



UNIVERSIDAD PERUANA
CAYETANO HEREDIA
ESCUELA DE POSTGRADO

Caracterización de nuevos marcadores genéticos
Microsatélites e identificación de SNP en el gen de
Tricohialina en alpacas (*Vicugna pacos*)

TESIS PARA OPTAR EL GRADO DE MAGÍSTER EN
BIOQUÍMICA Y BIOLOGÍA MOLECULAR

Bach. Irene Rosa Maria Delgado de la Flor Montauban

Lima - Perú

2014

Dr. José R. Espinoza Babilón

Asesor

Jurado de Tesis

Dra. Patricia Herrera Velit
Presidenta

Dr. Luis Destefano Beltrán
Vocal

Dr. Luis Aguilar Mendoza
Secretario

AGRADECIMIENTOS

Al Dr. José Espinoza B., asesor de tesis, por la oportunidad dada.

Al Dr. Jorge Rodríguez por el apoyo y la ayuda brindada.

A los miembros del jurado por las sugerencias aportadas.

A los miembros de la Unidad Biotecnología Molecular por toda la ayuda recibida, en especial a Teresa Barreto y Susana Castro.

A Juan Agapito del IPEN (Instituto Peruano de Energía Nuclear) por la ayuda prestada.

A CONCYTEC (Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica) que financió mis estudios de Maestría.

INDICE

I. Introducción	1
II. Planteamiento de la Investigación	3
1. Planteamiento del Problema	3
2. Marco Teórico	5
2.1 Marcadores Genéticos	5
2.1.1 Marcadores Microsatélites	5
2.1.2 Marcadores SNP	9
2.1.3 Parámetros Genéticos	13
2.1.3.1 Polimorfismo Genético	13
2.1.3.2 Probabilidad de Exclusión	13
2.1.3.3 Contenido de información polimórfica (PIC)	14
2.1.3.4 Heterocigosidad	14
2.1.3.5 Equilibrio de HARDY-WEINBERG	15
2.1.3.6 Equilibrio y desequilibrio de ligamiento	16
2.2 Tricohialina	17
3. Justificación del Estudio	21
III. Objetivos	23

	24
IV. Metodología	
1. Caracterización de marcadores microsatélites	24
1.1 Localización del estudio	24
1.2 Muestras de Animales	24
1.3 Elección de loci microsatélites	25
1.4 Tamizado de loci microsatélites con capacidad de amplificación en alpacas	27
1.4.1 Amplificación de ADN microsatélite	27
1.4.2 Selección de loci microsatélites para marcado con fluorocromos	28
1.4.3 Caracterización de los loci microsatélites	29
1.5 Reacción de PCR múltiple	29
2. Caracterización del gen de Tricohialina (TCHH)	29
2.1 Muestras de Animales	29
2.2 Alineamiento de secuencias	31
2.3 Diseño de cebadores	31
2.4 Amplificación del fragmento del gen de Tricohialina	32
2.5 Purificación del fragmento	32
2.6 Secuenciamiento del fragmento del gen Tricohialina	33
2.7 Edición de Secuencias	33
2.8 Análisis de SNP de Tricohialina	33
2.9 Verificación de la presencia de SNP por Clonación	34
2.10 Análisis del fragmento amplificado del gen de Tricohialina en alpacas Suri y Huacaya	35

V. Resultados	37
Microsatélites	37
Tricohialina	43
VI. Discusión	57
VII. Conclusiones	67
VIII. Recomendaciones	68
IX. Bibliografía	69

LISTA DE ABREVIATURAS

A	Diversidad alélica.
ADN	Ácido desoxirribonucleico.
dNTP	2'-deoxinucleotidos 5'-trifostato
FAO	Organización de las Naciones Unidas para la Agricultura y la Alimentación
h	Diversidad génica.
H	Heterocigosidad.
H _E	Heterocigosidad esperada.
H _O	Heterocigosidad observada.
LB	Medio de cultivo Luria-Bertoni
MgCl ₂	Cloruro de Magnesio
Mg ⁺² .	Magnesio
mg	Miligramos
ml	Mililitros
mM	Milimolar.
ng	Nanogramos
pb	Pares de bases
PCR	Reacción en cadena de la polimerasa.

PE	Probabilidad de exclusión.
PE1	Probabilidad de exclusión total para excluir un candidato a pariente dado solo el genotipo de la cría.
PE2	Probabilidad de exclusión total para excluir un candidato a pariente dado el genotipo de la cría y de otro pariente.
PE3	Probabilidad de exclusión para un par de candidatos a parientes.
PEA	Probabilidad de exclusión acumulada o total.
PIC	Contenido de información polimórfica.
SNPs	Polimorfismos de nucleótidos únicos.
SSRs	Simple secuencias repetidas.
STRs	Repeticiones cortas en serie.
TBE	Buffer Tris-Borato-EDTA
U	Unidad.
μl	Microlitro.
μM	Micromolar.
μg	Microgramo.

RESUMEN

El presente trabajo tiene como objetivo la identificación y caracterización de marcadores genéticos microsatelites y polimorfismos de nucleótido simple (SNP) y evaluar su posible rol en el fenotipo Suri y Huacaya de alpacas (*Vicugna pacos*) de los departamentos de: Puno, Cuzco, Huancavelica, Junin, Apurimac y Arequipa. Se determinaron 6 loci microsatélites: MNA0295, MNA0218, MNA0366, MNA0388, MNA0351 y MNA0394 y para 3 SNPs: A57G, A227G y A458G de un fragmento del extremo 3' terminal (3'UTR) del gen de Tricohialina. Los marcadores microsatélites mostraron ser altamente informativos ($PIC > 0.7$; $A > 6$; $H_o > 0.586$, $H_e > 0.691$; $PE_{acumulada} = 0.9942$). Los 3 SNP fueron identificados en un fragmento de 465bp perteneciente a la región 3'UTR del gen de Tricohialina. Los marcadores SNPs A57G y A458G son altamente informativos ($PIC=0.374$, $H_o > 0.351$, $H_e=0.5$) y se encuentran en desequilibrio de ligamiento genotípico ($p = 0$). Mientras que el marcador SNP A227G es poco informativo ($PIC=0.077$, $H_o =0.083$, $H_e=0.081$) y su alelo de menor frecuencia se encuentra presente departamentos de Huancavelica y Arequipa. Análisis de asociación entre los marcadores SNP y los fenotipos Suri y Huacaya no mostraron diferencias significativas para genotipos ($p>0.5$), haplotipos ($p>0.5$), ni en los análisis de múltiples componentes. En conclusión estos resultados sugieren que los SNP identificados del gen de Tricohialina no se encuentran relacionados a los fenotipos Suri y Huacaya.

Palabras claves: alpaca, microsatelites, SNP

ABSTRACT

The aim of this Project is the identification and characterization of new microsatellites and SNP genetic markers and evaluate their possible role in alpaca's (*Vicugna pacos*) breeds phenotypes: Suri and Huacaya from Puno, Cuzco, Huancavelica, Junin, Apurimac and Arequipa. Six microsatellite loci: MNA0295, MNA0218, MNA0366, MNA0388, MNA0351 and MNA0394 and 3 SNPs: A57G, A227G and A458G from the 3'UTR part of the trichohyalin gene were identified. The microsatellite markers were highly informative ($PIC > 0.7$, $A > 6$; $H > 0.586$, $He > 0.691$; $PE = 0.9516$). The 3 SNPs were identified in a 465bp fragment belonging to 3'UTR region from the Trichohyalin gene. The A57G and A458G SNP markers are highly informative ($PIC = 0.374$, $Ho > 0.351$, $He = 0.5$) and are in genotypic linkage disequilibrium ($p=0$). The A227G SNP marker is less informative ($PIC = 0.077$, $Ho = 0.083$, $He = 0.081$) and the less frequent allele is present only in 2 populations: Huancavelica and Arequipa. An Association analysis between SNP markers and the Huacaya and Suri phenotypes showed no significant difference for genotypes frequencies ($p > 0.5$), the haplotype frequencies ($p > 0.5$) and Factorial Correspondence Analysis (FCA). In conclusion these results suggest that the SNP identified in the trichohyalin gene are not related to Suri or Huacaya phenotypes.

Key words: Alpaca, microsatellite, SNP

I. INTRODUCCIÓN

Los marcadores genéticos son herramientas de utilidad en estudios de conservación, ecología, filogenia, evolución (Bruford y col, 2003), asociación y enfermedades (Kluth y Distl, 2013). En los últimos años, con el desarrollo de nuevas tecnologías y disminución de los costos, la generación de estos marcadores se ha simplificado, lo que se refleja en el aumento en el número de ellos.

De los diferentes tipos de marcadores genéticos existentes, los microsatélites y los SNPs son de utilidad en diferentes estudios realizados a la fecha. Los marcadores microsatélite poseen la ventaja de ser polimórficos lo que reduce el número de marcadores necesarios para realizar un análisis; los SNPs han cobrado una mayor relevancia en la actualidad, dado que se pueden generar y analizar una gran cantidad de ellos obteniéndose una gran cantidad de información (Morin y col, 2004).

Para poder rastrear la segregación de una característica de tipo cuantitativa es necesario el uso de un gran número de marcadores genéticos. En alpacas este número es bastante escaso.

La alpaca (*Vicugna pacos*) es un animal de importancia socioeconómica para el sector alto andino del Perú. Del comercio de los camélidos sudamericanos depende más de 15 mil familias ubicadas en los departamentos más pobres del país (MINAGRI, 2013). El Perú tiene más de 3 millones de alpacas, de las cuales casi el 100% de ellas habitan la zona de la sierra (INEI, 2013).

La alpaca es un camélido sudamericano perteneciente al Orden Artiodactyla que habita a más de 4000 msnm a temperaturas de entre -15 y 20°C. (Pérez-Cabal y col, 2010). Existen dos *razas* de alpacas: la Suri y la Huacaya las cuales se diferencian por su tipo de fibra. La *raza* Huacaya presenta una fibra compacta, lisa y muy risada mientras que en la *raza* Suri la fibra es lustrosa, sedosa y menos rizada. (Presciuttini y col, 2010)

En los últimos años se ha producido un descenso en la calidad de la fibra de alpaca debido principalmente a un mal criterio e inadecuado control en la selección y manejo reproductivo (Bonavia D, 1996) Para poder mejorar este recurso es necesario realizar estudios genéticos para definir los genes relacionados con una mejor calidad de fibra, y para poder identificar estos genes es necesario un gran cantidad de marcadores genéticos.

El presente estudio busca generar nuevos marcadores genéticos, microsatélites y SNPs, para la alpaca (*Vicugna pacos*), caracterizarlos en una población de alpacas del Perú y evaluar los marcadores SNPs de Trichialina en individuos de la raza Suri y Huacaya.

II PLANTEAMIENTO DE LA INVESTIGACIÓN

1. Planteamiento del Problema

La fibra de alpaca es un producto de gran importancia económica para la zona altoandina del Perú. En los últimos años se ha producido un descenso en la calidad de la fibra de alpaca debido principalmente a un mal criterio e inadecuado control en la selección y manejo reproductivo (Bonavia D, 1996). La calidad de la fibra se determina en base a los siguientes indicadores: el diámetro de la fibra (MDF), el coeficiente de variación del diámetro de fibra (CVDF), el índice de curvatura (IC) y la longitud de la mecha (LM) (Candido y Gutiérrez, 2011). El valor de la fibra de alpaca está dado principalmente por el diámetro de la fibra, a menor diámetro mayor valor tiene el producto. Esto se debe a que la sensación de confort que ejercen los tejidos sobre la piel está determinada por el grosor de la fibra (Frank, 2006).

La calidad de la fibra es una característica de tipo cuantitativa, es decir es un rasgo heredable determinado por la contribución de más de un gen, lo cual complica la identificación de asociaciones de tipo fenotipo-genotipo.

Las técnicas que permiten rastrear la segregación de una característica cuantitativa emplean un gran número de marcadores genéticos; en alpacas existe escasa información de este tipo. En la actualidad hay alrededor de 100 marcadores genéticos reportados y de utilidad en alpacas, constituyendo un número insuficiente para poder evaluar una característica fenotípica de tipo cuantitativa (Sasse y col , 2000; McPhartlan y col, 1998; Obreque y col, 1998; Penedo y

col 1998; Penedo y col 1999a; Penedo y col 1999b; Bustamante 2003). Además se desconoce la posición de estos marcadores en el genoma de la alpaca lo cual limita su uso.

A la fecha, se encuentran publicados en el NCBI (National Center of Biotechnological Information) los contigs generados por el proyecto de secuenciamiento del genoma de alpaca del Centro de Secuenciamiento Genómico de la Universidad de Washington y el Instituto Broad. Sin embargo estos datos solo presentan una cobertura de 2x.

Para poder determinar asociaciones de tipo fenotipo-genotipo que ayuden a establecer mejores criterios de control y manejo reproductivo que conlleven a una mejor en la calidad de los diferentes productos comerciales generados a partir de la alpaca, es necesario la generación de nuevos marcadores genéticos que provean una mayor cobertura del genoma.

Con la finalidad de generar estos nuevos marcadores genéticos se siguieron distintas aproximaciones según el marcador genético: microsatelites y SNP. La primera aproximación incluyó el empleo de marcadores genéticos microsatélites putativos, estos fueron seleccionados de un artículo científico que publicó cebadores diseñados en base a la información publicada en el NCBI sobre el genoma de la alpaca. Estos marcadores eran específicos para la especie y no habían sido probados.

La segunda aproximación busca marcadores SNP en el gen de Tricohialina, cuya proteína se encuentra involucrada en la formación y estructura del pelo. Al ser un gen codante, la presencia de un SNP en este gen indicaría, no solo variación genética, sino también la existencia de polimorfismos en la proteína que podrían estar relacionados a la finura de fibra o tipo de fibra, lisa o crespa.

2. Marco Teórico

2.1 Marcadores genéticos

2.1.1 Marcadores Microsatélites

Definición

Los marcadores genéticos microsatélites, conocidos también como SSR (simple sequence repeats) o STR (short tandem repeats), son secuencias repetitivas de hasta 6 pares de bases que se encuentran distribuidas a lo largo de todo el genoma eucarionte (Zou y col, 2005; Freeland, 2005; Hancock, 1991; Roizès, 2000). Frecuentemente exhiben un gran número de alelos por locus y altos grados de heterocigocidad (Hamilton, 2009) (Figura 1)

Características

Los marcadores microsatélites son de tipo codominantes, altamente polimórficos, presentan una herencia de tipo mendeliana simple y una distribución frecuente en el genoma (Freeland, 2005).



Figura 1. Esquematización de tres loci microsatélites (TA)_n, (TAA)_n y (GC)_n. Se aprecia que el locus 1 (TA) se encuentra en estado de homocigosis, mientras que los locus 2 (TAA)₇, (TAA)₈ y el locus 3 (GC)₇, (GC)₈ se encuentran en estado de heterocigosis. Modificado de Freeland (2005)

Pueden ser di-,tri-, tetra-, penta-, o hexanucleotidos dependiendo del número de bases nitrogenadas que conforman la secuencia repetitiva (Freeland, 2005). Las repeticiones dinucleótidas se usan con mayor frecuencia a pesar de que presentan bandas “stutter”. Las bandas “stutter” se producen por el deslizamiento de la polimerasa durante el proceso de amplificación generando fragmentos de menor tamaño que dificultan la identificación del alelo real (Guichoux y col, 2011). Son secuencias dinucleótidas entre el 30% y el 60% del total de microsatélites siendo, en el genoma de los vertebrados, las repeticiones (AC) n y (GT) n las más frecuentes (Hancock, 2003).

Esta última característica, tipo de repetición, junto con el tipo de secuencia en la que se encuentra el marcador, codante o no codante, influyen la distribución de los microsatélites a lo largo del genoma. Es más frecuente encontrarlos en zonas no codantes (Chistiakov y col, 2006) cumpliéndose esto en la mayoría de los casos. Una excepción son los microsatélites trinucleótidos debido a que estos no alteran el marco de lectura de un gen (Hancock, 2003).

