

**UNIVERSIDAD NACIONAL AGRARIA  
LA MOLINA**

**ESCUELA DE POSGRADO  
DOCTORADO EN CIENCIA ANIMAL**



**“IDENTIFICACIÓN DE POLIMORFISMOS DE NUCLEÓTIDO  
SIMPLE (PNSs) POR SECUENCIAMIENTO DE BIBLIOTECAS DE  
REPRESENTACIÓN REDUCIDA EN ALPACAS (*Vicugna pacos*)  
HUACAYA”**

**Presentada por:**

**MARCOS CALDERÓN MONTES**

**TESIS PARA OPTAR EL GRADO DE DOCTOR  
*DOCTORIS PHILOSOPHIAE* EN CIENCIA ANIMAL**

**Lima - Perú**

**2022**

Document Information

<b>Analyzed document</b>	Tesis Calderon 2022 20_06.docx (D140997552)
<b>Submitted</b>	6/22/2022 2:05:00 PM
<b>Submitted by</b>	Gustavo A. Gutierrez Reynoso
<b>Submitter email</b>	gustavogr@lamolina.edu.pe
<b>Similarity</b>	3%
<b>Analysis address</b>	gustavogr.unalm@analysis.urkund.com

Sources included in the report

**W** URL: <https://doi.org/10.3389/fgene.2021.723181>  
 Fetched: 6/22/2022 2:06:00 PM

**W** URL: <https://www.mdpi.com/2073-4425/12/2/291/htm>  
 Fetched: 6/22/2022 2:05:00 PM

**Universidad Nacional Agraria La Molina / TESIS - JORGE SEDANO.docx**

**SA** Document TESIS - JORGE SEDANO.docx (D140997003)  
 Submitted by: gustavogr@lamolina.edu.pe  
 Receiver: gustavogr.unalm@analysis.urkund.com

**W** URL: <https://doi.org/10.1186/1471-2164-11-469>  
 Fetched: 6/22/2022 2:06:00 PM

**SA** **Tesis - 11-11-19.docx**  
 Document Tesis - 11-11-19.docx (D58778556)

**W** URL: <https://doi.org/10.1186/s12864-020-06783-9>  
 Fetched: 6/22/2022 2:06:00 PM

Entire Document

UNIVERSIDAD NACIONAL AGRARIA LA MOLINA

ESCUELA DE POSGRADO

DOCTORADO EN CIENCIA ANIMAL

"IDENTIFICACIÓN DE POLIMORFISMOS DE NUCLEÓTIDO SIMPLE (PNSs) POR SECUENCIAMIENTO DE BIBLIOTECAS DE REPRESENTACIÓN REDUCIDA EN ALPACAS (Vicugna pacos) HUACAYANA"

Presentado por: MARCOS CALDERON MONTES

TESIS PARA OPTAR EL GRADO DE DOCTORIS PHILOSOPHIAE EN

CIENCIA ANIMAL

Lima - Perú

2021 DEDICATORIA

A Dios, por darme la oportunidad de vivir y por estar conmigo en cada paso que doy, por fortalecer mi corazón e iluminar mi mente para lograr mis objetivos, y el haber llegado hasta este momento.

A mis padres Elsa y Alejandro; por brindarme la libertad de seguir mis ideales orientándome y ofreciendo su apoyo incondicional. Gracias por el esfuerzo realizado para concluir una más de mis metas.

A mis hermanos Viviana, Carlos, José y María mil gracias por brindarme su apoyo y creer en mí por cumplir una de las metas familiares. Para mi motor y motivo Matias Aaron Calderon Cruz.

AGRADECIMIENTO

Al PhD. Federico Abel Ponce de León Bravo y PhD. Gustavo Gutiérrez Reynoso, por su apoyo incondicional y acertada dirección como patrocinadores del presente trabajo.

A todos los miembros de mi comité consejero: PhD. J. Alberto Barrón López, PhD. Juan Chávez Cossío y PhD. Javier Ñaupari Vásquez.

Al Proyecto 028-2016-INIA-PNIA/UPMSI/IE "Identificación de polimorfismos de nucleótidos simples (PNSs) y desarrollo de un Chip de PNSs para la implementación de tecnologías avanzadas de genómica en alpacas".

Al Convenio de Subvención N°178-2015-FONDECYT: Programa de Becas de Doctorado en Ciencia Animal por el financiamiento de mis estudios de doctorado.

A los productores de alpaca en la Cooperativa Comunal San Pedro de Racco, GACOCEN, Fundo PACOMARCA- INCATOPS SA y al INIA- Quimsachata.

Al Vicerrectorado de Investigación por el uso del Servidor HPC Bioinformática – UNALM.

Al proyecto de cooperación institucional entre el VLIR-UOS (Bélgica) y la UNALM.

A todos mis compañeros(as), amigos(as), con quienes compartimos gratos momentos y experiencias académicas.

ÍNDICE GENERAL

I. INTRODUCCIÓN 1 II. REVISIÓN BIBLIOGRÁFICA 3 2.1. LA ALPACA 3 2.1.1. Población de alpaca 3 2.1.2. Características fenotípicas de la alpaca 5 2.2. GENOMA DE LA ALPACA 6 2.3. BIBLIOTECA DE REPRESENTACIÓN REDUCIDA 6 2.4. IDENTIFICACIÓN DE POLIMORFISMOS DE NUCLEÓTIDO SIMPLE EN ANIMALES DE GRANJA 16 2.6. VALIDACIÓN DE POLIMORFISMOS DE NUCLEÓTIDO SIMPLE EN ANIMALES DE GRANJA 16 2.6.1. Identificación de PNSs 16 2.6.2. Selección de PNSs 16 2.6.3. Cobertura de PNSs seleccionados en el genoma de referencia VicPac 3.1 25 3.2.3 Construcción de la micromatriz de PNSs para alpaca 27 3.2.4. Comparación de genotipado por secuenciamiento y genotipado por micromatriz 30 3.3.3 Estructura genética de la población 31 IV. RESULTADOS Y DISCUSIÓN 33 4.1. BIBLIOTECA DE REPRESENTACIÓN REDUCIDA DE ALPACA 33 4.2. IDENTIFICACIÓN DE POLIMORFISMOS DE NUCLEÓTIDO SIMPLE EN ANIMALES DE GRANJA 33 4.3. VALIDACIÓN DE POLIMORFISMOS DE NUCLEÓTIDO SIMPLE EN ANIMALES DE GRANJA 33 4.4. COMPARACIÓN DEL GENOTIPADO POR SECUENCIAMIENTO ENTRE EL GENOTIPADO POR MICROMATRIZ DE PNS DE ALPACA 39 4.4.1. Identificación de PNSs 39 4.4.2. Selección de PNSs 39 4.4.3. Cobertura de PNSs seleccionados en el genoma de referencia VicPac 3.1 33 4.4.4. Comparación de genotipado por secuenciamiento y genotipado por micromatriz 33 4.4.5. Estructura genética de la población 33 4.5. COMPARACIÓN DEL GENOTIPADO POR SECUENCIAMIENTO ENTRE EL GENOTIPADO POR MICROMATRIZ DE PNS DE ALPACA 45 4.5.1. Identificación de PNSs 45 4.5.2. Selección de PNSs 45 4.5.3. Cobertura de PNSs seleccionados en el genoma de referencia VicPac 3.1 45 4.5.4. Comparación de genotipado por secuenciamiento y genotipado por micromatriz 45 4.5.5. Estructura genética de la población 45 4.6. RELACIONES GENÓMICAS ENTRE LOS MIEMBROS DEL TRÍO 46 4.6.1. Identificación de PNSs 46 4.6.2. Selección de PNSs 46 4.6.3. Cobertura de PNSs seleccionados en el genoma de referencia VicPac 3.1 46 4.6.4. Comparación de genotipado por secuenciamiento y genotipado por micromatriz 46 4.6.5. Estructura genética de la población 46 4.7. PEDIGRÍ Y RELACIONES GENÓMICAS DE MEDIOS HERMANOS PATERNOS 47 4.7.1. Identificación de PNSs 47 4.7.2. Selección de PNSs 47 4.7.3. Cobertura de PNSs seleccionados en el genoma de referencia VicPac 3.1 47 4.7.4. Comparación de genotipado por secuenciamiento y genotipado por micromatriz 47 4.7.5. Estructura genética de la población 47 4.8. RELACIONES GENÓMICAS ENTRE LOS MIEMBROS DEL TRÍO 46 Cuadro 12: Pedigrí y relaciones genómicas de medios hermanos paternos 47

ÍNDICE DE CUADROS

Cuadro 1: Número de animales por región y granja para la colecta de muestras de sangre 21 Cuadro 2: Ejemplo de la puntuación de localización de PNS. 26 Cuadro 3: Número de tríos (padre, madre y hijo) alineados con el genoma de referencia VicPac 3.1 33 Cuadro 5: Número de PNS seleccionados por puntuación de ubicación del primer y segundo conjunto 35 Cuadro 6: Cobertura del primer y segundo conjunto de cromosoma incluidos en la micromatriz 40 Cuadro 8: Número de fragmentos que contienen un PNS por intervalos de longitud de fragmentos. 42 Cuadro 9: Número de PNS seleccionados por puntuación de localización de PNS. 46 Cuadro 10: Número de PNS seleccionados por puntuación de ubicación del primer y segundo conjunto 46 Cuadro 11: Número de PNS seleccionados por puntuación de ubicación del primer y segundo conjunto 46 Cuadro 12: Pedigrí y relaciones genómicas de medios hermanos paternos 47

ÍNDICE DE FIGURAS

**UNIVERSIDAD NACIONAL AGRARIA  
LA MOLINA**

**ESCUELA DE POSGRADO**

**DOCTORADO EN CIENCIA ANIMAL**

**“IDENTIFICACIÓN DE POLIMORFISMOS DE NUCLEÓTIDO  
SIMPLE (PNSs) POR SECUENCIAMIENTO DE BIBLIOTECAS DE  
REPRESENTACIÓN REDUCIDA EN ALPACAS (*Vicugna pacos*)  
HUACAYA”**

**TESIS PARA OPTAR EL GRADO DE DOCTOR**

***Doctoris Philosophiae***

**Presentada por:**

**MARCOS CALDERÓN MONTES**

**Sustentada y aprobada ante el siguiente jurado:**

Ph.D. José Alberto Barrón López  
**PRESIDENTE**

Ph.D. Gustavo Gutiérrez Reynoso  
**ASESOR**

Ph.D. Federico Abel Ponce de León Bravo  
**CO-ASESOR**

Ph.D. Juan Francisco Chávez Cossio  
**MIEMBRO**

Ph.D. Javier Naupari Vásquez  
**MIEMBRO**

Dra. Marcia Marisol Paredes Peralta  
**MIEMBRO EXTERNO**

## **DEDICATORIA**

A Dios, por darme la oportunidad de vivir y por estar conmigo en cada paso que doy, por fortalecer mi corazón e iluminar mi mente para lograr mis objetivos, y el haber llegado hasta este momento tan importante de mi formación profesional.

A mis padres Elsa y Alejandro; por brindarme la libertad de seguir mis ideales orientándome y ofreciendo su apoyo incondicional. Gracias por el esfuerzo realizado para concluir una más de mis metas que también es suya.

A mis hermanos Viviana, Carlos, José y María mil gracias por brindarme su apoyo y creer en mí por cumplir una de las metas familiares. Para mi motor y motivo Matías Aarón Calderón Cruz.

## **AGRADECIMIENTO**

Al PhD. Federico Abel Ponce de León Bravo y PhD. Gustavo Gutiérrez Reynoso, por su apoyo incondicional y acertada dirección como patrocinadores del presente trabajo.

A todos los miembros de mi comité consejero: PhD. J. Alberto Barrón López, PhD. Juan Chávez Cossío y PhD. Javier Ñaupari Vásquez.

Al Proyecto 028-2016-INIA-PNIA/UPMSI/IE “Identificación de polimorfismos de nucleótidos simples (PNSs) y desarrollo de un Chip de PNSs para la implementación de tecnologías avanzadas de mejora genética en alpacas Huacaya” por financiamiento de la investigación.

Al Convenio de Subvención N°178-2015-FONDECYT: Programa de Becas de Doctorado en Ciencia Animal por el financiamiento de mis estudios de doctorado.

Al Programa de Investigación y Proyección Social en Mejoramiento Animal, al Instituto de Investigación de Bioquímica y Biología Molecular (IIBBM) de la Universidad Nacional Agraria La Molina y al Programa de Investigación y Proyección Social en Ovinos y Camélidos Sudamericanos.

A la Cooperativa Comunal San Pedro de Racco, Ganadería y Comercio del Centro S.R.L. (GACOCEN), Sociedad Ganadera Muñoz (SOGAMU) S.A.C., Centro de Investigación y Producción (CIP) Quimsachata del INIA y Fundo PACOMARCA.

Al Vicerrectorado de Investigación por el uso del Servidor HPC Bioinformática – UNALM.

Al proyecto de cooperación institucional entre el VLIR-UOS (Bélgica) y la UNALM.

Al Mg Sc. Eudosio Veli por su trabajo en la preparación de las muestras de ADN.

Al Sr. Godoy Muñoz y al Ing. Franklin Muñoz por el apoyo en la colección de muestras y recopilación de datos de las alpacas de SOGAMU.

Al Mg. Sc. Mayra Mendoza y al Mg. Sc. Gonzalo Gonzales por su apoyo en el IIBBM.

Al Ing. Luis Vela e Ing. Álvaro Fernández por su apoyo en diferentes etapas de la investigación, especialmente en la colección de muestras y en el IIBBM.

Al Mg Sc. Jorge Mendoza por su apoyo en la colección de muestras y recopilación de datos de las alpacas en Cerro de Pasco.

A todos mis compañeros(as), amigos(as), con quienes compartimos gratos momentos y experiencias académicas.

## ÍNDICE GENERAL

I.	INTRODUCCIÓN .....	1
II.	REVISIÓN BIBLIOGRÁFICA .....	3
2.1.	LA ALPACA .....	3
2.1.1.	Población de alpaca .....	3
2.1.2.	Características fenotípicas de la alpaca .....	5
2.2.	GENOMA DE LA ALPACA .....	6
2.3.	BIBLIOTECAS DE REPRESENTACIÓN REDUCIDA DE ADN.....	8
2.4.	POLIMORFISMO DE NUCLEÓTIDO SIMPLE .....	11
2.5.	MICROMATRIZ DE PNSs .....	14
2.7.1	Micromatriz con la tecnología BeadArray .....	15
2.7.2	Micromatriz con la tecnología GeneChip.....	16
2.7.3	Micromatrices de polimorfismos de nucleótido simple en animales de granja.....	16
2.6.	VALIDACIÓN DE LA MICROMATRIZ .....	18
III.	MATERIALES Y MÉTODOS .....	19
3.1	LOCALIZACIÓN .....	19
3.2	CONSTRUCCIÓN DE LA MICROMATRIZ DE PNSs DE ALPACA .....	21
3.2.1	Generación de bibliotecas de representación reducida de ADN .....	21
a.	Animales .....	21
b.	Colección de muestras de sangre y extracción de ADN .....	21
c.	Construcción de bibliotecas de representación reducida y secuenciación .....	22
3.2.2	Análisis bioinformático para la selección y cobertura de PNSs .....	23
a.	Identificación de PNSs.....	23
b.	Selección de PNSs .....	24

c.	Cobertura de PNSs seleccionados en el genoma de referencia <i>VicPac</i>	
	3.1	25
	3.2.3	27
3.3	VALIDACIÓN DE LA MICROMATRIZ DE PNSs DE ALPACA	28
	3.3.1	28
	3.3.2	30
	3.3.3	31
IV.	RESULTADOS Y DISCUSIÓN	33
	4.1.	33
	4.2.	34
	4.3.	39
	4.4.	45
	4.5.	48
	4.6.	49
V.	CONCLUSIONES	58
VI.	RECOMENDACIONES	59
VII.	REVISIÓN BIBLIOGRÁFICA	60
VIII.	ANEXOS	75



## ÍNDICE DE CUADROS

Cuadro 1: Número de animales por región y granja para la colecta de muestras de sangre .....	21
Cuadro 2: Ejemplo de la puntuación de localización de PNS. ....	26
Cuadro 3: Número de tríos (padre, madre, progenie) de alpacas y número de muestras de ADN por regiones .....	29
Cuadro 4: Cantidad de PNSs identificados por BRR alineados con el genoma de referencia <i>VicPac 3.1</i> .....	33
Cuadro 5: Número de PNS seleccionados por puntuación de ubicación del primer y segundo conjunto .....	35
Cuadro 6: Cobertura del primer y segundo conjunto de 80k de PNSs seleccionado por cromosoma en el <i>VicPac 3.1</i> .....	36
Cuadro 7: Número de PNS seleccionados para la micromatriz.....	39
Cuadro 8: Cobertura de los PNSs por cromosoma incluidos en la micromatriz .....	40
Cuadro 9: Número de fragmentos que contienen un PNS por intervalos de longitud de fragmentos.....	42
Cuadro 10: Cobertura del genoma de la alpaca con PNS incluidos en la micromatriz..	43
Cuadro 11: Pedigrí y relaciones genómicas entre los miembros del trío .....	46
Cuadro 12: Pedigrí y relaciones genómicas de medios hermanos paternos .....	47
Cuadro 13: Índice de fijación ( $F_{st}$ ) entre poblaciones de alpacas. ....	55

## ÍNDICE DE FIGURAS

Figura 1: Alpaca de la raza Huacaya. ....	5
Figura 2: Elaboración de bibliotecas de representación reducida .....	9
Figura 3: Polimorfismo de nucleótido simple .....	11
Figura 4: Genotipado del genoma con una micromatriz .....	15
Figura 5: Ubicación geográfica de Pacamarca e INIA-Quimsachata.....	20
Figura 6: Ubicación geográfica de GACOCEN y San Pedro de Racco. ....	20
Figura 7: Bandas de ADN de las muestras de GACOCEN y PACOMARCA en geles de agarosa al 0.8 por ciento teñidos con gel red. ....	22
Figura 8: Análisis de bioinformática para la identificación de PNSs.....	24
Figura 9: Flujo del análisis de bioinformática para la selección de PNSs.....	25
Figura 10: Distribución de los PNSs seleccionados del primer y segundo conjunto (80k) por cromosoma de alpaca, con una densidad promedio de 40 PNSs/Mb por cromosoma. ....	38
Figura 11: Distribución de PNSs seleccionados del primer y segundo conjunto (80k) por la frecuencia del alelo menor (FAM). ....	38
Figura 12: Densidad de PNSs incluidos en la micromatriz por cromosoma de alpaca. .	43
Figura 13: Estructura de la población basada en el genotipado por la micromatriz (Pacamarca-rojo, Quimsachata-verde, Racco-azul y Gacocen-negro). ....	50
Figura 14: Estructura genética de la población basada en el genotipado por secuenciamiento con resultados de PstI-MspI (Pacamarca-rojo, Quimsachata-verde, Racco-azul y Gacocen-negro). ....	50
Figura 15: Estructura genética de la población basada en el genotipado por secuenciamiento con resultados de ApeK1 (Pacamarca-rojo, Quimsachata-verde, Racco-azul y Gacocen-negro). ....	51
Figura 16: Mapa de calor de las relaciones genómicas entre animales basado en el genotipado por la micromatriz Gacocen (abajo a la izquierda), Pacamarca y Quimsachata (área central) y Racco (arriba a la derecha).....	52
Figura 17: Mapa de calor de las relaciones genómicas entre animales basada en el genotipado por secuenciamiento con resultados de PstI-MspI (Gacocen (abajo a la izquierda), Pacamarca y Quimsachata (área central) y Racco (arriba a la derecha).....	52

Figura 18: Mapa de calor de las relaciones genómicas entre animales basada en el genotipado por secuenciamiento con resultados de ApeK1 (Gacocen (abajo a la izquierda), Pacamarca y Quimsachata (área central) y Racco (arriba a la derecha).....	53
Figura 19: Coeficiente de consanguinidad (F) de alpacas genotipado por micromatriz.	54
Figura 20: Nivel de heterocigosidad de alpacas genotipado por micromatriz. ....	54
Figura 21: Análisis de diferenciación genética de las granjas de alpaca por individuo utilizando el software ADMIXTURE (K = 4). ....	55

## RESUMEN

El sustento de los pequeños productores depende de sus rebaños de alpacas y de la producción de fibra. La mejora genética de las características de la fibra incrementaría sus beneficios económicos y calidad de vida. La incorporación de la tecnología de marcadores moleculares podría superar las limitaciones actuales para la implementación de programas de mejoramiento genético. Por lo tanto, el objetivo de este proyecto fue el descubrimiento del polimorfismo de nucleótido simple (PNS) mediante la secuenciación de bibliotecas de representación reducida (BRR) de alpaca. Se utilizó una muestra de 150 alpacas Huacaya de cuatro granjas, dos en Puno y dos en Cerro de Pasco para el descubrimiento de PNS por genotipado por secuenciación (GBS). Se produjeron 02 BRR por animal, mediante la digestión del ADN con ApeK1 y con Pst1-Msp1. Se usaron diez genomas de alpaca, secuenciados a profundidades entre 12X y 30X, y el genoma de referencia VicPac3.1 para las alineaciones con las lecturas de los BRR. 76,508 PNSs fueron incluidos en la micromatriz de PNSs de alpaca, de los cuales 302 PNS se encontraban en genes candidatos para la calidad y el color de la fibra. Los PNSs de la micromatriz cubrió el 90.5 por ciento de la longitud del genoma con una densidad de  $39 \pm 2.51$  PNS/Mb de ADN en un intervalo medio de  $26.45 \pm 18.57$  kbp. El rendimiento se evaluó con el genotipado de 30 tríos y comparados con su pedigrí, asimismo se comparó los genotipos por micromatriz y GBS. Se observaron valores de concordancia de 0.93 y 0.94 para los PNS generados por ApeK1 y Pst1-Msp1. La disponibilidad de PNSs en esta micromatriz facilitará los estudios de asociación de todo el genoma, la selección asistida por marcadores y, con el tiempo, la selección genómica.

Palabras clave: Polimorfismos de nucleótido simple (PNSs), micromatriz, alpaca, Vicugna pacos, genes de fibra.

## ABSTRACT

Small farm producers' sustenance depends on their alpaca herds and the production of fiber. Genetic improvement of fiber characteristics would increase their economic benefits and quality of life. The incorporation of molecular marker technology could overcome current limitations for the implementation of genetic improvement programs. Hence, the aim of this project was the discovered of single nucleotide polymorphism (SNP) by sequencing reduced representation libraries (RRB) of alpaca. A sample of 150 Huacaya alpacas from four farms, two each in Puno and Cerro de Pasco were used for SNP discovery by genotyping by sequencing (GBS). The RRB, two per animal, were produced after DNA digestion with ApeK1 and double digestion with Pst1-Msp1. Ten alpaca genomes, sequenced at depths between 12X to 30X, and the VicPac3.1 reference genome were used for read alignments. 76,508 SNPs were included in the alpaca SNP microarray, where 302 SNPs were located in candidate genes for fiber quality and color. The microarray SNPs covered 90.5 percent of the genome length with a density of about  $39 \pm 2.51$  SNPs/Mb of DNA at an average interval of  $26.45 \pm 18.57$  kbp. The performance was evaluated by genotyping 30 family trios and comparing them to their pedigrees, as well as comparing microarray to GBS genotypes. Concordance values of 0.93 and 0.94 for ApeK1 and Pst1-Msp1 generated SNPs were observed. Availability of this microarray SNPs will facilitate genome-wide association studies, marker-assisted selection and, in time, genomic selection.

Keywords: Single nucleotide polymorphisms (SNPs), microarrays, alpaca, Vicugna pacos, fiber genes.

## I. INTRODUCCIÓN

La población de alpacas en Perú es de 3'685,516 unidades (CENAGRO 2012), lo que representa el 85 por ciento de la población mundial de esta especie doméstica. Las alpacas son aprovechadas por su producción de fibra, lo que ubica al Perú como el primer productor y exportador mundial de este producto para la industria textil (Quispe *et al.* 2009). La mejora genética de la producción y calidad de fibra podría contribuir a incrementar los beneficios económicos y mejorar la calidad de vida de los productores. Sin embargo, quedan por superar varios desafíos para la implementación de programas de mejora genética. Por nombrar algunos, la dispersión de pequeños productores que crían alrededor del 90 por ciento de la población de alpaca en condiciones de manejo extensivo sin implementación de manejo reproductivo y sin registros de producción. Por lo tanto, los programas de mejoramiento genético a nivel de pequeños productores son inexistentes. Algunos de estos desafíos pueden resolverse con el desarrollo y la implementación de tecnologías nuevas junto con programas de capacitación.

En la última década se han producido varios avances en la comprensión de la organización del genoma de la alpaca (Richardson *et al.* 2019). Además, se han desarrollado micromatrices de polimorfismo de nucleótido simple (PNS) para varias especies de animales domésticos con fines de mejora genética. En cada caso, los investigadores se propusieron identificar un gran número de polimorfismos de nucleótido simple (PNSs) localizados en intervalos equidistantes a lo largo del genoma.

El uso de una micromatriz de PNSs en alpacas permitiría la comparabilidad de los datos generados por estudios independientes, brindando una relevancia sobre la información genética y realizar futuras investigaciones. Las ventajas y desventajas inherentes a la micromatriz de PNS y el genotipado por secuenciamiento (GBS) no son el alcance de este estudio. La metodología GBS permite identificar PNSs ubicados aleatoriamente en todo el genoma e identifica PNSs raros con mayor frecuencia. La precisión de GBS en la lectura de genotipos heterocigotos depende de la profundidad de lectura de los PNS por estudio de población, por lo que se requieren más análisis bioinformáticos. Sin embargo,

GBS tiene la ventaja de ser más rentable que la genotipificación de micromatriz. Fan *et al.* (2020) han confirmado con éxito las relaciones evolutivas entre los camélidos sudamericanos basadas en GBS. En general, el uso de un método sobre el otro depende de los objetivos de cada estudio.

La tecnología de la inseminación artificial con semen diluido y el desarrollo de la tecnología de conservación de semen a largo plazo ampliará el uso de reproductores seleccionados (Ordoñez *et al.* 2013, Huanca 2015). De manera similar, el desarrollo de una micromatriz de PNSs facilitaría la implementación de programas de mejora genética, por esta razón el objetivo del presente estudio fue la identificación de polimorfismos de nucleótido simple por secuenciamiento de bibliotecas de representación reducida de ADN en alpacas Huacaya.

## II. REVISIÓN BIBLIOGRÁFICA

### 2.1. LA ALPACA

La alpaca es un animal perteneciente al orden Artiodactyla, suborden Tylopoda y a la familia *Camelidae*. El antecesor de esta familia se originó en Norteamérica, con la división de la familia *Camelidae* en 2 tribus, la tribu Camelini que son los actuales camellos, el dromedario (*Camelus dromedarius*), el bactraciano (*Camelus bactrianus*) y el bactriano salvaje (*Camelus ferus*), estos se distribuyen en Eurasia y África y la tribu Lamini donde se encuentran la llama (*Lama glama*), el guanaco (*Lama guanicoe*), la vicuña (*Vicugna vicugna*) y la alpaca (*Vicugna pacos*), quienes se distribuyen en el sur de América (Wheeler 1995, Wheeler *et al.* 2004).

Un estudio a través de un análisis filogenético, confirmó que la alpaca fue domesticada de la vicuña y la llama del guanaco, en el proceso de domesticación el flujo de alelos trasladados a lo que se denomina introgresión fue mucho mayor en el genoma de la alpaca (36 por ciento) que en la llama (5 por ciento). Sin embargo, también se reporta que algunos de los segmentos introgresados permitieron nuevas vías de selección, por ejemplo, para el color del pelaje, las características de las fibras y la adaptación a la altitud (Fan *et al.* 2020).

#### 2.1.1. Población de alpaca

La población de alpaca en el Perú es de 3'685,516 ejemplares, que pertenecen a 82,459 productores agropecuarios. Las principales regiones que concentran la producción de alpaca son: Puno (39.6 por ciento), Cusco (14.7 por ciento), Arequipa (12.7 por ciento), Huancavelica (8.3 por ciento), Apurímac (5.9 por ciento), Ayacucho (2.8 por ciento) y Pasco (1.8 por ciento) (CENAGRO 2012). Según estimaciones realizadas por el Ministerio de Agricultura y Riego (MINAGRI 2020), la población mundial de alpacas



ascendió a más de 6 millones de unidades, siendo Perú el de mayor población con 4 millones 492 mil 025 ejemplares. La población de alpacas en el Perú representa el 85 por ciento de la población mundial de esta especie doméstica. Las alpacas están especializadas en la producción de la fibra llegando a posicionar al Perú como el primer productor y exportador mundial de este producto para la industria textil (Quispe *et al.* 2009).

La crianza de alpacas constituye una de las actividades productivas y económicas más importantes que se desarrolla en la zona alto andina y de ella depende entre el 70 a 80 por ciento del ingreso familiar anual (Fernández-Baca 2005). Las alpacas de raza Huacaya, representan el 80.4 por ciento, seguida de la raza Suri con 12.2 por ciento y animales cruzados que representan el 9.2 por ciento de la población de alpacas del Perú (CENAGRO 2012).

La cría de la alpaca Huacaya está dirigida a la producción de fibra; sin embargo, en los rebaños de pequeños y medianos productores, no se emplean técnicas de manejo adecuadas, entre éstas, la selección genética. En los rebaños de grandes cooperativas, asociaciones, o empresas transformadoras, la selección genética se realiza dentro de los rebaños principalmente en base a rasgos fenotípicos, observándose que algunas empresas están trabajando con evaluaciones genéticas en base a registros de control de producción y genealogía (Cruz 2018).

Si bien, el producto principal de la crianza de alpaca es su fibra, la cual posee buenas características textiles como suavidad y longitud, la mayoría de los rebaños requiere mejorar la finura de sus vellones. Esta fibra de alpaca se comercializa principalmente a mercados extranjeros, generando ingreso de divisas al país. Por tanto, la mejora de la calidad de sus productos, a través del mejoramiento genético, contribuirá al incremento de los ingresos de los criadores y de su comercio (Quispe *et al.* 2009).

Los camélidos sudamericanos, incluida la alpaca, presentan características biológicas con ventajas para afrontar los futuros escenarios climáticos de la zona alto andina. Se predice para las zonas de puna una disminución y un cambio de régimen de las precipitaciones, así como una elevación de la temperatura; lo que conllevaría a una menor disponibilidad de agua y al cambio de vegetación de gramíneas por especies arbustivas (Carazas 2007). En comparación con ovinos y vacunos, por origen, los camélidos andinos estarían en ventaja biológica para enfrentar este escenario. Investigaciones realizadas han

demostrado que tienen una mayor capacidad de aprovechamiento de pastos con alto contenido de fibra celulósica, y menores requerimientos de consumo de materia seca y agua (Cerón 2015).

### **2.1.2. Características fenotípicas de la alpaca**

Existen dos razas de alpacas y se diferencian por sus características fenotípicas. La raza Huacaya (Figura 1) es caracterizada por presentar cabeza pequeña, comprimida, tiene un copete de pelos que llega hasta los ojos y los cubre completamente, las orejas son pequeñas, verticales y puntiagudas; el peso adulto fluctúa entre los 56 y 62 kg; los adultos presentan una altura a la cruz de 0.8 a 1 metro. La raza Suri, fenotípicamente esta raza se caracteriza por ser más pequeña que la Huacaya, además es menos robusta y susceptible a los cambios de ambiente; la fibra carece de rizos, pero es de mayor longitud que la Huacaya; la fibra cae paralelamente al cuerpo, dejando una línea superior descubierta de fibra (Cruz 2018).



**Figura 1: Alpaca de la raza Huacaya.**  
Fuente: Adaptado de Cruz (2018).

## 2.2. GENOMA DE LA ALPACA

El genoma se describe como el total de ADN (ácido desoxirribonucleico), representados en secuencias de nucleótidos, que está presente en el núcleo de la célula, incluyendo el ADN nuclear y mitocondrial. La secuencia de nucleótidos del ADN que contienen información genética para generar variedades de ARN se denomina gen. En particular, el ARNm transcrito llevado a un proceso de traducción generara una proteína específica (Griffiths *et al.* 2015).

A partir del 2002, se inició al estudio del genoma de la alpaca, después de secuenciar un genoma de alpaca con baja cobertura (2X), paralelamente con 21 especies de mamíferos, con el objetivo de implementar y mejorar la anotación del genoma humano a través de la genómica comparativa. Asimismo, se tuvo en cuenta la familia *Camelidae* con la finalidad de incrementar el número de especies con un genoma disponible para estudios evolutivos y comparativos (Raudsepp 2012).

En el 2008, se presentó el ensamblaje *VicPac 1.0*, secuenciado con la tecnología de Sanger con cobertura de 2.5X de una alpaca hembra (Carlotta). Tres años después, en el 2011 se reportó a una cobertura de 22X, *VicPac 2.0.2* (*VicPac 1.0* y se adiciono la tecnología de secuenciamiento de segunda generación), el ensamblado cuenta 276,726 scaffolds y un total de  $2.1 \times 10^9$  pb realizado por la universidad de Washington, disponible en la plataforma del National Center for Biotechnology Information (NCBI). En el año 2014, se reportó un genoma ensamblado a 72.5X denominado *Vi\_Pacos 1.0*, elaborado por la Universidad de Mongolia (IMAU) y BGI (China) utilizando la plataforma HiSeq 2000 de Illumina; el ensamblado cuenta con 52,275 scaffolds y un total de  $2.01 \times 10^9$  pb (Wu *et al.* 2014).

El año 2019 se reportó el ensamblado a 260X denominado *VicPac 3.1*, (NCBI, GCA\_000164845.4), realizado por un consorcio de investigación liderado por la Universidad de Deakin. Este ensamblado cuenta con 77,389 scaffolds, de los cuales 88 scaffolds están asignados a cromosomas que representan el 76 por ciento de la longitud del genoma de referencia, la diferencia representada por 77,301 scaffolds, no está asignada a cromosomas (Richardson *et al.* 2019).

Adicionalmente se cuenta con una biblioteca de cromosoma artificial bacteriano denominada BAC CHORI-246, elaborada con muestras biológicas de la alpaca Carlotta y está compuesta por un total de 202,752 clones (Ávila 2014).

Por otro lado, varios estudios citogenéticos en la familia *Camelidae*, han demostrado que existe una similitud en la conformación del cariotipo entre sus especies. El cariotipo de la alpaca está conformado por  $2n=74$  cromosomas (Balmus *et al.* 2007). Las técnicas de bandeos cromosómicos como los patrones de banda G (Giemsa), C (Heterocromatina Centromérica) y NOR (Región Organizadora Nuclear) permitieron identificar que los pares cromosómicos autosómicos del 1 al 20 son de tipo acrocéntrico–subterminal y del 21 al 36 son metacéntrico–submetacéntrico y los cromosomas sexuales X e Y son de tipo metacéntrico y acrocéntrico, respectivamente (Bianchi *et al.* 1986).

Se han identificado regiones cromosómicas homologas entre camélidos, bovinos y humanos por Zoo-FISH (Balmus *et al.* 2007), y varios genes y marcadores moleculares han sido localizado en cromosomas mediante hibridación fluorescente in situ (FISH), lo que permitió el desarrollo del primer mapa citogenético integrado del genoma de alpaca. El mapa citogenético está conformado por 230 marcadores moleculares ordenados y distribuidos en los 37 pares de cromosomas. Adicionalmente, 86 genes de alpacas fueron mapeados en el dromedario, observándose secuencias sinténicas entre ambos, afirmando que ambas especies son muy similares en su genoma (Ávila *et al.* 2015). Se ha reportado una nueva versión del mapa citogenético de la alpaca con 281 marcadores (Pauciullo *et al.* 2019, Mendoza *et al.* 2019, Mendoza *et al.* 2020).

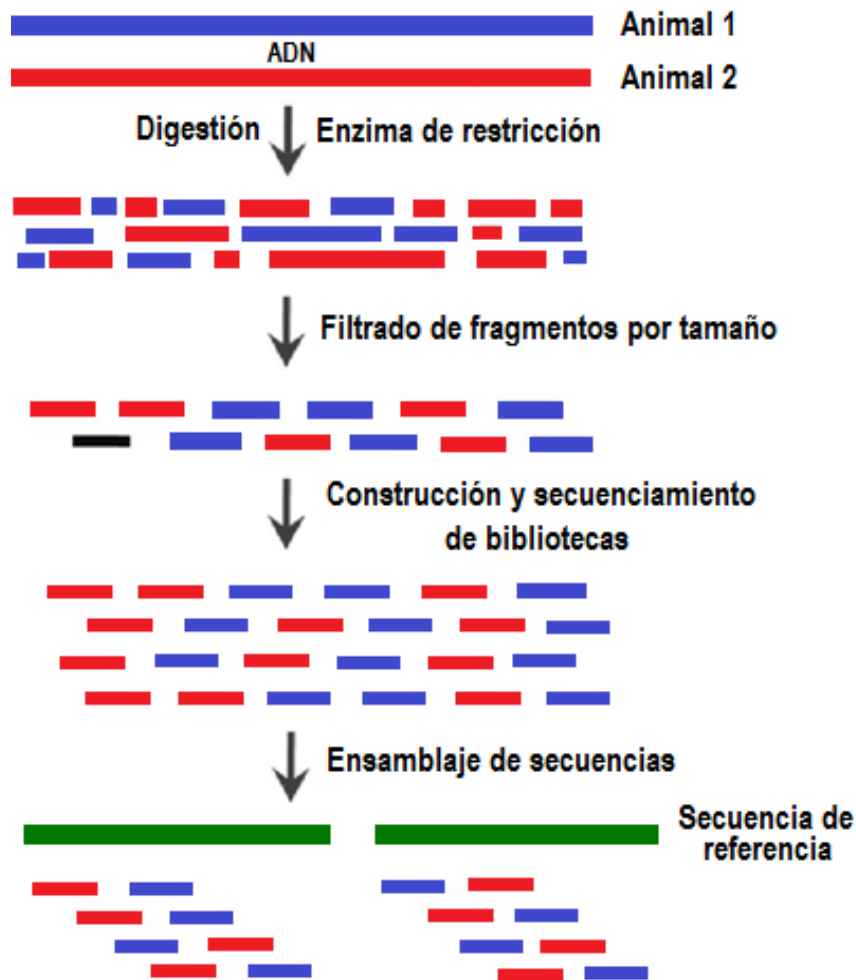
Se desarrollaron diferentes metodologías para encontrar genes asociados a caracteres productivos o enfermedades en genomas de diferentes especies utilizando marcadores moleculares (Ponce de León y Gutiérrez 2020). Los marcadores moleculares son una secuencia de ADN utilizada para identificar la ubicación de una región específica de interés dentro del genoma y a la vez identificar variantes (polimorfismos) a nivel de ADN entre 2 a más individuos de la misma especie, y que pueden estar localizados en regiones no codificantes o codificantes. Entre los marcadores moleculares que se utilizaron en estudios de animales de granja tenemos: Polimorfismos de longitud en fragmentos de restricción, microsatélites y actualmente los polimorfismos de nucleótido simple (Dekkers 2012, Al-Samarai y Al-Kazaz 2015), siendo el más eficiente el PNS a

comparación de microsatélites para estudios de diversidad genética (Zimmerman *et al.* 2020).

El estudio del genoma de la alpaca ha progresado lentamente en comparación con otras especies de animales domésticos (Ponce de León y Gutiérrez 2020). Existen pocos trabajos que relacionen características fenotípicas de interés a genes y/o marcadores en su genoma, así como evaluar la diversidad genética a nivel de genes específicos (Lado *et al.* 2020a). Un estudio inicial, consideró 140 alpacas de ambas razas (Huacaya y Suri) donde evaluaron 69 marcadores moleculares tipo microsatélite, con la finalidad de investigar la asociación con el diámetro de fibra. De los 69 marcadores se identificaron 4 microsatélites (*LCA68*, *VOLP59*, *LCA90* y *GLM6*) asociados significativamente para el carácter de diámetro de fibra con el valor de cría (Paredes *et al.* 2014).

### **2.3. BIBLIOTECAS DE REPRESENTACIÓN REDUCIDA DE ADN**

Una biblioteca de representación reducida (BRR) de ADN está compuesta por un grupo de fragmentos filtrados de diferentes longitudes de ADN (Altshuler *et al.* 2000). La construcción de las bibliotecas de representación reducida de ADN se realiza mediante el uso de enzimas de restricción (ERs). Se usan las ERs para reconocer y cortar en secuencias específicas de 4-6 nucleótidos y obtener el mayor número de fragmentos posibles de los que se seleccionan fragmentos de ADN que estén entre 100 a 300 nucleótidos (Figura 2). Este método es rápido, extremadamente específico, altamente reproducible y puede alcanzar regiones importantes del genoma (Metzker 2010).



**Figura 2: Elaboración de bibliotecas de representación reducida**  
Fuente: Adaptado de Fan *et al.* (2010)

Cada biblioteca de representación reducida representa una fracción del total de ADN de un individuo en particular, lo que constituye una reducción del genoma, gracias a enzimas de restricción que realizan cortes específicos en el ADN reduciendo la complejidad y eliminando las regiones repetidas o poco informativas (Altshuler *et al.* 2000). La creación de BRR ha sido uno de los métodos altamente eficientes para la detección e identificación de PNSs a lo largo del genoma, aplicado en aves, porcinos y bovinos (Van Tassell *et al.* 2008, Wiedmann *et al.* 2008, Fan *et al.* 2010, Du *et al.* 2012, Bovo *et al.* 2015, Zhu *et al.* 2016), en humanos (Altshuler *et al.* 2000), soya (Wu *et al.* 2010), pavos (Aslam *et al.* 2012), conejos (Bertolini *et al.* 2014) y trucha arco iris (Sánchez *et al.* 2009).

Por ejemplo, en el cerdo Wiedmann *et al.* (2008) y Ramos *et al.* (2009) secuenciaron BRR para la identificación de PNS que se utilizaron para construir la plataforma de genotipado Illumina PorcineSNP60 BeadChip. También, se identificaron alrededor de 40k PNS en el genoma del cerdo (Bovo *et al.* 2015). En comparación con otros métodos, la BRR basado en la digestión ADN con enzimas de restricción es relativamente factible y flexible, especialmente para aquellas especies sin genoma de referencia (Wiedmann *et al.* 2008). Las BRR, también permitió la identificación hasta 47k PNS con una tasa de validación del 48 por ciento en un genoma de trucha arco iris (Sánchez *et al.* 2009).

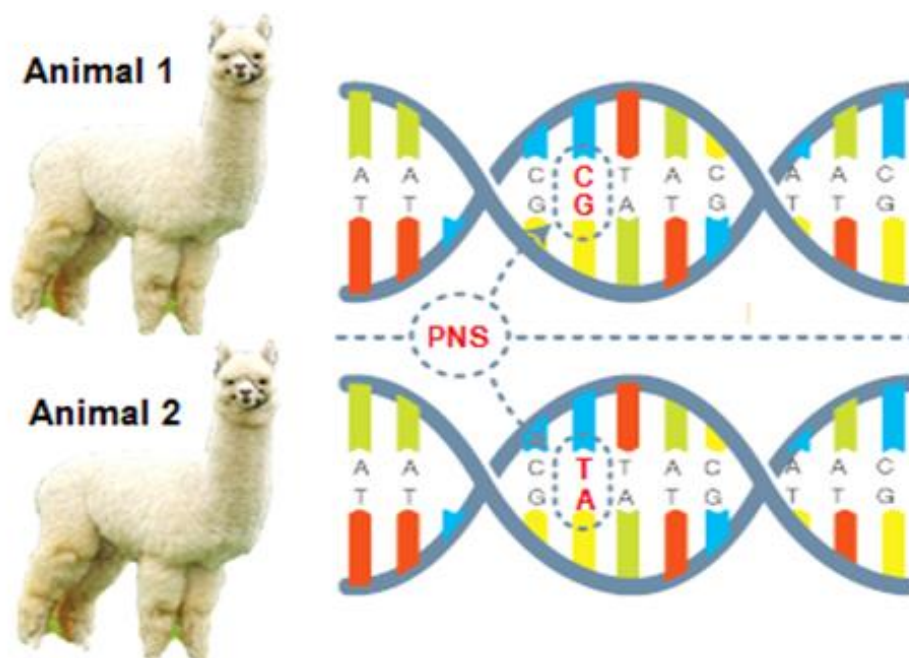
El uso de BRR combinado con la secuenciación de alto rendimiento es una forma eficaz para descubrir PNS en individuos o poblaciones, con una alta cobertura y especificidad, llegando a identificar más de 257k de PNS en humano (Du *et al.* 2012). Basado en la construcción y secuenciación de BRR generados de dos grupos de porcinos italianos Large White con valor genético estimado (VGE) para grosor de la grasa dorsal (GGD) extremo y divergente (50 con el VGE GGD más negativo y 50 con el VGE GGD más positivo), una estrategia para la identificación de polimorfismos a una fracción reducida del costo requerido para la secuenciación individual (Bovo *et al.* 2015).

La secuenciación de BRR de diferentes especies proporcionaron la identificación de una gran cantidad de PNS y la estimación de la frecuencia de los alelos (Van Tassell *et al.* 2008). La secuenciación de BRR también se ha utilizado con éxito para el descubrimiento de genes (Timko *et al.* 2008), análisis de metilación (Meissner *et al.* 2005) y caracterización genómica (Paterson 2006). Este método permite analizar y estudiar varios animales no relacionados de diferentes razas y secuenciar a una profundidad suficiente para una identificación confiable de PNS con un MAF alto (Wiedmann *et al.* 2008).

A través del análisis de secuencias basadas en BRR y alineadas con el genoma de referencia *VicPac 3.1* (Richardson *et al.* 2019) se identificaron una gran cantidad de PNSs y genes candidatos en alpacas (Fan *et al.* 2020). Además, se realizó un análisis filogenético sobre las relaciones evolutivas entre los camélidos sudamericanos que confirmó que la llama fue domesticada del guanaco y la alpaca de la vicuña. (Fan *et al.* 2020).

## 2.4. POLIMORFISMO DE NUCLEÓTIDO SIMPLE

El polimorfismo de nucleótido simple (PNS) representa una variación en una secuencia de ADN que afecta al cambio de un solo nucleótido (Figura 3). Una gran cantidad de PNSs existen a lo largo del genoma y para que los PNSs sean considerados marcadores genéticos un alelo debe estar presente en al menos el 1 por ciento de la población, deben ser estables genéticamente y poseer una baja tasa de mutaciones (Brookes 1999).



**Figura 3: Polimorfismo de nucleótido simple**

Fuente: Adaptado Griffiths *et al.* (2015)

La gran mayoría de los PNSs son bi-alélicos, es decir, tienen dos alelos los cuales están representados por una sustitución de una base por otra presente en cualquier posición de un segmento de ADN, incluyendo las transiciones purina-purina (A-G) o pirimidina-pirimidina (C-T) y las transversiones purina-pirimidina o pirimidina-purina (A-C, A-T, G-C o G-T) (Rao y Gu 2008). Los PNSs se encuentran en gran número en todo el genoma, razón por la que se considera como el marcador de elección para la evaluación de diversos estudios moleculares (Piro 2021). El desarrollo exitoso de estudios de asociación entre PNSs y características de importancia económica en animales de granja será posible cuando un mayor número de PNSs sean identificados, abarcando de manera representativa todo el genoma (Wang 2002).



Los marcadores moleculares tipo PNSs han adquirido importancia porque son fáciles de evaluar e interpretar por su distribución a lo largo del genoma, se calcula que en células eucariotas se presentan uno por cada mil pares de bases en una cadena de ADN (Lado *et al.* 2020a). Además, se han desarrollado micromatrices para diferentes especies con las cuales se pueden analizar simultáneamente miles de PNSs en una sola reacción (Ponce de León y Gutiérrez 2020). Asimismo, se incrementó la eficiencia para identificar los marcadores, se redujo el costo de genotipificación por marcador y por animal obteniendo una cobertura mayor del genoma (Viale *et al.* 2017). Además, el conjunto de PNSs facilita el análisis de haplotipos (combinación de alelos de PNSs en un determinado cromosoma) y poder evaluar el grado de diferenciación entre poblaciones (Vignal *et al.* 2002) y entre individuos (Chen *et al.* 2013).

Los PNSs han sido ampliamente usados en la identificación y determinación de parentescos en diferentes especies de animales, como *Bos taurus* (Heaton *et al.* 2002), y su asociación a genes de interés en *Ovis aries* (Kijas *et al.* 2009) y *Sus scrofa* (Hidalgo *et al.* 2013). También, se han reportado en ganado Simmental la asociación de PNSs a caracteres de carcasa y calidad de carne (Wu *et al.* 2014); mapeo de genes relacionados al color de fibra en equinos (McCue *et al.* 2012); asociación de PNSs y haplotipos con el color de pelaje y rasgos en cabras Mohair (Nazari-Ghadikolaei *et al.* 2018); y, la generación de borradores de genomas en especies relacionadas a las ovejas (Miller *et al.* 2015). También han sido utilizados en estudios para el mejor entendimiento de la evolución de las especies, formación y domesticación de las razas, elaboración de nuevas teorías en genética de poblaciones y mejorar el desarrollo del mejoramiento genético animal (Fan *et al.* 2010, Carhuaricra *et al.* 2022).

Existen pocos estudios de PNSs en alpacas, un estudio de simulación pudo determinar que un panel de 100 PNSs con valores de la frecuencia del alelo menor (MAF) de 0.35 proporciona una asignación de paternidad casi perfecta para los sistemas de apareamiento tradicionales (controlado, alternado y múltiple) y que se requieren aproximadamente 40 PNSs para dar un poder de exclusión comparable a un panel de 10 microsatélites (Paucar 2011). Además, se han realizado algunos estudios en la identificación de marcadores moleculares y su asociación con las características de la fibra, así Delgado de la Flor (2014) realizó la identificación de PNSs en el gen de Tricohialina y estudió su posible relación con el fenotipo Huacaya y Suri, donde reportó la ubicación de dos PNS que se encuentran en desequilibrio de ligamento, sin embargo, no encontró relación con el

fenotipo Huacaya o Suri. También, Foppiano (2016) ubicó 27 PNSs en los genes *KRTAP1-2* (10), *KRTAP6-1* (9), *KRTAP9-2* (3), *KRTAP11-1* (6) y *KRTAP13-1* (4) todos estos genes de proteínas asociadas a la queratina, sin embargo, no encontró asociación entre el gen *KRTAP11-1* y el diámetro de fibra en alpaca.

