	3	rs41255685	LGB
BTA 4	1	rs43375517	ARL4A	4	rs43691049	LGB	
	2	rs42213673	ETV1	5	rs109625649	LGB	
	3	rs41595314	SNX13	1	DGAT1_p.Lys232Ala	DGAT1	
	4	rs29004488	LEP	2	rs41256919	MAF1	
BTA 5	1	rs41655901	GALNT6	3	rs41579049	5-OPase	
	2	rs29014633	CACNG2	4	rs41624797	PTK2	
	3	rs41652648	ITPR2	5	rs55617160	NIBP	
	4	rs41592942	GUCY2C	6	rs41579063	BIG1	
	5	OLR1_g.8232C>A	OLR1	7	rs41628862	BIG1	
BTA 6	1	rs29018333	LOC536367	8	rs41587081	ZFHX4	
	2	rs43702337	ABCG2	BTA 15	1	rs42575474	ZNF215
	3	rs41622323	PKD2	1	STAT5_g.9501G>A	STAT5	
	4	SPP1_11740	OPN	2	GH_335	GH	
	5	OPN_g.8514C>T	OPN	3	GH_299G-C	GH	
	6	OPN3907	OPN	4	GH_51	GH	
	7	rs109579682	PPARGC1A	5	FASN_g.16024G>A	FASN	
	8	PPARGC1A_230	PPARGC1A	6	FASN_g.17924G>A	FASN	
	9	PPARGC1A_337	PPARGC1A	1	GHR_140	GHR	
BTA 6	10	rs29018853	LEC3	BTA 20	2	rs29018751	NIPBL
	11	CSN1_AB	CSN1S1	3	rs135164815	PRLR	
	12	CSN1_p.Ala53ThrP	CSN1S1	1	UTMP_g.1296G>A	UTMP	
	13	CSN1_p.SerP66Leu	CSN1S1	2	rs41644615	SERPINA5	
	14	rs43703010	CSN1S1	BTA 22	1	rs41256920	LTF
	15	rs43703013	CSN2	2	rs43706485	LTF	
	16	rs43703011	CSN2	1	PRL_g.8398G>A	PRL	
	17	rs43706475	CSN3	2	PRL_in1_188	PRL	
	18	rs43703015	CSN3	BTA 24	1	rs41645253	MGC139244
	19	rs43703017	CSN3	BTA 26	1	rs41255693	SCD1

Figura I.5. Cobertura cromosómica alcanzada con los SNPs analizados (continuación).

En muchas oportunidades el tambo de extracción no coincide con el tambo de nacimiento ya que los animales son trasladados al ser destetados. En la población con la que se trabajó, sólo el 3,8% de las vacas permanecieron en su tambo de nacimiento luego del destete. Sin embargo, durante su vida fértil y su período de lactancia se ubican en el tambo de extracción

correspondiente, llamado de esta forma por ser el tambo en el cual se produjo la extracción de sangre para la genotipificación.

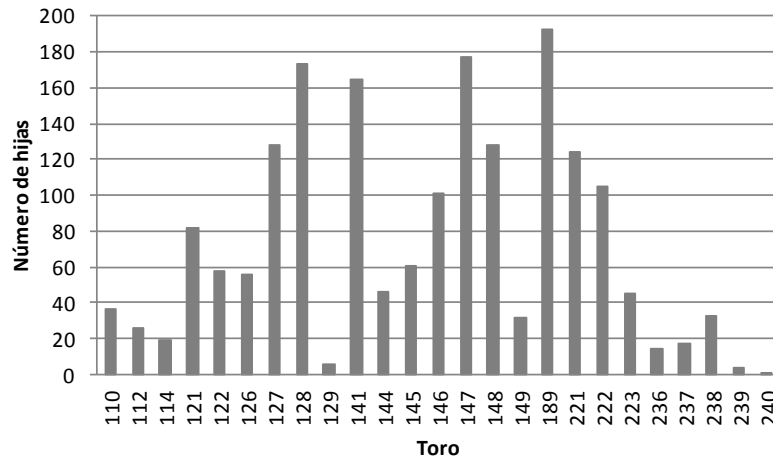


Figura I.6. Distribución de hijas entre los toros analizados.

La designación de los toros corresponde al RP propio de cada uno de ellos. En el gráfico se indica el número de vacas que son hijas de cada toro.

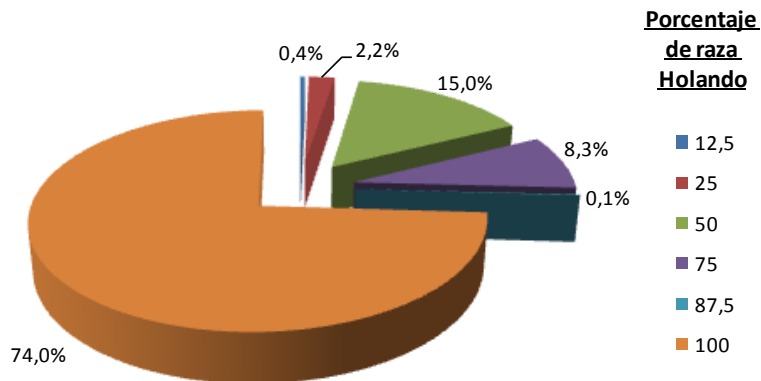


Figura I.7. Distribución de razas en las vacas analizadas.

La distribución de razas se expresa como porcentaje de raza Holando.

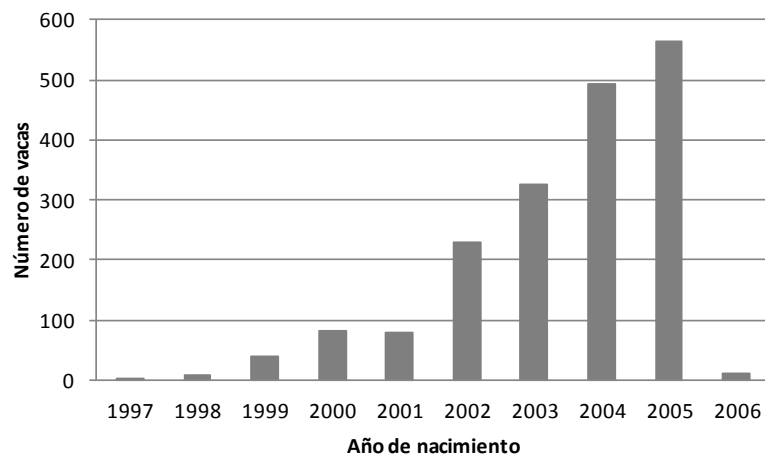


Figura I.8. Distribución de años de nacimiento de las vacas analizadas.

Número de vacas de la población analizada que nacieron entre los años 1997 y 2006.

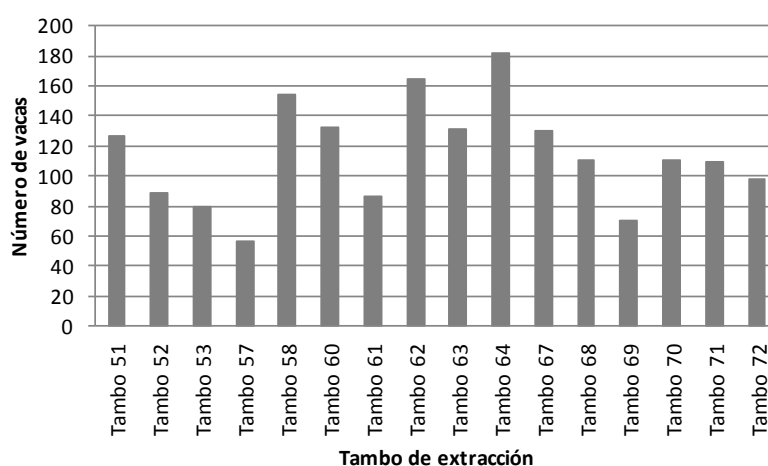


Figura I.9. Distribución de vacas según el tambo de extracción.

Número de vacas analizadas que pertenecen a cada uno de los 16 tambos evaluados.

5. Genotipificación.

5.A. Calidad de la genotipificación.

En primer lugar, antes de realizar el análisis de los genotipos obtenidos para las vacas de raza Holando y cruza, se calcularon una serie de parámetros que brindan información acerca de la calidad de las reacciones de genotipificación propias del sistema comercial utilizado. Los parámetros mencionados se exponen a continuación.

Tasas de genotipificación (*Call rate*).

- **de SNPs:** En las Tablas I.8 y I.9 se expone la tasa de genotipificación obtenida para cada uno de los SNPs analizados en el primer y segundo *kits* SNPLex®, respectivamente. Como puede observarse en dicha tabla, 5 SNPs del primer *kit* y 1 del segundo presentan una tasa de genotipificación inferior a 0,5; es decir, para más del 50% de las muestras procesadas no fue posible asignar un genotipo en esos *loci*.
- **de muestras:** Para el primer *kit*, 20 (1,1%) de las 1.856 muestras analizadas presentaron una tasa de genotipificación inferior a 0,5; mientras que en el segundo *kit* sucedió lo mismo con 9 (1,5%) de las 581 muestras. Estas muestras fueron descartadas y no consideradas para ningún cálculo posterior.

Porcentaje de éxito en las reacciones de genotipificación.

Para el primer *kit*, el éxito en las reacciones de genotipificación fue del 94,1%, para 1.836 muestras de ADN y 34 SNPs; mientras que para el segundo *kit* fue del 88,4%, considerando 572 muestras de ADN y 22 SNPs.

SNP	Tasa de genotipificación	SNP	Tasa de genotipificación
rs43375517	0,98	rs43703010	0,95
rs42213673	0,98	LGBprom_37	0,40
rs41595314	0,98	rs41255685	0,77
rs29004488	0,96	rs43691049	0,00
OLR1_g.8232C>A	0,88	rs109625649	0,83
rs43702337	0,98	DGAT1_p.Lys232Ala	0,59
OPN3907	0,00	GH_51	0,98
SPP1_11740	0,92	GH_335	0,95
OPN_g.8514C>T	0,90	FASN_g.16024G>A	0,00
rs109579682	0,98	FASN_g.17924G>A	0,75
PPARGC1A_230	0,97	STAT5_g.9501G>A	0,01
PPARGC1A_337	0,97	rs135164815	0,97
rs43706475	0,96	GHR_140	0,97
rs43703017	0,97	UTMP_g.1296G>A	0,93
rs43703015	0,98	rs41256920	0,97
rs43703011	0,97	rs43706485	0,92
rs43703013	0,94	PRL_in1_188	0,98
CSN1_AB	0,87	PRL_g.8398G>A	0,98
CSN1_p.Ala53ThrP	0,99	rs41255693	0,97
CSN1_p.SerP66Leu	0,98		

Tabla I.8. Tasa de genotipificación de los SNPs analizados en el primer *kit* SNPLex®.

En negrita se indican los SNPs con tasa de genotipificación inferior a 0,5.

Reproducibilidad en la asignación de genotipos.

Este parámetro fue evaluado con el primer *kit* SNPLex®, obteniéndose una reproducibilidad del 100% calculada para 65 muestras por duplicado.

SNP	Tasa de genotipificación	SNP	Tasa de genotipificación
rs41586699	0,98	rs41624797	0,96
rs41631818	0,87	rs55617160	0,98
rs41637121	0,69	rs41628862	0,96
rs29014633	0,86	rs41579063	0,83
rs41655901	0,87	rs41587081	0,98
rs41652648	0,99	rs41579049	0,72
rs41592942	0,87	rs42575474	0,86
rs29018333	0,83	GH_299G-C	0,67
rs41622323	0,85	rs29018751	0,93
rs29018853	0,97	rs41644615	0,84
rs41569023	0,41	rs41645253	0,82
rs41256919	0,92		

Tabla I.9. Tasa de genotipificación de los SNPs analizados en el segundo *kit* SNPlex®.

En negrita se indica el SNP con tasa de genotipificación inferior a 0,5.

Errores mendelianos.

Entre los genotipos obtenidos para los SNPs evaluados en el primer *kit* se hallaron 777 errores mendelianos de un total de 62.424 genotipos, es decir, el 1,24% de los genotipos asignados a las vacas evaluadas resultaron ser inconsistentes con los genotipos paternos. Estos genotipos fueron eliminados y los *loci* correspondientes fueron tenidos en cuenta en la imputación (ver *ítem* 6 de la sección de Resultados de este Capítulo).

En el caso del segundo *kit* no pudieron determinarse los errores mendelianos puesto que no se contaba con los genotipos de los padres (ni de las madres) de las vacas analizadas.

5.B. Genotipos obtenidos.

El resumen de los genotipos obtenidos en el primer y segundo *kits* se presenta en las Tablas I.10 y I.11, respectivamente, en las que se exponen, para cada SNP, las frecuencias genotípicas, alélicas y la frecuencia alélica mínima (FAM). Estos parámetros fueron determinados discriminando a los animales por raza, es decir, para las vacas de raza Holando por un lado, y para las cruzas con Jersey por otro. Para su cálculo se utilizaron 1.363 animales Holando y 468 cruzas HxJ para los SNPs evaluados en el primer *kit* y 352 vacas Holando y 220 cruzas HxJ para los evaluados en el segundo *kit*.

SNP	Holando						Cruzas Holando x Jersey					
	Frecuencias genotípicas			Frecuencias alélicas		EHW (p)	Frecuencias genotípicas			Frecuencias alélicas		EHW (p)
	AA	AB	BB	A	B		AA	AB	BB	A	B	
rs43375517	0,40	0,48	0,12	0,64 (G)	0,36 (C)	0,23	0,52	0,40	0,08	0,72 (G)	0,28 (C)	0,72
rs42213673	0,36	0,50	0,14	0,61 (C)	0,39 (T)	0,16	0,52	0,42	0,06	0,73 (C)	0,27 (T)	0,20
rs41595314	0,85	0,14	0,01	0,92 (A)	0,08 (T)	<0,05	0,89	0,10	0,01	0,94 (A)	0,06 (T)	<0,05
rs29004488	0,33	0,51	0,16	0,58 (C)	0,42 (T)	0,10	0,22	0,55	0,23	0,49 (C)	0,51 (T)	<0,05
OLR1_g.8232C>A	0,39	0,48	0,13	0,63 (C)	0,37 (A)	0,29	0,55	0,41	0,04	0,75 (C)	0,25 (A)	<0,05
rs43702337	0,98	0,02	0,00	0,99 (A)	0,01 (C)	0,71	0,98	0,02	0,00	0,99 (A)	0,01 (C)	0,89
SPP1_11740	0,47	0,44	0,09	0,69 (A)	0,31 (C)	0,55	0,64	0,34	0,02	0,81 (A)	0,19 (C)	<0,05
OPN_g.8514C>T	0,26	0,50	0,24	0,51 (C)	0,49 (T)	0,91	0,37	0,49	0,14	0,62 (C)	0,38 (T)	0,47
rs109579682	0,60	0,36	0,04	0,78 (G)	0,22 (A)	0,18	0,61	0,34	0,05	0,78 (G)	0,22 (A)	0,80
PPARGC1A_230	0,88	0,12	0,00	0,94 (G)	0,06 (A)	0,05	0,40	0,55	0,05	0,68 (G)	0,32 (A)	<0,05
PPARGC1A_337	0,85	0,15	0,00	0,92 (C)	0,08 (T)	0,12	0,89	0,10	0,01	0,94 (C)	0,06 (T)	0,16
rs43706475	1,00	0,00	0,00	1,00 (T)	0,00 (G)	0,99	1,00	0,00	0,00	1,00 (T)	0,00 (G)	NC
rs43703017	0,79	0,20	0,01	0,89 (A)	0,11 (G)	0,54	0,87	0,13	0,00	0,93 (A)	0,07 (G)	0,45
rs43703015	0,83	0,17	0,00	0,91 (C)	0,09 (T)	0,07	0,49	0,46	0,05	0,72 (C)	0,28 (T)	<0,05
rs43703011	0,38	0,49	0,13	0,63 (C)	0,37 (A)	0,05	0,47	0,43	0,10	0,68 (C)	0,32 (A)	0,84
rs43703013	0,89	0,10	0,01	0,94 (C)	0,06 (G)	<0,05	0,93	0,06	0,01	0,96 (C)	0,04 (G)	<0,05
CSN1_AB	1,00	0,00	0,00	1,00 (T)	0,00 (A)	0,99	0,99	0,01	0,00	1,00 (T)	0,00 (A)	0,94
CSN1_p.Ala53ThrP	1,00	0,00	0,00	1,00 (G)	0,00 (A)	<0,05	1,00	0,00	0,00	1,00 (G)	0,00 (A)	NC
CSN1_p.SerP66Leu	1,00	0,00	0,00	1,00 (C)	0,00 (T)	0,99	1,00	0,00	0,00	1,00 (C)	0,00 (T)	NC
rs43703010	0,94	0,06	0,00	0,97 (A)	0,03 (G)	0,25	0,51	0,45	0,04	0,74 (A)	0,26 (G)	<0,05
rs41255685	0,86	0,02	0,12	0,88 (C)	0,12 (T)	<0,05	0,93	0,01	0,06	0,94 (C)	0,06 (T)	<0,05
rs109625649	0,85	0,14	0,01	0,92 (T)	0,08 (C)	0,26	0,80	0,16	0,04	0,89 (T)	0,11 (C)	<0,05

Tabla I.10. Resumen de genotipos del primer *kit* SNPlex®.

Frecuencias genotípicas y alélicas (FAM en negrita) obtenidas para cada SNP, discriminando a los animales por raza. En las celdas de las frecuencias alélicas se indican entre paréntesis los nucleótidos correspondientes a cada uno de los alelos. El equilibrio de Hardy-Weinberg se expresa como el nivel de significación p obtenido al realizar un test de hipótesis χ^2 . En negrita se indican aquellos SNPs que no se hallan en EHW en cada una de las subpoblaciones analizadas. Los valores fueron expresados con dos cifras decimales.

NC: no calculado (por incluir un cociente con denominador cero).

SNP	Holando						Cruzas Holando x Jersey					
	Frecuencias genotípicas			Frecuencias alélicas		EHW (p)	Frecuencias genotípicas			Frecuencias alélicas		EHW (p)
	AA	AB	BB	A	B		AA	AB	BB	A	B	
DGAT1_p.Lys232Ala	0,43	0,45	0,12	0,65 (GC)	0,35 (AA)	0,99	0,45	0,43	0,12	0,66 (GC)	0,34 (AA)	0,63
GH_51	0,29	0,51	0,20	0,54 (T)	0,46 (C)	0,50	0,09	0,55	0,36	0,37 (T)	0,63 (C)	<0,05
GH_335	0,73	0,24	0,03	0,85 (A)	0,15 (C)	<0,05	0,83	0,15	0,02	0,91 (A)	0,09 (C)	<0,05
FASN_g.17924G>A	0,53	0,41	0,06	0,74 (G)	0,26 (A)	<0,05	0,70	0,29	0,01	0,85 (G)	0,15 (A)	<0,05
rs135164815	0,73	0,25	0,02	0,96 (AC)	0,14 (GT)	0,31	0,80	0,19	0,01	0,89 (AC)	0,11 (GT)	0,45
GHR_140	0,68	0,28	0,04	0,82 (T)	0,18 (A)	0,09	0,76	0,22	0,02	0,87 (T)	0,13 (A)	0,81
UTMP_g.1296G>A	0,48	0,45	0,07	0,70 (G)	0,30 (A)	<0,05	0,14	0,61	0,25	0,45 (G)	0,55 (A)	<0,05
rs41256920	0,73	0,26	0,01	0,86 (A)	0,14 (C)	0,06	0,31	0,56	0,13	0,59 (A)	0,41 (C)	<0,05
rs43706485	0,71	0,27	0,02	0,84 (G)	0,16 (C)	0,84	0,41	0,49	0,10	0,66 (G)	0,34 (C)	0,15
PRL_in1_188	0,47	0,44	0,09	0,69 (T)	0,31 (C)	0,31	0,46	0,48	0,06	0,70 (T)	0,30 (C)	<0,05
PRL_g.8398G>A	0,82	0,17	0,01	0,90 (G)	0,10 (A)	0,81	0,72	0,26	0,02	0,85 (G)	0,15 (A)	0,87
rs41255693	0,68	0,28	0,04	0,82 (C)	0,18 (T)	0,21	0,80	0,19	0,01	0,90 (C)	0,10 (T)	0,65

Tabla I.10. Resumen de genotipos del primer *kit* SNPlex® (continuación).

SNP	Holando						Cruzas Holando x Jersey					
	Frecuencias genotípicas			Frecuencias alélicas		EHW (p)	Frecuencias genotípicas			Frecuencias alélicas		EHW (p)
	AA	AB	BB	A	B		AA	AB	BB	A	B	
rs41586699	0,47	0,45	0,08	0,69 (T)	0,31 (C)	0,28	0,65	0,34	0,01	0,82 (T)	0,18 (C)	0,06
rs41631818	0,39	0,48	0,13	0,63 (A)	0,37 (C)	0,53	0,25	0,49	0,26	0,49 (A)	0,51 (C)	0,89
rs41637121	0,42	0,49	0,09	0,67 (A)	0,33 (G)	0,13	0,28	0,56	0,16	0,56 (A)	0,44 (G)	0,14
rs29014633	0,51	0,40	0,09	0,72 (A)	0,28 (T)	0,61	0,61	0,36	0,03	0,79 (A)	0,21 (T)	0,34
rs41655901	0,25	0,54	0,21	0,52 (C)	0,48 (T)	0,13	0,04	0,48	0,48	0,28 (C)	0,72 (T)	<0,05

Tabla I.11. Resumen de genotipos del segundo *kit* SNPlex®.

Frecuencias genotípicas y alélicas (FAM en negrita) obtenidas para cada SNP, discriminando a los animales por raza. En las celdas de las frecuencias alélicas se indican entre paréntesis los nucleótidos correspondientes a cada uno de los alelos. El equilibrio de Hardy-Weinberg se expresa como el nivel de significación p obtenido al realizar un test de hipótesis χ^2 . En negrita se indican aquellos SNPs que no se hallan en EHW en cada una de las subpoblaciones analizadas. Los valores fueron expresados con dos cifras decimales.

NC: no calculado (por incluir un cociente con denominador cero).

SNP	Holando						Cruzas Holando x Jersey					
	Frecuencias genotípicas			Frecuencias alélicas		EHW (p)	Frecuencias genotípicas			Frecuencias alélicas		EHW (p)
	AA	AB	BB	A	B		AA	AB	BB	A	B	
rs41652648	0,82	0,17	0,01	0,91 (A)	0,09 (G)	0,98	0,52	0,45	0,03	0,75 (A)	0,25 (G)	<0,05
rs41592942	0,39	0,46	0,15	0,62 (G)	0,38 (A)	0,70	0,66	0,33	0,01	0,82 (G)	0,18 (A)	<0,05
rs29018333	0,87	0,13	0,00	0,93 (G)	0,07 (A)	0,78	0,63	0,35	0,02	0,80 (G)	0,20 (A)	0,16
rs41622323	0,88	0,11	0,01	0,94 (G)	0,06 (A)	0,35	0,90	0,10	0,00	0,95 (G)	0,05 (A)	0,49
rs29018853	0,46	0,45	0,09	0,69 (C)	0,31 (A)	0,47	0,16	0,58	0,26	0,45 (C)	0,55 (A)	<0,05
rs41256919	0,34	0,53	0,13	0,61 (G)	0,39 (A)	0,06	0,14	0,53	0,33	0,41 (G)	0,59 (A)	0,15
rs41624797	0,81	0,18	0,01	0,90 (G)	0,10 (A)	0,42	0,81	0,19	0,00	0,90 (G)	0,10 (A)	0,12
rs55617160	0,81	0,18	0,01	0,90 (G)	0,10 (A)	0,86	0,82	0,18	0,00	0,91 (G)	0,09 (A)	0,14
rs41628862	0,53	0,34	0,13	0,70 (C)	0,30 (G)	<0,05	0,05	0,52	0,43	0,31 (C)	0,69 (G)	<0,05
rs41579063	0,56	0,35	0,09	0,74 (A)	0,26 (T)	0,13	0,05	0,60	0,35	0,35 (A)	0,65 (T)	<0,05
rs41587081	0,54	0,43	0,03	0,75 (T)	0,25 (G)	<0,05	0,75	0,24	0,01	0,87 (T)	0,13 (G)	0,35
rs41579049	0,55	0,40	0,05	0,75 (T)	0,25 (C)	0,20	0,69	0,31	0,00	0,84 (T)	0,16 (C)	<0,05
rs42575474	0,74	0,24	0,02	0,86 (G)	0,14 (A)	0,96	0,35	0,57	0,08	0,64 (G)	0,36 (A)	<0,05
GH_299G-C	0,95	0,05	0,00	0,97 (C)	0,03 (G)	0,65	0,36	0,61	0,03	0,66 (C)	0,34 (G)	<0,05
rs29018751	0,66	0,31	0,03	0,82 (A)	0,18 (G)	0,43	0,15	0,68	0,17	0,49 (A)	0,51 (G)	<0,05
rs41644615	0,32	0,48	0,20	0,56 (C)	0,44 (T)	0,59	0,41	0,53	0,06	0,67 (C)	0,33 (T)	<0,05
rs41645253	0,41	0,43	0,16	0,63 (G)	0,37 (T)	0,18	0,15	0,55	0,30	0,43 (G)	0,57 (T)	0,10

Tabla I.10. Resumen de genotipos del segundo *kit* SNPlex® (continuación).

Dado que los SNPs pertenecientes a los genes codificantes de caseínas fueron seleccionados de modo tal de poder identificar las diferentes variantes caseínicas, se calcularon las frecuencias de tales variantes en cada una de las subpoblaciones analizadas. Este resultado se presenta en la Tabla I.12.

5.C. Equilibrio de Hardy-Weinberg.

La población en estudio en este trabajo de tesis no cumple con los postulados que establece el principio de Hardy-Weinberg (enunciados en el Anexo VI) ya que los cruzamientos en la misma no son al azar y ocurre permanentemente migración (traslado de animales a/desde otros tambos del establecimiento o fuera del establecimiento). Además, aunque no existe un programa de selección específico, puede considerarse que tiene lugar cierto proceso

de selección artificial ya que para los cruzamientos se utilizan, en general, los animales de mejor aptitud productiva.

Gen	Variante	Holando	Cruzas Holando x Jersey
CSN1S1	A	0,00	0,00
	B	0,97	0,73
	C	0,03	0,26
CSN2	A ¹	0,32	0,28
	A ²	0,63	0,70
	B	0,05	0,02
CSN3	A	0,80	0,66
	B	0,09	0,28
	E	0,11	0,07

Tabla I.12. Variantes caseínicas.

Frecuencia de presentación de las distintas variantes correspondientes a los genes de caseínas para los animales de raza Holando y cruzas Holando x Jersey. Los valores fueron expresados con dos cifras decimales.

A pesar de lo mencionado en el párrafo anterior, se realizaron estimaciones del EHW para cada uno de los SNPs analizados, puesto que el hecho de que varios *loci* cercanos se encuentren fuera del equilibrio de Hardy-Weinberg puede dar un indicio de la existencia de desequilibrio de ligamiento entre ellos. La comprobación de desviación del EHW también constituye un control del sistema de genotipificación ya que la ausencia de desequilibrio podría estar sugiriendo una falla en el sistema. En las Tablas I.10 y I.11 se indica el nivel de significación (p) del *test* de hipótesis del EHW para cada uno de los SNPs analizados en el primer y segundo *kits*, respectivamente.

Es destacable mencionar que, tanto para los animales de raza Holando como para las cruzas estudiadas, 3 SNPs pertenecientes al *cluster* de las caseínas presentan un $p < 0,05$ por lo que no estarían en EHW. Teniendo en cuenta la proximidad física entre dichos SNPs (lo que determina que pertenezcan a un mismo *cluster*), este resultado confirmaría la existencia de desequilibrio de ligamiento entre los *loci* implicados.

6. Determinación de fases de haplotipos e imputación de genotipos faltantes.

Debido a que posteriormente al uso del sistema SNPlex® se pudo disponer de los genotipos del Bovine SNP50 v2 BeadChip (Illumina Inc., San Diego, California), los *loci* con genotipo no asignado por ninguno de los dos *kits* y aquellos del primer *kit* eliminados por presentar inconsistencias mendelianas pudieron ser imputados (ver Capítulo II) al reunir todos los genotipos generados mediante ambos sistemas de genotipificación y someterlos, mediante la utilización del programa BEAGLE (Browning y Browning, 2007; Browning y Browning, 2009), al proceso de construcción de haplotipos y asignación de fases. La mayor densidad de marcadores alcanzada de esta manera hace posible realizar una estimación de haplotipos con mayor precisión.

DISCUSIÓN.

Uno de los objetivos de este capítulo fue la aplicación de una metodología para la selección de genes candidatos relacionados con características productivas y atributos de calidad de leche en bovinos lecheros, y la identificación de marcadores moleculares en esos genes candidatos seleccionados. La búsqueda en bibliografía y en bases de datos internacionales de las publicaciones más relevantes relacionadas con la producción y composición de leche en el ganado bovino y los QTLs, genes y marcadores moleculares asociados a estas características, ha conducido a la selección de 27 genes candidatos. De éstos, 5 (18,5%) corresponden a genes codificantes de proteínas relacionadas al fenotipo estudiado que presentan diferentes variantes genéticas, 8 (29,6%) son genes con variantes alélicas que fueron reportadas en bibliografía como asociadas al rasgo en estudio, 2 (7,4%) consisten en genes con perfiles de expresión asociados a la variación fenotípica del rasgo analizado, 10 (37,0%) son genes localizados en QTLs asociados al fenotipo en estudio, 1 (3,7%) es un gen regulado post-transcripcionalmente por miRNAs expresados en la glándula mamaria y el último (3,7%) es un gen cuya expresión alterada en animales *knock out* o transgénicos resultó en fenotipos específicos asociados al rasgo en estudio. Algunos de estos genes candidatos fueron identificados siguiendo más de una estrategia de búsqueda, o, siguiendo un único abordaje de búsqueda, se hallaron en múltiples publicaciones. Veinte de estos genes (*LEP*, *OLR1*, α -*LA*, *ABCG2*, *SPP1*, *PPARGC1A*, *CSN2*, *CSN3*, *CSN1S1*, *CSN1S2*, *LGB*, *DGAT1*, *STAT5A*, *GH*, *FASN*, *GHR*, *PRLR*, *LTF*, *PRL* y *SCD*) también fueron seleccionados como genes candidatos para producción lechera por Ogorevc y col. (2009), mientras que 14 (*LEP*, α -*LA*, *CSN2*, *CSN3*, *CSN1S1*, *CSN1S2*, *LGB*, *DGAT1*, *STAT5A*, *GH*, *GHR*, *PRLR*, *LTF* y *PRL*) fueron asimismo considerados por Kamiński y col. e incluidos en el diseño de los *chips* MilkProtChip y/o MilkBovExp para la genotipificación de SNPs pertenecientes a genes candidatos relacionados a la producción lechera bovina (Kamiński y col., 2005; Kamiński y col., 2008). La mayor densidad de genes candidatos se halló en los BTAs 6 y 14, correspondiéndose esto con la gran cantidad de QTLs reportados para caracteres productivos en estos cromosomas: para el BTA 6 fueron reportados 21 QTLs relacionados con rendimiento de leche, 29 con grasa en la leche y 43 con proteína en la leche, mientras que para el BTA 14 se reportaron 5, 26 y 13 QTLs relacionados con dichos caracteres, respectivamente (Fuente: <http://genomes.sapac.edu.au/bovineqtl/>, Bovine QTL Viewer).

Posteriormente, la búsqueda de los SNPs reportados para cada gen candidato así como de SNPs asociados al fenotipo analizado localizados fuera de los genes candidatos previamente seleccionados, y también la detección de nuevos SNPs por secuenciación directa de una muestra de la población en estudio, nos permitió hallar 128 SNPs candidatos a ser analizados. Estos SNPs fueron ordenados según su prioridad para ser genotipificados en base al tipo y localización de la mutación. A aquellos polimorfismos consistentes en mutaciones no sinónimas o sin sentido que presentaban la posibilidad de afectar la función o la expresión de una proteína por ubicarse en dominios catalíticos, estructurales claves o bien regulatorios de la expresión de un gen, se les asignó una prioridad mayor para ser genotipificados. Por otro lado, a los SNPs consistentes en mutaciones sinónimas o localizadas en regiones intergénicas o en intrones les fue asignada una menor prioridad. Finalmente, luego de esta preselección, 62

SNPs fueron incluidos en el diseño del sistema de genotipificación elegido. Estos marcadores no se hallan homogéneamente distribuidos en el genoma bovino, dado que hay cromosomas que albergan varios de ellos (como los BTAs 5, 6, 11, 14 y 19, con 5 o más SNPs cada uno), mientras que otros están desprovistos de SNPs (BTAs 2, 3, 7, 8, 9, 10, 12, 13, 16, 17, 18, 25, 27, 28, 29, X). A pesar de lo mencionado previamente, 56 de los 62 SNPs pudieron ser considerados para el análisis posterior dado que los 6 restantes presentaron problemas en la asignación de genotipos. Por lo tanto, la aplicación de la metodología empleada para la selección de genes candidatos y de marcadores moleculares de tipo SNP relacionados con características de producción y composición de la leche bovina condujo a la obtención de un panel de 56 SNP, utilizado en la evaluación del genotipo de bovinos lecheros.

En relación a los 9 nuevos SNPs detectados en la población de toros secuenciada, es importante mencionar que 8 de los mismos no figuraban en bases de datos al momento del análisis, aunque durante la escritura de esta tesis fueron reportados. Por lo tanto, al haber sido detectados en una población diferente a la utilizada en este trabajo, por un lado confirman el resultado obtenido aquí, y por otro, validan su utilización en estudios genéticos posteriores. Se trata de los SNPs ABCG2_190, SPP1_11740, PPARGC1A_230, PPARGC1A_337, LGBprom_37, LGB_in3ex4_134, GH_335 y PRL_in1_39, para los que fueron asignados los códigos *rs* rs473562396, rs132812135, rs133669403, rs17870811, rs41255680, rs378251685, rs109191047 y rs384868312, respectivamente.

El otro objetivo del capítulo era la determinación del genotipo de los SNPs seleccionados en un rodeo comercial de razas Holando y cruza Holando x Jersey y su caracterización en función de distintos parámetros poblacionales. La genotipificación fue realizada satisfactoriamente dado que el éxito en las reacciones implicadas fue superior al 94 y al 88% para los SNPs del primer y segundo *kit*, respectivamente y, además, la reproducibilidad en la asignación de genotipos fue del 100% (calculada con el primer *kit*). Otros parámetros que brindan información acerca de la calidad de las reacciones de genotipificación son las tasas de genotipificación de las muestras y de los SNPs. Sin considerar los SNPs con tasa de genotipificación < 0,5 que tuvieron problemas en la asignación de genotipos, el 94% y el 81% de las muestras evaluadas presentaron una tasa de genotipificación > 0,8 en el caso del primer y segundo *kit*, respectivamente, mientras que el 91% y el 87% de los SNPs analizados en el primer y segundo *kit*, respectivamente presentaron una tasa de genotipificación > 0,8. De la observación de las frecuencias genotípicas y alélicas (en especial, de la frecuencia alélica mínima) calculadas para cada subpoblación y para cada SNP, y de los haplotipos hallados en cada subpoblación y las frecuencias en las que se presentan los mismos, puede establecerse el aporte de cada raza a la variabilidad alélica de la población en su conjunto. Los valores de FAMs difieren apreciablemente para algunos SNPs entre los animales Holando y las cruza HxJ, llegando en ocasiones a ser menos frecuente en una subpoblación el alelo que en la otra es más frecuente. Las frecuencias genotípicas y alélicas mínimas obtenidas para los SNPs analizados también por otros autores se corresponden con las reportadas en poblaciones de la misma raza o cruza de razas (Giblin y col., 2010; Buchanan y col., 2003; Khatib y col., 2006; Schennink y col., 2009; Komisarek y Dorynek, 2009; Leonard y col., 2005). En relación a las variantes de caseínas que predominan en la población analizada, éstas muestran una

tendencia coincidente con la publicada a nivel internacional (variante B para α 1-caseína-, variante A² para β -caseína y variante A para κ -caseína; Farrell y col., 2004; Lien y col., 1999). Otro parámetro analizado fue la existencia o no de equilibrio de Hardy-Weinberg en la población. Siete y 16 de los 34 SNPs evaluados con el primer *kit* se encontraron fuera del EHW para los animales Holando y las cruzas HxJ, respectivamente, mientras que sucedió lo mismo con 2 y 11 de los 22 SNPs evaluados con el segundo *kit* para las vacas Holando y las cruzas HxJ, respectivamente. Aunque el número de SNPs fuera del EHW es alto, se trató de un resultado esperable puesto que la población estudiada no respeta los postulados básicos del principio de Hardy-Weinberg (los cruzamientos no se dan al azar, hay migración y selección a favor de un fenotipo particular). Sin embargo, para aquellos SNPs cercanos en su ubicación cromosómica como es el caso de los pertenecientes al *cluster* de caseínas, el hecho de que se hallen fuera del EHW puede sugerir la existencia de desequilibrio de ligamiento entre ellos. Por otro lado, para los SNPs evaluados en el primer *kit* se determinó la presencia de errores mendelianos, es decir, de inconsistencias genotípicas en los núcleos familiares padre-madre-hija, hallándose este tipo de errores en el 1,24% de los genotipos asignados. Esto puede deberse a una asignación errónea de paternidades por parte del establecimiento al que pertenecen los animales (de hecho, el establecimiento reconoce que suceden este tipo de problemas en alrededor del 20% de los animales), a un error al registrar los animales, a errores en el procesamiento de las muestras (intercambio de tubos correspondientes a muestras diferentes), o bien a errores en la genotipificación o en la asignación de los genotipos obtenidos. Los genotipos que presentaron inconsistencias mendelianas, así como los no asignados por el sistema comercial utilizado, fueron imputados en el Capítulo II de este trabajo de tesis, en el proceso de determinación de haplotipos y asignación de fases, al considerar, junto a los genotipos obtenidos con el sistema SNPlex®, la información genotípica obtenida a través de la genotipificación con un *chip* de mediana densidad (50K). La mayor densidad de genotipos aportados por la genotipificación mediante el barrido del genoma respecto de la obtenida únicamente con el sistema SNPlex®, permitió alcanzar una precisión aceptable en la imputación de los genotipos faltantes y la determinación de fases haplotípicas.

La información genotípica obtenida a través del abordaje con genes candidatos será suplementada en el Capítulo II con los genotipos obtenidos a través de un microarreglo (*chip*) comercial de mediana densidad, para ser luego sometida a un análisis de asociación en el Capítulo III de este trabajo de tesis.

CAPÍTULO II

GENOTIPIFICACIÓN DE MEDIA DENSIDAD DE
MARCADORES MOLECULARES DISTRIBUIDOS EN
BLOQUES DE GENES CANDIDATOS

En este capítulo se describe el análisis realizado con los genotipos obtenidos mediante un microarreglo comercial de mediana densidad de marcadores. El trabajo se centra, en particular, en los cromosomas en los que se concentra la mayor cantidad de genes candidatos para caracteres relacionados con la producción de leche en bovinos.

El capítulo comienza con una introducción general a la genotipificación mediante barrido del genoma en la que se describen las principales aplicaciones de los microarreglos de mediana y alta densidad de marcadores así como los criterios más comúnmente utilizados para realizar el control de calidad de la genotipificación, luego se describen los materiales y métodos utilizados para el desarrollo y análisis de los genotipos obtenidos, y finalmente se exponen los resultados a través del análisis de su calidad o confiabilidad y de una descripción de distintos parámetros propios de la población estudiada. El capítulo finaliza con la discusión y exposición de las conclusiones obtenidas.

INTRODUCCIÓN.

1. Genotipificación de baja, media y alta densidad.

Los avances en las metodologías utilizadas en genética molecular así como de las herramientas bioinformáticas necesarias para el análisis de grandes volúmenes de datos que tuvieron lugar en las últimas décadas, introdujeron cambios en las evaluaciones genéticas del ganado a través del uso de marcadores de ADN en la identificación de genes o regiones genómicas responsables del control de determinadas características fenotípicas.

Actualmente, la tecnología de microarreglos (*microarrays* o *chips*) permite evaluar hasta cientos de miles de SNPs simultáneamente con alta precisión y a un costo accesible. Existen plataformas de baja, media y alta densidad de marcadores que permiten realizar la genotipificación de algunos miles, decenas de miles y cientos de miles de SNPs, respectivamente. La mayoría de los estudios reportados en literatura utilizan productos de dos plataformas comerciales: Illumina (San Diego, California, EE.UU.) y Affymetrix (Santa Clara, California, EE.UU.). La plataforma Affymetrix consiste en pequeñas secuencias de ADN dispuestas puntualmente en el *chip*, que reconocen los alelos específicos de un SNP. Los distintos alelos o nucleótidos son detectados por hibridación diferencial de la muestra de ADN. Por otro lado, la plataforma Illumina utiliza una tecnología basada en cuentas o *beads* (esferas que contienen ADN unido a su superficie) que soportan secuencias de ADN algo más largas que las de Affymetrix para detectar los diferentes alelos de un SNP, lo que provee mayor especificidad a la reacción de hibridación (Bush y Moore, 2012). Además de la tecnología utilizada, otros parámetros a tener en cuenta son la cantidad y el origen de los SNPs que incluye cada plataforma. Esto puede ser de relevancia según la población de animales con la que se trabaje.

Hasta el momento, las principales aplicaciones de los microarreglos de mediana y alta densidad de marcadores consisten en su utilización en estudios de asociación a nivel del genoma completo y para estimar el efecto de SNPs y luego el mérito genético de los animales, con el fin de ser utilizados en selección genómica.

En este trabajo de tesis se utilizó el microarreglo de mediana densidad BovineSNP50 v2 BeadChip (Illumina Inc., San Diego, California, EE.UU.). Esta plataforma evalúa 54.609 SNPs distribuidos en los 29 autosomas y en los cromosomas sexuales del genoma bovino. Los marcadores incluidos en el diseño del *chip* fueron seleccionados a partir de bases de datos de libre acceso, y el panel de SNPs obtenido provee genotipos robustos, reproducibles y polimórficos para una amplia variedad de razas bovinas, tanto para carne como para leche - Hereford, Charolais, Holando, Norwegian Red, Angus, Jersey, Brown Swiss, Simmental, Brahman, entre otras- (Matukumalli y col., 2009).

2. Control de calidad de la genotipificación.

Tradicionalmente, en una genotipificación a baja escala, con baja densidad de marcadores, como la descrita en el Capítulo I de este trabajo de tesis, la asignación de

genotipos se lleva a cabo tras realizar una inspección manual de las intensidades de las señales obtenidas para cada muestra y cada marcador analizado. Sin embargo, en la genotipificación con media y alta densidad de marcadores esta inspección manual deja de ser práctica, por lo que se recurre a algoritmos matemáticos computacionales.

Las plataformas comerciales que evalúan simultáneamente hasta cientos de miles de SNPs realizan la genotipificación a través de un proceso que implica la evaluación de las intensidades de fluorescencia de sondas complementarias a los posibles alelos de regiones específicas del genoma. La transformación de las intensidades de hibridación de cada *locus* en genotipos requiere el uso de algoritmos que comparan la intensidad relativa de las señales fluorescentes entre los posibles alelos para decidir cuál es el genotipo más probable (Teo, 2010). Estos algoritmos evalúan los perfiles de intensidad en múltiples muestras simultáneamente y, en base a las posiciones y características de las nubes de genotipos, estiman la probabilidad de los tres posibles genotipos (los dos homocigotas y el heterocigota) para cada muestra en cada SNP. Utilizan, además, un valor de corte o umbral de probabilidad a partir del cual los genotipos se aceptan y son asignados, y por debajo del cual los genotipos se clasifican como no asignados o perdidos. Del valor umbral aplicado dependerá la tasa de asignación de genotipos así como la calidad de los mismos: si el umbral es demasiado bajo es probable que se asignen genotipos erróneos, por el contrario, si es muy alto, es probable que queden sin asignar una gran cantidad de genotipos, reduciéndose de esta forma la cobertura genómica así como el poder del estudio para detectar asociaciones con caracteres fenotípicos de interés (Anderson y col., 2010). La interpretación incorrecta de los perfiles de intensidad de nubes de genotipos ambiguas o atípicas puede conducir a una asignación errónea de genotipos.

Debido a lo expuesto anteriormente, una vez asignados los genotipos, es de crucial importancia analizar una serie de parámetros con el objetivo de detectar posibles errores en la genotipificación y eliminar muestras o marcadores con genotipos dudosos. Se debe tener en cuenta que con cada SNP que se remueve del estudio se pierde una potencial asociación a algún QTL de interés, por lo que el impacto de esta remoción puede ser mayor que el de la remoción de un individuo. Por este motivo, Anderson y col. (2010) sugieren implementar el control de calidad de la genotipificación en primer lugar sobre los individuos y luego sobre los marcadores, de modo de maximizar el número de SNPs que permanezcan en el estudio.

A continuación se describen los aspectos a considerar para la obtención de una base de datos con genotipos de alta calidad en el uso de un *chip*. En la Figura II.1 se presenta esquemáticamente el diagrama de flujo del análisis.

2.A. Control de calidad sobre individuos.

El control de calidad sobre muestras individuales consiste básicamente en analizar los parámetros mencionados a continuación.

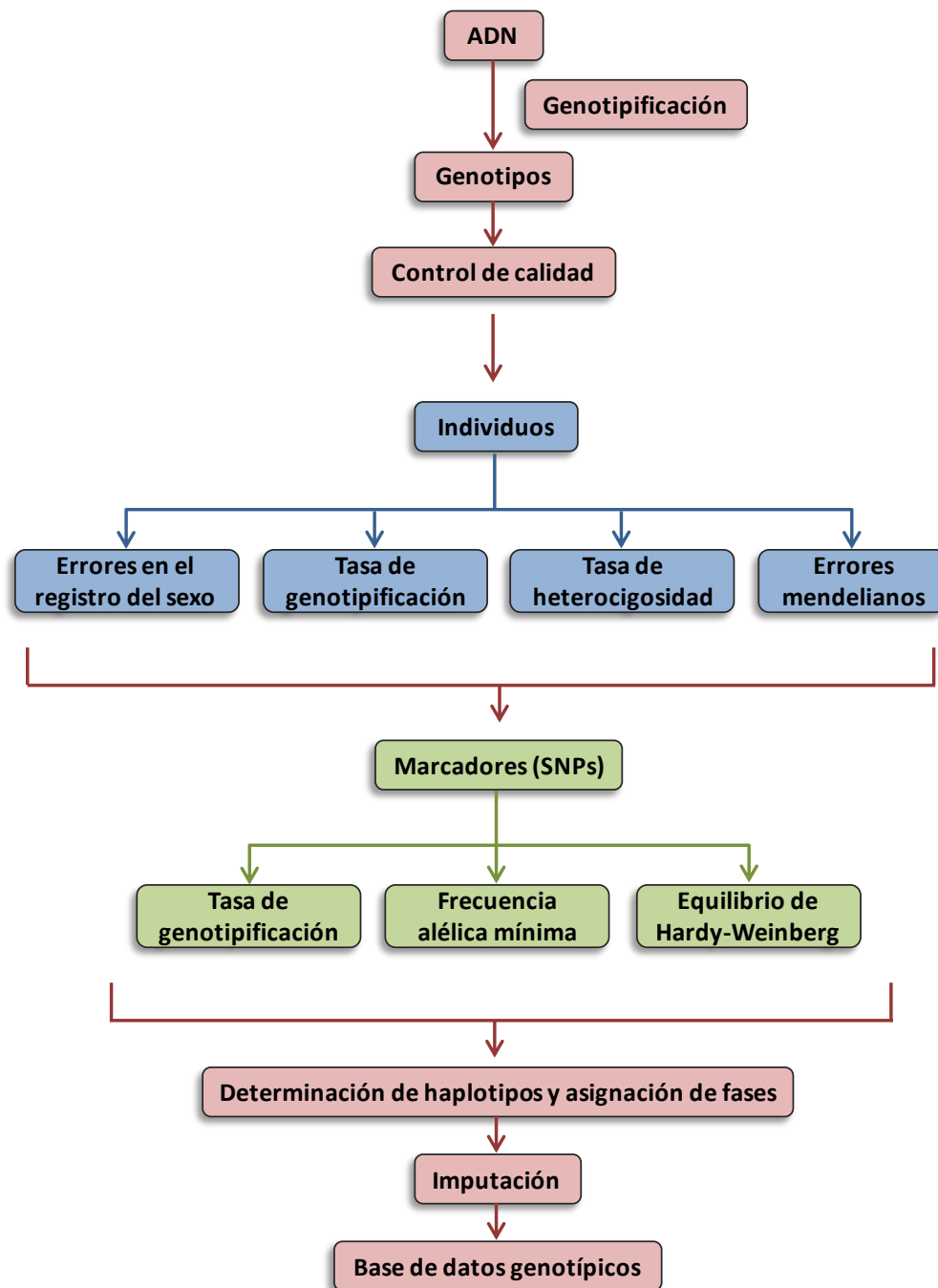


Figura II.1. Diagrama de flujo del análisis.

Aspectos a considerar en el análisis de genotipos con el fin de obtener una base de datos genotípicas de alta calidad.

2.A.1. Información relacionada al sexo.

Los genotipos de los SNPs localizados en el cromosoma X se utilizan para identificar individuos que presenten discordancias entre el sexo reportado y el correspondiente según los genotipos obtenidos. Como los machos poseen sólo una copia del cromosoma X, no pueden ser heterocigotas para ningún marcador que se halle fuera de la región pseudoautosómica del

cromosoma Y. Con el objetivo de detectar individuos con información sexual discordante, normalmente se calcula la tasa de homocigosidad para todos los SNPs pertenecientes al cromosoma X para cada individuo genotipificado y luego se comparan los valores obtenidos con los esperados. En humanos, se espera que los hombres presenten una tasa de homocigosidad = 1 en estos SNPs (aunque este valor puede ser algo menor debido a errores en la genotipificación) y las mujeres un valor menor a 0,2 (Anderson y col., 2010). A menos que el sexo del individuo pueda ser correctamente identificado utilizando la información genotípica disponible o, habiéndose detectado una discrepancia, pueda confirmarse que el sexo fue registrado incorrectamente, todos los individuos con información sexual discordante deberían ser removidos del análisis.

2.A.2. Tasas de genotipificación y de heterocigosidad.

La calidad de la muestra de ADN puede tener un gran efecto sobre la tasa de genotipificación y la precisión en la asignación de genotipos. Las muestras de ADN de mala calidad (medida a través de la relación de absorbancias 260nm/280nm y 260nm/230nm), de baja concentración o con ADN degradado, presentan frecuentemente una baja tasa de genotipificación, con alta proporción de genotipos incorrectamente asignados. La proporción de genotipos no asignados para cada muestra de ADN, así como su tasa de heterocigosidad son parámetros indicativos de la calidad de la muestra de ADN. Normalmente, los individuos con más de 5 - 10% de genotipos no asignados son removidos del análisis (Hayes y col., 2009; Jiang y col., 2010; Strucken y col., 2011) para evitar que estas muestras que han sufrido un proceso de genotipificación poco confiable aumenten la posibilidad de incurrir en asociaciones falsamente positivas en los análisis posteriores. También debe inspeccionarse la heterocigosidad media que presentan los individuos (excluyendo aquella correspondiente a los SNPs de los cromosomas sexuales) con el objetivo de identificar individuos con una proporción excesiva o muy baja de genotipos heterocigotas, dado que una proporción excesiva de los mismos puede indicar contaminación en la muestra de ADN, mientras que una proporción muy baja de heterocigotas sugiere existencia de consanguinidad (*inbreeding*).

2.A.3. Inconsistencias mendelianas.

Otro parámetro que determina la permanencia de una muestra (o del genotipo asignado a un *locus* en particular) en el análisis posterior o su remoción del mismo, es el número de inconsistencias mendelianas detectadas en dúos padre-hijo o tríos madre-padre-hijo. Los genotipos incompatibles en núcleos familiares son generalmente removidos del análisis (calificados como no asignados y considerados para su imputación). En la literatura se han reportado diversos criterios para decidir si un individuo que presenta inconsistencias genotípicas con sus padres es considerado o no para su estudio. A modo de ejemplo, Jiang y col. (2010) removieron de su análisis aquellos animales que superaban un umbral de 2% de errores mendelianos en 10.000 *loci* de SNPs elegidos aleatoriamente para los que los dúos toro-cría fueran homocigotas. Por otro lado, Wiggans y col. (2009) compararon los genotipos de 56.947 SNPs de cada animal que analizaron con los de sus padres, determinaron el número de veces que los SNPs homocigotas diferían en los dúos padre-hijo y, ante la presencia de más

de 200 discrepancias, consideraron la existencia de un conflicto padre-hijo y comprobaron la asignación de paternidades.

2.B. Control de calidad sobre marcadores.

El control de calidad sobre los marcadores de tipo SNPs analizados consiste en evaluar los parámetros que se mencionan en los párrafos siguientes.

2.B.1. Tasa de genotipificación.

La tasa de genotipificación alcanzada por cada SNP es generalmente el primer filtro que se impone en la selección de los marcadores a considerar para el análisis posterior. Una alta proporción de genotipos no asignados para un SNP dado (baja tasa de genotipificación) sugiere la presencia de perfiles de intensidad confusos que resultan en nubes de genotipos superpuestas o ambiguas o con características de agrupamiento atípicas, que al algoritmo de asignación de genotipos le cuesta interpretar (Teo, 2010). La remoción de SNPs con una alta proporción de genotipos no asignados reduce la posibilidad de que un SNP con genotipificación dudosa sea retenido en los análisis posteriores. Normalmente, los SNPs que presentan una tasa de genotipificación menor a 0,80 - 0,95 son removidos del análisis (Bolormaa y col., 2010; Schopen y col., 2011; Strucken y col., 2011).

2.B.2. Frecuencia alélica mínima (FAM).

Los SNPs monomórficos no son informativos en estudios genéticos de asociación por no presentar diferencias genotípicas. Por otro lado, las variantes genéticas que presentan una baja FAM resultan en bajos recuentos genotípicos para al menos uno de los tres posibles genotipos, produciendo nubes genotípicas difusas que pueden conducir a una asignación de genotipos poco fiable (Teo, 2010). Normalmente se remueven los SNPs con FAMs menores a 1 - 5% (Chan y col., 2008; Guo y col., 2012; Hayes y col., 2009; Mai y col., 2010) dado que éstos son más propensos a ser afectados por errores en el proceso de genotipificación y son generalmente poco informativos en los estudios de asociación.