El número de veces que la secuencia repetitiva se repite es un factor de importancia debido a que afecta el comportamiento del modelo mutacional del marcador. Generalmente los loci con mayor número de repeticiones presentan una mayor tasa mutacional lo que sugeriría la presencia de un mayor polimorfismo (Guichoux y col, 2011).

Otra característica de importancia es que las secuencias repetitivas no siempre son continuas, de acuerdo a esto existen marcadores que se denomina de repetición perfecta, cuando los bloques de secuencias repetidas se encuentran alineados uno detrás del otro, o de repetición imperfecta, cuando esta secuencia se ve interrumpida por otra diferente. Los marcadores microsatélites de repetición imperfecta presentan la dificultad de una falta de equivalencia entre

el largo del fragmento y la secuencia del amplicón, por lo que un tamaño de este último puede corresponder a varias secuencias (Guichoux y col, 2011).

Estabilidad

Lo microsatélites presentan una tasa mutacional de $10^{-4} - 10^{-5}$ eventos por locus por replicación la cual es mucho mayor que la del resto del genoma (10^{-6}) (Freeland, 2005). Sin embargo esta tasa mutacional no es constante y varía de una especie a otra.

Se postula que este alto grado de mutación se debe a errores de alineamiento entre las hebras y a un resbalamiento de la ADN polimerasa durante la replicación del ADN (Freeland, 2005). A diferencia del caso de una hebra de secuencia no repetitiva, la secuencia repetitiva de un microsatélite favorece, durante el proceso de síntesis de la hebra complementaria, la generación de un desfase al momento del alineamiento ocasionando la síntesis de un mayor o menor número de repeticiones, generando la aparición de nuevos alelos en un marcador microsatélite.

La recombinación es otra posible causa de la variación del número de repeticiones presentes en un microsatélite; al comparar los rangos de mutación de los microsatélites entre células en proceso de mitosis y células en proceso de meiosis en levadura, se observa que son muy similares, sin embargo el grado de recombinación en células en meiosis es mucho mayor (Hancock, 2003).

Usos

Los marcadores microsatélites se emplean en estudios de genética poblacional (deWoody y col, 1995), variabilidad poblacional (Maudet y col, 2002), genética de la conservación (Cañon, 2000), relaciones de parentesco (Agapito y col, 2008; Aderson y col, 2002) y generación de mapas de ligamiento en mamíferos (Kemp y col, 1995).

También son de utilidad en el manejo de poblaciones de animales ganaderos puesto que permiten cuantificar los niveles de diversidad genética en las poblaciones, documentar la presencia de introgresión entre poblaciones y asignar individuos genéticamente similares a ciertas poblaciones, razas o especies. (Bruford y col, 2003)

2.1.2 Marcadores SNP

Definición

Se denomina SNP al cambio de una base nitrogenada en un secuencia de ADN (Vignal y col, 2002; Lui, 2007), o a un sitio en una secuencia de ADN, que presenta más de un nucleótido en una población (Morin y col, 2004). Una sustitución de nucleótido única será considerada un SNP, si el alelo de menor frecuencia se encuentra en una frecuencia alélicas mayor al 1% (Brookes, 1999; Vignal y col, 2002) (Figura 2)

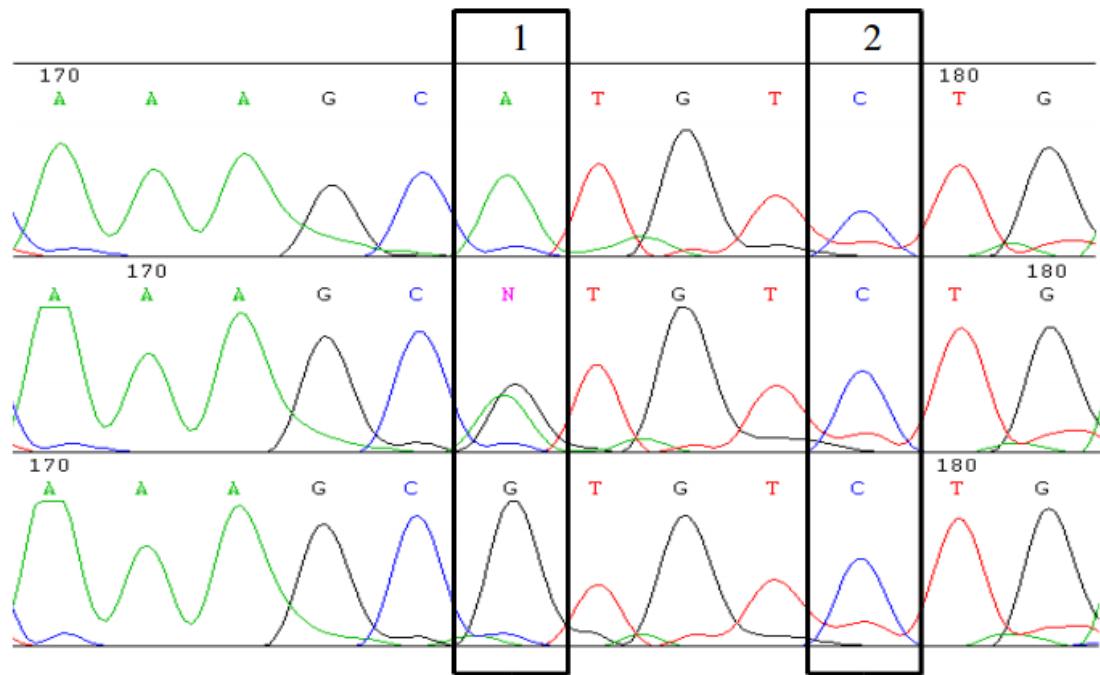


Figura 2. SNP. Identificación de SNP mediante el alineamiento de secuencias obtenido a partir de secuenciamiento directo de producto de PCR. 1. Señala un SNP. Al comparar las secuencias se observa el cambio de 1 base. Las secuencias superior e inferior son homocigotas para la base nitrogenada mientras que la secuencia del medio es heterocigota. 2. La secuencia presenta la misma base nitrogenada en los tres individuos (Vignal y col, 2002).

Características

Los SNP son considerados marcadores bialélicos principalmente por dos razones: a) la baja frecuencia de sustitución de nucleótidos únicos ($1 \times 10^{-9} - 5 \times 10^{-9}$ por nucleótidos por año), razón por la cual se generan los SNP (Li y col, 1981; Marínez-Arias y col, 2001) y b) el sesgo en los tipo de mutación, ocurre con mayor frecuencia una transición que una transversión. En roedores y en humanos la tasa de transición/transversión es de 1.4 (Collins y Jukes, 1994) mientras que en aves la tasa es mayor que en mamíferos con valores de 2.3 (Smith y col, 2001) a 4 (Kim y col, 2002) a partir de ESTs y 2.36 en secuencias no codantes (Vignal y col, 2002).

Son marcadores codominantes que presentan una herencia de tipo mendeliana simple (Lui, 2007). Se pueden encontrar SNP espaciados entre 300bp y 1000bp uno del otro (Morin, 2004). Se localizan en regiones codantes y no codantes del genoma pero se encuentran con mayor frecuencia en las regiones no codantes (Syvanen, 2001). Los SNP en zonas no codantes son importantes en estudios genómicos comparativos o evolutivos, los SNP presentes en zonas regulatorias pueden afectar la tasa de transcripción y aquellos presentes en zonas codantes pueden alterar la estructura y con ellos la función de las proteínas (Kim y Misra, 2007).

Los SNP presentan un menor valor de contenido de información polimórfica (PIC), lo que los hace menos informativos a nivel individual, sin embargo esto se ve balanceado debido a una mayor abundancia de ellos (Lui, 2007).

Uso

El empleo de SNPs presenta ciertas limitaciones como el bajo nivel de polimorfismo en relación a otros marcadores genéticos empleados, lo costoso de la técnica y una limitada disponibilidad de bases de datos e información, a excepción del caso del hombre (Gill, 2001). Sin embargo en la actualidad es una técnica cada vez más empleada. Los SNP se emplean en mapeo genético, estudios de asociación, genética poblacional y evolutiva (Syvanen, 2001), genética forense (Sobrino y col, 2005) ecología y conservación (Morin, 2004)

2.1.3 Parámetros Genéticos

2.1.3.1 Polimorfismo genético

El Polimorfismo genético es la presencia de dos o más alelos asociados a un locus, en donde el alelo más raro debe de presentarse en una frecuencia mayor al 1%. El polimorfismo genético se genera principalmente debido a mutaciones puntuales, inserciones, deleciones, conversión génica y recombinación intralélica (Nei y Kumar, 2000).

2.1.3.2 Probabilidad de exclusión (PE)

Se denomina probabilidad de exclusión (PE) a la capacidad de un marcador genético para excluir a un padre falsamente asignado en términos de probabilidad (Weir, 1996). La probabilidad de exclusión se encuentra determinada por el número de alelos y de las frecuencias alélicas del marcador y no por las frecuencias genotípicas observadas (Butler, 2005).

Jamieson y Taylor (1997) señalan 4 fórmulas distintas para determinar la probabilidad de exclusión (PE), en base a los datos disponibles para el análisis:

a) PE1: Probabilidad de exclusión para un candidato a pariente.

$$PE1 = 1 - 2 \sum_{i=1}^n p_i^2 + \sum_{i=1}^n p_i^3 + 2 \sum_{i=1}^n p_i^4 - 3 \sum_{i=1}^n p_i^5 - 2 \left(\sum_{i=1}^n p_i^2 \right)^2 + 3 \sum_{i=1}^n p_i^2 \sum_{i=1}^n p_i^3$$

Donde: p_i es la frecuencia del alelo I y n es el número de alelos en un locus.

b) PE2: Probabilidad de exclusión para candidato a pariente dado el genotipo de un pariente conocido del sexo opuesto.

$$PE2 = 1 - 4 \sum_{i=1}^n p_i^2 + 2 \left(\sum_{i=1}^n p_i^2 \right)^2 + 4 \sum_{i=1}^n p_i^3 - 3 \sum_{i=1}^n p_i^4$$

Donde: p_i es la frecuencia del alelo I y n es el número de alelos en un locus.

c) PE3: Probabilidad de exclusión para un par de candidatos a parientes.

$$PE3 = 1 + 4 \sum_{i=1}^n p_i^4 - 4 \sum_{i=1}^n p_i^5 - 3 \sum_{i=1}^n p_i^6 - 8 \left(\sum_{i=1}^n p_i^2 \right)^2 + 8 \left(\sum_{i=1}^n p_i^2 \right) \left(\sum_{i=1}^n p_i^3 \right) + 2 \left(\sum_{i=1}^n p_i^3 \right)^2$$

Donde: p_i es la frecuencia del alelo I y n es el número de alelos en un locus.

d) PEA = Probabilidad de exclusión acumulada o total.

$$PE \text{ acumulada o total} = 1 - (1 - P_1) (1 - P_2) (1 - P_3) \cdots (1 - P_k)$$

Donde: P_k es la probabilidad de exclusión del locus K.

2.1.3.3 Contenido de información polimórfica (PIC)

El contenido de información polimórfica (PIC) constituye una medida de la informatividad y polimorfismo de un locus genético, (Botstein y col, 1980). El valor de PIC es determinado mediante la siguiente fórmula:

$$PIC = 1 - \left(\sum_{i=1}^n p_i^2 \right) - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2 p_i^2 p_j^2$$

Donde: p_i es la frecuencia del alelo I, p_j es la frecuencia del alelo J y n es el número de alelos en un locus (Botstein y col., 1980).

2.1.3.4 Heterocigosidad

La heterocigosidad (H) indica la proporción de individuos heterocigotos de una población. Un alto valor de heterocigosidad indica una alta diversidad alélica presente en la muestra analizada (Butler, 2005). Existen dos indicadores de heterocigosidad: la heterocigosidad esperada (HE) y la heterocigosidad observada (HO).

La heterocigosidad esperada (HE) o diversidad génica (h) (Nei, 1973), es la probabilidad de que dos alelos escogidos al azar en una población sean diferentes (Weir, 1996).

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

Donde: p_i es la frecuencia del alelo I y n es el número de alelos en un locus (Nei, 1973).

La HE indica el nivel de heterocigosidad esperado para una población que se encuentre en equilibrio de Hardy-Weinberg (Freeland, 2005).

La heterocigosidad observada (HO) indica la proporción de individuos heterocigotos observados para un locus en particular en una población (Weir, 1996).

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

2.1.3.5 Equilibrio de HARDY-WEINBERG

En 1908 Godfrey Hardy y Wilhelm Weinberg formularon independientemente una relación que puede ser empleada para predecir frecuencias alélicas según las frecuencias genotípicas, o a la inversa, determinar frecuencias genotípicas según las frecuencias alélicas (Hamilton, 2009). Esta propiedad fue denominada Principio de Hardy –Weinberg (Nei y Kumar, 2000).

Esta relación se encuentra determinada por la fórmula (Freeland, 2005):

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

Una población que se encuentra en equilibrio de Hardy – Weinberg presenta ciertas características: la especie es diploide, presenta reproducción sexual, las generaciones son discretas, hay panmixia, es decir, la unión de los gametos se produce al azar, el tamaño de la población es grande, no hay migración, no hay mutaciones o su probabilidad es baja y no se ve

afectada por selección natural (Freeland, 2005; Hamilton, 2009). Mientras estas características se cumplan los valores de frecuencias alélicas se mantendrán constantes a lo largo de las generaciones (Hamilton, 2009).

Desviaciones de equilibrio de Hardy y Weinberg pueden deberse a selección, migración, subdivisión poblacional no detectada (Efecto Wahlund) o simplemente error en el muestreo (Nei y Kumar, 2000).

El Efecto Wahlund consiste en que la población muestreada presenta un exceso de homocigotos. Debido a que en ciertas ocasiones es difícil determinar los límites entre una población y otra, al momento de realizar el muestro, individuos de dos poblaciones diferentes son clasificados como una sola. Si las dos poblaciones presentan frecuencias alélicas diferentes se observará una mayor proporción de individuos homocigotos lo que podría llevar a asumir que la población no se encuentra en equilibrio de Hardy-Weinberg (Freeland, 2005).

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

2.1.3.6 Equilibrio y desequilibrio de ligamiento

El Equilibrio de ligamiento consiste en que los alelos de loci diferentes se segregan de manera independiente durante la reproducción, mientras que el desequilibrio de ligamiento implica una asociación no al azar entre los alelos de loci diferentes (Freeland, 2005).

El desequilibrio de ligamiento mide las diferencias observadas entre las frecuencias gaméticas observadas y esperadas. En individuos con reproducción asexual o clonal los valores de desequilibrio de ligamiento son mayores que en aquellos individuos que presentan reproducción sexual. Esto se debe a que en estos últimos ocurre el proceso de recombinación el cual rompe el desequilibrio de ligamiento (Hamilton, 2009)

2.2 Tricohialina

La Tricohialina (TCHH) es una proteína que se encuentra principalmente en la vaina radicular interna (IRS) del folículo piloso y en la médula del pelo, donde participa en la formación y desarrollo de la fibra o pelo (O'Keefe y col, 1993) (Figura 3).

Esta proteína es rica en residuos de arginina, glutamina/ácido glutámico, leucina y lisina, y su tamaño varía de acuerdo a la especie (O'Keefe y col, 1993). Presenta dos modificaciones post-traduccionales: los residuos de arginina son transformados en citrulina por acción de la enzima peptidil arginina deaminasa y además esta proteína es un sustrato de la enzima transglutaminasa que forma enlaces isopeptídicos de tipo ϵ (γ -glutamil) lisina produciendo filamentos similares a los de tipo intermedio (Alibardi, 2004).

La proteína varía considerablemente de tamaño de una especie a otra en el rango de 190 a 220 kD. Esta proteína se encuentra formada por un número variable de repeticiones de 23 aminoácidos los cuales no se encuentran distribuidos de forma homogénea a lo largo del gen. Sin embargo se ha podido utilizar anticuerpos policlonales antitricohialina de origen ovino para identificar la proteína en otras especies indicando la conservación de la estructura dimensional

de la proteína. Este anticuerpo se une a la región C-terminal de la proteína que presenta una secuencia muy conservada (Rogers y col, 1991).

