UNIVERSIDAD DE LA REPÚBLICA

FACULTAD DE AGRONOMÍA

Identificación de polimorfismos de un solo nucleótido asociados a la resistencia a parásitos gastrointestinales en ovinos Corriedale

por

Lic. en Ciencias Biológicas Nicolás Grasso

TESIS presentada como uno de los requisitos para obtener el título de *Magister* en Ciencias Agrarias opción Ciencias Animales

Montevideo

Uruguay

Diciembre 2013

Tesis aprobada por el tribunal integrado por Dr. Silvia Llambí, Graciela García, y Lucía Gutiérrez, el 26 de diciembre de 2013. Autor: Lic. en Ciencias Biológicas Nicolás Grasso. Director Ing. Agr. Gabriel Ciappesoni (Ph.D.), Mejoramiento Genético Animal, Programa Nacional de Carne y Lana. INIA - Uruguay, Co-director Dr. Gonzalo Rincón, Departamento de Ciencia Animal, Universidad de California, Davis, USA y Co-directora Lic. Eileen Armstrong, MSc. PhD. del Área Genética, Departamento de Genética y Mejora Animal, Facultad de Veterinaria, Universidad de la República.

AGRADECIMIENTOS Y DEDICATORIAS

A mis abuelos, particularmente a mi abuelo Tabaré y a mi madrina Sara.

Quisiera expresar mi profundo y sincero agradecimiento a todas las instituciones y personas que hicieron posible que este trabajo se llevara adelante, el cual hoy culmina en mi tesis de maestría:

Al Instituto Nacional de Investigación Agropecuaria (INIA) por el financiamiento del proyecto.

A mi tutor Gabriel Ciappesoni, quien confió en mí para la realización de este trabajo a través del proyecto Fondo Concursable Interno (FCI_07_0_00). Por "seguirme" y apoyarme en las propuestas en el proyecto, y también por "bajarme a tierra" varias veces. Agradecerle por sobre todo, por la enorme paciencia y apoyo que me brindó a su manera durante este proceso de crecimiento, personal, académico y profesional.

A mi cotutor Gonzalo Rincón, por su gran apoyo en el análisis del los datos, por su tiempo y paciencia.

A mi cotutora Eileen Armstrong, quien además de colaborar con este proyecto de maestría, siempre tuvo las palabras justas para mantenerme motivado.

A Elly Navajas, quien también supo tenerme paciencia en los manuscritos, y brindarme increíbles consejos, apoyo y demás.

A Juan F Medrano y Alma Islas por haberme abierto las puertas de la Universidad de Davis, y brindado numerosos conocimientos.

A Daniel "Peto" Castells, por la toma de muestras, quien como bien dijo seguramente después de aguantarme se termine ganando el novel de la paz.

A Diego Gimeno e Ignacio Aguilar por colaborar siempre con el análisis de los datos.

A Virginia Goldberg, con quien compartimos viajes para tomar muestras y en el análisis de datos.

A Wanda Iriarte, quien además de ser una excelente compañera de trabajo es una increíble amiga. Con quien supimos compartir largas horas de laboratorio. A Fernando Macedo, Florencia Pierucconi, Andrea Vergara, Carlos Monzalvo y Pablo Peraza, por haber contribuido en el trabajo.

A mi gran familia, mis padres Juan y Cristina, y mi hermano, María y Nacho, por haberme apoyado y entendido en este proceso. A mi abuela, tíos y primos.

A la familia que uno elige, mis amigos, quienes supieron entender y escucharme durante todo este tiempo. Agradecerle particular a mi amigo y "hermano" Milton y a Maru, quienes me dieron el título de padrino de esta personita que me cambio la vida, mi ahijado Valentino.

ÍNDICE

RESUMEN

En Uruguay, se ha seleccionado de forma divergente una población Corrriedale en base a la diferencia esperada en la progenie para el conteo de huevos por gramo de materia (DEP de HPG) logrando líneas R y S. En el presente trabajo, 54 R y 44 S fueron seleccionados para ser genotipados con el OvineSNP50 en base a valores extremos de DEP de HPG y un coeficiente de parentesco promedio menor a 0,04. Asimismo, se secuenció el genoma completo (WGS) de pooles de 4 animales R y S respectivamente, y se realizó el secuenciado completo del RNA mensajero de los tejidos blanco y potencialmente relacionados (mucosa de yeyuno, duodeno, íleon, abomaso y nodos linfáticos abomasales). Para comprobar el desempeño del OvineSNP50, 3 poblaciones fueron genotipadas (Merino= 110, Corriedale=108 y Criollo=10). La estructura poblacional y las distancias genéticas, fueron calculadas mediante el programa STRUCTURE, análisis de componentes principales (PCA) e índice de fijación (F_{ST}). Se tomaron como SNP raza específicos a los que estaban fijados de forma diferencial en cada raza. Un conjunto de 18.181 marcadores del OvineSNP50 fueron suficientes para explicar la estructura poblacional de las tres razas mediante el STRUCTURE y PCA. Las razas Merino y Corriedale presentaron 89.4% y 86.9% de SNP polimórficos respectivamente y una distancia genética moderada. Sin embargo, Criollo mostró niveles de 69% y una distancia genética mayor (F_{ST}=0,17 para las dos razas). Asimismo, los niveles de polimorfismo detectados con el OvineSNP50 lo harían una herramienta apta para llevar a cabo otro tipo de estudios (por ejemplo estudios de asociación). Los análisis de PCA y STRUCTURE mostraron la estructura de las líneas divergentes Corriedale (54 R y 44 S genotipados), mostrando una distancia genética moderada (F_{ST} =0,042). Se identificaron 8 regiones genómicas como asociadas a DEP de HPG mediante regresiones simples. Estas fueron evaluadas en detalle gracias al WGS de cada línea y RNA-seq. De esta manera se identificaron nuevos genes candidatos y se identificaron nuevos SNP.

Palabras clave: ovinos, parásitos, OvineSNP50, RNA-seq, secuenciado de todo el genoma.

Identification of single nucleotide polymorphism associated with resistance to gastrointestinal parasites in Corriedalle sheep

SUMMARY

In Uruguay, divergent selection lines have been developed since 1998 in Corriedale sheep, based on expected progeny difference for fecal egg count (EPD FEC), leading to R and S lines. In the present study, a total of 54 R and 44 S sheep were selected by extreme values of EPD FEC and inbreeding coefficient (<0.04) to be genotyped with the OvineSNP50 BeadChip. Two whole genome sequencing (WGS) were performed from pooled DNA of 4 R and 4 S individuals and transcriptome sequencing (RNA-Seq). Six libraries of mRNA were performed of parasites target tissues and related ones (jejunum mucosa, duodenum, ileum, abomasum and abomasal lymph nodes), and mapped all together. In order to test the OvineSNP50 performance, three breeds (Merino n=110, Corriedale n=108 and Creole n=10) were genotyped using the OvineSNP50 beadchip®. Genetic diversity was evaluated by the minor allele frequency (MAF) comparison between the three breeds. Population structure and genetic distance were assessed using STRUCTURE software, principal component analysis (PCA) and fixation index (F_{ST}). Fixed markers (MAF=0) that were different among breeds, were identified as specific breed markers. Using a subset of 18,181 SNPs, PCA and STUCTURE analysis were able to explain population stratification within breeds. Merino and Corriedale showed 89.4% and 86% of polymorphic SNPs, respectively, and a F_{ST} of 0.08 between them. In contrast, Creole only showed 69% polymorphic SNPs and a larger genetic distance from the other two breeds (F_{ST}=0.17 for both breeds). Levels of SNP polymorphism could be sufficient to carry out other studies. PCA and STUCTURE analysis were able to explain population stratification within the Corriedale divergent lines (54 R and 44 S genotyped) and a moderate genetic distance (F_{ST}=0.042). Association regression analysis indicated significant association for 8 genomic regions to EPD FEC. Those regions were deeply evaluated by combining WGS and RNA-Seq. Novel candidates genes were indentified, some of them were showing expression and new SNPs were also described.

Keywords: ovine, parasites, OvineSNP50, RNA-seq, Whole genome sequencing.

1 INTRODUCCIÓN

1.1 PRODUCCIÓN OVINA EN URUGUAY

Actualmente en Uruguay existe un stock aproximado de 7,5 millones de cabezas ovinas distribuidas en 24000 establecimientos (Dirección General De Servicios Ganaderos (DI.CO.SE), 2011). La producción ovina (carne, lana y cuero) es el cuarto sector de exportación del Uruguay (Montossi *et al.*, 2005), convirtiéndose en un rubro de alta importancia económica y social para el país. En consecuencia, existen más de 50.000 puestos de trabajo directamente ligados al rubro ovino (ej. esquiladores, transportistas, obreros textiles, obreros de la industria cárnica, servicios conexos, etc.) (Montossi *et al.*, 2005).

1.2 PARÁSITOS GASTROINTESTINALES EN URUGUAY

En Uruguay, al igual que en el resto de las áreas templadas del mundo, las parasitosis gastrointestinales (PGI) causadas por nemátodos constituyen una de las principales limitantes sanitario-económicas para la producción ovina en sistemas pastoriles (Perry *et al.*, 2002).

Las condiciones climáticas dadas en Uruguay, marcan variabilidad de mayor o menor intensidad en la parasitosis por nemátodos gastrointestinales, pero difícilmente se registran períodos prolongados libres de desafío parasitario. Estas características determinan un amplio número de especies parasitarias con una prevalencia notoriamente mayor de *Haemonchus contortus* (Nari *et al.*, 1977). Un estudio realizado en Uruguay por Nari *et al.* (1977) muestra la prevalencia de nemátodos adultos que parasitan animales de recría a lo largo del año y donde la presencia de *Haemonchus contortus* es la más alta (43%) ocupando el segundo lugar el género *Trichostrongylus* spp (38%). El impacto económico de las PGI está dado tanto por el costo que significan la prevención y tratamiento de estas infecciones,

como por las pérdidas productivas que ocasionan por morbilidad y mortalidad de los animales infectados (Sykes, 1994). El control de la parasitosis ha dependido en gran medida del tratamiento antihelmíntico, sin embargo éste ha llevado al desarrollo de resistencia por parte de los parásitos a los químicos utilizados y la consiguiente pérdida de efectividad.