Por otro lado, Rodríguez (2014) comparó poblaciones de alpacas de diferentes regiones utilizando ADN mitocondrial y nuclear, donde reportó moderada variabilidad en 8 PNSs en los genes *K33A*, *KAP1.1* y *KAP11.1*. Mamani *et al.* (2017) reportó 50,686 PNSs en alpacas a partir de un panel celular híbrido de alpaca/hámster y una micromatriz de alta densidad de bovino, concluyendo que es posible el uso de micromatrices de bovinos en alpacas.

Fernández *et al.* (2019), predejon la ubicación de 14 PNS en genes de queratina del epitelio de la piel de tipo 1, 23 PNS en genes de queratina del epitelio de la piel de tipo 2, 4 PNS en genes de queratina de la corteza de la fibra de tipo 1 y 6 PNS en genes de queratina de la corteza de la fibra de tipo 2; genes candidatos para la calidad de la fibra. Además, More *et al.* (2019) identificó 400 PNSs en alpacas utilizando una micromatriz de PNS de bovino de alta densidad, de los cuales 292 PNS fueron asignados a cromosoma de la alpaca.

Chandramohan *et al.* (2015) investigaron y caracterizaron el gen *MC1R*, su transcripción y el polimorfismo de nucleótido simple asociado con el color de fibra blanco. Además, reportaron la presencia de dos transcritos, denominados F1 y F2, que varían solo en la longitud de sus secuencias de la región no traducida (UTR) 5'-terminal y presentaban una expresión específica de color, el transcrito F1 del gen *MC1R* fue común en los fenotipos de alpaca de color café, blanco y negro; mientras que el transcrito F2 del gen *MC1R* fue específico de la alpaca blanca. Además, al secuenciar el gen *MC1R* en alpaca blanca y de color se identificó un total de 12 PNSs, de los cuales 9 se encontraron en fragmentos codificantes y 3 en no codificantes.

Se han reportado PNS en genes candidatos para el color de la fibra en alpacas (*MC1R*, *ASIP*, *KIT*). Feeley y Munyard (2009) identificaron 21 PNSs, Guridi *et al.* (2011) encontraron 18 PNSs y Chandramohan *et al.* (2015) identificaron 12 PNSs en el gen *MC1R*. Por otro lado, Feeley *et al.* (2011) reportaron 10 PNSs y Chandramohan *et al.*

(2013) identificaron dos PNSs en el gen *ASIP*. Además, Jones *et al.* (2019) identificaron 47 PNSs en el gen *KIT*, gen candidato para el fenotipo gris en alpacas.

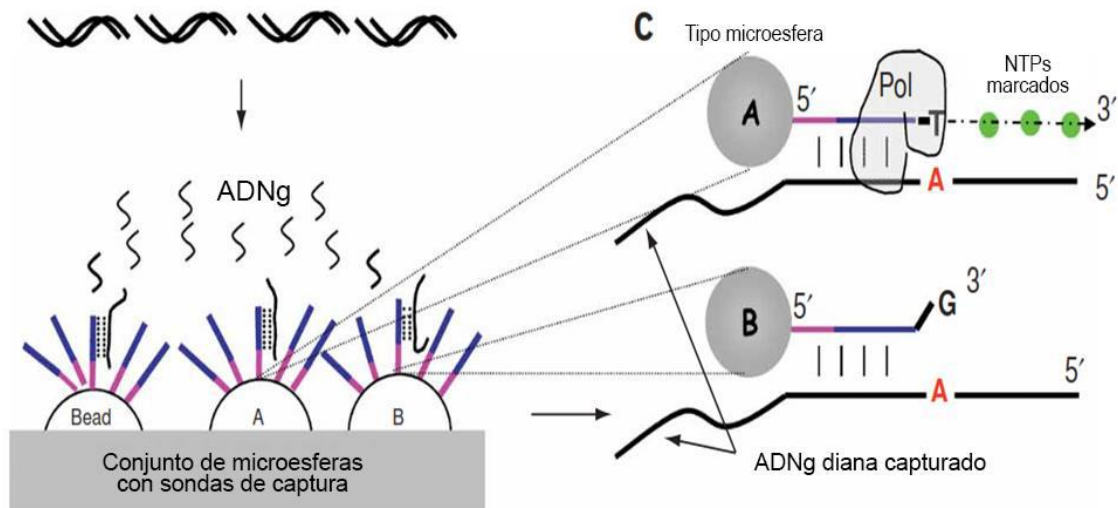
Figuerola (2022) identificó 67 PNSs dentro de genes de proteína asociados a la queratina, para ello utilizó el alineamiento de nueve genomas secuenciados y 300 bibliotecas de representación reducida de ADN de alpacas. Asimismo, confirmó la presencia de 32 PNSs en una población de 150 alpacas, para ello consideró la frecuencia del alelo menor  $\geq 0.05$ , una puntuación de Illumina  $\geq 0.6$  y la tasa de genotipado  $> 45$  por ciento por PNS.

## 2.5. MICROMATRIZ DE PNSs

La micromatriz de PNSs, consiste en múltiples fragmentos de ADN que se encuentra adheridos a un soporte pequeño fabricado en plástico, sílice o vidrio. Los fragmentos que se adhieren son oligonucleótidos conocidos como sondas, estableciendo un patrón de orden. Estas sondas son secuencias de nucleótidos y pueden ser sintetizadas *in vitro*, amplificadas por PCR o clonadas. Cada sonda adherida permite genotipar un PNS y contiene por lo menos 50 nucleótidos que terminan en el nucleótido polimórfico del PNS en la posición 3' (Dalma-Weiszhausz *et al.* 2006).

El ADN genómico para ser analizado, el primer paso es cortar en pequeños fragmentos y los fragmentos de interés pueden ser amplificados por PCR. Estas secuencias amplificadas serán marcadas con un colorante fluorescente para que a continuación sean hibridadas con la micromatriz (Ravi *et al.* 2014). La fluorescencia se generará en una mínima y mayor intensidad según sea que la secuencia analizada concuerde o no con las sondas de la micromatriz. Esta micromatriz puede analizar miles de miles de PNS, por animal, en una simple reacción (Brown 2007).

Una micromatriz incluye secuencias sintetizadas de 36 a 50 nucleótidos que corresponden a las secuencias flanqueantes de cada PNS denominada sonda. La hibridación de las sondas incorporadas que contiene la micromatriz con el ADN de la muestra a analizar, está sujeto a una condición específica y requerimientos muy estrictos (Figura 4). Para la cuantificación y detección en la hibridación de la sonda con la muestra objetivo tiene que estar etiquetado con el fluoróforo, estas señales ayudan a determinar la relación activa de las secuencias del ácido nucleico en el objetivo (Ravi *et al.* 2014).



**Figura 4: Genotipado del genoma con una micromatriz**

Fuente: Gunderson *et al.* (2005)

### 2.7.1 Micromatriz con la tecnología BeadArray

En el sistema de micromatrices Illumina (tecnología BeadArray), la superficie de cada matriz, puede contener millones de sondas de PNSs (<https://www.illumina.com/science/technology/microarray.html>), esta micromatriz se basa en el ensamblaje aleatorio de beads en micropocillos ubicados al final de un haz de fibra óptica (Michael *et al.* 1998). A inicios del año 2000 una micromatriz tuvo una capacidad de contener hasta 50,000 beads y con una longitud de 3 micras de diámetro. Los beads se distribuyen entre 1,520 tipos de beads, cada tipo de bead representa una secuencia de sonda de oligonucleótidos diferente (Oliphant *et al.* 2002).

La tecnología Bead-Array se basa en la precisión, incluida la reproducibilidad, la correlación de hebras, la concordancia con otros métodos de genotipado y la coherencia de la herencia mendeliana; cada medida tiene fortalezas y debilidades (Oliphant *et al.* 2002, Shen *et al.* 2005). La relación entre la precisión y la tasa de genotipado se considera como una métrica de rendimiento de genotipado. La tasa de genotipado es la fracción de llamadas de genotipos entre el total de genotipos posibles. En un estudio del cromosoma 20 del humano, se reportó 1,585,219 genotipos llamados de un total de 1,585,740, obteniéndose una tasa de genotipado del 99.97 por ciento (Fan *et al.* 2003). Illumina ha desarrollado un software de genotipado que realiza el llamado de todos los genotipos y brinda un puntaje de calidad (More *et al.* 2022).

### **2.7.2 Micromatriz con la tecnología GeneChip**

El sistema de micromatrices de Affymetrix (tecnología GeneChip) utiliza matrices fabricadas por síntesis directa de oligonucleótidos (sondas) en la superficie del vidrio utilizando la tecnología fotolitográfica empleada en la industria de los semiconductores (Gautier *et al.* 2004). La técnica y característica principal de un GeneChip es la utilización de un par de secuencias (probe pairs) cada una de ellas presentan una característica específica: la primera tiene exactamente la secuencia correspondiente al PNS, llamada Perfect Match Probe (PM) y la segunda ha sufrido un cambio en el nucleótido central de la probe sequence, y recibe el nombre de Mismatch Probe (MM) (Shapero *et al.* 2014; Dalma-Weiszhausz *et al.* 2006). De esta manera se obtiene un balance óptimo entre la sensibilidad y la especificidad del chip ([https://www.affymetrix.com/about\\_affymetrix/outreach/educator/index.affx](https://www.affymetrix.com/about_affymetrix/outreach/educator/index.affx)).

### **2.7.3 Micromatrices de polimorfismos de nucleótido simple en animales de granja**

Se han desarrollado micromatrices de PNSs para varias especies de animales domésticos. En cada caso, los investigadores se propusieron identificar un gran número de PNS informativos localizados a intervalos equidistantes a lo largo del genoma (Ramos *et al.* 2009, Matukumalli *et al.* 2009, Groenen *et al.* 2011, Darrier *et al.* 2019, Sabahat *et al.* 2020). La aplicación del genotipado con una micromatriz de alta densidad puede usarse para estudios genómicos y puede mejorar la comprensión de los procesos evolutivos en los animales de granja (Mustafa *et al.* 2018).

La micromatriz de bovino de 50k PNS, se elaboró en base a PNSs identificados específicamente para la micromatriz y adicionalmente PNS obtenidos de publicaciones científicas y disponibles en bases de datos. Los PNSs fueron identificados mediante el secuenciamiento y alineación de tres BRR de ADN provenientes de tres grupos de animales. El primer grupo fue de 15 vacas Holstein, el segundo de 35 toros Angus y el tercero de dos toros de cada una de las razas Charolais, Gelbvieh, Hereford, Limousin, Red Angus y Simmental. El propósito que tuvieron para usar estos dos grupos de razas puras y un grupo de razas mixtas fue para mejorar la probabilidad de identificar PNSs de secuencia entre animales dentro de una raza y entre razas (Matukumalli *et al.* 2009).

Se utilizó un enfoque similar para el desarrollo de la micromatriz de porcinos de 60k PNS, se utilizó 19 BRR generados a partir de 158 muestras de ADN de cuatro razas de porcinos (Duroc, Landrace, Large White, Piétrain) y una población de porcinos silvestres (Wild Board) procedentes de EE. UU., Países Bajos, Dinamarca, Europa y Japón. Esta estrategia permitió identificar 372k de PNSs, de los cuales se seleccionaron 64k para su inclusión en la micromatriz (Ramos *et al.* 2009).

De manera similar, la micromatriz de aves de 60k PNS fue desarrollada, para ello se utilizó 25 muestras de ADN obtenidas de cada una de 2 líneas de tipo carne y 2 líneas de tipo ponedoras, a partir de estas muestras, generaron 4 BRR con mezclas de ADN de cada línea de aves, luego de secuenciar y filtrar las secuencias obtenidas, identificaron 561k PNSs, de los cuales seleccionaron 61k PNSs para su inclusión en la micromatriz (Groenen *et al.* 2011)

Una micromatriz de alta densidad comercial de PNSs de bovino (777k BeadChip), fue utilizada para genotipar 20 alpacas y 25 dromedarios con el propósito de identificar PNS que podrían estar presente en el genoma de la alpaca y dromedario, el resultado demostró que la micromatriz de bovino pudo genotipar con éxito un 10.5 por ciento y 20.7 por ciento de los PNSs en alpaca y dromedario, respectivamente (Bertolini *et al.* 2016).

Las micromatrices de PNSs se han aplicado en estudios de asociación del genoma completo (GWAS) y permitieron incrementar la precisión en la identificación de QTLs para rasgos de interés en animales domésticos (Piro 2021). Estos hallazgos facilitarán la comprensión de la arquitectura genética de rasgos complejos en animales domésticos y la buena práctica de los programas de mejoramiento genético (Zhang *et al.* 2012).

La información de PNS obtenida del genotipado permitiría el cálculo de las relaciones genómicas (Yang *et al.* 2011), permitiendo corregir la información genealógica. Debido que en muchas poblaciones es bien conocido que el pedigrí registrado contiene errores (Oliehoek y Bijma 2009). En alpacas se ha identificado errores de asignación en el pedigrí con 4.44 por ciento de error de maternidad y 15.5 por ciento error de paternidad en alpacas (Yalta *et al.* 2014), siendo un punto crítico en la asignación de los supuestos padres dentro del sistema de identificación y recolección de datos genealógicos de los productores de alpacas (Morón *et al.* 2020).

## 2.6. VALIDACIÓN DE LA MICROMATRIZ

Para la validación del rendimiento de un panel de 64,232 PNS, genotiparon 158 individuos de los cinco grupos que se usaron para el descubrimiento de PNS, evaluaron la intensidad de señal y separación de grupos. De los 64k PNS se genotiparon con éxito 62,121 PNS (97.5 por ciento) y el restante de 1,611 PNS (2.5 por ciento) no se pudieron genotipar de manera confiable, asimismo los PNS que se genotiparon con éxito, 58,994 PNS fueron polimórficos, lo que indica que la tasa de conversión de PNS fue del 94 por ciento. Además, para 57,109 PNS el FAM fue > 5 por ciento y para 58,737 PNS el FAM fue > 1 por ciento. El FAM promedio para todos los PNS genotipados con éxito fue 0.274, observándose que más del >95 por ciento de los PNS predichos fueron validados (Ramos *et al.* 2009).

Asimismo, para la validación de micromatrices en diferentes especies se ha utilizado frecuentemente el genotipado y el análisis de la transmisión mendeliana, las secuencias que se transmiten son específicas por cada individuo. Se realizó el genotipado de PNSs en bovinos con una micromatriz y el análisis de transmisión mendeliana entre progenitores y crías lo cual obtuvo una fidelidad en la transmisión entre pares padre-cría de 99.94 por ciento de concordancia, y para los tríos padre-madre-cría de 99.95 por ciento (Matukumalli *et al.* 2009). Por otro lado, se observó que una micromatriz de 60k PNSs obtuvo > 94 por ciento de PNSs que se segregan en una variedad de razas en aves y que pueden ser genotipado de manera confiable (Groenen *et al.* 2011).

Para evaluar el rendimiento de las micromatrices de PNS en bovinos BeadChip BovineHD de Illumina (777,962 PNS) y Axiom Genome-Wide BOS 1 de Affymetrix (648,874 PNS), se genotiparon muestras de ADN de 10 Holstein y 6 Jersey. Con ello se obtuvo una tasa de genotipado de 99,7 por ciento y 98,5 por ciento para las micromatrices de Illumina y Affymetrix, respectivamente (Rincón *et al.* 2011).

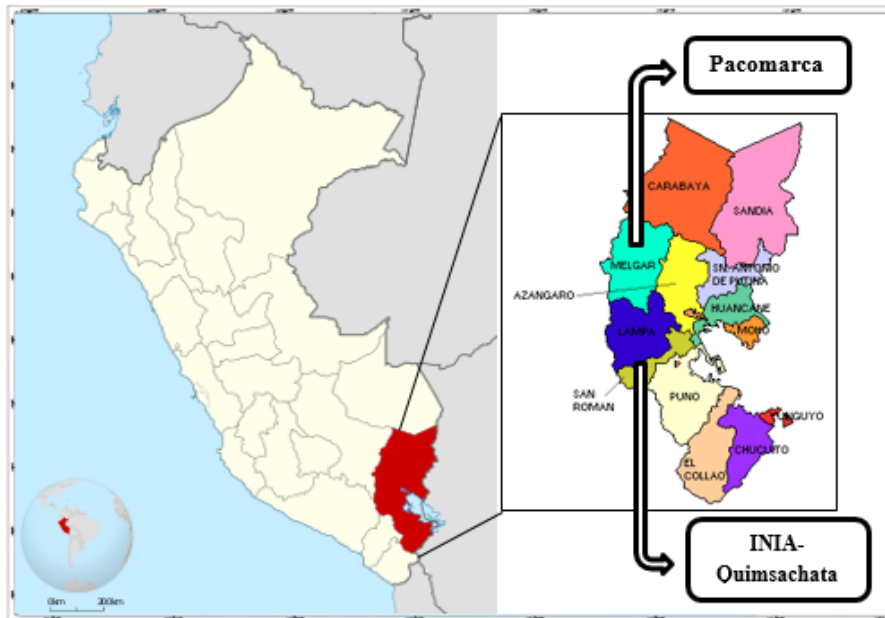
### **III. MATERIALES Y MÉTODOS**

#### **3.1 LOCALIZACIÓN**

Las actividades de investigación fueron ejecutadas en las regiones de Puno, Pasco y Lima. Las muestras biológicas fueron tomadas en dos granjas de Puno: Centro de Investigación y Producción de Camélidos de Quimsachata del Instituto Nacional de Innovación Agraria (INIA) y Fundo Pacamarca de INCA TOPS S.A., y en dos granjas de Pasco: Ganadería y Comercio del Centro S.R.L. (GACOCEN) y Cooperativa Comunal San Pedro de Racco.

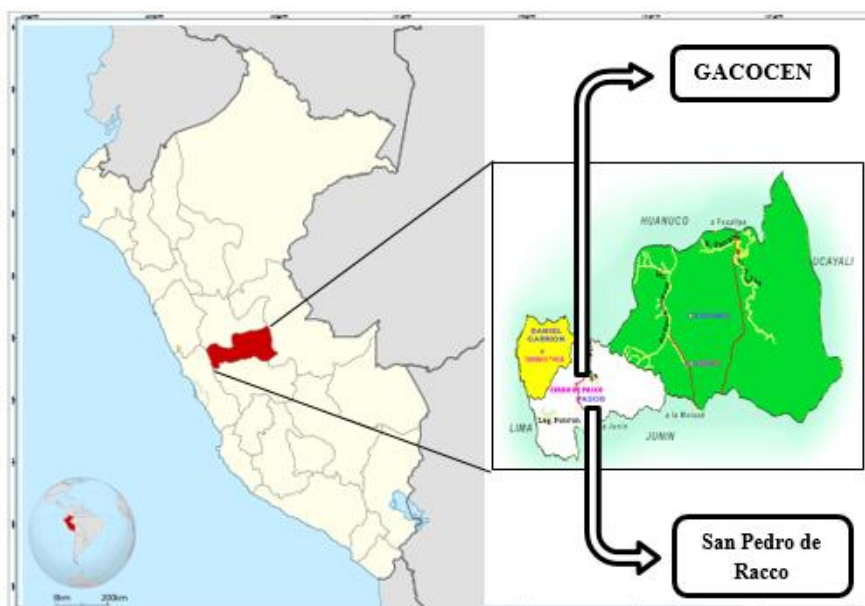
El Centro de Investigación y Producción de Camélidos de Quimsachata del INIA esta ubicado en el distrito de Santa Lucía, provincia de Lampa, región Puno, a una altitud aproximada que varía desde 3910 a 4680 msnm, una precipitación pluvial menor a 700 mm al año, una temperatura anual promedio de 22°C, una zona agroecológica de puna y el Fundo Pacamarca está ubicado en el distrito de Llalli, provincia de Melgar, región Puno, a una altitud aproximada que varía desde 3630 a 4320 msnm, una precipitación pluvial menor a 700 mm al año, una temperatura anual promedio de 17°C, una zona agroecológica de puna (Figura 5).





**Figura 5: Ubicación geográfica de Pacamarca e INIA-Quimsachata.**  
 Fuente: Adaptado de Google Maps (2021)

La granja GACOCEN se encuentra ubicada en la Comunidad Campesina de Lancari, distrito de Tinyahuarco, provincia de Pasco y región Pasco, a una altitud aproximada que varía desde 4320 a 4850 msnm y la Cooperativa Comunal San Pedro de Racco se encuentra ubicada en el distrito de Simón Bolívar, provincia de Pasco y región Pasco, a una altitud aproximada que varía desde 4310 a 4740 msnm, con una temperatura que oscila de 10°C a -11°C y es una zona agroecológica de puna (Figura 6).



**Figura 6: Ubicación geográfica de GACOCEN y San Pedro de Racco.**  
 Fuente: Adaptado de Google Maps (2021)

En la región Lima se realizó la extracción de ADN en el Instituto de Bioquímica y Biología Molecular de la Universidad Nacional Agraria La Molina y el Laboratorio de Biología Molecular y Genómica del INIA.

### **3.2 CONSTRUCCIÓN DE LA MICROMATRIZ DE PNSs DE ALPACA**

#### **3.2.1 Generación de bibliotecas de representación reducida de ADN**

##### **a. Animales**

Se colectaron 150 muestras de sangre de alpacas Huacaya de vellón blanco procedentes de cuatro granjas de las regiones de Pasco y Puno (Cuadro 1). Estas granjas fueron seleccionadas por contar con la información disponible de producción y genealogía.

Para la elección de las alpacas dentro de cada granja se tuvo en cuenta los siguientes criterios: información disponible de genealogía, no deben ser emparentados, información de finura de la fibra de segunda esquila, libre de enfermedades, raza Huacaya, vellón blanco y sexo hembra.

**Cuadro 1: Número de animales por región y granja para la colecta de muestras de sangre**

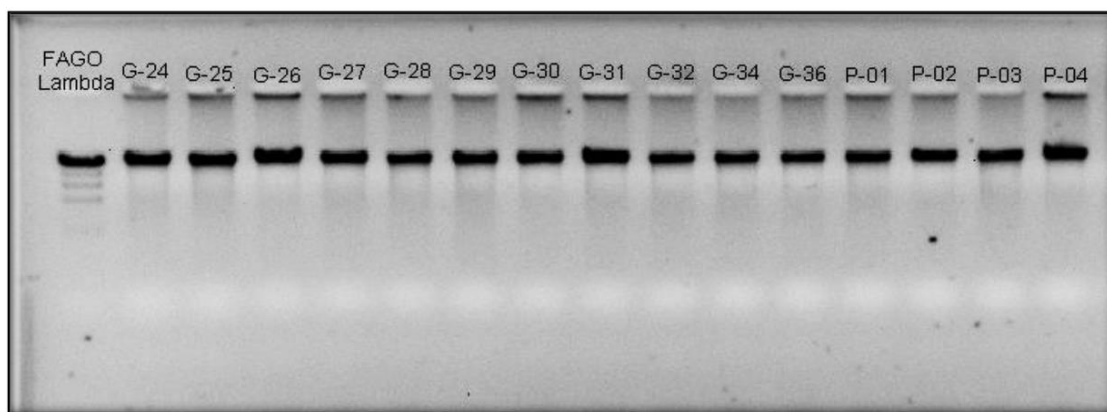
	<b>Región</b>	<b>Cantidad</b>
Pasco	San Pedro de Racco	25
	GACOCEN	25
Puno	Pacomarca	50
	INIA-Quimsachata	50
Total alpacas		150

##### **b. Colección de muestras de sangre y extracción de ADN**

Las muestras de sangre de 150 alpacas fueron recolectadas en tubos Vacutainer®, los tubos contenían anticoagulante EDTA K2 10.8 mg (Ácido Etilen Diamino Tetracético Dipotásico) para preservar las muestras de sangre, luego se realizó la extracción de ADN.

El método de extracción de ADN se realizó, siguiendo las especificaciones comerciales del kit *PureLink™. Genomic DNA Mini kit* (Invitrogen. California, Estados Unidos), adicionalmente se utilizó Buffer TE 20:5, SDS (10 por ciento), acetato de potasio 3M y etanol absoluto para el proceso de evaluación y obtención de una buena concentración y calidad de ADN. Luego de la extracción del ADN, se realizó la cuantificación de ácidos nucleicos, mediante espectrofotometría.

Para verificar la pureza e integridad del ADN extraído se verifico en dos procesos diferentes. La pureza fue verificada en 2 µl del ADN genómico, mediante la medición de la absorbancia a 260/280 nanómetros. La integridad del ADN fue evaluada por electroforesis en gel de agarosa (0.8 por ciento), a 120 voltios en buffer 1X TBE (Tris, Borato y EDTA) por 30 min. La observación de una sola banda sin fragmentación determina el grado de integridad del ADN. Los gels de agarosa fueron fotografiados bajo luz UV con un sistema de foto-documentación, como se muestra en la Figura 7. Debido que en algunas muestras biológicas de ADN extraído, se observaron la presencia de contaminantes, se procedió a una segunda purificación de ADN.



**Figura 7: Bandas de ADN de las muestras de GACOGEN y PACOMARCA en gels de agarosa al 0.8 por ciento teñidos con gel red.**

**c. Construcción de bibliotecas de representación reducida y secuenciación**

Las muestras de ADN extraído fueron liofilizadas en un concentrador plus (Eppendorf®), en tubos de microcentrífuga de 0.6 ml a una temperatura de 45°C por 90 minutos. Las muestras previamente liofilizadas, fueron enviadas a la empresa AgResearch Limited (Nueva Zelanda) para la preparación de 300 bibliotecas de representación reducida y el secuenciamiento de ADN.

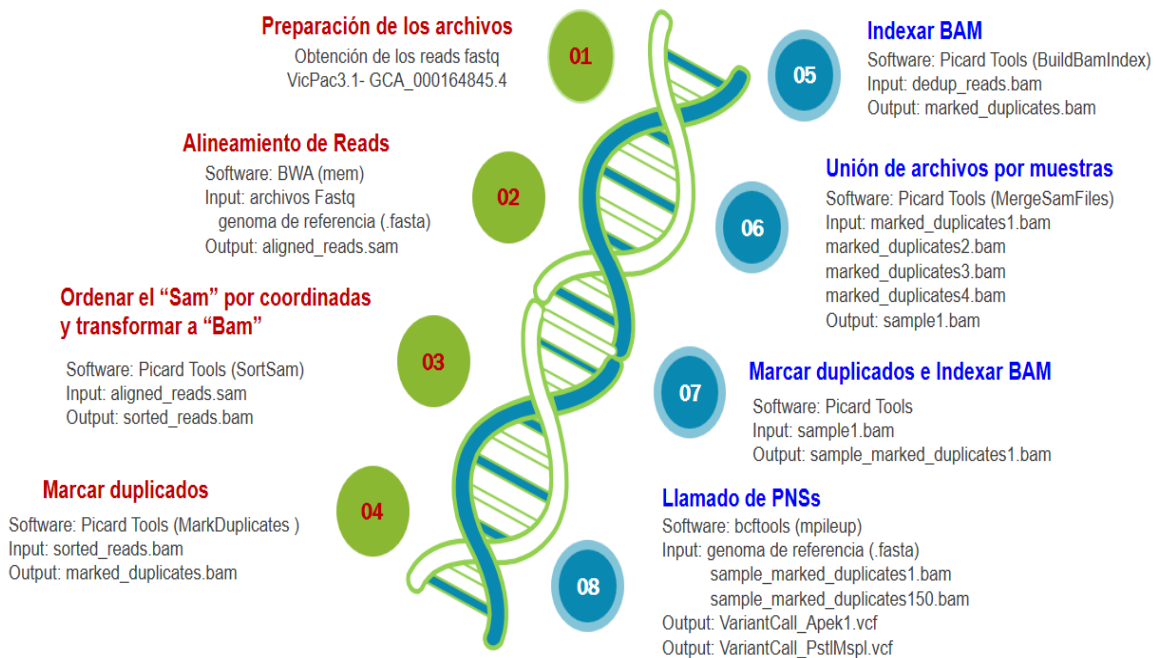
Se prepararon 150 bibliotecas de representación reducida de ADN resultantes de la digestión con la enzima ApeKI y 150 bibliotecas de representación reducida de ADN resultantes de la digestión con la combinación enzimática de las enzimas PstI y MspI. El rango del tamaño de fragmentos de ADN (lecturas) de las BRR fueron entre 100 a 200 pb. La secuenciación de los fragmentos de ADN, fue mediante el uso del secuenciador HiSeq 2500 a una profundidad de lectura media de ~6X por biblioteca.

### **3.2.2 Análisis bioinformático para la selección y cobertura de PNSs**

#### **a. Identificación de PNSs**

Todas las lecturas (reads) generadas con la enzima de ApeKI y en combinación del PstI-MspI se sometieron a control de calidad mediante el software FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Se utilizaron lecturas que superaban la puntuación de calidad de Illumina sobre la secuencia, mayor al umbral de 20, para el análisis e identificación de PNS (Ramos *et al.* 2009).

Para la identificación de PNSs (Figura 8) se utilizaron las lecturas que pasaron el control de calidad de las 300 BRR de ADN y se alinearon con el genoma de referencia *VicPac 3.1* (Richardson *et al.* 2019) disponible en NCBI (GCA\_000164845.4). Se utilizaron herramientas bioinformáticas como Alineador Burrows-Wheeler (BWA), Picard tools, SamTools y BCFTools. BWA y SamTools para el alineamiento y la indexación con el genoma de referencia (Li y Durbin 2010, Li *et al.* 2009); Picard tools para la creación de secuencia "diccionario", ordenamiento, marcación de duplicados, indexación y unión de archivos (Picard Toolkit 2019); BWA para el alineamiento de lecturas y BCFTools para la identificación y llamado de PNSs (BCFTools 2019).

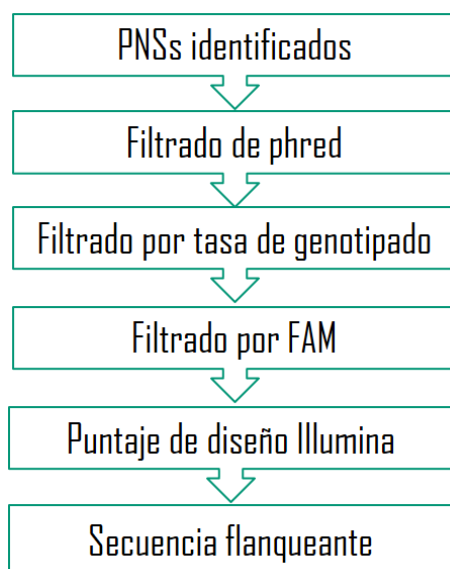


**Figura 8: Análisis de bioinformática para la identificación de PNSs**

## b. Selección de PNSs

La selección de PNSs de alpaca (Figura 9) se realizó utilizando los parámetros de puntuación de calidad con escala phred ( $>10$ ), tasa de genotipado ( $\geq 0.45$ ), frecuencia de alelo menor (entre 0.05 y 0.50), puntaje de diseño Illumina ( $\geq 0.6$ ) y ausencia de otros PNSs dentro de la secuencia flanqueante del PNS identificado (secuencia flanqueante de 35 a 60 pb orientado en la dirección 5' y 35 a 60 pb en la dirección 3').

La puntuación de calidad con escala phred representa la probabilidad de que la base identificada mediante el alineamiento no se dé al azar, con una precisión del 99 por ciento de que la base sea correcta. La tasa de genotipado estima el porcentaje de individuos donde se identifica el PNS. La frecuencia del alelo menor (FAM) se consideró entre 0.05 y 0.50, valores cercanos a 0 indican que el PNS es muy raro y existe baja probabilidad de que sea identificada en la población. El puntaje de diseño Illumina indica la probabilidad de éxito de la detección del PNS en la micromatriz. La ausencia de otros PNSs dentro de la secuencia del PNS aumenta la precisión de hibridación de la sonda de la micromatriz con la secuencia de ADN de la muestra a ser genotipada. Los procedimientos para la selección de los PNSs se ejecutaron mediante scripts programados en R (R Core Team 2020) (Anexo 1).



**Figura 9: Flujo del análisis de bioinformática para la selección de PNSs.**

**c. Cobertura de PNSs seleccionados en el genoma de referencia *VicPac 3.1***

Para la evaluación de PNSs seleccionados a lo largo de todo el genoma de referencia *VicPac 3.1*, se utilizó la puntuación de localización de cada PNS en 40 kpb, los PNS seleccionados son aquellos localizados en la posición media dentro y a cada 40 kpb. Esta puntuación fue definida para cada intervalo equidistante de 40 kpb a lo largo de cada scaffold. El scaffold forma parte de la secuencia del genoma como una serie de fragmentos no continuos.

La puntuación de localización de cada PNS se calcula de acuerdo a la posición del PNS dentro del fragmento. Un PNS con puntuación alta se localiza en la posición media o más próxima a la posición media del fragmento de 40 kpb, y el PNS con puntuación baja se localiza alejado de la posición media y se aproxima a los extremos del fragmento. Para calcular la puntuación de localización se utilizó la fórmula, adaptada de Matukumalli *et al.* (2009):

$$\text{Puntuación de Localización} = (E - S) - (2a - (E + S))$$

S = es la posición inicial del fragmento

E = es la posición final del fragmento

a = posición del PNSs

**Cuadro 2: Ejemplo de la puntuación de localización de PNS.**

<b>Scaffold</b>	<b>Fragmento</b>	<b>PNS por fragmento</b>	<b>Puntuación</b>
ABRR03000001.1	1	2	39820
ABRR03000001.1	1		30752
ABRR03000001.1	2	4	11248
ABRR03000001.1	2		35814
ABRR03000001.1	2		22238
ABRR03000001.1	2		19074
ABRR03000001.1	3	2	22514
ABRR03000001.1	3		22264
ABRR03000001.1	4	2	2612
ABRR03000001.1	4		2408

En el Cuadro 2, la columna fragmento nos indica el número correlativo de los fragmentos de 40 kbp que conforman un scaffold (fragmento 1, fragmento 2, fragmento 3.... fragmento n) y la columna PNS por fragmento es la cantidad de PNSs por cada fragmento de 40 kbp. Por ejemplo, el scaffold ABRR03000001.1 contiene 4 fragmentos y dentro del fragmento 2 se encuentran 4 PNSs de los cuales el segundo PNS tiene la puntuación más alta a comparación de los otros 3 PNSs en este fragmento, por lo tanto, se localiza próximo a la posición media del fragmento. De acuerdo a la puntuación de cada PNSs se seleccionó de los fragmentos 1, 2, 3 y 4 a los PNSs con puntuación de 39,820, 35,814, 22,514 y 2,612 respectivamente.

Con el objetivo de aumentar la densidad de PNS para la micromatriz, se seleccionó un segundo PNS dentro de cada fragmento de 40 kbp. Para ello, cada fragmento de 40 kbp se dividió en cinco subfragmentos de 8 kbp, codificados como subfragmento 1 (de 0 a 8 kbp), subfragmento 2 (de 8 a 16 kbp), hasta subfragmento 5 (de 32 a 40 kbp). Después de esto, se identificó el subfragmento de 8 kbp que contenía el primer PNS seleccionado. Según el subfragmento que contenía el primer PNS seleccionado, se seleccionó un segundo PNS en otro subfragmento de la siguiente manera: si el primer PNS estaba ubicado en el subfragmento 1, el segundo PNS se seleccionó para el subfragmento 3. Si el primer PNS estaba ubicado en el subfragmento 2, el segundo PNS se identificó para el subfragmento 4. De esta manera, los segundos PNS siempre se identificaron después de cada dos subfragmentos dentro de cada fragmento de 40 kbp. Este procedimiento de selección se realizó con un R script (R Core Team 2020) (Anexo 1).

Por lo tanto, para la construcción de la micromatriz de PNS de alpaca, se generaron dos conjuntos de PNS. El primer conjunto identificó PNS distantes a intervalos de 40 kpb en todo el genoma, y el segundo conjunto aumentó la densidad y redujo la distancia promedio entre PNS.

Para calcular la longitud cromosómica en pares de bases cubiertos con PNS a intervalos fijos, calculamos la longitud en pares de bases de todos los fragmentos que contienen un PNS por scaffold y por cromosoma. Luego, se realizó un proceso similar para todos los scaffolds no asignados a los cromosomas. El porcentaje de cobertura se calculó en función de la longitud del cromosoma y del genoma según lo informado para el genoma de referencia *VicPac 3.1*

Adicionalmente, también se incluyó un tercer conjunto de PNS ubicados en genes candidatos para la calidad y el color de la fibra (*KRT*, *KRTAP*, *ASIP*, *MC1R*, *TYRP1*) anotados en el genoma de referencia *VicPac 2.0.2* y *VicPac 3.1*. Para identificar este tercer conjunto de PNS se utilizó la secuencia de diez genomas de alpaca Huacaya (códigos de acceso NCBI SRR13340600 a SRR13340605, SRX2065156, SRX2065157, SRX2065161 y SRX682159) secuenciados a una profundidad aproximada de 30X y la secuencia de las 300 BRR, las cuales se alinearon con el genoma de referencia *VicPac 3.1*. Solo se seleccionaron los PNS con una puntuación de Illumina  $\geq 0,60$  y sin otros PNS en sus secuencias flanqueantes.

La lista de los tres conjuntos de PNS se envió a NEOGEN (Lincoln, NE, EE.UU.) para una evaluación adicional a fin de identificar el conjunto final de PNS. Este conjunto fue sometido a una nueva evaluación de algoritmos basados en la micromatriz de Affymetrix (CA, EE.UU.).

### **3.2.3 Construcción de la micromatriz de PNSs para alpaca**

La selección final de los PNSs que fueron incluidos en la micromatriz fue realizada por el personal técnico de Affymetrix, donde llegaron a considerar todo el conjunto de sondas para el caso de identificar PNS que no son recomendables, para ello se realizó un análisis de comprobaciones:



- Identificar polimorfismo en la secuencia flanqueante del PNS.
- Utilizar la secuencia genómica para identificar sondas dentro de las regiones altamente repetitivas y degradar tales sondas.
- Identificar PNS que comparten las mismas sondas, ya que no podemos determinar el genotipo de forma inequívoca da para tales PNSs.

Luego del análisis de comprobaciones, se consideró la base de recomendación para la selección de PNSs fueron: i) si un PNS tiene la misma recomendación para una cadena, recomendamos elegir el que tiene el valor más alto de pconvert, ii) un PNS se recomienda si el pconvert es  $> 0.6$ , no hay wobble\_count y el polycount = 0, iii) un PNS es no recomendado si presenta al menos un duplicate\_count  $> 0$ , polycount  $> 0$ , pconvert  $< 0.4$ , la wobble\_distance  $< 21$  o el wobble\_count  $\geq 3$ , y todos los otros PNS se consideran neutrales. Donde:

- Pconvert es la probabilidad esperada de conversión basada en la cadena directa de la secuencia.
- Wobble\_count es el número de polimorfismos intermedio dentro de las 24 bases.
- Polycount es el número de instancias donde 24 bases en cada lado del PNS coinciden con otras secuencias en la lista, pero los alelos difieren.
- Duplicate\_count es el número de instancias donde 24 bases en cada lado del PNS (incluidos los alelos) coinciden con otras secuencias en la lista. Tenga en cuenta que duplicatecount = 1 significa que hay otra secuencia coincidente.
- Wobble\_distance es la distancia al polimorfismo intermedio más cercano.

### **3.3 VALIDACIÓN DE LA MICROMATRIZ DE PNSs DE ALPACA**

#### **3.3.1 Validación de la micromatriz por genotipificación de tríos de alpacas**

Se colectaron muestras de sangre de 30 tríos de alpacas procedentes de dos regiones geográficas alto andinas de Pasco y Puno. Cada trio estuvo conformado por el padre, la madre y la cría que fueron identificados en base a su genealogía. La distribución y cantidades de la colecta de muestras de sangre se muestra en el Cuadro 3. Para la selección y colecta de las muestras del animal se tomó en cuenta la información disponible de producción y genealogía. La extracción y evaluación de calidad del ADN fue realizada

según lo descrito anteriormente en la sección 3.2.1 ítems b (Colección de muestras de sangre y extracción de ADN).

**Cuadro 3: Número de tríos (padre, madre, progenie) de alpacas y número de muestras de ADN por regiones**

Región	Cantidad de tríos	Cantidad de muestras
Pasco	10	30
Puno	20	60
Total	30	90

Para la validación de la micromatriz de PNSs de alpaca, se realizó la genotipificación del grupo de tríos (padre, madre y cría) con la micromatriz de PNS de alpaca. A través del genotipado se analizó en el grupo de tríos la transmisión de herencia mendeliana de los PNSs.

El genotipado de los 30 tríos de alpacas se analizó empleando el software Axiom Analysis Suite v.4.0.3 (Affymetrix) y posteriormente se ejecutó el análisis de relaciones genómicas (G) entre progenitores y crías. Para el cálculo de la matriz de relación genómica, se eliminaron los genotipos de PNS con bajas tasas de genotipado (Menor a 95 por ciento) y los PNSs monomórficos.

Los genotipos por PNS fueron codificados a partir de la frecuencia de alelos dando lugar al alelo de mayor frecuencia para el alelo de referencia y el alelo de menor frecuencia para el alelo alternativo. Los genotipos para cada PNS se codificaron como 0, 1 y 2 para AA, AB y BB, respectivamente, donde A es el alelo de referencia y B es el alelo alternativo. Se utilizó el programa GCTA (Yang *et al.* 2011) para el cálculo de la matriz de relaciones genómicas, la relación entre el individuo  $j$  y  $k$  se estima con la siguiente ecuación:

$$G_{jk} = \frac{1}{N} \sum_{i=1}^N \frac{(x_{ij} - 2p_i)(x_{ik} - 2p_i)}{2p_i(1 - p_i)}$$

$x_i$  = número de copias del alelo de referencia para el  $i$ -ésimo PNS;

$x_{ij}$  = número de copias del alelo de referencia para el  $i$ -ésimo PNS del  $j$ -ésimo individuo;

$x_{ik}$  = número de copias del alelo de referencia para el  $i$ -ésimo PNS del  $k$ -ésimo individuo;  
 $p_i$  = Frecuencia del alelo de referencia;  
 $N$  = Número de PNS

Los resultados de la genotipificación se compararon con la información de genealogía, información procedente de cada granja por región.

La tasa de conversión de PNSs, una medida del éxito de genotipificación, es definida como la proporción de todos los PNSs genotipados que muestran agrupaciones claras de genotipificación (Hoffman *et al.* 2013).

### **3.3.2 Comparación de genotipado por secuenciamiento y genotipado por micromatriz**

Para evaluar la confiabilidad del genotipado por la micromatriz (GM) y el genotipado por secuenciamiento (GBS), comparamos el genotipado de ambos métodos en las 150 muestras de alpaca que fueron utilizadas para identificar los PNSs. Dado que se espera que el conjunto de PNS incluidos en la micromatriz esté presente en cada uno de los genotipos de GBS en las 150 alpacas, se utilizó el conjunto de PNS de la micromatriz como referencia para la comparación.

El control de calidad del genotipado obtenidos por GM se realizó con el software Axiom Analysis Suite v.4.0.3 (Affymetrix, CA, EE. UU.). Se realizó un primer control de calidad de la muestra que incluyó los siguientes parámetros: a) el control de calidad del plato  $\geq 0.82$ , b) la tasa de llamadas de control de calidad  $\geq 97$  por ciento, c) porcentaje de muestras que pasaron el control de calidad por plato y tasa de llamada del control de calidad por placa  $\geq 95$  por ciento, y d) la tasa de llamadas promedio para muestras que pasaron  $\geq 98.5$ . Luego el control de calidad del PNS, incluyó la tasa de llamadas  $\geq$  del 97 por ciento y los mejores PNS recomendados de acuerdo a los criterios de Affymetrix para la determinación del genotipo (homocigoto para alelo A, alelo B o heterocigoto). Los criterios que Affymetrix consideró fueron: Resolución alta y monomórfica (MonoHighResolution), Resolución alta y polimórfica (PolyHighResolution) y Ausencia del Homocigoto para el alelo menor (NoMinorHom). De 150 muestras de alpacas genotipadas por GM solo 145 muestras pasaron el control de calidad.

El control y filtrado de calidad de los genotipos obtenidos por GBS se realizó manualmente (Microsoft Excel) en que se consideró un mínimo de tres lecturas de secuenciación, por alelo, por PNS y por animal, requeridos para la llamada del genotipo por GBS. Se tomaron en cuenta el conjunto de PNSs de referencia y número de muestra genotipadas por la GM.

Luego para efectuar la comparación de genotipado, cada genotipo de PNS del conjunto de datos de GBS y GM se codificó con 0, 1 y 2 para AA, AB y BB, respectivamente, donde A representa el alelo de referencia y B el alelo alternativo. La base de datos final contenía 75,577 PNS (57,697 PNS identificados de ApeK1 y 17,880 PNS de PstI-MspI de las BRR) que son comunes en ambos conjuntos de datos (GBS y GM).

Los genotipos de GBS filtrados se compararon, por parejas, con los genotipos de GM, por animal para determinar el nivel de concordancia entre ambos métodos de genotipado. Las comparaciones se realizaron utilizando BCFtools. Cada comparación se expresó como el porcentaje de concordancia entre ambos métodos de genotipado.

### **3.3.3 Estructura genética de la población**

Para evaluar la estructura genética de la población de los animales muestreados, se utilizó los resultados de genotipado obtenidos por GM y GBS por separado. Para el GM, se utilizaron 76,508 PNSs y para GBS se utilizaron 3'015,363 y 1'065,353 PNSs de ApeK1 y MspI-PstI, respectivamente. Para ambos métodos de genotipado, se calculó la matriz de relación genómica como se describió anteriormente teniendo en cuenta los PNS con una tasa de llamada  $\geq 0.90$  y FAM  $\geq 0.05$ . El análisis de componentes principales se utilizó para inferir la diferenciación de la población (Patterson *et al.* 2006; Deniskova *et al.* 2019). La relación genómica fue calculada usando la formula descrita anteriormente utilizando el programa GCTA (Yang *et al.* 2011). El mapa de calor de las relaciones genómicas se generó con el software RStudio Team (RStudio Team, 2020) (Anexo 2).

El coeficiente de consanguinidad y la heterocigosidad individual para el genotipado con la micromatriz se calcularon utilizando datos de 145 alpacas. Se excluyeron los PNS ubicados en el cromosoma X, así como los PNS con tasas de llamada  $< 0.90$  y FAM  $< 0.05$ . Después de la evaluación de PNS basada en el desequilibrio de ligamiento, evaluado con PLINK v1.90p, se conservaron 29,059 PNSs para el análisis.

Para el cálculo del coeficiente de heterocigosidad y coeficiente de consanguinidad con los datos de GBS, se utilizaron 2,452,915 y 884,297 PNS de ApeK1 y Mspl-PstI, respectivamente. Un mínimo de tres lecturas de secuenciación, por alelo, por PNS y por animal, fueron requeridos por llamada de genotipo por GBS. El coeficiente de heterocigosidad individual se calculó con PLINK v1.90p y representa la proporción de loci heterocigotos del total de loci genotipados por muestra animal (Chang *et al.* 2015). El coeficiente de consanguinidad (F) se calculó con GCTA (Yang *et al.* 2011), donde la estimación se basa en la correlación entre las uniones de gametos:

$$\hat{F}_i^{III} = [x_i^2 - (1 + 2p_i)x_i + 2p_i^2]/h_i \text{ y } var(\hat{F}_i^{III}|F)$$

$$\hat{F}_i^{III} = 1 + 2(1 - 2h_i)F/h_i - F^2$$

$\hat{F}_i^{III}$  es un estimador insesgado de F en el sentido de que  $E(\hat{F}_i^{III}|F) = F$ . Para múltiples PNS, promedia las estimaciones sobre todos los PNS o:

$$\hat{F} = 1/N \sum_{i=1}^N \hat{F}_i$$

$x_i$  = número de copias del alelo de referencia para el i-ésimo PNS;  
 $h_i = 2 p_i(1 - p_i)$ ;  
 $p_i$  = frecuencia del alelo de referencia;  
N = número de PNS.

Para estimar la diferenciación genética entre poblaciones de los animales muestreados, se utilizó el programa ADMIXTURE (Alexander *et al.* 2009, Skotte *et al.* 2013, Alexander *et al.* 2020). Este programa estima la ascendencia de individuos no emparentados y proporciona una medida de combinación entre poblaciones. Además, el índice de fijación ( $F_{st}$ ) es calculado asumiendo la existencia de dos poblaciones, de tal manera que los individuos son distribuidos en cada población de acuerdo a los datos de genotipado, para cada autovector. Para el análisis de la combinación, se utilizó solo los genotipos generados por la micromatriz de PNS para los animales muestreados.

## IV. RESULTADOS Y DISCUSIÓN

### 4.1. BIBLIOTECAS DE REPRESENTACIÓN REDUCIDA DE ADN

Las lecturas (reads) generadas utilizando las enzimas ApeK1 y la combinación de PstI-Mspl se sometieron a un control de calidad mediante el software FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>, consultado el 18 de octubre de 2020). Se utilizaron  $3.45 \times 10^8$  lecturas que superaban el umbral de 20 para la calidad de secuencia para identificar los PNSs (Ramos *et al.* 2009).

El análisis bioinformático para identificar PNSs mediante el alineamiento de las lecturas de calidad de los 300 BRR con el genoma de referencia *VicPac 3.1*, permitió identificar 3'427,450 y 1'297,789 PNSs de las BRR de ApeK1 y PstI-Mspl, respectivamente (Cuadro 4). Después de eliminar una copia de PNS duplicado o presentes en fragmentos generados por ambas enzimas se obtuvo 4'283,956 PNSs únicos.