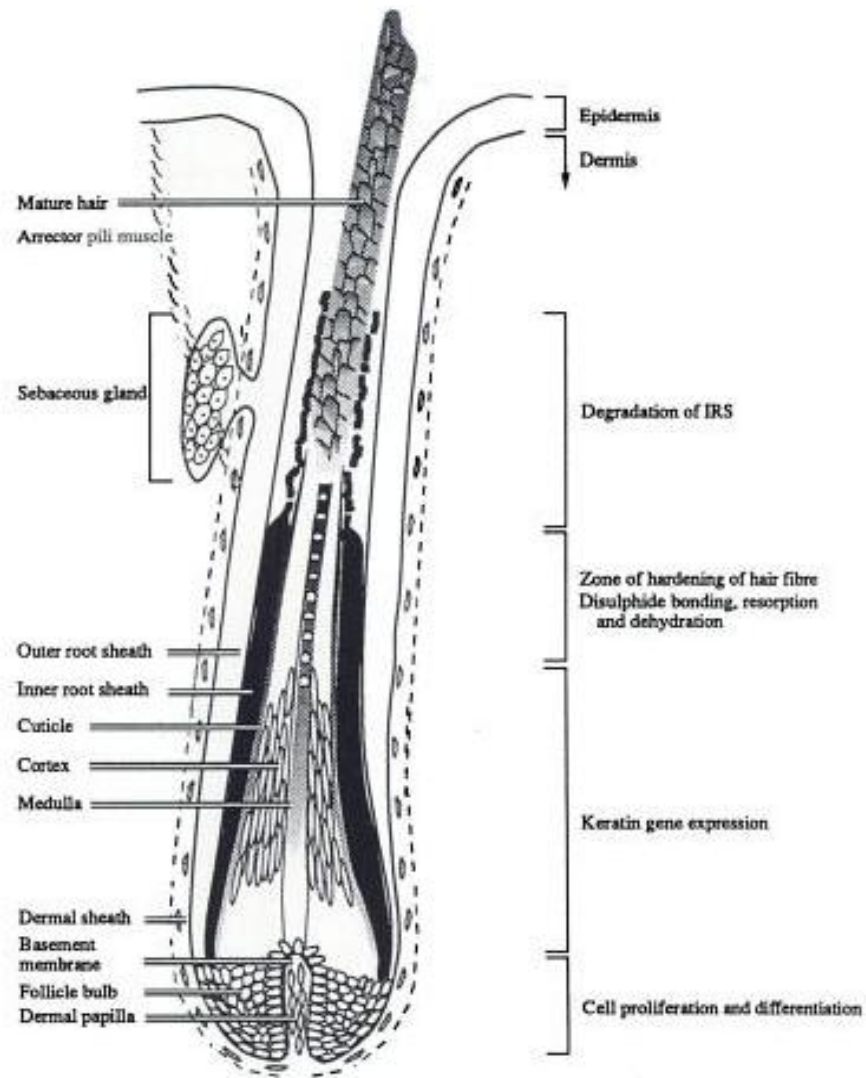


Figura 3. Representación esquemática del folículo del pelo identificando las diferentes estructuras que participan en su desarrollo (Powell y Roger, 1997) En la grafica se observa el IRS (Inner root sheath) y la médula (medulla) donde se encuentra la Tricohialina.

El gen de trichialina es un gen de copia única, con bajo contenido de timina, de aproximadamente 9 kb y un mRNA de aproximadamente 6.9 kb en humanos. Presenta 2 intrones y 3 exones; en humanos, el exón 1 contiene alrededor de 54 pares de bases y no se transcribe, el exón 2 contiene alrededor de 169 pares de bases y en él se encuentra el codón de inicio y por último el exón 3 contiene aproximadamente 6609 pares de bases, 5553 de la región codante incluyendo y el resto pertenece a la región 3' no codante (Lee y col, 1993).

2.2.1 El crecimiento del pelo y la Trichialina

El desarrollo del pelo es un proceso cíclico conformado por 3 fases: Anagen, Catagen y Telogen. La fase Anagen, es la fase de crecimiento. Las células del folículo del pelo crecen hacia la base formando la matriz de células que proliferarán y se diferenciarán originando distintos tipos de células que conforman el folículo en desarrollo, como el tallo del pelo y vaina radicular interna (IRS). La fase Catagen es la fase de regresión, el folículo se separa de la base y la fase Telogen es la fase de reposo (Shimomura y Christiano, 2010).

El IRS es un cilindro de células que rodean la fibra en desarrollo. Presenta una función de tipo estructural de soporte y dirección de la formación de la nueva fibra. Las células del IRS se encuentran en la zona periférica del bulbo del folículo piloso, estas células maduran, endureciéndose y formando una estructura rígida, mientras se desplazan hacia la parte superior del folículo degenerándose antes de llegar a la superficie por un proceso aún desconocido. En estadios tempranos del desarrollo de las células del IRS se observa la aparición de gránulos de TCHH lo cuales se encuentran cercanamente relacionados a la formación de filamentos de 8-10 nm de diámetro que aparecen conforme las células del IRS van migrando hacia la superficie.

Eventualmente estos gránulos de trichialina desaparecen conforme los filamentos formados se van alineando paralelos a la fibra en crecimiento (Fietz 1990).

La trichialina también se encuentra presente en la médula de la fibra en donde pasa por un proceso de maduración similar al que ocurre en las células IRS. Los gránulos de TCHH presentes en la zona medular no forman filamentos sino se encuentran como masas amorfas de proteínas las cuales, al separarse la fibra de la papila dermal, coalescen formando el interior duro de la médula (Fietz y col, 1990). Existen 2 tipos de fibra: la fibra fina o no medulada y la fibra medulada, que suele ser más gruesa. La fibra medulada, como su nombre lo indica, presenta una médula en la zona central la cual está hecha de Trichialina (Powel y Rogers, 1997) (Figura 4).

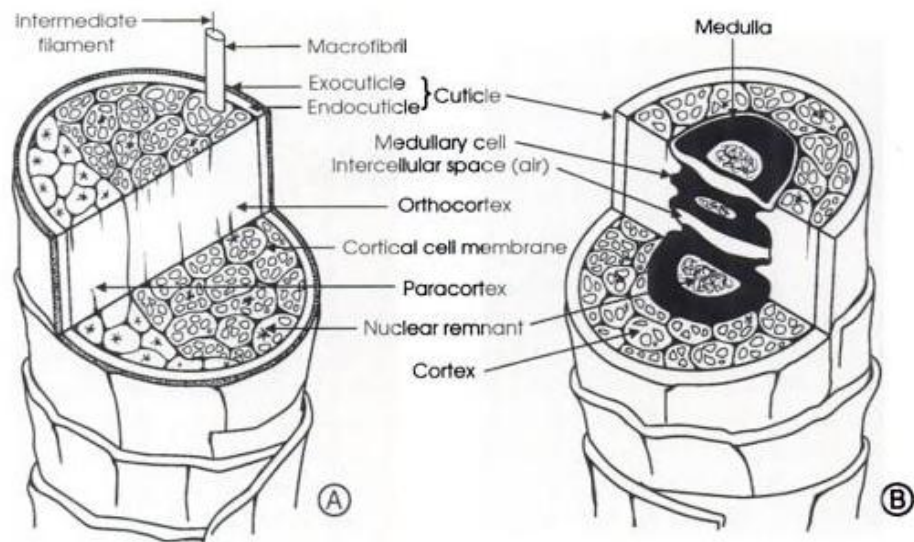


Figura 4. Diagrama de las estructuras celulares que conforman el pelo. A. Una fibra fina. B. Una fibra medulada. En la medula de la fibra se encuentra la Trichialina. (Powell y Rogers 1997)

En humanos, los estudios señalan una asociación entre la TCHH y la morfología del pelo (lacio, ondeado o crespo). Se ha identificado la presencia de un QTL el cual determina el 6% de la varianza ($p=1.5 \times 10^{-31}$) encontrada para la morfología del pelo en poblaciones de origen europeo. Este QTL se encuentra en la región 1q21.3 donde se ubica el gen de TCHH en humanos (Medland, 2009). Esto indicaría una posible relación entre el gen de la Tricohialina y la calidad de la fibra en alpaca y la diferencia entre la fibra de las alpacas Suri o Huacaya.

3. Justificación del Estudio

El Perú produce anualmente 4.4 miles de toneladas de fibra de alpaca con un promedio de 1.7 Kg de fibra por animal lo que lo convierte en el mayor productor de fibra de alpaca en el mundo (MINAG, 2009). Con un precio de 5.51 soles por libra de fibra de alpaca el valor de este producto es mayor que el de fibra de llama (2.3 soles por libra de fibra) o de lana de oveja (1.82 soles por libra de lana) (MINAG, 2009) lo que señala a este recurso como un factor económico de importancia especialmente para el sector alpaquero de las zonas altoandinas del país.

A su vez, el Perú es el tercer país exportador de fibra de alpaca. En el 2012 el Perú exportó 44.81 millones de dólares americanos en fibra de alpaca (COMTRADE, 2012), a diciembre de ese mismo año el precio FOB de la fibra de alpaca fue de US\$ 13.29 dólares americanos por kilogramo de fibra (SIICEX, 2012).

Según el IV Censo Nacional Agropecuario realizado el 2012, el Perú cuenta con 3,685,516 alpacas, lo que equivale a un aumento del 50.2% en el número de individuos en relación al censo anterior realizado en 1994. De estos individuos casi el 100% se concentra en la Sierra, donde la mayor proporción de alpacas pertenece a la raza Huacaya (78.9%) (INEI, 2013).

Más de 15 mil familias pertenecientes a comunidades campesinas ubicadas en departamentos en situación de pobreza o extrema pobreza dependen de la ganadería de camélidos, pues esta genera entre el 70% y el 80 % de su ingreso anual. (MINAGRI, 2013)

En ausencia de un genoma completo de alpaca que nos permita hacer una búsqueda global de información, la generación de una gran cantidad de marcadores genéticos permitiría establecer asociaciones de tipo genotipo-fenotipo con características de producción como finura de fibra, producción de carne o tipo de fibra.

La identificación de marcadores genéticos de importancia para los productos generados por la ganadería de alpacas brindará datos de utilidad para establecer mejores criterios de crianza y desarrollar mejores programas de manejo y empadre de animales que permitan mejorar la calidad de los productos y con ello el precio que recibe el productor.

III OBJETIVOS

1. Objetivo General

Generar nuevos marcadores genéticos, microsatélites y SNP, en alpacas (*Vicugna pacos*)

2. Objetivos Específicos

Evaluar 61 marcadores microsatélites en alpacas.

Caracterizar los nuevos marcadores microsatélites generados.

Evaluar la capacidad de multiplexado de los marcadores microsatelites generados.

Evaluar la presencia de SNP en un fragmento del gen de Trichialina.

Caracterizar los SNP encontrados en una población de alpacas.

Evaluar los marcadores SNP generados en alpacas de la raza Suri y Huacaya

IV. METODOLOGÍA

1. Caracterización de marcadores microsatélites

1.1 Localización del estudio

El presente estudio se realizó en las instalaciones de la Unidad de Biotecnología Molecular de la Universidad Peruana Cayetano Heredia (UBM-UPCH) ubicado en el distrito de San Martín de Porres, provincia de Lima, departamento de Lima.

1.2 Muestras de Animales

Se trabajó con muestras de ADN de alpacas (*Vicugna pacos*) pertenecientes a la Unidad de Biotecnología Molecular de la Universidad Peruana Cayetano Heredia (Anexo 1). Se emplearon un total de 35 muestras de ADN de alpacas Huacaya color blanco. De este total 30 fueron no emparentadas provenientes de 6 localidades (Anexo 2): i) Unidad de Producción de Cochas - SAIS Tupac Amaru, distrito de Canchayllo, provincia de Jauja, departamento de Junín, ii) Sector A de Carnicería - CC.CC. Santa Bárbara, distrito de Huancavelica, provincia de Huancavelica, departamento de Huancavelica, iii) Sector B de Carnicería - CC.CC. Santa Bárbara, distrito de Huancavelica, provincia de Huancavelica, departamento de Huancavelica, iv) Sector Pallcapampa - CC.CC. Santa Bárbara, distrito de Huancavelica, provincia de Huancavelica, departamento de Huancavelica, v) Sector de Lachocc - Centro de Investigación de Camélidos Sudamericanos Lachocc – Universidad de Huancavelica, distrito de

Huancavelica, provincia de Huancavelica, departamento de Huancavelica, vi) Sector Malkini, distrito de Muñani, provincia de Melgar, departamento de Puno.

El número de individuos evaluados en cada departamento se encuentra en correlación con la distribución de los individuos en los departamentos (Tabla 1)

Tabla 1: Número de individuos utilizados por departamento geográfico para la caracterización de los microsatélites

Departamento	Distribución de las poblaciones de alpacas en el Perú (%)	# de individuos
Puno	58	15
Cusco	11.9	4
Huancavelica	11.4	4
Apurimac	2.9	2
Junin	1.9	2
Arequipa	8.1	3

Las tres muestras restantes pertenecieron a una familia de alpacas, padre, madre y cría, pertenecientes al fundo Malkini (Puno) cuyo parentesco fue determinado mediante una prueba de ADN.

1.3 Elección de loci microsatélites

Se seleccionaron 61 loci microsatélites a partir de los reportados Reed y Chaves (2008). La selección se basó en el tipo de repetición, loci microsatélites dinucleótidos de repetición CA y en la capacidad de multiplexado de los marcadores determinada en base al tamaño esperado del fragmento determinado por el autor. (Tabla 2).

Tabla 2. Lista de cebadores seleccionados en base a lo reportado por Reed y Chaves (2008)

	Forward	Reverse
MNA0199	CTCCTTCCTGAAACCACATTC	CCTTTGCACCTTCTTTCCAG
MNA0272	AAACCCCTCAAGCCCTTATG	CCCTAGGAGCAATGGTTCAG
MNA0314	GGACTGTGGTACGGGGAAC A	GCTGTCAGAGGTGGAGGTG
MNA0422	GCTTCTCCTCCTACCCCATC	TGAATGAACCAGGGTGTGTG
MNA0234	AACTTTCTGCTTTTGAAAGTTG	TTTCTCCCTTCCTCCCTTTC
MNA0291	TGAATGGATATAGAAAATGTGGTG	TGTGGCACATGGTAGGATTC
MNA0295	TGCACTCCTTCCTTGTTTCC	ATCACACACGCACATGCAC
MNA0310	TTGTGCTAAAACCCAGAAAGC	GCAGGATAATAAGCCCAATTTT
MNA0356	TCTCCTCTTCCATCCCTGTC	ATGACCCTTGTCCTCCATCTG
MNA0341	TTCTCAGGGCTTTCCATACAG	TTCGTAATCCCTTCCCATTC
MNA0236	GCTTGGGTCTTAGGGAGAGG	AATTGAGGGGTGACTGCAAG
MNA0409	TCATTCCCATATATTTTATTAGG	TAGACCCAAGGGGCTATCTC
MNA0227	CTTCGTCGTTTAGGGTCTGG	TTGCGAAAATCTTGATGATGG
MNA0414	CACCCCAAAGACCACACAG	TTGCCTACACCAATTCTCCTG
MNA0205	CATCAACAGATGAATGGATGAAG	TTCTTTAGGTCAATTCATGTTGC
MNA0351	GGGTATTCAAAACAGAGTTGCAG	AAGATGCCAATCAGGCAAAG
MNA0218	GTCGGCTCTCTGTCCTCTCC	AGATCCGGCCAACAAGTATG
MNA0366	ACCCACAAAGACAGCAGGAC	ATGGAGAGGGTCAGGTTTCC
MNA0235	AATCTTAAGGGGCACACAGC	ATTTGCGGACTGTTCTTTGC
MNA0247	TCTGAGTCCATCCATGTTGC	GTCCATCGACAGATGAGTGG
MNA0305	TGTAAGCCCTGACAGTGTGG	GGTCCCTTTGGCTATTTACTTG
MNA0379	TCCTCCAGACAATTCAGAAGC	TGGCACTTCTCTCCGGACTA
MNA0217	GCTCATCCCATGAAATGAC	CTCCCTCTGCAAGTGGAATG
MNA0318	TGGGCCATCTCTTAGCAGAC	AAGCAACTCCAGCTCTGTGC
MNA0286	TGGAACACATGGAACAGGAC	GCTTGGGTGCAGAAGGTATC
MNA0401	CCAGAGGCAATTTATCGTACATC	TCAGAAGGAATGGGTGTGAAC
MNA0242	CCAGCATCAGCACAGAATG	TGCCACAAAATGCAGTAAG
MNA0280	TATCCACAACAGCCAAGACG	ATGCCCTCAAGGTCAATCC
MNA0300	GCAAAGTCATGGAAGCAACC	TGCTGCAGTGGCATTATTTT
MNA0388	TGGGAAATGTTCCCTCTTTG	GTGGGGCAGGAAAACCTC
MNA0233	TCAGAGCCTAAGGAACCAATG	TCCTCAAACCTTGAGAGATTTTCAG
MNA0284	TGCAGATCTGTGAAAGAGAGC	CAACAAGCCATGAAGAGCAC
MNA0221	TAGGATGGGGACATGTTTGG	TCAGCCATTACACACTCAAGC
MNA0224	CTGAGACAAAGGCACACAGC	TTCAATTTCCCAAAGGAAGTC
MNA0301	AGCACAGTGGGAAACAAAAG	CTGCCTTTTGATTGGAGGTC
MNA0408	TGGAAACAACCCAAGTGTCC	GCCTAACACCCTCCAAGTCC
MNA0265	TGCACCTCAATGTTTACAGC	CTGCAAATGGCATTATTTTCG
MNA0377	CAGAACTTGCTACTGTTTCTTCA	CCTTGTTGGTAATCATTTCTCT
MNA0258	AAGATAAATGTGCCTCTCAGTTCAC	ATGTTGTCACAAACGGCAAG
MNA0329	TTAATGGACCATGACCTCCAG	CGAGATACCAGGAGCAAAGC
MNA0246	TCATAACCAGGAAATCACTGC	TGCCAAAGAAGAAAATTTGACC
MNA0328	TCCCTTTAGGGTGACTCCTG	TGGGCATTTTAAGCAATGAAG
MNA0357	GCTGCAAATGGCATTATTTT	AGCCAAGACATGGAAGCAAC