2.B.3. Desviación del equilibrio de Hardy-Weinberg (EHW).

Matemáticamente, se define que un SNP está en equilibrio de Hardy-Weinberg cuando la probabilidad de observar un genotipo particular es igual a la probabilidad de observar los dos alelos independientemente. El análisis de la existencia de desviación del EHW implica comparar las frecuencias genotípicas observadas con las esperadas en base a las frecuencias alélicas de la población, mediante una prueba de bondad de ajuste χ^2 de Pearson o un *test* exacto de Fisher sobre la hipótesis nula de que $p^2 + 2pq + q^2 = 1$, donde p y q son las dos frecuencias alélicas (Emigh, 1980). Las desviaciones del EHW pueden deberse a consanguinidad, estratificación poblacional, selección (Balding, 2006) o deriva genética. En el caso de la genotipificación mediante barrido del genoma, dado el alto número de SNPs que se analizan y la estructura de la población estudiada normalmente, la conformidad con el principio de HW no es estricta. A pesar de esto, como los algoritmos matemáticos empleados

en las plataformas comerciales de genotipificación a escala masiva normalmente introducen errores que producen distribuciones genotípicas que violan significativamente el EHW (Teo y col., 2007), la mayoría de los estudios de asociación a nivel de genoma completo (GWAS) no consideran los marcadores que presentan una gran desviación del mismo, dado que esto puede sugerir una asignación de genotipos incorrecta (Anderson y col., 2010). Una gran desviación del EHW también puede indicar que el SNP se encuentre duplicado en el genoma en lugar de corresponder a un único *locus* (Wiggans y col., 2009). El umbral de significación p para declarar que un SNP se halla apartado del EHW varía ampliamente, habiéndose reportado valores entre 0,01 y 1.10^{-6} en literatura (Guo y col., 2012; Jiang y col., 2010). En el caso de los GWAS, debido a lo expuesto anteriormente, se utilizan umbrales algo más laxos ($p_{HWE} < 1.10^{-4}$ a 1.10^{-7}) (Teo, 2010).

3. Determinación de haplotipos y asignación de fases.

Un haplotipo es una secuencia de alelos que se ubican en el mismo cromosoma, es decir, que fueron heredados del mismo progenitor. Dado que los genotipos obtenidos en los ensayos de genotipificación consisten en pares de alelos desordenados, al determinar los haplotipos materno y paterno de un individuo debe inferirse la fase en la que se encuentran dichos alelos (Browning, 2008). La asignación de fases de haplotipos en individuos no emparentados se basa en la frecuencia haplotípica poblacional, siendo mejor la estimación cuanto mayor es el tamaño de la población analizada. Por otro lado, las fases de los haplotipos de individuos emparentados pueden determinarse con alta precisión en base a la información de la identidad por descendencia (IBD, *identity-by-descent*) existente entre pares de alelos de dichos individuos. Sin embargo, cuando la información de IBD en las familias no logra determinar con certeza las fases, la frecuencia en la que se presentan los diferentes haplotipos en la población puede utilizarse para completar la estimación de las mismas (Browning y Browning, 2012).

Se han desarrollado varios programas para estimar las fases haplotípicas. Éstos se basan en la utilización del algoritmo EM (de “esperanza-maximización”, que trata a todas las posibles configuraciones haplotípicas como igualmente probables *a priori*) y de los métodos basados en coalescencia y modelos ocultos de Markov (que reconocen qué nuevos haplotipos pueden ser derivados de los anteriores por proceso de recombinación y mutación). Entre los diversos programas estadísticos desarrollados para la determinación de fases haplotípicas se pueden mencionar los siguientes: PHASE (Stephens y col., 2001), fastPHASE (Scheet y Stephens, 2006), DualPHASE (Druet y Georges, 2010), MaCH (Li y col., 2010), IMPUTE v2 (Howie y col., 2009), BEAGLE (Browning y Browning, 2007; Browning y Browning, 2009). Éstos difieren en la precisión en la determinación de fases, en su performance frente a distintos tamaños poblacionales y densidades de marcadores, en el tiempo computacional que emplean y en su capacidad para incluir en el análisis información familiar (tríos padre-madre-hijo y dúos padre-hijo o madre-hijo).

4. Imputación de genotipos faltantes.

La imputación de genotipos faltantes tiene distintas aplicaciones en el contexto de los GWAS. En primer lugar, puede utilizarse para asignar genotipos en aquellos *loci* cuyos

genotipos observados no superaron el control de calidad realizado y fueron calificados como no asignados. En segundo lugar, pueden imputarse genotipos de marcadores que no fueron evaluados en el estudio, a través de la utilización de un panel de referencia. Asimismo, puede utilizarse la imputación con el fin de combinar resultados de dos o más estudios de genotipificación que implicaron diferentes marcadores, para esta aplicación también debe utilizarse un panel de referencia (Browning, 2008). En último lugar, en GWAS realizados con familias de individuos, cuando un miembro de la familia no está disponible para su genotipificación, puede imputarse su genotipo a partir de los genotipos de los restantes miembros de su familia. En este trabajo de tesis en particular se utilizará la imputación con el objetivo de estimar los genotipos faltantes (que no superaron el control de calidad o que no fueron genotipificados mediante el sistema de genotipificación de genes candidatos empleado en el Capítulo I) a partir de los genotipos de marcadores de tipo SNP cercanos y de las secuencias haplotípicas presentadas dentro y entre familias.

Los seis programas capaces de determinar fases haplotípicas mencionados anteriormente pueden también ser utilizados en la imputación de genotipos faltantes. La selección del más adecuado depende de la población a utilizar y de la localización de los marcadores cuyos genotipos se desea imputar, dado que los programas difieren en su capacidad para realizar la imputación de marcadores del cromosoma X, trabajar con individuos agrupados en familias y aceptar genotipos especificados sin un 100% de certeza (Marchini y Howie, 2010).

OBJETIVOS ESPECÍFICOS.

- Determinar los genotipos de los SNPs incluidos en un microarreglo de mediana densidad localizados en las regiones cromosómicas que contienen los genes candidatos seleccionados en el Capítulo I, en un rodeo comercial de razas Holando y cruza Holando x Jersey.
- Realizar un control de la calidad y un análisis estadístico descriptivo de los genotipos obtenidos.

MATERIALES.

1. Animales.

Se trabajó con un rodeo comercial de 998 bovinos de razas Holando, Jersey y cruza Holando x Jersey pertenecientes a 25 familias de medias hermanas paternas, de 16 tambos de la empresa Las Taperitas S.A., ubicados en la región centro de la provincia de Santa Fe.

La población estudiada consistió en 29 toros (24 de raza Holando y 5 de raza Jersey) y 969 vacas Holando y distintos grados de cruce Holando x Jersey.

Todas las vacas poseen registros genealógicos completos y datos fenotípicos para producción (Kg) y composición de leche (Kg y % de grasa y Kg y % de proteína) -una descripción de los datos fenotípicos puede hallarse en el Capítulo III-.

2. ADN.

La extracción del ADN utilizado, a partir de muestras de sangre de vacas y de semen de toros, así como también la determinación de su concentración y calidad, se realizaron según la forma descrita en la sección Materiales del Capítulo I de este trabajo de tesis.

MÉTODOS.

1. Genotipificación.

La determinación de los genotipos de toros y crías se realizó mediante el *chip* de mediana densidad BovineSNP50 v2 BeadChip (Illumina Inc., San Diego, California, EE.UU.). Se utilizaron 300 ng de ADN genómico de cada uno de los animales estudiados. Esta cantidad de ADN, proveniente de soluciones de concentración mínima 20 ng/ μ l y relaciones de absorbancias 260nm/280nm y 260nm/230nm mínimas de 1,6 y 1,7 respectivamente, fue colocada en los pocillos de microplacas de 96 posiciones y luego secada en un ciclador térmico Eppendorf Mastercycler durante 4 horas a 45°C. Las 11 placas con ADN seco obtenidas de esta manera fueron cubiertas con *film* y enviadas a la empresa GeneSeek (GeneSeek Inc., Lincoln, Nebraska, EE.UU.) para su posterior genotipificación mediante la plataforma de Illumina antes mencionada.

En el presente trabajo de tesis se trabajó con aquellos *loci* ubicados en bloques de genes candidatos, estando dichos bloques definidos por los SNPs pertenecientes a genes candidatos genotipificados en el Capítulo I, separados por 70 cM como máximo. En estos bloques se incluyeron los SNPs del *chip* contenidos entre los SNPs de genes candidatos, así como también aquellos flanqueantes de los SNPs de genes candidatos ubicados en los extremos de cada bloque en 10 cM hacia ambos lados.

2. Cálculo de factores descriptivos de los SNPs evaluados.

Con la información provista por la empresa GeneSeek acerca de la ubicación de los SNPs evaluados mediante la plataforma de genotipificación utilizada (de acuerdo al ensamblado UMD 3.1) y con los genotipos obtenidos para cada uno de los SNPs en la población analizada, se calcularon una serie de parámetros descriptivos de los SNPs contenidos en los bloques definidos previamente. Éstos consistieron en el número de SNPs de genes candidatos y aquellos pertenecientes al *chip* contenidos en cada uno de los bloques encontrados, la distancia promedio entre ellos y el porcentaje de heterocigosidad que presentaron los mismos en la población genotipificada.

Para facilitar la adquisición de la información requerida para estos cálculos, a partir de los archivos de resultados provistos por la empresa que realizó la genotipificación se creó una base de datos en MySQL (Ariel Amadio, comunicación personal).

En los casos en que fue posible, se localizaron los SNPs del *chip* con posición indefinida a través de la búsqueda de su código *rs* en la base de datos de acceso público Ensembl (www.ensembl.org).

3. Control de calidad de la genotipificación.

Con los objetivos de realizar un control de la calidad de la genotipificación y descartar aquellos genotipos poco confiables o incorrectamente asignados, se determinaron la proporción de SNPs y de muestras con genotipos asignados (tasa de genotipificación) y las

inconsistencias mendelianas entre los genotipos de los integrantes de cada familia (tríos madre-padre-hija o dúos padre-hija). Estas determinaciones se realizaron a través de las distintas aplicaciones del programa PLINK (Purcell y col., 2007).

Además, con el propósito de detectar cualquier intercambio ocurrido por error entre muestras de animales de distinto sexo se calculó la tasa de homocigosidad para los SNPs del cromosoma X y se la comparó con los valores esperados para machos y hembras. Para cada animal, esta tasa de homocigosidad fue calculada mediante la fórmula que se indica a continuación:

$$\text{Tasa de homocigosidad} = \frac{\text{Número de SNPs homocigotas en el cromosoma X}}{\text{Número de SNPs del cromosoma X con genotipo asignado}}$$

4. Comprobación de asignación de paternidades.

Con el objetivo de comprobar la consistencia genotípica entre padres e hijas se estimó la identidad por descendencia (IBD, *identity-by-descent*) existente entre pares de animales a través de la aplicación “--genome” del programa PLINK (Purcell y col., 2007).

5. Cálculo de factores descriptivos de los genotipos obtenidos.

A través de las aplicaciones “--freq” y “--hardy” del programa PLINK (Purcell y col., 2007) se determinó la frecuencia alélica mínima para cada SNP y se analizó la existencia de equilibrio de Hardy-Weinberg en la población estudiada, respectivamente. Se definieron luego valores umbrales para las tasas de genotipificación de las muestras y los SNPs, para las FAMs presentadas por los SNPs, así como para el valor de significación p del *test* exacto descrito e implementado por Wigginton y col. (2005), aplicado al comprobar desviaciones del EHW, de modo de decidir qué marcadores considerar en el análisis de asociación.

En la Figura II.2 se muestra una porción del archivo PED utilizado por el programa PLINK, con el objetivo de visualizar los genotipos que se consideran para la determinación de las tasas de genotipificación de SNPs y muestras.

Una vez descartados los SNPs y las muestras de ADN con genotipos de baja calidad y en el intento de dilucidar la estructura de la población en estudio, se determinó la existencia de subpoblaciones mediante el programa *Structure* v.2.3.4 (Pritchard y col., 2000; Falush y col., 2003). Esta determinación se realizó a partir de los genotipos correspondientes a 2.888 SNPs aproximadamente independientes (es decir, en equilibrio de ligamiento). Dichos SNPs fueron extraídos del conjunto de 44.124 SNPs genotipificados a lo largo de todos los cromosomas que superaron los umbrales de calidad preestablecidos (tasa de genotipificación = 0,9, FAM = 0,01 y p para EHW = 1.10^{-6}) y se obtuvieron mediante la aplicación “--indep-pairwise” del programa PLINK (Purcell y col., 2007).

Los parámetros del programa *Structure* utilizados se especifican a continuación:

- *Burnin length*: 10.000;
- *MCMC reps after burnin*: 10.000;
- *Admixture model*: considera que los animales pueden tener antecesores comunes;
- *LOCPRIOR*: este parámetro le permite al programa utilizar, además de la información genética, información adicional (en este caso, el porcentaje de raza Holando correspondiente a cada animal) en la determinación de la estructura poblacional;
- *Allele frequencies correlated*: este modelo asume que cada una de las subpoblaciones presentes en la muestra analizada han sufrido deriva genética a partir de la población ancestral en forma independiente;
- *Compute probability for the data*: permite la estimación del número de subpoblaciones presentes en la muestra analizada.

El programa se ejecutó con los parámetros especificados anteriormente para diferentes valores de K (número de subpoblaciones), evaluándose de 2 a 5 subpoblaciones.

Genotipos considerados
para el cálculo de la tasa de
genotipificación de un SNP

↓

28705	28705	122	6717	2	0	BB	AB	AB	AA	BB	BB	AB
29279	29279	114	8628	2	0	BB	BB	BB	AB	BB	AB	AB
29905	29905	121	20525	2	0	BB	AB	BB	AA	BB	BB	BB
30281	30281	127	14177	2	0	BB	AB	AB	AA	BB	BB	AB
30657	30657	112	18661	2	0	BB	BB	BB	AB	BB	BB	AB
30667	30667	128	18523	2	0	BB	BB	BB	AB	BB	BB	AB
30855	30855	112	12997	2	0	BB	BB	AB	AB	BB	BB	AA
31165	31165	126	17239	2	0	BB	BB	AB	AB	BB	BB	00
31211	31211	112	10369	2	0	AB	AB	AB	AB	BB	BB	AB
31341	31341	114	14946	2	0	BB	BB	AB	AB	BB	AB	AA
31631	31631	128	24480	2	0	BB	BB	BB	AA	BB	BB	00
31755	31755	127	22605	2	0	BB	BB	BB	AB	BB	BB	AB
32087	32087	126	21507	2	0	BB	BB	BB	BB	BB	AB	AB
32233	32233	122	7777	2	0	BB	BB	BB	AA	BB	BB	BB
32481	32481	126	22483	2	0	BB	BB	BB	BB	BB	AB	AB
32599	32599	126	24373	2	0	AB	AB	AB	BB	BB	AA	AB
32775	32775	121	5864	2	0	BB	AB	AB	AA	BB	BB	BB
32837	32837	126	23729	2	0	BB	AA	BB	AA	BB	BB	AA

Genotipos
considerados para el
cálculo de la tasa de
genotipificación de
una muestra →

Figura II.2. Esquema de los genotipos considerados para el cálculo de las tasas de genotipificación.

Porción del archivo PED utilizado por el programa PLINK, donde se indican algunos de los genotipos considerados para el cálculo de la tasa de genotipificación de un SNP (recuadro vertical azul) y algunos de los considerados para el cálculo de la tasa de genotipificación de la muestra de RP 31165 (recuadro horizontal rojo).

La primer columna corresponde a un identificador de familia, la segunda al código de identificación del animal (RP), la tercera al RP de su padre, la cuarta al RP de su madre, la quinta al código del sexo (2=hembra), la sexta al fenotipo (para estos cálculos no se incluyeron los fenotipos en el PED) y de la séptima columna en adelante se exponen los genotipos para cada SNP (2 alelos/SNP para cada animal, con codificación A/B y 0 para los genotipos no asignados).

Asimismo, con el objetivo de corroborar la estructura de la población hallada por *Structure*, se realizó la estimación de la identidad por estado (IBS, *identity-by-state*) existente entre los alelos de los animales tomados de a pares mediante las aplicaciones “*--genome*”, “*--read-genome*”, “*--cluster*” y “*--mds-plot4*” del programa PLINK (Purcell y col., 2007).

6. Determinación de haplotipos, asignación de fases e imputación de genotipos faltantes.

La determinación de los haplotipos y las fases presentes en cada bloque de genes candidatos (definidos en el punto 1) se realizó a través del programa BEAGLE (Browning y Browning, 2007; Browning y Browning, 2009).

Se determinaron las fases de los haplotipos correspondientes a los 922 animales (895 vacas y 27 toros) que superaron el control de calidad. Para ello se construyeron archivos con información genotípica de animales pertenecientes a tríos (padre-madre-hija) y dúos (padre-hija), así como archivos de animales genotipificados pero con padres no genotipificados. Asimismo, se creó un archivo por cada uno de los bloques de genes candidatos definidos en el punto 3. Por lo tanto, la información genotípica correspondiente a cada uno de los 15 bloques se organizó en tres archivos conteniendo 18 tríos, 864 dúos y 18 animales no pertenecientes a familias genotipificadas, respectivamente. En dichos archivos fueron incluidos, además, los genotipos para los 56 SNPs de genes candidatos obtenidos a través de la estrategia de genotipificación detallada en el Capítulo I de este trabajo de tesis.

Los tres archivos de genotipos (de tríos, dúos y animales no relacionados) correspondientes a cada uno de los bloques de genes candidatos fueron procesados simultáneamente por el programa BEAGLE, junto a un archivo con información sobre los marcadores utilizados (que incluía el nombre de los SNPs, su posición en pb y sus alelos posibles). Los argumentos utilizados en la línea de comando se especifican a continuación:

- *redundant=true*, para que los archivos de resultados contengan los dos haplotipos en fase correspondientes a cada animal.
- *niterations=20*, el número de iteraciones del algoritmo de asignación de fases se cambió de 10 (valor predeterminado por *default*) a 20 con el fin de aumentar la precisión en las determinaciones.
- *nsamples=4*, el número de pares de haplotipos a muestrear para cada animal en cada iteración del algoritmo de asignación de fases que se utilizó fue el valor predeterminado por *default*.

Simultáneamente, con el mismo programa, fueron imputados los genotipos faltantes (no asignados, de aquellos SNPs que no superaron el control de errores mendelianos y de los SNPs que no fueron genotipificados mediante el sistema de genotipificación SNPlex®).

RESULTADOS.

1. Descripción de la población genotipificada.

La genotipificación fue realizada sobre una población de 998 bovinos de razas Holando, Jersey y cruza Holando x Jersey pertenecientes a 25 familias de medias hermanas paternas, de 16 tambos de la cuenca lechera central de la provincia de Santa Fe.

La población estudiada consistió en 29 toros (24 de raza Holando y 5 de raza Jersey) y 969 vacas (703 de raza Holando y 266 cruza HxJ). El promedio de hijas por toro fue de 39, presentando un rango de variación de 1 a 125. A modo descriptivo de la población de vacas, a continuación se grafica el número de vacas analizadas que son hijas de cada toro (Figura II.3), el porcentaje en que están representados en la población la raza Holando y los distintos grados de cruce con Jersey (Figura II.4), la distribución de vacas según el año de nacimiento (Figura II.5) y su distribución en base al tambo de pertenencia, denominado “tambo de extracción” (Figura II.6).

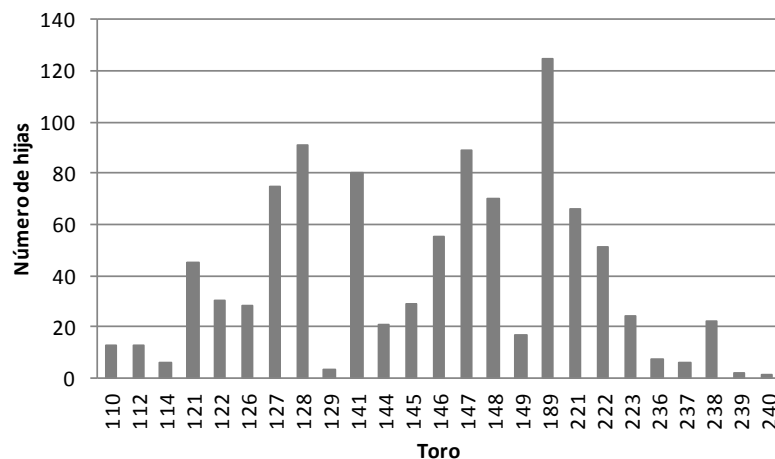


Figura II.3. Distribución de hijas entre los toros analizados.

La designación de los toros corresponde al RP propio de cada uno de ellos. En el gráfico se indica el número de vacas que son hijas de cada toro.

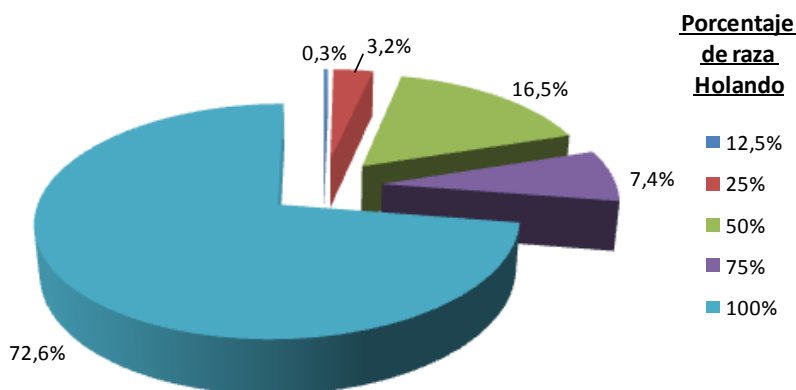


Figura II.4. Distribución de razas en las vacas analizadas.

La distribución de razas se expresa como porcentaje de raza Holando.

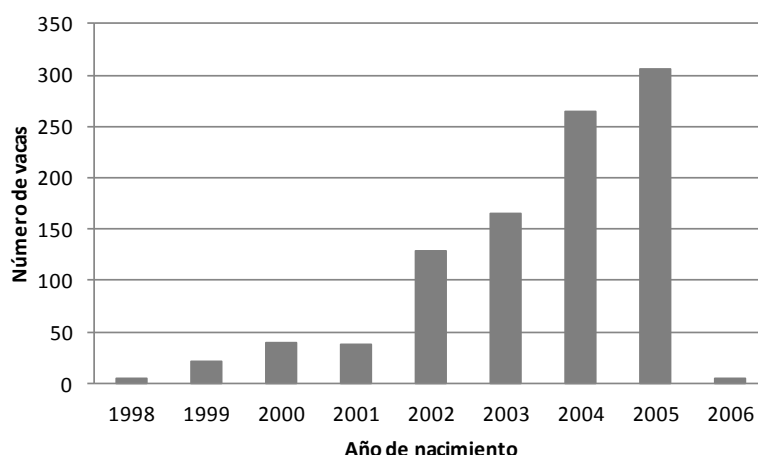


Figura II.5. Distribución de años de nacimiento de las vacas analizadas.

Número de vacas de la población analizada que nacieron entre los años 1998 y 2006.

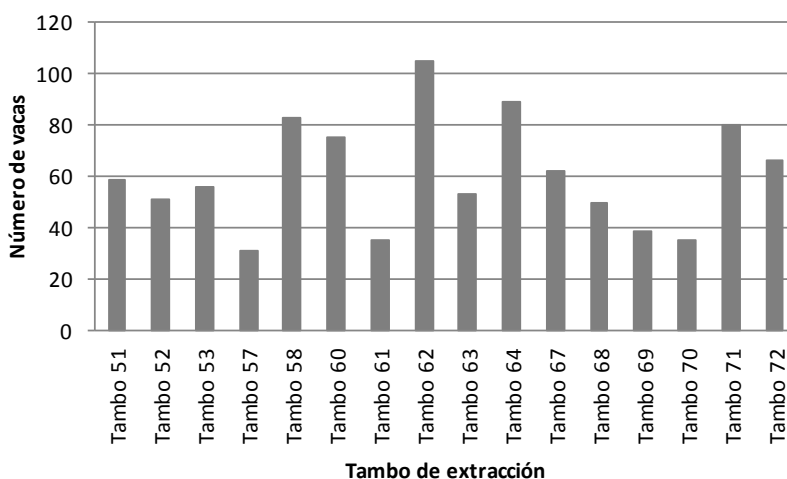


Figura II.6. Distribución de vacas según el tambo de extracción.

Número de vacas analizadas que pertenecen a cada uno de los 16 tambos evaluados.

2. Descripción de los SNPs incluidos en el *chip*.

El microarreglo de mediana densidad BovineSNP50 v2 BeadChip (Illumina Inc., San Diego, California, EE.UU.) evalúa 54.609 SNPs distribuidos en los 29 autosomas y en los cromosomas sexuales del genoma bovino, siendo el número de SNPs por cromosoma el que se indica en la Tabla II.1. Esta distribución corresponde a una distancia promedio entre SNPs de 50.114 pb a lo largo del genoma, con 21 SNPs por centiMorgan en promedio.

Los SNPs pertenecientes al BTA 0 son aquellos con posición indeterminada, según Illumina, en el ensamblado UMD 3.1.

BTA	Nº de SNPs	BTA	Nº de SNPs	BTA	Nº de SNPs
0	548	11	2295	22	1324
1	3431	12	1773	23	1092
2	2829	13	1850	24	1312
3	2550	14	1831	25	1004
4	2570	15	1762	26	1116
5	2271	16	1726	27	981
6	2575	17	1600	28	981
7	2352	18	1376	29	1087
8	2430	19	1420	X	1170
9	2095	20	1568	Y	1
10	2206	21	1483		

Tabla II.1. Número de SNPs por cromosoma, analizados por el BovineSNP50 v2 BeadChip®.

En primer lugar, previamente a la construcción de los bloques de SNPs, se intentó localizar los 548 SNPs sin localización definida. De éstos, fue posible ubicar sólo 12 a través de la búsqueda de los mismos mediante su código de identificación *rs* en la base de datos de la página web www.ensembl.org, utilizando el ensamblado UMD 3.1. Los 536 SNPs restantes, por no contar con código *rs*, no pudieron ser ubicados, por lo que no fueron utilizados en los análisis subsiguientes.

3. Descripción de los bloques de genes candidatos.

A partir de los SNPs genotipificados por el sistema SNPlex® se construyeron bloques de genes candidatos, es decir, regiones cromosómicas que contienen SNPs de genes candidatos relacionados con producción y composición de leche en bovinos, a los que se les sumaron SNPs evaluados mediante el *chip*. Para ser incluidos en un bloque, los SNPs del *chip* debían cumplir con la condición de estar localizados entre dos SNPs de genes candidatos separados por menos de 70 cM, o ser flanqueantes en 10 cM a los SNPs extremos de dichos bloques. En la construcción de los bloques se consideró la equivalencia 1 cM = 1.000.000 pb.

De esta forma se definieron 15 bloques pertenecientes a 14 cromosomas, los que contienen 12.510 SNPs, de los cuales 12.447 fueron genotipificados únicamente mediante el *chip*, 50 únicamente mediante el sistema SNPlex®, y 13 SNPs fueron genotipificados por ambos sistemas. En la Tabla II.2 se exponen la cantidad y las principales características de los bloques definidos en cada uno de los cromosomas que albergan genes candidatos. A modo de ejemplo, el bloque del BTA 4 contiene 2.000 SNPs, de los cuales 4 fueron genotipificados mediante el sistema SNPlex® y 1.996 mediante el *chip*, no habiendo en este caso SNPs genotipificados por ambos sistemas. Por otro lado, el bloque correspondiente al BTA 24 contiene 391 SNPs genotipificados mediante el *chip*, uno de los cuales, además, también fue genotipificado mediante el sistema SNPlex®.

De los 12 SNPs con posición indefinida que pudieron ser localizados en el genoma bovino a través de su código *rs*, 4 pertenecen a los bloques definidos previamente.

La determinación de los SNPs localizados en las regiones cromosómicas que definen cada uno de los bloques a partir de la totalidad de SNPs analizados por el *chip* fue facilitada a través de consultas realizadas en el sistema de gestión de bases de datos MySQL.

BTA	Ubicación de bloques (pb)	Nº de SNPs	Nº de SNPs		Distancia promedio entre SNPs (pb)	Nº promedio de SNPs/cM	Porcentaje de heterocigosidad promedio
			del SNPlex	del <i>chip</i>			
1	135.098-67.191.387	1.391	2	1391	48.242	21	30,7
	134.622.487-154.567.741	490	1	490	40.788	25	30,2
4	10.324.170-103.287.645	2.000	4	1996	46.505	22	29,0
5	18.726.210-110.237.957	1.617	5	1613	56.629	18	29,0
6	23.408.172-97.374.303	1.500	19	1484	49.344	20	30,5
11	76.488.136-107.246.475	630	5	625	48.900	20	29,9
14	1.435.005-52.150.706	1.179	9	1173	43.052	23	31,1
15	36.410.640-56.383.289	396	1	395	50.564	20	30,4
19	33.112.218-61.395.535	686	6	680	41.290	24	32,0
20	21.938.158-49.111.029	538	3	536	50.601	20	29,0
21	49.706.980-69.650.321	420	2	419	47.597	21	29,2
22	43.569.849-61.378.199	435	2	433	41.033	24	29,9
23	25.109.188-45.106.522	399	2	397	50.245	20	30,9
24	12.099.822-32.057.788	391	1	391	51.174	20	30,0
26	11.201.198-31.049.759	438	1	437	45.420	22	29,1

Tabla II.2. Bloques de genes candidatos.

Ubicación de los bloques definidos en cada uno de los cromosomas que albergan genes candidatos, cantidad y origen de los SNPs que contienen, distancia promedio entre los mismos, número promedio de SNPs/cM y porcentaje de heterocigosidad promedio en cada bloque. 1 cM = 1.000.000 pb.

4. Control de calidad de la genotipificación.

Tasa de genotipificación (*Call rate*).

Todos los bloques, a excepción del correspondiente al BTA 24, presentan **SNPs** con una **tasa de genotipificación** que varió entre 0 y 1. Por el contrario, los SNPs contenidos en el bloque del BTA 24 tomaron valores entre 0,69 y 1. El porcentaje de SNPs con tasa de genotipificación > 0,9 en cada uno de los bloques supera el 94% (Tabla II.3).

Bloque	Porcentaje de SNPs	Bloque	Porcentaje de SNPs
1.A	94,1	19	96,8
1.B	98,4	20	95,0
4	95,3	21	97,6
5	96,3	22	98,6
6	95,2	23	95,2
11	97,1	24	94,9
14	96,7	26	97,0
15	97,0		

Tabla II.3. Porcentaje de SNPs con tasa de genotipificación > 0,9 en cada bloque.

Por otro lado, la **tasa de genotipificación para las muestras** varió entre 0,745 y 0,996 y el 97,6% de las muestras (975 animales -26 toros y 949 vacas-) presentaron valores $\geq 0,9$. Estas tasas de genotipificación fueron calculadas considerando los genotipos asignados para cada una de las muestras únicamente en los *loci* correspondientes a los SNPs pertenecientes a los bloques definidos en el punto 3.

Información relacionada al sexo.

La tasa de homocigosidad para los SNPs del cromosoma X presentada por los 29 toros genotipificados se halló en el rango 0,81 - 0,99, siendo su promedio de 0,96. Este valor indica una concordancia entre el sexo determinado por el genotipo y el registrado para estos animales, puesto que se espera que los machos presenten una tasa de homocigosidad = 1.

Por otro lado, las 969 vacas genotipificadas presentaron una tasa de homocigosidad para los SNPs del cromosoma X en el rango 0,64 - 0,99, siendo su promedio de 0,73. Este valor no se corresponde con el esperado según el trabajo de Anderson y col. (2010) para individuos de sexo femenino (tasa de homocigosidad < 0,2).

Errores mendelianos.

La determinación de errores mendelianos en los genotipos asignados a los 994 animales pertenecientes a familias constituidas por tríos madre-padre-hija o dúos padre-hija se realizó utilizando la información genotípica disponible para los 54.073 SNPs con posición definida evaluados por el *chip*. No se utilizaron en este caso los SNPs pertenecientes a bloques únicamente, para evitar la pérdida de información que implicaría no considerar más del 75% de los SNPs genotipificados.

Se obtuvieron 700.994 errores mendelianos distribuidos en todos los cromosomas y las familias genotificadas. Es decir, el 1,3% de los genotipos, dentro de los 53.748.562 genotipos posibles, presentaron inconsistencias genotípicas entre padres e hijas.

El análisis de los archivos de resultados obtenidos condujo a la observación de la existencia de entre 0 y 10.301 errores mendelianos por animal. Se calculó entonces la frecuencia de aparición de dichos errores en la población analizada y se graficó esta frecuencia vs. el número de errores por animal, observándose la distribución que se indica en la Figura II.7. A partir de esta distribución se estableció empíricamente un valor umbral de 70 para la variable “número de errores mendelianos/animal” por debajo del cual los errores detectados fueron calificados como genotipos no asignados. Todos aquellos animales (n=226) que poseían un número de errores mendelianos superior a este umbral fueron analizados con mayor profundidad (ver siguiente sección “Comprobación de la asignación de paternidades”).

Antes de calificar como no asignados los genotipos con inconsistencias mendelianas correspondientes a los animales con menos de 70 errores, se analizó, para cada SNP y cada familia genotificada, si existía algún toro con más del 50% de sus hijas con errores mendelianos en el mismo SNP. En caso de cumplirse esta condición, se calificó como no asignado el genotipo del toro en ese SNP en lugar de los de las hijas.

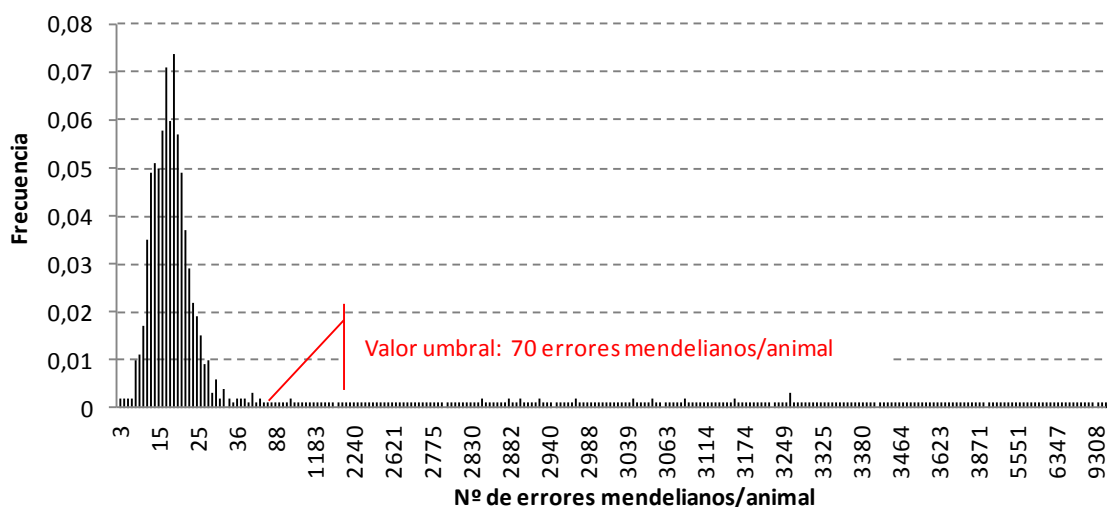


Figura II.7. Distribución de errores mendelianos en la población genotificada.

En la figura se indica el valor umbral considerado en el análisis (ver texto).

5. Comprobación de la asignación de paternidades.

Los números de errores mendelianos superiores al umbral establecido en el punto anterior sugerían la posibilidad de que no existiera relación de parentesco entre la vaca que presentaba esa cantidad de errores y su supuesto padre. Por tal motivo se decidió comprobar la consistencia genotípica entre padres e hijas a través de la estimación de la identidad por descendencia (IBD) existente entre los alelos de los animales tomados de a pares. Dichos pares de animales estaban compuestos por las 226 vacas que presentaban más de 70 errores

mendelianos y cada uno de los 29 toros utilizados normalmente en las inseminaciones que tienen lugar en el establecimiento al que pertenecen los animales en estudio.

Se consideraron como toros más probables de ser los padres de cada una de las vacas analizadas aquellos que presentaban el menor valor de probabilidad de que ningún alelo sea idéntico por descendencia en el archivo de resultados obtenido con el programa PLINK. Mediante este procedimiento, se les reasignó el padre a 202 vacas, es decir, al 88,5% de las vacas que presentaron más de 70 errores mendelianos.

Con las paternidades reasignadas se detectaron nuevamente los errores mendelianos con la misma metodología, obteniéndose esta vez 152.630 errores. Al analizar los archivos de resultados se observó que 80 vacas aun presentaban más de 70 errores. Entre ellas existían 13 dúos madre-hija y 48 hijas del toro 110 (muchas de éstas figuraban como hijas de otros toros en los registros genealógicos, pero debido a lo expuesto en el párrafo anterior se les había reasignado el padre).

Cada dúo madre-hija, correspondiente a una familia nuclear madre-padre-hija, fue analizado independientemente de modo de determinar, según los códigos de los errores mendelianos presentados, si éstos se debían mayoritariamente a inconsistencias genéticas entre madre e hija o entre padre e hija. En todos los casos se observó que la mayoría de los errores se debían a inconsistencias genéticas entre madre e hija, por tal motivo se decidió desvincular a estas vacas de sus madres, es decir, ignorar su supuesta relación de parentesco para realizar los cálculos siguientes.

Por otro lado, se decidió eliminar a la familia del toro 110 del análisis posterior, es decir, no se consideró al toro 110 (que presentaba una tasa de genotipificación $< 0,90$) ni a sus hijas con más de 70 errores mendelianos (48 hijas) para realizar las determinaciones posteriores. Asimismo, el resto de las vacas que presentaron una cantidad de errores mendelianos superior al umbral establecido (6 vacas) tampoco fueron consideradas para las siguientes estimaciones.

Luego de las correcciones, se repitió el análisis con los genotipos de 939 animales (915 vacas y 24 toros) para 54.073 SNPs. Esta vez se obtuvieron 16.616 errores mendelianos, lo que corresponde al 0,03% de los genotipos posibles.

6. Cálculo de factores descriptivos de los genotipos obtenidos.

6.A. Frecuencia alélica mínima y equilibrio de Hardy-Weinberg.

Los factores descriptivos fueron calculados para todos aquellos SNPs pertenecientes a bloques de genes candidatos cuya tasa de genotipificación superara el valor 0,9 (11.960 SNPs). Asimismo, se estableció un valor mínimo de 0,9 para la tasa de genotipificación de los animales para ser incluidos en el mismo cálculo. Sin embargo, se hizo una excepción y se consideró para los cálculos al toro 221 a pesar de presentar una tasa de genotipificación de 0,82. Se tomó esta decisión por ser el 221 un toro de raza Jersey que posee un alto número de hijas

genotipificadas (65 por el *chip* y 124 por el sistema SNPlex®). Se incluyeron, entonces, 976 animales en los cálculos.

Las FAMs obtenidas presentaron valores entre 0 y 0,5, siendo su promedio de 0,22. El valor umbral elegido para este parámetro fue de 0,01, lo que determinó que fueran eliminados los 1.747 SNPs que presentaron valores inferiores a este umbral, permaneciendo entonces 10.213 SNPs para ser considerados en análisis posteriores.

La determinación de la desviación del equilibrio de Hardy-Weinberg se realizó utilizando un valor umbral de 1.10^{-6} para el nivel de significación p obtenido bajo la hipótesis nula de que la población tiene proporciones de Hardy-Weinberg. Esto determinó que fueran eliminados los 31 SNPs que presentaron valores inferiores a dicho umbral, permaneciendo entonces 10.182 SNPs para ser considerados en análisis posteriores.

En la Tabla II.4 se indica el número de SNPs pertenecientes a cada uno de los bloques de genes candidatos que fueron descartados por presentar valores inferiores a los umbrales establecidos para tasa de genotipificación, FAM y nivel de significación p en la comprobación de la desviación del EHW.

Bloque	Tasa de genotipificación	FAM	EHW	SNPs remanentes
1.A	82 (5,9%)	163 (12,5%)	7 (0,6%)	1139 (81,9%)
1.B	8 (1,6%)	64 (13,3%)	2 (0,5%)	416 (84,5%)
4	93 (4,7%)	294 (15,4%)	2 (0,1%)	1607 (80,5%)
5	60 (3,7%)	266 (17,1%)	6 (0,5%)	1281 (79,4%)
6	71 (4,8%)	199 (14,1%)	2 (0,2%)	1212 (81,7%)
11	18 (2,9%)	89 (14,7%)	1 (0,2%)	517 (82,7%)
14	39 (3,3%)	164 (14,4%)	2 (0,2%)	968 (82,5%)
15	12 (3,0%)	47 (12,3%)	2 (0,6%)	334 (84,6%)
19	22 (3,2%)	95 (14,4%)	1 (0,2%)	562 (82,6%)
20	27 (5,0%)	64 (12,6%)	1 (0,2%)	444 (82,8%)
21	10 (2,4%)	73 (17,8%)	0 (0%)	336 (80,2%)
22	6 (1,4%)	75 (17,6%)	1 (0,3%)	351 (81,1%)

Tabla II.4. Número de SNPs eliminados según distintos filtros y SNPs remanentes.

Número de SNPs pertenecientes a cada uno de los bloques de genes candidatos que fueron removidos y no considerados para el análisis posterior por no superar los umbrales establecidos de tasa de genotipificación, FAM y nivel de significación en los tests que analizan la desviación del EHW. Junto al número de SNPs eliminados se indica el porcentaje que representan dichos SNPs del total de SNPs contenidos inicialmente en cada bloque, para cada una de las etapas de control de calidad realizadas. Asimismo, se expone el número de SNPs de cada bloque que superaron el control de calidad y serán considerados en futuros análisis, así como el porcentaje que éstos representan del total de SNPs iniciales en cada bloque. La última fila de la tabla indica los valores totales, sumados a lo largo de cada una de las columnas.

Valores umbral aplicados: tasa de genotipificación =0,9; FAM =0,01; p para EHW = 1.10^{-6} .

Bloque	Tasa de genotipificación	FAM	EHW	SNPs remanentes
23	19 (4,8%)	46 (12,2%)	2 (0,6%)	330 (83,1%)
24	20 (5,1%)	50 (13,5%)	1 (0,3%)	320 (81,8%)
26	13 (3,0%)	58 (13,7%)	1 (0,3%)	365 (83,5%)

Tabla II.4. Número de SNPs eliminados según distintos filtros y SNPs remanentes (continuación).

La proporción de SNPs conservados luego de realizar el control de calidad de la genotipificación es algo superior a la publicada en literatura, lo que es esperable dados los bajos umbrales establecidos para la FAM y el nivel de significación p del test utilizado al comprobar la existencia de EHW en la población.

6.B. Estructura poblacional.

La estructura de la población estudiada se analizó a través de la determinación de la existencia o no de subpoblaciones. Los resultados obtenidos permitieron inferir la existencia de cuatro grupos o subpoblaciones dentro de la población analizada. En la Figura II.8 se representan los coeficientes de pertenencia a cada una de las subpoblaciones, estimados para cada uno de los animales. En dicha figura, cada animal se representa mediante una línea vertical particionada en K segmentos coloreados que indican la fracción de pertenencia a cada una de las K subpoblaciones inferidas, estimada para cada individuo.

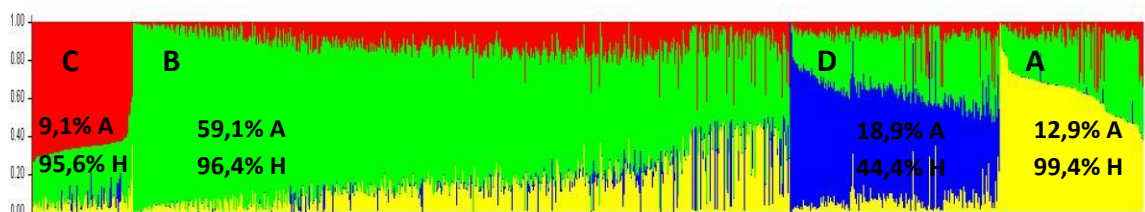


Figura II.8. Subpoblaciones halladas por el programa *Structure*.

Fracción de pertenencia de cada animal a cada una de las poblaciones (A, B, C, D) halladas por el programa *Structure*. Se indican el porcentaje de los animales evaluados pertenecientes a cada una de las subpoblaciones (% A) y el porcentaje de raza Holando promedio calculado para cada subpoblación (% H) en base al pedigrí de los animales.

En paralelo se realizó la estimación de la identidad por estado (IBS) existente entre los alelos de los animales tomados de a pares. La probabilidad IBS también se calculó para determinar la estructura poblacional en base, en este caso, al promedio de la proporción de alelos compartidos en los *loci* genotipificados entre todos los pares posibles de animales con tasa de genotipificación > 0,9. La aplicación “--mds-plot4” utilizada consiste en un análisis multivariado de descomposición de la varianza en cuatro componentes, de los cuales C1 y C2 son los que absorben el mayor porcentaje de variabilidad. Los valores de C1 y C2 fueron

conservados para realizar un gráfico C1 vs. C2 en el que se distinguen con diferentes colores las cuatro subpoblaciones definidas por el programa *Structure* (Figura II.9).

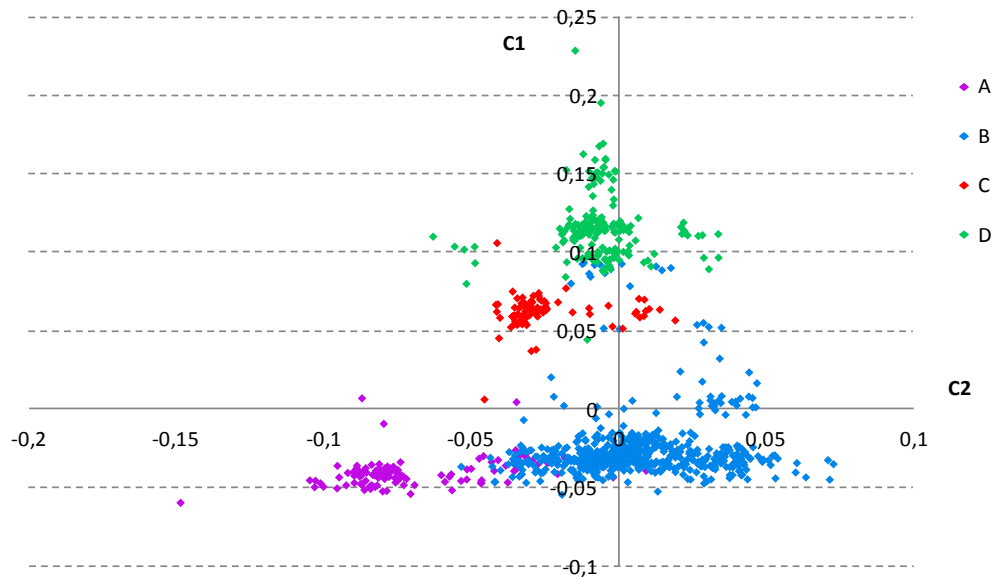


Figura II.9. Subpoblaciones halladas.

Valores estimados para C1 vs. C2 asignados por el programa PLINK a cada animal analizado. En colores se indican las cuatro subpoblaciones definidas por el programa *Structure*.

7. Determinación de haplotipos, asignación de fases e imputación de genotipos faltantes.

La imputación de genotipos faltantes se realizó en forma simultánea a la asignación de fases haplotípicas. Se imputaron de esta forma 90.649 genotipos faltantes, que representan el 0,96% de los genotipos correspondientes a 922 animales y 10.227 SNPs.

De los 10.227 SNPs de bloques de genes candidatos considerados en la determinación de fases haplotípicas y en la imputación de genotipos faltantes, 10.169 fueron genotificados únicamente a través del *chip* de mediana densidad, 45 únicamente a través del sistema de genotipificación SNPlex® y 13 mediante ambos sistemas.

DISCUSIÓN.

El propósito de la genotipificación con una densidad media de marcadores distribuidos por todo el genoma bovino fue obtener los genotipos correspondientes a todos aquellos *loci* ubicados dentro o en las proximidades de genes candidatos relacionados con producción y calidad de leche, para complementar la información genotípica obtenida en el Capítulo I.

A partir de las posiciones en el genoma bovino de los SNPs de genes candidatos genotipificados en el capítulo anterior se definieron los límites de 15 bloques de genes candidatos distribuidos en 14 cromosomas. A 4 SNPs genotipificados por el *chip* contenidos en dichos bloques fue posible asignarles una localización precisa en base a su código de identificación *rs*, a pesar de que en los archivos provistos por la empresa que realizó la genotipificación los mismos no contaban con una ubicación definida en el genoma bovino. Los 536 SNPs sin localización definida según los archivos provistos por Illumina y que no contaban con código *rs* no fueron utilizados en los análisis subsiguientes. Cada uno de los bloques fue caracterizado en base a su longitud en pares de bases, a la cantidad de SNPs pertenecientes a genes candidatos y al *chip* utilizado que comprendía, al porcentaje de heterocigosidad promedio de dichos SNPs y a la distancia media entre los mismos. Se obtuvieron bloques de varias Mb de longitud, como el correspondiente al BTA 4, de 92,96 Mb, que alberga 2.000 SNPs, y otros pequeños, como el del BTA 22, de 17,81 Mb de longitud, con sólo 435 SNPs. Para todos los bloques, el porcentaje de heterocigosidad media y la distancia media a la que se encontraban los SNPs fue muy similar, lo que indica una distribución uniforme de genotipos y de marcadores, respectivamente.

La existencia de errores en la asignación del género de los animales o bien el intercambio entre muestras de animales de distinto sexo durante el procesamiento de las mismas fueron chequeados a través del cálculo del porcentaje de homocigosidad presentado por los SNPs pertenecientes al BTA X. Este tipo de errores puede tener lugar en muestras marcadas como masculinas que presenten un exceso de heterocigosidad o en muestras marcadas como femeninas que muestren un exceso de homocigosidad en dichos marcadores (Qu y col., 2011). La tasa de homocigosidad presentada para los SNPs del BTA X por los toros genotipificados se halló dentro de los valores esperados para machos (cerca a 1). Este valor indica una concordancia entre el sexo determinado por el genotipo y el registrado para estos animales, ya que se espera que los machos presenten una tasa de homocigosidad = 1. Sin embargo, para el caso de las vacas estudiadas, la tasa de homocigosidad media para estos marcadores fue superior a la esperada (0,73 vs. 0,20 o inferior según Anderson y col., 2010). Esta diferencia puede deberse a que los valores esperados para la tasa de homocigosidad de SNPs del cromosoma X son altamente dependientes de la especie y la subpoblación analizada. Es probable, entonces, que las tasas presentadas por el ganado lechero no se correspondan con las esperadas para humanos según Anderson y col. (2010). De hecho, el tamaño de la población efectiva en ganado lechero es mucho menor que para humanos y, por lo tanto, las tasas de homocigosidad esperadas deberían ser mucho mayores tanto para el cromosoma X como para los autosomas (Carl A. Anderson, comunicación personal). Sumado a lo anterior, distintos trabajos han reportado la obtención de resultados ambiguos al utilizar únicamente la

heterocigosidad de los marcadores localizados en el cromosoma X en el intento de inferir la identidad de género (Laurie y col., 2010; Qu y col., 2011). Por este motivo dichos trabajos proponen utilizar, junto a la heterocigosidad presentada por los SNPs del cromosoma X, información relacionada con la intensidad que alcanzaron. Probablemente esta información adicional brinde una estimación más precisa del sexo de los animales estudiados.

Los SNPs analizados fueron sometidos a un control de calidad de la genotipificación con el objetivo de descartar aquellos SNPs que presentaron alta probabilidad de haber sufrido errores en el proceso de genotipificación o en la asignación de genotipos. El control de calidad mencionado incluyó la detección de errores mendelianos en las familias constituidas por madre-padre-hija o padre-hija, según los datos genotípicos de los que se disponía. La primera determinación condujo a la detección de 700.994 errores mendelianos, lo que representa el 1,3% de los genotipos. Luego de analizar la asignación de paternidades a través de la estimación de la probabilidad IBD existente entre los alelos de los animales tomados de a pares, se les reasignó el padre a 202 vacas, asignándoles como padre el toro que presentó el menor valor de que ningún alelo sea IBD. La reasignación de padres, la desvinculación de 13 vacas de sus madres (que presentaban más de 70 errores mendelianos) ignorando su supuesta relación de parentesco y la eliminación de los animales con un número de errores mendelianos superior al umbral establecido condujo a la reducción del número de dichos errores, detectándose finalmente sólo 16.616 errores mendelianos (valor que representa el 0,03% de los genotipos). Estos genotipos que presentaron inconsistencias familiares se clasificaron como no asignados y fueron luego imputados mediante el programa BEAGLE.

En este trabajo, por lo tanto, se les reasignaron los padres al 20,8% de las vacas analizadas. Este porcentaje se corresponde con los valores reportados en literatura, dado que la tasa de identificación errónea de padres en distintas poblaciones de bovinos a nivel mundial llega hasta casi un 25% (Banos y col., 2001). Ron y col. (1996) reportaron una tasa del 5,2% para la población Holstein de Israel; en la misma población, Weller y col. (2004) reportaron una tasa del 11,7% de paternidades erróneamente asignadas; Visscher y col. (2002) hallaron errores de pedigrí en aproximadamente el 10% de la población del Reino Unido analizada mediante microsatélites; Beechinor y Kelly (1987) hallaron un 8% y un 20% de registros de paternidad errónea en terneros hembras y machos de Irlanda, respectivamente, usando polimorfismos de grupos sanguíneos y proteicos; mientras que Geldermann y col. (1986) reportaron tasas del 4% al 23% en Alemania. Aunque algunos de los valores reportados son bajos en relación al obtenido en el presente trabajo de tesis, hay que tener en cuenta que la probabilidad de detectar paternidades incorrectamente asignadas es función del polimorfismo de los *loci* genotipificados. Debido al bajo polimorfismo que presentan los marcadores tradicionales como los grupos sanguíneos y las proteínas de la leche, las estimaciones realizadas mediante estos marcadores subestimaban las verdaderas tasas de identificación incorrecta de padres (Ron y col., 1996). Con el advenimiento de la PCR y el uso de marcadores de mayor polimorfismo como los microsatélites y los SNPs, la probabilidad de detectar errores de paternidad aumenta, aún cuando se analizan dúos padre-hijo en lugar de tríos madre-padre-hijo (Ron y col., 1993). Christensen y col. (1982) enunciaron una serie de razones para explicar los errores en los registros de paternidades, entre las que se encuentran las

siguientes: errores en el rotulado de las pajuelas de semen, inseminación de vacas preñadas por una inseminación previa, errores al ingresar en el registro de inseminación el número o nombre correspondiente al toro utilizado, e intercambio de terneros en el establecimiento ganadero al momento del nacimiento.