1.3 EVALUACIÓN GENÉTICA PARA HPG Y NÚCLEO CIEDAG-SUL.

El método más difundido en Uruguay para la identificación de ovinos potencialmente resistentes es el recuento de huevos por gramo de materia fecal (HPG). Para este análisis, se extraen dos muestras de materia fecal de infecciones parasitarias desarrolladas en forma natural y correspondiente a dos ciclos parasitarios independientes, separados entre sí por un tratamiento antihelmíntico; obteniéndose el HPG 1 y 2 respectivamente.

En nuestro país, el HPG es incluido en las Evaluaciones Genéticas Poblacionales de las razas Merino y Corriedale, contándose anualmente con estimaciones del valor genético de los animales a través de las DEP (Diferencia Esperada en la Progenie) para HPG. En ambas razas se han estimado valores moderados de heredabilidad (0.18) (Ciappesoni *et al.*, 2010, Castells *et al.*, 2002), lo que permitiría progresar genéticamente mediante selección. Sin embargo, es difícil incluir este carácter en los programas actuales de mejora, dado que la obtención de registros fenotípicos requiere permitir la expresión de la enfermedad en los animales. Esto implica pérdidas económicas para los cabañeros y por ende, reticencias de los mismos a la toma de registros.

Esta característica ha sido incluida en las evaluaciones genéticas de la raza Corriedale desde el comienzo en el año 1994. En base a esta información se forma en el año 1998, el núcleo resistente a PGI en el Centro de Investigación y Experimentación “Dr. Alejandro Gallinal” (CIEDAG), del Secretariado Uruguayo de la Lana (SUL), ubicado en Cerro Colorado, Departamento de Florida. Posteriormente

en el 2003 se comenzó la selección divergente dando origen a la línea susceptible. Actualmente el núcleo está compuesto por 150 vientres en la línea resistente (R) y 100 en la susceptible (S), encarneradas cada año con 5-6 carneros de cada línea. El único criterio de selección empleado todos los años es la DEP de HPG en la recría. Ambas líneas son manejadas en conjunto, permitiendo de esta manera la evaluación genética. Las tendencias genéticas obtenidas muestran una clara divergencia entre ambas líneas (Castells y Gimeno, 2011).

1.4 SELECCIÓN ASISTIDA POR MARCADORES: SNP

Para caracteres difíciles de medir, como la resistencia a enfermedades, la identificación de marcadores moleculares asociados a la variación del carácter resulta especialmente útil, ya que su aplicación en esquemas de selección asistida por marcadores o MAS (de su sigla en inglés, *Marker Assisted Selection*) facilita la identificación de animales resistentes directamente a partir de sus genotipos, aumentando la eficiencia de la selección (Beuzen *et al.*, 2000). Sin embargo, la aplicación de MAS para características complejas, ha estado limitada por la pequeña proporción de la variación fenotípica que explican la mayoría de los marcadores moleculares identificados (Kemper *et al.*, 2011, Hayes y Goddard, 2001).

La secuenciación de genomas de varias especies ha permitido identificar polimorfismos de nucleótido simple (*Single Nucleotide Polymorphisms, SNP*). Estas variaciones son las más comunes en las secuencias de DNA. Su abundancia y distribución en los genomas, y las nuevas tecnologías capaces de genotipar miles de SNP a la vez, han hecho a estos marcadores una herramienta por excelencia para los estudios de asociación y selección genómica (SG) (Ding y Jin, 2009). La aplicación más impactante de esta tecnología en el mejoramiento genético animal es la SG, donde se considera esta gran cantidad de información molecular para caracterizar la variación genética total de los individuos. En algunos casos, esto resulta en una estimación más exacta del mérito genético en animales jóvenes y en una mayor tasa

de ganancia genética en comparación con la MAS y la selección clásica (Meuwissen, 2008; Schaeffer, 2006).

1.5 GENOTIPADO MASIVO DE ALTO RENDIMIENTO DE SNP

Las tecnologías de genotipado masivo de alto rendimiento (BeadChip), fueron puestas a disposición de los investigadores en el año 2005 por la compañía Illumina. Las primeras plataformas lanzadas por esta empresa fueron desarrolladas para su aplicación en seres humanos; sin embargo, en la actualidad existen para numerosas especies. Recientemente se desarrolló para la especie ovina el OvineSNP50 Beadchip (Illumina), con 54,241 SNP con una distancia promedio de 46 kb entre cada marcador. Este chip, fue creado en conjunto por varios organismos como parte del International Sheep Genomics Consortium (ISGC). El OvineSNP50 Beadchip ha sido utilizado por el ISGC en 64 razas de ovinos, mostrando su efectividad (OvineSNP50 Datasheet - Illumina, 2010).

Sin embargo, pocos estudios de asociación en caracteres complejos realizados con este tipo de tecnologías, han logrado explicar el efecto de los SNP sobre un fenotipo (Hayes *et al.*, 2010).

1.6 TÉCNICAS DE SECUENCIADO DE PRÓXIMA GENERACIÓN (NGS): SECUENCIADO DEL GENOMA COMPLETO (WHOLE GENOME SEQUENCING) Y SECUENCIADO DEL RNA (RNA SEQUENCING, RNA-SEQ)

En términos generales, el DNA genómico completo es fragmentado en pequeños segmentos los cuales son ligados a adaptadores, para luego ser secuenciados de forma aleatoria y simultánea (Pareek *et al.*, 2011). El resultado primario de las NGS son secuencias cortas (50 a 500 pares de bases continuas) llamadas *reads*, las cuales son unas de las mayores limitantes de esta tecnología,

debido a los grandes volúmenes de datos y consecuentemente de sus requerimientos informáticos. El tamaño de estos *reads*, es un aspecto extremadamente relevante para la cobertura (Wang *et al.*, 2010), la cual se define como el número de *reads* que se solapan unos con otros en una región genómica específica. Para lograr un ensamblado preciso de las secuencias es crítico obtener una cobertura suficiente (Wang *et al.*, 2010).

Estas tecnologías pueden ser utilizadas para el secuenciado *de novo* de genomas o para el re-secuenciado de una muestra en particular y su consiguiente ensamblado con un genoma de referencia. En la actualidad el mayor uso de las NGS en el estudio de caracteres complejos, involucra la identificación de SNP y su posterior asociación con el fenotipo de interés (Day-Williams y Zeggini, 2011).

Las tecnologías de secuenciado de próxima generación han provisto oportunidades sin precedentes para la investigación genómica incluyendo el perfil de la expresión génica, anotación de genomas, descubrimiento de pequeños RNA y detección de expresión diferencial (Bentley 2008). Dentro de los diferentes enfoques, el secuenciado del RNA *(RNA sequencing, RNA-seq)* es un método nuevo muy poderoso para mapear y cuantificar transcriptomas, permitiendo analizar la expresión global en diferentes tejidos. Recientemente, esta técnica ha sido utilizada para detectar SNP, de forma eficiente y precisa, en regiones de transcriptos de diferentes especies (Morin *et al.*, 2010).

El RNA-Seq utiliza la tecnología de secuenciado de alto rendimiento desarrollada recientemente. A grandes rasgos, una población de RNA (RNA total, fraccionado o mensajero) es convertida en librerías de fragmentos de cDNA con adaptadores unidos a uno o ambos extremos. Cada molécula, con o sin amplificación, es luego secuenciada a alto rendimiento desde un extremo (*single end sequencing*) o ambos (*pair end sequencing*), obteniendo en esta etapa los *reads*. Generalmente, estos tienen un tamaño de 30 a 400 bp dependiendo de la tecnología de secuenciado utilizada. Los *reads* resultantes son ensamblados entre sí para formar *contigs*, los cuales son luego alineados contra un genoma de referencia. De esta manera se

obtiene la estructura de la transcripción del tejido analizado y/o diferentes niveles de expresión de los genes obtenidos (Wang *et al.*, 2010).

1.7 HIPÓTESIS

La variabilidad genética existente tras 10 años de selección divergente en la población Corriedale, puede haber tenido como resultado diferentes mecanismos para la resistencia y susceptibilidad a parásitos gastrointestinales. La variabilidad genética anteriormente mencionada, podría deberse a regiones genómicas que fueron seleccionadas a favor y en contra para el carácter en estudio, llevando consigo la diferencia a nivel de marcadores moleculares entre las líneas en estudio. Dichos marcadores con efecto sobre la población podrían ser detectados con las técnicas planteadas en el presente trabajo.

1.8 OBJETIVO GENERAL

A partir de la identificación de SNP asociados a la resistencia/susceptibilidad a parásitos gastrointestinales mediante el genotipado masivo, se determinarán regiones genómicas que podrán ser evaluadas en detalle gracias al resultado del secuenciado genómico completo y el RNA-seq. Esto proveerá una mayor resolución de estas regiones, permitiendo determinar un mayor número de SNP potencialmente asociados al carácter en estudio.

1.9 OBJETIVOS ESPECÍFICOS

- Evaluar el desempeño del OvineSNP50 en tres poblaciones comerciales locales Corriedale, Merino y Criollo: frecuencias alélicas, estructura poblacional, diferencias genéticas entre razas.

- Identificar SNP asociados a la resistencia a PGI mediante el genotipado masivo en Corriedale.

- Determinar regiones genómicas que podrán ser evaluadas en detalle gracias al resultado del secuenciado genómico completo y el RNA-seq, obteniendo como resultado nuevos SNP y genes candidatos para el carácter en estudio.

- Obtener un set de marcadores moleculares candidatos para diagnosticar la R/S a PGI en ovinos de la raza Corriedale en Uruguay.