**Cuadro 4: Cantidad de PNSs identificados por BRR alineados con el genoma de referencia *VicPac 3.1***

Bibliotecas de Representación Reducida	Nº PNSs
Apek1	3'427,450
PstI-Mspl	1'297,789

Se requiere la identificación de un gran número de PNSs para el desarrollo de una micromatriz, a partir de ello se pueden aplicar una serie de filtros para seleccionar PNSs que se puedan incluir en la micromatriz. Para identificar un gran número de PNSs, otros grupos de investigación (Groenen *et al.* 2011; Ramos *et al.* 2009; Matukumalli *et al.* 2009) utilizaron muestras de animales de diferentes razas y/o de diferentes líneas de selección (Groenen *et al.* 2011) que ya mostraban una diversidad fenotípica significativa y posiblemente diversidad genética.

En el caso de la alpaca, solo existen dos razas, la Huacaya y Suri, que representan el 80.4 por ciento y el 12.2 por ciento de la población de alpacas, respectivamente. Los cruces entre razas y especies representan el 7.3 por ciento (Fernández-Baca 2005). Con el objetivo de generar la mayor cantidad posible de PNSs para producir una micromatriz de mediana densidad se eligieron granjas elites de alpaca Huacaya de vellón blanca para la identificación de animales con registros de peligre y muestras de diámetro de fibra evaluadas a la segunda esquila. No se incluyeron en la muestra alpacas Suri ni alpacas con vellón de color.

Además, debido a la limitada información sobre diversidad genética de la población de alpacas en Perú, se determinó seleccionar muestras de animales criados extensivamente y considerar la ubicación de la granja, como mejores alternativas para garantizar cierta diversidad genética entre los animales. Por lo tanto, se seleccionó alpacas Huacaya de dos regiones diferentes y dos granjas por región para la identificación de PNS.

#### **4.2. SELECCIÓN DE PNS PARA LA MICROMATRIZ**

El uso de los parámetros de filtros usados para la selección de PNSs redujo el conjunto original de PNSs únicos a 513,467. De estos, se seleccionaron 45,156 PNS considerando una puntuación de calidad con escala de phred  $>10$ , tasa de genotipado  $\geq 0.45$ , frecuencia de alelo menor entre 0.05 - 0.50, puntaje de diseño Illumina  $\geq 0.6$  y la ausencia de otros PNSs dentro de la secuencia flanqueante al PNS de 35 pb y 40 pb, distribuidos a intervalos de aproximadamente 40 kpb (ronda 1) a lo largo del genoma de la alpaca.

Para aumentar el número de PNSs dentro de cada fragmento de 40 kbp, se seleccionaron otros PNS en rondas posteriores (rondas 2-6), variando en dos parámetros de calidad, la tasa de genotipado  $\geq 0.15$  -  $\geq 0.45$  y la frecuencia de alelo menor entre 0.01 - 0.50 generando 51,772 PNS seleccionados como nuestro primer conjunto de PNS para la micromatriz (Cuadro 5).

Asimismo, para incrementar la densidad de PNSs, se identificó el segundo conjunto de 28,429 PNS (Cuadro 5) seleccionados con los mismos parámetros de calidad que el primer conjunto de PNS pero que estaban dentro de cada fragmento de 40 kpb. Estos dos

conjuntos representan 80,201 nuevos PNS de alpaca (EMBL-EBI, código de acceso: ERZ1694265).

**Cuadro 5: Número de PNS seleccionados por puntuación de ubicación del primer y segundo conjunto**

Ronda	Phred Score	TG	FAM	Puntaje Illumina	Longitud de Secuencia Flanqueante	Numero de PNSs del 1° Conjunto	Numero de PNSs del 2° Conjunto
1	>10	≥0.45	0.05–0.50	≥0.60	40	45,156	17,148
2	>10	≥0.45	0.05–0.50	≥0.60	35	1319	1876
3	>10	≥0.15	0.04–0.50	≥0.60	40	4027	6734
4	>10	≥0.15	0.04–0.50	≥0.60	35	320	628
5	>10	≥0.15	0.01–0.039	≥0.60	40	829	1821
6	>10	≥0.15	0.01–0.039	≥0.60	35	121	222
Total						51,772	28,429

Además, se propuso la inclusión en la micromatriz de un conjunto de PNSs ubicados en los exones e intrones de genes candidatos para la calidad y el color de la fibra (EMBL-EBI, código de acceso: ERZ1694264). Estos PNSs fueron identificados en el presente estudio en los genes: *KRTs* (205), *KIT* (18), *KRTAPs* (56), *ASIP* (2), *MC1R* (1), *TYRPI* (1). De manera similar, se incluyen los PNS que estaban reportados en la literatura para *KRTs* (16) (Fernández *et al.* 2019), *MC1R* (1) (Chandramohan *et al.* 2015) y *TYRPI* (2) (Feeley *et al.* 2016) (Anexo 3).

Todos los scaffolds asignados a cromosomas, que representan aproximadamente el 76 por ciento del genoma (Richardson *et al.* 2019), contienen al menos un PNS por cada  $25.92 \pm 14.57$  kbp con una densidad de 40 PNSs/Mb por cromosoma (Figura 10). Los scaffolds no localizados, que representan aproximadamente el 24 por ciento del genoma, en promedio, contienen un PNS por cada  $23.69 \pm 24.58$  kpb. Sobre la longitud total de los scaffolds no localizados solo el 78.2 por ciento tiene PNSs asignados. De esta manera, calculamos que el 92 por ciento de la longitud total del genoma contiene un PNS por fragmento de ADN promedio de  $25.41 \pm 17.41$  kbp (Cuadro 6). La distribución de PNSs por FAM se muestra en la Figura 11. Los PNSs con  $FAM \geq 0.10$  representan el 71.9 por ciento del total.

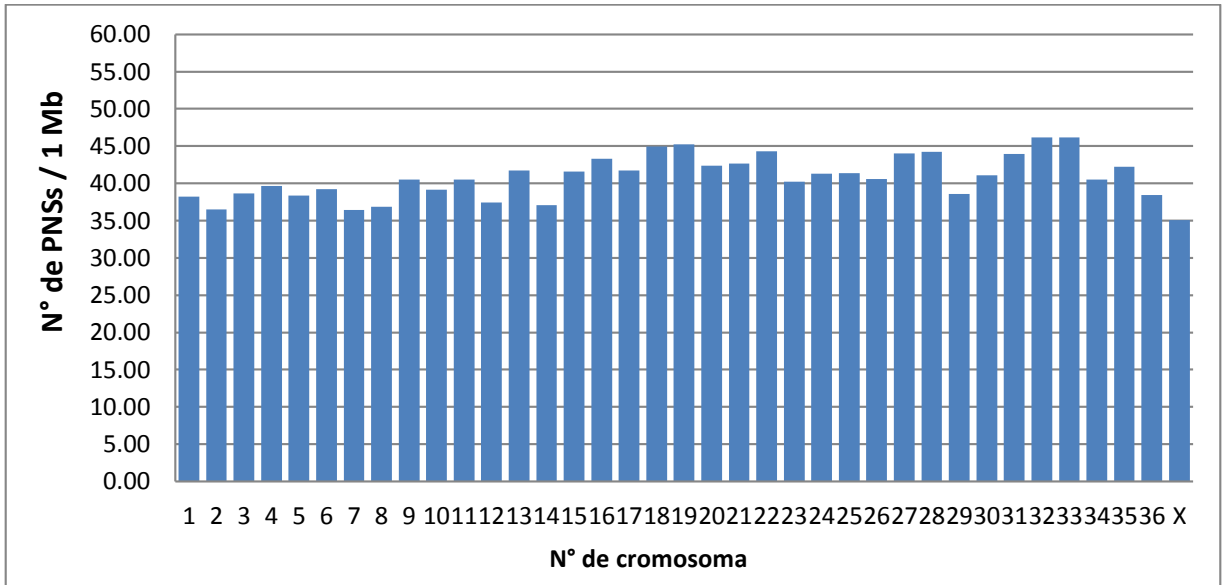


**Cuadro 6: Cobertura del primer y segundo conjunto de 80k de PNSs seleccionado por cromosoma en el *VicPac 3.1***

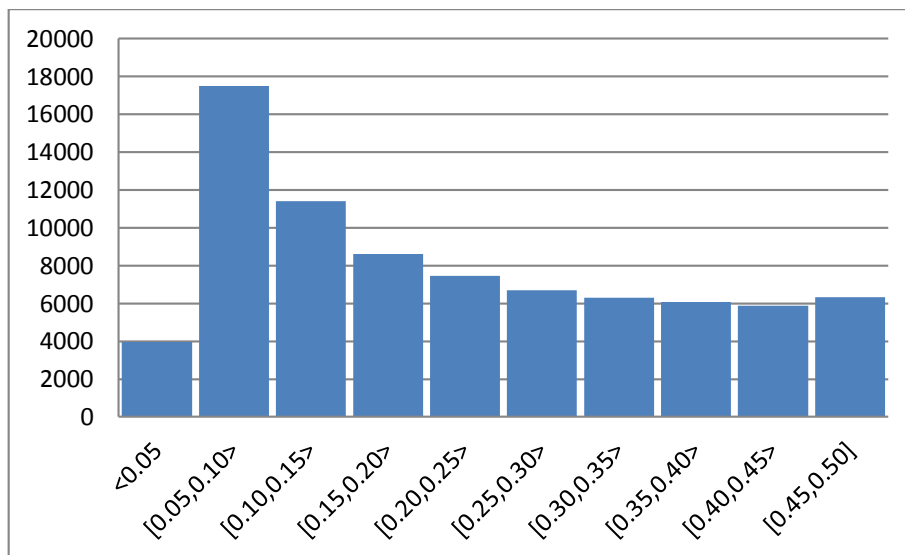
<b>Cromosoma</b>	<b>Número de PNSs</b>	<b>Número Fragmentos de 40 kpb</b>	<b>Promedio de Distancia Entre Marcadores</b>	<b>StdDev de Distancia Entre Marcadores</b>	<b>Cobertura Cromosomal con PNSs (pb)</b>	<b>Longitud del Cromosoma <i>VicPac 3.1</i> (pb)</b>	<b>Cobertura Cromosomal con PNSs ( por ciento)</b>
1	3693	2416	27343.15	16053.23	96557607	101041233	95.56
2	4132	2829	29371.84	18814.36	113130620	121370620	93.21
3	3077	1992	27087.39	15922.71	79643794	83363794	95.54
4	2499	1578	26258.07	14878.86	63076945	65636945	96.10
5	3553	2316	27092.82	15244.33	92634254	96274254	96.22
6	2820	1800	26515.69	14775.38	71951714	74791714	96.20
7	1065	731	29256.18	17064.91	29240000	31168711	93.81
8	2518	1710	27900.53	15440.99	68390077	70270077	97.32
9	2103	1299	25547.32	13996.72	51921482	53761482	96.58
10	1496	955	26452.18	15074.95	38182034	39582034	96.46
11	2957	1828	25345.51	13610.08	73019631	74979631	97.39
12	1747	1169	28029.14	16813.58	46706614	48986614	95.35
13	2491	1493	24471.38	12355.45	59685320	61008235	97.83
14	2315	1561	28972.18	18366.23	62424741	67111318	93.02
15	1321	795	24527.24	12483.52	31772912	32418436	98.01
16	1664	963	23450.67	11767.82	38426118	39074364	98.34
17	1941	1162	24171.00	11320.66	46480000	46944759	99.01
18	1338	744	22344.37	9736.16	29750177	29910177	99.47
19	1077	596	22290.07	9267.70	23822313	24022313	99.17
20	1608	948	23932.14	11824.77	37920000	38741345	97.88
21	1240	727	23778.91	11378.08	29070537	29520914	98.47

<<Continuación>>

22	1122	635	22747.28	9890.59	25362599	25522599	99.37
23	1163	724	25303.64	12820.82	28920657	29440657	98.23
24	751	455	24420.85	11778.63	18186407	18346407	99.13
25	2388	1444	25184.43	14055.49	57737896	60195357	95.92
26	1113	686	25123.85	12859.22	27410361	27987978	97.94
27	994	565	22813.30	9904.99	22600000	22699463	99.56
28	710	403	22762.83	10427.74	16082605	16162605	99.51
29	979	634	26578.24	14157.49	25358588	26137040	97.02
30	533	325	24564.71	12134.35	12962202	13130742	98.72
31	594	339	22895.98	10136.03	13522737	13602737	99.41
32	1042	565	21792.13	9073.34	22567807	22732685	99.27
33	749	406	21702.86	8330.38	16221182	16261182	99.75
34	881	544	25076.72	12789.24	21737801	22097801	98.37
35	768	456	24019.86	11701.03	18199644	18484027	98.46
36	187	122	27950.16	18305.86	4860638	5377765	90.38
X	1128	782	30261.25	20496.52	31242392	34309508	91.06
Total Cromosoma	61757	38697	25924.45	14572.56	1546780406	1602467523	96.52
Scaffolds No Asignados	18444	13090	23692.03	24584.76	404377893	517133374	78.20
<b>Total</b>	80201	51787	25411.06	17418.28	1951158299	2119600897	92.05



**Figura 10: Distribución de los PNSs seleccionados del primer y segundo conjunto (80k) por cromosoma de alpaca, con una densidad promedio de 40 PNSs/Mb por cromosoma.**



**Figura 11: Distribución de PNSs seleccionados del primer y segundo conjunto (80k) por la frecuencia del alelo menor (FAM).**

### 4.3. CONSTRUCCIÓN DE LA MICROMATRIZ DE PNSs DE ALPACA

La información de 80,503 PNSs fueron enviados a NEOGEN (Lincoln, NE, EE. UU.). El número final de PNSs seleccionados para su inclusión en la micromatriz fue de 76,508 PNSs, 302 controles de PNS duplicados y 100 PNSs controles de Affymetrix (CA, EE. UU.) (Cuadro 7).

**Cuadro 7: Número de PNS seleccionados para la micromatriz**

<b>Descripción de PNSs</b>	<b>N° de PNSs enviados a NEOGEN</b>	<b>N° de PNSs incluidos en la micromatriz</b>
1° Conjunto	51,772	49,282
2° Conjunto	28,429	26,924
Genes Candidatos	302	302
Controles		100
Control Duplicados		302
Total de PNSs	80,503	76,910

La micromatriz construida incluye 76,508 PNSs, que cubren el 95.2 por ciento de la longitud de los scaffolds asignados a cromosomas y el 76.1 por ciento de los scaffolds no asignados que representan el 90.5 por ciento de la longitud del genoma de la alpaca cubierto por PNSs a una distancia promedio de  $26.58 \pm 18.57$  kpb entre sí (Cuadro 8).

Otros 302 PNSs ubicados en las secuencias de genes candidatos para la calidad y el color de la fibra también se incluyen en la micromatriz de PNSs (Anexo 3). El mayor número de PNSs (58,110) se ubican a intervalos de longitud de fragmentos entre 10 y 30 kbp dentro del genoma de referencia (Cuadro 9).

**Cuadro 8: Cobertura de los PNSs por cromosoma incluidos en la micromatriz**

<b>Cromosoma</b>	<b>Número de PNSs</b>	<b>Número fragmentos de 40 kpb</b>	<b>Promedio de distancia entre marcadores</b>	<b>Desviación estándar de distancia entre marcadores</b>	<b>Cobertura por cromosoma con PNSs (pb)</b>	<b>Longitud del cromosoma <i>VicPac 3.1</i> (pb)</b>	<b>Cobertura por cromosoma con PNSs ( por ciento)</b>
1	3585	2377	28166.87	17194.23	94997607	101041233	94.02
2	4027	2795	30137.68	19576.45	111770620	121370620	92.09
3	2960	1966	28150.53	16727.25	78616390	83363794	94.31
4	2409	1564	27206.95	15661.65	62526287	65636945	95.26
5	3443	2287	27958.41	16071.04	91474254	96274254	95.01
6	2709	1768	27602.15	15914.77	70671714	74791714	94.49
7	1031	720	30220.98	17900.23	28800000	31168711	92.40
8	2431	1687	28899.02	16354.66	67470077	70270077	96.02
9	2031	1286	26431.61	15000.54	51406823	53761482	95.62
10	1441	945	27461.81	16092.56	37782034	39582034	95.45
11	2818	1795	26585.10	14905.72	71719875	74979631	95.65
12	1811	1159	27038.60	18079.15	46306614	48986614	94.53
13	2350	1478	25918.86	13590.82	59093595	61008235	96.86
14	2220	1529	30186.43	19858.38	61160000	67111318	91.13
15	1251	782	25899.67	13897.13	31252912	32418436	96.40
16	1684	951	23169.37	14201.84	37984166	39074364	97.21
17	1846	1153	25414.90	12548.38	46120000	46944759	98.24
18	1264	736	23652.50	11281.00	29430177	29910177	98.40
19	1013	589	23698.33	10993.81	23542313	24022313	98.00
20	1520	934	25317.69	13365.25	37360000	38741345	96.43

<<Continuación>>

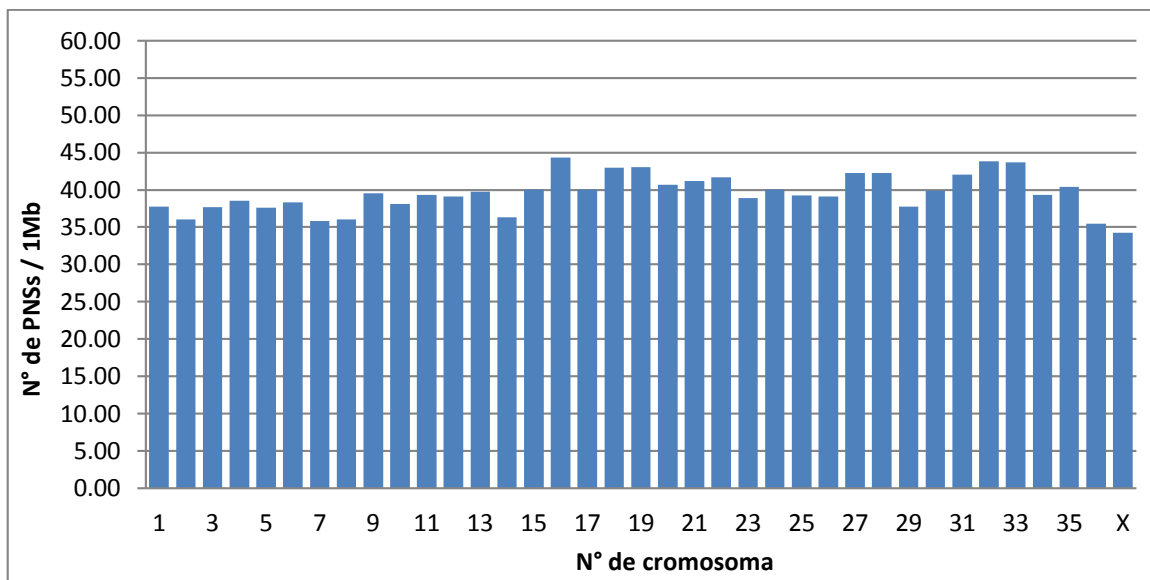
21	1192	724	24736.45	12198.65	28950537	29520914	98.07
22	1043	626	24470.23	11664.63	25002599	25522599	97.96
23	1115	717	26367.16	13773.23	28662452	29440657	97.36
24	718	449	25543.25	12896.59	17946407	18346407	97.82
25	2233	1422	26932.57	15652.00	56857896	60195357	94.46
26	1056	676	26479.96	14449.58	27010361	27987978	96.51
27	941	557	24098.22	11529.09	22280000	22699463	98.15
28	664	394	24339.77	12600.45	15722605	16162605	97.28
29	949	628	27368.18	14678.90	25120000	26137040	96.11
30	511	321	25597.38	13293.95	12802202	13130742	97.50
31	563	336	24156.69	11352.00	13402737	13602737	98.53
32	974	556	23313.55	11436.66	22207807	22732685	97.69
33	694	397	23354.08	10674.67	15880000	16261182	97.66
34	834	530	26472.94	15045.44	21200000	22097801	95.94
35	718	445	25692.55	14772.80	17759644	18484027	96.08
36	168	119	31111.19	19620.55	4740638	5377765	88.15
X	1080	767	31591.58	21235.14	30642392	34309508	89.31
Scaffolds asignados a cromosomas*	59297	38165	26992.36	15753.59	1525673735	1602467523	95.21
Scaffolds No Asignados*	17211	12491	25159.67	25991.20	393461101	517133374	76.09
Total	76508	50656	26580.09	18571.29	1919134836	2119600897	90.54

\* Asignación de scaffolds a cromosomas de alpaca según Richardson *et al.* (2019).

**Cuadro 9: Número de fragmentos que contienen un PNS por intervalos de longitud de fragmentos.**

<b>Intervalo de longitud de fragmentos (kpb)</b>	<b>Número de fragmentos que contienen un PNS identificado en el presente estudio.</b>	<b>Número de fragmentos con un PNS incluido en la micromatriz</b>
≥ 700–800	1	1
≥ 600–700	0	1
≥ 500–600	6	5
≥ 400–500	3	3
≥ 300–400	10	10
≥ 200–300	29	35
≥ 100–200	315	375
≥ 90–100	210	243
≥ 80–90	302	366
≥ 70–80	541	696
≥ 60–70	1075	1285
≥ 50–60	2770	3011
≥ 40–50	5999	6682
≥ 30–40	10,848	11,069
≥ 20–30	21,145	19,146
≥ 10–20	32,282	29,070
≥ 0–10	4683	4510
<b>Total</b>	<b>80,201</b>	<b>76,508</b>

Todos los scaffold asignados a cromosomas, que representan el 76 por ciento del genoma de la alpaca, tienen en promedio el 95.2 por ciento de su longitud cubierta, en promedio, con al menos un PNS por  $26.99 \pm 15.75$  kpb. La densidad media de PNSs por cromosomas es de  $39 \pm 2.51$  PNS/Mb (Figura 12). Los scaffolds no asignados representan aproximadamente el 24 por ciento del genoma y tienen, en promedio, el 76.1 por ciento de su longitud total cubierta con al menos un PNS por  $25.15 \pm 25.99$  kpb. Por lo tanto, 90.5 por ciento de los  $2.1 \times 10^9$  pares de bases del genoma de referencia *VicPac 3.1* contiene, en promedio, un PNS por cada  $26.58 \pm 18.57$  kpb representado en la micromatriz alpaca (Cuadro 10).



**Figura 12: Densidad de PNSs incluidos en la micromatriz por cromosoma de alpaca.**

**Cuadro 10: Cobertura del genoma de la alpaca con PNS incluidos en la micromatriz**

Scaffolds	Numero de PNSs	Numero de fragmentos de 40 kpb	Promedio del intervalo entre PNSs	Longitud cubierta por PNSs (pb)	Longitud <i>VicPac 3.1</i> (pb)	Longitud del genoma cubierto con PNSs en por ciento ( <i>VicPac 3.1</i> )
Asignados a Cromosoma	59,297	38,165	26,992	1,525,673,735	1,602,467,523	95.21
No Asignados	17,211	12,491	25,160	393,461,101	517,133,374	76.09
Total	76,508	50,656	26,580	1,919,134,836	2,119,600,897	90.54

El uso de la información de 10 genomas de alpaca y 300 BRR secuenciados a profundidades de 30X y 6X, respectivamente, acoplados a filtros de calidad apropiados para identificar PNS con FAM por encima de 0.05, permitió la identificación de un gran número de PNS para la micromatriz. El enfoque que se usó para la identificación de PNSs fue similar a otros enfoques (Groenen *et al.* 2011, Ramos *et al.* 2009, Matukumalli *et al.* 2009) pero con una mayor profundidad de secuenciación. De esta manera se identificaron  $4.28 \times 10^6$  PNS, de los cuales se seleccionaron 513,467 PNSs de *novo*, con los criterios mínimos de calidad requeridos que incrementan las probabilidades de ser PNSs verdaderos.



Un aspecto del desarrollo de la micromatriz fue asegurar que los PNSs estén ubicados uniformemente a intervalos de entre 30 y 40 kbp a lo largo del genoma. La distancia del intervalo de la micromatriz promedio entre PNSs es  $26.58 \pm 18.57$  kbp. Sin embargo, hay 12,713 PNSs ubicados en fragmentos del genoma mayores de 40 kbp, de los cuales 410 se ubican en fragmentos mayores de 100 kbp. La longitud del genoma de los 88 scaffolds localizados es 1.602 Gb (Richardson *et al.* 2019), de los cuales 0.076 Gb no tienen un PNSs asignado. Asimismo, la longitud del genoma de los 77,302 scaffolds no localizados es de 0.517 Gb (Richardson *et al.* 2019), y 74,228 de estos (0.124 Gb) no tienen un PNS asignado. Entre este último grupo, 72,387 scaffolds tienen menos de 5 kbp y 1841 más de 5 kbp.

La longitud total del genoma no cubierto con PNSs de la micromatriz fue el 9.46 por ciento (Cuadro 10). Dado que muchos de los cromosomas del *VicPac 3.1* (Richardson *et al.* 2019) permanecen fragmentados e incompletos, la mayoría de los scaffolds no localizados sin PNS pueden representar, una vez localizados, espacios más grandes entre los PNSs.

De los 76,508 PNSs, la mayoría de los PNSs (63,795 o el 83 por ciento) se localizan a intervalos entre 10 kbp y 40 kbp, con solo 55 PNS a intervalos  $\geq$  de 200 kbp. Teniendo en cuenta todos los scaffolds asignados a los cromosomas, la densidad cromosómica media de los PNS es de  $39 \pm 2.51$  PNS / Mb. Además, 36,993 PNS (52 por ciento) están ubicados en secuencias de intrones o exones dentro del intervalo de 14,096 en genes y/o loci anotados en *VicPac 3.1*. El FAM promedio de los PNS fue de  $0.215 \pm 0.139$ , donde el 94 por ciento de los PNS tienen FAM  $\geq 0.05$ . La distribución de los FAM de PNS se presenta en la Figura 11.

Se ha reportado que los genes *MC1R* (receptor de melanocortina 1) y *ASIP* (proteína de señalización de agutí) regulan el color de la fibra de alpaca (Feeley y Munyard 2009, Feeley *et al.* 2011). Otros genes del color de la fibra en mamíferos, tales como la tirosinasa relacionada con la proteína 1 (*TYRPI*), tirosina quinasa del receptor (*KIT*) de oncogén y el *KIT* de oncogén ligado (*KITLG*), han sido propuestos como reguladores de color de la fibra de alpaca (Feeley *et al.* 2016, Jones *et al.* 2019). Asimismo, algunos genes de la queratina (*KRT*) y de la proteína asociada a la queratina (*KRTAP*) son genes candidatos importantes para la calidad del vellón y la fibra (Allain y Renieri, 2010). Estos últimos

genes se encuentran, en grupos, en el cromosoma 12 (25 *KRT* genes) y el cromosoma 16 (22 *KRT* y 2 *KRTAP* genes) de la alpaca (Richardson *et al.* 2019).

El conjunto de 302 PNSs de genes candidatos incluidos en la micromatriz obtuvo un FAM promedio de  $0.1685 \pm 0.1361$ . Doce de estos PNSs fueron monomórficos, 175 mostraron FAM inferiores a 0.01 y 187 PNSs mostraron FAM entre 0.1 y 0.5 en los 145 animales evaluados de la población de muestra, genotipados con la micromatriz. Con la excepción de los 16 PNSs reportados en la literatura para estos genes candidatos (Chandramohan *et al.* 2013, Feeley *et al.* 2016, Fernández *et al.* 2019), todos los demás PNSs son nuevos.

En este estudio se presenta la selección de 80,201 PNSs de alpaca y el desarrollo de la primera micromatriz de PNSs de alpaca que contiene 76,508 PNS (Calderon *et al.* 2021, Anexo 9). Esta micromatriz será útil para la identificación de animales, la confirmación de pedigrís, estudios de asociación de PNSs a características fenotípicas usando el genoma completo, así como para la selección asistida por marcadores moleculares y la selección genómica. Prevemos que esta herramienta facilitará el desarrollo de programas de mejoramiento genético de alpacas Huacaya blanco a nivel de los pequeños productores.

#### **4.4. VALIDACIÓN DE LA MICROMATRIZ DE PNSs DE ALPACA**

Para la validación se analizó la concordancia de los valores de relación genómica obtenido a través del genotipado por la micromatriz de PNSs de alpaca para cada integrante de los tríos (padre, madre y progenie). Se genotificaron 68 animales que componen 30 tríos de alpaca, para evaluar la confiabilidad de la micromatriz en la evaluación de paternidad. Solo 68,700 PNSs pasaron el control de calidad del genotipo, realizado por el software Axiom Analysis Suite (Affymetrix, CA, EE. UU.). Después de la eliminación de los PNSs monomórficos y de bajo FAM ( $<0.05$ ), se utilizaron 61,086 PNSs para la construcción de la matriz de relación genómica (G). Las relaciones genómicas promedio entre padre y progenie y entre madre y progenie fueron de 0.485 y 0.489, respectivamente (Cuadro 11).

**Cuadro 11: Pedigrí y relaciones genómicas entre los miembros del trío**

<b>Progenie</b>	<b>Padre</b>	<b>Madre</b>	<b>Gpp</b>	<b>Gmp</b>
P-74	P-264	P-14	0.500	0.511
P-75	P-256	P-533	0.504	0.490
P-76	P-252	P-295	0.481	0.474
P-84	P-256	P-305	0.504	0.494
P-86	P-250	P-456	0.467	0.483
P-88	P-263	P-460	0.446	0.463
P-92	P-51	P-529	0.019*	0.480
P-105	P-50	P-364	0.482	0.476
P-113	P-52	P-329	-0.045*	0.072*
P-117	P-260	P-275	0.520	0.478
P-120	P-258	P-337	0.478	0.498
P-126	P-252	P-385	0.478	0.477
P-142	P-261	P-421	0.480	0.475
P-144	P-255	P-507	0.481	0.479
P-237	P-255	P-489	0.462	0.473
P-433	P-264	P-34	0.517	0.516
P-441	P-50	P-31	0.495	0.490
P-482	P-260	P-119	0.511	0.494
P-547	P-251	P-395	0.467	0.462
S-25	S-133	S-15	0.499	0.533
S-54	S-136	S-192	0.457	0.473
S-55	S-133	S-173	0.497	0.530
S-56	S-133	S-105	0.483	0.508
S-77	S-136	S-124	0.107*	0.487
S-83	S-136	S-96	0.463	0.481
S-85	S-136	S-68	0.528	0.515
S-89	S-133	S-97	0.456	0.475
S-191	S-133	S-174	0.108*	0.485
S-218	S-128	S-163	0.469	0.500
P-104	P-252	P-350	0.477	0.488

**Nota:** G = Relación genómica; Gpp = Relación genómica de padre y progenie; Gmp = Relación genómica de madre y progenie.

\* Valores de G que no corresponden con la información del pedigrí

Veintiséis tríos (86 por ciento) registrados con pedigrí eran concordantes con su matriz de relaciones genómicas calculadas. Los valores G de cuatro pares padre-progenie y un par madre-progenie no eran concordantes con la información registrada de su pedigrí. Según los valores de G, se identificó a una progenie a la que se le había asignado padre y madre incorrectamente. Dentro del grupo de tríos se identificó a medios hermanos por padre, donde se observaron que los valores G de medios hermanos estaban

relacionados con los valores G del padre y progenie con valores  $G < 0.10$  (Cuadro 12). Los valores  $G < 0.10$  representaron la no concordancia con la información registrada del pedigrí, asimismo, se confirmó la asignación errónea de progenie a un padre y del mismo modo, supuestos medios hermanos no fueron confirmados por sus valores G.

**Cuadro 12: Pedigrí y relaciones genómicas de medios hermanos paternos**

Padre	Medios hermanos paternos		Gmhp
P-255	P-144	P-237	0.216
P-256	P-75	P-84	0.291
P-260	P-117	P-482	0.297
P-264	P-74	P-433	0.281
P-50	P-105	P-441	0.240
P-252	P-76	P-126	0.231
		P-104	0.268
	P-126	P-104	0.245
S-133	S-25	S-55	0.288
		S-56	0.259
		S-89	0.197
		S-191*	0.072 <sup>&amp;</sup>
	S-55	S-56	0.332
		S-89	0.164
		S-191*	0.059 <sup>&amp;</sup>
	S-56	S-89	0.261
		S-191*	0.063 <sup>&amp;</sup>
	S-89	S-191*	0.002 <sup>&amp;</sup>
S-136	S-77*	S-54	0.052 <sup>&amp;</sup>
		S-83	0.044 <sup>&amp;</sup>
		S-85	0.060 <sup>&amp;</sup>
	S-54	S-83	0.226
		S-85	0.224
	S-83	S-85	0.234

**Nota:** G = Relación genómica; Gmhp = Relación genómica de medios hermanos paternos.

\* Valores G que no corresponden con la información del pedigrí de padre y progenie.

& Valores G que no corresponden a la información del pedigrí de los medios hermanos.

Entre estos tríos, había cinco machos con dos progenies de medios hermanos cada uno y tres machos con tres, cuatro y cinco progenies de medios hermanos cada uno, respectivamente.

Para el rendimiento de la micromatriz, se evaluó el genotipado del grupo de animales que conforman los tríos (68 animales) para las relaciones de paternidad. Se observaron tasas de genotipado del 98,9 por ciento para ambos grupos de animales (68 animales y 145 animales) y tasas de conversión de 89,8 por ciento y 98,8 por ciento para 68 animales del grupo de tríos y para 145 de los 150 animales de la muestra original, respectivamente. Estas tasas de conversión son comparables a las tasas de conversión obtenidas con la micromatriz para otras especies domésticas (Groenen *et al.* 2011, Ramos *et al.* 2009, Matukumalli *et al.* 2009).

Las relaciones genómicas calculadas en 4 de los 30 pares padre-progenie no eran concordantes con la anotación del pedigrí de su padre. Esto equivale a una tasa de error del 13 por ciento, que fue similar a las tasas de error de anotación de padres en el pedigrí para otras especies (Oliehoek y Bijma 2009). Asimismo, se observó que una madre se asignó erróneamente a un registro de pedigrí de progenie y que una de las treinta progenes analizadas tenía ambos progenitores, padre y madre, asignados incorrectamente a su registro de pedigrí. El error de anotación de la madre en alpacas es raro y los errores en los que ambos padres están anotados incorrectamente son aún más raros (Oliehoek y Bijma 2009). De manera similar, entre los grupos de medios hermanos, existía dos medios hermanos asignados incorrectamente a un grupo.

Estos resultados resaltan las dificultades que encuentran las granjas para el mantenimiento de buenos registros de genealogía y la utilidad de la información de un gran número de PNSs, contenidas en la micromatriz, para corregir la relación de parentesco entre individuos en ausencia de registros genealógicos.

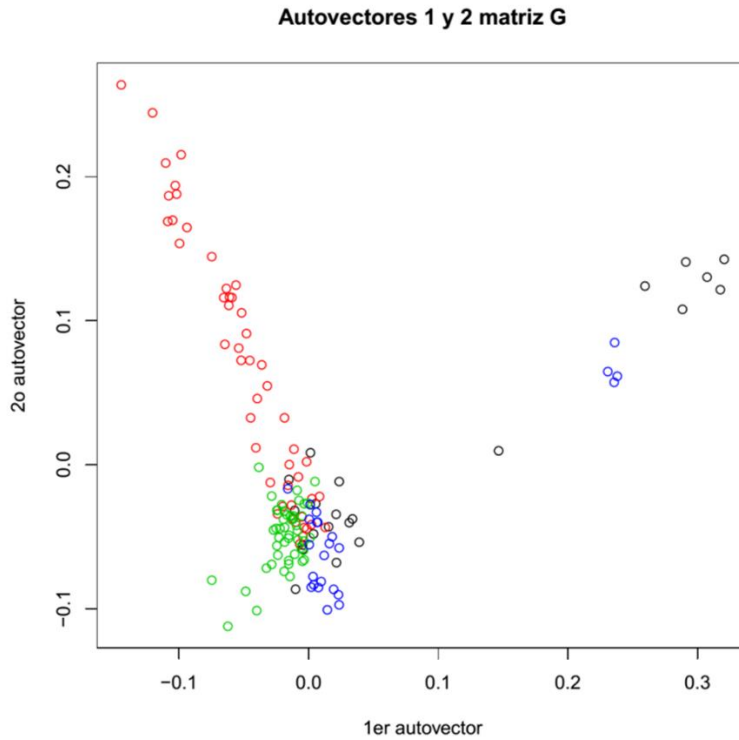
#### **4.5. COMPARACIÓN DEL GENOTIPADO POR SECUENCIAMIENTO CON EL GENOTIPADO POR MICROMATRIZ**

El genotipado por micromatriz (GM) se comparó con el genotipado por secuenciamiento (GBS) para lo cual se utilizó las muestras de ADN de alpacas que sirvieron para la identificación de los PNSs. Para la comparación de los resultados de genotipado entre los métodos GBS y GM, se utilizó la información del genotipado de 76,508 PNSs de las 145 muestras de ADN de alpacas. Las comparaciones de genotipado entre GBS y GM incluyeron 57,697 y 17,880 PNS de ApeK1 y Pst1-Msp1 de las BRR, respectivamente.

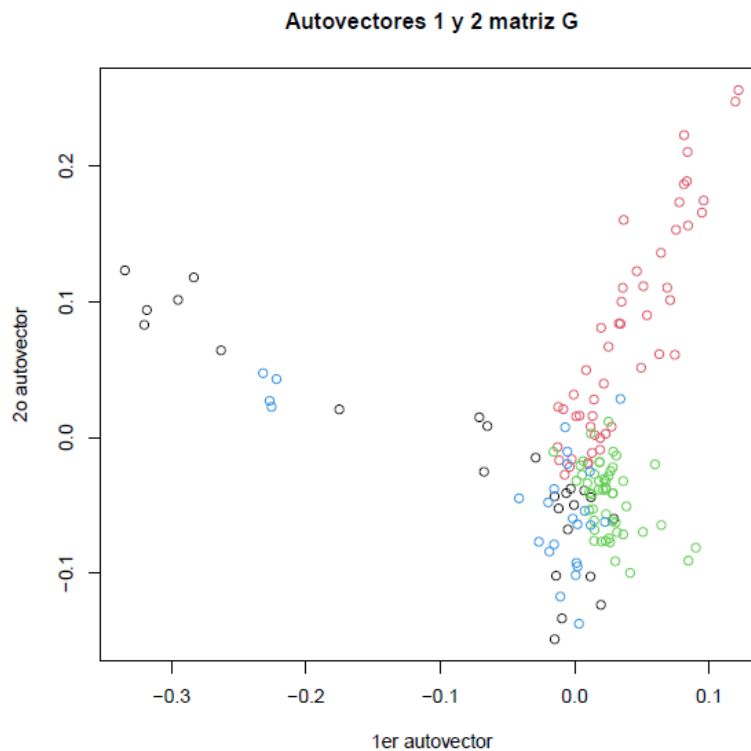
La comparación de los resultados del genotipado por GM y GBS, se realizó en base de las llamadas de genotipo por animal resultantes de GBS y GM para cada PNS. Solo se compararon los PNSs que fueron genotipados por ambos métodos. Se observó que la concordancia de genotipado de los PNSs entre el GM y GBS para el grupo de PNSs generados por la enzima ApeK1 fue de 0.93, y para los PNS generados por las enzimas MspI-PstI fue de 0.94 (Anexos 4 y 5). Estos niveles de concordancia se consideran buenos cuando tenemos en cuenta los criterios utilizados para las llamadas de PNS por GBS basadas en un mínimo de tres lecturas de secuenciación por animal. Este último criterio no es lo suficientemente robusto para las llamadas de heterocigotos que dependen de una mayor profundidad de secuencia de lectura para determinar el genotipo.

#### **4.6. ESTRUCTURA GENÉTICA DE LA POBLACIÓN**

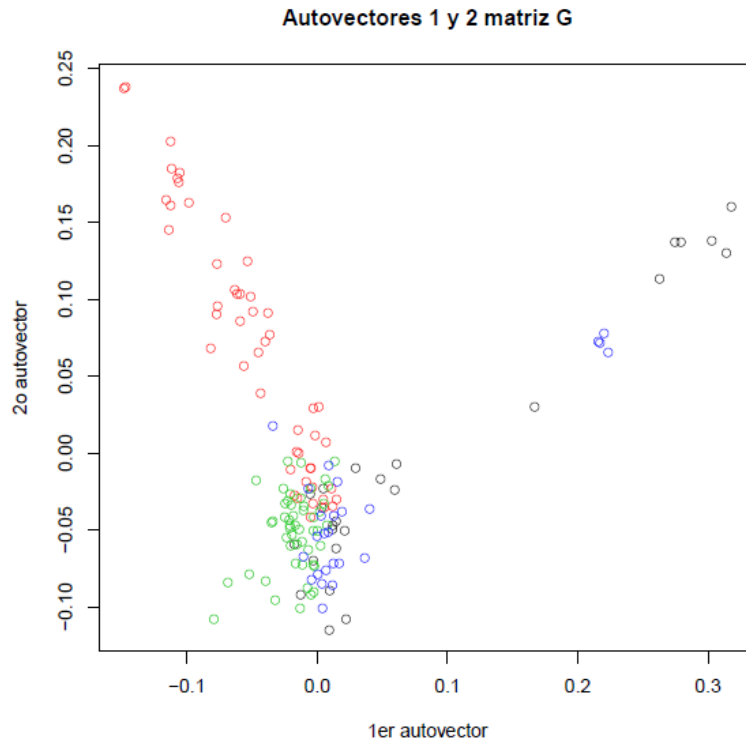
Ambos métodos de genotipado (GM y GBS) se obtuvo los valores de autovectores (Anexo 6) de la estructura genética de la población similares de los animales muestreados (Figura 13, 14 y 15). Los animales de ambas granjas de Puno (Pacomarca y Quimsachata) presentaron una mínima proporción de PNSs similares y el aislamiento del grupo de animales de Pacamarca. Sin embargo, una parte de los animales de Cerro de Pasco (Gacocen y Racco) mostraron mayor proporción de PNSs similares con los animales de Quimsachata. El grupo de animales más diferenciado fue el grupo Pacamarca y una parte de los animales de Racco. Quimsachata muestra más diferenciación de Pacamarca que las otras granjas (Figura 13, 14 y 15).



**Figura 13: Estructura de la población basada en el genotipado por la micromatriz (Pacomarca-rojo, Quimsachata-verde, Racco-azul y Gacocen-negro).**



**Figura 14: Estructura genética de la población basada en el genotipado por secuenciación con resultados de PstI-MspI (Pacomarca-rojo, Quimsachata-verde, Racco-azul y Gacocen-negro).**

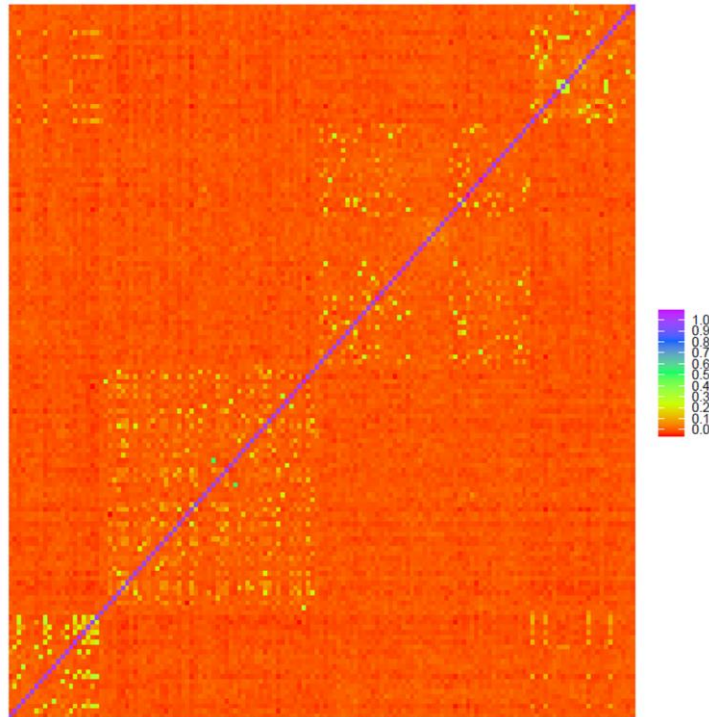


**Figura 15: Estructura genética de la población basada en el genotipado por secuenciamiento con resultados de ApeK1 (Pacomarca-rojo, Quimsachata-verde, Racco-azul y Gacocen-negro).**

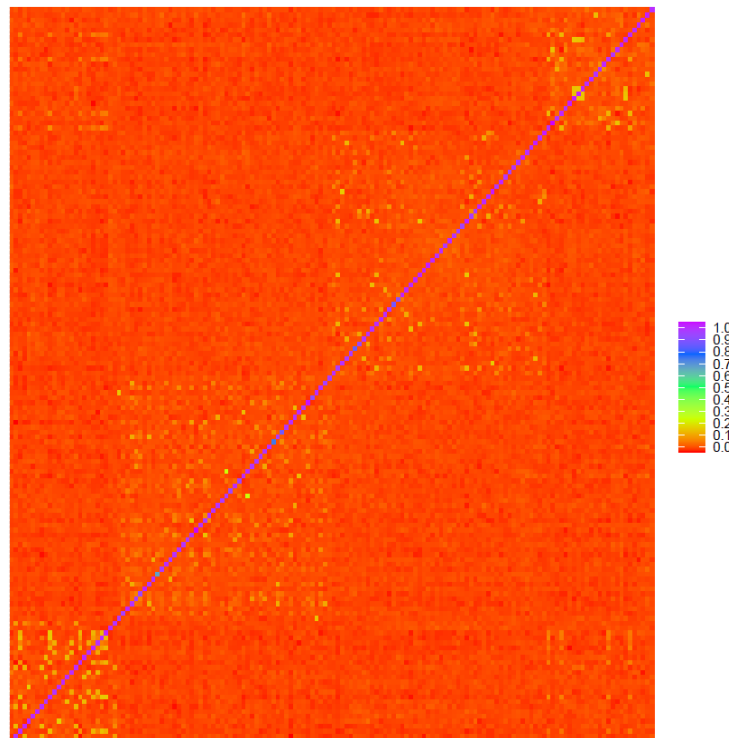
En el mapa de calor de las relaciones genómicas (Figura 16, 17 y 18) se observó cuatro bloques de subpoblación sobre la línea diagonal (Ejes de la X e Y representado por PNSs por animal). Donde Gacocen se ubica abajo a la izquierda, Pacamarca y Quimsachata en el área central y Racco en la parte de arriba a la derecha. Se observó que existen relaciones de parentesco entre algunos animales de Gacocen y Racco por tanto la diferenciación genética es menor entre ellas. Con respecto a los animales de Pacamarca y Quimsachata se observó que existen relaciones de parentesco entre algunos animales dentro de cada granja, pero no entre granjas.

La proporción de valores en el rango de 0.20 a 0.35 de relaciones genómicas dentro de cada granja fue 11.23 por ciento para Gacocen, 5 por ciento para Racco, 1.45 por ciento para Quimsachata y 0.82 por ciento para Pacamarca. Además, el número de muestras de las granjas de Cerro de Pasco son más pequeñas a comparación de las granjas de Puno, por lo cual hay que tomar con precaución estos resultados.

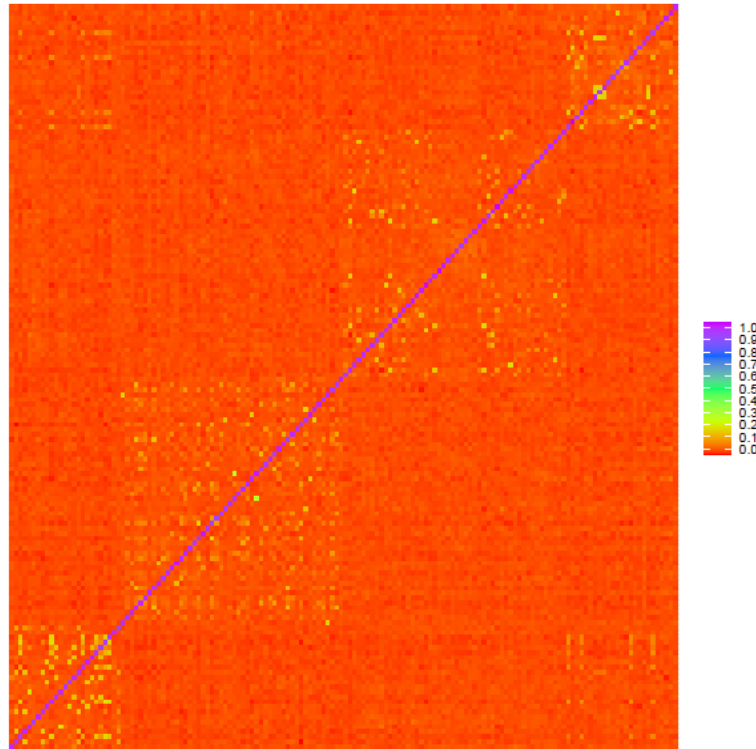




**Figura 16:** Mapa de calor de las relaciones genómicas entre animales basado en el genotipado por la micromatriz Gacocen (abajo a la izquierda), Pacamarca y Quimsachata (área central) y Racco (arriba a la derecha).

