



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Dirección General de Estudios de Posgrado

Facultad de Ciencias Biológicas

Unidad de Posgrado

Caracterización de marcadores moleculares de genes involucrados en la estructura y desarrollo de la fibra de alpaca y su potencial asociación con el diámetro de la fibra

TESIS

Para optar el Grado Académico de Doctor en Ciencias Biológicas

AUTOR

William Herminio SALAS CONTRERAS

ASESOR

Gustavo Augusto GUTIÉRREZ REYNOSO

Lima, Perú

2019

DEDICATORIA

A mí querida mamá Marina, mi tía María, mis hijos: Yesenia y Keny y; mis hermanos: Zoila, Gladys y Carlos

AGRADECIMIENTO

Al Doctor José Espinoza Babilón, docente principal de la Universidad Peruana Cayetano Heredia, por permitirme realizar mi tesis en el laboratorio que él dirige.

Al Ph. D. Gustavo Augusto Gutiérrez Reynoso, docente de la Universidad Nacional Agraria de la Molina, por su gentil apoyo en el asesoramiento de mi tesis.

Al Dr. Jorge Enrique Rodríguez Bailón, docente de la Universidad Peruana Cayetano Heredia, por su invaluable apoyo en el análisis bioinformático de mi tesis

Al M. Sc. Diego Veliz Otani, docente de la Universidad Peruana Cayetano Heredia, por su apoyo en el análisis estadístico de mi tesis.

A la Universidad Nacional de Huancavelica, por permitirme utilizar las alpacas del Centro de Investigación Lachocc.

A Javier Carhuapoma Boza, técnico de campo del Centro de Investigación Lachocc de la UNH por su apoyo desinteresado en la toma de muestras de sangre.

LISTA DE ABREVIATURAS

A	Diversidad alélica
ADN	Acido desoxirribonucleico
ADNmt	ADN mitocondrial
ARN	Ácido ribonucleico
ARNm	ARN mensajero
ARNr	ARN ribosomal
CONACS	Consejo Nacional de Camélidos Sudamericanos
D_A	Distancia genética de Nei
D_s	Distancia genética estándar de Nei
dATP	2-deoxiadenosina 5-trifosfato
dCTP	2-deoxicitidina 5-trifosfato
dGTP	2-deoxiguanosina 5-trifosfato
dTTP	2-deoxitimidina 5-trifosfato
dNTP	2-deoxinucleotido 5-trifosfato
EDTA	Ácido etilen diamino tetra acético
EST	marcador de secuencia expresada (expressed sequence tag)
<i>EDAR</i>	Gen regulador del desarrollo de la fibra (<i>Ectodisplasia A Regulator</i>)
FAO	Organización de las Naciones Unidas para la Agricultura y la Alimentación
F_{JK}	Coeficiente de parentesco o coancestría
F, F_i, F_{is}	Índice de fijación o coeficiente de consanguinidad
F_{IT}	Coeficiente global de consanguinidad
F_{ST}	Coeficiente de diferenciación
G_{ST}	Análogo del indicador F_{ST} descrito por Nei (1973)
GWAS	Estudio de asociación del genoma completo (<i>Genome-Wide Association Study</i>)
h	Diversidad genética
H	Heterocigosidad
H_E	Heterocigosidad esperada
H_o	Heterocigosidad observada
H_s	Heterocigosidad esperada dentro de una población

H _T	Heterocigosidad esperada total
H ₀	Hipótesis nula
HWE	Equilibrio de Hardy-Weinberg
<i>HOXC13</i>	Gen regulador del desarrollo de la fibra (<i>Homeobox 13</i>)
K	Número de poblaciones
KRT	Queratina
KAP	Proteína asociada a queratina
<i>KRT33</i>	Gen estructural de la queratina 33
<i>KRT34</i>	Gen estructural de la queratina 34
<i>KRT31</i>	Gen estructural de la queratina 31
<i>KAP1</i>	Proteína asociada a queratina 1
k	Número de alelos de un locus en una población
K ₃ E	Potasio-EDTA
LD	Desequilibrio de ligamiento
Ln	Logaritmo natural
MAF	Frecuencia del alelo menor (<i>Minor allele frequency</i>)
MAS	Selección asistida por marcadores (<i>Marker assisted selection</i>)
MC	Algoritmo de la cadena de Markov
MCMC	Método de Cadena de Markov-Monte Carlo
mM	Milimolar
MINAGRI	Ministerio de Agricultura y Riego
n	Número de alelos en un locus
N _m	Número efectivo de migrantes
ng	Nanogramo
nm	Nanómetro
N _m	Número de migrantes
ns	No significativo estadísticamente
NCBI	National Center for Biotechnology Information
OFDA	Analizador óptico del diámetro de fibra (<i>Optical fibre diameter analyser</i>).
PCR	Reacción en cadena de la polimerasa
PE	Probabilidad de exclusión
pH	potencial de hidrogeniones

pi	Frecuencia del alelo i
PIC	Contenido de información polimórfica
pb	Pares de bases
QTL	Loci para caracteres cuantitativos
Rho _{IS}	Análogo del indicador F _{IS} descrito por Slatkin (1995)
SNP	Polimorfismos de nucleótidos simples
SSM	Error en pareado de bases (Slipped strand mispairing)
SSR	Simple secuencias repetidas
STR	Repeticiones cortas en serie
Tris-HCl	Tris-Ácido clorhídrico
TE	Tris-EDTA
TBE	Tris borato-EDTA
T _m	Temperatura de alineamiento
TSS	Sitio de inicio de transcripción (<i>Transcription start site</i>)
Θ	Valor theta, análogo al F _{ST} descrito por Weir y Cockerham (1984)
μL	Microlitro
μM	Micromolar
μg	Microgramo
UNH	Universidad Nacional de Huancavelica
UBM	Unidad de Biotecnología Molecular
UPCH	Universidad Peruana Cayetano Heredia

ÍNDICE

CARÁTULA	i
DEDICATORIA	ii
AGRADECIMIENTO	iii
LISTA DE ABREVIATURAS	iv
ÍNDICE	vii
ÍNDICE DE TABLAS	xi
ÍNDICE DE FIGURAS	xii
RESUMEN	xiii
ABSTRACT	xiv
I. INTRODUCCIÓN	1
II. ANTECEDENTES	3
2.1. La alpaca	3
2.1.1. Distribución	3
2.1.2. Origen evolutivo de la alpaca	4
2.2. Caracterización molecular	5
2.3. Marcadores moleculares para estudios de diversidad genética	5
2.3.1. Microsatélite	6
a. Distribución, composición y frecuencia	6
b. Tasa de mutación de los microsatélites	7
2.3.2. Polimorfismo de nucleótidos únicos (SNP)	8
2.4. Diversidad genética	9
2.4.1. Estimadores de la diversidad genética	10
a. Frecuencia alélica y diversidad alélica	10
b. Heterocigosidad	10
c. Equilibrio de Hardy-Weinberg	11
d. Coeficiente de consanguinidad (F o F_{IS})	12
2.5. Estructura genética	12
2.5.1. Factores que influyen en la diversidad y estructura genética	13
a. Deriva genética	13
b. Flujo genético y migración	13

c. Efecto Wahlund	15
2.5.2. Métodos de estimación de la estructura genética en poblaciones	15
a. Estadístico F	15
b. Análisis de estructura	17
✓ Análisis Bayesiano	17
✓ Cadena de Markov-Monte Carlo	17
2.6. Estudios previos sobre diversidad genética en alpacas	18
III. HIPÓTESIS Y OBJETIVOS	21
3.1. Hipótesis	21
3.2. Objetivo	21
IV. MATERIALES Y MÉTODOS	22
4.1. Localización del estudio	22
4.2. De los animales	22
4.3. Colección de muestras de sangre	23
4.4. Extracción de ADN genómico	23
4.5. Cuantificación del ADN genómico	23
4.6. Amplificación de ADN microsatélites	23
4.7. Caracterización de la diversidad genética de las alpacas	
Huacaya del centro de investigación Lachocc-UNH	24
4.7.1. Análisis de variabilidad genética	24
4.7.2. Análisis de estructura	
25 Identificación de marcadores genéticos tipo SNP en genes que regulan el desarrollo de la fibra o codifican proteínas estructurales de la fibra de alpaca	25
4.8.1. Selección de genes candidatos	25
4.8.2. Identificación de polimorfismos en genes codificantes mediante análisis de tres genomas (<i>in silico</i>)	26
a. Predicción de los efectos de las variantes genéticas	27
4.8.3. Diseño de cebadores para amplificar los genes candidatos	28

4.8.4.	Estandarización de las reacciones de amplificación de los 3 genes candidatos	28
4.8.5.	Secuenciamiento mediante método Sanger	29
4.8.6.	Caracterización de marcadores SNP de los 3 genes candidatos en alpacas de Lachocc-UNH	29
	a. Cálculo del número muestral	29
	b. Indicadores de variabilidad genética de los genes <i>EDAR</i> , <i>HOXC13</i> y <i>KRT31</i>	30
4.9.	Determinación de la asociación genética entre marcadores tipo SNP y diámetro de fibra en alpacas	30
4.9.1.	Análisis de la potencia estadística	30
4.9.2.	Análisis de asociación entre genes y diámetro de fibra	31
	a. Descripción del modelo estadístico	31
V.	RESULTADOS	32
5.1.	Caracterización de la diversidad genética de una población de 96 alpacas Huacaya del Centro de Investigación Lachocc-UNH empleando 10 marcadores microsatélites	32
5.1.1.	Análisis de variabilidad genética	32
5.1.2.	Análisis de estructura genética	33
5.2.	Identificación de marcadores genéticos tipo SNP en 3 genes candidatos, dos que regulan el desarrollo de la fibra (<i>HOXC13</i> y <i>EDAR</i>) y uno que codifica una proteína estructural de la fibra de alpaca (<i>KRT31</i>)	35
5.2.1.	Selección de genes candidatos	35
5.2.2.	Elección de cebadores para los 3 genes candidatos	35
5.2.3.	Secuenciamiento e identificación de variantes en los genes <i>EDAR</i> , <i>HOXC13</i> y <i>KRT31</i>	36
5.2.4.	Análisis de variabilidad genética en los genes <i>EDAR</i> , <i>HOXC13</i> y <i>KRT31</i>	37
5.3.	Evaluación de la presencia de asociación genética entre los polimorfismos genéticos identificados en los genes candidatos y el carácter diámetro de la fibra de alpaca.	40

5.3.1.	Análisis de la potencia estadística	40
5.3.2.	Análisis de asociación entre los genes <i>EDAR</i> , <i>HOXC13</i> y <i>KRT31</i> con el diámetro de fibra de alpaca	40
VI.	DISCUSIÓN	45
VII	CONCLUSIONES	50
VIII.	RECOMENDACIONES	51
IX.	REFERENCIAS BIBLIOGRÁFICAS	52
X.	ANEXOS	62
	ANEXO 1: Ficha de evaluación	63
	ANEXO 2: Registro de características de las alpacas Huacaya	64
	ANEXO 3: Relación de los 14 cebadores	67
	ANEXO 4: Relación de los 22 genes en rumiantes	68
	ANEXO 5: Grafico del outlier del alelo T/T en el diámetro de fibra	69
	ANEXO 6: Genotipos del gen <i>EDAR</i>	70
	ANEXO 7: Genotipos del gen <i>KRT31</i>	74
	ANEXO 8: Genotipos del gen <i>HOXC13</i>	77

ÍNDICE DE TABLAS

Tabla 1.	Variantes en regiones no codantes del gen candidato <i>EDAR</i>	31
Tabla 2.	ariantes sinónimas del gen candidato <i>EDAR</i>	31
Tabla 3.	Variantes no sinónimas de los genes candidatos <i>EDAR</i> y <i>HOXC13</i>	31
Tabla 4.	Número de alelos (A), heterocigosidad observada (H_o), heterocigosidad esperada (H_E), p-valoré (valor de probabilidad) para HWE (Deficiencia heterocigotos), F_{IS} (coeficiente de endogamia) y Rho_{IS} para 10 microsatélites en población de alpacas de Lachocc (N=96).	36
Tabla 5 .	Ubicación de los transcritos codificados por los genes regulatorios <i>EDAR</i> y <i>HOXC13</i> y el estructural <i>KRT31</i> en el genoma de alpaca.	38
Tabla 6.	Características de los cebadores utilizados para la amplificación de los genes <i>EDAR</i> , <i>HOXC13</i> y <i>KRT31</i>	39
Tabla 7.	Frecuencias alélicas de los sitios variables presentes en el gen <i>EDAR</i> (N=86)	40
Tabla 8.	Frecuencias alélicas de los sitios variables presentes en el gen <i>HOXC13</i> (N=91)	41
Tabla 9.	Frecuencias alélicas de los sitios variables presentes en el gen <i>KRT31</i> (N=92)	41
Tabla 10.	Valores de heterocigosidad observada (H_o), heterocigosidad esperada (H_E), Equilibrio de Hardy Weinberg y F_{IS} para los SNP del gen <i>EDAR</i>	42
Tabla 11.	Valores de heterocigosidad observada (H_o), heterocigosidad esperada (H_E), Equilibrio de Hardy Weinberg y F_{IS} para el gen	43
Tabla 12.	Valores de heterocigosidad observada (H_o), heterocigosidad esperada (H_E), Equilibrio de Hardy Weinberg y F_{IS} para el gen <i>KRT31 HOXC13</i>	43
Tabla 13.	Análisis de asociación del diámetro de la fibra y SNPs presentes en el gen <i>EDAR</i> usando un modelo aditivo	45
Tabla 14.	Análisis de asociación del diámetro de la fibra y SNPs presentes en el gen <i>HOXC13</i> usando un modelo aditivo	47
Tabla 15.	Análisis de asociación del diámetro de la fibra y SNPs presentes en el gen <i>KRT31</i> usando un modelo aditivo	47

ÍNDICE DE FIGURAS

Figura 1.	Diagrama mostrando la ubicación de tres loci microsatélites	06
Figura 2.	Representación de un SNP (G*A) en una secuencia de ADN. Modificado de Liu (2007).	08
Figura 3.	Flujograma para elección de genes candidatos	27
Figura 4-7	Resultados gráficos de los análisis de estructura obtenidos por el programa Structure v2.3.	36
Figura 8.	Potencia estadística del análisis de asociación genética bajo el modelo aditivo.	44

RESUMEN

El objetivo del presente estudio fue identificar y validar polimorfismos de nucleótidos únicos (SNP) en tres genes, dos de los cuales codifican para proteínas funcionales (*EDAR* y *HOXC13*) y uno para proteína estructural (*KRT31*) en fibra y, evaluar su posible asociación con el diámetro de fibra de alpaca. Un total de 96 alpacas reproductoras no emparentadas fueron elegidas a partir de la evaluación fenotípica mediante los principales indicadores raciales y textiles de un total de 800 alpacas. El análisis de diversidad genética basada en genotipificación de 10 microsatélites (*YWLL08*, *YWLL29*, *YWLL36*, *YWLL40*, *YWLL44*, *LCA05*, *LCA08*, *LCA19*, *LCA37*, *LCA66*) reporta un total de 128 alelos con niveles de variabilidad genética elevada y baja consanguinidad ($H_E > 0.68$, $F_{IS} < 0.025$) y ausencia de estructuración poblacional ($K=2$) para la población en estudio. Estos tres genes fueron elegidos a partir de 22 genes funcionales y estructurales previamente seleccionados y anotados a partir de 4 genomas de referencia. Un análisis de variabilidad genética, identifico 34 variantes tipo SNP distribuidos en los genes *EDAR* (21 / 61.7%), *HOXC13* (5 / 14.8%) y *KRT31* (8 / 23.5%). Todos los genes estuvieron en HWE y mostraron niveles moderados de variabilidad genética ($H_E > 0.3$), los genes *EDAR* y *KRT31* mostraron niveles de consanguinidad baja ($F_{IS} < 0.02$) en contraste con el gen *HOXC13* ($F_{IS} = 0.108$). Un análisis de asociación genética entre marcadores SNP de los genes *EDAR*, *HOXC13*, *KRT31* y diámetro de fibra mediante regresión múltiple bajo modelo aditivo indican que no se detectó asociación con dicho carácter.

Palabras clave: Alpaca, SNP, microsatélite, *HOXC13*, *EDAR*, *KRT31*, variabilidad, asociación.

ABSTRACT

The aim of the present study was to identify and validate single nucleotide polymorphisms (SNP) in three genes, two of which code for functional proteins (*EDAR* and *HOXC13*) and one for structural protein (*KRT31*) in fiber, and to evaluate their possible association with the alpaca fiber diameter. 96 non-related breeding alpacas were chosen using a phenotypic evaluation by the main racial and textile indicators from 800 alpacas. The genetic diversity analysis based on genotyping of 10 microsatellites (*YWLL08*, *YWLL29*, *YWLL36*, *YWLL40*, *YWLL44*, *LCA05*, *LCA08*, *LCA19*, *LCA37*, *LCA66*) reports a total of 128 alleles with high levels of genetic variability and low consanguinity ($H_E > 0.68$, $F_{IS} < 0.025$) and absence of population structure ($K = 2$) for the population under study. These three genes were chosen from 22 functional and structural genes previously selected and annotated from 4 reference genomes. An analysis of genetic variability, identified 34 SNP-type variants distributed in the *EDAR* (21 / 61.7%), *HOXC13* (5 / 14.8%) and *KRT31* (8 / 23.5%) genes. All the genes were in HWE and showed moderate levels of genetic variability ($H_E > 0.3$), the *EDAR* and *KRT31* genes showed low inbreeding levels ($F_{IS} < 0.02$) in contrast to the *HOXC13* gene ($F_{IS} = 0.108$). An analysis of genetic association between SNP markers of *EDAR*, *HOXC13*, *KRT31* genes and fiber diameter by multiple regressions under the additive model indicate the absence of association with that character.

Key word: Alpaca, SNP, microsatélite, *HOXC13*, *EDAR*, *KRT31*, variability, association.

I. INTRODUCCIÓN

La alpaca (*Vicugna pacos*) es un animal económicamente importante para las familias altoandinas del Perú debido a que de ésta se obtiene carne y fibra (FAO, 2005). El Perú posee aproximadamente tres millones y medio de alpacas y constituye cerca del 90% de las alpacas a nivel mundial (FAO, 2005), siendo más de 120,000 productores asociados a esta actividad representando el 11% de la población humana altoandina (CONACS, 2004). Sin embargo existe el problema en cuanto a la mala calidad del diámetro de la fibra lo cual reduce el precio de ésta afectando los ingresos económicos a los productores de alpaca (Frank *et al.*, 2006).

La producción de la alpaca en el Perú se encuentra en manos de pequeños productores altoandinos quienes carecen de conocimiento técnico respecto a la producción de alpacas lo cual no permite convertirla en una actividad rentable, además de ser carentes de economía lo que finalmente se traduce en una explotación de baja calidad tanto en carne como en fibra. En cuanto al diámetro de la fibra en alpacas, es consecuencia de la ausencia de programas efectivos de mejoramiento genético, caracterizados por un casi nulo control de apareamientos, ausencia de registros productivos y genealógicos que dan como resultado la producción de crías de diferentes calidades respecto al diámetro de fibra, en consecuencia el diámetro de la fibra de alpaca es muy variable y suele presentar valores intermedios al de la vicuña (13 μm) y llama chaccu (29 μm) (Valbonesi *et al.*, 2010)

El diámetro de la fibra de alpaca oscila generalmente entre 18 y 27 μm (Paredes *et al.*, 2013) con ligeras diferencias entre regiones, por citar Arequipa y Puno con 18 μm (Braga *et al.*, 2007), y Huancavelica con valores promedio de 22 μm (Montes *et al.*, 2013).

Existe una hipótesis referente a que la calidad variable de la fibra de la alpaca se deba a la posible hibridación o mestizaje con la llama (*Lama glama*) (Kadwell *et al.*, 2001). Sin embargo no se ha determinado si el mestizaje de

llama con alpaca sucedió luego de la conquista o si la llama formó parte del proceso de domesticación inicial (Wheeler, 2001).

La presencia de momias de alpacas y llamas de calidades extraordinarias de fibra (Wheeler, 2001), sugieren que la alta calidad de la fibra de los camélidos sudamericanos puede ser alcanzada con la implementación de programas de reproducción y mejoramiento genético.

En consecuencia el presente trabajo de investigación propone como objetivo general, identificar polimorfismos genéticos potencialmente asociados al diámetro de la fibra en alpacas.

II. ANTECEDENTES

2.1. La alpaca

La alpaca (*Vicugna pacos*) es un mamífero ungulado perteneciente al Phylum Vertebrata, Orden Artiodactyla, Suborden Tilopoda, Familia Camelidae, Género Vicugna y habita en las zonas altoandinas del Perú, Chile, Bolivia y Argentina (Gentry *et al.*, 2004).

Evolutivamente está emparentada con la vicuña aunque en las poblaciones actuales hay una introgresión genética con la llama (*Lama glama*) siendo la fecha de su domesticación hace miles de años, prueba de ello podemos encontrar restos arqueológicos en la cultura Mochica (Wheeler, 2001)

La relación con los otros camélidos ha sido controvertida durante muchos años por lo que se le denominó inicialmente *Lama guanicoe* primeramente, luego *Lama pacos* y finalmente cuando se determinó que era descendiente filogenético de la vicuña se le denominó *Vicugna pacos* lo cual se determinó gracias al análisis del ADN (Wheeler, 2001)

Las alpacas pesan entre 45 y 77 kilos y su altura a la cruz es de 90 centímetros y poseen un diámetro de fibra desde 18 a 28 micras

2.1.1. Distribución

La alpaca se distribuye desde las mesetas de Argentina, Bolivia, Chile, Ecuador hasta los andes del Perú, en donde converge la mayor población de alpacas ((Wheeler, 2001).