2 GENOMIC VARIATION AND POPULATION STRUCTURE DETECTED BY SINGLE NUCLEOTIDE POLYMORPHISM ARRAYS IN CORRIEDALE, MERINO AND CREOLE SHEEP

A.N. Grasso[1]*, V. Goldberg[1], E.A. Navajas[2], W. Iriarte[2], D. Gimeno[3], I. Aguilar[1], J.F. Medrano[4], G. Rincón[4], C.G. Ciappesoni[1]

[1] Meat and Wool National Program, Instituto Nacional de Investigación Agropecuaria, Las Brujas, Canelones, 90200 Uruguay.

[2] Biotechnology Unit, Instituto Nacional de Investigación Agropecuaria, Las Brujas, Canelones, 90200 Uruguay.

[3]Department of Animal Genetic Improvement, Secretariado Uruguayo de la Lana, Montevideo, 11800 Uruguay.

[4]Department of Animal Science, University of California Davis, CA 95616 USA.

SNP set for sheep breeds

Keywords: population stratification, OvineSNP50, principal component analysis, fixation index, allele frequencies, breed composition.

* Corresponding author. Tel.: +59899736233.

E-mail address: nicograsso26@gmail.com

Enviado a Genetics and Molecular Biology (formerly named Revista Brasileira de Genética/Brazilian Journal of Genetics - ISSN 0100-8455) is published by the Sociedade Brasileira de Genética (Brazilian Society of Genetics). Bajo el formato solicitado por la revista.

2.1 **ABSTRACT**

The aim of this study was to investigate the genetic diversity within and between three breeds: Corriedale, Merino and Creole. Sheep from three breeds (Merino n=110, Corriedale n=108 and Creole n=10) were genotyped using the Illumina OvineSNP50 beadchip®. Genetic diversity was evaluated by the minor allele frequency (MAF) comparison between breeds. Population structure and genetic distance were assessed using STRUCTURE software; principal component analysis (PCA) and fixation index (F_{ST}). Fixed markers (MAF=0) that were different among breeds, were identified as specific breed markers. Using a subset of 18181 SNPs, PCA and STUCTURE analysis were able to explain population stratification within breeds. Merino and Corriedale showed high levels of polymorphism (89.4% and 86% of polymorphic SNPs, respectively), and a moderate genetic distance (F_{ST}=0.08) between them. In contrast, Creole only 69% polymorphic SNPs and a larger genetic distance from the other two breeds (F_{ST}=0.17 for both breeds). Hence, a subset of molecular markers present in the OvineSNP50 is informative enough for breed assignment and population structure analysis of commercial and Creole breeds.

2.2 **INTRODUCTION**

Natural and artificial selection processes in domestic animals took place in a relatively short period of time (10000 years) (Zeder, 2008) leading to a wide range of phenotypes with large genetic differences between breeds (Kijas *et al.,* 2009). Genomic information and the vast databases of polymorphisms that are available, provide a unique opportunity to study loci that were under selection between breeds in this recent history (Hayes *et al.*, 2008; Flori *et al.*, 2009). Several genomes have been sequenced providing the information to identify single nucleotide polymorphisms (SNP) (Ding and Jing, 2009), which are the most common variations in DNA sequences. Their abundance, genome-wide distribution, and availability of a large number of commercial platforms for High-Throughput SNPs genotyping, make SNPs an ideal marker for different genomic studies (Ding and Jin, 2009).

The OvineSNP50 beadchip® (OvineSNP50), which was developed by Illumina in collaboration with the International Sheep Genomics Consortium (ISGC), is a genome-wide genotyping array for the ovine genome. This beadchip contains 54,241 SNPs which were chosen for being uniformly widespread across the ovine genome (with an average gap size and distance of 50.9 kb and 46 kb, respectively) and for their high levels of polymorphism in the evaluated breeds. These SNPs were tested in more than 75 economically important sheep breeds by the ISGC. The chip features autosomal, mitochondrial and sex-linked SNPs (OvineSNP50 Datasheet - Illumina 2010).

The information provided by the OvineSNP50 has many applications such as identification of quantitative trait loci, genome-wide association studies (GWAS), characterization of genetic variability among breeds, genomic selection and genetic comparison between breeds (Alam *et al.,* 2011; Kijas *et al.,* 2009). Differences in genotyping results reported among breeds (Kijas *et al.*, 2012) may have unfavourable

effects on these applications (i.e. contribution of population stratification to generate false positive results in GWAS; Kemper *et al.*, 2011). For that reason, many studies have focused on the evaluation of the genetic relatedness and substructure within livestock populations in GWAS (Zenger *et al.*, 2007). Given the current efforts on developing GWAS in sheep populations, the aim of this study was to investigate the genetic diversity and population structure within and among Corriedale, Merino and Creole breeds, using the OvineSNP50.

2.3 MATERIALS AND METHODS

2.3.1 Selected animals

Data of 228 animals were analyzed including 98 Corriedale and 100 Merino lambs, and 10 sires of each of the three breeds included in this study. Corriedale lambs belong to divergent selection lines for resistance versus susceptibility to gastrointestinal parasites, developed by the Uruguayan Secretariat of Wool (SUL) (Castells and Gimeno, 2011). Merino lambs were selected from the Fine Merino Nucleus, located at the Experimental Station Tacuarembó of the National Agricultural Research Institute (INIA) of Uruguay (Montossi *et al.*, 2005).

In order to minimise the relationship in the dataset, Merino and Corriedale lambs with average relatedness coefficient value (AR) lower than 0.04 were selected. The AR values were calculated using ENDOG v4.6 software (Gutiérrez and Goyache, 2005). Similarly, 10 Corriedale and 10 Merino sires that were not related for at least two generations were genotyped. Finally, 10 Creole sires were selected from the San Miguel National Park in Rocha, Uruguay. Since there are no pedigree records of the Creole flock, selection criterion was based on maximizing the observed phenotypic variation (wool colour and presence of one or two pairs of horns).

2.3.2 **DNA extraction, genotyping and quality control**

DNA was isolated from blood samples and then genotyped using the OvineSNP50. Only samples that presented call rates higher than 80% were considered for further analysis. Likewise, SNPs with call rates lower than 90% were removed. Only autosomal markers were evaluated, to avoid X chromosome heterozygosity between females and males.

2.3.3 **SNPs frequencies**

Allele frequencies were calculated with SVS Golden Helix software for each breed to assess the OvineSNP50 performance. Rare and fixed alleles were determined by the minor allele frequency (MAF), with MAF values lower than 0.01 and 0, respectively. Highly polymorphic alleles were those with MAF values between 0.3 and 0.5. In this way, a comparative analysis was performed to identify which SNPs were fixed or polymorphic in each breed's genome.

2.3.4 **Genetic population structure**

Subdivisions between and within the populations were evaluated with the linkage model of the STRUCTURE software version 2.3.3 (Pritchard *et al.*, 2000), using a range of possible clusters (K) from two to nine. Ten runs of 10000 iterations after a burn-in period of 10000 iterations were performed for each K. Most probable K was identified by the ΔK method (Evanno *et al.*, 2005). Population structure was also inferred by principal components analysis (PCA) from the marker data (Patterson *et al.*, 2006). The genetic distances between breeds were calculated using the Fixation Index (F_{ST}) (Weir and Cockerman, 1984), which assess the reduction in genotypic heterozygosity. This parameter can range from zero (no genetic divergence between the subpopulations or from the ancestral population) to one (complete isolation of the subpopulations from each other and the overall population). Moderate and large

distances are those in the ranges of F_{ST} values from 0.05 to 0.15 and 0.15 to 0.25, respectively.

The subset of autosomic SNPs used in these analysis (STRUCTURE, PCA and F_{ST}) for each breed were in moderate linkage disequilibrium (r^r<0.4), with MAF>0.01, were not in Hardy Weinberg equilibrium and common for the three breeds. These criteria were applied to identify informative molecular markers for each breed.

2.4 **RESULTS AND DISCUSSION**

2.4.1 **Genotypes and quality control**

Most samples (96.9%) presented call rates higher than 80%. Only seven samples had lower call rates and therefore excluded from the analysis. Overall, 107 Corriedale, 104 Merino and 10 Creole samples were used for the present study. Despite the differences among the three breeds, our results showed that OvineSNP50 had a coverage similar to the expected in the design of the beadchip (average gap size of 50.9 kb, OvineSNP50 Datasheet - Illumina 2010), as well as a good genotyping quality given the high sample and marker call rates (Table 1). Although neither Corriedale nor Creole were included among the genotyped breeds in the validation (OvineSNP50 Datasheet - Illumina 2010), the results were promising.

Currently, not all SNPs in public databases have been validated and their levels of polymorphisms are unknown for many sheep breeds. It has been previously reported a low genotyping performance of the OvineSNP50 in non-commercial sheep (Miller *et al.*, 2011). However, in this study the OvineSNP50 showed similar coverage and genotyping quality in Creole sheep and commercial breeds (Table 1).

Table 1. Number of autosomal SNPs (52413) for each breed with call rate higher than 90% and SNP and genomic coverage (average gap and density) given by the OvineSNP50 beadchip.

	SNP with CR>90%	Average Gap (Kb)	Density (Kb)
Corriedale	48,862	50.46	50.67
Creole	51,287	50.56	50.77
Merino	49,955	50.55	50.76

2.4.2 **Fixed SNPs**

The number of fixed markers varied among breeds with values of 874, 1449 and 13790 SNPs for Corriedale, Merino and Creole sheep, respectively. Ninety three percent of Corriedale SNPs with MAF=0 were in common with Creole sheep and 69% with Merino. In Merino, 90% of the fixed molecular markers were also fixed in Creole sheep. Comparing the three breeds, there were a total of 585 fixed SNPs in common (Figure 1).

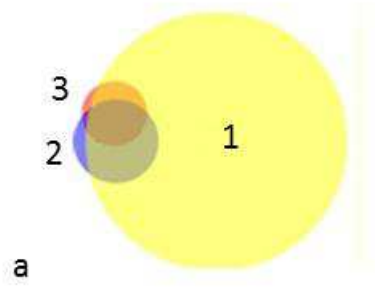

Figure 1. Venn diagram showing fixed SNPs (MAF=0) (a.) in Uruguayan Creole (1), Merino (2) and Corriedale (3) sheep.