Según Weller y col. (2004), la segunda razón enunciada podría explicar no más de un 20% de los casos de asignación errónea de paternidad, mientras que tanto la primera como la tercera razón serían las que explicarían la mayor parte de los errores de pedigrí detectados. En nuestro caso, consideramos que cualquiera de las cuatro razones enunciadas podría explicar el valor hallado. Siempre que sea detectado un error en la asignación de paternidades debe intentar corregirse puesto que estos errores pueden interferir en la estimación de parámetros genéticos como la heredabilidad y en la de los valores de cría de los animales, resultando en valores sesgados.

Otros controles realizados a los genotipos obtenidos consistieron en el cálculo de la tasa de genotipificación que presentaron las muestras y los SNPs, la determinación de la frecuencia alélica mínima de cada SNP y el análisis de la existencia o no de equilibrio de Hardy-Weinberg en la población estudiada. Los valores umbrales establecidos para cada uno de estos parámetros se correspondieron con los publicados a nivel internacional. Todas aquellas muestras y SNPs con tasa de genotipificación superior a 0,9 permanecieron en el análisis, siendo este umbral el utilizado en varios estudios reportados en literatura (Jiang y col., 2010; Hayes y col., 2009). El valor umbral elegido para la FAM fue de 0,01; es decir, los SNPs que presentaron una FAM menor a este valor no se consideraron en los análisis posteriores. Se estableció un valor bajo para este parámetro para conservar en el análisis a aquellos SNPs que presentaran alelos con baja frecuencia en la población, dado que estos alelos podrían ser responsables de la variación en algún rasgo fenotípico de interés. Si se hubiese utilizado un umbral más alto para la FAM como 0,02 o 0,03 (reportados, entre otros, por Bolormaa y col. (2010) y Jiang y col. (2010), respectivamente) se hubiesen descartado variantes alélicas raras con potenciales asociaciones al rasgo estudiado.

Por último, el valor umbral establecido para el nivel de significación p obtenido bajo la hipótesis nula de que la población tiene proporciones de Hardy-Weinberg fue de 1.10^{-6} . Dicho valor es el más bajo de los reportados en bibliografía para este parámetro (Jiang y col., 2010). Fue elegido para descartar los SNPs que más se aparten del EHW dado que esto sugiere la existencia de un error en la genotipificación o en la asignación de genotipos (Anderson y col., 2010). No se estableció un umbral mayor puesto que, como se mencionó en el Capítulo I, la población estudiada no se ajusta a los postulados básicos del principio de Hardy-Weinberg (los cruzamientos no se dan al azar, hay migración y selección a favor de un fenotipo particular), por lo que no es esperable que todos los SNPs se hallen en EHW.

Luego de la eliminación de SNPs en base a los distintos filtros mencionados previamente, se mantuvieron para ser considerados en el análisis posterior 10.182 SNPs (81,7%) evaluados mediante el *chip*, de los 12.460 incluidos en los bloques de genes candidatos. La proporción de SNPs conservados luego de realizar el control de calidad de la genotipificación es algo superior a la publicada en estudios de asociación a nivel de genoma completo en ganado vacuno,

siendo esta última de alrededor del 70%. Hayes y col. (2009) conservaron el 67,2% de los SNPs inicialmente genotipificados para el cálculo de valores de cría a ser utilizados en la implementación de metodologías de selección genómica en Australia. Utilizando el mismo microarreglo de 56.947 marcadores que el grupo de investigación mencionado previamente, VanRaden y col. (2009) conservaron el 67,5% de los SNPs, Wiggans y col. (2009) el 71,8% y Harris y col. (2008) el 77,5%. Otros autores, luego de aplicar distintos filtros en el proceso del control de calidad de la genotipificación con microarreglos de mediana densidad, mantuvieron similares porcentajes de marcadores: 69,7%; 72,5% y 73,3% de los SNPs genotipificados (Bolormaa y col., 2010; Jiang y col., 2010; Pimentel y col., 2010; respectivamente). El valor obtenido en este trabajo de tesis fue algo superior a los enumerados previamente debido a los bajos umbrales establecidos para la FAM y el nivel de significación p del *test* utilizado al comprobar la existencia de EHW en la población, lo que evitó que se descartaran demasiados SNPs para el análisis posterior.

El análisis de la estructura de la población estudiada (o “estratificación poblacional”) reveló la existencia de cuatro subpoblaciones, dado que $K=4$ resultó ser el menor valor de K que capturó la presencia de estructura a partir los datos analizados. Una de ellas se diferencia de las otras tres por presentar, en promedio, el menor porcentaje de raza Holando, siendo éste de 44,4%, mientras que las otras subpoblaciones presentan valores superiores al 95%. El programa utilizado para inferir la estructura poblacional implementa un modelo basado en un método de agrupamiento en base a datos genotípicos de marcadores no ligados, asumiendo que existen K subpoblaciones, cada una de ellas con frecuencias alélicas características en cada *loci*. Los individuos de la muestra analizada son asignados probabilísticamente a cada subpoblación según las frecuencias alélicas que presenten en cada *loci*. El programa también asume que dentro de cada subpoblación los *loci* están en equilibrio de Hardy-Weinberg. En la Figura II.9, tres de las subpoblaciones graficadas no se hallan bien delimitadas entre sí y, a su vez, los animales incluidos en ellas presentan un porcentaje de raza Holando similar. Esto puede deberse a la existencia de animales emparentados en la población analizada, debido a que múltiples miembros de una familia en la muestra viola las asunciones del modelo y puede conducir a una sobreestimación del número de subpoblaciones presentes, especialmente al utilizar el modelo de frecuencias alélicas correlacionadas. La existencia de estratificación poblacional debe chequearse y ser tomada en cuenta en todo análisis de asociación entre datos genotípicos y registros fenotípicos, dado que la ocurrencia de distintas frecuencias alélicas características de la composición racial de las diferentes subpoblaciones puede conducir a asociaciones erróneas si todos los animales son analizados sin tener en cuenta su origen (Pearson y Manolio, 2008).

La selección del programa a utilizar en la determinación de fases haplotípicas y en la imputación de genotipos faltantes se basó en la capacidad del mismo de incluir información genealógica, es decir, en la posibilidad de organizar la población analizada en familias compuestas por dúos o tríos de animales. Se le dio especial importancia a este aspecto puesto que los animales con los que se trabajó contaban con información de pedigrí, estando la mayor parte de ellos asentados en los registros de la Asociación de Criadores de Holando Argentino (ACHA). Además, la mayoría de los animales genotipificados podían organizarse en familias

compuestas por dúos o tríos. De los seis programas mencionados en la introducción de este capítulo, sólo DualPHASE y BEAGLE brindaban dicha posibilidad. El procesamiento de los genotipos obtenidos previamente mediante el programa BEAGLE condujo a la estimación de las dos fases haplotípicas correspondientes a cada uno de los animales, así como a la imputación de los genotipos faltantes, que representaban el 0,96% de los genotipos de SNPs correspondientes a la población analizada.

A partir de la descripción de los pasos seguidos para realizar el control de calidad de los genotipos obtenidos, se puede observar lo laborioso que resulta construir una base de genotipos confiables, dado el tiempo que implica su obtención y la cantidad de variables a tener en cuenta a la hora de analizar posibles errores en la asignación de genotipos o en la estructura familiar. Asimismo, se debe tener presente que en muchos casos los programas utilizados para el análisis de datos asumen cosas que no siempre se cumplen en los datos reales. Es importante recordar esto al momento de elegir el programa a utilizar, de modo de escoger aquél que mejor se adecúe a los datos disponibles, así como también al hacer interpretaciones a partir de los resultados obtenidos luego de su aplicación, para no extraer conclusiones erróneas.

Los genotipos resultantes de las metodologías de obtención y selección aplicadas en este capítulo, correspondientes a *loci* de genes candidatos (o cercanos a los mismos) complementan la información genotípica obtenida en el Capítulo I y fueron sometidos a un análisis de asociación a caracteres fenotípicos, cuyos resultados se muestran en el próximo capítulo de este trabajo de tesis.

CAPÍTULO III

**ANÁLISIS DE ASOCIACIÓN CON PRODUCCIÓN DE LECHE
ACUMULADA A LOS 305 DÍAS**

En el tercer capítulo de este trabajo de tesis se hará un análisis descriptivo de los datos fenotípicos de producción de la leche bovina y posteriormente se buscarán asociaciones entre la información genotípica obtenida en los capítulos anteriores y los registros fenotípicos de la primera lactancia de las vacas genotipificadas.

El capítulo comienza con una introducción en la que se describen las principales características de las lactancias del ganado bovino, las metodologías empleadas para estimar la producción de leche acumulada a 305 días de lactancia, los efectos a tener en cuenta en la construcción de modelos para la estimación de los valores de cría y los análisis de asociación más utilizados en la búsqueda de variantes alélicas asociadas a un determinado fenotipo. Luego continúa con la descripción de los materiales y métodos utilizados en la aplicación de modelos que ajusten los datos fenotípicos disponibles y en el estudio de asociación con la información genotípica obtenida previamente. Seguidamente se exponen los resultados obtenidos en las etapas de utilización de registros fenotípicos, previas al estudio de asociación, así como también aquellos del estudio de asociación en sí mismo. El capítulo finaliza con una discusión de los resultados obtenidos y las conclusiones extraídas a partir de los mismos.

INTRODUCCIÓN.

1. Curvas de lactancia del ganado bovino.

1.A. Definición y características generales.

Las curvas de lactancia describen la evolución a lo largo del tiempo de la producción de leche de un animal. En el ganado bovino la lactancia tipo o *standard* (Figura III.1) tiene una duración de 305 días, con 60 días de secado, lo que daría un intervalo teórico de un año entre partos (Buxadé Carbó, 1996). Las curvas de lactancia típicas para la producción de leche en el ganado bovino muestran un pico entre las 4 y 8 semanas después del parto, seguido de una disminución diaria de la producción hasta que la vaca es secada o la lactancia se termine por causas naturales (Keown y col., 1986; Macciotta y col., 2005). A la capacidad de un animal de sostener su producción de leche luego de haber ocurrido el pico en su curva de lactancia se le denomina persistencia (Cupps, 1966). Se considera que una vaca es persistente, por lo tanto, si es capaz de prolongar en el tiempo la producción de leche máxima que alcanzó en el pico de su lactancia.

De igual manera, puede confeccionarse una curva cualitativa en función de la composición de la leche (cantidad de grasa y proteína), ya que ésta también varía a lo largo del tiempo.

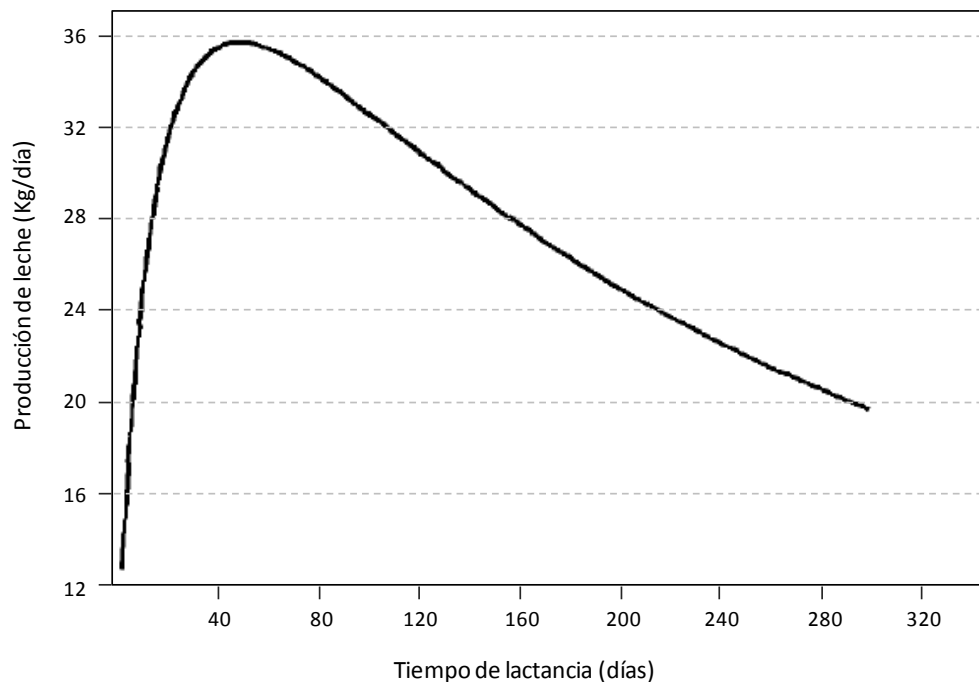


Figura III.1. Forma de una curva de lactancia *standard*.

Los valores de producción indicados en la figura son aproximados. Adaptación de Freeze y Richards, 1992.

Distintos factores dependientes del animal y de su entorno, sin embargo, pueden hacer que la curva de lactancia de un animal particular difiera significativamente de la curva de lactancia típica. De acuerdo a Buxadé Carbó (1996) los factores más importantes son:

1.A.1. Factores dependientes del animal.

- Factor genético: tiene gran influencia en la calidad y composición de la leche;
- Número de lactancia: el máximo de producción tiene lugar generalmente entre el tercer y quinto parto. Se trata de un factor importante ya que la producción puede aumentar hasta un 40% de la primera a la segunda lactancia. A partir de los 8-9 años de edad (sexta-séptima lactancia), la vaca empieza a reducir progresivamente su eficiencia en relación a la cantidad y calidad de la leche que produce.

En relación a la composición de la leche, sus componentes orgánicos, especialmente el porcentaje de grasa y de lactosa, van disminuyendo paulatinamente con la edad de los animales. El porcentaje de proteína total disminuye en menor proporción ya que el descenso en caseínas se compensa con una elevación en el contenido de proteínas del suero. A partir de la quinta lactancia, los cambios en la composición química de la leche tienden a ser mínimos.

- Edad al primer parto: tiene influencia sobre el nivel productivo de las vacas, aumentando la producción de leche de la primera lactancia con dicha edad. Sin embargo, si se retrasa el momento del primer parto (a más de 24 meses), la producción total de la vaca en toda su vida productiva tiende a disminuir, y si se adelanta demasiado (a menos de 18-20 meses), la producción se ve afectada negativamente en la primera lactancia, y puede ser inferior en lactancias posteriores.
- Estado de gestación: al final de la gestación (alrededor del séptimo mes) se produce una disminución en la producción de leche dado que el feto aumenta sus necesidades alimentarias.
- Estado de desarrollo: hace referencia a la condición corporal (peso corporal, cantidad de tejido secretor en las ubres) de las vacas y a la cantidad de grasa en reserva que pueden movilizar.
- Estado sanitario: es esencial que el animal se encuentre sano, en caso contrario la secreción láctea se verá afectada.

1.A.2. Factores dependientes del entorno que rodea al animal.

- Efectos ambientales: la estación del año en la que se produce el parto y la temperatura y humedad medioambientales durante el período de lactancia tienen influencia en la calidad y composición de la leche.

En sistemas pastoriles, las curvas de lactancia de vacas que tienen cría en otoño, aunque presentan en general una menor producción de leche en el pico de la lactancia que las de vacas que tienen cría en primavera, muestran mayores rendimientos en las fases media y tardía de la misma. Además, estas vacas también presentan lactancias de mayor duración. Esto determina que los rendimientos totales durante su período de lactancia sean mayores que los de las vacas que tienen cría en primavera (García y Holmes, 2001). Por otro lado, las temperaturas extremas y los climas secos se asocian a una disminución de la cantidad y calidad de la leche. Los efectos climáticos (como temperatura, humedad y fotoperíodo) pueden ser

separados en aquellos que actúan directamente sobre el animal y los que actúan indirectamente sobre la cantidad y calidad de las pasturas (García y Holmes, 1999).

- Alimentación: tiene influencia directa tanto sobre la calidad como la cantidad de leche producida.
- Período seco: se relaciona con el estado de reservas del animal en el momento del parto. La ubre necesita un período seco que regenere el tejido secretor, en general se recomiendan 60 días.

1.B. Modelos matemáticos utilizados para describir curvas de lactancia.

Las curvas de lactancia son descritas por medio de funciones matemáticas, con el objetivo de predecir el rendimiento diario de los animales durante el período de duración de su lactancia con un error mínimo, de modo de elucidar el patrón de producción lechera en presencia de una alta variación local debida al efecto del ambiente (Olori y col., 1999).

Se han propuesto diversos modelos matemáticos para describir la producción de leche a lo largo de la lactancia de animales domésticos. La utilidad de los mismos depende de su capacidad para reproducir la evolución del proceso biológico de producción de leche y de la forma en que ajustan los factores que lo afectan. No todos los modelos matemáticos se adecúan a una curva de lactancia típica con sus respectivas fases secuenciales de producción ascendente, máxima y descendente. Los parámetros de un modelo adecuado deben reflejar las influencias de factores genéticos, fisiológicos, productivos, ambientales y las interacciones entre ellos (Quintero y col., 2007). Estos modelos brindan la posibilidad de incluir en las evaluaciones, a través de la interpolación y extrapolación, animales con registros perdidos o lactancias incompletas (Wasike y col., 2011).

La bibliografía sobre la precisión de las distintas funciones matemáticas que se han aplicado en la descripción de las curvas de lactancia no es totalmente coincidente en cuanto a la función que ofrece el mejor ajuste de los registros obtenidos en distintos estudios (Sherchand y col., 1995; Olori y col., 1999; Silvestre y col., 2006; Beribe y col., 2013). Esto se debe a que la bondad de ajuste de los modelos depende no sólo de la función matemática que los describe, sino también de tendencias individuales que muestra la producción lechera diaria, las que varían entre distintos animales y entre distintas lactancias del mismo animal. Las características de la curva de lactancia están afectadas, asimismo, por variables ambientales como el día de lactancia en que se realizó el primer control lechero, el año, la estación y el número de parto, la edad del animal al momento del parto y su estado de gestación (Silvestre y col., 2006; Buxadé Carbó, 1996). Por lo tanto, para encontrar el mejor modelo estimador de curvas de lactancia, deben probarse distintos modelos y seleccionar aquél que mejor ajuste los registros de producción disponibles para la población en estudio. En el Anexo VII se describen brevemente las funciones matemáticas que utilizan los diferentes modelos para curvas de lactancia en el ganado vacuno.

2. Estimación de valores de cría. Modelo animal.

En producción animal generalmente se describe una observación a través de un modelo matemático lineal que incluye aquellos factores que se piensa tienen una influencia sobre la variable estudiada (Cardellino y Rovira, 1987, Cap. 17). En producción de carne y leche, el modelo animal es el más utilizado por describir mejor las características de interés económico. Éste, es un modelo lineal en el que el valor de cría de un animal se incluye como una variable explicativa para un rasgo fenotípico de interés. El valor de cría, a su vez, es el efecto aditivo que ejerce el genotipo de un individuo sobre la expresión del rasgo en relación a la media fenotípica poblacional (Wilson y col., 2010). Para un único carácter, el modelo animal más simple puede escribirse como:

$$y_i = \mu + a_i + e_i$$

donde y_i es el registro u observación fenotípica, μ es la media de las observaciones, a_i es el valor de cría o valor genético aditivo de los animales y e_i el efecto residual o medioambiental asociado con la i -ésima observación. Los términos residuales se asume que no están correlacionados entre individuos y que se distribuyen normalmente con una media = 0 y una varianza V_R . Este modelo puede ampliarse y tratarse como un modelo mixto que contenga efectos fijos y aleatorios. El uso de modelos mixtos hace posible descomponer la varianza fenotípica en varianzas que provienen de fuentes genéticas y medioambientales.

Normalmente, un factor se considera “fijo” si todos sus niveles (valores que puede adoptar) están presentes en los datos y si el objetivo es determinar el efecto de cada uno de estos niveles sobre la media fenotípica de la población. Por otro lado, un factor se considera “aleatorio” cuando los niveles presentes en los datos son una muestra de una población mayor sobre la cual se intenta hacer inferencias (Wilson y col., 2010). A modo de ejemplo, el efecto genético aditivo se considera aleatorio dado que en los datos normalmente no están representados todos los genotipos posibles, y se intenta usar la información de la que se dispone para hacer inferencias acerca de la varianza genética en la población. En general, se asume que los efectos aleatorios siguen una distribución normal con $\mu=0$ y varianza conocida (Weller, 2001). Algunos efectos, sin embargo, pueden ser tratados como fijos o aleatorios. Es el caso del año en que se realiza la medida fenotípica, o el año de nacimiento de los animales. Al considerar estos efectos como fijos, se estimarían los efectos específicos de cada año presente en los datos, sobre la media fenotípica. Mientras que al considerarlos como aleatorios se tendrá una estimación de cuánta varianza fenotípica puede ser explicada por el efecto del año, la que puede ser extrapolada a un conjunto mayor de años, distintos a los presentes en los datos (Wilson y col., 2010).

A través de la utilización del modelo animal puede hacerse la estimación de valores de cría (EBVs, *Estimated Breeding Values*). Este valor se obtiene generalmente a partir de análisis basados en REML (Máxima Verosimilitud Restringida, *Restricted -Residual- Maximum Likelihood*) como el mejor predictor lineal de mínima varianza (BLUP, *best linear unbiased predictor*) de los valores de cría verdaderos. Los EBVs se expresan como la diferencia entre la genética de un individuo con la base genética contra la cual fue comparada. Pueden ser positivos, negativos o cero, y son reportados en la unidad de medida correspondiente a cada

característica: kilos para producción de leche, o porcentaje de grasa para composición de grasa en la leche, por ejemplo.

La aplicación de estos modelos es interesante en el caso de caracteres de expresión limitada a uno de los sexos, como es la producción de leche, donde el valor genotípico de los machos no puede ser evaluado por su fenotipo, pues no lo presentan. Se debe entonces utilizar información proveniente de hembras relacionadas genéticamente con el individuo en cuestión, como hermanas, medio-hermanas, hijas, madre, etc. Aunque no toda la información tiene igual peso: cuanto más lejano es el parentesco, menor es el peso relativo que se le da a la información (Cardellino y Rovira, 1987, Cap. 8). Puede estimarse, por lo tanto, el mérito genético de los machos a través de la capacidad de producción de sus hijas.

Algunos de los programas ampliamente usados para trabajar con modelos animales son: ASReml y ASReml-R (Gilmour y col., 2009), WOMBAT (Meyer, 2007) y MCMCglmm-R (Hadfield, 2010). En este trabajo de tesis se utilizó el *software* WOMBAT para la estimación de los valores de cría para producción de leche, usando como punto de partida la producción de leche a 305 días de la población de animales con la que se trabaja.

3. Estudios de asociación.

Los estudios de asociación genética se clasifican en aquellos llevados a cabo sobre regiones/genes candidatos o a nivel de genoma completo (GWAS, *Genome-Wide Association Studies*). Ambas estrategias suponen la genotipificación de marcadores en un gran número de individuos implicados en diferentes tipos de diseños experimentales, como ser caso/control, o con registros para caracteres cuantitativos continuos. Los estudios sobre genes candidatos se basan en una hipótesis previa que sugiere un potencial rol del gen sobre un fenotipo particular. Un gen o región se selecciona como candidato en base a su función biológica o a su ubicación en una región implicada en un estudio de asociación o ligamiento previo. En un estudio de asociación de este tipo el objetivo principal es lograr la mayor cobertura posible de variación genética dentro de regiones específicas. Por el contrario, los GWAS utilizan SNPs distribuidos a lo largo de todo el genoma. Este tipo de estudios son útiles en la generación de hipótesis, pero raramente proveen la misma cobertura de una región candidata que un estudio de genes candidatos bien diseñado (Pettersson y col., 2009).

El objetivo de los análisis de asociación a nivel poblacional es la identificación de patrones de polimorfismos que varíen sistemáticamente entre individuos con diferentes fenotipos y que puedan, por lo tanto, constituir alelos responsables o asociados a los fenotipos observados.

A continuación se hará una introducción al concepto de desequilibrio de ligamiento y una breve exposición sobre los análisis de asociación *unilocus* y *multilocus*.

3.A. Desequilibrio de ligamiento.

El desequilibrio de ligamiento es una propiedad de los *loci* que se hallan sobre una porción contigua de la secuencia genómica, que describe el grado en el que un alelo de un *locus* es heredado o correlacionado con un alelo de otro *locus* en la población. Se relaciona con

el concepto de ligamiento cromosómico (*chromosomal linkage*), donde dos marcadores sobre un cromosoma permanecen físicamente unidos a través de las generaciones en una familia. Los eventos de recombinación que ocurren en una familia de generación a generación separan los segmentos cromosómicos. Este efecto se amplifica a través de las generaciones, y en una población de tamaño fijo con apareamientos al azar, los eventos repetidos de recombinación al azar separarán los segmentos contiguos de cromosomas (que contienen alelos ligados) hasta que eventualmente todos los alelos de la población estén en equilibrio de ligamiento o sean independientes (Bush y Moore, 2012).

Se han propuesto varias medidas del desequilibrio de ligamiento, todas ellas relacionadas con la diferencia entre la frecuencia observada de co-ocurrencia de dos alelos (es decir, de un haplotipo constituido por dos marcadores) y la frecuencia esperada considerando los dos marcadores independientes. Las dos medidas más comúnmente utilizadas son D' y r^2 :

$$D' = \left[\begin{array}{l} \frac{\pi_{AB}\pi_{ab} - \pi_{Ab}\pi_{aB}}{\min(\pi_A\pi_b, \pi_a\pi_B)} \text{ si } \pi_{AB}\pi_{ab} - \pi_{Ab}\pi_{aB} > 0 \\ \frac{\pi_{AB}\pi_{ab} - \pi_{Ab}\pi_{aB}}{\min(\pi_A\pi_b, \pi_a\pi_B)} \text{ si } \pi_{AB}\pi_{ab} - \pi_{Ab}\pi_{aB} < 0 \end{array} \right]$$

$$r^2 = \frac{(\pi_{AB}\pi_{ab} - \pi_{Ab}\pi_{aB})^2}{\min(\pi_A\pi_b\pi_a\pi_B)}$$

donde π_{AB} , π_{ab} , π_{Ab} y π_{aB} corresponden a las frecuencias de los haplotipos AB , ab , Ab y aB , respectivamente, y π_A , π_b , π_a y π_B son las frecuencias de los alelos A , b , a y B , respectivamente.

D' es una medida que se relaciona con los eventos de recombinación entre marcadores y oscila entre 0 y 1. Un valor de D' de 0 indica equilibrio de ligamiento completo, lo que implica la existencia de recombinación frecuente entre los dos marcadores e independencia estadística bajo el principio del equilibrio de Hardy-Weinberg. Un $D'=1$ indica completo desequilibrio de ligamiento, es decir, la ausencia de recombinación entre los dos marcadores en la población. Por otro lado, el desequilibrio de ligamiento también puede reportarse en términos de r^2 , una medida estadística de correlación. Altos valores de r^2 indican que dos SNPs llevan información similar, dado que un alelo del primer SNP se observa frecuentemente con un alelo del segundo SNP, por lo que basta con que sólo uno de los dos SNPs sea genotipificado para capturar la variación alélica.

La presencia de desequilibrio de ligamiento crea dos posibles alternativas para un estudio de asociación: la asociación directa y la asociación indirecta entre un SNP dado y el SNP responsable de la alteración en el fenotipo estudiado (Figura III.2). En el primer caso, el SNP que causa o tiene influencia sobre el fenotipo observado es genotipificado en el estudio e identificado como asociado estadísticamente con el carácter de interés. En el segundo caso, por el contrario, aunque el SNP que influencia el fenotipo no está genotipificado, si lo está un *tag SNP* en desequilibrio de ligamiento con el anterior, y este último es el que presenta una asociación estadísticamente significativa con el fenotipo (Bush y Moore, 2012).

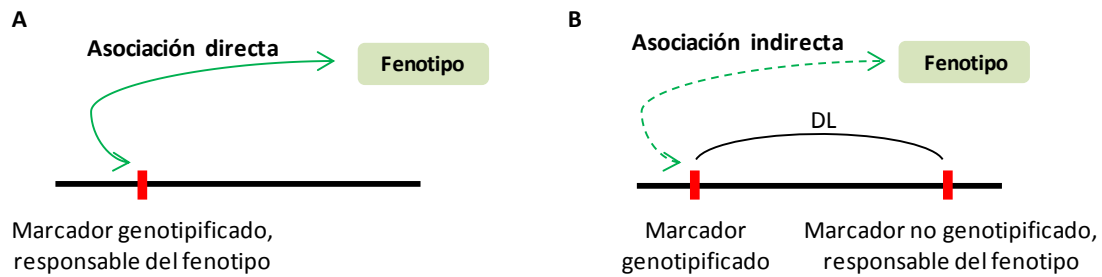


Figura III.2. Posibles alternativas para un estudio de asociación.

A. Asociación directa. **B.** Asociación indirecta.

DL: desequilibrio de ligamiento.

3.B. Análisis de asociación *unilocus*.

Estos análisis consisten en una serie de *tests* estadísticos *unilocus* que examinan independientemente cada SNP en busca de una asociación con el fenotipo. Pueden evaluarse asociaciones alélicas o genotípicas. Los estudios de asociación alélica estudian la asociación entre un alelo de un SNP y el fenotipo. Los estudios de asociación genotípica, en cambio, estudian la asociación entre genotipos (o clases genotípicas) y el fenotipo.

Los caracteres cuantitativos, como la producción de leche, se analizan generalmente a través de modelos lineales generalizados como el Análisis de Varianza (ANOVA). La hipótesis nula de un ANOVA que evalúa un único SNP consiste en la ausencia de diferencia entre las medias del carácter estudiado para los distintos grupos genotípicos. Las asunciones de los análisis lineales generalizados y del ANOVA son: la distribución normal de los valores fenotípicos, la igualdad de varianza entre cada grupo genotípico y la independencia entre los grupos. En los GWAS normalmente se utilizan modelos aditivos. Los mismos asumen, para un *locus* determinado con alelos alternativos A y a, que existe un incremento uniforme y lineal en la variación fenotípica que implica el alelo A de modo que, si la variación para el genotipo Aa es k , la implicada por el genotipo AA es $2k$. El efecto genético de interés (el número de copias de un alelo en particular de un SNP determinado) se incluye, entonces, como un efecto fijo en el modelo de regresión.

En los análisis de asociación se deben tener en cuenta factores con conocida influencia sobre el carácter estudiado como, por ejemplo, sexo y edad. Los ajustes por estos factores, aunque reducen las asociaciones erróneas, implican el uso de grados de libertad, lo que puede impactar en el poder estadístico del modelo. Uno de los factores más importantes a considerar en los análisis genéticos es la existencia de estructura poblacional. Como se mencionó en la discusión del Capítulo II de este trabajo de tesis, la variabilidad de frecuencias alélicas características de distintas composiciones raciales puede conducir a asociaciones erróneas si los animales de diferentes subpoblaciones son analizados sin tener en cuenta su origen racial. Una forma menos evidente de estructura poblacional está dada por la existencia de relaciones de parentesco entre los animales en estudio. Este problema puede superarse a través de la

inclusión en el modelo estadístico de un término que tenga en cuenta el efecto de todos aquellos genes que afectan el carácter, el “efecto poligénico” (Goddard y Hayes, 2009). Este término se incluye como un efecto aleatorio y en su estructura de covarianza está implicado el parentesco o relación genética entre cada par de individuos.

La estructura poblacional debiera ser evaluada en todo estudio de asociación a través de la observación de la distribución de los *tests* estadísticos (valores p) generados a partir de cada uno de los análisis de asociación realizados (uno por cada SNP examinado) y la estimación de su desviación de la distribución nula (aquella esperada bajo la hipótesis nula de que ningún SNP se asocia al carácter fenotípico) en un gráfico cuantil-cuantil o “*QQ-plot*” (Pearson y Manolio, 2008). En estos gráficos, las desviaciones de la diagonal correspondiente a la línea de identidad sugieren la existencia de *loci* fuertemente asociados al carácter, o bien de estructura en la población analizada. Si en los análisis de asociación no se toma en cuenta la estructura poblacional se aumenta la posibilidad de encontrar asociaciones falsamente positivas, dificultando la interpretación estadística de los resultados. Si se dispone de información genética de un gran número de marcadores, los resultados del análisis pueden ajustarse por efectos no específicos a través del método de Control Genómico -*Genomic Control*, GC- (Tsepilov y col., 2013). En los modelos aditivos, el factor de inflación por control genómico, constante λ (Devlin y Roeder, 1999), puede ser estimado empíricamente a partir de *loci* no asociados al carácter bajo estudio. El GC en este tipo de modelos ampliamente usados en GWAS tiene dos aplicaciones: en primer lugar, el coeficiente λ se usa para corregir el *test* estadístico, haciendo la interpretación de los valores p estadísticamente válida; en segundo lugar, λ resulta un importante indicador de la bondad del modelo utilizado en el análisis de asociación (Tsepilov y col., 2013). A pesar de no disponerse de un umbral establecido, un valor relativamente grande de λ (el valor deseado es 1) sugiere que el modelo elegido para el análisis considera pobremente la estructura genética presente en la muestra. En ese caso, el modelo debería ser revisado o reemplazado.

3.B.1. Correcciones para análisis de múltiples marcadores.

En un estudio de asociación, el análisis estadístico realizado se considera significativo y, por ende, se rechaza la hipótesis nula que asume ausencia de asociación entre fenotipo y genotipo, si el valor p que se genera para cada SNP examinado es menor a un valor α predefinido, generalmente de 0,05. Esto implica que el 5% de las veces la hipótesis nula se rechaza siendo en realidad verdadera, detectándose por lo tanto un falso positivo. Esto es válido para el caso de análisis estadísticos simples o *unilocus*; sin embargo, cuando se analizan varios marcadores como por ejemplo en los GWAS, se llevan a cabo cientos a miles de análisis, cada uno con su propia probabilidad de detección de falsos positivos. Por lo tanto, la probabilidad acumulada de encontrar uno o más falsos positivos a lo largo de todo el análisis de múltiples marcadores es mucho mayor. Para minimizar esta probabilidad se han propuesto varias alternativas, algunas de las cuales se describen brevemente a continuación:

- Corrección de Bonferroni. Esta corrección ajusta el valor α , de $\alpha=0,05$ a $\alpha=0,05/k$, donde k es el número de análisis estadísticos realizados. Se trata de la corrección más conservativa, dado que asume que cada uno de los análisis de asociación de un GWAS son independientes,

lo que no siempre es cierto dado el desequilibrio de ligamiento existente entre los distintos marcadores incluidos en un GWAS (Bush y Moore, 2012).

- Determinación de la tasa de falsos positivos (*False Discovery Rate*, FDR). FDR consiste en la estimación de la proporción de los resultados significativos (generalmente considerando un $\alpha=0,05$) que son falsos positivos (Benjamini y Hochberg, 1995). Provee también, por lo tanto, una estimación del número de resultados verdaderamente significativos, entre aquellos considerados significativos.
- Permutación. Las permutaciones consisten en reasignar al azar los fenotipos de cada individuo a otro individuo de la base de datos, interrumpiendo de esta manera cualquier asociación genotipo-fenotipo que pudiera tener lugar en la misma. Cada reasignación al azar de los datos, por lo tanto, genera una base de datos con la misma estructura de desequilibrio de ligamiento que la original pero que satisface la hipótesis nula de ausencia de asociación con el fenotipo (Balding, 2006), es decir, representa un posible muestreo de individuos bajo la hipótesis nula. Este proceso se repite un número predefinido de veces (N) de modo de generar una distribución empírica con resolución N. Este procedimiento tiene la desventaja de ser algo demandante desde el punto de vista computacional para los casos en que se trabaje con gran número de individuos y marcadores.

3.C. Análisis de asociación *multilocus*.

Consisten en estudios de asociación que, en lugar de analizar marcadores individuales, examinan las interacciones entre variantes genéticas a lo largo del genoma. En estos estudios son las combinaciones de marcadores las que son sometidas a análisis de asociación con fenotipos determinados. Se trata de estudios con mayores requerimientos computacionales, estadísticos y logísticos en relación a los estudios de asociación *unilocus* explicados previamente.

OBJETIVOS ESPECÍFICOS.

- Estimar el Valor de Cría para producción de leche para cada animal, a partir de los valores estimados de la producción de leche acumulada a 305 días durante la primera lactancia.
- Obtener SNPs asociados a la producción de leche acumulada a los 305 días integrando la información molecular, genealógica y fenotípica en el análisis.

MATERIALES Y MÉTODOS.

1. Información fenotípica.

El rasgo fenotípico del cual se partió fue la producción de leche de la primera lactancia acumulada a 305 días, cuyos valores fueron estimados por María José Beribe (Beribe y col., 2013) y que resumidamente se describen en los puntos 1 y 2 siguientes.

1.A. Descripción de características fenotípicas.

Se utilizaron los datos fenotípicos de producción (Kg) de leche registrados por el establecimiento Las Taperitas S.A. para las 1.842 vacas genotipificadas en los capítulos anteriores. Estas medidas se registraron dos veces al día (a la mañana y a la tarde) a intervalos desiguales a lo largo de las lactancias. Con estos registros se realizó un análisis descriptivo a través de la observación del número de lactancias registradas para cada vaca, de la cantidad de registros para la primera lactancia y del número de vacas por familia con registros para la primera lactancia. Luego, se establecieron una serie de condiciones que los animales y sus registros fenotípicos debían cumplir para ser incluidos en el análisis posterior de estimación de la producción de leche durante los primeros 305 días de la primera lactancia (Beribe MJ, comunicación personal). Dichas condiciones fueron las siguientes:

- Poseer al menos 4 controles lecheros durante la primera lactancia;
- Haberse registrado dichos controles lecheros como máximo a los 400 días en lactancia;
- Haberse registrado el primer control lechero como máximo a los 60 días de ocurrido el parto que da comienzo a la primer lactancia.

Al analizar la base de registros de datos fenotípicos provista por el establecimiento propietario de los animales evaluados se tuvieron en cuenta las reglas, estándares y guías para el registro de producción lechero, aprobadas en junio de 2012 por el Comité Internacional para el Registro de Animales (*International Committee for Animal Recording*, ICAR). El ICAR es la organización a nivel mundial responsable de la estandarización de la identificación y del registro y evaluación del rendimiento del ganado vacuno, caprino, ovino y bufalino, a través de la definición de normas para la medida de caracteres de importancia económica (*International agreement of recording practices*, 2012).

1.B. Modelado de curvas de lactancia. Estimación de la producción de leche a 305 días.

En base a un trabajo previo en el que se evaluó la capacidad predictiva de distintos modelos de regresión aleatoria para describir la curva de lactancia de bovinos de raza Holando y cruza Holando x Jersey pertenecientes al establecimiento Las Taperitas S.A. (Beribe y col., 2013), las curvas de lactancia de los animales utilizados en este trabajo de tesis se modelaron a través de la utilización de polinomios de Legendre. En particular, se utilizaron polinomios de Legendre de grado 3. Con éstos se estimó la producción diaria en la primera lactancia de cada una de las vacas que cumplieron las condiciones enunciadas en el punto anterior (1.628) y

luego se sumaron los valores obtenidos durante 305 días para obtener la estimación de la producción de leche acumulada a 305 días.

2. Información genealógica.

Los animales genotipificados poseían registros genealógicos completos hasta la cuarta generación. A pesar de esto, en este trabajo de tesis se consideraron en el pedigrí sólo dos generaciones, es decir, se trabajó con familias compuestas por padre-madre-hija, o padre-hija únicamente, dado que no se disponía de material biológico del resto de las generaciones del cual extraer y genotipificar el ADN.

3. Estimación de valores de cría.

Los valores de cría (EBV, *Estimated Breeding Values*) de los animales se estimaron mediante la aplicación del programa WOMBAT (Meyer, 2007). Se utilizó el siguiente modelo lineal mixto:

$$y_{ijklm} = \mu + pH_i + a\tilde{no}nac_j + tambo_k + EIL-AIL_l + a_m + e_{ijklm}$$

donde y_{ijklm} es el registro u observación fenotípica (producción de leche acumulada a 305 días), μ es la media de las observaciones, pH_j es el efecto racial, $a\tilde{no}nac_j$ es el efecto del año de nacimiento de la vaca, $tambo_k$ es el efecto del tambo, $EIL-AIL_l$ es el efecto combinado de estación y año de inicio de la lactancia, a_m es el valor de cría o valor genético aditivo de los animales y e_{ijklm} es el efecto residual. Todos los efectos incluidos en el modelo, a excepción del a_m , se consideraron fijos. Por otro lado, el valor genético aditivo de los animales se consideró un efecto aleatorio, definiéndose para el mismo una estructura de covarianza NRM (*Numerator Relationship Matrix*), lo que indica que el efecto se distribuye en forma proporcional al numerador de la matriz de parentesco.

Para poder ser incluidos en el archivo de pedigrí requerido por el programa WOMBAT, se recodificaron los animales de modo que cada animal tuviera un código mayor al de sus progenitores. Luego, para cada uno de los animales, se trabajó con la información registrada correspondiente a: tambo (en el que se realizó el primer control lechero, siendo 18 tambos); día, mes y año de inicio de lactancia (siendo 5 los años en los que las vacas iniciaron su lactancia -2004 a 2008-); fecha de nacimiento; porcentaje de raza Holando y producción de leche acumulada a 305 días. Esta información se reorganizó de la manera que se indica a continuación:

- En base a los valores de las variables “día de inicio de lactancia” y “mes de inicio de lactancia” se determinaron las estaciones de inicio de lactancia: verano, otoño, invierno y primavera.
- Se definieron los grupos de contemporáneos (conjunto de animales que tuvieron las mismas oportunidades para desempeñarse) para el efecto *EIL-AIL*, tras realizar el agrupamiento estación de inicio de lactancia-año de inicio de lactancia, resultando en 19 grupos diferentes.

- El porcentaje de raza Holando se agrupó en 4 clases: 100, 75, 50 y 25. En esta última categoría se incluyeron todos aquellos animales que poseían 25 o 12,5% de raza Holando. No se consideró 12,5% como una categoría aparte dado que sólo 7 animales presentaban este porcentaje.
- La fecha de nacimiento de las vacas se reorganizó bajo la categoría “año de nacimiento” (*añonac*), definiéndose 5 clases (años 2000-2001, 2002, 2003, 2004 y 2005-2006).

Con el fin de estimar los componentes de varianza y localizar el máximo de la función de verosimilitud se utilizaron los comandos *--aireml* (que utiliza el algoritmo AI, *average information*) y *--emalg* (que utiliza el algoritmo EM, *expectation maximization*) del programa WOMBAT (Meyer, 2007). Los valores iniciales para los componentes de varianza (incluidos en el archivo de parámetros provisto al programa) fueron 100.000 tanto para la varianza del valor genético aditivo de los animales como para la varianza residual. A partir de dichos valores iniciales el programa realiza iteraciones hasta encontrar los componentes de varianza más apropiados al modelo empleado. Posteriormente, con los valores obtenidos a partir de los comandos mencionados, se utilizó la aplicación *--blup* con el objetivo de hallar los EBVs de cada uno de los animales incluidos en el archivo de pedigrí. De esta forma, con el modelo se estimaron los valores de cría de las 1.628 vacas que contaban con registros de producción acumulada a 305 días, tras el modelado de sus curvas de lactancia, como también los de sus padres. Paralelamente, se obtuvieron los aportes de los efectos fijos.

Con el objetivo de comprobar si el aporte de un determinado efecto (fijo o aleatorio) en el modelo era significativo o no, se aplicó el modelo en presencia y en ausencia de los efectos a verificar y se compararon los coeficientes de verosimilitud obtenidos en cada caso (*maximum log likelihood*) a través de una prueba de razón de verosimilitud LRT (*likelihood ratio test*). Para ello, se calculó el doble de la diferencia entre el logaritmo de la verosimilitud obtenido al incluir el efecto y el obtenido en ausencia del efecto a revisar. Este resultado se asume que sigue una distribución χ^2 . Entonces, teniendo en cuenta los grados de libertad (el número de parámetros adicionales que se agregan al modelo más complejo respecto al más sencillo de los que se están comparando), se obtuvo el valor de significación p correspondiente a la comparación realizada. En base al valor p obtenido y habiendo definido un valor $\alpha=0,05$; se decidió si se justificaba incluir o no cada uno de los efectos considerados en el modelo.

Además, a partir del efecto animal o valor genético aditivo (V_A) y de la varianza fenotípica (V_p) estimados por el programa, se realizó la estimación de la heredabilidad en sentido estricto como $h^2=V_A/V_p$.

4. Información genotípica.

Se utilizaron los genotipos obtenidos a partir de la genotipificación llevada a cabo según los abordajes de genes candidatos y barrido del genoma descritos en los Capítulos I y II de este trabajo de tesis.

5. Análisis de asociación.

El análisis de asociación entre las distintas variantes alélicas de los SNPs pertenecientes a los bloques de genes candidatos definidos en el Capítulo II y los valores de cría estimados para producción de leche se realizó mediante la utilización del paquete GenABEL de R (Aulchenko y col., 2007). Se utilizó, en particular, la estrategia denominada FASTA (*FAMILY Based Score Test Approximation*; Chen y Abecasis, 2007). Este método hace uso en primer lugar de un modelo aditivo poligénico (Thompson y Shaw, 1990), para ello utiliza la aplicación “*polygenic*” del programa, que reconoce y corrige por la estructura poblacional presente en la muestra analizada, debida a la presencia de animales emparentados. La matriz de varianza/covarianza utilizada se estimó a partir de la matriz de parentesco genómico, es decir, aquella obtenida a partir de los genotipos de los SNPs de autosomas mediante la aplicación “*ibs*” (con la opción *weight=“freq”*) del programa. Para construir la matriz de parentesco no se utilizó la información de pedigrís disponible, dado que se ha reportado que los métodos que usan información genómica para inferir relaciones genéticas entre individuos son superiores a aquellos que utilizan únicamente la información provista por el pedigrí (Amin y col., 2007). En segundo lugar, se realizó el análisis de asociación utilizando los genotipos de cada uno de los SNPs pertenecientes a bloques de genes candidatos (considerados independientemente) y los EBVs obtenidos anteriormente mediante la función “*mmscore*” del programa. Finalmente, se utilizó la estrategia de Control Genómico (GC) a través de la estimación del factor *lambda* (λ) (mediante la aplicación “*lambda*”) para corregir y mejorar el análisis. Este factor es la media de los valores estadísticos obtenidos para cada SNP analizado, dividida por la media esperada de los valores estadísticos bajo la hipótesis nula de ausencia de asociación, asumiendo que el número de asociaciones verdaderas es muy pequeño en comparación con el número de análisis realizados (Clarke y col., 2011; Minozzi y col., 2013).

Para realizar el análisis de asociación mediante la estrategia FASTA-GC se utilizaron los 821 animales que superaron el control de calidad descrito en el Capítulo II de este trabajo de tesis. Por otro lado, los SNPs sobre los que se evaluó la asociación con el fenotipo en estudio fueron los 10.227 SNPs de bloques de genes candidatos que también superaron el control de calidad realizado (10.182 genotificados mediante el *chip* de mediana densidad BovineSNP50 v2 BeadChip de Illumina y 45 mediante el sistema “SNPlex® Genotyping System” de Applied Biosystems). No obstante, la matriz de parentesco utilizada fue construida a partir de los 43.313 SNPs pertenecientes a autosomas, genotificados mediante el *chip*, que superaron los umbrales establecidos en el control de calidad (tasa de genotificación > 0,9; FAM > 0,01; EHW, $p > 1.10^{-6}$).

Otras aplicaciones del programa GenABEL utilizadas (que no fueron mencionadas previamente) fueron las siguientes:

- “*convert.snp.tped*” con el fin de convertir la información genotípica desde el formato de PLINK al de GenABEL;
- “*load.gwa.data*” para cargar la información genotípica y fenotípica en el área de trabajo;
- “*as.dist*” y “*cmdscales*” con el objetivo de determinar los dos componentes principales resultantes del análisis de la matriz de parentesco y graficarlos uno en función del otro, de

modo de localizar cada una de las muestras en el espacio y observar la presencia o ausencia de *clusters* o agrupaciones de muestras que sugirieran la presencia de estructura en la población;

- “*kmeans*” para determinar las subpoblaciones observadas en el gráfico de componentes principales;
- “*qtscore*” con el fin de realizar el análisis de asociación sin considerar la estructura poblacional ni el efecto poligénico, para luego comparar la corrección del factor de inflación lograda con la utilización de la aplicación *mmscore*;
- “*descriptives.scan*” (con la opción *sortby="Pc1df"*) para ordenar los resultados del análisis de asociación en base a su valor *Pc1df* (valor *p* corregido por el factor λ) en orden decreciente.

Mediante el programa STATISTICA 7.0 se comprobó que los EBVs se ajustasen a una distribución normal de modo de poder ser utilizados como carácter fenotípico en el análisis de asociación.

Al realizar el análisis de asociación, aquellos SNPs que presentaron valores *p* corregidos por el factor λ inferiores a 0,001 fueron considerados asociados al carácter estudiado. Estos SNPs fueron ubicados en el contexto génico (ensamblado UMD 3.1) y su *rs* fue determinado a través de la base de datos SNPchiMp versión 2 (Nicolazzi y col., 2014), disponible *online* (<http://bioinformatics.tecnoparco.org/SNPchimp/>).

6. Determinación del desequilibrio de ligamiento.

El desequilibrio de ligamiento (DL) existente entre pares de SNPs se calculó a través de la estimación de *D'* con el programa Haploview (Barrett y col., 2005). Para ello se utilizó un archivo constituido por los 821 animales genotificados y los EBVs para producción de leche durante su primera lactancia. Dichos animales se organizaron como tríos padre-madre-hija en los casos en que fue posible, con el fin de considerar la estructura familiar de la población. La estimación de *D'* se realizó para aquellos pares de SNPs separados entre sí por menos de 500 Kb, que superaron el control de calidad de la genotipificación realizado en el Capítulo II.

Posteriormente se definieron bloques de SNPs en DL mediante el algoritmo de intervalos de confianza implementado en Haploview (Gabriel y col., 2002). El mismo consiste, brevemente, en la generación de intervalos de confianza del 95% para *D'* y en la clasificación de cada par de SNPs según se encuentre o no en fuerte DL (esto se cumple si el límite superior de dicho intervalo de confianza es mayor a 0,98 y el límite inferior es mayor a 0,7). Un bloque haplotípico de SNPs en DL se define luego como una región en la que el 95% de los pares de SNPs informativos se halla en fuerte DL o, dicho de otra forma, en una región en la que sólo un 5% de los pares de SNPs exhibe fuerte evidencia de recombinación histórica (Khatkar y col., 2007). Para la determinación de los bloques se ignoraron aquellos SNPs con $MAF < 0,01$.

Una vez definidos los bloques de SNPs en DL para cada bloque de genes candidatos, se evaluó si de los SNPs que resultaron significativos tras el análisis de asociación del *ítem* anterior existía al menos un par que pertenecieran al mismo bloque y, por lo tanto, brindaran la misma información.

RESULTADOS.

1. Descripción de características fenotípicas.

En la base de datos consistente en los registros de producción de las 1.842 vacas genotipificadas en los capítulos anteriores, el número de lactancias registradas para cada vaca varió en un rango de 1 a 10, y el número de controles lecheros por lactancia en un rango de 1 a 26. A partir de esta base de datos inicial, se construyó una base de datos reducida, consistente en los registros de producción de primera lactancia, obtenidos de las 1.842 vacas genotipificadas en los capítulos anteriores. De esta base de datos se eliminaron los registros y animales que no cumplían las condiciones enunciadas en la sección de Materiales y Métodos de este capítulo, permaneciendo controles lecheros correspondientes a la primera lactancia de 1.628 animales en la base, para el modelado de la curva de lactancia (Beribe MJ, comunicación personal). Dichos registros se obtuvieron durante un período de 5 años, entre los años 2004 y 2008, y el número de registros promedio por vaca varió entre 4 y 14. El número de vacas por familia varió en un rango de 1 a 194, siendo 25 las familias incluidas en la base. En la Figura III.3 se grafica el número de vacas incluidas en la base que son hijas de cada toro.

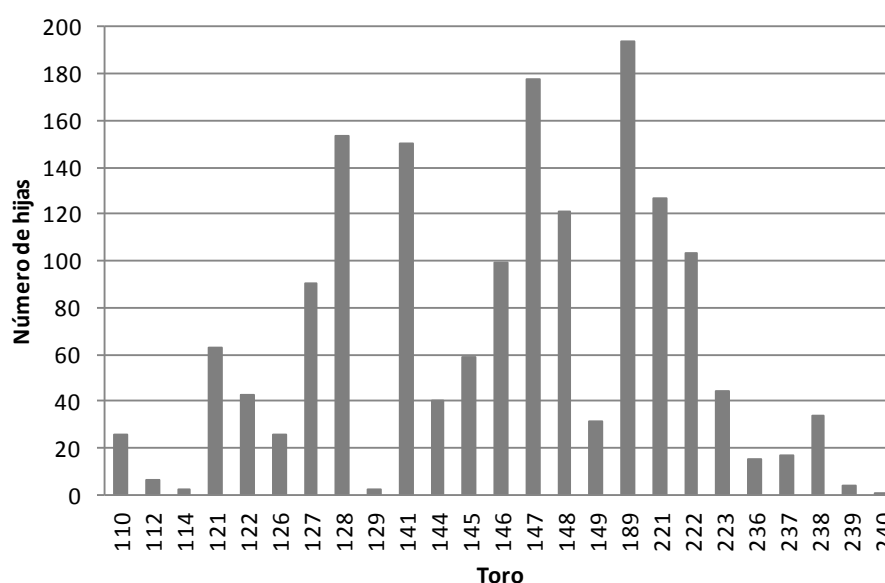


Figura III.3. Distribución de hijas por familia.

Toro: número que corresponde al RP de cada uno de ellos.

2. Estimación de la producción de leche a 305 días.

El rendimiento diario promedio tomó valores entre 8,69 y 29,53 Kg de leche, habiéndose registrado una producción mínima diaria de 5,21 Kg y una producción máxima de 34,69 Kg. Al sumar las producciones diarias de cada vaca durante 305 días, se obtuvo una producción de leche acumulada que varió en un rango de 2.650,64 - 9.007,13 Kg, con un promedio de 5.960,46 Kg.