MNA0384	AACACTGCGCCTAACCAATC	CTTTCTGGCCCTCTTTCTCC
MNA0262	TAGAAGGAGGGTGCCCTGTA	GATGGCTTCAGGGAATCAGA
MNA0323	CCATGGGGCTTTCTGGTAG T	CAATTTCTGGCAAAGACC
MNA0208	CCAACCCTCCAGGTGAGTC	TGTGACTGTTGGAGGAGGAG
MNA0275	TTTGAGTCACGGATTCAGC	CAAAGGGACCAGTAATTTGTGAG
MNA0345	GAGAGCTAAGATTTGGAGTCAGTTC	CCCTGTTTTGATTGAGATGTTG
MNA0398	CATCCCTATTGGTGCAAATACC	CCCAATGTTCATAGCAGCAC
MNA0309	GCCTCTCTTGACTAAACCTTCC	TTACCAGCAGCCAGGACAG
MNA0391	GCAGATGGTAAATGGGATGG	TACCTGGTTCCCACTTCCAC
MNA0269	ACTCTCTAGCTCCATTACATTG	GAACCTCAGGCTTCATAGCAG
MNA0322	ATATTATGCATACCATACACACGAG	CACAGCCCCAGACAGCTAC
MNA0358	CACGTGGAATCACACTGCAC	TCTTCATCCAGCCTCTCCTG
MNA0394	GGCATTTACAAACCCAGAATAG	AGAACGCCAGAAAGCCATC
MNA0392	AAGTAACTCTCCTTGGGTCAGC	TGGATGTGATTGTGGACACC
MNA0337	TCCAAAGTGTGTTCCCAAATAAC	TCCAAATCTTGGGCTTTTCAG
MNA0294	AAGCCAGATACTAAAGAGATTTGC	CACGCTTGGTGTGTATGAGG
MNA0338	TGATCTAGACATTAGAGTGTGAAACG	GCCTGATCCTTAGTTGTTTAAGTG
MNA0238	GCCAAGACATGGAAGTGACC	TTCTACATAACTCGGTTTCATTATTG

1.4. Tamizado de loci microsatélites con capacidad de amplificación en alpacas

1.4.1 Amplificación de ADN microsatélite

Se amplificaron 61 loci microsatélites (Tabla 1) mediante reacción en cadena de la polimerasa (PCR) utilizando 2 muestras de ADN genómico de alpaca. La PCR se realizó siguiendo las especificaciones de Reed y Chaves (2008). Las condiciones empleadas fueron: un volumen final de 10 uL conteniendo 5 ng de ADN genómico, 1X Buffer PCR, 1.5 mM de Mg^{+2} , 0.2 mM de dNTPs, 0.5 U DNA polimerasa Hot Star Taq Plus (Qiagen), 5 pmoles de cada cebador.

Las condiciones de temperatura empleadas inicialmente fueron la siguientes: 1 ciclo de 95°C por 5 minutos, 35 ciclos de 94°C por 30 segundos, temperatura de alineamiento por 30 segundos, 72°C por 30 segundos y 1 ciclo de 72°C por 5 minutos. Se evaluaron distintas temperaturas de alineamiento en el rango de 60°C a 52°C, debido a que Reed y Chaves,

reportan como temperatura de alineamiento para los cebadores diseñados 58°C y 56°C. Se buscó determinar las condiciones óptimas para la PCR. La PCR se realizó en un termociclador modelo 9700 GeneAmp® (Applied Biosystem).

Los productos de PCR fueron separados mediante electroforesis en gel de agarosa al 2%, TBE 0.5X (Tris-borato 0.045 M y EDTA 0.001M) a 40 V por 90 minutos. La visualización de los fragmentos se realizó mediante tinción con una solución de Bromuro de Etidio (0.5 ug/mL) posterior a la electroforesis. La presencia de productos de amplificación determinó la capacidad de amplificación para cada uno de los marcadores seleccionados.

1.4.2 Selección de loci microsatélites para marcado con fluorocromos

Se seleccionaron 12 loci microsatélites para ser marcados con un fluorocromo en el extremo 5' del cebador (6FAM™, HEX™, TAMRA™). Estos marcadores fueron seleccionados en base a los siguientes criterios: (i) amplificación específica, (ii) amplificación a una temperatura de alineamiento de 58°C.

Se volvió a optimizar las condiciones de la PCR para los 12 microsatelites marcados empleando un termociclador modelo Veriti ® (Applied Biosystem®). Para ello se emplearon 3 muestras de ADN de una familia de alpacas, padre, madre y cría. Los productos de PCR fueron separados por electroforesis en capilar basada en fluorescencia empleando un analizador genético ABI PRISM 3130 Genetic Analyser® (Applied Biosystems). El tamaño de los alelos para cada loci microsatélite fue determinado mediante el programa Genemapper® v.4.0 (Applied Biosystems).

1.4.3 Caracterización de los loci microsatélites

Los loci microsatélites marcados con fluorocromos fueron amplificados en 30 muestras (Hale y col., 2012) de ADN genómico de alpaca de distintas localidades para determinar su informatividad. Se realizó una PCR en base a las condiciones previamente establecidas y se determinaron los valores de i) contenido de información polimórfica (PIC); ii) heterocigosidad (HE), iii) probabilidad de exclusión (PE), iv) número de alelos (A), para cada marcador.

1.5 Reacción de PCR múltiple

Seis loci microsatélites fueron amplificados en 2 reacciones de PCR múltiple utilizando Qiagen Multiplex PCR Kit® (Qiagen) y un termociclador Veriti® (Applied Biosystem®) siguiendo las condiciones de temperatura establecidas por el fabricante.

Los productos de PCR fueron separados en un Analizador Genético ABI 3130 DNA sequencers® (Applied Biosystems). El tamaño de los alelos para cada loci microsatélite fue determinado mediante el programa Genemapper® v.4.0 (Applied Biosystems).

2. Caracterización del gen de Trichialina (TCHH)

2.1 Muestras de Animales

Para la identificación de los SNP se emplearon 48 muestras de ADN genómico de alpaca Huacaya pertenecientes a la Unidad de Biotecnología Molecular (UBM-UPCH), provenientes de distintas poblaciones localizadas en 6 departamentos del Perú (Anexo 2). El número de individuos tomados para cada departamento se encuentra en proporción con la

distribución de la población de alpacas en el Perú. En la siguiente tabla se indica la localidad geográfica y el número de muestras usadas (Tabla 3).

Tabla 3: Número de individuos utilizados por departamento geográfico para la caracterización de los SNP

Departamento	Distribución de las poblaciones de alpacas en el Perú (%)	# de individuos
Puno	58	20
Cusco	11.9	9
Huancavelica	11.4	8
Apurímac	2.9	4
Junín	1.9	3
Arequipa	8.1	4

Para la comparación de los SNP encontrados entre alpacas de la raza Huacaya y la raza Suri, se emplearon 48 muestras de ADN genómico de alpaca Huacaya pertenecientes a la Unidad de Biotecnología Molecular (UBM-UPCH), provenientes de distintas poblaciones localizadas en 5 departamentos del Perú : Puno, Cuzco, Huacavelica, Apurimac y Junin; y 48 muestras de ADN genómico de alpaca Suri pertenecientes a la Unidad de Biotecnología Molecular (UBM-UPCH), provenientes de 2 poblaciones localizadas los departamentos de Cuzco y Puno.

2.2 Alineamiento de secuencias

Se emplearon 180 secuencias de cDNA de la biblioteca de piel de alpaca del UBM que presentaron un alto valor de identidad con el gen de tricohialina mediante el uso del programa BLAST.

Las secuencias encontradas fueron alineadas para formar un contig empleando el programa DNAbaser y el Mega5. Mediante el programa BLAST, el contig obtenido fue comparado con la información del genoma de la alpaca, publicada en el NCBI por el Proyecto de Secuenciamiento del Genoma completo de la alpaca del Brod Institute de la Universidad de Washington. El Proyecto de Secuenciamiento aún se encuentra en la fase de ensamblaje de los contigs y tiene una cobertura del genoma de 2X.

A su vez también se comparó el contig del gen de Tricohilina de alpaca con las secuencias de DNA genómico del gen de este mismo gen pertenecientes a humano, ratón (*Mus musculus*) y bovino (*Bos Taurus*) presentes en el GeneBank.

2.3 Diseño de cebadores

Se diseñaron cebadores específicos para amplificar un fragmento de aproximadamente 350bp dentro de la secuencia consenso obtenida en base al alineamiento anterior. Para ellos se empleó el programa Primer3 (<http://frodo.wi.mit.edu/>).

Mediante el programa Primer Blast (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>) se evaluaron los cebadores y su especificidad de amplificación en relación a todas las secuencias del genoma de la alpaca que se encuentran publicadas. El programa indicó que los cebadores deberían de amplificar un solo fragmento.

2.4 Amplificación del fragmento del gen de Tricohialina

Se determinaron las condiciones óptimas para los cebadores diseñados mediante el empleo de la técnica de reacción en cadena de la polimerasa (PCR) que permitió amplificar un fragmento del gen de Tricohialina.

La PCR se realizó con el kit Hot Star Taq® Plus DNA Polymerase (Qiagen®) y un termociclador modelo Veriti® (Applied Biosystem®).

Se identificó la temperatura de alineamiento óptima realizando una curva de temperatura de alineamiento desde 55°C hasta 65°C en intervalos de 2°C, y la concentración óptima de Magnesio realizando una curva de magnesio evaluando las concentraciones 2.0mM, 2.5mM y 3.0 mM.

Los productos de PCR fueron separados mediante una electroforesis en gel de agarosa al 1%, TBE 0.1X (Tris-borato 0.045 M y EDTA 0.001M) a 40 V por 90 minutos.

La visualización de los fragmentos se realizó mediante tinción con una solución de Bromuro de Etidio (0.5 ug/ mL) posterior a la electroforesis. La presencia de productos de amplificación determinó la capacidad de amplificación del marcador.

2.5 Purificación del fragmento amplificado del gen de Tricohialina

Se purificó los fragmentos de PCR obtenidos mediante la técnica de EXO-SAP. Se agregó 0.25U de Exonucleasa I (EXO) y 2.5U de fosfatasa alcalina de camarón (SAP) por cada 5ul de producto de PCR bajo las siguientes condiciones termales: 1 ciclo de 37°C por 15 minutos y 1 ciclo de 80°C por 15 minutos.

2.6 Secuenciamiento del fragmento amplificado del gen *Tricohialina*

Los productos de PCR fueron enviados a secuenciar a la empresa Macogen (Korea), donde fueron secuenciados usando ABI PRISM Big Dye Terminador Cycle Sequencing Ready Reaction Kit y un secuenciador automático ABI PRISM 3730 Genetic Analyser®

Los productos fueron secuenciados en ambos sentidos empleando los cebadores TchhF y TchhR.

2.7 Edición de Secuencias obtenidas del gen de *Tricohialina*.

Las secuencias obtenidas fueron editadas mediante los programas DNAbaser y Mega 5 (Tamura, 2011) a través de una inspección visual, comparación entre duplicados, alineamiento de todas las secuencias y la confirmación de QV mayores a 20 para los polimorfismos identificados. Se eliminaron segmentos iniciales y finales los cuales generan distorsión o “background” debido a su menor calidad.

2.8 Análisis de SNP de *Tricohialina*

Se determinó los valores de frecuencias alélicas, número de alelos por locus, heterocigocidad observada y esperada (Nei, 1973), contenido de información polimórfica (Botstein et al 1980), probabilidad de exclusión individual y acumuladas (Jamieson y Taylor 1997) y la probabilidad de identidad (Waits y col, 2001) empleando el programa Cervus v3.0 (Kalinowski y col, 2007)

Desviaciones del equilibrio de Hardy-Weinberg (deficiencia o exceso de heterocigotos) para cada loci y población en estudio fueron estimados mediante la prueba de U (Score test) y el método de Fisher, utilizando el algoritmo Cadena de Markov (MC) [dememorization:10 000, batches:100, itineration per batch: 10000] para estimar el valor de probabilidad (Guo y Thompson, 1992), disponible en el programa GENEPOPv4.0 (Raymond y Rousset, 1995)

Un análisis de desequilibrio de ligamiento genotípico entre pares de loci fue estimado mediante la prueba G exacta (exact G test) y el método de Fisher utilizando el algoritmo Cadena de Markov (MC) [dememorization:10 000, batches:100, itineration per batch: 10000] para estimar el valor de probabilidad (Guo y Thompson, 1992), disponible en el programa GENEPOPv4.0 (Raymond y Rousset, 1995)

2.9 Verificación de la presencia de SNP por Clonación

Se clonó el fragmento del gen de Tricohialina de 2 individuos: uno heterocigoto para los tres SNP y uno heterocigoto para la delección.

Se empleó el kit de clonación TOPO cloning vector de Invitrogen para células electrocompetentes siguiendo el protocolo del fabricante. Para cada individuo se generaron 3 placas de medio LB con ampicilina (100ng/ul) a las que se les aplicaron distintos volúmenes (15ul, 50ul, 80ul) del producto de la clonación. Estas placas fueron cultivadas a 37°C toda la noche.

Mediante la técnica de PCR se comprobó la presencia del fragmento del gen de Tricohialina en 10 colonias picadas al azar de las placas generadas por la clonación. Se emplearon los cebadores específicos para el fragmento insertado siguiendo las condiciones previamente definidas.

Se distinguió la presencia del fragmento realizando una electroforesis horizontal en geles de agarosa al 1.5%, buffer TBE 1X, a 90V por 30 minutos. Se tiñó el gel con una solución de Bromuro de Etidio (0.5 ug/ ml)

Se seleccionaron 5 colonias de aquellas que presentaron una amplificación positiva del fragmento de interés, estas fueron cultivadas a 37°C todo la noche en medio LB con ampicilina (100ng/ul). A partir de estos cultivos se extrajo ADN plasmídico mediante el kit Wizard Plus SV Miniprep kit de extracción (Promega) siguiendo el protocolo del fabricante.