Figure 1 shows there were 99, 99 and 11190 fixed SNPs that were unique for Corriedale, Merino and Creole, respectively. These molecular markers could be useful when it comes to assign either breed identity or the proportion of a certain breed in a given animal. Wiggans *et al.* (2010) reported that a set of 622 SNPs is being used to determine breed identity, as one step in the quality control process for the USDA genotype database for dairy cattle (~200 monomorphic SNPs for each breed). In pigs, Ramos *et al.* (2011) searched for breed specific SNPs in five breeds (Duroc, Landrace, Large White, Pietrain and Wild Boar). In this case, a breed specific marker was defined as a SNP for which one of the alleles was detected only in one breed (a fixed SNP). A total of 193 SNPs from the PorcineSNP60 beadchip were confirmed by Ramos *et al.* (2011) as being breed specific and then validated using 490 animals of the five breeds. Four breed assignment validation tests were performed and the results indicated that for all methods tested, 99% of the animals were correctly assigned. In this way, fixed SNP found in this study as breed specific for Corriedale, Merino and Creole sheep could be used for breed identification after the validation in an independent group of animals.

2.4.3 **Frequencies of polymorphic and very polymorphic SNPs**

The percentages of polymorphic SNPs (MAF>0.01) present in Merino and Corriedale breeds were 89.4% and 86%, respectively (Table 2). However, only 69% of the SNPs were polymorphic in Creole sheep. Highly polymorphic SNPs (MAF>0.3) by breed follow a similar trend to polymorphic SNPs. The results in Table 2 revealed that 50.1% and 50.9% of highly polymorphic SNPs were observed in Corriedale and Merino animals, respectively. In contrast, Creole had lower number of highly polymorphic SNPs (36%; Table 2).

Table 2. Minor allele frequency (MAF) of Corriedale, Merino and Creole. All referring to autosomal SNPs shared by the three breeds. Min and Max are the minimum and maximum MAF values. SD: standard deviation.

	MAF					
	Alleles	Min	Max	Median	Mean	SNPs
Corriedale	All	0.00	0.50	0.28	0.27	48,862
	Rare	0.00	0.01	0.00	0.00	2,150
	High polymorphic	0.30	0.50	0.40	0.40	24,484
Merino	All	0.00	0.50	0.29	0.27	49,955
	Rare	0.00	0.01	0.00	0.00	1,715
	High polymorphic	0.30	0.50	0.40	0.40	25,449
Creole	All	0.00	0.50	0.20	0.19	51,287
	Rare	0.00	0.01	0.00	0.00	14,056
	High polymorphic	0.30	0.50	0.40	0.39	18,447

In general, Corriedale and Merino breeds had high levels of polymorphism. These results for Merino were expected because it was included in the design/validation of the OvineSNP50. Moreover, the results obtained in Merino and Corriedale (Table 2) were similar to those observed in the six breeds (Awassi, Merino, Poll Dorset, Romney, Scottish Blackface and Texel) sequenced during the OvineSNP50 development (Gutiérrez-Gil *et al*., 2012). The 50 Merino animals that were used in the beadchip design showed 1.3% and 48.5% of rare and highly polymorphic SNPs, respectively. In addition, highly polymorphic and rare SNPs percentages in the other commercial breeds ranged from 37.8% to 47.1%, and 3.2% to 13.9% respectively (Gutiérrez-Gil *et al*., 2012), supporting the results presented in our study. Nevertheless, there were more rare SNPs in Uruguayan Merino than in the Merino subsample analysed by Gutiérrez-Gil *et al*. (2012). In the case of the Corriedale breed, although it was not directly considered in the design, it is a composite breed

that resulted from the crossbreeding between Merino and Lincoln (Mendoza, 1968). High levels of polymorphisms were also reported by Gutiérrez-Gil *et al.* (2012) for four Spanish breeds with the OvineSNP50. They found that 95.6 to 98.5% of the SNPs were polymorphic and 46.5% to 50.6% were highly polymorphic. These values are slightly higher than those found in Merino and Corriedale, probably because Gutiérrez-Gil *et al*. (2012) included all SNPs, not only the autosomal SNPs.

Animals used in this study were selected using minimum AR as selection criterion, which could have contributed to the high percentage of polymorphic markers by avoiding highly related and inbred animals. The AR may be a useful tool at the time of designing GWAS to maximize the genomic information content and reduce the risk of spurious associations.

The Creole sheep had significantly lower percentages of polymorphic and very polymorphic SNPs. As the OvineSNP50 was initially intended for detecting differences between domestic sheep breeds, the selection of SNPs could have been biased towards sites with recent mutations (Miller *et al*., 2011). This could be one reason explaining the lower polymorphism in the Creole sheep, as many of the sites would exist in their ancestral monomorphic state. Nevertheless, the lower level of polymorphism in this case could be also explained by the small size of the population they belong to ($n<200$), and the small sample of Creole animals with genomic information. In addition, other unknown events in the history of this Creole population, such as any severe population bottleneck, could have also contributed to a lower percentage of polymorphic SNPs. The ten Creole animals in this study were not selected by AR, and in spite of these facts they presented relatively high levels of polymorphism.

In summary, the levels of polymorphism of the molecular markers in the OvineSNP50 would be sufficient to perform GWAS and genomic selection in local commercial breeds and Creole sheep. Many GWAS in sheep using a similar number of informative SNPs have been reported recently. Johnston *et al*. (2011) conducted a GWAS in non-commercial Soay sheep using 36000 SNPs from the OvineSNP50 and

indentified more SNPs in and around Horn gene (RXFP2), contributing to unravel the genetic basis of this trait. A total of 49233 SNPs were used in a multi-breed GWAS for the identification of genomic regions associated with susceptibility to and control of Ovine Lentivirus (White *et al*., 2012), although the number of informative SNPs was variable among breeds (Rambouillet 50,275; Polypay 50,264; Columbia 44,258). Significant findings on the genetics of litter size in sheep were reported by Våge *et al*. (2013), who included 47,986 SNPs in the analysis (after allowing for 5% missing data per SNP and 1% MAF). They found a missense mutation in growth differentiation factor 9 that was strongly associated with the number of lambs born in Norwegian White Sheep.

2.4.4 **Population structure**

A total of 18,181 autosomal SNPs were present in the three breeds and were informative for the population structure study (Table 3). Figure 2 (A,B,C) shows the results of population structure by STRUCTURE and PCA. STRUCTURE clustering results, using Bayesian methods, are presented in Figure 2A, assuming 4 hypothetical populations (K=4), determined by ΔK method (Figure 2B). For K=4, the three breeds were clearly differentiated. Creole animals were separated in a central cluster (Figure 2A). Merino breed was represented by one main cluster with low levels of admixture, while Corriedale individuals were mostly grouped in two different clusters, showing higher levels of admixture.

Table 3. Subsets of the 52413 autosomal SNPs per breed that meet the established criteria for the population structure analysis.

	Corriedale	**Creole**	**Merino**
MAF>0.01	50,418	38,623	50,950
CR>0.9, MAF>0.01	46,696	37,725	48,533
HW	52,327	51,287	51,730
r^r<0.4	33,595	15,196	38,806
Totals*	30,761	37,725	35,999
Subset of SNPs		18,181	

MAF Minor allele frequency
CR call rate
HW Hardy Weinberg equilibrium
r^2 Linkage disequilibrium

STRUCTURE methods allow the assignment of individuals to groups based on their genetic similarities, providing information about the number of ancestral populations underlying the observed genetic diversity. In fact, the Corriedale representative clusters were also present in Merino with low ancestry coefficient, supporting the possible common origin. A similar pattern was observed in Corriedale with some features of the Merino cluster. Besides these similarities, Corriedale shows two clusters indicating genetic substructuring that resulted from using Corriedale animals that belong to divergent selection lines.

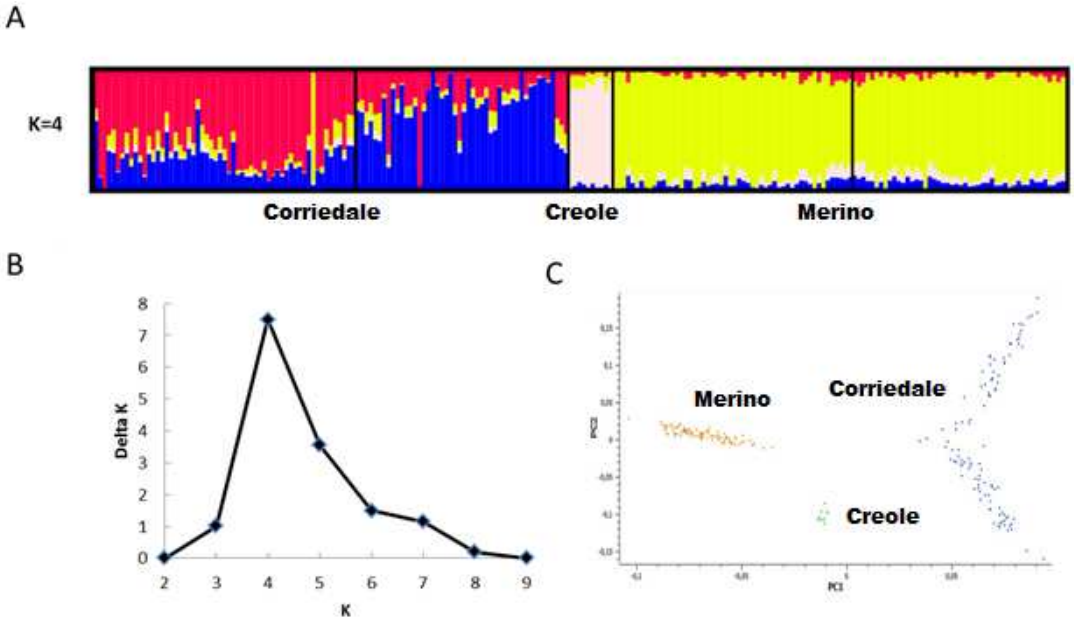

Figure 2. Population structure analysisestimated for 18181 SNPs with four cluster (K=4). A) Each individual is represented by a vertical line, which is divided by K colours segments representing the estimated fraction belonging to each cluster. The black lines separate individuals from different subpopulations. B) Delta K graphic showing that the peak (the best K) is 4. C) First and second principal components. Each point represents an animal, and individuals are coloured according to the breed. Each cluster is labelled with the name of the breed.