3. Estimación de los valores de cría para producción de leche.

Las pruebas de razón de verosimilitud realizadas para inspeccionar el aporte de los efectos considerados en el modelo fueron en todos los casos significativas ($p < 0,01$). Por lo tanto, existen evidencias para concluir que la inclusión de los efectos fijos *pH*, *añonac*, *tambo* y *EIL-AIL* mejora significativamente el modelo.

La heredabilidad (en sentido estricto) $-h^2$, estimada como la proporción de la varianza fenotípica causada por la variación genética aditiva (Goddard y Hayes, 2009), para la producción de leche acumulada a 305 días fue 0,208.

Luego, se hallaron los EBVs de cada uno de los animales incluidos en el archivo de pedigrí. Se estimaron los valores de cría de las 1.628 vacas que contaban con registros de producción de leche acumulada a 305 días, así como los de sus padres. Además, se obtuvieron los aportes de los efectos fijos (ver Anexo VIII).

En la Figura III.4 se grafica el rango de EBVs obtenido para cada familia analizada, es decir, se indica la distribución de EBVs para las hijas de cada uno de los toros padres de familia, así como el EBV para cada toro, estimado en base a los registros de producción de sus hijas. En la Tabla III.1 se muestran el número de hijas por toro (sólo aquellos con 10 o más hijas) y los porcentajes de hijas con EBVs negativos y positivos dentro de cada familia.

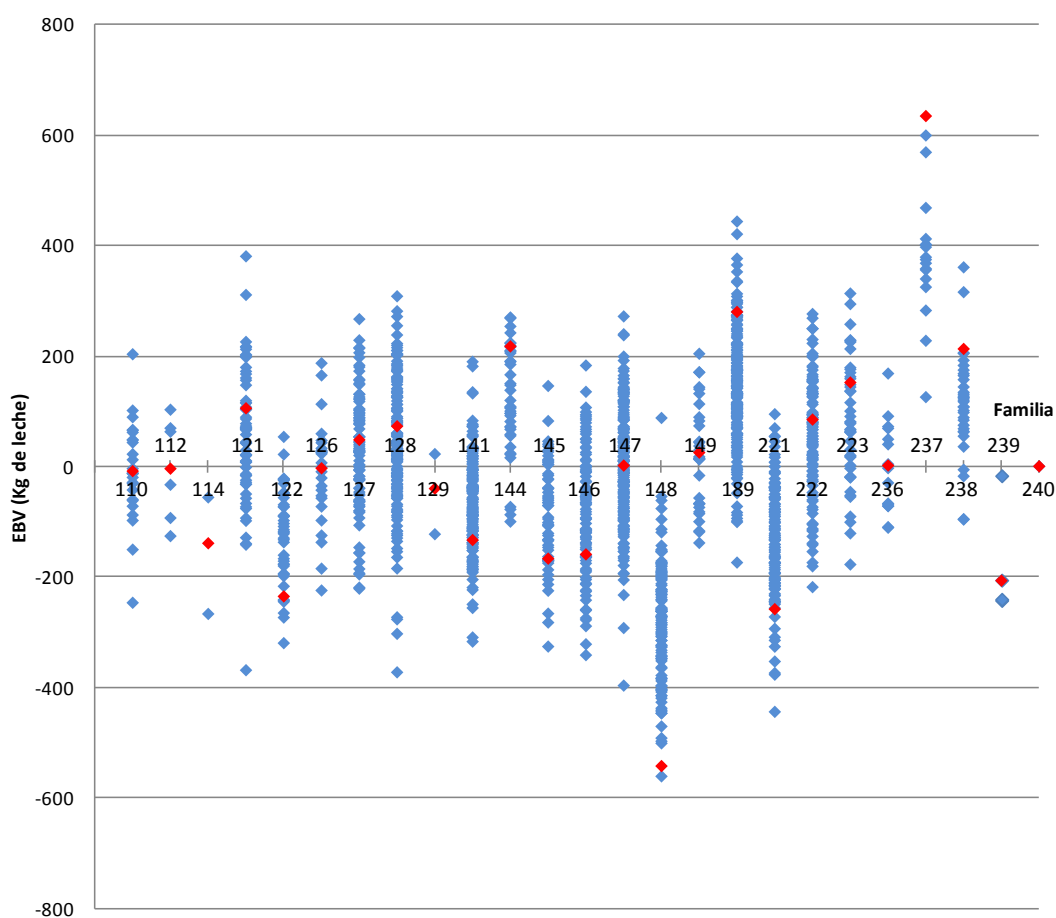


Figura III.4. Distribución de los valores de cría estimados por familia.

EBVs de las hijas de cada uno de los toros padres de familia (en azul), y EBV para cada toro (en rojo), estimado este último en base a los registros de producción de sus hijas.

Toro	Número de hijas	Porcentaje de EBVs negativos	Porcentaje de EBVs positivos
110	26	57,7	42,3
121	63	30,2	69,8
122	43	95,4	4,6
126	26	50,0	50,0
127	90	41,1	58,9
128	153	36,0	64,0
141	150	80,0	20,0
144	40	10,0	90,0
145	59	76,3	23,7
146	99	76,8	23,2
147	178	50,6	49,4
148	121	99,2	0,8
149	31	38,7	61,3
189	194	8,8	91,2
221	127	92,1	7,9
222	103	29,1	70,9
223	44	25,0	75,0
236	15	53,3	46,7
237	17	0,0	100
238	34	11,8	88,2

Tabla III.1. Porcentajes de EBVs negativos y positivos por familia.

Número de hijas por familia (para aquellas con más de 10 hijas), y porcentajes de EBVs negativos y positivos correspondientes a cada familia. Los toros se designan con su número de RP correspondiente.

4. Análisis de asociación.

En primer lugar, se comprobó que los EBVs para producción de leche se ajustasen a una distribución normal de modo de poder ser utilizados como carácter fenotípico en el análisis de asociación. Para ello se utilizó el programa STATISTICA 7.0, graficándose el histograma correspondiente a los EBVs de los 821 animales genotipificados que superaron el control de calidad realizado en el Capítulo II de este trabajo de tesis (Figura III.5). Los valores de EBVs de los 821 animales se ajustaron a una distribución normal con media= -5,799 Kg y desvío estándar= 161,42 Kg.

Luego, con el fin de evaluar la estructura presente en la población utilizada, tras construir la matriz de parentesco, se determinaron y graficaron los dos componentes principales (Figura III.6). Se observó la existencia de dos grupos bien diferenciados: uno de ellos, conformado por los 192 animales con porcentajes de raza Holando en el rango de 0 a 50% (Grupo 1), y el otro, por los 629 animales con porcentajes de Holando entre 75 y 100% (Grupo 2).

A los efectos de mantener el mayor poder estadístico se decidió trabajar con la población completa y realizar el análisis de asociación teniendo en consideración la estructura poblacional existente. Para ello, en la estimación del modelo poligénico se incluyó como covariable la primera componente del análisis de componentes principales (PCA) de los genotipos.

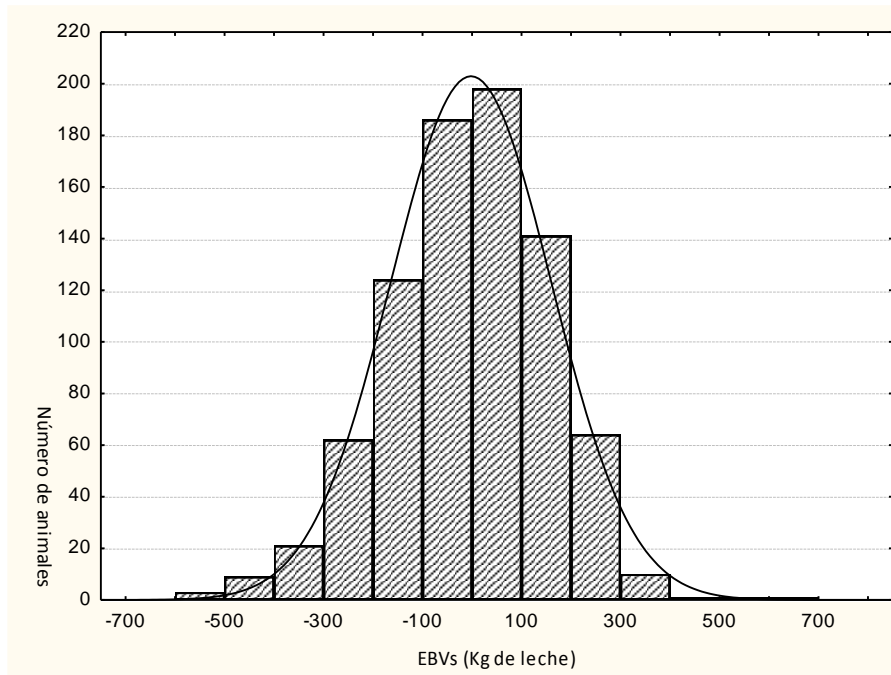


Figura III.5. Ajuste de los valores de EBV a una distribución normal.
 Histograma de los EBVs de los 821 animales genotipificados. La curva normal a la cual se ajustan los valores se indica con trazo negro continuo.

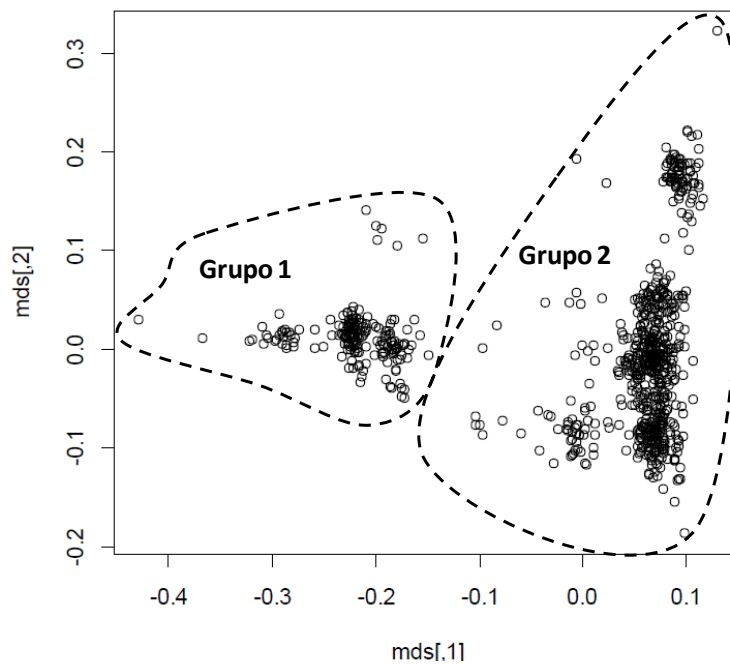


Figura III.6. Gráfico de componentes principales.
 Subpoblaciones en las que puede dividirse la población analizada (Grupos 1 y 2).

Luego, con el programa GenABEL, se realizó el análisis de asociación sobre 821 animales y 10.227 SNPs utilizando como información fenotípica los EBVs para producción de leche. La desviación de la diagonal correspondiente a la línea de identidad en el gráfico cuantil-cuantil (*Q-Q plot*) obtenido al no considerar el efecto poligénico (Figura III.7.A) indica la existencia de estructura en la población analizada, la cual se logra corregir al incluir el efecto poligénico y la primera componente principal del PCA de los genotipos en el modelo (Figura III.7.B).

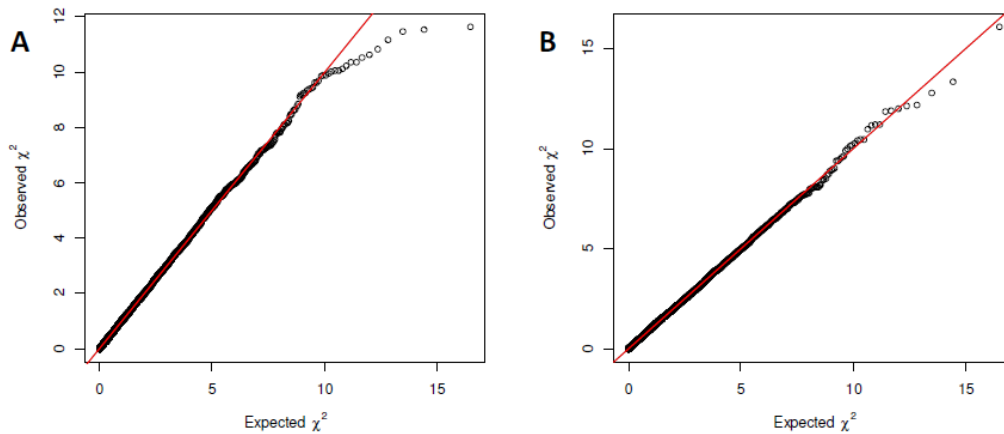


Figura III.7. Gráficos cuantil-cuantil resultantes del análisis de asociación.

Desviación de la distribución de los valores p generados en los análisis de asociación para cada SNP examinado (en negro), respecto de la distribución esperada en ausencia de asociación al carácter fenotípico (en rojo), sin (A) y con (B) corrección por la estructura poblacional.

Con el fin de visualizar de forma clara los resultados del análisis de asociación, se graficó el logaritmo negativo de los valores p obtenidos para cada SNP analizado (corregidos por el factor λ correspondiente), en función de la localización de los SNPs en el genoma (Gráfico de Manhattan). En el mismo (Figura III.8) puede observarse la localización de los SNPs asociados a la producción de leche en la población analizada, es decir, de aquellos SNPs con valores p corregidos inferiores a 0,001 ($-\log p_{\text{corregido}} > 3$).

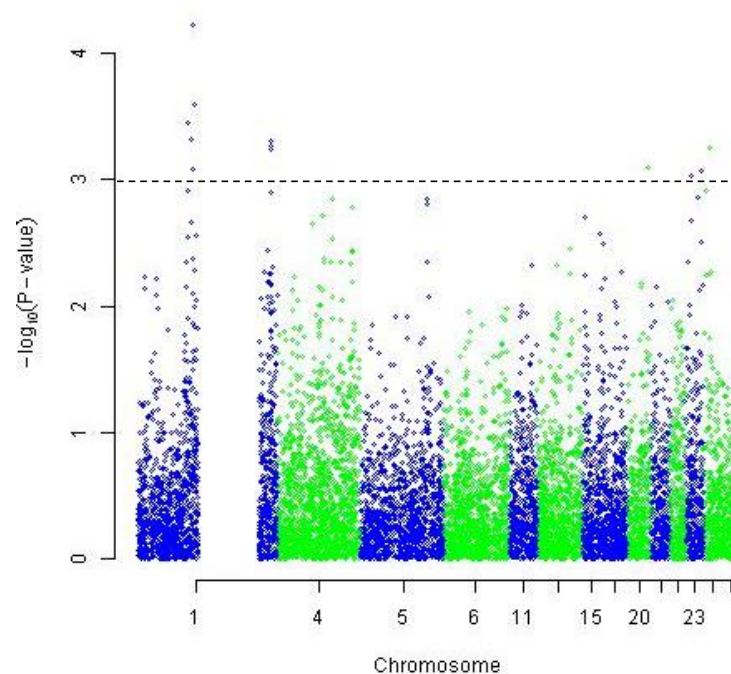


Figura III.8. Gráfico de Manhattan resultante del análisis de asociación.

Aquellos SNPs con $-\log(p\text{-valor})$ por encima de 3 fueron considerados asociados al carácter estudiado. En el eje de abscisas se indican los distintos bloques de genes candidatos analizados, que se corresponden con los cromosomas en los que se localizan. El espacio en blanco en el cromosoma 1 se debe a que en este cromosoma se evaluaron dos bloques de genes candidatos, separados entre sí por una región sin analizar.

Se hallaron 12 SNPs asociados a la producción de leche, correspondientes a los bloques localizados en los cromosomas 1 (bloques 1.A y 1.B), 20, 23 y 24 (ver Tabla III.2 en hoja siguiente). Estos SNPs fueron ubicados en el contexto génico (ensamblado UMD 3.1) a través de la base de datos SNPchiMp v.2 (Nicolazzi y col., 2014).

5. Determinación del desequilibrio de ligamiento y definición de bloques de SNPs en fuerte DL.

Luego de realizar la estimación del DL existente entre pares de SNPs separados entre sí por menos de 500 Kb bajo la forma de D', se definieron los bloques haplotípicos o bloques de SNPs en fuerte DL, de acuerdo al algoritmo de intervalos de confianza implementado en el programa Haploview.

En la Tabla III.3 se indica el número de bloques haplotípicos hallados en cada uno de los bloques de genes candidatos estudiados.

Bloque	SNPs	Bloques haplotípicos	SNPs significativos
1.A	1.139	55	5
1.B	416	28	3
4	1.611	105	0
5	1.285	80	0
6	1.227	81	0
11	519	26	0
14	974	62	0
15	335	16	0
19	566	29	0
20	447	30	1
21	337	16	0
22	353	24	0
23	332	9	2
24	320	15	1
26	366	20	0
Total	10.227		12

Tabla III.3. Bloques haplotípicos en desequilibrio de ligamiento.

Bloques de genes candidatos, número de SNPs que los componen, bloques haplotípicos y número de SNPs que resultaron significativos al realizar el análisis de asociación. En la última fila se expone la sumatoria de los SNP totales y de los significativos.

Una vez definidos los bloques de SNPs en DL para cada bloque de genes candidatos, se evaluó si los SNPs que resultaron significativos del análisis de asociación realizado previamente pertenecían o no a alguno de dichos bloques. Sólo un SNP, el ARS-BFGL-NGS-94997, resultó incluido en uno de los bloques haplotípicos definidos por Haploview.

SNP	Orden	rs	Ubicación	Valor p corregido	Cambio nucleotídico	Localización	Gen
ARS-BFGL-NGS-85105	3	rs110823196	1:55832662	$3,5 \cdot 10^{-4}$	C/T	intergénica	
ARS-BFGL-NGS-94997	4	rs43239003	1:59754059	$4,8 \cdot 10^{-4}$	A/G	intergénica	
Hapmap59475-rs29014975	10	rs29014975	1:60411957	$8,2 \cdot 10^{-4}$	C/T	intergénica	
BTB-00030894	1	rs43237323	1:61749066	$6,1 \cdot 10^{-5}$	A/G	intergénica	
Hapmap51585-BTA-32601	2	rs41629099	1:62104120	$2,6 \cdot 10^{-4}$	C/T	intergénica	
BTB-01301015	6	rs42429246	1:147255491	$5,4 \cdot 10^{-4}$	A/G	génica	<i>PCBP3</i>
BTB-00043331	8	rs43253207	1:147787147	$5,8 \cdot 10^{-4}$	A/G	génica	<i>PCNT</i>
BTB-00043033	5	rs43250616	1:148009894	$4,9 \cdot 10^{-4}$	A/G	génica	<i>S100B</i>
ARS-BFGL-NGS-47270	9	rs42501206	20:46500354	$8,1 \cdot 10^{-4}$	A/T	intergénica	
ARS-BFGL-NGS-15262	12	rs111013945	23:31312887	$9,3 \cdot 10^{-4}$	C/T	intergénica	
ARS-BFGL-NGS-70977	11	rs109351365	23:42647945	$8,5 \cdot 10^{-4}$	C/T	intergénica	
Hapmap50588-BTA-118970	7	rs41622951	24:18599946	$5,6 \cdot 10^{-4}$	C/T	intergénica	

Tabla III.2. SNPs que resultaron significativos al realizar el análisis de asociación.

En la tabla se indican el nombre del SNP (ordenados según su ubicación en el genoma), su número de orden según valores decrecientes de $-\log(\text{valor } p_{\text{corregido}})$, su código *rs*, su localización en el genoma bovino según el ensamblado UMD 3.1 (BTA:pb), su valor p corregido resultante del análisis de asociación, el cambio nucleotídico implicado y su ubicación génica o intergénica. En los casos en que los SNPs se hallen dentro de un gen, se indica el nombre del gen correspondiente.

La Figura III.9 muestra una porción del mapa de disequilibrio de ligamiento para el bloque de genes 1.A en la que se representa el bloque haplotípico 47 y se indica el SNP ARS-BFGL-NGS-94997 contenido en el mismo.

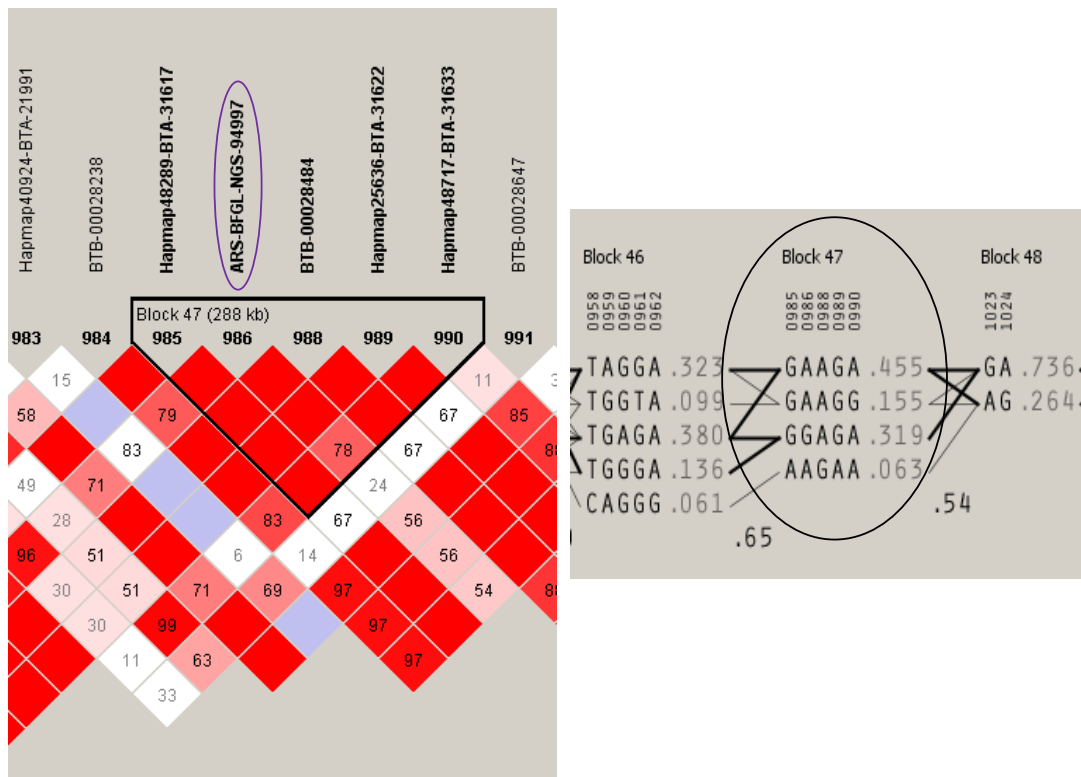


Figura III.9. Porción del mapa de disequilibrio de ligamiento del bloque 1.A de genes candidatos.

Bloque haplotípico del bloque 1.A que contiene uno de los SNPs que resultaron asociados al carácter estudiado. Se indica su longitud en Kb, así como los distintos genotipos hallados en la población estudiada en dicha región y las frecuencias en las que se presentan los mismos.

En el mapa de disequilibrio de ligamiento, a mayor intensidad de color rojo en los cuadros, mayores valores de D' . Los triángulos que rodean y agrupan cuadros representan los bloques haplotípicos obtenidos mediante el algoritmo de intervalos de confianza implementado en Haploview.

Por lo tanto, la asociación encontrada entre el EBV para producción de leche y el SNP ARS-BFGL-NGS-94997 podría extenderse a la zona del BTA 1 de aproximadamente 0,29 cM comprendida entre los SNPs Hapmap48289-BTA-31617 (localizado en 1:59683842) y Hapmap48717-BTA-31633 (localizado en 1:59971879), que constituyen los límites del bloque haplotípico definido por el programa Haploview mostrado en la Figura III.9.

DISCUSIÓN.

En este capítulo se partió de los valores predichos de la producción de leche a 305 días obtenidos por Beribe y col. (2013).

Los animales se organizaron en grupos de contemporáneos, y de esta forma fueron incluidos en el modelo utilizado para la obtención de los EBVs. Dicha organización tuvo en cuenta que los grupos quedaran representados por un número mínimo de animales, de modo de extraer conclusiones insesgadas de sus efectos. Dado que la estimación de los efectos de estos grupos de contemporáneos es simultánea con la predicción de los valores de cría, el registro erróneo de los factores que integran un grupo de contemporáneos afecta la predicción de dichos valores de cría. Por tal motivo, con el fin de minimizar la inclusión de errores, para el caso de las vacas con registros de producción que fueron genotipificadas mediante el *chip* BovineSNP50 v2 BeadChip (Illumina Inc., California), se utilizó la información de genealogía obtenida luego de realizar la reasignación de paternidades, así como la composición racial (expresada a través del porcentaje de raza Holando) que surgió de dicha reasignación. De las 1.628 vacas con registros de producción lechera, sólo 856 fueron genotipificadas mediante el *chip*, por lo que las 772 restantes, utilizadas en la estimación de los valores de cría, pudieron haber incorporado cierto error en la estimación si contaban con algún registro de paternidad erróneo proveniente del establecimiento de origen. En este punto, cabe recordar que a partir de las inconsistencias mendelianas detectadas en las distintas familias genotipificadas mediante el *chip* utilizado, se les reasignaron los padres al 20,8% de las vacas analizadas (discutido en el Capítulo II).

El rango de EBVs obtenido para cada familia es bastante amplio y se observa que todas las familias estudiadas, a excepción de las de los toros 114, 237 y 239, presentan hijas con valores de EBV positivos y negativos, es decir, con un mérito genético aditivo opuesto para el rasgo estudiado. Esto podría deberse a que, aunque todas las vacas de una familia comparten el padre, el aporte genético materno es diferente para cada una de ellas, lo que impacta en la estimación del valor de cría.

La heredabilidad estimada para la producción de leche acumulada a 305 días fue de 0,208. Este valor se halla dentro de los valores esperados para el rasgo. Cardellino y Rovira (1987, Cap. 5) en la clasificación que hacen de los grupos de caracteres de producción según la magnitud de sus valores de heredabilidad ubican la producción de leche en la categoría de heredabilidad media a alta, estableciendo un rango de 0,2 - 0,4 para la producción de leche bovina. No obstante, aunque los valores en general se pueden extrapolar entre poblaciones con similar estructura genética, historia y expuestas a medioambientes similares, debe recordarse que la heredabilidad de una característica dada es un valor relativo y no absoluto que se aplica a una población en particular.

En la actualidad existe una variedad de programas para realizar análisis de asociación (GenABEL -Aulchenko y col., 2007-, MERLIN -Abecasis y col., 2002-, TASSEL -Bradbury y col., 2007-, entre otros). Dado que muchos de estos programas presentan una potencia y un nivel de error de tipo I similar y proporcionan similares inferencias en cuanto a la localización de señales y valores de $-\log_{10}(\text{valor-}p)$ alcanzados, la decisión de qué programa usar se basa

generalmente en consideraciones prácticas como facilidad de uso y eficiencia computacional (Eu-ahsunthornwattana y col., 2014). En este trabajo de tesis se utilizó el paquete GenABEL de R. Su elección se basó en que este programa es conveniente para ser utilizado en GWAS con caracteres cuantitativos o binarios de individuos emparentados, es exitoso en el control genómico del factor de inflación λ en niveles adecuados y ofrece mayor potencia que los análisis de asociación tradicionales (Eu-ahsunthornwattana y col., 2014). En particular, se utilizó la estrategia FASTA-GC implementada en el paquete, incluyendo en la estimación del modelo poligénico la primera componente principal del análisis de componentes principales de los genotipos como covariable. La ventaja de utilizar esta estrategia reside en que toma en consideración la estructura de la población analizada debida tanto a las diferentes razas implicadas como a la presencia de animales emparentados.

La evidencia de dos grupos claramente diferenciados en el gráfico de componentes principales de la Figura III.6, correspondientes a 192 animales de hasta 50% de raza Holando por un lado y 629 animales predominantemente Holando por el otro, sugiere la existencia de estructura poblacional dada por el origen racial de los animales bajo estudio. Aunque la división de la población en dos subpoblaciones y el análisis de asociación sobre cada una de ellas es una opción válida, en nuestro caso implica una reducción importante en el número de animales incluidos en cada análisis, especialmente en el caso del grupo compuesto por 192 animales, lo que conlleva la pérdida de potencia estadística. Por este motivo se trabajó sobre la población completa teniendo en cuenta la estructura poblacional detectada, realizando el análisis de asociación como se mencionó en el párrafo anterior.

Otro tipo de estructura poblacional en la población analizada está determinada por la existencia de relaciones de parentesco entre los animales bajo estudio. Por ello se utilizó la estrategia FASTA-GC del programa GenABEL, que consiste en un análisis de asociación para poblaciones compuestas por individuos emparentados. Esto permitió hallar asociaciones entre los genotipos de SNPs y el fenotipo estudiado, libres de errores debidos a la estructura de parentesco existente en la población.

En este trabajo se identificaron 12 SNPs asociados al valor de cría estimado para producción de leche. Estos SNPs se ubican en regiones no contiguas del genoma bovino, en particular, en los BTAs 1, 20, 23 y 24. Este resultado indica que el fenotipo estudiado es influenciado por polimorfismos de distintas regiones del genoma. Este resultado no es de extrañar para el fenotipo analizado, dado que, aunque según el carácter puede haber genes con grandes efectos, más del 50% de la varianza genética reside en un gran número de regiones distribuidas a lo largo del genoma con magnitudes de efectos consistentes con un modo poligénico de herencia, acercándose al modelo infinitesimal de la genética cuantitativa (Dekkers, 2012). Otros trabajos, a su vez, reportaron asociaciones entre SNPs de distintos cromosomas y el rendimiento de leche bovina: Jiang y col. (2010) reportaron 20 SNPs asociados significativamente al fenotipo estudiado en los BTAs 1, 3, 5, 9, 14 y 26; mientras que Pimentel y col. (2010) detectaron 22 SNPs asociados a la producción de leche, localizados en los BTAs 2, 3, 4, 5, 6, 10, 13, 17, 19, 20, 21, 26 y 28. Raven y col. (2014), por su parte, realizaron un GWAS utilizando genotipos de 17.925 animales de razas Holstein y Jersey, identificando 5 SNPs asociados a la producción de leche, ubicados en los cromosomas 5, 6, 14 y 20. Más aun, actualmente (junio de 2015) han sido reportados (www.animalgenome.org/QTldb/) 1.043

QTLs asociados al rendimiento o producción de leche, distribuidos en todos los cromosomas del genoma bovino.

Sólo uno de los SNPs que resultaron significativos en el análisis de asociación resultó incluido en uno de los bloques haplotípicos (bloques de SNPs en desequilibrio de ligamiento) definidos por Haploview. Por lo tanto, al no hallarse SNPs que compartieran un mismo bloque haplotípico, puede concluirse que los SNPs significativos hallados en cada bloque de genes candidatos brindan información única.

En el BTA 1 se detectaron 8 marcadores SNPs significativamente asociados al carácter bajo estudio en dos regiones: 1:55832662-62104120 y 1:255491-148009894. En la primera de estas regiones, fueron reportados previamente 3 QTLs asociados a la producción lechera de bovinos Holando, en un análisis de asociación utilizando 4.856 SNPs (Daetwyler y col., 2008) y en mapeos de QTLs mediante genotipificación de microsatélites (Nadesalingam y col., 2001; Georges y sol., 1995). Mientras que, cercanos a la segunda región del BTA 1 mencionada, existen 4 QTLs asociados a la producción lechera reportados (Jiang y col., 2010; Daetwyler y col., 2008; Bagnato y col., 2008). En el BTA 20, cercanos al SNP rs42501206, Arranz y col. (1998) y Bagnato y col. (2008) mapearon dos QTLs asociados a la producción de leche trabajando con genotipos de 215 y 172 microsatélites, respectivamente. Por otro lado, los SNPs del BTA 23 que resultaron significativos en este trabajo de tesis, están comprendidos en dos QTLs detectados previamente por Bennewitz y col. (en los años 2003 y 2004). Finalmente, el SNP rs41622951 del BTA 24 no se halla comprendido ni se localiza en las cercanías de ningún QTL detectado previamente para el fenotipo bajo estudio.

A pesar de haberse identificado 12 SNPs asociados al EBV para producción de leche, debe tenerse en cuenta que los valores p corregidos resultantes del análisis de asociación que los caracterizan son algo mayores que los valores p alcanzados en análisis similares reportados en literatura ($p < 5 \times 10^{-5}$, Minozzi y col., 2011; Minozzi y col., 2013; Alpay y col., 2014). Esto puede deberse al número de animales con el que se trabajó, a los umbrales establecidos para realizar el control de calidad de los genotipos, o a la estimación de los EBVs, la cual incorporó fenotipos de animales cuyas relaciones de parentesco no fueron verificadas por no haber sido genotipificados mediante el *chip* de mediana densidad utilizado. No obstante, García-Ruiz y col. (2014) utilizaron el mismo valor umbral que el usado en este trabajo ($p < 0,001$) en un análisis de asociación con caracteres de producción y composición de leche y recuento de células somáticas sobre bovinos de raza Holando, hallando 251 marcadores significativamente asociados a los caracteres bajo estudio.

Recientemente, Ekine y col. (2014) realizaron simulaciones de análisis de asociación con pedigrís humanos y de cerdo bajo cuatro escenarios diferentes y concluyeron que los valores de cría estimados a partir de información familiar no deberían usarse en GWAS por reducir su poder, aumentar su tasa de falsos positivos y subestimar los efectos de los QTLs hallados en estos análisis. No obstante, ninguno de los cuatro modelos utilizados en su trabajo coincide con el que se usó en este trabajo de tesis dado que, aun el más parecido, utiliza la estrategia GRAMMAR del paquete GenABEL de R para estimar los valores de cría (a través de un modelo poligénico, en la primera etapa) y luego usa los EBVs obtenidos como fenotipo en el estudio de asociación (en la segunda etapa). Mientras que, en este trabajo de tesis, los EBVs, estimados

de forma independiente a través de un modelo lineal mixto, fueron sometidos a un análisis de asociación que, además de considerar el efecto poligénico, incluyó la primera componente principal del análisis de componentes principales de los genotipos como covariable, y se hizo mediante la estrategia FASTA en lugar de la GRAMMAR de GenABEL. De todas formas, Ekine y col. (2014) marcan una diferencia entre los casos en los que los individuos tienen asociado un valor fenotípico propio, y aquellos en los que la medida de su valor fenotípico es indirecta por establecerse a partir del fenotipo de la progenie. Sugieren entonces, que en este último caso, como es el del ganado lechero, la contribución al EBV de familiares con menor grado de parentesco que el de sus crías directas no es significativa y que por lo tanto, el uso de los EBVs como fenotipo en los estudios de asociación no sería tan perjudicial como el de los casos que presentan en su trabajo.

En el capítulo siguiente se discutirán las estrategias seguidas para la selección de animales y regiones a secuenciar, en el intento de profundizar el conocimiento de las regiones que contienen los polimorfismos responsables o contribuyentes a la generación del rasgo estudiado. Las regiones a secuenciar seleccionadas fueron aquellas que comprenden o se hallan cercanas a los SNPs asociados al carácter.

CAPÍTULO IV

SECUENCIACIÓN DE REGIONES CANDIDATAS EN
ANIMALES CON FENOTIPOS CONTRASTANTES

En el cuarto y último capítulo de este trabajo de tesis se explorará, por medio de la secuenciación de regiones específicas, la existencia de diferencias a nivel del ADN (*i.e.* nuevos SNPs y/o haplotipos) que pudieran asociarse con las diferencias observadas en los EBVs para producción de leche acumulada a 305 días.

Dado que en este capítulo se utilizan conceptos enunciados y discutidos en los capítulos anteriores, éste no contará con una introducción teórica como los otros. Se dará comienzo, por lo tanto, con los objetivos planteados y se continuará con los materiales y métodos utilizados, donde se expondrán las estrategias seguidas para la selección de animales y de regiones candidatas a secuenciar. Luego se mostrarán las similitudes y diferencias de las secuencias obtenidas entre los animales de bajo y alto EBV para producción de leche y, finalmente, en la discusión de los resultados obtenidos se expondrán las conclusiones alcanzadas.

OBJETIVOS ESPECÍFICOS.

- Seleccionar los animales con fenotipos extremos de EBVs para producción de leche acumulada a 305 días durante la primera lactancia considerando dos situaciones, dentro de familias y entre familias.
- Seleccionar regiones candidatas a secuenciar, en base al resultado obtenido del estudio de asociación del capítulo anterior.
- Secuenciar los animales en las regiones seleccionadas y por comparación de las secuencias obtenidas, hacer un análisis exploratorio en busca de polimorfismos que pudieran contribuir a las diferencias en los valores de EBV para producción de leche de la población bajo estudio.

MATERIALES Y MÉTODOS.

1. Selección de animales con fenotipos extremos.

La selección de animales con fenotipos extremos, es decir, con valores de cría ubicados en las colas opuestas de la curva de distribución de EBVs, se realizó inter e intrafamiliarmente.

1.A. Selección interfamiliar.

En base a la distribución de los EBVs de las 1.628 vacas con registros productivos para la primera lactancia con las que se trabajó en el capítulo anterior, se determinó qué animales presentaban valores de EBV por encima de 3 desvíos estándar (DE), así como también aquellos con valores comprendidos entre 2 y 3 DE, hacia cada lado de la media. De esta preselección se escogieron, posteriormente, cantidades equivalentes de animales con alto y bajo EBV para producción de leche de primera lactancia acumulada a 305 días.

1.B. Selección intrafamiliar.

Para cada una de las familias con hijas con valores de cría ubicados por encima de 2 DE hacia cada lado de la media de la distribución de los EBVs de las 1.628 vacas mencionadas previamente, se graficó su propia distribución de valores de EBV. Luego se calculó la media y el desvío estándar de cada una de estas distribuciones y se determinaron los animales con valores de EBV por encima de 2 DE hacia cada lado de la media, para cada una de estas familias. Finalmente, se seleccionó una familia representativa del extremo de valores negativos de EBV y otra, representativa del extremo de valores positivos de EBV de la distribución interfamiliar. Esta selección se basó en el número de hijas de cada familia que presentaba valores de EBV por encima de 2 DE a cada lado de la media.

2. Selección de regiones a secuenciar.

Como punto de partida para la determinación de las regiones a secuenciar se utilizaron los SNPs y las regiones flanqueantes a los mismos, que fueron detectados como significativamente asociados al fenotipo estudiado en el capítulo anterior. En primer lugar, se ubicaron en el contexto génico dichos SNPs. Para ello se utilizó la base de datos SNPchiMp versión 2 (Nicolazzi y col., 2014), la cual redirecciona al usuario a Ensembl (www.ensembl.org) o al NCBI (www.ncbi.nlm.nih.gov) a fin de visualizar el SNP buscado en el genoma y establecer si se encuentra dentro o se halla cercano a algún gen de interés. Luego, se determinó qué genes se hallaban a ± 100 Kb de dichos SNPs, por considerar que a esta distancia se mantiene el desequilibrio de ligamiento, como nuestro grupo probó en el trabajo de Nani y col. (2015).

Posteriormente, se realizó una búsqueda bibliográfica para determinar la función y patrón de expresión de los genes hallados a ± 100 Kb de los SNPs significativamente asociados al EBV para producción de leche y se buscó, asimismo, si existían trabajos publicados en los que se reportara una asociación entre polimorfismos presentes en dichos genes y el fenotipo estudiado. Por otro lado, estos genes fueron sometidos a dos tipos de análisis con el programa *Ingenuity® Pathway Analysis* (IPA®): un *Pathway Analysis* (que conecta las vías metabólicas en las que intervienen los genes analizados) y un *Network Analysis* (que conecta los genes

analizados en base a funciones comunes). Con la información obtenida se elaboró un orden de prioridades de genes y regiones a secuenciar.

3. Secuenciación.

3.A. ADN de animales con fenotipos extremos.

El ADN utilizado fue extraído a partir de muestras de sangre fresca anticoaguladas con EDTA, mediante *kit* comercial (AxyPrep Blood Genomic DNA Miniprep Kit, Axygen Biosciences; Union City, California, EE.UU.) según el protocolo provisto por el fabricante. Posteriormente, su concentración y calidad se determinaron a través de la lectura de su absorbancia a 260 nm y la relación de absorbancias 260nm/280nm y 260nm/230nm respectivamente, utilizando un espectrofotómetro NanoDrop 1000 Thermo Scientific (Wilmington, Delaware, EE.UU.).

3.B. Reacción de secuenciación.

En primer lugar, se diseñaron *primers* (sobre la secuencia del ensamblado UMD3.1) para amplificar las regiones seleccionadas previamente. Luego se utilizaron entre 7 y 50 ng de ADN genómico de cada uno de los animales seleccionados, como material de partida en la reacción de PCR. La secuenciación de las regiones seleccionadas se realizó posteriormente mediante la utilización de la química BigDye® Terminator Cycle Sequencing Kit de Applied Biosystems (Foster City, California, EE.UU.). La descripción de la técnica así como la lista de los *primers* utilizados en las reacciones de PCR y secuenciación se detallan en el Anexo IX.

Los productos de la reacción de secuenciación se resolvieron utilizando un secuenciador de capilares automático ABI 3130xl (Applied Biosystems, California, EE.UU.). Las secuencias obtenidas se procesaron utilizando el programa PreGap4 (Staden Package; Staden y col., 1998) con el fin de eliminar regiones de baja calidad. Posteriormente se ensamblaron mediante el programa Gap4 (Staden Package; Staden y col., 1998), obteniendo finalmente una secuencia consenso, corregida para cada fragmento y animal.

4. Comparación de las secuencias obtenidas.

Las secuencias obtenidas de los animales con un alto EBV para producción de leche fueron comparadas con aquellas obtenidas de los animales que presentaban bajo EBV para el mismo fenotipo, tanto entre como dentro de familias, a través de la utilización del programa CLUSTAL Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>), en un análisis exploratorio que intentaba hallar los polimorfismos que pudieran contribuir a la generación del rasgo estudiado. Para cada región analizada, en el alineamiento se incluyó como secuencia de referencia la que figura en la base de datos del NCBI (www.ncbi.nlm.nih.gov) correspondiente al ensamblado UMD 3.1 y como SNPs de referencia aquellos reportados en Ensembl (www.ensembl.org). En el análisis, se determinó si los SNPs previamente reportados estaban presentes en la población bajo estudio, si en la población existía algún SNP nuevo y si un alelo dado de algún SNP mostraba mayor frecuencia de aparición entre los animales de un fenotipo particular (alto o bajo EBV para producción de leche acumulada a 305 días).

RESULTADOS.

1. Selección de animales con fenotipos extremos.

La selección de animales con valores de cría ubicados en los extremos de la distribución fenotípica, se realizó inter e intrafamiliarmente. A continuación se exponen los resultados obtenidos en cada caso.

1.A. Selección interfamiliar.

A partir de los EBVs para producción de leche de las 1.628 vacas con registros productivos para la primera lactancia estimados en el capítulo anterior, se graficó su distribución (Figura IV.1) y se calculó su media y su desvío estándar, siendo éstos -10,66 y 156,15 Kg, respectivamente.

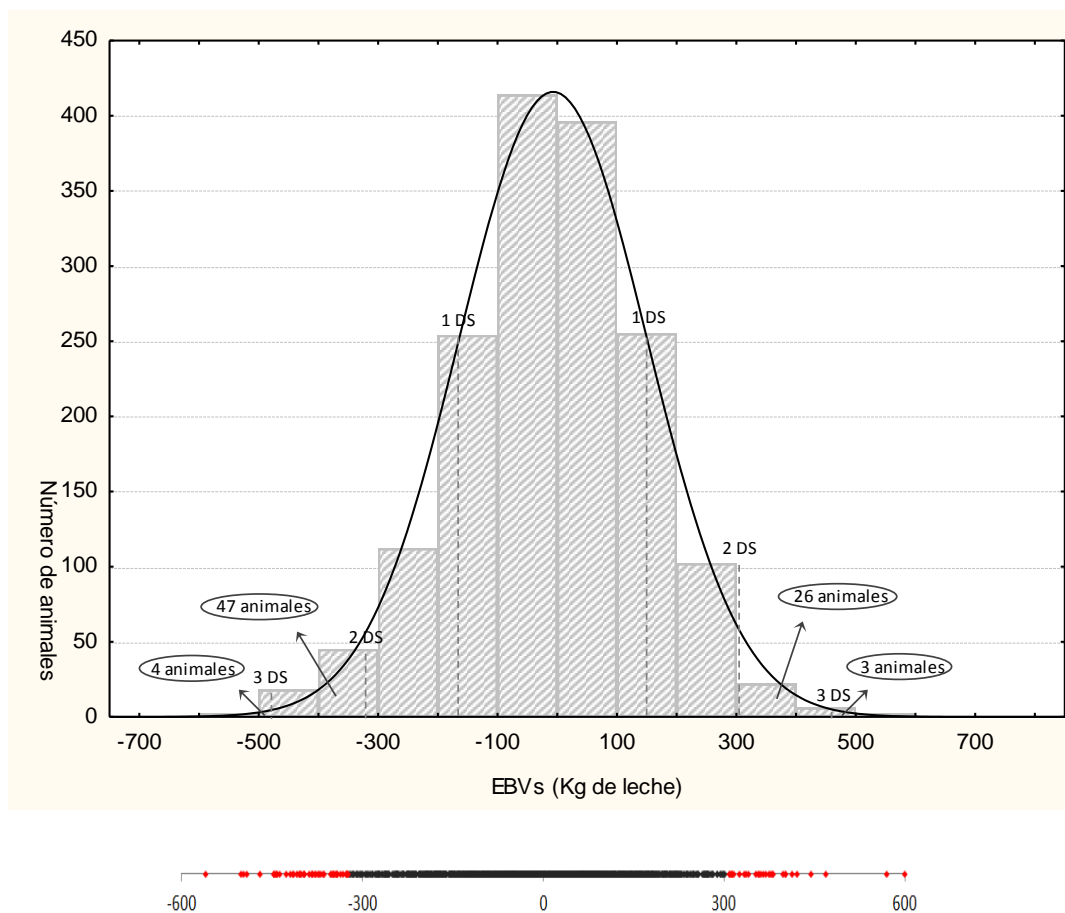


Figura IV.1. Distribución de EBVs interfamiliar.

Histograma de los valores de EBVs correspondientes a las 1.628 vacas con registros productivos para la primera lactancia. Sobre éste, se indica el número de animales con EBVs por encima de 2 y 3 DE hacia cada lado de la media. La curva normal a la cual se ajustan los valores se indica con trazo negro continuo.

Por debajo del histograma se grafica la distribución de valores de EBV, en la que cada punto corresponde a un animal diferente. En negro se indican las vacas con valores de cría ubicados entre -2DE y 2DE respecto de la media, y en rojo los animales con valores de EBV por encima y debajo de 2DE y -2DE, respectivamente.

En el extremo izquierdo de la distribución (EBVs negativos), se observó que 4 vacas presentaban valores de EBV por debajo de 3 DE, mientras que 47 presentaban valores entre 2 y 3 DE. Por otro lado, entre los animales ubicados en el extremo derecho de la distribución, 3 superaban los 3 DE, mientras que 26 presentaban valores entre 2 y 3 DE. Las familias a las que pertenecían las 51 vacas con bajo EBV así como las 29 vacas con alto EBV para producción de leche se indican en la Tabla IV.1.

Luego, se seleccionaron 24 vacas con los valores de EBV más alejados de la media para su secuenciación en regiones específicas. Es decir, se escogieron las 12 que presentaban los menores valores de cría y las 12 con los mayores valores de cría. El criterio utilizado para escoger este número de animales se basó en considerar los percentiles 0,5 y 99,5 de la distribución (lo que supone como mínimo 9 vacas de bajo EBV y 9 vacas de alto EBV).

1.B. Selección intrafamiliar.

Con el fin de realizar la selección de animales según esta estrategia se determinaron las familias que conformaban los grupos de animales de bajo (inferior a 2 DE) y alto (superior a 2 DE) EBV para producción de leche en la distribución de la Figura IV.1. A su vez, se computó el número de hijas de cada una de estas familias presente en ambos grupos (Tabla IV.1).

Bajo EBV para producción lechera		Alto EBV para producción lechera	
Familia (toro)	Número de hijas	Familia (toro)	Número de hijas
121	1	121	2
128	1	128	1
145	1	189	9
146	1	223	1
147	1	237	14
148	41	238	2
221	5		

Tabla IV.1. Composición de los grupos de animales de bajo y alto EBV para producción de leche de primera lactancia.

Familias que conforman los grupos de alto y bajo EBV ($\pm 2DE$) para producción lechera de la población analizada y número de hijas correspondiente a cada familia.

Para cada una de las familias que mejor representaban cada grupo de animales se graficó la distribución de los EBVs de todas sus hijas (incluyendo en el gráfico las hijas con valores de EBV comprendidos entre -2 DE y 2 DE respecto de la media en la Figura IV.1). Se graficaron, entonces, los histogramas correspondientes a los EBVs de las hijas de los toros 148 y 221 (que representan al grupo de animales de bajo EBV) así como de los toros 189 y 237 (que representan al grupo de animales de alto EBV). Luego se calcularon la media y el desvío estándar de estas distribuciones y se determinó el número de hijas de cada una de estas familias con valores de EBV por encima de 2 DE hacia cada lado de la media. Dado que este número fue mayor para las familias de los toros 148 y 189 que para las de los toros 221 y 237, se decidió trabajar con las primeras dos familias. Las distribuciones de los EBVs para producción de leche de las hijas de los toros 148 y 189 se exponen en la figura IV.2.

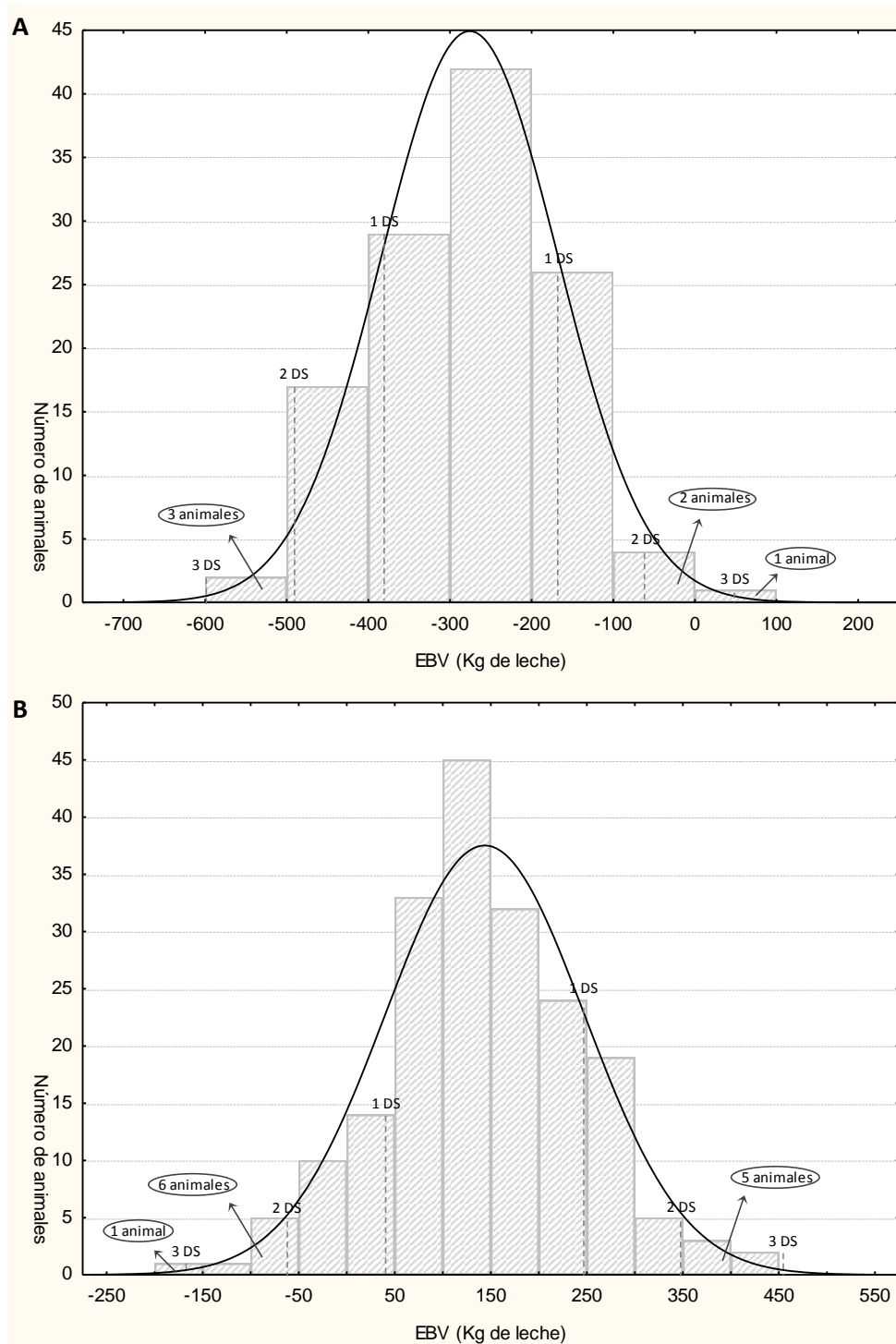


Figura IV.2. Distribución de EBVs intrafamiliar.

Histogramas de los EBVs para producción de leche durante la primera lactancia correspondientes a las hijas de los toros de RP 148 (A) y RP 189 (B). Se indica el número de animales con EBVs por encima de 2 y 3 DS hacia cada lado de la media. En ambos casos, la curva normal a la cual se ajustan los valores se muestra con trazo negro continuo.

Para la secuenciación en regiones específicas, se escogieron las 6 vacas con EBVs más extremos de la familia del toro 148 (3 con EBV alto y 3 con EBV bajo), mientras que fueron 10

las hijas del toro 189 seleccionadas (las 5 con valores más extremos hacia cada lado de la media).

En resumen, considerando las distribuciones de los valores de cría tanto inter como intrafamiliares, se seleccionaron 34 animales para la posterior secuenciación de regiones específicas (Figura IV.3).