El Perú cuenta con 3.3 millones de alpacas y en consecuencia es la población más grande a nivel mundial, y se distribuyen en las zonas alto andinas del Perú a más de 3,400 metros de altitud bajo sistemas de explotación extensivo con bajos niveles de producción (FAO, 2005). La mayor población de alpacas se encuentra en la región de Puno con 58% seguido por Cusco con 12% y luego el departamento de Huancavelica con 11% (FAO, 2005).

2.1.2. Origen evolutivo de la alpaca

Los camélidos sudamericanos se clasifican juntamente con los de Asia dentro de la familia Camelidae (Wheeler, 2001). Esta se subdivide en dos subfamilias, Lamini y Camelini, la subfamilia Camelini se dirigió hace aproximadamente 11 millones de años a Norte-américa y la subfamilia Lamini hacia Sudamérica. La subfamilia Lamini compone 4 especies de los cuales dos son silvestres: el guanaco (*Lama guanicoe*) y la vicuña (*Vicugna vicugna*) y dos son domesticas la alpaca (*Vicugna pacos*) y la llama (*Lama glama*), por lo que el origen de la alpaca es un tema polémico y se han propuesto diferentes hipótesis sobre su origen indicando que derivan del guanaco, llama o vicuña (Wheeler, 2001).

Estudios arqueozoológicos y análisis filogenético utilizando marcadores microsatélites sugieren que la alpaca es producto de la domesticación de la vicuña en los andes del Perú hace 7,000 años (Kadwell *et al.*, 2001), sin embargo, se han reportado elevados grados de introgresión con alelos del guanaco y la llama lo cual abre la hipótesis de una posible contribución en el proceso de domesticación.

Actualmente se reconocen dos tipos de alpacas las que se diferencian fundamentalmente por las características externas del vellón:

La alpaca Huacaya que posee una fibra de buena densidad, suavidad, lustre y rizos que le confieren un aspecto esponjoso, y la alpaca Suri que posee una fibra que forma rulos con un lustro más notable y brillante que en la Huacaya. La fibra de la alpaca Huacaya crece en forma perpendicular al cuerpo y de la Suri en forma paralela a su cuerpo (CONACS, 2003)

2.2. Caracterización molecular

La caracterización molecular se realiza fundamentalmente para explorar la diversidad genética dentro y entre poblaciones y para determinar relaciones genéticas entre ellas, específicamente para diferenciar un tipo de otro, identificar localizaciones geográficas y los individuos dentro de ellas, asimismo, proporcionan información sobre la evolución (árboles filogenéticos), rastreo de genes portadores conocidos, identificar relaciones de parentesco, entre otras aplicaciones (Kadweel *et al.*, 2001)

2.3. Marcadores moleculares para estudios de diversidad genética

Los marcadores moleculares son secuencias de ADN de localización definida en el genoma (locus específico), son secuencias polimórficas insensibles al medio ambiente y son de herencia conocida (Agarwal *et al.*, 2008), estos pueden ser de dos tipos: marcadores tipo I, los que están relacionados a secuencias codificantes del genoma y, marcadores tipo II, los que se encuentran en regiones anónimas o no codificantes (Chauhan, 2010)

La diversidad fenotípica entre organismos es consecuencia de las diferencias en las secuencias de ADN y de los efectos ambientales, cada individuo de una especie (a excepción de los individuos monocigóticos) posee una secuencia de ADN única, esta diferencia en el ADN es generada por mutaciones ya sea por inserción, delección, duplicación o inversión de fragmentos de ADN de diversas longitudes. Las variaciones de ADN se consideran neutras cuando no originan cambios en el fenotipo y por consiguiente no están sometidas a selección, también pueden suceder mutaciones en regiones reguladoras que pueden afectar la expresión génica como por ejemplo la activación, desactivación o sobreexpresión de los genes (Kadweel *et al.*, 2001).

Los marcadores moleculares más importantes que se usan en la actualidad para fines de investigación en mamíferos son los microsatélites y los polimorfismos de nucleótidos únicos (Single nucleotide polymorphism, SNP).

2.3.1. Microsatélites

Los microsatélites, también llamados secuencias simples repetidas (SSR) o secuencias cortas repetidas en tándem (STR) son secuencias cortas de ADN que consisten en repeticiones de 1 a 6 pares de bases de longitud localizadas a lo largo del genoma de los organismos eucariotas localizado en el ADN del núcleo, mitocondrias y los cloroplastos (Schlotterer, 2001) (figura 1)



Figura 1. Diagrama mostrando la ubicación de tres loci microsatélites (TA)_n, (TAA)_n y (GC)_n en un segmento cromosomal. Se aprecia el grado de homocigosidad para el locus 1 (TA)₄, y heterocigosidad para el locus 2 (TAA)₇, (TAA)₈ y el locus 3 (GC)₇, (GC)₈ Modificado por Freeland (2005)

a. Distribución, composición y frecuencia

La distribución de los microsatélites es relativamente uniforme dentro del genoma eucariota (Toth *et al.*, 2000), sin embargo, se observa una mayor frecuencia en regiones no codificantes comparados con las regiones codificantes en aproximadamente 15% en vertebrados (Chistiakov *et al.*, 2006). La frecuencia de distribución de los microsatélites en el genoma tiene relación con el tipo de repetición y la composición de la secuencia repetida (Hancock, 1991) siendo inversamente proporcional con el número de repeticiones (Schlotterer, 2001), las repeticiones nucleotídicas son las más abundantes (67% del total de microsatélites) (Chistiakov *et al.*, 2006), mientras que las repeticiones tri, tetra, penta y hexa nucleotídicas son alrededor de 1.5 veces menos frecuentes (Toth *et al.*, 2000). En el genoma de los vertebrados las repeticiones nucleotídicas más frecuentes son (AC)_n (GT)_n y (AT)_n (Eisen, 1991) mientras que las repeticiones tri nucleotídicas más comunes son las (CAG)_n y (AAT)_n (CCG)_n (Toth *et al.*, 2000).

Los microsatélites tienen mayor utilidad frente a otros marcadores genéticos dado que tienen el más alto grado de polimorfismo, se segregan de manera mendeliana y son codominantes (Freeland, 2005).

Además de la clasificación respecto a la cantidad de nucleótidos que se repiten también se pueden clasificar en microsatélites perfectos o puros, cuando las unidades nucleotídicas no son interrumpidas en su longitud por ejemplo (CACACACACA)₆, también se pueden identificar los microsatélites compuestos los cuales contienen dos clases de secuencias nucleotídicas (TATATATAGAGAGAGA) simbolizado como (TA)₄, (GA)₄ y finalmente los microsatélites interrumpidos los cuales tienen repeticiones separadas por un máximo de cuatro pares de bases (GCGCGCGCCTTTTGGCGCGGC) o (GC)₅ TTTT (GC)₄ (Butler, 2005).

b. Tasa de mutación de los microsatélites

Los microsatélites poseen tasas de mutación mayores que otros tipos de secuencias de ADN (Freeland, 2005) en genomas eucariotas es de 5×10^{-6} (Brookes, 1999). La inestabilidad de las mutaciones de los microsatélites puede ser considerada como un balance entre la generación de errores durante la replicación y la corrección de algunos de estos errores por los sistemas de corrección celulares (Ellegren, 2000).

La tasa de mutación de los microsatélites difiere entre y dentro de una misma especie siendo los factores determinantes la longitud del microsatélite y el tipo de repeticiones (Schlotterer, 2001). Los microsatélites con un elevado número de repeticiones causan una mayor tasa de mutaciones, de la misma manera los microsatélites con repeticiones dinucleótidas poseen tasas de mutación 1.5 a 2 veces mayores que los tri y tetra nucleótidos (Eisen, 1991).

Se postula que las mutaciones en microsatélites son producidas por mecanismos de deslizamiento de la ADN polimerasa durante la replicación del ADN molde y en consecuencia mal emparejamiento de la hebra o SSM (Slipped Strand Mispairing) (Hancock, 1991). El SSM ocurre en un elevado grado en los microsatélites debido a la presencia de repeticiones, incrementando la posibilidad de un mal alineamiento durante la replicación de la cadena del ADN molde (Eisen, 1991) lo cual se traduce en cambios en el número de repeticiones.

El SSM sucede en tres pasos: i) deslizamiento de la ADN polimerasa durante la replicación, ii) un error en el realineamiento del ADN molde en las cadenas replicadas y iii) la continuación de la replicación de los moldes mal alineados (Eisen, 1991).

2.3.2. Polimorfismo de nucleótidos únicos (SNP)

Los SNP son cambios de una simple base en una secuencia del ADN (Liu, 2014). Los SNP presentan diferentes alternativas de secuencia (alelos) en individuos de una o varias poblaciones dadas (figura 2) sin embargo, para ser considerado un SNP este debe presentar frecuencias alélicas mayores al 1% en la población (Vignal *et al.*, 2002).

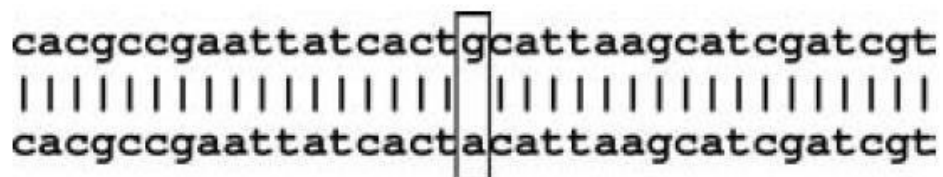


Figura 2. Representación de un SNP (G*A) en una secuencia de ADN. Modificado de Liu (2007).

A los SNP se consideran como marcadores principalmente bialelicos (Vignal *et al.*, 2002) sin embargo, en humanos se ha

demostrado la existencia de SNP tri y tetra alélicos pero con una frecuencia muy rara al punto de casi no existir (Brookes, 1999).

Los SNP son bialelicos en la práctica debido a dos razones principales por la baja frecuencia de sustituciones de nucleótidos únicos que va de 1×10^{-9} a 5×10^{-9} por nucleótido por año en posiciones neutrales en mamíferos (Liu *et al.*, 2007) y al sesgo en las mutaciones con una mayor frecuencia de transiciones sobre transversiones.

Los SNP se heredan como marcadores bialelicos codominantes en forma mendeliana simple. El grado de polimorfismo medido como contenido de información polimórfica (PIC) no es alto como en los microsatélites, pero su gran abundancia en el genoma balancea esta deficiencia (Liu, 2014).

Los SNP son el polimorfismo más abundante en el genoma de cualquier tipo de organismo y constituye el último nivel de variación genética entre individuos (Vignal *et al.*, 2002). Del total de variación genética en humanos, un 90% corresponden a SNP. En humanos se han reportado en la base de datos 5 millones de SNP de los cuales 4 millones han sido validados (Sobrino *et al.*, 2005). Los SNP se pueden encontrar en regiones codificantes y no codificantes del genoma sin embargo existe una mayor densidad de SNP en regiones no codificantes (Syvanen, 2001).

2.4. Diversidad genética

La diversidad genética se define como la variación o diferencias genéticas dentro de una población o especie (Freeland, 2005).

Constituye la característica más importante en cualquier población por ser primordial para su adaptación a cambios en el medio ambiente, capacidad de sobrevivencia a largo plazo y evolución (Nei, 2000).

La caracterización de la diversidad genética de una población es importante porque permite conocer su capacidad de evolución y en animales domésticos su capacidad de selección. Actualmente se cuentan con marcadores genéticos tales como los microsatélites y los SNP que

permiten cuantificar y determinar la variabilidad genética de las poblaciones (Nei, 2000).

2.4.1. Estimadores de la diversidad genética

La diversidad genética se estima utilizando marcadores genéticos, los cuales se basan principalmente en frecuencias alélicas y frecuencias genotípicas siendo los principales estimadores: la diversidad alélica (A), heterocigosidad esperada (H_E), heterocigosidad observada (H_o), el coeficiente de consanguinidad (F_{IS}) y el equilibrio de HWE (Freeland, 2005)

a. Frecuencia alélica y diversidad alélica

La frecuencia relativa de un alelo en particular en una población, es considerada un indicador fundamental en estudios evolutivos, porque un cambio genético en una población es usualmente descrito por un cambio en las frecuencias de sus alelos (Nei, 2000). La diversidad alélica (A) constituye simplemente el número promedio de alelos por locus (Nei, 2000).

b. Heterocigosidad

La heterocigosidad (H) constituye la proporción de individuos heterocigotos en la población. Un valor de heterocigosidad elevado significa que existe una elevada diversidad alélica presente en la muestra analizada (Butler, 2005).

La heterocigosidad puede ser dividida en dos indicadores: la heterocigosidad esperada (H_E) y la heterocigosidad observada (H_o).

La heterocigosidad esperada (H_E), es definida como la probabilidad de que dos alelos escogidos al azar en una población en un locus sean diferentes (Weir y Cockerham, 1984).

$$H_E = 1 - \sum_{i=1}^n p_i^2$$

Dónde:

p_i , es la frecuencia del alelo i y, n es el número de alelos en un locus (Nei, 1972)

La heterocigosidad observada (H_o) es definida como la proporción de individuos heterocigotos observados para un locus en particular (Weir y Cockerham, 1984).

$$H_o = \frac{\text{Número de individuos heterocigotos}}{\text{Número total de individuos}}$$

c. Equilibrio de Hardy-Weinberg

En organismos con reproducción sexual donde el apareamiento ocurre al azar, los genotipos son producidos por la unión al azar de los gametos masculino y femenino, por lo tanto, bajo ciertas condiciones las frecuencias genotípicas dentro de una población pueden seguir un patrón predecible basado en las frecuencias alélicas. Condiciones como apareamiento al azar dentro de la población o panmixia, ausencia de selección, ausencia de migración, tasa de mutación despreciable, segregación alélica mendeliana y tamaño efectivo poblacional infinito permiten predecir las frecuencias de los genotipos presentes dentro de una población en base a las frecuencias alélicas (Freeland, 2005).

Esta propiedad fue reportada independientemente por Godfrey Hardy, matemático inglés y Wilhem Weinberg, físico alemán en 1908, por lo cual fue denominada principio de Hardy-Weinberg (Nei, 2000). Una población que presenta las condiciones antes mencionadas se denomina una población en equilibrio de Hardy-Weinberg. (HWE).

Las desviaciones del equilibrio HWE pueden deberse a la selección, migración, subdivisión poblacional no detectada (efecto Wahlund), baja variabilidad de los loci, poblaciones pequeñas (menores de 20 individuos) o simplemente error en el muestreo (Nei, 2000).

Las desviaciones de las proporciones de Hardy-Weinberg pueden ser medidas por un simple parámetro denominado índice de fijación o F (Wright, 1969).

d. Coeficiente de consanguinidad (F o F_{IS})

Consanguinidad se define como el resultado del apareamiento entre individuos estrechamente relacionados. La forma más intensa de consanguinidad es originada por autofecundación y es dada en individuos hermafroditas. La consanguinidad puede ser medido mediante dos importantes coeficientes: el coeficiente de parentesco o coancestría (F_{JK}) y el coeficiente de consanguinidad (F , F_I o F_{IS}) (Maynard, 1999).

El coeficiente de parentesco o coeficiente de coancestría (F_{JK}) constituye la probabilidad que dos genes homólogos en dos individuos J y K seleccionados al azar sean idénticos por descendencia (Pritchard *et al.*, 2000). El concepto de identidad por descendencia se define como la propiedad que dos alelos dados sean copias de un solo alelo que existió en un ancestro reciente (Freeland, 2005).

El coeficiente de consanguinidad (F , F_I o F_{IS}) constituye la probabilidad que dos alelos en un locus de un individuo sean idénticos por descendencia (Freeland, 2005). El coeficiente de consanguinidad describe la proporción por la cual la heterocigosidad disminuye en relación con una población que presenta apareamientos al azar asumiendo que ambas presentan la misma frecuencia génica inicial (Maynard, 1999).

2.5. Estructura genética

La estructura genética de una población permite entender como la variabilidad genética se distribuye dentro de poblaciones y subpoblaciones de individuos de una misma especie (Templeton, 2006).

2.5.1. Factores que influyen en la estructura genética

La estructura genética está influenciada por múltiples factores siendo los principales la deriva genética, el flujo genético (migración), efecto Wahlund entre otros (Templeton, 2006).

a. Deriva genética

La deriva genética es un proceso que causa cambios en las frecuencias alélicas en una población como resultado del azar (Freeland, 2005). La deriva genética no posee una dirección determinada, los cambios generados en la frecuencia alélica no permiten un retorno a las frecuencias alélicas ancestrales debido a que dichos cambios son acumulativos en el tiempo, la deriva genética es un mecanismo de la evolución en el que las frecuencias alélicas de una población cambian a lo largo de varias generaciones debido al azar (error de muestreo) la deriva genética sucede en todas las poblaciones cuyo tamaño no es infinito pero sus efectos son más fuertes en poblaciones pequeñas llegando al extremo de fijar o extinguir un alelo en la población, por tanto provocando la pérdida de variación genética dentro de una población finita (Templeton, 2006).

También se puede definir como el producto de las diferencias en el éxito reproductivo (mayor número de crías) de algunos individuos en comparación con otros (Templeton, 2006).

b. Flujo genético

Se denomina flujo genético a la transferencia del material genético de una generación a otra siguiente entre poblaciones por medio de la reproducción de individuos e implica la dispersión de nuevas variantes genéticas entre poblaciones diferentes (Freeland, 2005). La intensidad del flujo genético de una población es importante para la diferenciación genética y tiende a homogenizar las poblaciones (Freeland, 2005).

Los indicadores de variabilidad como son la deriva genética, la diversidad genética, la diferenciación poblacional y el número efectivo de la población (N_E) mantienen una relación estrecha con el flujo genético.

En consecuencia poblaciones con flujo genético constante tienden a ser poblaciones uniformes con niveles bajos de diferenciación poblacional debido a un efecto reducido de la deriva genética, mientras que poblaciones con bajo flujo genético tienden a ser poblaciones aisladas con niveles elevados de diferenciación poblacional, y puede ser medido marcando a los individuos o usando un método indirecto denominado “número efectivo de migrantes” (N_{em}) (Freeland, 2005).

Los grandes cambios de la frecuencia genética de una generación a otra siempre se deben a la migración como factor de la dinámica poblacional. El hecho de intercambiar genes y con la intensidad que se presenta permite cambiar la frecuencia genética en cada generación subsiguiente (Altukhov, 2006).

El apareamiento al azar propiamente dicho es una cuestión ideal. En poblaciones reales, los individuos se reproducen principalmente con individuos de poblaciones vecinas. Existen tres modelos que tratan de explicar este fenómeno reproductivo: el modelo isla de Wright, el modelo de “stepping-stone” y el modelo continuo (Maynard, 1999).

El modelo isla de Wright, propone que si una población está dividida en subpoblaciones, una pequeña fracción de estos individuos tienen la capacidad de migrar indistintamente a cualquier subpoblación vecina sin importar su dirección.

El modelo “stepping-stone”, es parecido al modelo de isla con la diferencia que los migrantes pueden moverse solo en dirección de la subpoblación vecina.

El modelo continuo, asume migración continua de individuos dentro de la gran población, propone la ausencia de subpoblaciones pero asume que la distancia de recorrido o de dispersión es corta.

c. Efecto Wahlund

Se denomina Efecto Wahlund al exceso de la frecuencia de homocigotos en comparación con la proporción de Hardy-Wienberg debido a la subdivisión intrapoblacional (Nei, 2000). Si los datos genéticos de dos o más poblaciones con diferentes frecuencias alélicas son combinadas, es posible apreciar un efecto Wahlund, lo cual se traduce en un incremento en la proporción de homocigotos en la muestra agregada en comparación con una muestra disgregada (Freeland, 2005).

2.5.2. Métodos de estimación de la estructura genética en poblaciones

Existen diferentes métodos para estimar la estructuración genética en poblaciones, los más utilizados son: el estadístico F y el análisis de estructura.

a. Estadístico F

El estadístico F o estadístico de Wright, usa el coeficiente de consanguinidad para dividir la variación genética dentro y entre poblaciones (Freeland, 2005).

El estadístico F cuenta con tres indicadores: F_{IS} , F_{ST} y F_{IT} (Altukhov, 2006) los cuales son definidos en términos de heterocigosidad (Nei, 2000).

El F_{IS} , es el coeficiente de endogamia de un individuo (**I**) respecto de la subpoblación (**S**) también denominado coeficiente de consanguinidad, constituye la probabilidad de que dos alelos en un locus sean idénticos por descendencia (Maynard, 1999). El F_{IS} refleja la cantidad de consanguinidad presente dentro de una subpoblación (Freeland, 2005) y se determina mediante la siguiente formula:

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

Dónde:

H_o, es la heterocigosidad observada dentro de una población y

H_s, es la heterocigosidad esperada dentro de una población.

