



UNIVERSIDAD NACIONAL DE LA PLATA
FACULTAD DE CIENCIAS VETERINARIAS



**ANÁLISIS DE LA VARIABILIDAD DE GENES IMPLICADOS
EN LA RESPUESTA INMUNE EN POBLACIONES EQUINAS
DE ARGENTINA MEDIANTE DETECCIÓN DE
POLIMORFISMOS DE NUCLEÓTIDO SIMPLE.**

Trabajo de Tesis realizado como requisito para optar al título de

DOCTOR EN CIENCIAS VETERINARIAS

Autor: Lic. Claudia Corbi Botto

DIRECTOR: Dra. Silvina Díaz

CODIRECTOR: Dra. Pilar Peral García

LUGAR DE TRABAJO: Instituto de Genética Veterinaria «Ing. Fernando Noel Dulout» IGEVET, Facultad de Ciencias Veterinarias, Universidad Nacional de La Plata

MIEMBROS DEL JURADO:

Dr. Carlos Alberto Fossati

Dra. Mariana E. Kienast

Dr. Marcelo Miragaya



La Plata, Argentina, 2020

*A mi madre, que se lo merece,
a mi esposo Manuel por acompañarme siempre
y a nuestro hijo Alfonso que llegó para
hacernos más felices que nunca.*

Reconocimiento

A la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC-PBA) y al Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) por sus programas de becas que permitieron que continúe ampliando mi formación académica.

A la Universidad Nacional de La Plata y especialmente a la Facultad de Ciencias Veterinarias, por recibirme en su Casa de Estudios donde pude llevar adelante el Doctorado.

A los miembros del Jurado de esta tesis, quienes desinteresadamente aceptaron ser parte de este proceso de formación, invirtiendo su tiempo en el seguimiento y evaluación de cada informe, así como en la revisión y corrección de este manuscrito.

Un reconocimiento a todos los propietarios y establecimientos que desinteresadamente facilitaron muestras para poder realizar este trabajo de tesis. Especialmente al Instituto Nacional de Tecnología Agropecuaria de Leales, Tucumán.

Finalmente, a la hermosa ciudad de La Plata, que supo ser mi hogar durante todos estos años. Quedarán en mi corazón tus diagonales, el olor a tilo al caer la tarde en verano y el verde eterno de Gonnet...

Agradecimientos

Sin lugar a dudas hacer un Doctorado es algo muy intenso y, a mi entender, es imposible sacarlo adelante en solitario. En ese sentido, quiero agradecer a quienes de una u otra manera fueron parte de esta etapa tan importante y me ayudaron a lograrlo.

A mi directora de tesis Silvina Díaz, por todas las horas y esfuerzo dedicadas a formarme, pero también por las charlas, los mates, las risas y por trasmitirme tu pasión por los caballos. ¡Gracias por tu tenacidad y por todo este aprendizaje!

A Pilar Peral García, codirectora de esta tesis y directora del IGEVET. Gracias por abrirme las puertas del instituto y siempre estar dispuesta a escuchar mis inquietudes y necesidades. Gracias por tu calidez y cariño.

A todos los integrantes de la inolvidable familia IGEVET.

A Guillermo Giovambattista, por el buen humor, la buena disposición para compartir tus conocimientos y por el entusiasmo de ir siempre a más.

A mis compañeros de Cátedra por enseñarme a enseñar y brindarme su apoyo. Especialmente a Egle Villegas Castagnasso y a Martín Vicente: con Uds. todo es mejor.

A quienes en estos años compartí mates, almuerzos en IGEVET y alguna salida por la ciudad: a Julieta Corvi, Gabriela Rudd Garcés (mi cubana favorita), Beatriz Zandoná, Elisa Córdoba, Cecilia Bruno, Suani Pacheco, Nadia Castillo.

A Víctor «*yo siempre ando bien*» Sabatini, gracias por la buena onda y por haberme ayudado en momentos clave.

A Mónica Carino por estar siempre dispuesta a ayudarme en el laboratorio, enseñarme tanto y compartir tu experiencia conmigo.

A mi coterráneo Andrés Rogberg Muñoz por estar siempre dispuesto a compartir tus conocimientos desinteresadamente y ayudarme en lo que estuviera a tu alcance.

A Hernán Morales, gracias por la buena onda, por todo tu trabajo y tiempo para ayudarme con los análisis bioinformáticos. También a Hernán Olivera por su ayuda en el procesamiento de datos genealógicos.

A Euge Zappa: gracias por tu compañerismo, tu empatía y ayuda incondicional.

A Sebastián Sadaba, mi *coequiper* en esta aventura como doctorandos, por los muestreos, los asados. Gracias por ser siempre genuino y amable conmigo.

A Agustín Falomir, Julián Crespi y Diego Posik, quienes sin duda hicieron más divertidos los días en el IGEVET. Gracias, por los mates, las risas, los chistes.

A Laurita Barrientos, gracias por tu energía, el aguante y por tu carcajada sanadora.

A mi entrañable amiga Eglita Villegas Castagnasso por estar siempre ahí, por contenerme, acompañarme y escucharme, por hacerme sentir todo tu cariño.

A mi amiga de fierro Noe Leyún, quien a pesar de la distancia estuvo siempre acompañándome y además me dio la ahijada más linda.

A mis grandes amigas Ale Kröger y Claudia Elizondo, gracias por palpar esta etapa conmigo y siempre creer en mí.

A mis amigos en Uruguay: Ine Carrocio, Sole Guasco Pía Grignola, Fede Bialade, Sole Doldan, Chofa Ramirez, Shei Giacamann, Nani Elizondo y Mauri Díaz, siempre presentes desde el otro lado del charco, organizando para vernos cada vez que iba a Montevideo. También a Luchita Negro y a Mari Pereira por el aguante y aliento desde el otro hemisferio.

A mi familia por estar siempre presente, no importa desde donde, siempre interesándose por mí, haciéndome llegar su cariño, apoyo y contención.

A mis padrinos por alentarme siempre y guiarme cuando lo necesité.

A mi madre, quien me acompaña siempre, quien es faro en la oscuridad y puerto seguro donde sé que siempre puedo volver. Gracias ma, por quererme tanto y hacérmelo sentir.

A Manu, gracias por ser mi refugio todos estos años. Por tu familia quienes siempre se preocuparon por mí y me recibieron de la mejor manera. Tus padres, hermanos, abuelos, tíos y primos que me hicieron sentir parte de la familia, siempre. Gracias por elegirme en mis luces, pero también en mis sombras. Por siempre alentarme a seguir adelante, motivándome a ser mi mejor versión y alegrándote por mis logros. Gracias infinitas por nuestro hijo Alfonso, quien alegra todos nuestros días y es lo más lindo que tenemos. Gracias por tu amor, tu mirada bondadosa y por ser mi cable a Tierra.

Esta etapa fue mucho más que un doctorado... ¡me estalla el corazón al ver todo lo que coseché en este camino!

¡Gracias!

Citas bibliográficas correspondientes a las publicaciones parciales del trabajo de tesis:

Corbi Botto CM, Sadaba SA, Zappa ME, Peral García P, Díaz S. (2016). «Análisis de la diversidad genética de una población de caballos Criollo Argentino mediante polimorfismos de nucleótido simple de los genes IL12B y TNF- α .» *Analecta Veterinaria*, 36(2): 11-15.

Corbi-Botto CM, Sadaba SA, Zappa ME, Peral-García P, Díaz S. (2017). «Nonsynonymous changes of equine lentivirus receptor-1 (ELR1) gene in amino acids involved in the interaction with equine infectious anemia virus (EIAV)». *Research in Veterinary Science*, 112, 185-191.

Corbi-Botto CM, Morales-Durand H, Zappa ME, Sadaba SA, Peral-García P, Giovambattista G, Díaz S. (2019). «Genomic structural diversity in Criollo Argentino horses: Analysis of copy number variations» *Gene*, 695, 26-31.

Índice de contenidos

Sección	Página
Resumen	1
Abstract	2
1.Introducción	3
1.1 Origen del caballo Criollo Argentino	4
1.2 Definición de raza	5
1.3 Estándar o modelo racial del caballo Criollo Argentino	5
1.4 Actualidad del caballo Criollo Argentino	7
1.4.1 Registros genealógicos	10
1.5 La genética del caballo Criollo Argentino	11
1.5.1 Estado del arte	11
1.5.2 Variabilidad genética y genómica	12
2. Hipótesis	16
3. Objetivos	17
3.1 Objetivo general	18
3.2 Objetivos específicos	18
<i>CAPÍTULO 1</i>	
<i>Variabilidad en genes implicados en la respuesta inmune en poblaciones equinas de Argentina</i>	19
1. Introducción	20
1.1 Genes implicados en la respuesta inmune	21
1.2 Marcadores genéticos analizados	23
1.2.1 Gen <i>ELA - DRA</i>	23
1.2.2 Genes <i>ELA - DRB</i>	24
1.2.3 Gen del factor de necrosis tumoral alfa	24
1.2.4 Gen de la interleuquina 12	25
1.2.5 Gen del receptor de lentivirus equino 1	26

2. Materiales y métodos	28
2.1 Material animal	29
2.1.1 Toma de muestras y extracción de ADN	30
2.1.2 Evaluación de la calidad y cantidad de ADN	31
2.2 Amplificación por reacción en cadena de la polimerasa	31
2.2.1 Verificación y purificación de los productos de PCR	33
2.3 Métodos de genotipificación	35
2.3.1 Secuenciación directa	35
2.3.2 Pirosecuenciación®	35
2.3.3 Genotipificación por PCR - RFLP	39
2.4 Marcadores analizados	40
2.4.1 Gen <i>ELA - DRA</i>	40
2.4.2 Gen <i>ELA - DRB2</i>	41
2.4.3 Gen <i>ELA - DRB3</i>	45
2.4.4 Gen de la subunidad p40 de la interleuquina 12	46
2.4.5 Gen del factor de necrosis tumoral alfa	48
2.4.6 Gen del receptor de lentivirus equino 1	51
2.5 Análisis bioinformático	55
2.6 Medidas de variabilidad genética	57
2.6.1 Frecuencias alélicas y equilibrio de Hardy-Weinberg	57
2.6.2 Heterocigosis	58
2.6.3 Estadísticos F de Wright	60
2.7 Análisis estadístico poblacional	62
2.8 Análisis filogenético	63
2.9 Análisis de ligamiento	63
3. Resultados	64
3.1 Detección de SNPs en los genes inmunes	65

3.1.1 Gen <i>ELA-DRA</i>	65
3.1.2 Genes <i>ELA-DRB2</i> y <i>ELA-DRB3</i>	66
3.1.3 Gen de la subunidad p40 de la interleuquina 12	72
3.1.4 Gen del factor de necrosis tumoral alfa	73
3.1.5 Gen del receptor de lentivirus equino 1	78
3.1.5.1 SNP <i>rs68535250</i> del exón 4 del gen <i>ELR1</i>	78
3.1.5.2 Exones 3 y 5 del gen <i>ELR1</i>	81
3.1.5.2.1 Análisis filogenético de las secuencias del exón 5 de <i>ELR1</i>	92
4. Discusión	94
4.1 Genes del ELA	95
4.2 Genes de citoquinas	97
4.3 <i>ELR1</i>	101
CAPÍTULO 2	
<i>Variabilidad genómica en caballos Criollo Argentino</i>	108
1. Introducción	109
1.1 Arreglos de ADN para análisis genómico	110
1.2 Antecedentes de datos genómicos en razas de caballos	111
2. Materiales y métodos	113
2.1 Material animal	114
2.2 Análisis genealógico	115
2.3 Genotipificación con arreglo de mediana densidad	117
2.3.1 Control de calidad de la genotipificación	118
2.3.2 Control de calidad de los datos	118
2.3.3 Análisis de la diversidad genómica	119
2.3.3.1 Heterocigosis esperada	119
2.3.3.2 Equilibrio de Hardy-Weinberg	120
2.3.3.3 Estadísticos F de Wright	120

2.3.4 Desequilibrio de ligamiento	121
2.3.5 Detección y análisis de variaciones estructurales	122
2.3.5.1 Análisis funcional y enriquecimiento génico	123
2.4 Análisis genético interracial	124
2.4.1 Muestras y conjunto de datos	124
2.4.2 Diferenciación genética	126
2.4.3 Análisis de varianza molecular	126
2.4.4 Análisis de la estructura poblacional	127
3. Resultados	129
3.1 Análisis genealógico	130
3.2 Análisis genómico	132
3.2.1 Diversidad genómica y equilibrio de Hardy-Weinberg	132
3.2.2 Análisis de consanguinidad	134
3.2.2.1 Endogamia individual genómica	134
3.2.2.2 Estimación de los coeficientes de parentesco genómico	135
3.2.3 Desequilibrio de ligamiento	137
3.2.4 Variaciones genómicas estructurales	138
3.2.4.1 Análisis funcional y enriquecimiento génico	142
3.3 Análisis genético interracial	146
3.3.1 Diferenciación genética	146
3.3.2 Análisis de la varianza molecular	147
3.3.3 Análisis de la estructura poblacional	149
4. Discusión	153
4.1 Caracterización genómica de los caballos Criollo Argentino	154
4.2 Diversidad genómica de los caballos Criollo Argentino	156
4.3 Variabilidad genómica estructural	158
4.4 Análisis genómico interracial	160

5. Discusión General	164
6. Conclusiones	170
6.1 Conclusiones del Capítulo 1	171
6.2 Conclusiones del Capítulo 2	172
6.3 Conclusión final	173
7. Bibliografía	174
8. Anexos	200

Índice de tablas	Pág.
Capítulo 1	
Tabla 1.1 Cebadores utilizados para la amplificación por PCR del exón 2 del gen <i>ELA-DRA</i>	39
Tabla 1.2. Cebadores utilizados para la amplificación por PCR anidada del exón 2 del gen <i>ELA-DRB2</i>	41
Tabla 1.3. Definición de alelos del gen <i>ELA-DRB2</i> por patrones RFLP.....	43
Tabla 1.4. Cebadores utilizados para la amplificación por PCR anidada del exón 2 del gen <i>ELA-DRB3</i>	45
Tabla 1.5. Cebadores utilizados para la amplificación por PCR del exón 5 del gen <i>IL12p40</i>	46
Tabla 1.6. Cebadores utilizados para la amplificación por PCR anidada del gen <i>TNF-α</i>	49
Tabla 1.7. Cebadores utilizados para la amplificación por PCR del exón 4 del gen <i>ELR1</i>	52
Tabla 1.8. Cebadores utilizados para la amplificación por PCR de los exones 3 y 5 del gen <i>ELR1</i>	53
Tabla 1.9. Posiciones de nucleótido variables en las secuencias del exón 2 de <i>ELA-DRB3</i> de los caballos Criollo Argentino.....	68
Tabla 1.10. Frecuencias de los alelos <i>ELA-DRA</i> , <i>ELA.DRB2</i> y <i>ELA-DRB3</i> observados en los caballos Criollo Argentino.....	70
Tabla 1.11. Frecuencias alélicas y parámetros de diversidad genética de los 3 SNPs del promotor de <i>TNF-α</i> en distintas poblaciones equinas de Argentina.....	75
Tabla 1.12. Frecuencias de los haplotipos detectados en la región promotora del gen <i>TNF-α</i> mediante pirosecuenciación en poblaciones equinas de Argentina.....	76
Tabla 1.13. Frecuencias y diversidad genética del SNP <i>rs68535250</i> del exón 4 del gen <i>ELR1</i>	79
Tabla 1.14. Análisis del SNP <i>rs68535250</i> del exón 4 del gen <i>ELR1</i> .	79
Tabla 1.15. Polimorfismos detectados por secuenciación en los exones 3 y 5 del gen <i>ELR1</i>	83

Tabla 1.16. Frecuencias de las variantes alélicas definidas para el exón 3 del gen <i>ELR1</i>	87
Tabla 1.17. Frecuencias de las variantes alélicas definidas para el exón 5 del gen <i>ELR1</i>	90
Capítulo 2	
Tabla 2.1. Valores medios de los parámetros de diversidad analizados por cromosoma en la población de caballos Criollo Argentino.....	132
Tabla 2.2. CNVRs detectadas en la población de caballos Criollo Argentino.....	138
Tabla 2.3. Análisis de enriquecimiento funcional de los genes presentes en las CNVRs en la población de caballos Criollo Argentino.....	145
Tabla 2.4. Análisis de varianza molecular (AMOVA) entre poblaciones equinas de diversas razas.....	148
Índice de figuras	Pág.
Capítulo 1	
Figura 1.1. Ejemplar de la raza Criollo Argentino.....	7
Figura 1.2. Distribución y porcentaje de cabañas de Caballo Criollo Argentino por provincia.....	9
Figura 1.3. Esquema representativo de la cascada de reacciones que ocurren durante la Pirosecuenciación ^R	36
Figura 1.4. Esquema representativo de la purificación del producto de PCR biotinilado.....	37
Figura 1.5. Patrones de restricción posibles para la definición de alelos del exón 2 de <i>ELA-DRB2</i>	42
Figura 1.6. Diseño experimental de la amplificación del exón 5 de <i>IL12p40</i>	47
Figura 1.7. Diseño experimental utilizado para amplificar la región promotora del <i>TNF-α</i>	48

Figura 1.8. Representación esquemática de la estructura del gen equino <i>ELR1</i>	51
Figura 1.9. Pirogramas de los genotipos detectados para el exón 2 del gen <i>ELA-DRA</i> en caballos Criollo Argentino.....	64
Figura 1.10. Fragmentos de PCR del exón 2 de <i>ELA-DRB2</i> analizados por PCR-RFLP con tres enzimas para realizar un rastreo de variabilidad.....	66
Figura 1.11. Pirogramas del exón 5 del gen <i>IL12p40</i> . En azul se observan los distintos genotipos detectados.....	71
Figura 1.12. Genotipos detectados en el gen <i>TNF-α</i> mediante Pirosecuenciación.....	74
Figura 1.13. Histogramas teóricos y pirogramas obtenidos para cada genotipo del SNP: rs68535250 del exón 4 del gen <i>ELR1</i>	78
Figura 1.14. Electroferograma parcial de la secuenciación del exón 5 del gen <i>ELR1</i>	81
Figura 1.15. Alineamiento nucleotídico de las variantes alélicas del exón 3 del gen <i>ELR1</i>	86
Figura 1.16. Alineamiento nucleotídico de las variantes alélicas del exón 5 del gen <i>ELR1</i>	89
Figura 1.17. Árbol de <i>Neighbor-joining</i> de las variantes del gen equino <i>ELR1</i> y sus ortólogos construido con el software MEGA7.....	92
Capítulo 2	
Figura 2.1 Esquema de los criterios aplicados para la selección de muestras a incluir en el genotipado con el arreglo GGP Equine70k..	114
Figura 2.2 Contribución de los ancestros principales a la población actual de caballos criollos.....	130
Figura 2.3 Valores del coeficiente de endogamia individual (F) respecto a la homocigosis observada en los 24 caballos Criollo Argentino analizados.....	134
Figura 2.4 Coeficiente de parentesco genómico estimado versus la proporción de SNPs que no comparten identidad por estado (IBS0).....	135

Figura 2.5 Decaimiento del desequilibrio de ligamiento (r^2) promedio entre pares de marcadores en función de la distancia genética evaluada cada 100 kb en la población de caballos Criollo Argentino.....	136
Figura 2.6 Distribución cromosómica de las CNVRs génicas e intergénicas.....	139
Figura 2.7 Mapa de las CNVRs detectadas en la población de 24 caballos Criollo Argentino.	140
Figura 2.8 Clasificación por ontología génica de los genes de las CNVRs en caballos Criollo Argentino utilizando el software PANTHER.....	142
Figura 2.9 Valores de diferenciación genética (F_{ST}) calculados en las 9 poblaciones equinas tomadas de a pares, con el DS3 conteniendo 19.614 SNPs autosómicos.....	147
Figura 2.10 Resultado para 2 valores de K del agrupamiento Bayesiano realizado con el software fastStructure utilizando 19.614 SNPs, en los 200 caballos de las 9 poblaciones equinas analizadas.....	150
Figura 2.11 Análisis de componentes principales en base a 38.199 SNPs en las 9 poblaciones equinas.....	152
Índice de anexos	Pág.
Anexo 1 Purificación de las reacciones de pcr por precipitación con PEG.....	201
Anexo 2 Mapa de filtros para definición de los datasets 1 y 2.....	203
Anexo 3 Mapa de filtros para definición de los datasets 3 y 4.....	204
Anexo 4 Resultados complementarios.....	205

Abreviaturas y símbolos

ABS: Sitio de unión a antígeno (en inglés, *Antigen Binding Site*).

ACCC: Asociación de Criadores de Caballos Criollos.

ADN: Ácido desoxirribonucleico.

AIE: Anemia infecciosa equina.

CNV: Variación del número de copias (en inglés, *Copy Number Variation*).

CNVR: Región de variaciones del número de copias (en inglés, *Copy Number Variation Region*).

CRD: Dominios ricos en cisteína (en inglés, *cysteine-rich domains*).

DRA: Gen *ELA-DRA*.

DRB: Genes *ELA-DRB*.

ELA: Complejo Mayor de Histocompatibilidad equino (en inglés, *Equine Lymphocyte Antigen*).

ELR1: Gen del receptor de lentivirus equino 1 (en inglés, *Equine Lentivirus Receptor 1*).

HWE: Equilibrio de Hardy-Weinberg (en inglés, *Hardy-Weinberg Equilibrium*).

IL-12: Gen de la interleuquina 12.

INTA: Instituto Nacional de Tecnología Agropecuaria.

ISAG: Sociedad Internacional de Genética Animal.

LD: Desequilibrio de ligamiento (en inglés, *Linkage disequilibrium*)

MAF: Frecuencia del alelo menor (en inglés, *Minor Allele Frequency*),

MHC: Complejo Mayor de Histocompatibilidad.

mM: Milimolar.

Mpb: Millón de pares de bases.

ng: Nanogramo.

pb: Pares de bases.

PCR: Reacción en cadena de la polimerasa (en inglés, *Polymerase Chain Reaction*).

RRGG: Registros genealógicos de la Sociedad Rural Argentina.

SNP: Polimorfismo de nucleótido simple (en inglés, *Single Nucleotide Polymorphism*).

TNF- α : Factor de necrosis tumoral alfa (en inglés, *Tumor Necrosis Factor Alpha*).

TNFRSF: Superfamilia de receptores del factor de necrosis tumoral.

μ L: Microlitro.

μ M: Micromolar.

VAIE: Virus de la anemia infecciosa equina.

Glosario

Acervo genético: conjunto de alelos presentes en una población.

Alelos: variantes de un gen.

Call rate: valor mínimo aceptado en el control de calidad de datos.

Cebador: secuencia corta de ADN de cadena simple complementaria al ADN blanco.

Endogamia: apareamientos entre individuos emparentados.

Fenotipo: características observables de un organismo.

Genoma: información genética total de un organismo.

Haplobloque: número mínimo de SNPs que explican los haplotipos comunes en una secuencia.

Haplotipos: combinación de alelos de diferentes loci que se heredan juntos.

Marcadores genéticos: variaciones en la secuencia de ADN, con modo de herencia conocido y rastreables en la población.

Microsatélites: marcador genético que consiste en secuencias de 2 - 6 nucleótidos repetidos en tándem. Son polimorfismos donde los alelos se definen por la diferencia en el número de repeticiones.

Mutaciones genéticas: variaciones en la secuencia de ADN.

Población: en términos genéticos refiere a un grupo de individuos de una especie o raza que se aparean entre sí.

Polimorfismo: variación en la secuencia de ADN con frecuencia en la población de al menos un 1 %.

Recursos zoogenéticos: especies o razas animales que son de interés económico, científico y cultural para la agricultura.

Resumen

La raza de caballos Criollo Argentino es un patrimonio ganadero local y representa un recurso único en cuanto a la identidad y al sistema productivo de Argentina. El objetivo de esta tesis doctoral fue caracterizar el caballo Criollo Argentino mediante abordajes complementarios. En el Capítulo 1, se detectaron y analizaron los polimorfismos de 6 genes implicados en la respuesta inmune (*ELA-DRA*, *ELA-DRB2*, *ELA-DRB3*, *TNF- α* , *IL12* y *ELR1*) en 96 caballos criollos y en 130 caballos de poblaciones equinas de Argentina. La variabilidad en los genes del ELA fue elevada (24 alelos, 15 nuevos); en *IL-12B* se detectaron dos alelos; se detectaron 4 haplotipos en *TNF- α* , uno observado exclusivamente en el Criollo Argentino (TCT, $f = 0,038$); el gen *ELR1* mostró variabilidad a nivel de la secuencia codificante (21 SNPs, 14 alelos), uno de ellos con un probable efecto deletéreo en la función de la proteína predicha. En el Capítulo 2, el análisis genealógico mostró que el 36,9 % de los animales están emparentados, con un grado reducido de consanguinidad ($F = 0,041$). A partir de la población de referencia ($N = 280$), se estimó un número efectivo de fundadores de 157,7 con 50 ancestros principales contribuyendo con sus genes a la población criolla actual. La caracterización genómica con el arreglo de ADN de 65.175 SNPs en 24 animales representativos de la raza, indicó variabilidad media a elevada ($He = 0,347$) y consanguinidad genómica baja ($F = - 0,009$). La variación genómica estructural reflejó un enriquecimiento en genes vinculados a funciones de defensa/inmunidad en los caballos criollos. En la comparación genómica interracial y en los análisis de estructuración poblacional se determinó que el caballo Criollo Argentino es una entidad genética definida, con alta variabilidad y baja endogamia a nivel genético, genealógico y genómico. Estos resultados destacan la importancia de esta raza y su acervo genético para el estudio de caracteres y patologías de la especie equina.

PALABRAS CLAVE: equinos, genealogías, SNP, citoquinas, genómica.

Abstract

The Criollo Argentino horse breed is a local livestock heritage that represents a unique resource in terms of the identity and the productive system of Argentina. The aim of this doctoral dissertation was to characterize the Criollo Argentino horse by means of complementary approaches. In Chapter 1, six genes implied in the immune response were analyzed (*ELA-DRA*, *ELA-DRB2*, *ELA-DRB3*, *TNF- α* , *IL12* and *ELR1*) through the detection and analysis of polymorphisms in 96 Criollo horses and 130 horses from horse herds of Argentina. High variability was observed in *ELA* genes (24 alleles, 15 of them novel); 2 alleles were detected in *IL-12B*; 4 haplotypes were detected in *TNF- α* , one of them only observed in Criollo Argentino horses (TCT, $f = 0.038$); *ELR1* gene was highly variable in its coding regions (21 SNPs, 14 alleles), one of which could have a deleterious effect in the protein function. In Chapter 2, pedigree analysis showed that 36,9 % of the horses are inbred, with a low level of inbreeding ($F = 0.041$). For the reference population ($N = 280$), an effective number of founders of 157,7 and 50 major ancestors that contributed their genes to the current Criollo population, were estimated. Genomic characterization in 24 representative horses of the breed, using a DNA array containing 65157 SNPs, revealed moderate to high variability ($He = 0.347$) and low genomic inbreeding ($F = -0.009$). Structural genomic variation showed enrichment in genes related to defense/immunity functions. Among breed genomic comparison and population structuring analyses, showed the Criollo Argentino horse as a defined genetic entity, with high variability and low inbreeding at the genetic, genealogical and genomic level. These results highlight the importance of this breed and its gene pool for the study of characters and pathologies of the equine species.

KEY WORDS: horse, genealogy, SNP, cytokines, genomics.

1. INTRODUCCIÓN

1. Introducción general

1.1 Origen del caballo Criollo Argentino

La especie *Equus caballus* pertenece a la familia *Equidae*, dentro del género *Equus* y el orden *Perisodactyla*, cuyo representante más antiguo se remonta a 55 millones de años (Orlando *et al.*, 2009). El origen del caballo doméstico es probablemente el más difícil de definir entre las especies domésticas. Sin embargo, en el caso de las poblaciones criollas de América Latina se puede fijar su origen en el ganado caballar introducido por los conquistadores españoles durante el siglo XVI (Rodero *et al.*, 1992). Desde la zona del Caribe los caballos tomaron dos rutas: una hacia México y California, la otra hacia Colombia y Venezuela, continuando luego hacia la región rioplatense (Primo, 2004).

El caballo Criollo Argentino actual es descendiente directo de los caballos traídos al territorio del Virreinato de la Plata en el año 1534 por Don Pedro de Mendoza en su primera expedición, y resultado de años de reproducción en libertad, así como de siglos de adaptación a condiciones ambientales muy diversas. Esta adaptación a las condiciones de vida del medio ambiente le permitió desarrollar sus grandes cualidades de resistencia a las enfermedades (Dowdall, 1982), posicionando al caballo Criollo Argentino como la herramienta de trabajo para muchos productores en este país.

En 1911 comienza una etapa conocida como Recuperación Racial, en la que el Dr. Emilio Solanet y otros criadores comenzaron a buscar

linajes de caballos Criollos puros con el objetivo de recuperar la raza y aumentar su base genética. Finalmente, en 1922 se redacta el estándar de la Raza Criolla y en 1923 se funda la Asociación de Criadores de Caballos Criollos (ACCC - <https://www.caballoscriollos.com/sitio/>).

1.2 Definición de raza

De acuerdo con la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO, www.fao.org), los recursos genéticos de los animales de granja son aquellas especies animales y sus poblaciones utilizadas o que pueden utilizarse con fines de productividad alimentaria y agrícola. Por otra parte, la FAO define una raza como aquel grupo de animales domésticos identificables por exploración visual, cuyas características externas semejantes permiten separarlo fenotípicamente dentro de una misma especie. Finalmente, una población en términos genéticos hace referencia a todos los individuos pertenecientes a una especie que se reproducen entre sí, transmitiendo la variabilidad genética de generación en generación (FAO, www.fao.org).

1.3 Estándar o modelo racial del caballo Criollo Argentino

El caballo Criollo Argentino conserva características a nivel morfológico, de sus antecesores españoles, presentando formas y medidas armoniosas e intermedias (eumétrico y mesoformo), que se

corresponde a un caballo de silla. Presenta buena musculatura, espalda fuerte, gran amplitud de pecho, cabeza de base ancha y vértice fino con perfil preferentemente rectilíneo o subconvexilíneo (Figura 1). En cuanto al pelaje, se aceptan todos los pelajes a excepción del «tobiano» y el «pintado». Con respecto al carácter, es un caballo activo, enérgico y dócil, ágil y rápido en sus movimientos. Esta raza se identifica por su rusticidad, longevidad, fertilidad, resistencia, valentía, poder de recuperación y aptitud para trabajos ganaderos. En la actualidad, la cría de caballos criollos en Argentina está orientada a la conformación de un caballo de silla con gran aptitud para el trabajo ganadero y que conserve las cualidades adaptativas producto de sus años de cría en libertad.



Figura 1.1. Ejemplar de la raza Criollo Argentino. Fotografía tomada en la Cabaña de Caballos Criollos - Asociación Cooperadora INTA, Leales, Tucumán.

1.4 Actualidad del caballo Criollo Argentino

La raza Criolla es una de las razas referentes de la especie equina en Argentina, y, por lo tanto, un patrimonio ganadero local que representa un recurso único en cuanto a la identidad y al sistema productivo del país. Actualmente existen en Argentina aproximadamente 900 criadores activos, concentrados principalmente en la provincia de Buenos Aires (Figura 2), inscribiéndose aproximadamente 6.000 potrillos por año, lo que

hace a la ACCC la más numerosa (ACCC, <http://www.caballoscriollos.com>). El caballo de la raza Criolla fue declarado como Caballo Nacional y patrimonio cultural de la Argentina por el Senado y Cámara de Diputados de la Nación Argentina Ley N.º 27414 (75-S.-2017).

En Argentina, desde 1999, por iniciativa de la Federación Ecuestre Argentina, el 20 de septiembre se celebra el Día Nacional del Caballo en conmemoración de la llegada, en 1925, de Gato y Mancha a Nueva York, con el profesor suizo Aimé Félix Tschiffely. Estos caballos argentinos conquistaron el récord mundial de distancia y altura de la época, al recorrer 21.500 kilómetros a lo largo de todo el continente y alcanzar 5.900 metros sobre el nivel del mar en el paso El Cóndor (Bolivia). Los animales eran criollos, tenían 18 años de edad y habían sido comprados por Emilio Solanet al cacique tehuelche Liempichín, en Chubut. Eran, además, descendientes directos de los caballos traídos por Pedro de Mendoza en 1535.

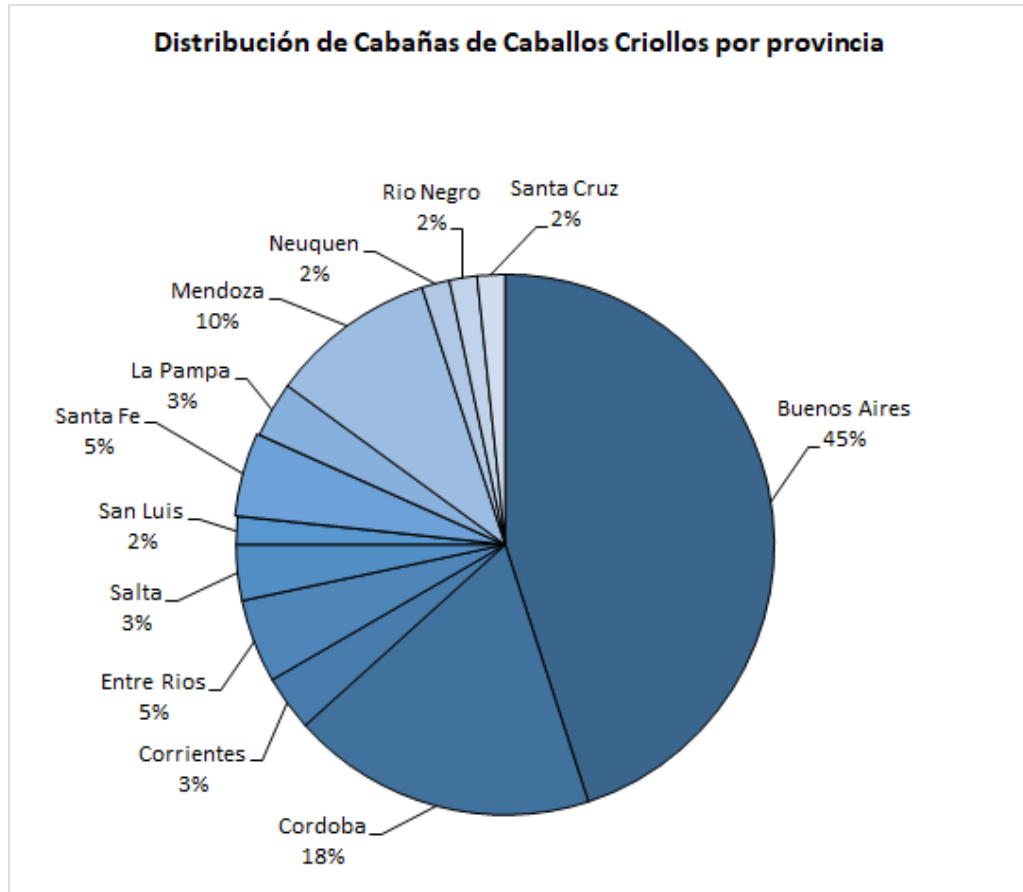


Figura 1.2. Distribución y porcentaje de cabañas de Caballo Criollo Argentino por provincia. Fuente: <http://www.caballoscriollos.com>

1.4.1 Registros genealógicos

Los datos genealógicos se han utilizado como estrategia para evaluar la diversidad genética dentro de poblaciones y razas de caballos domésticos (Głażewska *et al.*, 2007; Leroy *et al.*, 2013), por medio de la estimación del coeficiente de consanguinidad definido como la probabilidad de que dos alelos de un locus dado sean idénticos por descendencia (Malécot, 1948); por medio del parentesco (coeficiente de consanguinidad de la descendencia hipotética de dos individuos) (Caballero y Toro, 2000), y por el número efectivo del tamaño poblacional y de los ancestros. El conocimiento incompleto de las genealogías puede limitar tales estrategias (Boichard *et al.*, 1997), por lo que se pueden investigar las bases de datos genealógicos para obtener los datos completos de los caballos. Es bien conocida la historia de los caballos Criollos y su origen, sin embargo, la contribución genética de estos caballos en Argentina permanece con frecuencia en el registro social y folclórico de la transmisión de los nombres de las líneas de sangre entre los criadores de la raza. Los registros genealógicos de la raza Criolla en Argentina son mantenidos por la Sociedad Rural Argentina (SRA), y para la identificación e inscripción de los ejemplares de la Raza Criolla se utiliza la identificación por ADN y microchips en la provincia de Buenos Aires.

1.5 La genética del caballo Criollo Argentino

1.5.1 Estado del arte

La caracterización de una población es un paso fundamental para establecer estrategias de selección, conservación y utilización. El conocimiento de la variabilidad genética entre individuos de una especie o raza resulta relevante para comprender aquellas vías biológicas que tienen como resultado final una determinada característica fenotípica, como, por ejemplo, los caracteres genéticos que determinan los procesos biológicos y celulares que subyacen a una condición patológica.

Las primeras investigaciones en genética de caballos criollos argentinos se iniciaron hace varias décadas, y consistieron en la caracterización genética mediante la detección de grupos sanguíneos y polimorfismos bioquímicos de proteínas séricas (Peral García, 1994; Peral García *et al.*, 1996; Díaz, 1997; Villegas Castagnasso, 2005). Posteriormente la identificación genética se realizó con marcadores del tipo microsatélite (Kienast, 2004), recomendados por la Sociedad Internacional de Genética Animal (ISAG). Además, se realizaron estudios en genes del Complejo Principal de Histocompatibilidad (MHC) equino (*Equine Lymphocyte Antigen - ELA*), que condujeron a detectar y caracterizar nuevos alelos, identificar SNPs (en inglés, *Single Nucleotide Polymorphisms*, Polimorfismos de Nucleótido Simple) en regiones que regulan la expresión de determinados genes, así como demostrar las

relaciones evolutivas entre los mismos (Díaz *et al.*, 2001; Díaz, 2003; Díaz *et al.*, 2005; Sadaba, 2018).

1.5.2 Variabilidad genética y genómica

La identificación de variantes en el ADN que contribuyen a las enfermedades y rasgos específicos ha sido un objetivo central en la genética humana, y ha cobrado importancia creciente en especies domésticas y pecuarias (*Online Mendelian Inheritance in Animals*, OMIA, <http://omia.org/>). La variabilidad genética de una población puede ser determinada a escala genómica teniendo en cuenta que, en la secuencia de ADN, las variaciones o SNPs son posiciones genómicas en las que coexisten dos nucleótidos distintos (alelos) y cada uno aparece en una proporción significativa (> 1 %) de la población (polimorfismos). Los SNPs son la fuente de variación genética más común e informativa en los individuos de una especie. Si bien la gran mayoría de los sitios genómicos (99 %) de cada individuo tiene el mismo residuo nucleotídico en ambos cromosomas homólogos (LaFramboise, 2009), el número de SNPs publicados en las bases de datos supera el millón y medio en humanos, y corresponden a una frecuencia aproximada de un SNP cada 2000 pb (Kruglyak y Nickerson, 2001; Zhao *et al.*, 2003), en tanto que el genoma equino tiene un tamaño aproximado de 2,67 Gb con una frecuencia de un SNP cada 1500 pb (Wade *et al.*, 2009).

Entre las variaciones genéticas que subyacen al espectro de fenotipos diversos, se encuentran aquellos cambios o mutaciones genéticas que determinan la aparición de una característica que puede alterar el estado normal de salud de un organismo, que puede causar una enfermedad, y que pueden transmitirse de generación en generación, es decir, son hereditarias. Partiendo de la premisa de que la expresión diferencial de algunos genes podría tener consecuencias en el proceso de las enfermedades, es importante profundizar en el conocimiento de la variación de las secuencias en los genes implicados en la respuesta inmune (RI). En humanos se ha reportado que la variación genética afecta la capacidad de producción de citoquinas (Li *et al.*, 2016) y que la ocurrencia de SNPs en estos genes podría ocasionar variaciones tanto en las frecuencias alélicas y haplotípicas, así como en el desequilibrio de ligamiento (Lazarus *et al.*, 2002). De esta manera, el estudio de la diversidad genética en estos *loci* podría aportar información valiosa respecto a la susceptibilidad o resistencia de una población (Janova *et al.*, 2013).