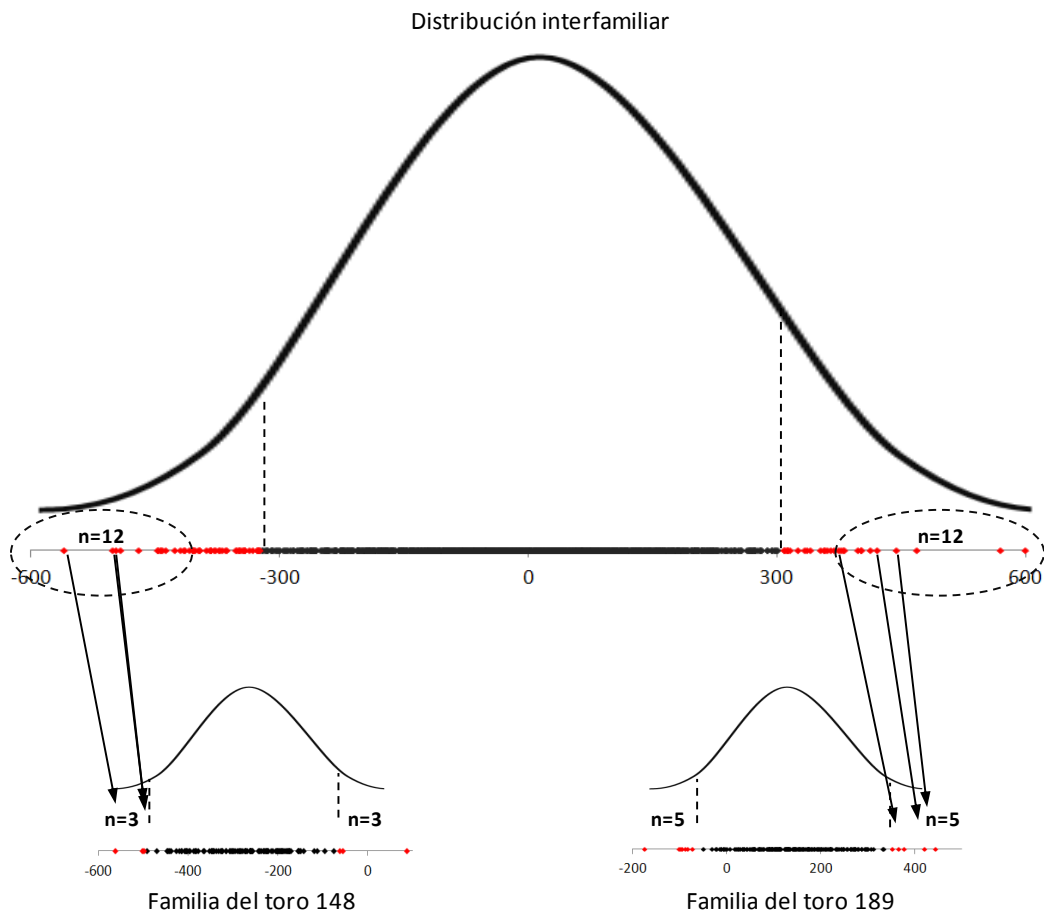


Figura IV.3. Animales seleccionados para la secuenciación de regiones específicas.

Cantidad de animales seleccionados a partir de las distribuciones interfamiliar e intrafamiliares correspondientes a las familias de los toros 148 y 189. En todos los casos, se indica con rojo los animales con valores de cría por encima y por debajo de 2DE y -2DE, respectivamente. De la distribución interfamiliar fueron seleccionadas las 12 vacas con valores más extremos de EBV hacia cada lado de la media, mientras que de las distribuciones de las familias de los toros 148 y 189 se seleccionaron las 3 y 5 vacas con valores más extremos de EBV hacia cada lado de la media, respectivamente. Las tres vacas de EBV más bajo seleccionadas de la distribución de la familia del toro 148 están incluidas en el grupo de animales seleccionado de la cola izquierda de la distribución interfamiliar. Mientras que tres de las cinco vacas seleccionadas de la distribución de la familia del toro 189 están incluidas en el grupo de animales seleccionado de la cola derecha de la distribución interfamiliar.

2. Selección de genes y regiones génicas a secuenciar.

Los 12 SNPs que resultaron significativamente asociados al EBV para producción de leche al realizar el análisis de asociación en el Capítulo III se ubicaron en el contexto génico mediante la utilización de la base de datos SNPchiMp versión 2 (Nicolazzi y col., 2014). La visualización de los SNPs en el genoma permitió establecer si se encontraban dentro o cercanos a algún gen de interés. Los genes cercanos que se tuvieron en cuenta fueron aquellos localizados a ± 100 Kb de los SNPs analizados. En la Tabla IV.2 se exponen los genes cercanos o aquellos que contienen los SNPs que resultaron asociados al carácter bajo estudio.

Además, con el objetivo de buscar más información que pudiera vincular estos genes y/o las funciones de las proteínas codificadas por los mismos entre sí, se utilizó el programa *Ingenuity® Pathway Analysis* (IPA®). Mediante este programa se analizaron los genes incluidos en la Tabla IV.2, así como otros 5 genes que, si bien no se encuentran separados por menos de 100 kb de ninguno de los SNPs asociados al fenotipo, se localizan cercanos a los mismos, entre los SNPs BTB-01301015 y BTB-00043331 (ambos separados entre sí por 531 kb). Estos 5 genes se mencionan a continuación:

- *COL6A1: collagen, type VI, alpha 1* (BTA 1: 147.404.396-147.423.644)
- *COL6A2: collagen, type VI, alpha 2* (BTA 1: 147.542.825-147.572.432)
- *FTCD: formiminotransferase cyclodeaminase* (BTA 1: 147.582.685-147.598.147)
- *SPATC1L: spermatogenesis and centriole associated 1-like* (BTA 1: 147.600.282-147.615.974)
- *LSS: lanosterol synthase (2,3-oxidosqualene-lanosterol cyclase)* (BTA 1: 147.623.691-147.654.154)

Se hicieron dos tipos de análisis: *Pathway Analysis* (conecta las vías metabólicas en las que intervienen los genes analizados) y *Network Analysis* (conecta los genes analizados en base a funciones comunes). El *Pathway Analysis* no dio resultados concluyentes: habiendo analizado pocos genes, ningún agrupamiento entre ellos resultó significativo. No obstante, el *Network Analysis* agrupó a 6 de los 20 genes analizados (*LSS*, *FTCD*, *S100B*, *AKR1C3*, *COL6A1* y *COL6A2*) bajo la calificación “Desórdenes hereditarios. Desórdenes esqueléticos y musculares. Producción de energía.” (Figura IV.4).

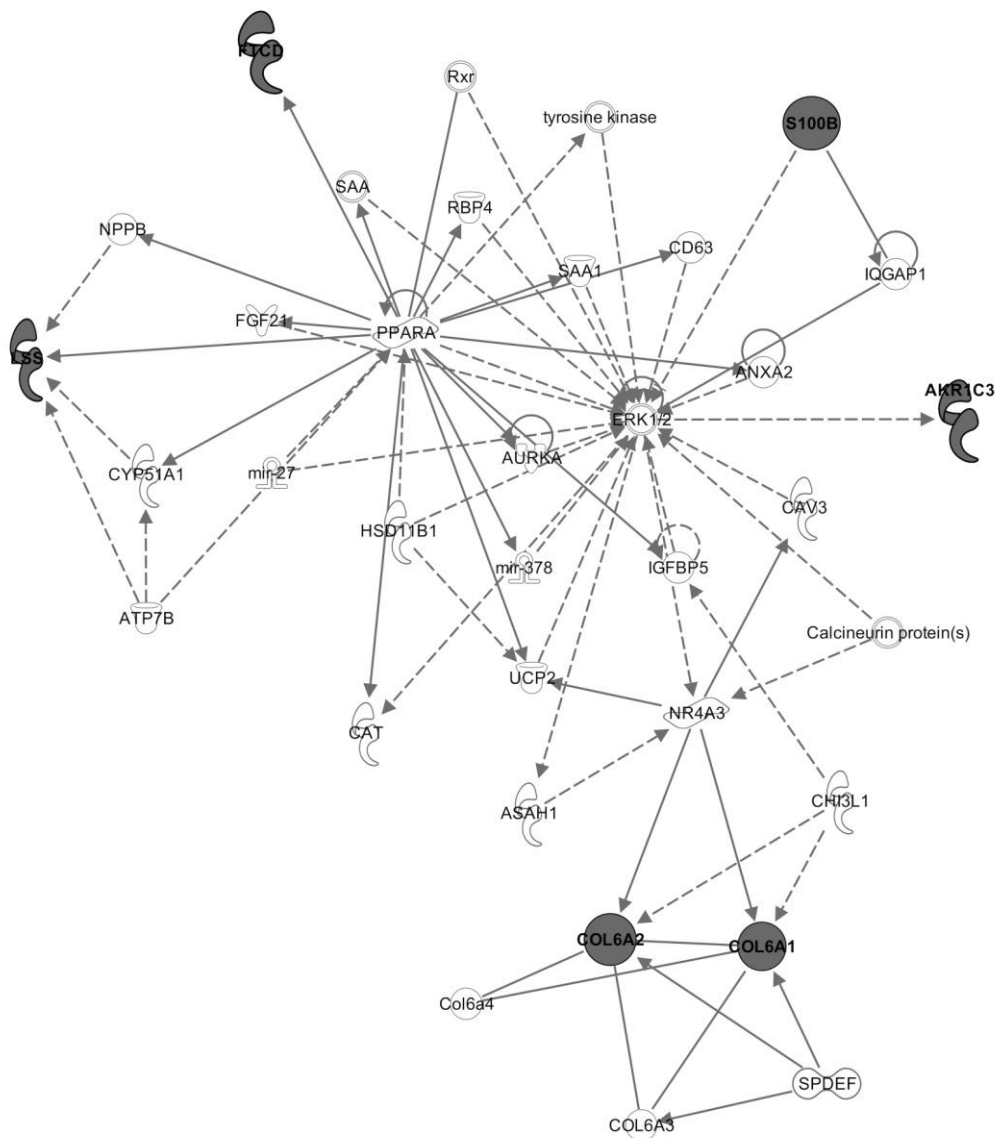
SNP	rs	Ubicación	Localización	Gen
BTB-00030894	rs43237323	1:61749066	intergénica	
Hapmap51585-BTA-32601	rs41629099	1:62104120	intergénica	<i>AKR1C1</i> (1: 62,162,949-62,171,172)
ARS-BFGL-NGS-85105	rs110823196	1:55832662	intergénica	
ARS-BFGL-NGS-94997	rs43239003	1:59754059	intergénica	<i>ZBTB20</i> (1: 59,420,222-59,721,760)
BTB-00043033	rs43250616	1:148009894	génica	<i>S100B</i> (1: 148,009,651-148,016,981); <i>DIP2A</i> (1: 147,856,757-147,977,430); <i>PRMT2</i> (1: 148,057,295-148,082,061)
BTB-01301015	rs42429246	1:147255491	génica	<i>PCBP3</i> (1: 147,132,679-147,359,843)
Hapmap50588-BTA-118970	rs41622951	24:18599946	intergénica	
BTB-00043331	rs43253207	1:147787147	génica	<i>PCNT</i> (1: 147,736,441-147,840,896); <i>YBEY</i> (1: 147,707,850-147,712,115)
ARS-BFGL-NGS-47270	rs42501206	20:46500354	intergénica	
Hapmap59475-rs29014975	rs29014975	1:60411957	intergénica	
ARS-BFGL-NGS-70977	rs109351365	23:42647945	intergénica	<i>RANBP9</i> (23: 42,736,705-42,795,080); <i>CCDC90A</i> (23: 42,547,988-42,567,105)
ARS-BFGL-NGS-15262	rs111013945	23:31312887	intergénica	<i>HMGNA4</i> (23: 31,322,596-31,330,574); <i>BTN</i> (23: 31,343,551-31,373,121); <i>BTN1A1</i> (23: 31,361,715-31,367,294); <i>ABT1</i> (23: 31,278,371-31,281,323); <i>ZNF322A</i> (23: 31,206,277-31,216,020)

Tabla IV.2. Genes cercanos a los SNPs asociados a la producción de leche.

En la tabla se indican los SNPs ordenados según valores decrecientes de $-\log(\text{valor-}p_{\text{corregido}})$, su código rs, su localización en el genoma bovino según el ensamblado UMD 3.1 (BTA:pb), su ubicación génica o intergénica y los genes que se localizan cerca de ellos. En los casos en que los SNPs se hallen dentro de un gen, se indica el nombre del gen correspondiente en negrita.

Abreviaturas: *AKR1C1*: aldo-keto reductase family 1, member C1 (dihydrodiol dehydrogenase 1; 20-alpha (3-alpha)-hydroxysteroid dehydrogenase)-like, *ZBTB20*: zinc finger and BTB domain containing 20, *S100B*: S100 calcium binding protein B, *DIP2A*: disco-interacting protein 2 homolog A, *PRMT2*: arginine N-methyltransferase 2, *PCBP3*: poly(rC)-binding protein 3, *PCNT*: pericentrina, *YBEY*: ybeY metalloproteinase (putative), *RANBP9*: RAN binding protein 9, *CCDC90A*: coiled-coil domain-containing protein 90A, mitochondrial, *HMGNA4*: high mobility group nucleosome-binding domain-containing protein 4, *BTN*: butyrophilin, *BTN1A1*: butyrophilin subfamily 1 member A1, *ABT1*: activator of basal transcription 1, *ZNF322A*: zinc finger protein 322.

Path Designer Network 1



© 2000-2014 Ingenuity Systems, Inc. All rights reserved.

Figura IV.4. Gráfico correspondiente al Network Analysis realizado por IPA®.

En la figura se indican en color gris los genes cercanos a los SNPs que resultaron significativos en el análisis de asociación, que fueron agrupados por IPA®.

Se indagó luego, en bibliografía, la función y el perfil de expresión de los genes localizados a ± 100 Kb de los SNPs significativos, así como de los que resultaron agrupados en el análisis por IPA®, en busca de algún indicio de su relación directa o indirecta con la producción de leche.

S100B es una proteína de unión a calcio, originalmente aislada del sistema nervioso, donde se halla en alta concentración en las células gliales. Está presente en distintos fluidos biológicos, como en el líquido cefalorraquídeo, sangre, orina, líquido amniótico y leche del ser humano (Gazzolo y col., 2003). Su concentración en la leche es marcadamente más elevada que en el resto de los fluidos, siendo mayor en la leche madura que en el calostro (Gazzolo y

col., 2004; Li y col., 2011). En la leche, esta proteína es sintetizada en células epiteliales mamarias y linfocitos, y actuaría como una citoquina con rol neurotrófico que ejercería un efecto estimulante de la maduración cerebral en niños alimentados con leche materna. En la leche humana, su concentración es tres veces más alta que en la de bovinos, mientras que éstos son los mamíferos que presentan una mayor concentración de S100B en comparación con la de patos, ovejas y cabras (Galvano y col., 2009).

PRMT2 (Protein arginine N-methyltransferase 2, también denominada **HRMT1L1**) es una enzima que cataliza la transferencia de grupos metilo desde la S-adenosilmetionina a residuos de arginina de numerosos sustratos. En bovinos se ha reportado su expresión en corazón, hígado, bazo, pulmón riñón, músculo, intestino y glándula mamaria, siendo mayor en este último tejido. Más aun, se ha observado una mayor expresión de *PRMT2* en la glándula mamaria con mastitis o con cáncer que en el tejido sano (Li y col., 2014; Zhong y col., 2012). Entre los procesos celulares en los que participa se incluyen la función pulmonar, la respuesta inflamatoria, la promoción de la apoptosis y la regulación de la señalización de la leptina. En un análisis de asociación con caracteres de producción lechera en ganado Holstein de China, Li y col. (2014) encontraron asociaciones entre dos SNPs localizados en *PRMT2*, uno intrónico y otro correspondiente a la región 3'-UTR del gen, y el contenido de proteína y la producción de leche, respectivamente.

LSS (Lanosterol synthase) es una enzima que cataliza la conversión del 2,3-oxidoescualeno en lanosterol, un paso clave en la biosíntesis del colesterol y de otros esteroides. Se ha descrito en la membrana de los glóbulos de grasa de la leche, siendo mayor su expresión en aquellos glóbulos aislados el día 7 de lactancia que en el calostro del primer día post-parto en bovinos (Reinhardt y Lippolis, 2008).

Los genes **BTN (Butyrophilin)** constituyen un subgrupo de al menos 10 genes de la familia de las Inmunoglobulinas que fueron identificados en humanos, ratones y otras especies. Entre ellos, se pueden distinguir los siguientes genes: *BTN1A1*, altamente expresado en el epitelio secretor de la glándula mamaria durante la lactancia y también, aunque en bajos niveles, en otros tejidos como timo y bazo (Smith y col., 2010); *BTL-II*, expresado predominantemente en el músculo esquelético y el intestino; *ERMAP*, en células eritroides; *BTN2A1* y *2* y *BTN3A1*, *2* y *3* ubicuamente expresados. Las evidencias sugieren que los dominios estructurales de las proteínas BTN tendrían tanto funciones universales como específicas de tejido (Ogg y col., 2004).

La proteína **BTN1A1** fue originalmente denominada "butirofilina" para reflejar su asociación específica con la grasa de la leche. Es una proteína transmembrana expresada en la superficie apical de las células epiteliales secretoras mamarias y la principal proteína que constituye la membrana de los glóbulos grasos de la leche bovina. La expresión del gen que la codifica se correlaciona con el inicio de la secreción de grasa en la leche hacia el final de la preñez y se mantiene a lo largo de la lactancia (Mather y Jack, 1993). Aunque su función no está totalmente elucidada, su perfil de expresión sugiere un rol importante durante la lactancia. Ogg y col. (2004) llevaron a cabo estudios con ratones deficientes en *BTN1A1*, observando seriamente comprometida la regulación de la secreción de glóbulos de grasa en la leche, así como la supervivencia y el peso al destete de las crías de estos animales. Los

resultados obtenidos por este grupo de trabajo sugieren la posibilidad de que *BTN1A1* actúe como una proteína estructural implicada directamente en el transporte y/o secreción de las gotas de lípidos a la leche o como un receptor de señalización a través de su unión a la enzima xantina deshidrogenasa/oxidasa. Por otro lado, Smith y col. (2010) han reportado la capacidad de *BTN1A1* de inhibir las células T, por lo que argumentan que esta proteína también podría tener un papel importante en la inmunidad neonatal. El efecto del gen *BTN1A1* sobre caracteres relacionados con la producción de leche en bovinos no ha sido ampliamente estudiado, aunque se ha reportado una asociación entre el polimorfismo K468R del gen y el contenido proteico de la leche en bovinos Holstein-Friesian, así como una tendencia a dicha asociación en Jersey (Suchocki y col., 2010). En el mismo trabajo se evidenció un efecto epistático significativo en el contenido de proteína en la leche entre los genes *LEPR* y *BTN1A1* en la raza Holstein-Friesian.

La información obtenida sobre el resto de los genes no sugiere una relación clara con la glándula mamaria ni con las vías metabólicas de producción de leche ni de síntesis de alguno de sus componentes. Por lo tanto, se seleccionaron tres genes para su posterior secuenciación: *PRMT2*, *BTN1A1* y *S100B*.

El criterio de selección de regiones específicas dentro de estos genes para su posterior secuenciación se basó en la asignación de mayor prioridad a aquellas regiones que contuvieran exones y/o marcadores polimórficos previamente reportados en la literatura como asociados al carácter fenotípico estudiado, así como también a aquellos genes cuya expresión proteica mostrara sólida evidencia de estar implicada en características productivas o composicionales de la leche. De esta forma, se seleccionó la región del gen *PRMT2* que comprende parte de la región 3'UTR y una porción de su intrón flanqueante, que contiene al SNP *rs42869038*, reportado como asociado a la producción de leche bovina (Li y col., 2014). Del gen *BTN1A1* fueron seleccionadas dos regiones: la primera, *BTN1A1_A*, que consiste en parte de la región 5'UTR, el exón 1, parte del exón 2 y el intrón ubicado entre dichos exones, y la segunda, *BTN1A1_B*, que incluye el exón 3 y parte de los intrones que lo flanquean. Finalmente, del gen *S100B* se seleccionó una única región para su secuenciación, la que comprende el exón 2, parte de la región 3'UTR y parte del intrón flanqueante.

3. Secuenciación.

Las reacciones de PCR mediante las cuales se amplificaron cada una de las regiones seleccionadas para secuenciar fueron exitosas, lo que demuestra que el diseño de los *primers* utilizados fue satisfactorio. Asimismo, las secuencias obtenidas de las regiones seleccionadas fueron en general de buena calidad para todos los animales analizados. En la Figura IV.5 se muestran, a modo de ejemplo, los amplicones obtenidos de la región *S100B* para 12 de los 34 animales seleccionados para secuenciar, así como también una porción de la secuencia de esta misma región obtenida para la vaca de RP 47995.

De las regiones *PRMT2*, *BTN1A1_A* y *BTN1A1_B* pudo obtenerse la secuencia de alrededor del 90% del fragmento amplificado por PCR, mientras que de la región *S100B* sólo se obtuvo el 76,6% del amplicón correspondiente. Este último valor es algo inferior a lo esperado, y se debe a que el fragmento del gen *S100B* amplificado contiene una región poli-A a una distancia de

583 bases del extremo 5', lo que dificulta la secuenciación de esta región, dando una secuencia de mala calidad que debe ser descartada. Por este mismo motivo sólo pudo utilizarse en la reacción de secuenciación el primer *forward*, dado que el *reverse* daba secuencias que resultaban de mala calidad prácticamente apenas comenzada la lectura (la región poli-A se halla a 84 pb del extremo 3' del fragmento amplificado). Por otro lado, las secuencias correspondientes a los *primers* fueron descartadas del análisis (para las cuatro regiones analizadas) por no poder buscarse mutaciones en estas regiones.

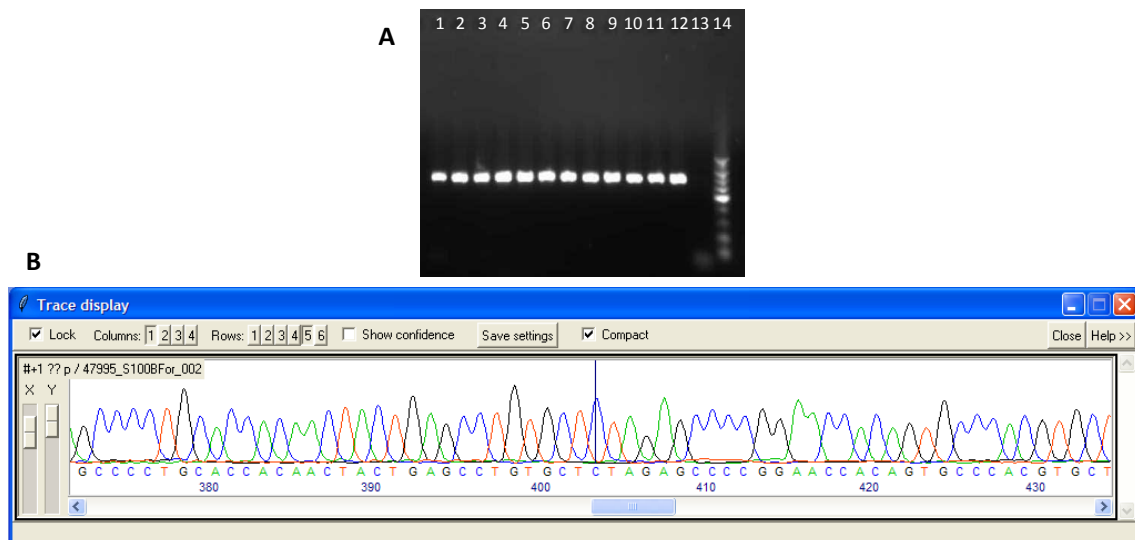


Figura IV.5. Visualización de productos de PCR y secuencias.

La figura **A** muestra los amplicones de 701 bp obtenidos de la región S100B para 12 animales (calles 1 a 12), el blanco de PCR (*non-template control*, calle 13) y el *standard* de PM utilizado como referencia (calle 14). La figura **B** representa la secuencia obtenida para una porción de 62 pb del gen *S100B* en la vaca de RP 47995.

4. Alineamiento y comparación de secuencias.

Para cada región analizada, se realizó el alineamiento de las secuencias de los animales seleccionados por poseer valores extremos del fenotipo estudiado, incluyéndose en el mismo la secuencia de referencia publicada en la base de datos del NCBI. Este análisis se realizó tanto inter como intrafamiliarmente. En primer lugar, la comparación de las secuencias obtenidas con la secuencia de referencia permitió determinar el número de SNPs previamente reportados presente en la población. No se halló, para ninguna de las regiones secuenciadas, ningún SNP nuevo presente en la población analizada. Luego, por comparación de las secuencias de los animales de alto EBV con las de los de bajo EBV se hizo un análisis exploratorio en busca de polimorfismos que podrían proponerse como candidatos a contribuir a las diferencias fenotípicas de la población. Los resultados obtenidos para cada una de las regiones secuenciadas se resumen en la Tabla IV.3.

Los tres SNPs hallados en la región seleccionada del gen *PRMT2* (rs42869037, rs42869038 y rs42869039) se localizan en la región 3'UTR, siendo uno de ellos el reportado por Li y col.

(2014) como asociado a la producción lechera (rs42869038). La localización en el genoma bovino, así como las frecuencias genotípicas y alélicas que caracterizan cada uno de los SNPs hallados en los animales seleccionados por presentar fenotipos extremos dentro de la distribución de EBVs interfamiliar se muestran en la Tabla IV.4.

Región	SNPs reportados	Alineamiento de secuencias		SNPs hallados
PRMT2	43	Interfamiliar		3
		intrafamiliar	toro 148	2
			toro 189	2
BTN1A1_A	59	Interfamiliar		13
		intrafamiliar	toro 148	13
			toro 189	12
BTN1A1_B	29	Interfamiliar		7
		intrafamiliar	toro 148	7
			toro 189	4
S100B	54	Interfamiliar		0
		intrafamiliar	toro 148	0
			toro 189	0

Tabla IV.3. Resultado del alineamiento de secuencias.

En la tabla se indica, para cada una de las regiones secuenciadas, el número de SNPs reportados en la base de datos Ensembl, así como el número de SNPs hallados luego de alinear y comparar las secuencias obtenidas tanto entre familias como dentro de las familias de los toros de RP 148 (Holando) y RP 189 (Jersey).

Por otro lado, al comparar las secuencias de la misma región, obtenidas de las 6 hijas del toro 148, se hallaron los SNPs rs42869037 y rs42869038, previamente reportados. Mientras que las 10 vacas seleccionadas hijas del toro 189 evidenciaron la presencia de los SNPs rs42869038 y rs42869039.

En las secuencias de la región BTN1A1_A de los animales seleccionados a partir de la distribución interfamiliar de EBVs se detectó la presencia de 13 SNPs (cuyos códigos *rs*, localización y frecuencias genotípicas y alélicas se indican en la Tabla IV.4), todos ellos reportados previamente, no habiéndose hallado ningún SNP nuevo en la población secuenciada. A modo de ejemplo, una porción del alineamiento de las secuencias obtenidas para esta región se expone en la Figura IV.6. Los mismos SNPs fueron detectados entre las vacas con fenotipos correspondientes a los extremos de la distribución de los EBVs de las hijas del toro de RP 148. Todos ellos, con excepción del SNP rs473043894, también se hallaron en las hijas secuenciadas de la familia del toro 189. Seis de los mencionados SNPs se ubican en exones del gen *BTN1A1* y, exceptuando el SNP rs132717963 que da lugar a una mutación sinónima, es decir, sin cambio del aminoácido codificado (prolina), el resto dan lugar a mutaciones con cambio de sentido puesto que las variaciones nucleotídicas que caracterizan a los SNPs provocan la aparición de codones que codifican para aminoácidos diferentes (Tabla IV.5).

Región	SNP	Variación nucleotídica	Ubicación	Frecuencias genotípicas		Frecuencias alélicas	
PRMT2	rs42869037	G/A	1:148081585 (3'UTR)	GG	0,04	G	0,05 / 0,14
				GA	0,12	A	0,95 / 0,86
				AA	0,83		
	rs42869038	C/T	1:148081691 (3'UTR)	CC	0,32	C	0,40 / 0,62
				CT	0,41	T	0,60 / 0,38
				TT	0,27		
rs42869039	G/A	1:148081719 (3'UTR)	GG	0,06	G	0,21 / 0,35	
			GA	0,47	A	0,79 / 0,65	
			AA	0,47			
BTN1A1_A	rs133113758	G/A	23:31362520 (exón)	GG	0,67	G	0,73 / 0,92
				GA	0,33	A	0,27 / 0,08
				AA	0		
	rs137057391	C/T	23:31362521 (exón)	CC	0,67	C	0,73 / 0,92
				CT	0,33	T	0,27 / 0,08
				TT	0		
	rs209464806	T/A	23:31362539 (exón)	TT	0,79	T	0,95 / 0,85
				TA	0,21	A	0,05 / 0,15
				AA	0		
	rs133513674	C/G	23:31362625 (intrón)	CC	0,67	C	0,73 / 0,92
				CG	0,33	G	0,27 / 0,08
				GG	0		
rs473043894	A/C	23:31362632 (intrón)	AA	0,92	A	0,95 / 0,96	
			AC	0,08	C	0,05 / 0,04	
			CC	0			
rs137658829	G/A	23:31362677 (intrón)	GG	0,64	G	0,71 / 0,92	
			GA	0,36	A	0,29 / 0,08	
			AA	0			
rs134495891	A/T	23:31362767 (intrón)	AA	0,64	A	0,71 / 0,92	
			AT	0,36	T	0,29 / 0,08	
			TT	0			

Tabla IV.4. Caracterización de los SNPs detectados en los animales seleccionados a partir de la distribución interfamiliar de EBVs.

SNPs detectados en cada una de las regiones secuenciadas junto a su código *rs*, la variación nucleotídica implicada, su localización en el genoma (BTA:pb), las frecuencias genotípicas halladas y las frecuencias alélicas que presentaron los animales de bajo (trazo normal) y alto (negrita) EBV para producción de leche. Los valores se redondearon a la segunda cifra decimal.

Región	SNP	Variación nucleotídica	Ubicación	Frecuencias genotípicas		Frecuencias alélicas	
BTN1A1_A	rs136752620	A/G	23:31362769 (intrón)	AA	0,64	A	0,71 / 0,92
				AG	0,36		
				GG	0	G	0,29 / 0,08
	rs135838239	C/A	23:31362930 (intrón)	CC	0,68		
				CA	0,32		
				AA	0	A	0,25 / 0,08
	rs134410305	C/G	23:31362933 (intrón)	CC	0,68		
				CG	0,32		
				GG	0	G	0,25 / 0,08
	rs109972486	C/A	23:31362975 (exón)	CC	0,24		
				CA	0,44		
				AA	0,32	A	0,42 / 0,65
	rs132717963	G/A	23:31362976 (exón)	GG	0,68		
				GA	0,32		
				AA	0	A	0,25 / 0,08
rs43706495	G/A	23:31363023 (exón)	GG	0,64	G		
			GA	0,36			
			AA	0	A	0,29 / 0,08	
rs137519381	T/G	23:31364336 (intrón)	TT	0,39			T
			TG	0,30			
			GG	0,30	G	0,50 / 0,43	
rs135176047	A/G	23:31364404 (intrón)	AA	0,32			A
			AG	0,40			
			GG	0,28	G	0,59 / 0,39	
rs43708474	G/C	23:31364476 (intrón)	GG	0,32			G
			GC	0,40			
			CC	0,28	C	0,59 / 0,39	
rs211141026	G/A	23:31364768 (exón)	GG	0,92			G
			GA	0,08			
			AA	0	A	0,05 / 0,04	
rs209257029	G/C	23:31364858 (exón)	GG	0,92			G
			GC	0,08			
			CC	0	C	0,05 / 0,04	

Tabla IV.4. Caracterización de los SNPs detectados en los animales seleccionados a partir de la distribución interfamiliar de EBVs (continuación).

Región	SNP	Variación nucleotídica	Ubicación	Frecuencias genotípicas		Frecuencias alélicas	
BTN1A1_B	rs43708439	T/A	23:31364916 (exón)	TT	0,40	T	0,41 / 0,61
				TA	0,24		
				AA	0,36	A	0,59 / 0,39
	rs208558228	G/A	23:31364917 (exón)	GG	0,92	G	0,95 / 0,96
				GA	0,08		
				AA	0	A	0,05 / 0,04

Tabla IV.4. Caracterización de los SNPs detectados en los animales seleccionados a partir de la distribución interfamiliar de EBVs (continuación).

En la segunda región secuenciada del gen *BTN1A1*, BTN1A1_B, se evidenció la presencia de 7 SNPs en *loci* previamente reportados (Tabla IV.4). Los mismos SNPs se hallaron presentes en las hijas con fenotipos extremos de la familia del toro de RP 148. Cuatro de estos SNPs se localizan en exones del gen *BTN1A1* y todos ellos dan lugar a mutaciones con cambio de sentido (Tabla IV.6). Por otro lado, entre las hijas secuenciadas de la familia del toro de RP 189 sólo se evidenció la presencia de los SNPs rs137519381, rs135176047, rs43708474 y rs43708439, estando sólo el último localizado en un exón del gen y dando lugar a la variación aminoacídica ácido aspártico/ácido glutámico.

Finalmente, en la base de datos Ensembl se reportan 54 SNPs localizados en la región seleccionada para secuenciar del gen *S100B*, sin embargo en los animales secuenciados no se detectó la presencia de ningún SNP.

```

RP 48519      TCCACATTCCGTCCGCCCTCTTGGTGTTTTGGCAGCTCCCTTTGACGTGATCGGACCCC
RP 54395      TCCACATTCCGTCCGCCCTCTTGGTGTTTTGGCAGCTCCCTTTGACGTGATCGGACCCC
RP 51695      TCCACATTCCGTCCGCCCTCTTGGTGTTTTGGCAGCTCCCTTTGACGTGATCGGACCCC
RP 48719      TCCACATTCCGTCCGMCCSTCTTGGTGTTTTGGCAGCTCCCTTTGACGTGATCGGACCCC
RP 54333      TCCACATTCCGTCCGMCCSTCTTGGTGTTTTGGCAGCTCCCTTTGACGTGATCGGACCCC
RP 52003      TCCACATTCCGTCCGCCCTCTTGGTGTTTTGGCAGCTCCCTTTGACGTGATCGGACCCC
RP 51533      TCCACATTCCGTCCGMCCSTCTTGGTGTTTTGGCAGCTCCCTTTGACGTGATCGGACCCC
RP 57327      TCCACATTCCGTCCGCCCTCTTGGTGTTTTGGCAGCTCCCTTTGACGTGATCGGACCCC
RP 54515      TCCACATTCCGTCCGMCCSTCTTGGTGTTTTGGCAGCTCCCTTTGACGTGATCGGACCCC
RP 52077      TCCACATTCCGTCCGCCCTCTTGGTGTTTTGGCAGCTCCCTTTGACGTGATCGGACCCC
RP 52137      TCCACATTCCGTCCGMCCSTCTTGGTGTTTTGGCAGCTCCCTTTGACGTGATCGGACCCC
RP 55743      TCCACATTCCGTCCGMCCSTCTTGGTGTTTTGGCAGCTCCCTTTGACGTGATCGGACCCC
-----
RP 46119      TCCACATTCCGTCCGCCCTCTTGGTGTTTTGGCAGCTCCCTTTGACGTGATCGGACCCC
RP 53815      TCCACATTCCGTCCGCCCTCTTGGTGTTTTGGCAGCTCCCTTTGACGTGATCGGACCCC
RP 42167      TCCACATTCCGTCCGCCCTCTTGGTGTTTTGGCAGCTCCCTTTGACGTGATCGGACCCC
RP 55131      TCCACATTCCGTCCGCCCTCTTGGTGTTTTGGCAGCTCCCTTTGACGTGATCGGACCCC
RP 55107      TCCACATTCCGTCCGCCCTCTTGGTGTTTTGGCAGCTCCCTTTGACGTGATCGGACCCC
RP 55897      TCCACATTCCGTCCGMCCSTCTTGGTGTTTTGGCAGCTCCCTTTGACGTGATCGGACCCC
RP 46607      TCCACATTCCGTCCGCCCTCTTGGTGTTTTGGCAGCTCCCTTTGACGTGATCGGACCCC
RP 54871      TCCACATTCCGTCCGCCCTCTTGGTGTTTTGGCAGCTCCCTTTGACGTGATCGGACCCC
RP 54225      TCCACATTCCGTCCGCCCTCTTGGTGTTTTGGCAGCTCCCTTTGACGTGATCGGACCCC
RP 57897      TCCACATTCCGTCCGCCCTCTTGGTGTTTTGGCAGCTCCCTTTGACGTGATCGGACCCC
RP 55133      TCCACATTCCGTCCGCCCTCTTGGTGTTTTGGCAGCTCCCTTTGACGTGATCGGACCCC
RP 47995      TCCACATTCCGTCCGMCCSTCTTGGTGTTTTGGCAGCTCCCTTTGACGTGATCGGACCCC
RP 53977      TCCACATTCCGTCCGCCCTCTTGGTGTTTTGGCAGCTCCCTTTGACGTGATCGGACCCC
BTN1_A_GenBank TCCACATTCCGTCCGCCCTCTTGGTGTTTTGGCAGCTCCCTTTGACGTGATCGGACCCC 538
***** ** *****

```

Figura IV.6. Alineamiento de secuencias de la región BTN1A1_A.

En la figura se expone, a modo de ejemplo, una porción del alineamiento de secuencias de la región secuenciada BTN1A1_A. La línea de guiones permite discriminar los animales de bajo (por encima) y alto (por debajo) EBV para producción de leche. La última secuencia corresponde a la secuencia de referencia publicada en la base de datos del NCBI. En la secuencia de referencia, los nucleótidos en azul indican la ubicación de los SNPs reportados en Ensembl para esta región (rs135838239 y rs134410305, presentes en la población secuenciada, y rs454841281, rs452104569, rs480976119, rs440818638, rs459373832 y rs448302814, ausentes en la población secuenciada). En rojo se indican los genotipos heterocigotas. Las estrellas significan identidad entre las secuencias alineadas en los *loci* en que se encuentran.

SNP	Variación aminoacídica	Transcripto
rs133113758	Ala10Thr	ENSBTAT00000045016
	Ala12Thr	ENSBTAT00000045021
rs137057391	Ala10Val	ENSBTAT00000045016
	Ala12Val	ENSBTAT00000045021
rs209464806	Phe16Tyr	ENSBTAT00000045016
	Phe18Tyr	ENSBTAT00000045021
rs109972486	Pro35Gln	ENSBTAT00000045016
	Pro37Gln	ENSBTAT00000045021
rs43706495	Arg51His	ENSBTAT00000045016
	Arg53His	ENSBTAT00000045021

Tabla IV.5. Variaciones aminoacídicas en la región BTN1A1_A.

Mutaciones con cambio de sentido (en los transcritos correspondientes) que surgen a partir de las variaciones nucleotídicas que caracterizan a cinco SNPs detectados en la población secuenciada en la región BTN1A1_A.

SNP	Variación aminoacídica	Transcripto
rs211141026	Arg181Lys	ENSBTAT00000045016
	Arg183Lys	ENSBTAT00000045021
rs209257029	Ser211Thr	ENSBTAT00000045016
	Ser213Thr	ENSBTAT00000045021
rs43708439	Asp230Glu	ENSBTAT00000045016
	Asp232Glu	ENSBTAT00000045021
rs208558228	Val231Ile	ENSBTAT00000045016
	Val233Ile	ENSBTAT00000045021

Tabla IV.6. Variaciones aminoacídicas en la región BTN1A1_B.

Mutaciones con cambio de sentido (en los transcritos correspondientes) que surgen a partir de las variaciones nucleotídicas que caracterizan a cuatro SNPs detectados en la población secuenciada en la región BTN1A1_B.

Posteriormente, y sólo para un análisis exploratorio, se calculó la diferencia entre las frecuencias alélicas halladas en los grupos de animales de bajo y alto EBV, para cada uno de los SNPs presentes en la población secuenciada, con el fin de determinar si un alelo dado de algún SNP mostraba mayor frecuencia de aparición entre los animales de un fenotipo particular (alto o bajo EBV para producción de leche acumulada a 305 días). A continuación se indican los nueve SNPs que evidenciaron una diferencia de al menos 0,2 entre las frecuencias mencionadas:

PRMT2: rs42869038

BTN1A1_A: rs137658829, rs134495891, rs136752620, rs109972486 y rs43706495

BTN1A1_B: rs135176047, rs43708474 y rs43708439

Para cada uno de estos SNPs se construyó una tabla de contingencia utilizando como variables los dos alelos del SNP y el EBV (bajo o alto) que presentaban los animales secuenciados. A continuación se planteó un contraste de hipótesis considerando:

- Hipótesis nula: no hay asociación entre las variables, es decir, el valor de cría para producción de leche es independiente de la portación de uno u otro alelo del SNP;
- Hipótesis alternativa: hay asociación entre las variables, es decir, el valor de cría que caracteriza a un animal es dependiente del alelo que presenta en el *locus* del SNP.

Las 9 tablas de contingencia se analizaron mediante el *test* exacto de Fisher (Fisher, 1922; Soper, 2015) o mediante el *test* de hipótesis χ^2 , según correspondía en base a los valores que presentaban las celdas de cada tabla. En todos los casos, los *tests* realizados condujeron a aceptar la hipótesis nula de independencia entre las variables consideradas (con un valor $p < 0,05$). Es decir, con un nivel de significación del 5% se tienen evidencias para concluir que no existe asociación entre el valor de cría para producción de leche y la portación de uno u otro alelo del SNP.

DISCUSIÓN.

En este capítulo se realizó un análisis exploratorio utilizando secuencias de regiones específicas de animales seleccionados, en el intento de hallar polimorfismos que pudieran vincularse con las diferencias observadas en dichos animales a nivel fenotípico.

Se seleccionaron 24 animales que presentaban EBVs para producción de leche acumulada a 305 días en los extremos de la distribución interfamiliar, construida a partir de 1.628 animales con registros productivos para la primera lactancia. Posteriormente, se seleccionaron las 6 vacas con EBVs más extremos de la familia del toro 148 (familia que representa al grupo de animales de bajo EBV de la distribución interfamiliar) y las 10 hijas con EBVs más extremos de la familia del toro de RP 189 (que representa al grupo de animales de alto EBV de la distribución interfamiliar). Dado que algunos de los animales seleccionados a partir de las distribuciones intrafamiliares ya habían sido escogidos de la distribución interfamiliar para la secuenciación de regiones específicas, el total de animales secuenciado fue 34.

Los genes considerados como potenciales candidatos para ser secuenciados en regiones específicas fueron aquellos localizados a una distancia de ± 100 Kb de los 12 SNPs que resultaron significativamente asociados al EBV para producción de leche en el capítulo anterior. Se consideró que hasta esta distancia se mantiene el desequilibrio de ligamiento en base a los resultados obtenidos por nuestro grupo en un trabajo previo sobre la misma población (Nani y col., 2015). En dicho trabajo se reportó la reducción del valor de r^2 promedio sobre todos los cromosomas con la distancia genética, hallándose un r^2 promedio de 0,40 para marcadores separados entre sí por hasta 10 kb, de 0,25 para marcadores entre 40 y 50 kb, y de 0,18 para aquellos separados por 100 kb. Más allá de esta distancia, el decaimiento del valor de r^2 fue más pronunciado. Estos valores son concordantes con algunos de los reportados en bibliografía. Sargolzaei y col. (2008) reportan considerable DL en ganado Holstein de Norteamérica a distancias menores a 100 kb. McKay y col. (2007) evaluaron el DL en ocho razas de ganado bovino (*Bos taurus*: Angus, Charolais, Dutch Black and White Dairy, Holstein, Japanese Black y Limousin y *Bos indicus*: Brahman y Nelore) utilizando 2.670 SNPs y hallaron un r^2 promedio de 0,15 - 0,20 al considerar pares de SNPs separados por una distancia de 100 kb. Por su parte, de Roos y col. (2008) evaluaron el DL en ganado de razas Holstein-Friesian, Angus y Jersey a través del cálculo de r^2 utilizando marcadores SNPs y hallaron un valor promedio de 0,14 a una distancia de 100 kb entre los marcadores. Asimismo, Marques y col. (2008) reportaron que el DL, analizado con algo más de 500 SNPs del BTA 14 de ganado Holstein y Angus, muestra niveles moderados hasta 100 kb y decae hasta valores basales a más de 500 kb. No obstante, debe tenerse en cuenta que existen variaciones en la extensión del desequilibrio de ligamiento publicado en distintos trabajos debido a que la estimación del mismo depende de varios factores como la historia y estructura de la población estudiada, el tamaño de la muestra, el tipo de marcador utilizado (microsatélites o SNPs), la densidad y la distribución de los marcadores, el método usado para la reconstrucción de los haplotipos, los umbrales de FAM y EHW utilizados al realizar el control de calidad de los SNPs y el uso de haplotipos de ambos padres o maternos únicamente (Bohmanova y col., 2010). En el trabajo de Nani y col. (2015) se evaluó el DL sobre los haplotipos obtenidos a partir de los animales fundadores de una población consistente en 544 Holando y cruza Holando x Jersey, utilizando

todos los pares posibles de SNPs de cada cromosoma, siendo el número total de SNPs analizados 38.872.

Se seleccionaron, entonces, 15 genes candidatos. Tres de ellos contienen tres de los 12 SNPs asociados previamente. Algunos de los genes hallados, junto a otros que, si bien no se encuentran separados por menos de 100 kb de ninguno de los SNPs asociados al fenotipo, se localizan muy cercanos a los mismos, fueron agrupados por el programa IPA® bajo la calificación “Desórdenes hereditarios. Desórdenes esqueléticos y musculares. Producción de energía” por poseer funciones comunes, relacionadas probablemente con la producción de energía en nuestro caso. Tras consultar en bibliografía la función y el perfil de expresión de los genes candidatos hallados, se decidió secuenciar *PRMT2*, *BTN1A1* y *S100B* por estar directa o indirectamente relacionados con la producción o composición de la leche (Li y col., 2014; Zhong y col., 2012; Smith y col., 2010; Mather y Jack, 1993; Ogg y col., 2004; Suchocki y col., 2010; Gazzolo y col., 2003; Gazzolo y col., 2004; Li y col., 2011; Galvano y col., 2009). Dentro de estos genes se seleccionaron regiones específicas, priorizándose la secuenciación de regiones que contuvieran exones y/o marcadores polimórficos previamente reportados en literatura como asociados al carácter fenotípico estudiado.

La comparación de las secuencias obtenidas para las regiones *PRMT2*, *BTN1A1_A*, *BTN1A1_B* y *S100B* con la secuencia de referencia correspondiente publicada en la base de datos del NCBI, permitió la determinación del número de SNPs reportados previamente que estaban presentes en la población secuenciada. Para las regiones *PRMT2*, *BTN1A1_A* y *BTN1A1_B* se obtuvo el 7%, el 22% y el 24% de los SNPs reportados, respectivamente, mientras que en las secuencias obtenidas de la región *S100B* no se observó la presencia de ningún SNP. Estos bajos porcentajes podrían deberse, en primer lugar, a que algunos de los SNPs reportados no estén presentes en las razas utilizadas en este trabajo de tesis, y en segundo lugar, a que el reducido número de animales secuenciados no haya permitido detectar SNPs caracterizados por presentar bajas FAMs. En el caso de *S100B* hay que tener en cuenta también que se obtuvo una secuencia de buena calidad en promedio de sólo el 76,6% del fragmento amplificado por PCR, y el cómputo de los SNPs reportados en la base de datos del NCBI se hizo considerando el amplicón en su totalidad.

Para cada uno de los SNPs hallados se realizó el cálculo de las frecuencias alélicas que presentó la población secuenciada, discriminando los animales de acuerdo a su valor de cría para producción de leche acumulada a 305 días. Entre los SNPs hallados en la región *PRMT2*, el único que presenta una diferencia de frecuencias alélicas apreciable (de 0,22) entre los animales de bajo y los de alto EBV es el SNP rs42869038. Este SNP fue reportado por Li y col. (2014) como asociado a la producción lechera de bovinos Holando. Sin embargo, mientras que en este trabajo de tesis se encontró una mayor frecuencia del alelo C entre los animales con alto EBV en relación a los de bajo EBV, Li y col. (2014) reportaron una mayor producción lechera en los animales con genotipo CT respecto de aquellos con genotipo CC. A pesar de esto, cabe recordar que el test de hipótesis χ^2 realizado a partir de la tabla de contingencia para este SNP demostró la independencia entre las variantes alélicas y el EBV.

Del gen *BTN1A1*, directamente relacionado con la composición grasa de la leche bovina, se seleccionaron dos regiones para su secuenciación. Al analizar las secuencias obtenidas de

una de ellas, BTN1A1_A, se detectaron 13 SNPs, 6 de los cuales se ubican en exones del gen. Cinco de estos SNPs dan lugar a mutaciones con cambio de sentido, y para 4 de ellos (rs133113758, rs137057391, rs132717963 y rs43706495) se obtuvo una diferencia de alrededor de 0,2 entre las frecuencias alélicas de los grupos de animales de bajo y alto EBV analizados. Las cinco mutaciones no sinónimas a las que dan lugar estos SNPs generan variaciones entre aminoácidos de similares características fisicoquímicas para el caso de los SNPs rs137057391, rs209464806 y rs43706495 (entre aminoácidos con cadena lateral hidrocarbonada, aromáticos y básicos, respectivamente), mientras que el SNP rs133113758 implica la sustitución de un aminoácido no polar alifático (alanina) por uno polar hidroxilado (treonina) y el rs109972486 la sustitución quizá más relevante a nivel estructural, del aminoácido cíclico no polar prolina por la amida del ácido glutámico, el aminoácido polar glutamina.

Por otro lado, de la región BTN1A1_B se obtuvieron 7 SNPs, 4 de los cuales corresponden a exones del gen. Los cuatro dan lugar a mutaciones con cambio de sentido, generando todos ellos variaciones entre aminoácidos de similares características fisicoquímicas: sustituciones entre aminoácidos básicos, hidroxilados, ácidos y con cadena lateral hidrocarbonada los SNPs rs211141026, rs209257029, rs43708439 y rs208558228, respectivamente. De los SNPs ubicados en exones, sólo el rs43708439 presenta una diferencia de 0,2 entre las frecuencias alélicas de los grupos de animales de bajo y alto EBV analizados.

El número de animales utilizado en este análisis exploratorio sólo permitiría encontrar polimorfismos que contribuyan marcadamente a la generación del rasgo estudiado, de forma tal que la frecuencia de aparición de un determinado alelo entre los animales de un fenotipo dado difiera notablemente de aquella que muestren los animales del fenotipo contrario. Aunque para los SNPs que presentaron valores considerables para la diferencia entre las frecuencias alélicas en animales con diferentes fenotipos no pudo demostrarse una asociación entre sus variantes alélicas y el valor de cría, las diferencias halladas sugieren la existencia de tendencias que se debería estudiar con mayor profundidad, es decir, analizar en una población mayor, más aun cuando los SNPs implicados se localizan en exones del gen y dan lugar a un cambio en la estructura primaria de la proteína que pueda alterar su función biológica.

CONSIDERACIONES FINALES

La identificación de genes involucrados en caracteres de herencia compleja en animales domésticos, como es el caso de los bovinos lecheros, no sólo contribuye a una mejor comprensión de estos caracteres sino que, además, implica su potencial beneficio para la ganadería. Tradicionalmente, la base genética de los caracteres de herencia compleja en estas especies ha sido estudiada sin la identificación de los genes implicados. La selección se ha basado en los valores de cría estimados a partir de registros fenotípicos y pedigrís, utilizando un modelo cuantitativo que es exitoso cuando el carácter tiene alta heredabilidad, pero presenta ciertas limitaciones si el carácter sólo puede ser medido en uno de los dos sexos (producción de leche), después de la muerte del animal (calidad de la carne), a una edad avanzada (longevidad), o si la medición del carácter es costosa (resistencia a enfermedades) (Goddard y Hayes, 2009).

El advenimiento de la genética molecular, que comenzó en la década del '70, brindó nuevas oportunidades de mejorar los programas de cría de ganado a través del uso de marcadores del ADN en la identificación de genes o regiones genómicas que controlan caracteres de interés (Dekkers, 2012). Esto se basa en que la identificación de genes y de polimorfismos vinculados con las diferencias observadas a nivel fenotípico permite seleccionar los animales portadores de los alelos deseados a una edad temprana y estimar los valores de cría con mayor precisión.

En este trabajo de tesis se realizó un análisis cuantitativo-molecular que permitió identificar SNPs asociados a la producción de leche en un rodeo comercial de raza Holando y cruza Holando x Jersey de la cuenca lechera central de la Argentina.