El F_{IT} , es el coeficiente de endogamia de un individuo (**I**) respecto a la población total (**T**), refleja la cantidad total de consanguinidad en una población, es considerado un coeficiente global de consanguinidad que compara la heterocigosidad de un individuo con el total de la población bajo consideración (Freeland, 2005). Se determina mediante la siguiente formula:

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

Dónde:

H_o , es la heterocigosidad observada dentro de una población y

H_T , es la heterocigosidad total esperada

El valor F_{IT} se encuentra influenciado por el F_{IS} y el F_{ST} (Freeland, 2005). Los indicadores F_{IS} y F_{IT} pueden ser negativos cuando la H_o es inusualmente elevada (Nei, 2000).

El F_{ST} , es la diferencia de las subpoblaciones (**S**) comparado al total de la población (**T**), es considerado como la probabilidad de que dos gametos tomados al azar dentro de una misma subpoblación sean idénticos por descendencia (Nei, 2000). El F_{ST} refleja la cantidad de consanguinidad en una población debido a la presencia de subpoblaciones. Se determina mediante la siguiente formula:

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

Dónde:

H_s , es la heterocigosidad esperada dentro de una subpoblación y

H_T , es la heterocigosidad total esperada.

El F_{ST} es la medida más usada para describir la diferenciación genética en poblaciones; mide cuan diferentes genéticamente son dos poblaciones y refleja la proporción de la variabilidad genética total que se debe a las diferencias netas entre las poblaciones (Excoffier, 2006).

Los valores de F_{ST} pueden variar entre 0 y 1. Si dos poblaciones presentan frecuencias alélicas iguales, entonces el valor de F_{ST} es igual a cero. En contraste, si dos poblaciones presentan frecuencias alélicas totalmente diferentes, entonces el valor de F_{ST} es igual a 1.

Valores de F_{ST} comprendido entre 0 a 0.049 se relacionan con un nivel bajo de diferenciación genética, entre 0.050 a 0.250 se relacionan con un nivel moderado de diferenciación genética, y valores mayores de 0.25 se relacionan con un nivel elevado de diferenciación genética (Freeland, 2005).

Existen análogos estadísticos a los indicadores propuestos por *Wright*, como el estadístico G (G_{ST}) de *Nei*, el estadístico θ de *Weir* y *Cockerham* y el estadístico R (R_{ST}) de *Slatkin*.

b. Análisis de estructura

El análisis de estructura permite identificar de forma probabilística poblaciones genéticamente distintas, asignar individuos a poblaciones, identificar migrantes, identificar individuos híbridos e identificar zonas de hibridación (Pritchard *et al.*, 2000).

✓ **Análisis Bayesiano**

Los métodos de análisis de estructura poblacional se basan en el modelo de agrupamiento bayesiano (o de máxima probabilidad) que utiliza el método de Cadena de Markov-Monte Carlo (MCMC) (Excoffier, 2006).

Este método usa información genotípica de múltiples loci (mínimo de 7), asumiendo la presencia de equilibrio de ligamiento y HWE entre los diferentes loci y poblaciones (Penedo *et al.*, 1999). Este modelo no asume mutación y puede ser aplicado a diferentes tipos de marcadores genéticos como microsatélites y SNP.

✓ **Cadena de Markov-Monte Carlo**

La Cadena de Markov es una metodología basada en el teorema de Bayes y permite generar distribuciones de probabilidad que son difíciles o imposibles de obtener a partir de ecuaciones analíticas.

La MCMC permite la simulación de un tipo especial de proceso estocástico conocido como Cadena de Markov (Hamilton, 2009), el cual genera una serie de variables aleatorias cuyo estado futuro depende del estado actual en cualquier punto de la cadena, permitiendo obtener muestras aleatorias del espacio muestral aún cuando este espacio es enorme (por ejemplo, miles de filogenias)

2.6. Estudios previos sobre diversidad genética

Ordoñez, (2017), en un trabajo de investigación realizado con un total de 113 muestras ovinos de diferentes subpoblaciones en la región Huancavelica, utilizando un panel de 11 marcadores microsatélites fluoromarcados (*CSRD247*, *HSC*, *OarAE129*, *MAF65*, *OarCP0049*, *OarFCB304*; *BM1818*, *INRA63*, *BM1258*, *OarFCB20*, *McM527*), encontró una $H_o=0,718$, $H_e=0,790$, además de un desequilibrio HW para el 36% de locus analizados y un $F_{ST}=0,016$ con los que se puede inferir que existe alta variabilidad genética, así como un moderado grado de endogamia en la población analizada

En el ser humano, la sustitución de un residuo aminoácido de valina por uno de alanina en la posición 370 de la proteína *EDAR*, que interviene en la regulación del desarrollo del pelo, está asociada a una variación de 4.2-4.5 μm en el diámetro del pelo (Fujimoto *et al.*, 2008). Asimismo, polimorfismos del SNP rs4752566 del gen *FGFR2* humano está asociado a un incremento de entre 1.4 y 3.4 μm en el diámetro del pelo (gen que codifica la proteína tirosín-kinasa que interviene en la transducción de señales de un factor de crecimiento celular) (Fujimoto *et al.*, 2009).

Estudios sobre diversidad genética en poblaciones de ovinos de Jordania reporta valores de H_e de 0.82, H_o de 0.68 y número de alelos por locus de 12.67 (Fujimoto *et al.*, 2009) en contraste con lo encontrado en ovinos de indonesia con número promedio de alelos de 6.1, H_o y H_e de 0.57 y 0.68 respectivamente (Fujimoto *et al.*, 2008).

En camélidos sudamericanos, estudios sobre diversidad genética basados en marcadores microsatélites, han sido realizados en Puno (Centro Experimental Illpa del INIA) en llamas, y en diferentes regiones del Perú en alpacas (Rodríguez, 2009). Los estudios realizados en Perú reportan una elevada variabilidad genética ($A = 12$; $H_e = 0.75$), elevada informatividad de los marcadores utilizados ($PIC = 0.72$) y una baja consanguinidad y diferenciación genética tanto en llamas como alpacas ($F_{is} < 0.02$; $F_{ST} < 0.02$) (Rodríguez, 2009) sugiriendo una débil estructura poblacional y un intenso flujo genético (Perez *et al.*, 2010).

Estudios de diversidad genética en poblaciones de animales domésticos en Sudamérica muestra moderada a elevada variabilidad, probablemente

al uso de razas autóctonas en la generación de nuevas razas o procesos de introgresión. Poblaciones de ovinos domésticos de Colombia presentaron valores de H_E y H_o de 0.44 y 0.67 respectivamente (Paredes *et al.*, 2011) similar a lo reportado para ovinos Corriedale en Uruguay con valores de H_o de 0.66 (Kadwell *et al.*, 2001).

Por otro lado, se han identificado genes reguladores que poseen variantes asociadas al diámetro del pelo y lana. Por ejemplo, en un estudio de asociación de genoma completo (GWAS, *Genome-Wide Association Study*) en ovejas merino de China, se identificó 9 SNP asociados al diámetro de la lana los cuales se ubican en o cerca de genes que codifican proteínas que regulan diferentes procesos, tales como la división celular, transporte de bases nitrogenadas glucosiladas o transducción de señales ((Wang *et al.*, 2014).

Estudios en cabras Angora, kilis y Honamli en Turquía reportan moderada a elevada variabilidad genética ($A = 15.65$, $H_E = 0.51$, $F_{IS} = 0.035$) (Cervantes *et al.*, 2010). Cabras criollas venezolanas presentan valores de H_E de 0.65 y número medio de alelos de 6.2 (Wang *et al.*, 2014).

Con la finalidad de contribuir a los programas de mejoramiento genético para camélidos sudamericanos, Yaltta, (2014), buscó determinar la variabilidad genética y el nivel de endogamia de dos rebaños de reproducción de alpacas (*Vicugna pacos*) huacayas blancas de reciente formación, pertenecientes al Centro Piloto de Mejoramiento Genético Munay Paqocha y el Fundo Itita, de la Sociedad Peruana de Criadores de Alpacas y Llamas (SPAR)-Puno; así como realizar su genotipificación e identificación de marcadores STR útiles para la determinación de pruebas de paternidad y parentesco para lo cual evaluó 10 marcadores STR (*LCA37*, *LCA90*, *LCA5*, *VOLP92*, *YWLL36*, *YWLL44*, *YWLL08*, *VOLP92*, *LCA94*, *LCA90*) a partir de sangre y folículos pilosos de 183 individuos colectados al azar, encontrando en la población valores de F_{IS} (0,016), y F_{ST} (0,003) que reflejan bajos niveles de endogamia. El rebaño del Fundo Itita presentó una mayor H_o (0,858) respecto a la H_E (0,848), mientras que por el contrario el rebaño del Centro Munay Paqocha presentó un menor valor de la H_o (0,815) respecto a la H_E (0,848), con una tendencia al déficit de heterocigotos.

Fernández, (2015), analizó 7286 ESTs de piel de una alpaca del tipo Huacaya, resultando en 2428 secuencias ensambladas, posteriormente realizó la búsqueda de homologías y la anotación de estas secuencias llegando a identificar 22 genes relacionados a la síntesis de la fibra (*KRT10, KRT25, KRT31, KRT14, KRT15, KRT5, KRT78, KRT81, KRT85, KRTAP3-2, KRTAP10, KRTAP11-1, NFSF12, sFRP4, RACK1, COL1A1, COL1A2, DAB2IP, GLI1, BMP1, BMP4* y *CTNNB1*) en los que observó 4 genes con alta expresión (*TNFSF12, sFRP4, COL1A1* y *COL1A2*) lo cual indica que el gen *KRT31* incluido en el presente trabajo está comprometido con la fibra en alpacas.

Morón, (2016), analizo 247 alpacas 15 marcadores microsatélites en Cerro de Pasco con el objetivo de estimar el nivel de endogamia por unidad de producción encontrando un alto nivel de endogamia en los animales estudiados.

Los marcadores microsatélites (*LCA99, LCA71, LCA94, LCA66, YWLL36, LCA05, LCA77, LCA19, YWLL40, YWLL29, YWLL44* y *LCA08*), están totalmente validados para fines de filiación y consanguinidad, tal como lo demuestra Yucra, (2017), quien determinó la filiación y nivel de consanguinidad de alpacas Huacaya mediante los marcadores citados producto del cual encontró valores de consanguinidad de 0.00 % y el parentesco medio de la población de 1.99 %.

Fopiano, (2016), evaluó al gen *KRTAP11-1* respecto a la asociación con el diámetro de fibra para lo cual utilizó el genotipo de 2 de los 6 marcadores (SNP330 y SNP 956) validados en este gen sin encontrar ninguna asociación con el diámetro de fibra

Veliz, (2017), realizó un análisis de asociación de 5 SNP en el gen *BMP4*, 9 SNP en el gen *KAP1-1* y 4 SNP en el gen *KAP13-1* en los cuales no encontró ninguna asociación con el diámetro de la fibra de alpaca. El SNP 3451703 en el gen *KAP1-1* fue el más próximo a alcanzar la significancia estadística, sin embargo, este efecto se debió a un único valor extremo en uno de los genotipos

III. HIPÓTESIS Y OBJETIVOS

3.1. Hipótesis:

Los genes *HOXC13*, *EDAR* y *KRT31*, poseen polimorfismos de nucleótidos únicos (SNP) potencialmente asociados al diámetro de la fibra de alpaca.

3.2. Objetivos:

El objetivo general fue identificar polimorfismos genéticos tipo SNP en los genes *HOXC13*, *EDAR* y *KRT31* potencialmente asociados al diámetro de la fibra en alpacas Huacaya.

Los objetivos específicos fueron:

1. Caracterizar la diversidad genética de la población de alpacas Huacaya a partir de una muestra de 96 alpacas del Centro de Investigación Lachocc-UNH empleando 10 marcadores microsatélites
2. Identificar marcadores genéticos tipo SNP en 3 genes candidatos, dos de los cuales regulan el desarrollo de la fibra de alpaca (*HOXC13* y *EDAR*) y uno codifica una proteína estructural de la fibra de alpaca (*KRT31*).
3. Evaluar la asociación genética entre los SNP identificados en tres genes candidatos y el carácter diámetro de fibra de alpaca.

IV. MATERIALES Y MÉTODOS

4.1. Localización del estudio

El presente estudio se realizó en el Centro de Investigación Lachocc de la Universidad Nacional de Huancavelica (UNH), ubicado en el distrito, provincia y departamento de Huancavelica y, en las instalaciones del Laboratorio de Investigación y Desarrollo (LID) de la Universidad Peruana Cayetano Heredia, distrito de San Martín de Porres, provincia y departamento de Lima.

4.2. De los animales

Se utilizaron 96 alpacas Huacaya de color blanco (10 machos y 86 hembras) seleccionadas a partir de una población de 800 alpacas pertenecientes al Centro de Investigación Lachocc de la UNH, a las cuales se evaluaron en base a los principales descriptores raciales (anexo 1) y anotadas en una ficha fenotípica (anexo 2) que incluye la apariencia general, edad, peso vivo, uniformidad del vellón, diámetro, rizo y densidad folicular.

Los animales seleccionados fueron identificados con aretes colocados en la oreja izquierda, procediéndose al registro fotográfico (cabeza y cuerpo completo).

Del total de animales seleccionados se procedió a la colecta de 50 gramos de fibra proveniente de la zona del costillar medio del animal. Las muestras de fibra fueron analizadas para la característica diámetro de fibra en el Laboratorio de Análisis de Fibra de la UNH utilizando un equipo OFDA (*optical fibre diameter analyser*).

4.3. Colección de muestras de sangre

Se colectó 4 mL de sangre completa de las 96 alpacas mediante punción en la vena yugular utilizando aguja de colección de dos puntas Vacutainer® 21G 1½ y tubos BD Vacutainer® con anticoagulante (K₃E: Potasio – EDTA al 15%). Las muestras colectadas fueron almacenadas a 4 °C hasta su procesamiento

4.4. Extracción de ADN

Se extrajo ADN genómico a partir de leucocitos (PBMC: peripheral blood mononuclear cell) obtenidos de sangre periférica de 96 alpacas utilizando el kit comercial GF-1 Nucleic Acid Extraction Kits de Vivantis Technologies siguiendo las instrucciones del fabricante. El ADN obtenido fue diluido con una solución TE 1X (10 mM Tris-HCl, 1 mM disodium EDTA, pH 8.0) hasta una concentración final 5 ng/μL y colocado en tubos de microcentrífuga de 1.5 mL y almacenados a -20 °C.

4.5. Cuantificación del ADN

Se determinó la concentración del ADN genómico obtenido mediante espectrofotometría en ng/μL utilizando un equipo NanoDrop™ 1000 (Thermo Scientific).

La pureza se midió basada en la absorbancia a 260nm a través de la proporción de los valores de absorbancia a 260nm/280nm y 260nm/230nm. La integridad del ADN se visualizó por electroforesis en gel de agarosa al 1% TBE 1X (89 mM Tris-borate, 2 mM EDTA, pH 8.3).

4.6. Amplificación de ADN microsatélite

Se amplificaron 10 loci microsatélites descritos para alpacas (*Vicugna pacos*) y llamas (*Lama glama*): YWLL08, YWLL29, YWLL36, YWLL40, YWLL44, LCA05, LCA08, LCA19, LCA37, LCA66, mediante PCR múltiple utilizando Qiagen Multiplex PCR kit® (Qiagen) y un Termociclador modelo 9700 Geneamp® (Applied Biosystem). La reacción de la PCR fue llevada a cabo en un volumen final de 10μL conteniendo: 10 ng de ADN genómico, 0.2 mM dNTP (Quiagen), 1x PCR buffer (20mM Tris HCl pH 8.4, 50 mM KCl) (Qiagen) y 0.2 μM de cada cebador (Invitrogen), Hot Star taq DNA polimerasa. Los cebadores de avance fueron marcados en el extremo 5' con fluoroforos (6FAM®, HEX® y NED®). El perfil del ciclado termal incluyó una denaturación inicial a 94°C, por 4 minutos, 30 ciclos compuesto de 3 fases (a 94°C 30 segundos, a 65°C 45 segundos y a 72°C 70 segundos) y un ciclo final de 72°C por 4 minutos. Los productos de la PCR fueron diluidos en formamida GENESCAN standard 500 ROX®

(1:10), denaturados por 3 minutos a 95 °C y separados por electroforesis capilar en un analizador genético ABI 3130 DNA sequencer® (Applied Biosystem). El tamaño de los alelos fueron calculados con el programa Genemapper v4.0™ utilizando métodos de determinación de tamaños de alelos de mínimos cuadrados de tercer orden.

4.7. CARACTERIZACIÓN DE LA DIVERSIDAD GENÉTICA DEL PLANTEL DE ALPACAS HUACAYA DEL CENTRO DE INVESTIGACIÓN LACHOCC-UNH

Se analizaron un total de 96 alpacas reproductoras utilizando 10 marcadores microsatélites *YWLL08*, *YWLL29*, *YWLL36*, *YWLL40*, *YWLL44* (Lang *et al.*, 1996) *LCA05*, *LCA08*, *LCA19*, *LCA37*, *LCA66* (Penedo *et al.*, 1999)) previamente validados para pruebas de determinación de parentesco (Agapito *et al.*, 2008).

4.7.1. Análisis de variabilidad genética

El número de alelos por locus, heterocigosidad observada y esperada (Nei, 1973), para la totalidad de alpacas genotipificadas (N=96) fueron determinadas utilizando el programa Cervus v3.0 (Kalinowski *et al.*, 2007).

Las desviaciones del equilibrio de Hardy-Weinberg (deficiencia o exceso de heterocigotos) para cada loci y población en estudio fueron estimados mediante la prueba U (Score test) y el método de Fisher, utilizando el algoritmo de Cadena de Markov (MC) (Dememorization: 10 000, batches: 100, itineration per batch: 10 000) para estimar el valor exacto de probabilidad, disponible en el programa Genepop v4.0 (Raymond y Rousset, 1995).

El grado de diferenciación genética se determinó mediante los estadísticos F (F_{IS}) (Weir y Cockerham, 1984). Se determinó el valor del coeficiente de consanguinidad F_{IS} e índice de fijación R_{hoIS} con la totalidad de los individuos utilizando el programa Genepop v4.0 (Raymond y Rousset, 1995)

4.7.2. Análisis de estructura

La presencia de estructuración poblacional fue determinada a partir de los genotipos multilocus (10 microsatelites) mediante el programa STRUCTURE 2.3 (Pritchard *et al.* 2000) bajo los siguientes parámetros: Burn in period de 1×10^6 , correspondiente al número de muestras aleatorias del espacio de parámetros en los cuales se debe alcanzar la convergencia hacia una zona de credibilidad posterior, numero de cadenas de Markov-Montecarlo (MCMC) posterior al burn in period de 5×10^6 , modelo de admixture y de frecuencias alélicas correlacionadas.

4.8. IDENTIFICACIÓN DE MARCADORES GENÉTICOS TIPO SNP EN GENES CANDIDATOS QUE REGULAN EL DESARROLLO DE LA FIBRA O CODIFICAN PROTEÍNAS ESTRUCTURALES DE LA FIBRA DE ALPACA

4.8.1. Selección de genes candidatos

A partir de un total de 22 genes relacionados a la generación de fibra en rumiantes y con el soporte de estudios previos se eligieron dos genes que codifican proteínas regulatorias (*EDAR* y *HOXC13*) y un gen que codifica una proteína estructural (*KRT31*). La ubicación de los genes fue obtenida de la anotación disponible de los tres genomas de alpaca disponibles en la Unidad de Biotecnología Molecular (UBM) de la Universidad Peruana Cayetano Heredia (UPCH). (<https://www.ncbi.nlm.nih.gov/bioproject/340289>). El genoma de referencia y su anotación fueron obtenidos de la base de datos del NCBI el mismo que cuenta con la accesión: vicugna-pacos 2.0.2 (Vicugna pacos, 2017) (Figura 3).

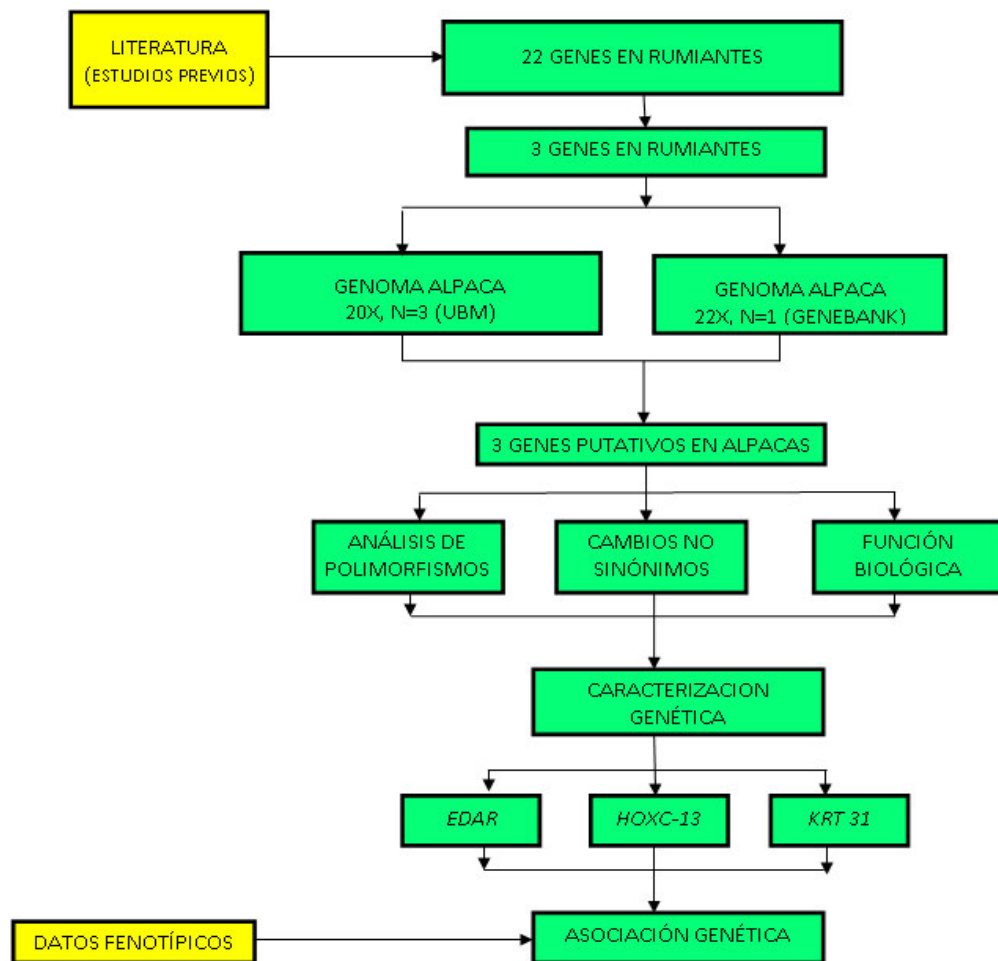


Figura 3. Flujograma para elección de genes candidatos

4.8.2. Identificación de polimorfismos y variantes codificantes mediante análisis de tres genomas en los tres genes candidatos(*in silico*)

Los polimorfismos genéticos fueron identificados mediante análisis bioinformático del genoma de 3 alpacas disponibles en la Unidad de Biotecnología Molecular de la Universidad Peruana Cayetano Heredia. Estos genomas fueron alineados al genoma de referencia versión VicPac 2.0.2 empleando el programa BWA (Burrows Wheeler Aligner) (Li, 2009).