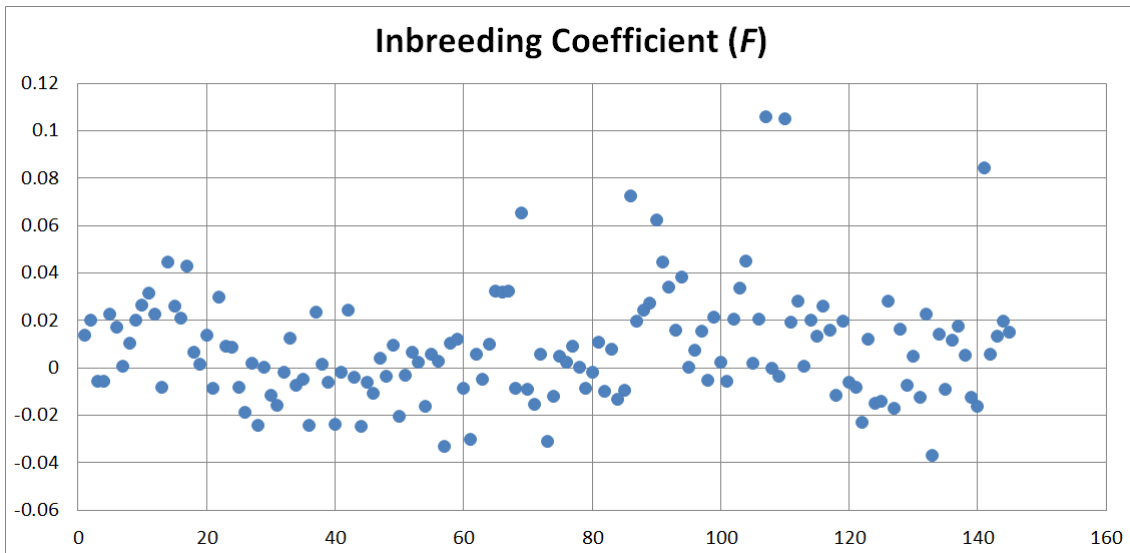


**Figura 17:** Mapa de calor de las relaciones genómicas entre animales basada en el genotipado por secuenciamiento con resultados de PstI-Mspl (Gacocen (abajo a la izquierda), Pacamarca y Quimsachata (área central) y Racco (arriba a la derecha)).

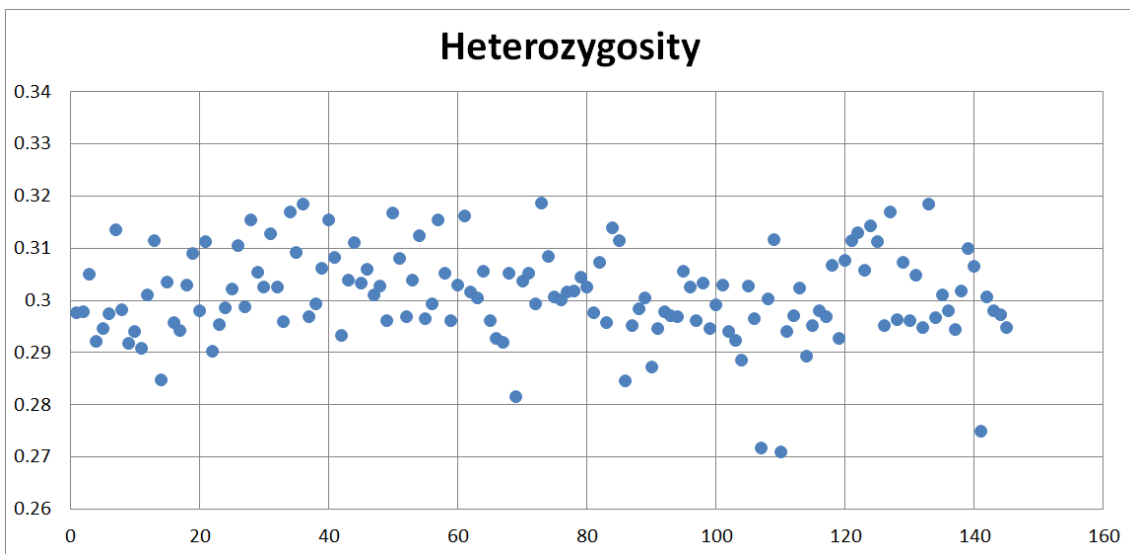


**Figura 18:** Mapa de calor de las relaciones genómicas entre animales basada en el genotipado por secuenciamiento con resultados de ApeK1 (Gacocen (abajo a la izquierda), Pacomarca y Quimsachata (área central) y Racco (arriba a la derecha)).

El promedio del coeficiente de consanguinidad (F) calculado con los datos de GM fue  $0.004 \pm 0.02$ , y usando el genotipado GBS para PNS MspI-PstI y ApeK1 fue  $0.1060 \pm 0.033$  y  $0.158 \pm 0.018$ , respectivamente (Figura 19, Anexo 7). El coeficiente promedio de heterocigosidad individual para el genotipado por la micromatriz fue  $0.27 \pm 0.01$ , y para el genotipado GBS para PNS MspI-PstI y ApeK1 fue  $0.1327 \pm 0.037$  y  $0.0947 \pm 0.0225$ , respectivamente (Figura 20, Anexo 8).



**Figura 19: Coeficiente de consanguinidad (F) de alpacas genotipado por micromatriz.**



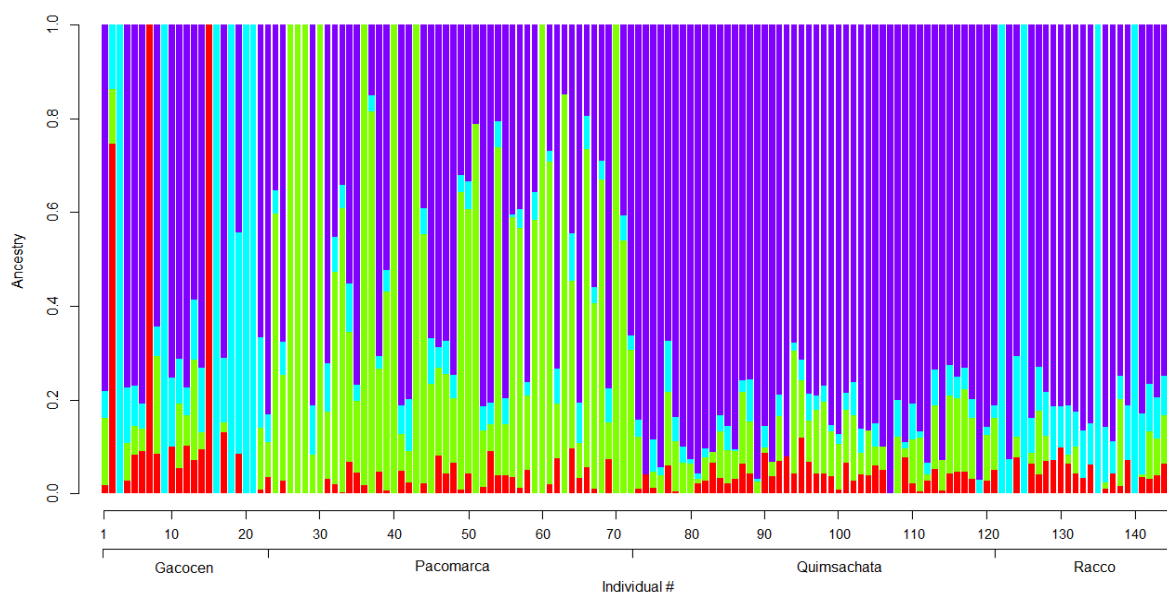
**Figura 20: Nivel de heterocigosidad de alpacas genotipado por micromatriz.**

La divergencia genética de los cuatro grupos de animales se basó en el uso de 76,206 PNS en el análisis. El grado de divergencia entre poblaciones fue estimado utilizando el índice de fijación ( $F_{st}$ ) (Cuadro 13). En relación a la diferenciación genética de Gacocen con las otras poblaciones, se observó relativamente menos divergencia con Racco ( $F_{st} = 0.134$ ), que con Pacamarca ( $F_{st} = 0.156$ ) y con Quimsachata ( $F_{st} = 0.177$ ). De manera similar, Pacamarca y Quimsachata son más divergentes ( $F_{st} = 0.098$ ) comparado con Pacamarca y Racco ( $F_{st} = 0.053$ ) y con Quimsachata y Racco ( $F_{st} = 0.075$ ).

La diferenciación genética entre las poblaciones se puede visualizar en la Figura 21 que muestra el análisis de ascendencia por individuo para los 145 animales considerando un valor de  $K = 4$ , con error de validación cruzada de 0.53902.

**Cuadro 13: Índice de fijación ( $F_{st}$ ) entre poblaciones de alpacas.**

	Quimsachata	Racco	Gacocen
Pacomarca	0.098	0.053	0.156
Quimsachata		0.075	0.177
Racco			0.134



**Figura 21: Análisis de diferenciación genética de las granjas de alpaca por individuo utilizando el software ADMIXTURE ( $K = 4$ ).**

El análisis con el software ADMIXTURE proporcionó información sobre la diferenciación genética entre los grupos de animales y permitió evaluar las diferencias genéticas, con cautela, entre regiones y entre granjas. Tanto las estructuras de población, como los mapas de calor de relaciones genómicas obtenidos por separado para GM y GBS (ApeK1 y Pst1-Msp1) indican que existe mínima diferenciación genética para algunos animales entre la región de Puno y la región de Cerro de Pasco.

La población de animales más diferenciado fue el de Pacamarca. Las otras tres granjas mostraron mínima proporción de PNSs similares en la población, con la excepción de algunos animales de Racco (Figura 13). Racco tiene mayor proporción de PNSs similares con las granjas de Pacamarca y Quimsachata que con Gacocen. Por lo tanto, parece que las granjas de Cerro de Pasco han adquirido animales y/o han utilizado reproductores de la región de Puno para el mejoramiento genético. Sin embargo, estas observaciones no se pueden generalizar porque la muestra de animales por granja se priorizó en el contenido de la información de genealogía y producción (primera y segunda esquila). Por otro lado, los mapas de calor de relaciones genómicas entre animales muestran un mayor número de animales emparentados entre las poblaciones de Gacocen y Racco ubicadas en Cerro de Pasco que las granjas de Pacamarca y Quimsachata ubicadas en Puno.

El promedio del coeficiente de consanguinidad fue de 0.004 y el promedio de heterocigosidad fue de 0.27 cuando se evaluó con datos de GM. Estos valores parecen estar en contradicción porque 0.27 de heterocigosidad implica 0.73 de homocigosidad, y por lo tanto se espera que el valor de consanguinidad sea superior al 0.004 observado. Sin embargo, el coeficiente de consanguinidad y la homocigosidad de múltiples loci no miden lo mismo. La consanguinidad mide en un individuo si los PNSs son idénticos por descendencia, y la homocigosidad mide la posesión de dos alelos idénticos dentro de cada loci (Hartl *et al.* 1997).

En un estudio para evaluar la relación entre el coeficiente de consanguinidad (F) y la heterocigosidad de múltiples locus (MLH) (Slate *et al.* 2004), se encontró que MLH era un mal indicador de F incluso en poblaciones donde la consanguinidad era similar. Estos autores también indicaron que sus hallazgos eran consistentes con otros hallazgos que no detectaron una correlación significativa entre H y F en grandes poblaciones de apareamiento aleatorio (Houle 1989, Savolainen y Hedrick 1995) o en poblaciones estructuradas (Whitlock 1993). Además, indicaron que el marcador de heterocigosidad no proporciona una estimación sólida de la heterocigosidad en todo el genoma, sino que solo puede reflejar la heterocigosidad para los loci vinculados.

En base a lo anterior, la heterocigosidad de 0.27 medida a partir de los genotipos de los 145 animales es quizás una consecuencia de la selección preferible de PNS con alto FAM. De hecho, los PNS de la micromatriz tienen FAM promedio de  $0.215 \pm 0.139$  donde más de 31,000 PNS presentan  $FAM > 0.25$ . Otro factor a considerar es que

alrededor de 60 animales mostraron valores  $F$  negativos.  $F$  solo puede ser negativo si se observan más genotipos heterocigotos de los esperados (Van Der Auwera y O'connor 2020). Por tanto, muchos de los PNS evaluados no se encontrarían en equilibrio de Hardy-Weinberg (Vieira *et al.* 2013). En general, la heterocigosidad observada de nuestra muestra no se puede utilizar como predictor del coeficiente de consanguinidad.

El desarrollo de la micromatriz de alpaca con 76k PNS representa un importante paso para avanzar en los estudios genéticos de los camélidos sudamericanos (CSA). El uso de la micromatriz para el genotipado de otros CSA aún no se ha probado. Sin embargo, se espera que una gran proporción de los PNS de alpaca se conserve en las otras especies de camélidos.

## V. CONCLUSIONES

Bajo las condiciones en la cual se desarrolló el estudio se concluye que:

1. La construcción de 300 bibliotecas de representación reducida, a partir de la digestión de ADN con la enzima ApeK1 y con las enzimas Pst1-Msp1 de 150 muestras de ADN de alpacas Huacaya, permitió identificar 4'283,956 PNSs.
2. La aplicación de diferentes filtros de calidad permitió asignar 76,508 PNSs a la primera micromatriz de PNSs de alpaca, que incluyo a 302 PNSs localizados en secuencias codificantes de genes candidatos para la calidad y el color de la fibra.
3. La micromatriz fabricada cuenta una cobertura estimada del 90.5 por ciento del genoma de referencia VicPac 3.1. Cada PNS se encuentra a una distancia promedio de  $26.58 \pm 18.57$  kpb.
4. La aplicación de los resultados del genotipado de PNSs con la micromatriz permitió verificar la información de genealogía con el análisis de relaciones genómicas y calcular el coeficiente de consanguinidad. Asimismo, permitió observar diferenciación genética entre las poblaciones y la existencia de relaciones de parentesco entre individuos dentro de una población y entre poblaciones

## **VI. RECOMENDACIONES**

- 1.** Continuar el genotipado de otras poblaciones de alpaca por región y futuras generaciones para la renovación de marcadores moleculares tipo PNS en alpacas, identificado mediante bibliotecas de representación reducida.
- 2.** Elaborar proyectos para verificar e implementar los registros de genealogía en un análisis utilizando la micromatriz para verificar las relaciones de parentesco en alpacas.
- 3.** Implementar metodologías para la aplicación y uso de la selección asistida por marcadores y la selección genómica como alternativa para aumentar el progreso genético.
- 4.** Desarrollar estudios de asociación de características productivas con marcadores moleculares de tipo PNS en alpacas.



## VII. REVISIÓN BIBLIOGRÁFICA

Alexander, D.H.; Novembre, J.; Lange, K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.*, 19, 1655–1664.

Alexander, D.H.; Shringarpure, S.S.; Novembre, J.; Lange, K. 2020. Admixture 1.3 Software Manual 2020. Available online: <http://dalexander.github.io/admixture/admixture-manual.pdf> (accessed on 13 November 2020).

Altshuler, D.; Pollara, V.J.; Cowles, C.R.; VAN Etten, W.J.; Baldwin, J.; Linton, L.; Lander, E.S. 2000. An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature*. 2000, 407 (6803): 513-6. 10.1038/35035083.

Allain, D.; Renieri, C. 2010. Genetics of fibre production and fleece characteristics in small ruminants, Angora rabbit and South American camelids. *Animal* 2010, 4, 1472–1481.

Al-Samarai, F.R.; Al-Kazaz, A.A. 2015. Applications of Molecular Markers in Animal Breeding: A review. *American Journal of Applied Scientific Research* 1(1): 1-5.

Aslam, M.L.; Bastiaansen, J.W.; Elferink, M.G. ET AL. 2012. Whole genome SNP discovery and analysis of genetic diversity in Turkey (*Meleagris gallopavo*). *BMC Genomics* 13, 391. <https://doi.org/10.1186/1471-2164-13-391>

Ávila, F. 2014. Comparative mapping of the alpaca genome. PhD Dissertation in Biomedical Sciences. Texas A&M University, EE.UU. 179 p.

Ávila, F.; Baily, M.P.; Perelman, P.; Das, P.J.; Pontius, J.; Chowdhary, R.; Owens, E.; Johnson, W.E.; Merriwether, D.A.; Raudsepp, T. 2015. A comprehensive whole-genome integrated cytogenetic map for the alpaca (*Lama pacos*). *Cytogenetic Genome Research* 144: 193-204.

Balmus, G.; Trifonov, V.A.; Biltueva, L.S.; O'Brien, P.C.M.; Alkalaeva, E.S.; FU, B.; Skidmore, J.A.; Allen, T.; Graphodatsky, A.S.; Yang, F.; et al. 2007. Cross-species chromosome painting among camel, cattle, pig and human: Further insights into the putative Cetartiodactyla ancestral karyotype. *Chromosom. Res.* 2007, 15.

Bcftools. Doc. 2019. Available online: <http://samtools.github.io/bcftools> (accessed on 18 February 2019).

Bertolini, F.; Schiavo, G.; Scotti, E.; Ribani, A.; Martelli, P.L.; Casadio, R.; Fontanesi, L. 2014. High-throughput SNP discovery in the rabbit (*Oryctolagus cuniculus*) genome by next-generation semiconductor-based sequencing. *Anim Genet. Apr*; 45(2):304-7. doi: 10.1111/age.12121. Epub 2014 Jan 21. PMID: 24444082.

Bertolini, F.; Elbeltagy, A.; Ponce de Leon, F.A.; Gutierrez, G.A.; Rothschild, M. 2016. Applicability of using bovine, ovine and caprine SNP chips for alpaca and dromedary genomic studies. 35th International Society for Animal Genetics Conference. Utah. p. 85.

Bianchi, N.O.; Larramendy, M.L.; Bianchi M.S.; Cortes, L. 1986. Karyological conservatism in South American camelids. *Experientia* 42: 622-624.

Bovo, S.; Bertolini, F.; Schiavo, G.; Mazzoni, G.; Dall'olio, S.; Fontanesi, L. 2015. Reduced Representation Libraries (RRLs) from DNA Pools Analysed with Next Generation Semiconductor Based-Sequencing to Identify SNPs in Extreme and Divergent Pigs for Back Fat Thickness. Hindawi Publishing Corporation International Journal of Genomics Volume 2015, Article ID 950737, 8 pages <http://dx.doi.org/10.1155/2015/950737>.

Brookes, A.J. 1999. The essence of SNPs. *Gene* 234 (2): 177-186.

Brown, TA. 2007. *Genomes 3*. Third ed. New York. Wiley. 69-70 p.

Calderon, M; More, MJ; Gutierrez, GA; Ponce de León, FA. 2021. Development of a 76k Alpaca (*Vicugna pacos*) Single Nucleotide Polymorphisms (SNPs) Microarray (en línea). *Genes* 12(2):291. DOI: <https://doi.org/10.3390/genes12020291>.

Carazas, Y. 2007. Línea de base: estrategias de mitigación en comunidades campesinas criadoras de alpaca. En: Familias alpaqueras enfrentando al cambio climático. Soluciones Prácticas. Moya, Enrique y Torres, Juan (Ed). 110 pp.

Carhuaricra, H.D.; Orozco-Terwengel, P.; Rosadio, A.R.; Rojas, D.F. y Maturrano, H.L. 2022. Diversidad genética y estructura poblacional del ovino Junín mediante el uso de microarreglos de alta densidad de marcadores polimórficos de nucleótido simple (SNP). Revista de Investigaciones Veterinarias del Perú, 33(1), e21459. <https://doi.org/10.15381/rivep.v33i1.21459>

CENAGRO. Censo Nacional Agropecuario .2012. CENAGRO Resultados Definitivos: IV Censo Nacional Agropecuario-2012; Instituto Nacional de Estadística e Informática (INEI): Julio, Perú, 2013; Volume 63, p. 18. Available online: <https://sinia.minam.gob.pe/documentos/resultados-definitivos-iv-censo-nacional-agropecuario-2012-0> (accessed on 13 November 2020).

Cerón, M.E. 2015. Diversidad microbiana del estómago de los camélidos sudamericanos. VII Congreso mundial en Camélidos Sudamericanos Puno – Perú 2015. <https://www.researchgate.net/publication/306392853>.

Chandramohan, B.; Renieri, C.; La Manna, V.; La Terza, A. 2013. The alpaca agouti gene: Genomic locus, transcripts and causative mutations of eumelanic and pheomelanic coat color. Gene 2013, 521.

Chandramohan, B.; Renieri, C.; La Manna, V.; La Terza, A. 2015. The Alpaca Melanocortin 1 Receptor: Gene Mutations, Transcripts, and Relative Levels of Expression in Ventral Skin Biopsies. Sci. World J. 2015.

Chang, C.C.; Chow, C.C.; Tellier, L.C.; Vattikuti, S.; Purcell, S.M.; Lee, J.J. 2015. Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience 2015, 4.

Chen, W.; LI, B.; Zeng, Z.; Sanna, S.; Sidore, C.; Busonero, F.; Kang, H.M.; Li, Y.; Abecasis, G.R. 2013. Genotype calling and haplotyping in parent-offspring trios. *Genome research*, 23(1), 142–151. <https://doi.org/10.1101/gr.142455.112>

Cruz, A.L. 2018. Parámetros genéticos de caracteres funcionales y secundarios en alpacas. Universidad Complutense de Madrid Facultad de Veterinaria. Departamento de Producción Animal. Repositorio, 1-171.

Dalma-Weiszhausz, D.D.; Warrington, J.; Tanimoto, E.Y.; Miyada, C.G. 2006. The affymetrix GeneChip platform: an overview. *Methods Enzymol.* 2006;410:3-28. doi: 10.1016/S0076-6879(06)10001-4. PMID: 16938544.

Darrier, B.; Russell, J.; Milner, S.G.; Hedley, P.E.; Shaw, P.D.; Macaulay, M.; Ramsay, L.D.; Halpin, C.; Mascher, M.; Fleury, D.L.; et al. 2019. A Comparison of Mainstream Genotyping Platforms for the Evaluation and Use of Barley Genetic Resources. *Front. Plant Sci.* 2019, 10.

Dekkers J.C. 2012. Application of genomics tools to animal breeding. *Current genomics*, 13(3), 207–212. <https://doi.org/10.2174/138920212800543057>.

Delgado de la Flor, I. 2014. Microsatélites e identificación de SNP en el gen de *Trichialina* en alpacas (*Vicugna pacos*). Tesis Magíster en Bioquímica y Biología Molecular. Lima, Perú, UPCH. 78 p.

Deniskova, T.; Dotsev, A.; Lushihina, E.; Shakhin, A.; Kunz, E.; Medugorac, I.; Reyer, H.; Wimmers, K.; Khayatzadeh, N.; Sölkner, J.; et al. 2019. Population Structure and Genetic Diversity of Sheep Breeds in the Kyrgyzstan. *Front. Genet.*, 10.

Du, Y.; Jiang, H.; Chen, Y.; LI, C.; Zhao, M.; Wu, J.; Qiu, y.; LI, Q.; Zhang, X. 2012. Comprehensive evaluation of SNP identification with the Restriction Enzyme-based Reduced Representation Library (RRLs) method. *BMC Genomics* 2012, 13:77.

Fan, J.B.; Oliphant, A.; Shen, R.; Kermani, B.G.; Garcia, F.; Gunderson, K.L.; Hansen, M.; Steemers, F.; Butler, S.L.; Deloukas, P.; Galver, L.; et al. 2003. Highly parallel SNP

genotyping. Cold Spring Harb Symp Quant Biol. 2003;68:69-78. doi: 10.1101/sqb.2003.68.69. PMID: 15338605.

Fan, B.; Zh, D.U.; Gorbach, D.M.; Rothschild, M.F. 2010. Development and Application of High-density SNP Arrays in Genomic Studies of Domestic Animals. Department of Animal Science and Center for Integrated Animal Genomics, Iowa State University, Ames, IA 50011, USA. Asian-Aust. J. Anim. Sci. Vol. 23, No. 7: 833 – 847.

Fan, R.; Gu, Z.; Guang, X.; Marín, J.C.; Varas, V.; González, B.A.; Wheeler, J.C.; Hu, Y.; Li, E.; Sun, X.; et al. 2020. Genomic analysis of the domestication and post-Spanish conquest evolution of the llama and alpaca. Genome Biol. 2020, 21.

Feeley, N.L.; Bottomley, S.; Munyard, K.A. 2011. Three novel mutations in ASIP associated with black fibre in alpacas (*Vicugna pacos*). The Journal of Agricultural Science 149(4):529-538. DOI: <https://doi.org/10.1017/S0021859610001231>.

Feeley, N.L.; Bottomley, S.; Munyard, K.A. 2016. Novel mutations in *Vicugna pacos* (alpaca) *Tyrp1* are not correlated with brown fibre colour phenotypes. Small Rumin. Res. 2016, 143.

Feeley, N.L.; Munyard, K.A. 2009. Characterisation of the melanocortin-1 receptor gene in alpaca and identification of possible markers associated with phenotypic variations in colour. Anim. Prod. Sci. 2009, 41, 675–681.

Fernández-Baca, S. 2005. Situación actual de los camélidos sudamericanos en Perú. In Proyecto de Cooperación Técnica en Apoyo de la Crianza y Aprovechamiento de los Camélidos Sudamericanos en la Región Andina TCP/RLA/2914; Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO): Julio, Perú; Volume 63.

Fernández, A.G.; Gutiérrez, G.A.; Ponce de León, F.A. 2019. Identificación bioinformática de Polimorfismos de Nucleótido Simple (PNSs) en genes candidatos para las características de la fibra en alpacas (*Vicugna pacos*). Rev. Peru. Biol. 2019, 26.

Figuroa, D. 202). Identificación bioinformática de polimorfismos de nucleótido simple (PNS) en genes de proteínas asociadas a queratinas en alpacas (*Vicugna pacos*). Tesis para optar el título profesional de Ingeniero Zootecnia, Universidad Nacional Agraria La Molina. Lima, Perú.

Foppiano, F. 2016. Caracterización de marcadores genéticos en genes que codifican a proteínas asociadas a queratina y evaluación de la asociación del gen KRTAP11-1 al diámetro de fibra en alpaca (*Vicugna pacos*) siguiendo una aproximación de gen candidato. Tesis de Maestría. Universidad Peruana Cayetano Heredia, Perú.

Gautier, L.; Cope, L.; Bolstad, B.M.; Irizarry, R.A. 2004. Affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*. Feb 12;20(3):307-15. doi: 10.1093/bioinformatics/btg405. PMID: 14960456.

Griffiths, A.; Wessler, S.; Carroll, S.; Doebley, J. 2015. *Introduction to Genetic Analysis*. Eleventh ed. New York, USA, W. H. Freeman and Company. 868 p.

Groenen, M.A.; Megens, H.-J.; Zare, Y.; Warren, W.C.; Hillier, L.W.; Crooijmans, R.P.; Vereijken, A.; Okimoto, R.; Muir, W.M.; Cheng, H.H. 2011. The development and characterization of a 60K PNS chip for chicken. *BMC Genomics* 2011, 12.

Gunderson, K.L.; Steemers, F.J.; Lee, G.; Mendoza, L.G.; Chee, M.S. 2005. A genome-wide scalable SNP genotyping assay using microarray technology. *Nature Genetics* 37(5): 549-554.

Guridi, M.; Soret, B.; Alfonso, L.; Arana, A. 2011. Single nucleotide polymorphisms in the Melanocortin 1 Receptor gene are linked with lightness of fibre colour in Peruvian Alpaca (*Vicugna pacos*). *Anim. Genet.* 2011, 42.

Hartl, D.L.; Clark, A.G. 1997. *Principles of Population Genetics*, 4th ed.; Sinauer: Sunderland, MA, USA, 1997; ISBN 9780878933082.

Heaton, M.P.; Harhay, G.P.; Bennett, G.L.; Stone, R.T.; Grosse, W.M.; Casas, E.; Keele, J.W.; Smith, T.P.; Chitko-Mckown, C.G.; Laegreid, W.W. 2002. Selection and use of

SNP markers for animal identification and paternity analysis in U.S. beef cattle. *Mammalian Genome* 13: 272-281.

Hidalgo, A.M.; Lopes, P.S.; Paixão, D.M. 2013. Fine mapping and single nucleotide polymorphism effects estimation on pig chromosomes 1, 4, 7, 8, 17 and X. *Genetics and Molecular Biology* 36(4): 511-519.

Houle, D. 1989. Allozyme associated heterosis in *Drosophila melanogaster*. *Genetics* 1989, 123, 789–801.

Huanca, W. 2015. Reproductive biotechnologies in domestic South American camelids as alternatives for genetic improvement. *Arch. Latinoam. Prod. Anim.* 2015, 23, 1–4.

Jones, M.; Sergeant, C.; Richardson, M.; Groth, D.; Brooks, S.; Munyard, K. 2019. A non-synonymous PNS in exon 3 of the KIT gene is responsible for the classic grey phenotype in alpacas (*Vicugna pacos*). *Anim. Genet.* 2019, 50.

Kijas, J.W.; Townley, D.; Dalrymple, B.P.; Heaton, M.P. 2009. A Genome Wide Survey of SNP Variation Reveals the Genetic Structure of Sheep Breeds. *PLoS ONE* 4(3): 1-13.

Lado, S.; Elbers, J.P.; Rogers, M.F.; Melo-Ferreira, J.; Yadamsuren, A.; Corander, J.; et al. 2020a. Nucleotide Diversity of Functionally Different Groups of Immune Response Genes in Old World Camels Based on Newly Annotated and Reference-Guided Assemblies. *BMC Genomics*. 21, 606. doi:10.1186/s12864-020-06990-4

Li, H.; Durbin, R. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 2010, 26.

Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009, 25.

Mamani, C.; Gutiérrez, G.; Ponce De León, A. 2017. Identificación de polimorfismos de nucleótido simple en alpaca (*vicugna pacos*) usando un panel de células híbridas irradiadas alpaca/hámster. *Revista RICBA* 1(2): 92-95.

Matukumalli, L.K.; Lawley, C.T.; Schnabel, R.D.; Taylor, J.F.; Allan, M.F.; Heaton, M.P.; O'connell, J.; Moore, S.S.; Smith, T.P.L.; Sonstegard, T.S.; et al. 2009. Development and Characterization of a High Density PNS Genotyping Assay for Cattle. *PLoS ONE* 2009, 4.

McCue, M.E.; Bannasch, D.L.; Petersen, J.L. 2012. A High Density SNP Array for the Domestic Horse and Extant Perissodactyla: Utility for Association Mapping, Genetic Diversity, and Phylogeny Studies. *PLoS Genetics* 8(1): 1-14.

Meissner, A.; Gnirke, A.; Bell, G.W.; Ramsahoye, B.; Lander, E.S.; Jaenisch, R. 2005. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* 2005, 33: 5868-5877. 10.1093/nar/gki901.

Mendoza, M.N.; Raudsepp, T.; Alshanbari, F.; Gutiérrez, G.; Ponce de León, F.A. 2019. Chromosomal Localization of Candidate Genes for Fiber Growth and Color in Alpaca (*Vicugna pacos*). *Front. Genet.* 2019, 10.

Mendoza, M.N.; Raudsepp, T.; More, M.J.; Gutiérrez, G.A.; Ponce de León, F.A. 2020. Cytogenetic Mapping of 35 New Markers in the Alpaca (*Vicugna pacos*). *Genes (Basel)*. 2020, 11, 522.

Metzker, M. 2010. Sequencing technologies - the next generation. *Nat Rev Genet* 11, 31–46. <https://doi.org/10.1038/nrg2626>

Michael, K.L.; Taylor, L.C.; Schultz, S.L.; Walt, D.R. 1998. Randomly ordered addressable high-density optical sensor arrays. *Anal Chem.* Apr 1;70(7):1242-8. doi: 10.1021/ac971343r. PMID: 9553489.



Miller, J.M.; Moore, S.M.; Stothard, P.; Liao, X.; Coltman, D.W. 2015. Harnessing crossspecies alignment to discover SNPs and generate a draft genome sequence of a bighorn sheep (*Ovis canadensis*). *BMC Genomics* 16(1): 397.

MINAGRI. (2020). Ministerio de desarrollo agrario y riego. Dirección general de estadística, seguimiento y evaluación de políticas (DGESEP). Anuario Estadístico: Producción ganadera y avícola 2020.

More, M.; Gutiérrez, G.; Rothschild, M.; Bertolini, F.; Ponce de León, F.A. 2019. Evaluation of PNS Genotyping in Alpacas Using the Bovine HD Genotyping Beadchip. *Front. Genet.* 2019, 10.

Morón, J.A.; Veli, E.A.; Membrillo, A.; Paredes, M.M.; Gutiérrez, G.A. 2020 Genetic diversity and validation of a microsatellite panel for parentage testing for alpacas (*Vicugna pacos*) on three Peruvian farms. *Small Rumin. Res.* 2020, 193, 106246.

Mustafa, H.; Khan, W.A.; Heather, H.J.; Kuthu, Z.H.; Euisoo, K.; Ajmal, A.; Ain, N.U.; Sonstegard, T.S. 2018. Performance of bovine high-density SNPs genotyping array in indigenous Pakistani cattle breeds. *Pure and Applied Biology*. Vol. 7, Issue 1, pp221-226. <http://dx.doi.org/10.19045/bspab.2018.70026>

More, M. 2022. Identificación de polimorfismos de nucleótido simple y su asociación con el diámetro de fibra en alpacas huacaya (*Vicugna pacos*). Tesis de doctorado en Ciencia Animal, Universidad Nacional Agraria La Molina. Lima, Perú.

Nazari-Ghadikolaei, A.; Mehrabani-Yeganeh, H.; Miarei-Aashtiani, S.R.; Staiger, E.A.; Rashidi, A.; Y Huson, H.J. 2018. Genome-Wide Association Studies Identify Candidate Genes for Coat Color and Mohair Traits in the Iranian Markhoz Goat. *Front. Genet.* 9:105. DOI: 10.3389/fgene.2018.00105.

Oliehoek, P.A.; Bijma, P. 2009. Effects of pedigree errors on the efficiency of conservation decisions. *Genet. Sel. Evol.* 2009, 41.

Oliphant, A.; Barker, D.L.; Stuelpnagel, J.R.; Chee, M.S. 2002. BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. *Biotechniques*. 2002 Jun; Suppl:56-8, 60-1. PMID: 12083399.

Ordoñez, C.; Cucho, H.; Ampuero, E.; Antezana, W.; Cayo, S. 2013. Inseminación Artificial de Alpacas con Semen Fresco, Refrigerado y Descongelado Colectado por Electroeyaculación. *Spermova* 2013, 3, 65–66.

Pauciullo, A.; Shuiep, E.T.; Ogah, M.D.; Cosenza, G.; DI Stasio, L.; Erhardt, G. 2019. Casein Gene Cluster in Camelids: Comparative Genome Analysis and New Findings on Haplotype Variability and Physical Mapping. *Front Genet*. 2019, 10, 748.

Paredes, M.M.; Membrillo, A.; Gutiérrez, J.P.; Cervantes, I.; Azor, P.J.; Morante, R.; Alonso, A.; Molina, A.; Muñoz-Serrano, A. 2014. Association of microsatellite markers with fiber diameter trait in Peruvian alpacas (*Vicugna pacos*). *Livestock Science* 161:6-16.

Paterson, A.H. 2006. Leafing through the genomes of our major crop plants: strategies for capturing unique information. *Nat Rev Genet*. 2006, 7: 174-184. 10.1038/nrg1806.

Patterson, N.; Price, A.L.; Reich, D. 2006. Population Structure and Eigenanalysis. *PLoS Genet*. 2006, 2.

Paucar, R. 2011. Utilidad de marcadores SNP en la mejora genética de poblaciones altoandinas de alpacas. Tesis Máster Agrobiotecnología. Universidad Pública de Navarra. 40 p.

Picard Toolkit. Broad Institute, GitHub Repository. 2019. Available online: <http://broadinstitute.github.io/picard/> (accessed on 18 January 2019).

Piro M. 2021. Aspects of Molecular Genetics in Dromedary Camel. *Front. Genet.*, 21 October 2021. <https://doi.org/10.3389/fgene.2021.723181>

Ponce de León, F.A.; Gutierrez, G.A. 2020. Genómica y producción animal. Revista Peruana De Biología, 27(1), 015–020. <https://doi.org/10.15381/rpb.v27i1.17574>.

Quispe, E.C.; Rodríguez, T.C.; Iñiguez, L.R.; Mueller, J.P. 2009. Producción de fibra de alpaca, llama, vicuña y guanaco en Sudamérica. Anim. Genet. Resour. Inf. 2009, 45, 1–14.

R Core Team. 2020. R: A Language and Environment for Statistical Computing. Available online: <http://www.R-project.org/> (accessed on 18 October 2020).

Ramos, A.M.; Crooijmans, R.P.M.A.; Affara, N.A.; Amaral, A.J.; Archibald, A.L.; Beever, J.E.; Bendixen, C.; Churcher, C.; et al. 2009. Design of a High Density PNS Genotyping Assay in the Pig Using PNSs Identified and Characterized by Next Generation Sequencing Technology. PLoS ONE 2009, 4.

Rao, D.C.; Gu C.C. 2008. Genetic dissection of complex traits. Academic Press. London, UK. 760p.

Raudsepp, T. 2012. Camelid genomics and its applications in clinical cases. Proceedings from the International Camelid Health Conference for Veterinarians. Marzo 21-25, Columbus, Ohio, USA.

Ravi, I.; Baunthiyal, M.; Saxena, J. 2014. Advances in Biotechnology. Springer India. DOI: 10.1007/978-81-322-1554-7\_6.

Richardson, M.F.; Munyard, K.; Croft, L.J.; Allnut, T.R.; Jackling, F.; Alshanbari, F.; Jevit, M.; Wright, G.A.; Cransberg, R.; Tibary, A.; et al. 2019. Chromosome-Level Alpaca Reference Genome VicPac3.1 Improves Genomic Insight Into the Biology of New World Camelids. Front. Genet. 2019, 10.

Rincón, G.; Weber, K.L.; Eenennaam, A.L.; Golden, B.L.; Medrano, J.F. 2011. Hot topic: performance of bovine high-density genotyping platforms in Holsteins and Jerseys. J Dairy Sci. 2011 Dec;94(12):6116-21. doi: 10.3168/jds.2011-4764. PMID: 22118099.

Rodríguez, J. 2014. Caracterización de la variabilidad genética en regiones anónimas y codantes en alpacas (*Vicugna pacos*) peruanas mediante el análisis de ADN mitocondrial y ADN nuclear. Tesis de doctorado en Bioquímica y Biología Molecular. Lima Perú.

RStudio Team. 2020. RStudio: Integrated Development for R. Available online: <http://www.rstudio.com/> (accessed on 18 August 2020).

Sabahat, S.; Brauning, R.; Clarke, SM.; Nadeem, A.; Thomson, PC.; Khatkar, MS. 2020. SNP discovery and population structure analysis in Lassi and Marecha camel breeds using a genotyping by sequencing method. *Anim Genet.* 2020 Aug;51(4):620-623. doi: 10.1111/age.12953. Epub 2020 May 18. PMID: 32421863.

Savolainen, O.; Hedrick, P. 1995. Heterozygosity and fitness: No association in Scots pine. *Genetics* 1995, 140, 755–766.

Shapero, M.; Gardner, S.; Mittmann, M.; Mcloughlin, K.; Jaing, C.; Bellon, L.; Slezak, T. 2014. Microbial detection using Affymetrix' Axiom® Genotyping Solution. Affymetrix, Biology for a better world.

Shen, R.; Fan, J.B.; Campbell, D.; Chang, W.; Chen, J.; Doucet, D.; Yeakley, J.; Bibikova, M.; Wickham Garcia, E.; McBride, C.; Steemers, F.; Garcia, F.; Kermani, B.G.; Gunderson, K.; Oliphant, A. 2005. High-throughput SNP genotyping on universal bead arrays. *Mutat Res.* 2005 Jun 3;573(1-2):70-82. doi: 10.1016/j.mrfmmm.2004.07.022. PMID: 15829238.

Skotte, L.; Korneliussen, T.S.; Albrechtsen, A. 2013. Estimating Individual Admixture Proportions from Next Generation Sequencing Data. *Genetics* 2013, 195, 693.

Slate, J.; David, P.; Dodds, K.G.; Veenvliet, B.A.; Glass, B.C.; Broad, T.E.M. 2004. Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: Theoretical expectations and empirical data. *Heredity (Edinb).* 2004, 93, 255–265.

Sánchez, C.C.; Smith, T.P.; Wiedmann, R.T.; Vallejo, R.L.; Salem, M.; Yao, J.; Rexroad, C.E. 3RD. 2009. Single nucleotide polymorphism discovery in rainbow trout by deep sequencing of a reduced representation library. *BMC Genomics*. 2009 Nov 25;10:559. doi: 10.1186/1471-2164-10-559. PMID: 19939274; PMCID: PMC2790473.

Timko, M.P.; Rushton, P.J.; Laudeman, T.W.; Bokowiec, M.T.; Chipumuro, E.; Cheung, F.; Town, C.D.; Chen, X. 2008. Sequencing and analysis of the gene-rich space of cowpea. *BMC Genomics*. 2008, 9: 103-10.1186/1471-2164-9-103.

Van Der Auwera, G.A.; O'connor. B.D. 2020. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (1st Edition)*. O'Reilly Media.

Van Tassell, C.P.; Smith, T.P.; Matukumalli, L.K.; Taylor, J.F.; Schnabel, R.D.; Lawley, C.T.; Haudenschild, C.D.; Moore, S.S.; Warren, W.C.; Sonstegard, T.S. 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nat Methods*. 2008, 5: 247-252. 10.1038/nmeth.1185.

Viale, E.; Zanetti, E.; Özdemir, D.; Broccanello, C.; Dalmaso, A.; De Marchi, M.; Cassandro, M. 2017. Development and validation of a novel SNP panel for the genetic characterization of Italian chicken breeds by next-generation sequencing discovery and array genotyping. *Poult Sci*. 2017 Sep 1;96 (11):3858-3866. DOI: 10.3382/ps/pex238. PMID: 29050437.

Vieira, F.G.; Fumagalli, M.; Albrechtsen, A.; Nielsen, R. 2013. Estimating inbreeding coefficients from NGS data: Impact on genotype calling and allele frequency estimation. *Genome Res*. 2013 Nov;23(11):1852-61. doi: 10.1101/gr.157388.113. Epub 2013 Aug 15. PMID: 23950147; PMCID: PMC3814885.

Vignal, A.; Milan, D.; Sancristobal, M.; Eggen, A. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. *Genet Sel Evol*. 34: 275-305.

Wang, J. 2002. An estimator for pairwise relatedness using molecular markers. *Genetics* 160, 1203–1215.

Wheeler, J.C. 1995. Evolution and present situation of the South American *Camelidae*. *Biological Journal of the Linnean Society*, vol. 54, n° 3, p. 271-295.

Wheeler, J.C.; Chikhi, L.; Bruford, M.W. 2004. Genetic Analysis of the Origins of Domestic South American Camelids. *Documenting Domestication: New Genetic and Archaeological Paradigms*.

Whitlock, M. 1993. Lack of correlation between heterozygosity and fitness in forked fungus beetles. *Heredity (Edinb)*., 70, 574–581.

Wiedmann, R.T.; Smith, T.P.; Nonneman, D.J. 2008. SNP discovery in swine by reduced representation and high throughput pyrosequencing. *BMC Genet.* 2008, 9: 81- PubMed Central Article PubMed.

Wu, H.; Guang, X.; Al-Fageeh, M.B.; Cao, J.; Pan, S.; Zhou, H.; Zhang, L.; Zhang, S.; Wang, J.; Yin, Y.; Yang, H.; Al-Swailem A.M.; Wang, J. 2014. Camelid genomes reveal evolution and adaptation to desert environments. *Nature Communications* 5: 5188.

Wu, X.; Ren, C.; Joshi, T. et al. 2010. SNP discovery by high-throughput sequencing in soybean. *BMC Genomics* 11, 469. <https://doi.org/10.1186/1471-2164-11-469>

Wu, Y.; Fan, H.; Wang, Y.; Zhang, L.; Gao, X.; Chen, Y.; Li, J. 2014. Genome-Wide Association Studies Using Haplotypes and Individual SNPs in Simmental Cattle. *PLoS ONE* 9(10): 1-11

Yalta, C.; Sotil, G.; Veli, E. 2014. Variabilidad genética y detección de error en filiación utilizando microsatélites en dos rebaños de alpacas Huacaya (*Vicugna pacos*). *Salud y Tecnología Veterinaria*. 2. 134. [10.20453/stv.2014.2253](https://doi.org/10.20453/stv.2014.2253)

Yang, J.; Lee, S.H.; Goddard, M.E.; Visscher, P.M. 2011. GCTA: A Tool for Genome-wide Complex Trait Analysis. *Am. J. Hum. Genet.* 2011, 88.

Zhang, H.; Wang, Z.; Wang, S.; Li, H. 2012. Progress of genome wide association study in domestic animals. *Journal of Animal Science and Biotechnology* 2012, 3:26.

Zhu, F.; Cui, Q.; Hou, Z. 2016. SNP discovery and genotyping using Genotyping-by-Sequencing in Pekin ducks. *Scientific Reports*. 6:36223.