Ello implicó llevar a cabo un flujo de actividades en las áreas de biología molecular, estadística, bioinformática, entre otras, que resumidamente produjeron los siguientes resultados:

- se empleó una metodología de selección de genes candidatos y de marcadores moleculares de tipo SNP relacionados con características de producción y composición de la leche bovina que condujo a la obtención de un panel de 62 SNPs en 43 genes candidatos para ser utilizado en la evaluación del genotipo de bovinos lecheros;
- se detectaron 9 polimorfismos nuevos, no incluidos en las bases de datos de SNPs ni reportados en la bibliografía consultada, al secuenciar regiones específicas de genes candidatos en los toros padres de la población analizada;
- se determinó el genotipo de los marcadores moleculares seleccionados y se caracterizó a la población bajo estudio según distintos parámetros genéticos;
- se definieron los límites de 15 bloques de genes candidatos distribuidos en 14 cromosomas (BTAs 1, 4, 5, 6, 11, 14, 15, 19, 20, 21, 22, 23, 24 y 26) y se determinaron los genotipos de los 12.460 SNPs localizados en dichos bloques e incluidos en un microarreglo comercial de mediana densidad;

- se realizó un control de calidad de los genotipos obtenidos, que consistió en el cálculo de la tasa de genotipificación que presentaron las muestras y los SNPs, la determinación de la frecuencia alélica mínima de cada SNP, el análisis de la existencia o no de equilibrio de Hardy-Weinberg en la población estudiada y la detección de errores mendelianos en las familias constituidas por madre-padre-hija o padre-hija. Este control de la genotipificación determinó que se mantuvieran 10.182 SNPs para ser considerados en el análisis posterior;
- se analizó la asignación de paternidades a través de la estimación de la probabilidad IBD existente entre los alelos de los animales tomados de a pares, lo que condujo a la reasignación de padres al 20,8% de las vacas analizadas;
- se analizó la estratificación poblacional presente en la muestra analizada, asignando probabilísticamente los individuos a cada subpoblación según las frecuencias alélicas que presentarían en cada *locus*. Se reveló de esta forma la existencia de cuatro subpoblaciones: una de ellas con un porcentaje de raza Holando de 44,4% y las restantes, con animales emparentados entre sí, con valores superiores al 95%;
- se predijo el Valor de Cría Estimado para producción de leche acumulada a 305 días durante la primera lactancia de 1.628 vacas y se estimó su heredabilidad, siendo su valor de 0,208;
- integrando la información genotípica, fenotípica y genealógica, se realizó un análisis de asociación a través de un modelo aditivo poligénico, estimando la matriz de varianza/covarianza utilizada a partir de la matriz de parentesco genómico y tomando en consideración la estructura de la población analizada, debida tanto a las diferencias raciales como a la presencia de animales emparentados. Del análisis se encontraron 12 SNPs asociados al valor de cría estimado para producción de leche, ubicados en regiones no contiguas del genoma bovino, en los BTAs 1, 20, 23 y 24;
- se seleccionaron 34 vacas con EBVs extremos considerando las distribuciones de EBVs inter e intrafamiliares para secuenciar regiones específicas de los genes *PRMT2*, *BTN1A1* y *S100B*, genes seleccionados por su cercanía en el genoma a los SNPs significativos resultantes del estudio de asociación y por estar directa o indirectamente relacionados con la producción o composición de la leche;
- en las secuencias de las regiones *PRMT2*, *BTN1A1_A* y *BTN1A1_B* se hallaron 3, 7 y 13 SNPs de los reportados en las secuencias de referencia publicadas en la base de datos del NCBI, respectivamente, mientras que en las secuencias obtenidas de la región *S100B* no se observó la presencia de ningún SNP previamente reportado;
- en las regiones estudiadas del gen *BTN1A1* se hallaron 9 SNPs que generan mutaciones con cambio de sentido, dando lugar dos de ellos a variaciones entre aminoácidos de diferentes características químicas;
- en el análisis exploratorio realizado no se encontró una marcada diferencia de frecuencias alélicas entre animales con diferentes fenotipos (bajo y alto EBV para producción de

leche) para ninguno de los SNPs hallados en las regiones secuenciadas, sin embargo algunos SNPs evidenciaron diferencias de alrededor de 0,2 entre dichas frecuencias.

En base a los resultados obtenidos en este trabajo se puede concluir que el proceso de búsqueda bibliográfica y secuenciación que en primer lugar condujo a la selección de genes y marcadores candidatos y posteriormente permitió definir bloques o regiones del genoma a genotipificar, a pesar de ser una tarea que insume gran cantidad de tiempo (lectura de trabajos científicos, consulta de bases de datos de QTLs y SNPs, selección de regiones candidatas a secuenciar, diseño de *primers*, puesta a punto de reacciones de PCR y secuenciación, análisis de secuencias obtenidas), brinda la posibilidad de acotar las regiones del genoma sobre las que se realizará un estudio de asociación, resultando en una búsqueda de marcadores asociados al fenotipo bajo estudio más dirigida/limitada que la que podría hacerse omitiendo esta etapa previa, considerando para el análisis el genoma en su totalidad. No obstante, a partir de la selección de genes y marcadores candidatos, se definieron 15 bloques de genes candidatos distribuidos en 14 cromosomas, es decir, casi la mitad de los cromosomas del genoma bovino fue considerada para el análisis de genotipos. El hecho de que una porción tan significativa del genoma haya sido considerada para el análisis ulterior está de acuerdo con el número de QTLs asociados a la leche bovina (7.579) que han sido reportados hasta el momento (www.animalgenome.org/QTLdb/), en particular, al número de QTLs asociados al rendimiento o producción de leche bovina (1.043). Estos 1.043 QTLs están distribuidos en los 29 autosomas del genoma y en el cromosoma X y contienen gran cantidad de SNPs. Esta influencia de polimorfismos de distintas regiones del genoma sobre el carácter estudiado ratifica la herencia poligénica del mismo. En contraposición con el contenido de grasa de la leche, carácter para el cual la mutación K232A del gen *DGAT1* explica un gran porcentaje de la varianza fenotípica, para el caso de la producción de leche bovina (Grisart y col., 2002), en particular, no se ha hallado hasta el momento un polimorfismo de gran efecto, sino que la varianza explicada por cada uno de los QTLs/polimorfismos asociados a este fenotipo es baja.

A partir de lo expuesto se considera que la estrategia de búsqueda y definición de regiones genómicas candidatas es interesante y probablemente más eficiente cuando el fenotipo estudiado está influenciado/determinado por QTLs o polimorfismos de gran efecto si la información quisiera ser usada directa e inmediatamente en un esquema de mejoramiento animal. Mientras que, debido al descenso de los costos que está teniendo el genotipado con los microarreglos de SNPs de mediana o alta densidad, para caracteres de herencia típicamente cuantitativa, un GWAS convencional ofrecería ciertas ventajas por utilizar información a nivel del genoma completo.

Sin embargo, la asociación encontrada de 12 SNPs con el valor de cría estimado para producción de leche de primera lactancia demuestra la utilidad de esta metodología para la detección de información a nivel del ADN implicada en un carácter fenotípico complejo. Si bien en las regiones exploradas no se encontraron nuevos SNPs, esto tiene al menos dos consideraciones: la primera se relaciona con el "*background*" genético Holando estudiado, que sin duda es el resultado del proceso de "*holsteinization*" comenzado en la década del '80 a

nivel mundial y continuado en la actualidad en el país por medio de la importación de semen, embriones y animales principalmente de Estados Unidos y Canadá. Por lo tanto, no es sorprendente que los SNPs significativamente asociados a producción de leche sean en su mayoría aquellos reportados en otros estudios de la raza Holstein. En segundo lugar, debe considerarse que la secuenciación de regiones de genes candidatos y el número de animales secuenciados en lugar de ser “masivos” fueron muy limitados y, por lo tanto, es probable que existan variaciones en la población estudiada que no hayan sido detectadas en el análisis realizado.

En resumen, este estudio aporta datos originales de interés para la caracterización genotípica de los bovinos de Argentina de raza Holando y cruza Holando x Jersey y constituye el primer estudio llevado a cabo en el país que, en el caso de querer utilizarse la metodología de Selección Genómica en mejoramiento de bovinos para leche, podría contribuir a la formación de la población de referencia (*training population*).

ANEXOS

ANEXO I.

Raza Holando.

La raza Holando, también llamada Holstein, fue introducida en nuestro territorio desde Holanda en el año 1880. En ese entonces la raza tenía doble propósito: carne y leche. Por sus características únicas de color, fortaleza, adaptación a diferentes climas y producción lechera, la Holando empezó a diferenciarse de las demás razas, y pronto comenzó a expandirse por otros países. Hoy, el Holando es aquí como en el resto del mundo, la raza lechera por excelencia.

La hembra Holando tiene una alzada mediana de 1,40 a 1,50 metros, con una gran capacidad en su arco costal y rumen, que le permite digerir el gran volumen de forraje ingerido. Su peso adulto está entre los 600 y 650 Kg. Se caracteriza por poseer manchas bien delimitadas negras en un manto blanco, aunque su pelaje también puede ser blanco y rojo; esta última coloración representa adaptabilidad a climas cálidos. Su vientre, patas y cola son blancos (Fuente: Asociación de Criadores de Holando Argentino, <http://www.acha.org.ar/>).

Posee netas características lecheras como piel fina, huesos chatos, cuello fino, morro ancho y buen sistema mamario. En esta raza los terneros son de gran tamaño ya que pesan entre 42 y 45 Kg al nacimiento.

Se destacan por producir altos volúmenes de leche (en promedio 25 litros por día) con un 3,9% de grasa butirosa. Es la raza lechera que tiene el menor porcentaje de grasa, pero la mayor producción total. A pesar de ser la principal raza lechera del mundo, estos ejemplares muestran más incidencia de problemas reproductivos y sanitarios que otras razas de aptitud láctea.

Raza Jersey.

La raza Jersey es originaria de la pequeña isla de Jersey en el Canal de la Mancha. No hay seguridad en cuanto a cuáles fueron las razas originarias que la conformaron, pero se aceptan como las más probables el ganado negro pequeño de Gran Bretaña y el colorado grande de Francia. En 1784, se comenzó a exportar ejemplares a Inglaterra y las pruebas realizadas allí confirmaron su superioridad en el rendimiento de grasa, único componente valorado en esa época. La demanda de la vaca Jersey fue en aumento y durante el siglo XVIII muchos ejemplares llegaron a Sudáfrica, Austria, Tasmania y especialmente a Nueva Zelanda, en donde la rápida dispersión la hizo constituir el 80% del rodeo lechero de ese país. Las zonas templadas de América del Sur también la incorporaron: Argentina en 1909, Uruguay en 1910 y Brasil algo después.

Es la raza productora de leche de menor tamaño. Su peso a edad adulta oscila entre los 350 y 450 o 500 kg y tiene una alzada de 1,25 metros. Su pelaje es de color variable, desde el bayo claro al casi negro, pasando por el tostado y con menor frecuencia el grisáceo. El pelaje de la cabeza y el cuello es más oscuro encontrándose un anillo claro alrededor del hocico. Su perfil es cóncavo con frente ancha, cara corta y descarnada de pezuñas, borla y mucosidades

oscuras, lo que le confiere una alta adaptabilidad a climas cálidos. Esta raza se distingue de las demás razas lecheras por su temperamento manso y afectivo.

Es una raza precoz. Su velocidad de desarrollo y su pubertad temprana hacen posible que a los 14 meses en promedio, con un peso aproximado de 250 kilogramos, estén aptas para el primer servicio, pariendo por primera vez alrededor de los 24 meses. Debido a un canal de parto amplio y fácilmente dilatable y al bajo peso del ternero Jersey al nacer (entre 25 y 35 Kg), esta raza tiene mínimos problemas durante el parto. Además, tiene alta fecundidad con intervalos cortos entre partos, lo que la lleva a lograr más terneros que otras razas durante el período fértil. La conformación de la ubre y de sus patas la convierten en una vaca con alta vida útil en el tambo, ya que fácilmente supera las 8 a 10 lactancias (Fuente: Asociación Argentina de Criadores de Jersey, <http://www.jerseyargentina.com.ar/>).

La vaca Jersey se adapta rápidamente a los distintos tipos de climas y suelos. Es muy resistente al estrés calórico. La disminución de la producción por calor comienza a una temperatura 5°C mayor en la raza Jersey que en otras razas lecheras.

Experiencias realizadas en Nueva Zelanda, Gran Bretaña, Canadá y Estados Unidos, han demostrado que esta raza es un 30% más eficiente en la conversión de pasto a leche que las demás razas lecheras. Esto da la posibilidad de manejarla sobre pasturas de menor volumen forrajero o aumentando la carga animal por hectárea. Su producción se destaca por el alto contenido en grasa butirosa (alrededor del 5,3%) y proteínas, y su menor volumen en comparación con la raza Holando. Por estos motivos, la Jersey es una raza orientada en forma exclusiva hacia la producción de leche.

Cruza Holando x Jersey.

Haciendo este cruzamiento se logra un animal capaz de producir mayor cantidad de litros de leche que una vaca Jersey y mayor cantidad de sólidos en la leche que una Holando, siendo por lo tanto, un animal de producción con características deseables intermedias a las dos razas que le dan origen. Otras ventajas que se obtienen al hacer este cruzamiento son las siguientes:

- Se mejora la performance reproductiva: aumenta el porcentaje de preñez, se acortan los intervalos parto-concepción y parto-parto y disminuye la cantidad de partos asistidos (Ficha Técnica Nº 19, INTA Lechero).
- Disminuyen los niveles de células somáticas y los problemas de salud en general, incrementándose la vida productiva de los animales en el rodeo.
- Al ser las cruza animales de menor tamaño, se puede incrementar la carga animal por hectárea.
- Son animales que se alimentan a un costo menor.
- Son animales más dóciles.

ANEXO II

1. Estructura de la glándula mamaria.

La ubre de la vaca está diseñada para producir y ofrecer al ternero recién nacido un fácil acceso a la leche. Se encuentra suspendida por fuera de la pared del abdomen posterior y no se encuentra fijada, soportada o protegida por ninguna estructura ósea. Está constituida por cuatro glándulas mamarias o "cuartos". Cada cuarto es una unidad funcional en sí misma que opera en forma independiente y drena la leche por medio de su propio canal. Generalmente, los cuartos posteriores son ligeramente más desarrollados y producen más leche (60%) que los cuartos anteriores (40%).

La ubre es una glándula exocrina, ya que la leche es sintetizada en células especializadas agrupadas en alvéolos y luego excretada fuera del cuerpo por medio de un sistema de conductos.

La unidad funcional de producción de leche es el alvéolo. Éste consiste en una capa de células secretoras que rodean una cavidad central o lumen hacia donde es secretada la leche. Rodeando el alvéolo se hallan capilares sanguíneos y células mioepiteliales (Figura A.II.1). Las funciones del alvéolo son las siguientes:

- Remover los nutrientes de la sangre.
- Transformar estos nutrientes en leche.
- Descargar la leche dentro del lumen.

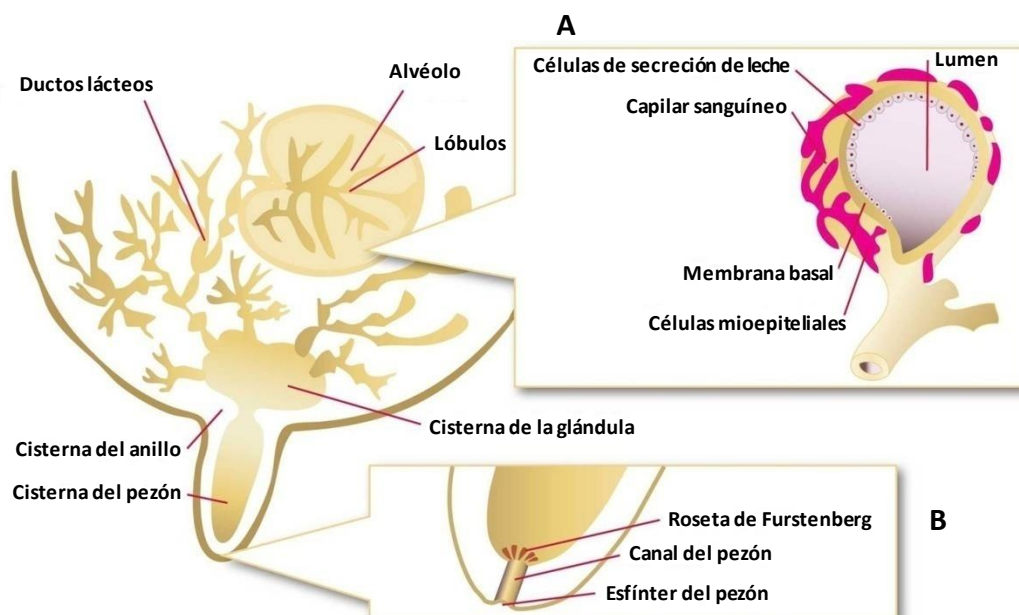


Figura A.II.1. Estructuras secretoras y colectoras de leche.

En el recuadro A se muestra con mayor detalle la estructura del alvéolo, mientras que en B se amplía la porción extrema del pezón.

Adaptado de: www.ucv.ve/fileadmin/user.../Anatomia_Glandula_Mamaria.

La leche abandona el lumen por medio de un tubo colector. Un grupo de 10 a 100 alvéolos que drenan por medio de un conducto común constituyen un lóbulo. A su vez, los lóbulos se encuentran organizados en unidades de mayor tamaño, que descargan la leche dentro de un gran conducto colector que conduce a la cisterna de la glándula (Figura A.II.1).

Los conductos forman canales de drenaje en los que la leche se acumula entre los ordeños. Cuando las células mioepiteliales que recubren el alvéolo y los pequeños conductos se contraen en respuesta a la hormona oxitocina se produce el reflejo de bajada de leche y la leche fluye hacia la cisterna de la glándula. Esta cisterna descansa directamente encima del pezón de la glándula.

El pezón está recubierto por una piel suave y está muy innervado e irrigado. En su extremo externo, se cierra con un anillo de músculo liso o esfínter llamado canal del pezón (Figura A.II.1). El pezón constituye una barrera para las células invasoras. La preservación de su estructura normal es esencial para mantener los mecanismos de defensa contra las bacterias productoras de mastitis.

En la superficie de la ubre existen receptores nerviosos sensibles al contacto y a la temperatura. Durante la preparación de la ubre para el ordeño, estos receptores son estimulados y tiene lugar el reflejo que permite la liberación de leche.

La producción de leche demanda gran cantidad de nutrientes que son llevados a la ubre por la sangre. Además, la sangre transporta hormonas que controlan el desarrollo de la ubre, la síntesis de leche y la regeneración de células secretoras entre lactancias (durante el período seco).

2. Metabolismo durante la gestación y el inicio de la lactancia.

Durante la preñez del animal, su metabolismo se vuelve anabólico de modo de poder cubrir el gran requerimiento nutricional que tendrá durante la lactancia. Las adaptaciones que sufre la vaca lechera para que esto ocurra tienen lugar desde la gestación temprana.

Con el inicio de la lactancia aumentan los requerimientos de agua y nutrientes como glucosa, aminoácidos y ácidos grasos, para ser utilizados en la producción de leche. En este momento el metabolismo general cambia nuevamente, en este caso de anabólico a catabólico, de modo de proveer a la glándula mamaria de los nutrientes necesarios. Las reservas acumuladas durante el período de gestación son movilizadas durante la lactancia. En el tejido adiposo aumenta la lipólisis y disminuye la lipogénesis, se eleva la producción de glucosa por el hígado, aumenta la absorción de minerales en el intestino y la utilización de nutrientes se redirecciona hacia la ubre. Estos mecanismos de adaptación están regulados por el sistema endócrino (Svennersten-Sjaunja y Olssen, 2005).

El uso de glucosa por la célula secretora. A pesar de que la glucosa en la dieta se fermenta totalmente en el rumen a ácido graso volátil (acético, propiónico y butírico), durante el período de lactancia la ubre requiere grandes cantidades de este azúcar. Para cubrir esta necesidad, el hígado transforma el ácido propiónico nuevamente en glucosa que es

transportada por la sangre hacia la ubre, donde es asimilada por las células secretoras. Una vez allí, la glucosa puede ser utilizada como fuente de energía para las células, en el proceso de síntesis de lactosa, o como fuente del glicerol necesario para la síntesis de triglicéridos.

3. Leche de vaca.

La leche es un líquido de color blanco opaco de composición y estructura compleja. La materia grasa se encuentra en emulsión, las proteínas constituyen una suspensión, mientras que los restantes componentes (lactosa, otras sustancias nitrogenadas, minerales, etc.) están disueltos. A continuación se indican las principales características relacionadas con la composición de la leche y se hace una breve reseña del origen de cada componente de la misma.

3.A. Materia grasa.

La materia grasa de la leche se presenta principalmente en forma de glóbulos de sustancias liposolubles y consiste en su mayor parte (98%) en triglicéridos. Los triglicéridos de la leche contienen una alta proporción de ácidos grasos saturados (principalmente ácidos mirístico, C14:0; palmítico, C16:0; y esteárico, C18:0), un alto contenido de ácidos grasos monoinsaturados (ácido oleico, C18:1 cis-9) y un bajo contenido de ácidos grasos poliinsaturados (Jensen, 2002; O'Donnell-Megaró y col., 2011). Una muy pequeña fracción de los lípidos de la leche está constituida por partículas dispersas conteniendo una mezcla de diglicéridos, triglicéridos y lípidos complejos.

Se han identificado más de 150 ácidos grasos en la leche, muchos de ellos esenciales, como los ácidos linoleico y linolénico. Los ácidos grasos de cadena corta (de 4 a 12 átomos de carbono) y algunos de cadena media (de 14 a 16 átomos de carbono) se sintetizan en la glándula mamaria a partir de los ácidos acético e hidroxibutírico presentes en la sangre y provenientes de la fermentación de los alimentos en el rumen. Esta síntesis representa el 40% en peso de los ácidos grasos secretados en la leche. Entre el 17 y el 45% de la grasa en la leche se forma del acetato, mientras que entre el 8% y el 25% proviene del butirato. Por otro lado, el glicerol necesario para la síntesis de triglicéridos deriva de la glucosa.

Los ácidos grasos de cadena larga (de 18 a 22 átomos de carbono) y algunos de cadena media son captados directamente de la sangre y, en parte, desaturados en la ubre. En este sentido, el tipo de grasa que ingiere la vaca a través de su dieta posee una influencia muy importante en la composición de la grasa de la leche que producirá.

En la Tabla A.II.1 (Taverna, Proyecto Nacional de Lechería, INTA) se muestra el perfil de ácidos grasos de la leche.

Desde los puntos de vista biológico y económico, la grasa de la leche tiene un gran aporte al valor nutricional de la leche ya que es la fracción que más contribuye a la densidad energética de la leche entera y tiene un rol esencial en muchas de las propiedades físicas, calidad de manufactura y características organolépticas de los productos derivados de la leche de vaca (Harvatine y col., 2009).

Ácidos grasos	Concentración (g/100g de ácidos grasos totales)
Saturados de cadena corta (C_4 - C_{12})	9,4
Saturados de cadena media y larga (C_{14} - C_{18})	51,8
Insaturados	28,1
Poliinsaturados	4,2

Tabla A.II.1. Perfil de ácidos grasos de la leche de vaca producida en la cuenca lechera central de la Argentina.

Se considera ácidos grasos insaturados a aquellos con un doble enlace en su molécula y poliinsaturados a los que presentan más de un doble enlace.

3.B. Fracción nitrogenada.

La fracción nitrogenada de la leche está compuesta en un 95% por proteínas y en un 5% por nitrógeno no proteico (Tabla A.II.2).

En la leche producida por vacas sanas el 80% de las proteínas está representado por las caseínas. Este valor es similar en las distintas razas lecheras y permanece constante durante todo el período de lactancia, a excepción de los primeros días debido al alto contenido de inmunoglobulinas del calostro. El 20% restante está constituido por las proteínas solubles o del suero, siendo las principales la β -lactoglobulina y la α -lactoalbúmina.

La síntesis de las proteínas de la leche tiene lugar principalmente en la ubre, y se realiza a partir de los aminoácidos provenientes de la sangre. Una pequeña fracción (del 5 al 10%) es tomada directamente de la sangre y está constituida por seroalbúmina e inmunoglobulinas. La permeabilidad de las células secretoras para las inmunoglobulinas es alta durante la síntesis del calostro, pero decrece rápidamente al avanzar en la lactancia.

El nitrógeno no proteico consiste principalmente en urea, aunque también se observan en esta fracción compuestos residuales de los procesos de síntesis y aminoácidos libres.

3.C. Lactosa.

La lactosa es un disacárido compuesto por glucosa y galactosa en unión β -1,4 que se encuentra exclusivamente en la leche de los mamíferos. Se sintetiza en la ubre a partir de la glucosa de la sangre en una reacción catalizada por la enzima lactosa sintasa.

La secreción de lactosa dentro de la cavidad del alvéolo incrementa la concentración de sustancias disueltas, es decir, aumenta la presión osmótica, respecto de la superficie de las células secretoras que mira hacia la membrana basal, en contacto con los capilares sanguíneos. Como resultado, la concentración de sustancias disueltas a cada lado de las células secretoras tiende a balancearse atrayendo agua desde la sangre y mezclándola con otros componentes que se encuentran en la cavidad de los alvéolos. Para la leche normal, el equilibrio se alcanza cuando existe 4,5 a 5% de lactosa en la leche. Por lo tanto, la lactosa regula la cantidad de agua que entra al lumen del alvéolo y, en consecuencia, el volumen de leche producido.

	Promedio (g/Kg)	Valor relativo (%)
Fracción nitrogenada	32,00	100
Proteína verdadera		
A. Caseínas	25,00	78,2
α_{s1} -caseína	9,00	36
α_{s2} -caseína	2,50	10
β -caseína	8,50	34
κ -caseína	3,20	13
γ -caseínas	1,75	7
B. Proteínas del lactosuero	5,40	16,8
Albúminas		
β -lactoglobulina	2,70	50
α -lactoalbúmina	1,20	22
Seroalbúmina	0,25	5
Inmunoglobulinas	0,65	12
Peptosa-peptona	0,60	10
Nitrógeno no proteico	1,60	5

Tabla A.II.2. Fracción nitrogenada de la leche.

Adaptación de Alais, 1985.

3.D. Minerales.

Aunque la fracción mineral constituye alrededor del 0,7% de la materia seca de la leche, posee una gran importancia nutricional, principalmente por su contenido de calcio y fósforo. Los minerales pasan de la sangre a la leche mediante sistemas de transporte activos.

3.E. Otros componentes.

La leche contiene gran cantidad de otros componentes en muy pequeñas concentraciones, muchos de los cuales participan en los procesos de transformación y/o degradación de la leche (enzimas) o tienen valor nutricional (vitaminas).

Las vitaminas hidrosolubles de los grupos B y C provienen de la biosíntesis de las bacterias del rumen y están presentes en la fase acuosa. Por el contrario, las liposolubles pertenecientes a los grupos A, E y D, están asociadas a la materia grasa y presentan una concentración variable según la alimentación del animal.

4. Proteínas relacionadas con la composición y calidad de la leche.

Entre las proteínas relacionadas directa o indirectamente con la composición y calidad de la leche bovina pueden encontrarse proteínas con función estructural, nutritiva, enzimas pertenecientes a rutas metabólicas, proteínas involucradas en vías de transducción de señales, etc. A continuación se enumerarán y describirán las principales características de diferentes proteínas que componen la leche bovina o que participan en los procesos de su producción y secreción.

4.A. Proteínas que componen la leche bovina.

Las proteínas que componen la leche bovina pueden dividirse en dos grandes grupos: la fracción caseínica y las proteínas del suero de leche. La inducción de los genes que codifican estas proteínas está bajo el control de un complejo proceso multihormonal que involucra esencialmente a la prolactina, glucocorticoides, insulina, hormona de crecimiento y otros factores lactogénicos. A continuación se hará una breve reseña de las proteínas y variantes proteicas que componen cada uno de estos grupos.

4.A.1. Caseínas.

Las caseínas de la leche del género *Bos* se definieron originalmente como aquellas fosfoproteínas capaces de precipitar de la leche cruda descremada por acidificación a pH 4,6 a 20°C (Farrell y col., 2004). Según su estructura primaria pueden ser clasificadas en cuatro familias: α_{s1} -, α_{s2} -, β - y κ -caseínas. Todas ellas son codificadas por genes ubicados en el cromosoma 6 del genoma bovino. La caracterización e identificación de los miembros de cada familia puede realizarse, entre otras técnicas, mediante electroforesis en geles de agarosa o poliacrilamida.

Alfa S1-caseínas (α_{s1} -CN).

La familia de las α_{s1} -caseínas representa alrededor del 40% de la fracción caseínica de la leche bovina. Está compuesta por diferentes variantes denominadas A, B, C, D, E, F, G y H. La proteína de referencia en esta familia es la α_{s1} -caseína variante B, que consta de 199 aminoácidos. El resto de las variantes se diferencian de la variante B por deleciones o sustituciones que tienen lugar en su secuencia (Tabla A.II.3). La variante A ha sido descrita en las razas Holstein Friesian, Red Holstein y German Red; la variante B es la predominante en *Bos taurus*; la variante C en *Bos indicus* y *Bos grunniens*, la variante D se ha encontrado en varias razas de Francia e Italia, así como también en animales de raza Jersey de los Países Bajos; y la variante E está presente en *Bos grunniens*. Las variantes más recientemente identificadas son las variantes F, G y H. La variante F fue descrita en ganado German Black and White y la G fue descubierta en vacas Italian Brown (Farrell y col., 2004).

Alfa S2-caseínas (α_{s2} -CN).

La familia de las α_{s2} -caseínas representa alrededor del 10% de la fracción caseínica de la leche bovina y es la más hidrofílica de todas las caseínas. Esta familia está compuesta por cuatro variantes denominadas A, B, C y D (Tabla A.II.3). La variante de referencia es la A, consistente en una cadena polipeptídica de 207 aminoácidos y un puente disulfuro interno.

Esta variante es la más frecuentemente observada en razas occidentales. La variante B se presenta, aunque con baja frecuencia, en el ganado cebú de Sudáfrica; la variante C en Nepal y Mongolia; y la variante D se ha encontrado en las razas Vosgienne, Montbeliarde y en tres razas españolas (Farrell y col., 2004). Esta familia de proteínas ha mostrado poseer actividad antibacteriana (Zucht y col., 1995).

Beta-caseínas (β -CN).

La familia de β -caseínas es la más hidrofóbica y abundante de todas las caseínas ya que constituye alrededor del 45% de la fracción caseínica de la leche (Farrell y col., 2004). La proteína de referencia en esta familia es la β -caseína variante A², consistente en una cadena polipeptídica con 209 aminoácidos. Las variantes restantes (A¹, A³, B, C, D, E, F, G, H¹, H² e I) difieren de la variante A² en algunos aminoácidos hallados en posiciones específicas (Tabla A.II.3).

Kappa-caseínas (κ -CN).

Esta familia representa la menor porción de la fracción caseínica de la leche bovina. Se aísla de la leche como una mezcla de polímeros unidos por puentes disulfuro intermoleculares. Está compuesta por once variantes: A, B, C, E, F¹, F², G¹, G², H, I y J, de las cuales la variante A, con 169 aminoácidos, es la de referencia. Las sustituciones aminoacídicas que dan lugar al resto de las variantes se detallan en la Tabla A.II.3. La variante A es la que predomina en la mayoría de las razas lecheras, a excepción del ganado Jersey (Farrell y col., 2004). Las distintas variantes proteicas de esta familia sufren modificaciones post-traduccionales como glicosilación y fosforilación (Vreeman y col., 1986).

4.A.2. Proteínas del suero.

La denominación “proteínas del suero de leche” hace referencia al conjunto de proteínas de la leche que permanecen solubles luego de la precipitación de las caseínas a pH 4,6 y 20°C (Farrell y col., 2004). Entre ellas se encuentran la β -lactoglobulina, la α -lactoalbúmina, la seroalbúmina, las inmunoglobulinas, la lactoferrina y la fracción proteosa-peptona. Asimismo, permanecen solubles luego de la precipitación de caseínas, y por lo tanto forman parte de este grupo, pequeños fragmentos que surgen de la proteólisis de caseínas, así como también proteínas que constituyen la membrana de los glóbulos de grasa.

Beta-lactoglobulina (β -LGB).

La β -lactoglobulina es la principal proteína del suero. Se han descrito once variantes de esta proteína (A, B, C, D, E, F, G, H, I, J y W), siendo las variantes A y B las más abundantes en la mayoría de las razas (Farrell y col., 2004). A pesar de estar presente en la leche de varias especies, principalmente de rumiantes, la β -lactoglobulina se halla ausente en la leche de roedores, conejos e, incluso, en la humana (Martin y col., 2002). La proteína de referencia en esta familia es la variante B, que posee 162 aminoácidos. Las demás variantes difieren de esta última por sustituciones aminoacídicas en posiciones específicas (Tabla A.II.4). El gen que codifica la β -lactoglobulina bovina se encuentra ubicado en el cromosoma 11.

Proteína	Variante	Posición y aminoácido en la proteína												
		14-26		53	51-58			59	66		192			
α_{s1} -CN	A	Delección												Glu
	B			Ala				Gln	SerP					Glu
	C													Gly
	D			ThrP										Glu
	E							Lys						Gly
	F								Leu					Glu
	G													Glu
	H				Delección									Glu
α_{s2} -CN	A	33		47			51-59			130				
	B	Glu		Ala									Thr	
	C	Gly		Thr									Ile	
	D						Delección							
β -CN	A ¹	18	25	35	36	37	67	72	88	93	106	122	137/138	152
	A ²	SerP	Arg	SerP	Glu	Glu	Pro	Gln	Leu	Met	His	Ser	Leu/Pro	Pro
	A ³										Gln			
	B						His					Arg		

Tabla A.II.3. Diferencias aminoacídicas entre variantes caseínicas.

Los aminoácidos en la variante de referencia de cada familia se escriben en negrita (Fuente: Farrell y col., 2004).

Proteína	Variante	Posición y aminoácido en la proteína													
		18	25	35	36	37	67	72	88	93	106	122	137/138	152	
β-CN	C			Ser		Lys	His								
	D	Lys													
	E				Lys										
	F						His							Leu	
	G						His					Leu			
	H ¹		Cys						Ile						
	H ²							Glu		Leu					
	I									Leu					
	κ-CN	A	Arg		Arg		Ser		Thr		Thr		Asp		Ser
B										Ile		Ala			
C				His											
E														Gly	
F ¹												Val			
F ²		His								Ile		Ala			
G ¹				Cys						Ile		Ala			
G ²												Ala			
H										Ile					
I						Ala									
J										Ile		Ala		Arg	

Tabla A.II.3. Diferencias aminoacídicas entre variantes caseínicas (continuación).

Alfa-lactoalbúmina (α -LA).

La α -lactoalbúmina es una proteína globular de 123 aminoácidos con tres variantes (A, B y C) que se diferencian entre sí de acuerdo a los aminoácidos que contienen en posiciones específicas (Tabla A.II.4). La proteína de referencia en esta familia es la variante B, siendo la variante predominante en la leche del ganado *Bos taurus* y estando también presente en el *Bos indicus*. Por otro lado, la variante A ha sido hallada en *Bos indicus* y la variante C en *Bos javanicus* (Farrell y col., 2004). El gen codificante de esta proteína se ubica en el cromosoma 5 bovino.

La α -lactoalbúmina de la leche es una importante fuente proteica ya que tiene un alto contenido de aminoácidos esenciales. Además, tiene un rol clave como regulador de la producción de lactosa en la glándula mamaria y en la secreción de leche. Esta proteína interactúa con la enzima ubicuamente expresada β -1,4-galactosiltransferasa en el aparato de Golgi de las células del epitelio mamario para formar el complejo lactosa sintasa. En este complejo la α -lactoalbúmina posibilita la captura de la molécula de glucosa e inhibe la unión de N-acetilglucosamina a la β -1,4-galactosiltransferasa de modo de modificar la especificidad de sustrato de la enzima permitiendo la formación de lactosa a partir de glucosa y UDP-galactosa (Figura A.II.2). Ratones *knocked out* para el gen que codifica la α -lactoalbúmina producen una leche carente de lactosa y extremadamente viscosa que no puede ser removida de la glándula mamaria por las crías (Stinnakre y col., 1994; Stacey y col., 1995).

La leche bovina contiene α -lactoalbúmina en una concentración de 1,2 a 1,5 g/L. Esta concentración disminuye hacia el final de la lactancia, lo que se correlaciona con el descenso de la concentración de la lactosa en la leche.

Seroalbúmina (SA).

Esta proteína representa el 1,5% de las proteínas de la leche y aproximadamente el 8% de las proteínas del suero de leche. Consta de 583 residuos aminoacídicos.

La seroalbúmina se halla en todos los tejidos y secreciones del cuerpo y tiene un rol importante en el transporte, metabolismo y distribución de moléculas que actúan como ligandos. También contribuye a la presión osmótica de la sangre y proporciona protección contra radicales libres.

Proteína	Variante	Posición y aminoácido en la proteína											
		45	50	56	59	64	70	78	108	118	126	129	158
β-LGB	A					Asp				Val			
	B	Glu	Pro	Ile	Gln	Gly	Lys	Ile	Glu	Ala	Pro	Asp	Glu
	C				His								
	D	Gln											
	E												Gly
	F		Ser									Tyr	Gly
	G							Met					Gly
	H					Asp	Asn			Val			
	I								Gly				
	J										Leu		
	W			Leu									
α-LA					10					?			
	A				Gln								
	B				Arg					Asp			
	C									Asn			

Tabla A.II.4. Diferencias aminoacídicas entre variantes de β-lactoglobulina y α-lactoalbúmina.

Los aminoácidos en la variante de referencia de cada familia se escriben en negrita (Fuente: Farrell y col., 2004).

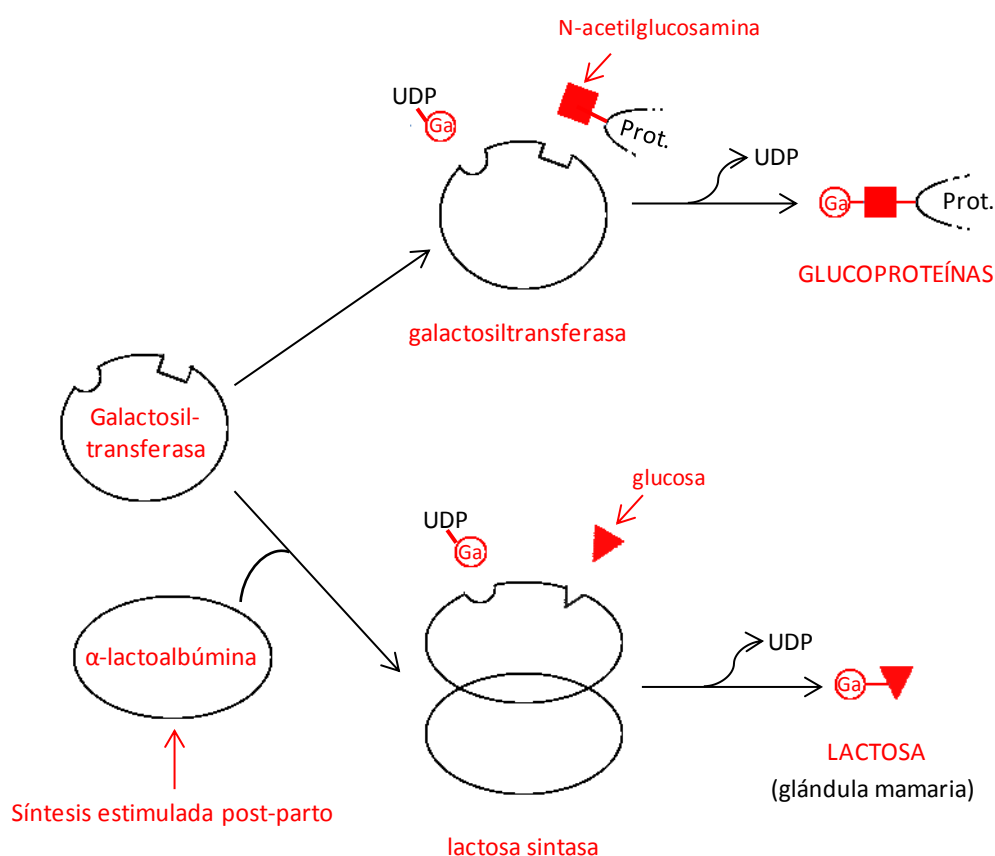


Figura A.II.2. Lactosa sintasa.

En la glándula mamaria, la α -lactoalbúmina, cuya síntesis está estimulada por los niveles hormonales post-parto, modifica la especificidad de la galactosiltransferasa que actúa como lactosa sintasa.

Inmunoglobulinas (Igs).

La fracción de inmunoglobulinas representa aproximadamente el 1% del total de proteínas de la leche o alrededor del 6% de las proteínas del suero.

La estructura básica de las inmunoglobulinas del ganado vacuno es similar a la de las de origen humano, siendo polímeros compuestos por cuatro cadenas polipeptídicas (dos pesadas y dos livianas) unidas por enlaces disulfuro inter e intramoleculares. Tanto la cadena pesada como las livianas λ y κ contienen una región constante (C) que determina el isotipo de Ig (Ig A, Ig D, Ig E, Ig G o Ig M) y un dominio variable (V) que determina la especificidad de unión al antígeno. De los cinco isotipos de inmunoglobulinas presentes en el ganado vacuno, tres han sido identificados en la leche bovina: Ig G, Ig A e Ig M. A su vez, la Ig G presenta tres subclases, denominadas Ig G1, Ig G2 e Ig G3. Los cambios en las concentraciones y proporciones relativas de las distintas Igs en el calostro en relación a la leche durante la lactancia se muestran en la Tabla A.II.5. La Ig G, y en particular la Ig G1, es la principal Ig encontrada tanto en el calostro como en la leche bovina.

Ig	Concentración (mg/ml)		
	Calostro	Leche madura	Suero
Ig G1	46,4	0,58	11,2
Ig G2	2,87	0,05	9,2
Ig A	5,36	0,1	0,37
Ig M	6,77	0,09	3,1

Tabla A.II.5. Concentraciones de Inmunoglobulinas en la leche bovina.

Lactoferrina o lactotransferrina (LTF).

La lactoferrina se encuentra en la mayoría de los fluidos biológicos, incluyendo los gránulos de los neutrófilos y secreciones como lágrimas, saliva, leche, bilis, etc. En la leche humana alcanza altas concentraciones: 7-8 g/L en el calostro y 1-2 g/L en la leche madura (Lönnerdal e Iyer, 1995), mientras que en la leche bovina su concentración varía entre 20 y 200 µg/L y aumenta notablemente en respuesta a estados inflamatorios o infecciosos. Su secreción tiene lugar durante el desarrollo funcional normal del epitelio mamario y durante la lactancia en respuesta a hormonas como la prolactina.

Este polipéptido de 689 aminoácidos adopta un plegamiento particular formando dos lóbulos globulares y presenta diferentes grados de glicosilación en cinco sitios de la proteína. La composición de los glicanos es compleja y varía en las distintas etapas de la lactancia.

La lactoferrina tiene un rol importante en el sistema inmune innato en la defensa contra la infección y la inflamación. Su actividad antimicrobiana radica en su capacidad de secuestrar hierro removiéndolo del medio e impidiendo de esta forma su uso como nutriente por parte de bacterias.

4.B. Proteínas que participan en los procesos de producción y secreción de la leche bovina.

4.B.1. Endocrinología de la producción de leche.

El período de lactancia puede dividirse según Svennersten-Sjaunja y Olsson (2005) en cuatro etapas consecutivas: desarrollo de la glándula mamaria, lactogénesis, mantenimiento de la producción de leche (galactopoyesis) e involución de la glándula. Cada una de ellas tiene lugar bajo un estricto control hormonal. A continuación se mencionarán las hormonas implicadas en cada una de las etapas y se realizará una breve descripción de la función que ejercen.

Desarrollo de la glándula mamaria.

Durante el período de gestación, la proliferación del epitelio mamario es dependiente del estrógeno y la progesterona. El estrógeno es esencial para la morfogénesis ductal, mientras que la progesterona juega un papel crítico en el desarrollo de lóbulos y alvéolos. El estrógeno, además, estimula la secreción del factor de crecimiento insulínico tipo 1 (IGF-1, *Insulin-like*

Growth Factor) a partir de las células estromales de la ubre, provocando el crecimiento de células epiteliales.

Lactogénesis y galactopoyesis.

Durante estas etapas, la producción de leche es controlada por las hormonas lactogénicas prolactina y somatotropina (GH). Ambas hormonas son esenciales para que ocurra la transición de una glándula mamaria proliferativa a una productora de leche. La prolactina actúa en forma directa sobre la glándula, así como también a través de factores epiteliales mamarios activando numerosos factores de transcripción, entre ellos, el STAT5.

En estas etapas también están implicados factores de crecimiento locales que modulan la supervivencia y la apoptosis de la glándula. Muchos de ellos son citoquinas, como los factores de crecimiento epidérmico (EGFs) y los miembros de la familia de factores de crecimiento transformante (TGF- α y TGF- β).

Involución de la glándula.

El término “involución” hace referencia a la regresión gradual que sufre la glándula mamaria al finalizar una lactancia normal. La serie de eventos que tienen lugar en esta etapa, que se halla bajo control endócrino, es importante debido al impacto que tienen sobre la próxima lactancia. La muerte de células epiteliales por apoptosis que ocurre durante esta etapa se asocia a una disminución de los niveles de prolactina, GH e IGF-1. Se ha propuesto que la hormona GH estimula la síntesis de IGF-1 y que la prolactina optimiza su acción al reprimir la función de la proteína de unión de IGF IGFBP5 (miembro 5 de la familia de IGF *binding proteins*). IGFBP5 es un inhibidor de la acción de IGF-1 ya que, al unirse a esta molécula, impide que se una a su receptor (Svennersten-Sjaunja y Olsson, 2005). Por tal motivo, se considera que la disminución en los niveles de prolactina constituye la principal señal para el control de la muerte celular durante el proceso de involución de la glándula mamaria.

4.B.2. Endocrinología de la secreción de leche.

En las vacas lecheras, alrededor del 80% de la leche almacenada en una ubre llena se encuentra en el compartimento alveolar. El reflejo de “bajada de leche” se activa al liberarse a la circulación la hormona hipotalámica oxitocina. En respuesta a esta hormona, las células mioepiteliales que rodean los alvéolos se contraen y la leche drena hacia la cisterna de la glándula.

Luego de la liberación de oxitocina tiene lugar la liberación de otras dos hormonas a la circulación: la prolactina y el cortisol. La prolactina estimula el metabolismo en las células epiteliales de la ubre y la transcripción de genes codificantes de proteínas de la leche como la α -lactoalbúmina, la cual es crítica para la síntesis de lactosa y, por ende, para la producción de leche. El cortisol, por su parte, ejerce un efecto general sobre el metabolismo del animal durante el período de lactancia, pero también se requiere para mantener la actividad secretora de las células epiteliales (Svennersten-Sjaunja y Olsson, 2005).

ANEXO III.

Metodologías de genotipificación de SNPs.

Existen varias técnicas para genotipificar SNPs. La mayoría de ellas consiste en la generación de productos de reacción específicos de alelo, seguida de la separación, detección e identificación de los mismos. A continuación se presenta un breve resumen de las técnicas utilizadas para genotipificar SNPs.

1. Técnicas de hibridización directa.

Las técnicas de hibridización se basan en la diferente estabilidad que presentan los híbridos entre oligonucleótidos cortos y sus secuencias diana según presenten una complementariedad de bases perfecta o no. La hibridización con oligonucleótidos específicos de alelo (ASO, *allele-specific oligonucleotide*) puede realizarse en los formatos denominados “*dot blot*” o “*reverse dot blot*” (Sylvänen, 2001). El formato *dot blot* consiste en disponer las muestras de ADN a analizar sobre una superficie adecuada y sobre éstas realizar la hibridización con ASO. Este formato suele utilizarse en humanos para evaluar la presencia de alelos causantes de enfermedades en muestras de varios pacientes simultáneamente. Por el contrario, el formato *reverse dot blot* parte de ASOs inmovilizados en una superficie sobre la cual se agregan las muestras de ADN para su hibridización. Este último formato es más adecuado que el anterior para analizar gran cantidad de SNPs en paralelo, y es el precursor de los métodos utilizados por los microarreglos (*chips*) de alta densidad.

La utilización de ASOs como sondas de hibridización o *primers* de PCR es el principio que utilizan los métodos de genotipificación de SNPs denominados “homogéneos”, puesto que no implican etapas de separación de reactivos/fases (Landegren y col., 1998).

Existen dos metodologías basadas en la hibridización diferencial que permiten discriminar distintos alelos durante una reacción de PCR, sin separaciones ni lavados previos: las técnicas que utilizan sondas TaqMan o las que utilizan *Molecular Beacons*. Ambas metodologías se basan en el principio de transferencia de energía con emisión de fluorescencia. La fluorescencia se detecta como resultado del cambio en la distancia física a la que se encuentran un fluoróforo y una molécula *quencher* que tiene lugar al producirse la hibridización entre la sonda ASO y su secuencia diana perfectamente complementaria (Sylvänen, 2001).

2. Técnicas que involucran la generación y separación de productos específicos de alelo.

Digestión con enzimas de restricción. Si el SNP a estudiar altera el sitio de reconocimiento de una endonucleasa de restricción, puede usarse la técnica PCR-RFLP para su genotipificación. La discriminación alélica se realiza en este caso evaluando la diferencia en la movilidad electroforética de los productos de PCR digeridos (Landegren y col., 1998).

Conformación del ADN de hebra simple. El polimorfismo de conformación de hebra simple (SSCP) se basa en la conformación específica de plegamiento que adoptan las hebras simples de ADN en condiciones desnaturizantes. Esta técnica consiste en la amplificación por PCR de

una secuencia de ADN que contiene al SNP en estudio, con su consiguiente desnaturalización y corrida electroforética. Una diferencia de una única base en fragmentos de ADN de hasta 300 pb produce un cambio en la conformación del ADN que puede ser detectado al realizar una electroforesis en gel de poliacrilamida en condiciones desnaturalizantes (Vignal y col., 2002).

PCR alelo-específica. La amplificación selectiva de ambos alelos de un SNP puede realizarse mediante una reacción de PCR que utilice dos *primers* alelo-específicos que contengan en la posición terminal 3' o cercano a ella un nucleótido complementario a cada una de las dos bases características del SNP, junto a un *primer* reverso común. Cada primer específico posee una "cola" de tamaño diferente, que permite la diferenciación de los productos mediante electroforesis. La técnica se basa en que únicamente los oligonucleótidos perfectamente complementarios a su secuencia diana promoverán la extensión con dNTPs por la ADN polimerasa.

Reacción de ligación de oligonucleótidos (OLA, *Oligonucleotide Ligation Assay*). Esta técnica de genotipificación de SNPs se basa en la especificidad de la ADN ligasa por secuencias perfectamente complementarias en el sitio de ligación al unir covalentemente dos oligonucleótidos adyacentes hibridizados a una secuencia de ADN determinada (Syvänen, 2001). Los oligonucleótidos que utiliza se diseñan de modo que hacia un lado del SNP se encuentre un ASO marcado de forma diferente para cada alelo del SNP, y del otro lado un oligonucleótido común.

Pirosecuenciación. Esta tecnología de secuenciación consiste en la incorporación de nucleótidos complementarios a una secuencia de ADN molde en una reacción catalizada por la ADN polimerasa y mediada por *primers* específicos. Dicha reacción puede ser monitoreada a través de la medida de la liberación de pirofosfato (Vignal y col., 2002). La comparación de la secuencia obtenida con una secuencia de referencia permite detectar la presencia de SNPs.

El desarrollo reciente de la tecnología de secuenciación masiva de nueva generación (NGS, *Next-generation sequencing technologies*) brinda la posibilidad de obtener la secuencia de todo el genoma y su utilización en evaluaciones genéticas de rutina y en programas de selección genómica. Pérez-Enciso y col. (2015) evaluaron por simulación las ventajas potenciales del uso de la secuenciación parcial o completa respecto de la genotipificación de SNPs mediante *microarrays*. Su trabajo reveló que el incremento en la respuesta a la selección realizada utilizando información de la secuencia del genoma completo frente a la realizada con datos resultantes de la genotipificación de alta densidad de SNPs no es significativo, a menos que se cuente con información biológica precisa y correcta relacionada al carácter bajo estudio que pueda ser incluida en la predicción del mérito genético de los animales.

ANEXO IV.

Protocolos utilizados.

1. Técnica de fenol-cloroformo para la extracción de ADN a partir de muestras de semen.

En primer lugar, se centrifugaron 250 μ l de semen a 6.000 rpm durante 4 minutos, descartándose el sobrenadante obtenido. El *pellet* se resuspendió en 1 ml de buffer STE (0,1M NaCl, 0,05M Tris, 0,001M EDTA) y se centrifugó a 6.000 rpm durante 4 minutos, descartándose nuevamente el sobrenadante. Posteriormente se agregaron 500 μ l de *buffer* de digestión (0,13M NaCl, 0,01M Tris, 0,01M EDTA, 2% SDS) al *pellet* obtenido y, sobre éste, 22 μ l de DTT 0,1M y 27 μ l de proteinasa K (20 mg/ml) y esto se incubó a 55°C durante 3 horas. Al finalizar la incubación se agregaron 500 μ l de fenol (a pH 7,8) para extraer el ADN. El contenido del tubo se mezcló por inversión del mismo y se centrifugó a 10.000 rpm durante 2 minutos. La fase acuosa superior fue transferida a un tubo que contenía 300 μ l de fenol y 300 μ l de cloroformo. El contenido se mezcló por inversión del tubo y se centrifugó nuevamente a 10.000 rpm durante 2 minutos. Luego la fase acuosa superior fue transferida a otro tubo, que contenía 700 μ l de cloroformo. El contenido se mezcló por inversión del tubo y se volvió a centrifugar a 10.000 rpm durante 2 minutos. La fase acuosa obtenida a partir de esta última centrifugación fue transferida a un tubo limpio y sobre ésta se agregaron 2 volúmenes de etanol frío al 95% para precipitar el ADN. Luego se centrifugó a 14.000 rpm durante 5 minutos, se descartó el sobrenadante y se secó el *pellet* obtenido. Finalmente, el ADN seco fue resuspendido en 80-150 μ l (dependiendo del tamaño del *pellet*) de *buffer* TE (0,01M Tris, 0,001M EDTA), dejándolo aproximadamente una hora a temperatura ambiente para su completa resuspensión.

2. Técnica de secuenciación de regiones seleccionadas de genes candidatos.

En primer lugar se amplificaron por reacción en cadena de la polimerasa (PCR) en la población de toros en estudio regiones de algunos genes candidatos seleccionadas de modo que incluyeran los SNPs descritos en literatura. Los volúmenes y concentraciones de cada reactivo de PCR fueron los siguientes:

- <i>Buffer</i> 10X	1,50 μ l
- dNTPs 2mM	1,50 μ l
- <i>Primer</i> F 10 μ M	0,30 μ l
- <i>Primer</i> R 10 μ M	0,30 μ l
- Enzima Paq	0,15 μ l
- ADN	entre 6 y 30 ng, según el gen a amplificar
- Agua bidestilada	hasta completar 15 μ l

En la reacción se utilizó la enzima Paq5000® DNA Polymerase 5U/ μ l (Agilent Technologies, Stratagene), junto a su *buffer* correspondiente. Los dNTPs usados fueron diluciones hechas a partir del dNTP set 100mM Solutions (dATP, dCTP, dGTP y dTTP) de Illustra, GE Healthcare. Las condiciones de PCR fueron las siguientes: desnaturalización del ADN, 3 min a 94°C; seguida de 35 ciclos de 20 seg a 94°C (desnaturalización), 20 seg a 60°C (*annealing*, hibridización de *primers*) y 1 min a 72°C (polimerización, extensión); y una etapa de extensión final de 5 min a

72°C. Las reacciones se realizaron en cicladores Eppendorf Mastercycler. El producto obtenido se visualizó en gel de agarosa al 1% por comparación con un estándar de tamaño conocido.