Cabe mencionar que el análisis bioinformático del presente estudio fue realizado con la colaboración del Mag. Diego Veliz, (2017) y el Dr. Jorge Rodríguez, (2015)

A partir de la información de los 3 genomas de alpaca se seleccionaron las variantes presentes en los genes *HOXC13*,

EDAR y *KRT31* haciendo un total de 14 potenciales variantes genéticas.

Se seleccionaron solamente las variantes con un índice de calidad Phred > 30; es decir, las variantes cuyos genotipos tuvieron una probabilidad mayor a 0.999 de ser correctos. La totalidad de variantes encontradas en regiones no codificantes y codificantes (sinónimas y no sinónimas) de los genes *EDAR*, *HOXC13* y *KRT31* se describen en las tablas 1, 2, 3 y 4

a. Predicción de los efectos de las variantes genéticas

Determina qué variantes genéticas podrían tener un mayor impacto sobre la función de las proteínas que las contienen. Para ello, se predijeron los cambios a nivel de secuencia de aminoácidos, de sitios de splicing y en secuencias cercanas a los sitios de inicio de la transcripción (TSS, *Transcription start site*) ya que podrían afectar la transcripción de los genes contiguos. Se empleó el paquete VariantAnnotation (Obenchain *et al.*, 2014) de la plataforma Bioconductor (Gentleman *et al.*, 2004) para seleccionar las variantes en sitios codificantes y las cercanas a los TSS y sitios de splicing.

Tabla 1. Variantes en regiones no codificantes del gen candidato

EDAR

Gen	Scaffold	Inicio	Fin	Transcrito	Posición*	Alelo de referencia	Alelo alternativo
<i>EDAR</i>	Scaffold212	2305064	2305064	XM_015248843.1	+98	G	C
		2305084	2305084		+78	G	A
		2305099	2305099		+63	T	C

*Posición respecto al sitio de inicio de transcripción.

Tabla 2. Variantes sinónimas del gen candidato *EDAR*

Gen	Transcrito	Hebra	Ubicación (CDS)*	Ubicación (proteína)	Alelo Referencia	Alelo Alterno	Codón Referencia	Codón Alterno	AA
<i>EDAR</i>	XM_015248843.1	-	39	13	C	A	CTG	CTT	L
		-	537	179	C	T	TCG	TCA	S
		-	660	220	A	G	GCT	GCC	A

*Posición respecto al sitio de inicio de transcripción

Tabla 3. Variantes no sinónimas de los genes candidatos *EDAR* y *HOXC13*

Gen	Transcrito	Hebra*	Ubicación (CDS)**	Ubicación (proteína)	Alelo de Referencia	Alelo Alterno	Codón de Referencia	Codón Alterno	AA de Referencia	AA Alterno
<i>EDAR</i>	XM_015248843.1	-	866	289	C	T	CGG	CAG	R	Q
		-	743	248	G	A	GCG	GTG	A	V
		-	1208	403	A	G	ATC	ACC	I	T
		-	755	252	C	G	CGA	CCA	R	P
		-	694	232	A	G	TGG	CGG	W	R
<i>HOXC13</i>	XM_015238026.1	+	8	3	T	C	CTA	CCA	L	P

*La hebra especificada hace referencia a la secuencia del scaffold en el genoma de referencia

**CDS: Secuencia codificante

Tabla 4. Variantes sinónimas del gen *KRT31*

Gen	Scaffold	Inicio	Fin	Transcrito	Posicion	Alelo de referencia	Alelo alternativo
<i>KRT31</i>	Scaffold	10534	10534	XM0152520059.1	+237	G	A
		10572	10572	XM0152520059.1	+275	C	T

4.8.3. Diseño de cebadores para amplificar los genes candidatos

Los cebadores fueron diseñados mediante la herramienta primer-BLAST del NCBI (Ye *et al.*, 2012), que incorpora el algoritmo del software PRIMER3 (Untergasser *et al.*, 2012), y posee la ventaja de analizar la especificidad de los cebadores diseñados mediante una PCR *in silico* y contrasta contra el genoma de referencia de la especie de interés para predecir si los cebadores podrían amplificar regiones inespecíficas del genoma. Se determinó que los cebadores debían cumplir los siguientes criterios: 18 a 24 nucleótidos de longitud, Tm de 50 a 60 grados centígrados, tamaños de producto de la PCR entre 400 y 900 pares de bases.

4.8.4. Estandarización de las reacciones de amplificación de los tres genes candidatos

Las reacciones de la PCR fueron estandarizadas realizando una curva de temperaturas de alineamiento que incluían 5 grados por debajo y 5 grados por encima de la temperatura de *melting* para cada par de cebadores. Se realizó una curva de concentraciones

de $MgCl_2$ (1.5 mM, 2.0 mM, 2.5 mM) para obtener las concentraciones óptimas de amplificación. Los productos de la PCR fueron evaluados mediante visualización en geles de agarosa al 1%.

4.8.5. Secuenciamiento mediante método Sanger

Una vez que se confirmó la amplificación mediante electroforesis en agarosa, las muestras fueron enviadas a la empresa MacroGen (Corea) para ser secuenciadas mediante el método Sanger (Dideoxy nucleotide chain termination) en ambas direcciones utilizando un analizador genético 3730XL (Applied Biosystems). Los archivos AB1 obtenidos fueron analizados con el software CodonCode. Primero se cortaron los extremos de las secuencias para maximizar las regiones con una probabilidad de error menor a 1%. Las secuencias recortadas con menos de 90% de bases con un Phred score < Q30 fueron descartadas por ser consideradas de baja calidad. Finalmente, se determinaron los genotipos analizando alineando y editando con el programa Codon Code Aligner.

4.8.6. Caracterización de marcadores SNP en los 3 genes candidatos en alpacas de Lachocc - UNH

a. Cálculo del número muestral

Se utilizó el programa MINSAGE (Gregorius, 1980) para el cálculo del mínimo número de animales que aseguren un 99% de confianza en la detección de un alelo con una frecuencia del 5% para marcadores bialélicos asumiendo desequilibrio de Hardy-Weinberg, ya que se consideró posible que la población manifieste consanguinidad o que los genes candidatos se encuentren bajo presión de selección. El número calculado fue de 90 animales por lo que se consideró un total de 96 animales para el análisis.

b. Indicadores de variabilidad genética en genes *EDAR*, *HOXC13* y *KRT31*

Los valores de heterocigosidad observada y esperada fueron calculados con el software Cervus (Kalinowski *et al.*, 2007). La prueba de hipótesis de equilibrio de Hardy-Weinberg se realizó con el programa Genepop 4.6 mediante el método de Rousset y Raymond (1995) que plantea como hipótesis alternativa que existe deficiencia de heterocigotos (Rousset, y Raymond, 1995). Se evaluó la presencia de consanguinidad por estado mediante el cálculo del indicador F_{IS} (Weir, y Cockerham, 1984)

4.9. DETERMINACIÓN DE LA ASOCIACIÓN GENÉTICA ENTRE MARCADORES TIPO SNP Y DIÁMETRO DE FIBRA EN ALPACAS.

4.9.1. Análisis de la potencia estadística

La potencia estadística es la probabilidad de que la hipótesis nula sea rechazada cuando la hipótesis alterna es verdadera. En consecuencia a medida que aumenta la potencia la posibilidad de que se presente un error del tipo II disminuye (Sham y Purcell, 2014).

Para este fin se realizó una simulación en R para calcular la potencia estadística que pueda detectar variantes genéticas asociadas al diámetro de la fibra de alpaca.

Esta simulación se hizo con datos de diámetro de fibra para 96 individuos, empleando una distribución normal con los mismos valores de media, varianza y desviación estándar (3.7 μm) observados en la población real, asumiendo que los SNP estuviera en desequilibrio de Hardy-Weinberg debido a la endogamia con diferentes frecuencias alélicas y diferentes valores genotípicos medios que influirían sobre el fenotipo.

La potencia estadística se calculó como la proporción de las veces en que la hipótesis nula fue rechazada con un nivel de significancia de 0.05.

4.9.2. Análisis de asociación entre genes y diámetro de fibra

El análisis de asociación genética se realizó utilizando una regresión lineal múltiple mediante el modelo aditivo incluyendo al sexo como covariable.

a. Descripción del modelo estadístico

$$Y = \beta_0 + \beta_1 X + \beta_2 S + e$$

Dónde:

- **Y:** Diámetro de la fibra.
- **X:** Variable que codifica el genotipo de tal modo que X toma valores de 0, 1 y 2 para los genotipos A_2A_2 , A_1A_2 y A_1A_1 ,
- **β_0 :** Intercepto de la regresión.
- **β_1 :** Coeficiente de regresión del genotipo. Es el efecto promedio de la sustitución alélica, representado por la letra " α " en la literatura de genética cuantitativa (Falconer, 1995).
- **β_2 :** Coeficiente de regresión de la variable sexo.
- **S:** Variable que representa el sexo, codificado como 1 para machos y 0 para hembras.
- **e:** Residual aleatorio. Se asume que sigue una distribución normal con media cero

Debido al gran número de pruebas de hipótesis realizadas (una por cada SNP por cada modelo por cada fenotipo examinado), se controló la probabilidad de errores tipo I mediante el método de corrección para pruebas de hipótesis múltiples de Benjamini-Hochberg (Noble, 2009). Los reportes estadísticos y el análisis de asociación se realizaron con el software estadístico R

V. RESULTADOS

5.1. CARACTERIZACIÓN DE LA DIVERSIDAD GENÉTICA DE UNA POBLACIÓN DE 96 ALPACAS HUACAYA DEL CENTRO DE INVESTIGACIÓN LACHOCC-UNH EMPLEANDO 10 MARCADORES MICROSATÉLITES

5.1.1. Análisis de variabilidad genética

Un total de 128 alelos microsateélites fueron identificados en 96 alpacas genotipificadas, con un número de alelos que varió entre 6 alelos para el locus *YWLL40* y 26 alelos para el locus *YWLL08* con un promedio de alelos por locus de 12.8.

Los indicadores de variabilidad genética de la población de Lachocc reportan una moderada a elevada variabilidad en todos los locus analizados con valores promedio de H_E de 0.8235 y H_o de 0.8044. La totalidad de loci a excepción de los loci *LCA5* y *LCA66* ($p < 0.05$) se encontraron en equilibrio de Hardy Weinberg. El locus *LCA5* y *LCA66* se encontraron en desequilibrio de Hardy Weinberg debido a deficiencia de heterocigotos ($p < 0.05$).

El análisis de consanguinidad por estado calculada mediante el indicador F_{is} y R_{hois} reporta un valor para la población de 0.023 y 0.021 respectivamente, indicando la presencia de bajos niveles de consanguinidad en la población evaluada, sin embargo algunos loci como el *LCA5* y *YWLL8* presentaron niveles de consanguinidad más elevados ($F_{is} = 0.072$ y 0.078).

Los principales indicadores de consanguinidad se detallan en la tabla 5.

Tabla 5. Número de alelos (A), heterocigosidad observada (H_o), heterocigosidad esperada (H_E), p-valué (valor de probabilidad) para HWE, F_{IS} (coeficiente de endogamia) y Rho_{IS} para 10 microsatélites en la población de alpacas de Lachocc (N=96).

Locus	A	H _o	H _E	HWE (p-value)	F _{IS}	Rho _{IS}
	10	0.708	0.762	0.027	0.072	-0.002
LCA08	9	0.820	0.826	0.572	0.007	-0.050
LCA19	13	0.742	0.764	0.179	0.030	0.093
LCA37	16	0.809	0.864	0.136	0.064	-0.095
LCA66	14	0.831	0.877	0.030	0.052	0.090
YWLL08	26	0.865	0.938	0.078	0.078	0.066
YWLL29	8	0.764	0.743	0.794	-0.028	-0.018
YWLL36	12	0.921	0.887	0.885	-0.038	-0.219
YWLL40	6	0.719	0.687	0.716	-0.046	-0.009
YWLL44	14	0.865	0.887	0.271	0.025	0.119
Global	12.8	0.804	0.823		0.023	0.021

5.1.2. Análisis de estructura genética

La población en estudio no mostró estructuración genética ($F_{ST} < 0.05$, $p\text{-valué} > 0.05$) valor que fue obtenido con el uso del programa STRUCTURE 2.3 simulando diferentes valores de K (Pritchard *et al.*, 2000).

Por su parte, el análisis de agrupamiento bayesiano también no indica la presencia de estructuración poblacional tal como se muestra en el gráfico de análisis de estructura genética para la hipótesis de 2 grupos ($k=2$) hasta 5 grupos ($k=5$), demostrando la ausencia de estructuración poblacional en la muestra de alpacas del Centro de Investigación Lachocc. (Figura 4-7).

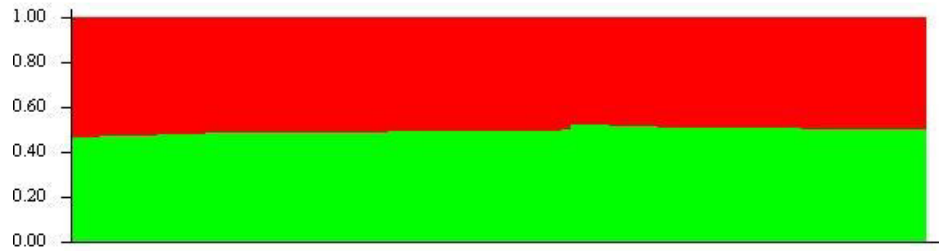


Figura 4 (K = 2)

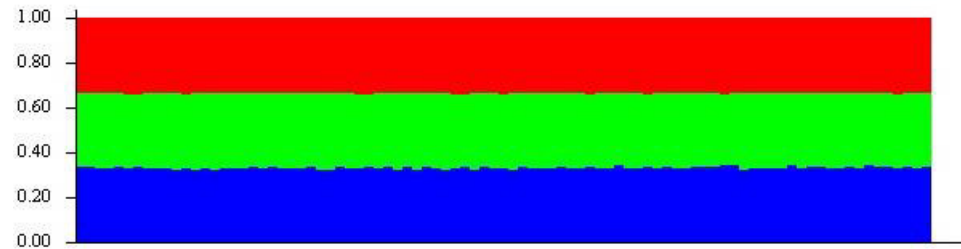


Figura 5 (K = 3)



Figura 6 (K = 4)

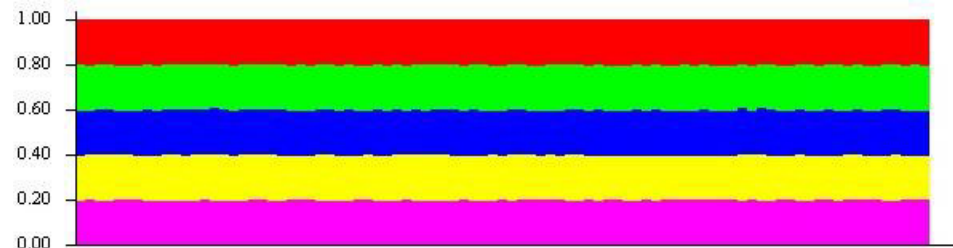


Figura 7 (K = 5)

Figura 4-7. Resultados gráficos del análisis de estructura obtenidos por el programa Structure v2.3. En las figuras 4, 5, 6 y 7 se muestran los resultados para K=2, K=3, K=4 y K=5 subpoblaciones respectivamente. Cada columna representa un individuo (la separación entre columnas no es visible). El eje "Y" representa la probabilidad de pertenecer a cada subpoblación.

5.2. IDENTIFICACIÓN DE MARCADORES GENÉTICOS TIPO SNP EN 3 GENES CANDIDATOS, 2 QUE REGULAN EL DESARROLLO DE LA FIBRA (HOXC13 Y EDAR) Y UNO QUE CODIFICA UNA PROTEÍNA ESTRUCTURAL DE LA FIBRA DE ALPACA (KRT31)

5.2.1. Selección de genes candidatos

Los genes candidatos seleccionados fueron el *HOXC13* y el *EDAR*, ambos de orden funcional en la generación de la fibra y el gen estructural *KRT31* encargado de la producción de proteínas para constitución de fibra.

Mediante análisis bioinformático los genes *EDAR*, *HOXC13* y *KRT31* fueron ubicados en los scaffolds y transcritos disponibles en la bases de datos del Genbank (Vicugna-pacos 2.0.2). Cada gen presentó solo un transcrito putativo (Tabla 6).

Tabla 6. Ubicación de los transcritos codificados por los genes regulatorios *EDAR* y *HOXC13* y el estructural *KRT31* en el genoma de alpaca.

Gen	Scaffold (N° de accesión)	Transcrito (N° de accesión)	Inicio	Fin	Tamaño (nt)	Hebra
<i>EDAR</i>	212	XM-015248843.1	2246405	2305162	58758	-
<i>HOXC13</i>	36	XM-015238026.1	11647938	11664806	16869	+
<i>KRT31</i>	450	XM-015252059.1	10297	54385	44089	-

5.2.2. Elección de los cebadores para los tres genes candidatos

De un total de 14 cebadores candidatos (anexo 3) luego de probar en una muestra de 10 alpacas Huacaya, se eligieron 6 cebadores para amplificación de los genes *EDAR*, *HOXC13* y *KRT31* con sus respectivos sitios de anclaje y longitud (Tabla 7).

Tabla 7. Características de los cebadores utilizados para la amplificación de los genes *EDAR*, *HOXC13* y *KRT31*

Gen	Cebador	Sentido	Secuencia (5'→3')	Longitud cebador	Tm	GC%	Tamaño amplicon	Inicio y fin
<i>EDAR</i>	<i>EDAR3</i>	Forward	AGCTCCCTGTCCTGCTTG	18	58	61	920	2251903-2251920
	<i>EDAR3</i>	Reverse	CCTCCTCTCCACACCTACCT	20	59	60		2252803-2252822
<i>HOXC13</i>	<i>HOXC4</i>	Forward	GGGGCGGGACAGAGAAAAG	20	61	60	581	11647632-11647651
	<i>HOXC4</i>	Reverse	GAGAACCTTGCAGCTGTGAC	20	59	55		11648212-11648193
<i>KRT31</i>	<i>KRT31</i>	Forward	GGAGAGCGAGGACTGC AA	18	59	61	281	10297-10315
	<i>KRT31</i>	Reverse	TC AAGATGGGG AATTCCAGTGG A	23	59	45		10872-10894

5.2.3. Secuenciamiento e identificación de SNP en los genes *EDAR*, *HOXC13* y *KRT31*

A partir del secuenciamiento, edición, y alineamiento se identificaron un total de 34 SNP, de las cuales 21 (61.7%) pertenecieron al gen *EDAR*, 5 (14.8%) al gen *HOXC13* y 8 (23.5%) al gen *KRT31*, los SNP identificados se describen en las tablas 8, 9 y 10.

Cabe mencionar que se maximizó la recuperación de fragmentos de ADN con probabilidades de error menores del 1% y solo se consideró secuencias con un valor de Phred score > Q30).