Figure 2C shows the results obtained by the PCA. The three breeds were clearly differentiated by the first two principal components, which were sufficient to account for the observed population structure, with findings very similar to those obtained by STRUCTURE (Figure 2A). Creole sheep were very tightly clustered, whereas Merino breed was located in a separate, but loosely cluster. The subdivision of the Corriedale cluster into two subpopulations was also showed by the PCA analysis.

Genetic relatedness and substructure within populations should be taken into account in order to avoid spurious associations in GWAS (Zenger *et al.*, 2007).The correction through population stratification (Lander and Schork, 1994) and the removal of outliers are alternatives to minimize the likelihood of false positive associations. The results of this study indicate that the OvineSNP50 provide useful information for identifying populations structures and outliers. Both PCA and STRUCTURE separated into two clusters animals of the same breed which belonged to divergent selection lines. In addition, there was an animal that was wrongly assigned to a breed which was clearly detected by both methods using the selected informative SNPs.

Genetic differences in population structure and variation detected from allele frequencies could be explained by the fact that Corriedale, Merino and Creole have undergone different selection processes. In this particular case, Corriedale flock has been divergently selected for resistance to parasites, whilst the breeding programme in the Merino nucleus has been focused on improving wool quality (Montossi *et al.*, 2005).In contrast, Creole sheep has not been under artificial selection since it was introduced in Uruguay in the XVIII century (Fernández, 2000). Nevertheless, more Creole animals should be evaluated.

A moderate genetic distance was observed (F_{ST}=0.08) between the two commercial breeds. Results from Delgado *et at.* (2011) also indicated that commercial cattle breeds showed low to moderate genetic similarity. On the other hand, distances were larger when Creole was compared with Corriedale and Merino with F_{ST} values of 0.16 and 0.15, respectively (Table 4). Nevertheless, both analyses confirmed that Corriedale and Merino were more related between them and Creole was more distant from the commercial breeds (Table 4). In fact, Mendoza (1968) indicated that Corriedale breed came from crossing Merino with Lincoln sheep 150 years ago. Moreover, Creole has not been under any defined selection process since it was introduced in Uruguay (Camacho *et al.*, 2000; Fernández, 2000). Thus, this 400-year-period of no artificial selection in Creole is more remote in time than the common origin between Merino and Corriedale.

Table 4. Genetic distances between breeds, expressed as Fixation index (F_{ST}, Weir & Cockerham, 1984) using a subset of 18,181 autosomal SNPs frequencies.

	W&C's F_{ST}		
Breed	Corriedale	Creole	Merino
Corriedale	0		
Creole	0.16	0	
Merino	0.08	0.15	0

Creole is supposed to be descendant of Churra, Manchega, Rasa or Canaria Spanish breeds. However, F_{ST} analysis between Churra and Merino showed a short distance (F_{ST}=0.023) when both breeds were compared (Gutiérrez-Gil *et al.*, 2012). In this study, Merino was more genetically distant from Creole, which has the Churra breed as one likely ancestor in South America. This larger distance suggests that Creole may not be a direct descendant of Churra, as it was proposed. Moradi *et al.* (2012) found low differentiation between sheep populations, with F_{ST} average values of 0.024 and 0.027 for the Zel-Lori Bakhtiari and HapMap data sets, respectively. These results are in agreement with a previous study in which 23 domestic breeds and two wild sheep subspecies were considered and revealed that sheep breeds showed generally low differentiation (Kijas *et al.*, 2009). Similar results have been reported when Creole cattle was compared to commercial cattle (Delgado *et al.*, 2011).

In conclusion, the high levels of polymorphisms of the OvineSNP50 in Corriedale, Merino and Creole sheep indicate that SNPs included in this chip could be used to characterize genetic diversity and clearly assign breed and detect population structure. This information could also become a helpful tool towards the

development of specific sets of SNPs for breed assignment (traceability) and parentage testing that could be used for commercial and conservation purposes. Results need to be validated with a representative number of animals to ensure accurate results previous to further applications.

2.5 **ACKNOWLEDGEMENTS**

The authors gratefully acknowledge DMV Daniel Castells (Uruguayan Secretariat of Wool) and DMV. Fernando Macedo (Universidad de la República) for their contribution to this study. Thanks also to staff at Glencoe Experimental Research Station (INIA) for data collection and flock management.

2.6 **REFERENCES**

Alam M *et al.* (2011). A Whole Genome Association Study to Detect Single Nucleotide Polymorphisms for Body Conformation Traits in a Hanwoo Population. Asian-Aust.*Journal of Animal Science*. 24,322-329.

Camacho Vallejo ME, Delgado Bermejo JV and Barba Capote CJ. (2000). Recursos genéticos de animales domésticos en España: situación actual y sistemas tradicionales de explotación. *Archivos de zootecnia.* Vol. 49, Nº 187.pags. 423-430 .

Castells D and Gimeno D. (2011). Selection of Corriedale sheep for resistance or susceptibility to nematode infection in Uruguay. XXIII Congreso Internacional de la Asociación Mundial para el avance de la Parasitología Veterinaria (WAVVP). Buenos Aires, Argentina 2011.

Delgado JV, Martínez AM, Acosta A, Álvarez LA, Armstrong E, Camacho E, *et al.* (2011). Genetic characterization of Latin-American Creole cattle using microsatellite markers. *Animal Genetics*. 43(1):2-10.

Ding C and Jin S. (2009). Single Nucleotide Polymorphisms. Human Press: Totowa, NJ. 578, 245-254.

Evanno G Regnaut Sand Goudet J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*. 14:2611–2620.

Fernández G. (2000). Situation of the local domestic genetic resources in Uruguay. *Archivos deZootecnia*. 49: 333-340.

Flori L, Fritz S, Jaffrézic, F, Boussaha M, Gut I, Heath S, Foulley JL and Gautier M. (2009). The genome response to artificial selection: a case study in dairy cattle. *PLoS One.* 4(8):e6595.

Gutiérrez JP and Goyache F. (2005). A note on ENDOG: a computer program for analysing pedigree information. *Journal of Animal Breeding Genetics*. 122**:** 357-360.

Gutiérrez-Gil B, García-Gámez E, Kijas JW, San Primitivo F, de la Fuente LF, Bayón Y, International Sheep Genomics Consortium (ISGC) and Arranz JJ. (2012). Preliminary analysis of selective sweeps in the Spanish Churra sheep. National Conference of Animal Genetic Improvement. Menorca (Spain), 04/31/2012.

Hayes BJ, Lien S, Nilsen H, Olsen HG, Berg P, MacEachern S, Potter S, Meuwissen, THE. (2008). The origin of selection signatures on bovine chromosome 6. *Animal Genetics*. 39(2):105-111.

Johnston SE, McEwan JC, Pickering NK, Kijas JW, Beraldi D, Pilkington JG, Pemberton JM, Slate J. 2011. Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. Mol Ecol. 20(12):2555-66.

OvineSNP50 Datasheet - Illumina. 2010. OvineSNP50 DNA Analysis Kit. Available: 12/05/2012 http://www.illumina.com/products/ovinesnp50_dna_analysis_kit.ilmn.

Kemper K, Emery D, Bishop S, Oddy H, Hayes BJ, Dominik S, Henshall Jand Goddard M. (2011). The distribution of SNP marker effects for faecal worm egg count in sheep, and the feasibility of using these markers to predict genetic merit for resistance to worm infections. *Genetics Research*. 93: 203-219.

Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, *et al.* (2012) Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. *PLoS Biology* 10(2).

Kijas JW, Townley D, Dalrymple BP, Heaton MP, Maddox JF, McGrath A, Wilson P, Ingersoll RG, McCulloch R, McWilliam S, Tang D, McEwan J, Cockett N, Oddy VH, Nicholas FW, and Raadsma H. (2009). A Genome Wide Survey of SNP Variation Reveals the Genetic Structure of Sheep Breeds. *PLoS*. 4, e4668.

Mendoza A. (1968) Raza Corriedale. En: Curso básico teórico práctico de lanares y lana. Comisión Honoraria de Mejoramiento de la Producción ovina. Ministerio de Ganadería y Agricultura, pp. 120-122.

Miller JM, Poissant, J, Kijas JW, and ColtmanDW. (2011). A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. *Molecular EcologyResources*. 11, 314–322.

Montossi F, De Barbieri I, Ciappesoni G, Ravagnolo O, De Mattos D, Pérez Jones J, Fros A, Grattarola M, Mederos A and Soares de Lima M. (2005). Núcleo fundacional de merino fino y superfino de la Unidad Experimental "Glencoe" - INIA Uruguay: una experiencia innovadora de mejoramiento genético asociativo y participativo. *Agrociencia*. IX-N°1. 609-616.

Moradi MH, Nejati-Javaremi A, Moradi-Shahrbabak M, Dodds KGand McEwan JC.(2012).Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. *BMC genetics*.13(1):10.

Patterson N, Price ALand Reich D.(2006). Population structure and eigenanalysis. *PLoS Genetics*. 2(12): e190.

Pritchard JK, Stephens M, and Donnelly P. (2000). Inference of population structure using multilocus genotype data. *Genetics*. 155:945–959.

Ramos M, Megens HJ, Crooijmans RPM, Schook LBand Groenen M. (2011). Identification of high utility SNPs for population assignment and traceability purposes in the pig using high-throughput sequencing. *Animal genetics*.42(6):613-20.

Våge DI, Husdal M, Kent MP, Klemetsdal G, Boman IA. 2013. A missense mutation in growth differentiation factor 9 (GDF9) is strongly associated with litter size in sheep. BMC Genet. 2;14:1.