Teniendo en cuenta las características genéticas que presenta el caballo Criollo Argentino a nivel de los marcadores analizados en estudios previos (Peral García, 1994; Díaz, 1997), así como sus rasgos de rusticidad y adaptabilidad, producto de períodos en que la raza estuvo sujeta a la selección natural en ambientes diversos, el estudio de esta raza aportará información de un valioso acervo genético.

En este trabajo de Tesis Doctoral, la caracterización se realizó mediante dos abordajes complementarios y no excluyentes. En principio, en el Capítulo 1: «Variabilidad en genes implicados en la respuesta inmune en poblaciones equinas de Argentina», se realiza la detección y el análisis de polimorfismos de genes candidatos direccionada a determinados genes implicados en la respuesta inmune en el caballo Criollo Argentino y en otras poblaciones equinas de Argentina. Se establece como comparación la situación de la variabilidad genética observada en grupos de caballos con diagnóstico de Anemia Infecciosa Equina, dado que se realiza la descripción y análisis genético del receptor de lentivirus 1, encargado de la entrada del virus.

En segunda instancia, en el Capítulo 2: «Variabilidad genómica en caballos Criollo Argentino», se realiza el abordaje desde una perspectiva a nivel genómico global, para evaluar la diversidad del caballo Criollo Argentino tanto a nivel poblacional como interracial mediante la descripción de sus características genómicas y la comparación con los genomas de otras razas equinas. Este abordaje a nivel genómico es el primero que se realiza en una raza caballar de Argentina.

2. HIPÓTESIS

2. Hipótesis

La presencia de polimorfismos de nucleótido simple (SNPs) dentro de los genes inmunes permite caracterizar la variabilidad genética en recursos zoogenéticos.

3. OBJETIVOS

3. Objetivos

3.1 Objetivo general

Caracterizar la diversidad genética en el caballo Criollo Argentino mediante el análisis de la variabilidad de los genes de citoquinas, utilizando estrategias de detección de SNPs, y realizar el análisis comparativo inter e intrarracial con otras razas y poblaciones equinas.

3.2 Objetivos específicos

- 3.2.1 Relevar las bases de datos de secuencias de referencia en busca de SNPs informativos de citoquinas (reportados y validados en especies modelo y en equinos), y diseñar una estrategia de tipificación para los SNPs seleccionados.
- 3.2.2 Tipificar a nivel genético molecular los SNPs informativos de citoquinas seleccionados en 3.2.1, en los caballos de la raza Criollo Argentino y otras poblaciones equinas.
- 3.2.3 Tipificar y analizar a nivel genómico los caballos Criollo Argentino empleando un arreglo de mediana densidad.
- 3.2.4 Realizar el análisis genómico comparativo con caballos de diferentes razas y poblaciones equinas.

Capítulo 1

Variabilidad en genes implicados en la respuesta inmune en poblaciones equinas de Argentina

1. INTRODUCCIÓN

1. Introducción

1.1 Genes implicados en la respuesta inmune

El sistema inmune (SI) se define como el conjunto de tejidos, células y moléculas especializadas en defender al organismo de las infecciones causadas por diversos agentes patógenos. El SI consta de dos partes principales: el sistema inmune innato (SII) y el sistema inmune adaptativo (SIA). El SII es la primera línea de defensa contra los agentes patógenos, se caracteriza por no tener memoria y sus principales elementos son: barreras físicas como la piel y las mucosas, componentes humorales (sistema del complemento, interferón α y β , proteínas de fase aguda), células fagocíticas (neutrófilos, macrófagos, dendríticas), eosinófilos y células NK (*natural killer*). El SIA en cambio, es específico y presenta memoria, es decir, ante la exposición a un patógeno ya enfrentado previamente, es capaz de reconocerlo específica y rápidamente para iniciar la respuesta inmune necesaria, que presenta componentes celulares, los linfocitos B y T, y humorales tales como inmunoglobulinas y citoquinas.

El Complejo Mayor de Histocompatibilidad (MHC, *Major Histocompatibility Complex*) en el caballo es denominado complejo ELA (*Equine Lymphocyte Antigen*) y comprende una región genómica de aproximadamente 6 Mb en el cromosoma equino 20 (ECA20), donde se localizan genes estrechamente ligados que cumplen un rol preponderante en el desarrollo de la respuesta inmune tanto innata

como adaptativa (Holmes *et al.*, 2019). Este grupo de genes son altamente polimórficos y se encuentran organizados en tres regiones no solapadas denominadas genes de clase I (MHC-I), clase II (MHC-II) y clase III (MHC-III). Las moléculas codificadas por los genes de clase I y II son glicoproteínas de membrana encargadas de presentar péptidos de antígenos a los linfocitos T. Las moléculas de clase I se encuentran presentes en todas las células nucleadas donde cumplen la función de presentar péptidos derivados de proteínas intracelulares a los linfocitos T CD8⁺ y están formadas por dos cadenas glicoproteicas: una transmembrana, denominada α y otra citoplasmática denominada β microglobulina. Por otra parte, las moléculas de clase II se expresan en células dendríticas, macrófagos y linfocitos B y se encargan de presentar antígenos derivados de proteínas extracelulares a los linfocitos colaboradores (T-CD4⁺) y citotóxicos (T-CD8⁺). Estas glicoproteínas están compuestas por dos cadenas transmembrana, α y β , unidas mediante enlaces no covalentes y que conforman el sitio de unión a antígeno (ABS, en inglés, *Antigen Binding Site*). Los genes que codifican estas cadenas se distribuyen en cuatro loci *DQA*, cuatro loci *DQB*, seis loci *DRB* y un locus *DRA* (Viluma *et al.*, 2017). Los genes de clase III son funcionalmente heterogéneos; entre ellos se encuentran las proteínas del complemento tales como C2, C4, y el Factor de Necrosis Tumoral Alfa (TNF α - en inglés, *Tumor Necrosis Factor Alpha*).

1.2 Marcadores genéticos analizados

1.2.1 Gen *ELA-DRA*

El gen *DRA* se encuentra representado por una única copia en el genoma equino y codifica la cadena α del receptor DR. El exón 2 es el más variable del gen *DRA* (Albright-Fraser *et al.*, 1996; Brown *et al.*, 2004; Díaz *et al.*, 2008; Janova *et al.*, 2009; Kamath y Getz, 2011) y codifica aminoácidos asociados con el ABS. En contraste a lo observado para otras especies de mamíferos donde el locus *DRA* está altamente conservado (Andersson *et al.*, 1986; Yuhki *et al.*, 1997; Smith *et al.*, 2005), en el caballo doméstico se han reportado los siguientes alelos: *ELA-DRA*0101* (L47174), *ELA-DRA*0201* (M60100), *ELA-DRA*0301* (L47172), *Eqca-DRA*00101* (JQ254079-80), *Eqca-DRA*00102* (JQ254081), *Eqca-DRA*00103*(JQ254082), *Eqca-DRA*00104* (JQ254083), *ELA-DRA*JBH11* (AJ575295) (Fraser y Bailey, 1996; Brown *et al.*, 2004; Janova *et al.*, 2009; Miller *et al.*, 2017). De esta manera, los distintos alelos del gen *ELA-DRA* se definen por las variaciones del exón 2 (los más frecuentes *ELA-DRA*1*, *ELA-DRA*2* y *ELA-DRA*3*, Andersson *et al.*, 2012).

1.2.2 Genes *ELA-DRB*

El gen *DRB* codifica la cadena β del receptor DR y se ha reportado que la región *ELA* clase II presenta seis loci: tres funcionales (*Eqca-DRB1*, *Eqca-DRB2* y *Eqca-DRB3*) y tres pseudogenes (Viluma *et al.*, 2017). Los alelos reportados por Albright-Fraser (1998) son: *Eqca-DRB*5* (L76978), *Eqca-DRB*7* (L76973), *Eqca-DRB*8* (L76975), *Eqca-DRB*4* (L77079); Hedrick *et al.*, 1999: *Eqpr-DRB*1* (AF084187), *Eqpr-DRB*4* (AF084190), *Eqpr-DRB*5* (AF084191), *Eqpr-DRB*6* (AF084192); por otro lado, Díaz *et al.*, (2001): *Eqca-DRB*13* (AF170067); Kumar *et al.*, (2014): *EqcaMH3.1* (KJ596500), *EqcaMH4.1* (KJ596499); y finalmente Miller *et al.*, (2017): *Eqca-DRB2*00101* (JQ254090), *Eqca-DRB2*00102* (JQ254091), *Eqca-DRB2*00201* (JQ254092), *Eqca-DRB2*00301* (JQ25408), *Eqca-DRB2*00401* (JQ254093).

1.2.3 Gen del factor de necrosis tumoral alfa (TNF- α)

El factor de necrosis tumoral alfa (TNF- α) es una potente citoquina que cumple un rol crucial en la regulación de las respuestas inmunes e inflamatorias (Brown *et al.*, 2006). El gen que codifica esta molécula se encuentra en el cromosoma 20 equino (ECA20) (Bailey *et al.*, 1995; Lindgren *et al.*, 2001), comprende 2784 pb y presenta 4 exones. La región del gen *TNF* presenta muchos polimorfismos incluyendo

microsatélites y SNPs (Matiašovic *et al.*, 2002; Horřín *et al.*, 2004). Varios estudios han demostrado que los polimorfismos del gen *TNF- α* están asociados con distintas condiciones inflamatorias frente a la infección o el ejercicio, y numerosa evidencia sugiere que las diferencias de expresión se deben a 3 SNPs localizados en la región promotora del gen (5' UTR, en inglés, *untranslated region*), denominados: SNP₋₉₅ C/T, SNP₋₈₉ C/T, y en SNP₋₈₅ T/C (Tornquist *et al.*, 1997; Moore *et al.*, 2003; Liburt *et al.*, 2010; Arfuso *et al.*, 2020; Mostafavi *et al.*, 2020).

1.2.4 Gen de la interleuquina 12 (IL-12)

La interleuquina 12 (IL-12) es una citoquina proinflamatoria que cumple un papel central en la coordinación de la inmunidad innata y adaptativa. La molécula de IL-12 es un heterodímero compuesto por una cadena de 35 kDa (p35 o IL-12 α) y una cadena de 40 kDa (p40 o IL-12 β). En trabajos previos se ha reportado un polimorfismo de nucleótido simple en la subunidad p40 equina con diferencias significativas en las frecuencias alélicas para distintas razas de caballos (Horin *et al.*, 2004), definiendo la presencia de dos alelos.

1.2.5 Gen del receptor de lentivirus equino 1 (ELR1)

Dentro del análisis de genes potencialmente relevantes en la respuesta inmune de los caballos, y ante la disponibilidad de muestras con estatus sanitario de AIE conocido, se llevó a cabo el estudio del gen que codifica para el receptor de lentivirus (ELR1), identificado como el mediador de la entrada del virus de la anemia infecciosa equina (VAIE) a las células blanco (Zhang *et al.*, 2005).

El receptor ELR1 es una proteína transmembrana de tipo I expresada en la superficie de los macrófagos equinos y está codificada por el gen *TNFRSF14*. Este gen, de aquí en adelante nombrado como *ELR1*, fue mapeado en el cromosoma 2 equino, abarca 6.605 pb, presenta 9 exones y pertenece a la Superfamilia de Receptores del Factor de Necrosis Tumoral (TNFRSF, Zhang *et al.*, 2005). Los receptores del tipo TNF (TNFR) constituyen una superfamilia de proteínas solubles o unidas a membrana caracterizadas por presentar entre uno y seis dominios ricos en cisteínas (en inglés *cysteine-rich domains*, *CRD*). Los motivos canónicos de ELR1 están definidos por enlaces disulfuro intracatenarios generados por residuos de cisteína altamente conservados en las cadenas del receptor (Smith *et al.*, 1994). Los ligandos de la superfamilia del TNF tienen roles pivotaes en la organización y funcionamiento del sistema inmune. Estos comparten motivos estructurales comunes, el dominio de homología del TNF (en

inglés, *TNF homology domain, THD*), que se une a los dominios ricos en cisteínas (CRDs) de los receptores TNF. Los CRDs están compuestos por módulos estructurales, cuya variación en la cantidad y tipo le confiere heterogeneidad a la familia (Bodmer *et al.*, 2002). Los pliegues proteicos del THD y el CRD también se encuentran en otras familias proteicas, abriendo la posibilidad de que el modo de interacción entre el TNF y los receptores del TNF podrían estar conservados en otros contextos. Las actividades de los miembros del TNFRSF comprenden un amplio rango de funciones biológicas tales como proliferación, supervivencia, diferenciación y apoptosis celular, por lo que regulan la fisiología normal del sistema inmune.

En este primer capítulo de la tesis se abordarán los objetivos específicos:

3.2.1 Relevar las bases de datos de secuencias de referencia en busca de SNPs informativos de citoquinas (reportados y validados en especies modelo y en equinos), y diseñar una estrategia de tipificación para los SNPs seleccionados.

3.2.2 Tipificar a nivel genético molecular los SNPs informativos de citoquinas seleccionados en 3.2.1, en los caballos de la raza Criollo Argentino y otras poblaciones equinas.

2. MATERIALES Y MÉTODOS

2. MATERIALES Y MÉTODOS

El presente estudio fue autorizado por el Comité Institucional para el Cuidado y Uso de Animales de Laboratorio (CICUAL) de la Facultad de Ciencias Veterinarias de la Universidad Nacional de La Plata, Argentina (Nro. de Protocolo 89-2-18T).

2.1 Material animal

Para la primera parte de este trabajo de tesis se utilizaron un total de 226 muestras equinas de pelo y/o sangre, obtenidas de distintas poblaciones de la República Argentina. Los equinos muestreados fueron registrados con los siguientes datos cuando estaban disponibles: nombre completo, fecha de nacimiento, sexo, pelaje, condición sanitaria, relaciones de parentesco. Se realizó la reconstrucción de las genealogías a partir de los datos de pedigrí de los registros genealógicos (RRGG) de la Sociedad Rural Argentina (<https://www.sra.org.ar/>) para posterior análisis en conjunto con los resultados obtenidos a nivel molecular. Las muestras de sangre y pelo de los caballos de la raza Criollo Argentino (N = 96) provienen de la Cabaña de Caballos Criollos – Asociación Cooperativa INTA, Leales, Tucumán, y de la Cabaña La Clarita, Pergamino, Buenos Aires. Además, se utilizaron muestras de sangre, pelo y ADN de caballos de la provincia de Chaco (N = 43), mestizos (N = 63) y de la raza Árabe (N = 24).

2.1.1 Toma de muestras y extracción de ADN

Las muestras de sangre entera total se obtuvieron por punción yugular con aguja descartable estéril 18 g en tubos de 5 mL conteniendo anticoagulante (EDTA 6 %) y se mantuvieron refrigeradas para extraer la interfase que contiene la capa leucocitaria (*buffy coat*). La extracción de ADN genómico a partir de las muestras de sangre se realizó mediante la utilización del reactivo DNAzol[®] (Invitrogen Carlsbad, CA, EE. UU.) según las indicaciones del fabricante (Ausubel *et al.*, 1990). Este reactivo contiene detergente de tiocianato de isoguanidina, el cual lisa las células, desnaturaliza las proteínas, hidroliza el ARN y permite la precipitación selectiva del ADN genómico utilizando etanol (Chomczynski *et al.*, 1997).

Las muestras de pelo se recolectaron arrancando de raíz mechones de 30-40 pelos del nacimiento de la cola o crin del caballo, verificando la presencia de folículo piloso. Estas muestras se almacenaron y conservaron a temperatura ambiente en sobres de papel cerrados y rotulados. El ADN se extrajo mediante la técnica de hidróxido de sodio/ácido clorhídrico (NaOH/HCl). Brevemente, el protocolo de extracción consistió en la incubación de 3 a 5 bulbos pilosos en 50 μ L de solución de lisis (NaOH 200 mM) a 95 °C por 10 minutos en tubos de microcentrífuga. Inmediatamente después, esta solución se neutralizó con 50 μ L de una solución 200 mM HCl – TRIS (pH 8,5). Finalmente, los tubos se centrifugaron y el sobrenadante se trasvasó a un nuevo tubo previamente rotulado y las soluciones de ADN resultantes se almacenaron en congelador a -20 °C hasta su utilización.

2.1.2 Evaluación de la calidad y cantidad de ADN

La integridad del ADN genómico se evaluó mediante electroforesis en geles de agarosa al 1 % en buffer TBE 1X (0,089 M Tris; 0,089 M ácido bórico; 0,002 M EDTA). Las corridas electroforéticas se realizaron a 100 voltios durante 35 minutos, sembrando 4 μ L de muestra y 2 μ L de buffer de siembra 6X (azul de bromofenol 0,25 %, xileno cianol 0,25 % y glicerol 30 %). Los geles se revelaron mediante tinción con bromuro de etidio (0,5 μ g/mL) durante al menos 15 minutos y posteriormente se visualizaron las bandas de ADN en transiluminador de luz ultravioleta ($\lambda \approx 300$ nm).

La concentración del ADN extraído tanto de sangre como de folículo piloso, se midió por espectrofotometría a una longitud de onda de 260 nm utilizando los equipos Nanovue Plus (GE Healthcare) y/o Sinergy HT Take3 (Biotek™). La pureza del ADN se evaluó en función del cociente de absorbancias a 260 y 280 nm (A_{260}/A_{280}), considerando aceptables los valores en el rango entre 1,8 y 2,0 (Gallagher, 1994). El ADN extraído se conservó en tubos de tipo Eppendorf en congelador a -20 °C hasta su utilización.

2.2 Amplificación por Reacción en Cadena de la Polimerasa (PCR)

La técnica de Reacción en Cadena de la Polimerasa (en inglés, *PCR*, *Polymerase Chain Reaction*; Mullis y Faloona, 1987) es una herramienta central de la biología molecular que permite obtener un gran número de copias de un fragmento de ADN de interés a partir de una

única copia. Para amplificar la región del gen objetivo se utilizan pequeños oligonucleótidos, denominados cebadores, diseñados específicamente para obtener la región de interés. Estos cebadores flanquean la secuencia blanco y son complementarios a las bases de la misma, por lo que en cada ciclo repetido de amplificación se generan nuevas copias del fragmento de interés. La reacción consiste en ciclos donde se alternan etapas a diferentes temperaturas, en los que las cadenas de la doble hélice se separan (desnaturalización), luego los cebadores se aparean base a base con la secuencia complementaria (hibridación), y finalmente el agregado de nucleótidos a la nueva cadena, en el extremo 3'OH (extensión).

Los cebadores para la amplificación por PCR de los distintos fragmentos de interés se diseñaron utilizando el programa Primer3 (<http://bioinfo.ut.ee/primer3-0.4.0/>) y las secuencias de referencia equinas fueron depositadas en las bases de datos de acceso público disponibles en internet (www.ensembl.org). Los cebadores específicos se diseñaron teniendo en cuenta distintas estrategias metodológicas que involucraron, en algunos casos, PCR semi-anidado con cebadores de referencia o PCR-anidado con cebadores diseñados específicamente a partir de las secuencias obtenidas en la primera amplificación.

2.2.1 Verificación y purificación de los productos de PCR

Los productos de PCR se evaluaron mediante electroforesis en geles de poliacrilamida al 6 % (en una proporción acrilamida: bisacrilamida de 19:1,) en tampón TBE 1X (Tris-Cl 44,5 mM, ácido bórico 44,5 mM, EDTA 1 mM, pH = 8,4). Las corridas electroforéticas se realizaron a 170 voltios durante 30-40 minutos, sembrando 2 μ L del producto de PCR y 1 μ L de buffer de siembra 6X (azul de bromofenol 0,25 %, xileno cianol 0,25 % y glicerol 30 %). Finalizada la electroforesis, los geles se sometieron a tinción con bromuro de etidio (0,5 mg/ml) durante al menos 15 minutos y se visualizaron con transiluminador de luz ultravioleta ($\lambda \approx 300$ nm).

Con el fin de eliminar el exceso de nucleótidos y cebadores procedentes de la reacción de amplificación, los productos de PCR que presentaron una banda única en la verificación con geles de poliacrilamida, fueron sometidos a reacciones de purificación. Se utilizó el kit comercial *AccuPrep® PCR Purification Kit* (BIONEER, Corea) siguiendo las instrucciones del fabricante. Este método de purificación se basa en la adsorción selectiva del ADN a filtros de sílica fijados en columnas de unión. El uso de sales caotrópicas favorece la unión del ADN a la columna de sílica, mientras que proteínas y sales no son adsorbidas y atraviesan la columna de purificación al realizar lavados con buffer.

Alternativamente, se realizó la purificación de los amplificados mediante la técnica de precipitación diferencial utilizando polietilenglicol

(PEG) 8000 (ANEXO 1) al 20 % y NaCl 2,5 M (Lis, 1980). Esta mezcla de PEG precipita selectivamente al ADN dejando cebadores, nucleótidos residuales y productos de PCR truncados en el sobrenadante, el cual es descartado cuidadosamente luego de un paso de centrifugación.

2.3 Métodos de genotipificación

2.3.1 Secuenciación directa

Para la detección de las variantes de los exones 3 y 5 del gen *ELR1*, del exón 2 del gen *ELA-DRB3*, así como la verificación de los alelos del exón 2 del gen *ELA-DRB2* se utilizó la tecnología de secuenciación directa. Se purificaron los productos de PCR positivos para el fragmento del gen analizado, y la secuenciación se realizó directamente a partir de los mismos, utilizando uno de los cebadores empleados para generar el fragmento por PCR. En la secuenciación directa cada producto de PCR se incorpora en una reacción junto con dideoxinucleótidos (ddNTPs) marcados con fluorescencia, que generan fragmentos de distinto tamaño terminado en cada una de las bases de la secuencia. Los fragmentos marcados son entonces sometidos a una electroforesis capilar donde migran en función de su tamaño mientras que su emisión fluorescente será captada por un detector y visualizada bajo la forma de picos en un electroferograma. Las reacciones de secuenciación se realizaron en secuenciador automático ABI3730XLs (Applied Biosystems) en servicio externo (Macrogen Inc., Corea).

2.3.2 Pirosecuenciación®

Este método permite la detección en tiempo real de la polimerización de fragmentos de ADN de hasta 20 bases y consiste en ciclos donde se dispensan de a una por vez las distintas bases. Para llevar adelante esta técnica es necesario amplificar por PCR el fragmento de interés utilizando

uno de los cebadores (*forward* o *reverse*) marcado con biotina y además diseñar un cebador interno de secuenciación de manera que su extremo 3' se hibride una o pocas bases previas al sitio polimórfico.

La Pirosecuenciación[®] se basa en la detección de la señal luminosa generada por una reacción quimioluminiscente que se inicia a partir del pirofosfato inorgánico (PPi) liberado durante la adición de nucleótidos complementarios a la cadena en crecimiento (Figura 1.3). Cada evento de incorporación es acompañado por la liberación de PPi en una cantidad equimolar a la cantidad incorporada de nucleótido. A partir de la generación de PPi se desencadena una cascada enzimática que conduce finalmente a la producción de luz visible. Primero, la enzima ATP sulfurilasa en presencia de adenosina 5' fosfosulfato (APS), convierte el PPi en ATP. Este ATP media la conversión de luciferina en oxiluciferina por parte de la enzima luciferasa, generando luz visible en intensidad proporcional a la cantidad de ATP. La luz producida en la reacción catalizada por luciferasa es detectada por una cámara CCD (en inglés *charge-coupled device*, dispositivo de carga acoplada) y se visualiza en el pirograma como un pico cuya altura es proporcional al número de nucleótidos incorporados. Debido a que el orden de dispensación de nucleótidos es conocido y constante, es posible detectar así la secuencia de la cadena molde. Los nucleótidos no incorporados, así como el ATP no utilizado, son degradados por la enzima apirasa; deteniendo la emisión de luz y restaurando las condiciones para dispensar un nuevo nucleótido en la reacción. En la genotipificación mediante Pirosecuenciación[®], cada

combinación alélica (homocigota, heterocigota) genera un patrón de picos específico comparado con las otras dos variantes (Ahmadian *et al.*, 2000a) permitiendo determinar el genotipo.

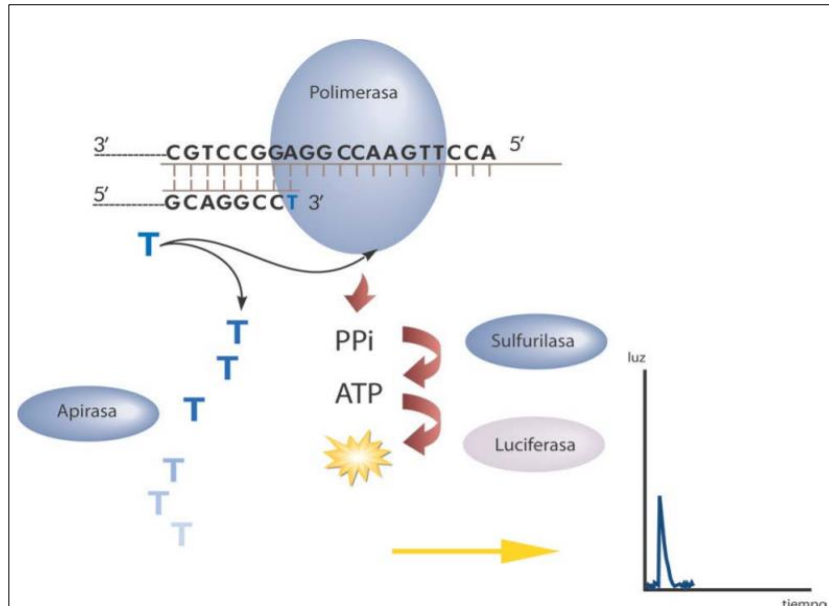


Figura 1.3. Esquema representativo de la cascada de reacciones de la Pirosecuenciación[®].

Los SNPs de los genes *TNF- α* , *IL12p40*, *ELR1* y *ELA-DRA*, se detectaron por Pirosecuenciación[®] según el procedimiento recomendado (Ronaghi, 2001). Previo a la pirosecuenciación se realizó la purificación del producto de PCR biotinilado en equipo PyroMark Prep Workstation (Biotage AB, Uppsala, Suecia).

Para esto se mezclaron 20 μL del amplificado con 40 μL de buffer de *binding*, 17 μL de agua destilada y 3 μL de perlas de sefarosa recubiertas con estreptavidina. Luego, se desnaturalizó el ADN y se realizaron sucesivos lavados con etanol al 70 %, NaOH 0,2 M, agua destilada y buffer de lavado, obteniéndose la cadena simple biotinilada que se utilizó como molde para la pirosecuenciación (Figura 1.4).

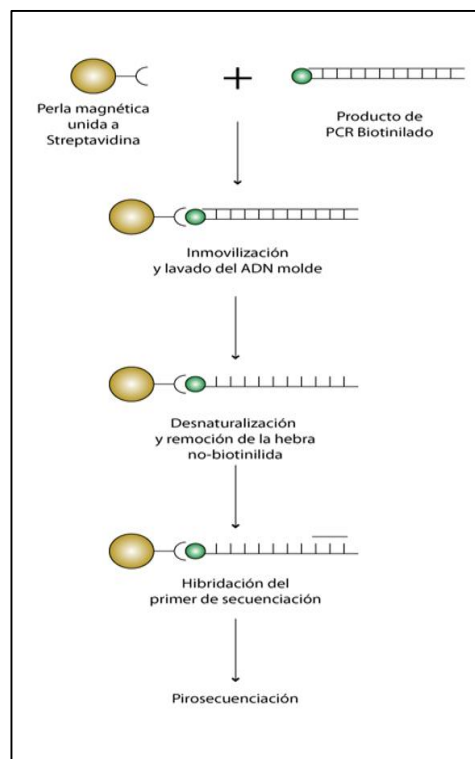


Figura 1.4. Esquema representativo de la purificación del producto de PCR biotinilado en equipo PyroMark Prep Workstation (Biotage AB, Uppsala, Suecia).

Los complejos biotina:estreptavidina fueron incubados en una solución de 45 µL de buffer de *annealing* y 0,3 µM del cebador interno de secuenciación, diseñado específicamente para cada polimorfismo. Finalmente se realizó la reacción de pirosecuenciación en un equipo PSQ® 96 MA (Pyrosequencing AB, Uppsala, Suecia), utilizando el kit PyroMark Gold Q96 (Qiagen, Hilden, Alemania) y el programa PSQ® 96 MA 2.1.1 (Pyrosequencing AB, Uppsala, Suecia) para diseñar el cebador interno y la lectura de los pirogramas obtenidos.

2.3.3 Genotipificación por PCR-RFLP

La técnica de PCR-RFLP (en inglés; *Restriction Fragment Length Polymorphism*) permite la genotipificación de polimorfismos mediante la digestión con endonucleasas de restricción de un producto de PCR. Las enzimas de restricción poseen sitios de reconocimiento y corte específico en la secuencia del ADN, por lo que, cuando ocurre un cambio de una base, como un SNP, se crean o pierden sus sitios de reconocimiento. Como consecuencia, al digerir un fragmento de ADN que presenta variantes alélicas de secuencia, se obtienen fragmentos de ADN de distinto tamaño que al ser visualizados mediante electroforesis en gel permiten detectar los distintos alelos y determinar el genotipo de los individuos analizados.

2.4 Marcadores analizados

2.4.1 Gen *ELA-DRA*

La reacción de PCR se llevó a cabo en un volumen final de 30 μ L conteniendo buffer 1X, 2 mM de $MgCl_2$, 2 mM de dNTPs, 0,6 μ M de cada cebador (Tabla 1.1), 1 U de KAPA *Taq* polimerasa de ADN (Kapa Biosystems, EEUU) y ~50 ng de ADN. El programa de amplificación consistió en un ciclo de 5 minutos de desnaturalización inicial a 95 °C, 32 ciclos de 30 segundos de desnaturalización a 95 °C, 30 segundos de hibridación a 58 °C, 1 minuto de replicación a 72 °C, y una extensión final de 10 minutos a 72 °C.

Tabla 1.1

Cebadores utilizados para la amplificación por PCR del exón 2 del *gen ELA-DRA*.

Cebador	Secuencia	T_{hib} (°C)	Amplicón (pb)
DRABe1 Biot	5'- GGATCCAGGCTGAGTTCTATCTG-3'		
DRABe2	5'- GGCTTAAGAGTGTTGTTGGAGCGCTT-3'	58	249
DRA int	5'- GAGCCTCAAAGCTGG-3'		

Nota. Se indica la temperatura de hibridación (T_{hib}) y el tamaño del amplicón resultante.

2.4.2 Gen *ELA-DRB2*

Para la amplificación del exón 2 del gen *ELA-DRB2* se utilizó la técnica de PCR anidada, utilizando los cebadores DRB2i1F y STR1-2R para la primera ronda y DRB2i1F y DRB2e2R para la segunda ronda (Tabla 1.2). La primera reacción de PCR se llevó a cabo en un volumen final de 25 μ L conteniendo buffer 1X, 2 mM $MgCl_2$ (25 mM), 0,08 mM dNTPs, 0,2 μ M de cada cebador, 0,5 U de Taq polimerasa de ADN (Fermentas, MD, EEUU) y 50 ng de ADN. El programa de amplificación consistió en un ciclo de 5 minutos de desnaturalización inicial a 95 °C, 30 ciclos de 30 segundos de desnaturalización a 95 °C, 30 segundos de hibridación a 58 °C, 60 segundos de replicación a 72 °C, y una extensión final de 10 minutos a 72 °C. Como producto de la primera ronda de PCR se obtuvo un fragmento de ADN de 464 pb.

La reacción de PCR de la segunda ronda se llevó a cabo en un volumen final de 50 μ L conteniendo buffer 1X, 2,0 mM de $MgCl_2$, 0,08 mM de dNTPs, 0,2 μ M de cada cebador, 1 U de Taq polimerasa de ADN (Fermentas, MD, EEUU) y 50 ng del producto de PCR de la primera ronda. El programa de amplificación consistió en un ciclo de 5 minutos de desnaturalización inicial a 95 °C, 30 ciclos de 30 segundos de desnaturalización a 95 °C, 30 segundos de hibridación a 58 °C, 60 segundos de replicación a 72 °C, y una extensión final de 10 minutos a 72 °C. El producto de PCR obtenido tras la segunda ronda de amplificación fue de 293 pb.

Tabla 1.2

Cebadores utilizados para la amplificación por PCR anidada del exón 2 del gen ELA-DRB2.

Cebador	Secuencia	T_{hib} (°C)	Amplicón (pb)
DRB2i1F	5'- CTCTTCTCAGGAGCCCTCTCT - 3'		464
STR 1-2R	5'- TCACACGGTCCCTTGTA CT -3'		
DRB2i1F	5'- CTCTTCTCAGGAGCCCTCTCT - 3'	58	293
DRB2e2R	5'- CTGACGCCGTAGTTGTGTCT - 3'		

Nota. Se indica la temperatura de hibridación (T_{hib}) y el tamaño del amplicón resultante.

La genotipificación del exón 2 de *ELA-DRB2* se realizó mediante la técnica de PCR - RFLP explicada previamente (Figura 1.5). Para la misma, 10 µL del producto de la segunda ronda de PCR se incubaron *overnight* a una temperatura de 37 °C, con 2,5 U de las enzimas de restricción *HaeIII*, *MspI*, *RsaI*, en un volumen final de 20 µL. Los fragmentos obtenidos tras la digestión fueron sometidos a electroforesis en geles de poliacrilamida al 8 %, utilizando bromuro de etidio para su revelado y visualizados con transiluminador UV. En la figura 1.5 y en la tabla 1.3 se muestran los patrones de restricción esperados para cada variante alélica.

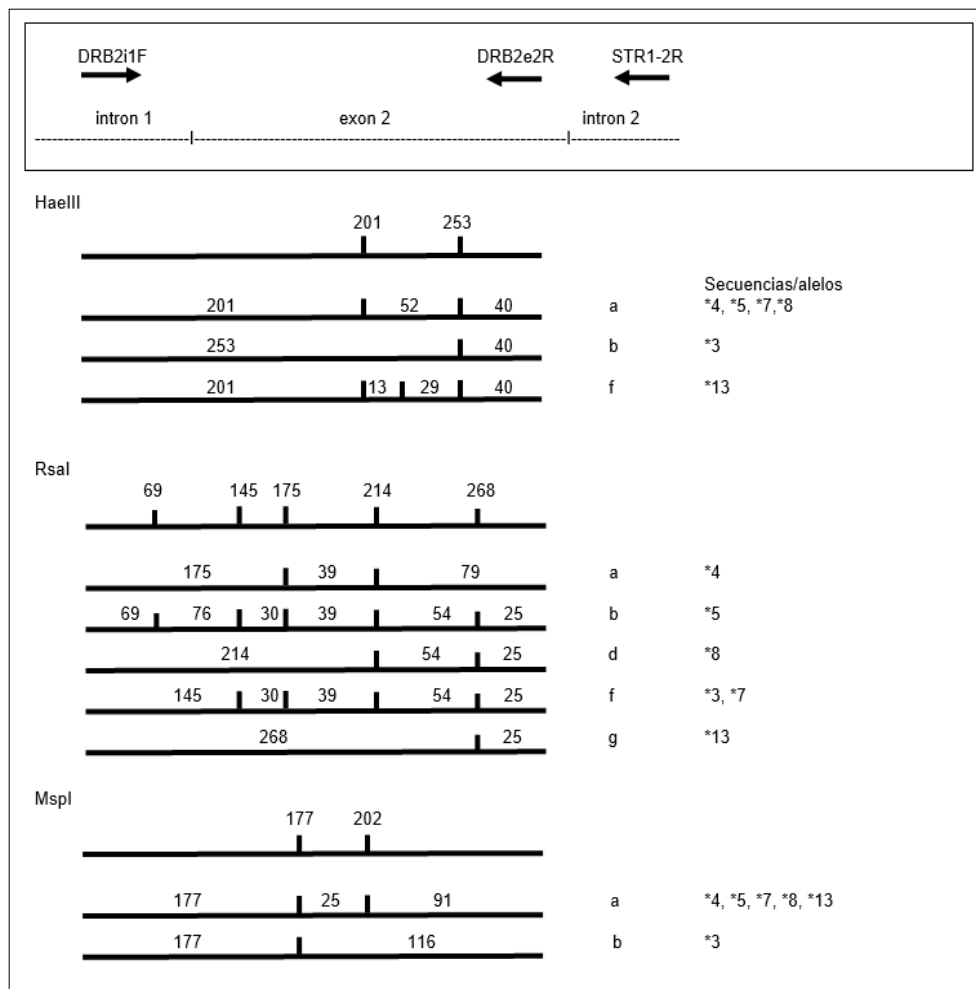


Figura 1.5. Patrones de restricción posibles para la definición de alelos del exón 2 de *ELA-DRB2*.

Tabla 1.3

Definición de alelos del gen ELA-DRB2 por patrones RFLP.

Alelo	No. Acceso Genebank	<i>HaeIII</i>	<i>RsaI</i>	<i>MspI</i>
ELA-DRB*3	L25646	b	f	b
ELA-DRB*4	L77079	a	a	a
ELA-DRB*5	L76978	a	b	a
ELA-DRB*7	L76973	a	f	a
ELA-DRB*8	L76975	a	d	a
ELA-DRB*13	AF170067	f	g	a

2.4.3 Gen *ELA-DRB3*

Para la amplificación del exón 2 del gen *ELA-DRB3* se utilizó la técnica de PCR anidada, utilizando los cebadores DRB3i1F y STR3-4R para la primera ronda y DRB3i1F y DRB3e2R para la segunda ronda (Tabla 1.4). La primera reacción de PCR se llevó a cabo en un volumen final de 25 μ L conteniendo buffer 1X, 2 mM $MgCl_2$ (25 mM), 0,08 mM dNTPs, 0,2 μ M de cada cebador, 0,5 U de Taq polimerasa de ADN (Fermentas, MD, EEUU) y 50 ng de ADN. El programa de amplificación consistió en un ciclo de 5 minutos de desnaturalización inicial a 95 °C, 30 ciclos de: 30 segundos de desnaturalización a 95 °C, 30 segundos de hibridación a 58 °C, 60 segundos de replicación a 72 °C, y una extensión final de 10 minutos a 72 °C. Como producto de la primera ronda de PCR se obtuvo un fragmento de ADN de 443 pb.

La segunda ronda de PCR se llevó a cabo en un volumen final de 50 μ L conteniendo buffer 1X, 2,0 mM de $MgCl_2$, 0,08 mM de dNTPs, 0,2 μ M de cada cebador, 1 U de Taq polimerasa de ADN (Fermentas, MD, EEUU) y 50 ng del producto de PCR de la primera ronda. El programa de amplificación consistió en un ciclo de 5 minutos de desnaturalización inicial a 95 °C, 10 ciclos de 30 segundos de desnaturalización a 95 °C, 30 segundos de hibridación a 58 °C, 60 segundos de replicación a 72 °C, y una extensión final de 10 minutos a 72 °C. El producto de PCR obtenido tras la segunda ronda de amplificación fue de 318 pb.

Tabla 1.4

Cebadores utilizados para la amplificación por PCR anidada del exón 2 del gen ELA-DRB3.

Cebador	Secuencia	T_{hib} (°C)	Amplicón (pb)
DRB3i1F	5'- GTCGCCATCCATCTCTTCTCG- 3'	58	443
DRB3e2R	5'-AGACACA AACTACGGCGTCAG- 3'		
DRB3i1F	5'- GTCGCCATCCATCTCTTCTCG- 3'		
STR3-4R	5'- CTCGCTCACAGTCCTACACA -3'		318

Nota. Se indica la temperatura de hibridación (T_{hib}) y el tamaño del amplicón resultante.

2.4.4 Gen de la subunidad p40 de la interleuquina 12 (IL12p40)

La amplificación del exón 5 del gen *IL12p40* se realizó mediante PCR en termociclador MaxyGene™ (Axygene Scientific, CA, EEUU) utilizando un cebador directo específicamente diseñado para amplificar un fragmento de 201 pb (Tabla 1.5), junto con el cebador reverso biotinilado reportado previamente por Horřin *et al.*, (2004), (Figura 1.6).

Tabla 1.5

Cebadores utilizados para la amplificación por PCR del exón 5 del gen IL12p40.