Tabla 8. Frecuencias alélicas de los sitios variables presentes en el gen *EDAR* (N=86)

	*Coordenada	Alelo mayor	Alelo menor	**MAF
1	2252010	C	T	0.227
2	2252111	A	T	0.087
3	2252123	C	T	0.186
4	2252227	T	A	0.122
5	2252231	G	A	0.227
6	2252234	G	A	0.052
7	2252304	C	T	0.023
8	2252343	G	A	0.041
9	2252364	G	C	0.18
10	2252376	G	A	0.134

11	2252425	G	A	0.308
12	2252434	G	A	0.047
13	2252459	G	A	0.163
14	2252558	G	A	0.017
15	2252582	C	T	0.157
16	2252613	A	C	0.145
17	2252618	C	T	0.221
18	2252646	C	T	0.041
19	2252745	C	T	0.122
20	2252747	G	T	0.006
21	2252791	G	A	0.081

*Coordenadas en el scaffold 212

** Frecuencia del alelo menor

Tabla 9. Frecuencias alélicas de los sitios variables presentes en el gen *HOXC13* (N=91)

	*Coordenada	Alelo mayor	Alelo menor	**MAF
1	11647923	A	T	0.236
2	11647945	C	T	0.385
3	11648009	C	T	0.006
4	11648027	A	T	0.029
5	11648042	G	C	0.017

*Coordenadas en el scaffold 36

** Frecuencia del alelo menor

Tabla 10. Frecuencias alélicas de los sitios variables presentes en el gen *KRT31* (N=92)

	*Coordenada	Alelo mayor	Alelo menor	**MAF
1	10534	G	A	0.072
2	10572	C	T	0.106
3	10575	G	A	0.044
4	10628	G	A	0.217
5	10657	G	A	0.317
6	10963	T	C	0.315
7	10995	A	G	0.292
8	11007	T	C	0.32

*Coordenadas en el scaffold 450 (NW 00588315.1)

** Frecuencia del alelo menor

5.2.4. Análisis de variabilidad genética en los genes *EDAR*, *HOXC13* y *KRT31*

Un total de 68 alelos distribuidos en 34 variantes tipo SNP fueron identificados en los 3 genes analizados, de los cuales 21 variantes (61.7%) pertenecieron al gen *EDAR*, 5 variantes (14.8%) al gen *HOXC13* y 8 variantes (23.5%) al gen *KRT31*. Los indicadores de informatividad para marcadores SNP revelan una moderada informatividad de los 34 SNP identificados.

La totalidad de loci en los genes *EDAR*, *HOXC13* y *KRT31* se encontraron en equilibrio de Hardy Weinberg ($p > 0.01$).

Los SNP de los genes *KRT31* y *EDAR* presentaron valores de F_{IS} cercanos a 0, lo que indica que no hay deficiencia de heterocigotos con respecto a lo esperado en una población con reproducción aleatoria y sin estructuración poblacional. Estos resultados concuerdan con lo observado en el valor agregado de los 10 microsatélites que mostró un valor de F_{IS} de 0.02 aproximadamente, en contraste con lo observado en el gen *HOXC13* ($F_{IS} = 0.108$).

La heterocigosidad observada (H_o), heterocigosidad esperada (H_e), Equilibrio de Hardy Weinberg y F_{IS} para los genes *EDAR*, *HOXC13* y *KRT31* se detallan en las tablas 11, 12 y 13 respectivamente.

Tabla 11. Valores de heterocigosidad observada (H_o), heterocigosidad esperada (H_e), Equilibrio de Hardy-Weinberg y F_{IS} para los SNP del gen *EDAR*

	*Posición	Heterocigosidad Observada	Heterocigosidad Esperada	p- valué (HWE)	F_{IS}
1	2252010	0.36	0.353	1	-0.022
2	2252111	0.174	0.16	1	-0.09
3	2252123	0.256	0.305	0.15	0.161
4	2252227	0.244	0.216	0.35	-0.133
5	2252231	0.291	0.353	0.12	0.177
6	2252234	0.105	0.1	1	-0.049
7	2252304	0.047	0.046	1	-0.018
8	2252343	0.081	0.079	1	-0.037
9	2252364	0.267	0.297	0.45	0.101
10	2252376	0.267	0.233	0.34	-0.149

11	2252425	0.407	0.429	0.62	0.051
12	2252434	0.093	0.089	1	-0.043
13	2252459	0.209	0.274	0.04	0.238
14	2252504	0.081	0.079	1	-0.037
15	2252558	0.035	0.034	1	-0.012
16	2252582	0.291	0.266	0.68	-0.093
17	2252613	0.244	0.25	1	0.023
18	2252618	0.419	0.346	0.06	-0.21
19	2252646	0.081	0.079	1	-0.037
20	2252745	0.244	0.216	0.35	-0.133
21	2252791	0.163	0.15	1	-0.083
	Fis global				-0.002

*coordenadas del SNP respectivo en el scaffold 212

Tabla 12. Valores de heterocigosidad observada (H_o), heterocigosidad esperada (H_E), Equilibrio de Hardy-Weinberg y F_{IS} para el gen *HOXC13*

	*Posición	Heterocigosidad Observada	Heterocigosidad Esperada	p- valué (HWE)	F_{IS}
1	11647923	0.341	0.363	0.5673	0.062
2	11647945	0.418	0.476	0.2714	0.123
3	11648009	0.011	0.011	No	NA
4	11648027	0.034	0.056	0.0573	0.387
5	11648042	0.034	0.034	1	-0.012
	FIS global				0.108

*Coordenadas del SNP respectivo en el scaffold 36

Tabla 13. Valores de heterocigosidad observada (H_o), heterocigosidad esperada (H_E), Equilibrio de Hardy-Weinberg y F_{IS} para el gen *KRT31*

	*Posición	Heterocigosidad Observada	Heterocigosidad Esperada	p- valué (HWE)	F_{IS}
1	10534	0.144	0.135	1	-0.072
2	10572	0.167	0.19	0.243	0.123
3	10575	0.067	0.085	0.15	0.22
4	10628	0.344	0.341	1	-0.009
5	10657	0.433	0.435	1	0.004
6	10963	0.427	0.434	1	0.016
7	10995	0.449	0.416	0.607	-0.081
8	11007	0.348	0.438	0.086	0.205
	Fis global				0.038

*Coordenadas del SNP respectivo en el scaffold 450

5.3. EVALUACIÓN DE LA ASOCIACIÓN GENÉTICA ENTRE LOS POLIMORFISMOS GENÉTICOS IDENTIFICADOS EN LOS GENES CANDIDATOS Y EL CARÁCTER DIÁMETRO DE LA FIBRA DE ALPACA.

5.3.1. Análisis de la potencia estadística

La potencia estadística estimada fue de 80% asumiendo que el locus candidato tiene efectos puramente aditivos sobre el fenotipo, tal como se muestra en la figura 8.

En consecuencia se determinó que utilizando una regresión lineal simple es posible detectar un efecto del gen de 5% con una potencia estadística de 0.8. Esta potencia se alcanza siempre que se utilice un marcador donde el alelo menos representado tenga una frecuencia igual o mayor a 0.5.

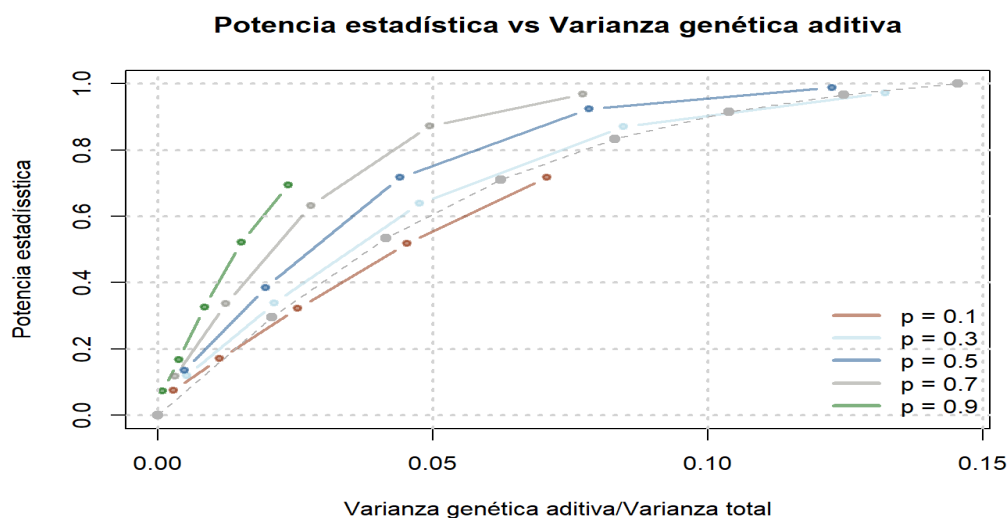


Figura 8. Potencia estadística del análisis de asociación genética bajo el modelo aditivo.

5.3.2. Análisis de asociación entre los genes *EDAR*, *HOXC13* y *KRT31* con el diámetro de fibra en alpacas

Se utilizó el modelo aditivo para evaluar la asociación de los 34 SNP de los genes *EDAR*, *HOXC13* y *KRT31*. Los resultados indican la ausencia de asociación significativa ($p > 0.05$) entre los SNP de los genes *EDAR*, *HOXC13* y *KRT31* y la característica

diámetro de fibra de alpaca Huacaya tal como se muestra en las tablas 14 (Gen *EDAR*), 15 (Gen *HOXC13*) y 16 (Gen *KRT31*).

Tabla 14. Análisis de asociación del diámetro de la fibra y SNPs presentes en el gen *EDAR* usando un modelo aditivo

SNP ubicación	Efecto	Valor estimado	Error estándar	p-valué (HWE)	R ²
2252010	intercepto	23.887	0.536	0	0.019
	Efecto aditivo	-0.084	0.693	0.903	0.019
	sexo	-1.714	1.323	0.198	0.019
2252111	intercepto	23.904	0.468	0	0.020
	Efecto aditivo	-0.309	1.066	0.772	0.020
	sexo	-1.725	1.321	0.195	0.020
2252123	intercepto	23.831	0.490	0	0.019
	Efecto aditivo	0.046	0.687	0.946	0.019
	sexo	-1.713	1.327	0.200	0.019
2252227	intercepto	24.474	1.717	0	0.021
	Efecto aditivo	-0.354	0.942	0.707	0.021
	sexo	-1.738	1.322	0.192	0.021
2252231	intercepto	24.697	1.080	0	0.028
	Efecto aditivo	-0.540	0.631	0.394	0.028
	sexo	-1.832	1.322	0.169	0.028
2252234	intercepto	21.205	2.522	0	0.032
	Efecto aditivo	1.393	1.311	0.291	0.032
	sexo	-1.693	1.311	0.200	0.032
2252304	intercepto	23.894	0.438	0	0.022
	Efecto aditivo	-0.894	1.921	0.642	0.022
	sexo	-1.750	1.322	0.189	0.022
2252343	intercepto	29.076	2.815	0	0.059
	Efecto aditivo	-2.720	1.448	0.063	0.059
	sexo	-1.793	1.293	0.169	0.059
2252364	intercepto	24.061	1.263	0	0.020
	Efecto aditivo	-0.128	0.715	0.857	0.020
	sexo	-1.731	1.329	0.196	0.020
2252376	intercepto	22.173	1.660	0	0.032
	Efecto aditivo	0.955	0.915	0.299	0.032
	sexo	-1.514	1.323	0.255	0.032

2252425	intercepto	23.656	0.931	0	0.020
	Efecto aditivo	0.140	0.607	0.818	0.020
	sexo	-1.730	1.324	0.195	0.020
2252434	intercepto	22.168	2.684	0	0.024
	Efecto aditivo	0.879	1.388	0.527	0.024
	sexo	-1.685	1.317	0.204	0.024
2252459	intercepto	23.390	1.270	0	0.021
	Efecto aditivo	0.269	0.703	0.703	0.021
	sexo	-1.634	1.331	0.223	0.021
2252558	intercepto	26.198	4.341	0	0.023
	Efecto aditivo	-1.198	2.203	0.587	0.023
	sexo	-1.656	1.320	0.213	0.023
2252582	intercepto	23.787	0.493	0	0.020
	Efecto aditivo	0.202	0.830	0.808	0.020
	sexo	-1.733	1.325	0.194	0.020
2252613	intercepto	23.409	0.480	0	0.058
	Efecto aditivo	1.466	0.788	0.066	0.058
	sexo	-1.591	1.294	0.222	0.058
2252618	intercepto	23.647	0.541	0	0.023
	Efecto aditivo	0.467	0.778	0.549	0.023
	sexo	-1.763	1.320	0.185	0.023
2252646	intercepto	23.892	0.441	0	0.021
	Efecto aditivo	-0.565	1.477	0.702	0.021
	sexo	-1.684	1.319	0.205	0.021
2252745	intercepto	23.765	0.480	0	0.021
	Efecto aditivo	0.354	0.942	0.707	0.021
	sexo	-1.738	1.322	0.192	0.021
2252747	intercepto	23.640	0.433	0	0.057
	Efecto aditivo	2.657	1.449	0.070	0.057
	sexo	-1.791	1.295	0.170	0.057
2252791	intercepto	25.935	2.038	0	0.032
	Efecto aditivo	-1.14	1.088	0.298	0.032
	sexo	-1.637	1.312	0.215	0.032

Tabla 15. Análisis de asociación del diámetro de la fibra y SNPs presentes en el gen *HOXC13* usando un modelo aditivo

SNP ubicación	Efecto	Valor estimado	Error estándar	p-valué (HWE)	R ²
11647923	intercepto	24.152	0.511	0	0.026
	Efecto aditivo	-0.924	0.635	0.149	0.026
	sexo	-0.583	1.254	0.644	0.026
11647945	intercepto	23.379	0.583	0	0.010
	Efecto aditivo	0.458	0.547	0.404	0.010
	sexo	-0.727	1.272	0.568	0.010
11648009	intercepto	23.738	0.439	0	0.012
	Efecto aditivo	3.361	3.855	0.385	0.012
	sexo	-0.628	1.289	0.627	0.012
11648027	intercepto	23.565	0.424	0	0.096
	Efecto aditivo	4.163	1.416	0.004	0.096
	sexo	-0.871	1.233	0.481	0.096
11648042	intercepto	22.594	4.463	0	0.004
	Efecto aditivo	0.605	2.265	0.789	0.004
	sexo	-0.695	1.296	0.592	0.004

Tabla 16. Análisis de asociación del diámetro de la fibra y SNPs presentes en el gen *KRT31* usando un modelo aditivo

SNP ubicación	Efecto	Valor estimado	Error estándar	p-valué (HWE)	R ²
10534	intercepto	21.954	2.124	0	0.016
	Efecto aditivo	1.006	1.134	0.377	0.016
	sexo	-1.159	1.217	0.343	0.016
10572	intercepto	23.979	0.470	0	0.015
	Efecto aditivo	-0.735	0.871	0.401	0.015
	sexo	-1.170	1.221	0.340	0.015
10575	intercepto	26.457	2.404	0	0.021
	Efecto aditivo	-1.379	1.230	0.265	0.021
	sexo	-1.139	1.201	0.347	0.021
10628	intercepto	23.959	1.165	0	0.007
	Efecto aditivo	-0.098	0.686	0.885	0.007

	sexo	-1.006	1.211	0.408	0.007
10657	intercepto	22.684	0.933	0	0.027
	Efecto aditivo	0.803	0.599	0.183	0.027
	sexo	-0.824	1.202	0.495	0.027
10963	intercepto	22.689	0.939	0	0.025
	Efecto aditivo	0.799	0.602	0.188	0.025
	sexo	-0.758	1.261	0.548	0.025
10995	intercepto	23.629	0.568	0	0.008
	Efecto aditivo	0.298	0.650	0.647	0.008
	sexo	-0.917	1.265	0.470	0.008
11007	intercepto	24.724	0.854	0	0.023
	Efecto aditivo	-0.680	0.549	0.218	0.023
	sexo	-0.881	1.256	0.484	0.023

VI DISCUSIÓN

Uno de los pilares para la mejora de la calidad de los productos de origen animal constituye los programas de mejoramiento genético. En el Perú, durante varias décadas, los sistemas de producción de alpacas vienen desarrollando e implementando diferentes programas de mejora genética para características de importancia económica y textil de la fibra (CONACS, 2003), por lo que es fácil deducir que las poblaciones de alpacas vienen siendo sometidas a fuerte presión de selección para dichas características (Brenes *et al.*, 2001).

Estos procesos de selección y mejora genética probablemente se relacionen con una pérdida drástica de los niveles de variabilidad genética e incremento considerable de los niveles de consanguinidad a nivel genómico afectando tanto a regiones de ADN codificantes como secuencias anónimas (Bertorelle *et al.*, 2009), tal como se ha observado en otras especies de animales domésticos sometidos a fuerte presión de selección para características de orden productivo, siendo necesario evaluar la variación genética inicial en cualquier población a ser considerada como punto de partida en los procesos de selección.

Poblaciones con niveles de variabilidad adecuados permiten llevar a cabo procesos de selección exitosos, constituyendo la variabilidad un potencial determinante para la mejora genética animal (Avisé *et al.*, 2004)

El uso de métodos de selección más eficientes y exactos que incluyen tecnologías de ADN se han venido generando y utilizando en diferentes especies domésticas. La selección asistida por marcadores de ADN, viene siendo utilizada como una herramienta para mejora genética en ganado bovino, ovino e incluso porcino (Marin, 2013)

El análisis de 10 microsatélites en la población de alpacas reproductoras de Lachocc indica valores bajos de F_{IS} (0.023) y R_{hoIS} (0.021) sugiriendo la presencia de adecuada variabilidad genética lo que indica que no hay deficiencia de heterocigotos con respecto a lo esperado bajo reproducción aleatoria en ausencia de estructuración poblacional similar a lo reportado para otras poblaciones de alpacas reproductoras a nivel nacional (Rodríguez *et al.*, 2009).

En un esfuerzo por seleccionar y evaluar genes con potencial importancia en características de utilidad textil, nuestro estudio evaluó un total de 3 genes regulatorios y estructurales asociados a la formación y/o constitución de fibra de alpaca, a partir de 22 genes que regulan el desarrollo del pelo en rumiantes por medio del análisis del genoma de 3 alpacas secuenciadas en un estudio previo y un genoma disponible en bases de datos (Vicugna-pacos 2.0.2). Los genes *HOXC13*, *EDAR* y *KRT31* fueron considerados como los mejores candidatos para un estudio de asociación genética.

El gen *EDAR*, codifica una proteína llamada receptor de “ectodisplasina A” la cual cumple un papel importante en el desarrollo de la fibra o pelo antes del nacimiento del animal, siendo crítico para las interacciones entre dos capas de células embrionarias llamadas ectodermo y mesodermo. La interacción ectodermo y mesodermo es esencial para la formación de varias estructuras embrionarias incluidas la piel, el cabello, las uñas, los dientes y las glándulas sudoríparas. En humanos existe una variante en el gen *EDAR* (370: V>A) asociada a un incremento de 4.2-4.5 micras en el diámetro del pelo (Fujimoto *et al.*, 2008), además existe reportes que indican su intervención en el desarrollo del pelo (Millar, 2002).

El gen *HOXC13* codifica un factor de transcripción que regula clústeres de genes de queratina y proteínas asociadas a queratina (Tkatchenko *et al.*, 2001), tanto en humanos como en ratones. La alteración de los niveles de expresión del *HOXC13* producen alteraciones fenotípicas en el pelaje de ratones, como alopecia y fragilidad del pelo, entre otros (Tkatchenko *et al.*, 2001).

Los genes estructurales como queratinas (*KRT*) y proteínas asociadas a queratinas (*KRTAP*) han sido ampliamente analizados en los estudios de asociación genética en otras especies, pero muy poco se ha investigado sobre estos genes estructurales en alpacas. El gen *KRT31*, en humanos codifica a la proteína queratina tipo 1 cuticular Ha1 y forma parte estructural del citoesqueleto, y tiene actividad como molécula estructural en la formación del pelo (Fopiano, 2016)

Los resultados de variabilidad genética en los 3 genes analizados en la población de alpacas de Lachocc indican que los genes *EDAR* y *KRT31* presentaron valores de F_{IS} muy cercanos a cero, lo que concuerda con lo

observado en el análisis de microsatélites. En contraste, el gen *HOXC13* presentó un F_{IS} elevado (0.108). Este déficit de heterocigotos no puede ser atribuido al efecto Wahlund causado por la estructuración poblacional, ya que los otros dos genes presentaron valores de F_{IS} bajos; además, el análisis combinado de los 10 microsatélites arrojó valores de F_{IS} muy bajos.

El déficit de heterocigotos en *HOXC13* no se debería a fenómenos demográficos (consanguinidad o estructuración poblacional), ya que de ser así, los otros dos genes y los 10 microsatélites también hubieran presentado valores de F_{IS} elevados. Una causa más posible del déficit de heterocigotos en *HOXC13* sería la naturaleza estocástica de la transmisión de gametos. Este fenómeno puede ocasionar que algunas regiones del genoma presenten deficiencia o exceso de heterocigotos por puro azar (Veliz, 2017).

El análisis de la potencia estadística mostró que el tamaño muestral empleado en el estudio de asociación genética permitió alcanzar un poder por encima de 0.8 para detectar SNP que expliquen más del 5% de la varianza fenotípica (figura 8). La potencia estadística alcanzado en este estudio sería suficiente para detectar ese potencial gen con un efecto tan grande o si estuviera en LD con alguno de los SNP evaluado en nuestro análisis de asociación.

La ecuación de regresión empleado en este estudio ha sido utilizado anteriormente para identificar SNP asociados a 8 rasgos cuantitativos en una población humana coreana (Cho *et al.*, 2009), y para identificar marcadores asociados a los niveles de expresión de alrededor de más de 1000 genes en población humana europea quienes utilizaron el software especializado PLINK que está orientado a estudios de asociación genética (Purcell *et al.*, 2007) y ha sido recomendado para estudios de asociación en individuos no emparentados en ausencia de estructuración poblacional (Neale, 2008).