Weir BS and Cockerham CC.(1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*. Vol. 38, No. 6, pp. 1358-1370.

Wiggans GR, VanRaden PM, Bacheller LR, Tooker ME, Hutchison JL, Cooper TAand Sonstegard TS.(2010). Selection and management of DNA markers for use in genomic evaluation. *Journal of Dairy Science*. 93:2287–2292.

White SN, Mousel MR, Herrmann-Hoesing LM, Reynolds JO, Leymaster KA, Neibergs HL, Lewis GS, Knowles DP. 2012. Genome-wide association identifies multiple genomic regions associated with susceptibility to and control of ovine lentivirus. PLoS One.;7(10):e47829.

Zeder MA.(2008). Domestication and early agriculture in the Mediterranean Basin: origins, diffusion, and impact. PNAS. 105:11597–11604.

Zenger KR, Khatkar MS, Cavanagh JA, Hawken RJ and Raadsma HW. (2007). Genome-wide genetic diversity of Holstein Friesian cattle reveals new insights into Australian and global population variability, including impact of selection. *Animal Genetics* 38: 7–14.

3 CANDIDATE GENES INVOLVED IN RESISTANCE/SUSCEPTIBILITY FOR GASTROINTESTINAL PARASITES IN CORRIEDALE SHEEP

A.N. Grasso[1], V. Goldberg[1], E.A. Navajas[2], P. Peraza[2], D. Castells[3], I. Aguilar[1], W. Iriarte[1], G. Rincon[4], D. Gimeno[3], J.F. Medrano[4], G. Ciappesoni[1].

[1] Meat and Wool National Program, Instituto Nacional de Investigación Agropecuaria, Las Brujas, Canelones, 90200, Uruguay.

[2] Biotechnology Unit, Instituto Nacional de Investigación Agropecuaria, Las Brujas, Canelones, 90200, Uruguay.

[3] Department of Animal Genetic Improvement, Secretariado Uruguayo de la Lana, Montevideo, 11800, Uruguay.

[4] Department of Animal Science, University of California Davis, CA 95616 USA.

Resistance to parasites in sheep

Key words: parasites resistance, sheep, RNA-seq, whole genome sequencing, Illumina BeadChip.

* Corresponding author. Tel.: +59899736233.

E-mail address: nicograsso26@gmail.com

A ser enviado a Genetics and Molecular Biology (formerly named Revista Brasileira de Genética/Brazilian Journal of Genetics - ISSN 0100-8455) is published by the Sociedade Brasileira de Genética (Brazilian Society of Genetics). Bajo el formato solicitado por la revista.

ABSTRACT

Gastrointestinal parasites (GIP) are a serious limitation affecting sheep production worldwide. Because of the selection complexities for resistance/susceptibility (R/S) to GIP, genetic markers could simplify assessment of R/S candidates in breeding programs. In Uruguay, divergent selection lines have been developed since 1998 in Corriedale sheep, based on expected progeny difference for fecal egg count (EPD FEC), leading to R and S lines. In the present study, a total of 54 R and 44 S sheep were selected by extreme values of EPD FEC and inbreeding coefficient (<0.04) to be genotyped with the Illumina OvineSNP50 BeadChip. Additionally, two whole genome sequencing (WGS) were performed from pooled DNA of 4 R and 4 S individuals and transcriptome sequencing (RNA-Seq) from messenger RNA (mRNA) was performed in six pools of mRNA of target parasites tissues and related ones (jejunum mucosa, duodenum, ileum, abomasum and abomasal lymph nodes). Population structure was assessed using STRUCTURE software, principal component analysis (PCA) and fixation index (F_{ST}). PCA and STUCTURE analysis were able to explain population stratification within lines and a moderate genetic distance (F_{ST}=0.042) between them. Association regression analysis indicated significant association for 8 genomic regions to EPD FEC. Those regions were deeply evaluated by combining WGS and RNA-Seq. Novel candidates genes were indentified, some of them showing expression, and new SNPs were also found.

3.1 **INTRODUCTION**

Parasitism, and particularly gastrointestinal parasites (GIP), are one of the most serious limitations affecting small ruminant worldwide production. In Uruguay, GIP represent one of the main sanitary and economic restriction for sheep production in pasture based systems (Castells *et al.,* 1995). The economic impact of GIP is also given by the prevention costs (i.e. antihelminthics), and production losses due to morbidity and mortality of infected animals (Sykes, 1994, Cesar *et al.,* 2010). Additionally, anthelmintics are rapidly becoming ineffective due to selection pressure over parasite populations leading to increase proportions of individual nematodes being able to resist those treatments (Cesar *et al.,* 2010). *Haemonchus contortus* is a blood-sucking parasite of the abomasum of sheep and is the most predominant parasite genus present in Uruguay (Castells, 2002). Chronic blood loss caused by *H. contortus* can result in anaemia, anorexia, reduction in body weight and wool growth (Karlsson and Greeff 2012, Simpson, 2000). Its high fecundity leads to rapid contamination of pastures with larvae and high levels of ingestion can lead to death (Karlsson and Greeff, 2012).

The resistance/susceptibility (R/S) to GIP is a difficult trait to measure due toof Fecal Egg Count (FEC) complexity and because infected animals are required, thus the availability of genetic markers would simplify assessment of candidates for selection in breeding programs.

Microsatellites (STRs) have been one of the main described markers which might be associated with quantitative trait loci (QTL) for parasites resistance in sheep. In fact, different significant associations between STRs and those QTLs have been reported in many chromosomes like OAR1, OAR3, OAR5, OAR6, OAR11, OAR12 and OAR15 (Silva *et al.*, 2012, Beh *et al.,* 2002, Benavides *et al.*, 2002, Hulme, *et al.*, 1993). On the other hand, single nucleotide polymorphisms (SNPs) are the most common DNA sequence variation in mammalian genomes. The abundance and genomic wide distribution of SNPs have resulted in their adoption as the marker of choice for association studies and identifying genes associated with complex

diseases. Numerous SNPs have been identified in the genomes of domestic animals. Moreover, a large number of commercial platforms are available for high throughput SNP genotyping (Ding and Jing, 2009). Recent attention has been focused on using dense panels of SNPs that are widespread throughout a genome. These SNP BeadChips are now commercially available for many animal species like sheep (54,000 OvineSNP50) (Zhang *et al*., 2012). The development of genomic tools such as the OvineSNP50 BeadChip would allow researchers to perform genome-wide association study (GWAS), improve the resolution of QTL, estimate allele frequency and effect of a SNP in sheep breeds.

Next-generation sequencing (NGS) can be used both for *de novo* sequencing of genomes (requiring sequence assembly) and for sample re-sequencing (Whole genome sequencing, WGS) that compares the resulting data to the reference sequence to discover variations present in the sample (Day-Williams and Zeggini, 2011). In addition to genomic variants, NGS is an ideal platform to investigate gene expression by RNA sequencing. Currently, the most widely used application of NGS in complex trait genetic studies involves SNP discovery and genotyping (Day-Williams and Zeggini, 2011).

Therefore, the aim of this study was to determine SNPs and genes associated with resistance/susceptibility to GIP in Corriedale sheep by using genome wide association studies, whole genome sequencing and RNA sequencing.

3.2 MATERIALS AND METHODS

3.2.1 Phenotypic measurements in Training population

The present work was performed in a Corriedale flock belonging to Ciedag, a research station of the Uruguayan Secretariat of Wool (SUL) which was divergently selected for R/S to GIP since 1998. The selection is based on the prediction of the genetic merit for this trait of each animal, in terms of the Expected Progeny Differences (EPD) for Fecal worm Egg Count (FEC).

The sampling protocol for the FEC genetic evaluation consists of two samples recorded after weaning, under natural mixed-species parasite challenge on pasture, as is described in Goldberg *et al.* (2011). The genetic evaluation is performed by a repeatability model as described by Ciappesoni *et al.* (2010) to obtain the FEC EPD. FEC was determined using the modified McMaster technique (Whitlock, 1948), with a sensibility of 100.

3.2.2 **DNA extraction and genotyped animals**

The animal resource used in this study comprised 325 sheep (216 R and 109 S) of three generations (2008, 2009 and 2010) from the divergent lines. Genomic DNA was extracted from blood of all sheep using (Medrano *et al.* 1990) protocol (with modifications). Animals were selected to be genotyped with the Illumina OvineSNP50 BeadChip according to three criteria.(1)-Animals were considered Resistant either Susceptible for this research by extreme values of EPD FEC. (2)- An average relatedness coefficient (AR) lower than 0.04 (ENGOG v4.6 software), in order to prevent false association results by inbreeding. Therefore 54 R and 44 S were selected from the whole resource. (3)- Only samples that presented call rates higher than 90% were considered for further analysis. Likewise, SNPs with call rates lower than 90% and minor allele frequency (MAF) lower than 0.01 were removed. Only autosomal markers were evaluated, to avoid X chromosome heterozygosity between females and males.

3.2.3 **Genetic population structure**

Subdivisions between lines were evaluated with the linkage model from the STRUCTURE software version 2.3.3 (Pritchard *et al.,* 2000), using a range of possible clusters (K) from two to nine. Ten runs of 10.000 iterations after a burn-in period of 10.000 iterations were performed for each K. Most probable K was identified by the ΔK method (Evanno *et al.,* 2005). Population structure was also

inferred by principal components analysis (PCA) from the marker data (see Patterson *et al.*, 2006). The genetic distances between lines were calculated using the Fixation Index (F_{ST}) (Weir and Cockerman, 1984), which assess the reduction in genotypic heterozygosity. The subset of autosomic SNPs (r^2<0.4) used in these analysis (STRUCTURE, PCA and F_{ST}) were in moderate linkage disequilibrium. These criteria were applied to achieve informative molecular markers for each line. Another F-statistics was also calculated to evaluate the level of inbreeding in each line (F_{IS}, Weir and Cockerham). To estimate F_{IS}, a set of SNPs that were in moderate linkage disequilibrium (r^2<0.4) typical of R and S were used.