Una alícuota del amplicón se secuenció en ambos sentidos utilizando los mismos *primers* que en la reacción de PCR, según el método de dideoxinucleótidos de Sanger. La reacción de secuenciación se realizó en un ciclador Perkin Elmer GeneAmp PCR System 9600 de acuerdo a las siguientes condiciones: 1 min a 96°C, seguido de 30 ciclos consistentes en 10 seg a 96°C, 5 seg a 50°C y 4 min a 60°C. Los volúmenes de los reactivos utilizados en la reacción figuran a continuación:

- <i>Mix</i>	0,50 µl
- <i>Buffer 5X</i>	1,75 µl
- <i>Primer</i>	1,00 µl
- ADN (dilución)	1,00 µl
- Agua bidestilada	5,75 µl (para llevar a un volumen final de 10 µl)

La mezcla de reacción (*mix*) fue la BigDye® Terminator Cycle Sequencing Kit de Applied Biosystems (California, EE.UU.) y el *buffer 5X* correspondiente el BigDye® Terminator v3.1 del mismo fabricante. Como molde se utilizó una dilución 1/10 del amplicón obtenido en la primera reacción de PCR.

En la Tabla A.IV.1 se muestra la secuencia de cada uno de los *primers* utilizados en las reacciones de PCR y secuenciación, su temperatura de *melting* y la longitud del amplicón para cada una de las regiones génicas amplificadas.

Gen	Primers	Tm (°C)	Amplicón (bp)
ABCG2	ABCG2_F: 5'-TCACGAGACTGTCAGGGACTT-3'	60	369
	ABCG2_R: 5'-CAATCCTCAGCTCCCTCAG-3'	60	
OPN	SPP1_F: 5'-GCCACAGAGGAGGACTTCAC-3'	60	764
	SPP1_R: 5'-CACAAGTATAGCACCAGACTCA-3'	60	
	OPN_F: 5'-GCTACCCCTAACTTCTGTTC-3'	59	424
	OPN_R: 5'-TCGTGGGGTTTTGAGAGAT-3'	59	
PPARGC1A	PPARGC1A_F: 5'-GGATCCCGCTTCTCATACTC-3'	60	414
	PPARGC1A_R: 5'-TCCAGGGGCTACTCAGTCAT-3'	60	
CSN1	CSN1_ex4_F: 5'-GGCCTGGTAAAGAAGGAACC-3'	60	469
	CSN1_ex4_R: 5'-TGCACAAAACAGCGTAAGTG-3'	59	
	CSN1_ex17_F: 5'-TGTGATGCGAATAGCCATGT-3'	60	460
	CSN1_ex17_R: 5'-CACTGCTCCACATGTTCTCTG-3'	60	
CSN3	CSN3_ex4F: 5'-TAGGTCACCTGCCCAAATTC-3'	60	462
	CSN3_ex4R: 5'-ATTAGCCATTTCGCCTTCT-3'	60	

Tabla A.IV.1. Primers usados en las reacciones de PCR y secuenciación, temperatura de *melting* de cada uno y longitud del amplicón esperado.

Abreviaturas: ABCG2: miembro 2 de la subfamilia G de la familia de transportadores ATP-binding cassette, OPN: Osteopontina, PPARGC1A: Coactivador 1-α del receptor y activado por proliferadores peroxisómicos, CSN1: α-caseína, CSN3: κ-caseína, β-LGB: β-lactoglobulina, DGAT1: *AcylCoA:diacylglycerol acyltransferase 1*, GH: Hormona de crecimiento, GHR: Receptor de hormona de crecimiento, PRLR: Receptor de prolactina, LTF: Lactoferrina, PRL: Prolactina, Tm: Temperatura de *melting*.

Gen	Primers	Tm (°C)	Amplicón (bp)
<i>B-LGB</i>	LGBprom_F : 5'-GCAGAGCCCTCGATACTGAC-3'	60	615
	LGBprom_R : 5'-CATGGTCTGGGTGACAATGA-3'	60	
	LGB_in2ex3_F: 5'-GGGTCAGAGTGCAGGAGAGA-3'	60	402
	LGB_in2ex3_R: 5'-CACCATCGATCTTGAACACC-3'	59	
	LGB_in3ex4_F: 5'-GAAGGGGCCACACCATTTT-3'	62	435
	LGB_in3ex4_R: 5'-CCGGTATATGACCACCCTCT-3'	59	
<i>DGAT1</i>	DGAT1_F: 5'-GCACCATCCTCTTCTCAAG-3'	57	411
	DGAT1_R: 5'-GGAAGCGCTTTCGGATG-3'	56	
<i>GH</i>	GH_F: 5'-CAACAGATGGCTGGCAACTA-3'	60	436
	GH_R: 5'-GCTTAGCCAGGAGAATGCAC-3'	60	
<i>GHR</i>	GHR_F: 5'-GTGGCTATCAAGTAAAATCATTGAC-3'	60	342
	GHR_R: 5'-ACTGGGTTGATGAAACACTTCACTC-3'	60	
<i>PRLR</i>	PRLR_inF : 5'-GTGCTCCATGGCAAGAGAAG-3'	61	446
	PRLR_inR : 5'-CACTAACCCCAACCTTCCAC-3'	59	
	PRLR_S18N_F: 5'-TCAGCAAGGAGCAAGAATCA-3'	60	452
	PRLR_S18N_R: 5'-TGGAAGTCAGAGCATGGTGA-3'	60	
<i>LTF</i>	LTF_F: 5'-GTCTCCACCCCACTCTTC-3'	59	419
	LTF_R: 5'-CGTGTCAAACCTGCCTGCTG-3'	62	
<i>PRL</i>	PRL_in1_F: 5'-TTTGCAGAACACAGGAGCAC-3'	60	489
	PRL_in1_R: 5'-ACGCAAAGGTTTCTTCTGGA-3'	60	

Tabla A.IV.1. *Primers* usados en las reacciones de PCR y secuenciación, temperatura de *melting* de cada uno y longitud del amplicón esperado (continuación).

ANEXO V.

1. Estrategias utilizadas en la selección de genes candidatos.

La búsqueda de genes candidatos se realizó siguiendo las seis estrategias mencionadas en la sección 5.A. de la introducción del Capítulo I de este trabajo de tesis. A continuación se describen las principales características funcionales de las proteínas codificadas por los genes seleccionados según los distintos abordajes de búsqueda bibliográfica, constituyendo un resumen de las razones que justifican su elección.

1.A. Genes codificantes de proteínas relacionadas al fenotipo estudiado que presenten diferentes variantes genéticas.

Caseínas.

Las α_{s1} - (CSN1S1), α_{s2} - (CSN1S2), β - (CSN2) y κ - (CSN3) caseínas se presentan en distintas variantes que surgen como consecuencia de deleciones o sustituciones aminoacídicas en posiciones determinadas dentro de la secuencia primaria de dichas proteínas.

Los genes codificantes de las caseínas bovinas se localizan en el cromosoma 6 formando un *cluster*. Su inducción se lleva a cabo bajo un complejo proceso multihormonal que involucra esencialmente a la prolactina, a glucocorticoides como también a la insulina y otros factores lactogénicos como la hormona de crecimiento. La expresión de los cuatro genes de las caseínas se da en forma coordinada y dependiente del tejido y estadio en el que se encuentra el organismo. Existe gran cantidad de evidencia que demuestra que la expresión de estos genes en la mayoría de las especies de mamíferos requiere un conjunto de factores de transcripción comunes.

Los distintos miembros de cada una de las familias de caseínas, al poseer una estructura primaria ligeramente diferente entre sí, difieren en su capacidad de unirse formando dímeros o estructuras de mayor número de unidades monoméricas, en su capacidad de unión de iones divalentes como Ca^{++} , en su hidrofobicidad, su actividad antimicrobiana, etc. Algunas de estas propiedades determinan la producción de una leche de mejor calidad composicional, con mayor actividad antibacteriana o con mejores aptitudes para la industria quesera. Por ejemplo, los quesos producidos a partir de leches con la variante C de la α_{s1} -caseína forman una cuajada más firme que los producidos a partir de leches conteniendo la variante B (Sadler y col., 1968; adaptado de Farrell y col., 2004). Es reconocido que la variante B de la κ -caseína aporta una calidad superior a la leche. Los quesos producidos a partir de leches con el genotipo BB para el *locus* de la κ -caseína presentan un mayor contenido proteico, un mayor rendimiento y mejor calidad que aquellos producidos a partir de leches con κ -caseína AA o AB (Martin y col., 2002).

Beta-lactoglobulina.

La β -lactoglobulina se presenta en once formas o variantes que son resultado de variaciones aminoacídicas en diferentes posiciones de la proteína. Las variantes genéticas A y B son las predominantes en la mayoría de las razas vacunas y la presencia de una u otra afecta marcadamente las propiedades de la leche debido a las diferentes propiedades fisicoquímicas

de las mismas. Se han reportados efectos cuantitativos de las variantes más comunes de esta proteína sobre la producción de leche y la aptitud en la industria quesera. La variante B de la β -lactoglobulina se asocia con un mayor contenido de caseínas y grasa en la leche bovina, mientras que la leche de vacas Holando homocigotas para la variante A de esta proteína contiene mayor cantidad de proteínas del suero y proteína total que aquellas de vacas con las otras variantes (Aleandri y col., 1990).

1.B. Genes con variantes alélicas que hayan sido reportadas en bibliografía como asociadas al rasgo en estudio.

Hormona de crecimiento (GH, Growth Hormone).

La hormona de crecimiento o somatotropina regula el crecimiento posnatal y el metabolismo y, por lo tanto, tiene un rol fundamental en el desarrollo de la glándula mamaria, el control de la lactancia, y los procesos de crecimiento y fertilidad en el ganado (Mullen y col., 2011). El gen de la hormona de crecimiento bovina (*bGH*) se localiza en el cromosoma 19 y codifica una proteína de alrededor de 22 kDa de peso molecular.

Mullen y col. (2011) han encontrado una asociación significativa entre un SNP (GH33) localizado en la región promotora del gen de la somatotropina y caracteres productivos, habiendo utilizado 610 vacas lecheras Holstein-Friesian.

Receptor de hormona de crecimiento (GHR, Growth Hormone Receptor).

Varios estudios han reportado asociaciones significativas entre polimorfismos en la región 5'UTR del gen del receptor de la somatotropina bovina y la producción de leche y el tamaño corporal en el ganado vacuno Holstein-Friesian (Waters y col., 2011).

El gen que codifica el receptor de la hormona de crecimiento bovina se localiza en el cromosoma 20. Una sustitución del nucleótido T por A en el exón VIII tiene como consecuencia el reemplazo de una fenilalanina por un residuo polar sin carga de tirosina en la posición 279 de la proteína (F279Y). El residuo de fenilalanina se ubica en el dominio transmembrana del receptor y es conservado entre varios mamíferos, entre los cuales se encuentran el ser humano, conejo, ratón, rata, perro, oveja. El polimorfismo F279Y ha sido asociado a efectos sobre el rendimiento y la composición de la leche en bovinos Holstein-Friesian y Jersey (Blott y col., 2003) y, en particular, al porcentaje de grasa y de proteína en la leche del ganado Finnish Ayrshire (Viitala y col., 2006).

Receptor 1 de LDL oxidada (OLR1, Oxidized LDL Receptor 1)

OLR1 es un receptor presente en la superficie de las células del endotelio vascular que tiene por función unir, internalizar y degradar proteolíticamente las lipoproteínas de baja densidad (LDLs) oxidadas (Mehta y Li, 2002). También remueve de la circulación células viejas y apoptóticas (Oka y col., 1998).

El gen *OLR1* bovino se localiza en el BTA 5 y codifica una proteína de 270 aminoácidos. La secuenciación de los productos de retrotranscripción y posterior PCR (RT-PCR) obtenidos a

partir del ARN total extraído de varios tejidos de bovinos adultos y fetos permitió la detección de varios SNPs en el gen *OLR1*: 2 SNPs en el exón 4, 5 en el intrón 4 y uno en la región 3'UTR. Este último SNP (C223A, o *OLR1*_{g.8232C>A}) mostró efectos significativos sobre el contenido y porcentaje de grasa en la leche de animales Holando de Norteamérica, estando el alelo C asociado a un aumento significativo de estos caracteres (Khatib y col., 2006). El mecanismo por el cual el polimorfismo C223A influenciaría la producción de grasa en la leche estaría basado en la eficiencia de remoción de las LDL oxidadas de la corriente sanguínea. El menor porcentaje de grasa observado en la leche de animales homocigotas para el alelo A de este SNP sería consecuencia de una expresión disminuida del gen *OLR1*, lo que conduciría a una insuficiente reducción de la concentración plasmática de las LDL oxidadas. Dado que los lípidos oxidados inhiben la expresión de los genes que codifican la lipoproteín lipasa y transportadores de ácidos grasos, su aumento en la circulación conduciría a una menor captación de ácidos grasos por la glándula mamaria y a la consiguiente disminución de la concentración de triglicéridos en la leche (Komisarek y Dorynek, 2009). En particular, el polimorfismo C223A se asocia al contenido de ácidos grasos de cadena larga en los triglicéridos de la leche, que son justamente los captados directamente de la sangre (Schennink y col., 2009).

ABCG2 (también llamado *BCRP* por codificar la BCRP, *Breast Cancer Resistance Protein*).

La proteína ABCG2 es un miembro de la familia ABC (*ATP Binding Cassette*) de transportadores transmembrana (específicamente, es el miembro 2 de la subfamilia G de los transportadores *ATP-binding cassette*). Está involucrada en el transporte activo de una amplia variedad de drogas, carcinógenos y toxinas de la dieta a nivel de la membrana plasmática de las células del intestino, riñón y otros órganos. El análisis de la presencia de la proteína ABCG2 en diferentes estadios del desarrollo mamario de ratones, vacas y humanos reveló que, si bien no se expresa en organismos vírgenes, su expresión se induce en la etapa final de la preñez y, especialmente, durante la lactancia. Al término del período de lactancia, durante la involución de la glándula, la expresión de *ABCG2* disminuye rápidamente (Jonker y col., 2005). Esta proteína es responsable de la secreción dependiente de ATP de sustancias de importancia clínica y toxicológica y de algunos nutrientes menores como vitamina B2 o colesterol a la leche (van Herwaarden y col., 2007; Farke y col., 2008).

El gen que codifica la proteína ABCG2 bovina se localiza en el BTA 6, dentro de la región correspondiente a un QTL que afecta la producción y composición de la leche. Este gen presenta un SNP en el exón 14 consistente en una mutación de A por C, causante de la sustitución no conservativa tirosina/serina en la posición 581 de la proteína (Y581S) correspondiente a la quinta región extracelular de la misma (Cohen-Zinder y col., 2005), que sería responsable del efecto del QTL. Este cambio aminoacídico afecta la función transportadora de ABCG2 e influencia significativamente las características relacionadas a la producción y composición de la leche del ganado Holstein-Friesian israelí (Cohen-Zinder y col., 2005) y Norwegian Red (Olsen y col., 2007). El alelo A, codificante de tirosina, es el que presenta mayor frecuencia poblacional y se asocia a un mayor porcentaje de grasa y proteína y a un menor rendimiento de leche.

Sintasa de ácidos grasos (*FASN*, *Fatty Acid Synthase*)

La enzima FASN es un complejo homodímero que interviene en la síntesis *de novo* de ácidos grasos de cadena larga, al catalizar la formación de palmitato a partir de acetil-coenzima A y malonil-coenzima A en presencia de NADPH. El gen *FASN* bovino está localizado el cromosoma 19 y está regulado a nivel transcripcional y post-transcripcional. Uno de los principales factores de transcripción implicados en su regulación es el SP1 (Schweizer y col., 2002), cuyo sitio de unión al ADN se ubica en la región 5'UTR.

Dado el rol central que tiene FASN en la lipogénesis *de novo*, es un gen candidato para el contenido de grasa en la leche de mamíferos. Además, varios QTLs que afectan el contenido de grasa en la leche y caracteres relacionados han sido localizados en el BTA 19 en la región del gen *FASN*. Se ha reportado la existencia de SNPs a lo largo de todo el gen *FASN*; algunos de ellos fueron asociados a características composicionales de la leche. Por secuenciación directa de productos obtenidos por PCR a partir de animales Holstein-Friesian, Roy y col. (2006) detectaron un SNP en el exón 1 no traducido del gen *FASN* ($FASN_{g.763G>C}$) y otro en el exón 34 ($FASN_{g.16009A>G}$, renombrado $FASN_{g.16024G>A}$ por Schennink y col., 2009). La sustitución G>C del SNP $FASN_{g.763G>C}$ altera el sitio de unión para el factor de transcripción SP1: el alelo G produce un plegamiento de mayor estabilidad en la región 5'UTR del ARNm de *FASN* y mantiene el sitio de unión a SP1, mientras que el alelo C pierde este sitio. Este SNP, por lo tanto, puede afectar la expresión proteica a través de la modulación de la tasa de transcripción del gen o de la estabilidad del ARNm de *FASN*. El alelo C se ha asociado a un alto contenido de grasa en la leche bovina. La mutación A>G en el exón 34 del gen *FASN* genera una sustitución no conservativa de treonina por alanina en la posición 1953 de la proteína (p.Thr1953Ala), estando el alelo G asociado a un mayor contenido de grasa en la leche. Morris y col. (2007) identificaron 5 SNPs en el gen *FASN*, dos de ellos asociados significativamente con el contenido de ácidos grasos de cadena media y larga en la leche ($FASN_{g.15531C>A}$ y $FASN_{g.15603G>A}$). Schennink en el año 2009 reportó una asociación entre el SNP $FASN_{g.17924A>G}$ identificado previamente por Morris y el porcentaje de grasa en la leche de animales Holstein-Friesian. Este SNP es exónico y provoca una sustitución no conservativa de un residuo de treonina por uno de alanina en la proteína FASN, lo que influenciaría la actividad de la enzima.

Esteaoril-Coenzima A desaturasa 1 (*SCD1*, Stearoyl-Coenzyme A Desaturase 1)

La enzima SCD, también llamada **delta-9 desaturasa**, tiene un rol clave en la biosíntesis de ácidos grasos monoinsaturados. Se localiza en el retículo endoplasmático y cataliza la inserción de un doble enlace entre los átomos de carbono 9 y 10 en un amplio espectro de ácidos grasos de cadena media y larga. En particular, es responsable de la conversión de los ácidos palmítico y esteárico en sus formas monoinsaturadas. En la glándula mamaria bovina, la SCD1 participa en la síntesis de alrededor del 50% del ácido oleico y hasta el 90% del ácido linoleico conjugado cis-9, trans-11 presentes en la leche (Rincón y col., 2012). En bovinos han sido identificadas y caracterizadas dos isoformas: SCD1 y SCD5. SCD5 se expresa principalmente en el cerebro aunque también lo hace en la glándula mamaria, mientras que la expresión de SCD1 tiene lugar en adipocitos, glándula mamaria y en menor medida en otros tejidos. La actividad mamaria de esta enzima en animales de raza Jersey es menor que en los de raza Holando (Beaulieu y Palmquist, 1995; Drackley y col., 2001).

El gen *SCD* es un componente clave en la vía de señalización promovida por la leptina. Se localiza en el BTA 26 y codifica una proteína de 355 aminoácidos. Varios SNPs han sido detectados a lo largo de la secuencia del gen en ganado Holstein, Jersey y Brown Swiss (Rincon y col., 2012). Uno de ellos causa una sustitución no conservativa de alanina a valina en el residuo 293 de la proteína (A293V) y ha sido asociado a la composición de ácidos grasos en la leche (Conte y col., 2010; Mele y col., 2007). A nivel del ADN genómico dicho SNP se denomina *SCD1_{g.10329C>T}*, siendo el alelo C el que codifica una alanina y el T el que codifica una valina en la posición 293 de la proteína. Se ha demostrado que el alelo C influencia positivamente la incorporación de dobles enlaces a ciertos ácidos grasos como el C14:0 y el C10:0 para formar sus respectivos productos monoinsaturados (Moioli y col., 2007).

Alfa-lactoalbúmina (α -LA)

La α -lactoalbúmina es una proteína esencial para la biosíntesis de lactosa en la glándula mamaria ya que aumenta la afinidad de la β -1,4-galactosiltransferasa por el sustrato glucosa, permitiendo la formación de lactosa (Brew y Hill, 1975).

Se ha demostrado que la sustitución de un nucleótido (A/G) en la posición +15 del ARNm de la α -lactoalbúmina se halla asociada a un incremento en la producción de leche por vacas de raza Holando. Las vacas con el alelo A presentan una mayor producción de leche, grasa y proteína, mientras que el alelo G se asocia a un aumento del porcentaje de grasa y proteína (Bleck y Bremel, 1993).

Leptina (*LEP*)

La leptina es una hormona producida principalmente por el tejido adiposo, aunque también se expresa en placenta, glándula mamaria, músculo esquelético, mucosa gástrica y cerebro. Su expresión está regulada por la masa y el tamaño de las células del tejido adiposo y por hormonas como insulina, hormona de crecimiento y glucocorticoides. Esta hormona tiene un rol importante en el control del peso corporal, la homeostasis energética, el apetito, la función inmune y la reproducción. Las concentraciones de leptina en los fluidos biológicos y en los tejidos varían a lo largo de la preñez y el período de lactancia. Durante la preñez se observa hiperleptinemia debida al aumento gestacional de la expresión de este gen en el tejido adiposo. La placenta de rumiantes presenta una muy baja expresión del gen codificante de la leptina a lo largo de la gestación, por lo que la producción placentaria no contribuye, como en el caso de humanos y ratones, a la hiperleptinemia observada. Hacia el momento del parto las concentraciones de leptina disminuyen debido, en parte, a los cambios que ocurren en el metabolismo energético antes del inicio de la lactancia (Liefers y col., 2005). En las etapas tempranas de la lactancia, las vacas lecheras se encuentran en un estado de balance energético negativo en el que las grasas almacenadas y los nutrientes ingeridos son utilizados por el proceso energéticamente demandante de lactogénesis, manteniendo niveles plasmáticos bajos de leptina hasta el tercer mes de lactancia aproximadamente.

En bovinos, el gen de la leptina se localiza en el BTA 4, contiene 3 exones y codifica una proteína de 167 aminoácidos. Se han detectado polimorfismos tanto en el promotor como en los intrones y exones del gen. En bibliografía existen reportes de asociaciones entre

polimorfismos en los genes de leptina y del receptor de leptina y las concentraciones plasmáticas de la hormona durante la preñez, la producción de leche y la ingestión de alimento en ganado lechero Holstein-Friesian. Liefers y col. (2005) hallaron en el promotor del gen tres SNPs asociados a fertilidad, balance energético y porcentaje de proteína en la leche. Buchanan y col. (2003) midieron el efecto del genotipo de un SNP del exón 2 del gen de leptina consistente en una transición C por T que codifica una sustitución de arginina por cisteína en la posición 25 del polipéptido (R25C). Este grupo de investigación encontró asociaciones significativas entre el genotipo TT y el aumento en el rendimiento de leche y proteína en la leche. Giblin y col. (2010) hallaron asociaciones significativas entre SNPs ubicados en el promotor del gen (LEP-2470, LEP-11238 y LEP-963) y en la región codificante (T7F y R25C) y caracteres relacionados a la producción lechera. En particular, el alelo T del SNP R25C se asoció con un aumento del rendimiento de leche y de la proteína en la leche y un descenso en la concentración de grasa en la leche.

1.C. Genes con perfiles de expresión asociados a la variación fenotípica del rasgo analizado.

Proteínas de leche uterina (*UTMP*, *Uterine Milk Proteins*).

Las UTMP son proteínas secretadas por vacas preñadas cuya función estaría relacionada con el mantenimiento de la preñez, la inhibición de proteasas, el control del crecimiento y sobrevivencia del feto y la supresión del sistema inmune materno. En ovejas, las UTMP son las principales proteínas detectadas en el fluido uterino desde el día 30 de preñez hasta su término (Moffatt y col., 1987).

Khatib y col. (2007) investigaron la distribución de transcritos del gen *UTMP* en varios tejidos bovinos fetales y adultos. En los fetos analizados, los transcritos de *UTMP* se hallaron principalmente en los cotiledones, mientras que en vacas preñadas se encontraron predominantemente en el endometrio, ovario y carúncula. La expresión mayoritaria del gen *UTMP*, localizado en el BTA 21, en tejidos reproductivos es consistente con la importancia del rol que juega esta proteína en el éxito reproductivo.

En bovinos lecheros se han descrito asociaciones en la región del gen *UTMP* con la vida productiva de los animales y la composición de su leche. La vida productiva se define como la cantidad de meses durante los cuales una vaca se encuentra en período de lactancia a lo largo de su vida, con límites de 10 meses por lactancia y hasta los siete años de vida del animal. En particular, un polimorfismo de tipo SNP A/G en posición 1296 del gen se ha asociado con la vida productiva, incrementándose ésta ante la presencia del alelo G (Khatib y col., 2005).

Prolactina (*PRL*)

La prolactina es una hormona polipeptídica con múltiples funciones, secretada por la glándula pituitaria anterior (Bole-Feysot y col., 1998). Está involucrada en el desarrollo de la glándula mamaria y en el inicio y mantenimiento de la lactancia. Su rol crítico en estos procesos fue demostrado en experimentos con disrupción del gen, en los que se vio afectado el desarrollo de la glándula mamaria, la lactogénesis y la expresión de genes relacionados a la producción de leche (Horseman y col., 1997). En base a esta información, el gen *PRL* es un

fuerte candidato funcional en análisis de asociación con caracteres productivos de bovinos lecheros. La variación alélica en la secuencia estructural o regulatoria del gen podría tener efectos directos o indirectos sobre la producción y composición de la leche (Brym y col., 2005).

El gen de la prolactina bovina se localiza en el BTA 23 y consiste en 5 exones que codifican una proteína de 199 aminoácidos. Se han reportado varios polimorfismos a lo largo de su secuencia, entre ellos PRL_{g.8398G>A}, el que ha mostrado análisis de asociación inconsistentes en diferentes poblaciones analizadas (Brym y col., 2005; Schennink y col., 2009).

1.D. Genes localizados en QTLs asociados al fenotipo en estudio.

ARL4A

La proteína ARL4 (*ADP-ribosylation factor-like 4*) ha sido incluida en la familia de los factores de ADP-ribosilación (ARFs, *ADP-ribosylation factors*) debido a su similitud con los mismos. Los ARFs constituyen, a su vez, una subfamilia de las GTPasas relacionadas con la proteína Ras y poseen funciones variadas ya que intervienen en la regulación del tráfico de vesículas, la endocitosis y la actividad de la fosfolipasa D. La ARL4A en particular, es una proteína altamente conservada que participa en las vías de transporte vesicular endocíticas y exocíticas en diferentes compartimentos celulares (Jacobs y col., 1999). Posee una alta expresión en la glándula mamaria bovina durante el período de lactancia.

XR_027435.1

Gen predicho por análisis computacional, similar al codificante de la proteína ribosomal S4.

ETV1 (ets variant 1)

El gen *ETV1* codifica un factor de transcripción localizado en el núcleo y la mitocondria que no se expresaría en la glándula mamaria.

SNX13 (Sorting nexin 13)

SNX13 pertenece a la familia de las proteínas SNX, implicadas en el tráfico intracelular. Esta proteína ha sido detectada en la glándula mamaria.

MEOX2 (Mesenchyme homeobox 2)

Las proteínas MEOX2 se localizan en el núcleo celular y actúan como factor de transcripción. Durante la lactancia, se observa una expresión elevada de estas proteínas en la glándula mamaria bovina (Rincón y col., 2009).

Coactivador 1- α del receptor y activado por proliferadores peroxisómicos (PPARGC1A, Peroxisome Proliferator Activated Receptor Gamma, Coactivator 1 Alpha)

El *PPARGC1A* es un candidato posicional y funcional para un QTL localizado en la región central del BTA 6. El gen consta de 13 exones que codifican una proteína de 797 aminoácidos. A pesar de tratarse de una proteína ubicua, la mayor expresión del gen tiene lugar en hígado,

riñón, glándula tiroides y glándula pituitaria. Se ha observado una alta similitud entre las secuencias de ADNc y proteína del *PPARGC1A* de bovinos y su homólogo de humano, rata y ratón (Weikard y col., 2005).

La proteína *PPARGC1A* tiene una función clave en la activación de numerosos receptores nucleares de hormonas y factores de transcripción que regulan la homeostasis de energía celular. Además, media la expresión de genes implicados en la termogénesis adaptativa, el metabolismo oxidativo, la adipogénesis y la gluconeogénesis.

En el período de lactancia la disponibilidad de glucosa dependiente de la continua gluconeogénesis hepática es un factor limitante en la producción de leche. De hecho, durante la lactancia, para satisfacer los requerimientos de la glándula mamaria, la producción hepática de glucosa aumenta hasta siete veces en relación a la del período seco (Bell y Bauman, 1997). En la glándula mamaria el 60-70% de la glucosa es utilizada en la síntesis de lactosa y el 20-30% entra en la vía de las pentosas fosfato para generar el NADPH_2 que se utiliza como poder reductor en la síntesis de ácidos grasos de la leche. Por lo tanto, la disponibilidad de glucosa es limitante tanto para la síntesis de lactosa como para la de los ácidos grasos presentes en la leche. Debido a su función en la regulación de vías relacionadas a la homeostasis energética y a su capacidad para coordinar el proceso de adaptación metabólica en hígado, tejido graso y músculo, el *PPARGC1A* puede considerarse como un mediador general de las demandas metabólicas que acompañan el inicio y la progresión de la lactancia en las vacas lecheras.

Weikard y col. (2005) secuenciaron el locus del gen *PPARGC1A* y hallaron 11 polimorfismos, ubicados en exones, intrones, en la región 3'UTR y hacia la región 5' del gen. Uno de ellos, el SNP c.1892+19T>C en el intrón 9 del gen, se asoció con el contenido de grasa de la leche de animales de raza German Holstein. En particular, el alelo C se encontró asociado a una menor producción de grasa en la leche. Los investigadores argumentan que esta variación en la región no codificante del gen podría tener efectos sobre el procesamiento del ARNm, la eficiencia del *splicing* o el nivel de transcriptos, lo que afectaría la respuesta a la demanda incrementada de glucosa por la glándula mamaria durante la lactancia. Sin embargo, la asociación anterior no pudo ser reproducida en ganado Hosltein-Friesian (Schennink y col., 2009; Komisarek y Dorynek, 2009).

Osteopontina (*OPN*, también denominada *SPP1*)

La osteopontina es una glicoproteína fosforilada que se expresa en numerosos tejidos y está implicada en las interacciones célula-matriz extracelular y en la señalización celular. La proteína se halla presente en la leche, el plasma y la orina bovinas. Tiene un rol clave en el desarrollo de la glándula mamaria, la ramificación del sistema epitelial de ductos mamarios y la lactancia. El principal origen de esta proteína reside en las células epiteliales de la glándula mamaria y los monocitos y macrófagos presentes en la leche. En la leche cruda bovina fue detectada en una concentración de 8 mg/L (Bayless y col., 1997).

El gen *OPN* es, por ende, un gen candidato funcional y posicional ya que se localiza en el BTA 6, en la región que alberga un QTL para caracteres relacionados a la producción lechera.

Leonard y col. (2005) detectaron un SNP en el intrón 4 del gen, consistente en una sustitución de citosina por timina y hallaron una asociación entre el alelo C y un incremento en los porcentajes de proteína y grasa en la leche de animales de raza Holando. Schnabel y col. (2005) identificaron 9 polimorfismos en el gen *OPN*. Uno de ellos, el OPN3907, es un indel localizado en dirección 5' del sitio de inicio de la transcripción del gen, en una región que contiene elementos regulatorios específicos de tejido. Este indel tiene lugar en una región de secuencia simple compuesta por una repetición de timinas y se caracteriza por presentar dos alelos consistentes en 9 (T9) o 10 (T10) timinas contiguas. El mismo grupo de investigación halló efectos estadísticamente significativos del alelo T9 sobre el aumento en la producción de leche y el descenso en los porcentajes de grasa y de proteína en la leche de animales de raza Holando. Sin embargo, estos resultados no pudieron ser reproducidos en ganado de raza Norwegian Red (Olsen y col., 2007).

Poliquistina 2 (*PKD2*)

Un tercer gen candidato posicional y funcional potencialmente responsable de los efectos de QTLs del BTA 6 relacionados a producción y composición lechera es el *PKD2*. Este gen codifica una proteína integral de membrana involucrada en la homeostasis del calcio intracelular y en otras vías de transducción de señales (Nauli y col., 2003). Dado que el calcio es el principal contribuyente a la osmolaridad de la leche, diferencias en la expresión de *PKD2* podrían afectar el contenido de agua en la leche reduciendo indirectamente los porcentajes de grasa y proteína en la misma (Olsen y col., 2007).

***DGAT1* (*AcylCoA:diacylglycerol acyltransferase 1*)**

El gen *DGAT1* se localiza en el cromosoma 14 bovino y codifica una proteína de 489 aminoácidos. Este gen fue considerado un fuerte candidato para explicar los efectos del QTL relacionado al contenido de grasa de la leche mencionado anteriormente, debido a que el 98% de los lípidos de la leche son triglicéridos, a que la enzima *DGAT1* cataliza el paso final en la síntesis de triglicéridos y a que en ratones *knock out* para el gen *DGAT1* se ve inhibida la lactancia (Smith y col., 2000). El efecto del QTL fue explicado a través de la detección de una mutación consistente en la sustitución del dinucleótido AA por GC en el exón 8 del gen, que causa una sustitución no conservativa K232A del aminoácido lisina cargado positivamente por un residuo neutro e hidrofóbico de alanina (Grisart y col., 2002).

Receptor de prolactina (*PRLR*).

El receptor de prolactina es una proteína transmembrana perteneciente a la familia de receptores de citoquinas. Presenta dos isoformas que se originan por *splicing* alternativo del mismo gen y difieren en la longitud y la secuencia de su porción citoplasmática. Las hormonas capaces de unirse al *PRLR* son denominadas colectivamente hormonas lactogénicas y sus principales funciones son la estimulación del desarrollo de la glándula mamaria durante la preñez y la regulación de la lactancia luego del parto. Estudios en ratones con el gen *PRLR* deletado han permitido observar alteraciones importantes a nivel de la reproducción de las hembras principalmente, viéndose afectadas la frecuencia de ovulación, la tasa de fertilización,

la receptividad uterina a la implantación del embrión y el desarrollo de la glándula mamaria y de la lactancia (Ormandy y col., 1997).

El gen codificante del receptor de prolactina bovino, localizado en el BTA 20, constituye un gen candidato a considerar en la selección de marcadores a incluir en estudios de asociación con caracteres relacionados a producción y calidad de leche, dado que es un regulador clave del proceso de reproducción. La sustitución de dos nucleótidos contiguos en la porción del gen *PRLR* codificante del péptido señal de la proteína genera un SNP (S18N) que origina un cambio de serina por asparagina en el péptido. Este polimorfismo fue asociado en ganado Finnish Ayrshire a la producción de leche y de grasa y proteína en la leche (Viitala y col., 2006).

1.E. Genes regulados post-transcripcionalmente por miRNAs expresados en el/los órganos relacionados al rasgo estudiado.

Lactoferrina (*LTF*)

El miRNA miR-214 se ha identificado en líneas celulares humanas de cáncer de mama. Este miRNA es perfectamente complementario a una secuencia conservada localizada en la región 3' no traducida (3' UTR, *untranslated region*) del gen de la lactoferrina humano y de otros organismos, entre ellos los bovinos. La transfección de miR-214 en la línea celular HC11 del epitelio mamario de ratón resulta en el bloqueo de la estimulación del gen de la lactoferrina y la represión de la expresión proteica desencadenada por tratamiento con estradiol o prolactina (Liao y col., 2010). Estos resultados demuestran que la traducción del ARNm de lactoferrina está bajo regulación directa del miR-214 en las células epiteliales de la glándula mamaria.

1.F. Genes cuya expresión alterada en animales *knock out* o transgénicos haya resultado en fenotipos específicos asociados al rasgo en estudio.

Transductor de señales y activador de la transcripción 5A (*STAT5A*, *Signal Transducer and Activator of Transcription 5A*)

Las proteínas STAT5 pertenecen a la familia de proteínas STATs, involucradas en la proliferación, diferenciación y apoptosis celular. Existen dos subtipos STAT5 íntimamente relacionados que se denominan STAT5A y STAT5B y poseen un 96% de homología en su secuencia proteica. La activación de estos factores de transcripción se da en respuesta a una gran variedad de citoquinas como prolactina, hormona de crecimiento, eritropoyetina, trombopoyetina e interleuquinas-2, -3 y -12. Al unirse estas citoquinas a sus respectivos receptores celulares, la quinasa JAK2 asociada a los receptores fosforila las proteínas STAT5 en un residuo crítico de tirosina, activándolas. Las proteínas STAT5A y STAT5B activadas forman homo- y hetero-dímeros que se translocan al núcleo donde se unen a secuencias específicas de ADN denominadas GAS (secuencias activadas por interferón- γ) y estimulan la transcripción de genes diana (Miyoshi y col., 2001; Hennighausen y Robinson, 2008).

Mientras que STAT5A es la isoforma de STAT5 prevalente en el tejido mamario, STAT5B es más abundante en músculo e hígado. La activación de STAT5A aumenta fuertemente durante la preñez cuando se elevan los niveles de prolactina y lactógeno placentario para inducir el desarrollo de los alvéolos mamarios y la transcripción de genes codificantes de las proteínas de la leche. De hecho, la señalización vía prolactina y STAT5A es crítica para la proliferación y diferenciación de los alvéolos mamarios durante la preñez. Dado que la ausencia completa de proteínas STAT5 causa muerte perinatal (Cui y col., 2004), para evaluar el desarrollo histológico y funcional del epitelio mamario en estadíos adultos se han realizado transplantes del epitelio y transfecciones con transgenes con el gen *STAT5* deletado y sólo activos en células del epitelio mamario. Las células epiteliales carentes del gen *STAT5*, si bien son capaces de formar ductos normales, no desarrollan alvéolos ni expresan genes codificantes de proteínas de la leche. La delección del gen *STAT5A* causa displasia mamaria con atenuación del desarrollo alveolar y disminución de la capacidad de lactancia (Liu y col., 1997), mientras que la delección de *STAT5B* no afecta el desarrollo de la glándula mamaria (Udy y col., 1997). Estos estudios demuestran que la proteína STAT5A está implicada en el desarrollo de la glándula mamaria adulta y en la lactogénesis.

El gen *STAT5A* bovino se localiza en el BTA 19 y contiene 19 exones que codifican una proteína de 794 aminoácidos. A lo largo del gen existe gran cantidad de polimorfismos, entre ellos un SNP en el intrón 9 (*STAT5A*_{g.9501G>A}) que ha sido asociado a la producción de leche y contenido de grasa y proteína en la leche de la primera y segunda lactancias de vacas de raza Jersey por Brym y col. (2004). En estudios independientes con distintas poblaciones también se han encontrado asociaciones significativas entre dicho SNP y la producción de leche, producción de proteína en la leche y porcentaje de grasa en la leche (He y col., 2007) y la producción de leche a 305 días y el porcentaje de proteína en la leche (Bao y col., 2008) de ganado Holando. El alelo A de este SNP fue asociado con un mayor porcentaje de grasa en animales de raza Jersey (Brym y col., 2004) y con mayor contenido de ácidos grasos de cadena media saturados y menor contenido de insaturados en animales Holstein-Friesian (Schennink y col., 2009).

2. Selección de SNPs.

A continuación se expone la lista de marcadores obtenidos según la estrategia de búsqueda bibliográfica de SNPs asociados al fenotipo en estudio, localizados en regiones intergénicas o en genes predichos *in silico* (Kolbehdari y col., 2009; Jiang y col., 2010). Para cada uno de los caracteres testeados se indican el cromosoma y el gen (entre paréntesis) al que pertenece cada SNP.

- Producción de leche.

BTA 1: rs41631818 (*MGC128242*), rs41629125 (*ITGB5*)

BTA 3: Hapmap38643-BTA-95454 (*LOC534011*)

BTA 5: rs41655901 (*GALNT6*), rs41592943 (*GUCY2C*)

BTA 8: rs41658330 (*FANCC*)

BTA 11: rs41569023 (*ROCK2*)

BTA 14: ARS-BFGL-NGS-34135 (*CYHR1*), rs17870736 (*VPS28*), **rs41256919 (*MAF1*)**, Hapmap25486-BTC-072553 (*GML*), ARS-BFGL-NGS-100480 (*NIBP*), rs41633631 (*TG*), **rs41628862 (*BIG1*)**, **rs41587081 (*ZFHX4*)**

BTA 18: rs41581694 (*FOXA3*)

BTA 19: rs41577598 (*BAIAP2*)

BTA 21: rs41608371 (*FBLN5*), **rs41644615 (*SERPINA5*)**

- Contenido de proteína en la leche.

BTA 1: **rs41586699 (*PDE9A*)**, BFGL-NGS-113002 (*DIP2A*), **rs41631818 (*MGC128242*)**, rs41629125 (*ITGB5*)

BTA 3: rs41589462 (*KCND3*), Hapmap38643-BTA-95454 (*LOC534011*),

BTA 8: rs29011990 (*MGC155155*)

BTA 11: **rs41569023 (*ROCK2*)**

BTA 14: ARS-BFGL-NGS-100480 (*NIBP*), **rs41628862 (*BIG1*)**

BTA 18: rs41581694 (*FOXA3*)

BTA 21: **rs41644615 (*SERPINA5*)**

BTA 26: rs41648723 (*CTBP2*)

BTA 28: rs41606880 (*JDP1*)

- Contenido de grasa en la leche.

BTA 1: **rs41637121 (*APP*)**, **rs41631818 (*MGC128242*)**, rs41629125 (*ITGB5*), rs41634488 (*LOC786403*)

BTA 4: rs41588659 (*COL1A2*)

BTA 5: **rs41652648 (*ITPR2*)**, rs41591894 (*ITPR2*), **rs41592942 (*GUCY2C*)**

BTA 14: ARS-BFGL-NGS-34135 (*CYHR1*), rs17870736 (*VPS28*), **rs41579049 (*5-OPase*)**, rs41580517 (*KCNQ3*)

BTA 21: rs41644615 (*SERPINA5*)

BTA 24: **rs41645253 (*MGC139244*)**

BTA 28: rs41653440 (*PSAP*), rs41569649 (*CAMK2G*)

- Porcentaje de proteína en la leche.

BTA 4: rs41650658 (*NRCAM*)

BTA 5: **rs29014633 (*CACNG2*)**, rs41590827 (*RAC2*)

BTA 6: **rs29018333 (*GRID2*)**, **rs41622323 (*PKD2*)**

BTA 10: rs41593881 (*HIF1A*)

BTA 13: rs29021058 (*PLCG1*)

BTA 14: ARS-BFGL-NGS-34135 (*CYHR1*), rs17870736 (*VPS28*), Hapmap27703-BTC-053907 (*NIBP*)

BTA 17: rs41570561 (*SCARB1*)

BTA 20: BFGL-NGS-118998 (*GHR*), ARS-BFGL-NGS-38482 (*RICTOR*), rs41581059 (*LOC100138964*), rs41942492 (*NIPBL*), **rs29018751 (*NIPBL*)**, rs42954630 (*NIPBL*)

BTA 22: rs41637636 (*SLC38A3*)

- Porcentaje de grasa en la leche.

BTA 2: rs41617243 (*KBTBD10*)

BTA 5: rs41652649 (*ITPR2*)

BTA 6: **rs29018853** (***LEC3***)

BTA 8: rs43499009 (*NFIB*)

BTA 11: ARS-BFGL-NGS-26919 (*LOC526069*)

BTA 14: ARS-BFGL-NGS-34135 (*CYHR1*), rs17870736 (*VPS28*), **rs41256919** (***MAF1***), Hapmap25384-BTC-001997 (*MAPK15*), Hapmap24715-BTC-001973 (*MAPK15*), ARS-BFGL-NGS-101653 (*EEF1D*), ARS-BFGL-NGS-26520 (*ZC3H3*), Hapmap25486-BTC-072553 (*GML*), Hapmap30374-BTC-002159 (*RHPN1*), ARS-BFGL-NGS-33248 (*PTK2*), **rs41624797** (***PTK2***), Hapmap26072-BTC-065132 (*EIF2C2*), ARS-BFGL-NGS-56327 (*NIBP*), ARS-BFGL-NGS-100480 (*NIBP*), **rs55617160** (***NIBP***), UA-IFASA-5765 (*NIBP*), ARS-BFGL-BAC-25166 (*NIBP*), Hapmap27703-BTC-053907 (*NIBP*), ARS-BFGL-NGS-3571 (*COL22A1*), BFGL-NGS-118478 (*COL22A1*), Hapmap30091-BTC-005211 (*LOC618755*), UA-IFASA-6647 (*KHDRBS3*), Hapmap32948-BTC-047992 (*KHDRBS3*), rs41579049 (*5-OPase*), rs41633631 (*TG*), rs41567322 (*TG*), rs41580517 (*KCNQ3*), **rs41579063** (***BIG1***), **rs41587081** (***ZFHX4***)

BTA 17: rs41639879 (*LOC505156*), rs41570561 (*SCARB1*)

Los SNPs seleccionados a partir de esta lista para ser incluidos en el diseño de los kits de genotipificación se indican en negrita en la lista anterior y, además, se mencionan en el Capítulo I, punto 2.C. de la sección de Resultados.

ANEXO VI.**Equilibrio de Hardy-Weinberg.**

El principio de Hardy-Weinberg establece que una población se halla en equilibrio, es decir, sus frecuencias genotípicas permanecen constantes de generación en generación, si cumple los postulados que se mencionan a continuación (Thompson & Thompson, 2008).

- La población es grande y los apareamientos ocurren al azar con respecto al *locus* en cuestión.
- Las frecuencias alélicas permanecen constantes a lo largo del tiempo puesto que no existe deriva génica, la tasa de mutación es inapreciable, no hay selección contra un genotipo determinado, ni migración significativa de individuos con frecuencias alélicas muy diferentes a las de la población en cuestión.

Dada una frecuencia alélica mínima = q para un *locus* bialélico que se halla en EHW, las probabilidades de los tres genotipos posibles (aa, Aa y AA) son $(1-q)^2$, $2q(1-q)$ y q^2 , respectivamente (Anderson y col., 2010).

ANEXO VII.

Modelos matemáticos utilizados para describir curvas de lactancia.

La primera referencia al modelado de la curva de lactancia se atribuye a Brody y col. (1924). Debido a dificultades para realizar los cálculos y limitaciones computacionales, los primeros modelos utilizados consistieron en transformaciones logarítmicas de funciones lineales exponenciales y polinómicas, entre otras. Posteriormente, hacia fines de la década de 1980 y durante la década de 1990, se utilizaron modelos no lineales (Silvestre y col., 2006).

Entre los modelos que se han utilizado en la estimación de curvas de lactancia para predecir la producción total de leche a partir de registros de producción parciales, quizás los más utilizados en el ganado vacuno son la función gamma incompleta de Wood (Wood, 1967), el modelo exponencial de Wilmink (Wilmink, 1987), el de regresión polinómica de Ali y Schaeffer (Ali y Schaeffer, 1987) y los polinomios de Legendre. Los tres primeros comparten la característica de haber sido desarrollados específicamente para modelar curvas de lactancia. En las últimas décadas comenzaron a aplicarse herramientas matemáticas generales, como los *splines* (curva determinada por varias curvas polinómicas sucesivas) y los polinomios de Legendre en el modelado de estas curvas (White y col., 1999; Macciotta y col., 2005).

Los modelos que describen las curvas de lactancia son funciones $y=f(t)$, en donde y es la producción diaria de leche de un animal, y t el día de lactancia, es decir, el tiempo transcurrido en la lactancia desde el momento del parto expresado en días. A continuación se exponen las funciones matemáticas que utilizan los modelos que han sido más utilizados en el ganado vacuno.

- Modelo de Wood (Wood, 1967)

$$y_t = at^b e^{-ct} + e_t$$

donde a es un parámetro que representa el rendimiento al comienzo de la lactancia, b y c son parámetros que se asocian a las porciones ascendente y descendente de la curva de lactancia, respectivamente, y e_t es el error residual del modelo.

- Modelo de Wilmink (Wilmink, 1987)

$$y_t = a + be^{-kt} + ct + e_t$$

donde a , b y c son coeficientes de regresión asociados con el nivel de producción, el incremento en la producción antes del pico y el descenso de producción después del pico, respectivamente, k es un parámetro que se relaciona con el momento en que tiene lugar el pico de la curva y generalmente asume un valor fijo que se determina en un análisis previo, y e_t es el error residual del modelo.

- Función de regresión polinómica de Ali y Schaeffer (Ali y Schaeffer, 1987)

$$y_t = a + b(t/305) + c(t/305)^2 + d[\ln(305/t)] + e[\ln(305/t)]^2 + e_t$$

donde a , b , c , d y e son coeficientes de regresión asociados con la producción en el pico de la curva (a), en la porción descendente (b y c) y en la porción ascendente de la misma (d y e), y e_t es el error residual del modelo.

- Polinomios de Legendre (Silvestre y col., 2006)

Los polinomios de Legendre (LEG) son funciones polinómicas de grado n y dominio $n+1$. La ecuación que describe una observación única puede escribirse de la siguiente manera:

$$y_t = \sum_{i=0}^n \alpha_i \phi_i(w) + e_t$$

donde α_i es un coeficiente del modelo y w una función de tiempo estandarizada que varía entre -1 y 1, y se define como:

$$w = 2 \left(\frac{t - t_{min}}{t_{max} - t_{min}} \right) - 1$$

siendo t_{min} y t_{max} los días de lactancia más temprano y más tardío, respectivamente.

Por otro lado, $\phi_n(w)$ se define como:

$$\phi_n(w) = \sqrt{\frac{2n+1}{2}} P_n(w)$$

donde $P_n(w)$ es un polinomio de grado n y $\phi_n(w)$ es el polinomio normalizado.

Los primeros cinco polinomios de Legendre, definidos como funciones de unidades estandarizadas de tiempo se exponen a continuación:

$$P_0(w) = 1 \quad P_1(w) = w \quad P_2(w) = \frac{1}{2} (3w^2 - 1)$$

$$P_3(w) = \frac{1}{2} (5w^3 - 3w) \quad P_4(w) = \frac{1}{8} (35w^4 - 30w^2 + 3)$$

Las funciones de grado 2, 3 y 4 (LEG2, LEG3 y LEG4) requieren un mínimo de 4, 5 y 6 registros por lactancia, respectivamente.

En todos los modelos enunciados se asume que los errores residuales son independientes y están homogéneamente distribuidos, con un valor promedio de cero y una varianza constante.

La determinación de la bondad de ajuste y comparación de los modelos pueden realizarse según distintos criterios, siendo el análisis de los errores residuales el método más utilizado. Estos errores se definen, para un registro dado de producción diaria, como la diferencia entre la producción de leche observada y la predicha por el modelo utilizado (Silvestre y col., 2006).

Para un modelo dado, el valor residual para un día particular se calcula como el promedio de los residuos correspondientes a la lactancia de cada uno de los animales estudiados.

Los modelos descritos anteriormente para el modelado de las curvas de producción de leche también pueden ser utilizados para describir la evolución de la producción de grasa y proteína en la leche a través del transcurso de la lactancia. En la literatura se han reportado curvas para los componentes de la leche con distintas formas: continuamente decreciente (Schutz y col., 1990; Stanton y col., 1992) y con la forma típica de las curvas de producción de leche (Pollott, 2004).

ANEXO VIII.

Aportes de efectos fijos al valor de cría estimado para producción de leche.

En el modelo utilizado en la estimación de los valores de cría para producción lechera mediante el programa WOMBAT (Meyer, 2007) se incluyeron cuatro efectos fijos: *PH* (porcentaje de raza Holando), *añonac* (año de nacimiento), *tambo* y *EIL-AIL* (estación y año de inicio de la lactancia). En la Tabla A.VIII.1 se indican los aportes de cada uno de ellos al EBV.

Efecto	Clase	Aporte
<i>PH</i>	100	474,576
	75	198,771
	50	-192,573
	25	-480,775
<i>añonac</i>	2000-2001	266,486
	2002	225,233
	2003	39,9708
	2004	-207,274
	2005-2006	-324,416
<i>tambo</i>	51	179,025
	52	-258,375
	53	-195,759
	55	352,901
	57	80,0366
	58	471,689
	60	-259,647
	61	4,13011
	62	16,1890
	63	685,041
	64	-434,613
	67	-946,621
	68	-243,690
69	350,549	
70	256,824	
71	-84,2737	
72	-10,8514	
73	37,4456	

Tabla A.VIII.1. Aportes de efectos fijos a la producción de leche.

En la tabla se indican los aportes de los efectos fijos *PH* (porcentaje de raza Holando), *añonac* (año de nacimiento), *tambo* y *EIL-AIL* (estación y año de inicio de la lactancia) a la producción de leche, como el resultado de la estimación por mínimos cuadrados generalizados.

Efecto	Clase	Aporte
	Grupo 1	-209,643
	Grupo 2	-329,920
	Grupo 3	-122,778
	Grupo 4	-63,2981
	Grupo 5	726,417
	Grupo 6	-508,432
	Grupo 7	28,5454
	Grupo 8	-261,874
	Grupo 9	34,8518
<i>EIL-AIL</i>	Grupo 10	806,329
	Grupo 11	-226,118
	Grupo 12	-223,073
	Grupo 13	-185,443
	Grupo 14	205,640
	Grupo 15	636,202
	Grupo 16	-358,806
	Grupo 17	-198,856
	Grupo 18	-197,991
	Grupo 19	448,248

Tabla A.VIII.1. Aportes de efectos fijos a la producción de leche (continuación).

ANEXO IX.

Técnica de secuenciación de regiones seleccionadas en animales con fenotipos extremos.

En primer lugar se amplificaron por reacción en cadena de la polimerasa (PCR) en los 34 animales seleccionados por presentar EBVs ubicados en los extremos de la distribución fenotípica (tanto inter como intrafamiliarmente), cuatro regiones correspondientes a tres genes candidatos: *S100B*, *BTN1A1* y *PRMT2*. Estos genes fueron seleccionados por localizarse cerca de los SNPs que resultaron significativamente asociados al EBV para producción de leche en el Capítulo III, y por poseer un perfil de expresión o función que sugería un posible rol en las diferencias observadas a nivel fenotípico.