Se han realizado varios estudios que evalúan si variantes en genes estructurales se encuentran asociadas al diámetro de la fibra o a otras características morfológicas de la fibra de alpaca o lana de oveja. Por ejemplo, anteriormente se ha analizado si genes que codifican queratinas y proteínas asociadas a queratina poseen variantes asociadas al diámetro de la lana de oveja. Al respecto McLaren *et al.*, (1997) resalta un estudio que encontró evidencia de ligamiento entre el diámetro de la lana y el gen que codifica la proteína estructural *HGT-KAP6-1* (proteína asociada a queratina 6-1 alta en

glicina y tirosina). Otro estudio encontró ligamiento entre el diámetro de la lana de oveja y la región 200cM del cromosoma 3 de oveja (Bidinost *et al.*, 2008). Los autores de ese trabajo plantearon que el gen *KRT2*, que se localiza en esa zona, sería responsable del ligamiento; sin embargo, esta hipótesis no fue evaluada. Es necesario mencionar que en esta región genómica se han identificado dos genes reguladores asociados al rizado de la lana (Wang *et al.*, 2014), uno de los cuales (*GPRC5A*) regularía la diferenciación epitelial durante la etapa embrionaria (Cheng, y Lotan, 1998). Asimismo se tiene conocimiento de dos trabajos con el mismo objetivo, uno que buscó genes asociados al diámetro de la fibra de alpaca enfocándose en proteínas asociadas a queratina (Fopiano, 2016), y otro que buscó identificar el locus responsable del fenotipo Suri tomando como candidato al gen *TCHH*, el cual codifica la proteína estructural tricohialina que está asociada al rizado del pelo en humanos (Delgado de la Flor, 2014), ambos estudios no encontraron asociación estadísticamente significativa entre los genes y el diámetro de fibra.

Los estudios de asociación de genoma completo (GWAS, *Genome-Wide Association Study*) de genes estructurales es el mejor candidato para estudios de asociación genética. Por ejemplo, en un estudio de asociación de genoma completo (GWAS) para caracteres de lana en ovejas, se identificaron 15 SNP asociados al rizado y 9 SNP asociados al diámetro de la lana, ninguno de estos SNP se ubica en genes que codifican proteínas estructurales (Wang *et al.*, 2014). En contraste, de 10 SNP intragénicos (3 asociados al diámetro y 7 al rizado de la lana) encontrados por Fujimoto *et al.*, 2008, se ubican en genes que codifican proteínas con funciones regulatorias (transporte de bases nitrogenadas y división del ciclo celular)

Además, el alelo T del SNP rs4752566 del gen *FGFR2* (Receptor tipo 2 del factor de crecimiento de fibroblastos) humano está asociado a un incremento de 1.4-3.4 μm en el diámetro del pelo (Fujimoto *et al.*, 2009).

A la fecha hay insuficientes estudios en alpacas de genes estructurales y funcionales siendo este uno de los pocos estudios disponibles para asociación genética con el carácter diámetro de fibra.

En el presente estudio solo una variante rara (11648027, MAF = 0.029) en el gen *HOXC13* mostró una aparente asociación con el diámetro de la fibra de alpaca, sin embargo, este fenómeno se debió a un único outlier, tal como se

observa en el gráfico de cajas del anexo 5, por lo tanto, esta asociación sería espuria. Esta interpretación concuerda con lo observado en los demás polimorfismos de *HOXC13*, los cuales presentaron valores de MAF más elevados, y por lo tanto mayor número de observaciones para la categoría genotípica menos frecuente, lo que permitió una evaluación más precisa de sus posibles efectos sobre el diámetro. Cabe destacar que, si la variante en la posición 11648027 realmente estuviera asociada al diámetro, se esperaría que los SNP cercanos también presenten asociación debido al desequilibrio de ligamiento (Kimura, 1969).

Estos resultados tienen similitud con el trabajo de Fopiano, (2016), que al evaluar al gen *KRTAP11-1* (SNP330 y SNP 956) respecto a la asociación con el diámetro de fibra no encontró ninguna asociación con este carácter.

De la misma forma, Veliz, (2017), en un análisis de asociación de 5 SNP en el gen *BMP4*, 9 SNP en el gen *KAP1-1* y 4 SNP en el gen *KAP13-1* no encontró asociación entre el diámetro de la fibra de alpaca. El SNP 3451703 en el gen *KAP1-1* fue el más próximo a alcanzar la significancia estadística, sin embargo, este efecto se debió a un único valor extremo en uno de los genotipos.

En cuanto al análisis de los genes *HOXC13*, *KRT31* y *EDAR* no mostró ninguna asociación real con el diámetro de la fibra (tabla 14, 15 y 16).

Dado el poder estadístico alcanzado en este estudio, se puede concluir que es muy improbable que los polimorfismos con la mayor contribución sobre el diámetro de la fibra de alpaca se encuentren en la proximidad de alguno de los SNP analizados, estos resultados muestran la dificultad que suelen presentar los estudios de genes candidatos para encontrar asociaciones genotipo-fenotipo.

VII. CONCLUSIÓN

- Se encontraron 34 variantes tipo SNP distribuidos en los genes *EDAR* (21 / 61.7%), *HOXC13* (5 / 14.8%) y *KRT31* (8 / 23.5%).
- No se encontró estructuración poblacional ni endogamia en la muestra de alpacas Huacaya utilizadas en la presente investigación
- No se encontró asociación entre el diámetro de la fibra de alpaca y los fragmentos de los genes *KRT31*, *HOXC13* y *EDAR* analizados.
- Se observó una elevada variabilidad genética ($H_E=0.82$) y bajos niveles de consanguinidad $F_{IS}=0.023$ en población de alpacas de Lachocc

VIII RECOMENDACIONES

1. Consideración otros genes candidatos que tengan asociación con el diámetro de fibra
2. Se recomienda reemplazar los estudios de genes candidatos por estudios de asociación de genoma completo
3. Mejorar el modelo de asociación corrigiendo por edad
4. Usar muestreo estratificado en las alpacas para el carácter diámetro de fibra

VIII. REFERENCIAS BIBLIOGRÁFICAS

- Agapito, J., Rodríguez, J., Herrera-Velit, P., Timoteo, O., Rojas, P. (2008). Parentage testing in alpacas (*Vicugna pacos*) using semi-automated fluorescent multiplex PCRs with 10 microsatellite markers. *Anim. Genet.* 9(2):201-3.
- Agarwal, M., Shrivastava, N., Padh, H. (2008). Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Rep.* 27:617-631.
- Avise, J., (2004). *Molecular markers, natural history, and evolution*. 2da Ed. Sinauer Associates, Inc. Publishers. USA: 161-169.
- Altukhov, Y. (2006). *Intraspecific Genetic Diversity. Monitoring, Conservation and Management*. Springerlink, Alemania: 5-40.
- Bertorelle, G., Bruford, M., Hauffe, H., Rizzoli, A., Vernesi, C. (2009). *Population genetics for animal conservartion*. Cambridge University Press:1-365.
- Bidinost, F., Roldan, D., Doderó, A. (2008). Wool quantitative trait loci in Merino sheep. *Small Rumin Res.* Jan;74(1-3):113-8.
- Braga, W., Leyva, V., Cochran, R. (2007). The effect of altitude on alpaca (*Lama pacos*) fiber production. *Small Rumin Res.* Apr;68(3):323-8
- Brenes, E., Madrigal, F., Pérez & K. Valladares. (2001). *El Cluster de los Camélidos en Perú: Diagnóstico Competitivo y Recomendaciones Estratégicas*. Instituto Centroamericano de Administración de Empresas <http://www.caf.com/attach/4/default/CamelidosPeru.pdf>.
- Brookes, A., (1999). The essence of SNPs. *Gene.* 234(2):177-86.

- Butler, J., (2005). Forensic DNA typing. Biology, technology and genetics of STR markers. 2da. Ed. Elseiver Academic Press. USA: 455-539. Camelid Sci [Internet]. Available from: http://www.academia.edu/4059361/South_American_Camelids_-_Past_Present_and_Future

- Cervantes, I., Pérez-Cabal, M., Morante, R. (2010). Genetic parameters and relationships between fibre and type traits in two breeds of Peruvian alpacas. *Small Rumin Res.* Jan;88(1):6-11

- Chauhan, T., Rajiv, K. (2010). Molecular markers and their applications in fisheries and aquaculture. *Advances in Bioscience and Biotechnology.* 1:281-291.

- Cheng, Y., Lotan, R. (1998). Molecular Cloning and Characterization of a Novel Retinoic Acid-inducible Gene That Encodes a Putative G Protein-coupled Receptor. *J Biol Chem.* Dec 25;273(52)

- Cheung, V., Spielman, R., Ewens, K. (2005). Mapping determinants of human gene expression by regional and genome-wide association. *Nature.* Oct 27;437(7063)

- Chistiakov, D., Hellemans, B., Volckaert, F. (2006). Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. *Aquaculture,* 255:1-29.

- Cho, S., Go, M., Kim, Y., Heo, Y. (2009). A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat Genet.* May;41(5)

- CONACS, (2003). Población de Alpacas por departamentos. Programa de Camélidos domésticos. Lima-Perú.

- CONACS, (2004). Resumen de las actividades realizadas en el programa de repoblamiento de alpacas. Ministerio de Agricultura, Lima, Perú.

- Delgado de la Flor Montauban I., (2014). Caracterización de nuevos marcadores genéticos Microsatélites e identificación de SNP en el gen de Trichialina en alpacas (*Vicugna pacos*); Available from: <http://repositori.concytec.gob.pe/handle/CONCYTEC/74> Ecuador y Perú [Internet]. Available from: http://www.unido.org/fileadmin/import/58563_camelidos_final.pdf

- Eisen, J., (1991). Mechanistic basic for microsatellite instability. In *Microsatellites. Evolution and applications*. Ed. Goldstein, D., Schlötterer, C. Oxford University Press. Oxford: 34-48.

- Ellegren, H. (2000). Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in genetics*, 10(12):552-558

- Excoffier, L., Heckel, G. (2006). Computer programs for population genetics data analysis: a survival guide. *Nature review genetics*, 7:745-758.

- Falconer, D., Mackay, P. (1995). *Introduction to Quantitative Genetics*. 4edit. Longman; 480 p.

- FAO-Organización de las Naciones Unidas para la Agricultura y la Alimentación. (2005). Situación actual de los camélidos sudamericanos en Perú. Proyecto de cooperación técnica en apoyo a la crianza y aprovechamiento de los camélidos sudamericanos en la región andina

- Fernández, D., (2015). Búsqueda de genes relacionados a la síntesis de la fibra y marcadores SSR en los ESTs de piel de alpaca *Vicugna pacos*. Tesis Universidad Nacional mayor de San Marcos. Lima

- Foppiano, F., (2016). Caracterización de marcadores genéticos en genes que codifican a proteínas asociadas a queratina y evaluación de la asociación del gen KRTAP11-1 al diámetro de fibra en alpaca (*Vicugna pacos*) siguiendo una aproximación de gen candidato. Available from: <http://190.116.48.45/handle/upch/127>

- Frank, E., Hick, M., Gauna, D., Lamas, H., Renieri, C. (2006). Phenotypic and genetic description of fibre traits in South American domestic camelids (llamas and alpacas). *Small Rumin Res.* Feb;61(2-3)

- Freeland, J., (2005). *Molecular Ecology*. John Wiley y Sons Inc. USA: 110-111.

- Fujimoto, A., Kimura, R., Ohashi, J. (2008). A scan for genetic determinants of human hair morphology: EDAR is associated with Asian hair thickness. *Hum Mol Genet.* Mar 15;17(6)

- Fujimoto, A., Nishida, N., Kimura, R. (2009). FGFR2 is associated with hair thickness in Asian populations. *J Hum Genet.* Jul 10; 54(8):461-5.

- Gentleman, R., Carey, V., Bates, D. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5(10)

- Gentry, A., Clutton-Brock, J., Groves, C. (2004) The naming of wild animal species and their domestic derivatives. *Journal of Archaeological Science*; 31: 645-651

- Gregorius, H., (1980). The probability of losing an allele when diploid genotypes are sampled. *Biometrics.* Dec;36(4)

- Hamilton, M. (2009). *Population Genetics*. John Wiley and sons, Ltd, Publications: 235-280.

- Hancock, J., (1991). Microsatellites and other simple sequences: genomic context and mutational mechanisms. In: Microsatellites. Evolution and applications. Ed. Goldstein, D., Schlötterer, C. Oxford University Press. Oxford: 1-9. Jul 13]. Available from: http://genome.sph.umich.edu/wiki/Variant_Normalization

- Kadwell, M., Fernandez, M., Stanley, F., Wheeler, J., Rosadio, R. (2001). Genetic analysis reveals the wild ancestors of the llama and the alpaca. *Soc. Biol Sci.* Dec 22;268

- Kalinowski, S., Taper, ML., Marshall, T. (2007). Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment. *Mol Ecol.* Mar 1;16(5)

- Kijas, J., Lenstra, J., Hayes, B. (2012). Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. *PLOS Biol.* Feb 7;10(2)

- Kimura, M. (1969). The Number of Heterozygous Nucleotide Sites Maintained in a Finite Population Due to Steady Flux of Mutations. *Genetics.* Apr; 61(4):893–903.

- Lang, M., Wang, Y., (1996). Fifteen polymorphic dinucleotide microsatellites in llamas and alpacas. *Anim Genet.* Aug 1: 27(4)

- Liu, F., Tian, C., Zhang, P., Huang, H. (2007). Candidate gene analysis of high quality merino sheep. *Yi Chuan.*29 (1)

- Liu, F. (2014). Polymorphisms of KAP6, KAP7, and KAP8 genes in four Chinese sheep breeds. *Genet Mol Res.*

- MARÍN, P., Cardona-Cadavid, H. (2013). Genómica en la producción animal. *Revista Colombiana De Ciencia Animal - RECIA*, 5(2), 497-518. <https://doi.org/10.24188/recia.v5.n2.2013.460>

- Maynard, J., (1999). Evolutionary Genetics. 2ed. Oxford University Press: 139-160.
- McLaren, R., Rogers, G., Davies, K. (1997). Linkage mapping of wool keratin and keratin-associated protein genes in sheep. Mamm Genome. Dec;8(12):938–40. mensalis) fleece; Available from: http://www.academia.edu/11839660/Variation_of_commercially_important_characteristics_among_sampling_sites_for_vicu%C3%B1a_Vicugna_vicugna_mensalis_fleeces
- Millar, S., (2002). Molecular mechanisms regulating hair follicle development. J Invest Dermatol. Feb;118(2)
- Montes, M., Quicaño, I., Quispe, R., Quispe, E., Alfonso, L. (2013). Quality characteristics of Huacaya alpaca fibre produced in the Peruvian Andean Plateau region of Huancavelica. Span J Agric Res. 6(1)
- Moron, B., (2016). Verificación de la genealogía en alpacas (Vicugna pacos) usando un panel de microsatélites <http://repositorio.lamolina.edu.pe/handle/UNALM/2706>
- Neale, B., Ferreira, M., Medland, S. (2008). Statistical Genetics: Gene Mapping through Linkage and Association. 1 edition. New York: Taylor & Francis; 608 p.
- Nei, M. (1972). Genetic distances between populations. American naturalist, 106p.
- Nei, M., Kumar, S. (2000). Molecular Evolution and Phylogenetics. Oxford University Press. USA: 231-275.
- Noble, W., (2009). How does multiple testing corrections work? Nat Biotechnol. Dec; 27(12)

- Obenchain, V., Lawrence, M., Carey, V. (2014). VariantAnnotation: a Bioconductor package for exploration and annotation of genetic variants. *Bioinformatics*. Jul 15;30(14)
- Ordoñez, A., (2017). Caracterización molecular de ovinos (*ovis aries*) criollos de Huancavelica utilizando marcadores microsatélite, Tesis, ing. Zootecnista.
- Paredes, M., Membrillo, A., Azor P. (2013). Genetic and phenotypic variation in five populations of Huacaya Alpacas (*Vicugna pacos*) from Peru. *Small Rumin Res.* Abril;111(1–3)
- Paredes-P, M., Alonso-M, A., (2011). Genetic Parameters and Fixed Effects Estimation for Fibre Traits in Alpaca Huacaya (*Lama pacos*). *J Anim Vet Adv.* Nov 1:10 (11)
- Peraza, P., Rincón, G., Ravagnolo, O. (2013). Desarrollo de un multiplex de microsatélites para diagnóstico de paternidad en ovinos Corriedale del Uruguay *Agrociencia Uruguay* vol.17 no.1 Montevideo jun.
- Pérez-Cabal, M., Cervantes, I., Morante, R., Burgos, A. (2010). Analysis of the existence of major genes affecting alpaca fiber traits. *J Anim Sci.* Dec; 88(12)
- Pritchard, J., Stephens, M., Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*. Jun;155(2):945–59.
- Purcell, S., Neale, B., Todd-Brown, K., Ferreira, M. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet.* Sep;81(3)
- Raymond, M., Rousset, F. (1995). Genepop (Version 1.2): Population Genetics Software for Exact Tests and Ecumenicism. *J Hered.* May 1;86(3)

- Rodríguez, J., Flor, D de la., Barreto, T., Timoteo, O., Herrera-Velit, P. Espinoza, R. (2009). Genetic diversity in alpaca (*Vicugna pacos*) populations using 10 microsatellite markers. In Available from: http://inis.iaea.org/Search/search.aspx?orig_q=RN:41046519
- Rousset, F., Raymond, M. (1995). Testing Heterozygote Excess and Deficiency. *Genetics*. Aug;140(4)
- Schlake, T., (2005). FGF signals specifically regulate the structure of hair shaft medulla via IGF-binding protein 5. *Dev Camb Engl*. Jul;132(13)
- Schlötterer, C., Harr, B. (2001). Microsatellite Instability. *Encyclopedia of life sciences*. Nature Publishing Group:1-4.
- Sham, P., Purcell, S. (2014). Statistical power and significance testing in large-scale genetic studies. *Nat. Rev. Genet*. May;15(5)
- Sobrino, M., Carracedo, A. (2005). SNP in forensic genetics:a review on SNP typing methodologies. *Forensic science international*. 154:181-194
- Strand, M., Prolla, T., Liskay, R. (1993). Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature*, 365:274-276.
- Syvanen, A. (2001). Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature Review Genetics*. 2(12): 930-942.
- Tan, A., Abecasis, G., Kang, H. (2015). Unified representation of genetic variants. *Bioinformatics*. Jul 1;31(13):2202-4.
- Templeton, A., (2006). *Population Genetics and Microevolutionary Theory*. John Wiley y Sons, New Jersey: 82-246

- Tkatchenko, A., Visconti, R., Shang, L. (2001). Overexpression of Hoxc13 in differentiating keratinocytes results in downregulation of a novel hair keratin gene cluster and alopecia. *Dev Camb Engl.* May;128(9):1547–58.
- Toth, G., Gaspari, Z., Jurka, J. (2000). Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Research*, 10:967-981.traits in young Australian alpacas. <http://www.aaabg.org/livestocklibrary/1999/AB99111.pdf>. 13:468–71.
- Untergasser, A., Cutcutache, S., Koressar, T., Ye, J. (2012). Primer -3 new capabilities and interfaces. *Nucleic Acids Res.* Aug; 40(15)
- Valbonesi, A., Cristofanelli, S., Pierdominici, F. (2010). Comparison of Fiber and Cuticular Attributes of Alpaca and Llama Fleeces. *Text Res J.* Mar 1;80(4)
- Veliz, D. (2017). Búsqueda de genes asociados a rasgos seleccionados para la producción de fibra de alpaca. Tesis para optar el grado de maestro en bioquímica y biología molecular. Universidad Peruana Cayetano Heredia, Lima-Peru.
- Vignal, A., Milan, D., SanCristobal, M., & Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics selection evolution GSE*, 34(3).
- Wang,, Z., Zhang, H., Yang, H., Wang, S., Rong, E. (2014). Genome-Wide Association Study for Wool Production Traits in a Chinese Merino Sheep Population. *PLOS ONE.* Sep 30;(9).
- Weir, B., Cockerham, C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38:1358-1370.
- Wheeler, J., Kadwell, M., Fernandez, M., Rosadio, R. (2001). Genetic analysis reveals the wild ancestors of the llama and the alpaca. *Proceeding of the Royal Society B: Biological Sciences* 268 (1485)

- Wright, S., (1969). Evolution and the genetics of populations. Vol. 2. The theory of gene frequencies. University of Chicago Press.
- Yalta, C., (2014). Variabilidad genética poblacional de alpacas Vicugna pacos determinada por marcadores microsatélites en el Centro Piloto Munay Paqocha y del Fundo Itita, Puno-Perú. <http://cybertesis.unmsm.edu.pe/handle/cybertesis/3600>
- Ye, J., Coulouris, G., Zaretskaya, I. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics. Jun 18;13 (1):134.
- Yucra, A., (2017). Determinación de filiación mediante marcadores microsatélite en alpacas Huacaya del anexo Quimsachata INIA – Puno. Tesis, Universidad Nacional del Altiplano, Puno
- Zhou, H., Gong, H., Yan, W., Luo, Y. (2012). Identification and sequence analysis of the keratin-associated protein 24-1 (KAP24-1) gene homologue in sheep. Gene 10:511(1): 62-65.