Cebador	Secuencia	T_{hib} (°C)	Amplión (pb)
IL12Be6F	5'-CACGACCTTACTCCAAAGCCCA-3'		
IL12Be6-R-Biot	5'-CTTGCCCTGGACCTGAATAGAG-3'	58	205
IL12Be6-int	5'-TTCCATAGTCAAACCAG-3'		

Nota. Se indica la temperatura de hibridación (T_{hib}) y el tamaño del amplicón resultante.

La reacción de PCR se llevó a cabo en un volumen final de 30 µL conteniendo buffer 1X, 2 mM de MgCl₂, 2 mM de dNTPs, 0,6 µM de cada cebador, 1 U de *Taq* polimerasa de ADN (Fermentas, MD, EEUU) y ~50 ng de ADN. El programa de amplificación consistió en un ciclo de 5 minutos de desnaturalización inicial a 95 °C, 32 ciclos de 30 segundos de desnaturalización a 95 °C, 30 segundos de hibridación a 58 °C, 1 minuto de replicación a 72 °C, y una extensión final de 10 minutos a 72 °C.

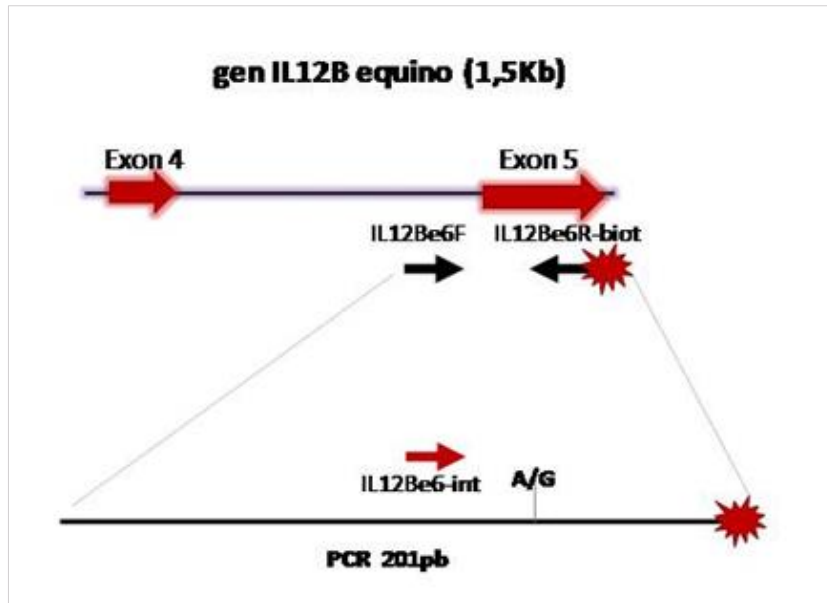


Figura 1.6. Diseño experimental para amplificar el exón 5 de *IL12p40*. Se muestra el SNP de interés A/G, el par de cebadores utilizado en la PCR (flechas negras) y el cebador interno utilizado en la pirosecuenciación (flecha roja).

2.4.5 Gen del factor de necrosis tumoral alfa (TNF- α)

La amplificación de la región promotora del gen *TNF- α* se realizó mediante la estrategia de PCR anidada, una variante de la PCR convencional, que tiene mayor sensibilidad y especificidad. Esta estrategia consistió en dos rondas de amplificación (Figura 1.7), cada una con dos juegos de cebadores distintos (Tabla 1.6), uno de ellos marcado con biotina (TNF-E1R-Biot) para obtener el producto de cadena simple en el siguiente paso de purificación. En la primera ronda se utilizaron un par de cebadores externos (TNF α -F2, TNF-P1R1), a partir de los que se

obtuvo un producto de PCR de 320 pb, donde se encuentra la secuencia de interés. En la segunda ronda de amplificación, utilizando este producto de PCR como molde y los cebadores internos (TNF α -F2, TNF-E1R1-Biot), se obtiene la región específica (197 pb) que se utilizó como molde para la pirosecuenciación.

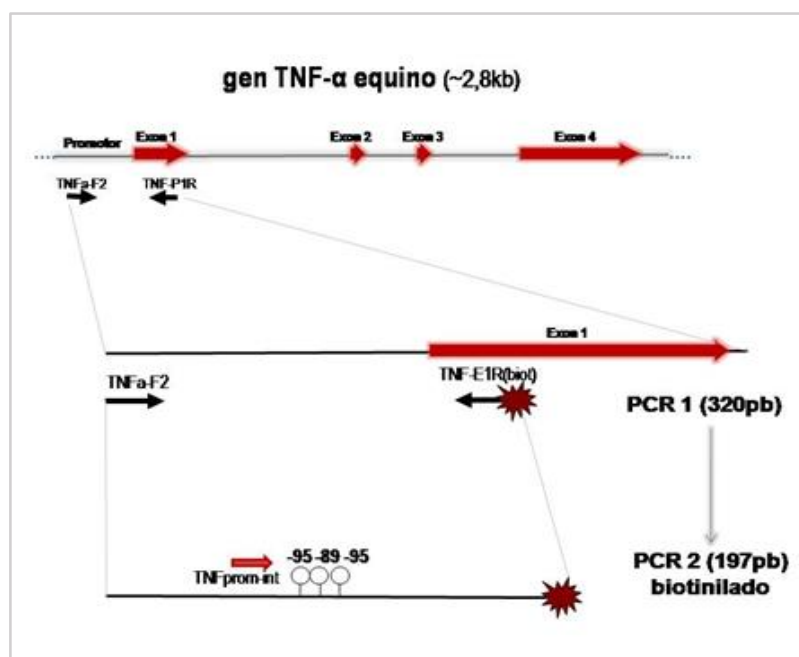


Figura 1.7. Diseño experimental utilizado para amplificar la región promotora del TNF- α . Se representan los 3 SNPs de interés (círculos), así como los pares de cebadores utilizados para cada ronda de PCR (flechas negras) y el cebador interno utilizado en la pirosecuenciación.

La reacción de PCR de la primera ronda se llevó a cabo en un volumen final de 25 μ L conteniendo buffer 1X, 2 mM MgCl₂, 0,6 mM

dNTPs, 0,4 μ M de cada cebador, 0,5 U de *Taq* polimerasa de ADN (Fermentas, MD, EEUU) y 50 ng de ADN. El programa de amplificación consistió en un ciclo de 5 minutos de desnaturalización inicial a 95 °C, 20 ciclos de 30 segundos de desnaturalización a 95 °C, 30 segundos de hibridación a 58 °C, 45 segundos de replicación a 72 °C, y una extensión final de 10 minutos a 72 °C.

Tabla 1.6

Cebadores utilizados para la amplificación por PCR anidada del gen TNF- α .

Cebador	Secuencia	T_{hib} (°C)	Amplicón (pb)
TNF α -F2	5'-GCCAGGAGAGAGAGAAGCAA-3'		
TNF α -P1R1	5'-CGATCACCCCAAAGTGCAAG-3'		320
TNF α -F2	5'-GCCAGGAGAGAGAGAAGCAA-3'	58	
TNF E1R Biot	5'-CTCCTCTGCCAGCTCCAC-3'		197
TNFprom-int	5'-gAAATAACCTCTCAgA-3'		

Nota. Se indica la temperatura de hibridación (T_{hib}) y el tamaño del amplicón resultante.

La reacción de PCR de la segunda ronda se llevó a cabo en un volumen final de 30 μ L conteniendo buffer 1X, 1,25 mM de MgCl₂, 0,13 mM de dNTPs, 0,4 μ M de cada cebador, 0,6 U de *Taq* polimerasa de ADN (Fermentas, MD, EEUU) y 50 ng del producto de PCR de la primera

ronda. El programa de amplificación consistió en un ciclo de 3 minutos de desnaturalización inicial a 95 °C, 45 ciclos de 30 segundos de desnaturalización a 95 °C, 45 segundos de hibridación a 60 °C, 45 segundos de replicación a 72 °C, y una extensión final de 10 minutos a 72 °C.

2.4.6 Gen del receptor de lentivirus equino 1 (ELR1)

La búsqueda de polimorfismos en los exones 3, 4 y 5 del gen ELR1 se realizó por dos técnicas distintas: Pirosecuenciación[®] y secuenciación directa. En primer lugar, debido a la escasa información disponible en ese momento en las bases de datos en cuanto a polimorfismos reportados para este gen, se seleccionó el SNP G > A ubicado en la posición 47311347 de ECA2 (*rs68535250*), ya que se localiza en una región codificante (exón 4) y genera un cambio de aminoácido en la proteína resultante (Figura 1.8). La reacción de PCR para el exón 4 se llevó a cabo en un volumen final de 30 µL conteniendo buffer 1X, 2,5 mM de MgCl₂, 2 mM de dNTPs, 0,6 µM de cada cebador (Tabla 1.7), 0,6 U de *Taq* polimerasa de ADN (Genbiotech, Argentina) y 50 ng de ADN. El programa de amplificación consistió en un ciclo de 3 minutos de desnaturalización inicial a 95 °C, 45 ciclos de 30 segundos de desnaturalización a 95 °C, 45 segundos de hibridación a 60 °C, 45 segundos de replicación a 72 °C, y una extensión final de 10 minutos a 72 °C.

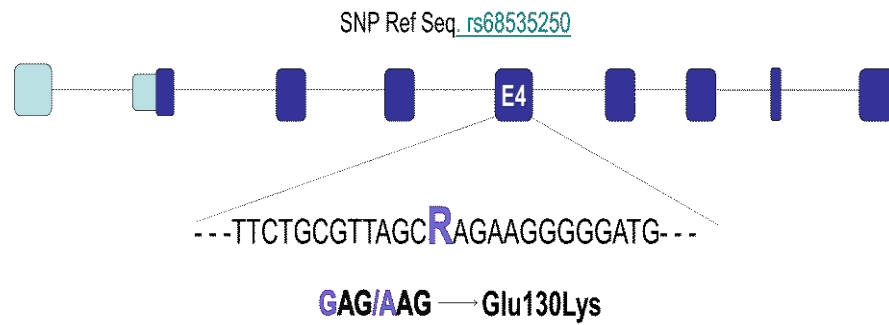


Figura 1.8. Representación esquemática de la estructura del gen equino *ELR1*. Los exones codificantes se indican en color azul. La secuencia corresponde a las bases que rodean al SNP G > A en la posición 47311347 de ECA2.

Para tipificar este polimorfismo se utilizó la técnica de pirosecuenciación para la cual se diseñaron específicamente tanto el par de cebadores para la amplificación por PCR, así como el cebador interno requerido para la tipificación (Tabla 1.7).

Tabla 1.7

Cebadores utilizados para la amplificación por PCR del exón 4 del gen ELR1.

Cebador	Secuencia	T_{hib} (°C)	Amplicón (pb)
ELR1F	5'-CGAGCACGGAAAACACTGAG-3'		
ELR1-R-biot	5'-TCTCTCCTGTACCCTCTGGC-3'	58	125
ELR1-617-int	5'-CACTTCTGCGTTA-3		

Nota. Se indica la temperatura de hibridación (T_{hib}) y el tamaño del amplicón resultante.

Para la tipificación de los exones 3 y 5, se utilizó una estrategia de secuenciación directa a partir de los productos de PCR. Se diseñaron dos pares de cebadores específicos (Tabla 1.8) para cada exón a partir de la secuencia de referencia ENSECAG00000008751, con el programa Primer3 v. 0.4.0 (<http://bioinfo.ut.ee/primer3-0.4.0/primer3>).

Tabla 1.8

Cebadores utilizados para la amplificación por PCR de los exones 3 y 5 del gen ELR1.

Cebador	Secuencia	T_{hib} (°C)	Amplicón (pb)
ELR1e3F	5'- CCCATTTACAGAGGCTGAT-3'	58	196
ELR1e3R	5'- CCTACCTGGACTGCATTTGG-3'		
ELR1e5F	5'- ATTGTCCCAGTGTCCCTCAG-3'	58	476
ELR1e5R	5'- CGGCTCACCTCTCTCCTGTA-3'		

Nota. Se indica la temperatura de hibridación (T_{hib}) y el tamaño del amplicón resultante

La reacción de PCR para el exón 3 se llevó a cabo en un volumen final de 30 µL conteniendo buffer 1X, 2,5 mM de MgCl₂, 2 mM de dNTPs, 0,6 µM de cada cebador, 0,6 U de *Taq* polimerasa de ADN (Genbiotech, Argentina) y 50 ng de ADN.

La reacción de PCR para el exón 5 se llevó a cabo en un volumen final de 30 µL conteniendo buffer 1X, 1,67 mM de MgCl₂, 0,2 mM de dNTPs, 0,2 µM de cada cebador, 1 U de *Taq* polimerasa de ADN (Genbiotech, Argentina) y 50 ng de ADN. El programa de amplificación para los exones 3 y 5 consistió en un ciclo de 5 minutos de desnaturalización inicial a 95 °C, 35 ciclos de 30 segundos de desnaturalización a 95 °C, 45 segundos de hibridación a 58 °C, 1 minuto de replicación a 72 °C, y una extensión final de 10 minutos a 72 °C.

2.5 Análisis bioinformático

Las secuencias de ADN obtenidas se compararon con aquellas reportadas en el Genbank (*National Institute of Health, USA*), utilizando el programa BLAST® (*Basic Local Alignment Search Tool - National Center for Biotechnology Information, U.S. National Library of Medicine*, <https://blast.ncbi.nlm.nih.gov/Blast.cgi>), para confirmar su origen. Las secuencias de ADN se alinearon con el programa CLUSTAL Omega - Multiple sequence alignment (*European Bioinformatics Institute*, <https://www.ebi.ac.uk/Tools/msa/clustalo/>), para poder analizar la ubicación de los sitios polimórficos. Los SNPs detectados se compararon con los depositados en las bases de datos públicas (ENTREZ SNPdb, <http://www.ncbi.nlm.nih.gov/snp>).

En la comparación y análisis de las secuencias obtenidas para los genes *ELA-DRB2* y *ELA-DRB3*, se utilizaron las secuencias de las bases de datos del *National Center for Biotechnology Information* (NCBI) GeneBank (<http://www.ncbi.nlm.nih.gov/>) y de la base de datos *Immuno Polymorphism Database* (IPD) desarrollada por el grupo *HLA Informatics* del Instituto de Investigación Anthony Nolan y el Instituto Europeo de Bioinformática (<https://www.ebi.ac.uk/ipd/>). Las secuencias obtenidas a partir del producto de PCR secuenciado, se analizaron en forma comparativa con los alelos de las bases de datos, se depuraron las secuencias de baja calidad y se detectaron los polimorfismos (SNPs – *Single Nucleotide Polymorphism*) por observación directa de los

cromatogramas para detectar la posición heterocigota. Las variaciones de secuencia se registraron según el código de ambigüedades IUPAC para nucleótidos (Johnson, 2010).

En las secuencias de los genes *DRB*, la estimación de los haplotipos se realizó en dos etapas: primero, de forma intragénica, es decir, se contabilizaron todos los SNPs y se identificaron así las posiciones variables dentro de la secuencia del exón 2. Tomando como referencia las secuencias de los alelos reportados, se identificaron las combinaciones presentes en los animales homocigotas y de los heterocigotas en un solo locus, resolviendo así las primeras combinaciones haplotípicas de la población, para luego identificar y asignar los alelos a cada individuo.

El análisis predictivo del efecto potencial de los cambios de aminoácidos en la secuencia predicha del receptor ELR se realizó con el software PROVEAN (en inglés *Protein Variation Effect Analyzer*; Choi y Chan, 2015). Para valorar el efecto neutral en la función biológica de la proteína se utilizó el valor de corte por defecto (*cut-off score* = -2,5).

2.6. Medidas de variabilidad genética

2.6.1 Frecuencias alélicas y equilibrio de Hardy-Weinberg

La composición genética de una población analizada mediante marcadores moleculares se describe mediante la determinación de los alelos presentes y sus frecuencias relativas en la muestra de la población. El cálculo de las frecuencias alélicas se hace en base al recuento directo de los alelos presentes para cada polimorfismo/SNP. La ocurrencia de cierto genotipo en la población, es decir, la frecuencia genotípica, se determina de la misma manera.

El análisis del equilibrio Hardy-Weinberg (en inglés, *Hardy-Weinberg Equilibrium*, HWE) se realiza para determinar si la población se encuentra en equilibrio de frecuencias génicas y genotípicas. La ley se aplica para poblaciones diploides con reproducción sexual y asume los siguientes supuestos:

- Generaciones discretas, es decir que no exista solapamiento entre las mismas
- Apareamientos al azar
- Tamaño poblacional infinito
- Ausencia de mutación, selección y migración

El principal postulado de la ley es que, bajo las condiciones antes estipuladas y luego de un ciclo de apareamiento al azar, las frecuencias

genotípicas de un locus individual se fijarán en un valor de equilibrio particular, es decir, permanecerán constantes. Asimismo, esta ley especifica que dichas frecuencias de equilibrio se pueden representar en función de las frecuencias alélicas para ese locus mediante la siguiente ecuación:

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

Donde p y q son las frecuencias alélicas para el sistema bialélico considerado, p^2 y q^2 son las frecuencias de los genotipos homocigotas y $2pq$ la frecuencia genotípica heterocigota. Esta fórmula puede ser generalizada para más de dos alelos, siempre cumpliendo que la sumatoria de todas las frecuencias, tanto alélicas como genotípicas, sea igual a 1.

Los análisis estadísticos para determinar presencia o ausencia de equilibrio Hardy-Weinberg se basan en la comparación de las frecuencias genotípicas observadas con las que se esperarían en equilibrio. Si los valores observados y esperados no difieren significativamente, se concluye que la población se encuentra en equilibrio para ese locus en ese momento particular (Ayala, 1982; Hartl, 1988).

2.6.2 Heterocigosis

Un indicador de la variación genética de una población es la Heterocigosis (H), la que representa la probabilidad de que un individuo

presente dos alelos diferentes (es decir, sea heterocigota) para ese locus. Está fuertemente relacionada con el número de alelos presentes y su frecuencia en la población. La heterocigosis es una medida de la variación genética de la población en estudio y adopta valores entre 0 y 1. Cuanto más próxima sea a 1, mayor será la diversidad genética presente y mayor será es la información que aporte el marcador acerca de la transmisión de alelos a la siguiente generación.

Mediante el recuento directo de los genotipos heterocigotas en el total de la población se obtiene la proporción observada de los mismos, es decir, la heterocigosis observada (H_o). La heterocigosis esperada (H_e) de un locus se calcula a partir de las frecuencias alélicas aplicando la siguiente ecuación:

$$H_e = 1 - \sum_{i=1}^k x_i^2$$

Donde x_i es la frecuencia del alelo i y k es el número de alelos (Nei, 1973). La H_e es la heterocigosis que habría en la población si los individuos se aparean al azar y existiera equilibrio de Hardy-Weinberg.

La comparación de H_o y H_e es un indicador de si la población se encuentra en equilibrio Hardy-Weinberg para ese locus; si es así entonces H_e coincidirá con H_o . En el caso de no encontrar tal coincidencia entre los valores de heterocigosis, la población en estudio presenta un

exceso ($H_o > H_e$) o un déficit ($H_o < H_e$) de heterocigotas, que provoca la desviación de los valores esperados (Ayala, 1982; Hartl, 1988).

2.6.3 Estadísticos F de Wright

Los estadísticos F jerárquicos de Sewall Wright (1969), también llamados índices de fijación, describen el nivel de heterocigosis en una población, más específicamente cuánto ésta se reduce en relación a lo esperado según el equilibrio Hardy-Weinberg.

La endogamia puede medirse entonces en términos de la reducción en la proporción de genotipos heterocigotas, es decir, en función de H. De esta manera los estadísticos F también constituyen una medida de la endogamia según la ecuación:

$$F = (H_o - H_e) / H_o$$

En la estructura de una población existen tres niveles distintos de complejidad: el individual (I), el subpoblacional (S) y el de la población total (T). Donde:

H_i : heterocigosis de un individuo en la población.

H_s : heterocigosis esperada de un individuo en una subpoblación que presenta cruzamientos al azar.

H_T : heterocigosis esperada de un individuo en la población total (suma de todas las subpoblaciones) con cruzamientos al azar, es decir, en equilibrio Hardy-Weinberg.

El coeficiente de endogamia F_{IS} mide la reducción de la heterocigosis de un individuo debido a la existencia de apareamientos no al azar (endogamia) dentro de su subpoblación.

$$F_{IS} = 1 - (H_o / H_s)$$

El F_{IS} es una correlación estadística que toma valores desde -1 hasta +1. Un F_{IS} con signo negativo sugiere un exceso de heterocigotas, mientras que aquel mayor que 0 y por lo tanto de signo positivo, indica una deficiencia de estos (o exceso de homocigotas). La magnitud de esta tendencia es expresada por el valor absoluto del índice, según los siguientes rangos:

0	< F_{IS} < 0,05	bajo
0,06	< F_{IS} < 0,15	medio
0,16	< F_{IS} < 0,25	alto
0,25	< F_{IS}	muy alto

El efecto de la subdivisión de la población se mide por un parámetro denominado “índice de fijación” (F_{ST}), que es la reducción promedio en la heterocigosis de una subpoblación debido a la deriva génica:

$$F_{ST} = 1 - H_s / H_T$$

Se utiliza como medida de distancia genética (Wright, 1969; Ayala, 1982; Hartl, 1988).

2.7 Análisis estadístico poblacional

Las frecuencias alélicas y genotípicas de los marcadores genéticos utilizados se calcularon con el software GENEPOP 4.0 (Rousset 1997, 2008). La diversidad genética se estimó mediante el número de alelos (N_a), la heterocigosis observada (h_o) y esperada (h_e) para cada locus y la heterocigosis promedio esperada sobre todos los locus (H_e). El número de alelos se determinó por conteo directo, mientras que la h_e y h_o se estimaron de acuerdo a la fórmula propuesta por Nei (1978) usando el programa de análisis genéticos poblacionales ARLEQUIN 3.5 (Schneider *et al.*, 2000). Las posibles desviaciones del equilibrio de Hardy-Weinberg (HWE) se estimaron para cada locus y para cada grupo mediante el índice F_{IS} y el test exacto de Fisher (Wright 1965) incluidos en el programa GENEPOP 4.0. El desequilibrio de ligamiento (fase de ligamiento) entre los marcadores genéticos se reconstruyó por el método bayesiano implementado en el software PHASE v2.1.1 (Li y Stephens, 2003;

Crawford *et al.*, 2004) utilizando las opciones predeterminadas. Las fases y el desequilibrio de ligamiento entre los loci se visualizaron por medio del programa HAPLOVIEW v3.31 (Barrett *et al.*, 2005).

2.8 Análisis filogenético

El análisis filogenético de las secuencias de los exones 3 y 5 se realizó con el paquete informático MEGA versión 7.0.14 (Tamura *et al.*, 2013), utilizando el método de *Neighbour Joining* y 1000 repeticiones (*bootstrap* = 1000).

2.9 Análisis de ligamiento

Para los marcadores *ELA-DRA*, *ELR1* y *TNF- α* se realizó la reconstrucción de las fases de ligamiento de los SNPs mediante el programa PHASE (Li & Stephens, 2003), utilizando las opciones por defecto ($p = q = 90\%$). Una vez determinadas las fases de ligamiento más probables se llevó a cabo el análisis de desequilibrio de ligamiento (LD) y la visualización de las combinaciones haplotípicas utilizando el programa HAPLOVIEW v 4.2 (Barrett *et al.*, 2005).

3. RESULTADOS

3. RESULTADOS

3.1 Detección de SNPs en los genes inmunes

3.1.1 Gen *ELA-DRA*

Para el análisis de los SNPs del gen *ELA-DRA* se tipificó un grupo de 18 caballos Criollos emparentados conformado por un padrillo y 17 productos (medio hermanos paternos). La pirosecuenciación del marcador *ELA-DRA* (Figura 1.9) permitió detectar tres alelos: *ELA-DRA*0101*, *ELA-DRA*0201* y *ELA-DRA*0301*, cuyas frecuencias se muestran en la Tabla 1.10. El valor de heterocigosis observada resultó de 0,330.

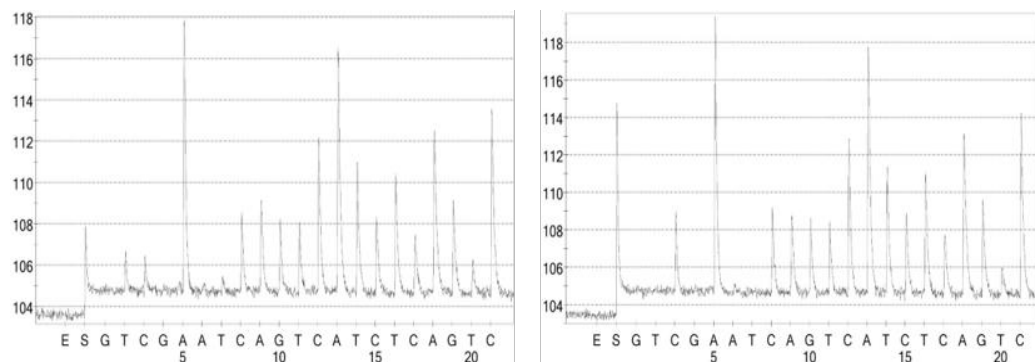


Figura 1.9. Pirogramas de los genotipos detectados para el exón 2 del gen *ELA-DRA* en caballos Criollo Argentino. Sobre el eje horizontal se indica el orden de adición de los nucleótidos. El eje vertical indica la intensidad del pico lumínico correspondiente a la base incorporada.

3.1.2 Genes *ELA-DRB2* y *ELA-DRB3*

La tipificación de los alelos del exón 2 del gen *ELA-DRB2* se realizó por PCR-RFLP (Francino *et al.*, 1997; Díaz *et al.*, 2001; 2003). Posteriormente, se obtuvieron las secuencias de los animales que presentaron diferentes alelos definidos por RFLP (ver tabla 1.7 en Materiales y Métodos). Los mismos productos fueron digeridos con las tres enzimas, de manera que en el gel se pueden leer los diferentes patrones para cada animal y de esta manera definir el alelo presente (Figura 1.10). La nomenclatura empleada para los patrones se detalla en los Materiales y Métodos del Capítulo 1 (Figura 1.8). La secuenciación de los animales que presentaron patrones diferentes permitió verificar la presencia de los sitios de corte de las enzimas, y de los polimorfismos reportados en diferentes posiciones del exón 2. Finalmente, el análisis de estas secuencias permitió comprobar la presencia en los caballos Criollos de los alelos *DRB2*7*, *DRB2*3*, y de tres variantes denominadas provisoriamente *DRB2*3.2*, *DRB2*4.2* y *DRB2*7.2*.

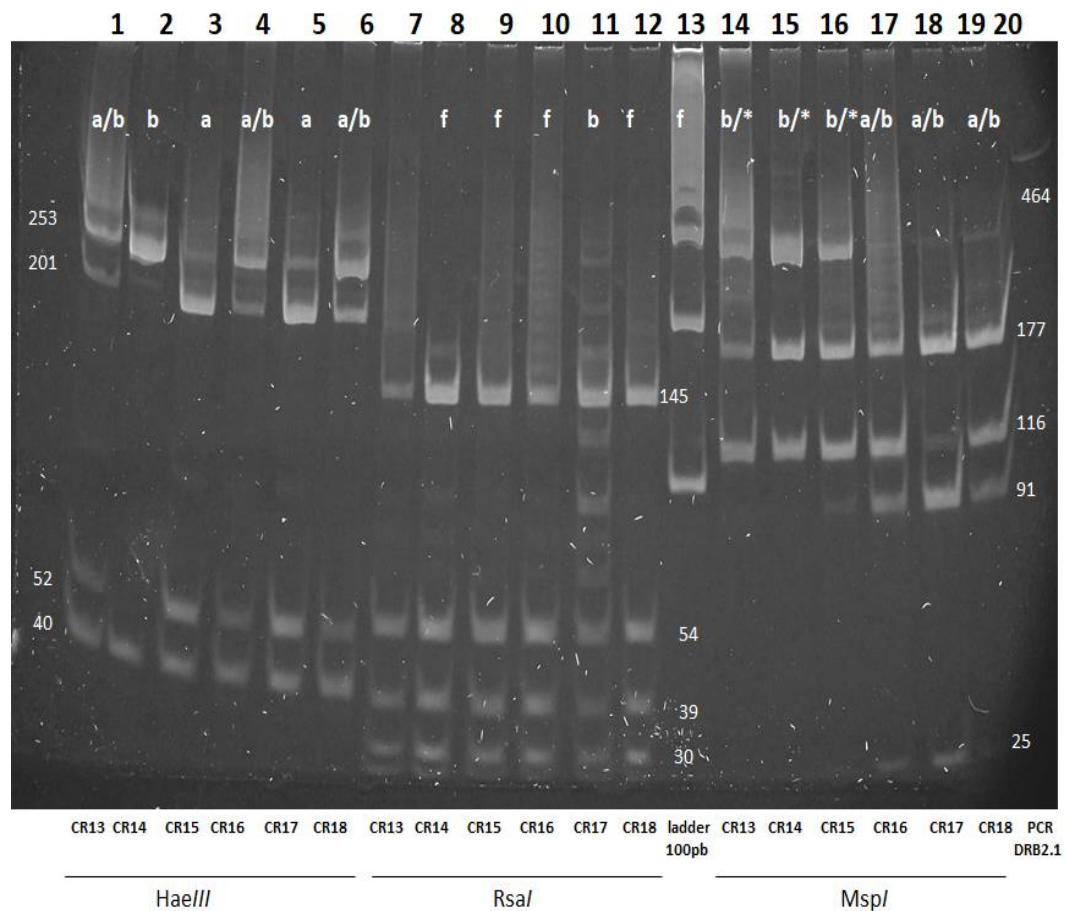


Figura 1.10. Fragmentos de PCR del exón 2 de *ELA-DRB2* analizados por PCR-RFLP con las enzimas *HaeIII*, *RsaI* y *MspI*. En la parte superior se indican los patrones de RFLP detectados; los números a los lados de la figura indican los tamaños de los fragmentos obtenidos para cada patrón.

Dada la complejidad en la lectura de los patrones de restricción observados, la tipificación del exón 2 de *ELA-DRB3* se realizó por secuenciación directa. Las secuencias obtenidas corresponden a 23 caballos Criollo Argentino pertenecientes a grupos familiares de dos y tres generaciones. El análisis de las secuencias por medio de alineamientos

múltiples con secuencias de referencia de los alelos conocidos de *ELA-DRB3*, reveló la presencia de 26 sitios variables. Las posiciones de los sitios variables, los tipos de cambios de bases y los cambios de aminoácidos en cada una se indican en la Tabla 1.9.

Tabla 1.9

Posiciones de nucleótido variables en las secuencias del exón 2 de *ELA-DRB3* de los caballos Criollo Argentino.

Posición genómica Eca20		34274390	34274392	34274394	34274407	34274414	34274417	34274418	34274423	34274428	34274429	34274436	34274442	34274443	34274484	34274521	34274523	34274524	34274573	34274574	34274586	34274587	34274588	34274594	34274605	34274614	34274615
Tipo de cambio		G/	C/	G/	T/	A/	A/	G/	A/	G/	G/	G/	T/	A/	G/	G/	A/	G/	C/	A/	T/	A/	C/	A/	C/	A/	G/
ELA-DRB*9		G	C	G	T	A	A	G	A	G	G	G	T	A	G	G	A	G	C	A	T	A	C	A	C	A	G
CR11	DRB*9	G	C	G	T	A	A	G	A	G	G	G	T	A	G	G	A	G	C	A	T	A	C	A	C	A	G
CR16	DRB*9	G	C	G	T	A	A	G	A	G	G	G	T	A	G	G	A	G	C	A	T	A	C	A	C	A	G
CR78	DRB*18	G	C	G	T	A	A	G	A	G	G	G	T	A	G	G	A	G	C	A	T	A	C	A	C	G	A
CR77	DRB*24	C	C	G	C	A	T	C	A	G	G	G	T	A	C	G	A	G	G	G	T	A	C	A	C	A	G
CR81	DRB*25	C	C	G	T	A	T	C	A	G	G	G	T	A	G	C	G	A	G	G	G	T	G	A	C	A	G
CR01	DRB*26	C	T	G	T	A	T	C	G	G	G	G	T	A	C	C	G	G	G	G	T	G	A	C	A	A	G
CR05	DRB*27	C	C	G	C	A	T	C	G	G	G	C	C	T	C	C	G	G	G	G	T	A	C	A	C	A	G
CR06	DRB*28	C	T	G	C	A	T	C	G	T	G	G	T	T	C	C	G	G	G	G	T	G	A	C	A	A	G
ELA-DRB*1		G	C	G	T	A	A	G	A	G	G	G	T	A	G	C	G	A	C	A	T	A	C	A	C	A	G
ELA-DRB*6		C	T	G	C	G	T	C	G	T	A	C	C	T	C	G	A	G	G	G	T	G	G	G	A	A	G
CR04a	DRB*19	C	T	G	C	G	T	C	G	T	A	C	C	T	C	G	A	A	G	G	T	G	G	G	A	A	G
CR75	DRB*20	C	T	G	C	G	T	C	G	T	A	C	C	T	C	C	G	G	G	G	T	G	G	G	A	A	G
CR04b	DRB*11	C	T	G	C	G	T	C	G	T	A	C	C	T	C	C	G	A	G	G	T	G	A	A	A	A	G
ELA-DRB*11		C	T	G	C	G	T	C	G	T	A	C	C	T	C	C	G	A	G	G	T	G	A	S	A	A	G
Eapr-DRB*2		C	T	G	C	G	T	C	G	T	A	C	C	T	C	C	G	A	G	G	T	G	A	A	A	A	G
CR02	DRB*11	C	T	G	C	G	T	C	G	T	A	C	C	T	C	C	G	A	G	G	T	G	A	A	A	A	G
CR03	DRB*11	C	T	G	C	G	T	C	G	T	A	C	C	T	C	C	G	A	G	G	T	G	A	A	A	A	G
CR07	DRB*11	C	T	G	C	G	T	C	G	T	A	C	C	T	C	C	G	A	G	G	T	G	A	A	A	A	G
CR12	DRB*11	C	T	G	C	G	T	C	G	T	A	C	C	T	C	C	G	A	G	G	T	G	A	A	A	A	G
CR13	DRB*11	C	T	G	C	G	T	C	G	T	A	C	C	T	C	C	G	A	G	G	T	G	A	A	A	A	G
CR14	DRB*11	C	T	G	C	G	T	C	G	T	A	C	C	T	C	C	G	A	G	G	T	G	A	A	A	A	G
CR84	DRB*11	C	T	G	C	G	T	C	G	T	A	C	C	T	C	C	G	A	G	G	T	G	A	A	A	A	G
CR09	DRB*21	C	T	G	C	G	T	C	G	T	A	C	C	T	C	G	A	G	G	T	G	A	A	A	A	A	G
CR85	DRB*22	C	T	A	C	G	T	C	G	T	A	C	C	T	C	C	G	A	G	G	T	G	A	A	A	A	G
CR16	DRB*23	C	T	A	C	G	T	C	G	T	A	C	C	T	C	C	A	A	G	G	T	G	G	G	A	A	G
CR17	DRB*23	C	T	A	C	G	T	C	G	T	A	C	C	T	C	C	A	A	G	G	T	G	A	A	A	A	G
CR15	DRB*15	C	T	G	C	G	T	C	G	T	A	C	C	T	C	C	A	A	G	G	T	G	G	G	A	A	G
CR10	DRB*15	C	T	G	C	G	T	C	G	T	A	C	C	T	C	C	A	A	G	G	T	G	G	G	A	A	G
DRB3.2.n1	DRB*15	C	T	G	C	G	T	C	G	T	A	C	C	T	C	C	A	A	G	G	T	G	G	G	A	A	G
DRB3.2.n2	DRB*16	C	T	G	C	G	T	C	G	T	A	C	C	T	C	G	G	G	G	G	T	G	G	G	A	A	G
DRB3.2.n3	DRB*17	C	T	G	T	G	T	C	G	G	A	C	C	T	C	C	A	A	G	G	T	A	C	A	C	A	G

Nota. Se muestran solamente los sitios variables de las secuencias y su posición en ECA20. Cada secuencia representa un alelo reportado con su respectiva nomenclatura. Los colores diferencian sitios conservados (celeste) de los variables (amarillo) entre las diferentes secuencias alélicas. Los alelos designados con *nº indican las secuencias alélicas nuevas, y la numeración es correlativa al último alelo reportado en el gen *ELA-DRB3*.

El análisis comparativo de las secuencias obtenidas con los alelos reportados para *E. caballus* y *E. przewalski* mostró que las secuencias detectadas tienen 100 % de identidad con *Eqca-DRB*9*, *Eqca-DRB*6* y *Eqca-DRB*11* del caballo doméstico, *Eqpr-DRB*2* de *E. przewalski*, y se detectaron 11 secuencias nuevas que no han recibido nomenclatura estandarizada en *E. caballus* (Tabla 1.9). En los animales heterocigotas, los alelos nuevos se identificaron por comparación con las secuencias conocidas y las secuencias nuevas en animales homocigotas. Los alelos más frecuentes fueron *Eqca-DRB*9* y *Eqca-DRB*11*, en tanto que los restantes alelos mostraron frecuencias de entre 2 y 8 % (Tabla 1.10), y un valor observado de heterocigosis de 0,130. No se detectaron deleciones o inserciones. Los cambios de nucleótidos consistieron en 14 transiciones (Ts): 11 A/G (R) + 3 T/C (Y), y 12 transversiones (Tv): 8 G/C (S) + 2 T/G (K) + 2 A/T (W). Estos cambios ocurrieron en 1.^a posición (10), 2.^a posición (11) y 3.^a posición (11) del codón, lo que, a nivel de la secuencia predicha de aminoácidos, permitió identificar 8 sustituciones sinónimas (S) y 19 sustituciones no sinónimas (NS).

Tabla 1.10.

Frecuencias de los alelos ELA-DRA, ELA-DRB2 y ELA-DRB3 observados en los caballos Criollo Argentino.

Gen	N	Frecuencia
<i>ELA-DRA</i>		
DRA*0101	39	0,813
DRA*0201	8	0,167
DRA*0301	1	0,020
<i>ELA-DRB2</i>		
DRB2*3	21	0,583
DRB2*7	7	0,194
DRB2*3.2	1	0,028
DRB2*4.2	1	0,028
DRB2*7.2	5	0,139
DRB2*8	1	0,028
<i>ELA-DRB3</i>		
DRB3.2.6	1	0,021
DRB3.2.9	7	0,146
DRB3.2.11	15	0,313
DRB3*18	2	0,042
DRB3*19	1	0,021
DRB3*20	1	0,021
DRB3*21	2	0,042
DRB3*22	2	0,042
DRB3*23	4	0,083
DRB3*24	2	0,042
DRB3*25	2	0,042
DRB3*26	2	0,042
DRB3*27	2	0,042
DRB3*28	1	0,021

3.1.3 Gen de la subunidad p40 de la interleuquina 12 (IL12p40)

La pirosecuenciación permitió detectar exitosamente tanto en la población de caballos Criollo Argentino (N = 85) como en la de caballos chaqueños (N = 45) el SNP del exón 5 de *IL-12p40* (Figura 1.11). El alelo *IL12p40*-G resultó el más frecuente ($f = 0,772$) y se detectaron los tres genotipos posibles: GG, AG, AA, con frecuencias genotípicas de: 0,593; 0,367 y 0,04, respectivamente. La población se encuentra en equilibrio de Hardy-Weinberg ($p = 0,546$). La heterocigosis esperada para el locus *IL12p40* fue de $He = 0,350$ y la heterocigosis observada fue de $Ho = 0,380$.

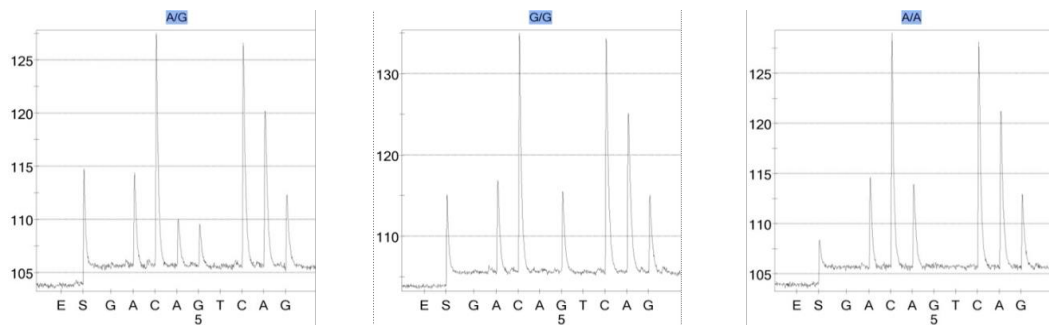


Figura 1.11. Pirogramas del exón 5 del gen *IL12p40*. En azul se observan los distintos genotipos detectados. Sobre el eje horizontal se indica el orden de adición de los nucleótidos. El eje vertical indica la intensidad del pico lumínico correspondiente a la base incorporada.