Se verificó el éxito de la extracción mediante un electroforesis horizontal el geles de agarosa al 1%, TBE 1X , a 90V por 45 minutos. Se tiño el gel en una solución de Bromuro de Etidio (0.5 ug/ml)

El producto obtenido fue enviado a secuenciar a la empresa Macrogen (Korea) empleando los cebadores M13. Las secuencias obtenidas fueron analizadas mediante los programas Mega y DNAbaser (Kumar y col, 2012).

2.10 Análisis del fragmento amplificado del gen de Tricohialina en alpacas Suri y Huacaya

Se amplificó el fragmento de Tricohialina para 48 alpacas de la raza Suri y 48 de la raza Huacaya. Se comprobó la presencia del producto de amplificación mediante electroforesis horizontal bajo las condiciones anteriormente establecidas y los productos de PCR fueron enviados a secuenciar a la empresa Macrogen en Korea. El análisis de las secuencia se realizó empleando los programas Mega 5 (Tamura, 2011), Cervus (Kalinowski et al, 2007) y Genepop(Raymond y Rousset, 1995).

Se empleó el programa Arlequin para definir los haplotipos y se realizó un análisis estadístico empleando la prueba de chi-cuadrado y un Análisis Factorial de Correspondencia (AFC) empleando el programa Genetix (Belkhir y col, 2004)

V. RESULTADOS

Microsatélites

De los 61 loci microsatélites evaluados se logró amplificar 60 de ellos a diferentes temperaturas de alineamiento: 15 loci amplificaron a una temperatura de alineamiento de 58°C, 13 loci amplificaron a 56°C, 27 loci amplificaron a 54°C y 5 loci amplificaron a una temperatura de alineamiento de 52°C. (Tabla 4)

Tabla 4: Loci microsatélites que presentaron amplificación separados de acuerdo a su temperatura de alineamiento.

Temperatura de alineamiento					No amplificó
58°C	56°C	54°C		52°C	
MNA0295	MNA0234	MNA0199	MNA0246	MNA0236	MNA0280
MNA0341	MNA0291	MNA0272	MNA0262	MNA0409	
MNA0351	MNA0414	MNA0314	MNA0323	MNA0221	
MNA0218	MNA0217	MNA0422	MNA0275	MNA0384	
MNA0366	MNA0318	MNA0310	MNA0345	MNA0269	
MNA0235	MNA0286	MNA0356	MNA0398		
MNA0247	MNA0224	MNA0227	MNA0309		
MNA0242	MNA0301	MNA0205	MNA0391		
MNA0300	MNA0265	MNA0305	MNA0358		
MNA0388	MNA0258	MNA0379	MNA0337		
MNA0233	MNA0328	MNA0401	MNA0294		
MNA0329	MNA0208	MNA0284	MNA0338		
MNA0357	MNA0392	MNA0408	MNA0238		
MNA0322		MNA0377			
MNA0394					

Las condiciones de amplificación establecidas fueron las siguientes:

Se trabajó con un volumen final de 10 μ l conteniendo 10ng de ADN genómico, 1X Buffer PCR, 2.0 mM de Mg^{+2} , 0.2 mM de dNTPs, 0.5 U DNA polimerasa Hot Star Taq Plus (Qiagen), 5 pmoles de cada cebador.

Las condiciones termales empleadas fueron: 1 ciclo de 95°C por 5 minutos, 35 ciclos de 95°C por 30 segundos, temperatura de alineamiento por 30 segundos, 72°C por 30 segundos y 1 ciclo de 72°C por 5 minutos.

Se seleccionaron 12 loci microsatélites que cumplieron con los criterios de selección para ser marcados con fluorocromos: MNA0295, MNA0341, MNA0251, MNA0218, MNA0366, MNA0235, MNA0247, MNA0242, MNA0388, MNA0233, MNA0357 y MNA0394.

Los doce loci microsatélites seleccionados fueron alineados de acuerdo al tamaño del fragmento determinado por el autor (Reed y Chaves, 2008) y se solicitaron los cebadores marcados con los fluorocromos HEX, FAM y TAMRA de manera que pudieran ser amplificados en una reacción múltiple. (Tabla 5)

Tabla 5: Loci microsatélite y tipo de fluorocromo

Loci	Tamaño del fragmento (pb)	Fluorocromo
MNA0295	110	FAM™
MNA0341	113	HEX™
MNA0351	140	TAMRA™
MNA0218	141	HEX™
MNA0366	141	FAM™
MNA0235	142	FAM™
MNA0247	142	HEX™
MNA0242	173	TAMRA™
MNA0388	173	HEX™
MNA0233	176	FAM™
MNA0357	191	HEX™
MNA0394	200	TAMRA™

Las condiciones de amplificación establecidas para los cebadores marcados fueron las siguientes:

Se trabajó con un volumen final de 10 ul conteniendo 5ng de ADN genómico, 1X Buffer PCR, 2.0 mM de Mg^{+2} , 0.2 mM de dNTPs, 0.5 U DNA polimerasa Maxima HS Taq(Fermentas), 5 pmoles de cada cebador para aquellos loci marcados con FAM y HEX y 7 pmol de cada cebador para aquellos loci marcados con TAMRA.

Las condiciones de temperatura empleadas fueron: 1 ciclo de 95°C por 4 minutos, 27 ciclos de: 95°C por 30 segundos, temperatura de alineamiento por 30 segundos, 72°C por 30 segundos.

Seis loci microsatélites fueron excluidos debido a distintos problemas al visualizar los picos en el secuenciador: MNA0341, MNA0242, MNA0247 y MNA0357 presentaron una morfología de pico diferente de la esperada para un locus microsatélite dinucleotido, lo que imposibilitaba la identificación de los alelos presentes, MNA0235 presento un alelo nulo y MNA0233 al analizarse varias muestras para este marcador, se observó que la morfología del pico era inconstante por lo que fue eliminado.

Para los 6 loci microsatélites restantes se estimaron los valores de Heterocigocidad observada, Heterocigocidad esperada, PIC y Probabilidad de Exclusión (Tabla 6) y equilibrio de Hardy-Weinberg.

Tres de loci microsatélites se encontraron en equilibrio de Hardy-Weinberg: MNA0218, MNA0366 y MNA0351. De los otros 3 marcadores MNA0394 presentó un déficit de heterocigotos (p-value:0.0037) y MNA0295 y MNA0388 presentaron un exceso de heterocigotos (p-value:0.0472, p-value:0.0041).

Se realizó un análisis de desequilibrio de ligamiento genotípico pareado para los loci microsatélites evaluados y ninguno de ellos presentó desequilibrio de ligamiento (p-value > 0.05).

Tabla 6: Estimadores de información de los loci microsatélites.

Locus	A	Rango (pb)	HObs	HExp	PIC	PE1	PE2	PE3	P(ID)
MNA0295	6	102 -116	0.9	0.781	0.731	0.373	0.552	0.736	0.91
MNA0218	9	118-158	0.724	0.827	0.793	0.472	0.647	0.833	0.945
MNA0366	6	131-147	0.786	0.749	0.701	0.34	0.522	0.716	0.896
MNA0388	8	157-179	0.933	0.763	0.711	0.354	0.531	0.718	0.899
MNA0351	12	135-183	0.857	0.849	0.816	0.51	0.678	0.855	0.954
MNA0394	9	157-201	0.586	0.691	0.659	0.3	0.492	0.709	0.876
Promedio	8.3	-	0.798	0.777	0.735	-	-	-	-
PE total	-	-	-	-	-	0.9516	0.9942	0.9998	0.9999

HObs: Heterocigocidad Observada

HExp: Heterocigocidad Esperadas

PIC: Índice de contenido polimórfico

PE1: Probabilidad de exclusión para un candidato a pariente

PE2: Probabilidad de exclusión para un candidato a pariente dado el genotipo de un pariente conocido del sexo opuesto

PE3: Probabilidad de exclusión para un par de candidato a parientes

P(ID): Probabilidad de Identidad

Los 6 loci microsatélites identificados se amplificaron en dos reacciones de PCR múltiple bajo (Tabla 7) las siguientes condiciones: 20ng de ADN genómico, 1X Buffer PCR, 3.0 mM de $MgCl_2$, 0.2 mM de dNTPs, 1.0 U DNA polimerasa Hot Star Taq (Quiagen) 2 pmoles de cada cebador para aquellos loci marcados con FAM y HEX y 3 pmol de cada cebador para aquellos loci marcados con TAMRA.

Las condiciones termales empleadas fueron: 1 ciclo de 95°C por 15 minutos, 25 ciclos de: 94°C por 30 segundos, 58°C por 90 segundos, 72°C por 60 segundos y 1 ciclo de 60°C por 30 minutos.

Tabla 7: Paneles de PCR múltiple para los loci microsatélites

Panel	Loci	Tamaño del fragmeto (pb)	Fluorocromo
1	MNA0351	135 – 183	TAMRA™
	MNA0218	118 – 158	HEX™
	MNA0366	131 – 147	FAM™
2	MNA0295	102 – 116	FAM™
	MNA0388	157 – 183	HEX™
	MNA0394	157 – 201	TAMRA™

Tricohialina

Alineamiento de secuencias

La secuencia de Tricohialina, obtenida a partir de la información de la biblioteca de piel de alpaca preparada en la Unidad de Biotecnología Molecular – UPCH, al ser comparada con el genoma de la alpaca publicado alineó al extremo del contig: Vicugna pacos contig 1.390472, whole genome shotgun sequence (e-value: 0)

El contig fue comparado con las secuencias publicadas en el GenBank mediante un el programa Blast. Se obtuvo que la secuencia con la que presenta mayor similitud pertenece a un mRNA de piel de cerdo (*Sus scrofa*) (e-value:2e-71) seguido del mRNA de Tricohialina de vacuno (*Bos taurus*) (e-value:1e-47).

Diseño de cebadores

Se alineó la secuencia de interés con el contig 1.390472 y se diseñaron cebadores para amplificar un fragmento (490bp) que cubriera toda la secuencia de interés. (Figura 5) (Tabla 8)

Tabla 8: Secuencia de los cebadores diseñados para Tricohialina.

	Inicio	Largo	Tm	GC%	Secuencia
Cebador izquierdo	2072	20	60.25	55	CAGGTCCCCACTCTGCTAAA
Cebador derecho	2562	20	60.14	55	TTTGCTCTCTGGTGCAGATG

Las condiciones de amplificación establecidas fueron las siguientes:

Se trabajó con un volumen final de 10 μ l conteniendo 10ng de ADN genómico, 1X Buffer PCR, 3.0 mM de Mg^{+2} , 0.2 mM de dNTPs, 0.5 U DNA polimerasa Hot Star Taq Plus (Qiagen), 2 pmoles de cada cebador.

Las condiciones de temperaturas empleadas fueron: 1 ciclo de 95°C por 5 minutos, 32 ciclos de 95°C por 30 segundos, 65°C por 30 segundos, 72°C por 30 segundos y 1 ciclo de 72°C por 5 minutos.

Producto de amplificación

Se logró amplificar un fragmento de 465bp en el cual se identificaron 3 SNP y una posible delección (Figura 6). Los SNP se presentaron en la posición A57G, A227G, A458G y los tres presentan una transición adenina-guanina. La posible delección se encuentra en la posición 72 y comprende la presencia o ausencia de una Adenina dentro de un bloque de repeticiones de 9 adeninas.

TGCTAAAATTAACAAAGGTAGTACTATTCTCTAGCTCATGAAAGACTGGGGGAAGA	<u>A</u> AAAC	60
TGAGCACTACAGACTACAGCAGACACAGTTTTAAAGTTTGACCTCTAGAGTTCTTTGAGA		120
AACACCTTACTTTTGCTTCATGAGGTTTTAAATGTTTATGAAGTTAAGTGT	<u>AAAAAAAAAA</u>	180
TCGAATTCAAATTCTAAGCCTTCAAGCTAAACCTTATTCTTAACAG	<u>A</u> TATTTATTCCAGT	240
TCTGGGCCGATTTACGAGGGGCTCTTATGGCGAAGCCATTCCCCCTGCACTATTCAAAG		300
GCTGTTGTAAAAATTCAGTCCCCTCACAGAATCTGCTAATTATGCTACAAGAAGTGAGAG		360
GCTATTTGGAGCTAAATATTAATAAAGCAATAAATGAACGTGTTCCCCAAACAAGACTCC		420
TCAAATTAAATGGCTTATGGTTTTAAAAATGCACACC	<u>G</u> CATCTGC	465

Figura 6. Secuencia obtenida a partir del alineamiento de las secuencia de Tricohialina y la posición de los SNP encontrados. Los SNP se encuentran señaladas en color rojo y la fila de Adeninas se encuentra subrayada.

Se detallan los siguientes estimadores del nivel de información del marcador: frecuencias alélicas ($f(a)$) (Tabla 9), número de alelos, heterocigocidad observada (HObs), heterocigocidad esperada (HEsp), índice de contenido polimórfico (PIC) y probabilidad de exclusión (PE) (Tabla 10).

Tabla 9: Frecuencias alélicas para los marcadores SNP presentes en el gen de Tricohialina. $f(a)$ es la frecuencia alélica

Loci	Alelo	$f(a)$
A57G	A	0.527
	G	0.473
A227G	A	0.9583
	G	0.0417
A458G	A	0.5303
	G	0.4697

Tabla 10. Estimadores de información de los SNP identificados para el gen de Trichialina.

Locus	Nº de alelos	n	HObs	HExp	PIC	PE1	PE2	PE3	P(ID)
A57G	2	37	0.351	0.505	0.374	0.124	0.187	0.281	0.624
A227G	2	48	0.083	0.081	0.077	0.003	0.038	0.072	0.15
A458G	2	33	0.394	0.506	0.374	0.124	0.187	0.281	0.624

n: numero muestral

HObs: Heterocigocidad Observada

HExp: Heterocigocidad Esperadas

PIC: Indice de contenido polimórfico

PE1: Probabilidad de exclusión para un candidato a pariente

PE2: Probabilidad de exclusión para un candidato a pariente dado el genotipo de un pariente conocido del sexo opuesto

PE3: Probabilidad de exclusión para un par de candidato a parientes

P(ID): Probabilidad de identidad

El análisis individual de los marcadores SNP indicó que estos se encontraron en equilibrio de Hardy y Weinberg ($p > 0.05$) para la población muestreada. Sin embargo un análisis global de todos los locus determinó que no se encuentran en equilibrio de Hardy y Weinberg, sino un déficit de heterocigotos (p-value: 0.0154)

El análisis de disequilibrio de ligamiento genotípico entre pares de loci indicó que el marcador A227G es independiente de los marcadores A57G (p-value:0.768)) y A458G (p-value:0.606). Sin embargo A57G y A458G presentan disequilibrio de ligamiento genotípico (p-value:0).

La clonación del fragmento amplificado de un individuo heterocigoto para los 3 SNP proveyó el fragmento completo y confirmó la presencia de los 3 SNP. El fragmento completo obtenido a partir de la clonas fue comparado con el obtenido a partir del alineamiento de las 48 secuencias y con la secuencia original. (Figura 7)

Una vez confirmada la secuencia obtenida, ésta fue comparada con la secuencia del gen de Trichialina reportado para humanos por Lee y col en 1993; la secuencia alineó en la zona 3' UTR del gen.