3.2.4 **Association studies**

Before the association study a linkage disequilibrium pruning was performed, in order to avoid redundant markers information. Association analysis was developed using SVS Golden Helix Regression module from helix tree to test SNPs associations with phenotypic variables. Additive and recessive association models were used, and for each one case/control (S/R) and quantitative (EPD FEC) studies were carried out. Only SNPs with –log p-value>5 and False discovery rate (FDR) <0.2 were considered as associated.

3.2.5 **Whole genome sequencing and RNA sequencing**

A total of 8 animals were selected for genome sequencing, 4 R and 4 S. The selection criteria were: extreme values of FEC EPD, females, DNA quality (checked in Nanodrop® N8000 and agarose gel 1%), and AR lower than 0.04. DNA was sent to the Department of Animal Science, UC Davis, CA, USA, where two DNA pools for libraries construction were made for each line and contained equal amounts of DNA from each individual. Libraries were sent to Berkeley, CA, USA facilities where 100 bp pair end sequencing was performed.

The CLC Genomics Workbench (CLC Bio, Aarhus, Denmark) Reference Mapping function was used to assemble the trimmed reads to the OAR_V2.0 reference assembly. The SNP detection function was used in CLC Genomics Workbench, based on the neighborhood quality standard (NQS) algorithm, using default parameters.

Two animals were selected for RNA-sequencing, based on the same selection criteria previously described. Tissues collection occurred after euthanasia and involved the removal of jejunum, duodenum, ileum and abomasum mucosa and abomasal lymph nodes (one sample per tissue per animal). Samples were placed in an RNA stabilization reagent (RNAlater; Qiagen, Hilden, Germany) for storage at -80°C. Total RNA was isolated from the tissues using TRIZOL INVITROGEN according to the manufacturer's instructions. The concentration and integrity of each RNA sample was determined by an Automated Electrophoresis System Experion^{TM} by BIO-RAD. One library for each candidate tissue, and an one extra library in pool from all tissues were performed for both animals (R and S). All libraries were prepared using the mRNA-Seq sample prep Kit (Illumina, San Diego, CA USA) according to the manufacturer's instructions, and send to the Sequencing facilities of the Davis University CA. Sequencing was performed on a Genome Analyzer IIx (Illumina) for 40 cycles. Obtained 40bp reads were mapped all together to OAR_V2 with the CLC Genomics Workbench Reference Mapping function.

All experiments were conducted following national regulations for the care and use of animals and approved by the Honorary Committee of Animal Experimentation (CHEA) of the University of the Republic of Uruguay.

3.2.6 **Integrated bioinformatics: Association studies, Whole genome sequencing, RNA sequencing and databases.**

Results from the association studies were used to find candidate genes, genome regions and novel SNPs that contribute to R/S. All information from association studies, whole genome sequencing and RNA-seq was combined in the CLC

Genomics Workbench Genomics Gateway from CLC Bio. All data were tied to a genomic coordinate that was represented as tracks: ovine V2.0 reference genome sequence (track 1), whole genome read mapping from R (track 2) and S (track 3), annotation (track 4), RNA-seq read mapping (track 5) and SNP probes from the OvineSNP50 (track 6). In order to visualize several tracks together, a Track List was created.

In addition, the Sheep Genome Browser OarV2.0 from the Livestock Industry division of CSIRO (Commonwealth Scientific and Industrial Research Organization, http://www.livestockgenomics.csiro.au/sheep/oar2.0.php) was also used. It provides public access to sheep genome browsers annotated with SNP, Marker and reference sequence annotations from sheep and cattle. For genomics regions with no annotation but with gene expression, a BLAST search was performed in order to detect gene identity.

3.3 **RESULTS**

3.3.1 **DNA extraction and Genotyping**

A total of 98 sheep were genotyped using the Illumina OvineSNP50 BeadChip. Call rates for all samples were over 90%, so none was discarded. After quality filters described in methods section were applied, a total of 48,645 SNPs remained. An additional subset of 1,548 SNPs with an observed low minor allele frequency (MAF) (<0.1%) were also discarded. Thus, a total of 46,662 autosomal SNPs were maintained for the subsequent analysis.

3.3.2 **Genetic population structure**

To study the population stratification using F_{ST}, PCA and STRUCTURE, a subset of SNPs were selected that not exceeded an r^2 linkage equilibrium threshold of 0.4. This preserved 28,653 SNPs from the 46,662 for further population analysis.

Differentiation between lines based on SNP allele frequencies was tested with the F_{ST} statistic, which describes the proportion of variance within a species that is due to population subdivision. In general the differentiation between subpopulations was low, with an F_{ST} value of 0.042.

Principal components analysis (PCA) was used to explore the relationship and subdivision between R and S lines groups. The two largest components (PCA1 and PCA2) clustered animals from the same line together, indicating a clear genetic division among R and S divergent lines (Fig. 1). The assignment of animals into the two groups was considered during the analysis. In fact, three animals that were supposed to belong to the R cluster were located in the S cluster (Fig 1 red dots marked with an arrow in the green cluster) and the same case occurred with three S sheep (green dots marked with an arrow in the red cluster). These animals were considered as outliers and were not taken into account for association analysis. These animals could be outliers probably to a wrong identification or mislabeling.

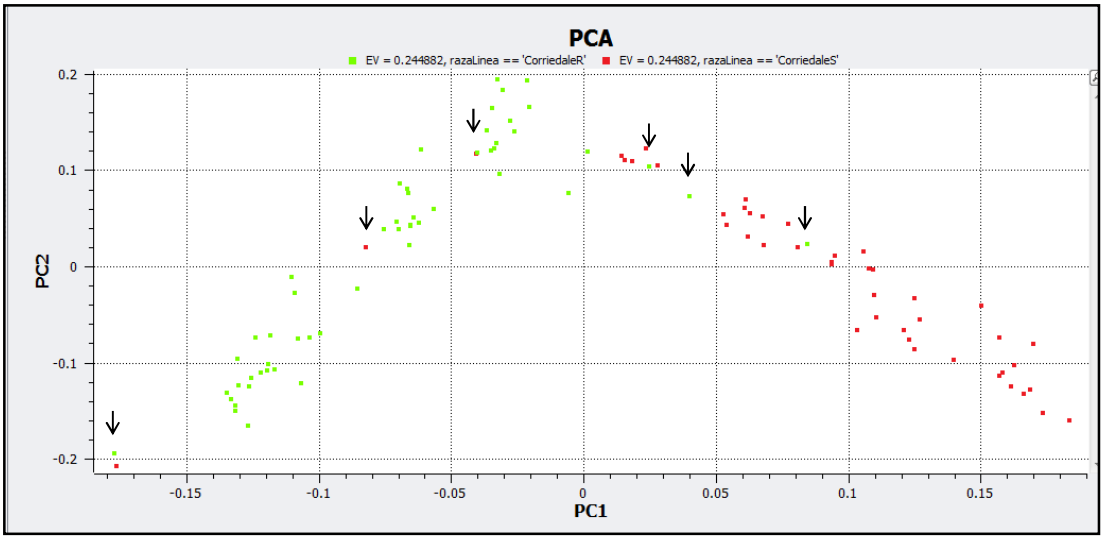

Figure 1. Evaluation of population stratification by Principal Component Analysis using 28,653 SNPs with moderate linkage disequilibrium (R2<0.4) in Corriedale divergent lines selected by resistance and susceptibility to parasites. Susceptible and resistant animals are represented by green and red dots respectively. Animal categorized as outliers are signaled by a black arrow.

A model-based Bayesian clustering (STRUCTURE) was also performed to determine population admixture and structure. It partitioned the genomic information given by the SNPs of each individual into a number of components (K) represented by K colors. The best K was determined by ΔK method resulting in K=4. Exploratory analysis of K=2 and K=3 were also assessed. At K = 2, the R and S appear to carry the same two ancestral components that are both present in each line, but R line is mainly represented by the green cluster (0.704) meanwhile S is by the blue one (0.763) (Fig. 2). At K = 3, both lines still are represented by a major cluster each, nevertheless they keep sharing a cluster that was presented in equal proportions.

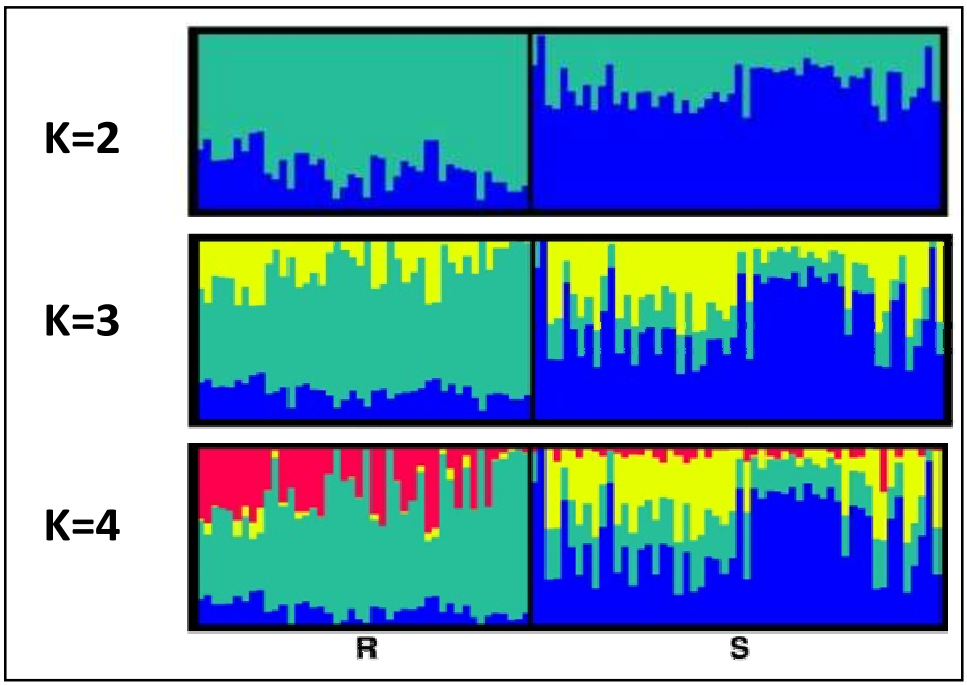

Figure 2. Bar Plot-Population structure analysis estimated with 28.653 SNPs with moderate LD (R2<0.4) for two to four clusters. Each individual is represented by a vertical line, which is divided by K colors. Each color segment represents the estimated fraction of belonging to each cluster. The black line separates individuals from different lines, identified by R= Resistant and S =Susceptible.