Por otra parte, en la población de caballos chaqueños, el alelo *IL12p40-G* también resultó el de mayor frecuencia ($f = 0,598$) y se detectaron los tres genotipos posibles del SNP con frecuencias de: 0,358 (GG), 0,481 (AG) y 0,162 (AA). La heterocigosis esperada fue de $H_e = 0,490$, mientras que la heterocigosis observada fue de $H_o = 0,510$. La población se encuentra en equilibrio de Hardy-Weinberg ($p = 1$).

3.1.4 Gen del factor de necrosis tumoral alfa (TNF- α)

Para el análisis del gen *TNF- α* , se tipificaron 85 caballos de la raza Criollo Argentino pertenecientes a la Cabaña de Caballos Criollos de INTA Leales, del Instituto de Investigación Animal del Chaco Semiárido (IIACS), localizada en Tucumán y 43 caballos mestizos de la provincia del Chaco. La amplificación por PCR anidada de la región promotora del gen *TNF- α* fue exitosa en ambas poblaciones, obteniendo en cada ronda los fragmentos de interés con el tamaño esperado. Luego, la pirosecuenciación permitió detectar los 3 SNPs: *TNF-_{95C/T}*, (posición 20:31358687, rs69281192C > T), *TNF-_{89C/T}* (posición 20:31358693, rs909253 T > C) y *TNF-_{85C/T}* (posición 20:31358697, rs397263474 C > T) de la región promotora del gen *TNF- α* (Figura 1.12) tanto en la población de caballos Criollo Argentino como en la de caballos chaqueños.

En la población de Criollos, la H_e para el SNP *TNF-_{95C/T}* resultó de $H_e = 0,461$ con una $H_o = 0,426$; mientras que el SNP *TNF-_{89C/T}* presentó una $H_e = 0,124$ y $H_o = 0,132$ y para la posición *TNF-_{85C/T}* se observaron

los valores de 0,457 y 0,294 para He y Ho respectivamente. Por otra parte, los animales criollo chaqueño analizados, mostraron para el SNP TNF^{-95C/T} un valor de He = 0,391 con una Ho = 0,444; mientras que para la posición TNF^{-85C/T} se observaron los valores de 0,333 y 0,422 para He y Ho respectivamente. En este grupo de caballos la posición TNF^{-89C/T} resultó monomórfica para el alelo T.

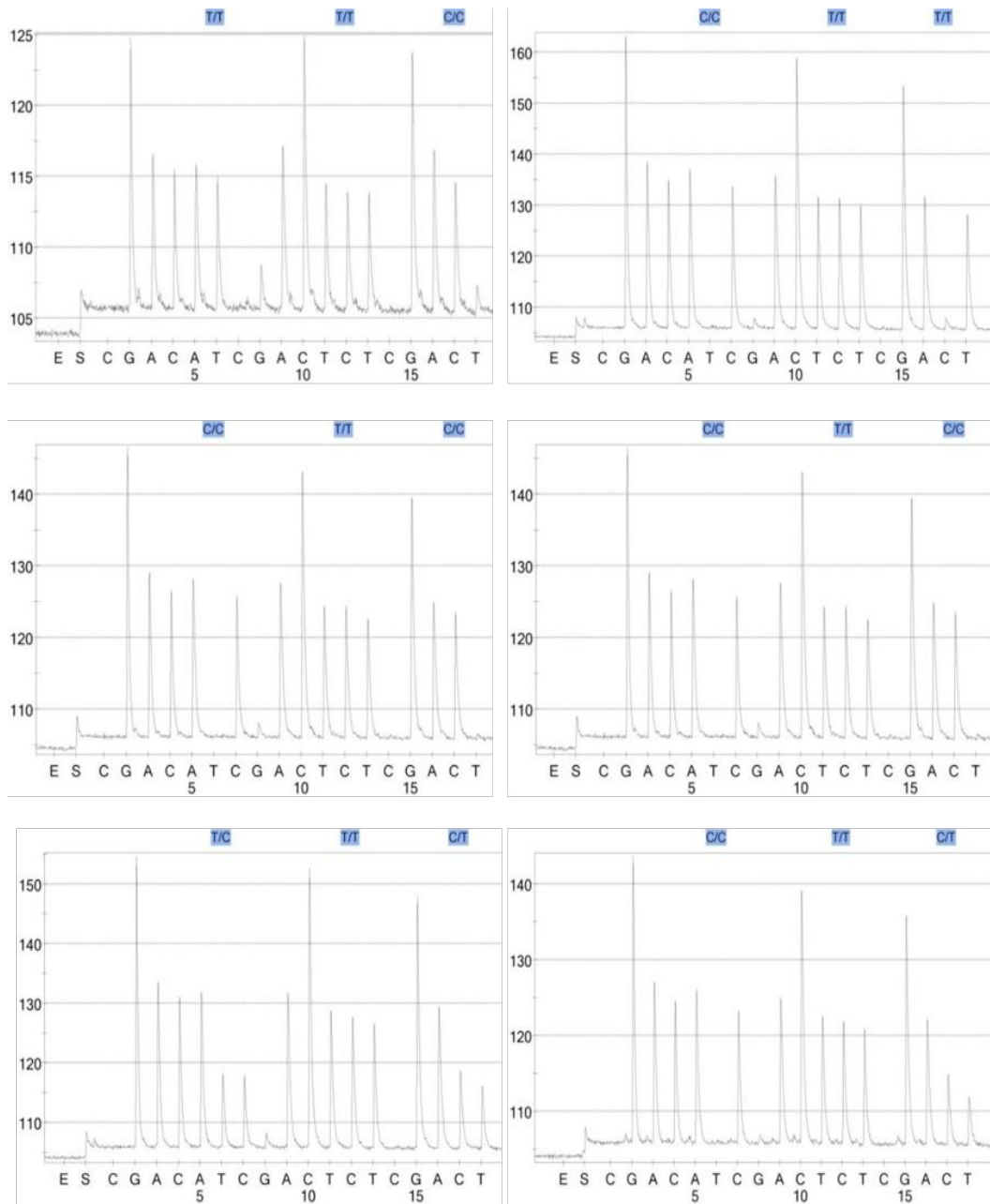


Figura 1.12. Genotipos detectados en el gen $TNF-\alpha$ mediante pirosecuenciación. En color se indican los genotipos correspondientes a los tres SNPs de la región promotora de $TNF-\alpha$. Sobre el eje horizontal se indica el orden de adición de los nucleótidos. El eje vertical indica la intensidad del pico lumínico correspondiente a la base incorporada.

En la tabla 1.11 se muestra el análisis genético comparativo interracial, para el que se incluyeron datos de 179 caballos de distintas razas (SA, SPC, APP, CRxPR) obtenidos en trabajos realizados previamente en el grupo Equinos (Francisco, 2014; Kalemkerian, 2014) de la línea de investigación de animales domésticos del IGEVET.

Tabla 1.11

Frecuencias alélicas y parámetros de diversidad genética de los 3 SNPs del promotor de TNF- α en distintas poblaciones equinas de Argentina.

Raza / población	SNP-95			SNP-89		SNP-85	
	N	C	T	T	C	C	T
Criollo Argentino	85	0,640	0,360	0,934	0,066	0,647	0,353
Criollo Chaqueño	43	0,733	0,267	1,000	0,000	0,789	0,211
Silla Argentino	122	0,467	0,533	0,988	0,012	0,816	0,184
CRxPR*	10	0,800	0,200	1,000	0,000	0,850	0,150
SPC†	11	0,227	0,773	1,000	0,000	0,000	1,000
Appaloosa	18	0,562	0,438	1,000	0,000	0,750	0,250
Árabe	18	0,833	0,167	1,000	0,000	0,889	0,111
Total	307	0,558	0,442	0,982	0,018	0,783	0,217
He promedio		0,493		0,035		0,340	
HWE		0,0313		1,000		0,0948	

Nota: *CRxPR: Criollo Argentino x Percherón; †SPC: Sangre Pura de Carrera. p -valor

HWE 0,001

La reconstrucción de las fases de ligamiento y la asignación de probabilidad de los haplotipos del promotor de *TNF- α* , mostró en la población de caballos Criollos, cinco haplotipos posibles siendo las combinaciones: TTC, CTT y CTC las más frecuentes (Tabla 1.12). Además, se detectó la variante C en la posición *TNF*_{-89C/T} (Figura 1.11), generando una combinación haplotípica poco frecuente (TCT, $f = 0,038$). Por otra parte, la población de caballos chaqueños presentó tres haplotipos, siendo la combinación CTC la más frecuente (Tabla 1.12).

Tabla 1.12.

*Frecuencias de los haplotipos inferidos con el software PHASE v2.1.1 en la región del promotor del gen *TNF- α* en poblaciones equinas de Argentina.*

Haplotipo	CR	CRxPR	SA	AR	APP	CH	SPC
TTC	0,397	0,189	0,507	0,122	0,437	0,274	0,773
CTC	0,274	0,661	0,296	0,767	0,313	0,524	0,227
CTT	0,265	0,139	0,163	0,066	0,250	0,202	0,000
TTT	0,000	0,011	0,022	0,045	0,000	0,000	0,000
TCT	0,038	0,000	0,000	0,000	0,000	0,000	0,000
CCT	0,026	0,000	0,000	0,000	0,000	0,000	0,000

Nota. Los valores resaltados en **negrita** corresponden a los haplotipos que solo se observaron en los caballos Criollo Argentino (CR). CRxPR: Criollo Argentino x Percherón; SA: Silla Argentino; AR: Árabe; APP: Appaloosa; CH: Criollo Chaqueño; SPC: Sangre Pura de Carrera.

3.1.5 Gen del receptor de lentivirus equino 1 (ELR1)

3.1.5.1 SNP *rs68535250* del exón 4 del gen *ELR1*

Con el objetivo de analizar el SNP G > A en la posición 47311347 de ECA2 (*rs68535250*) del gen *ELR1*, se tipificaron caballos Criollo Argentino (N = 50), Árabe (N = 24) y mestizos (N = 63) mediante la técnica de pirosecuenciación. En la Figura 1.13 se observan los pirogramas correspondientes a los genotipos detectados para la variante: A/A, A/G, G/G. Las frecuencias génicas así como los parámetros de diversidad genética se resumen en la Tabla 1.13.

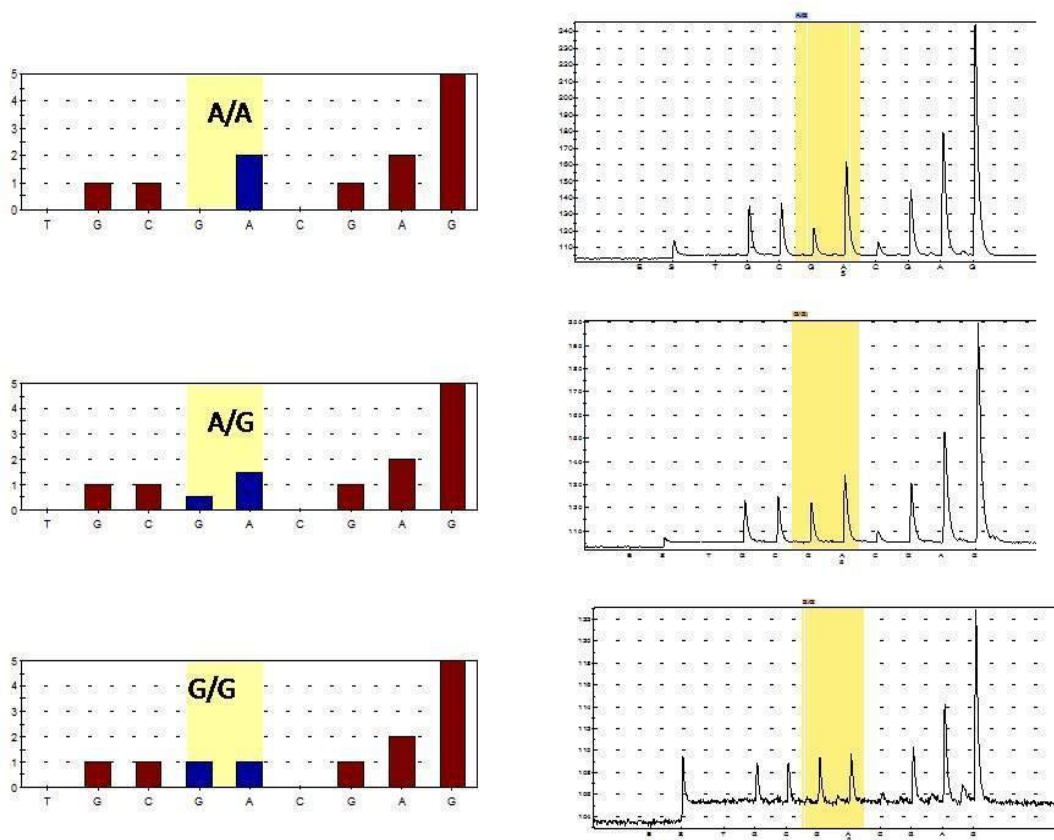


Figura 1.13. Histogramas teóricos (izquierda) y pirogramas obtenidos (derecha) para cada genotipo del SNP: *rs68535250* del exón 4 del gen *ELR1*. En cada esquema teórico (izquierda) se puede observar la secuencia de lectura, el orden de dispensación de nucleótidos del equipo y el genotipo detectado.

Tabla 1.13

Frecuencias y diversidad genética del SNP rs68535250 del exón 4 del gen ELR1.

Población	Frecuencias alélicas			Diversidad genética			
	N	A	G	He	Ho	FIS	HWE
Criollos	50	0,330	0,670	0,447	0,620	-0,394	0,008*
Árabes	24	0,563	0,437	0,503	0,542	-0,079	1,000
Mestizos	63	0,159	0,841	0,269	0,222	0,176	0,165

Nota. * valor significativo de p ($\alpha = 0,05$).

El SNP *rs68535250* también fue analizado considerando el estatus sanitario respecto a AIE (test de Coggin's) en el grupo de caballos mestizos (N = 63). Así, se dividió la población en dos grupos: caballos seropositivos (N = 26) y seronegativos (N = 37) para AIE. En la tabla 1.14 se muestran los valores estimados de *Odds ratio* (OR) y el p -valor del test de Fisher.

Tabla 1.14.

Análisis del SNP rs68535250 del exón 4 del gen ELR1 en caballos seropositivos y seronegativos para VAIE.

SNP	VAIE + (N=37)	VAIE - (N=26)	OR	95 % IC	Fisher p -valor
A	0,18	0,13	0,313	0,115 – 0,852	0,023
G	0,82	0,87	3,191	1,174 – 8,673	0,023

Nota: VAIE: Virus Anemia Infecciosa Equina; OR: *Odds Ratio*; IC: intervalo de confianza.

3.1.5.2 Exones 3 y 5 del gen *ELR1*

Para analizar la variabilidad genética a nivel del exón 3 (ENSECAE00000042690) y del exón 5 (ENSECAE00000042908) del gen *ELR1*, se rastrearon nuevos SNPs en 47 caballos mestizos no emparentados testeados serológicamente para AIE. El diagnóstico serológico fue realizado mediante el test de inmunodifusión en agar gel (AGID) indicado por la Oficina Internacional de Epizootias (OIE) para comercio de caballos (test de Coggin's) en laboratorios habilitados (FCV-UNLP y SENASA) por el Servicio Nacional de Sanidad Agroalimentaria (SENASA). Las muestras se encontraban disponibles en el banco de ADN del IGEVET y provenían de muestreos realizados en la provincia de Chaco y Buenos Aires. Se incluyeron 16 caballos seropositivos, 31 caballos seronegativos y un burro libre de anemia para el estudio.

La región de interés del exón 3 consistió en 151 pb: 42 pb pertenecientes al intrón 2 y 109 pb correspondientes al exón 3 completo. Por otro lado, la región del exón 5 comprendió 303 pb: 159 pb del intrón 4 y 144 pb del exón 5. En la figura 1.14 se presenta uno de los electroferogramas obtenidos como resultado de la secuenciación directa del exón 5.

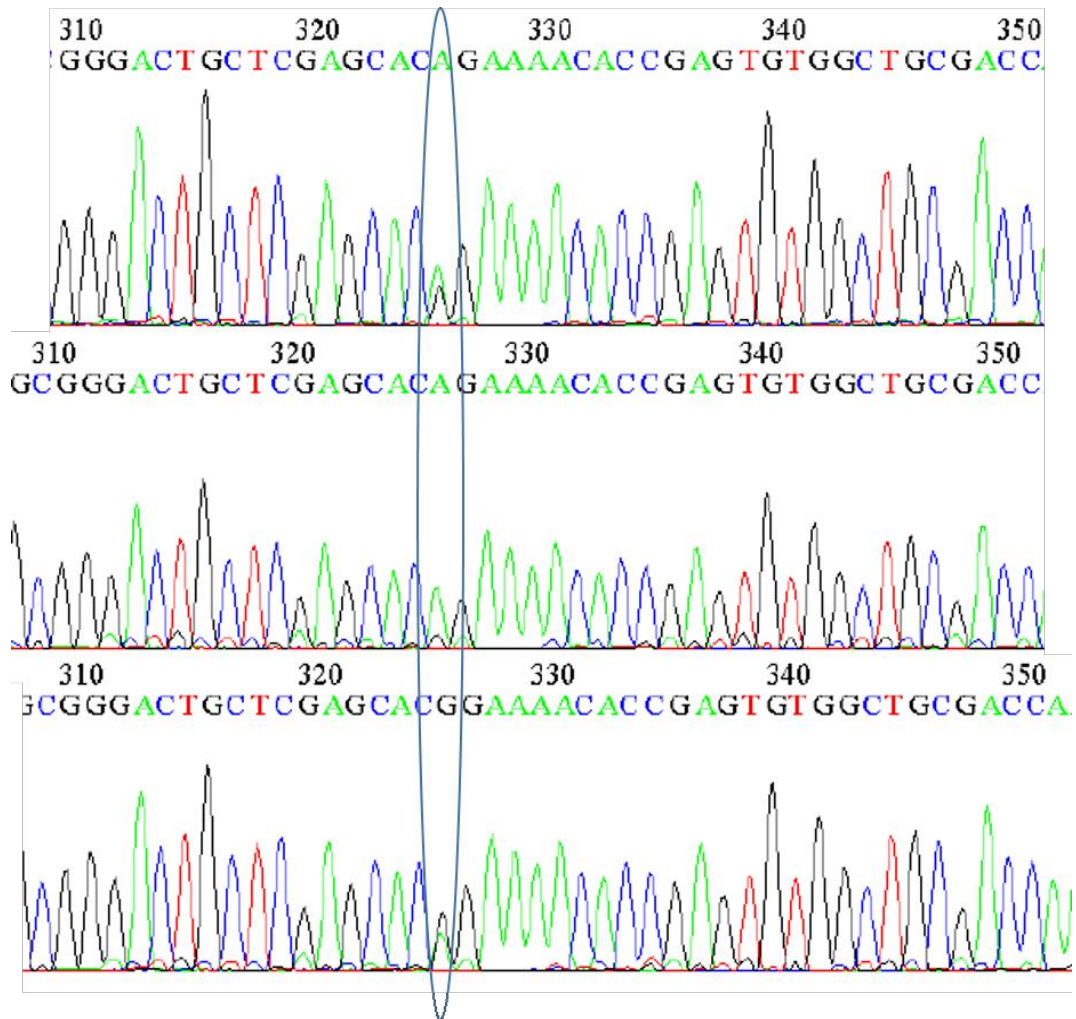


Figura 1.14. Electroferograma parcial de la secuenciación del exón 5 de *ELR1*. Se muestra el SNP G > A detectado en la posición 47311301.

En total se identificaron 21 SNPs, 10 en las secuencias intrónicas (un SNP cada 19 nucleótidos), ocho en el exón 5 (un SNP cada 19 nucleótidos), y tres en el exón 3 (un SNP cada 36 nucleótidos) del gen *ELR1*. El análisis de la secuencia nucleotídica del exón 3 mostró tres transiciones: dos SNPs reportados previamente (*rs395939045* y *rs397088947*) y un SNP nuevo C > T en la posición 47309385 del cromosoma 2 (Tabla 1.15). Este SNP fue validado en la población por detección tanto en individuos no emparentados homocigotas y heterocigotas. Por otro lado, el análisis de las secuencias nucleotídicas del exón 5 mostró ocho SNPs: dos previamente reportados (*rs68535248* y *rs68535250*) y seis nuevos SNPs, representados por cinco transiciones: 47311282 G > A (*rs1095048803*), 47311284 C > T (*rs1095048804*), 47311323 G > A (*rs1095048807*), 47311295 A > G (*rs1095048805*), 47311301 A > G (*rs1095048806*) y una transversión: 47311358 T > G (*rs1095048808*) (Tabla 1.15). Además, una transición C > T en la posición 47309360 del intrón 2 se detectó únicamente en un burro. Los SNPs nuevos detectados en esta tesis fueron depositados en la base de datos de variaciones cortas (dbSNP) del NCBI (www.ncbi.nlm.nih.gov/project/SNP) donde luego del proceso de curado del sitio se le asignó un número de referencia de SNP definitivo (rs).

Tabla 1.15.

Polimorfismos detectados por secuenciación en los exones 3 y 5 del gen ELR1.

Región	Posición*	SNP	Tipo	MAF	p-valor	Cambio AA	ID
intrón 2	47309353	A/G	NC	NA			rs1095048794
intrón 2	47309360	C/T [†]	NC	NA			NR
exón 3	47309385	T/C	S	0,224	0,839		rs1095048796 ‡
exón 3	47309412	G/A	S	0,414	8,09E ⁻⁰⁵		rs395939045
exón 3	47309451	C/T	S	0,172	0,046		rs397088947
intrón 4	47311106	A/G	NC	NA			rs1095048797 ‡
intrón 4	47311118	C/G	NC	NA			rs1095048798 ‡
intrón 4	47311124	C/T	NC	NA			rs397236404
intrón 4	47311135	A/C	NC	NA			rs1095048799 ‡
intrón 4	47311155	C/T	NC	NA			rs1095048800 ‡
intrón 4	47311162	A/G	NC	NA			rs395177919
intrón 4	47311175	C/T	NC	NA			rs1095048801 ‡
intrón 4	47311193	A/C	NC	NA			rs1095048802 ‡
exón 5	47311282	G/A	NS	0,039	0,080	Arg108Lys	rs1095048803 ‡
exón 5	47311284	C/T	NS	0,026	1,000	Arg109Trp	rs1095048804 ‡
exón 5	47311295	G/A	S	0,013	1,000		rs1095048805 ‡
exón 5	47311301	G/A	S	0,224	0,001		rs1095048806 ‡
exón 5	47311310	C/T	S	0,000	1,000		rs68535248
exón 5	47311323	G/A	NS	0,013	1,000	Asp122Asn	rs1095048807 ‡
exón 5	47311347	A/G	NS	0,079	0,772	Glu130Lys	rs68535250
exón 5	47311358	T/G	NS	0,079	0,016	Asp133Glu	rs1095048808 ‡

Nota: *Posición respecto a la del acceso ENSECAG00000008751 de la base de datos de ENSEMBL. †Variante observada solo en burros. ‡SNPs detectados en esta tesis. NC: no codificante, NR, no reportado en dbSNP, NS: no sinónima, S: sinónima. MAF: frecuencia del alelo menor (en inglés *Minor Allele Frequency*), en **negrita**: alelo menor, NA: no analizado, rs: número referencia del SNP en SNPdb (RefSNP). Significancia probabilística: p -valor = 0,05.

Los tres SNPs analizados en la región del exón 3 presentaron valores de MAF entre 0,172 y 0,414 (Tabla 1.15). El alelo más frecuente para la transición C > T descrita en la posición 47309385 del exón 3 fue el T, con una frecuencia de 0,776 (MAF = 0,224). El análisis de las secuencias predichas de aminoácidos mostró que los tres SNPs causan sustituciones sinónimas.

En cuanto al exón 5, se detectaron ocho SNPs: seis no reportados previamente (cinco transiciones y una transversión) y dos transiciones ya reportadas en trabajos anteriores (*rs68535248C/T* y *rs68535250A/G*). Los valores de MAF se situaron entre 0,013 y 0,224. El SNP *rs68535248* resultó monomórfico, siendo el alelo C el único detectado en los caballos analizados. Los SNPs (*rs1095048803*, *rs1095048804G/A*, *rs1095048807G/A*, *rs1095048808T/G* y *rs68535250A/G*) conducen a cambios aminoacídicos no sinónimos.

El análisis del efecto de las sustituciones aminoacídicas con el software PROVEAN mostró que las variantes Arg108Lys (*rs1095048803*

G/A), Asp122Asn (*rs1095048807* G/A), Glu130Lys (*rs68535250* A/G), Asp133Glu (*rs1095048808* T/G) tienen un puntaje indicativo de efecto neutral sobre la función biológica de la proteína (rango: -1,388-0,565). En contraste, el análisis de la sustitución Arg109Trp mostró que podría tener un efecto deletéreo sobre la función proteica (puntaje PROVEAN = -4,357, cut-off = -2,5).

El análisis de desequilibrio de ligamiento (LD) y de los haplotipos permitió identificar seis variantes alélicas en el exón 3 (Figura 1.15) las cuales se nombraron como: *ELR1e3-002*, *ELR1e3-003*, *ELR1e3-004*, *ELR1e3-005*, *ELR1e3-006* y *ELR1e3-007* (Números de acceso: KT956427-32, Tabla 1.16).

```

                                24          29
                                A  L  Y  L  L  L  L  G  S  F  R  Y  T  L
ELR1e3-001  CTGCTGCTCCCTCCCAACGCCCGCTCTGTCCCTCCAG GCC TTG TAT CTC CTC CTC CTG GGG TCC CCG CGC TAC ACC CTG
ELR1e3-002  .....-.....A.....T.....
ELR1e3-003  .....-.....A.....T.....
ELR1e3-004  .....-.....A.....T.....
ELR1e3-005  .....-.....A.....T.....
ELR1e3-006  .....-.....
ELR1e3-007  .....-.....T.....
----- intron -----}----- exon -----

```

```

                                38          40          51
                                A  T  P  Q  C  K  E  E  E  Y  P  V  G  T  E  C  C  P  K  C  S  P
ELR1e3-001  GCA ACG CCC CAG TGC AAA GAG GAG GAG TAC CCA STG GGG ACT GAG TGC TGC CCC AAA TGC AGT CCA GGTAGG
ELR1e3-002  ..G ..... ..C .....
ELR1e3-003  ..G ..... ..C .....
ELR1e3-004  ..G ..... ..C .....
ELR1e3-005  ..... ..C .....
ELR1e3-006  ..... ..C .....
ELR1e3-007  ..... ..C .....
----- exon -----|-intron-

```

Figura 1.15. Alineamiento nucleotídico de las variantes alélicas del exón 3 del gen *ELR1*. En la primera línea se muestra la secuencia aminoacídica predicha, los números indican la posición de cada aminoácido. En amarillo se indican las cisteínas conservadas dentro del dominio rico en cisteínas 1 (CRD1).

Tabla 1.16

Frecuencias de las variantes alélicas definidas para el exón 3 del gen ELR1.

Variante alélica	N° Acceso	VAIE negativo	VAIE positivo
		(N=18)	(N=11)
<i>ELR1e3-001</i>	ENSECAE0000042690	0,000	0,136
<i>ELR1e3-002</i>	KT956427	0,500	0,454
<i>ELR1e3-003</i>	KT956428	0,111	0,000
<i>ELR1e3-004</i>	KT956429	0,056	0,000
<i>ELR1e3-005</i>	KT956430	0,000	0,227
<i>ELR1e3-006</i>	KT956431	0,111	0,091
<i>ELR1e3-007</i>	KT956432	0,222	0,091
He		0,740	0,692
Ho		0,273	0,333
HWE*		0,000	0,000
Diferenciación génica*			0,003
Diferenciación genotípica*			0,039

Nota: Distribución de las frecuencias de las variantes alélicas definidas mediante el programa PHASE en caballos negativos y positivos para VAIE. HWE = Equilibrio de Hardy-Weinberg. *Nota probabilística: *p-valor* = 0,05.

El análisis de los ocho SNPs detectados en el exón 5 mostró ocho combinaciones haplotípicas diferentes. Sin embargo, solo cuatro de ellas (Figura 1.16) fueron validadas por su ocurrencia en individuos heterocigotas y homocigotas, así como en productos de PCR

independientes. Estas combinaciones haplotípicas validadas conformaron cuatro variantes alélicas denominadas: *ELR1e5_001.1*, *ELR1e5_002.1*, *ELR1e5_002.2* y *ELR1e5_003.1*, cuyas frecuencias se muestran en la tabla 1.17.

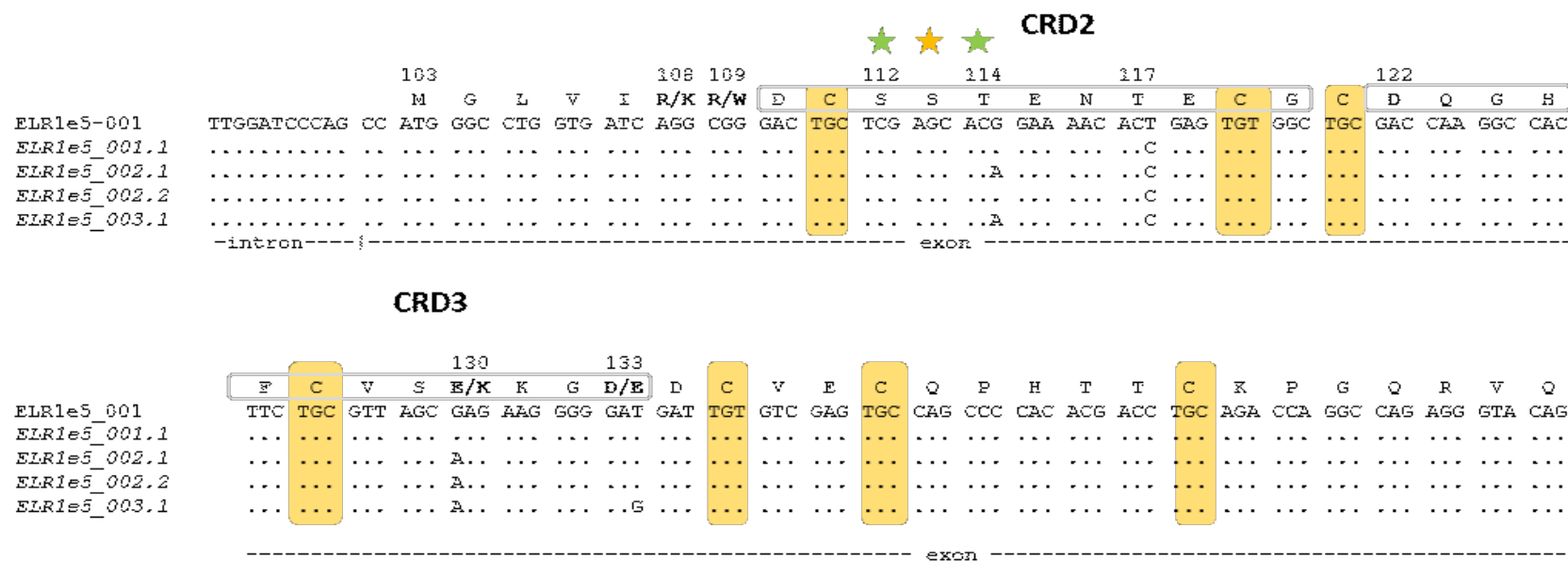


Figura 1.16. Alineamiento nucleotídico de las variantes alélicas del exón 5 del gen *ELR1*. En la primera línea se muestra la secuencia aminoacídica predicha, los números indican la posición de cada aminoácido. En amarillo se indican las cisteínas conservadas dentro de los dominios ricos en cisteínas 2 y 3 (CRD2, CRD3).

Tabla 1.17.

Frecuencias de las variantes alélicas definidas para el exón 5 del gen ELR1.

Variante alélica	N° acceso	VAIE-negativo (N=25)	VAIE-positivo (N=16)
<i>ELR1e5_001</i>	ENSECAE0000042908	0,000	0,000
<i>ELR1e5_001.1</i>	KT932625	0,071	0,000
<i>ELR1e5_002.1</i>	KT932626	0,190	0,094
<i>ELR1e5_002.2</i>	KT932627	0,595	0,781
<i>ELR1e5_003.1</i>	KT932628	0,048	0,094
<i>ELR1e5_004</i>		0,000	0,031
<i>ELR1e5_008</i>		0,024	0,000
<i>ELR1e5_009</i>		0,024	0,000
<i>ELR1e5_010</i>		0,048	0,000
He		0,656	0,333
Ho		0,320	0,125
HWE*		0,000	0,030
Diferenciación génica*			0,021
Diferenciación genotípica*			0,063

Nota: Distribución de las frecuencias de las variantes alélicas definidas mediante el programa PHASE en caballos negativos y positivos para VAIE. HWE = Equilibrio de Hardy-Weinberg. *Nota probabilística: *p*-valor 0,05.

Cada SNP se evaluó mediante un análisis de asociación de marcador único. Solo el SNP previamente reportado (*rs68535250*) resultó estadísticamente significativo ($p = 0,026$). Esta transición A > G causa un

cambio de ácido glutámico a lisina y se presentó en las variantes: *ELR1e5_002.1*, *ELR1e5_002.2* y *ELR1e5_003.1*.

3.1.5.2.1 Análisis filogenético de las secuencias del exón 5 de *ELR1*

La construcción de árboles filogenéticos con el método de *Neighbour Joining* mostró que las variantes alélicas de caballo se agruparon juntas en un *cluster* único incluyendo la secuencia de *Equus asinus* y de forma separada de otras especies. Debido a que los cambios nucleotídicos detectados en el exón 3 fueron de tipo sinónimo, y por ende no producen cambios aminoacídicos en los productos proteicos predichos, el análisis filogenético se realizó únicamente para el exón 5.

Las secuencias aminoacídicas predichas a partir de las variantes detectadas en el exón 5 del gen *ELR1* fueron comparadas con sus ortólogos en varias especies (Figura 1.17). El análisis involucró 17 secuencias aminoacídicas y el set de datos final consistió en 47 posiciones aminoacídicas. La identidad aminoacídica varió de 94 a 98 % entre las variantes equinas de *ELR1*; mientras que, al comparar con otras especies, fue mayor cuando se comparó con ortólogos como perro y gato (64 – 79 %) que cuando se contrastó con otros ortólogos mamíferos (< 55 %).

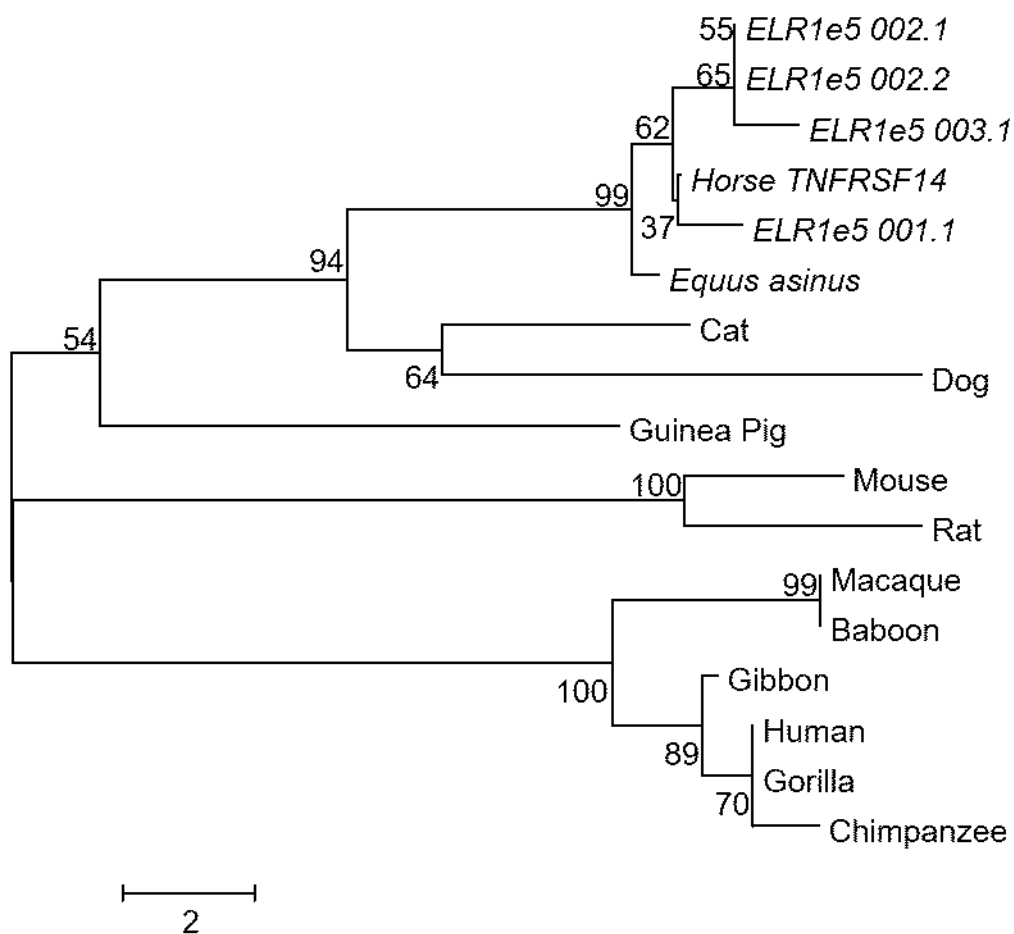


Figura 1.17. Árbol de *Neighbor-joining* de las variantes del gen equino *ELR1* y sus ortólogos construido con el software MEGA7 (Kumar *et al.*, 2016). Los números al lado de las ramas indican el test de *bootstrap* con 1000 repeticiones. El árbol está dibujado a escala con la longitud de las ramas en las mismas unidades que aquellas de las distancias evolutivas utilizadas para inferir el árbol filogenético. Las distancias evolutivas fueron calculadas utilizando el método del número de diferencias (Kimura, 1980) y están en las unidades del número de aminoácidos por secuencia.

4. DISCUSIÓN

4. Discusión

4.1. Genes del ELA

La tipificación y análisis de los alelos de tres genes estructurales de la región de Clase II del ELA mediante métodos moleculares permitió identificar el polimorfismo en caballos de la raza Criollo Argentino, y de este modo, caracterizar su variabilidad intrarracial, identificando un total de 23 alelos en los tres genes DR analizados (*ELA-DRA*, *ELA-DRB2* y *ELA-DRB3*).

La tipificación del gen *ELA-DRA* por pirosecuenciación permitió identificar 3 alelos en el exón 2, con una distribución de frecuencias similar a aquellas reportadas previamente para otras razas (Díaz *et al.*, 2008), con predominancia del alelo *ELA-DRA*0101*. Si bien la cantidad de alelos detectados es baja en relación a otros genes del MHC (Miller *et al.*, 2017), el polimorfismo que presenta el gen *DRA* equino es característico de esta especie. En este sentido, Janova (2009) demostró que el polimorfismo puede estar vinculado a la selección positiva en el locus de *DRA*. La identificación de las variaciones de los genes *ELA-DRB2* y *ELA-DRB3* se realizó con diferentes metodologías, y permitió detectar un grado variable de polimorfismo (4 alelos *ELA-DRB2* y 15 alelos *ELA-DRB3*), adecuando las posibilidades metodológicas disponibles y la adquisición de nueva información para el desarrollo de técnicas más precisas. Por lo expuesto, los caballos criollos mostraron numerosos alelos, pero un grado reducido de diversidad medida como heterocigosis por gen (13 - 35 %).