ALINEAMIENTOTCH	-----TGCTAAAATTAACAAAGGTAGTACTATTCTCTAGCTCATGAAAGACT	47
CLONASTCH	CAGGTCCCCACTCTGCTAAAATTAACAAAGGTAGTACTATTCTCTAGCTCATGAAAGACT	60

ALINEAMIENTOTCH	GGGGGAAGAAAAC TGAGCACTACAGACTACAGCAGACACAGTTTTAAAGTTTGACCTCTA	107
CLONASTCH	GGGGGAAGAAAAC TGAGCACTACAGACTACAGCAGACACAGTTTTAAAGTTTGACCTCTA	120

ALINEAMIENTOTCH	GAGTTCTTTGAGAAACACCTTACTTTTGCTTCATGAGGTTTTAAATGTTTATGAAGTTAA	167
CLONASTCH	GAGTTCTTTGAGAAACACCTTACTTTTGCTTCATGAGGTTTTAAATGTTTATGAAGTTAA	180

ALINEAMIENTOTCH	GTGTAAAAAAAATCGAATTCAAATTC TAAGCCTTCAAGCTAAACCTTATTCTTAACAGA	227
CLONASTCH	GTGTAAAAAAAATCGAATTCAAATTC TAAGCCTTCAAGCTAAACCTTATTCTTAACAGA	240

ALINEAMIENTOTCH	TATTTATTCCAGTTCTGGGCCGATTTACGAGGGGCTCTTATGGCGAAGCCATTCCCCCT	287
CLONASTCH	TATTTATTCCAGTTCTGGGCCGATTTACGAGGGGCTCTTATGGCGAAGCCATTCCCCCT	300

ALINEAMIENTOTCH	GCACTATTCAAAGGCTGTTGTAAAAATTCAGTCCCCTCACAGAATCTGCTAATTATGCTA	347
CLONASTCH	GCACTATTCAAAGGCTGTTGTAAAAATTCAGTCCCCTCACAGAATCTGCTAATTATGCTA	360

ALINEAMIENTOTCH	CAAGAAGTGAGAGGCTATTTGGAGCTAAATATTAATAAAGCAATAAATGAACGTGTTCCC	407
CLONASTCH	CAAGAAGTGAGAGGCTATTTGGAGCTAAATATTAATAAAGCAATAAATGAACGTGTTCCC	420

ALINEAMIENTOTCH	CAAACAAGACTCCTCAAATTAATGGCTTATGGTTTTAAAAATGCACACCGCATCTGC --	465
CLONASTCH	CAAACAAGACTCCTCAAATTAATGGCTTATGGTTTTAAAAATGCACACCGCATCTGCAC	480

ALINEAMIENTOTCH	-----	
CLONASTCH	CAGAGAGCAAA	491

Figura 7. Alineamiento de las secuencias de TCH obtenidas de la clonación vs la obtenida del alineamiento. La secuencia obtenida a partir de la clonación es: alineamientotch y la obtenida a partir de las clonaciones es: clonastch. Los asteriscos (*) indican que las bases son iguales. Los números al extremo derecho indican el tamaño en pares de bases.

Se identificaron los 3 marcadores SNP en las posiciones previamente establecidas tanto en los individuos de la raza Suri como en los de la raza Huacaya. Las frecuencias alélicas establecidas por raza se detallan en la tabla siguiente (Tabla 11):

Tabla 11: Frecuencias alélicas obtenidas de los marcadores SNP: A57G, A227G y A458G del gen de Tricohialina, en las poblaciones de Suri y Huacaya

Loci	Alelo	Frecuencia alélica	
		Suri	Huacaya
A57G	A	0.5857	0.5429
	G	0.4143	0.4571
A227G	A	0.9787	0.9583
	G	0.0213	0.0417
A458G	A	0.6296	0.6167
	G	0.3704	0.3833

Los marcadores mostraron encontrarse en equilibrio de Hardy Weinberg para ambos grupos ($p > 0.05$). Un análisis de equilibrio de Hardy Weinberg global indicó lo mismo ($p > 0.05$), salvo en el caso del marcadores A57G el cual parece presentar un déficit de heterocigotos ($p > 0.0450$).

El análisis de disequilibrio de ligamiento genotípico corroboró los resultados previamente obtenidos para ambos grupos, el marcador A227G es independiente de los otros

dos marcadores (A57G, p: 0.916; A458G, p:0.894) mientras que A57G Y A458G se encuentran en desequilibrio de ligamiento genotípico (p:infinito).

Las frecuencias genotípicas fueron establecidas para cada uno de los individuos por marcador SNP analizado. Se emplearon solo aquellos individuos en los que se pudo definir los genotipos durante el proceso de análisis de las secuencias (Tabla 12).

Tabla 12: Frecuencia genotípicas observadas de los marcadores SNP: A57G, A227G y A458G del gen de Tricohialina, en las poblaciones Suri y Huacaya

SNP	A57G		A227G		A458G	
	Suri	Huacaya	Suri	Huacaya	Suri	Huacaya
AA	13	13	45	44	11	13
AG	15	12	2	4	12	11
GG	7	10	0	0	4	6
Total	35	35	47	48	27	30

Se realizó una prueba de chi-cuadrado (χ^2) para definir si las frecuencias genotípicas (Tabla 12), para cada uno de los marcadores SNP evaluados, eran similares entre las poblaciones Suri y Huacaya. Para los 3 marcadores SNP se obtuvo que la frecuencia de los

genotipos de las poblaciones de Suri y Huacaya eran los mismos (A57G , $0.7 > p > 0.5$; A227G, $0.7 > p > 0.5$; A458G, $0.95 > p > 0.90$).

Se determinaron los haplotipos presentes en los dos grupos evaluados, Suri y Huacaya (Tabla 13). Debido a que para ciertos individuos fue imposible identificar algunos de los alelos el programa Arlequin determinó haplotipos incompletos lo cuales clasificó como diferentes (Tabla 13).

Tabla 13: Frecuencia de haplotipos de los marcadores SNP: A57G, A227G y A458G del gen de Trichialina para las poblaciones Suri y Huacaya.

Haplotipos	Huacaya	Suri
111	34	33
212	23	20
?1?	25	24
121	3	1
21?	9	9
11?	1	6
?2?	1	0
12?	0	1

Análisis anteriores señalan que los dos SNP de los extremos (A57G y A58G) se encuentran en desequilibrio de ligamiento genotípico, por lo que se puede inferir algunos de los

alelos faltantes en los haplotipos, lo que permite inferir los haplotipos completos de algunos individuos (Tabla 14). Se analizaron solo los 35 individuos cuyos haplotipos completos eran conocidos (Haplotipos : 111, 212 y 121), no fueron incluidos en el análisis los individuos con haplotipo ?1? debido a que fue imposible determinar si los alelos faltantes son Adeninas o Guaninas.

Tabla 14: Frecuencias Haplotípicas inferidas de los marcadores SNP: A57G, A227G y A458G del gen de Trichialina para las poblaciones Suri y Huacaya.

Haplotipos	Hucaya	Suri
111	35	39
212	32	29
121	3	2
?1?	24	24
?2?	1	-

Se comparó las frecuencias haplotípicas corregidas de ambos grupos, Suri y Huacaya, empleando una prueba de ji-cuadrado (χ^2) del total de individuos muestreados. Los resultados de la prueba señalan que ambas poblaciones presentan la misma frecuencia de haplotipos (p-value: $0.975 > p > 0.950$).

Se realizó un análisis factorial de correspondencia (AFC) para establecer si existe alguna diferenciación genética entre las poblaciones de Suri y Huacaya. En Análisis Factorial de Correspondencia de 2 dimensiones se observa la ausencia de un patrón de diferenciación entre los grupos Suri y Huacaya(Figura 8). En la Gráfica los 2 ejes explican el 26.54% y el 43.72% de la variabilidad analizada indicando que entre ellos ocupan el 70% de la variación observada en la muestra.

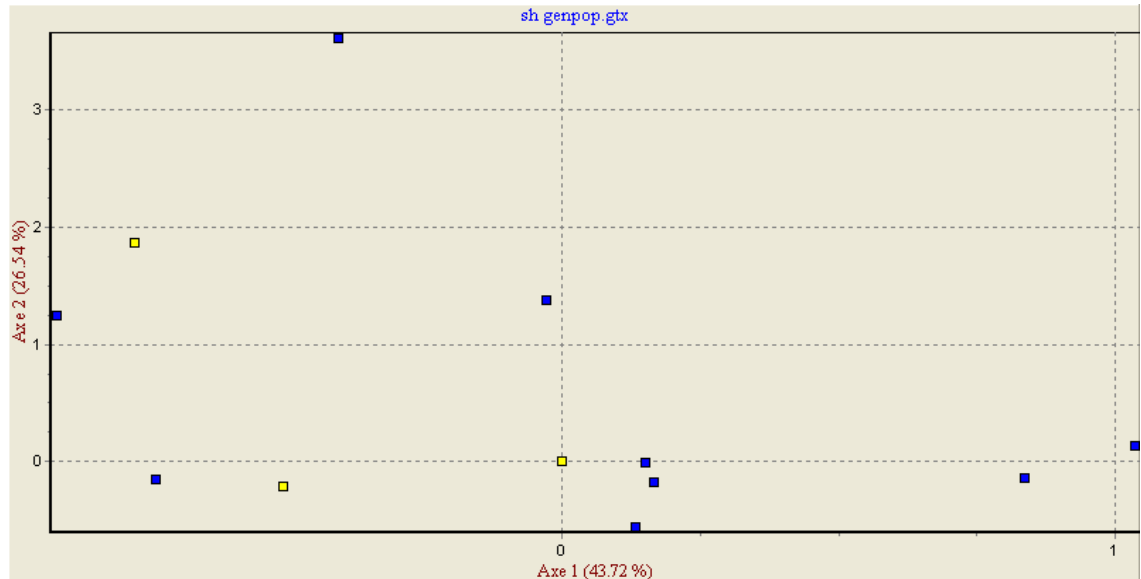


Figura 8: Análisis factorial de Correspondencia de 2 dimensiones las poblaciones de Suri y Huacaya. Cuadrados blancos alpacas Suri. Cuadrados negros alpacas Huacaya.

En la Figura 9 los 3 ejes explican el 26.54%, el 43.72% y el 20.68% de la variabilidad analizada indicando que entre ellos ocupan el 91% de la variación observada en la muestra. Aquí tampoco se observa diferenciación entre Suri y Huacaya.

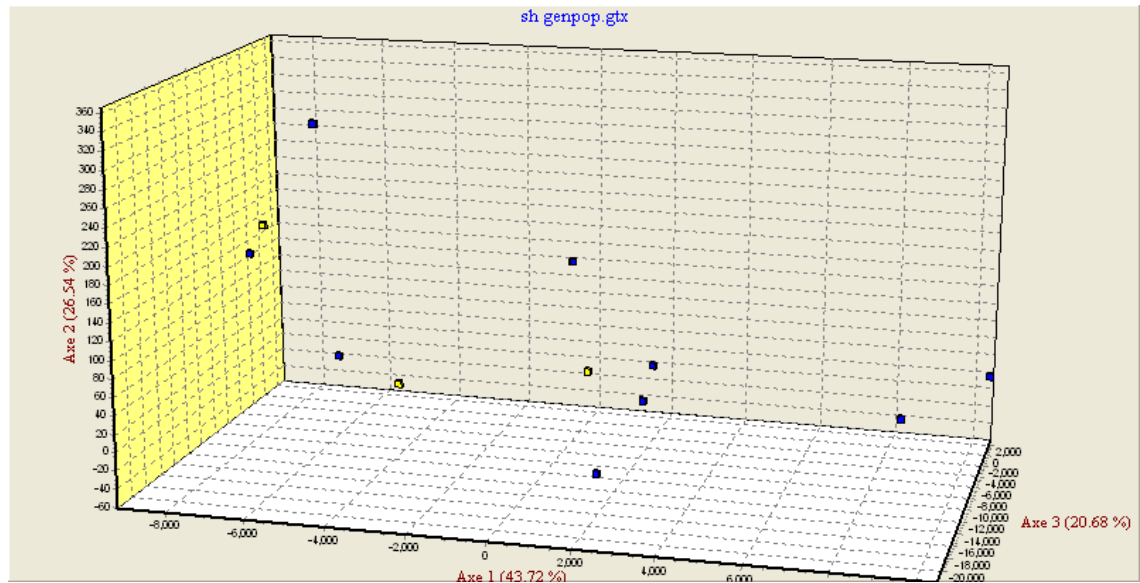


Figura 9: Análisis factorial de Correspondencia de 3 dimensiones las poblaciones de Suri y Huacaya. Cuadrados blancos alpacas Suri. Cuadrados negros alpacas Huacaya.

VI. DISCUSIÓN

Los marcadores genéticos se emplean en la construcción de mapas genéticos (Stauba y col, 1996), análisis de asociación (Hirschhorn y col, 2002), descubrimiento de nuevas drogas (Voiseg y Morris, 2008), pronóstico de cáncer (Petijean y col, 2007), para la identificación de individuos y sus interacciones entre ellos (Sunnucks, 2000) y la identificación de QTL (Seaton y col, 2002). Los marcadores genéticos permiten rastrear el patrón de segregación de características cuantitativas siendo en nuestro caso de particular interés aquellas de tipo comercial como la finura de fibra, su grosor o la coloración en alpacas puesto que estos rasgos le confieren valor económico al producto.

La fibra de alpaca es un producto comercial de importancia socioeconómica para el Perú ya que cuenta con más de 3 millones y medio de alpacas, las cuales casi un 100% se encuentran distribuidas a lo largo de los Andes Peruanos (INEI, 2013). Para las familias y comunidades campesinas de esta zona, la ganadería de camélidos sudamericanos es su principal fuente de ingreso (MINAGRI, 2013). En comparación a la fibra de llama o la lana de oveja, la fibra de alpaca presenta un mayor precio en el mercado (MINAG 2009) siendo el Perú el tercer país exportador de fibra de alpaca en el mundo.

El valor de la fibra de alpaca está dado principalmente por el diámetro de la fibra, a menor diámetro mayor valor tiene el producto. Esto se debe a que la sensación de confort que

ejercen los tejidos sobre la piel está determinada principalmente por el grosor de la fibra (Frank y col, 2006). En los últimos años se ha producido un descenso en la calidad de la fibra de alpaca debido principalmente a un mal criterio e inadecuado control en la selección y manejo reproductivo (Bonavia D, 1996). Por ello es necesario identificar marcadores genéticos que permitan rastrear aquellos genes que se encuentren relacionados a características de interés, como la finura de fibra, de manera que puedan generarse programas de mejoramiento genético que permitan aumentar la calidad del producto y con ello su precio.

Con esa finalidad, este proyecto buscó generar dos tipos de marcadores genéticos, microsatélites y SNP, que pudieran servir para rastrear la característica finura de fibra. Para la generación estos dos tipos de marcadores se emplearon dos aproximaciones diferentes.

En el caso de los marcadores microsatélites, estos fueron obtenidos a partir de una publicación (Reed y Chaves, 2008), en ella se encuentran los cebadores para loci microsatélites putativos generados a partir de la información existente en el Gene Bank. De los 61 loci seleccionados amplificaron 60, lo cual es acorde a lo esperado pues los cebadores diseñados son específicos para alpaca. Sin embargo, debido a que la temperatura de alineamiento a la cual funcionaron la mayoría de cebadores era baja (Tabla 4) y a que la caracterización de los loci microsatélite se realizó mediante un secuenciador automático en donde la temperatura de alineamiento de la reacción de PCR es importante para la especificidad del cebador; se seleccionaron aquellos marcadores con temperatura de alineamiento de 58°C y que además presentaran amplificación específica en geles de agarosa.

Doce loci microsatélite cumplieron estas condiciones. Estos loci fueron marcados con tres fluorocromos diferentes de acuerdo al tamaño de fragmento amplificado (pb) de manera que pudieran ser amplificados en reacciones de PCR múltiples (Tabla 5). De los 12 loci probados 6 fueron excluidos debido a problemas en la morfología del pico. MNA0341 presentó un patrón trialélico, tres picos de igual tamaño, lo cual imposibilita determinar cuál es el alelo real (Guichoux y col, 2011) MNA0242 y MNA0247 no presentaron la morfología de un marcador dinucleótido. MNA0357 fue inespecífico, se observaron varios picos. MNA0233 al ser evaluado en varias muestras presentó un cambio en la morfología del pico que impedía determinar el alelo. MNA0235 presentó un alelo nulo. Los alelos nulos ocurren por la presencia de una mutación en la zona flanqueante donde alinea el cebador, esto puede solucionarse rediseñando los cebadores empleados (Guichoux y col, 2011).