Zimmerman, S.J.; Aldridge, C.L.; Oyler-McCance, S.J. 2020. An empirical comparison of population genetic analyses using microsatellite and SNP data for a species of conservation concern. *BMC Genomics* 21, 382 (2020). <https://doi.org/10.1186/s12864-020-06783-9>

## VIII. ANEXOS

### ANEXO 1: R scripts para la selección de PNSs para la Micromatriz

```
## Base de datos
```

```
# Genoma ensamblado VicPac3.1
```

```
assembly<-read.table("Vp3_assembly.txt",header=TRUE,sep="\t")
```

```
#SNPs identified based on sequencing with Phred Score>10
```

```
q10<-read.csv("VariantCall_vp3.1_APM_q10_sort_u.csv",header=TRUE)
```

```
#Missing genotyping rate and minor allele frequency
```

```
KGD<-read.csv("APM_freq_lmiss.csv",header=TRUE)
```

```
# Puntaje Illumina
```

```
Illumina<-read.csv("Illumina_Score.csv",header=TRUE)
```

```
#SNPs without other SNPs within flanking sequences
```

```
FSS40<-read.csv("snp_vp3.1_apm_FS_FSS40_SU_uniq.bed",header=FALSE,sep=" ")
```

```
FSS35<-read.csv("snp_vp3.1_apm_FS_FSS35_SU_uniq.bed",header=FALSE,sep=" ")
```

```
## Manejo de datos
```

```
FSS40a<-FSS40[,c(1,3)]
```

```
FSS40a_col<-c("CHROM","POS")
```

```
colnames(FSS40a)<-FSS40a_col
```

```
FSS40a$CHROM_POS<-paste(FSS40$V1,"_",FSS40$V3,sep="")
```

```
write.csv(FSS40a,"FSS40a.csv",row.names=FALSE)
```

```
FSS35a<-FSS35[,c(1,3)]
```

```
FSS35a_col<-c("CHROM","POS")
```



```

colnames(FSS35a)<-FSS35a_col

FSS35a$CHROM_POS<-paste(FSS35$V1,"_",FSS35$V3,sep="")

write.csv(FSS35a,"FSS35a.csv",row.names=FALSE)

q10_KGD<-merge(x=q10, y=KGD, by.x="X.CHROM_POS", by.y="NAME")

q10_KGD_Illumina<-merge(x=q10_KGD, y=Illumina, by.x="X.CHROM_POS",
by.y="NAME")

q10_KGD_Illumina_FSS40<-merge(x=q10_KGD_Illumina, y=FSS40a,
by.x="X.CHROM_POS", by.y="CHROM_POS")

q10_KGD_Illumina_FSS35<-merge(x=q10_KGD_Illumina, y=FSS35a,
by.x="X.CHROM_POS", by.y="CHROM_POS")

library(dplyr)

library(ggplot2)

library(ggpubr)

library(psych)

### Selección de Rondas

# Ronda 1

filter_CR45_MAF550_IScore60_FSS40<-
subset(q10_KGD_Illumina_FSS40,q10_KGD_Illumina_FSS40$F_MISS<0.55 &
q10_KGD_Illumina_FSS40$MAF>=0.05 & q10_KGD_Illumina_FSS40$MAF<=0.5 &
q10_KGD_Illumina_FSS40$Final_Score>=0.6)

filter2_CR45_MAF550_IScore60_FSS40<-
distinct(filter_CR45_MAF550_IScore60_FSS40, X.CHROM, POS.x, .keep_all= TRUE)

```

```

table_CR45_MAF550_IScore60_FSS40<-
filter2_CR45_MAF550_IScore60_FSS40[,c(2:3,10,14)]

q10_CR45_MAF550_IScore60_FSS40_assembly<-
merge(x=table_CR45_MAF550_IScore60_FSS40,y=assembly,by.x="X.CHROM",by.y="G
enBank_Accn")

write.csv(q10_CR45_MAF550_IScore60_FSS40_assembly[,c("X.CHROM","POS.x","Assi
gned_Molecule","Assigned_Molecule_Location_Type","Sequence_Length")], "q10_CR45_
MAF550_IScore60_FSS40_assembly.csv",row.names=FALSE)

q10_CR45_MAF550_IScore60_FSS40_assembly$pos4k<-
(q10_CR45_MAF550_IScore60_FSS40_assembly$POS.x)/40000

q10_CR45_MAF550_IScore60_FSS40_assembly$gpos<-
ceiling(q10_CR45_MAF550_IScore60_FSS40_assembly$pos4k)

q10_CR45_MAF550_IScore60_FSS40_assembly$coverage <-
if_else((q10_CR45_MAF550_IScore60_FSS40_assembly$gpos*40000) <
q10_CR45_MAF550_IScore60_FSS40_assembly$Sequence_Length, 40000,
q10_CR45_MAF550_IScore60_FSS40_assembly$Sequence_Length-
((q10_CR45_MAF550_IScore60_FSS40_assembly$gpos-1)*40000))

q10_CR45_MAF550_IScore60_FSS40_assembly$$S<-
(40000*(q10_CR45_MAF550_IScore60_FSS40_assembly$gpos-1))+1

q10_CR45_MAF550_IScore60_FSS40_assembly$E<-
40000*(q10_CR45_MAF550_IScore60_FSS40_assembly$gpos)

q10_CR45_MAF550_IScore60_FSS40_assembly$EmS<-
q10_CR45_MAF550_IScore60_FSS40_assembly$E-
q10_CR45_MAF550_IScore60_FSS40_assembly$$S

q10_CR45_MAF550_IScore60_FSS40_assembly$aas<-
2*q10_CR45_MAF550_IScore60_FSS40_assembly$POS.x

q10_CR45_MAF550_IScore60_FSS40_assembly$EMS<-
q10_CR45_MAF550_IScore60_FSS40_assembly$E+q10_CR45_MAF550_IScore60_FSS4
0_assembly$$S

q10_CR45_MAF550_IScore60_FSS40_assembly$Score<-
q10_CR45_MAF550_IScore60_FSS40_assembly$EmS-
(abs(q10_CR45_MAF550_IScore60_FSS40_assembly$aas-
q10_CR45_MAF550_IScore60_FSS40_assembly$EMS))

q10_CR45_MAF550_IScore60_FSS40_assembly$FSS<-"40pb"

```

```

q10_CR45_MAF550_IScore60_FSS40_assembly$Round<-1

q10_CR45_MAF550_IScore60_FSS40_assembly$f<-
ceiling((q10_CR45_MAF550_IScore60_FSS40_assembly$pos4k-
trunc(q10_CR45_MAF550_IScore60_FSS40_assembly$pos4k))*5)

q10_CR45_MAF550_IScore60_FSS40_assembly$find<-
if_else(q10_CR45_MAF550_IScore60_FSS40_assembly$f<=3,q10_CR45_MAF550_IScor
e60_FSS40_assembly$f+2,q10_CR45_MAF550_IScore60_FSS40_assembly$f-2)

write.csv(q10_CR45_MAF550_IScore60_FSS40_assembly[,c("X.CHROM","POS.x","Assi
gned_Molecule","Assigned_Molecule_Location_Type","Sequence_Length","pos4k","gpos",
"coverage","S","E","EmS","aa","EMS","Score","FSS","Round","f","find")],q10_CR45_M
AF550_IScore60_FSS40_assembly_coverage.csv",row.names=FALSE)

q10_CR45_MAF550_IScore60_FSS40_assembly2<-
q10_CR45_MAF550_IScore60_FSS40_assembly[order(q10_CR45_MAF550_IScore60_FS
S40_assembly$X.CHROM,q10_CR45_MAF550_IScore60_FSS40_assembly$gpos,q10_CR
45_MAF550_IScore60_FSS40_assembly$Score,decreasing = TRUE),]

write.csv(q10_CR45_MAF550_IScore60_FSS40_assembly2[,c("X.CHROM","POS.x","Ass
igned_Molecule","Assigned_Molecule_Location_Type","Sequence_Length","pos4k","gpos
","coverage","S","E","EmS","aa","EMS","Score","FSS","Round","f","find")],q10_CR45_
MAF550_IScore60_FSS40_assembly2_coverage.csv",row.names=FALSE)

q10_CR45_MAF550_IScore60_FSS40_assembly3<-
q10_CR45_MAF550_IScore60_FSS40_assembly2[!duplicated(q10_CR45_MAF550_IScor
e60_FSS40_assembly2[,c("X.CHROM","gpos")]), ]

write.csv(q10_CR45_MAF550_IScore60_FSS40_assembly3[,c("X.CHROM","POS.x","Ass
igned_Molecule","Assigned_Molecule_Location_Type","Sequence_Length","pos4k","gpos
","coverage","S","E","EmS","aa","EMS","Score","FSS","Round","f","find")],q10_CR45_
MAF550_IScore60_FSS40_assembly3_coverage.csv",row.names=FALSE)

# Ronda 2

filter_CR45_MAF550_IScore60_FSS35<-
subset(q10_KGD_Illumina_FSS35,q10_KGD_Illumina_FSS35$F_MISS<0.55 &
q10_KGD_Illumina_FSS35$MAF>=0.05 & q10_KGD_Illumina_FSS35$MAF<=0.5 &
q10_KGD_Illumina_FSS35$Final_Score>=0.6)

```

```

filter2_CR45_MAF550_IScore60_FSS35<-
distinct(filter_CR45_MAF550_IScore60_FSS35, X.CHROM, POS.x, .keep_all= TRUE)

table_CR45_MAF550_IScore60_FSS35<-
filter2_CR45_MAF550_IScore60_FSS35[,c(2:3,10,14)]

q10_CR45_MAF550_IScore60_FSS35_assembly<-
merge(x=table_CR45_MAF550_IScore60_FSS35,y=assembly,by.x="X.CHROM",by.y="G
enBank_Accn")

write.csv(q10_CR45_MAF550_IScore60_FSS35_assembly[,c("X.CHROM","POS.x","Assi
gned_Molecule","Assigned_Molecule_Location_Type","Sequence_Length")], "q10_CR45_
MAF550_IScore60_FSS35_assembly.csv",row.names=FALSE)

q10_CR45_MAF550_IScore60_FSS35_assembly$pos4k<-
(q10_CR45_MAF550_IScore60_FSS35_assembly$POS.x)/40000

q10_CR45_MAF550_IScore60_FSS35_assembly$gpos<-
ceiling(q10_CR45_MAF550_IScore60_FSS35_assembly$pos4k)

q10_CR45_MAF550_IScore60_FSS35_assembly$coverage <-
if_else((q10_CR45_MAF550_IScore60_FSS35_assembly$gpos*40000) <
q10_CR45_MAF550_IScore60_FSS35_assembly$Sequence_Length, 40000,
q10_CR45_MAF550_IScore60_FSS35_assembly$Sequence_Length-
((q10_CR45_MAF550_IScore60_FSS35_assembly$gpos-1)*40000))

q10_CR45_MAF550_IScore60_FSS35_assembly$$S<-
(40000*(q10_CR45_MAF550_IScore60_FSS35_assembly$gpos-1))+1

q10_CR45_MAF550_IScore60_FSS35_assembly$E<-
40000*(q10_CR45_MAF550_IScore60_FSS35_assembly$gpos)

q10_CR45_MAF550_IScore60_FSS35_assembly$EmS<-
q10_CR45_MAF550_IScore60_FSS35_assembly$E-
q10_CR45_MAF550_IScore60_FSS35_assembly$$S

q10_CR45_MAF550_IScore60_FSS35_assembly$a<-
2*q10_CR45_MAF550_IScore60_FSS35_assembly$POS.x

q10_CR45_MAF550_IScore60_FSS35_assembly$EMS<-
q10_CR45_MAF550_IScore60_FSS35_assembly$E+q10_CR45_MAF550_IScore60_FSS3
5_assembly$$S

q10_CR45_MAF550_IScore60_FSS35_assembly$Score<-
q10_CR45_MAF550_IScore60_FSS35_assembly$EmS-
(abs(q10_CR45_MAF550_IScore60_FSS35_assembly$a-
q10_CR45_MAF550_IScore60_FSS35_assembly$EMS))

```

```

q10_CR45_MAF550_IScore60_FSS35_assembly$FSS<-"35pb"
q10_CR45_MAF550_IScore60_FSS35_assembly$Round<-2
q10_CR45_MAF550_IScore60_FSS35_assembly$f<-
ceiling((q10_CR45_MAF550_IScore60_FSS35_assembly$pos4k-
trunc(q10_CR45_MAF550_IScore60_FSS35_assembly$pos4k))*5)
q10_CR45_MAF550_IScore60_FSS35_assembly$find<-
if_else(q10_CR45_MAF550_IScore60_FSS35_assembly$f<=3,q10_CR45_MAF550_IScor
e60_FSS35_assembly$f+2,q10_CR45_MAF550_IScore60_FSS35_assembly$f-2)
write.csv(q10_CR45_MAF550_IScore60_FSS35_assembly[,c("X.CHROM","POS.x","Assi
gned_Molecule","Assigned_Molecule_Location_Type","Sequence_Length","pos4k","gpos",
"coverage","S","E","EmS","aa","EMS","Score","FSS","Round","f","find")],q10_CR45_M
AF550_IScore60_FSS35_assembly_coverage.csv",row.names=FALSE)

```

```

q10_CR45_MAF550_IScore60_FSS35_assembly2<-
q10_CR45_MAF550_IScore60_FSS35_assembly[order(q10_CR45_MAF550_IScore60_FS
S35_assembly$X.CHROM,q10_CR45_MAF550_IScore60_FSS35_assembly$gpos,q10_CR
45_MAF550_IScore60_FSS35_assembly$Score,decreasing = TRUE),]
write.csv(q10_CR45_MAF550_IScore60_FSS35_assembly2[,c("X.CHROM","POS.x","Ass
igned_Molecule","Assigned_Molecule_Location_Type","Sequence_Length","pos4k","gpos",
"coverage","S","E","EmS","aa","EMS","Score","FSS","Round","f","find")],q10_CR45_
MAF550_IScore60_FSS35_assembly2_coverage.csv",row.names=FALSE)

```

```

q10_CR45_MAF550_IScore60_FSS35_assembly3<-
q10_CR45_MAF550_IScore60_FSS35_assembly2[!duplicated(q10_CR45_MAF550_IScor
e60_FSS35_assembly2[,c("X.CHROM","gpos")]), ]
write.csv(q10_CR45_MAF550_IScore60_FSS35_assembly3[,c("X.CHROM","POS.x","Ass
igned_Molecule","Assigned_Molecule_Location_Type","Sequence_Length","pos4k","gpos",
"coverage","S","E","EmS","aa","EMS","Score","FSS","Round","f","find")],q10_CR45_
MAF550_IScore60_FSS35_assembly3_coverage.csv",row.names=FALSE)

```

# Ronda 3

```

filter_CR15_MAF450_IScore60_FSS40<-
subset(q10_KGD_Illumina_FSS40,q10_KGD_Illumina_FSS40$F_MISS<0.85 &

```

```

q10_KGD_Illumina_FSS40$MAF>=0.04 & q10_KGD_Illumina_FSS40$MAF<=0.5 &
q10_KGD_Illumina_FSS40$Final_Score>=0.6)

filter2_CR15_MAF450_IScore60_FSS40<-
distinct(filter_CR15_MAF450_IScore60_FSS40, X.CHROM, POS.x, .keep_all= TRUE)

table_CR15_MAF450_IScore60_FSS40<-
filter2_CR15_MAF450_IScore60_FSS40[,c(2:3,10,14)]

q10_CR15_MAF450_IScore60_FSS40_assembly<-
merge(x=table_CR15_MAF450_IScore60_FSS40,y=assembly,by.x="X.CHROM",by.y="G
enBank_Accn")

write.csv(q10_CR15_MAF450_IScore60_FSS40_assembly[,c("X.CHROM","POS.x","Assi
gned_Molecule","Assigned_Molecule_Location_Type","Sequence_Length")], "q10_CR15_
MAF450_IScore60_FSS40_assembly.csv",row.names=FALSE)

q10_CR15_MAF450_IScore60_FSS40_assembly$pos4k<-
(q10_CR15_MAF450_IScore60_FSS40_assembly$POS.x)/40000

q10_CR15_MAF450_IScore60_FSS40_assembly$gpos<-
ceiling(q10_CR15_MAF450_IScore60_FSS40_assembly$pos4k)

q10_CR15_MAF450_IScore60_FSS40_assembly$coverage <-
if_else((q10_CR15_MAF450_IScore60_FSS40_assembly$gpos*40000) <
q10_CR15_MAF450_IScore60_FSS40_assembly$Sequence_Length, 40000,
q10_CR15_MAF450_IScore60_FSS40_assembly$Sequence_Length-
((q10_CR15_MAF450_IScore60_FSS40_assembly$gpos-1)*40000))

q10_CR15_MAF450_IScore60_FSS40_assembly$$<-
(40000*(q10_CR15_MAF450_IScore60_FSS40_assembly$gpos-1))+1

q10_CR15_MAF450_IScore60_FSS40_assembly$E<-
40000*(q10_CR15_MAF450_IScore60_FSS40_assembly$gpos)

q10_CR15_MAF450_IScore60_FSS40_assembly$EmS<-
q10_CR15_MAF450_IScore60_FSS40_assembly$E-
q10_CR15_MAF450_IScore60_FSS40_assembly$$

q10_CR15_MAF450_IScore60_FSS40_assembly$a<-
2*q10_CR15_MAF450_IScore60_FSS40_assembly$POS.x

q10_CR15_MAF450_IScore60_FSS40_assembly$EMS<-
q10_CR15_MAF450_IScore60_FSS40_assembly$E+q10_CR15_MAF450_IScore60_FSS4
0_assembly$$

```

```

q10_CR15_MAF450_IScore60_FSS40_assembly$Score<-
q10_CR15_MAF450_IScore60_FSS40_assembly$EmS-
(abs(q10_CR15_MAF450_IScore60_FSS40_assembly$aa-
q10_CR15_MAF450_IScore60_FSS40_assembly$EMS))

q10_CR15_MAF450_IScore60_FSS40_assembly$FSS<-"40pb"

q10_CR15_MAF450_IScore60_FSS40_assembly$Round<-3

q10_CR15_MAF450_IScore60_FSS40_assembly$f<-
ceiling((q10_CR15_MAF450_IScore60_FSS40_assembly$pos4k-
trunc(q10_CR15_MAF450_IScore60_FSS40_assembly$pos4k))*5)

q10_CR15_MAF450_IScore60_FSS40_assembly$find<-
if_else(q10_CR15_MAF450_IScore60_FSS40_assembly$f<=3,q10_CR15_MAF450_IScor
e60_FSS40_assembly$f+2,q10_CR15_MAF450_IScore60_FSS40_assembly$f-2)

write.csv(q10_CR15_MAF450_IScore60_FSS40_assembly[,c("X.CHROM","POS.x","Assi
gned_Molecule","Assigned_Molecule_Location_Type","Sequence_Length","pos4k","gpos",
"coverage","S","E","EmS","aa","EMS","Score","FSS","Round","f","find")],q10_CR15_M
AF450_IScore60_FSS40_assembly_coverage.csv",row.names=FALSE)

q10_CR15_MAF450_IScore60_FSS40_assembly2<-
q10_CR15_MAF450_IScore60_FSS40_assembly[order(q10_CR15_MAF450_IScore60_FS
S40_assembly$X.CHROM,q10_CR15_MAF450_IScore60_FSS40_assembly$gpos,q10_CR
15_MAF450_IScore60_FSS40_assembly$Score,decreasing = TRUE),]

write.csv(q10_CR15_MAF450_IScore60_FSS40_assembly2[,c("X.CHROM","POS.x","Ass
igned_Molecule","Assigned_Molecule_Location_Type","Sequence_Length","pos4k","gpos
","coverage","S","E","EmS","aa","EMS","Score","FSS","Round","f","find")],q10_CR15_
MAF450_IScore60_FSS40_assembly2_coverage.csv",row.names=FALSE)

q10_CR15_MAF450_IScore60_FSS40_assembly3<-
q10_CR15_MAF450_IScore60_FSS40_assembly2[!duplicated(q10_CR15_MAF450_IScor
e60_FSS40_assembly2[,c("X.CHROM","gpos")]), ]

write.csv(q10_CR15_MAF450_IScore60_FSS40_assembly3[,c("X.CHROM","POS.x","Ass
igned_Molecule","Assigned_Molecule_Location_Type","Sequence_Length","pos4k","gpos
","coverage","S","E","EmS","aa","EMS","Score","FSS","Round","f","find")],q10_CR15_
MAF450_IScore60_FSS40_assembly3_coverage.csv",row.names=FALSE)

```

# Ronda 4

```
filter_CR15_MAF450_IScore60_FSS35<-
subset(q10_KGD_Illumina_FSS35,q10_KGD_Illumina_FSS35$F_MISS<0.85 &
q10_KGD_Illumina_FSS35$MAF>=0.04 & q10_KGD_Illumina_FSS35$MAF<=0.5 &
q10_KGD_Illumina_FSS35$Final_Score>=0.6)

filter2_CR15_MAF450_IScore60_FSS35<-
distinct(filter_CR15_MAF450_IScore60_FSS35, X.CHROM, POS.x, .keep_all= TRUE)

table_CR15_MAF450_IScore60_FSS35<-
filter2_CR15_MAF450_IScore60_FSS35[,c(2:3,10,14)]

q10_CR15_MAF450_IScore60_FSS35_assembly<-
merge(x=table_CR15_MAF450_IScore60_FSS35,y=assembly,by.x="X.CHROM",by.y="G
enBank_Accn")

write.csv(q10_CR15_MAF450_IScore60_FSS35_assembly[,c("X.CHROM","POS.x","Assi
gned_Molecule","Assigned_Molecule_Location_Type","Sequence_Length")], "q10_CR15_
MAF450_IScore60_FSS35_assembly.csv",row.names=FALSE)

q10_CR15_MAF450_IScore60_FSS35_assembly$pos4k<-
(q10_CR15_MAF450_IScore60_FSS35_assembly$POS.x)/40000

q10_CR15_MAF450_IScore60_FSS35_assembly$gpos<-
ceiling(q10_CR15_MAF450_IScore60_FSS35_assembly$pos4k)

q10_CR15_MAF450_IScore60_FSS35_assembly$coverage <-
if_else((q10_CR15_MAF450_IScore60_FSS35_assembly$gpos*40000) <
q10_CR15_MAF450_IScore60_FSS35_assembly$Sequence_Length, 40000,
q10_CR15_MAF450_IScore60_FSS35_assembly$Sequence_Length-
((q10_CR15_MAF450_IScore60_FSS35_assembly$gpos-1)*40000))

q10_CR15_MAF450_IScore60_FSS35_assembly$$S<-
(40000*(q10_CR15_MAF450_IScore60_FSS35_assembly$gpos-1))+1

q10_CR15_MAF450_IScore60_FSS35_assembly$E<-
40000*(q10_CR15_MAF450_IScore60_FSS35_assembly$gpos)

q10_CR15_MAF450_IScore60_FSS35_assembly$EmS<-
q10_CR15_MAF450_IScore60_FSS35_assembly$E-
q10_CR15_MAF450_IScore60_FSS35_assembly$$S

q10_CR15_MAF450_IScore60_FSS35_assembly$aa<-
2*q10_CR15_MAF450_IScore60_FSS35_assembly$POS.x
```



```
q10_CR15_MAF450_IScore60_FSS35_assembly$EMS<-
q10_CR15_MAF450_IScore60_FSS35_assembly$E+q10_CR15_MAF450_IScore60_FSS35_assembly$S
```

```
q10_CR15_MAF450_IScore60_FSS35_assembly$Score<-
q10_CR15_MAF450_IScore60_FSS35_assembly$EmS-
(abs(q10_CR15_MAF450_IScore60_FSS35_assembly$aa-
q10_CR15_MAF450_IScore60_FSS35_assembly$EMS))
```

```
q10_CR15_MAF450_IScore60_FSS35_assembly$FSS<-"35pb"
```

```
q10_CR15_MAF450_IScore60_FSS35_assembly$Round<-4
```

```
q10_CR15_MAF450_IScore60_FSS35_assembly$f<-
ceiling((q10_CR15_MAF450_IScore60_FSS35_assembly$pos4k-
trunc(q10_CR15_MAF450_IScore60_FSS35_assembly$pos4k))*5)
```

```
q10_CR15_MAF450_IScore60_FSS35_assembly$find<-
if_else(q10_CR15_MAF450_IScore60_FSS35_assembly$f<=3,q10_CR15_MAF450_IScore60_FSS35_assembly$f+2,q10_CR15_MAF450_IScore60_FSS35_assembly$f-2)
```

```
write.csv(q10_CR15_MAF450_IScore60_FSS35_assembly[,c("X.CHROM","POS.x","Assigned_Molecule","Assigned_Molecule_Location_Type","Sequence_Length","pos4k","gpos","coverage","S","E","EmS","aa","EMS","Score","FSS","Round","f","find")],q10_CR15_MAF450_IScore60_FSS35_assembly_coverage.csv",row.names=FALSE)
```

```
q10_CR15_MAF450_IScore60_FSS35_assembly2<-
q10_CR15_MAF450_IScore60_FSS35_assembly[order(q10_CR15_MAF450_IScore60_FSS35_assembly$X.CHROM,q10_CR15_MAF450_IScore60_FSS35_assembly$gpos,q10_CR15_MAF450_IScore60_FSS35_assembly$Score,decreasing = TRUE),]
```

```
write.csv(q10_CR15_MAF450_IScore60_FSS35_assembly2[,c("X.CHROM","POS.x","Assigned_Molecule","Assigned_Molecule_Location_Type","Sequence_Length","pos4k","gpos","coverage","S","E","EmS","aa","EMS","Score","FSS","Round","f","find")],q10_CR15_MAF450_IScore60_FSS35_assembly2_coverage.csv",row.names=FALSE)
```

```
q10_CR15_MAF450_IScore60_FSS35_assembly3<-
q10_CR15_MAF450_IScore60_FSS35_assembly2[!duplicated(q10_CR15_MAF450_IScore60_FSS35_assembly2[c("X.CHROM","gpos")]), ]
```

```
write.csv(q10_CR15_MAF450_IScore60_FSS35_assembly3[,c("X.CHROM","POS.x","Assigned_Molecule","Assigned_Molecule_Location_Type","Sequence_Length","pos4k","gpos","coverage","S","E","EmS","aa","EMS","Score","FSS","Round","f","find")],q10_CR15_MAF450_IScore60_FSS35_assembly3_coverage.csv",row.names=FALSE)
```

#Round 5

```
filter_CR15_MAF150_IScore60_FSS40<-
subset(q10_KGD_Illumina_FSS40,q10_KGD_Illumina_FSS40$F_MISS<0.85 &
q10_KGD_Illumina_FSS40$MAF>=0.01 & q10_KGD_Illumina_FSS40$MAF<=0.5 &
q10_KGD_Illumina_FSS40$Final_Score>=0.6)

filter2_CR15_MAF150_IScore60_FSS40<-
distinct(filter_CR15_MAF150_IScore60_FSS40, X.CHROM, POS.x, .keep_all= TRUE)

table_CR15_MAF150_IScore60_FSS40<-
filter2_CR15_MAF150_IScore60_FSS40[,c(2:3,10,14)]

q10_CR15_MAF150_IScore60_FSS40_assembly<-
merge(x=table_CR15_MAF150_IScore60_FSS40,y=assembly,by.x="X.CHROM",by.y="GenBank_Accn")

write.csv(q10_CR15_MAF150_IScore60_FSS40_assembly[,c("X.CHROM","POS.x","Assigned_Molecule",
"Assigned_Molecule_Location_Type","Sequence_Length")], "q10_CR15_MAF150_IScore60_FSS40_assembly.csv",row.names=FALSE)

q10_CR15_MAF150_IScore60_FSS40_assembly$pos4k<-
(q10_CR15_MAF150_IScore60_FSS40_assembly$POS.x)/40000

q10_CR15_MAF150_IScore60_FSS40_assembly$gpos<-
ceiling(q10_CR15_MAF150_IScore60_FSS40_assembly$pos4k)

q10_CR15_MAF150_IScore60_FSS40_assembly$coverage <-
if_else((q10_CR15_MAF150_IScore60_FSS40_assembly$gpos*40000) <
q10_CR15_MAF150_IScore60_FSS40_assembly$Sequence_Length, 40000,
q10_CR15_MAF150_IScore60_FSS40_assembly$Sequence_Length-
((q10_CR15_MAF150_IScore60_FSS40_assembly$gpos-1)*40000))

q10_CR15_MAF150_IScore60_FSS40_assembly$$S<-
(40000*(q10_CR15_MAF150_IScore60_FSS40_assembly$gpos-1))+1

q10_CR15_MAF150_IScore60_FSS40_assembly$E<-
40000*(q10_CR15_MAF150_IScore60_FSS40_assembly$gpos)

q10_CR15_MAF150_IScore60_FSS40_assembly$EmS<-
q10_CR15_MAF150_IScore60_FSS40_assembly$E-
q10_CR15_MAF150_IScore60_FSS40_assembly$$S

q10_CR15_MAF150_IScore60_FSS40_assembly$aa<-
2*q10_CR15_MAF150_IScore60_FSS40_assembly$POS.x
```

```
q10_CR15_MAF150_IScore60_FSS40_assembly$EMS<-
q10_CR15_MAF150_IScore60_FSS40_assembly$E+q10_CR15_MAF150_IScore60_FSS40_assembly$S
```

```
q10_CR15_MAF150_IScore60_FSS40_assembly$Score<-
q10_CR15_MAF150_IScore60_FSS40_assembly$EmS-
(abs(q10_CR15_MAF150_IScore60_FSS40_assembly$aa-
q10_CR15_MAF150_IScore60_FSS40_assembly$EMS))
```

```
q10_CR15_MAF150_IScore60_FSS40_assembly$FSS<-"40pb"
```

```
q10_CR15_MAF150_IScore60_FSS40_assembly$Round<-5
```

```
q10_CR15_MAF150_IScore60_FSS40_assembly$f<-
ceiling((q10_CR15_MAF150_IScore60_FSS40_assembly$pos4k-
trunc(q10_CR15_MAF150_IScore60_FSS40_assembly$pos4k))*5)
```

```
q10_CR15_MAF150_IScore60_FSS40_assembly$find<-
if_else(q10_CR15_MAF150_IScore60_FSS40_assembly$f<=3,q10_CR15_MAF150_IScore60_FSS40_assembly$f+2,q10_CR15_MAF150_IScore60_FSS40_assembly$f-2)
```

```
write.csv(q10_CR15_MAF150_IScore60_FSS40_assembly[,c("X.CHROM","POS.x","Assigned_Molecule","Assigned_Molecule_Location_Type","Sequence_Length","pos4k","gpos","coverage","S","E","EmS","aa","EMS","Score","FSS","Round","f","find")],q10_CR15_MAF150_IScore60_FSS40_assembly_coverage.csv",row.names=FALSE)
```

```
q10_CR15_MAF150_IScore60_FSS40_assembly2<-
q10_CR15_MAF150_IScore60_FSS40_assembly[order(q10_CR15_MAF150_IScore60_FSS40_assembly$X.CHROM,q10_CR15_MAF150_IScore60_FSS40_assembly$gpos,q10_CR15_MAF150_IScore60_FSS40_assembly$Score,decreasing = TRUE),]
```

```
write.csv(q10_CR15_MAF150_IScore60_FSS40_assembly2[,c("X.CHROM","POS.x","Assigned_Molecule","Assigned_Molecule_Location_Type","Sequence_Length","pos4k","gpos","coverage","S","E","EmS","aa","EMS","Score","FSS","Round","f","find")],q10_CR15_MAF150_IScore60_FSS40_assembly2_coverage.csv",row.names=FALSE)
```

```
q10_CR15_MAF150_IScore60_FSS40_assembly3<-
q10_CR15_MAF150_IScore60_FSS40_assembly2[!duplicated(q10_CR15_MAF150_IScore60_FSS40_assembly2[,c("X.CHROM","gpos")]), ]
```

```
write.csv(q10_CR15_MAF150_IScore60_FSS40_assembly3[,c("X.CHROM","POS.x","Assigned_Molecule","Assigned_Molecule_Location_Type","Sequence_Length","pos4k","gpos","coverage","S","E","EmS","aa","EMS","Score","FSS","Round","f","find")],q10_CR15_MAF150_IScore60_FSS40_assembly3_coverage.csv",row.names=FALSE)
```

#Ronda 6

```
filter_CR15_MAF150_IScore60_FSS35<-
subset(q10_KGD_Illumina_FSS35,q10_KGD_Illumina_FSS35$F_MISS<0.85 &
q10_KGD_Illumina_FSS35$MAF>=0.01 & q10_KGD_Illumina_FSS35$MAF<=0.5 &
q10_KGD_Illumina_FSS35$Final_Score>=0.6)

filter2_CR15_MAF150_IScore60_FSS35<-
distinct(filter_CR15_MAF150_IScore60_FSS35, X.CHROM, POS.x, .keep_all= TRUE)

table_CR15_MAF150_IScore60_FSS35<-
filter2_CR15_MAF150_IScore60_FSS35[,c(2:3,10,14)]

q10_CR15_MAF150_IScore60_FSS35_assembly<-
merge(x=table_CR15_MAF150_IScore60_FSS35,y=assembly,by.x="X.CHROM",by.y="GenBank_Accn")

write.csv(q10_CR15_MAF150_IScore60_FSS35_assembly[,c("X.CHROM","POS.x","Assigned_Molecule",
"Assigned_Molecule_Location_Type","Sequence_Length")], "q10_CR15_MAF150_IScore60_FSS35_assembly.csv",row.names=FALSE)

q10_CR15_MAF150_IScore60_FSS35_assembly$pos4k<-
(q10_CR15_MAF150_IScore60_FSS35_assembly$POS.x)/40000

q10_CR15_MAF150_IScore60_FSS35_assembly$gpos<-
ceiling(q10_CR15_MAF150_IScore60_FSS35_assembly$pos4k)

q10_CR15_MAF150_IScore60_FSS35_assembly$coverage <-
if_else((q10_CR15_MAF150_IScore60_FSS35_assembly$gpos*40000) <
q10_CR15_MAF150_IScore60_FSS35_assembly$Sequence_Length, 40000,
q10_CR15_MAF150_IScore60_FSS35_assembly$Sequence_Length-
((q10_CR15_MAF150_IScore60_FSS35_assembly$gpos-1)*40000))

q10_CR15_MAF150_IScore60_FSS35_assembly$$S<-
(40000*(q10_CR15_MAF150_IScore60_FSS35_assembly$gpos-1))+1

q10_CR15_MAF150_IScore60_FSS35_assembly$E<-
40000*(q10_CR15_MAF150_IScore60_FSS35_assembly$gpos)

q10_CR15_MAF150_IScore60_FSS35_assembly$EmS<-
q10_CR15_MAF150_IScore60_FSS35_assembly$E-
q10_CR15_MAF150_IScore60_FSS35_assembly$$S

q10_CR15_MAF150_IScore60_FSS35_assembly$aa<-
2*q10_CR15_MAF150_IScore60_FSS35_assembly$POS.x
```

```
q10_CR15_MAF150_IScore60_FSS35_assembly$EMS<-
q10_CR15_MAF150_IScore60_FSS35_assembly$E+q10_CR15_MAF150_IScore60_FSS35_assembly$S
```

```
q10_CR15_MAF150_IScore60_FSS35_assembly$Score<-
q10_CR15_MAF150_IScore60_FSS35_assembly$EmS-
(abs(q10_CR15_MAF150_IScore60_FSS35_assembly$aa-
q10_CR15_MAF150_IScore60_FSS35_assembly$EMS))
```

```
q10_CR15_MAF150_IScore60_FSS35_assembly$FSS<-"35pb"
```

```
q10_CR15_MAF150_IScore60_FSS35_assembly$Round<-6
```

```
q10_CR15_MAF150_IScore60_FSS35_assembly$f<-
ceiling((q10_CR15_MAF150_IScore60_FSS35_assembly$pos4k-
trunc(q10_CR15_MAF150_IScore60_FSS35_assembly$pos4k))*5)
```

```
q10_CR15_MAF150_IScore60_FSS35_assembly$find<-
if_else(q10_CR15_MAF150_IScore60_FSS35_assembly$f<=3,q10_CR15_MAF150_IScore60_FSS35_assembly$f+2,q10_CR15_MAF150_IScore60_FSS35_assembly$f-2)
```

```
write.csv(q10_CR15_MAF150_IScore60_FSS35_assembly[,c("X.CHROM","POS.x","Assigned_Molecule","Assigned_Molecule_Location_Type","Sequence_Length","pos4k","gpos","coverage","S","E","EmS","aa","EMS","Score","FSS","Round","f","find")],q10_CR15_MAF150_IScore60_FSS35_assembly_coverage.csv",row.names=FALSE)
```

```
q10_CR15_MAF150_IScore60_FSS35_assembly2<-
```

```
q10_CR15_MAF150_IScore60_FSS35_assembly[order(q10_CR15_MAF150_IScore60_FSS35_assembly$X.CHROM,q10_CR15_MAF150_IScore60_FSS35_assembly$gpos,q10_CR15_MAF150_IScore60_FSS35_assembly$Score,decreasing = TRUE),]
```

```
write.csv(q10_CR15_MAF150_IScore60_FSS35_assembly2[,c("X.CHROM","POS.x","Assigned_Molecule","Assigned_Molecule_Location_Type","Sequence_Length","pos4k","gpos","coverage","S","E","EmS","aa","EMS","Score","FSS","Round","f","find")],q10_CR15_MAF150_IScore60_FSS35_assembly2_coverage.csv",row.names=FALSE)
```

```
q10_CR15_MAF150_IScore60_FSS35_assembly3<-
```

```
q10_CR15_MAF150_IScore60_FSS35_assembly2[!duplicated(q10_CR15_MAF150_IScore60_FSS35_assembly2[c("X.CHROM","gpos")]), ]
```

```
write.csv(q10_CR15_MAF150_IScore60_FSS35_assembly3[,c("X.CHROM","POS.x","Assigned_Molecule","Assigned_Molecule_Location_Type","Sequence_Length","pos4k","gpos","coverage","S","E","EmS","aa","EMS","Score","FSS","Round","f","find")],q10_CR15_MAF150_IScore60_FSS35_assembly3_coverage.csv",row.names=FALSE)
```

```
#Reportes
```

```
# Primer conjunto de PNSs
```

```
pR1<-  
read.csv("q10_CR45_MAF550_IScore60_FSS40_assembly3_coverage.csv",header=TRUE)
```

```
pR2<-  
read.csv("q10_CR45_MAF550_IScore60_FSS35_assembly3_coverage.csv",header=TRUE)
```

```
pR3<-  
read.csv("q10_CR15_MAF450_IScore60_FSS40_assembly3_coverage.csv",header=TRUE)
```

```
pR4<-  
read.csv("q10_CR15_MAF450_IScore60_FSS35_assembly3_coverage.csv",header=TRUE)
```

```
pR5<-  
read.csv("q10_CR15_MAF150_IScore60_FSS40_assembly3_coverage.csv",header=TRUE)
```

```
pR6<-  
read.csv("q10_CR15_MAF150_IScore60_FSS35_assembly3_coverage.csv",header=TRUE)
```

```
pR<-rbind(pR1,pR2,pR3,pR4,pR5,pR6)
```

```
write.csv(pR,"pR.csv",row.names=FALSE)
```

```
FirstSet<-pR[!duplicated(pR[c("X.CHROM","gpos")]), ]
```

```
write.csv(FirstSet[,c("X.CHROM","POS.x","Assigned_Molecule","Assigned_Molecule_Location_Type","Sequence_Length","pos4k","gpos","coverage","S","E","EmS","aa","EMS","Score","FSS","Round","f","find")],"FirstSet.csv",row.names=FALSE)
```

```
# Segundo conjunto de PNSs
```

```
sR1<-  
read.csv("q10_CR45_MAF550_IScore60_FSS40_assembly2_coverage.csv",header=TRUE)
```

```
sR2<-  
read.csv("q10_CR45_MAF550_IScore60_FSS35_assembly2_coverage.csv",header=TRUE)
```

```

sR3<-
read.csv("q10_CR15_MAF450_IScore60_FSS40_assembly2_coverage.csv",header=TRUE)

sR4<-
read.csv("q10_CR15_MAF450_IScore60_FSS35_assembly2_coverage.csv",header=TRUE)

sR5<-
read.csv("q10_CR15_MAF150_IScore60_FSS40_assembly2_coverage.csv",header=TRUE)

sR6<-
read.csv("q10_CR15_MAF150_IScore60_FSS35_assembly2_coverage.csv",header=TRUE)

sR<-rbind(sR1,sR2,sR3,sR4,sR5,sR6)

write.csv(sR,"sR.csv",row.names=FALSE)

ssRa<-sR[!duplicated(sR[c("X.CHROM","POS.x")]), ]

write.csv(ssRa[,c("X.CHROM","POS.x","Assigned_Molecule","Assigned_Molecule_Location_Type","Sequence_Length","pos4k","gpos","coverage","S","E","EmS","aa","EMS","Score","FSS","Round","f","find")],"ssRa.csv",row.names=FALSE)

List<-FirstSet[,c("X.CHROM","gpos","find")]

ssRb<-
merge(x=List,y=ssRa,by.x=c("X.CHROM","gpos","find"),by.y=c("X.CHROM","gpos","f"))

write.csv(ssRb[,c("X.CHROM","POS.x","Assigned_Molecule","Assigned_Molecule_Location_Type","Sequence_Length","pos4k","gpos","coverage","S","E","EmS","aa","EMS","Score","FSS","Round","find","find.y")],"ssRb.csv",row.names=FALSE)

ssRc<-ssRb[order(ssRb$Round,ssRb$X.CHROM,ssRb$gpos,ssRb$Score),]

write.csv(ssRc[,c("X.CHROM","POS.x","Assigned_Molecule","Assigned_Molecule_Location_Type","Sequence_Length","pos4k","gpos","coverage","S","E","EmS","aa","EMS","Score","FSS","Round","find","find.y")],"ssRc.csv",row.names=FALSE)

SecondSet<-ssRc[!duplicated(ssRc[c("X.CHROM","gpos","find")]), ]

write.csv(SecondSet,"SecondSet.csv",row.names=FALSE)

```

## ANEXO 2: R scripts para el mapa de calor de las relaciones genómicas

```
data<-read.csv("relationship.csv",header=TRUE,sep=";")
```

```
library("ggplot2")
```

```
ggp<-ggplot(v2a, aes(x = id1, y = id2, fill = rel)) +  
  geom_tile() +  
  scale_fill_gradientn(colours=rainbow(5),breaks =  
c(0.00,0.10,0.20,0.30,0.40,0.50,0.60,0.70,0.80,0.90,1.00)) +  
  labs(fill = "") +  
  theme(axis.title.x=element_blank(),  
        axis.title.y=element_blank(),  
        axis.text.x=element_blank(),  
        axis.text.y=element_blank(),  
        axis.ticks.x=element_blank(),  
        axis.ticks.y=element_blank())
```

```
ggp
```

```
pdf('heatmap.pdf')
```

```
ggp
```

```
dev.off()
```



**ANEXO 3:** Lista de PNS ubicados dentro o cercano a las secuencias codificantes de genes candidatos para la calidad y el color de la fibra que también se incluyen en la micromatriz.

N°	ID UNALM	ID Affymetrix	Cromosoma	Número Acceso Scaffold (VicPac3.1)	Posición Scaffold (VicPac3.1)	Secuencia del PNS	Alelo VicPac 3.1	Alelo Alternativo	FAM	Nombre Gen VicPac 3.1	Descripción Gen VicPac3.1	Referencia
1	UNALM_VP A_2786320	AX-417368864	12	ABRR03077 305.1	1883387	CTGAGTGCTCCCTTGCAGCTGTGAG CACCAGCCACACCAGCGTCAGTGGA AGCAGTAGCC[G/A]TGGAGGCGGTG GTTATAGCtctggcggcgggcggtacAGC TACAGCTCCGGGGGCGG	G	A	0.311	KRT1	keratin 1	Fernández et al., 2019
2	UNALM_VP A_4284507	AX-417369384	16	ABRR03077 387.1	640265	TTCCCTCTCATTCCAGAAAGCATCAT GCAATatcctctcctttgtttttacagTGAC[T/A]CTTTGGATGAAAGTTCATACGTGAT GACAGCCTCCAAATCTCAAGCACCA TCAACTGATT	T	A	0.253	LOC107034533 KRT12	uncharacterized LOC107034533, keratin 12	Fernández et al., 2019
3	UNALM_VP A_4284441	AX-417368856	12	ABRR03077 305.1	1880504	GGAGGATTTCCGGAACAAGTAAGG ATAAGTGGCTGGGTAGCTCTTGACT TTCGTCTAAA[G/A]TATTTTGTGAA AATGGCAAGCTAAGACTATGAACAG GATGTAACAGTAACTTTTCTTAGG	G	A	0.49	KRT1	keratin 1	Fernández et al., 2019
4	UNALM_VP A_3426264	AX-417369011	16	ABRR03077 387.1	639822	TACCTTCCTTGATAAAAGATCATAG AATGCTGAGCTGAACAACGGCACCA CAGAGAGGCA[T/C]ACTGCGGCTCA GTACTAGAAAAACCAACCGCTCCAG TTCTGGTATCAGTGCCCCCTCGCTA	T	C	0.247	LOC107034533 KRT12	uncharacterized LOC107034533, keratin 12	Fernández et al., 2019
5	UNALM_VP A_4284466	AX-417276562	16	ABRR03077 387.1	153608	CAACAGGTGCCTCACACCACTGGCA ACTTCAGATTCCCAGACAGGACAGC CTCCAGCCCC[G/A]GACCCCTTCTTC	G	A	0.118	LOC102527828	keratin type I cytoskeletal 14	Fernández et al., 2019

						TAGCTCTGAATGCCCGAGTGGAGC CGATCCATGAGAAGAaggcctcc							
6	UNALM_VP A_4284464	AX- 417272667	16	ABRR03077 387.1	132824	GGCCCTGGGCCACTTAGAATTGATG GCAGAAGGTCCCTTTTCAGCATCAA GGCATCTGAG[C/T]GGGCCTGGTACC CTGAATCTCTTATTCCCTGCAAGCAT TCTTTGCACTCAGCTTCTCTCA	C	T	0.348	LOC102 527556 LOC102 530093	keratin type I cytoskeletal 17, keratin type I cytoskeletal 16	Fernández et al., 2019	
7	UNALM_VP A_4284533	AX- 417324032	16	ABRR03077 387.1	284447	gatatttcttacattttataataCCGTGTTGTGTAG AGGGTATATTTGCACATGCTACCT[C /T]ATCTGGTCTTGCAATTACCACCAG GTAGGTAGGATCAGTATTCTCCTGG GTCACCAGCTA	C	T	0.128	LOC102 525174	keratin type I cuticular Ha1	Fernández et al., 2019	
8	UNALM_VP A_2788475	AX- 417368896	12	ABRR03077 305.1	2011690	TGGCCTCACATTTGATAAAGTGCAG CTTCTCCTCCACTACAGGGTATAAA ACACATGCAC[G/A]GGGGGCCACTA TCTTGTGCATGATGGTTATGCATCAT TAATCCAATGATCCAATCCAATC	G	A	0.237	KRT5	keratin 5	Fernández et al., 2019	
9	UNALM_VP A_4284442	AX- 417368858	12	ABRR03077 305.1	1880767	ATTTTAGTTCAGAAAGCTAGATTTT AGGCCTCACCTCCTCTGAATCTCCTA ATCACTTCT[G/A]ATTTCTTGCAGG TATGAGGATGAGATCAATAAGCGGA CAAATGCCGGAGAATGAATTTGT	G	A	0.003	KRT1	keratin 1	Fernández et al., 2019	
10	UNALM_VP A_4284461	AX- 417368974	12	ABRR03077 305.1	2248503	TGCCGCTGGGCCAGAGGTGTGCAG AGAGGAGCTCCCCAGTAGCTGCACT CAGGACCCTT[C/T]CCTGCATCCCCT TCCAGGTCCTCCTCAGCCTGAGCCT GGAGCTCAGCCTGGCCTCCCAG	C	T	0.345	LOC102 545800 KRT7	keratin type II cuticular Hb1- like, keratin 7	Fernández et al., 2019	
11	UNALM_VP A_2792019	AX- 417368978	12	ABRR03077 305.1	2305246	TGGCCATAACAAATGCTACCTAAG GCCTGGGCTAGAAGCACCCCTAGAG AGCCTTGTCT[A/G]GAACACCCcatgcc caccacccccacaacATGGTGACCAAGGA AGCCCCAAGCTGCATC	A	G	0.451	KRT80	keratin 80	Fernández et al., 2019	

12	UNALM_VP A_2792021	AX- 417369366	12	ABRR03077 305.1	2305519	TTTTCTCAGCGATCTCTGTAAAGTTC ACTTGAAGATCAACCAAAACCCTTC ATCAGGAGA[G/A]CCCCAAAGGGC AGCTGGCCATGGAGCTCAGAGCCCT GCTGGCTCCAGGGTGGAGACAGG	G	A	0.439	KRT80	keratin 80	Fernández et al., 2019
13	UNALM_VP A_4284451	AX- 417369326	12	ABRR03077 305.1	2054369	CAGTGGCAGAGCCTTAGGAAGTGGC CTCAGCTCCACCGGGGGCAGCAGTT TCACCATTAA[G/A]TAcactgccacctctct ccagcaggaaGAGCTACAAGCATTGAAG TCTTCCTGTTGGT	G	A	0.01		keratin 6*	Fernández et al., 2019
14	UNALM_VP A_4284452	AX- 417368902	12	ABRR03077 305.1	2103969	TTGAAGTACAGAAATAGATTGCCTG TGGATGACACCAGGGTTCTATGCTA CCATCAGGAA[A/T]ACAGGCTCCTTC TAGCTTTCCATTCTGCCATTCTTAAT GCATTGCTTTAATCTTCATGGT	A	T	0.429	KRT82	keratin 82	Fernández et al., 2019
15	UNALM_VP A_4284457	AX- 417368915	12	ABRR03077 305.1	2177420	TGCATGACCAGAGTCCTTGTTGTG GTTACTCAGGGAGACCCAGGTGATG AAGGGGAGAG[A/G]CCCTGTTCTG GGGTGATCCCAGCTGGCTGGGAGAG ACTGGATACATATAGGAAATGAGC	A	G	0.12	LOC116 282606 KRT83	uncharacterized LOC11628260 6, keratin 83	Fernández et al., 2019
16	UNALM_VP A_4284471	AX- 417288309	16	ABRR03077 387.1	169440	GTTCTTCTGCATAGAGGCAGCTGGG AAAGCTCCTGTACACAAGGGTCTGA TGGATTGCGG[G/A]TTTCTTCCTCCC ATGGGGGGAACCTCAGGGCATTGG GCTTTGGGAAAGGAGGATGGTAA	G	A	0.22	KRT9	keratin 9	Fernández et al., 2019
17	UNALM_VP A_3756988	AX- 417369284	4	ABRR03002 523.1	8493095	TTTCcttggggcagggcaggagagctTTGGTGA AGTGGATTTCTCTCACGAGGGACCA GC[A/G]TTTCTCACGTGGCACAGGTA CCACCTGCTGCAGCTGGAGAGAGAC ATGCAGGTACGCACC	A	G	0.399	TYRP1	tyrosinase related protein 1	Feeley et al., 2016
18	UNALM_VP A_4284431	AX- 417369283	4	ABRR03002 523.1	8491826	GGTTCTGCATGGGACTCCACACACT CAACAGAATCCTAAATCATCTGAGC TGCTTAGAGA[G/A]GCTAAGAAATG	G	A	0.176	TYRP1	tyrosinase related protein 1	Feeley et al., 2016

						TATTTGATGTAAAGAGTTGAAATTG ATGCTCCAGTCTTCTGACTAATAA						
19	UNALM_VP A_4284416	AX- 417368777	na	ABRR03000 014.1	137832	CTTGTGGAGCTGGGCGATGCCCCGG GCATGCTGGCACGCCCCGGGCCAGCA TGTGGACATA[T/C]AGCACCGCCATG AGCGCCAGCATGGCTACAAAAAAG CTGACGAGACAGAGGAGGACGGCT	T	C	0.091	LOC102 529783	tubulin beta-3 chain	Chandramo han <i>et al.</i> 2015
20	UNALM_VP A_4284526	AX- 417368770	19	ABRR03000 009.1	11522586	CGGCCTGGTTGTAAATCACTCCTGC CTCCTTTGCCAGGAATTCCTAGGTC AACTTTAGAC[C/G]CTGAGTTGTGAC TTTGGTTTCAACTCAGTGAGTGACA CGTCTTGGCCATTAACCTGCCTC	C	G	0.037		agouti signaling protein*	
21	UNALM_VP A_185691	AX- 417368773	19	ABRR03000 009.1	21510575	ACCCACAATAGAGACAGAAGGGAA ATCTAACAGGTTCTTGGAGGAGTTG CTTCTCAGGCT[T/C]CCTTCATCCCTG GGCTTTTCGTCAGGTGCTAGGTGGC TGCAGGCAGTGAGGAAGCACAGG	T	C	0.19	ASIP	agouti signaling protein	
22	UNALM_VP A_559752	AX- 417369264	2	ABRR03000 026.1	38135351	GCTTTGCCGCGCTCGGTGCACTTGG GCGAGAGCAGGAACGTGGAACAGA GCTCCGGTCCC[G/A]GCGCAGCCACC GCGATGAGAGGCGCTCGCGGCGCCT GGGATTTCTCTTCGTCCTGCTTC	G	A	0.287	KIT	KIT proto-oncogene receptor tyrosine kinase	
23	UNALM_VP A_4284417	AX- 417368781	2	ABRR03000 026.1	38167617	CCCAGGATTTGTCAAATGGACTTTT GAGACCCTGGGTGAGCTGAGTGAGA ACAGACATGC[G/A]GAATGGATCAT GGAGAAAGCAGAGGCCACGAATAC GGGCAATTACACGTGCACCAACAAA	G	A	0.341	KIT	KIT proto-oncogene receptor tyrosine kinase	
24	UNALM_VP A_4284418	AX- 417368784	2	ABRR03000 026.1	38167701	GGCCACGAATACGGGCAATTACACG TGCACCAACAAAGGCGGCTTCAGCA CTTCCACCTA[T/C]GTGTTTGTGAGA GGTAAATGCTTGGCTTTCTGTGATG ATCTGCTTGGAGAACTTTCTGTGTC	T	C	0.341	KIT	KIT proto-oncogene receptor tyrosine kinase	