Esta situación de aparente variabilidad reducida puede deberse, en primera instancia, a que los animales analizados para los genes del ELA pertenecen a familias de medio hermanos paternos. Si bien el modelo familiar es útil para confirmar o descartar alelos o resolver combinaciones haplotípicas mediante el análisis de segregación, opera por otro lado, en detrimento de la variabilidad observada. De hecho, en estudios previos del ELA se utilizan grupos familiares para resolver las fases de ligamiento de los haplotipos mientras que se utiliza un segundo grupo de caballos no emparentados para explorar y capturar la mayor diversidad genética (Sadeghi *et al.*, 2017). Vale la pena mencionar que las nuevas secuencias deben ser validadas en otras poblaciones y razas equinas, y ser sometidas a los estándares de nomenclatura internacional.

Si bien los cebadores utilizados amplifican solo los alelos correspondientes a uno u otro gen, es decir, son gen-específico, evitando la amplificación cruzada entre secuencias alélicas que tienen porciones altamente conservadas entre los distintos genes DR, con la técnica de RFLP los patrones de restricción mostraron un número limitado de alelos. Por esta razón, la tipificación final se realizó mediante secuenciación directa, y se verificó la presencia de los sitios de corte de las enzimas, y de los polimorfismos reportados y nuevos en diferentes posiciones del exón 2, para el gen *DRB2* y para todos los animales en *ELA-DRB3*.

En estos genes, cuyos productos participan en la interacción y reconocimiento antigénico, la variabilidad observada toma relevancia al considerar en conjunto el acervo génico de los caballos criollos. Puede

también surgir el planteo si verdaderamente el alto número de alelos detectados en estos genes refleja de alguna forma las mejoras metodológicas empleadas para la detección de estos polimorfismos; sin embargo, se han empleado las mismas técnicas en la detección de alelos de los genes DR en otras razas, tales como el caballo Árabe, donde se describieron 27 alelos en total (Sadaba 2018).

4.2. Genes de citoquinas

Los avances en el conocimiento de las citoquinas y los genes relacionados con la respuesta inmune han llevado a estrategias novedosas para el tratamiento de enfermedades inflamatorias, autoinmunes y la respuesta a enfermedades infecciosas. Los polimorfismos de nucleótido simple son las variaciones genéticas más comunes descritas en los genes de citoquinas (Merino *et al.*, 2013). Así, por ejemplo, el polimorfismo en el gen de *IL12p40* fue descrito en estudios de caballos de Europa, definiendo la presencia de dos alelos y mostrando diferencias significativas en las frecuencias alélicas para distintas razas de caballos (Horin *et al.*, 2004). En forma adicional, en los caballos Sangre Pura de Carrera, no se detectaron polimorfismos para el gen *IL12p40* (Horin *et al.*, 2004). En los caballos criollos, si bien el alelo G fue el más frecuente, se detectaron los 3 genotipos posibles, y una heterocigosis esperada de 0,350. El genotipo GG fue el más frecuente tanto en criollos como en los caballos criollo chaqueño, estos últimos con un valor de heterocigosis mayor ($He = 0,490$). En contraste, en un estudio

reciente (Al Abri *et al.*, 2020) donde se secuenciaron 6 caballos de distintas razas, el genotipo más frecuente para este SNP resultó ser el AA. Esto podría deberse a la cantidad de animales que habitualmente se utilizan en estudios de NGS (en inglés *Next Generation Sequencing*, Secuenciación de próxima generación), así como a las diferencias entre las razas analizadas.

En el factor de necrosis tumoral (TNF- α), la presencia de variabilidad en la región del promotor del gen se reportó previamente por Brown (2006), con una distribución de algunos haplotipos del *TNF- α* limitada a razas como *Old Kladruber*, *Akhal-Teke*, Sangre Pura de Carrera, Árabes, *Standardbred*. En los caballos Criollo Argentino analizados en esta tesis, los resultados evidencian la frecuencia incrementada de determinadas combinaciones de SNPs (haplotipos) y la ausencia de otras. La asignación de los haplotipos, sobre la base de la frecuencia de los SNPs, demostró una muy baja frecuencia de los haplotipos CCT (< 4 %) y TTT (< 0,2 %). Sin embargo, al combinar esta información con el análisis de segregación de los individuos emparentados, se confirmó la ausencia de animales con estos dos haplotipos. Por otra parte, entre los haplotipos que presentan la variante C en la posición TNF-₈₉ (CCT, TCT, CCC) y de acuerdo con las probabilidades de ocurrencia estimadas mediante el software PHASE v2.1.1 (Li y Stephens, 2003) el haplotipo TCT sería el único probable en esta población. Vale la pena notar que la presencia de la variante C en esta posición no había sido descrita previamente en ninguna raza equina, sólo se encuentra reportada en burros y cebras de

Grant (Brown *et al.*, 2006). Las distintas frecuencias de los haplotipos obtenidas podrían deberse a la presencia de distintos tipos de presión selectiva sobre ciertos SNPs (Brown *et al.*, 2006). De hecho, las razas con alto impacto de selección, tales como la raza Árabe y Sangre Pura de Carrera, presentaron un número menor de combinaciones haplotípicas (Tabla 1.12), mientras que otras poblaciones con evidencias de escasos o prácticamente nulos procesos de selección, presentaron al menos 4 combinaciones haplotípicas. Dada la importancia que tiene la secuencia del promotor en la función regulatoria del gen, el desequilibrio de ligamiento observado podría determinar la eficiencia en la afinidad e interacción con factores de transcripción específicos, que, a su vez, podrían condicionar la intensidad de la respuesta inmune en las que intervienen estas citoquinas (Yilmaz *et al.*, 2005).

La presencia de polimorfismos en los genes *IL-12p40* y *TNF- α* , mediadores proinflamatorios con una amplia variedad de funciones en la respuesta inmune, podría, de cierta forma, reflejar la historia del caballo Criollo Argentino en términos de la interacción de los caballos originarios con un entorno ambiental diverso, en el que también estuvieron expuestos a un amplio rango de patógenos. Los avances en el conocimiento de las citoquinas y los genes relacionados con la respuesta inmune han llevado a estrategias novedosas para el tratamiento de enfermedades inflamatorias, autoinmunes y la respuesta a enfermedades infecciosas (Vandenbroeck y Goris, 2003).

El aporte de nueva información detallada en las razas para las que este tipo de variaciones no ha sido descrito previamente, posibilita, entre otros aspectos, tener un buen punto de partida como base para discriminar entre factores genéticos y no genéticos. Por ejemplo, Patin *et al.*, (2018) identificaron en humanos los efectos relativos del ambiente y los factores genéticos en la variación de los parámetros de las células inmunes de la respuesta innata y adaptativa en condiciones homeostáticas. Combinando datos fenotípicos con tipificación genómica de ADN, describieron los factores no genéticos que afectan la variación en los parámetros de las células inmunes, y hallaron 15 loci que mostraron enriquecimiento genético con variantes asociadas a enfermedades. Finalmente, demostraron que la inmunidad innata está más fuertemente controlada por las variaciones genéticas, y que la inmunidad adaptativa, está en mayor parte conducida por factores no genéticos, principalmente exposición ambiental, estableciendo la importancia fundamental de la inmunidad innata en la susceptibilidad a enfermedades comunes.

Uno de los objetivos de este estudio consistió en la descripción de las variaciones en genes relacionados con la respuesta inmune. Esta información sin dudas es relevante como una base informativa para relacionar la presencia de variaciones genéticas con susceptibilidad a, predisposición, o riesgo a enfermedades infecciosas, sumando a antecedentes tales como el vínculo reportado entre la evolución de la infección por VAIE y la alteración del perfil de moléculas pro inflamatorias (IL-1 α , IL-1 β , IL-6, IL-10 y TNF- α) (Covaleda *et al.*, 2010), o aumento en

la expresión de estas citoquinas en la fase aguda de la enfermedad (Cook *et al.*, 2013). Por tanto, el conjunto de datos reportados en este trabajo aporta conocimientos acerca de genes de la RI que han sido parcialmente estudiados, y otros genes para los cuales no se conocía su relación con determinadas enfermedades. No obstante, se requiere de un diseño muestral que excede los objetivos de este trabajo, que solamente buscó describir el caudal de variaciones de un conjunto de genes indicativos, seleccionados por su función potencial en condiciones homeostáticas y de defensa del organismo ante este tipo de enfermedades. La descripción de variaciones genéticas en genes inmunes en la raza criolla, caballos con una historia de rusticidad y adaptación conocida, toma relevancia en tanto y en cuanto permita continuar investigando para establecer relaciones estadísticas con la presencia o ausencia de determinadas variantes genéticas.

4.3. Receptor equino del lentivirus 1 (*ELR1*)

La incorporación en el trabajo de tesis del análisis del gen del receptor equino de lentivirus 1 (*ELR1*), respondió a la inquietud surgida ante el conocimiento de que es el único receptor encargado de la interacción de la partícula viral de AIE con las células del caballo. Además, si bien no se disponía de información bibliográfica que acreditara la presencia de variaciones alélicas del receptor, la bibliografía relevada indicaba el uso de los miembros de la familia de receptores TNFRSF como receptores virales (Qian *et al.*, 2015; Hehlhans y Pfeffer, 2005).

Estos motivos, sumados a la disponibilidad de muestras con diagnóstico serológico para VAIE, motivó la inclusión del gen *ELR1* en este trabajo de tesis.

El estudio de *ELR1*, identificado como el único receptor involucrado en la interacción entre la partícula viral de VAIE y las células del hospedador (Zhang *et al.*, 2005), se orientó a conocer la secuencia de los exones 3 y 5, que codifican para dos regiones estructurales relevantes: el CRD1, que media el contacto directo entre la partícula viral y el receptor propiamente dicho, y el CRD2, que influye en la afinidad de unión a la proteína de superficie. Se obtuvo información novedosa de la secuencia de *ELR1* en 47 caballos con acervo genético diverso, entre los que se incluyeron caballos positivos y negativos para VAIE. Sin embargo, teniendo en cuenta que este estudio es descriptivo, y si bien ha permitido demostrar la existencia de variaciones alélicas en regiones del gen *ELR1* relevantes para su función, y en grupos heterogéneos de caballos, el análisis aun debería extenderse a otros SNPs en este gen, que podrían revelar conocimiento relevante relacionado con la susceptibilidad a la enfermedad. También, hacer extensivo el estudio a otros grupos de caballos expuestos a VAIE, mediante un análisis de asociación adecuado, que excede el alcance de este trabajo. Por tanto, se reporta la diferenciación genética observada entre grupos, no así una asociación estricta entre marcadores y condición ante AIE.

El *ELR1* juega un papel dual al actuar como receptor de VAIE, pero también en condiciones fisiológicas como receptor de citoquinas. En este

sentido, el hallazgo de isoformas de ELR1 y la evidencia de que la forma soluble de ELR1 pueden competir con su forma asociada a membrana, implica la participación de esta proteína tanto en la infectividad viral como en la respuesta del huésped (Lin *et al.*, 2013). Estas observaciones sugieren que las diferentes variantes alélicas podrían ser relevantes en términos de la respuesta inmune del huésped y la eficiencia de infección del virus.

Al considerar las 7 variantes alélicas detectadas en el exón 3 de *ELR1* y al subdividir la muestra en caballos positivos y negativos para VAIE, la diferenciación genética entre los dos grupos resultó significativa (Tabla 1.16). Sin embargo, es importante remarcar que los SNPs detectados no causan cambios en la secuencia aminoacídica (Tabla 1.15). Estos resultados son en cierto modo esperables, teniendo en cuenta la función predicha de los dominios codificados por el exón 3: una porción del péptido señal y el CRD1, fundamentales en condiciones fisiológicas del huésped. Por otro lado, la función del CRD1 es crucial en términos de la capacidad de unión de ELR1 al gp90 del VAIE (Zhang *et al.*, 2008), por lo que variaciones en la secuencia de aminoácidos de este dominio podrían repercutir en la unión efectiva de ELR1 al gp90 del VAIE.

El exón 5 de *ELR1* codifica el CRD2 y parte del dominio CRD3. A pesar de que CRD2 no es crucial en términos de unión a gp90 de VAIE (Zhang *et al.*, 2008), la composición aminoacídica podría tener impacto en la conformación del receptor y la interacción con la proteína gp90 y, por tanto, influir en la afinidad de unión de la proteína. Teniendo esto en

mente, la ocurrencia de cinco cambios no sinónimos podría estar reflejando los efectos de la variación antigénica, de acuerdo con la noción de que las presiones de selección podrían ser ejercidas por el antígeno y, en consecuencia, conducir a variaciones en la secuencia del receptor. Además, otras fuerzas selectivas podrían estar dirigiendo la ocurrencia y perduración de ciertos niveles de polimorfismo de las variantes de la proteína gp90 de VAIE, más allá de la inmunidad adaptativa (Mealey *et al.*, 2004).

El análisis de los posibles efectos de los cambios aminoacídicos predichos en el exón 5 de *ELR1* mostró dos tipos de consecuencias para las sustituciones no sinónimas: a) un efecto neutral en la función biológica de la proteína y b) un probable efecto deletéreo en la función proteica (Choi y Chan, 2015). Entre este último, el SNP *rs1095048804*, detectado en la secuencia codificante correspondiente a CRD2, está predicho que cause un cambio no sinónimo Arg109Trp con un puntaje PROVEAN = -4,357 que implica que la sustitución afecta la función o estructura proteica. Esto podría estar relacionado con las distintas propiedades del grupo R de estos dos aminoácidos. El cambio de un aminoácido básico como la arginina por un residuo no polar, aromático y voluminoso como el triptófano, y el probable cambio subsecuente en la estructura secundaria proteica, podría interferir con las funciones de los dominios. De hecho, se ha reportado previamente que cambios aminoacídicos que generen impedimento estérico entre el receptor y las glicoproteínas de superficie del virus, resultan en un bloqueo de la entrada

del mismo, confiriendo una suerte de mecanismo de defensa (Bibollet-Ruche *et al.*, 2019). En consecuencia, el papel de este cambio aminoacídico en la integridad y/o funcionalidad del receptor debería investigarse en más profundidad en relación a la interacción funcional y unión a gp90 de VAIE, ya que los cambios funcionales en proteínas que interactúan con ciertos virus podrían resultar en una rápida evolución que podría explicar las diferencias en la respuesta e incluso la progresión de la enfermedad (Barreiro y Quintana-Murci, 2010).

Según Bibollet-Ruche *et al.*, (2019), en lentivirus como el SIV (en inglés; *Simian Immunodeficiency Virus*), la diversidad en los receptores CD4 de las células blanco reduce el ingreso del virus inhibiendo la infección viral y confiriendo cierta protección al hospedador. Asimismo, algunas de las variantes alélicas estimadas para el exón 5 de *ELR1* ocurrieron exclusivamente en el grupo de caballos positivos o negativos para VAIE, pero no en ambos (Tabla complementaria 2, ver ANEXO 4). Estas observaciones plantean la interrogante de si estas variantes podrían estar relacionadas a la progresión de la enfermedad en cuanto a la integridad del receptor y/o modificación de las vías de señalización. No obstante, son necesarios estudios adicionales en otros grupos de caballos afectados y no afectados para VAIE para validar esta hipótesis.

Finalmente, en el análisis comparativo de las secuencias aminoacídicas de *ELR1* entre distintas especies de mamíferos, se observó que ciertos residuos se mantienen conservados a lo largo de todas las variantes alélicas detectadas. Por ejemplo, los CRDs están

conservados en las especies analizadas, con CRD2 y CRD3 presentando los tres enlaces disulfuro intradominio esperados para miembros de la superfamilia de TNFR. Entre las variantes equinas del exón 5, los SNPs *rs1095048805* y *rs1095048806* causan sustituciones sinónimas en dos residuos cruciales: Ser112 y Thr114, los cuales forman una interacción hidrofílica con gp90 por comparación con otros complejos lentivirus – receptor (Qian *et al.*, 2015).

En conjunto, estas observaciones sugieren que a pesar de la ocurrencia de variantes alélicas de *ELR1* en caballos (3 en el exón 3 y 4 en el exón 5), los residuos críticos para la función de la proteína madura están conservados, lo que se explicaría por las presiones selectivas a favor de preservar la función específica de los miembros de TNFRSF y la respuesta inmune del huésped. Sin embargo, las variantes genéticas pueden algunas veces modificar la función biológica de los productos génicos o los niveles de expresión y pueden estar asociados con la susceptibilidad a enfermedades. A pesar de que las poblaciones equinas presentan amplia diversidad genética a lo largo del genoma, la mayor parte de estas variaciones pueden no afectar la función de los genes. De la misma manera, la variación alélica de *ELR1* podría conducir a situaciones diversas y no necesariamente excluyentes: primero, aumentar la diversidad genómica global dentro de una población equina dada; segundo, disminuir la variabilidad del repertorio por la presencia de individuos que carecen de alelos clave y, finalmente, actuar como una posible barrera para la interacción y la afinidad de unión durante las

infecciones virales. Hasta donde sabemos, estos resultados son el primer reporte de la existencia de variantes alélicas que involucran algunos residuos cruciales en ELR1. Esta información podría ser importante como paso inicial para evaluar la relevancia funcional y la relación con la infección por VAIE y la progresión de la enfermedad, así como el desarrollo de estrategias de prevención.

Capítulo 2

Variabilidad genómica en caballos Criollo Argentino

1. INTRODUCCIÓN

1. Introducción

1.1 Arreglos de ADN para análisis genómico

El conocimiento genómico en equinos ha avanzado desde el establecimiento del Taller Internacional para Mapeo Genético Equino y posteriormente con el Proyecto Genoma Equino (*Horse Genome Project*, <http://www.uky.edu/Ag/Horsemap/>). Los primeros mapas genéticos construidos en caballos (Guerin *et al.*, 1999; Milenkovic *et al.*, 2002; Chowdhary *et al.*, 2003; Penedo *et al.*, 2005) permitieron el mapeo de varias enfermedades equinas de origen genético (Valberg *et al.*, 2001; Tryon *et al.*, 2007). En el año 2009, se consigue finalmente el primer ensamblaje del genoma equino de referencia (Wade *et al.*, 2009), y esto, junto al desarrollo de nuevas tecnologías a nivel masivo y de cobertura genómica, ha permitido el avance tanto en términos de información como de tecnologías disponibles para la especie equina. Actualmente, a doce años de la secuenciación del genoma de la yegua purasangre *Twilight*, se encuentran disponibles dos versiones del ensamblaje del genoma equino de referencia: EquCab2.0 (Wade *et al.*, 2009) y EquCab3.0 (Kalbfleisch *et al.*, 2018), así como la secuencia del cromosoma Y (Janečka *et al.*, 2018) y varias plataformas de análisis masivo como los arreglos de ADN y los estudios de secuenciación global del genoma (WGS, en inglés, *Whole Genome Sequencing*).

Con la disponibilidad de información de la secuencia genómica de referencia se espera identificar rasgos genómicos funcionales comunes a

los mamíferos y obtener un mejor conocimiento de las enfermedades que afectan a los equinos. La formación de razas modernas y la presencia de enfermedades con mayor incidencia dentro de ciertas razas (Brosnahan *et al.*, 2010) sugieren que el mapeo puede resultar informativo considerando las diferentes razas de caballos. Además, como los caballos comparten un número de ciertas condiciones médicas (como alergias, artritis, enfermedades respiratorias), el mapeo de genes de enfermedades usando poblaciones equinas podría a su vez beneficiar a la salud humana.

1.2 Antecedentes de datos genómicos en razas de caballos

La disponibilidad de un genoma de referencia (Wade *et al.*, 2009) dio comienzo a la era genómica y con esta las posibilidades de estudio de las distintas poblaciones y razas equinas aumentaron exponencialmente. Desde entonces, se han desarrollado dos arreglos de ADN de mediana densidad: uno con 54.602 SNPs y otro posterior que cuenta con 65.157 SNPs (Illumina) y uno de alta densidad con 670.805 marcadores (Affymetrix) para el genotipado de la especie equina (Schaefer *et al.*, 2017). La tecnología de arreglos de ADN permitió realizar diferentes estudios a nivel genómico como el análisis de la diversidad (McCue *et al.*, 2012; Petersen *et al.*, 2013a; Wang *et al.*, 2014), huellas de selección (Petersen *et al.*, 2013b; Metzger *et al.*, 2015; Grilz-Seger *et al.*, 2019),

estudio de enfermedades (Orr *et al.*, 2010; Go *et al.*, 2012; Han *et al.*, 2020; Rockwell *et al.*, 2020).

El aporte de este estudio en la caracterización de los caballos Criollo Argentino con la visión genómica, es un paso importante y con muchas ventajas futuras para iniciar estudios aplicados en esta raza. Si bien el desarrollo de los estudios genómicos con arreglos de ADN se inició hace varios años, el caballo criollo no ha sido incluido como raza de referencia en las iniciativas internacionales, por lo que este aporte es decisivo en cuanto al conocimiento de un recurso propio y particular de nuestro país.

En este segundo capítulo de la tesis se abordarán los objetivos específicos:

3.2.3 Tipificar y analizar a nivel genómico los caballos Criollo Argentino empleando un arreglo de ADN de mediana densidad.

3.2.4 Realizar el análisis genómico comparativo con caballos de diferentes razas y poblaciones equinas.

2. MATERIALES Y MÉTODOS

2. MATERIALES Y MÉTODOS

2.1 Material animal

Para la segunda etapa de la tesis se utilizaron 24 muestras de pelo de caballos puros de la raza Criollo Argentino (CR) seleccionadas (Figura 2.1) del banco de muestras de CR del IGEVET (N = 150), provenientes de muestreos realizados en las provincias de Tucumán (Cabaña de Caballos Criollos – Asociación Cooperativa INTA, Leales) y Buenos Aires (Cabaña La Clarita, Pergamino). Los equinos muestreados fueron registrados con los siguientes datos cuando disponibles: nombre completo, fecha de nacimiento, sexo, pelaje, condición sanitaria, relaciones de parentesco. Con el objetivo de minimizar niveles de parentesco cercanos, se realizó la reconstrucción de las genealogías a partir de los datos de pedigrí de los Registros Genealógicos de la Sociedad Rural Argentina (<http://www.sra.org.ar/rrgg/>).

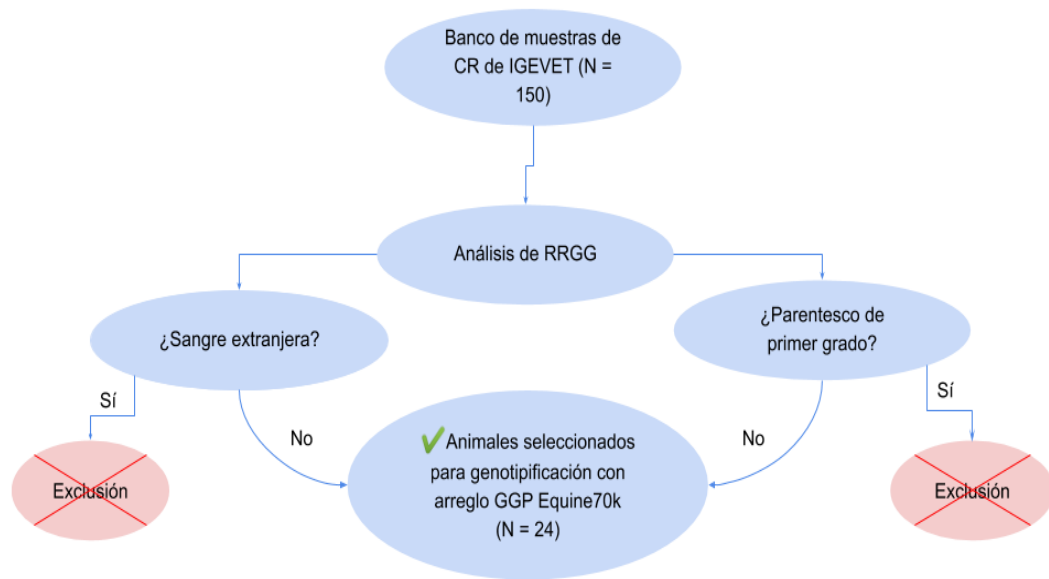


Figura 2.1 Esquema de los criterios aplicados para la selección de muestras a incluir en el genotipado con el arreglo GGP Equine70k.

2.2 Análisis genealógico

El análisis genealógico se realizó utilizando los datos de caballos Criollo Argentino registrados en la Sociedad Rural Argentina (www.sra.org.ar) y el software PEDIG (Boichard, 2002). A partir de la lista de muestras de la raza Criollo Argentino disponibles en el banco de ADN del IGEVET se recuperaron los pedigríes de la base datos de la SRA conteniendo información desde el año 1944 hasta 2012. Los registros redundantes fueron excluidos aplicando un filtro por nombre y año de nacimiento; y se completaron los parentescos faltantes entre animales.

Finalmente, este rastreo devolvió un pedigrí revisado con 968 registros de caballos nacidos en Argentina entre 1971 y 2001, a partir de los que se calculó la población de referencia. Aquellos individuos sin padres conocidos fueron tomados como fundadores, mientras que aquellos animales —fundadores o no— con el mayor potencial de contribuir genéticamente a la población de referencia se definieron como ancestros (Boichard *et al.*, 1997). Para la población de referencia se calculó el número efectivo de animales (N_e), el número efectivo de fundadores (f_e) (James, 1972) y el número efectivo de ancestros (f_a) (Boichard *et al.*, 1997). El número efectivo de fundadores (f_e) se define como el número de fundadores que, a igual contribución, producen la diversidad genética existente en una población (Lacy, 1989). Por otra parte, el f_a se define como el número mínimo de ancestros, no necesariamente fundadores, que explica la variabilidad genética total de una población (Boichard *et al.*, 1997). Este parámetro es complementario a f_e ya que explica la pérdida de diversidad genética debida a los cuellos de botella que haya experimentado la población en estudio (por ejemplo, generados por el uso desbalanceado de reproductores en la población estudiada, inseminación artificial, etc.). Para evitar sobreestimaciones de la diversidad genética, para calcular f_a se debe considerar únicamente la contribución de un ancestro que no se explica por otros ancestros considerados previamente (contribución genética marginal).

El análisis de consanguinidad se abordó tomando como base dos tipos de datos: el coeficiente de consanguinidad (F) por individuo se

estimó a partir de los datos de pedigrí, y la estimación de la probabilidad de identidad por estado (en inglés, *identity by state*, IBS) compartida y de los coeficientes de parentesco se realizó a partir de los genotipos obtenidos de los 61.175 SNPs, utilizando el software King 2.1 (Manichaikul *et al.*, 2010).

2.3 Genotipificación con arreglo de mediana densidad

Las muestras de pelo de los animales seleccionados se enviaron a Neogen Corporation (Geneseek; Lincoln, NE, EEUU) para la extracción de ADN y genotipificación. Como plataforma de genotipificación se utilizó el microarreglo de mediana densidad GGP Equine70k (Illumina, San Diego, USA) que contiene 65.157 SNPs distribuidos uniformemente y derivados de la colección de EquCab2.0 compilada por *The Broad Institute at MIT and Harvard*. Este conjunto de sondas fue obtenido a partir de la secuenciación de 7 razas equinas representativas de caballos modernos y ancestrales. Este arreglo de segunda generación presenta sondas con un espaciado promedio de 40 kb distribuidas en los 31 autosomas y ha sido validado en 15 razas modernas de caballos con un valor promedio de frecuencia del alelo menor de 0,21 (<https://genomics.neogen.com/en/ggp-equine70k>).

2.3.1 Control de calidad de la genotipificación

El control de calidad de los datos obtenidos a partir de un arreglo de SNPs juega un papel central con respecto a la precisión y exactitud de los análisis a realizar posteriormente. Para el análisis inicial de los datos genómicos, los datos crudos obtenidos a partir del microarreglo de SNPs fueron extraídos utilizando el software GenomeStudio (Illumina) con un umbral mínimo de *GenCall* (GC) de 0,15 como sugiere el fabricante para productos con tecnología *Infinium*. El valor GC no es una probabilidad sino un puntaje mediante el cual se ranquean y excluyen genotipos fallidos. De esta manera, se obtuvieron los archivos «.PED», el cual contiene los genotipos y «.MAP» que contiene la posición genómica de los SNPs. A continuación, se detalla el control de calidad y procesamiento de datos realizado con el software PLINK v1.9 (Purcell *et al.*, 2007).

2.3.2 Control de calidad de los datos

En primer lugar, se aplicó a los archivos .PED y .MAP el filtro recomendado para evaluar la calidad global por muestra denominado *call rate* de muestra, definido como la proporción de genotipos asignados exitosamente por muestra sobre el número total de SNPs del conjunto de datos. El valor mínimo de *call rate* por muestra utilizado para todos los análisis de esta tesis fue de 90 % y para implementar este filtro se aplicó el parámetro *--mind* 0.10 de PLINK (Purcell *et al.*, 2007). Una vez excluidas las muestras que no cumplieran con este criterio, se prosiguió

con el control de calidad de los SNPs. Se estableció como valor de corte un *call rate* de 90 %, para lo que se aplicó el parámetro *--geno 0.10* con el que se filtraron aquellas SNPs con error de genotipado por encima del 10 % en el total de las muestras. De esta manera, se definió el «DATASET 1» (DS1) conteniendo 61175 SNPs (ANEXO 2).

2.3.3 Análisis de la diversidad genómica

Para el cálculo de medidas básicas de diversidad genómica se utilizó el DS1 y con el software PLINK v1.9 (Purcell *et al.*, 2007) se determinaron: la heterocigosis esperada (H_e), equilibrio de Hardy-Weinberg (HWE), estadístico F de Wright.

2.3.3.1 Heterocigosis esperada

Un indicador de la variación genética de una población es la Heterocigosis (H), la cual representa la probabilidad de que un individuo sea heterocigota para ese locus. Está fuertemente relacionada con el número de alelos presentes y su frecuencia en la población. La heterocigosis es una medida de la variación genética de la población en estudio y adopta valores entre 0 y 1. Cuanto más próxima sea a 1, mayor será la diversidad genética presente y mayor será es la información que aporte el marcador acerca de la transmisión de alelos a la siguiente

generación. Las heterocigosis esperada (H_e) y heterocigosis observada (H_o) fueron calculadas mediante el comando (`--hwe`).

2.3.3.2 Equilibrio de Hardy-Weinberg

La Ley de Hardy-Weinberg es un concepto central en genética de poblaciones y permite inferir qué sucederá con las frecuencias genotípicas de cierto locus tras una generación de apareamiento al azar. La ley se aplica para poblaciones diploides con reproducción sexual, asumiendo los siguientes supuestos: generaciones discretas, es decir que no exista solapamiento entre las mismas; la población es panmíctica, es decir que los apareamientos sean al azar; tamaño poblacional infinito de manera de minimizar las diferencias entre individuos; ausencia de mutación, selección y migración, o sea que no haya ganancia ni pérdida de alelos. De esta manera, bajo las condiciones antes estipuladas y luego de un ciclo de apareamiento al azar, la población se mantendrá en equilibrio génico. El equilibrio de Hardy-Weinberg fue evaluado mediante el comando (`--hwe`).

2.3.3.3 Estadísticos F de Wright

Los estadísticos F jerárquicos de Sewall Wright (1969), también llamados índices de fijación, describen el nivel de heterocigosis en una población, más específicamente cuánto se reduce la heterocigosis en

relación a lo esperado según el equilibrio Hardy-Weinberg. La endogamia puede medirse entonces en términos de la reducción en la proporción de genotipos heterocigotas, es decir, en función de H . De esta manera los estadísticos F también constituyen una medida de la endogamia según la ecuación: $F = (H_o - H_e) / H_o$. Los coeficientes individuales de consanguinidad (F) fueron estimados utilizando el comando (`--het`) de PLINK v1.9 (Purcell *et al.*, 2007).

2.3.4 Desequilibrio de ligamiento

Se define el desequilibrio de ligamiento como la asociación no al azar de alelos en dos o más loci diferentes (inter-alélicas) y se evidencia por una alteración en las frecuencias esperadas según la Ley de distribución independiente Mendel. Físicamente, el LD se debe a dos alelos ubicados muy próximos entre sí haciendo muy poco probable la ocurrencia de recombinación, lo que conduce a que estos se hereden juntos, «en bloque». La variación del LD fue evaluada mediante el cálculo del valor de r^2 entre todos los pares de SNPs en cada cromosoma equino, tomando distancias entre 100 y 1000 Kb, en ventanas de a 100 Kb mediante los comandos `--r2` y `--ld-window-kb` de PLINK v1.9 (Purcell *et al.*, 2007). Mediante el parámetro `--recode tab` de PLINK se generaron los archivos `.ped` y `.info` de cada cromosoma, los cuales fueron utilizados para poder visualizar los haplotipos mediante la función del gráfico LD del programa Haploview (Barret *et al.*, 2005). El análisis se realizó por

cromosoma debido a que la gran cantidad de información implica un elevado costo en términos computacionales debido a la alta utilización de memoria y entonces sólo fuese posible cargar de a un cromosoma por vez. Para estos análisis se utilizó el «DATASET 2» (DS2), conteniendo 44606 SNPs (ANEXO 2).

2.3.5 Detección y análisis de variaciones estructurales

La detección de la variación del número de copias (en inglés, *Copy Number Variation*, CNV) en los cromosomas autosómicos se realizó mediante el software PennCNV (Wang *et al.*, 2007), utilizando el DS1 conteniendo 61.175 SNPs. En primer lugar, se obtuvo la información de intensidad de la señal, expresada como el log ratio de R (LRR) y la frecuencia del alelo B (en inglés, *B Allele Frequency*, BAF) de cada SNP con el software Genome Studio 2.0 de Illumina. Los archivos conteniendo la frecuencia del alelo B en la población (en inglés, *Population Frequency of the B allele*, PFB) se generaron en base al valor promedio de BAF de cada marcador en la población. A fin de corregir el efecto del contenido en GC, se generó un archivo modelo GC equino calculando el contenido GC en 1 Mb de la región genómica alrededor de cada marcador (500 Kb hacia cada lado). Las ondas genómicas se ajustaron usando la opción --gcmode1, según Wang *et al.*, (2012). Las muestras que presentaron una desviación estándar del LRR > 0,35, BAF > 0,05 y un factor de contenido de gc > 0,01 fueron excluidas. Finalmente, las CNV se definieron como

segmentos conteniendo tres o más SNP consecutivos. Se definió una región de CNVs (en inglés, *Copy Number Variation Region*, CNVR) como la superposición de al menos un par de bases entre CNVs como se reportó en estudios previos (Redon *et al.*, 2006; Liu *et al.*, 2013). El software CNVRuler (Kim *et al.*, 2012) fue implementado para la derivación de la CNVR y para minimizar la sobreestimación del tamaño de la CNVR. Las CNVR fueron clasificadas como eventos de ganancia y pérdida de acuerdo con la ocurrencia de duplicaciones o eliminaciones dentro de la región definida, respectivamente. Para minimizar falsos positivos en la detección de las CNVR, se implementó el parámetro del umbral regional de densidad (recurrencia) y las regiones genómicas con una densidad de CNV por debajo del 10 % se excluyeron durante el proceso de empalme del software CNVRuler usando el valor por defecto (0,1). Además, se usó la opción «región separada ganancia/pérdida» para concatenar la CNV en función del genotipo.

2.3.5.1 Análisis funcional y enriquecimiento génico

El porcentaje de enriquecimiento génico en los cromosomas equinos fue calculado como se describió anteriormente (Ghosh *et al.*, 2014). El análisis del contenido de genes en la CNVR se realizó con el programa BioMart (Kasprzyk *et al.*, 2011) y utilizando la base de datos Ensembl Genes 93 - Horse (EquCab 2.0). También se implementó la minería de datos manual como análisis complementario del enriquecimiento de genes

en la CNVR usando el buscador NCBI Genome Data Viewer (<http://www.ncbi.nlm.nih.gov/genome/gdv/>).

La lista de genes recuperada con BioMart fue remitida al sistema de clasificación PANTHER (en inglés *Protein ANalysis THrough Evolutionary Relationships*) (Mi *et al.*, 2016) con el fin de agrupar y clasificar genes según el proceso biológico, la función molecular y la clase proteica. También se consultó la base de datos OMIA (en inglés *Online Mendelian Inheritance in Animals*) para los genes equinos obtenidos (<http://omia.angis.org.au/>). Dado que el número de genes equinos anotados en la base de datos *Gene Ontology* (GO) es todavía limitado, la lista de genes equinos ENSEMBL obtenida fue convertida a su gen ortólogo humano ENSEMBL usando el software Biomart.

Posteriormente, la lista obtenida de homólogos humanos se estableció como *background* para el análisis de GO y de las vías metabólicas realizado con el software DAVID v6.7 (en inglés *Database for Annotation, Visualization and Integrated Discovery*) (Huang *et al.*, 2009a, 2009b).

2.4 Análisis genético interracial

2.4.1 Muestras y conjuntos de datos

Para el análisis de la diversidad genómica interracial, se incorporaron datos de 176 caballos de 8 razas: Andaluz (AND, N = 18),

Árabe (ARA, N = 24), Islandés (ICE, N = 25), Miniatura (MIN, N = 21), Percherón (PRC, N = 23), Paso Peruano (PER = 21), Shetland (SHE, N = 27) y Sangre Pura de Carrera (SPC, N = 17) tipificados con el arreglo SNP50 Beadchip (Illumina) por Petersen *et al.*, (2013a) y de acceso libre en el Repositorio de Datos de la Comunidad del Programa Nacional de Investigación del Genoma Animal (*National Animal Genome Research Program*, NAGPR, <https://www.animalgenome.org/>). Utilizando el entorno de programación RStudio (RStudio Team, 2020), se obtuvieron las sondas en común entre los arreglos GGP Equine70k y SNP50 Beadchip a partir de las cuales se construyó el archivo para el análisis comparativo interracial mediante el software PLINK (Purcell *et al.*, 2007). Finalmente, el conjunto de datos consistió en los genotipos de 200 caballos pertenecientes a 9 razas equinas en 42.357 sondas distribuidas en los 31 cromosomas autosómicos (ANEXO 3).

Previo al análisis de la diferenciación genética interracial y de la estructura poblacional se utilizó el comando `--indep-pairwise (50 5 0,2)` en PLINK v1.9 (Purcell *et al.*, 2007) para depurar el set de datos y evitar la influencia de los *clusters* de SNPs en los subsiguientes análisis (Zheng *et al.*, 2012; Moorjani *et al.*, 2013). Este parámetro considera la correlación entre cada par de SNPs en ventanas de 50 SNPs, removiendo un par de SNPs si el LD entre ellos es mayor que 0,2. Luego, desplaza esa ventana 5 SNPs y repite el proceso. De esta manera, el DATASET 3 (DS3) quedó conformado por los genotipos de 200 caballos pertenecientes a 9 razas

equinas, filtrados por *--maf* 0,05, desequilibrio de ligamiento (*--indep-pairwise*), conteniendo 19.614 SNPs (ANEXO 3).

2.4.2 Diferenciación genética

Para analizar la diferenciación genética entre las razas se estimó el índice F_{ST} entre las 9 poblaciones tomadas de a pares, a partir del DS3 con el software ARLEQUIN v3.5 (Excoffier *et al.*, 2010), utilizando la distancia de Reynolds (Reynolds *et al.*, 1983) y 20000 permutaciones para el test de significancia ($\alpha = 0,05$).

2.4.3 Análisis de varianza molecular (AMOVA)

El análisis de la diversidad genética dentro y entre las poblaciones fueron estimados mediante un análisis jerárquico de la varianza molecular (AMOVA) (Excoffier *et al.*, 1992) utilizando el programa ARLEQUIN v3.5 (Excoffier *et al.*, 2010) y el DS3. El nivel de significancia para los componentes de la varianza fue evaluado implementando 1000 permutaciones para obtener el índice de diferenciación genética entre poblaciones (F_{ST}), el coeficiente de endogamia en las poblaciones (F_{IS}) y el coeficiente de endogamia total (F_{IT}).

2.4.4 Análisis de la estructura poblacional

La estructura de la población se evaluó mediante el programa fastSTRUCTURE v2.3.4 (Raj *et al.*, 2014), el cual implementa un algoritmo rápido para analizar conjuntos de datos grandes como los provenientes de genotipos de SNPs. Este programa utiliza un modelo de análisis Bayesiano en el que se asumen K poblaciones ancestrales para explicar la diversidad genética observada. Los datos filtrados por LD (DS3), se utilizaron para correr el programa fastSTRUCTURE utilizando la configuración por defecto y el modelo de ancestría *admixture* con correlación de frecuencias alélicas y un rango de valores de K entre 2 y 15. El análisis fue realizado con un período inicial de *burnin* de 50000 iteraciones para minimizar el efecto de la configuración inicial, seguidas de 2000000 de repeticiones de cadenas de Markov Monte Carlo (MCMC) para cada valor de K .