Los otros 6 loci microsatélites fueron probados en 30 muestras de ADN de alpaca de diferentes zonas del Perú. Los 6 marcadores presentaron un alto número de alelos (entre 6-12) al igual que valores altos de PIC (entre 0.659-0.816) y de Heterocigocidad esperada y observada (entre 0.691-0.849 y 0.586-0.9 respectivamente) (Tabla 6). Esto se asemeja a lo obtenido en otra especie doméstica, el cerdo del norte de la India donde también se observa un alto número de alelos (5-12), de PIC (PIC: 0.70-0.87) y de Heterocigocidad esperada y observada (0.75- 0.89 y 0.48-0.84 respectivamente) (Kaul y col, 2001). Al compararlo a los datos obtenidos en vacunos (*Bos taurus*) estos últimos presentan menor número de alelos (entre 3-10) y menores valores de PIC (PIC entre 0.16-0.77) (Vaiman y col, 1994). En el caso de las cabras (*Capra hircus*) un análisis de 22 marcadores microsatélites en 4 razas de importancia económica mostraron valores de número de alelos (entre 2 y 10), heterocigocidad esperada (entre 0.147-0.850) y PIC (entre 0.303 – 0.851) que abarcan desde valores altos hasta bajos, pudiendo llegar a ser mayores o menores que los datos obtenidos en este proyecto (Luikart y col, 1999). Esto indica que los

valores obtenidos para los 6 marcadores se encuentran dentro lo esperado en comparación con las demás especies domésticas.

De los 6 loci microsatelites probados 3 de ellos: MNA0218, MNA0366 y MNA0351 se encuentran en equilibrio de Hardy-Weinberg. Los otros 3 no se encontraron en equilibrio de Hardy – Weinberg dos de ellos MNA0295 y MNA0388 presentaron un exceso de heterocigotos (p-value: 0.0472 y p-value: 0.0041) mientras que MNA0394 presentó un déficit de ellos (p-value: 0.0037). Esto puede deberse a que el número muestral con el que se trabajó es muy pequeño (n: 30) o, en el caso de déficit de heterocigotos, a que las diferentes poblaciones con las que se trabajó presentan distintos valores de frecuencias alélicas de manera que se observaría una mayor proporción de homocigotos o efecto Wahlund (Freeland, 2005).

El análisis de desequilibrio de ligamiento genotípico determinó que los loci microsatélites eran independientes los unos de los otros, lo cual sugiere que se encuentran en regiones cromosómicas alejadas o en cromosomas diferentes. Esto concuerda con lo mencionado por Reed y Chaves en su artículo, donde indican que los marcadores con los que se trabajó en este proyecto se encontrarían en cromosomas diferentes. Esto lo determinaron comparando la secuencias de los loci microsatelites con el genoma bovino.

Los dos paneles de PCR múltiple que fueron probados amplificaron correctamente (Tabla 7). El uso de una PCR múltiple puede aumentar la capacidad de genotipificación mediante la reducción del trabajo de laboratorio y la disminución del uso reactivos, sin comprometer los resultados. Además se necesita una menor cantidad de ADN y se obtiene una

mayor cantidad de información por reacción (Guichoux y col 2011). Nuestro laboratorio ha generado un panel de 10 marcadores microsatélites multiplexados para pruebas de paternidad en alpacas (Agapito y col, 2008) el cual presenta una probabilidad de exclusión total de 0.999 y un valor de PIC promedio de 0.7951, siendo ambos valores mayores a los obtenidos para los 6 marcadores analizados (PIC promedio:0.735, PE acumulada: 0.994) (Tabla 6). Al comparar los valores de PIC y PE individuales se observa que los nuevos marcadores generados presentan valores mayores en ambos indicadores. Para esta comparación se tomó como criterio que ambos marcadores presentaran el mismo número de alelos, puesto que a mayor número de alelos más informativo y mayor es su probabilidad de exclusión. Esta diferencia entre los valores de PIC y PE, a pesar de presentar el mismo número de alelos, puede deberse a la diferencia existente entre la cantidad de individuos evaluados, habiéndose empleado un mayor número de individuos para caracterizar el panel de 10 marcadores (n=329) en comparación a los empleados en este proyecto (n=30). El valor de la frecuencia alélica es dependiente del tamaño de la muestra y los valores de PIC y de PE dependen a su vez de la frecuencia alélica.

Este es el primer trabajo que busca caracterizar la Tricohialina, una proteína constituyente de la fibra de alpaca. Sobre la secuencia genómica parcial del gen de esta proteína se buscó identificar mutaciones que califiquen como marcadores SNPs. La Tricohialina tiene una función conservada entre las especies, sin embargo no así su tamaño que varía en un rango de 190-220kDa, esta variación de tamaño se debe a que la proteína presenta un número variable de repeticiones de 23 aminoácidos los cuales se encuentran distribuidas de forma homogénea lo largo del gen (Rogers y col, 1991). Esta variación de tamaño indicaría diferencias en la secuencia del gen dependiendo de la especie. En humanos el gen de Tricohialina presenta más de 9 kb y se encuentra compuesto por 3 exones de los cuales solo dos son traducidos. (Lee y col, 1993).

La Unidad de Biotecnología Molecular de Universidad Peruana Cayetano Heredia posee una biblioteca de cDNA de piel de alpaca de la cual se obtuvo la secuencia del fragmento del gen de Tricohiliana. Esta secuencia fue comparada con la información existente en el GeneBank publicada por el proyecto de secuenciamiento del genoma de la alpaca y alineó con el contig 1.390472 en el extremo final. Se diseñaron cebadores para amplificar el fragmento identificado a partir de la biblioteca de piel debido a que el proyecto de genoma completo de alpaca solo tiene una resolución de 2x lo que no asegura la veracidad de la secuencia.

Los cebadores amplificaron un fragmento de 465pb dentro del cual se identificaron 3 SNP en la posición 57, 227 y 476 (Figura 6) A57G y A476G presentaron valores moderados de PIC, Heterocigocidad observada y Heterocigocidad esperada (Tabla10). Para un marcador bialélico el valor máximo que se puede obtener de PIC es de 0.374 y este valor siempre será menor que el de heterocigocidad. (Hildebrand y col, 1992). A227G presentó valores bajos de PIC, Heterocigocidad observada y Heterocigocidad esperada (Tabla10). Esto último se debe a que el alelo de menor frecuencia (Guanina) se presentó en una frecuencia de 0.0417, además los individuos portadores de este alelo pertenecieron en su mayoría (75%) a las muestras tomadas de la población de Huancavelica y el resto a Arequipa.

Los marcadores A57G y A476G presentaron un número muestral menor (n: 37 y 33 respectivamente) debido a la presencia de una fila de 9 adeninas a partir de la posición 172 luego de la cual algunas secuencias presentaron una superposición de picos impidiendo determinar las bases de los extremos. Esto puede deberse a un error de la polimerasa al

encontrar la fila de adeninas o a un posible polimorfismo en el número de adeninas entre un alelo y el otro lo que genera la sobreposición de los picos (Anonimo, 2014).

Los tres marcadores SNP se encuentran en equilibrio de Hardy-Weinberg para la población muestreada ($p > 0.05$). El análisis de desequilibrio de ligamiento genotípico entre pares de loci indicó que el marcador A227G es independiente de los otros dos marcadores: A57G ($p\text{-value}: 0.768$) y A476G ($p\text{-value}: 0.606$). Sin embargo, estos dos últimos sí presentan desequilibrio de ligamiento genotípico ($p\text{-value}: 0.00$). Una explicación a esto podría ser el número muestral pequeño de ambas muestras o el error introducido por la fila de adeninas pues, debido a ella, son las mismas muestras que faltan para los dos marcadores. Es interesante notar que estos dos últimos SNP se encuentran a los extremos con A227G posicionado al medio. La falta de desequilibrio de ligamiento entre estos y A227G podría deberse a que este último es de introducción reciente lo cual se ve corroborado con el hecho de que el alelo de menor frecuencia para el SNP A227G se encuentra presente solo en las poblaciones de Huancavelica y Arequipa.

La secuencia de Tricohialina obtenida a partir de los datos fue comparada con la secuencia del gen de Tricohialina en humanos y alineo en la región 3'UTR. La región 3' UTR en una zona del mRNA la cual es transcrita pero no traducida, que se encuentra luego del codón de terminación de la proteína. Esta región participa en varios procesos regulatorios como estabilidad y poliadenilación del mRNA, su traducción y localización, por lo que es crítica en la determinación del destino de esta hebra simple (Barret y col, 2012).

La región 3'UTR presenta una mayor extensión que la 5'UTR y se encuentra más expuesta a variaciones evolutivas, lo que le da un mayor potencial para la generación de elementos reguladores. Se cree existe una relación entre el largo de la región y la edad evolutiva de la secuencia, el tamaño de la secuencia aumenta con el tiempo, y posiblemente también con la complejidad del organismo. Esta relación sugiere que la región 3'UTR tendría un gran potencial regulador de la expresión de los genes, pues al ser variable, permite la generación de nuevos factores reguladores (Mazumder y col 2003). Sin embargo aunque presenta una gran variabilidad también se encuentra en ella regiones altamente conservadas. La región 3' UTR presenta algunos de los elementos más conservados dentro del genoma de los mamíferos (Barret y col, 2012). Esto podría explicar la alta similitud obtenida entre el fragmento amplificado del gen de Trichohialina en alpaca y la secuencia del gen de Trichohialina humana a pesar de la diferencias evolutivas entre las especies y la posible diferencia en el tamaño de la proteína.

Se buscó determinar la presencia de asociación entre los marcadores SNP identificados en el gen de Trichohialina de alpacas y los fenotipos Suri y Huacaya. Para ello primero fue necesario comprobar si los 3 marcadores SNP, previamente identificados en alpacas Huacaya, estaban presentes en alpacas Suri. Los marcadores A57G y A476G presentaron un número muestral menor (Suri n: 35 - 27 y Huacaya n: 35-30 respectivamente) debido a la presencia de una fila de 9 adeninas a partir de la posición 172 que generó una pérdida de la secuencia. En los datos analizados se comprobó la presencia de los 3 SNP en las muestras de ADN genómico de alpacas Suri.

Luego se evaluó si los marcadores SNPs se encontraban en equilibrio Hardy y Weinberg para la población de Suri mostrada. Cada uno de los 3 SNPs evaluados se encontró en equilibrio de Hardy y Weinberg ($p > 0.05$). Sin embargo al realizar un análisis de equilibrio de Hardy-Weinberg global el marcador A57G parece presentar un déficit de Heterocigotos ($p > 0.0450$). Se corroboró que cada uno de los 3 SNPs presentes en el fragmento del gen de Trichialina amplificado se encuentra en equilibrio de Hardy y Weinberg para las alpacas Suri.

Para determinar si los marcadores presentaban herencia independiente entre ellos se realizó una prueba de desequilibrio de ligamiento genotípico. Se observó que en Suri los marcadores A57G y A458G se encuentran en desequilibrio de ligamiento, mientras que A227G es independiente. Estos datos concuerdan con los antes obtenidos para alpacas Huacaya en donde también se observó que los marcadores A57G y A458G se encuentran en desequilibrio de ligamiento ($p: \infty$) y A227G es independiente (A57G, $p: 0.916$; A458G, $p: 0.894$).

El patrón de segregación de los fenotipos Suri y Huacaya en alpacas aun no ha sido identificado de forma certera, se cree que estos fenotipos se encuentran determinados por 2 loci que interaccionan presentándose el fenotipo Huacaya solo cuando ambos loci son homocigotos recesivos (Renari, 2011). El análisis de asociación realizado busca definir la existencia de asociación entre los SNPs identificados en el gen de Trichialina con las diferencias fenotípicas Suri - Huacaya. Los 3 análisis realizados, frecuencia de genotipos, frecuencia de haplotipos y Análisis de correspondencia Factorial mostraron que no hay diferencias significativas entre los dos fenotipos. Los SNPs identificados en la región 3'UTR del gen de Trichialina no se encuentran relacionados a los fenotipos Suri y Huacaya.

Este proyecto caracterizó 6 nuevos marcadores microsatelites y 3 marcadores SNPs de utilidad en alpaca; sin embargo a la fecha existen nuevas tecnologías más efectivas para el desarrollo de un mayor número de marcadores genéticos. Técnicas como Genotipificación por Secuenciamiento (Genotyping by sequencing, GBS) ha mostrado ser efectivas para identificar nuevos marcadores en plantas y animales. En maiz permitió identificar 200 000 marcadores en un corto periodo de tiempo y a un bajo precio (Elshire y col, 2011). En bovinos se identificaron y genotipificaron mas de 50 000 SNPs con esta técnica, que en comparación con el BovineSNP50BeadChip, analiza un número similar de marcadores a un precio menor por muestra (De Donato y col, 2013). Esta técnica puede ser aplicada tanto en especies con genoma completo como en especies sin genoma publicado, lo que la hace un herramienta util en estudios de especies domesticas como en especies silvestre; además los datos obtenidos permiten realizar estudios de diversidad genetica, evolución, filogenia y asociación genotipo-fenotipo (Narum y col, 2013).

VII. CONCLUSIONES

Los marcadores microsatélites caracterizados (n=6) mostraron ser altamente informativos ($PIC > 0.7$) lo que los hace aptos para estudios de identificación de individuos, pruebas de paternidad y análisis de estructura poblacional.

La capacidad de multiplexado de los 6 marcadores generados en 2 reacciones de PCR múltiple facilita el empleo rápido, económico y standarizado de los marcadores en futuros estudios.

Los 3 marcadores SNP: A57G, A227G y A458G, se encuentran presentes en la región 3'UTR del gen de Tricohialina donde podrían encontrarse relacionados a la regulación de la expresión del gen.

Los análisis de asociación realizados entre los marcadores SNPs de Tricohialina y los fenotipos Suri y Huacaya sugieren que ninguno de los SNPs se encuentra relacionado a las diferencias existentes entre los 2 fenotipos.

VIII. RECOMENDACIONES

Se recomienda determinar los estimadores de variabilidad de los 45 cebadores restantes previamente evaluados.

Caracterizar los 3 SNP encontrados para el gen de Tricohialina empleando un mayor número de individuos.

Identificar la secuencia completa del gen Tricohialina con la finalidad de buscar SNP en zonas codantes.