25	UNALM_VP A_4284419	AX- 417368786	2	ABRR03000 026.1	38171009	CCTGCCCTTGTACGGGAAAGAGGAC AGCGACACGCTGGTCCGCTGTCCTC TGACGGACCC[G/A]GAGGTGACCCA TTACTCCCTGACCGGGTGCAGAGGGG AAACCTCTTCCCAAGGACCTGACG	G	A	0.318	KIT	KIT proto-oncogene receptor tyrosine kinase
26	UNALM_VP A_4284420	AX- 417369268	2	ABRR03000 026.1	38171120	GGACCTGACGTTTGTGCGCTGACCCC AAGGCTGGCATCACCATCAGAAACG TGAAGCGCGA[G/A]TATCATCGGCTT TGCTGCACTGCTCCGCGAACCAAG GGGGCAAGTCCGTGCTGTGCGCAG	G	A	0.105	KIT	KIT proto-oncogene receptor tyrosine kinase
27	UNALM_VP A_4284421	AX- 417368790	2	ABRR03000 026.1	38180967	GTACGAGGCATATCCCAAGCCTGAA CACCGACAGTGGGTGTATATGAACA GAACCTTCAC[G/A]GATAAGTGGGA AGATTATCCCAAGTCTGAGAACGAA AGTAACATCAGGTGAGAAGAGGGC	G	A	0.186	KIT	KIT proto-oncogene receptor tyrosine kinase
28	UNALM_VP A_4284422	AX- 417368792	2	ABRR03000 026.1	38183016	CACACCCTCCGAGGCTGGCCGTCCC CAGGTGACTGATGAGTTGCCCTTTC TTTGCAAGATA[C/T]GTAAGTGAACCTT CATCTCACAGATTAAGGGGACCG AAGGAGGCACCTACACGTTTCTC	C	T	0.155	KIT	KIT proto-oncogene receptor tyrosine kinase
29	UNALM_VP A_4284423	AX- 417369271	2	ABRR03000 026.1	38197246	CAGCTCCATCGATTACAGTGCGTTC AAGCACAATGGCACAGTGGAAATGTA GGGCTCACAA[C/T]GATGTGGGCAA GAGTTCTGCCTTTTTTAACCTTTCAT TTAAAGGTAACAGCAAaggtata	C	T	0.318	KIT	KIT proto-oncogene receptor tyrosine kinase
30	UNALM_VP A_4284424	AX- 417368796	2	ABRR03000 026.1	38198572	TTTGCTGATTGGTTTTGTGATCGCAG CTGGTATAATGTGTATCATCGTGAT GATTCTTAC[G/A]TACAAATATTTAC AGGTAAGTGTATTTGTTCTCTGCC CTGAACGTAGTAATGACTGAA	G	A	0.02	KIT	KIT proto-oncogene receptor tyrosine kinase
31	UNALM_VP A_4284527	AX- 417368798	2	ABRR03000 026.1	38203317	TCCTGCCATCATGGAAGACGACGAG CTGGCCCTCGACCTGGAAGACTTGC TGAGCTTTTC[T/C]TACCAGGTGGCA	T	C	0.088	KIT	KIT proto-oncogene receptor tyrosine kinase

						AAGGGCATGGCGTTCCTCGCTTCAA AGAATGTAAGTAGGGGTGATTCT					
32	UNALM_VP A_559877	AX- 417368800	2	ABRR03000 026.1	38210467	GCGCATCTTCCACACAGCCTCTGCTT GTCCACGAAGACGTCTGAGGCCAAA TGAGCATCT[G/A]GGGGGCTCCCA GCAAGAACAACCCctgttcttttggtttcatga tggattttgtt	G	A	0.362	KIT	KIT proto-oncogene receptor tyrosine kinase
33	UNALM_VP A_559878	AX- 417368802	2	ABRR03000 026.1	38210802	CTTTTCTCCGATTCTGTTTGAAAAG TGAGTTGGCAGTTTATTTGAATAC ATGTCAGTC[G/A]TCACCTTCGGTCC TGATAATCATCCACAATGGGGTGAT GCGGCAGCAGGATTAGAAGCTG	G	A	0.189	KIT	KIT proto-oncogene receptor tyrosine kinase
34	UNALM_VP A_4284425	AX- 417368804	2	ABRR03000 026.1	38210961	caaaCAAAGGGCAGAAGCCTATGAAT ATCAGGGCTCAAGAAATCTAGTAGT TCATGCTGG[A/G]AATGAGACGCCG TGGAAAAGTGCTCTCCGAGCATGAA TACAGCCTGCCGGGCATGAGCCG	A	G	0.48	KIT	KIT proto-oncogene receptor tyrosine kinase
35	UNALM_VP A_559879	AX- 417368806	2	ABRR03000 026.1	38211222	GGTGTAGTTTCTGGCTCAAGTTTGCT GCGTGTAGGCATAGATAAAGAGCCA GACAAGTTT[T/G]AAATCAACAGTtag tgctttgcttttttttaagaaaaaacataattgCCAA GCA	T	G	0.122	KIT	KIT proto-oncogene receptor tyrosine kinase
36	UNALM_VP A_4284426	AX- 417369278	2	ABRR03000 026.1	38211336	CCAAGCACAATCTTAACAAAAATCT CCTTTTGTACCGATGAACTTGCTCA ATACAGATA[G/A]TCCGGAACAAAC TGACCCACTTCAGAATGGCATTGTG TTCAATAGATTTGatgccatttg	G	A	0.345	KIT	KIT proto-oncogene receptor tyrosine kinase
37	UNALM_VP A_4284427	AX- 417369279	2	ABRR03000 026.1	38211528	GATTCTTCTGtagcaggaataaaatttagattcC ACCTCCATTGCAGGCATGTCCTGGA[ T/G]GCCGGGCCAGTATCTATATAAG TGTGTATGTGCACGCGGTGTGTAT GCTTGTAGACAAA	T	G	0.108	KIT	KIT proto-oncogene receptor tyrosine kinase

38	UNALM_VP A_4284428	AX- 417369280	2	ABRR03000 026.1	38211710	GGCTTTCGTCCGCAGCGCAGCTCCT GTTTCTCTTCACCTGGCTGTCTAGAG AAGCTTACC[A/C]GCAGTTACACAG AGGCGAATAGGAAGCGGAGTGCGT AAGTACCTATTTATTTGCAACCCA	A	C	0.149	KIT	KIT proto-oncogene receptor tyrosine kinase
39	UNALM_VP A_4284429	AX- 417368814	2	ABRR03000 026.1	38211834	GCACTTAAGGCACTCTGTTATTTATA CTCAACATACTGTACCTGTTCCGTA GACATTATT[G/C]ATGCTACTCTTCC CACcgaacatatttaaatttaccctTAGGCTGT AGCCTGGATAAC	G	C	0.357	KIT	KIT proto-oncogene receptor tyrosine kinase
40	UNALM_VP A_4284545	AX- 432742373	na	ABRR03000 580.1	9989	ACCAGAACCGCGAGTGCTGCCAGAG CAACCTGGAGCCGCTGTTTCGACGGC TACATCCAGA[C/T]GCTGCGGCGGG AGGCCGAGTGCGTGGAGGCCGACA GCGGGAGGCTGGCCTCGGAGCTCAA	C	T	0	LOC102 534162	keratin type II cuticular Hb1- like
41	UNALM_VP A_4284433	AX- 417369286	12	ABRR03077 305.1	1705701	CAGCAGAACCCCATGAGCTGCTCC ATCTGCATGGCGTAGCGGGCCTCCA CCTCCCTCAG[G/A]CTGTTCTCCAAG CTGGCCTTCTAtagaggcacagggaggagggt CAGGTGAGACAGAGTG	G	A	0.162	KRT18	keratin 18
42	UNALM_VP A_4284434	AX- 417368826	12	ABRR03077 305.1	1708744	GGAGGCCAGGACTCAGGTGCTGGGC CGTGGAGTGAGTCCGATCGGCAACT CCAGTTATAG[T/C]Tccgagggagggggtgg ggactggCCCGGAcgggccccgccccgccacca GAGTCGCCG	T	C	0.26	KRT18	keratin 18
43	UNALM_VP A_4284435	AX- 417369288	12	ABRR03077 305.1	1740005	GGTGAACCAGAGCCTGCTGAGCCCC CTCAAGCTGGAGGTGGACCCACCA TCCAGGCCGT[G/A]CGCACCCAGGA GAAGGAGCAGATCAAGACCCTCAA CAACAAGTTTGCCCTCCTTCATCGAC	G	A	0.027	KRT8	keratin 8
44	UNALM_VP A_4125267	AX- 432742280	12	ABRR03077 305.1	1787767	AGAGGCCCTTGCTCCCCCTGGTCT CCCGCATCCTCATCAGCCCCATCCCT GCAGGGCTG[G/A]CTCTGTGTCCACT	G	A	0.189	KRT78	keratin 78

						GCCCATAGCCCGAGCCAGCCCACAG GGGACTCTGCAGAAGTCAGCCG						
45	UNALM_VP A_4284548	AX- 432742281	12	ABRR03077 305.1	1787904	TGTTCCAGCCTTGCTCTGCATACTGA TTTGATAGCGCCCTGTCCTCCAGTCA GACAGCCT[C/T]GTCTTCTCCTCTCTC CACAGCCTTCCACTGGGCTGGGAAA GCTCCTGTATTTGGGTGGT	C	T	0.193	KRT78	keratin 78	
46	UNALM_VP A_2785370	AX- 417368829	12	ABRR03077 305.1	1792695	CAGGAGCCAGCCCAGCTGGGAAA GATAAAAAGCCCATCAGTGGCCAG CAGCTCACTGC[T/C]GGCCTGCGTTC CTGCCCTCTGCAACCATGAAGTCC TCTGTGTCTCGGCGGATGTTCTCC	T	C	0.057	KRT79	keratin 79	
47	UNALM_VP A_2785409	AX- 417368831	12	ABRR03077 305.1	1795183	GAGACCAAGTGGGCACTGCTGCAGG AGCAGGGCCAGAACTCAGGCGTCAC CAGGAAAAT[C/T]TGGAGCCCCTCT TTGAGAATTACATCAGAAACCTGCA GAGGAACCTGGACAGCCTCCTGA	C	T	0.182	KRT79	keratin 79	
48	UNALM_VP A_4284549	AX- 432741861	12	ABRR03077 305.1	1796377	AACGTCTTAAAAGGCAGAACTGGGC TGAGCCTGGGCTGAGTCTTCCCTGT GTAACAGGTA[C/T]GAGGATGAAAT CAACAAGCGCACTGGTGC GGAGAA CGAGTTTGTGGTGCTCAAGAAGGTG	C	T	0.25	KRT79	keratin 79	
49	UNALM_VP A_4284436	AX- 417368833	12	ABRR03077 305.1	1796585	CAactgctcctcctcaccaccaggatGTGGATG CCGCATACATGGGCCGAATGGATCT [G/A]CAAGGCAGAGTGC GCCTCCTG ACCGAGGAGATTGACTTCTGAGTC ATCTCTACGAGATG	G	A	0.277	KRT79	keratin 79	
50	UNALM_VP A_4284550	AX- 432741863	12	ABRR03077 305.1	1802776	ACTGACCCttctctggcctttctttccccaggA TGCTCTGGAGAATGTCCCAGTGCC[G/ A]TTAGCATCTGTGAGTATTTTGTG CCTCAGGCCAGCCCCAGTTCTGCT CGGTCCATTCA	G	A	0.068	KRT79	keratin 79	



51	UNALM_VP A_2785688	AX- 417369292	12	ABRR03077 305.1	1818282	GGCCCTGAAAGATGCCTACAGCAAG CGCACAGAGCTGGAAACCGCCCTGC AGAAGGCCAA[G/A]GAGGAGCTGGC CCGGATGCTGCGTGAATACCAGGAG CTCATGAGCGTGAAGTTGGCCTTG	G	A	0.108	KRT4	keratin 4	
52	UNALM_VP A_4284437	AX- 417369293	12	ABRR03077 305.1	1818869	TGGTGCAGGAGGCATCGGTGGAGG GCTAGGAAGCGGCTCTGGATTTGGC CTGGGCAGCGG[C/G]TTTGGCTCTGG CTCCGGAATTGGCTTTGGATTTGGT GGTGGTGTCTGTGGCAGTTCACG	C	G	0.047	KRT4	keratin 4	
53	UNALM_VP A_2785746	AX- 417369294	12	ABRR03077 305.1	1830081	TGAGCTGTGTGGCCCGCACGGCGGG AGCCGGAGGAGGGGCCTTTGGGTTC CGGAGCGGAG[T/C]GGGTGGCTTTG GCAGTCGCAGCCTCTACAACCTGGG TGGCAGCAAGAGCATCTCCATCAG	T	C	0.301	KRT3	keratin 3	
54	UNALM_VP A_4284528	AX- 417369295	12	ABRR03077 305.1	1830218	CGGGCTGGTGGCTTTGGGGGAGGGC GCAGCAGCTGTGTCAGTGGCTTTGG AGGTGGCTAC[G/C]GAGGTGGATTT GGTGGTGGCTTTGGTGGTGGCAGAG GACTGGGAGGTGGCTTTGGAGGGG	G	C	0.047	KRT3	keratin 3	
55	UNALM_VP A_4284438	AX- 197140402	12	ABRR03077 305.1	1830523	CCAGATTGGGCAAGTAAAGGCCAG GAGCGGGAGCAGATCAAGACCCTC ACAACAAGTT[C/T]GCCTCCTTCAT CGACAAGGTGAGGGAACCTTGACTTG CGCCTCCATTCTAGGCAGGGGC	C	T	0	KRT3	keratin 3	
56	UNALM_VP A_4284529	AX- 417368844	12	ABRR03077 305.1	1833237	CAGAGCCATGTCAGTGACACGTCCG TTGTCCTGTCCATGGACAACAACCG CTCCCTAGAC[C/T]TGGACAGCATCA TCGCCGAGGTCAAGGCCAGTATGA GGAGATCGCCAACCGCAGCCGGG	C	T	0.078	KRT3	keratin 3	
57	UNALM_VP A_4284439	AX- 417368846	12	ABRR03077 305.1	1858839	CCAGTCAGCGTTTAGCTCAAGGAGC AGGCGGGTTTACAGCGCTGGCTCCT CTGCGGGCTC[C/T]GGTGGTGGGAGT	C	T	0.098	LOC102 534737	keratin type II cytoskeletal 1b	

						CAGGCTGTGGGGTCTGTGTGCCTGG TCGGAGGGAGGTGTGGCGGTGGT						
58	UNALM_VP A_2786126	AX- 417369299	12	ABRR03077 305.1	1867238	GCAGCAGGCTGACCATGATGCTGCT CTTGACCCTCAGTACCAGGAGCTCC AGATCTCAGC[G/A]GGAAGACACGG AGATGAGCTGAAGACCAGCAAGAT GGAGATCTCGGAGCTCAACCTCACC	G	A	0.345	LOC102 534737	keratin type II cytoskeletal 1b	
59	UNALM_VP A_2786129	AX- 417368850	12	ABRR03077 305.1	1867616	CTGCAGCAGTCCAAGGAGGAGATG ACTCGGCTGCTGCGTGACTACCAGG CGCTGCTGGGA[G/A]TCAAGCTGGC CCTGGACGTGGAGATCGCCACCTAC CGCAAGCTGCTGGAGGGCGAGGAG T	G	A	0.169	LOC102 534737	keratin type II cytoskeletal 1b	
60	UNALM_VP A_2786161	AX- 432741865	12	ABRR03077 305.1	1870045	AGACTTGGATCTGAACCAGTCTGCA AGCTGCCTGTTGTTTGAAGGTAGC TTAATGCTCT[T/C]TCCCAGATGATG AAGTCCCTTCTGCCTCTTGATCTCTG TTATGCTGAGCACTTTCTGAGA	T	C	0.193	LOC102 534737	keratin type II cytoskeletal 1b	
61	UNALM_VP A_4284551	AX- 432741867	12	ABRR03077 305.1	1870379	GTCCTCTGGAGAGTGTGCTGTGACTCA TCATTTCCCTGGTTGAAGTTTGTACCT TGAAGGGAT[C/T]gccatttctgtgtttctacaG TAAATTTCCCTGTGATCTCAAACCTGTT GTTTGGTCAGAT	C	T	0.324	LOC102 534737	keratin type II cytoskeletal 1b	
62	UNALM_VP A_2786252	AX- 417369301	12	ABRR03077 305.1	1878645	TCCTTCCTACTCTCTCCAAGTCAACA TGAGTCGACACTTTAGCTCCAGGTC TGGGTACCG[T/C]TGCGGAAAGGGC TTCAGCTCTGGCTCCGCCGGGGTGG TCAGCTACCAGCGCAGATCCACC	T	C	0.331	KRT1	keratin 1	
63	UNALM_VP A_4284440	AX- 417369302	12	ABRR03077 305.1	1879180	AGTAAAGTCACAAGAAAGGGAGCA AATCAAGTCCCTCAACAACCGGTTT GCTTCCTTCAT[T/C]GACAAGGTGAG TCTCTCTGCTCAGGCACTGGGGGAG TTTCTCCTGGTCCATGTGGGATTG	T	C	0.155	KRT1	keratin 1	

64	UNALM_VP A_2786280	AX- 417369305	12	ABRR03077 305.1	1881534	ctttgtttatgcTATAAGGGTCACTGAAAT ATTCTGTCTTGCAGGAGCTGTCTCA GAT[G/A]CAGACTCATATCAGCGAA ACCAATGTCATCCTCTCCATGGACA ACAACCGCAGCCTGGAC	G	A	0.335	KRT1	keratin 1	
65	UNALM_VP A_4125462	AX- 417369306	12	ABRR03077 305.1	1882500	AGAGTGTATAACATGCTGGATTCTC CTTTCTCAGATCTCTTCGCTGCAGCA GTCCATCAG[C/T]GATGCTGAGCAGC GTGGTGAGAACGCCCTCAAAGATGC CCAGAAGAAGCTGGCTGAGCTG	C	T	0.233	KRT1	keratin 1	
66	UNALM_VP A_4284443	AX- 417369308	12	ABRR03077 305.1	1883583	AGGGGGCTCTGGAGGTAGCTTCAGC TCCTCCGGAGGCCGAGGGTCCAGCT CTGGGGGCAC[C/T]AAGACCTCTGGT GGCAGTTCCAGCGTGAAGTTTGT CCACCAGCTATTCCCGAGGAACC	C	T	0.071	KRT1	keratin 1	
67	UNALM_VP A_4284552	AX- 432742286	12	ABRR03077 305.1	1900671	CAGGCAGACCAAACCCATTAATAAG CGAGCTCACACCCATGCACACAACA CCTAGTTGGC[G/A]GATATAAAAGA GGACCATGTCAGTCTCCTGACAGCA GAATCGAGTCTCCAGCCTTAACGC	G	A	0.01	KRT2	keratin 2	
68	UNALM_VP A_4284444	AX- 417368868	12	ABRR03077 305.1	1900806	CACTTTTCTCTCTGGGCAAAGGCA TCATGAGTTGTCAGATCTCGTGCAA ATCTCGAGGA[A/G]taggaggaggagg ggaggattCCGGGGCTTCAGCAGTG GTT CAGCTGTGGTCTCCG	A	G	0.375	KRT2	keratin 2	
69	UNALM_VP A_4284445	AX- 417368870	12	ABRR03077 305.1	1901318	CTCCTGCAGCCTCTCAACGTGAAAG TTGACCCAGAAATACATAATGTGAA GTCCCAGGAG[C/T]GGGAGCAGATC AAAACCTCTCAACAACAAATTTGCTT CCTTCATCGACAAGGTGAGTCACG	C	T	0.044	KRT2 LOC102 533992	keratin 2, keratin type II cytoskeletal 73	
70	UNALM_VP A_2786701	AX- 417368872	12	ABRR03077 305.1	1902454	GGTGCTGCAGACCAAATGGGAGCTG CTGCAACAAATTGACGTCAGCACCC GTACCACCCG[C/T]CTGGACCCCTC	C	T	0.166	KRT2 LOC102 533992	keratin 2, keratin type II cytoskeletal 73	

						TTTCAGGAGTACATTGGCAAACACTGC AGACTTACGTGAACACGCTTTCT					
71	UNALM_VP A_2786731	AX- 417369312	12	ABRR03077 305.1	1903905	TACAAGAACAAGGTGGACCTGCAG GTCAAGGTGGACATGATAAACCAGG AGATTGACTTC[T/C]TGAAAATCCTC TTTGATGCGGTAAGGAGGCTGCTCC TAGAGAGATCCTTATTTCCACCCT	T	C	0.253	KRT2 LOC102 533992	keratin 2, keratin type II cytoskeletal 73
72	UNALM_VP A_4284446	AX- 417368875	12	ABRR03077 305.1	1907833	GCTCTGGCTCTGGTGGCGGGTATGG CTCTGGAGGAGGATATGGTTCTGGA GGAAAACACA[G/T]CTCTGGAGGCG GCTCCAGAGGAGTCTCTGGCTCTGG GGGAGGGTCTGGTTCTGGAGGAGG	G	T	0.174	KRT2 LOC102 533992	keratin 2, keratin type II cytoskeletal 73
73	UNALM_VP A_4284553	AX- 432741870	12	ABRR03077 305.1	1916595	ACAAGAGCCTCCTGGCCCCACTCAA CGTGGAAGTGGACCCCAAGATCCAG AAAGTTCGTG[C/A]GCAGGAGCGGG AGCAGATCAAGGCACTGAACAACA AGTTCGCCTCCTTCATCGACAAGGT	C	A	0.179	LOC102 533992	keratin type II cytoskeletal 73
74	UNALM_VP A_2787006	AX- 432742288	12	ABRR03077 305.1	1923592	CCGGCATGGGGATGACCTCAAATAC ACCAAGAACGAGATCTCAGAGCTGA CTCGGCTCAT[C/T]CAAAGGCTGCGC TCGGAGATCGAGAGTGTGAAGAAG CAGGTGAGGGGTTCTGGGAGTTCC	C	T	0.027	LOC102 533992	keratin type II cytoskeletal 73
75	UNALM_VP A_2787091	AX- 432741874	12	ABRR03077 305.1	1927748	ACCACACCTCTTCTGTATGTAGAGG TCAGCAGCTTGGAACACGGGGCTTT GAAGGTCCAG[C/T]GTCTTCCAGAGT CCCAAAGCATCTTTTGATCTGACTA CAAGAGATCCTTGCATGGACAGC	C	T	0.193	LOC102 533992	keratin type II cytoskeletal 73
76	UNALM_VP A_4284554	AX- 432741876	12	ABRR03077 305.1	1927869	AGGCCCCACCATACCAATACTGGA TTTGGTGACACGCAGGATTCATCTC TGTCCCCAT[C/T]CCCAGCAGAAGC ACCTTCTAGGTCGTGCTTTTCTCAGG TCAGTGCTCCAGTGAGTGGGCA	C	T	0.044	LOC102 533992	keratin type II cytoskeletal 73

77	UNALM_VP A_4284555	AX- 432741878	12	ABRR03077 305.1	1928450	AGTTCAAGTCAATAGACTTGAATTC AGATTCAGTCCCTGTGACTAAGTGA TGATCATGAG[C/T]GAAGCTCTCAAT CTTTTGACCCTCATTTGttttccagctataa aataagcaGAATTAC	C	T	0.213	LOC102 533992	keratin type II cytoskeletal 73
78	UNALM_VP A_4284556	AX- 432742292	12	ABRR03077 305.1	1928551	tataaataagcaGAATTACACCCCGGGT TGTTTTGTGCAATTCCTGGTAAAAA GTTCA[C/T]GCCTAATAAATGTTAGT GGAAATGGAATCTTGTCTTCAGTCT TTTTGGTTCAAAAATGTA	C	T	0.169	LOC102 533992	keratin type II cytoskeletal 73
79	UNALM_VP A_2787257	AX- 417368877	12	ABRR03077 305.1	1935471	AACCAAGTGGGAGCTGCTGCAGCAG CTGGACCTGAACAACCTGTAAGAACC ACTTGGAGCA[C/T]GTCTTCGAGGGC TACACCAGCAACCTGCGGAAGCAGC TGGAGGCGCTGTCCGGGGACAGG	C	T	0.216	KRT72	keratin 72
80	UNALM_VP A_4284447	AX- 417368879	12	ABRR03077 305.1	1939056	TGTCCCCCAGGAGATCGCTCAGATC CAGTCCCACATCAGCGACACATCTG TCGTCCTGTC[T/C]ATGGACAACAAC CGGGACCTGGACCTGGACAGCATCA TCGCCGAGGTCCGCGCCCAGTAC	T	C	0.108	KRT72	keratin 72
81	UNALM_VP A_2787355	AX- 417368881	12	ABRR03077 305.1	1939128	CCGGGACCTGGACCTGGACAGCATC ATCGCCGAGGTCCGCGCCCAGTACG AGGAGATCGC[C/G]CTGAAGAGCAA GGCCGAGGCCGAGGCGCTGTACCAG ACCAAGGTGGGCCGAGTGCCAGGT	C	G	0.26	KRT72	keratin 72
82	UNALM_VP A_4284448	AX- 417369317	12	ABRR03077 305.1	1943863	CTTCTGTAACCACAATCCCAGTCCC ATCCCCTCTTCCTGTTGCCCTCAGCA GGGCGGTGC[T/C]CTGGTGCAGGAG GTCCAGCTGTGGGCTCTTTGGTGA GCTCTCCTCCTCGATCACCTTTC	T	C	0.483	KRT72	keratin 72
83	UNALM_VP A_2787604	AX- 417369318	12	ABRR03077 305.1	1954820	CGTGGAGCTGGACCCCGAGATCCAG AAAGTGCGGGCGCAGGAGCGGGAG CAGATCAAGGC[G/A]CTGAACAACA	G	A	0.253	LOC102 525527	keratin type II cytoskeletal 74

						AGTTTGCCTCCTTCATTGACAAGGT GGGTCTCACAGAAGGATAAGACGGT						
84	UNALM_VP A_4284557	AX- 432741882	12	ABRR03077 305.1	1956141	GCAGAGGCCTCTGCTATCTTCTCTCC TTGCCGACAGGTACGAGGTGGAGAT TAATCGGCG[C/T]ACAGCAGCTGAG AATGAGTTTGTGGTGCTCAAGAAGG TAAGGGAGAAAAGAGCCTATGGG	C	T	0.047	LOC102 525527	keratin type II cytoskeletal 74	
85	UNALM_VP A_4284558	AX- 432741883	12	ABRR03077 305.1	1961927	TCCCATGGCACACTCGCGCCTCTTTA ACAAAGGGAATGTCAGATTCATGGT TCTTCTGAT[G/T]AGGGTAGCTGACT TAGGTTGCCTTTCTGTTAGTGGGTAT GTGGTTTCTCTTTGGGCCTCT	G	T	0.014	LOC102 525527	keratin type II cytoskeletal 74	
86	UNALM_VP A_4284449	AX- 417368886	12	ABRR03077 305.1	1973595	GAGCTCCGTCTCCACTGGAAGCATG AGCCGCCAATTCACCTGCAAGTCGG GAGCTGCTGC[C/G]AAGGGGGGCTT CAGTGGCTGCTCGGCTGTGCTCTCG GGGGCAGCTCGTCCTCCTACAGG	C	G	0.416	LOC102 533318	keratin type II cytoskeletal 71	
87	UNALM_VP A_2787770	AX- 417369320	12	ABRR03077 305.1	1977073	TGGGCAAACGAGGGTTCCAGTGAG ACGAATTAACCGATTGAGTCCAAT ATTCCCACAG[G/A]TACGAGGAAGA AATCAACAGGCGGACGGCTGCAGA GAACGAGTTTGTGCTGCTCAAgaag	G	A	0.01	LOC102 533318	keratin type II cytoskeletal 71	
88	UNALM_VP A_4284530	AX- 417369321	12	ABRR03077 305.1	2006972	GCCCAAATGCCCTTTGCTTCCTGGC ACCCTTTTCTGGGCTCCTCACTGCTC TGGTCCCC[G/A]GTGGTAAAGAGAA ACTTGCTAGCAGCATTGCCGTCTTG CTGCCAAGACATCAGGGCTGCA	G	A	0.152		keratin 5*	
89	UNALM_VP A_4125950	AX- 417368892	12	ABRR03077 305.1	2007271	TTGCGATGCCTGGGTAACAGAGCCA CCTTCTGCGTCTTGCTGTGCTCTGTT CTCTCCAGC[A/G]CCTCCCCACTGCC GGTGTCTAGTCATCTTGCTCCACTAG GAACAAGCCATCATGTCTCGC	A	G	0.027	KRT5	keratin 5	

90	UNALM_VP A_4284450	AX- 417368894	12	ABRR03077 305.1	2008612	TCTGTTTGAGCAGTACATCAACAAC CTCAGGAGGCAGCTGGACGGTATCC TGGGGGAGAG[A/G]GGCCGCTGGA CTCGGAGCTGAGGAACATGCAGGAC CTGGTGGAGGACTTCAAGAACAAG	A	G	0.01	KRT5	keratin 5	
91	UNALM_VP A_4284531	AX- 417369325	12	ABRR03077 305.1	2013407	TCAAATCAGCTTTGGGTCTCCACAG CCTGATCCTCAGTGACAGAAGAATC CAAAGTTTTT[C/T]GGTCTCTAAACC ATTAACACAGAGTCCCCATCCTAAT CCCATCAGGTCTGACTACCTCCA	C	T	0.061	KRT5	keratin 5	
92	UNALM_VP A_2789370	AX- 432741887	12	ABRR03077 305.1	2067464	ACCAGCAGCGCGAGGGCTGGCTTTG GGAGTCGCAGCCTCTATAACCTTGG GGGGACAAAG[C/A]GGGTCTCCATT GGTGGGTGTGGCAGCAGCTTCCGAA GTGGCTTTGGTGGCCGAGCAAGCG	C	A	0.044	LOC102 533079	keratin type II cytoskeletal 75	
93	UNALM_VP A_2789395	AX- 432741889	12	ABRR03077 305.1	2068423	GGCCTACATCAACGACCTCCGGCGG CAGCTGGACAGCATCACTGAGA GGGGCAGGCT[C/G]GATGCTGAGCT GAGGAACATGCAGGATGTTCGTGGA AGATTTCAAAGTCAGGTGAGTGTGA	C	G	0.476	LOC102 533079	keratin type II cytoskeletal 75	
94	UNALM_VP A_4284560	AX- 432741890	12	ABRR03077 305.1	2069883	CTACATGAACAAGGTGGAGCTGGAG TCCAAGGTCAACTGTCTGACTGATG AGATCCACTT[C/T]TTCCGGATGCTC TTTGAGGCAGTAAGAGCCTACAAGT GTTTTGGAGGGTAGCAAGGAGGG	C	T	0.007	LOC102 533079	keratin type II cytoskeletal 75	
95	UNALM_VP A_4284561	AX- 432742299	12	ABRR03077 305.1	2080250	TAGGCGGGCTTGGGACCCAGGCCTG CCCCAAGTCCACATTCCCAGGGCTG GTCCTGCCCT[G/C]GGTGGGGTGTTC CCCATGATGATGAGAGGGGCAGGA AGATGGGCTGTGGCCAGACATGCC	G	C	0.189	LOC102 532430	keratin type II cytoskeletal 5- like	
96	UNALM_VP A_4284562	AX- 432741894	12	ABRR03077 305.1	2080480	CATCCCTTTGACATTTTCACTCTTCT GACCCTCTGGCACCAGCACCCACCA TGACTCAAC[G/A]ATCCTCCATCACC	G	A	0.071	LOC102 532430	keratin type II cytoskeletal 5- like	

						ATCAAGTCAGGGGGCACTCGGAACT TCAGTGCATCCTCAGCTAGCCT					
97	UNALM_VP A_4284563	AX- 432741896	12	ABRR03077 305.1	2080707	TTTGGGGGTAGTAAGAGGATCTCCA TCGGTGGGAGCTACCGCTCCGGCCA GGCCAGCTTC[G/A]GAGGTGCTGGCT ATGGGCTGGGTCTTGGGGGCATTGG GTACAGGGTAGGGGGAGCCAGCA	G	A	0.017	LOC102 532430	keratin type II cytoskeletal 5- like
98	UNALM_VP A_2789653	AX- 432742302	12	ABRR03077 305.1	2082889	GCCCATGTTTGAAGCCTACATCAGC AACCTGAGGCGGCAGCTGGACAGCC TGGGTGGAGA[A/G]CGCTTGAGGCT GGAGACAGAGCTGAAGAGCATGCA GGATGTGGTGGAGGACTTCAAGATC	A	G	0.071	LOC102 532430	keratin type II cytoskeletal 5- like
99	UNALM_VP A_2789746	AX- 432742303	12	ABRR03077 305.1	2088330	GGACAACAACCGCAAGCTGGACCTG GACAGCATCATCGCTGAGGTCAAGG CCCAGTACGA[G/A]GACATCGCCAA CCGCAGCCGGGCTGAGGCTGAGTCC TGGTACCAGATCAAGGTGAGACAT	G	A	0.056	LOC102 532430	keratin type II cytoskeletal 5- like
100	UNALM_VP A_4126207	AX- 432742304	12	ABRR03077 305.1	2092593	AAACCCTGCCACATAAttcctctcacctct gtcCTCCTTGTCCTCCAGGCACCCA G[C/A]CATATCCCCAATAACTCCCA TCATCAGGATATTTGAATGAGAAAT CAACATCCTTCAA	C	A	0.395	LOC102 532430	keratin type II cytoskeletal 5- like
101	UNALM_VP A_4284564	AX- 432741904	12	ABRR03077 305.1	2092708	TTCAAATATTACCAGCTTGATTATT CATAGTAGATCAAATAATGTTATC AAGAACTCC[A/G]AAGTTCTCCGTGG CCCGGGGCTACATTACGTGAGAGTT CAAAGTCAAATTCCTTCCACAT	A	G	0.304	LOC102 532430	keratin type II cytoskeletal 5- like
102	UNALM_VP A_4284565	AX- 432741906	12	ABRR03077 305.1	2093640	AGACAGGAGAGTTCacggttcaaaaaaaaa gtcactcccGCTCAGTCAGTTCAAACCC[ G/A]TCCGAACTGCCTGCTCCTTGTGC TTGCCACTCCCTGAAATCGCTTCA ATAAACTGTCAC	G	A	0.007	LOC102 532430	keratin type II cytoskeletal 5- like



103	UNALM_VP A_4284566	AX- 432741908	12	ABRR03077 305.1	2107746	CCAAGCCAGACTCCCCTACCCCTGG TGGTTCCCCACCTGTCTTACCTCTGG CTCTCATGC[G/A]CCTAAGAAAAGTA AAGTTCTTTTAACATCCACCTCATA ATGCACATTTGCAAATATCCA	G	A	0.051	KRT82	keratin 82	
104	UNALM_VP A_4284453	AX- 417368904	12	ABRR03077 305.1	2116381	TGCCGCTCTTACCGCGTCAGCTCCA GTCGCCGCGTGGGCAACTTCAGCTC TTGCTCATCA[A/G]TAAGCCCCGGGA ACCTGAACCACTTCCGGGCCAGCTC TGCTCTCTGCAGGAGCGGGCCCA	A	G	0.348	KRT84	keratin 84	
105	UNALM_VP A_2790027	AX- 417368906	12	ABRR03077 305.1	2116602	AGTTGGCTTCAGTGCTGGCAGTGGG ATGGCCTTTGGTTTTGGTGATAGGA GTGGTGCTGG[T/C]CTGGGCTTTGGG CCCAGCAGTTGTTGGGTCTGGGGT TTGGAGCCGGCAGTGGCCTTGGC	T	C	0.169	KRT84	keratin 84	
106	UNALM_VP A_2790040	AX- 417369330	12	ABRR03077 305.1	2118076	GGTGCGTTTCTGGAACAGCAGAAT AAGCTCTTAGAGACCAAATGGAGCT TCCTCCAAGA[G/A]CAGAAATGTGC CAGGAGCAACCTGGAGCCCCTCTTC GAGA ACTACATCACCAACATGCGG	G	A	0.051	KRT84	keratin 84	
107	UNALM_VP A_4284454	AX- 167296131	12	ABRR03077 305.1	2120437	TCTCGCTGATCCTCCAACACAGGAA ATCCAGTTGCTGCAGTCACACATCT CAGAGACATC[G/A]GTCATTGTGAA GATGGACAACAGCCGAGACCTGGA CCTTGATGGGATCATCTGTGAAATC	G	A	0.112	KRT84	keratin 84	
108	UNALM_VP A_4284455	AX- 417369332	12	ABRR03077 305.1	2120524	AGACCTGGACCTTGATGGGATCATC TGTGAAATCAAGACCCAGTATGAAG AGATTGCCAG[G/A]CGCAGTCGGGC TGATGCTGAGGCCTGGTACCAGACC AAGGTGGGCAGGAGGgcgctgggg	G	A	0.159	KRT84	keratin 84	
109	UNALM_VP A_2790311	AX- 432742308	12	ABRR03077 305.1	2136186	CAAGTGGCAGTTCTACCAGAATCAG CGCTGCTGTGAGAGCAACCTGGAGC CGCTGTTCAA[T/C]GGCTACATCGAG	T	C	0.139	LOC102 531931	keratin type II cuticular Hb5	

						ACCCTGAGACGGGAGGCCGAGTGC GTGGAGGCCGACAGCGGGAGGCTG					
110	UNALM_VP A_4284567	AX- 432741912	12	ABRR03077 305.1	2137135	CTGCGCCTACTTGCGGAAGTCAGAC CTGGAGGCCAACGTGGAGGCCCTGG TGGAGGAGTC[A/C]AGCTTCCTGAA GCGCCTCTATGATGAGGTCAGGGGG CCCCTGCCAGCCAGCCAGGAAGGG	A	C	0.26	LOC102 531931	keratin type II cuticular Hb5
111	UNALM_VP A_4284568	AX- 432741914	12	ABRR03077 305.1	2143144	AAATGGACGTGGACATGGGGGAAC CACCCCTTGTCTGTATACCCTGCTGT TACTCCCTTG[G/A]TGCTAGGGGACA GCACAGTGGTAGCCAGGGAAaggagtg ggaggaggcaggactcaGAGG	G	A	0.098	LOC102 531690	keratin type II cuticular Hb5- like
112	UNALM_VP A_4284569	AX- 432741916	12	ABRR03077 305.1	2143403	GTGCCTGGTGGGACTCCTTCCCCTCT CAGAGCAGATCTGATGCCCTTGGAA CCTGACTTT[C/T]TCCCTGCTCAAAC TTGGCGGTATCCAGACTGAGATGTG CCCAGGATACAGCCCCTGTGGG	C	T	0.155	LOC102 531690	keratin type II cuticular Hb5- like
113	UNALM_VP A_4284570	AX- 432741918	12	ABRR03077 305.1	2144144	GCAGGTTTCGTTTATGCCACTTCCAC CTCTTATCTCTGTCAGGCTAGTAAA GTAGGATCT[A/G]TTAGAACCCCTCG TCCAGGAGCATTCTGATGAGGCAAG ATGCTTATTCTGAAGGAACAAC	A	G	0.142	LOC102 531690	keratin type II cuticular Hb5- like
114	UNALM_VP A_4284571	AX- 432741920	12	ABRR03077 305.1	2144607	TCCACGGCCAGAATTACCTGACCTG CACACAAAACCCAGTGACTCTCAA CTTCAGTTCT[C/T]AGAGgaatcacctggaa gcttggttAAACTCAGATTCATGGGCTCA CCCCTGAATTTTG	C	T	0.311	LOC102 531690	keratin type II cuticular Hb5- like
115	UNALM_VP A_4284572	AX- 432742314	12	ABRR03077 305.1	2145272	TAGTTGGGGAAAGAAAATAGCTAA GCAGTTTATATCATAAAGCCTTTAT ATTCTTTCCCC[G/A]ACAACCAGAGG AGAGAAATCAACAGAGCTGGCAAG CATGCCAAATCCAGGGTTTGCCTC	G	A	0.105	LOC102 531690	keratin type II cuticular Hb5- like

116	UNALM_VP A_4284573	AX- 432742315	12	ABRR03077 305.1	2152850	TGGCCACATCATCATACTGAGCCTT GATCTCGGCCACAACCGAGTCCATG TTGAGCTCCC[G/A]GCTGTTGTCCAT CTTGACCACCACCGAGGTGTCAGAG ATTTGTGACTGAAGGAGGCAGAT	G	A	0.22	LOC102 531690	keratin type II cuticular Hb5- like
117	UNALM_VP A_4284574	AX- 432741926	12	ABRR03077 305.1	2154672	GTGGAAAAAGTCACTCACCTCCTTC AGCACCACAAACTCATTCTCAGCCG TGGCCTTGAG[A/C]GCCACCTCTTCT TCATACCTGAGGAGAGAGAGGATA GGAGGAGAGGGCCCCCGGAATGAG	A	C	0.385	LOC102 531690	keratin type II cuticular Hb5- like
118	UNALM_VP A_4284575	AX- 432741928	12	ABRR03077 305.1	2160004	CCTTCCTAAGCCACAGCCCCAGGGA ACAACCCAGAGACCCACCTTGTCAA TGAAGGCGGC[G/A]AACCTGTTGTTG AGGCTCTTGATCTGCTCCTTCTCCTC GTGCTTACGCACTGCGCGGTG	G	A	0.487	LOC102 531690	keratin type II cuticular Hb5- like
119	UNALM_VP A_4284456	AX- 417369333	12	ABRR03077 305.1	2177183	TCTTCAGTGTGAGGAGATCAAGGCC ACAGTGATCCGGCAAGGGGAGACC CTGCGCCGCAC[C/G]AAGGAGGAGA TCAATGAGCTGAACCGCATGATCCA GAGGCTGACAGCCGAGGTGGAGAA C	C	G	0	LOC116 282606 KRT83	uncharacterized LOC116282606, keratin 83
120	UNALM_VP A_2790635	AX- 417368962	12	ABRR03077 305.1	2177581	GGCCCCCTAGAACTCCAAGCTGGAG GCGGCAGTGACCCAGTCTGAGCAGC AGGGCGAGGC[T/C]GCCATTAATGA TGCCCGCTGCAAGCTGGCTGAGCTG GAAGGCGCCCTGCAGAAGGCCAAG	T	C	0.216	LOC116 282606 KRT83	uncharacterized LOC116282606, keratin 83
121	UNALM_VP A_4284458	AX- 417369358	12	ABRR03077 305.1	2179366	TGTGCCTGGGAAGGGACCCCTCCCT TCTGCTGTTTTGGCCTTTGGCCCCAC CCAGGAAGA[A/G]GGGCTCTCCTTTC CTGGAGGGAGCACCCTGGACCCAC TCCCCTAACTCTCTGGCCTCTT	A	G	0.027	LOC116 282606 KRT83	uncharacterized LOC116282606, keratin 83
122	UNALM_VP A_4284576	AX- 432741930	12	ABRR03077 305.1	2185554	CAAGACCAATGAATCCAGCAAAAAt cagaaagtagaagaagatGAACCCTCGGAA	T	C	0.118	LOC102 546053	keratin type II cuticular Hb6

						CAA[T/C]AAAACGACTACTAGGAAA GGCTACAGAAAGAGGCCAAGATGC ACCCAGCCTCCACAGGA					
123	UNALM_VP A_2790990	AX- 432741933	12	ABRR03077 305.1	2224991	CTTGATCTGCTCCTTCTCCTGCTGCT TCACGCACTGCGCATTGGGGTTCGAC CTCCAGGTT[G/C]AGGGGTGCCAGG AGGCTCTCGTTGACAGACACGGTGG TGATGCTGGGAGGGCTGGGTCCA	G	C	0.231	LOC102 546053	keratin type II cuticular Hb6
124	UNALM_VP A_2790991	AX- 432742321	12	ABRR03077 305.1	2225056	GTGCCAGGAGGCTCTCGTTGACAGA CACGGTGGTGATGCTGGGAGGGCTG GGTCCACACA[C/T]GCCACCGGAGC GGTACCCGAATTTGCGGCTGCAGGA GCCGGTGCGGGAGCCCCGCAGAC	C	T	0.257	LOC102 546053	keratin type II cuticular Hb6
125	UNALM_VP A_4284577	AX- 432741937	12	ABRR03077 305.1	2225247	CAGCAGCGGCCGGGCCGGGGCCCG CAGGCAGAGGTGCAGCTGAAGGCG CCACCAGAACTG[C/T]ATCTCTGAGT CGTCCACAGGAGCTGGATTTGCAG ACTCCACTACGGGAGCTAAAGCAAG	C	T	0.274	LOC102 546053	keratin type II cuticular Hb6
126	UNALM_VP A_2791082	AX- 432741939	12	ABRR03077 305.1	2234883	GCCTCTCTGCTTCCCCCTCAGTGTG AGGAGATGAAGGCCACAGTGGCCC GGCCCCGGCAC[G/A]GCGAGACCCTG CGCCGCACCAAGGAGGAGATCAAC GAGCTGAGCCGCATGATCCAGAGGC	G	A	0.331	LOC102 545800	keratin type II cuticular Hb1- like
127	UNALM_VP A_4284578	AX- 432742324	12	ABRR03077 305.1	2235903	GATGAACTCCAAGCTGGGCCTGGAC ATCGAGATCGCCACCTACCAGCGCC TGCTGGAGGG[C/T]GAGGAGCAGAG GTGAGAATCACACCTGGGGGAGGGT ATCTCTTCATCTCCAAACTGTTC	C	T	0.204	LOC102 545800	keratin type II cuticular Hb1- like
128	UNALM_VP A_4284459	AX- 417368966	12	ABRR03077 305.1	2238726	accaggggctgagctggggctggCCCAGGCCTC AGAAAGAGGCTACTTTATTGGAAGC TA[A/T]TCTGACATCATTTTTTCAGAC TGCCTTGCAGTCTTGGGACCAGGGA TGGGACTTGTCTTGT	A	T	0.188	LOC102 545800 KRT7	keratin type II cuticular Hb1- like, keratin 7

129	UNALM_VP A_4284460	AX- 417369360	12	ABRR03077 305.1	2238799	TTTCAGACTGCCTTGCAGTCTTGGG ACCAGGGATGGGACTTGTCTTGTA CTATGGGGTC[A/G]GGGTAAGGGTC TCCCACACCCGGGGCTCAGTTGAGG GTGCTCCTGCCAGTAGCAGATGTG	A	G	0.39	LOC102 545800 KRT7	keratin type II cuticular Hb1- like, keratin 7
130	UNALM_VP A_2791188	AX- 417368970	12	ABRR03077 305.1	2244624	GGCCTCAGCCTCCGCCCGGCTGCGG TTGGCGATCTCCTCATACTGGGCCTT GACCTCAGC[A/G]ATGATGCCATCCA GGTCCAGGGAGCGGCTGTTGTCCAT GGACAGGACCACAGACGTGTCT	A	G	0.34	LOC102 545800 KRT7	keratin type II cuticular Hb1- like, keratin 7
131	UNALM_VP A_2791194	AX- 417369362	12	ABRR03077 305.1	2244693	ATCCAGGTCCAGGGAGCGGCTGTTG TCCATGGACAGGACCACAGACGTGT CTGAGATCTC[A/G]GACTGCAGCTCT TTGAACTCCTATAGGGACAGAACAG AATGTCAGGACCTGGGGCATAAC	A	G	0.193	LOC102 545800 KRT7	keratin type II cuticular Hb1- like, keratin 7
132	UNALM_VP A_2791918	AX- 417369364	12	ABRR03077 305.1	2295468	CGCTGGCGCTTCTGCAGAGCCAGG ACTCGGCTGCCTTTGACCTTGGGCA CCTCTACAAC[G/A]AGTACCAGGGC CGGCTGCAGGACGAACAGCACAAA CTCAGCCAGGAGCGGGGCCAGCTGG	G	A	0.108	KRT80	keratin 80
133	UNALM_VP A_2792039	AX- 417269813	12	ABRR03077 305.1	2307547	GACCTGAGCGGCATCGTGGAAGAG GTCAAGGCCCAATATGATGCCATCG CAGCTCGCAGC[C/T]TGGAAGAGGC CGAGGCGTACTCTCGGAGCCAGGTC TGGGGCTGATGGTGCAGGCTGGGGT	C	T	0.105	KRT80	keratin 80
134	UNALM_VP A_4284579	AX- 432741943	12	ABRR03077 305.1	2310300	CCTAAACTCACATCTCAAGCCTTGC CTTGCCTCTGCCTCGGAACCGAGGC GGGTGCCCTC[G/A]CTCACCCAACCC CTGCCAGCCACGGACCAGGGGCTAA TGGGGACTGTTTGGGCATCCTCC	G	A	0.003	KRT80	keratin 80
135	UNALM_VP A_4284580	AX- 432742326	12	ABRR03077 305.1	2310396	AATGGGGACTGTTTGGGCATCCTCC TGTATCTGGAGTTCCTGGGGAGGCT GGGTGCTGAG[G/A]ACCCCTGGCTTT	G	A	0.166	KRT80	keratin 80

						CCTGCACCACCTAGGCTCCCCAGAG ACCTGACAGTattaggagagag						
136	UNALM_VP A_2792073	AX- 432741947	12	ABRR03077 305.1	2311074	GGCCATGAAGAAGGCACGTCCACTC GGGTCTGGGAGCCTCTGAGCCCTTG AGGTCAGGCT[G/T]AGGGGCCCCAA TAACAGCCAAAGAGAAATCTGCTGC CAACAAGAAgatctgggctggga	G	T	0.446	KRT80	keratin 80	
137	UNALM_VP A_4284582	AX- 432741949	16	ABRR03077 387.1	112395	GCGAGAAGGAGACCATGCAGAACC TCAATGATCGCCTGGCCTCCTACTTG GAGAAGGTGC[G/A]TGCCCTGGAGG AGGCCAACGCCGACCTGGAGGTGA AGATCCGGGACTGGTACAAGAAGC A	G	A	0.22	LOC102 527297	keratin type I cytoskeletal 42-like	
138	UNALM_VP A_4284583	AX- 432742329	16	ABRR03077 387.1	112462	GGAGGAGGCCAACGCCGACCTGGA GGTGAAGATCCGGGACTGGTACAAG AAGCAGGGCCC[C/T]GGGCCTGCC GTGACTACAGCTCCTACTTTAAGAG ACCACTGAGGACCTTCGAAGCAAGG	C	T	0.041	LOC102 527297	keratin type I cytoskeletal 42-like	
139	UNALM_VP A_3421785	AX- 417270523	16	ABRR03077 387.1	124019	GCGAGGCCAGGTGGGTGGCGAGAT CAATGTGGAGATGGACGCTGCCCC GGCGTGGACCT[A/G]AGCCGCATCCT GAACGAGATGCGCGACCAGTACGA GAAGATCGCGGAGAAGAACCGCAA G	A	G	0.111	LOC102 527556	keratin type I cytoskeletal 17	
140	UNALM_VP A_3421821	AX- 371970733	16	ABRR03077 387.1	126108	TGGGGCGGAACCAGGGTCTTtggccc ccaccaccccgaaTGTCTGTCTCTTGCAG C[C/T]GTGACCACCCGCCAGGTGCGC ACCATTGTGGAAGAGGTCCAGGACG GCAGGGTCATCTCC	C	T	0.041	LOC102 527556	keratin type I cytoskeletal 17	
141	UNALM_VP A_4284462	AX- 417273181	16	ABRR03077 387.1	126186	GCGCACCATTGTGGAAGAGGTCCAG GACGGCAGGGTCATCTCCTCCCGCG AGCAGGTCCA[C/T]CAGACCACCCA	C	T	0.047	LOC102 527556	keratin type I cytoskeletal 17	