La estructura poblacional también fue evaluada mediante el análisis de componentes principales (PCA, en inglés *Principal Component Analysis*) que condensa la información de los genotipos en frecuencias alélicas para identificar las principales variables y explicar la varianza observada. De esta manera, si se eligen los dos componentes principales (PC1 y PC2), cada individuo es representado por un punto sobre los ejes de variación (Patterson *et al.*, 2006). El análisis se realizó con el comando *--pca* en el software PLINK, el cual construye una matriz de distancias genéticas entre pares de alelos con identidad en estado (en inglés *identity by state*, IBS). Para este análisis se utilizó el DATASET 4 (DS4) filtrado

únicamente por $MAF = 0,05$, conteniendo 38.199 SNPs. La representación gráfica de estos resultados se realizó en el entorno RStudio (RStudio Team, 2020).

3. RESULTADOS

3. RESULTADOS

3.1 Análisis genealógico

El análisis de los pedigríes se realizó con el fin de evaluar la variabilidad genética dentro de la población de caballos criollos a nivel genealógico, y con esta información de base, seleccionar un conjunto de individuos no emparentados y de diversas líneas de sangre definidas históricamente, para obtener un grupo de animales representativo de la raza para el análisis a nivel genómico. Así, la base de datos fue conformada a partir de todos los caballos la raza Criollo Argentino del banco de ADN del IGEVET, excluyendo registros redundantes (aplicando un filtro por nombre y año de nacimiento) y completando los parentescos faltantes entre animales. Se seleccionaron aquellos animales que cumplieran con los siguientes requisitos: menor grado de parentesco posible (se excluyeron caballos que compartieran antecesores en primer y segundo grado), presencia de sangres extranjeras, pertenencia a distintas líneas de sangre. Finalmente, la base de datos depurada devolvió un pedigrí revisado con 968 registros de caballos nacidos en Argentina conteniendo información desde el año 1944 hasta 2012. El análisis de consanguinidad de los 968 animales, mostró que 357 de los mismos están emparentados, con un valor promedio de consanguinidad de 0,041.

Por otra parte, se observó un aumento del valor promedio de la consanguinidad de 1,14 (generación 2) a 14,94 (generación 8). Se definió una única población de referencia (animales que potencialmente

contribuyeron a la siguiente generación), conformada por 280 caballos registrados en la raza Criollo Argentino, con el primer y último año de nacimiento en 1971 y 2001, respectivamente. El número estimado de fundadores fue de 235 individuos, con un número efectivo de fundadores (f_e) igual a 157,7. El análisis de la probabilidad del origen genético (*probability of gene origins*) mostró que 50 ancestros principales (f_a = número efectivo de ancestros) contribuyeron a la población de referencia (Figura 2.2).

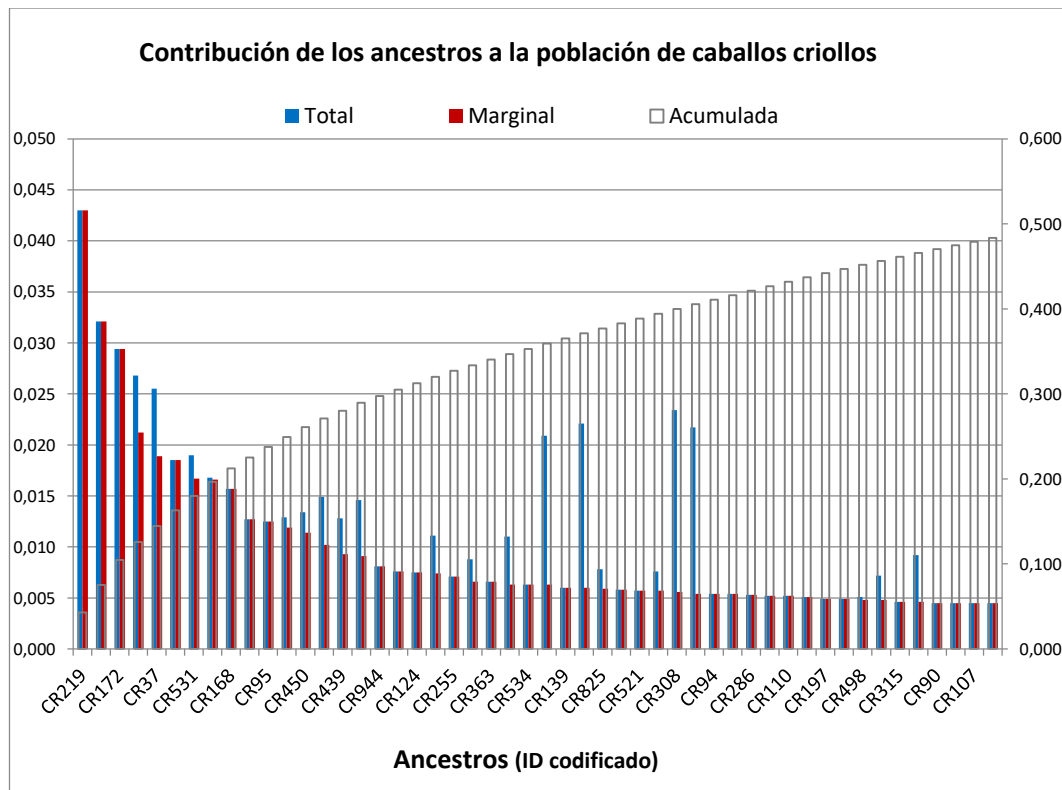


Figura 2.2. Contribución de los ancestros principales a la población actual de caballos criollos.

3.2 Análisis Genómico (genotipificación con arreglo de mediana densidad)

Luego del control de calidad aplicado a las muestras y a los datos crudos obtenidos con el arreglo de mediana densidad GGP Equine70k, la cantidad de SNPs informativos remanentes en los 24 caballos Criollo Argentino fue de 44.606 (MAF > 0,05).

3.2.1 Diversidad genómica y equilibrio de Hardy-Weinberg

En la Tabla 2.1 se observan los valores promedio obtenidos para los parámetros: heterocigosis esperada (H_e), heterocigosis observada (H_o), frecuencia del alelo menor (MAF), p -valor del Equilibrio de Hardy Weinberg (HWE), número de SNPs remanentes y la tasa de genotipado, para cada cromosoma en la población analizada. Se observó que la diversidad genética evaluada en términos de H_e fue menor en ECA11 ($H_e = 0,315$) y máxima en ECA22 ($H_e = 0,353$). En cuanto al valor de MAF, este varió entre 0,248 (ECA26) y 0,264 (ECA22), con un valor promedio de 0,254. Los marcadores analizados se encuentran en equilibrio de Hardy - Weinberg. La tasa de genotipado promedio fue de 0,994 (Tabla 2.1).

Tabla 2.1.

Valores medios de los parámetros de diversidad en el Criollo Argentino analizados por cromosoma a partir del DS1 conteniendo 61.175 SNPs.

ECA	Ho	He	MAF	HWE	N.º SNPs	Tasa de genotipado
1	0,349	0,344	0,253	0,713	3489	0,995
2	0,354	0,345	0,254	0,688	2493	0,992
3	0,344	0,349	0,258	0,677	2171	0,994
4	0,344	0,343	0,252	0,704	2126	0,994
5	0,347	0,344	0,251	0,699	2016	0,995
6	0,343	0,343	0,251	0,713	1793	0,995
7	0,353	0,345	0,254	0,717	1917	0,994
8	0,359	0,351	0,261	0,686	2015	0,994
9	0,349	0,348	0,257	0,693	1691	0,995
10	0,348	0,343	0,253	0,704	1670	0,994
11	0,320	0,315	0,260	0,715	1334	0,995
12	0,365	0,348	0,255	0,690	602	0,993
13	0,339	0,334	0,245	0,750	743	0,995
14	0,334	0,342	0,251	0,688	1891	0,994
15	0,355	0,343	0,252	0,707	1875	0,994
16	0,344	0,345	0,254	0,699	1817	0,994
17	0,348	0,341	0,249	0,692	1668	0,995
18	0,353	0,346	0,255	0,713	1568	0,995
19	0,342	0,347	0,257	0,692	1234	0,994
20	0,351	0,346	0,254	0,695	1272	0,994
21	0,339	0,340	0,251	0,705	1241	0,995
22	0,367	0,353	0,264	0,709	1038	0,993
23	0,367	0,350	0,258	0,723	1106	0,995
24	0,348	0,350	0,258	0,689	1006	0,994

25	0,354	0,348	0,251	0,712	778	0,993
26	0,341	0,340	0,248	0,702	668	0,994
27	0,344	0,346	0,253	0,712	813	0,995
28	0,342	0,344	0,253	0,675	905	0,995
29	0,362	0,346	0,253	0,709	564	0,995
30	0,348	0,335	0,251	0,697	628	0,995
31	0,346	0,347	0,259	0,682	474	0,992
Promedio	0,348	0,344	0,254	0,702	1439	0,994

Nota: Valores medios de heterocigosis esperada (He), heterocigosis observada (Ho), frecuencia del alelo menor (MAF), *p*-valor del Equilibrio de Hardy Weinberg (HWE), número de SNPs remanentes y la tasa de genotipado, para cada cromosoma en la población de Caballos Criollos analizada.

3.2.2 Análisis de consanguinidad

3.2.2.1 Endogamia individual genómica (F)

Los coeficientes de endogamia (F) fueron computados para cada individuo de la población mostrando valores entre -0,076 y 0,100 (Figura 2.3), con un valor medio de $-0,0097 \pm 0,0424$.

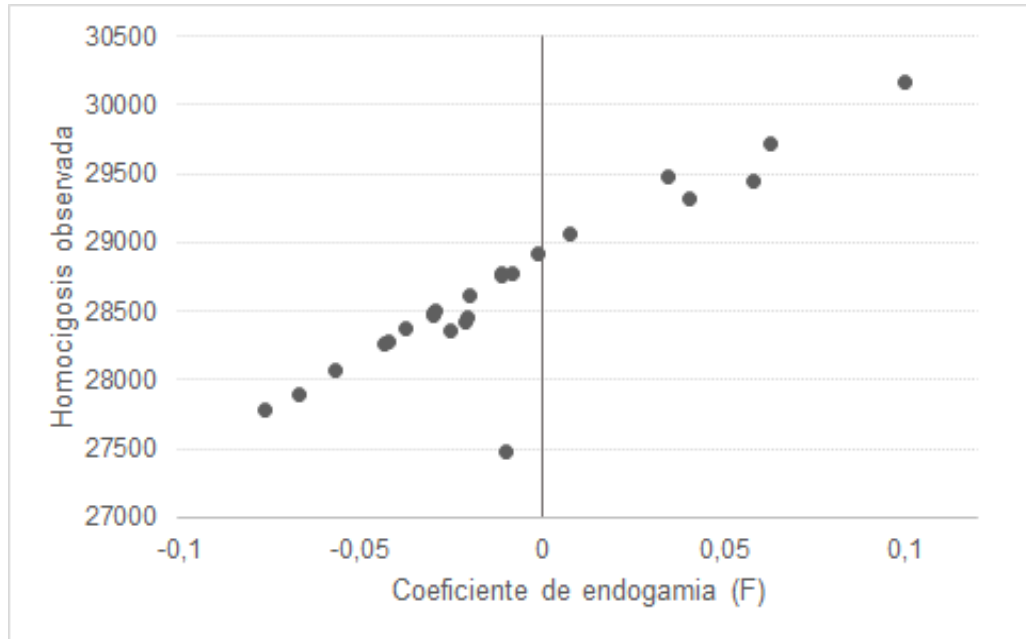


Figura 2.3. Valores del coeficiente de endogamia individual (F) respecto a la homocigosis observada en los 24 caballos Criollo Argentino analizados.

3.2.2.2 Estimación de los coeficientes de parentesco genómico

La estimación de los coeficientes de parentesco se realizó a partir de los 61.175 SNPs (DS1), por medio de la estimación del estado de identidad compartida (en inglés, *identity by state sharing, IBS*) con el software King v2.1 (Manichaikul *et al.*, 2010).

La figura 2.4 muestra el coeficiente de parentesco estimado en función de la proporción de identidad no compartida ($IBS = 0$). En este caso la estimación del coeficiente de parentesco involucra datos de los

SNPs tomando los individuos de a pares. Los coeficientes de parentesco negativos, indican que no hay relaciones de parentesco.

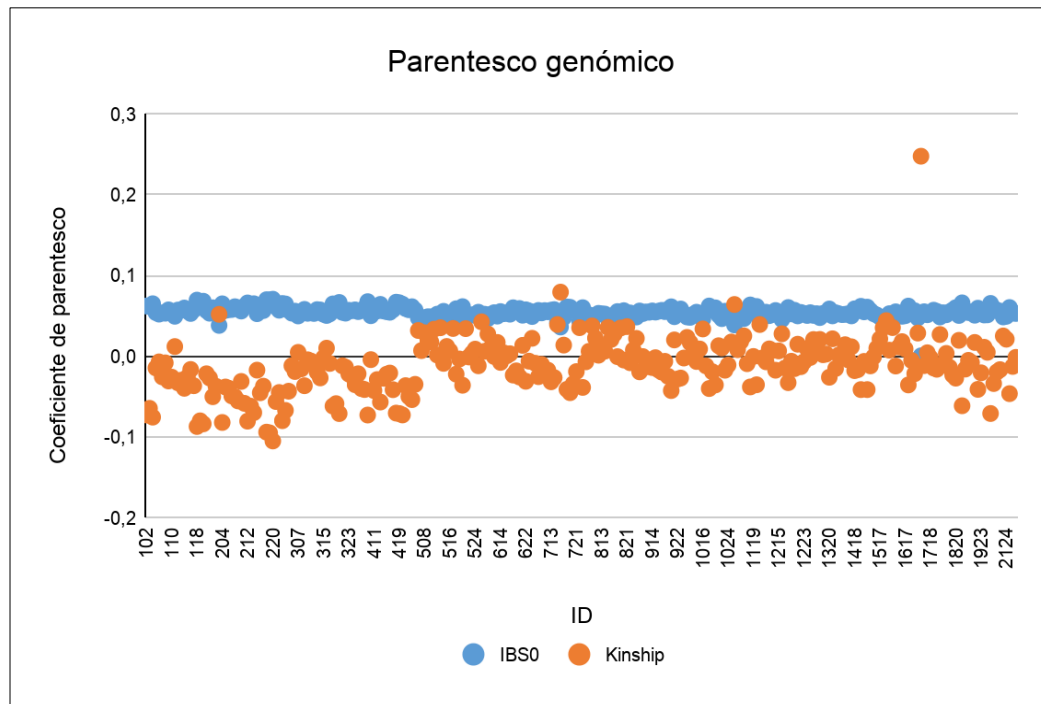


Figura 2.4. Coeficiente de parentesco genómico estimado versus la proporción de SNPs que no comparten identidad por estado (IBS0).

3.2.3 Desequilibrio de ligamiento (LD)

El desequilibrio de ligamiento decayó conforme aumentaba la distancia entre pares de SNPs, siendo el declive más marcado en el entorno de los primeros 200 kb. A partir de los 300 kb el valor medio de r^2 disminuyó lentamente hasta casi estabilizarse (Figura 2.5).

El análisis de haplotipos mostró que 44.606 SNPs fueron concentrados en 2522 haplobloques cuyos tamaños variaron en un rango entre 1 y 199,175 kb en los 31 autosomas.

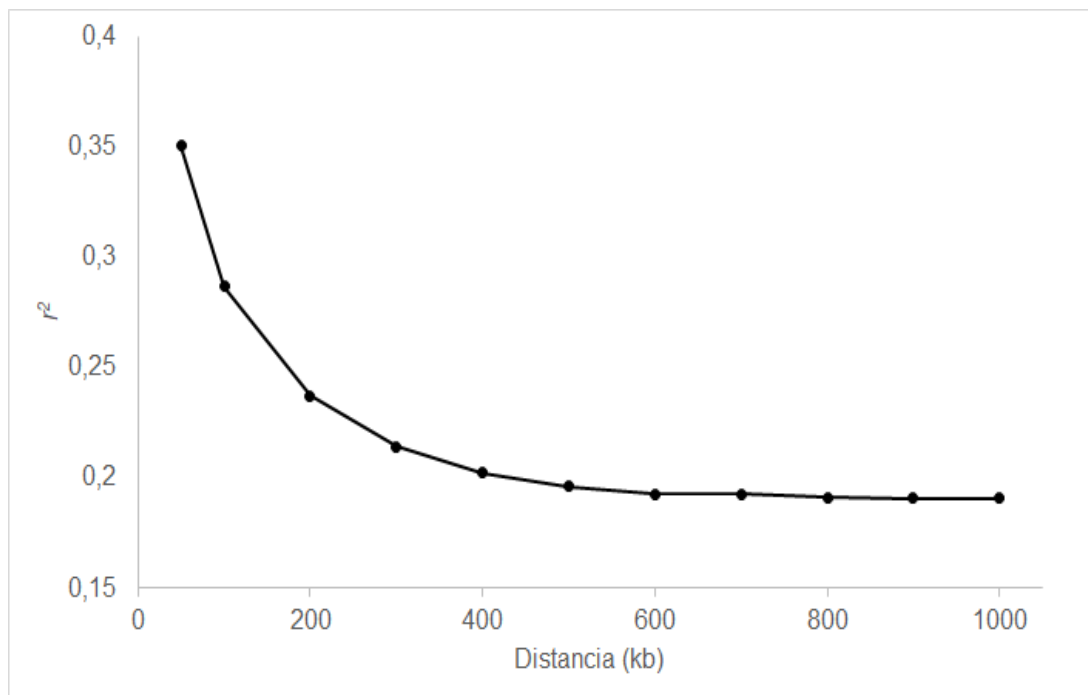


Figura 2.5. Decaimiento del desequilibrio de ligamiento (r^2) promedio entre pares de marcadores en función de la distancia genética evaluada cada 100 kb en la población de caballos Criollo Argentino.

3.2.4 Variaciones genómicas estructurales

Se detectaron 165 CNVs autosómicas en la población de 24 caballos Criollo Argentino, con un promedio de 6,8 detecciones por animal. Del total de las CNVs identificadas, 68 resultaron eventos del tipo pérdida, mientras que 97 representaron ganancia de copias. La longitud promedio de las CNVs fue 119 Kb, donde aproximadamente el 41 % de las CNVs varió entre 100 y 500 Kb, 18 % entre 50 y 100 Kb y un 18 % correspondían a CNVs pequeñas (<10 Kb). No se detectaron CNVs en los cromosomas: ECA5, ECA17, ECA23, ECA27, ECA29 y ECA30 (Tabla complementaria 1, ver ANEXO 4).

La agrupación de las CNVs solapantes resultó en la identificación de 87 CNVRs autosómicas, las cuales representan una cobertura de 13,69 Mb (0,60 %) del genoma equino (Tabla 2.2). El tamaño promedio de las CNVRs fue de 141,3 Kb, fluctuando entre 515 pb a 2 Mb. Aproximadamente el 70,5 % de las regiones presentaron un tamaño entre 50 y 500 Kb, mientras que el 19,3 % fueron más pequeñas que 10 Kb. Se distribuyeron en todos los autosomas a excepción de: ECA5, ECA17, ECA23, ECA27, ECA29 y ECA30. La cobertura en cada cromosoma varió entre 0,001 y 5,79 %, siendo ECA12 el que mostró el porcentaje más alto (Tabla 2.2). Con respecto al tipo de CNVR, 48 (55,2 %) correspondieron a eventos de ganancia (rojo), mientras que 39 (44,8 %) resultaron eventos de pérdida de número de copia (verde) (Figura 2.6). No se detectaron eventos del tipo mixto.

Tabla 2.2

CNVRs detectadas en la población de caballos Criollo Argentino.

ECA	CNVRs	Long. CNVR (pb)	Longitud ECA (Mpb)	%	Genes	CNVRs génicos	CNVRs inter- génicos
1	13	2706811	188,26	1,44	67	10	3
2	3	595945	121,35	0,49	3	1	2
3	8	1126356	121,35	0,93	26	7	1
4	2	585599	109,46	0,53	3	1	1
5	0	0	96,76	0	0	0	0
6	5	533224	87,23	0,61	17	5	0
7	5	895504	100,79	0,89	32	5	0
8	1	36693	97,56	0,04	1	1	0
9	1	515	85,79	0,001	1	1	0
10	5	498494	85,16	0,59	11	4	1
11	3	516190	61,68	0,84	8	2	1
12	2	2140457	36,99	5,79	74	1	1
13	3	410603	43,78	0,94	3	2	1
14	3	207909	94,60	0,22	11	2	1
15	4	277965	92,85	0,30	4	1	3
16	4	319821	88,96	0,36	9	4	0
17	0	0	80,72	0	0	0	0
18	7	722756	82,64	0,87	12	5	2
19	2	160167	62,68	0,26	1	1	1
20	4	336814	65,34	0,52	5	2	2
21	3	340637	58,98	0,58	4	2	1
22	1	33903	50,93	0,07	0	0	1
23	0	0	55,56	0	0	0	0
24	1	3238	48,29	0,01	1	1	0
25	2	732297	40,28	1,82	32	2	0
26	1	71260	43,15	0,17	4	1	0
27	0	0	40,25	0	0	0	0
28	2	274869	47,35	0,58	7	2	0
29	0	0	34,78	0	0	0	0
30	0	0	31,40	0	0	0	0
31	2	162862	26,00	0,63	3	2	0
Total	87	13690889	2280,92	19,44	339	65	22

La distribución cromosómica de las CNVRs génicas e intergénicas en el genoma equino se muestra en la figura 2.7. Mientras los cromosomas ECA6, ECA7, ECA8, ECA9, ECA16, ECA24, ECA25, ECA26, ECA28 y ECA31 presentaron solo CNVRs génicas, la única CNVR detectada en ECA22 fue de tipo intergénica. El resto de los cromosomas analizados presentaron CNVRs tanto génicas como intergénicas, mientras seis cromosomas (ECA5, ECA17, ECA23, ECA27, ECA29 y ECA30) no presentaron ninguna CNVR (Tabla 2.2, Figuras 2.6 y 2.7).

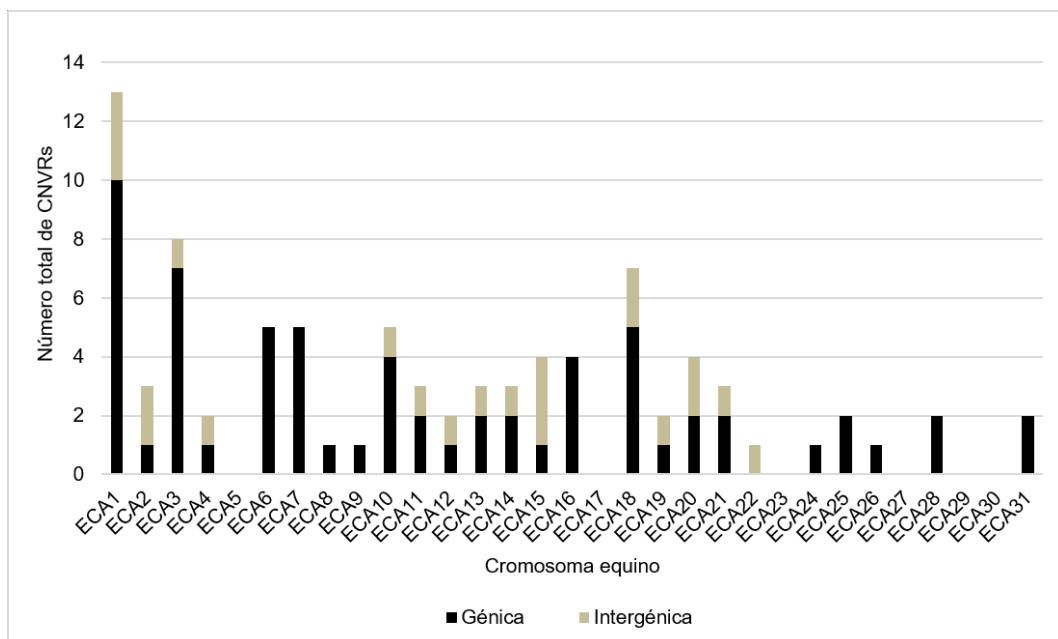


Figura 2.6. Distribución cromosómica de las CNVRs génicas (en negro) e intergénicas (en gris).

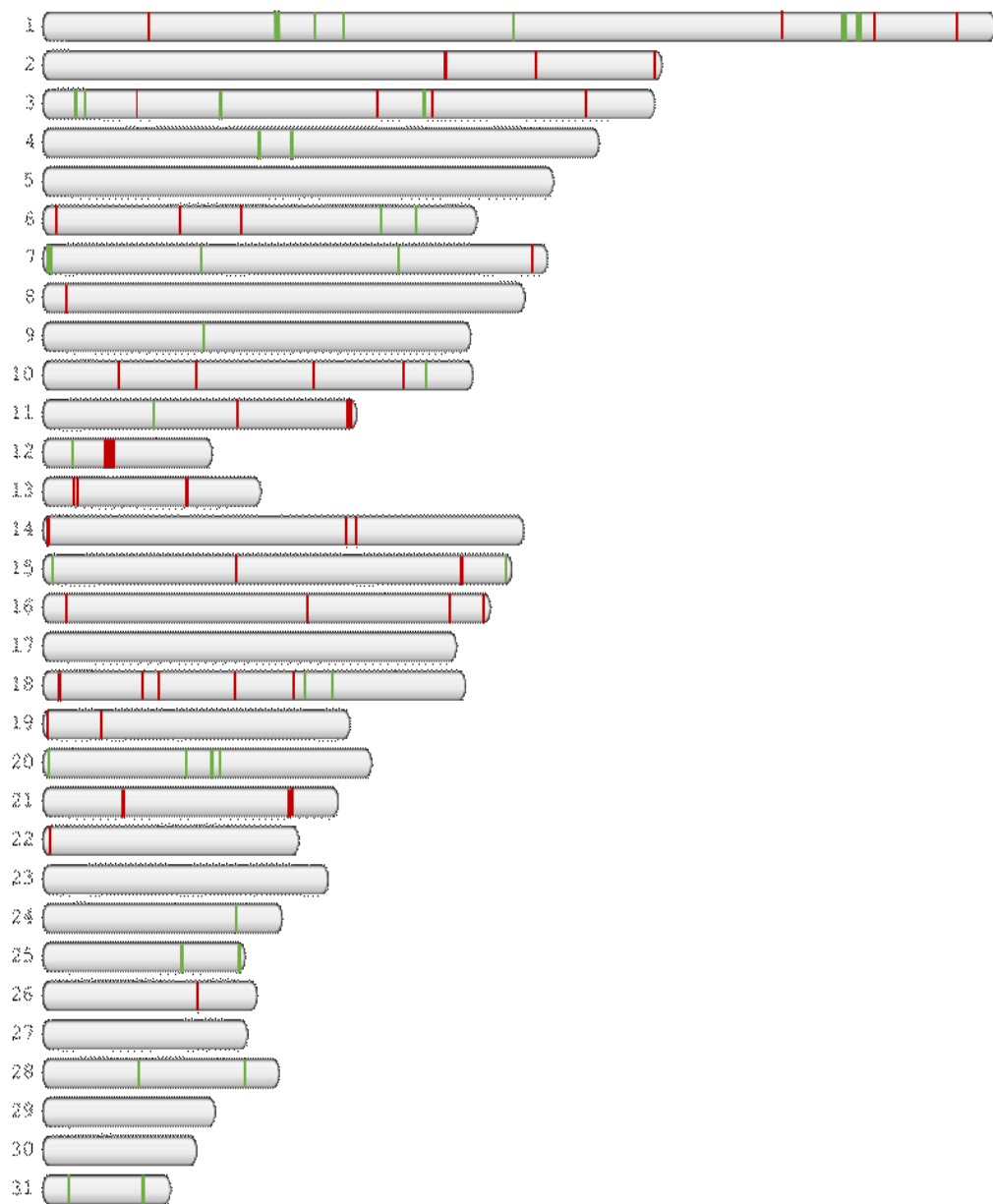


Figura 2.7. Mapa de las CNVRs detectadas en la población de 24 caballos Criollo Argentino. Las líneas en color rojo representan eventos de ganancia en el número de copias y las líneas en color verde corresponden a eventos de pérdida.

3.2.4.1 Análisis funcional y enriquecimiento génico

Todas menos una de las 87 CNVR detectadas se localizaron en regiones génicas, conteniendo al menos un gen equino reportado en la base de datos Ensembl (<https://www.ensembl.org>). Se extrajeron en total 337 genes, incluyendo 286 genes codificantes, 38 pseudogenes y 13 especies de ARNs no codificantes. El análisis con el software PANTHER (Figura 2.8) mostró que los mayores *clusters* de genes enriquecidos correspondían a la actividad de: transducción de señales (receptores olfatorios), actividades receptoras (receptores acoplados a proteína G, receptores de glutamato, receptores transmembrana, receptores tirosin quinasa, receptores de citoquinas) y actividad de unión a proteínas, ácidos nucleicos y cromatina. Por último, la clase proteica más representada (26,4 %) fue la de las proteínas de defensa, tales como la subfamilia de receptores de inmunoglobulinas, inmunoglobulinas y el antígeno del complejo mayor de histocompatibilidad (Figura 2.8).

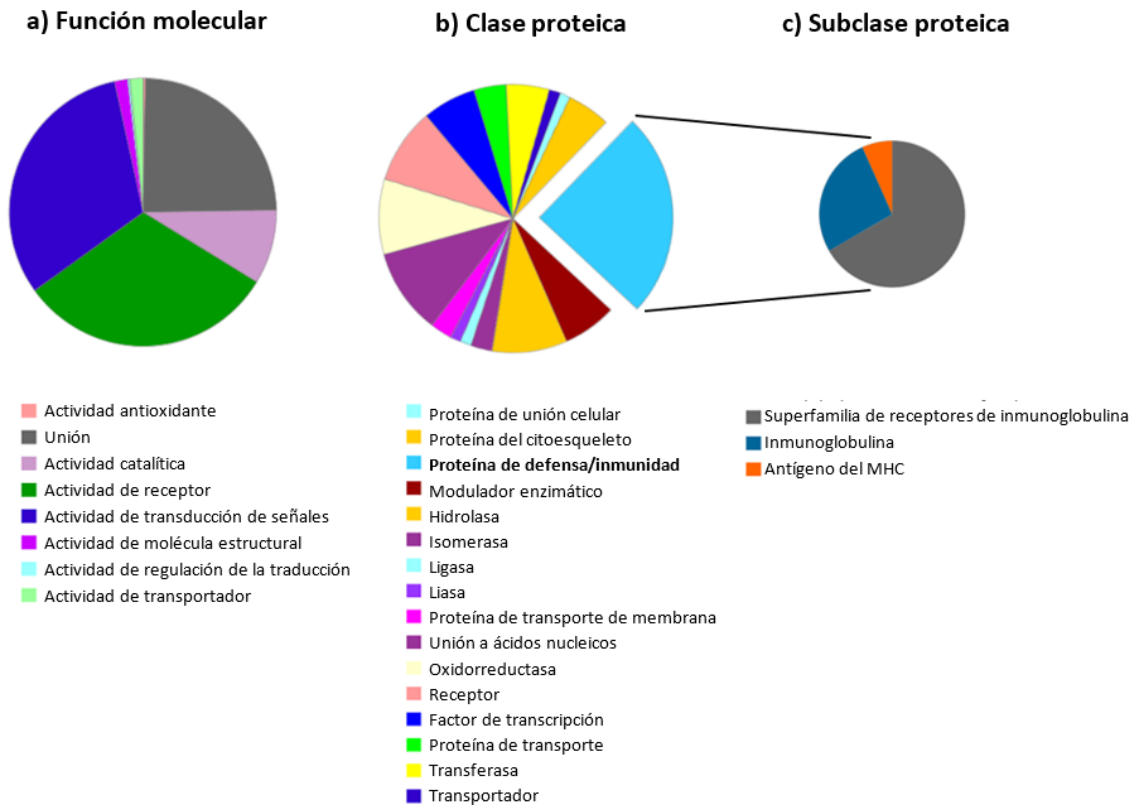


Figura 2.8. Clasificación por ontología génica de los genes de las CNVRs en caballos Criollo Argentino utilizando el software PANTHER. a) Función molecular, b) Distribución de las clases proteicas, c) Subclase proteica.

Con el fin de realizar un análisis más profundo, se evaluaron los ortólogos humanos y la lista de genes extraídos fue remitida al software en línea DAVID v6.7 (<https://david-d.ncicrf.gov/>). El análisis de ontología génica mostró que luego de aplicar la corrección de Bonferroni, 11 términos GO resultaron estadísticamente significativos, involucrando elementos relacionados con la percepción sensorial del olfato y estímulos químicos, cognición, procesos del sistema neurológico, vías de

señalización de proteínas del receptor acoplado a proteína G y actividad receptora olfativa, así como una vía KEGG (transducción olfativa), (Tabla 2.3).

Tabla 2.3

Análisis de enriquecimiento funcional de los genes presentes en las CNVRs.

Categoría	Término	Cantidad	%	p-valor	Bonferroni	FDR
Proceso celular	GO: 0007608 percepción sensorial del olfato	48	26,7	1,40E ⁻³⁷	1,10E ⁻³⁴	2,10E ⁻³⁴
Proceso celular	GO: 0007606 percepción sensorial de estímulos químicos	48	26,7	1,70E ⁻³⁵	1,40E ⁻³²	2,60E ⁻³²
Proceso celular	GO: 0050890 cognición	52	28,9	6,80E ⁻²⁷	5,60E ⁻²⁴	1,00E ⁻²³
Proceso celular	GO:0007600 percepción sensorial	49	27,2	3,30E ⁻²⁶	2,70E ⁻²³	5,00E ⁻²³
Proceso celular	GO:0050877 procesos del sistema nervioso	53	29,4	4,50E ⁻²²	3,80E ⁻¹⁹	7,00E ⁻¹⁹
Proceso celular	GO:0007186 vía de señalización de proteínas acopladas a proteína G	50	27,8	6,00E ⁻²¹	5,00E ⁻¹⁸	9,30E ⁻¹⁸
Proceso celular	GO:0007166 transducción de señal asociada a receptores de superficie celular	55	30,6	3,00E ⁻¹⁵	2,50E ⁻¹²	4,60E ⁻¹²
Componente celular	GO:0005886 membrana plasmática	72	40	1,70E ⁻¹⁰	3,00E ⁻⁰⁸	2,10E ⁻⁰⁷
Componente celular	GO:0016021 integral de membrana	76	42,2	2,00E ⁻⁰⁵	3,50E ⁻⁰³	2,50E ⁻⁰²
Componente celular	GO:0031224 intrínseco de membrana	77	42,8	4,00E ⁻⁰⁵	6,80E ⁻⁰³	4,90E ⁻⁰²
Función molecular	GO:0004984 actividad de receptor olfatorio	47	26,1	2,50E ⁻³⁴	6,30E ⁻³²	3,20E ⁻³¹
Vía KEGG	hsa04740 transducción olfatoria	45	25	1,70E ⁻²⁶	1,30E ⁻²⁴	1,80E ⁻²³

3.3 Análisis genético interracial

3.3.1 Diferenciación genética

La estimación del índice F_{ST} entre las 9 poblaciones tomadas de a pares mostró un rango de valores entre 0,036 y 0,193 (Figura 2.9). El menor valor de diferenciación se observó entre las razas Miniatura y Shetland ($F_{ST} = 0,036$), mientras que la mayor distancia se encontró entre el Shetland y el Sangre Pura de Carrera ($F_{ST} = 0,193$). Los caballos Criollo Argentino mostraron una mayor distancia genética respecto al Sangre Pura de Carrera ($F_{ST} = 0,128$), mientras que la menor distancia genética se observó con la raza Paso Peruano ($F_{ST} = 0,052$). El valor de F_{ST} a nivel global resultó de 0,104.

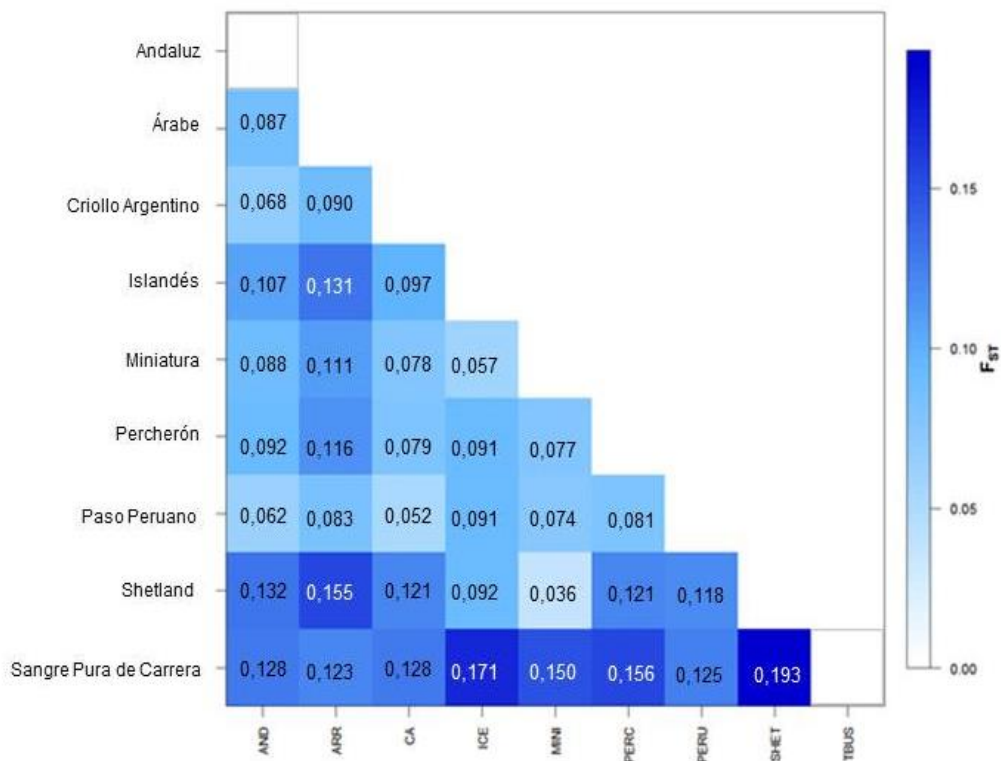


Figura 2.9. Valores de diferenciación genética (F_{ST}) calculados en las 9 poblaciones equinas tomadas de a pares, con el DS3 conteniendo 19.614 SNPs autosómicos. El gradiente de colores representa los valores ascendentes (oscuro) y descendentes (claro) del índice de fijación de Wright, indicados en los recuadros correspondientes.

3.3.2 Análisis de la varianza molecular (AMOVA)

El análisis de la varianza molecular (AMOVA) entre las 9 poblaciones de diferentes razas equinas (Tabla 2.4) reveló que un 10,43 % de la varianza corresponde a diferencias entre poblaciones (F_{ST}), un 1,09 % de la varianza se debe a diferencias dentro de cada población

(F_{IS}), y un 88,5 % de la variación se explica por la diferenciación alélica dentro de las poblaciones tomadas en conjunto (F_{IT}).

Tabla 2.4.

Análisis de varianza molecular (AMOVA) entre poblaciones equinas.

Fuente de variación	Grados de libertad	Suma de cuadrados	Componentes de varianza	Porcentaje de variación	Índices de fijación
Entre poblaciones	8	15093,76	355,75	10,43	$F_{ST} = 0,104^*$
Intrapoblacional	191	590878,35	37,31	1,09	$F_{IS} = 0,012$
Interalélica	200	603797,50	3018,99	88,50	$F_{IT} = 0,120^*$
Total	399	1345608,13	3412,04		

Nota: * $p < 0,001$.

3.3.3 Análisis de la estructura poblacional

La estructuración poblacional se analizó utilizando la hipótesis de 2 a 10 poblaciones (rango $K = 2 - K = 10$). Se tomó como valor óptimo $K = 7$ (Figura 2.10), valor en que las poblaciones pueden individualizarse claramente, y muestran componentes genéticos compartidos acordes a los registros filogenéticos y evolutivos.