IX. REFERENCIAS BIBLIOGRÁFICAS

1. Agapito J., Rodriguez J., Herrera-Velit P., Timoteo O., Rojas P., Boettcher P., Garcia F., Espinoza J. (2008) Parentage testing in alpacas (*Vicugna pacos*) by using semi-automated fluorescent multiplex PCRs with 10 microsatellite markers. *Animal genetics*, 39:201-203
2. Alibardi L. (2004) Fine structure of marsupial hairs, with emphasis on trichohyalin and the structure of the inner root sheath. *Journal of Morphology*, 261:390-402
3. Anderson J.D., Honycutt R.L., Gonzales R.A., Gee K.L., Skow L.C., Gallagher R.L., Honycutt D.A., deYoung R.W. (2002) Development of microsatellite DNA markers for the automated genetic characterization of White-Tailed Deer Populations. *Journal of Wildlife management*, 66: 67-74
4. Anonimo (2014) In Case of Trouble. DNA Sequencing Core. The University of Michigan. <http://seqcore.brcf.med.umich.edu/doc/dnaseq/trouble/badseq.html>
5. Barret L.W., Fletcher S., Wilton S.D. (2012) Regulation of eukaryotic gene expression by the untranslated gene region and other non-coding elements. *Cellular and Molecular Life Sciences*, 69:3613-3634
6. Belkhir K., Borsa P., Chikhi L., Raufaste N. & Bonhomme F. 1996-2004 GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France)
7. Bonavia D. (1996) Los Camelidos Sudamericanos: (una introducción a su estudio), *Travaux de l'I.F.E.A.* 93:843-853

8. Botstein D., White R., Skolnick M., Davis R (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American journal of human genetics*, 32:314-31
9. Brooks A.J. (1999) The essence of SNPs. *Gene*, 234 (2):177-186
10. Buford M.W., Bradley D.G., Luikart G. (2003) DNA markers reveal the complexity of Livestock domestication. *Nature Reviews genetics*, 4:900-908
11. Bustamante A.V., Mat M. L. , Zambelli A., Vidal-Rioja L. (2003) Isolation and characterization of 10 polymorphic dinucleotide microsatellite markers for llama and guanaco. *Molecular ecology Notes*, 3(1):68-69
12. Butler J.M. (2005) Forensic DNA typing. Biology, technology and genetics of STR markers. 2da. Ed. Elsevier Academic Press. USA: 455-539
13. Candido J.R., Gutiérrez G.A.(2011) Finura de los vellones del plantel de alpacas de la Sais Pachacutec del departamento de Junín-Perú. VII Congreso de la Asociación Latinoamericana de Especialistas en Pequeños Rumiantes y Camélidos Sudamericanos
14. Cañon J., Checa M., Carleos C., Vega-Pla J., Dunner S. (2000) The genetic structure of Spanish celtic horse breeds inferred from microsatellites data. *Animal Genetics*, 31: 39-48
15. Chistiakov D., Hellemans B., Volckaert F. (2006) Microsatellites and their genomic distribution, evolution, function and applications: A review with reference to fish genetics. *Aquaculture*, 255:1-29
16. Collins D.W., Jukes T.H. (1994) Rates of transition and transversion in coding sequences since the human-rodent divergence. *Genomics*, 20:386-396
17. COMTRADE (2012) UN COMTRADE database. <http://comtrade.un.org/>
18. De Donato M., Peters S.O., Mitchell S.E., Hussain T., Imumorin I.G. (2013) Genotyping-by-Sequencing (GBS): a novel, efficient and cost-effective genotyping method for cattle

- using next-generation sequencing. Plos One, (8)5: e62137. doi:10.1371/journal.pone.0062137
19. deWoody J.A., Honycutt R.L., Skow L.C. (1995) Microsatellite markers in white-tailed deer. *The Journal of Heredity*, 86:317-318
 20. Elshire R.J., Glaubitz J.C., Sun Q., Poland J.A., Kawamoto K., Buckler E.S., Mitchell S.E. (2011) A robust, simple Genotyping-by sequencing (GBS) approach for high diversity species. *Plos One* 6(5): e19379. doi:10.1371/journal.pone.0019379
 21. Fietz M.J., McLaughlan C.J., Campbell M.T., Rogers G.E. (1993) Analysis of the sheep trichohyalin gene: potential structural and calcium-binding roles of trichohyalin in the hair follicle. *The Journal of cell Biology*, 121:855-865
 22. Fietz M.J., Presland R.B., Rogers G.E. (1990) The cDNA-deduced amino acid sequence for Trichohyalin, a differentiation marker in the hair Follicle contains a 23 amino acid repeat. *The Journal of cell Biology*, 110:427-436
 23. Frank E., Hick M., Gauna C., Lamas H., Renieru C., Antonini M. (2006) Phenotypic and genetic description of the fibre traits in South American domestic camelids (llamas and alpacas). *Small Ruminant Research* 61:113-129
 24. Freeland J.R. (2005) *Molecular Ecology*. John Wiley and Sons. pg 51 – 53
 25. Gill P. (2001) An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *Int. J. Legal. Med.*, 114:204–210
 26. Gregorius H.G. (1980) The probability of losing an allele when diploid genotypes are sampled. *Biometrics*, 36, 643-652
 27. Guichoux E., Lagache L., Wagner S., Chaumeil P., Leger P., Lepais O., Lepoittevin C., Malausa T., Revardel E., Salin F., Petit R.J. (2011) Current trends in microsatellite genotyping. *Molecular Ecology Resources*, 11: 591-611

28. Guo S. W., Thompson E. A. (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, 48:361-372
29. Hale M.L., Bruge T.M., Stevees T.E. (2012) Sampling for Microsatellite-based population genetic studies: 25-30 individuals per population is enough to accurately estimate allele frequencies. *PLoS ONE* 7(9): e45170
30. Hamilton M.B. (2009) *Population Genetics*. Ed. Wiley-Blackwell, pg 9-52
31. Hancock J. (2003) Microsatellites and other simple sequences: genomic context and mutational mechanisms En: Goldstein D., Schlötterer C. *Microsatellites Evolution and Applications*. Oxford University Press pag 1-9
32. Hildebrand C.E., Torney D.C., Wagner R.P. (1992) Informativeness of Polymorphic DNA markers. *Los Alamos Science*, 20: 100-102
33. Hirschhorn J.N., Lohmueller K., Byrne E., Hirschhorn K. (2002) A comprehensive review of genetic association studies. *Genetics in Medicine*, 4:45-61
34. INEI (2013) IV Censo Nacional Agropecuario. Resultados definitivos. pg18
35. Jamieson A., Taylor S. (1997) Comparisons of three probability formulae for parentage exclusion. *Animal genetics*, 28:397-400
36. Kalinowski S.T., Taper M.L., Marshall T.C. (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular ecology*, 16:1099–106
37. Kaul R., Singh A., Vijh R. K., Tania M. S. and Behl R.(2001) Evaluation of the genetic variability of 13 microsatellite markers in native Indian pigs. *Journal of Genetics*, 80: 149–153
38. Kemp S., Hishida O., Wambugu J., Rink A., Longeri M., Ma R., Da Y., Lewin H., Barendse W., Teale A. (1995) A panel of polymorphic bovine, ovine and caprine microsatellite markers. *Animal Genetics*, 26: 299-306

39. Kim H., Schmidt C., Decker K., Emara M. (2002) Chicken SNP discovery by EST data mining, in: Plant, Animal & Microbe Genomes X, 12-16 January 2002, San Diego. Disponible: http://www.intl-pag.org/pag/10/abstracts/PAGX_P246.html
40. Kim S. Misra A. (2007) SNP Genotyping: technologies and Biomedical Applications. Annual review of biomedical Engineering,9:289-320
41. Kluth S., Distl O. (2013) Congenital Sensorineural deafness in Dalmatian Dogs associated with Quantitative Trait Loci. PLoS ONE 8(12): e80642. doi:10.1371/journal.pone.0080642
42. Kumar S., Stecher G., Peterson D., and Tamura K. (2012) MEGA-CC: Computing Core of Molecular Evolutionary Genetics Analysis Program for Automated and Iterative Data Analysis. Bioinformatics 28:2685-2686
43. Lee S.C., Kim I.G., Marekov L.N., O'Keefe E.J., Parry D.A., Steinert P.M. (1993) The structure of human trichohyalin. Potential multiple roles as a functional EF-hand-like calcium-binding protein, a cornified cell envelope precursor, and an intermediate filament-associated (cross-linking) protein. Journal of Biological Chemistry, 268: 12164-76
44. Li W., Gojobori T., Nei M. (1981) Pseudogenes as a paradigm of neutral evolution. Nature, 292:237-239
45. Liu Z. (2007) Single nucleotide polymorphism (SNP). In: Aquaculture genome technologies. Eds. Z. Liu. Blackwell Publishing. USA:59-72
46. Luikart G., Biju-Duval M.P., Ertugrul O., Zagdsuren Y., Maudet C., Taberlet P. (1999) Power of 22 microsatellite markers in fluorescent multiplexes for parentage testing in goats (*Capra hircus*). Animal Genetics, 30:431-438
47. Martinez-Arias R., Calafell F., Mateu E., Comas D., Andres A., Bertranpetit J. (2001) Sequence variability of a human pseudogene. Genome Res., 11:1071-1085
48. Maudet C., Miller C., Bassano B., Breitenomoseer-wursten C., Gauthier D., Luikart G. et al. (2002) Microsatellite DNA and recent stadistical methods in wildlife conservation

- management: applications in alpine Ibex (*Capra ibex(ibex)*). *Molecular Ecology*, 11:421-436
49. Mazumder B., Seshadri V., Fox P.L. (2003) Translational control by the 3'UTR: the ends specify the means. *TRENDS in Biochemical Science*, 28:91-98
 50. McPartlan H.C., Matthews M.E., Robinson N.A. (1998) Alpaca micro- satellites at the VIAS A1 and VIAS A2 loci. *Animal Genetics*, 29:158
 51. Medland S.E., Nyholt D.R., Painter J.N., McEvoy B.P., McRae A.F., Zhu G., Gordon S.D., Ferreira M.A., Wright M.J, Henders A.K., Campbell M.J., Duffy D.L., Hansell N.K., Macgregor S., Slutske W.S., Heath A.C., Montgomery G.W., Martin N.G. Common Variants in the trichohyalin gene are associated with straight hair in europeans. *The American Journal of Human Genetics*, 85: 750-755
 52. MINAG (2009) *Dinámica Agropecuaria 1997-2009*
 53. MINAGRI (2013) *Camelidos Sudamericanos*. <http://www.minagri.gob.pe/portal/sector-agrario/pecuaria/situacion-de-las-actividades-de-crianza-y-produccion/cam%C3%A9lidos-sudamericanos>
 54. Morin P.A., Luikart G., Wayne R.K. (2004) The SNP workshop group. SNPs in ecology, evolution and conservation. *TRENDS in Ecology and Evolution*, 19:208-216
 55. Narum S.R., Buerkle C.A., Davey J.W., Miller M.R., Hohenlohe P.A. (2013) genotyping-by-Sequencing in ecological and conservation genomics. *Molecular Ecology*, 22:2841-2847
 56. Nei M. (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the national academy of sciences USA*, 70:3321-3323
 57. Nei M., Kumar S. (2000) *Molecular Evolution and Phylogenetics*. Oxford University Press. USA: 231-275

58. O'Keefe E.J., Hamilton E.H., Lee S.C., Steinert P. (1993) Trichohyalin: A Structural Protein of Hair, Tongue, Nail and Epidermis. *The Society for Investigative Dermatology*. 101,65S-71s
59. Obreque V., Coogle L., Henney P., Bailey E., Mancilla R., Garcia-Huidobro J., Hinrichsen P., Cothran E. (1998) Characterisation of 10 polymorphic alpaca dinucleotide microsatellites. *Animal Genetics*, 29:461-462
60. Penedo M.C., Caetano A.R., Cordova K. (1999) Eight microsatellite markers for South American camelids. *Animal Genetics*, 30(2):166-7
61. Penedo M.C., Caetano A.R., Cordova K.I. (1998) Microsatellite markers for South American camelids. *Animal Genetics*, 29(5):411-2
62. Penedo M.C., Caetano A.R., Cordova K.I. (1999) Six microsatellite markers for South American camelids. *Animal Genetics*, 30(5):399
63. Pérez-Cabal M.A., Cervantes I., Morante R., Burgos A., Goyache F., Gutiérrez J.P. (2010) Analysis of the existence of major genes affecting alpaca fiber traits. *The Journal of Animal Science*, 88:3783-3788
64. Petitjean A., Achatz M.I.W., Borresen-Dale A.L., Hainaut P., Olivier M. (2007) TP53 mutation in human cancers: functional selection and impact on cancer prognosis and outcomes. *Oncogene*, 26:2157-2165
65. Powell B.C., Rogers G.E. (1997) The role of keratin proteins and their genes in the growth structure and properties of hair. En: Jalles P., Zehn H., Hocker H. *Formation and structure of the human hair*. Ed. Birkhäuser Verlag. pag 59-14z
66. Presciuttini S., Valbonesi A., Apaza N., Antonini M., Huanca T., Renieri C. (2010) Fleece variation in alpaca (*Vicugna pacos*): a two-locus model for the suri/Huacaya phenotype. *BMC Genetics* 11:70

67. Raymond M., Rousset F. (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, 86:248-249
68. Reed K. M., Chaves L. D. (2008). Simple Sequence Repeats for Genetic Studies of Alpaca. *Animal Biotechnology*, 19:4, 243 – 309
69. Reneri C., Valbonesi A., Antonini M., La Manna V., Huanca T., Apaza N., Presciuttini S., Asparrin M. (2011) Suri/Huacaya phenotype inheritance in alpaca (*Vicugna pacos*) En: *Fibre production in South American camelids and other fibre animal*. Pérez-Cabal M.A. y col. Wageningen Academic Publishers. Pag 25-34
70. Rogers G.E., Fietz M.J., Fratini A. (1991) Trichohyalin and Matrix Proteins. *Annals of the New York Academy of Science*. 642, 64-80
71. Rogers G.E., Powell B.C. (1993) Organization and expression of hair follicle genes. *The Journal of Investigative dermatology*, 101(1 Suppl):50S-55S
72. Roizès G. (2000) Identification of Microsatellite Markers: Screening for Repeat Sequence and Mapping Polymorphism. Eaton Publishing. 35-48
73. Sasse J., Mariasegaram M., Jahabar Ali M., Pullena Yegum S., Babu R., Kinne J., Wernery U. (2000) Development of a microsatellite parentage and identity verification test for dromedary racing camels. Abstracts of the 27th Conference of the international society of animal genetics (ISAG). Mineapolis. EE.UU
74. Seaton G, Haley C, Knott SA, Kearsey, Visscher PM. (2002) QTL Express: mapping quantitative trait loci in simple and complex pedigrees. *Bioinformatics Applications Note*, 18:1-2
75. Shimomura Y., Christiano A.M. (2010) Biology and Genetics of Hair. *Annual review of genomics and human genetics*, 11:109-132
76. SIICEX (2012) Producto/ Fibra de alpaca. Partidas arancelarias del producto, exportadas en los últimos años.

http://www.siicex.gob.pe/siicex/portal5ES.asp?_page_=172.17100&_portletid_=sfichaproducto=cc_fp_init&ppproducto=86&pnomproducto=Fibra

77. Smith E., Shi L., Drummond P., Rodriguez L., Hamilton R., Ramlal S., Smith G., Pierce K., Foster J. (2001) Expressed sequence tags for the chicken genome from a normalized 10-day-old White Leghorn whole embryo cDNA library: 1. DNA sequence characterization and linkage analysis. *Journal of Heredity*, 92:1-8
78. Sobrino A., Brión M., Carracedo A. (2005) SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Science International*, 154:181–194
79. Staub J.E., Serquen C., Gupta M. (1996) Genetic markers, map construction and their application in plant breeding. *HortScience* 13:729- 741
80. Sunnucks P. (2000) Efficient genetic markers for population biology. *Tree*, 15:199-203
81. Syvanen A. (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature Review Genetics*, 2(12):930-942
82. Syvanen A. (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature Review Genetics*, 2(12):930-942
83. Tamura K., Peterson D., Peterson N., Stecher G., Nei M., and Kumar S. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* 28: 2731-2739
84. Vaiman D., Mercier D., Moazami-Goudarzi K., Eggen A., Ciampolini R., Lépingle A., Velmala R., Kaukinen J., Varvio S.L., Martin P., y col (1994) A set of 99 cattle microsatellites: characterization, synteny mapping, and polymorphism. *Mammalian Genome*, 5: 288-297
85. Vignal A., Milan D., San Cristobal M., Eggen A. (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genet. Sel. Evol.*, 34:275-305

86. Voisey J., Morris C.P. (2008) SNP technologies for drug discovery: a current review. *Curr Drug Discov Technol*, 5(3):230-5
87. Waits, L. P., Luikart, G., Taberlet, P. (2001) Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. *Molecular ecology*, 10:249-256.
88. Weir, B. (1996) Genetic data análisis II. Methods for discrete population genetic data. Sinauer Associates Inc Publishers. Massachusetts: 150-156; 209-211
89. Wright, S. (1969) Evolution and the genetics of populations. Vol. 2. The theory of gene frequencies. University of Chicago Press

ANEXO 1

Cuantificación de ADN

Se determinó la calidad y cantidad del ADN genómico mediante espectrofotometría empleando el equipo NanoDrop 1000 (Thermo Scientific). La cuantificación se realizó a una longitud de onda de 260nm y la calidad del ADN fue determinada en base a los índices 260nm/280nm y 260nm/230nm. Todas las muestras de ADN fueron diluidas a una concentración de 10ng/ul.

ANEXO 2

Calculo del número muestral

Se empleó el numero muestral empleando el programa MINSAGE (Gregorius, 1980). Los parámetros empleados fueron una frecuencia alélica mínima de 0.05, un nivel de confianza del 99% y se asumió que los polimorfismos se encuentran distribuidos de manera uniforme encontrándose en equilibrio de Hardy -Weinberg.

Para los microsatelites también se tuvo en cuenta el artículo publicado por Hale y col en el 2012 donde establece que con 25- 30 individuos puedes estimar de manera precisa los valores de frecuencias alélicas.