						CTGAGGACCTGGCCAGCCTGGCCG ACCCCTGCCAGGGGTGAGGCCCC					
142	UNALM_VP A_3421944	AX- 432742330	16	ABRR03077 387.1	131579	ggcagaggcccagaggaggcaggtgaggggcctcagg tcacacagtgagCAAAGCAGAGC[A/G]GG GCCATGAACGCAGCTCTCCTGGCCA CAGCCAGCCATGCATGATCCCCGCA GCTGCACC	A	G	0.152	LOC102 527556 LOC102 530093	keratin type I cytoskeletal 17, keratin type I cytoskeletal 16
143	UNALM_VP A_4284463	AX- 417273457	16	ABRR03077 387.1	131867	TCTCCTAGCCTGAGTCACCACCCAC TTGAGCTGGgttttctctctcagcttcacATCCC [C/T]CAAAGCTGGGGTGGGAACCTCG GAGCAGCACATAGTGAATTGAGTTG ACCTTGGGCCGAAT	C	T	0.027	LOC102 527556 LOC102 530093	keratin type I cytoskeletal 17, keratin type I cytoskeletal 16
144	UNALM_VP A_4284465	AX- 417274340	16	ABRR03077 387.1	152807	AGTTTCCAGTAAAATATGACAGACG TGATGAGGCGGCGAAGAGGAGGGA TCTGGCTGGGA[G/A]TTGGCACTGGC ACCTGGATGACGAAAGCCAAGGGG AATGGAAAATGCCAGGCCCGCCCT	G	A	0.155	LOC102 527828	keratin type I cytoskeletal 14
145	UNALM_VP A_3422298	AX- 417274516	16	ABRR03077 387.1	153447	CCAACGCCGACCTGGAGGTGAAGAT CCGCGACTGGTACCAGAGGCAGCGT CCTGCTGAGA[T/C]CAAGGACTACA GCCCCTACTTCAAGACCATCGAGGA CCTGAGAAACAAGGTGGGTGGGCC	T	C	0.125	LOC102 527828	keratin type I cytoskeletal 14
146	UNALM_VP A_4284467	AX- 417278622	16	ABRR03077 387.1	155504	CGGCCTGCGCAGGGTGCTGGACGAG CTGACTCTGGCCAGGGCTGACCTGG AGATGCAGAT[A/C]GAGAGCCTGAA GGAGGAGCTGGCCTACCTGAGGAA GAACCACGAGGAGGTGAGAGCTAA A	A	C	0.171	LOC102 527828	keratin type I cytoskeletal 14
147	UNALM_VP A_4284468	AX- 417280422	16	ABRR03077 387.1	157115	aacaccctctccctcttcccagtgACCTCCTCC AGTCGCCAGATTTCGCACCAAAGT[C/ A]GTGGACGTGCACGATGGCAAGGT GGTGTCTCCACGAGCAGGTCCTT CGACCAAGAAC	C	A	0.034	LOC102 527828	keratin type I cytoskeletal 14

148	UNALM_VP A_3422452	AX- 432742331	16	ABRR03077 387.1	163952	CTTAACCTCAAGGCTGACTCTGCTC ACACCAACCGGCCCTCCTCAGGGGA TTGGGGGAGC[G/A]ttgggctggagagagaag gagctgctcccagccccaccctcccaatCCTGGG GCCttg	G	A	0.071	KRT9	keratin 9
149	UNALM_VP A_4284584	AX- 432742332	16	ABRR03077 387.1	164353	CCCAGGGCTCCACATGGGCTAGGCC CCTGGAAAAGGTCAACGCCATACCC ACTGCTGGGC[T/A]CTGCTGAAGAAC CAAGACAGTTGAGCCACTTGCCTTT CTGCCTGAAGGTCAGCTAATACT	T	A	0.034	KRT9	keratin 9
150	UNALM_VP A_4284585	AX- 432741959	16	ABRR03077 387.1	164588	CTGCCACGCCACTCATCCTTTCCATC AGACTGGAGACTCTGAGAGCAAGG ACTGGGCTCC[C/T]CATCAGACTGGG CTCTGTGAGTCCAGGGGCCAGTCTC CCTACCCACCCAGGCAGCTCCCC	C	T	0.142	KRT9	keratin 9
151	UNALM_VP A_3422455	AX- 432741961	16	ABRR03077 387.1	164698	GGCAGCTCCCCATCCTCTATCATA ACTCCTTGGTACTCCAGTCTAGCTC ACCCACAGT[G/A]GGCCCCAAGTGG GAGTGGAGAGGGGTGCAGACAGTA ATGATCCTCAAAGGAGGAGGTGAC	G	A	0.166	KRT9	keratin 9
152	UNALM_VP A_3422457	AX- 432742335	16	ABRR03077 387.1	165008	AAGGACGCTGAGCCTGGGGCAGCTT GGAGAAGCCCAGACACCTCCCGTAT CTCAGATCTC[C/A]TTAGAGTCTGTC TCAGGGAGACCAGGGCAGACAGAG GGCACACCCAGGCGTGCCACCATC	C	A	0.166	KRT9	keratin 9
153	UNALM_VP A_3422469	AX- 417283604	16	ABRR03077 387.1	167876	CAGTGTGGAGATGAACGCTGCTCCT GGCAGAGATCTACTAAGATCCTCA ATGACATGCG[T/C]GAAGAGTATGA GCAGCTCAGTGCTAAGAACCGAAGG GACATCGAGCAGCAATATGAGAGT	T	C	0.196	KRT9	keratin 9
154	UNALM_VP A_4284470	AX- 417289873	16	ABRR03077 387.1	168957	ATGTGTTATGGCCAGGAAGTGGAGT CCAACAACAAGGAGGTGACCAAGC TTCGGCACAGC[G/A]TCCAGGAGTTG	G	A	0.311	KRT9	keratin 9



						GAGATCGAGCTGCAGGCTCAGCTCA GCATGGTTGGTGTTCCTTCCCTGG						
155	UNALM_VP A_3422934	AX- 417293204	16	ABRR03077 387.1	202481	TGCCCTCTCTCAGCACCAAGCACGG CTGagactccttccctctccctgccctcagaaAGC[C/ T]GCCCTGGAAGGCACCCTGGCAGA AACGGAGGCTCGCTTTGGAGCCCAG CTGGCACAGATC	C	T	0.052	KRT19	keratin 19	
156	UNALM_VP A_3423037	AX- 417295077	16	ABRR03077 387.1	208313	CAGTGCCTGGGTCTCAGCCCTGCCC GTGCCTCCCACCAGCTTCTCTTGAC TGCTGAGAG[C/T]TTGGGGCTGCTGC CTGGCCATGAGCACCACATTTCTGC AGACTTCTTTCTCCAGCCTTGG	C	T	0.22	LOC102 528247	keratin type I cytoskeletal 15	
157	UNALM_VP A_3423039	AX- 417296396	16	ABRR03077 387.1	208384	TGCCTGGCCATGAGCACCACATTTT TGCAGACTTCTTTCTCCAGCCTTGGG GGTGGCTCT[A/T]CCCGAGGGGGTTC CCTCCTGGCTGGGGGAGGTGGCTTT GGCGGCGGGAGCCACTATGGGG	A	T	0.085	LOC102 528247	keratin type I cytoskeletal 15	
158	UNALM_VP A_4284586	AX- 432741965	16	ABRR03077 387.1	209390	TCCTGTTGACTGGGATGTTCTGTAG ATGCATGGCTGTGTCAAGGTGTGCA GGAGGCCATC[G/A]ACCATGAGCTG TCCACAGGGCAGGCTAGTGGAGTGC CCTGGAGATGGGCACCGAGCCCAC	G	A	0.118	LOC102 528247	keratin type I cytoskeletal 15	
159	UNALM_VP A_3423076	AX- 417304006	16	ABRR03077 387.1	212416	AGCCTCCCTGGGTGGTGGCAGCGGC GGCAAAGTCCGCATCAATGTTGAAG AGTCAGTGGA[T/C]GGAAAGGTGGT TTCTTCTCGAAAGAGAGAAATCTAA ATACCCAGTGCGAGAGAGACGCC	T	C	0.329	LOC102 528247	keratin type I cytoskeletal 15	
160	UNALM_VP A_4284473	AX- 417304457	16	ABRR03077 387.1	223012	cccacagGAGATGAAGGAGTTAGGCAA CCAGGTGGTTCGGCCAGGTCAACGTG GAGATGGA[C/T]GCCACCCCGGGCA TTGACCTGACCCGCGTGCTGGCGGA GATGAGGGAGCAGTACGAGGCC	C	T	0.01	LOC102 526268	keratin type I cytoskeletal 13	

161	UNALM_VP A_4284587	AX- 432742337	16	ABRR03077 387.1	235580	taaaaaaaaaacaacattctgagcctcagttctca atafTGGCCCCCTCCCAG[G/T]GTTGTG GTGAGGTGCACATGAAGTAATGAGG AGCTTGGCACATGAGAGATGTTCCA t gaa	G	T	0.233	LOC102 526014	keratin type I cuticular Ha6
162	UNALM_VP A_4284532	AX- 417304837	16	ABRR03077 387.1	235793	TGGGCAGACCAGGCTCCTCAGGGGA AGGCACAGAGAGAGGGCAGGTACC AAGTGTCATTC[G/A]TGAGGCCACTC CCAATTTATGGGATGTGGTGCAAAG ACTTTGAAGAGGcaataaaattca	G	A	0.018	LOC102 526014	keratin type I cuticular Ha6
163	UNALM_VP A_3423557	AX- 417306112	16	ABRR03077 387.1	236347	CTTTGGGAGCTGCTTGCCGGGTTCTT ATCTGTCTCTGGGTACCACTCCTCT GGCTTTGC[T/C]GGGAGCGGGGGCT GGTTCTGCGAGGGCTCCTTCAATGG CAATGAGAAGGAGACCATGCAG	T	C	0.014	LOC102 526014	keratin type I cuticular Ha6
164	UNALM_VP A_3423559	AX- 417306789	16	ABRR03077 387.1	236470	CCTGAACGACCGCCTGGCCA ACTACTGGAGAAGGTGCGGCAGCTGGAG CGGGACAACGC[G/A]GAGCTGGAGA GCCGCATCCGGGAGTGGTACGAGGC TCAGATCCCGTACATCTGCCCGGAC	G	A	0	LOC102 526014	keratin type I cuticular Ha6
165	UNALM_VP A_4284474	AX- 417308094	16	ABRR03077 387.1	239035	CGCTGTGACCTGGAGCGGCAGAACC AGGAGTACCAAGTGCTGCTGGACGT CAAGGCCCGC[C/T]TGGAGTCGGAG ATCGCCACCTACCGCCGCTGCTGG AGGGAGAAGACTGCAGGTGAGTGG	C	T	0.051	LOC102 526014	keratin type I cuticular Ha6
166	UNALM_VP A_4284475	AX- 417309886	16	ABRR03077 387.1	239917	TAACCGCTTTTACGCCACATAAAAC TGGGTTTCCCTAGAAATTTACCCAA TAAAGTGTGT[T/C]CTCTTGGCATAG CAA ACTGGCCTCTGTCTACTCTCTGT GGTTGGTTAGTGGGGGTGTCTG	T	C	0.027	LOC102 526014	keratin type I cuticular Ha6
167	UNALM_VP A_3423767	AX- 417312059	16	ABRR03077 387.1	245187	GGTGCCCTACCTGTGTCCTGACTAC CAGTCCTACTTTCTGACTATCGAGG AGCTTCAGAA[G/A]AAGGTGAGAGG	G	A	0.483	KRT35	keratin 35

						TTCAGATTAGCTTTGGCTGCCCTGG GAGCCTGGTGCCACCAGCCCTGGA						
168	UNALM_VP A_3423859	AX- 417311968	16	ABRR03077 387.1	246571	GGCTGACCTGGAGGCCAGGTGGAG TCCCTGAAGGAGGAGCTGCTCTGCC TTAAGAAGAA[T/C]CATGAGGAGGT GAGGATGGGGAGCAGGCCGAGGATG GGGGAGCAGGTGAGCATGGGGCTC A	T	C	0.132	KRT35	keratin 35	
169	UNALM_VP A_4284476	AX- 417312183	16	ABRR03077 387.1	246936	TAGACTCAACGTTGAGGTGGATGTG GCCCCACCTGTTGACCTGAACCGTG TTCTGGATGA[G/A]ATGAGATGCCA GTATGAGACCCTGGTGGAGAATAAC CGTCGGGATGCTGAAGACTGGTTC	G	A	0.331	KRT35	keratin 35	
170	UNALM_VP A_4284588	AX- 432742338	16	ABRR03077 387.1	256269	AGGGCAAGTGGTGAACAGATTCCCT GGGAGGGCGATTGGAATAGCTGGCC TTCCAGATTT[G/A]AATCTCATTAA TGATGAAGCGATGGTCAGGGAAAG AGACTTCCCTAGGGTGTAGAATTA	G	A	0.399	LOC116 283833	keratin type I microfibrillar 47.6 kDa-like	
171	UNALM_VP A_4284589	AX- 432741971	16	ABRR03077 387.1	256399	TTGAGTGCTCAGCCAGAGCCCTTAA TCTGGAAAAGGTCTGAGATCTAGT CTCAATTCTG[T/C]CATTGATCCATT ACATGACTCTAGGAggtccctccctgggcct cagtttctcatgtgt	T	C	0.203	LOC116 283833	keratin type I microfibrillar 47.6 kDa-like	
172	UNALM_VP A_4284477	AX- 417314108	16	ABRR03077 387.1	280362	AGCATTGAAAAACAGTGAGGAAGT AAAATGAAGAGAGACATGACTTCCT GTGTTACACAC[C/T]CTCTGTGGAACA GGTATAAATGCTCCCCAGAGGCTGA GAACTCATTGGCTCCTTGGCCAG	C	T	0.051		keratin 31*	
173	UNALM_VP A_4284478	AX- 417316374	16	ABRR03077 387.1	280521	AGCACCATGCCTTACAGCTTCTACC TGCCCAACCTGAGCTGCCGCTCCAG CTGCTCCTCC[C/T]GGCCCTGCGTGC CCCCAGCTGTCACGGCTGCACCCT GCCCCGGGCCTGCAACATCCCCA	C	T	0.136	LOC102 525174	keratin type I cuticular Ha1	

174	UNALM_VP A_3424322	AX- 417324648	16	ABRR03077 387.1	281472	CTCTGAGCCGATCCTTGTGCAGGTA TGAGACGGAGCTGGGCTTGC GGCGAG CTGGTGGAGT[T/C]GGACATAAATG GCCTGCGTAGGATTCTGGATGAGCT GACCTTGTGCAAGTCTGATCTGGA	T	C	0.171	LOC102 525174	keratin type I cuticular Ha1
175	UNALM_VP A_3424395	AX- 417327130	16	ABRR03077 387.1	285441	TCATGCACTAGAGCCCAGGGGTTGC CAGAGCAGCAAGGAGGCATGACCT GGGGTTCCAGA[G/A]CTTTGACGTGG CCCTGGTTTCATTCCGCTAAAAGGGC CTGAAAACAAAATTTGCATGGCTG	G	A	0.034	LOC102 525174	keratin type I cuticular Ha1
176	UNALM_VP A_4284590	AX- 432742340	16	ABRR03077 387.1	291116	CTGGAGAAGGTGCGGCAGCTGGAG CGGGACAACGCGGAGCTGGAGAGC CGCATCCGGGAG[T/C]GGTCCCAGC AGCAGGAGCCCCTGGTGTGCCCAA CTACCAGTCCACTTCCGGACCATC G	T	C	0	LOC102 525174 LOC102 528756	keratin type I cuticular Ha1, keratin type I cuticular Ha4
177	UNALM_VP A_4284479	AX- 417327491	16	ABRR03077 387.1	295100	TTCTAAACTTCTGTGTCTCCAGAC TCCCCTGCAACCCATGCGCCACCAC CAATGCTAG[C/T]AGCAACTCCTGCG GGCCCTGCTCTCAAAAACGTTGCTT TTAAAACTTGCATCCTCTTTGA	C	T	0.28	LOC102 525174	keratin type I cuticular Ha1
178	UNALM_VP A_4284591	AX- 432741975	16	ABRR03077 387.1	302386	GGAGAGCCGCATCCGGGAGCGGTCC CAGCAGCAGGAGCCCCTGGTGTGCC CCAATAACCA[G/A]TCCTACTTCCGG ACCATCGAGGAGCTCCAGCAGAAG GTGAGGGGCTGGGCACCCCGCTGC	G	A	0	LOC102 525174 LOC102 528504	keratin type I cuticular Ha1, keratin type I cuticular Ha1
179	UNALM_VP A_3424748	AX- 432742342	16	ABRR03077 387.1	308855	CACCCCTTGTGGCCCACGCTCCCGC ATCGGGCCCTGCAGCACCTTTGGGT GCTAGATATG[T/C]CGGGCTGAGCA GGAGA ACTGTGTGAAGACAGAAGA CCATCCACGGACCAGCCCTGCTCTC	T	C	0.051	LOC102 525174 LOC102 528504	keratin type I cuticular Ha1, keratin type I cuticular Ha1
180	UNALM_VP A_3424830	AX- 432742343	16	ABRR03077 387.1	321389	GGAGGTGTCCTTGCGGCAGCTGGTG GAGGCGGACCTGAACGGCCTGCGCA	C	T	0	LOC102 525174	keratin type I cuticular Ha1, keratin type I cuticular Ha3-I

						GGATCCTGGA[C/T]GAGCTGACCCTG TGCAAGTCCGACCTGGAGGCTCAGG TGGAGTCCCTGAAGGAGGAGCTG				LOC102 529018		
181	UNALM_VP A_4284593	AX- 432741982	16	ABRR03077 387.1	485030	AGGTCCAGGGGGAACAAGCAGACC TAACTCAGAACTTTTCCAAGCACC CAAATTCTCAG[C/A]CCAGCTTCTGA CATCATGGCCTGTTGCTCTACCAGCT TCTGTGGATTTCCCCTTGTTC	C	A	0.088	LOC102 528989 LOC102 536934	keratin high-sulfur matrix protein B2C, keratin high- sulfur matrix protein B2D- like	
182	UNALM_VP A_4284493	AX- 417368760	16	ABRR03077 387.1	540152	AACTGGAGTCTTTGTCCTTCCACACCT GCTAGGATGACTTCTGACTGTTCCC CCACCTGCT[G/A]CTCCTCCGAGTCC GGTGCTAAGTCTTCCGACGGTGCAA TTGCTTCCGCCTGTTCCATGGA	G	A	0.044	KRT40	keratin 40	
183	UNALM_VP A_4284494	AX- 417369367	16	ABRR03077 387.1	541254	AAAGACCTGTTTGTACATAATAAC TGACTTCCCAGATCCTGTGCACAAA GGCAGAGAAT[T/G]CTAGACTTGCTG TACAGCTTGACAACCTGCAAACCTGGC TGCTGATGACTTCAGGTCAAAGT	T	G	0.074	KRT40	keratin 40	
184	UNALM_VP A_4284495	AX- 417369368	16	ABRR03077 387.1	541670	GGTAGAGAATGACATCAGTGGCCTG CGTGGGATCCTGGGGGAGCTGACCT TGTGCAAATC[C/T]GACCTGGAGGCC CATGTGGAGTCTCTGAAGGAAGATC TCCTTTGCCTTAAGAACAATCAT	C	T	0.007	KRT40	keratin 40	
185	UNALM_VP A_3426104	AX- 417368986	16	ABRR03077 387.1	553786	AATAAGCAACTTCAACACCTGTCCC TCTCTGGATGATTGTGGCTGGTGTG GTGAGGGCAT[C/T]AACAGCTATGA GAAAGAGACCATGCAAATCTTGAAC GACCGCCTTGCTAACTACCTGGAA	C	T	0.155	KRT39	keratin 39	
186	UNALM_VP A_4284496	AX- 417368987	16	ABRR03077 387.1	554863	CTTCTTAATTCTAGATCTTGTGTACC AAGGCTGAGAATTCCAGACTGGTCT CACAAATTG[G/T]CAACACCAAACCT GGCTGCTGATGACTTGAGAGCCAAG TAAGCCCACCAACTCTGTTCGTT	G	T	0.287	KRT39	keratin 39	

187	UNALM_VP A_4284497	AX- 417368989	16	ABRR03077 387.1	555562	CTACATCTTAATGTCTTTCCGTGTCT CATTTTTAAAGGTGTGAAGCTGAGG TGTCCTAC[G/A]TCAGCTAGTGGAG GCAGATGCCAACAGCCTACAGCAGA TCCTGGATGCACTGACCCTGGG	G	A	0.155	KRT39	keratin 39	
188	UNALM_VP A_4284498	AX- 417368991	16	ABRR03077 387.1	558113	CCTTGTTGGCCCAGATCCAGTGTCT GATCGATTACCTAGAGGCTCAGCTG GCAGAGATCC[G/A]GGGTGCCCTGG AAAGACAAAACCAAGAATATGAGA TTCTGCTGGACGTCAGGTCCCGGCT	G	A	0.027	KRT39 LOC107 034533	keratin 39, uncharacterized LOC107034533	
189	UNALM_VP A_4284499	AX- 417369373	16	ABRR03077 387.1	558189	ACAAAACCAAGAATATGAGATTCTG CTGGACGTCAGGTCCCGGCTGGAGT GTGAGATTAG[A/C]ACATACCGCAG TCTTCTGGAGAGCTCGGATGGCAAG TACGTAAAATAAAAAGCAGCTTCTA	A	C	0.142	KRT39 LOC107 034533	keratin 39, uncharacterized LOC107034533	
190	UNALM_VP A_4284500	AX- 417369374	16	ABRR03077 387.1	559824	AGGATGGGAAAGTCATTTCTTCTTA CGAGCATGTGCAGCCTTGCTTCATC ACCAGAACTG[C/T]CAAACCTCTGACA TCCcaaggtgataaaaaaaaaaaggaccctcATTC ATGAAAGAAAGGC	C	T	0.074	KRT39 LOC107 034533	keratin 39, uncharacterized LOC107034533	
191	UNALM_VP A_4284501	AX- 417368997	16	ABRR03077 387.1	579395	AAATGGGAAGGAAATGATGCAGAA CCTCAACAGCCGCCTGGCCTCCTAC CTGGACAAGGT[A/G]CGCGCCCTGG AGGAGGCCAACACGAAACTCGAAA GTTGCATCTTCAAGTGGCATCAGCA G	A	G	0	LOC107 034533 KRT23	uncharacterized LOC107034533, keratin 23	
192	UNALM_VP A_3426177	AX- 417369376	16	ABRR03077 387.1	588826	TTCTCTTTCAGAAATCTTCTTTGGA AAGCATGTTATCGGAGACCCAGTCT CAGTACTCC[C/T]GCCAACTCCAGGA CATGCAACGGATCATCTCCCACCAC GAGGAGGAGCTGATGCAGCTAC	C	T	0.193	LOC107 034533 KRT23	uncharacterized LOC107034533, keratin 23	

193	UNALM_VP A_4284502	AX- 417369377	16	ABRR03077 387.1	588891	ACTCCAGGACATGCAACGGATCATC TCCCACCACGAGGAGGAGCTGATGC AGCTACGCCA[C/T]GACCTGGAGTGC CAGAGCAACGGATACAAGGTCCTCC TGGGCATCAAACCCACCTAGAG	C	T	0.01	LOC107 034533 KRT23	uncharacterized LOC107034533, keratin 23
194	UNALM_VP A_4284598	AX- 432741991	16	ABRR03077 387.1	622017	AGATTGCTGTACaaaggggaataaaaaaaaa aaaagccaggtggACTGTGGAGTCAATC[A /G]GAAAGTGAGATGTCAGCCCAGGA GGCGTTTCTTTCTGAAACTTTATA AAAGGTCAGCT	A	G	0.084	LOC107 034533 KRT20	uncharacterized LOC107034533, keratin 20
195	UNALM_VP A_4284599	AX- 432742351	16	ABRR03077 387.1	622090	TCAGCCCAGGAGGCGTTTCTTTCTTCT GGAACTTTATAAAAGGTCAGCTCA TCCACCAAGA[G/A]ATAGACTCTGCT CCATCCATCCTGAAGCTACAGCTGC TCCCTCCTAGAATCCCCGATGGA	G	A	0.179	LOC107 034533 KRT20	uncharacterized LOC107034533, keratin 20
196	UNALM_VP A_4284503	AX- 417369002	16	ABRR03077 387.1	622314	GGAGGCCACGGCACCCGCATCTCGA CCTCCAGACACATGGTGA ACTATGG GAGCGATCTC[G/A]CCACGGGGATC CTGTGTGCTGGCAATGAGAAGATGA CCATGCAGAGCCTAAACGACCGCC	G	A	0.282	LOC107 034533 KRT20	uncharacterized LOC107034533, keratin 20
197	UNALM_VP A_4284600	AX- 432741995	16	ABRR03077 387.1	629809	actGCCATCTTTACCACAAAATGCATT AATCACAAGTTACACACCACCTAAT ATCAATTC[G/A]TTAATGTTCCCAA CATTACAGCATATTTCAATTATACTA AACTGCTATACAGAGCAGAG	G	A	0.034	LOC107 034533 KRT20	uncharacterized LOC107034533, keratin 20
198	UNALM_VP A_4284601	AX- 432741997	16	ABRR03077 387.1	629901	ATTATACTAAACTGCTATACAGAGC AGAGGAGTGTATGCAA ACTGCAAAT GAACCTCTCT[C/T]TAATTCCTCTGT TTCTGGATTGCAAATCTCCTTTTCAG GGGTCCCTGGATTCTTGAGTC	C	T	0.093	LOC107 034533 KRT20	uncharacterized LOC107034533, keratin 20
199	UNALM_VP A_4284602	AX- 432741999	16	ABRR03077 387.1	630053	ATCAGTATAATTCTGCTGGCTTTAA GAGATATCTGCTTTTTGCCCTCATTG TCTCCACTG[T/C]CTTTAAAAGTCAG	T	C	0.293	LOC107 034533 KRT20	uncharacterized LOC107034533, keratin 20

						ATCAGCTCTTTGAATAATTTGAGAG TCAAATTCCAATCATAAATGCT					
200	UNALM_VP A_4284504	AX- 417369004	16	ABRR03077 387.1	636205	TGGAGGCCTTCTTTCTGgatcagaaaagga aactatgaaAAATCTTAACGACAGATTG GC[G/T]TCCTACCTGGATAAGGTTTCG AGCTCTAGAAGAGGCGAACACCGA GCTAGAAAACAAAATT	G	T	0.257	LOC107 034533 KRT12	uncharacterized LOC107034533, keratin 12
201	UNALM_VP A_4284505	AX- 417369006	16	ABRR03077 387.1	636818	ctGGTCTCTCACAACTTGCTTCCCTT GTTTTGCCAGATCATTTCTGCCAGCA TTGCAA[C/T]GCCAGCTCATCCTG CAGATCGACAATGCAAGACTAGCTG CCGAAGACTTCAGGATGAAG	C	T	0.044	LOC107 034533 KRT12	uncharacterized LOC107034533, keratin 12
202	UNALM_VP A_4284506	AX- 417369381	16	ABRR03077 387.1	637875	TGAGCTGGCCCTGTGCCAGAACGTG GAAGCTGACATCAACGGCCTGCGCA GGGTGCTAGA[T/C]GAAATGACCCT GGCCAGAGCCGACCTGGAGATGCA GATCGAGAATCTCCAGGAGGAGTTG	T	C	0.404	LOC107 034533 KRT12	uncharacterized LOC107034533, keratin 12
203	UNALM_VP A_3426259	AX- 417369010	16	ABRR03077 387.1	639391	TCCATTGTTACGTCTGTGTCATTCCC CTCCACGTGCAGAAGAAATCCCTG GAGGACTCG[T/C]TGGCCGAAACCG AGGCCGACTACTGCGGCCAGCTGTC CCAGGTGCAGCAGCTCATCGGCA	T	C	0.034	LOC107 034533 KRT12	uncharacterized LOC107034533, keratin 12
204	UNALM_VP A_4284603	AX- 432742355	16	ABRR03077 387.1	641072	TGATATATTAGTTTAAAtctccatttctgtttttt cattttctctggattcTCCATTAC[C/T]TTGA GTATTTTTTGGCCTTCTGAAACAAAA TTCAATTACAAGTCATTGTGGCCAT GTTGG	C	T	0.044	LOC107 034533 KRT12	uncharacterized LOC107034533, keratin 12
205	UNALM_VP A_4284604	AX- 432742356	16	ABRR03077 387.1	641211	CCTTACATCTTTCTGGACCTGGAGTA GTCTTTTAATAAACTTTCTTCTAGTG ATCTGGAA[T/C]ATGTTTTTAATCAC AAGACTTTGATCAAGTAGTACTGTT TTAATAAAGCTAATTTTAATG	T	C	0.044	LOC107 034533 KRT12	uncharacterized LOC107034533, keratin 12



206	UNALM_VP A_4284605	AX- 432742357	16	ABRR03077 387.1	641708	GGGAACGGTGATACAGCGAGATAG GGCAAACAGACGTTCAAAGGACACT TGGAGGAAGAA[A/G]TGAAGTGGTC AGGAGACAATCAAATGCTTCATACA GAAATCatccaggcagagaagggag	A	G	0.243	LOC107 034533 KRT12	uncharacterized LOC107034533, keratin 12
207	UNALM_VP A_4284606	AX- 432742007	16	ABRR03077 387.1	641792	ATGCTTCATACAGAAATCatccaggcaga gaagggagaaaggaatcGAGGAAGCACAT G[T/C]GAAAGGACACAGGGGCGTGA ACAAACATTGCTTTGATGGCCTCTCT CCTATCTGGCACCA	T	C	0.044	LOC107 034533 KRT12	uncharacterized LOC107034533, keratin 12
208	UNALM_VP A_4284508	AX- 417369385	16	ABRR03077 387.1	668129	AGGAGATGGTGGCCTTATCTCCGGA AATGAGAAAGTAACCATGCAGAATC TGAATGACCG[C/T]CTGGCTTCTAC TTGGACAAAGTTCGGGCTCTGGAAG AGTCAAACACTACGAGCTAGAAGGC	C	T	0.078	LOC107 034533 KRT10	uncharacterized LOC107034533, keratin 10
209	UNALM_VP A_4284509	AX- 417369017	16	ABRR03077 387.1	669613	CAGATACGAGAATGAGGTGACCCTG CACCAGAGTGTGGAGGCTGACATCA ATGGACTGCG[C/A]AGGGTGCTGGA CGAGCTGACCTTGACCAAGAGTGAC CTGGAGATGCAGATCGAGAGCCTG	C	A	0.24	LOC107 034533 KRT10	uncharacterized LOC107034533, keratin 10
210	UNALM_VP A_4284510	AX- 417369019	16	ABRR03077 387.1	670588	TTGGCAGAGACAGAAGGTTCGCTATT GTGTGCAGCTCTCACAAATTCAGTC CCAGATCACT[G/T]CTCTGGAAGAGC AGCTGCAACAGATTCGGGCTGAGAC CGAGTGCCAGAATGCCGAGTACC	G	T	0.071	LOC107 034533 KRT10	uncharacterized LOC107034533, keratin 10
211	UNALM_VP A_4284511	AX- 417369388	16	ABRR03077 387.1	671818	TACTAACAAAGCCAGAGTGATCAAG ACAATTATTGAGGAGGTGGCACCTG ATGGTAGAGT[C/T]CTTTCATCTGTG GTTGAATCAGAAACCAAGAAACACT ACTATTAAGCCGCATCAAGAGGA	C	T	0.057	KRT10	keratin 10
212	UNALM_VP A_4284512	AX- 417369389	16	ABRR03077 387.1	679516	AAGCCATCAGCACACAACATTCTGT TCTCCTTTGAAGTTCAGCTATCCAC AGACAAGGC[A/G]AACATTTCTCAC	A	G	0.213	KRT28	keratin 28

						ACTACTATGTCTCTCCGATTTTCTAG TGGGTCCAGGCATATTTGCTTG						
213	UNALM_VP A_4284513	AX- 417369390	16	ABRR03077 387.1	679813	CTTTGCTGGAAGTGAAGGGGACTC CTCTCTGGGAATGAGAAGGTGACCA TGCAGAATCT[T/C]AATGATCGCTTG GCATCCTACCTGGATAACGTGCGCG CTCTGGAGGATGCAAATGCTGAc	T	C	0.105	KRT28	keratin 28	
214	UNALM_VP A_4284514	AX- 417369025	16	ABRR03077 387.1	688279	CAGAACCACACTGGTAAAGACGGTG GTTGAAGAGATAGATCAACGTGGCA AAGTTCTTTC[A/G]TCGAGGGTTCAG TCCATAGAAGAAAAGACATCCAAA ATGACCAATGGCAAGATAGAACAG	A	G	0.129	KRT28	keratin 28	
215	UNALM_VP A_4284607	AX- 432742009	16	ABRR03077 387.1	689077	AACTCAGCCATTGCAGCAGAGAGCA ACCATAAACTATAGATAAGCCAAAG AGCCTGACTA[T/C]GTtctaataagtttatta tggacATTGAAATTTGAATGGCAAAT ACTTTCAATGTGT	T	C	0.169	KRT28	keratin 28	
216	UNALM_VP A_4284608	AX- 432742011	16	ABRR03077 387.1	696568	AGGCCAACGCCGACCTGGAGCAGA AGATCAAGGGCTGGTATGAGAAATT TGGGCCTGGTG[T/C]TTGCCGTGGTC TTGATCATGATTACAGCAGATACTT CCCAATCATTGACGACCTTAGGAA	T	C	0.439	LOC102 526397	keratin type I cytoskeletal 27	
217	UNALM_VP A_4284609	AX- 432742012	16	ABRR03077 387.1	697268	ACTCTCCATTTATCAGATAATTTCTG CAACGACAAGTAATGCCAATGTTGT CCTGCAAAA[T/C]GATAATGCAAGA CTAACAGCTGATGACTTCAAATAA AGTGAGTAGAGAAGAGACATCAA	T	C	0.122	LOC102 526397	keratin type I cytoskeletal 27	
218	UNALM_VP A_4284610	AX- 432742014	16	ABRR03077 387.1	708649	GTTTGAAAATGAGCTTGCTCTGCAT CAGAGTGTTGAGGCTGACATCCACG GTCTTCATCA[A/T]GTTATGGATGAG CTGACCCTTTGCACGACCCACCTGG AGATACAGTGTGACGCACTCAGT	A	T	0.145	KRT26	keratin 26	

219	UNALM_VP A_4284611	AX- 432742016	16	ABRR03077 387.1	708716	GATGAGCTGACCCTTTGCACGACCC ACCTGGAGATACAGTGTGACGCACT CAGTGAGGAG[T/C]TGATGTACCTCA AAAAAATCACGAGGAGGTAAGGA ACCCTGCGGAGTAACTCTGTATGA	T	C	0.146	KRT26	keratin 26	
220	UNALM_VP A_4284612	AX- 432742364	16	ABRR03077 387.1	711622	gaaagaaaaaatgcaataatcttctcttcacagAA AAAGCAAGGCCACATGTTA[C/T]AC ATCAAAAAGGCATGGGCCATAAAC TCTGAAAATCAAGTCAAAGGTATAG AAGGGTTT	C	T	0.139	KRT26	keratin 26	
221	UNALM_VP A_4284613	AX- 432742365	16	ABRR03077 387.1	713015	AGctcattttgcctttcattaaaacatttctgTTGCA AGTTATTTTTTGCTAAATC[C/G]GGT GCTTCTGAATTTCCATGGTGTGATC TCAATGTTACAGTCTTAAGGAGATG ctaata	C	G	0.368	KRT26	keratin 26	
222	UNALM_VP A_4283886	AX- 417369027	16	ABRR03077 387.1	722327	TGTTTTACTCTATGTAGGCAGAATTT CATCCGGAACACACTTGAGAATTGT CTATAAAAG[T/G]GGGACCAGGAGA GTGGGGATGGGAAGTCACTCCACCC CGGCTTTCAGAGACAGCTCTGCG	T	G	0.294	KRT25	keratin 25	
223	UNALM_VP A_4283887	AX- 417369393	16	ABRR03077 387.1	722399	AGTGGGGATGGGAAGTCACTCCACC CCGGCTTTCAGAGACAGCTCTGCGA CTCGCTCGCC[C/A]GGATACCATGTC TCTCCGACTTTCAGCGGATCCAGG AGGGCCGGTCCCCGTGCCACGGC	C	A	0.294	KRT25	keratin 25	
224	UNALM_VP A_4284515	AX- 417369030	16	ABRR03077 387.1	722699	ATGAGAAGGTGACCATGCAGAACCT CAACGACCGCCTGGCGTCCTACCTG GACAACGTGC[G/A]CGCCCTGGAGG AGGCCAACCCGACCTGGAGCAGA AGATCAAGGGCTGGTATGAGAAATA	G	A	0.297	KRT25	keratin 25	
225	UNALM_VP A_4284516	AX- 417369395	16	ABRR03077 387.1	723129	tttttggttgatAAGTTGCCTCTAAAAAG TCAAACCTCCCCATTTCTAGATCATCG CT[T/G]CCACCACAGCAATGCTAAC	T	G	0.034	KRT25	keratin 25	

						GCTGTTCTGCAGATCGATAATGCCA GGCTGACCGCTGATG						
226	UNALM_VP A_4284517	AX- 417369396	16	ABRR03077 387.1	726355	ACACTTTTTCTGCCCTAGGAAATG CAAGTTCTGCAGTGC GCGCGCCGGAG GCAACGTGAA[C/T]GTGGAGATGAA CGCGGCGCCCCGGCGTGGACCTCACG GTTCTGCTGAACAACATGCGGGCC	C	T	0.051	KRT25	keratin 25	
227	UNALM_VP A_3426357	AX- 417369397	16	ABRR03077 387.1	728281	GCAGCTGCACCAGGTCAGAACTGAG ACCGAGGGCCAGAACTGGAGTAC GAGCAGCTCCT[C/T]GACATCAAGGT ccacctggaaaaaagaattgagaccTACTGTCTC CTTATCGGTGGAGAT	C	T	0.034	KRT25	keratin 25	
228	UNALM_VP A_4284518	AX- 417369038	16	ABRR03077 387.1	729913	AAGAGAAATCTCAAAGCAATTAATC TGATACACGAGAGAGAGATATGTGCC AAACACTCTC[T/C]GAGAAGAGAAG GCATTACGTATCTGTCCGAAAAAT GAGCGAGTCTAAGAAGAAAGTCTG	T	C	0.199	KRT25	keratin 25	
229	UNALM_VP A_4284519	AX- 417369399	16	ABRR03077 387.1	730005	AAAATGAGCGAGTCTAAGAAGAAA GTCTGTCTGTTTGCTGTCTTTCTGT TAACTGAAAT[A/G]TAATCTCTATCT AGGTAACCAACAGTGGTATCTTCCC ATTCATCTGGAAAAACATTACAC	A	G	0.034	KRT25	keratin 25	
230	UNALM_VP A_3426394	AX- 417369042	16	ABRR03077 387.1	753269	TAGAGTGCATCACTGCAAGCACAGA TCTCTACTGATGCGGGGGCAGCCAG TTCCGCCAAG[A/G]ATGAGATAACA GAACTGAAACGTACTCTGCAAGCCC TGGAATCGAGCTGCAGTCCCAAC	A	G	0.125	KRT24	keratin 24	
231	UNALM_VP A_4284535	AX- 417369401	16	ABRR03077 387.1	753864	atgaaatgtttattgcttttctcttagAAAAGCTCCC TGGAAGGGACCTTGCTGACA[C/T] GGAGGCTGGCTACATGGCTCAGCTG TCACAAATTCAGATACAGATCAGCA GCCTGGAGGA	C	T	0.02	KRT24	keratin 24	

232	UNALM_VP A_4284614	AX- 432742022	16	ABRR03077 387.1	754817	TGCATCAATCTAGCCATCTTTCTGGA TGACTTCAGGGAAGAGTTCCATCC AGAAATAGA[T/C]GACTGACCGATTT TTCTGCatcctctcttttcttattaGAATGCTCT TGAAAAAGTTGA	T	C	0.02	KRT24	keratin 24	
233	UNALM_VP A_4284615	AX- 432742367	16	ABRR03077 387.1	754885	CGATTTTCTGCaTcctctcttttcttattaGAA TGCTCTTGAAAAAGTTGAAATGATG[ A/G]CTTTGCTCTAATTGTTTCTATGC TTCAGTAACTTACTTTTGCAAGTTC ACTAAATGatt	A	G	0.105	KRT24	keratin 24	
234	UNALM_VP A_4284616	AX- 432742368	16	ABRR03077 387.1	755547	AGGGTGTTTCATAGTGAAAGGCTTGA TTCCACAACAGTCAAGAGAAATACT AGAGAAATCT[G/T]ACAAGTCATGTT TGCAGAACACTTAGGAGACTGggtcaa aattaaatcttgggAAAGAA	G	T	0.017	KRT24	keratin 24	
235	UNALM_VP A_4284617	AX- 432742027	16	ABRR03077 387.1	755988	GATCACATGTTAGATCATCAGTCCC ACAAGCTGGTACTCAAGAAAAGAA CATGGGACTTG[G/A]CCTAAAAACC CAAATTAGAAATCACAAAAGGATCA ACCTGAATATAATTCATGTAAATTG	G	A	0.128	KRT24	keratin 24	
236	UNALM_VP A_4284618	AX- 432742029	16	ABRR03077 387.1	756354	gtagtttttttccaggcaCATTAATAAAAC ATAAAAGTGAACTGTAGGCACACT [G/A]TAACCACAGAACGAAGTTAAT AGCTAAGCACAAACCACTATCCAAA GGAAAAAGAGCCTC	G	A	0.059	KRT24	keratin 24	
237	UNALM_VP A_4284619	AX- 432742031	16	ABRR03077 387.1	756707	CCAGTCCTGTACCACCTGGGACCTA TGTTTACGGCTAGTCTTACTTACCCT GATATCAAG[G/A]CAAAAGTATGTG GTTACTCTGAAAATCATCTTGAGGT GAAATCCTCATTCTGGATTTCCT	G	A	0.017	KRT24	keratin 24	
238	UNALM_VP A_4284520	AX- 417369402	16	ABRR03077 387.1	784695	TATGGTTGTATCCAaggtgggaaaaagaac aaaaacctaCCACAAGTAGAGTTGGTTTT [C/G]TTTTACCTTCAGGTGAGTTTCT	C	G	0.476	KRT222	keratin 222	

						GATATAAGTAACTAACTGGAAAAG TGAAGGGCCATT						
239	UNALM_VP A_4284521	AX- 417369048	16	ABRR03077 387.1	785739	TGTCTGACTTGATGTGAAGATGACA TTAAGATGGACCAAGGGTGGGCCAT GCTGACTATC[G/A]CCTGTCCCTAAA TGAAAGTTATTAAGTACACTTGGTA TGACCAACATACGagtcttca	G	A	0.084	KRT222	keratin 222	
240	UNALM_VP A_4284472	AX- 417299547	16	ABRR03077 387.1	208605	TTTGGGCGGAGGCTTTGGAGGTGGC CTTGGAGGTGGCTTTGGTGGGGGTT TTGGTGGTGG[C/A]TTTGGTGA CTTC GGTGGTGGCGATGGTGGCCTCCTCT CCGGCAATGAGAAGATCACCATG	C	A	0.084	LOC102 528247	keratin type I cytoskeletal 15	
241	UNALM_VP A_4282715	AX- 417274646	16	ABRR03077 387.1	134675	ctcACACTGGGTGGGGCCTGCTAAGC TCCCAGATCCCTTTTCCCCGCTCCC ATCTCTCC[C/T]ACAGtctctctccttgc cATCCTCCTCTGGCCGCCAGACCCGG CCATCCTGAAG	C	T	0.348		keratin 16*	
242	UNALM_VP A_4284559	AX- 432741885	12	ABRR03077 305.1	1973481	TTGACTGGCCAGCCACgagccccagtgcc agcacatataAAGGAGGGCACCTGCCAGT C[C/T]TCACTGCAACCTGCAGAACCG TGTGGGAGTTTGCCTTCGTCCCTCCT GCGTCTGAGCTCC	C	T	0.044	LOC102 533318	keratin type II cytoskeletal 71	
243	UNALM_VP A_4126407	AX- 432741932	12	ABRR03077 305.1	2222425	GCGGCTGGCAATATCATCATACTGA GCCTTGATCTCAGCCACAATGCTGT CCATGTCTAG[G/A]TCCCGGCTGTTG TCCATCTTGACAATGATGGAGGTGT CTGAGATGTGGGCTTGGATGACT	G	A	0.223	LOC102 546053	keratin type II cuticular Hb6	
244	UNALM_VP A_4284469	AX- 417280622	16	ABRR03077 387.1	165428	CAGCTTTGGCCAAGGATCTGCGGGT GGCTTTAGTGGTGGTAGTTTTGGTA GTAGCTTTGG[C/T]GGAGGCTTTGGG GGCTTTGGGGCTGGCTTTGGTGGCG GTGTTGGAGATGGTGGCATTCTG	C	T	0.054	KRT9	keratin 9	

245	UNALM_VP A_4284405	AX- 417368738	1	ABRR03000 004.1	12376109	GTTACGCACATCTGTCTGTCTGGCA ACATGTCCTACAACACTGCTCCACAAG GAATTGCTCT[T/G]CCAGGCTGATTG GGGGACAATACTCTGTCCCCGTGAC CCCTGTTGTCACAGCTTCTACCC	T	G	0.445	LOC102 544905	keratin-associated protein 11- 1
246	UNALM_VP A_4284406	AX- 417368627	1	ABRR03000 004.1	12376800	CCTGGCCAGCACCAATCTTATTTTA AGGGTTTGATCACCGGTGGCACGTA TACCTCTGGA[T/G]GTTTCCAGAAAT TTACCACCCACGCGCAAGTCTCTAA TGGTTTTGGCATGTCTTGACCTT	T	G	0.443	LOC102 544905	keratin-associated protein 11- 1
247	UNALM_VP A_4284522	AX- 417369247	1	ABRR03000 004.1	12376918	CTTTCTGCTTTgtctggcttctgcttctgttTGT GCCTGTGAAAAAGGGAACCTTGCTCT[ C/T]GCTCTGTGTTTCTCAATAAACCT TCATTACTTGGCATTGCAAATGTGT GTCTCAACAGAG	C	T	0.132	LOC102 544905	keratin-associated protein 11- 1
248	UNALM_VP A_4284523	AX- 417369248	1	ABRR03000 004.1	12420470	GAACAAGCCACCCAGGCATCAGATG CAGAACTGGCACTCAGCTACAGCCA AAGCCAGATG[C/T]GTTTCTCTGATA CACCCGAATCTTACCAATGATGCAA GTAGTTTGTCTCTTACAGTTGCT	C	T	0.02	LOC102 544636	keratin-associated protein 7-1
249	UNALM_VP A_4284539	AX- 432741842	1	ABRR03000 004.1	12436171	TTTTTCTAGTGATTTGCTGAAATCCC CAAGGAGGCATGGAATCTTCTCTCG TGAGGCTGA[A/G]GGGTGCACCTCC ACATCCTCATCTGCTCCTCAACCTCC CAGTCGCTGATTTTCAGAGCCA	A	G	0.162	LOC116 280298	keratin-associated protein 8-1
250	UNALM_VP A_4284540	AX- 432742273	1	ABRR03000 004.1	12616355	TGTTGATCACATATTTTCTACTAATT TACTACTAGCATACTTCCTGAATCTTT GTTTCATT[A/G]GACGATGTGAGACT CTTGTAAGCAAATGACTTTGTTTTT GAGTCAGCATTCCCTATGATC	A	G	0.003	LOC116 286138	keratin-associated protein 20- 2-like
251	UNALM_VP A_4284541	AX- 432742274	1	ABRR03000 004.1	12660986	AGCCAGAAGACCAGTATCTTCCATA GCAGCATGGGCGGTAGCAGGCATA GCCATAGCCAC[C/T]GTAGCCACATC	C	T	0	LOC116 276588	keratin-associated protein 20- 2-like

						CATAGCCACATCCCAGGCCACCATA GCCATAGCCCAGGCCACCATAGTA					
252	UNALM_VP A_72783	AX- 432742275	1	ABRR03000 004.1	12746893	CTGTGGCTATGGTTGCCTCCGAGGT CTGGGCTATGGCTACGGCGCTGGCT ATGGTGGCTA[T/C]GGATACGGCTGC TACCGCCCGTGTTACTACGGAAGAT ACTTGTCTCTGGCTTCTACTGA	T	C	0.03	LOC116 280314	keratin-associated protein 19- 3-like
253	UNALM_VP A_4284542	AX- 432741850	1	ABRR03000 004.1	12746965	CTACCGCCCGTGTTACTACGGAAGA TACTTGTCTCTGGCTTCTACTGAGA AACCCGCA[T/C]GCAACCTTCCTTC CACGTGACTGAGAACTTCCTTCTTG CTGATTCTACAGTCTGGCCACT	T	C	0.135	LOC116 280314	keratin-associated protein 19- 3-like
254	UNALM_VP A_4284543	AX- 432742277	1	ABRR03000 004.1	12804959	GATTTTAGGAATATTTATGATTTTCA AATTCTCTGCTTCATGTCTCTTATTT GTGCTTTC[A/G]TAACTGTGGCATT CCCTATTCTTTTGAATAAATTTCCC TAAGATAAAACAGCA <sub>Atacat</sub>	A	G	0.007	LOC116 276605	keratin-associated protein 19- 1-like
255	UNALM_VP A_4284524	AX- 417368745	1	ABRR03000 004.1	12844094	GTAACCTCCGAGAGAGCGGGAGGA GAAGTTTCCAAATCTGCGGTTGAAA GACATGCTGTC[A/C]GGAGTTCTCAG TTCAGCTGAGTTGCAGGCGGATGTT CTGAGTGTGAACGTTACCTCCTAC	A	C	0.007		keratin-associated protein 15- 1-like*
256	UNALM_VP A_4284407	AX- 417368747	1	ABRR03000 004.1	12853114	aaaagatggcTGTACTGAGCAGAGACTTT GATCTCAACTCAGAAGGATCTAG ACCCTC[A/G]GCAAGTTGAATAATA GAAACCAGAGCTAGAGATTGGTCTA TAACACGATGACTGGCAGCT	A	G	0.098	LOC107 034113	keratin-associated protein 10- 1-like
257	UNALM_VP A_72912	AX- 417368749	1	ABRR03000 004.1	12853197	GAGCTAGAGATTGGTCTATAACACG ATGACTGGCAGCTGCTGGAGGTA GCAGGTTGGA[C/G]GGCAGAATCTG GATCCACAGCTCAGGGAGGGGAAG CCACAGACTCCATAACCCAGAGGTC	C	G	0.02	LOC107 034113	keratin-associated protein 10- 1-like