Los caballos Criollo Argentino muestran una estructura definida, y de acuerdo a la asignación del software, con elevada cercanía al Paso Peruano, con componentes compartidos también con Percherón y Miniatura. En la raza de caballos Miniatura, la asignación racial de los individuos evidenciaría componentes genéticos compartidos con varias razas.

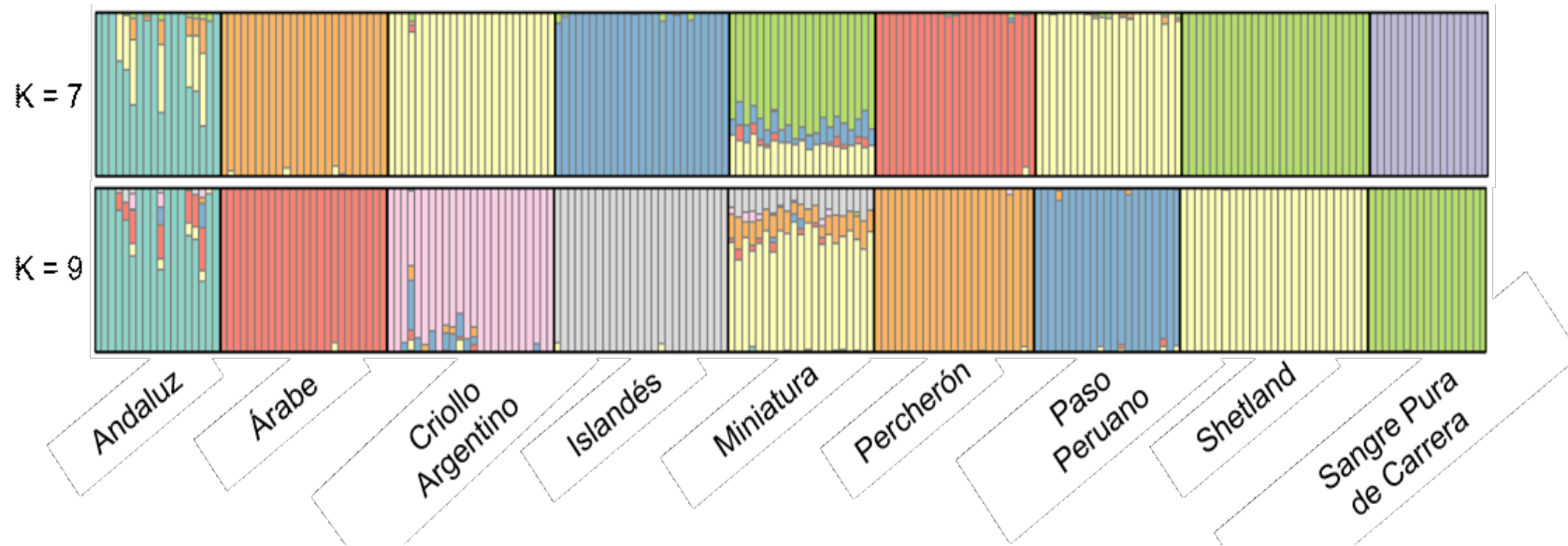


Figura 2.10. Resultado para 2 valores de K del agrupamiento Bayesiano realizado con el software fastStructure utilizando 19.614 SNPs, en los 200 caballos de las 9 poblaciones equinas analizadas. Cada animal está representado por una línea vertical única dividida en K colores, siendo K el número de grupos ancestrales asumidos y los colores muestran la proporción individual estimada de pertenencia al grupo. Cada población está separada de la siguiente por una línea negra.

La estructura poblacional fue evaluada, además, mediante un análisis de componentes principales a partir de los 38.199 SNPs (Figura 2.11). En la siguiente figura puede observarse que el primer y segundo componente (PC1 y PC2) representaron un 24,5 % y un 12 % de la varianza, respectivamente. Se observó que los caballos Criollo Argentino se agrupan en cercanías de los caballos de la raza Paso Peruano y el caballo Andaluz. Asimismo, se observó que el caballo Miniatura conformó un grupo junto con el caballo Islandés, separándose en tanto de la nube de puntos correspondiente a la raza Shetland y las demás razas incluidas.

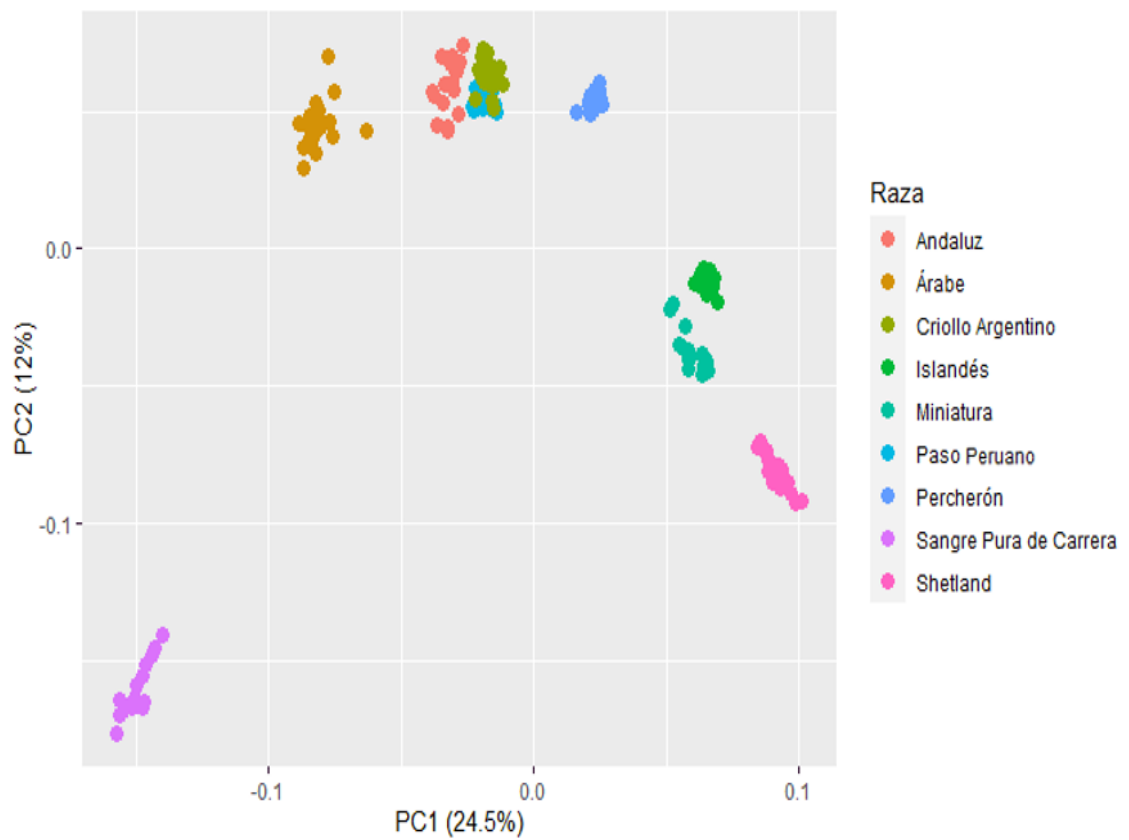


Figura 2.11. Análisis de componentes principales en base a 38.199 SNPs en las 9 poblaciones equinas. PC1: componente principal 1; PC2: componente principal 2. Entre paréntesis se indica el porcentaje de la varianza explicado por ese componente.

4. DISCUSIÓN

4. Discusión

4.1 Caracterización genómica de los caballos Criollo Argentino

El análisis genómico realizado en los caballos Criollo Argentino, implicó determinar el estado de variabilidad de 61.175 SNPs distribuidos por todo el genoma equino. Para este fin, los animales que se analizaron, se seleccionaron de manera tal que la muestra final conformada por 24 individuos, fuera representativa de la raza Criollo Argentino. El criterio general aplicado para la elección de los animales, como se mencionó previamente, consistió en tomar aquellos individuos menos emparentados, que permitieran evidenciar de forma no sesgada, la variabilidad genómica existente en la raza.

Como punto de partida, se analizó el pedigrí de cada animal, de un total de 150 caballos, y se realizó la reconstrucción de la genealogía completa, tomando como base los Registros Genealógicos (RRGG) de la Sociedad Rural Argentina (SRA, <http://www.sra.org.ar/rrgg/>). A partir de esta información, se buscó minimizar la presencia o aporte de sangres extranjeras, de manera de caracterizar los linajes argentinos más importantes desde la apertura inicial de los RRGG. Una vez seleccionados los individuos con el mayor porcentaje de sangre argentina, se excluyeron caballos que compartieran antecesores en segundo grado.

Finalmente, la muestra analizada quedó representada por individuos de pura raza Criollo Argentino, que no compartían antecesores de segundo grado. Los pedigríes recuperados de la base de datos de la SRA

conteniendo información de 968 caballos nacidos entre 1944 y 2012, permitieron definir la población de referencia, conformada por 280 caballos, en un período de 30 años (1971 - 2001). Para la población de referencia, se estimó un número efectivo de ancestros (f_a) de 50 individuos y un número efectivo de fundadores (f_e) de 157,7. El análisis de consanguinidad mostró que 357 de los 968 animales están emparentados, con un valor promedio de consanguinidad (F) igual a 0,041. Por otra parte, se observó un aumento del valor promedio de la consanguinidad de 1,14 (generación 2) a 14,94 (generación 8).

Los resultados obtenidos en el análisis de la endogamia individual genómica (F), respecto a la homocigosis observada, muestran un estado de endogamia bajo y con un leve exceso de heterocigotas, reflejado en el valor negativo obtenido. Si bien el valor obtenido para el coeficiente de endogamia a partir de los datos genealógicos no es exactamente igual al computado para cada individuo a partir de la tipificación genómica, la tendencia observada es la de un estado de endogamia bajo para la raza Criollo Argentino.

Considerando conjuntamente los resultados obtenidos para el coeficiente de endogamia a partir de los datos genealógicos y los valores computados para cada individuo a partir de la tipificación genómica, se puede decir que, si evaluamos la tendencia de los mismos, estos concuerdan entre sí y muestran un nivel bajo de endogamia para la raza. Las diferencias observadas podrían deberse a la naturaleza de los datos utilizados, o la profundidad del pedigrí (MacCluer *et al.*, 1983). No

obstante, el abordaje genómico es particularmente útil como estrategia complementaria al estudio genealógico especialmente en casos de genealogías complejas o en ausencia de registros de pedigrí (Li *et al.*, 2011).

Finalmente, la estimación de los coeficientes de parentesco, medida como el estado de identidad compartida (*IBS*) en contra de la proporción de identidad no compartida ($IBS=0$) corroboran (coeficientes de parentesco negativos) que entre los animales seleccionados como representativos, no existen relaciones de parentesco (*Figura 2.4.*). El conocimiento de las consecuencias de la consanguinidad y la endocría dan relevancia a la necesidad de preservar la variabilidad genética, como premisa esencial e importante, dado que es la base del progreso genético (De Rochambeau *et al.*, 2000) y especialmente para algunos tipos de razas locales con un censo reducido o que han experimentado cuellos de botella selectivos, ya sea natural o artificialmente. Los programas de cría deberían desarrollarse entonces, con el fin de seleccionar el stock de reproducción con el potencial genético más alto para rasgos de interés económico, manteniendo tanta variabilidad genética como sea posible por medio de procedimientos de selección óptimos.

4.2 Diversidad genómica de los caballos Criollo Argentino

En el análisis de la variabilidad genómica de los 24 caballos Criollo Argentino, la diversidad evaluada como H_e varió entre 0,315 (ECA 11) y

0,353 (ECA 22) (Tabla 2.1), por lo que la población presentó niveles altos de diversidad, como se desprende a partir de los valores obtenidos de heterocigosis y complementados por los bajos niveles de endogamia genómica observados (Figura 2.2). Los valores obtenidos superan lo reportado por Petersen *et al.*, (2013a), que utilizando un arreglo de ADN de mediana densidad para analizar 38 razas equinas, obtuvieron valores de He entre 0,236 en la raza Clydesdale y 0,337 en caballos Paint.

El desequilibrio de ligamiento (LD) se analizó en términos de su decaimiento, evaluado mediante el descenso en el valor del coeficiente de correlación (r^2) en función de la distancia genética entre pares de marcadores. Este análisis mostró que el Criollo Argentino presenta un nivel moderado de LD, con un marcado descenso de r^2 en los primeros 200 kb (Figura 2.5). Este valor es cercano a lo reportado para las razas más antiguas que presentan un declive más rápido en el entorno de los 50 - 100 kb (Wade *et al.*, 2009; Schurink *et al.*, 2019), pero sensiblemente menor a lo observado en razas de marcada selección como el Sangre Pura de Carrera, cuyo LD se extiende más allá de los 400 kb (Corbin *et al.*, 2010).

El valor obtenido podría deberse a que el LD dentro de las razas se conserva en distancias más largas en comparación con el análisis interracial (McCue *et al.*, 2012). Por otro lado, se ha reportado que a comienzos del S.XX las poblaciones de caballos criollos del Río de La Plata habrían sufrido un cuello de botella (Cortés *et al.*, 2017), lo cual podría explicar parcialmente la moderada extensión de LD. El primer

efecto de un evento como un cuello de botella es la reducción repentina en el tamaño de la población y en consecuencia una drástica reducción en la heterocigosis por locus (Nei *et al.*, 1975). Este tipo de evento demográfico ocasiona cambios en los patrones del desequilibrio de ligamiento a lo largo del genoma (Dapper y Payseur, 2017). Asimismo, el análisis de haplotipos mostró que los SNPs se concentraron en 2522 haplobloques cuyos tamaños variaron en un rango entre 1 y 199,175 kb en los 31 autosomas.

4.3 Variabilidad genómica estructural

El análisis de los 61.175 SNPs genotipados en la raza Criollo Argentino permitió la identificación de 165 CNV autosómicas, que fueron compiladas en 87 CNVR y comparadas con aquellas previamente informadas en estudios de CNV/CNVR en caballos (Huang *et al.*, 2009a; Doan *et al.*, 2012a, 2012b; Dupuis *et al.*, 2013; Metzger *et al.*, 2013; Ghosh *et al.*, 2014; Jun *et al.*, 2014; Wang *et al.*, 2014; Kader *et al.*, 2016). El porcentaje de cobertura del genoma (0,6 %) estuvo en concordancia con estudios que implementaron en la detección el arreglo Equine 70K SNP (Kader *et al.*, 2016) así como el análisis CGH (Ghosh *et al.*, 2014; Wang *et al.*, 2014). Además de la concordancia en la cobertura del genoma, variables como el recuento de CNVR, el tamaño del rango de CNVR y cromosoma más enriquecido (ECA12) presentaron valores concordantes con estudios previos que usaron similares plataformas de

genotipado (Dupuis *et al.*, 2013; Metzger *et al.*, 2013; Kader *et al.*, 2016). La distribución de ganancias en CNVR versus pérdidas en CNVR fue similar con respecto a otros informes de caballos (Dupuis *et al.*, 2013; Metzger *et al.*, 2013; Ghosh *et al.*, 2014; Wang *et al.*, 2014). Sin embargo, en nuestro estudio se observó un patrón invertido, es decir, los eventos de ganancias (55,2 %) prevalecieron sobre los de pérdidas (44,8 %). Considerando que el número de ganancias/pérdidas demostró ser bastante variable en estudios equinos, estas diferencias podrían ser parcialmente explicadas por el número variable de razas y la cantidad de muestras incluidas en cada estudio. Asimismo, se debe considerar la historia de la raza, ya que a menudo la CNV involucra largos segmentos de ADN, dando muestras de una posible huella de selección.

Finalmente, los diferentes algoritmos y filtros usados para la detección de las CNVs, así como la distribución y espaciado de las sondas en los arreglos de SNPs podrían hacer la diferencia con respecto a la cobertura genómica informada en los estudios que utilizan CGH y NGS.

En el análisis de enriquecimiento génico, los 337 genes detectados en las 87 CNVR están involucrados en un amplio espectro de funciones moleculares con significancia en las categorías: transducción de señales y actividad de receptor/unión y dentro de clase de proteína (Figura 2.8) las de tipo defensa/inmunidad (Tabla 2.3). El enriquecimiento de la transducción de señales estuvo primordialmente representado por receptores olfativos (RO), de los que se ha informado ampliamente en

estudios previos de CNV (Liu *et al.*, 2013; Paduel *et al.*, 2015; Keel *et al.*, 2016) y podría reflejar no solo la adaptación a estímulos ambientales sino también un progreso en términos evolutivos. Es interesante que el análisis del tipo de proteína mostró que las proteínas de defensa/inmunidad representó el 26,4 % de los genes, incluyendo la región V de la cadena alfa del receptor de las células T, el antígeno beta-2-microglobulina del complejo mayor de histocompatibilidad, los receptores tipo I y II de citoquina: la subunidad beta del receptor de interleuquina-10 y el receptor 2 de interferón alfa/beta.

Por otro lado, el análisis de ontología génica (GO) que usa ortólogos humanos en el software DAVID, presentó resultados similares a los obtenidos en caballos (Kader *et al.*, 2016), pero también en cerdos (Wang *et al.*, 2012), ganado (de Lemos *et al.*, 2018) y seres humanos (Redon *et al.*, 2006). Como establecieron Ghosh *et al.*, (2014), los estudios dirigidos a la CNV en el genoma equino han contribuido al conocimiento del rol de la CNV en la variación interindividual. Sin embargo, la aplicación biomédica eficiente de esta información requerirá la incorporación de datos procedentes de muchas más poblaciones e individuos y el uso de plataformas metodológicas comparables.

4.4 Análisis genómico interracial

Las poblaciones equinas comparadas, de acuerdo a los resultados obtenidos a partir del análisis de varianza molecular, se presentan

significativamente diferenciadas entre ellas ($F_{ST} = 0,104$; $p < 0,001$), pero a su vez, con homogeneidad en términos intrarraciales ($F_{IS} = 0,012$; $p = 0,200$). Además, el análisis mostró que la mayor parte de la variabilidad observada (88,5 %) se debe a diferencias dentro de los individuos ($F_{IT} = 0,120$; $p < 0,001$). Este contraste se ve reflejado en la marcada diferenciación que presentó el Criollo Argentino respecto al Sangre Pura de Carrera ($F_{ST} = 0,128$) y, por otro lado, la pequeña distancia respecto al Paso Peruano ($F_{ST} = 0,052$). Considerando que el parámetro F_{ST} es útil para tratar de comprender la historia de las poblaciones, estos valores podrían explicarse por el origen compartido entre la raza Paso Peruano y el Criollo Argentino, ambos descendientes de los caballos traídos por los españoles durante la colonización (Luis *et al.*, 2006).

El análisis de componentes principales (PCA) también respalda esta hipótesis, ya que en el mismo se observa la agrupación del Criollo Argentino no solo con el Paso Peruano y sino también con el Andaluz, con el que ya se ha reportado una relación cercana (Mirol *et al.*, 2002).

Finalmente, la evaluación de la estructura poblacional mediante el método de agrupamiento Bayesiano, mostró resultados alineados con lo observado en el PCA. Nuevamente se observó al Criollo Argentino y al Paso Peruano asignados a un mismo clúster hasta un valor de $K = 7$, lo que podría indicar la existencia de ascendencia compartida, y separándose en $K = 9$ (Figura 2.9). Asimismo, al forzar valores de K mayores a 7, se observó que tanto la raza Andaluza como el Criollo

Argentino comienzan a perder definición como entidades genéticas, motivo por el cual se tomó $K = 7$ como el valor más probable de K . Por último, destacar que el Criollo Argentino se presenta como una población genéticamente definida y distinta, lo cual es esperado en una raza con 400 años de historia y con registro genealógico cerrado.

En términos poblacionales, sobre todo de aquellas poblaciones sometidas a presiones selectivas, la consanguinidad dentro de la progenie de los individuos con potencial reproductivo puede ir en aumento, aún más que por puro efecto de la deriva genética, por lo que estos datos son importantes para realizar un monitoreo de los índices de variación genética que subyacen a la variación fenotípica. Considerando entonces tanto la selección como la conservación, los parámetros demográficos y genéticos también tienen un impacto en la evolución de la variabilidad genética y, al depender en gran medida del manejo de la población (Gutiérrez *et al.*, 2003; Goyache *et al.*, 2003), son particulares de cada grupo.

La estimación del tamaño efectivo de la población (N_e ; Falconer and Mackay, 1996) es un parámetro clave, para describir tanto la extensión de la diversidad genética en poblaciones animales, como para propósitos predictivos, por ejemplo, ayudando a prever cuándo un determinado apareamiento inadecuado puede conducir a una estructuración poblacional (Caballero y Toro, 2000). Por lo tanto, evaluar la evolución de la diversidad genética de las poblaciones a partir de datos de pedigrí (Boichard, 2002), otorga un valor agregado al análisis de datos

de los genomas de ciertos individuos representativos de la población, brinda información adecuada porque considera el entorno genético en el cual se está realizando el estudio y posiciona a la población con su propio acervo genético dentro del ambiente determinado en el que se desarrolla y evoluciona.

5. DISCUSIÓN GENERAL

5. Discusión General

La historia de la raza de caballos Criollo Argentino da cuenta de eventos de selección natural, donde los primeros caballos introducidos en América del Sur hace alrededor de 500 años, fueron los fundadores de un creciente número de animales que se reprodujeron en estado salvaje. Estas manadas originales, debieron responder a numerosos desafíos adaptativos (entornos cambiantes y diversos, situaciones climáticas extremas, amenazas depredadoras, disponibilidad de alimento, etc.) que probablemente modularon algunos aspectos fenotípicos como respuesta, y cuya consecuencia podría reflejarse en aspectos tales como la adaptación al clima, la eficiencia de la reproducción y la supervivencia de las crías, la resistencia a las enfermedades, y la apariencia conformacional, entre otros. En este contexto, esos mismos desafíos, pueden haber puesto a prueba al sistema inmunitario, y los mecanismos selectivos operantes, haber favorecido así a los ejemplares que portaban una mayor diversidad genética inmune, traducidos en la plasticidad adaptativa de la raza que se ve en la actualidad.

El componente genético de estas características, agrupados como variabilidad fenotípica (evidenciados, por ejemplo, en la diversidad de pelajes aceptados en el estándar de la raza), y la remarcable plasticidad de adaptación racial, ha sido un interrogante, cuyas respuestas parciales han sido respondidas mediante la descripción de los genes que determinan algunos colores de pelajes, la identificación de mutaciones que causan enfermedades de origen genético y conjuntos de genes

relacionados con caracteres complejos, como el rendimiento deportivo, tal y como se resume en la base de datos OMIA (<https://omia.org/home/>), en la que se observan 240 registros de caracteres relacionados con la herencia en los caballos.

Respecto a las características genéticas de esta raza tan emblemática, y como una pregunta recurrente que intenta responder qué hay en la genética de los caballos criollos que los hace tan particulares, se han realizado diversos estudios, que con el correr de los años, han incorporado abordajes metodológicos diversos, siempre tendientes a descubrir si los rasgos fenotípicos especiales de los caballos criollos pueden perpetuarse a través del manejo reproductivo y de la selección de los mejores ejemplares. Es así que se describió la variación intrapoblacional a nivel bioquímico (Peral Garcia, 1994; Peral Garcia *et al.*, 1996; Díaz, 1997; Díaz *et al.*, 2002), genético (Díaz *et al.*, 2002; Díaz, 2003; Corbi Botto *et al.*, 2016), e inmunogenético (Díaz *et al.*, 2003; Villegas-Castagnasso *et al.*, 2003; Díaz *et al.*, 2005; Díaz *et al.*, 2008) en los caballos Criollo Argentino, así como su singularidad, comparándolos con otras poblaciones de Criollos de América del Sur.

En este trabajo de Tesis Doctoral, la caracterización de la raza criolla se realizó mediante dos abordajes: la detección y el análisis de polimorfismos en genes de la respuesta inmune en el caballo Criollo Argentino, y posteriormente, a nivel genómico, se evaluó la diversidad del caballo Criollo Argentino tanto a nivel poblacional como interracial. En ambos casos, se realizó una comparación con otras razas equinas, tanto

poblaciones de Argentina como otras razas del mundo. El abordaje a nivel genómico, es el primero que se realiza en una raza caballar de Argentina.

El estudio de los SNPs en determinados genes relacionados con la RI puso en evidencia la presencia de variabilidad; tanto las citoquinas como los genes estructurales analizados, mostraron diversos alelos y diferencias poblacionales en la distribución de las variantes observadas.

La situación general en los genes del ELA analizados, es que las variaciones fueron más abundantes en porciones codificantes, específicamente en el segundo exón, en los sitios de interacción con antígenos, y se detectaron nuevas variantes alélicas en caballos criollos. Estos nuevos alelos enriquecen el acervo genético propio de la raza, y permiten hipotetizar acerca de su relación con características de resistencia de esta raza a condiciones sanitarias variables, principalmente a enfermedades infecciosas. Es de destacar, que la variabilidad de estos genes tiene relevancia documentada y validada en la plasticidad de las poblaciones animales para adaptarse a condiciones ambientales, incluyendo a la exposición a distintos tipos de patógenos.

Las variaciones observadas en el TNF- α se localizan en la región de regulación de la expresión. Esto podría vincularse con los efectos de un aumento o disminución marcado en los niveles de expresión, debido al rol central que juega esta citoquina tanto en la respuesta inmune como en los procesos inflamatorios, y refleja la complejidad de la comunicación célula-célula mediante la molécula de TNF- α (Kruglov *et al.*, 2008). En todos los

casos, sería importante determinar la relevancia funcional de la variabilidad observada a nivel de los SNPs de la región que controla la expresión de esta citoquina.

La información acerca de los factores genéticos que determinan caracteres fenotípicos en caballos no es abundante y en algunos casos no concluyente, sobre todo en relación a caracteres complejos como las enfermedades infecciosas. En muchos casos, la falta de información se remite al vínculo con el propio productor equino, quien es renuente a brindar datos que puedan aportar a realizar estudios genéticos completos y adecuados. No obstante, consideramos que el paso inicial importante es conocer el acervo genético de diversos grupos de caballos, ya sean razas o poblaciones mestizas. Este conocimiento preliminar puede ser la motivación inicial para mostrar al productor los beneficios potenciales de conocer el aporte genético de sus manadas, entre los que se cuentan: la posibilidad de predecir el valor productivo, seleccionar los mejores reproductores para la siguiente generación, reducir la frecuencia de aparición de enfermedades relacionadas con los genes y concomitantemente las pérdidas económicas asociadas y en definitiva, potenciar y dar un valor agregado a su producción.

De este modo, la diversidad genómica estructural posiciona al caballo Criollo como una extraordinaria fuente zoogenética no solo desde un enfoque genético de la población, sino también como una reserva de la diversidad genómica, conformacional y adaptativa.

La información obtenida en este trabajo de tesis es un aporte como punto de partida para diseñar planes de cría, manejo, mejora y conservación del ganado caballar. Por una parte, con el objetivo de salvaguardar la diversidad genética de la raza aquí descrita, es crucial minimizar los procesos erosivos, tales como el aumento de la endogamia o de variantes deletéreas. Por otra parte, para conocer la influencia de la genética de la raza en los caracteres deseados, específicamente, a los que se dirige la mejora y selección de los ejemplares.

6. CONCLUSIONES

6. Conclusiones

6.1 Capítulo 1

- Se diseñaron exitosamente estrategias biotecnológicas para la tipificación de SNPs de genes de la RI.
- Se determinaron las características genéticas de genes de la respuesta inmune (citoquinas, presentadores de antígeno, receptor de lentivirus) en caballos Criollo Argentino, con alta variabilidad comparado con otras razas equinas.
- Los genes del ELA mostraron variabilidad en la región codificante (exón 2), en los sitios de interacción con antígenos; es importante destacar que se detectaron nuevas variantes alélicas en caballos criollos, las cuales podrían ser una ventaja frente a la diversidad de patógenos.
- En el promotor de *TNF- α* , se detectó polimorfismo y un haplotipo poco frecuente, hasta ahora reportado únicamente en burros y cebras de Grant.
- Se identificaron 12 variantes alélicas en las regiones codificantes de *ELR1*. Las 7 variantes en el exón 3 no causan alteraciones en la proteína resultante; en contraste, una de las variantes del exón 5 podría tener un efecto deletéreo en la estructura predicha de la proteína.
- Los residuos críticos para la función de la proteína madura de ELR1 están conservados en equinos y otras especies. Esto podría

explicarse por la existencia de presiones selectivas a favor de preservar la función específica de los miembros de TNFRSF y la respuesta inmune del huésped.

6.2 Capítulo 2

- De acuerdo a su estructura genealógica y genómica, la población de caballos Criollo Argentino presenta un estado de diversidad aceptable con bajos índices de consanguinidad.
- El número de ancestros mayoritarios fue menor al número efectivo de fundadores, lo que podría indicar que la población sufrió un cuello de botella o uso intensivo de ciertos reproductores en su historia evolutiva.
- El genotipado genómico y los índices de diversidad evidencian una población de Criollo Argentino con elevada diversidad dentro de los individuos, diferenciados genéticamente de otras razas equinas.
- La estructuración poblacional fue concordante con los 400 años de historia de la raza y las características del registro genealógico del caballo Criollo Argentino.
- El desequilibrio de ligamiento evaluado en el Criollo Argentino entre pares de marcadores disminuyó conforme aumentaba la distancia entre ellos, estabilizándose en el entorno de los 200 kb, evidenciando posibles eventos selectivos.
- Se evidenció variabilidad estructural en la raza Criollo Argentino, y las 87 CNVRs identificadas están involucradas en: funciones moleculares

tales como transducción de señales (receptores olfativos) y en proteínas de defensa/inmunidad (26,4 % de los genes), lo que podría reflejar la adaptación a condiciones ambientales.

6.3 Conclusión Final

Los resultados obtenidos a partir del análisis genético, genómico y genealógico son un aporte a la caracterización de una de las razas emblemáticas del país: el caballo Criollo Argentino. La información obtenida en este trabajo de tesis es un aporte como punto de partida para diseñar planes de cría, de manejo, mejora genética y conservación del ganado caballar. Con el objetivo de salvaguardar la diversidad genética de la raza descrita, y, si bien los valores de endogamia aquí observados no son alarmantes, sería conveniente monitorear periódicamente a las poblaciones y así minimizar los procesos erosivos, tales como el aumento de la frecuencia de variantes deletéreas. Asimismo, el conocimiento de las características de la diversidad de la raza será útil para conocer la influencia de la genética de la raza en los caracteres deseados, específicamente, para plantear criterios y objetivos de selección para utilizar los ejemplares más adecuados para obtener la mejora esperada.

7. BIBLIOGRAFÍA

Bibliografia

Ahmadian A, Gharizadeh B, Gustafsson AC, Sterky F, Nyren P, Uhlen M, and Lundeberg J. Single-nucleotide polymorphism analysis by Pyrosequencing. *Anal Biochem* 2000a; 280: 103-10.

Al Abri MA, Holl HM, Kalla SE, Sutter NB, Brooks SA. Whole genome detection of sequence and structural polymorphism in six diverse horses. *PloS one* 2020; 15(4): e0230899.

Albright-Fraser DG, Reid R, Gerber V, Bailey E. Polymorphism of DRA among equids. *Immunogenetics* 1996; 43(5): 315-7.

Andersson L, Bohme J, Peterson P, Rask L. Genomic hybridization of bovine class II major histocompatibility genes: 2. Polymorphism of DR genes and linkage disequilibrium in the DQ- DR region. *Anim Genet* 1986; 17(2): 295-304.

Andersson LS, Swinbune JE, Meadows JR, Broström H, Eriksson S, Fikse WF, Lindgren G, et al. The same ELA class II risk factors confer equine insect bite hypersensitivity in two distinct populations. *Immunogenetics* 2012; 64(3): 201-8.

Arfuso F, Giannetto C, Fazio F, Panzera F, Piccione G. Training program intensity induces an acute phase response in clinically healthy horses. *J Equine Vet Sci* 2020; 102986.

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidmann JG, Struhl, K. DNAzol method. In: *Curr Protoc Mol Biol* 1990; 2(Suppl A), 1-5.

Ayala FJ. Population and Evolutionary Genetics. The Benjamin/Cummings Publishing, Menlo Park, California, USA, 1982.

Bailey E, Graves KT, Cothran EG, Reid R, Lear TL, Ennis RB. Synteny-mapping horse microsatellite markers using a heterohybridoma panel. *Anim Genet* 1995; 26: 177-80.

Barreiro LB, Quintana-Murci L. From evolutionary genetics to human immunology: how selection shapes host defense genes. *Nat. Rev. Genet* 2010; 11: 17–30.

Barrett JC, Fry B, Maller JDMJ, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; 21(2): 263-5.

Bibollet-Ruche F, Russell RM, Liu W, Stewart-Jones GB, Sherrill-Mix S, Li Y, Decker J M, et al. (2019). CD4 receptor diversity in chimpanzees protects against SIV infection. *Proc Natl Acad Sci USA* 2019; 116(8): 3229-38.

Bodmer JL, Schneider P, Tschopp J. The molecular architecture of the TNF superfamily. *Trends Biochem Sci* 2002; 27: 19–26.

Boichard D, Maignel L, Verrier E. The value of using probabilities of gene origin to measure genetic variability in a population. *Genet Sel Evol* 1997; 29: 5-23.

Boichard D. PEDIG: a fortran package for pedigree analysis suited for large populations. In: Proceedings of the 7th world congress on genetics applied to livestock production, Montpellier 19-23 August, 2002; 32: 525-8.

Brosnahan MM, Brooks SA, Antczak DF. Equine clinical genomics: A clinician's primer. *Equine Vet J.* 2010; 42(7): 658- 70.

Brown JJ, Thomson W, Clegg P, Eyre S, Kennedy LJ, Matthews J, Carter S, Ollier WER. Polymorphisms of the equine major histocompatibility complex class II DRA locus. *Tissue Antigens* 2004; 64(2): 173-9.

Brown JJ, Ollier WER, Thomson W, Matthews JB, Carter SD, Binns M, Pinchbeck G, Clegg PD. TNF- α SNP haplotype frequencies in equidae. *Tissue Antigens* 2006; 67(5): 377-82.

Caballero A, Toro MA. Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genet Res* 2000; 75(3): 331-43.

Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* 2015; 31(16): 2745-7.

Chomczynski P, Mackey K, Drews R, Wilfinger W. (1997). DNazol (R): A reagent for the rapid isolation of genomic DNA. *Biotechniques* 1997; 22(3): 550-3.

Chowdhary BP, Raudsepp T, Kata SR, Goh G, Millon LV, Allan V, Skow LC, et al. The first-generation whole-genome radiation hybrid map in the horse identifies conserved segments in human and mouse genomes. *Genome Res* 2003; 13: 74251.

Cook RF, Leroux C, Issel CJ. Equine infectious anemia and equine infectious anemia virus in 2013: A review. *Vet Microbiol* 2013; 167(1-2): 181–204.

Corbi Botto C, Sadaba SA, Zappa ME, Peral García P, Díaz S. Análisis de la diversidad genética de una población de caballos Criollo Argentino mediante polimorfismos de nucleótido simple de los genes IL12B y TNF- α . *Analecta Veterinaria* 2016; 36(2): 11-5.

Corbin LJ, Blott SC, Swinburne JE, Vaudin M, Bishop SC, Woolliams JA. (2010). Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Anim Genet* 2010; 41, 8-15.

Cortés O, Dunner S, Gama LT, Martínez AM, Delgado JV, Ginja C, Vega- Pla JL, et al. The legacy of Columbus in American horse populations assessed by microsatellite markers. *J Anim Breed Genet* 2017; 134(4): 340-50.

Covaleda L, Fuller FJ, Payne SL. EIAV S2 enhances pro-inflammatory cytokine and chemokine response in infected macrophages. *Virology* 2010; 397(1): 217-23.

Crawford DC, Bhangale T, Li N, Hellenthal G, Rieder MJ, Nickerson DA, Stephens M. Evidence for substantial fine-scale variation in recombination rates across the human genome. *Nat Genet* 2004; 36(7): 700-6.

Dapper AL, Payseur BA. Effects of Demographic History on the Detection of Recombination Hotspots from Linkage Disequilibrium. *Mol Biol Evol* 2017; 35(2): 335–53.

de Lemos MVA, Berton MP, de Camargo GMF, Peripolli E, de Oliveira Silva RM, Baldi F, et al. Copy number variation regions in Nellore cattle: evidences of environment adaptation. *Livest Sci* 2018; 207: 51-8.

De Rochambeau H, Fournet-Hanocq F, Khang JVT. Measuring and managing genetic variability in small populations. In *Annales de zootechnie* 2000; 49(2): 77-93.

Díaz S. Estudio comparativo de la variabilidad genética en poblaciones de equinos de diferentes razas mediante sus polimorfismos genéticos sanguíneos [tesis de grado]. Argentina, Universidad Nacional de Misiones, Facultad de Ciencias Exactas, Químicas y Naturales, 1997.

Díaz S, Giovambattista G, Dulout FN, Peral-Garcia P. Genetic variation of the second exon of ELA-DRB genes in Argentine Creole horses. *Anim Genet* 2001; 32(5):257-63.

Díaz S, Dulout FN, Peral-García P. Greater genetic variability in Argentine Creole than in Thoroughbred horses based on serum protein polymorphisms. *Genet Mol Res* 2002; 1(3): 261-5.

Díaz S. Estudio del polimorfismo y poligenismo de los loci DRB en caballos criollo argentino [tesis doctoral]. Argentina, Universidad Nacional de La Plata, Facultad de Ciencias Naturales y Museo, 2003.

Díaz S, Ripoli MV, Peral García P, Giovambattista G. Marcadores genéticos para resistencia y susceptibilidad a enfermedades infecciosas en animales domésticos. *Analecta Vet* 2005; 25(1): 40-52.

Díaz S, Echeverría MG, It V, Posik DM, Rogberg- Muñoz A, Pena NL, Peral-García P, Vega-Pla JL, Giovambattista G. Development of an ELA-DRA gene typing method based on pyrosequencing technology. *Tissue Antigens* 2008; 72(5): 464-8.

Doan R, Cohen N, Harrington J, Veazy K, Juras R, Cothran G, McCue ME, Skow L, Dindot SV. Identification of copy number variants in horses. *Genome Res* 2012a; 22, 899–907.

Doan R, Cohen N, Sawyer J, Ghaffari N, Johnson C, Dindot S. Whole genome sequencing and genetic variant analysis of a Quarter Horse mare. *BMC Genomics* 2012b; 13: 78.

Dowdall CR. Criando Criollos. Buenos Aires: Hemisferio Sur; 1982.

Dupuis MC, Zhang Z, Durkin K, Charlier C, Lekeux P, Georges M. Detection of copy number variants in the horse genome and examination of their association with recurrent laryngeal neuropathy. *Animal genetics* 2013; 44(2): 206-08.

Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 1992; 131 (2): 479–91.

Excoffier L, Lischer HEL. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 2010; 10: 564-7.

Falconer DS, Mackay TFC. *Introduction to Quantitative Genetics*, 4th ed. Longman Scientific & Technical, Burnt Mill, Harlow, England, 1996.

Francino O, Amills M, Sanchez A. Canine *Mhc* DRB1 genotyping by PCR–RFLP analysis. *Anim Genet* 1997; 28(1): 41-5.

Francisco E. Caracterización de los Haplotipos del Complejo Principal de Histocompatibilidad (ELA) en diferentes Razas Equinas. [Tesis Doctoral] Argentina, Universidad Nacional de La Plata, Facultad de Ciencias Veterinarias, 2014.

Fraser DG, Bailey E. Demonstration of three DRB loci in a domestic horse family. *Immunogenetics* 1996; 44(6):441-5.

Gallagher SR. Quantitation of DNA and RNA with absorption and fluorescence spectroscopy. In: Current Protocols in Molecular Biology. Edited by Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K. New York: Greene and Wiley-Interscience; 1994. p. A.3D.1–A.3D.8.

Ghosh S, Qu Z, Das PJ, Fang E, Juras R, Cothran EG, Raudsepp, T, et al. Copy number variation in the horse genome. PLoS Genet 2014; 10(10).

Go YY, Bailey E, Cook F, Timoney P, Coleman S, MacLeod J, Balasuriya UBR, et al. Genome-wide association study (GWAS) among four horse breeds identifies a common haplotype in equine chromosome 11 (ECA11) associated with the in vitro CD3+ T cell susceptibility/resistance to equine arteritis virus infection. J Equine Vet Sci 2012; 10(32): S74-S75.