ANEXOS

ANEXO 1

Ficha de evaluación

<https://www.youtube.com/watch?v=aQkH7b8YM>

UNIVERSIDAD NACIONAL DE HUANCAYELICA - PROYECTO FOCAM
IDENTIFICACION DE MARCADORES GENETICOS MOLECULARES SNP_s PARA CARACTERISTICAS DE
IMPORTANCIA ECONOMICA EN LA ALPACA HUACAYA*

FICHA DE REGISTRO ANIMAL															N°			
DATOS DEL ANIMAL																		
ID											REG				FECHA			
ARETE											MCR				PV (Kg)			
RAZA	HUACAYA		SURI								SEXO	♀	♂	CASTR	ESQUILADO	SI	NO	
											F NAC				EDAD	DL	2D	4D
COLOR	BLAN	NEG	GRIS	CAF/CL	CAF/OSC	LF	OTROS:											
DATOS DE ORIGEN																		
DEPART	PUN	ARE	CUSCO	JUNIN	HYCA	FUNDO												
DISTRITO											PROPE							
LOCAL											POBLAC	♀	♂					
COORD	S	N	E	O														
CARACTERISTICAS DEL VELLON																		
MDF											SD				CV %			
PF %											CF %				MEDUL			
UNIFORM	B	M	A	MANCH	SI	NO	OTRO											
PVES (g)											PVEL(g)				REN %			
GPI(mm)											SFD				Fines?	SRPP		
SANIDAD – REPRODUCCION																		
ENFERMEDADES										REPRODUCCION								
DIARREA	SI	NO	TRAT	SI	NO	ESPECIF				CONDICION								
ENTEROT	SI	NO	TRAT	SI	NO	ESPECIF				ED. EMP (m)								
NEUMON	SI	NO	TRAT	SI	NO	ESPECIF				N° EMP								
SARNAS	SI	NO	TRAT	SI	NO	ESPECIF				TEST (mm)								
FARAS	SI	NO	TRAT	SI	NO	ESPECIF				N° CRIAS								
PROBLEMAS CONGENITOS – GENETICOS										OTROS DATOS								
SARC	SI	NO	1	2				VACUNACION										
CRIFT	SI	NO	1	2				DESPAR - INT										
HERN	SI	NO	INGUI	UMBIL	OTROS				DESPAR - EXT									
PROGN	SI	NO	AVAN	MEDIO	LEVE				ASESORIA VET									
MICROT	SI	NO	AVAN	MEDIO	LEVE	COLECC-MUEST	SANGRE	HECES	PIEL	VELL								
REGISTRO DE CRIAS																		
ID	SEX		RAZA		PROGE	FNAC	COLOR	MDF	MICROCHIP									
	♀	♂	H	S														
	♀	♂	H	S														
	♀	♂	H	S														
	♀	♂	H	S														
	♀	♂	H	S														
	♀	♂	H	S														
	♀	♂	H	S														

ANEXO 2

Registro de características de las alpacas Huacaya de color blanco del presente trabajo de investigación

N°	DATOS DEL ANIMAL			CARACTERÍSTICAS DEL VELLÓN				PUNTAJE
	ARETE	EDAD	PESO VIVO	UNIFORMIDAD	DIÁMETRO	RIZO	DENSIDAD FOLICULAR	
1	6114	BLL 8A	70	ALTA	24	2X	ALTA	90
2	I 10361	BLL 4A	55	ALTA	20	1X	ALTA	90
3	12.01	2D	50	ALTA	20	2X	ALTA	89
4	29Y11	4D	50	ALTA	20	2X	ALTA	89
5	12.072	2D	65	ALTA	18	2X	MEDIA	88
6	043 C	DL Mayor	45	ALTA	22	2X	ALTA	88
7	12.069	DL menor	50	ALTA	20	2X	ALTA	87
8	S/A	4D	50	ALTA	20	2X	ALTA	87
9	02Y11	4D	45	ALTA	20	2X	ALTA	87
10	14.093	DL Menor	25	ALTA	20	2X	MEDIA	86
11	13.004	DL Mayor	40	MEDIA	20	1X	ALTA	86
12	12.067	2D	50	ALTA	24	2X	ALTA	85
13	S/A	4D	65	MEDIA	22	1X	ALTA	85
14	6.001	BLL8A	70	ALTA	26	2X	ALTA	85
15	10.001	BLL 5A	65	ALTA	24	1X	ALTA	85
16	14.145	DL Menor	28	ALTA	22	2X	ALTA	85
17	14.094	DL Menor	28	MEDIA	20	1X	ALTA	85
18	1853	BLL 5A	65	MEDIA	24	1X	ALTA	84
19	I 10356	4D	48	MEDIA	22	2X	ALTA	84
20	5.103	BLL 8A	70	ALTA	26	1X	MEDIA	84
21	13.001	DL Mayor	45	ALTA	22	1X	ALTA	84
22	14.085	DL Menor	30	ALTA	20	1X	ALTA	84
23	14.058	DL Menor	32	ALTA	24	2X	ALTA	84
24	11.051	BLL 5A	60	ALTA	22	2X	MEDIA	83
25	9060	BLL 5A	60	ALTA	24	1X	MEDIA	83
26	11120	4D	65	ALTA	21	2X	MEDIA	83
27	13Y11	BLL	65	MEDIA	24	2X	ALTA	83
28	14.092	DL Menor	25	ALTA	20	2X	MEDIA	83
29	13.111	DL Mayor	45	MEDIA	22	1X	ALTA	83
30	19	DL Menor	38	ALTA	18	1X	MEDIA	82
32	11.158	4D	65	ALTA	26	1X	MEDIA	82
33	9003	BLL 5A	60	ALTA	26	2X	ALTA	82
34	13.038	DL Mayor	45	MEDIA	22	2X	ALTA	82
35	24	2D	55	ALTA	20	1X	MEDIA	81
36	12.179	2D	45	MEDIA	22	2X	MEDIA	81
37	12.167	2D	50	MEDIA	24	1X	ALTA	81
38	12050	2D	45	ALTA	24	2X	MEDIA	81

39	A620	4D	60	ALTA	24	1X	ALTA	81
40	14.016	DL Menor	28	MEDIA	24	1X	ALTA	81
41	14.039	DL menor	28	MEDIA	24	1X	ALTA	81
42	110358	DL Mayor	50	MEDIA	22	1X	MEDIA	80
43	12.116	4D	45	MEDIA	20	1X	MEDIA	80
44	6.071	BLL 7A	70	ALTA	26	1X	ALTA	80
45	11.014	BLL9A	55	MEDIA	26	1X	ALTA	80
46	6.049	DL Menor	70	MEDIA	25	1X	ALTA	80
47	10040	BLL 6A	70	MEDIA	28	1X	ALTA	80
48	12.021	2D	55	MEDIA	22	1X	MEDIA	79
49	13.143	DL Mayor	45	MEDIA	24	1X	ALTA	79
50	13.029	DL Mayor	45	MEDIA	26	1X	ALTA	79
51	14.026	DL menor	30	MEDIA	22	1X	ALTA	79
52	11.083	4D	65	MEDIA	24	1X	MEDIA	78
53	7.039	BLL 5A	65	MEDIA	24	1X	MEDIA	78
54	12.172	2D	60	ALTA	24	1X	MEDIA	78
55	11.112	BLL	60	ALTA	24	1X	ALTA	78
56	6.033	BLL 7A	70	ALTA	26	1X	ALTA	78
57	14.111	DL Menor	28	MEDIA	22	1X	MEDIA	78
58	14.147	DL Menor	22	MEDIA	22	1X	MEDIA	78
59	14.119	DL Menor	28	MEDIA	22	1X	MEDIA	78
60	12.128	4D	45	MEDIA	22	1X	MEDIA	77
61	11.151	4D	55	ALTA	24	1X	MEDIA	77
62	11.65	4D	50	MEDIA	22	1X	MEDIA	77
63	12.183	2D	48	MEDIA	22	1X	MEDIA	77
64	11090	BLL 5A	65	MEDIA	24	1X	MEDIA	77
65	9.113	BLL	70	MEDIA	28	1X	MEDIA	77
66	7.024	BLL 7A	75	ALTA	26	1X	ALTA	77
67	14.103	DL Menor	28	MEDIA	22	1X	MEDIA	77
68	14.135	DL Menor	30	MEDIA	24	2X	MEDIA	77
69	14.001	DL Menor	30	MEDIA	22	1X	MEDIA	77
70	7.091	BLL 6A	70	MEDIA	24	1X	MEDIA	76
71	8.148	BLL 7A	70	MEDIA	24	1X	MEDIA	76
72	5.115	BLL 8A	65	MEDIA	28	1X	ALTA	76
73	110362	BLL	65	MEDIA	24	1X	MEDIA	76
74	11.044 C	4D	60	MEDIA	22	1X	MEDIA	76
75	05Y11	4D	65	MEDIA	24	1X	MEDIA	76
76	5.177	BLL 7A	70	ALTA	26	1X	ALTA	76
77	12.146	2D	60	ALTA	24	1X	ALTA	76
78	13.152	DL Mayor	45	MEDIA	26		MEDIA	76
79	14	DL Menor	28	MEDIA	24	1X	MEDIA	76
80	11.094	4D	60	ALTA	20	1X	MEDIA	75
81	11.065	4D	60	MEDIA	23	1X	MEDIA	75
82	9.053	BLL 6A	70	MEDIA	26	1X	MEDIA	75

83	7016	BLL 6A	70	ALTA	24	1X	ALTA	75
84	13.047	DL Mayor	50	MEDIA	24	1X	BAJA	75
85	90055	BLL 7A	65	MEDIA	26	1X	MEDIA	75
86	10.021	BLL 6A	65	MEDIA	26		MEDIA	75
87	11.129	BLL 5A	70	MEDIA	28	1X	ALTA	75
88	12145	2D	55	MEDIA	24		MEDIA	75
89	11133	BLL	65	ALTA	26	1X	MEDIA	75
90	A520	4D	60	MEDIA	24		MEDIA	75
91	13.108	DL Mayor	45	MEDIA	24		MEDIA	75
92	11.121	4D	60	MEDIA	28	1X	BAJA	74
93	5.084	BLL 7A	65	MEDIA	28		MEDIA	74
94	10.075	BLL 5A	65	MEDIA	26	1X	MEDIA	74
95	12.122	2D	60	ALTA	26		MEDIA	74
96	012.003 C	2D	60	MEDIA	22	1X	MEDIA	74
97	011.012 C	BLL 4A	60	ALTA	24	1X	MEDIA	74
98	012.035 C	2D	65	MEDIA	28		MEDIA	73
99	11.041	BLL 5A	70	MEDIA	26	1X	ALTA	72
100	9.009	BLL 7A	70	MEDIA	26		MEDIA	72
101	5.093	BLL 6A	70	BAJA	28		ALTA	71
102	A50034	BLL 8A	70	MEDIA	26	1X	ALTA	71
103	9.051	BLL	70	MEDIA	28		MEDIA	70
104	9.072	BLL	70	MEDIA	28		MEDIA	69
105	6.029	BLL 8A	70	MEDIA	28		MEDIA	66
106	11.115	BLL 4A	65	MEDIA	28		MEDIA	66
107	9.024	BLL 5A	65	MEDIA	28		MEDIA	59
108	8.024	BLL 6A	65	MEDIA	30		BAJA	58
109	7141	BLL 7A	70	MEDIA	30		ALTA	58
110	5.138	BLL 6A	65	MEDIA	30		BAJA	56
111	6.102	BLL 8A	70	MEDIA	30	1X	ALTA	56

ANEXO 3

Relación de los 14 cebadores

gen	Primer	Sentido	Secuencia (5'→3')	Long.	Tm	GC%	Tamaño
<i>EDAR</i>	<i>EDAR1</i>	Forward	TAGCAAACCAAGAGCCAGGG	20	59.9	55	833
	<i>EDAR1</i>	Reverse	TGGAAGTGGGAGGACTGACA	20	60.1	55	
	<i>EDAR2</i>	Forward	CTCTGGAGATGACACGGCTG	20	60.1	60	884
	<i>EDAR2</i>	Reverse	TCAGGATGGC TGGAAGTGGGA	20	60.1	55	
	<i>EDAR</i>	Forward	AGCTCCCTGTCCTGCTTG	18	58.9	61.1	920
	<i>EDAR3</i>	Reverse	CCTCCTCTCCACACCTACCT	20	59.6	60	
<i>HOXC13</i>	<i>HOXC2</i>	Forward	GGCCATTCCCTCCACTTGAT	20	59.7	55	450
	<i>HOXC2</i>	Reverse	AGAACCTTGACAGCTGTGACC	20	60.5	55	
	<i>HOXC3</i>	Forward	AGATGACGAACTGCTGGGT	20	59.3	50	385
	<i>HOXC3</i>	Reverse	CTGTGACCCTCCTGTGACCT	20	60.8	60	
	<i>HOXC4</i>	Forward	GGGGCGGGACAGAGAAAAAG	20	60.9	60	581
	<i>HOXC4</i>	Reverse	GAGAACCTTGACAGCTGTGAC	20	59.1	55	
<i>KRT31</i>	<i>KRT31</i>	Forward	GGAGAGCGAGGACTGCAA	18	59	61	281
	<i>KRT31</i>	Reverse primer	TCAAGATGGGGAATTCAGTGGA	22	59	45	

ANEXO 4

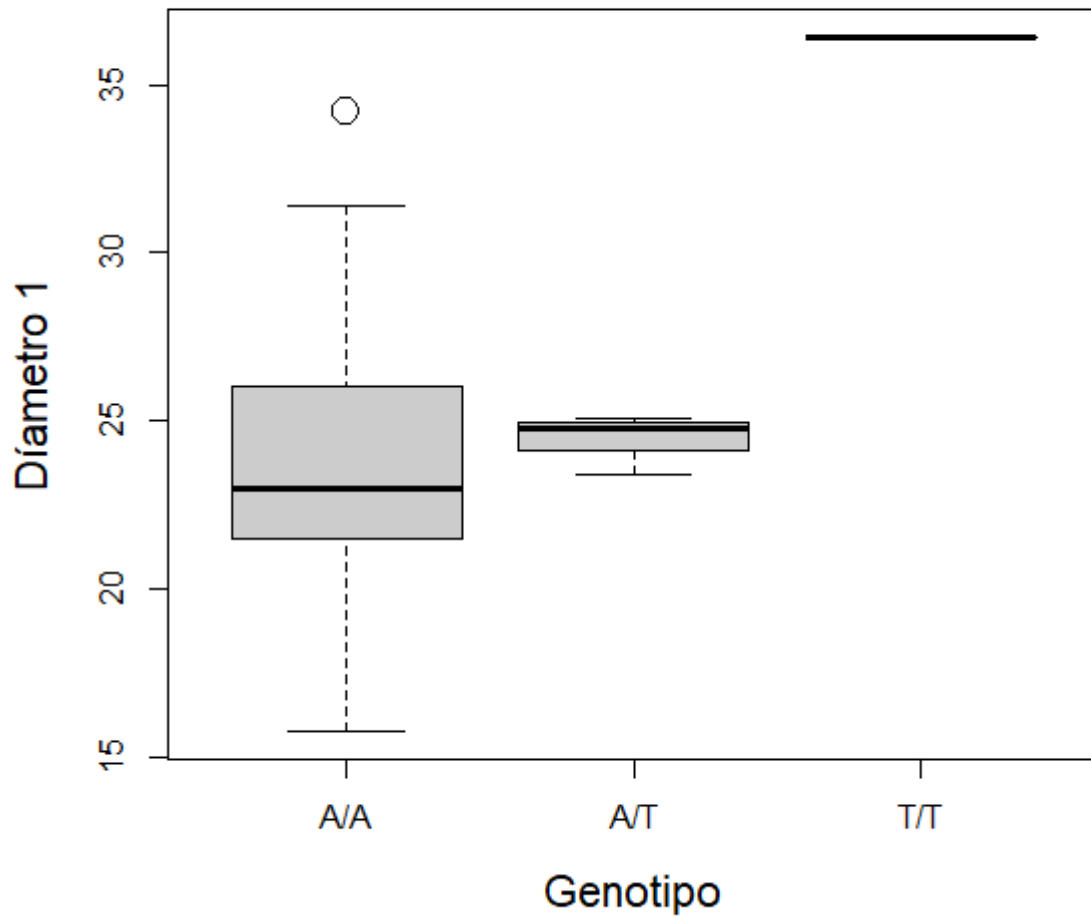
Genes en rumiantes

1	<i>K33-1</i>
2	<i>K33-2</i>
3	<i>KAP1-1</i>
4	<i>KAP11-1</i>
5	<i>KAP11-2</i>
6	<i>KAP11-3</i>
7	<i>KAP13-1</i>
8	<i>HGT-KAP6</i>
9	<i>KRT2</i>
10	<i>GPRC5A</i>
11	<i>TCHH</i>
12	<i>Fgfr2-IIIb</i>
13	<i>IGFB5</i>
14	<i>EDAR</i>
15	<i>KRT31</i>
16	<i>HOXC13</i>
17	<i>HOX3G</i>
18	<i>HOXA</i>
19	<i>BMP4</i>
20	<i>HOXD</i>
21	<i>EDARADD</i>
22	<i>EDA</i>

ANEXO 5

Gráfico del outlier del alelo T/T en el diámetro de fibra

HOXC13 - Variante 11648027



ANEXO 6

Genotipos del gen *EDAR*

2252010	2252111	2252123	2252227	2252231	2252234	2252304	2252343	2252364	2252376	2252425
C/C	A/A	C/T	T/T	A/G	G/G	C/C	G/G	G/C	G/G	A/A
T/T	A/A	C/C	T/T	G/G	A/G	C/C	G/G	G/G	G/G	G/G
C/T	A/T	C/C	T/T	G/G	G/G	C/C	G/G	G/G	A/G	G/G
C/C	A/A	C/C	A/T	G/G	G/G	C/C	G/G	G/G	G/G	G/G
C/C	A/A	C/T	A/T	A/G	G/G	C/C	G/G	G/C	G/G	A/G
C/T	A/A	C/T	T/T	A/G	G/G	C/C	G/G	G/C	G/G	A/G
C/C	A/A	C/C	T/T	G/G	G/G	C/C	G/G	G/G	G/G	A/A
C/T	A/A	C/C	T/T	G/G	A/A	C/C	G/G	G/G	A/G	G/G
C/C	A/A	C/C	A/T	G/G	G/G	C/C	G/G	G/G	A/G	G/G
C/C	A/A	C/C	T/T	G/G	G/G	C/C	G/G	G/G	G/G	A/A
C/C	A/A	C/C	T/T	A/G	G/G	C/C	A/G	G/G	A/G	G/G
C/T	A/A	C/C	T/T	G/G	G/G	C/C	G/G	G/G	G/G	A/G
C/C	A/A	C/T	T/T	A/G	G/G	C/T	G/G	G/C	G/G	A/G
C/C	A/A	T/T	T/T	A/A	G/G	C/C	G/G	C/C	G/G	A/A
C/C	A/A	C/C	T/T	G/G	G/G	C/C	G/G	G/G	G/G	A/G
C/C	A/A	C/C	T/T	G/G	G/G	C/C	G/G	G/G	G/G	A/A
C/C	A/A	C/C	T/T	A/G	G/G	C/C	A/G	G/G	G/G	A/G
C/T	A/A	C/C	A/T	G/G	A/G	C/C	G/G	G/G	G/G	G/G
C/C	A/A	C/C	A/T	G/G	G/G	C/C	G/G	G/G	G/G	A/G
C/C	A/A	C/C	A/T	G/G	G/G	C/C	G/G	G/G	A/G	G/G
C/C	A/A	C/T	T/T	A/G	G/G	C/C	G/G	G/C	G/G	G/G
C/C	A/A	T/T	T/T	A/A	G/G	C/C	G/G	C/C	G/G	A/A
C/T	A/A	C/C	T/T	G/G	A/G	C/C	G/G	G/G	G/G	G/G
C/C	A/A	C/T	T/T	A/G	G/G	C/C	G/G	G/C	A/G	A/G
C/T	A/A	C/C	T/T	G/G	G/G	C/C	G/G	G/G	G/G	G/G
T/T	A/T	C/C	T/T	G/G	A/A	C/C	G/G	G/G	G/G	G/G
C/C	A/A	C/T	A/T	A/G	G/G	C/C	G/G	G/C	G/G	A/G
C/C	A/A	C/C	A/T	G/G	G/G	C/C	G/G	G/G	A/G	G/G
C/C	A/A	C/C	T/T	A/G	G/G	C/C	A/G	G/G	G/G	A/G
C/T	A/A	C/T	T/T	A/G	A/G	C/C	G/G	G/C	G/G	A/G
C/T	T/T	C/C	T/T	G/G	G/G	C/C	G/G	G/G	G/G	G/G
C/C	A/A	C/T	T/T	A/G	G/G	C/C	G/G	G/C	G/G	G/G
C/C	A/A	C/C	A/T	G/G	G/G	C/C	G/G	G/G	G/G	A/G
C/C	A/A	C/C	A/T	G/G	G/G	C/C	G/G	G/G	A/G	G/G
C/T	A/A	C/T	T/T	A/G	A/G	C/C	G/G	G/C	G/G	A/G
C/T	A/T	C/C	T/T	G/G	G/G	C/C	G/G	G/G	G/G	G/G
C/T	A/A	C/C	T/T	G/G	G/G	C/C	G/G	G/G	A/G	G/G
C/C	A/A	C/C	T/T	G/G	G/G	C/C	G/G	G/G	A/G	A/G
C/T	T/T	C/C	T/T	G/G	G/G	C/C	G/G	G/G	A/G	G/G
C/C	A/A	C/T	T/T	A/G	G/G	C/C	G/G	G/C	G/G	G/G
C/C	A/A	C/C	T/T	A/G	G/G	C/T	G/G	G/G	G/G	G/G