258	UNALM_VP A_4284408	AX- 417368751	1	ABRR03000 004.1	12859761	TCTGTGGCTCCCCTTCACTGAGCTAT GGGTCCAATTTCTGGTACCCAATCA ACTTTCCTT[C/G]CAGAAGTTTCTGT TCATCTTGTTACTGGCCAATTTGTAG ATCTGGCTTCTACTGATTAAT	C	G	0.054	LOC107 034113 LOC102 532037	keratin-associated protein 10- 1-like, keratin-associated protein 13-1-like
259	UNALM_VP A_4284409	AX- 417368753	1	ABRR03000 004.1	12905040	AGGTCTAAGAGTATATATACTAG AAGGTCAGGTCGTGACACTGAAACT CAGTCTCCTC[A/G]GTGTAACCTCAGC TGAGCTCACACCTCTGTACCATG TCCTACAACCTGCTGCTCTGGAAA	A	G	0.024		keratin-associated protein 13- 2-like*
260	UNALM_VP A_73091	AX- 417368755	1	ABRR03000 004.1	12905326	GACCTCCACGCTCCGCAGTCCCTGC CAGACAACCTTATGCTGGGTCTCTGG GCTGCGGGTC[T/C]AGCAGAGGCTG CTCCCTGGGTTCTGGATCTAGAAGC TGCTGCTCCCTGGGCTATTGATAC	T	C	0.324		keratin-associated protein 13- 2-like*
261	UNALM_VP A_4284410	AX- 417369255	1	ABRR03000 004.1	12905564	CTTCCAGGAGCTGCCAGCCTTCTTG CTATGGGCCAGTTTATAGATCAGCT TTCTGTAAAT[C/T]AGCTTGTTGAGT TTCTGGACCTTCTGagcaaatgttcagtctc TTCTTAGAACTGCTA	C	T	0.064		keratin-associated protein 13- 2-like*
262	UNALM_VP A_4284411	AX- 417369256	1	ABRR03000 004.1	12923762	ATCTCCTGTCACCATGTCCCACATCT GCTGCTCTGGAaacttctctctgctccttgg[ G/A]GGCCACTTGCACTACCCAAGCT CTTCCTGTGGCTCCTCCTACCCAGC AACCTGATCTAC	G	A	0.044	LOC102 533350	keratin-associated protein 13- 1-like
263	UNALM_VP A_4284412	AX- 417368761	1	ABRR03000 004.1	12923825	CCACTTGCACTACCCAAGCTCTTCTT GTGGCTCCTCCTACCCAGCAACCT GATCTACAC[T/C]ACGGAcctetgctctccc agcacctgccagCTGAGCTCCTCTCTGTAC AGTGGCTGTCAG	T	C	0.375	LOC102 533350	keratin-associated protein 13- 1-like
264	UNALM_VP A_4284413	AX- 417368652	1	ABRR03000 004.1	12924045	CCCTGCTGGACAAATTATGCTGGGT CTCTGGGCTCTGGGTCCGGCAAAGG CTACTTCTTG[G/T]TCTATGGATCCA	G	T	0.024	LOC102 533350	keratin-associated protein 13- 1-like

						GAAGCTGTGGGTCCAGTGGATCTCTGGGTTCTGGATTCTTTGGCTTCC					
265	UNALM_VP A_73135	AX- 432742278	1	ABRR03000 004.1	12924160	GCTTCCCCCTCCCTGAGCTGTGGATCC TGCCATCCAATTTATGTGGTTTCTTG AAGCTGCC[A/G]AACTTCTTGTTACC AGCCAACCTGTAGATCTGGCCTCTT CTACTGATCACCTTAAGGGTT	A	G	0.014	LOC102 533350	keratin-associated protein 13- 1-like
266	UNALM_VP A_4284544	AX- 432742279	1	ABRR03000 004.1	12930908	AAAAGAGTCTTGAGAAGAAGCCAA CACCCAGCTTTCTGAGTATATAATCT CCTCGGTAAC[G/A]AATTTAATCTTC AGACCCAGCGTGTGGGTCCTTGTG ATCAACTTCAATCAAGGGTTTTA	G	A	0.341	LOC102 533595	keratin-associated protein 27- 1
267	UNALM_VP A_4284414	AX- 417368763	1	ABRR03000 004.1	12979560	AGGCTATTCTGGGGATTGCAGGGGC ATATGCTACAGAACCCACTGTATTA TCCCAGTGAC[T/G]CCTTCTGTTGCT CTTTGCTCCAGCGATGTAAGCCCTC CCTGGGTCTTGGATTGCCTAGT	T	G	0.186	LOC102 532787	keratin-associated protein 24- 1
268	UNALM_VP A_4284525	AX- 417368766	1	ABRR03000 004.1	12979659	CTTGGGTCTTGGATTGCCTAGTTGTT ACCAAGGAACCTCTGGCTTCTGGA TAACTGCCA[A/T]GAAACCTGTGGTG AAGCACCAATCTGTGAATCTCCCAG CTGTGAGCCCAGGACCTGCACC	A	T	0.027	LOC102 532787	keratin-associated protein 24- 1
269	UNALM_VP A_73251	AX- 417369260	1	ABRR03000 004.1	12979802	CTGGTGTGCCCTGCAGCTCTGCAGC AGTGGGCAAATAACGCAGTGCCTGT GAAACTACCA[A/G]TGTCGGACCCA GCCCCAGCTGCAACCCATGCACTCA GACCAAAGGGTATGTATCCGATTG	A	G	0.213	LOC102 532787	keratin-associated protein 24- 1
270	UNALM_VP A_4284536	AX- 432741830	1	ABRR03000 004.1	1556945	CTGAGCAAATGGCGGGTGTCTGTCA GGAATGGCCCTGCCCTCGCACCCCT GCTGTGCTTG[G/T]TCAACTGCTGGG GGTGGCCTGGCGTGGATGCCGCGGT CCATCCAAGAGTGGCAGCAGGAA	G	T	0	TSPEA R LOC107 033590	thrombospondin type laminin G domain and EAR repeats, keratin-associated protein 10- 8-like

271	UNALM_VP A_77106	AX- 432741832	1	ABRR03000 004.1	1557010	CTGCTGGGGGTGGCCTGGCGTGGAT GCCGCGTCCATCCAAGAGTGGCAG CAGGAACCCA[T/C]TGCGGGTTCTCA TGGGGGTGGCCTGCCTGGCCAGGAT GGCCCCAGCATGGAGGAGCAAAG	T	C	0.419	TSPEA R LOC107 033590	thrombospondin type laminin G domain and EAR repeats, keratin-associated protein 10- 8-like
272	UNALM_VP A_4284537	AX- 432742268	1	ABRR03000 004.1	1557431	ccccagccctgcegcCACCGTACTCAGA GCAGTAGGGATGACAGATGAAGAA GCTTA[T/C]GAAGTCGAGAAGTAAG GTCATCAGGCCAGGTAGGCAGGTCC TTCCCCGGGAGCAGAGAAG	T	C	0.426	TSPEA R LOC107 033590	thrombospondin type laminin G domain and EAR repeats, keratin-associated protein 10- 8-like
273	UNALM_VP A_77252	AX- 432742269	1	ABRR03000 004.1	1568912	CCATGGGTCCAGAGGCAGTTTTCTT GAGATGGACCCAGGGCTGCGTGCT GGAGCTCAGG[G/C]GGCTACAGGCC TCAGAGTGGGCCTGGCCTGTGGGGG TCACAGAGGCCAGCACCAGGCGTC	G	C	0.434	TSPEA R LOC107 033590	thrombospondin type laminin G domain and EAR repeats, keratin-associated protein 10- 8-like
274	UNALM_VP A_77253	AX- 432742270	1	ABRR03000 004.1	1568981	GCCTCAGAGTGGGCCTGGCCTGTGG GGGTCACAGAGGCCAGCACCAGGC GTCCCATTAC[C/T]CTGCCACCGTG AATCCACAGACAGAACAGGTCGTTG AGATGGAGGTGGCTTTATTAGGAC	C	T	0.034	TSPEA R LOC107 033590	thrombospondin type laminin G domain and EAR repeats, keratin-associated protein 10- 8-like
275	UNALM_VP A_4284430	AX- 417369282	na	ABRR03001 991.1	1519	CTTAGGGCTGGGCTGGCGTCGGGAG GAGGGCGGTGATGTCTGCAGCCTCC CTGCCCCGGG[G/T]GCCTTTATACCC GGGCCGTGGGCGTCCCAGCAGCACA GAAGCTCACCTGCTGACTTCCTC	G	T	0.331	LOC102 530622	keratin-associated protein 12- 2-like
276	UNALM_VP A_4284546	AX- 432742374	na	ABRR03001 991.1	807	acaatactcaataaaaaataagaattaaaaaatggag ggacACACTCCATTGAGAA[A/T]GCAT GATTTATTTGTACACATGGGATGT GGCTTCTGCCGGAGACAGAGAGGG ACCCTG	A	T	0.31	LOC102 530622	keratin-associated protein 12- 2-like
277	UNALM_VP A_4284547	AX- 432741840	na	ABRR03001 991.1	902	CCTGCCGGAGACAGAGAGGGACCCT GGGAACCAGGGCTTCCAGAGGAGA TGCGCAGATTC[A/G]TAGAAAGATC	A	G	0.314	LOC102 530622	keratin-associated protein 12- 2-like

						CAGATCGTCCTGCCACGCTCTTGCC CCAGAGTTCAGAGGGTTGGTTCATC						
278	UNALM_VP A_4284581	AX- 432742033	18	ABRR03077 307.1	8124973	GCCTCAACTGTCTCTGCTTTGAGATC AAGCTCCGATGATGACCCAGTGTCC CTGCCCCTT[G/T]GGGGAGCGGCCAC CCCCGGGGCCACATGCCCTGTTC CCTGAGGGGCCTCAGGACACCT	G	T	0	LOC116 284217	keratin-associated protein 10- 10	
279	UNALM_VP A_4284480	AX- 417329142	16	ABRR03077 387.1	354957	TCTGTGAGCCTGCTTGCTGTCAGCC AGTGTGCCCCACGCCTAGCTGCTGT TCATCTGTCC[G/A]CTCCGTGGCCAA TGGCCGCCAGTCTGTCTGCTGTGAG CCACCTTGTGAGCCACATTGTGA	G	A	0.007	LOC102 544939	keratin-associated protein 16- 1	
280	UNALM_VP A_3425413	AX- 417329721	16	ABRR03077 387.1	355358	CCGTCTACCTCGTGCCGACCTCTCTG CTGCCGCCAGGGTCTTCTGCATCT GTCATCTGC[C/A]GGCCAATTTGCTC TCGAACTTTCTACATACCCAGCTCCT GCAAACAACCATGCACTCCTT	C	A	0.115	LOC102 544939	keratin-associated protein 16- 1	
281	UNALM_VP A_3425425	AX- 417329829	16	ABRR03077 387.1	355598	cGCCGCCAGCCTGCATCACTTCTGT CTCTTACCGTCCCATCTGCACCCGCC TGTCTTGT[A/G]CTGACACCTGCAAA CGTGATTGCAAAAAGTCTACTTCCA GCCAACCGGATTGTGCTGACT	A	G	0.264	LOC102 544939	keratin-associated protein 16- 1	
282	UNALM_VP A_4283681	AX- 417332025	16	ABRR03077 387.1	355869	CTGCAGTGCACGCCTCTGACTACAG AGAAAATGTGACTGCCCTCCCCAAA GCTCATCTGA[C/T]ttagaatcttttcttgcac tATCACTCACCATCTGCTTATGCTTC AAAGAACTCAC	C	T	0.262	LOC102 544939	keratin-associated protein 16- 1	
283	UNALM_VP A_4284481	AX- 417334740	16	ABRR03077 387.1	361421	CTGCCAATCAACTTATTACCAACCT GTCTGCTACTTTTTTAAGTCTTGTC ATCAGTTCC[C/T]TGCGTGCCTGTGC CCCACCAGCTGTGTCcttgtgtttctgttcttg CAATCCTGCTTGC	C	T	0.142	KRTAP 29-1	keratin associated like protein 29-1	

284	UNALM_VP A_4284482	AX- 417339267	16	ABRR03077 387.1	361604	CCAGCCAGTGGCTAACCCTTGTCT GTAAAGAACCCTTGCAAACCAGCTT CCTGCAGCAC[C/T]GTCCTTCTGGC CAACCAACTTGTGGTGAACCTACTT CCTGCAATCAAAGTGCCTGCAA	C	T	0.142	KRTAP 29-1	keratin associated like protein 29-1
285	UNALM_VP A_4284592	AX- 432742344	16	ABRR03077 387.1	361748	CTGTGTGACAGGTTCTGGCAAATCA TCCAGTGGAGGTTCCAATCGCTTCC GAACCACTGC[T/C]CCAAGTCTGCCA GGCCAGCACCTGCTAGCCAACTTCC TGCCAACCCAGCCAGGAGTCCAG	T	C	0.429	KRTAP 29-1	keratin associated like protein 29-1
286	UNALM_VP A_4284483	AX- 417343124	16	ABRR03077 387.1	379872	CCACGTGCTGCCGCCCTAGCTGTGT GTCCAGCTGCTGCCAGCCCTCCTGC TGATCACCTC[A/T]CCAAGAGCCATC CCCTGCATCCAACACAATCTGTCAA CTGAGTTGCCGTTTTGGGGGCAA	A	T	0	LOC102 544126	keratin-associated protein 9- 2-like
287	UNALM_VP A_4284484	AX- 417342892	16	ABRR03077 387.1	400197	GCCTGTGTGCTGCCAACCTACCTGC TGTCGCCCCAGGTGCTGCAACTCCA GCTGCTGCCG[G/C]CCCATCTGCTGA GGATCCTCTTTTTGCTGAACTTCATT CCTCACCACCAGCCCTGAGCCA	G	C	0.405	LOC102 543867 LOC116 283835	keratin-associated protein 4- 1-like, uncharacterized LOC116283835
288	UNALM_VP A_4284485	AX- 417344581	16	ABRR03077 387.1	403985	ATTCTAacttttccatgttttaaatgaCCACCAAG TTTATTATCCTGTGACACTACCAAG[ G/C]CACTGGAGAAAGAGCCTCCATC TGCGTTCAGTACCACCCCTGCTTCCC AGAAATCTCTTT	G	C	0.49	LOC102 543867 LOC116 283835	keratin-associated protein 4- 1-like, uncharacterized LOC116283835
289	UNALM_VP A_4284534	AX- 417343990	16	ABRR03077 387.1	404135	GTTGGATCAAACCCCAAGTCTCTG GCTGTTACTTGCCTTGACCTTCAT AACTGCATG[T/C]TAACTATTTCTCA ATAATACTATCTTGATATCATAAATT TGIGTATCCTTCTTTACTTAT	T	C	0.061	LOC102 543867 LOC116 283835	keratin-associated protein 4- 1-like, uncharacterized LOC116283835
290	UNALM_VP A_4284486	AX- 417347426	16	ABRR03077 387.1	450514	ATAAAAGGGGACAGGTGTAAGGAG ACATTCAGACTCAAGAAACCACCCC TGTGGAAACCC[T/A]CTCGTGAACC	T	A	0.466	LOC102 529530	keratin-associated protein 4- 7-like

						CCAATCCTTGCAATTAACACCATGGT CAGCTCCTGCTGTGGCTCTGTCTG					
291	UNALM_VP A_4284487	AX- 417350058	16	ABRR03077 387.1	460798	GTAGACCCTCCTACTGTGCTGACCC TTAGGATACATGGAGCCTGGCTAAC ACATTCAGTG[G/T]CTTGGACCCATG ATTCCAATGAGTCCATTACTATCCC ACTGACTCTAGGAGCACATTCTG	G	T	0.101	LOC102 529530	keratin-associated protein 4- 7-like
292	UNALM_VP A_4284488	AX- 417356793	16	ABRR03077 387.1	461108	AGACCGGGTCCAGTTATACCCTGT ATTTTCTCCTAGAGTGAATTTCTTAT GTTTTGTTG[C/T]TTCTCTGCTTTCCA ATAAACTTTTCTGCACATAAGAATT CATTGGCAttgtttcatgat	C	T	0.101		keratin-associated protein 4- 7A-like*
293	UNALM_VP A_4284594	AX- 432741984	16	ABRR03077 387.1	497643	TGTAGAAGGTGAAGCCAAAACCTCAG AAACTTCTCTTAACAAACCAACTCT CAGCCCAACT[C/T]TTGATACCATGG CCTGCTGTTCCACCAGCTTCAACGG ATTCCCATCTGTTCTACTGGTG	C	T	0.118	LOC102 529268	keratin-associated protein 4- 3-like
294	UNALM_VP A_4284595	AX- 432741986	16	ABRR03077 387.1	501126	TTCATCCTGCAGTGAATATCTTGACT TTCCACTGTGGACACAGAAATGTTG CAAACCCAC[C/A]TGCTGATCATCAA ATTGCTCTGGATATACTATTTCTGAC TTTCTTGCCAGGTTCAAAT	C	A	0.274	LOC102 529268	keratin-associated protein 4- 3-like
295	UNALM_VP A_4284489	AX- 417358599	16	ABRR03077 387.1	515060	CTGTGAAAGTCCCTGCGAGCCCTGC TGTTAAGCAGCCACATCTTTGCCCT GGTTCAGTGA[C/T]GGGCTGCTCAAC GTCTTATGCATCCTCTGGCCTTCCAC ACGCAACTACACGTCGACGACT	C	T	0.068	LOC102 528725	keratin-associated protein 3-1
296	UNALM_VP A_4284490	AX- 417358985	16	ABRR03077 387.1	526383	CCCCGCCACCACCATCTGCTCCTTTG ATAAGTCCTGCCGGTGTGGAGTCTG CCTGCCAG[T/C]ACCTGCCACACA CGGTTTGGTTACTGGAGCCAACCTG CTGTGACAActgccccacc	T	C	0.387	LOC116 283838 LOC102 528209 LOC102 528468	uncharacterized LOC116283838, keratin- associated protein 3-3-like, keratin-associated protein 3- 3-like

297	UNALM_VP A_4284491	AX- 417363068	16	ABRR03077 387.1	532094	aaaccaaaccAATAAAAAATCACAGATT CCCAAGGCCATGGCTTGCTGTGTTT CCTGCGA[C/T]GGCTGCAGTGTTCGC ACCGGCCCCGCCACCACCATCTGCT CCTTTGATAAGTCCTGCCGG	C	T	0.007	LOC116 283838 LOC102 528209	uncharacterized LOC116283838, keratin- associated protein 3-3-like
298	UNALM_VP A_4284492	AX- 417367590	16	ABRR03077 387.1	532200	GATAAGTCCTGCCGGTGTGGAGTCT GCCTGCCCAGCACCTGCCACACAC GGTTTGGTTA[C/T]TGGAGCCAACCT GCTGTGACAActgccccccacctgccacaTT CCTCAGCCCTGTGTGC	C	T	0.062	LOC116 283838 LOC102 528209	uncharacterized LOC116283838, keratin- associated protein 3-3-like
299	UNALM_VP A_4284596	AX- 432742348	16	ABRR03077 387.1	532631	ATAGGTGGATGCCTACAGGCTTCCC AAAGTTCAATTCTTTCATATTGAAGT GTATCTTTC[T/C]TTGGGTGCTTTGG GAATTCCTGGGTGGTATCTTCTGGA AACTGAGGAGCTTCCTCATGA	T	C	0.294	LOC102 528209	keratin-associated protein 3- 3-like
300	UNALM_VP A_4284597	AX- 432741990	16	ABRR03077 387.1	532713	GGGTGGTATCTTTCTGGAAACTGAG GAGCTTCCTCATGATTATTCTAATAA ATTCTACAT[C/T]TCTGGCATAGCAC AAATGTTTGCAAttactctgtttattctgtaattC ATTAAATGT	C	T	0.151	LOC102 528209	keratin-associated protein 3- 3-like
301	UNALM_VP A_4284415	AX- 417368775	na	ABRR03000 014.1	126095	CTGGGCCTCGGATTCTTCCTCATCGT CTTCAatcatctctccctcctccteggCCGTGGC[ A/G]TCCTGGTACTGCTGGTACTCGG ACACCAGGTCGTTTCATGTTGCTCTCC GCTTCGGTGAAC	A	G	0.092	LOC102 529783	tubulin beta-3 chain
302	UNALM_VP A_4284432	AX- 417369285	4	ABRR03002 523.1	8506962	TCTCTGACATATTTTAGATATATCCA TATTTCCATTAGAAAATGCCCTATT GGACATAA[C/T]AGACAATACAATA TGGTGCCATTCTGGCCTCCAGTTACC AACATGGAAATGTTTGTTACT	C	T	0.436	TYRP1	tyrosinase related protein 1

**ANEXO 4:** Comparación de PNS entre el genotipado por GBS generados a partir de BRR con ApeK1 y el genotipado por micromatriz. La concordancia media entre ambos métodos de genotipado fue de 0.93

<b>Muestra</b>	<b>Número de PNSs discordantes</b>	<b>Número de PNSs concordantes</b>	<b>Número de PNSs comparados</b>	<b>Tasa de concordancia</b>
B95064	726	7981	8707	0.92
B95065	759	9003	9762	0.92
B95066	669	7449	8118	0.92
B95067	864	10154	11018	0.92
B95068	777	9815	10592	0.93
B95069	861	8164	9025	0.90
B95070	1105	13522	14627	0.92
B95071	1124	16710	17834	0.94
B95072	1187	16666	17853	0.93
B95073	1221	20882	22103	0.94
B95074	952	12993	13945	0.93
B95075	1335	15374	16709	0.92
B95076	764	7944	8708	0.91
B95077	1407	16887	18294	0.92
B95078	1017	12756	13773	0.93
B95079	1058	17289	18347	0.94
B95080	1329	19239	20568	0.94
B95081	1399	16386	17785	0.92
B95082	1131	17284	18415	0.94
B95083	1158	19296	20454	0.94
B95084	1251	21146	22397	0.94
B95085	1375	25469	26844	0.95
B95086	1279	22618	23897	0.95
B95087	1425	23682	25107	0.94
B95088	1170	19180	20350	0.94
B95089	920	11294	12214	0.92
B95090	1015	14950	15965	0.94
B95091	1369	17862	19231	0.93
B95092	1214	17368	18582	0.93
B95093	1248	20632	21880	0.94
B95094	990	12973	13963	0.93
B95095	1532	21323	22855	0.93
B95096	1168	18220	19388	0.94



B95097	1007	12736	13743	0.93
B95098	1562	20160	21722	0.93
B95099	1304	19310	20614	0.94
B95100	1017	15626	16643	0.94
B95101	1172	18396	19568	0.94
B95102	1137	12868	14005	0.92
B95103	1167	18344	19511	0.94
B95104	1163	18230	19393	0.94
B95105	1512	20113	21625	0.93
B95106	982	12232	13214	0.93
B95107	997	12811	13808	0.93
B95108	1180	21903	23083	0.95
B95109	1285	20291	21576	0.94
B95110	1070	18868	19938	0.95
B95111	1346	20549	21895	0.94
B95112	1034	15129	16163	0.94
B95113	844	12056	12900	0.93
B95114	1361	18973	20334	0.93
B95115	1206	15357	16563	0.93
B95116	1211	19342	20553	0.94
B95117	1175	16855	18030	0.93
B95118	1232	19573	20805	0.94
B95119	1089	13240	14329	0.92
B95120	1009	14324	15333	0.93
B95121	1201	21716	22917	0.95
B95122	1006	14574	15580	0.94
B95123	1187	17365	18552	0.94
B95124	1250	21946	23196	0.95
B95125	1092	14537	15629	0.93
B95126	1393	19372	20765	0.93
B95127	957	11094	12051	0.92
B95128	919	11976	12895	0.93
B95129	766	7970	8736	0.91
B95130	1042	14383	15425	0.93
B95131	1062	17427	18489	0.94
B95132	921	12765	13686	0.93
B95133	1075	17194	18269	0.94
B95134	1032	15180	16212	0.94
B95135	706	8536	9242	0.92
B95136	1117	16827	17944	0.94
B95137	1272	21373	22645	0.94

B95138	1039	14202	15241	0.93
B95139	910	12271	13181	0.93
B95140	1184	18774	19958	0.94
B95141	927	11833	12760	0.93
B95142	1231	17577	18808	0.93
B95143	933	12070	13003	0.93
B95144	1026	14717	15743	0.93
B95145	1117	15580	16697	0.93
B95146	1262	19301	20563	0.94
B95147	836	9038	9874	0.92
B95148	1074	15596	16670	0.94
B95149	890	11741	12631	0.93
B95150	1231	14758	15989	0.92
B95151	975	15374	16349	0.94
B95152	900	13128	14028	0.94
B95153	1068	15437	16505	0.94
B95154	1175	18694	19869	0.94
B95155	1160	19584	20744	0.94
B95156	1048	17190	18238	0.94
B95157	1085	13951	15036	0.93
B95158	1193	20585	21778	0.95
B95159	877	10555	11432	0.92
B95161	1002	13550	14552	0.93
B95162	1220	20227	21447	0.94
B95163	1122	17211	18333	0.94
B95164	1110	16591	17701	0.94
B95165	1245	16074	17319	0.93
B95166	1216	20370	21586	0.94
B95168	1285	21837	23122	0.94
B95169	531	6002	6533	0.92
B95170	1078	15230	16308	0.93
B95171	1362	16596	17958	0.92
B95172	899	10979	11878	0.92
B95173	1034	15292	16326	0.94
B95174	1081	16512	17593	0.94
B95175	1058	16152	17210	0.94
B95176	1302	20797	22099	0.94
B95177	1486	20874	22360	0.93
B95178	1007	15056	16063	0.94
B95179	1184	17779	18963	0.94
B95180	980	13716	14696	0.93

B95182	866	11773	12639	0.93
B95183	1304	20871	22175	0.94
B95184	1586	22011	23597	0.93
B95185	1262	21919	23181	0.95
B95186	1373	17647	19020	0.93
B95187	837	10851	11688	0.93
B95189	1487	21968	23455	0.94
B95190	1152	16285	17437	0.93
B95191	1580	21729	23309	0.93
B95192	1185	15525	16710	0.93
B95193	1261	20144	21405	0.94
B95194	1198	14587	15785	0.92
B95195	1048	16544	17592	0.94
B95196	1322	23435	24757	0.95
B95197	938	12056	12994	0.93
B95198	1335	22651	23986	0.94
B95200	1261	22734	23995	0.95
B95201	1556	22404	23960	0.94
B95202	1217	21428	22645	0.95
B95203	1221	21421	22642	0.95
B95204	1082	17005	18087	0.94
B95205	1212	21398	22610	0.95
B95206	1214	20056	21270	0.94
B95207	1358	23834	25192	0.95
B95208	1252	23057	24309	0.95
B95209	1157	21891	23048	0.95
B95210	1167	17149	18316	0.94
B95211	1052	15327	16379	0.94
B95212	1074	16966	18040	0.94
B95213	971	13549	14520	0.93

**ANEXO 5:** Comparación de PNS entre el genotipado por GBS generados a partir de BRR con Pst1-Msp1 y el genotipado por micromatriz. La concordancia media entre ambos métodos de genotipado fue de 0.94

<b>Muestra</b>	<b>Número de PNSs discordantes</b>	<b>Número de PNSs concordantes</b>	<b>Número de PNSs comparados</b>	<b>Tasa de concordancia</b>
B95064	447	5943	6390	0.93
B95065	397	5516	5913	0.93
B95066	498	8232	8730	0.94
B95067	438	6271	6709	0.93
B95068	336	4786	5122	0.93
B95069	449	4512	4961	0.91
B95070	580	8620	9200	0.94
B95071	353	5745	6098	0.94
B95072	425	5510	5935	0.93
B95073	124	952	1076	0.88
B95074	399	6185	6584	0.94
B95075	548	7177	7725	0.93
B95076	374	4817	5191	0.93
B95077	577	7271	7848	0.93
B95078	495	7206	7701	0.94
B95079	498	8522	9020	0.94
B95080	543	8121	8664	0.94
B95081	593	8216	8809	0.93
B95082	528	9655	10183	0.95
B95083	486	8633	9119	0.95
B95084	402	5748	6150	0.93
B95085	502	8248	8750	0.94
B95086	476	7744	8220	0.94
B95087	465	6961	7426	0.94
B95088	342	5164	5506	0.94
B95089	445	6715	7160	0.94
B95090	316	4535	4851	0.93
B95091	533	7087	7620	0.93
B95092	375	4236	4611	0.92
B95093	548	11002	11550	0.95
B95094	545	9699	10244	0.95
B95095	520	7054	7574	0.93
B95096	490	8143	8633	0.94

B95097	437	6913	7350	0.94
B95098	513	7229	7742	0.93
B95099	434	6008	6442	0.93
B95100	154	1340	1494	0.90
B95101	425	5551	5976	0.93
B95102	294	2816	3110	0.91
B95103	438	6364	6802	0.94
B95104	497	9284	9781	0.95
B95105	602	8676	9278	0.94
B95106	571	10935	11506	0.95
B95107	435	5877	6312	0.93
B95108	500	8729	9229	0.95
B95109	335	3901	4236	0.92
B95110	466	8893	9359	0.95
B95111	573	10300	10873	0.95
B95112	436	6206	6642	0.93
B95113	354	4847	5201	0.93
B95114	564	7887	8451	0.93
B95115	428	5137	5565	0.92
B95116	457	6904	7361	0.94
B95117	538	8177	8715	0.94
B95118	399	5366	5765	0.93
B95119	220	2437	2657	0.92
B95120	369	4627	4996	0.93
B95121	484	8323	8807	0.95
B95122	499	8781	9280	0.95
B95123	494	6388	6882	0.93
B95124	541	9627	10168	0.95
B95125	619	10285	10904	0.94
B95126	562	8412	8974	0.94
B95127	495	7644	8139	0.94
B95128	282	3033	3315	0.91
B95129	310	3179	3489	0.91
B95130	463	6920	7383	0.94
B95131	429	6497	6926	0.94
B95132	451	6520	6971	0.94
B95133	527	9049	9576	0.95
B95134	538	9434	9972	0.95
B95135	475	7799	8274	0.94
B95136	514	8710	9224	0.94
B95137	622	8478	9100	0.93

B95138	522	9517	10039	0.95
B95139	534	10194	10728	0.95
B95140	597	9603	10200	0.94
B95141	539	9443	9982	0.95
B95142	535	9152	9687	0.94
B95143	397	5592	5989	0.93
B95144	370	4757	5127	0.93
B95145	532	9365	9897	0.95
B95146	401	5508	5909	0.93
B95147	478	6355	6833	0.93
B95148	518	10494	11012	0.95
B95149	552	10119	10671	0.95
B95150	678	10279	10957	0.94
B95151	431	9074	9505	0.95
B95152	394	6337	6731	0.94
B95153	516	9730	10246	0.95
B95154	416	6143	6559	0.94
B95155	381	6401	6782	0.94
B95156	494	8502	8996	0.95
B95157	519	8945	9464	0.95
B95158	553	9964	10517	0.95
B95159	564	10123	10687	0.95
B95161	531	10229	10760	0.95
B95162	596	11011	11607	0.95
B95163	612	10564	11176	0.95
B95164	573	10341	10914	0.95
B95165	647	10791	11438	0.94
B95166	568	11313	11881	0.95
B95168	560	11115	11675	0.95
B95169	453	8483	8936	0.95
B95170	567	11173	11740	0.95
B95171	709	10952	11661	0.94
B95172	597	10353	10950	0.95
B95173	508	9077	9585	0.95
B95174	426	7235	7661	0.94
B95175	560	11068	11628	0.95
B95176	601	11012	11613	0.95
B95177	627	11031	11658	0.95
B95178	500	9832	10332	0.95
B95179	633	10851	11484	0.94
B95180	607	10250	10857	0.94

B95182	513	9943	10456	0.95
B95183	628	11070	11698	0.95
B95184	658	10214	10872	0.94
B95185	548	10549	11097	0.95
B95186	730	10669	11399	0.94
B95187	547	10522	11069	0.95
B95189	552	8809	9361	0.94
B95190	618	10953	11571	0.95
B95191	660	10188	10848	0.94
B95192	619	10847	11466	0.95
B95193	548	10511	11059	0.95
B95194	611	10268	10879	0.94
B95195	576	10208	10784	0.95
B95196	542	10554	11096	0.95
B95197	546	10664	11210	0.95
B95198	571	10931	11502	0.95
B95200	587	10993	11580	0.95
B95201	644	9273	9917	0.94
B95202	555	10668	11223	0.95
B95203	554	10130	10684	0.95
B95204	493	8333	8826	0.94
B95205	533	11277	11810	0.95
B95206	580	10884	11464	0.95
B95207	586	10545	11131	0.95
B95208	586	11069	11655	0.95
B95209	517	10655	11172	0.95
B95210	526	10368	10894	0.95
B95211	562	10397	10959	0.95
B95212	511	9556	10067	0.95
B95213	491	9545	10036	0.95

**ANEXO 6: Valores de distribución (Autovectores) de la estructura genética de la población genotipado por micromatriz.**

<b>ID</b>	<b>1er Autovector</b>	<b>2do Autovector</b>
P-24	-0.144588685	0.263852054
P-21	-0.120418236	0.244461864
P-44	-0.110271028	0.209567927
P-05	-0.098375709	0.215389158
P-09	-0.102895994	0.19395348
P-06	-0.108021505	0.186880037
P-56	-0.101886964	0.188032109
P-07	-0.108695378	0.169035494
P-48	-0.104916916	0.169852061
P-15	-0.093904827	0.164830026
P-32	-0.099591083	0.153719706
P-16	-0.074723984	0.14456352
P-54	-0.063497962	0.122458049
P-30	-0.065538577	0.116126624
P-51	-0.056076715	0.124802332
P-46	-0.061102619	0.116429728
P-02	-0.059183855	0.116159569
P-36	-0.061626467	0.110706562
P-12	-0.051753356	0.105546406
P-39	-0.06462755	0.08364269
P-43	-0.048006115	0.091158768
P-31	-0.053971025	0.081064359
P-40	-0.052123552	0.072564493
P-25	-0.045395598	0.072492892
P-58	-0.03620331	0.06940152
P-11	-0.031944968	0.054808313
P-19	-0.039631765	0.045993658
P-52	-0.044761443	0.032621938
P-59	-0.040710432	0.0118317
P-49	-0.018637855	0.032638954
Q-22	-0.038433262	-0.001753015
P-13	-0.011391511	0.010853339
P-17	-0.029619634	-0.012308159
P-04	-0.014954871	0.000152841
G-14	0.001387897	0.00841817
Q-45	-0.028503068	-0.02172829
G-09	-0.015228557	-0.010182525
P-26	-0.001642094	0.002130745



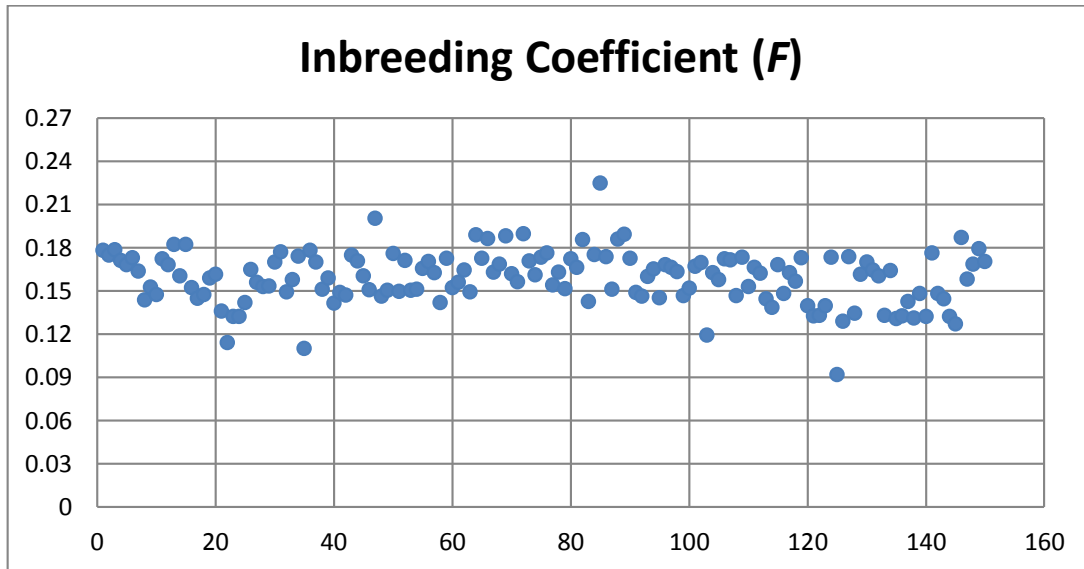
P-27	-0.015980073	-0.014248025
P-28	-0.007975022	-0.008420577
R-24	-0.016169908	-0.016660628
Q-32	-0.07470775	-0.080092738
Q-25	-0.020876086	-0.027679813
Q-26	-0.024532589	-0.031623009
Q-15	-0.023877123	-0.031378778
Q-43	-0.008852539	-0.017705914
P-41	-0.020011117	-0.028904152
P-14	-0.024042166	-0.034085284
Q-44	-0.019020625	-0.032051388
P-29	-0.013322322	-0.027976198
Q-05	0.004824673	-0.011607603
Q-30	-0.00747334	-0.024719681
Q-27	-0.027157432	-0.04527691
Q-01	-0.016728203	-0.034855663
Q-12	-0.01948607	-0.037827433
Q-49	-0.025153799	-0.044122469
G-01	-0.010854696	-0.031755278
Q-46	-0.013943368	-0.035701353
Q-39	-0.022124854	-0.044247663
Q-23	-0.011991747	-0.035450668
Q-41	-0.003472738	-0.027063513
Q-50	-0.018532876	-0.043487655
Q-29	-0.013054126	-0.038159176
P-33	-0.012128965	-0.037492178
Q-06	-0.01140299	-0.037141664
Q-36	-0.000857378	-0.026597992
P-10	0.002497834	-0.023512905
Q-28	-0.022846174	-0.050498121
R-12	0.001194524	-0.027492352
Q-16	0.001341231	-0.027823735
P-38	-0.010017851	-0.039751367
Q-38	-0.005252527	-0.035505707
P-08	0.00836266	-0.022016324
P-47	-0.005003958	-0.036085958
Q-42	-0.02452007	-0.056160156
G-12	0.005445926	-0.027022806
Q-20	-0.009458206	-0.041994746
Q-04	-0.015139091	-0.049018091
Q-08	-0.018504622	-0.053654066
G-25	0.02355122	-0.011623424

Q-10	-0.015526574	-0.050892667
Q-13	-0.008701146	-0.045594503
R-30	0.000384142	-0.037812812
R-28	0.005978418	-0.032906807
Q-34	-0.023749944	-0.062728038
Q-14	-0.032624367	-0.071768269
Q-11	-0.048577459	-0.0878386
P-01	-0.003463047	-0.043590373
Q-19	-0.028487232	-0.069131768
P-50	-0.00152318	-0.044439401
P-55	0.002095635	-0.041570777
Q-07	-0.008244367	-0.052139689
P-22	-0.00444799	-0.049715227
R-29	0.006369831	-0.039974746
R-31	0.007012677	-0.039908449
P-34	-0.006878676	-0.054978697
Q-24	-0.003615246	-0.052927122
Q-02	-0.062364682	-0.112147902
G-13	-0.004674206	-0.055686133
Q-17	-0.015447723	-0.066532405
Q-31	-0.010799768	-0.062092503
G-06	0.003713836	-0.047986339
Q-03	0.001116422	-0.050617569
Q-48	-0.005740099	-0.057729314
Q-40	-0.015339308	-0.068869855
G-07	-0.00446549	-0.058674192
Q-35	-0.018870089	-0.073975635
Q-18	-0.00399085	-0.059278448
G-05	0.021181671	-0.034375841
R-18	0.00041401	-0.055435366
P-23	0.012850927	-0.043485519
G-02	0.015396132	-0.043043599
Q-21	-0.040069001	-0.101230027
Q-09	-0.004833552	-0.067047708
Q-33	-0.003486398	-0.065955217
Q-37	-0.014384891	-0.077552074
R-27	0.018102411	-0.049925227
R-22	0.015972824	-0.054685252
G-20	0.03378682	-0.037550359
G-15	0.031301613	-0.040166849
R-16	0.011976821	-0.062923632
G-16	-0.010132056	-0.086412293

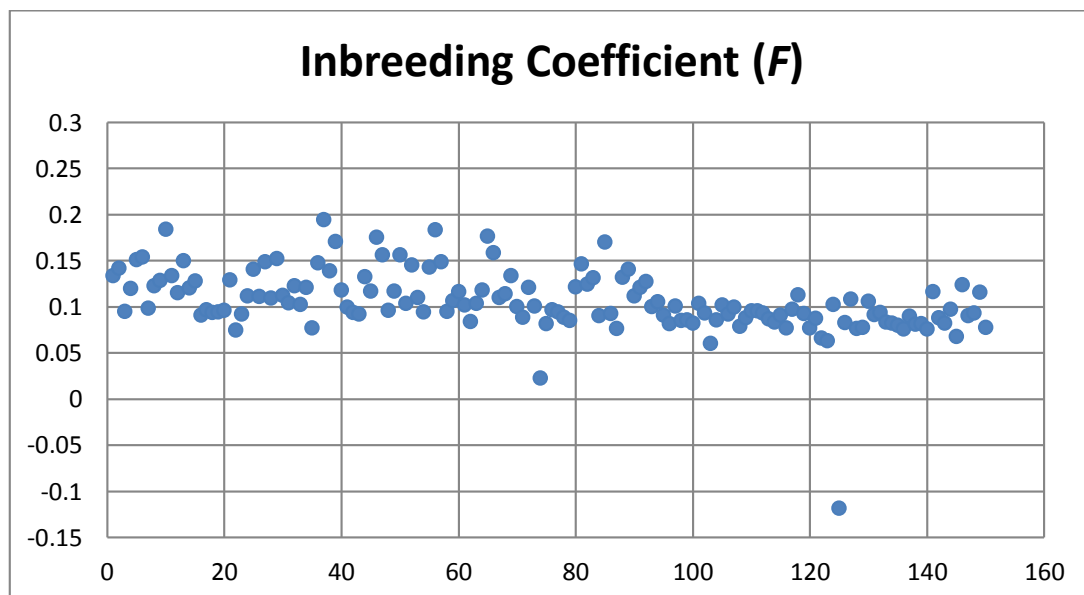
R-11	0.003357633	-0.077620685
R-08	0.023520573	-0.057736918
R-07	0.002031855	-0.085006254
R-23	0.003894408	-0.083262561
G-08	0.021432142	-0.067933185
R-20	0.00962402	-0.081088942
R-13	0.007468099	-0.085223497
G-11	0.038981453	-0.053757133
R-19	0.019234591	-0.086490787
R-25	0.022958348	-0.090212643
R-15	0.014361628	-0.100723547
R-14	0.023493214	-0.097207789
G-21	0.259387034	0.124137198
G-22	0.146515813	0.009728987
G-03	0.290944586	0.140881529
R-06	0.236029896	0.084914252
R-26	0.230761452	0.064743482
R-21	0.238054505	0.061535426
G-24	0.307362409	0.13033721
G-10	0.320520334	0.14271159
R-10	0.235640564	0.05730466
G-17	0.288367117	0.10796912
G-23	0.317564378	0.12162175

**ANEXO 7: Coeficiente de consanguinidad ( $F$ ) de alpacas genotipadas por GBS (PNS generados por ApeK1 y Pst1-Msp1).**

Coeficiente de consanguinidad ( $F$ ) de alpacas genotipadas por GBS (PNS generados por ApeK1).

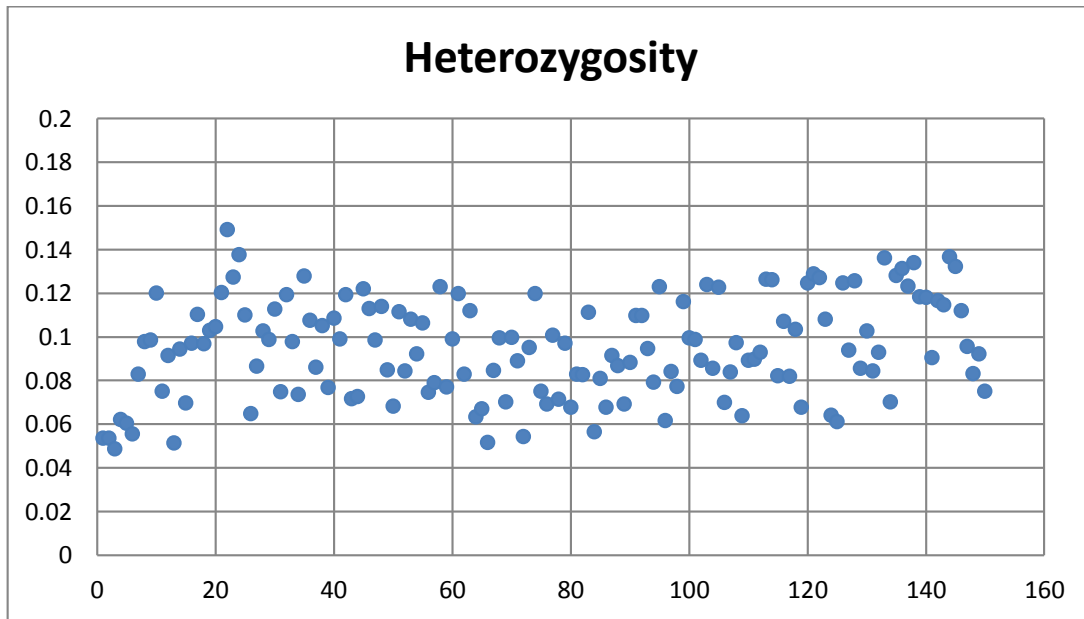


Coeficiente de consanguinidad ( $F$ ) de alpacas genotipadas por GBS (PNS generados por Pst1-Msp1).

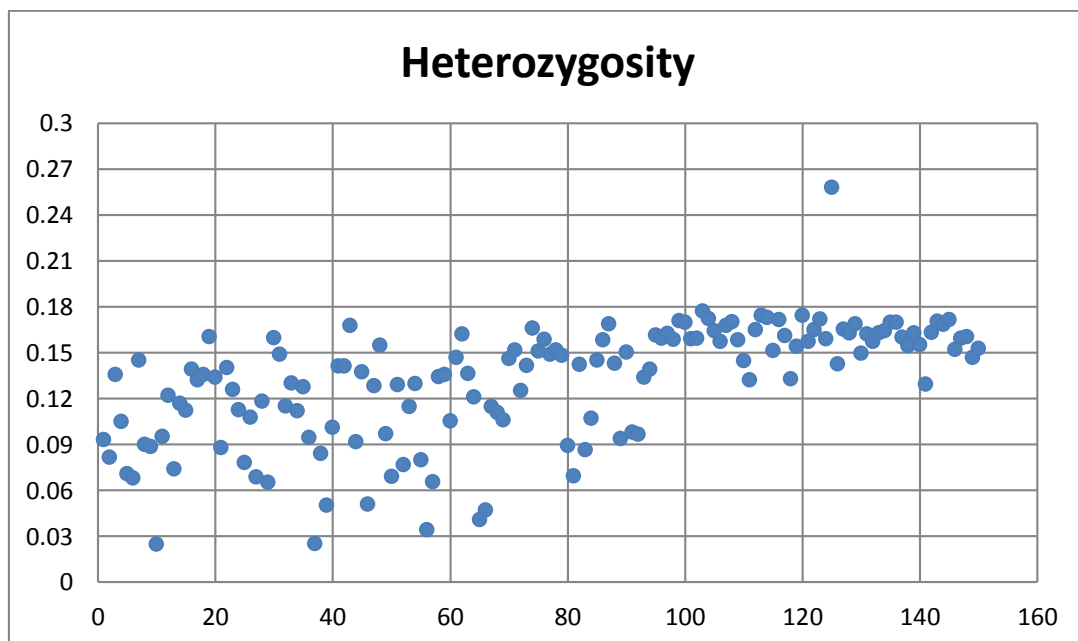


**ANEXO 8: Nivel de heterocigosidad de alpacas genotipadas por GBS. (PNSs generados por ApeK1 y Pst1-Msp1)**

Nivel de heterocigosidad de alpacas genotipadas por GBS. (PNSs generados por ApeK1)



Nivel de heterocigosidad de alpacas genotipadas por GBS. (PNSs generados por y Pst1-Msp1)



**ANEXO 9: Artículo publicado sobre el desarrollo de la micromatriz de alpaca.**

Development of a 76k Alpaca (*Vicugna pacos*) Single Nucleotide Polymorphisms (SNPs) Microarray (en línea). *Genes* 12(2):291. DOI: <https://doi.org/10.3390/genes12020291>.