Goyache F, Gutiérrez JP, Fernández I, Gómez E, Álvarez I, Díez J, Royo LJ. Using pedigree information to monitor genetic variability of endangered populations: the Xalda sheep breed of Asturias as an example. Journal of Animal Breeding and Genetics 2003; 120(2): 95-105.

Grilz-Seger G, Druml T, Neuditschko M, Mesarič M, Cotman M, Brem G. Analysis of ROH patterns in the Noriker horse breed reveals signatures of selection for coat color and body size. Anim Genet 2019; 50(4): 334-46.

Guerin G, Bailey E, Bernoco D, Anderson I, Antczak DF, Bell K, Bowling AT, et al. Report of the international equine gene mapping workshop: male linkage map. *Anim Genet* 1999; 30(5): 341-54.

Gutiérrez JP, Altarriba J, Díaz C, Quintanilla R, Cañón J, Piedrafita J. Pedigree analysis of eight Spanish beef cattle breeds. *Genet Sel Evol* 2003; 35(1): 1-21.

Han H, McGivney BA, Farries G, Katz LM, MacHugh DE, Randhawa IA, Hill EW. Selection in Australian Thoroughbred horses acts on a locus associated with early two-year old speed. *PloS one* 2020; 15(2): e0227212.

Hartl DL. A Primer of Population Genetics, 2nd ed., Sinauer Associates, Sunderland, Massachusetts, USA, 1988.

Hedrick PW, Parker KM, Miller EL, Miller PS. Major histocompatibility complex variation in the endangered Przewalski's horse. *Genetics* 1999; 152(4):1701-10.

Hehlgans T, Pfeffer K. The intriguing biology of the tumour necrosis factor/ tumour necrosis factor receptor superfamily: players, rules and the games. *Immunology* 2005; 115: 1–20.

Holmes CM, Violette N, Miller D, Wagner B, Svansson V, Antczak DF. MHC haplotype diversity in Icelandic horses determined by polymorphic microsatellites. *Genes Immun* 2019; 20(8): 660-70.

Horin P, Smola J, Matiasovic J, Vyskocil M, Lukeszova L, Tomanova K, Kralik P, Glasnak V, Schröfflelova D, Knoll A, Sedlinska M, Krenkova L, Jahn P. Polymorphisms in equine immune response genes and their associations with infections. *Mamm Genome* 2004; 15(10): 843-50.

Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009a; 37(1): 1-13.

Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc* 2009b; 4(1): 44-57.

James JW. Computation of genetic contributions from pedigrees. *Theor Appl Genet* 1972; 42: 272-3.

Janečka JE, Davis BW, Ghosh S, Paria N, Das PJ, Orlando L, Raudsepp T, et al. Horse Y chromosome assembly displays unique evolutionary features and putative stallion fertility genes. *Nature communications* 2018; 9(1): 1-15.

Janova E, Matiasovic J, Vahala J, Vodicka R, Van Dyk E, Horin P. Polymorphism and selection in the major histocompatibility complex DRA and DQA genes in the family Equidae. *Immunogenetics* 2009; 61(7): 513-27.

Janova E, Futas J, Klumplerova M, Putnova L, Vrtkova I, Vyskocil M, Horin P, et al. Genetic diversity and conservation in a small endangered horse population. *J Appl Genet* 2013; 54(3): 285-92.

Johnson AD. An extended IUPAC nomenclature code for polymorphic nucleic acids. *Bioinformatics (Oxford, England)*. 2010 May; 26(10):1386-9.

Jun J, Cho YS, Hu H, Kim HM, Jho S, Gadhvi P, Han K, et al. Whole genome sequence and analysis of the Marwari horse breed and its genetic origin. *BMC Genom* 2014; 15: S4.

Kader A, Liu X, Dong K, Song S, Pan J, Yang M, Ma Y, et al. Identification of copy number variations in three Chinese horse breeds using 70K single nucleotide polymorphism BeadChip array. *Anim Genet* 2016; 47(5): 560-9.

Kalbfleisch TS, Rice ES, DePriest MS, Walenz BP, Hestand MS, Vermeesch JR, MacLeod JN, et al. Improved reference genome for the domestic horse increases assembly contiguity and composition. *Commun Biol* 2018; 1(1): 1-8.

Kalemkerian PB. Determinación de la susceptibilidad/resistencia genética a la infección por el Virus de Arteritis Equina (AVE) mediante el estudio de genes del complejo principal de histocompatibilidad (MHC) equino. [Tesis Doctoral] Argentina, Universidad Nacional de La Plata, Facultad de Ciencias Veterinarias, 2014.

Kamath PL, Getz WM. Adaptive molecular evolution of the Major Histocompatibility Complex genes, DRA and DQA, in the genus Equus. BMC Evol Biol 2011; 11(1): 128.

Kasprzyk A. BIOMART: driving a paradigm change in biological data management. Database 2011; vol 2011.

Keel BN, Lindholm-Perry AK, Snelling WM. Evolutionary and functional features of copy number variation in the cattle genome. Front Genet 2016; 7: 207.

Kienast ME. (2004). Estudio de polimorfismos genéticos en poblaciones equinas de Argentina. [Tesis Doctoral] Argentina, Universidad Nacional de La Plata, Facultad de Ciencias Veterinarias, 2004.

Kim JH, Hu HJ, Yim SH, Bae JS, Kim SY, Chung YJ. CNVRuler: a copy number variation-based case-control association analysis tool. Bioinformatics 2012; 28(13): 1790-2.

Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 1980; 16(2): 111-20.

Kruglov AA, Kuchmiy A, Grivennikov SI, Tumanov AV, Kuprash DV, Nedospasov SA. Physiological functions of tumor necrosis factor and the consequences of its pathologic overexpression or blockade: mouse models. Cytokine Growth Factor Rev 2008; 19(3-4): 231-44.

Kruglyak L, Nickerson DA. Variation is the spice of life. *Nature genetics* 2001; 27(3): 234-6.

Kumar S, Patra MC, De BC, Brahma B, Mahanty S, Mishra P, Chopra M, Sharma RC, Manuja BK, Singh RK, De S. MHC-DRB polymorphism in Indian Equines. 2014.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Analysis version 7.0 for bigger datasets. *Mol Biol Evol* 2016; 33(7): 1870-4.

Lacy RC. Analysis of founder representation in pedigrees: founder equivalents and founder genome equivalents. *Zoo Biol* 1989; 8: 111-23.

LaFramboise T. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Res* 2009; 37(13):4181-93.

Lazarus R, Klimecki WT, Palmer LJ, Kwiatkowski DJ, Silverman EK, Brown A, Weiss ST, et al. Single-nucleotide polymorphisms in the interleukin-10 gene: differences in frequencies, linkage disequilibrium patterns, and haplotypes in three United States ethnic groups. *Genomics* 2002; 80(2): 223-8.

Leroy G, Mary-Huard T, Verrier E, Danvy S, Charvolin E, Danchin-Burge C, et al. Methods to estimate effective population size using pedigree data: Examples in dog, sheep, cattle and horse. *Genetics Selection Evolution* 2013; 45(1): 1.

Li MH, Strandén I, Tiirikka T, Sevón-Aimonen ML, Kantanen, J. A comparison of approaches to estimate the inbreeding coefficient and pairwise relatedness using genomic and pedigree data in a sheep population. *PLoS One* 2011; 6(11), e26256.

Li N, Stephens M. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* 2003; 165(4): 2213-33.

Li Y, Oosting M, Smeekens SP, Jaeger M, Aguirre-Gamboa R, Le KT, Netea MG, et al. A functional genomics approach to understand variation in cytokine production in humans. *Cell* 2016; 167(4): 1099-110.

Liburt NR, Adams AA, Betancourt A, Horohov DW, McKeever KH. Exercise- induced increases in inflammatory cytokines in muscle and blood of horses. *Equine Vet J* 2010; 42: 280-8.

Lindgren G, Breen M, Godard S, Bowling A, Murray J, Scavone M, Ellegren H, et al. Mapping of 13 horse genes by fluorescence in-situ hybridization (FISH) and somatic cell hybrid analysis. *Chromosome Res* 2001; 9: 53-9.

Lis JT. (1980). [42] Fractionation of DNA fragments by polyethylene glycol induced precipitation. In: *Methods in enzymology*. Academic Press; 1980. Vol. 65, p. 347-53.

Liu J, Zhang L, Xu L, Ren H, Lu J, Zhang X, Du L, et al. Analysis of copy number variation in the sheep genome using 50K SNP 338 BeadChip array. *BMC Genom* 2013; 14: 229.

Luis C, Bastos-Silveira C, Cothran EG, Oom MDM. Iberian origins of New World horse breeds. *Journal of Heredity* 2006; 97(2): 107-13.

MacCluer JW, Boyce AJ, Dyke B, Weitkamp LR, Pfennig DW, Parsons CJ. Inbreeding and pedigree structure in Standardbred horses. *J Hered* 1983; 74: 394-99.

Malécot G. *Mathématiques de l'hérédité*. Paris, Ed.Masson et Cie, 1948.

Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics* 2010; 26(22):2867-73.

Matiašovic J, Lukeszova L, Hořín P. Two bi-allelic single nucleotide polymorphisms within the promoter region of the horse tumour necrosis factor α gene. *Eur J Immunogenet* 2002; 29(4): 285-6.

McCue ME, Bannasch DL, Petersen JL, Gurr J, Bailey E, Binns MM, Leeb T, et al. A high density SNP array for the domestic horse and extant Perissodactyla: utility for association mapping, genetic diversity, and phylogeny studies. *PLoS Genet* 2012; 8(1): e1002451.

Mealey RH, Leib SR, Pownder SL, McGuire TC. Adaptive immunity is the primary force driving selection of equine infectious anemia virus envelope SU variants during acute infection. *J Virol* 2004; 78(17): 9295-305.

Merino AM, Zhang K, Kaslow RA, Aissani B. Structure of tumor necrosis factor-alpha haploblocks in European populations. *Immunogenetics* 2013; 65(7):543-52.

Metzger J, Philipp U, Lopes MS, da Camara Machado A, Felicetti M, Silvestrelli M, Distl O. Analysis of copy number variants by three detection algorithms and their association with body size in horses. *BMC Genom* 2013; 14(1): 487.

Metzger J, Karwath M, Tonda R, Beltran S, Águeda L, Gut M, Gut IG, Distl O. Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. *BMC Genom* 2015; 16(1): 764.

Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, Thomas PD. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic acids research* 2016; 45(D1), D183-D189.

Milenkovic D, Oustry-Vaiman A, Lear TL, Billault A, Mariat D, Piumi F, Guérin G, et al. Cytogenetic localization of 136 genes in the horse:

comparative mapping with the human genome. *Mamm Genome* 2002; 13(9): 524-34.

Miller D, Tallmadge RL, Binns M, Zhu B, Mohamoud YA, Ahmed A, Antczak DF, et al. Polymorphism at expressed DQ and DR loci in five common equine MHC haplotypes. *Immunogenetics* 2017; 69(3): 145-56.

Mirol PM, García PP, Vega-Pla JL, Dulout FN. Phylogenetic relationships of Argentinean Creole horses and other South American and Spanish breeds inferred from mitochondrial DNA sequences. *Anim Genet* 2002; 33(5): 356-63.

Moore BD, Balasuriya UB, Watson JL, Bosio CM, MacKay RJ, MacLachlan NJ. Virulent and avirulent strains of equine arteritis virus induce different quantities of TNF- α and other proinflammatory cytokines in alveolar and blood-derived equine macrophages. *Virology* 2003; 314(2): 662-70.

Moorjani P, Patterson N, Loh PR, Lipson M, Kisfali P, Melegh BI, Reich D, et al. Reconstructing Roma history from genome-wide data. *PLoS One* 2013; 8(3).

Mostafavi E, Esmailnejad B, Abtahi-Foroushani SM. Evaluation of cytokines and sialic acids contents in horses naturally infected with *Theileria equi*. *Comp Immunol Microbiol Infect Dis* 2020; 101453.

Mullis KB, Faloona FA. Specific Synthesis of DNA in vitro via a Polymerase-Catalyzed Chain Reaction. *Methods Enzymol* 1987; 155: 335-50.

Nei M, Maruyama T, Chakraborty R. The bottleneck effect and genetic variability in populations. *Evolution* 1975; 1-10.

Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 1978; 89(3): 583-90.

Orr N, Back W, Gu J, Leegwater P, Govindarajan P, Conroy J, Brama PAJ, et al. Genome-wide SNP association-based localization of a dwarfism gene in Friesian dwarf horses. *Animal genetics* 2010; 41: 2-7.

Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet* 2006; 2(12).

Patin E, Hasan M, Bergstedt J., et al. Natural variation in the parameters of innate immune cells is preferentially driven by genetic factors. *Nat Immunol* **19**, 302–314 (2018). <https://doi.org/10.1038/s41590-018-0049-7>

Penedo MCT, Millon LV, Bernoco D, Bailey E, Binns M, Cholewinski G, Tozaki T, et al. International Equine Gene Mapping Workshop Report: a comprehensive linkage map constructed with data from new markers and by merging four mapping resources. *Cytogene Genome Res* 2005; 111(1): 5-15.

Peral García P. Caracterización racial de equinos de pura raza criolla [Tesis Doctoral]. Argentina, Universidad Nacional de La Plata, Facultad de Ciencias Naturales y Museo, 1994.

Peral García P, Kienast M, Villegas Castagnasso EE, Díaz S, Dulout F. Genetic relationship between horse breeds through multivariate analysis. *Agro Sur* 1996.

Petersen JL, Mickelson JR, Cothran EG, Andersson LS, Axelsson J, Bailey E, McCue ME, et al. Genetic Diversity in the Modern Horse Illustrated from Genome-Wide SNP. *PLoS One* 2013a; 8(1): e54997.

Petersen JL, Mickelson JR, Rendahl AK, Valberg SJ, Andersson LS, Axelsson J, Brama P, et al. (2013b). Genome-wide analysis reveals selection for important traits in domestic horse breeds. *PLoS Genet* 2013b; 9(1): e1003211.

Primo AT. America: Conquista e colonizacao: A fantastica historia dos conquistadores ibericos e seus animais na era dos descobrimentos. Porto Alegre: Ed. Movimento, 2004.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; 81(3): 559-75.

Orlando L, Metcalf JL, Alberdi MT, Telles-Antunes M, Bonjean D, Otte M, Prado JL, et al. Revising the recent evolutionary history of equids using ancient DNA. *Proc Natl Acad Sci USA* 2009; 106(51): 21754-9.

Qian L, Han X, Liu X. Structural insight into equine lentivirus receptor 1. *Protein Sci* 2015; 24: 633–42.

Raj A, Stephens M, Pritchard JK. fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. *Genetics* 2014; 197: 573-89.

RedonR, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Hurles ME, et al. Global variation in copy number in the human genome. *Nature* 2006; 444: 444-54.

Reynolds J, Weir BS, Cockerham CC. Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* 1983; 105: 767–79.

Rockwell H, Mack M, Famula T, Sandmeyer L, Bauer B, Dwyer A, Bellone RR. Genetic investigation of equine recurrent uveitis in Appaloosa horses. *Anim Genet* 2020; 51(1): 111-6.

Rodero E, Rodero A, Delgado JV. El ganado andaluz primitivo y sus implicaciones en el descubrimiento de América. *Archivos de Zootecnia* 1992; 41: 383–400.

Ronaghi M. Pyrosequencing sheds light on DNA sequencing. *Genome Res* 2001; 11(1): 3-11.

Rousset F. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 1997; 145(4): 1219-28.

Rousset F. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Mol Ecol Resour* 2008; 8: 103-6.

RStudio Team RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. 2020; <http://www.rstudio.com/>.

Sadaba S. Estudio de la composición genética de caballos de biotipo deportivo de la raza Árabe en Argentina mediante la caracterización de haplotipos de clase II del MHC y haplotipos de ADN mitocondrial [tesis doctoral]. Argentina, Universidad Nacional de La Plata, Facultad de Ciencias Veterinarias, 2018.

Schaefer RJ, Schubert M, Bailey E, Bannasch DL, Barrey E, Bar-Gal GK, McCue ME, et al. Developing a 670k genotyping array to tag ~2M SNPs across 24 horse breeds. *BMC Genom* 2017; 18: 565.

Schneider S, Roessli D, Excoffier L. Arlequin: a software for population genetics data analysis. *User manual ver 2000*; 2: 2496-7.

Schurink A, Shrestha M, Eriksson S, Bosse M, Bovenhuis H, Back W, Ducro BJ, et al. The genomic makeup of nine horse populations sampled in the Netherlands. *Genes* 2019; 10(6): 480.

Smith CA, Farrah T, Goodwin RG. The TNF receptor superfamily of cellular and viral proteins. Activation, costimulation, and death. *Cell* 1994; 76: 959–62.

Smith DM, Lunney JK, Ho CS, Martens GW, Ando A, Lee JH, Chardon, P. Nomenclature for factors of the swine leukocyte antigen class II system, 2005. *Tissue Antigens* 2005; 66(6): 623-39.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 2013; 30: 2725-29.

Tornquist SJ, Oaks JL, Crawford TB. Elevation of cytokines associated with the thrombocytopenia of equine infectious anaemia. *J Gen Virol* 1997; 78(10): 2541-8.

Tryon RC, White SD, Bannasch DL. Homozygosity mapping approach identifies a missense mutation in equine cyclophilin B (PPIB) associated with HERDA in the American Quarter Horse. *Genomics* 2007; 90(1): 93-102.

Valberg SJ, Ward TL, Rush B, Kinde H, Hilaragi H, Nahey D, Mickelson JR. Glycogen branching enzyme deficiency in quarter horse foals. *J Vet Intern Med* 2001; 15(6): 572-80.

Vandenbroeck K, Goris A. Cytokine gene polymorphisms in multifactorial diseases: gateways to novel targets for immunotherapy? *Trends Pharmacol Sci* 2003; 24(6): 284-9.

Villegas-Castagnasso EE, Diaz S, Giovambattista G, Dulout FN, Peral-García P. Analysis of ELA-DQB exon 2 polymorphism in Argentine Creole horses by PCR-RFLP and PCR-SSCP. *J Vet Med A* 2003; 50(6): 280-5.

Villegas Castagnasso EE. Asociación entre marcadores genéticos y la producción de gonadotrofina coriónica equina (eCG). [Tesis Doctoral] Argentina, Universidad Nacional de La Plata, Facultad de Ciencias Naturales y Museo, 2005.

Viljuma A, Mikko S, Hahn D, Skow L, Andersson G, Bergström TF. Genomic structure of the horse major histocompatibility complex class II region resolved using PacBio long-read sequencing technology. *Sci Rep* 2017; 7(1): 1-12.

Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, Imsland F, Lindblad-Toh K, et al. Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* 2009; 326(5954): 865-7.

Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, Bucan M, et al. (2007). PennCNV: an integrated hidden Markov model designed for high-

resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* 2007;17(11): 1665-74.

Wang J, Jiang J, Fu W, Jiang L, Ding X, Liu JF y Zhang Q. A genome-wide detection of copy number variations using SNP genotyping arrays in swine. *BMC genomics* 2012; 13(1): 273.

Wang W, Wang S, Hou C, Xing Y, Cao J, Wu K, Zhou H, et al. Genome-wide detection of copy number variations among diverse horse breeds by array CGH. *PLoS One* 2014; 9(1): e86860.

Wright S. The interpretation of population structure by F-statistics with special regard to system of mating. *Evolution* 1965; 19: 395-420.

Wright S. The theory of gene frequencies. *Evolution and the Genetics of Populations* 1969; 2: 290-344.

Yilmaz V, Yentür SP, Saruhan-Direskeneli G. IL-12 and IL-10 polymorphisms and their effects on cytokine production. *Cytokine* 2005; 30(4), 188-94.

Yuhki N, O'Brien SJ. Nature and origin of polymorphism in feline MHC class II DRA and DRB genes. *The Journal of Immunology* 1997; 158(6): 2822-33.

Zhang B, Jin S, Jin J, Li F, Montelaro RC. A tumor necrosis factor receptor family protein serves as a cellular receptor for the macrophage-tropic equine lentivirus. *Proc Natl Acad Sci USA* 2005; 102: 9918-23.

Zhang B, Sun C, Jin S, Cascio M, Montelaro RC. Mapping of equine lentivirus receptor 1 residues critical for equine infectious anemia virus envelope binding. *J Virol* 2008; 82: 1204–13.

Zhao Z, Fu YX, Hewett-Emmett D, Boerwinkle E. Investigating single nucleotide polymorphism (SNP) density in the human genome and its implications for molecular evolution. *Gene* 2003; 312: 207-13.

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 2012; 28: 3326-8.

8. ANEXOS

ANEXO 1

PURIFICACIÓN DE LAS REACCIONES DE PCR POR PRECIPITACIÓN CON PEG

Este protocolo está ajustado para reacciones de PCR de 50 µL. Si el volumen de la PCR es diferente, modificar la escala manteniendo las proporciones.

Correr 5 µL de la PCR en un gel de agarosa para verificar la amplificación.

Al resto de la reacción de PCR agregarle 50 µL de una solución 20 % PEG – 2,5 M NaCl y mezclar bien.

Incubar a 37 °C por 15 minutos.

Centrifugar a alta velocidad (~15000 g) durante 15 minutos a temperatura ambiente.

Retirar el sobrenadante con pipeta y descartar. El pellet es incoloro y queda adherido a la pared del tubo.

Agregar 125 µL de etanol 70 %. Si se agrega el etanol en el fondo del tubo, centrifugar 2 minutos a 15000 g. Si se agrega el etanol suavemente por la pared del tubo, dejar reposar un minuto. Eliminar el sobrenadante y descartar. Eliminar la mayor cantidad posible del etanol.

Secar el pellet a 37 °C durante 10-15 minutos. Asegurarse de que no queden restos de etanol y que el pellet esté seco. Un exceso en el secado del pellet dificulta la resuspensión.

Disolver el producto de PCR en 20 µL de agua bidestilada estéril. Resuspender completamente el pellet (con pipeta o vórtex). Incubar por unos minutos a temperatura ambiente o a 37 °C.

Correr de 2 a 4 µL en un gel de agarosa durante 10 minutos para tener una idea aproximada del porcentaje de recuperación. Simultáneamente sembrar 20 y 100 ng de estándar de ADN.

SOLUCIÓN 20 % PEG – 2,5 M NaCl

Para 50 mL mezclar:

10,0 g polietilenglicol 8000 (también funciona con PEG 6000)

7,3 g de NaCl

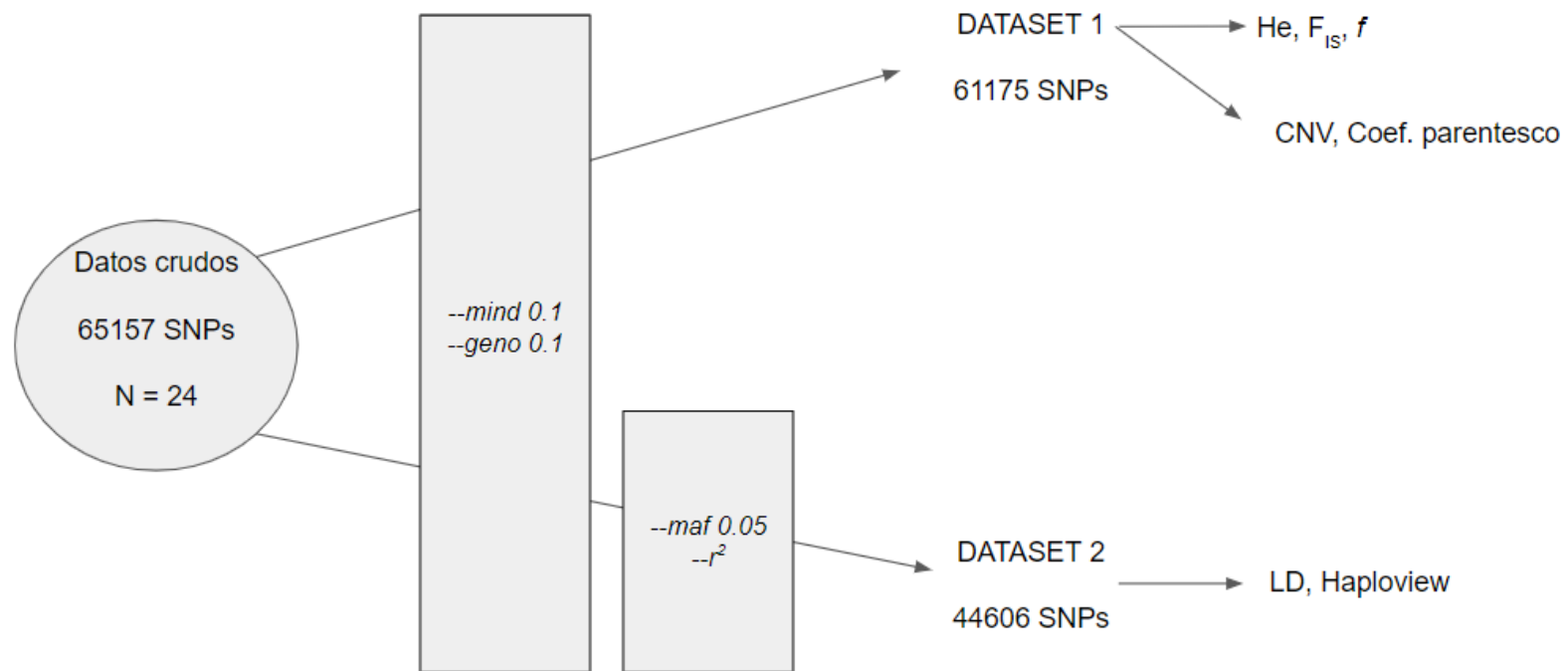
Agregar 40 mL de agua bidestilada. Agitar y disolver el PEG durante al menos 20 minutos. Se puede colocar la solución en un agitador a 37 °C.

Una vez disuelto el PEG (la solución se torna incolora), completar el volumen hasta 50 mL con agua bidestilada. Conservar a temperatura ambiente.

Modificado de Travis Glenn, Savannah River Ecology Lab, Department of Biological Sciences, Drawer E University of South Carolina, Columbia, SC.

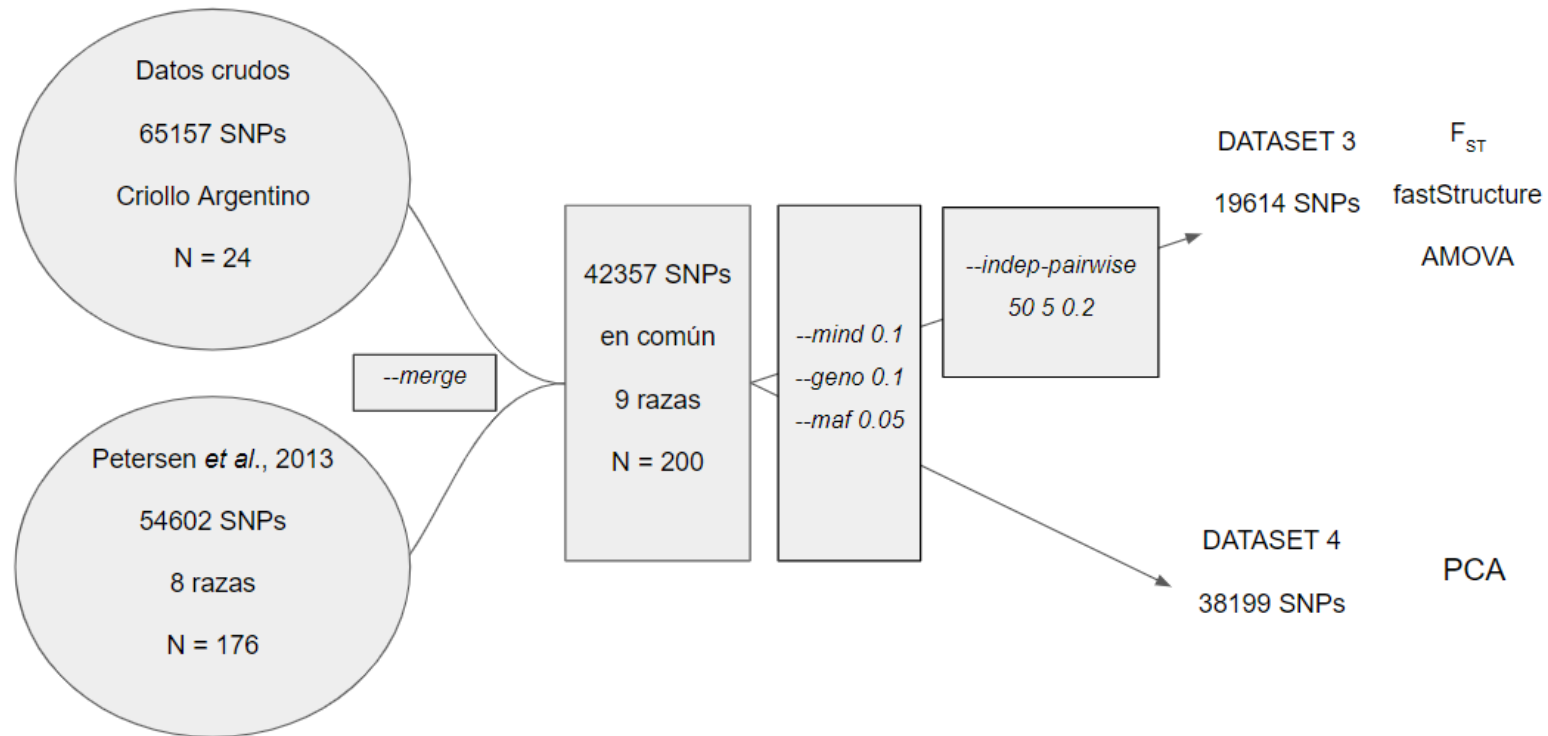
ANEXO 2

MAPA DE FILTROS PARA DEFINICIÓN DE LOS DATASETS 1 Y 2



ANE XO 3

MAPA DE FILTROS PARA DEFINICIÓN DE LOS DATASETS 3 Y 4



ANEXO 4

RESULTADOS COMPLEMENTARIOS

Tabla complementaria 1

Lista de las CNVs detectados en los autosomas de los caballos Criollo Argentino.

ECA	Inicio	Fin	Cantidad de SNPs	Tamaño*	Tipo
1	159296153	159325599	4	29447	pérdida
1	159674209	159749634	5	75426	pérdida
1	178798269	178815370	3	17102	ganancia
1	45738709	46181474	8	442766	pérdida
1	20505965	20797410	9	291446	ganancia
1	159674209	159897607	6	223399	pérdida
1	53271588	53343543	3	71956	pérdida
1	144642448	144655950	3	13503	ganancia
1	156599954	156784088	9	184135	pérdida
1	159674209	159897607	6	223399	pérdida
1	46026522	46181474	4	154953	pérdida
1	58999235	59066996	5	67762	pérdida
1	160036661	160102520	3	6586	pérdida
1	45738709	46083964	6	345256	pérdida
1	159296153	159749634	14	453482	pérdida
1	45969677	46181474	5	211798	pérdida
1	156135693	156305412	3	16972	pérdida
1	178798269	178815370	3	17102	ganancia
1	45969677	46181474	5	211798	pérdida
1	156449080	156784088	14	335009	pérdida
1	156599954	156784088	9	184135	pérdida
1	144277678	144512512	9	234835	ganancia
1	92052701	92265429	10	212729	pérdida
1	162835689	163018369	4	182681	ganancia
2	96664019	96702010	7	37992	ganancia
2	96664019	96702010	7	37992	ganancia
2	78756002	79049288	9	293287	ganancia
2	119912812	120125190	8	212379	ganancia
2	120048890	120125190	4	76301	ganancia
2	96701526	96754300	7	52775	ganancia

Anexos 206

3	65951617	66065643	4	114027	ganancia
3	106517004	106660146	5	143143	ganancia
3	18499566	18507627	3	8062	ganancia
3	65889054	65951800	3	62747	ganancia
3	74495490	74738129	4	24264	pérdida
3	6848717	6952969	6	104253	pérdida
3	8694423	8726437	3	32015	pérdida
3	76326251	76342093	4	15843	ganancia
3	65889054	65951800	3	62747	ganancia
3	76326251	76625704	8	299454	ganancia
3	35200976	35321182	9	120207	pérdida
3	35279221	35321182	8	41962	pérdida
4	48702295	48966208	7	263914	pérdida
4	42420512	42742198	7	321687	pérdida
6	39295651	39405007	3	109357	ganancia
6	73072905	73261049	8	188145	pérdida
6	2672935	2712701	5	39767	ganancia
6	2672935	2712701	5	39767	ganancia
6	66323494	66399351	10	75858	pérdida
6	73072905	73320179	10	247275	pérdida
6	27494541	27494650	3	110	pérdida
6	73072905	73320179	10	247275	pérdida
6	27430472	27491334	6	60863	ganancia
6	73072905	73261049	8	188145	pérdida
6	27494541	27494650	3	110	pérdida
6	27430472	27491334	6	60863	ganancia
7	95894473	96109416	5	214944	ganancia
7	73083306	73333458	6	250153	pérdida
7	73083306	73333458	6	250153	pérdida
7	73083306	73333458	6	250153	pérdida
7	1114451	1213480	3	9903	pérdida
7	185387	435683	9	250297	pérdida
7	73054093	73333458	7	279366	pérdida
7	34754706	34806577	4	51872	pérdida
8	4993812	5030505	3	36694	ganancia
8	4993812	5030505	3	36694	ganancia
8	4993812	5030505	3	36694	ganancia
9	31574454	31574969	5	516	pérdida
10	29963602	30019290	3	55689	ganancia
10	70451170	70653841	6	202672	ganancia
10	14618688	14773732	5	155045	ganancia
10	53261091	53276524	4	15434	ganancia
10	74942096	75011754	4	69659	pérdida
11	38181838	38200567	3	1873	ganancia

11	38181838	38200567	3	1873	ganancia
11	38181838	38200567	3	1873	ganancia
11	38181838	38200567	3	1873	ganancia
11	60340549	60802894	11	462346	ganancia
11	21120681	21155797	3	35117	pérdida
12	6059876	6126253	7	66378	pérdida
12	6059876	6126253	7	66378	pérdida
12	12520513	12594142	3	7363	ganancia
12	13014168	13422256	8	408089	ganancia
12	13014168	14407281	15	1393114	ganancia
12	12333201	13170959	12	837759	ganancia
12	13014168	13098923	3	84756	ganancia
12	12520513	12594142	3	7363	ganancia
12	13398392	13573356	5	174965	ganancia
12	14128309	14407281	4	278973	ganancia
12	6059876	6126253	7	66378	pérdida
12	6059876	6126253	7	66378	pérdida
12	12520513	13488187	13	967675	ganancia
13	7359845	7413137	4	53293	ganancia
13	28421959	28590293	5	168335	ganancia
13	28421959	28590293	5	168335	ganancia
13	6340547	6529524	5	188978	ganancia
13	28421959	28590293	5	168335	ganancia
13	28421959	28590293	5	168335	ganancia
13	7359845	7378465	3	18621	ganancia
13	28421959	28590293	5	168335	ganancia
13	28421959	28590293	5	168335	ganancia
14	59460837	59463001	3	2165	ganancia
14	59413745	59463001	4	49257	ganancia
14	1296596	1300439	5	3844	ganancia
14	59413745	59463001	4	49257	ganancia
14	61445385	61450197	4	4813	ganancia
14	59413745	59463001	4	49257	ganancia
14	61445385	61450197	4	4813	ganancia
14	1238596	1392437	11	153842	ganancia
15	37668975	37754882	3	85908	ganancia
15	81544884	81668605	4	123722	ganancia
15	37668975	37754882	3	85908	ganancia
15	2267229	2316484	3	49256	pérdida
15	90197083	90216165	3	19083	pérdida
15	37668975	37754882	3	85908	ganancia
15	37668975	37754882	3	85908	ganancia
15	37668975	37754882	3	85908	ganancia
15	2267229	2316484	3	49256	pérdida

Anexos 208

16	85926223	86077067	5	150845	ganancia
16	79424659	79511509	5	86851	ganancia
16	4488665	4506785	3	18121	ganancia
16	51722594	51786601	6	64008	ganancia
18	3286734	3432787	3	146054	ganancia
18	19901232	19902018	3	787	ganancia
18	49404497	49408946	3	445	ganancia
18	3156252	3286883	3	130632	ganancia
18	23152931	23167832	4	14902	ganancia
18	3286734	3432787	3	146054	ganancia
18	37545702	37589241	5	4354	ganancia
18	49404497	49408946	3	445	ganancia
18	23152931	23167832	4	14902	ganancia
18	37545702	37589241	5	4354	ganancia
18	49404497	49408946	3	445	ganancia
18	3286734	3432787	3	146054	ganancia
18	56779448	56889525	10	110078	pérdida
18	49404497	49408946	3	445	ganancia
18	51465789	51738258	8	27247	pérdida
19	12123487	12233905	3	110419	ganancia
19	1346278	1396027	3	4975	ganancia
19	12123487	12233905	3	110419	ganancia
20	27984017	28089836	6	10582	pérdida
20	34355448	34463425	6	107978	pérdida
20	1296566	1344089	3	47524	pérdida
20	1296566	1344089	3	47524	pérdida
20	1296566	1344089	3	47524	pérdida
20	32491252	32566747	3	75496	pérdida
21	48436447	48438741	4	2295	ganancia
21	48436447	48438741	4	2295	ganancia
21	47689624	47823967	5	134344	ganancia
21	15191571	15395571	8	204001	ganancia
21	15191571	15395571	8	204001	ganancia
22	1321505	1355408	3	33904	ganancia
24	36948203	36951441	3	3239	pérdida
24	36948203	36951441	3	3239	pérdida
25	37929574	38258772	12	329199	pérdida
25	26512888	26891173	10	378286	pérdida
25	37904760	38258772	13	354013	pérdida
26	29873480	29944740	3	71261	ganancia
28	24527605	24537607	3	10003	ganancia
28	39159039	39341709	5	182671	pérdida
28	18242886	18335085	7	922	pérdida
31	19536798	19644343	4	107546	pérdida

31	4829031	4884348	4	55318	pérdida
31	4829031	4884348	4	55318	pérdida

*No se muestran las CNVs en el cromosoma 27 debido a que no cumplían con el criterio de calidad.

Tabla complementaria 2

Combinaciones haplotípicas detectadas en el exón 5 de *ELR1*.

Variante	Haplotipo								VAIE-	VAIE+	χ^2	p-valor
	rs1095048803	rs1095048804	rs1095048805	rs1095048806	rs68535250	rs1095048807	rs68535250	rs1095048808				
ELR1e5_001	G	C	G	G	T	G	G	T	nd	nd	nd	nd
<i>ELR1e5_001.1</i>	G	C	G	G	C	G	G	T	0,071	0,000	2,382	0,1227
<i>ELR1e5_002.1</i>	G	C	G	A	C	G	A	T	0,190	0,094	1,343	0,2466
<i>ELR1e5_002.2</i>	G	C	G	G	C	G	A	T	0,595	0,781	2,868	0,0904
<i>ELR1e5_003.1</i>	G	C	G	A	C	G	A	G	0,048	0,094	0,613	0,4335
ELR1e5_004	G	C	G	G	C	A	A	T	0,000	0,031	1,330	0,2487
ELR1e5_008	G	C	A	A	C	G	A	G	0,024	0,000	0,772	0,3795
ELR1e5_009	A	C	G	G	C	G	G	T	0,024	0,000	0,772	0,3795
ELR1e5_010	A	T	G	G	C	G	G	T	0,048	0,000	1,566	0,2108
χ^2	2,382	1,566	0,772	0,568	nd	1,330	4,975	0,121				
p - valor	0,123	0,211	0,379	0,451	nd	0,249	0,026	0,727				

Se indican los test de asociación por marcador y los p - valores del test de asociación de haplotipos. *p-valor: 0,05.

Tabla complementaria 3

F_{IS} de cada población calculado con el software ARLEQUIN v3.5 y el DS3 conteniendo 19.614 SNPs.

Población	F_{IS}	p-valor
Andaluz	0,062	0,078
Árabe	0,035	0,202
Criollo Argentino	-0,00063	0,533
Islandés	0,008	0,387
Miniatura	0,004	0,430
Percherón	-0,0055	0,558
Paso Peruano	0,00094	0,440
Shetland	0,024	0,227
Pura Sangre de Carrera	-0,0188	0,709