C/T	A/A	C/T	T/T	A/G	G/G	C/C	G/G	G/C	G/G	A/G
C/C	A/A	C/C	T/T	G/G	G/G	C/C	G/G	G/G	G/G	G/G
C/T	A/A	C/C	T/T	G/G	G/G	C/C	G/G	G/G	G/G	G/G
C/T	A/T	C/C	T/T	G/G	G/G	C/C	G/G	G/G	G/G	G/G
C/C	A/A	C/C	T/T	G/G	G/G	C/C	G/G	G/G	G/G	A/G
C/C	A/A	C/T	A/T	A/A	G/G	C/C	G/G	G/C	G/G	A/G
C/T	T/T	C/C	T/T	G/G	G/G	C/C	G/G	G/G	G/G	G/G
C/C	A/A	C/C	T/T	G/G	G/G	C/C	G/G	G/G	A/G	G/G
T/T	A/T	C/C	T/T	G/G	G/G	C/C	G/G	G/G	G/G	G/G
C/C	A/A	C/T	T/T	A/A	G/G	C/T	G/G	G/C	G/G	A/G
C/C	A/A	C/T	T/T	A/G	G/G	C/C	G/G	G/C	A/A	A/G
C/C	A/A	C/C	T/T	G/G	G/G	C/C	G/G	G/G	A/G	A/G
C/T	A/A	C/C	T/T	A/G	G/G	C/C	A/G	G/G	G/G	G/G
C/C	A/A	T/T	T/T	A/A	G/G	C/C	G/G	C/C	G/G	A/A
C/T	A/A	C/C	A/T	G/G	G/G	C/C	G/G	G/G	G/G	G/G
C/C	A/A	C/T	T/T	A/A	G/G	C/C	A/G	G/C	G/G	A/G
C/C	A/A	C/C	A/T	G/G	G/G	C/C	G/G	G/G	G/G	A/G
C/C	A/A	C/C	A/T	G/G	G/G	C/C	G/G	G/G	G/G	A/G
C/C	A/A	C/C	T/T	G/G	G/G	C/C	G/G	G/G	A/G	G/G
C/C	A/A	C/C	A/T	G/G	G/G	C/C	G/G	G/G	G/G	A/G
T/T	A/T	C/C	T/T	G/G	G/G	C/C	G/G	G/G	G/G	G/G
C/C	A/A	C/T	T/T	A/A	G/G	C/C	A/G	G/C	G/G	A/G
C/C	A/A	C/T	A/T	A/G	G/G	C/C	G/G	G/C	G/G	A/G
C/C	A/A	C/T	T/T	A/G	G/G	C/C	G/G	G/C	G/G	A/A
C/C	A/A	C/C	T/T	G/G	G/G	C/C	G/G	G/G	G/G	A/G
C/T	A/A	C/C	T/T	G/G	G/G	C/C	G/G	G/G	G/G	A/G
C/C	A/A	C/C	T/T	G/G	G/G	C/C	G/G	G/G	A/G	A/G
C/T	T/T	C/C	T/T	G/G	G/G	C/C	G/G	G/G	A/G	G/G
C/T	T/T	C/C	A/T	G/G	G/G	C/C	G/G	G/G	G/G	G/G
C/C	A/A	C/C	T/T	G/G	G/G	C/T	G/G	G/G	A/G	G/G
C/C	A/A	T/T	T/T	A/A	G/G	C/C	G/G	C/C	G/G	A/A
C/C	A/A	C/T	A/T	A/G	G/G	C/C	G/G	G/C	G/G	G/G
C/T	A/A	C/T	T/T	A/G	G/G	C/C	G/G	G/C	G/G	A/G
C/T	T/T	C/C	T/T	G/G	G/G	C/C	G/G	G/G	A/G	G/G
C/C	A/A	C/C	A/T	G/G	G/G	C/C	G/G	G/G	A/G	G/G
C/C	A/A	C/C	T/T	G/G	G/G	C/C	G/G	G/G	A/G	G/G
C/C	A/A	C/C	A/T	G/G	G/G	C/C	G/G	G/G	G/G	G/G
C/T	A/A	C/C	T/T	G/G	A/A	C/C	G/G	G/G	G/G	A/G
C/T	A/T	C/C	T/T	G/G	G/G	C/C	G/G	G/G	G/G	G/G
C/T	A/A	C/C	T/T	G/G	A/A	C/C	G/G	G/G	G/G	A/G
C/T	A/T	C/C	A/T	G/G	G/G	C/C	G/G	G/G	G/G	G/G
C/C	A/A	T/T	T/T	A/A	G/G	C/C	G/G	C/C	G/G	A/G
C/T	T/T	C/C	T/T	G/G	G/G	C/C	G/G	G/G	G/G	G/G
C/C	A/A	C/T	T/T	A/G	G/G	C/C	G/G	G/C	A/G	A/G
C/T	A/A	C/C	T/T	G/G	G/G	C/C	G/G	G/G	A/G	G/G

2252434	2252459	2252558	2252582	2252613	2252618	2252646	2252745	2252747	2252791
G/G	A/G	G/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/C	A/A	C/T	C/C	C/C	G/G	G/G
A/A	G/G	G/G	C/C	A/C	C/C	C/C	C/C	G/G	A/A
G/G	G/G	G/G	C/T	A/C	C/T	C/C	C/T	G/G	G/G
G/G	A/G	G/G	C/T	A/A	C/C	C/C	C/T	G/G	G/G
G/G	A/G	G/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	T/T	A/A	T/T	C/C	T/T	G/G	G/G
G/G	G/G	G/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/C	A/A	C/T	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
G/G	A/G	G/G	C/C	A/A	C/T	C/C	C/C	G/G	G/G
G/G	A/A	G/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/C	A/A	C/C	C/T	C/C	G/G	G/G
G/G	G/G	A/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/C	A/A	C/T	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/T	A/A	C/T	C/C	C/T	G/G	G/G
G/G	G/G	G/G	C/T	A/A	C/C	C/C	C/T	G/G	G/G
G/G	G/G	G/G	C/T	A/A	C/T	C/C	C/T	G/G	G/G
G/G	A/G	G/G	C/T	A/A	C/C	C/C	C/C	G/G	G/G
G/G	A/A	G/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
G/G	A/G	G/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/C	A/C	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/C	A/A	C/C	C/C	C/C	G/G	A/A
G/G	A/G	G/G	C/T	A/A	C/T	C/C	C/T	G/G	G/G
G/G	G/G	G/G	C/T	A/A	C/T	C/C	C/T	G/G	G/G
G/G	G/G	G/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/C	C/C	C/C	C/T	C/C	G/G	G/G
G/G	A/G	G/G	C/T	A/A	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/T	A/A	C/T	C/C	C/T	G/G	G/G
G/G	G/G	G/G	C/T	A/A	C/T	C/C	C/T	G/G	G/G
G/G	A/G	G/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
A/G	G/G	G/G	C/C	A/C	C/C	C/T	C/C	G/G	A/A
G/G	G/G	G/G	C/C	A/A	C/T	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/C	C/C	C/C	C/C	C/C	G/G	G/G
G/G	A/G	G/G	C/T	A/A	C/C	C/T	C/C	G/G	G/G
G/G	G/G	G/G	C/C	A/A	T/T	C/C	C/C	G/G	G/G
G/G	A/G	G/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/C	A/C	C/C	C/C	C/C	G/G	G/G

G/G	G/G	G/G	C/C	C/C	C/C	C/C	C/C	G/G	A/A
G/G	G/G	G/G	C/C	A/A	C/C	C/C	C/C	G/G	A/A
G/G	G/G	A/G	C/C	A/A	C/C	C/T	C/C	G/G	G/G
G/G	A/G	G/G	C/T	A/A	C/T	C/C	C/T	G/G	G/G
A/A	G/G	G/G	C/C	A/C	C/C	C/T	C/C	G/G	A/A
G/G	G/G	G/G	C/C	C/C	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/C	A/C	C/T	C/C	C/C	G/G	A/A
G/G	G/G	G/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/C	A/A	C/T	C/C	C/C	G/G	A/A
G/G	A/A	G/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/T	A/A	C/C	C/C	C/T	G/G	G/G
G/G	A/G	G/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/T	A/A	C/C	C/C	C/T	G/G	G/G
G/G	G/G	G/G	C/T	A/A	C/C	C/C	C/T	G/G	G/G
G/G	G/G	G/G	C/C	A/A	T/T	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/T	A/A	C/C	C/C	C/T	G/G	G/G
A/G	G/G	G/G	C/C	A/C	C/T	C/C	C/C	G/G	A/A
G/G	A/G	G/G	C/C	A/A	C/T	C/C	C/C	G/G	G/G
G/G	A/G	G/G	C/T	A/A	C/T	C/C	C/T	G/G	G/G
G/G	G/G	A/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/C	A/C	C/C	C/C	C/C	G/G	G/G
A/G	G/G	G/G	C/T	A/C	C/T	C/C	C/T	G/G	A/A
G/G	G/G	G/G	C/C	A/A	C/T	C/C	C/C	G/G	G/G
G/G	A/A	G/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
G/G	A/G	G/G	T/T	A/A	C/T	C/C	C/T	G/G	G/G
G/G	A/G	G/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/C	C/C	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/T	A/A	C/T	C/C	C/T	G/T	G/G
G/G	G/G	G/G	C/C	A/C	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/T	A/A	C/T	C/T	C/T	G/G	G/G
G/G	G/G	G/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
A/G	G/G	G/G	C/T	A/C	C/C	C/C	C/C	G/G	A/A
G/G	G/G	G/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
A/G	G/G	G/G	C/T	A/A	C/T	C/C	C/T	G/G	A/A
G/G	A/A	G/G	C/T	A/A	C/C	C/C	C/C	G/G	G/G
A/A	G/G	G/G	C/C	C/C	C/C	C/C	C/C	G/G	A/A
G/G	A/G	G/G	C/C	A/A	C/C	C/C	C/C	G/G	G/G
G/G	G/G	G/G	C/C	A/A	C/T	C/C	C/C	G/G	G/G

ANEXO 7

Genotipos del gen *KRT31*

10534	10572	10575	10628	10657	10963	10995	11007
G/G	C/C	G/G	G/G	A/G	C/T	A/G	T/T
A/G	C/C	G/G	A/G	G/G	T/T	A/G	T/T
G/G	C/C	G/G	G/G	A/A	C/C	A/A	T/T
G/G	C/C	A/G	G/G	G/G	T/T	A/A	C/C
G/G	C/C	G/G	A/A	G/G	T/T	G/G	T/T
A/G	C/C	G/G	G/G	G/G	T/T	A/A	C/T
G/G	C/C	G/G	A/G	A/G	C/T	A/G	T/T
G/G	C/C	G/G	A/G	A/G	C/T	A/G	T/T
G/G	C/C	G/G	G/G	A/A	C/C	A/A	T/T
G/G	C/C	G/G	G/G	A/A	C/C	A/A	T/T
A/G	C/C	G/G	G/G	G/G	T/T	A/A	C/T
G/G	C/C	G/G	G/G	A/G	C/T	A/G	T/T
G/G	C/C	G/G	A/G	G/G	T/T	A/G	C/T
G/G	C/C	G/G	G/G	G/G	T/T	A/A	C/C
G/G	C/C	G/G	A/G	A/G	C/T	A/G	T/T
A/G	C/C	G/G	G/G	A/G	C/T	A/A	T/T
G/G	C/C	G/G	A/G	A/G	C/T	A/G	T/T
G/G	C/T	G/G	G/G	G/G	T/T	A/A	C/C
G/G	C/C	G/G	G/G	A/G	C/T	A/A	C/T
A/G	C/C	G/G	G/G	G/G	T/T	A/G	T/T
G/G	C/C	G/G	G/G	A/G	C/T	A/A	C/T
G/G	C/C	G/G	A/G	A/G	C/T	A/G	T/T
A/G	C/C	G/G	G/G	A/G	C/T	A/A	T/T
G/G	C/T	G/G	G/G	A/G	C/T	A/A	C/T
G/G	C/C	G/G	A/G	G/G	T/T	G/G	T/T
G/G	C/C	G/G	G/G	A/G	C/T	A/G	T/T
G/G	C/C	G/G	A/G	A/G	C/T	A/G	T/T
A/G	C/C	G/G	G/G	A/G	C/T	A/A	T/T
G/G	C/C	G/G	A/G	A/G	C/T	A/G	T/T
G/G	C/C	G/G	A/G	A/G	C/T	A/G	T/T
G/G	C/C	G/G	G/G	A/G	C/T	A/A	C/C
G/G	C/C	G/G	G/G	A/G	C/T	A/A	C/C
G/G	C/C	G/G	A/A	G/G	T/T	G/G	T/T

G/G	C/T	G/G	A/G	G/G	T/T	A/G	C/T
A/G	C/C	G/G	G/G	G/G	T/T	A/G	T/T
G/G	C/C	G/G	G/G	A/A	C/C	A/A	T/T
A/G	C/C	G/G	G/G	A/G	C/T	A/A	T/T
G/G	C/T	A/G	G/G	G/G	T/T	A/A	C/C
G/G	C/C	G/G	G/G	G/G	T/T	A/G	C/T
G/G	C/C	G/G	A/G	A/G	C/T	A/G	T/T
A/G	C/T	G/G	G/G	G/G	T/T	A/A	C/T
G/G	C/C	A/G	G/G	G/G	T/T	A/A	C/C
G/G	C/C	G/G	G/G	G/G	T/T	A/A	C/C
G/G	T/T	G/G	G/G	G/G	T/T	A/A	C/C
G/G	C/C	G/G	A/A	G/G	T/T	G/G	T/T
G/G	C/T	G/G	A/G	G/G	T/T	A/G	C/T
A/G	C/C	G/G	A/G	G/G	T/T	A/G	T/T
G/G	C/C	G/G	G/G	A/G	C/T	A/G	T/T
G/G	C/T	G/G	A/G	G/G	T/T	A/G	C/T
G/G	C/C	G/G	A/G	G/G	T/T	G/G	T/T
G/G	C/T	G/G	G/G	A/G	C/T	A/A	C/T
G/G	T/T	G/G	G/G	G/G	T/T	A/A	C/C
G/G	C/C	A/G	G/G	A/G	C/T	A/A	C/T
G/G	C/C	G/G	A/A	G/G	T/T	G/G	T/T
G/G	C/C	G/G	A/G	A/G	C/T	A/G	T/T
G/G	C/C	G/G	A/G	G/G	T/T	A/G	C/T
A/G	C/C	G/G	G/G	A/G	C/T	A/A	T/T
G/G	C/C	G/G	G/G	A/A	C/C	A/A	T/T
G/G	C/T	G/G	G/G	A/G	C/T	A/A	C/T
G/G	C/C	G/G	G/G	A/A	C/C	A/A	C/T
G/G	C/T	A/G	G/G	G/G	T/T	A/A	C/C
G/G	C/C	G/G	G/G	A/G	C/T	A/A	C/C
G/G	C/C	G/G	A/G	G/G	T/T	A/G	C/T
G/G	C/C	G/G	A/G	A/G	C/T	A/G	T/T
G/G	C/C	G/G	A/G	A/G	C/T	A/G	T/T
G/G	C/C	G/G	G/G	G/G	T/T	A/A	C/C
G/G	C/C	G/G	A/G	G/G	T/T	A/G	
G/G	C/C	G/G	A/G	G/G	T/T	A/G	
G/G	C/C	G/G	A/G	A/G	C/T	A/G	T/T
G/G	C/C	G/G	G/G	A/A	C/C	A/A	T/T
G/G	C/T	G/G	G/G	A/G	C/T	A/A	C/T
G/G	C/C	G/G	A/G	G/G	T/T	A/G	C/T

G/G	C/C	G/G	G/G	A/G	C/T	A/A	C/C
G/G	C/C	A/G	G/G	G/G	T/T	A/G	C/T
G/G	C/C	G/G	G/G	G/G	T/T	A/G	C/T
G/G	C/T	G/G	G/G	G/G	T/T	A/A	C/C
G/G	C/C	G/G	G/G	A/G	C/T	A/G	T/T
G/G	C/C	G/G	G/G	A/A	C/C	A/A	T/T
G/G	C/C	G/G	A/G	G/G	T/T	A/G	C/T
G/G	C/C	G/G	A/G	G/G	T/T	A/G	C/T
G/G	C/C	G/G	A/G	A/G	C/T	A/G	T/T
G/G	C/T	G/G	G/G	G/G	T/T	A/A	C/C
G/G	C/C	G/G	G/G	A/A	C/C	A/A	T/T
A/G	C/T	G/G	G/G	G/G	T/T	A/A	C/T
G/G	C/T	G/G	G/G	G/G	T/T	A/G	C/T
G/G	C/C	G/G	G/G	A/A	C/C	A/A	T/T
G/G	C/C	A/A	G/G	G/G	T/T	A/A	
G/G	C/C	G/G	A/G	A/G			
G/G	C/C	G/G	G/G	A/G	C/T	A/G	T/T

ANEXO 8

Genotipos del gen HOXC13

11647923	11647945	11648009	11648027	11648042
A/T	C/C	C/C	A/A	G/G
T/T	C/C	C/C	A/A	G/G
A/A	C/T	C/C	A/A	G/G
A/A	C/C	C/C	A/A	G/G
A/A	C/T	C/C	A/T	G/G
A/A	T/T	C/C	A/A	G/G
T/T	C/C	C/C	A/A	G/G
A/T	C/T	C/C	A/A	G/G
A/T	C/T	C/C	A/A	G/G
T/T	C/C	C/C	A/A	G/G
T/T	C/C	C/C	A/A	G/G
A/A	T/T	C/C	A/A	G/G
A/A	C/T	C/C	A/A	G/C
A/A	C/C	C/C	A/A	G/G
A/A	T/T	C/C	A/A	G/G
A/A	T/T	C/C	A/A	G/G
A/T	C/C	C/C	A/A	G/G
A/A	T/T	C/C	A/A	G/G
A/A	C/C	C/C	A/A	G/G
T/T	C/C	C/C	A/A	G/G
A/A	T/T	C/C	A/A	G/G
A/A	T/T	C/C	A/A	G/G
A/T	C/T	C/C	A/A	G/G
A/A	C/C	C/C	A/A	G/G
T/T	C/C	C/C	A/A	G/G
A/A	T/T	C/C	A/A	G/G
A/A	C/C	C/C	A/A	G/G
A/T	C/C	C/C	A/A	G/G
A/T	C/T	C/C	A/A	G/G
A/A	C/T			
T/T	C/C	C/C	A/A	G/G
A/A	C/C	C/C	A/A	G/G
A/A	T/T	C/C	A/A	G/G

A/A	C/C	C/C	A/A	G/G
A/A	T/T	C/C	A/A	G/G
A/T	C/T			
A/A	T/T	C/C	A/A	G/G
A/T	C/T	C/C	A/A	G/G
A/A	C/C	C/C	A/A	G/G
A/T	C/C	C/C	A/A	G/G
A/A	T/T	C/C	A/A	G/G
T/T	C/C	C/C	A/A	G/G
A/A	T/T	C/C	A/A	G/G
A/A	C/T	C/C	A/A	G/G
A/T	C/T	C/C	A/A	G/G
A/T	C/T	C/C	A/A	G/G
A/A	T/T	C/C	A/A	G/G
A/A	T/T	C/C	T/T	G/G
A/T	C/C	C/C	A/A	G/G
A/A	C/C	C/C	A/A	G/G
A/T	C/C	C/C	A/A	G/G
A/T	C/C	C/C	A/A	G/G
A/A	C/C	C/C	A/A	G/G
A/T	C/C	C/C	A/A	G/G
A/A	T/T			
A/A	C/C	C/C	A/A	G/G
A/T	C/C	C/C	A/A	G/C
A/A	C/T	C/C	A/A	G/G
A/A	T/T	C/C	A/A	G/G
A/T	C/T	C/C	A/A	G/G
A/T	C/T	C/C	A/A	G/G
A/A	C/T	C/C	A/A	G/G
A/A	C/C	C/C	A/A	C/C
A/T	C/T	C/C	A/A	G/G
A/A	T/T	C/C	A/A	G/G
A/A	T/T	C/C	A/A	G/G
A/T	C/T	C/C	A/A	G/G
A/A	C/C	C/C	A/A	G/G
A/A	T/T	C/C	A/A	G/G
A/A	C/T	C/C	T/T	G/G
A/A	C/T	C/C	A/A	G/G
A/A	T/T	C/C	A/A	G/G

A/A	T/T	C/C	A/A	G/G
A/A	C/T	C/C	A/A	G/G
A/A	T/T	C/T	A/A	G/G
A/A	T/T	C/C	A/A	G/G
A/A	C/T	C/C	A/A	G/G
A/A	T/T	C/C	A/A	G/G
A/T	C/C	C/C	A/A	G/G
T/T	C/C	C/C	A/A	G/G
A/T	C/C	C/C	A/A	G/G
A/T	C/T	C/C	A/A	G/G
A/A	T/T	C/C	A/A	G/G
A/A	T/T	C/C	A/T	G/G
A/A	T/T	C/C	A/A	G/G
A/T	C/C	C/C	A/A	G/G
A/T	C/C	C/C	A/A	G/G
A/A	C/C	C/C	A/A	G/G
A/T	C/T			
T/T	C/C	C/C	A/A	G/G
A/A	T/T	C/C	A/A	G/G