Los volúmenes y concentraciones de cada reactivo de PCR fueron los siguientes:

- Buffer 10X	1,50 μ l
- Cl ₂ Mg 25 mM	1,20 μ l
- dNTPs 2mM	1,50 μ l
- Primer F 10 μ M	0,50 μ l
- Primer R 10 μ M	0,50 μ l
- Enzima INBIO	0,06 μ l
- ADN	entre 7 y 50 ng, según el gen a amplificar
- Agua bidestilada	hasta completar 15 μ l

En la reacción se utilizó la enzima INBIO ADN Polimerasa 5U/ μ l (Kit T-Plus ADN Polimerasa 500U, INBIO Highway® Biología Molecular), junto a su *buffer* y solución de Cl₂Mg correspondientes. Los dNTPs usados fueron diluciones hechas a partir del dNTP set 100mM Solutions (dATP, dCTP, dGTP y dTTP) de Illustra, GE Healthcare. Las condiciones de PCR fueron las siguientes: desnaturalización del ADN, 3 min a 94°C; seguida de 35 ciclos de 30 seg a 94°C (desnaturalización), 30 seg a la temperatura óptima de *annealing* o hibridización de cada par de *primers* y 1 min a 72°C (polimerización, extensión); y una etapa de extensión final de 10 min a 72°C. Las reacciones se realizaron en cicladores Eppendorf Mastercycler. El producto obtenido se visualizó en gel de agarosa al 1% por comparación con un estándar de tamaño conocido.

Una alícuota del amplicón se secuenció utilizando los mismos *primers* que en la reacción de PCR, según el método de dideoxinucleótidos de Sanger. La reacción de secuenciación se realizó en un ciclador Perkin Elmer GeneAmp PCR System 9600 de acuerdo a las siguientes condiciones: 1 min a 96°C, seguido de 30 ciclos consistentes en 10 seg a 96°C, 5 seg a 50°C y 4 min a 60°C. Los volúmenes de los reactivos utilizados en la reacción figuran a continuación:

- Mix	0,50 μ l
- Buffer 5X	1,75 μ l
- Primer	1,00 μ l
- ADN (dilución)	2,00 μ l
- Agua bidestilada	4,75 μ l (para llevar a un volumen final de 10 μ l)

La mezcla de reacción (*mix*) fue la BigDye® Terminator Cycle Sequencing Kit de Applied Biosystems (California, EE.UU.) y el *buffer* 5X correspondiente el BigDye® Terminator v3.1 del mismo fabricante. Como molde se utilizaron diluciones de los amplicones obtenidos en la primera reacción de PCR, de modo de lograr soluciones de concentración 5 - 20 ng/ μ l, adecuadas para secuenciar. Las diluciones óptimas para cada uno de los productos de PCR obtenidos, así como la secuencia de los *primers* diseñados para las reacciones de PCR y secuenciación, su temperatura de *annealing* óptima y la longitud del amplicón para cada una de las regiones génicas amplificadas se muestran en la Tabla A.IX.I.

Gen	Primer amplificación	Ta (°C)	Amplicón (bp)	Cantidad de ADN (ng)	Primer secuenciación	Dilución secuenciación
PRMT2	PRMT2_F 5'-CTGGTCGGTGAGAACAGAGA-3'	60	701	50	ambos	1:15
	PRMT2_R 5'-AAACACAGCCCAGGTTTCAG-3'					
BTN1A1	BTN1A1_A_F 5'-TGAAGTTCCCGACAAACTCG-3'	65	710	7	ambos	1:20
	BTN1A1_A_R 5'-CGGTATTCAGCCATCTCCTC-3'					
	BTN1A1_B_F 5'-GCCCTAGAGATTTGACAGCA-3'	60	714	10	ambos	1:20
	BTN1A1_B_R 5'-ATCTGCCAGGAGGTTGTAGC-3'					
S100B	S100B_F 5'-GGTCAAACCTCCGTCTCTGCT-3'	65	701	10	Forward	1:15
	S100B_R 5'-ATAGCCTCCTGGTCACTGGA-3'					

Tabla A.IX.1. *Primers* usados en las reacciones de PCR y secuenciación.

En la tabla se indican las secuencias de los *primers* diseñados para las reacciones de PCR y secuenciación, su temperatura de *annealing* óptima (Ta), la longitud del amplicón, la cantidad de ADN molde en la PCR, el *primer* que se utilizó en la reacción de secuenciación y la dilución del producto de PCR utilizada en esta reacción.

BIBLIOGRAFÍA.

Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin – rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genetics*. 2002, 30:97-101.

Alais C. *Ciencia de la leche*. Editorial Reverté. 1985.

Aleandri R, Buttazzoni LG, Schneider JC, Caroli A, Davoli R. The effects of milk protein polymorphisms on milk components and cheese-producing ability. *Journal of Dairy Science*. 1990, 73:241-255.

Ali TE, Schaeffer LR. Accounting for covariances among test day milk yields in dairy cows. *Canadian Journal of Animal Science*. 1987, 67:637-644.

Alpay F, Zare Y, Kamalludin MH, Huang X, Shi X, Shook GE, Collins MT, Kirkpatrick BW. Genome-Wide Association Study of Susceptibility to Infection by *Mycobacterium avium* Subspecies *paratuberculosis* in Holstein Cattle. *PLoS ONE*. 2014, 9(12):e111704.

Amin N, van Duijn CM, Aulchenko YS. A Genomic Background Based Method for Association Analysis in Related Individuals. *PLoS ONE*. 2007, 2(12):e1274.

Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nature Protocols*. 2010, 5(9):1564-1573.

Arranz JJ, Coppieters W, Berzi P, Cambisano N, Grisart B, Karim L, Marcq F, Moreau L, Mezer C, Riquet J, Simon P, Vanmanshoven P, Wagenaar D, Georges M. A QTL affecting milk yield and composition maps to bovine chromosome 20: a confirmation. *Animal Genetics*. 1998, 29:107-115.

Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics*. 2007, 23(10):1294-1296.

Bagnato A, Schiavini F, Rossoni A, Maltecca C, Dolezal M, Medugorac I, Sölkner J, Russo V, Fontanesi L, Friedmann A, Soller M, Lipkin E. Quantitative trait loci affecting milk yield and protein percentage in a three-country Brown Swiss population. *Journal of Dairy Science*. 2008, 91:767-783.

Balding DJ. A tutorial on statistical methods for population association studies. *Nature Reviews, Genetics*. 2006, 7:781-791.

Banos G, Wiggans GR, Powell RL. Impact of paternity errors in cow identification on genetic evaluations and international comparisons. *Journal of Dairy Science*. 2001, 84:2523-2529.

Bao B, Fang XT, Chen H, Zhang RF, Yan LJ, Zhang HJ. Polymorphisms of STAT5A gene and its association with milk performance traits in Chinese Holstein cattle. *Scientia Agricultura Sinica*. 2008, 41(6):1872-1878.

Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005, 21(2): 263-265.

Bayless KJ, Davis GE, Meininger GA. Isolation and biological properties of osteopontin from bovine milk. *Protein Expression and Purification*. 1997, 9:309-314.

Beaulieu AD, Palmquist DL. Differential effects of high fat diets on fatty acid composition in milk of Jersey and Holstein cows. *Journal of Dairy Science*. 1995, 78(6):1336-44.

Beechinor JG, Kelly EP. Errors of identification amongst cattle presented as progeny of some bulls used in the artificial insemination service in Ireland. *Irish Veterinary Journal*. 1987. 41:348-353.

Bell AW, Bauman DE. Adaptations of glucose metabolism during pregnancy and lactation. *Journal of Mammary Gland Biology and Neoplasia*. 1997, 2:265-278.

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1995, 57:289-300.

Bennewitz J, Reinsch N, Grohs C, Levéziel H, Malafosse A, Thomsen H, Xu N, Looft C, Kühn C, Brockmann GA, Schwerin M, Weimann C, Hiendleder S, Erhardt G, Medjugorac I, Russ I, Förster M, Brenig B, Reinhardt F, Reents R, Averdunk G, Blümel J, Boichard D, Kalm E. Combined analysis of data from two granddaughter designs: A simple strategy for QTL confirmation and increasing experimental power in dairy cattle. *Genetics Selection Evolution*. 2003, 35:319-338.

Bennewitz J, Reinsch N, Guiard V, Fritz S, Thomsen H, Looft C, Kühn C, Schwerin M, Weimann C, Erhardt G, Reinhardt F, Reents R, Boichard D, Kalm E. Multiple Quantitative Trait Loci mapping with cofactors and application of alternative variants of the false discovery rate in an enlarged granddaughter design. *Genetics*. 2004, 168:1019-1027.

Beribe MJ, Carignano H, Roldán D, López-Villalobos N, Raschia MA, Poli M. Comparación de seis modelos de curvas de lactancia en bovinos Holando y Holando x Jersey en Argentina. Trabajo presentado en el IV Encuentro Iberoamericano de Biometría y XVIII Reunión Científica del GAB (Grupo Argentino de Biometría). Mar del Plata, Argentina. Septiembre de 2013.

Bleck GT, Bremel RD. Correlation of the alpha-lactalbumin (+15) polymorphism to milk production and milk composition of Holsteins. *Journal of Dairy Science*. 1993, Vol. 76 No. 8:2292-2298.

Blott S, Kim J-J, Moiso S, Schmidt-Küntzel A, Cornet A, Berzi P, Cambisano N, Ford C, Grisart B, Johnson D, Karim L, Simon P, Snell R, Spelman R, Wong J, Vilkki J, Georges M, Farnir F, Coppieters W. Molecular Dissection of a Quantitative Trait Locus: A Phenylalanine-to-Tyrosine Substitution in the Transmembrane Domain of the Bovine Growth Hormone Receptor Is Associated With a Major Effect on Milk Yield and Composition. *Genetics*. 2003, 163:253:266.

Bohmanova J, Sargolzaei M, Schenkel FS. Characteristics of linkage disequilibrium in North American Holsteins. *BMC Genomics*. 2010, 11:421.

Bole-Feysot Ch, Goffin V, Edery M, Binart N, Kelly PA. Prolactin (PRL) and its receptor: actions, signal transduction pathways and phenotypes observed in PRL receptor knockout mice. *Endocrine Reviews*. 1998, 19:225-268.

Bolormaa S, Pryce JE, Hayes BJ, Goddard ME. Multivariate analysis of a genome-wide association study in dairy cattle. *Journal of Dairy Science*. 2010, 93:3818-3833.

Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*. 1980, 32:314-331.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*. 2007, 23:2633-2635.

Brew K, Hill RL. Lactose biosynthesis. *Reviews of Physiology, Biochemistry and Pharmacology*. 1975; 72:105-58.

Brody S, Turner CW, Ragsdale AC. The relation between the initial rise and subsequent decline of milk secretion following parturition. *The Journal of General Physiology*. 1924, 6:541-545.

Brookes AJ. The essence of SNPs. *Gene*. 1999, 234:177-86.

Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *The American Journal of Human Genetics*. 2009, 84:210-223.

Browning SR, Browning BL. Haplotype phasing: Existing methods and new developments. *Nature Reviews, Genetics*. 2012, 12(10):703-714.

Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics*. 2007, 81:1084-1097.

Browning SR. Missing data imputation and haplotype phase inference for genome-wide association studies. *Human Genetics*. 2008, 124(5):439-450.

Brym P, Kamiński S, Ruś A. New SSCP polymorphism within bovine STAT5A gene and its associations with milk performance traits in Black-and-White and Jersey cattle. *Journal of Applied Genetics*. 2004, 45(4):445-52.

Brym P, Kaminski S, Wójcik E. Nucleotide sequence polymorphism within exon 4 of the bovine prolactin gene and its associations with milk performance traits. *Journal of Applied Genetics*. 2005, Vol. 45 No. 2:179-185.

-
- Buchanan FC, Van Kessel AG, Waldner C, Christensen DA, Laarveld B, Schmutz SM. Hot Topic: An Association Between a Leptin Single Nucleotide Polymorphism and Milk and Protein Yield. *Journal of Dairy Science*. 2003, Vol. 86 No. 10:3164-3166.
- Bush WS, Moore JH. Chapter 11: Genome-Wide Association Studies. *PLoS Computational Biology*. 2012, 8(12):e1002822.
- Buxadé Carbó, Carlos. *Zootecnia, bases de producción animal*. Tomo VII: Producción vacuna de leche y carne. Capítulo V: "La curva de lactación". Editorial: Mundi-Prensa. 1996. Páginas 91-102.
- Cardellino R, Rovira J. *Mejoramiento genético animal*. Editorial Hemisferio Sur. 1987. Cap. 3, pág. 48. Cap. 5, págs. 67-69. Cap. 8, pág. 121. Cap. 17, pág. 249.
- Chan EKF, Hawken R, Reverter A. The combined effect of SNP-marker and phenotype attributes in genome-wide association studies. *Animal Genetics*. 2008, 40:149-156.
- Chen WM, Abecasis GR. Family-based association tests for genome-wide association scans. *The American Journal of Human Genetics*. 2007, 81(5):913-26.
- Christensen LG, Madsen P, Petersen J. The influence of incorrect sire identification on the estimates of genetic parameters and breeding values. *Proceedings of the 2nd World Congress on Genetics Applied to Livestock Production*, Madrid, España. 1982, 200-208.
- Clarke GM, Anderson CA, Pettersson FH, Cardon LR, Morris AP, Zondervan KT. Basic statistical analysis in genetic case-control studies. *Nature Protocols*. 2011, 6(2):121-133.
- Cohen-Zinder M, Seroussi E, Larkin DM, Loo JJ, Everts-van der Wind A, Lee J-H, Drackley JK, Band MR, Hernandez AG, Shani M, Lewin HA, Weller JI, Ron M. Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. *Genome Research*. 2005, 15:936-944.
- Collins FS, Brooks LD, Chakravarti A. A DNA polymorphism discovery resource for research on human genetic variation. *Genome Research*. 1998, 8: 1229–1231.
- Conte G, Mele M, Chessa S, Castiglioni B, Serra A, Pagnacco G, Secchiari P. Diacylglycerol acyltransferase 1, stearoyl-CoA desaturase 1, and sterol regulatory element binding protein 1 gene polymorphisms and milk fatty acid composition in Italian Brown cattle. *Journal of Dairy Science*. 2010, 93(2):753-763.
- Cooper DN, Krawczak M. Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. *Human Genetics*. 1989, 83(2):181-8.
- Cowie AT. *Comparative Physiology of Lactation*. *Proceedings of the Royal Society of Medicine*. 1972, 65:1084-1085.
- Cui Y, Riedlinger G, Miyoshi K, Tang W, Li C, Deng CX, Robinson GW, Hennighausen L. Inactivation of Stat5 in mouse mammary epithelium during pregnancy reveals distinct

functions in cell proliferation, survival, and differentiation. *Molecular and Cellular Biology*. 2004, 24(18):8037-47.

Cupps PT. Breeds of dairy cattle. Introduction to Livestock Production. Ediciones Cole HH, segunda edición, 1966. Página 149.

Daetwyler HD, Schenkel FS, Sargolzaei M, Robinson JAB. A genome scan to detect quantitative trait loci for economically important traits in Holstein cattle using two methods and a dense single nucleotide polymorphism map. *Journal of Dairy Science*. 2008, 91:3225-3236.

De La Vega FM, Lazaruk KD, Rhodes MD, Wenz MH. Assessment of two flexible and compatible SNP genotyping platforms: TaqMan® SNP Genotyping Assays and the SNPLex™ Genotyping System. *Mutation Research*. 2005, 573:111-135.

de Roos APW, Hayes BJ, Spelman RJ, Goddard ME. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics*. 2008, 179:1503-1512.

Dekkers JC, Hospital F. The use of molecular genetics in the improvement of agricultural populations. *Nature Reviews Genetics*. 2002, 3(1):22-32.

Dekkers JCM. Application of Genomic Tools to Animal Breeding. *Current Genomics*. 2012, 13(3):207-212.

Devlin B, Roeder K. Genomic Control for Association Studies. *Biometrics*. 1999, 55:997-1004.

Drackley JK, Beaulieu AD, Elliott JP. Responses of milk fat composition to dietary fat or nonstructural carbohydrates in Holstein and Jersey cows. *Journal of Dairy Science*. 2001, 84(5):1231-7.

Druet T, Fritz S, Boichard D, Colleau JJ. Estimation of genetic parameters for quantitative trait loci for dairy traits in the French Holstein population. *Journal of Dairy Science*. 2006, 89(10):4070-4076.

Druet T, Georges M. A Hidden Markov Model Combining Linkage and Linkage Disequilibrium Information for Haplotype Reconstruction and Quantitative Trait Locus Fine Mapping. *Genetics*. 2010, 184:789-798.

Ekine CC, Rowe SJ, Bishop SC, de Koning D-J. Why breeding values estimated using familial data should not be used for genome-wide association studies. *G3: Genes-Genomes-Genetics*. 2014, 4:341-347.

Emigh TH. A comparison of tests for Hardy-Weinberg equilibrium. *Biometrics*. 1980, 36:627-642.

Enciclopedia Alfabética del Campo. EB Ediciones E.U. 2011, pp. 344-353.

Enciclopedia Bovina. Capítulo 11: Glándula mamaria y secreción láctea. Ramón Gasque Gómez. Facultad de Medicina Veterinaria y Zootecnia. Universidad Nacional Autónoma de México. Primera edición, 2008.

-
- Eu-ahsunthornwattana J, Miller EN, Fakiola M, Wellcome Trust Case Control Consortium, Jeronimo SMB, Blackwell JM, Cordell HJ. Comparison of methods to account for relatedness in Genome-Wide Association Studies with family-based data. *PLoS Genetics*. 2014, 10(7):e1004445.
- Falush D, Stephens M, Pritchard JK. Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. *Genetics*. 2003, 164: 1567-1587.
- Farke C, Meyer HH, Bruckmaier RM, Albrecht C. Differential expression of ABC transporters and their regulatory genes during lactation and dry period in bovine mammary tissue. *Journal of Dairy Science*. 2008, 75(4):406-14.
- Farrell Jr. HM, Jimenez-Flores R, Bleck GT, Brown EM, Butler JE, Creamer LK, Hicks CL, Hollar CM, Ng-Kwai-Hang KF, Swaisgood HE. Nomenclature of the Proteins of Cows' Milk – Sixth Revision. *Journal of Dairy Science*. 2004, Vol. 87. No. 6:1641-1674.
- Ficha Técnica Nº 19: El Uso de Cruzas en la Eficiencia Productiva de los Sistemas Lecheros. INTA Lechero, Sistemas de Producción. Junio 2011.
- Fisher RA. On the interpretation of X^2 from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society*. 1922, 85(1):87-94.
- Freeze BS, Richards TJ. Lactation curve estimation for use in economic optimization models in the dairy industry. *Journal of Dairy Science*. 1992, 75:2984-2989.
- Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, Defelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The structure of haplotype blocks in the human genome. *Science*. 2002, 296: 2225-2229.
- Galvano F, Frigiola A, Gagliardi L, Ciotti S, Bognanno M, Iacopino AM, Nigro F, Tina GL, Cavallaro D, Mussap M, Piva A, Grilli E, Michetti F, Gazzolo D. S100B milk concentration in mammalian species. *Frontiers in Bioscience (Elite edition)*. 2009, 1:542-546.
- García SC, Holmes CW. Effects of time of calving on the productivity of pasture-based dairy systems: A review. *New Zealand Journal of Agricultural Research*. 1999, 42:347-362.
- García SC, Holmes CW. Lactation curves of autumn- and spring-calved cows in pasture-based dairy systems. *Livestock Production Science*. 2001, 68:189-203.
- García-Ruiz A, Ruiz FJ, Van-Tassell CP, Montaldo HH. Genome regions associated to milk production traits and somatic cell score in the Mexican Holstein population. *Proceedings, 10th World Congress on Genetics Applied to Livestock Production*. 2014, Vancouver, Canadá.
- Gazzolo D, Bruschetini M, Lituania M, Serra G, Santini P, Michetti F. Levels of S100B protein are higher in mature human milk than in colostrum and milk-formulae milks. *Clinical Nutrition*. 2004, 23:23-26.
-

Gazzolo D, Monego G, Corvino V, Bruschetti M, Bruschetti P, Zelano G, Michetti F. Human milk contains S100B protein. *Biochimica et Biophysica Acta*. 2003, 1619:209-212.

Geldermann H, Pieper U y Roth B. Effects of marked chromosome sections on milk performance in cattle. *Theoretical and Applied Genetics*. 1985, 70:138-146.

Geldermann H, Pieper U, Weber WE. Effect of misidentification on the estimation of breeding value and heritability in cattle. *Journal of Animal Science*. 1986. 63:1759-1768.

Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R, Pasquino AT, Sargeant LS, Sorensen A, Steele MR, Zhao X, Womack JE, Hoeschele I. Mapping Quantitative Trait Loci Controlling Milk Production in Dairy Cattle by Exploiting Progeny Testing. *Genetics*. 1995, 139:907-920.

Georges M. Capítulo 15: Impact of high-throughput genotyping and sequencing on the identification of genes and variants underlying phenotypic variation in domestic cattle. En *Bovine Genomics*. Editor: James E. Womack. Wiley-Blackwell, Oxford, UK. 2012.

Giblin L, Butler ST, Kearney BM, Waters SM, Callanan MJ, Berry DP. Association of bovine leptin polymorphisms with energy output and energy storage traits in progeny tested Holstein-Friesian dairy cattle sires. *BioMed Central (BMC) Genetics*. 2010, 11:73.

Gilmour AR, Gogel BJ, Cullis BR, Thompson R. ASReml User Guide Release 3.0. VSN International Ltd, Hemel Hempstead, UK. 2009. www.vsnl.co.uk.

Goddard ME, Hayes BJ. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics*. 2009, 10:381-391.

Grisart B, Coppieters W, Farnir F, Karim L, Ford C, Berzi P, Cambisano N, Mni M, Reid S, Simon P, Spelman R, Georges M, Snell R. Positional Candidate Cloning of a QTL in Dairy Cattle: Identification of a Missense Mutation in the Bovine DGAT1 Gene with Major Effect on Milk Yield and Composition. *Genome Research*. 2002, 12:222-231.

Guillaume F, Fritz S, Boichard D, Druet T. Estimation by simulation of the efficiency of the French marker-assisted selection program in dairy cattle (*Open Access publication*). *Genetics, Selection, Evolution*. 2008, 40:91-102.

Guo J, Jorjani H, Carlborg Ö. A genome-wide association study using international breeding-evaluation data identifies major loci affecting production traits and stature in the Brown Swiss cattle breed. *BioMed Central (BMC) Genetics*. 2012, 13:82.

Hadfield JD. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*. 2010, 33(2):1-22.

Harris BL, Johnson DL, Spelman RJ. Genomic selection in New Zealand and the implications for national genetic evaluation. *Proceedings of the Interbull Meeting, Niagara Falls, USA*. 2008.

Harvatine KJ, Boisclair YR, Bauman DE. Recent advances in the regulation of milk fat synthesis. *Animal*. 2009, 3(1):40-54.

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science*. 2009, 92:433-443.

He F, Sun D, Yu Y, Wang Y, Zhang Y. SNPs detection of STAT5A gene and association with milk production traits in Holstein cattle. *Acta Veterinaria et Zootechnica Sinica*. 2007, 38(4):326-331.

Hennighausen L, Robinson GW. Interpretation of cytokine signaling through the transcription factors STAT5A and STAT5B. *Genes & Development*. 2008, 22:711-721.

Holliday R, Grigg GW. DNA methylation and mutation. *Mutation Research*. 1993, 285:61-67.

Horseman ND, Zhao W, Montecino-Rodriguez E, Tanaka M, Nakashima K, Engle SJ, Smith F, Markoff E, Dorshkind K. Defective mammopoiesis, but normal hematopoiesis, in mice with a targeted disruption of the prolactin gene. *The EMBO Journal*. 1997, 16(23):6926-35.

Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of Genome-Wide Association Studies. *PLoS Genetics*. 2009, 5(6):e1000529.

Hu ZL, Fritz ER, Reecy JM. AnimalQTLdb: a livestock QTL database tool set for positional QTL information mining and beyond. *Nucleic Acids Research*. 2007, Vol. 35, Database issue, D604-D609.

Hu Z-L, Park CA, Wu X-L, Reecy JM. Animal QTLdb: an improved database tool for livestock animal/QTL association data dissemination in the post-genome era. *Nucleic Acids Research, Database issue*. 2013, 41:D871-D879.

International agreement of recording practices. Guidelines approved by the general assembly held in Cork, Ireland on June 2012. International Committee for Animal Recording - ICAR.

Jacobs S, Schilf C, Fliegert F, Kolling S, Weber Y, Schürmann A, Joost HG. ADP-ribosylation factor (ARF)-like 4, 6, and 7 represent a subgroup of the ARF family characterization by rapid nucleotide exchange and a nuclear localization signal. *Federation of European Biochemical Societies (FEBS) Letters*. 1999, 456:384-388.

Jeffreys AJ, Wilson V, Thein SL. Hypervariable "minisatellite" regions in human DNA. *Nature*. 1985, 314:67-73.

Jensen R G. The composition of bovine milk lipids: January 1995 to December 2000. *Journal of Dairy Science*. 2002, 85:295-350.

Jiang L, Liu J, Sun D, Ma P, Ding X, Yu Y, Zhang Q. Genome Wide Association Studies for Milk Production Traits in Chinese Holstein Population. *PLOS One*. 2010, Vol. 5, Issue 10, e13661.

Jonker JW, Merino G, Musters S, van Herwaarden AE, Bolscher E, Wagenaar E, Mesman E, Dale TC, Schinkel AH. The breast cancer resistance protein BCRP (ABCG2) concentrates drugs and carcinogenic xenotoxins into milk. *Nature Medicine*. 2005, Vol. 11 No. 2:127-129.

Kamiński S, Ahman A, Ruść A, Wójcik E, Malewski T. MilkProtChip – a microarray of SNPs in candidate genes associated with milk protein biosynthesis – development and validation. *Journal of Applied Genetics*. 2005, 46(1):45-58.

Kamiński S, Malewski T, Ahman A, Wójcik E, Ruść A, Oleński K, Jakubczak A, Sazanov AA. Towards an Integrated Approach to Study SNPs and Expression of Candidate Genes Associated with Milk Protein Biosynthesis. *Russian Journal of Genetics*. 2008, 44(4):459-465.

Keown JF, Everett RW, Empet NB, Wadell LH. Lactation curves. *Journal of Dairy Science*. 1986, 69:769-781.

Khatib H, Heifetz E, Dekkers JC. Association of the protease inhibitor gene with production traits in Holstein dairy cattle. *Journal of Dairy Science*. 2005, 88:1208-1213.

Khatib H, Leonard SD, Schutzkus V, Luo W, Chang YM. Association of the OLR1 gene with milk composition in Holstein dairy cattle. *Journal of Dairy Science*. 2006, Vol. 89 No. 5:1753-1760.

Khatib H, Schutzkus V, Chang YM, Rosa JM. Pattern of Expression of the Uterine Milk Protein Gene and its Association with Productive Life in Dairy Cattle. *Journal of Dairy Science*. 2007, 90:2427-2433.

Khatkar MS, Thomson PC, Tammen I, Raadsma HW. Quantitative trait loci mapping in dairy cattle: Review and Meta-analysis. *Genetics, Selection, Evolution*. 2004, 36:163-190.

Khatkar MS, Zenger KR, Hobbs M, Hawken RJ; Cavanagh JAL, Barris W, McClintock AE, McClintock S, Thomson PC, Tier B, Nicholas FW, Raadsma HW. A Primary Assembly of a Bovine Haplotype Block Map Based on a 15,036-Single-Nucleotide Polymorphism Panel Genotyped in Holstein-Friesian Cattle. *Genetics*. 2007, 176:763-772.

Kolbehdari D, Wang Z, Grant JR, Murdoch B, Prasad A, Xiu Z, Marques E, Stothard P, Moore SS. A whole genome scan to map QTL for milk production traits and somatic cell score in Canadian Holstein bulls. *Journal of Animal Breeding and Genetics*. 2009, 126:216-227.

Komisarek J, Dorynek Z. Effect of ABCG2, PPARGC1A, OLR1 and SCD1 gene polymorphism on estimated breeding values for functional and production traits in Polish Holstein-Friesian bulls. *Journal of Applied Genetics*. 2009, Vol. 50 No. 2:125-132.

Koreth J, O'Leraly J, McGee J. Microsatellites and PCR Genomic Analysis. *Journal of Pathology*. 1996, 178:239-248.

Landegren U, Nilsson M, Kwok P-Y. Reading Bits of Genetic Information: Methods for Single-Nucleotide Polymorphism Analysis. *Genome Research*. 1998, 8:769-776.

Laurie CC, Doheny KF, Mirel DB, Pugh EW, Bierut LJ, Bhangale T, Boehm F, Caporaso NE, Cornelis MC, Edenberg HJ, Gabriel SB, Harris EL, Hu FB, Jacobs K, Kraft P, Landi MT, Lumley T, Manolio TA, McHugh C, Painter I, Paschall J, Rice JP, Rice KM, Zheng X, Weir BS. Quality control and quality assurance in genotypic data for genome-wide association studies. *Genetic Epidemiology*. 2010, 34(6): 591–602.

Leonard S, Khatib H, Schutzkus V, Chang YM, Maltecca C. Effects of the Osteopontin Gene Variants on Milk Production Traits in Dairy Cattle. *Journal of Dairy Science*. 2005, Vol. 88 No. 11:4083-4086.

Levinson G, Gutman GA. Slipped-strand mispairing: A major mechanism for DNA sequence evolution. *Molecular Biology and Evolution*. 1987, 4:203-221.

Li R, Xia W, Zhang Z, Wu K. S100B protein, brain-derived neurotrophic factor and glial cell line-derived neurotrophic factor in human milk. *PLoS ONE*. 2011, 6(6):e21663.

Li Y, Willer CJ, Ding J, Scaet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic Epidemiology*. 2010, 34:816-834.

Li Z, Zhai M, Wang H, Chen L, Wang L, Ru C, Song A, Liu X. Identification of splice variants, expression analysis and single nucleotide polymorphisms of the PRMT2 gene in dairy cattle. *Gene*. 2014, 539:37-43.

Liao Y, Du X, Lönnerdal B. miR-214 Regulates Lactoferrin Expression and Pro-Apoptotic Function in Mammary Epithelial Cells. *The Journal of Nutrition*. 2010, 140:1552-1556.

Liefers SC, Veerkamp RF, te Pas MFW, Delavaud C, Chilliard Y, Platje M, van der Lende T. Leptin promoter mutations affect leptin levels and performance traits in dairy cows. *Animal Genetics*. 2005, 36:111-118.

Lien S, Kantanen J, Olsaker I, Holm L-E, Eythorsdottir E, Sandberg K, Dalsgard B, Adalsteinsson S. Comparison of milk protein allele frequencies in Nordic cattle breeds. *Animal Genetics*. 1999, 30:85-91.

Liu X, Robinson GW, Wagner KU, Garrett L, Wynshaw-Boris A, Hennighausen L. Stat5a is mandatory for adult mammary gland development and lactogenesis. *Genes and Development*. 1997, 11:179-186.

Lönnerdal B, Iyer S. Lactoferrin: molecular structure and biological function. *Annual Review of Nutrition*. 1995, 15:93-110.

Macciotta NPP, Vicario D, Cappio-Borlino A. Detection of Different Shapes of Lactation Curve for Milk Yield in Dairy Cattle by Empirical Mathematical Models. *Journal of Dairy Science*. 2005, 88(3):1178-1191.

Mai MD, Sahana G, Christiansen FB, Guldbbrandtsen B. A genome-wide association study for milk production traits in Danish Jersey cattle using a 50K single nucleotide polymorphism chip. *Journal of Animal Science*. 2010, 88:3522-3528.

- Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nature Reviews, Genetics*. 2010, 11:499-511.
- Marques E, Schnabel RD, Stothard P, Kolbehdari D, Wang Z, Taylor JF, Moore SS. High density linkage disequilibrium maps of chromosome 14 in Holstein and Angus cattle. *BMC Genetics*. 2008, 9:45.
- Martin P, Szymanowska M, Zwierzchowski L, Leroux C. The impact of genetic polymorphisms on the protein composition of ruminant milks. *Reproduction Nutrition Development*. 2002, 42:433-459.
- Mather IH, Jack LJW. A review of the molecular and cellular biology of butyrophilin, the major protein of bovine milk fat globule membrane. *Journal of Dairy Science*. 1993, 76:3832-3850.
- Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, O'Connell J, Moore SS, Smith TPL, Sonstegard TS, Van Tassell CP. Development and Characterization of a High Density SNP Genotyping Assay for Cattle. *PLoS ONE*. 2009, 4(4):e5350.
- McKay SD, Schnabel RD, Murdoch BM, Matukumalli LK, Aerts J, Coppieters W, Crews D, Dias Neto E, Gill CA, Gao C, Mannen H, Stothard P, Wang Z, Van Tassell CP, Williams JL, Taylor JF, Moore SS. Whole genome linkage disequilibrium maps in cattle. *BMC Genetics*. 2007, 8:74.
- Mehta JL, Li D. Identification, regulation and function of a novel lectin-like oxidized low-density lipoprotein receptor. *Journal of the American College of Cardiology*. 2002, 39(9):1429-1435.
- Mele M, Conte G, Castiglioni B, Chessa S, Macciotta NPP, Serra A, Buccioni A, Pagnacco G, Secchiari P. Stearoyl-CoA desaturase gene polymorphism and milk fatty acid composition in Italian Holsteins. *Journal of Dairy Science*. 2007, 90:4458-4465.
- Meyer K. WOMBAT – A tool for mixed model analyses in quantitative genetics by REML. *Journal of Zhejiang University SCIENCE B*. 2007. 8(11):815-821.
- Minozzi G, Nicolazzi EL, Stella A, Biffani S, Negrini R, Lazzari B, Ajmone-Marsan P, Williams, JL. Genome Wide Analysis of Fertility and Production Traits in Italian Holstein Cattle. *PLoS ONE*. 2013, 8(11):e80219.
- Minozzi G, Nicolazzi EL, Strozzi F, Stella A, Negrini R, Ajmone-Marsan P, Williams, JL. Genome wide scan for somatic cell counts in Holstein bulls. *BMC Proceedings*. 2011, 5(Suppl 4):S17.
- Miyoshi K, Shillingford JM, Smith GH, Grimm SL, Wagner KU, Oka T, Rosen JM, Robinson GW, Hennighausen L. Signal transducer and activation of transcription (Stat) 5 controls the proliferation and differentiation of mammary alveolar epithelium. *The Journal of Cell Biology*. 2001, Vol. 155 No. 4:531-542.
- Moffatt J, Bazer FW, Hansen PJ, Chun PW, Roberts RM. Purification, secretion and immunocytochemical localization of the uterine milk proteins, major progesterone-induced proteins in uterine secretions of the sheep. *Biology of Reproduction*. 1987, 36:419-430.

Moioli B, Contarini G, Avalli A, Catillo G, Orrú L, De Matteis G, Masoero G, Napolitano F. Short communication: Effect of Stearoyl-Coenzyme A Desaturase Polymorphism on Fatty Acid Composition of Milk. *Journal of Dairy Science*. 2007, Vol. 90 No. 7:3553-3558.

Morris CA, Cullen NG, Glass BC, Hyndman DL, Manley TR, Hickey SM, McEwan JC, Pitchford WS, Bottema CDK, Lee MAH. Fatty acid synthase effects on bovine adipose fat and milk fat. *Mammalian Genome*. 2007, 18:64-74.

Mullen MP, Lynch CO, Waters SM, Howard DJ, O'Boyle P, Kenny DA, Buckley F, Horan B, Diskin MG. Single nucleotide polymorphisms in the growth hormone and insulin-like growth factor-1 genes are associated with milk production, body condition score and fertility traits in dairy cows. *Genetics and Molecular Research*. 2011, 10 (3):1819-1830.

Nadesalingam J, Plante Y, Gibson JP. Detection of QTL for milk production on Chromosomes 1 and 6 of Holstein cattle. *Mammalian Genome*. 2001, 12:27-31.

Nani JP, Raschia MA, Poli MA, Calvinho LF, Amadio AF. Genome-wide association study for somatic cell score in Argentinean dairy cattle. *Livestock Science*. 2015, 175:1-9 (<http://dx.doi.org/10.1016/j.livsci.2015.02.007>).

Nauli SM, Alenghat FJ, Luo Y, Williams E, Vassilev P, Li X, Elia AEH, Lu W, Brown EM, Quinn SJ, Ingber DE, Zhou J. Polycystins 1 and 2 mediate mechanosensation in the primary cilium of kidney cells. *Nature Genetics*. 2003, 33(2):129-137.

Nicolazzi EL, Picciolini M, Strozzi F, Schnabel RD, Lawley C, Pirani A, Brew F, Stella A. SNPchiMp: A database to disentangle the SNPchip jungle in bovine livestock. *BMC Genomics*. 2014, 15:123.

O'Donnell-Megaró AM, Barbano DM, Bauman DE. Survey of the fatty acid composition of retail milk in the United States including regional and seasonal variations. *Journal of Dairy Science*. 2011, 94(1):59-65.

Ogg SL, Weldon AK, Dobbie L, Smith AJH, Mather IH. Expression of butyrophilin (Btn1a1) in lactating mammary gland is essential for the regulated secretion of milk-lipid droplets. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*. 2004, 101(27):10084-10089.

Ogorevc J, Kunej T, Razpet A, Dovc P. Database of cattle candidate genes and genetic markers for milk production and mastitis. *Animal Genetics*. 2009, 40:832-851.

Oka K, Sawamura T, Kikuta K, Itokawa s, Kume N, Kita T, Masaki T. Lectin-like oxidized low-density lipoprotein receptor 1 mediates phagocytosis of aged/apoptotic cells in endothelial cells. *Proceedings of the National Academy of Sciences of USA*. 1998, 95:9535-9540.

Olori VE, Brotherstone S, Hill WG, McGuirk BJ. Fit of standard models of the lactation curve to weekly records of milk production of cows in a single herd. *Livestock Production Science*. 1999, 58:55-63.

Olsen HG, Nilsen H, Hayes B, Berg PR, Svendsen M, Lien S, Meuwissen T. Genetic support for a quantitative trait nucleotide in the ABCG2 gene affecting milk composition of dairy cattle. *BioMed Central (BMC) Genetics*. 2007, 8:32.

Ormandy CJ, Camus A, Barra J, Damotte D, Lucas B, Buteau H, Edery M, Brousse N, Babinet C, Binart N, Kelly PA. Null mutation of the prolactin receptor gene produces multiple reproductive defects in the mouse. *Genes and Development*. 1997, 11:167-178.

Pearson TA, Manolio TA. How to Interpret a Genome-wide Association Study. *The Journal of the American Medical Association (JAMA)*. 2008, 299(11):1335-1344.

Pérez-Enciso M, Rincón JC, Legarra A. Sequence- vs. chip-assisted genomic selection: accurate biological information is advised. *Genetics Selection Evolution*. 2015, 47:43.

Pettersson FH, Anderson CA, Clarke GM, Barrett JC, Cardon LR, Morris AP, Zondervan KT. Marker selection for genetic case-control association studies. *Nature Protocols*. 2009, 4(5):743-752.

Pimentel ECG, Bauersachs S, Tietze M, Simianer H, Tetens J, Thaller G, Reinhardt F, Wolf E, König S. Exploration of relationships between production and fertility traits in dairy cattle via association studies of SNPs within candidate genes derived by expression profiling. *Animal Genetics*. 2010, 42:251-262.

Pollott GE. Deconstructing milk yield and composition during lactation using biologically based lactation models. *Journal of Dairy Science*. 2004, 87:2375-2387.

Pritchard JK, Stephens M, Donnelly P. Inference of Population Structure Using Multilocus Genotype Data. *Genetics*. 2000, 155: 945-959.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*. 2007, 81:559-575.

Qu C, Schuetz JM, Min JE; Leach S, Daley D, Spinelli JJ, Brooks-Wilson A, Graham J. Cost-effective prediction of gender-labeling errors and estimation of gender-labeling error rates in candidate-gene association studies. *Frontiers in Genetics*. 2011, 2(31):1-10.

Quintero JC, Serna JI, Hurtado NA, Noguera RR, Cerón-Muñoz MF. Modelos matemáticos para curvas de lactancia en ganado lechero. *Revista Colombiana de Ciencias Pecuarias*. 2007, 20:149-156.

Raven LA, Cocks BG, Hayes BJ. Multibreed genome wide association can improve precisión of mapping causative variants underlying milk production in dairy cattle. *BMC Genomics*. 2014, 15:62.

Reinhardt TA, Lippolis JD. Developmental changes in the milk fat globule membrane proteome during the transition from colostrums to milk. *Journal of Dairy Science*. 2008, 91:2307-2318.

-
- Rincon G, Islas-Trejo A, Casellas J, Ronin Y, Soller M, Lipkin E, Medrano JF. Fine mapping and association analysis of a quantitative trait locus for milk production traits on *Bos taurus* autosome 4. *Journal of Dairy Science*. 2009, Vol. 92 No. 2:758-764.
- Rincon G, Islas-Trejo A, Castillo AR, Bauman DE, German BJ, Medrano JF. Polymorphisms in genes in the SREBP1 signalling pathway and SCD are associated with milk fatty acid composition in Holstein cattle. *The Journal of Dairy Research*. 2012, 79(1):66-75.
- Rocha JL, Baker JF, Womack JE, Sanders JO, Taylor JF. Statistical associations between restriction fragment length polymorphisms and quantitative traits in beef cattle. *Journal of Animal Science*. 1992, 70:3360-3370.
- Ron M, Band M, Wyler A, Weller JI. Unequivocal determination of sire allele origin for multiallelic microsatellites when only sire and progeny are genotyped. *Animal Genetics*. 1993. 24(3):171-176.
- Ron M, Blanc Y, Band M, Ezra E, Weller JI. Misidentification rate in the Israeli dairy cattle population and its implications for genetic improvement. *Journal of Dairy Science*. 1996, 79:676-681.
- Roy R, Ordovas L, Zaragoza P, Romero A, Moreno C, Altarriba J, Rodellar C. Association of polymorphisms in the bovine FASN gene with milk-fat content. *Animal Genetics*. 2006, 37:215-218.
- Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. Artemis: sequence visualization and annotation. *Bioinformatics*. 2000, 16:944-945.
- Sadler AM, Kiddy CA, McCann RE, Mattingly WA. Acid production and curd toughness in milks of different α 1-casein types. *Journal of Dairy Science*. 1968, 51:28-35.
- Sargolzaei M, Schenkel FS, Jansen GB, Schaeffer LR. Extent of linkage disequilibrium in Holstein cattle in North America. *Journal of Dairy Science*. 2008, 91:2106-2117.
- Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *The American Journal of Human Genetics*. 2006, 78:629-644.
- Schennink A, Bovenhuis H, Léon-kloosterziel KM, van Arendonk AM, Visker MHPW. Effect of polymorphisms in the FASN, OLR1, PPARGC1A, PRL and STAT5A genes on bovine milk-fat composition. *Animal Genetics*. 2009, 40:909-916.
- Schlötterer C, Tautz D. Slippage synthesis of simple sequence DNA. *Nucleic Acids Research*. 1992, 20:211-215.
- Schlötterer C. The evolution of molecular markers – just a matter of fashion? *Nature Reviews Genetics*. 2004, 5:63-69.
-

-
- Schnabel RD, Kim J-J, Ashwell MS, Sonstegard TS, Van Tassell CP, Connor EE, Taylor JF. Fine-mapping milk production quantitative trait loci on BTA 6: Analysis of the bovine osteopontin gene. *Proceedings of the National Academy of Sciences (PNAS)*. 2005, Vol. 102 No. 19:6896-6901.
- Schopen GCB, Visker MHPW, Koks PD, Mullaart E, van Arendonk JAM, Bovenhuis H. Whole-genome association study for milk protein composition in dairy cattle. *Journal of Dairy Science*. 2011, 94(6):3148-3158.
- Schrooten C. Tesis Doctoral: Genomic Variation in Dairy Cattle – Identification and Use. 2004. Animal Breeding and Genetics Group, Wageningen University, Wageningen and Holland Genetics BV, CR Delta VRV Holding BV, Arnhem. ISBN: 90-8504-009-4.
- Schutz MM, Hansen LB, Steuernagel GR. Variation of milk, fat, protein and somatic cells for dairy cattle. *Journal of Dairy Science*. 1990, 73:484-493.
- Schweizer M, Roder K, Zhang L, Wolf SS. Transcription factors acting on the promoter of the rat fatty acid synthase gene. *Biochemical Society Transactions*. 2002, 30:1070-1072.
- Sherchand L, McNew RW, Kellogg DW, Johnson B. Selection of a mathematical model to generate lactation curves using daily milk yields of Holstein cows. *Journal of Dairy Science*. 1995, 78:2507-2513.
- Silvestre AM, Petim-Batista F, Colaço J. The accuracy of seven mathematical functions in modeling dairy cattle lactation curves based on test-day records from varying sample schemes. *Journal of Dairy Science*. 2006, 89:1813-1821.
- Smith IA, Knezevic BR, Ammann JU, Rhodes DA, Aw D, Palmer DB, Mather IH, Trowsdale J. *BTN1A1*, the mammary gland butyrophilin, and *BTN2A2* are both inhibitors of T cell activation. *The Journal of Immunology*. 2010, 184:3514-3525.
- Smith SJ, Cases S, Jensen DR, Chen HC, Sande E, Tow B, Sanan DA, Raber J, Eckel RH, Farese RVJr. Obesity resistance and multiple mechanisms of triglyceride synthesis in mice lacking *DGAT1*. *Nature Genetics*. 2000, 25:87-90.
- Soller M, Beckmann JS. Restriction fragment length polymorphisms and genetic improvement. *Proc. 2nd World Congr. Genet. Appl. Livest. Prod., Madrid, Spain VI*: 396-404. 1982.
- Soper DS (2015). Fisher's Exact Test Calculator for a 2x2 Contingency Table [Software]. Disponible en <http://www.danielsoper.com/statcalc>.
- Stacey A, Schnieke A, Kerr M, Scott A, McKee C, Cottingham I, Binas B, Wilde C, Colman A. Lactation is disrupted by α -lactalbumin deficiency and can be restored by human α -lactalbumin gene replacement in mice. *Proceedings of the National Academy of Sciences, Agricultural Sciences*. 1995, 92:2835-2839.
- Staden R, Beal KF, Bonfield JK. The Staden Package. *Methods in Molecular Biology*. 1998, 132:115-130.
-

- Stanton TL, Jones LR, Everett RW, Kachman SD. Estimating milk, fat, and protein lactation curves with a test day model. *Journal of Dairy Science*. 1992, 75:1691-1700.
- Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*. 2001, 68:978-989.
- Stinnakre MG, Vilotte JL, Soulier S, Mercier JC. Creation and phenotypic analysis of α -lactalbumin-deficient mice. *Proceedings of the National Academy of Sciences, Agricultural Sciences*. 1994, 91: 6544-6548.
- Strucken EM, Bortfeldt RH, de Koning DJ, Brockmann GA. Genome-wide associations for investigating time-dependent genetic effects for milk production traits in dairy cattle. *Animal Genetics*. 2011, 43:375-382.
- Suchocki T, Komisarek J, Szyda J. Testing candidate gene effects on milk production traits in dairy cattle under various parametrizations and modes of inheritance. *Journal of Dairy Science*. 2010, 93:2703-2717.
- Svennersten-Sjaunja K, Olsson K. Endocrinology of milk production. *Domestic Animal Endocrinology*. 2005, 29:241-258.
- Syvänen A-C. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature Reviews, Genetics*. 2001, Vol. 2:930-942.
- Tabor HK, Risch NJ, Myers RM. Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nature Reviews, Genetics*. 2002, Vol. 3:391-397.
- Tautz D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research*. 1989, 17:6463-6471.
- Taverna MA, Charlón V, Cuatrín AL, Gaggiotti MC, Páez RB, Chavez MS. Composición química de la leche producida en la cuenca lechera central de la Argentina y Composición mineral de la leche producida en la cuenca lechera central de la Argentina. *Revista Argentina de Producción Animal*. 2001, Vol. 21, Supl. 1.
- Taverna MA. Composición química de la leche. Proyecto Nacional de Lechería, INTA.
- Teo YY, Fry AE, Clark TG, Tai ES, Seielstad M. On the usage of HWE for identifying genotyping errors. *Annals of Human Genetics*. 2007, 71:701-703.
- Teo YY. Exploratory data analysis in large-scale genetic studies. *Biostatistics*. 2010, 11(1):70-81.
- The Bovine Genome Sequencing and Analysis Consortium, Elsik CG, Tellam RL, Worley KC. The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution. *Science*. 2009, 324:522-528.
- The Bovine HapMap Consortium. Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. *Science*. 2009, 324(5926):528-532.

Thompson & Thompson. Capítulo 9: Variación genética en los individuos y las poblaciones: mutación y polimorfismo. En *Genética en medicina*. Editorial Elsevier. Séptima edición, 2008.

Thompson EA, Shaw RG. Pedigree analysis for quantitative traits: variance components without matrix inversion. *Biometrics*. 1990, 46:399-413.

Tobler AR, Short S, Andersen MR, Paner TM, Briggs JC, Lambert SM, Wu PP, Wang Y, Spoonde AY, Koehler RT, Peyret N, Chen C, Broomer AJ, Ridzon DA, Zhou H, Hoo BS, Hayashibara KC, Leong LN, Ma CN, Rosenblum BB, Day JP, Ziegle JS, De La Vega FM, Rhodes MD, Hennessy KM, Wenz M. The SNPlex Genotyping System: A Flexible and Scalable Platform for SNP Genotyping. *Journal of Biomolecular Techniques*. 2005, Vol. 16 Issue 4:396-404.

Tsepilov YA, Ried JS, Strauch K, Grallert H, van Duijn CM, Axenovich TI, Aulchenko YS. Development and Application of Genomic Control Methods for Genome-Wide Association Studies Using Non-Additive Models. *PLoS ONE*. 2013, 8(12): e81431.

Udy GB, Towers RP, Snell RG, Wilkins RJ, Park SH, Ram PA, Waxman DJ, Davey HW. Requirement of STAT5b for sexual dimorphism of body growth rates and liver gene expression. *Proceedings of the National Academy of Sciences (PNAS)*. 1997, 94(14):7239-7244.

van Herwaarden AE, Wagenaar E, Merino G, Jonker JW, Rosing H, Beijnen JH, Schinkel AH. Multidrug transporter ABCG2/breast cancer resistance protein secretes riboflavin (vitamin B2) into milk. *Molecular and Cellular Biology*. 2007, 27(4):1247-1253.

VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, Schenkel F. Invited review: Reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science*. 2009, 92:16-24.

Vignal A, Milan D, SanCristobal M, Eggen A. A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*. 2002, 34:275-305.

Viitala S, Szyda J, Blott S, Schulman N, Lidauer M, Mäki-Tanila A, Georges M, Vilkki J. The Role of the Bovine Growth Hormone Receptor and Prolactin Receptor Genes in Milk, Fat and Protein Production in Finnish Ayrshire Dairy Cattle. *Genetics*. 2006, 173:2151-2164.

Visscher PM, Woolliams JA, Smith D, Williams JL. Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection. *Journal of Dairy Science*. 2002, 85:2368-2375.

Vreeman HJ, Visser S, Slangen CJ, Van Riel JAM. Characterization of bovine κ -casein fractions and the kinetics of chymosin-induced macropeptide release from carbohydrate-free and carbohydrate-containing fractions determined by high-performance gel-permeation chromatography. *Biochemical Journal*. 1986, 240:87-97.

Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lander ES. Large-scale

identification, mapping and genotyping of single nucleotide polymorphisms in the human genome. *Science*. 1998, 280:1077-1082.

Wang X, Wurmser C, Pausch H, Jung S, Reinhardt F, Tetens J, Thaller G, Fries R. Identification and Dissection of Four Major QTL Affecting Milk Fat Content in the German Holstein-Friesian Population. *PLoS ONE*. 2012, 7(7): e40711.

Wasike CB, Khi AK, Peters KJ. Modelling of lactation curves of dairy cows based on monthly test day milk yield records under inconsistent milk recording scenarios. *Animal*. 2011, 5(11):1780-1790.

Waters SM, McCabe MS, Howard DJ, Giblin L, Magee DA, MacHugh DE, Berry DP. Associations between newly discovered polymorphisms in the *Bos taurus* growth hormone receptor gene and performance traits in Holstein-Friesian dairy cattle. *Animal Genetics*. 2011, 42(1):39-49.

Weikard R, Kühn C, Goldammer T, Freyer G, Schwerin M. The bovine PPARGC1A gene: molecular characterization and association of an SNP with variation of milk fat synthesis. *Physiological Genomics*. 2005, 21:1-13.

Weller JI, Feldmesser E, Golik M, Tager-Cohen I, Domochoovsky R, Alus O, Ezra E, Ron M. Factors affecting incorrect paternity assignment in the Israeli Holstein population. *Journal of Dairy Science*. 2004, 87:2627-2640.

Weller JI. Quantitative trait loci analysis in animals. CAB International. Cromwell Press. 2001. Cap.I: "Historical Overview", págs. 6-9. Cap.III:"Random and Fixed Effects, the Mixed Model", pág. 30.

White IMS, Thompson R, Brotherstone S. Genetic and environmental smoothing of lactation curves with cubic splines. *Journal of Dairy Science*. 1999, 82:632-638.

Wiggans GR, Sonstegard TS, VanRaden PM, Matukumalli LK, Schnabel RD, Taylor JF, Schenkel FS, Van Tassell CP. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. *Journal of Dairy Science*. 2009, 92:3431-3436.

Wigginton JE, Cutler DJ, Abecasis GR. A Note on Exact Tests of Hardy-Weinberg Equilibrium. *The American Journal of Human Genetics*. 2005, 76:887-893.

Wilmink JBM. Adjustment of test day milk, fat and protein yield for age, season and stage of lactation. *Livestock Production Science*. 1987, 16:335-348.

Wilson AJ, Réale D, Clements MN, Morrissey MM, Postma E, Walling CA, Kruuk LEB, Nussey DH. An ecologist's guide to the animal model. *Journal of Animal Ecology*. 2010. 79:13-26.

Wood PDP. Algebraic model of the lactation curve in cattle. *Nature*. 1967, 216:164-165.

Zhong J, Cao RX, Zu XY, Hong T, Yang J, Liu L, Xiao XH, Ding WJ, Zhao Q, Liu JH, Wen GB. Identification and characterization of novel spliced variants of PRMT2 in breast carcinoma. The FEBS (Federation of European Biochemical Societies) Journal. 2012, 279:316-335.

Zucht HD, Raida M, Adermann K, Mägert HJ, Forssmann WG. Casocidin-I: a casein- α_{s2} derived peptide exhibits antibacterial activity. Federation of European Biochemical Societies (FEBS) Letters. 1995, 372:185-188.