Is not until K=4 that R carry an ancestral component largely absent from S animals, and the same occurs with cluster that is almost exclusive in the S line (Figure 2) This likely reflects the past genetic contribution of a line into the other and that divergent selection led to markedly genetic differences. Finally F_{IS} values for R and S were -0.05 and -0.05 for both lines, reflecting no inbreeding.

3.3.3 **Association studies**

The profiles of the P-values (in terms of –log10(P)) of the tested SNPs for FEC EPD in the association analysis are shown as Manhattan plots in supplementary material (SS1). A total of 390 SNPs for FEC EPD were detected by the four association

models based on the previously defined criteria for association. Nevertheless none of those SNP showed high effect over the trait in study (data not shown). Therefore, among these significant SNPs, 17 SNP with the most significant association value (-log10 P value) were used to indentify 8 candidate chromosomes regions to be deeply evaluated (Table 4). A summary of the 17 SNPs identified, their name, their map locations and their -log P-values are reported in Table 1. False discovery rate of these SNP ranged from $5.72E^{-18}$ to $6.58E^{-04}$.

.

It is interesting to note that SNPs significant for one model were in general significant for the other models(although the significance levels varied with model). Allele substitution effect varied between models, but is in agreement with the effect of the SNP in the population studied. With the exception of Model A, SNP in all other models show low effects. The assumption that is made in Model A, where R and S animals are consider as case controls, and EPD FEC is taken under a recessive approach, is the cause that allele substitution effects are that high.

Table 1. Results from the genotype association test showing the 17 SNP with highest levels of association that defined the genomic region to be deeply evaluated. Genotype association models: A- Recessive case/control, B-Recessive EDP FEC, C-Additive model case control and D- Additive model EPD FEC.

		Association Models							
OAR	SNP	Model A		Model B		Model C		Model D	
		-log10P	Ase	-log10P	Ase	-log10P	Ase	-log10P	Ase
1	OAR1_150528301,1	10.54	4.34	12.63	0.06	11.03	2.55	13.98	0.34
	OAR1_213617854,1	--	--	9.56	0.07	9.58	2.63	10.25	0.30
2	s41345,1	--	--	13.22	0.06	16.70	4.07	16.99	0.35
	OAR2_43175231,1	--	--	12.15	0.07	16.10	4.05	17.91	0.36
	OAR2_43304314,1	9.01	-37.87	8.53	0.07	16.48	-4.37	19.27	-0.37
3	OAR3_227769239,1	10.59	3.83	11.17	0.06	11.37	2.42	12.58	0.31
	OAR3_235574907,1	--	--	10.54	0.07	12.11	2.49	13.95	0.32
4	OAR4_50370017,1	8.81	4.04	8.73	0.07	10.02	2.29	9.58	0.29
	s54836,1	9.08	-37.31	7.91	0.07	11.94	-2.89	12.51	-0.33
13	OAR13_9727532,1	11.15	4.44	11.97	0.06	11.18	2.38	11.96	0.31
	s14792,1	9.08	-37.31	7.56	0.07	12.34	-2.69	13.11	-0.32
15	OAR15_12066235,1	8.69	-3.93	8.95	0.07	11.50	-2.45	13.01	-0.32
	OAR15_12128190,1	--	--	6.26	0.08	14.12	-3.27	15.46	-0.34
20	OAR20_38864697,1	13.92	4.96	15.92	0.06	13.36	2.75	13.86	0.31
	OAR20_31086401,1	8.81	3.52	8.72	0.07	13.00	2.54	14.09	0.31
25	OAR25_38239942_X,1	8.44	3.20	10.21	0.07	11.72	2.42	14.12	0.32
	s50241,1	10.54	4.34	12.10	0.07	13.01	2.67	15.36	0.33

Note: -log10P value from the genotype association test

Ase allele substitution effect

3.3.4 **Whole genome sequencing and RNA sequencing**

Sequences of 100 bp reads from each pool were trimmed and mapped to the autosomes and X chromosome of the reference sheep genome (OAR V.2). A total of 84.3% and 84.3% of the reads from R and S pools respectively were aligned to the reference genome, for both lines (Table 2), resulting in an average of sequence coverage of the sheep's genome of 10,24x (min 0 max 896,601) and 11.71x (min 0 max 1,118,430) for R and S respectively.

Table 2. Mapping results from the whole genome sequencing analysis of two DNA pools of 4 resistant and 4 susceptible sheep to gastrointestinal parasites. The genome assembly was performed with the OAR V.2 sheep genome.

Reads	**susceptible Pool**	**resistant Pool**
Totals	357,752,380	312,811,656
Mapped	301,937,019	263,941,871
Unmapped	55,815,361	48,869,785

From the total of the RNA-seq reads, 34% did not map to the reference genome. The average coverage achieved was 1.39x (min 0 and max 429,504).

Table 3. RNA sequencing results combined from 40bp read pools from jejunum, duodenum, ileum and abomasum mucosa and abomasal lymph nodes. The mapping was assessed with the OAR V.2 sheep genome.

Reads	RNA-seq Pool
Total	140.820.015
Mapped	92.360.108
Unmapped	48.459.907

3.3.5 **Integrated bioinformatics: Association studies, Whole genome sequencing, RNA sequencing and databases.**

Data from the association studies, whole genome sequencing, and RNA-seq was combined and complemented with the annotation and reference of the genome OAR V2.0. This track list, lead us to explore the regions determined by the associated SNP, finding out the nearest annotated or only expressed gene, or both (Table 4, Figure 3a).

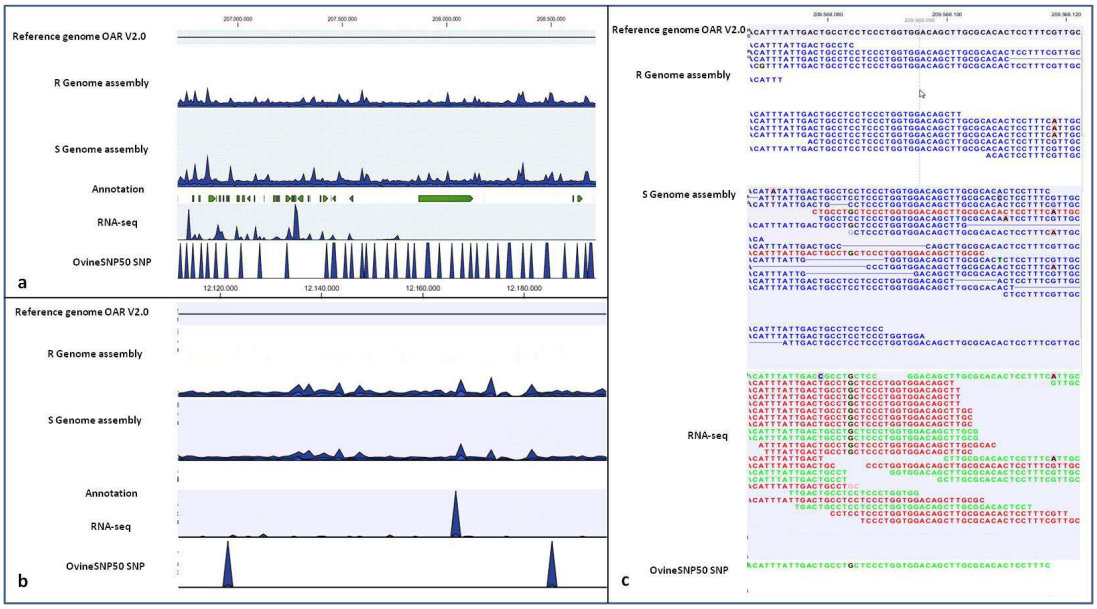

Figure 2. a- Illustrative figure of the track list of the combined techniques (integrative approach). A region of the ovine genome is shown, the first track corresponds to the reference genome OAR V2.0, second and third track to the resistant and susceptible sheep genome assembly, fourth to the annotation, fifth to the RNA-seq results, and sixth to the SNPs from the Illumina OvineSNP50 BeadChip. b- Integrated results from association studies, whole genome sequencing from resistant and susceptible sheep, RNA-seq, annotation, and OvineSNP50 Illumina probes from associated SNP. This genome region is determined by a SNP of the OAR15, showing no annotated genes although there is gene expression. c- One of the associated SNPs from the OAR3, which is in a region with expression but no annotation, showing that there are genotypic differences between the R and S genome assembly.

Genes with no annotation but with high RNA-seq coverage (Fig 3) where identified using BLAST of the reads.

3.4 **DISCUSSION**

The aim of this research was to detect associated SNPs, genome regions and candidate genes to resistance/susceptibility to gastrointestinal parasites in Corriedale divergent selection lines for this trait. There are several reports of QTL and candidate gene detection, but most of them used crossed breeds and STR as markers. To our knowledge, there are only three works (Riggio *et al.*, 2013, Salle *et al.*, 2012, Kemper *et al.*, 2011) that used high throughput SNPs genotyping, and all of them used different approaches. The present research contains novel integrative approaches that would be useful in subsequent investigations aiming to resolve the effects of molecular markers on susceptibility or resistance to GIP in sheep.

3.4.1 **Genetic population structure**

A total of 28,653 SNPs were able to explain population structure of divergent lines determined by F_{ST}, STRUCTURE and PCA methods. Genetic differences and variation detected from allele frequency between R and S lines could be explained by the fact that Corriedale flock has been divergently selected for resistance to parasites. Moderate genetic distance was observed between R and S lines. This genetic variation reached after ten years of divergent selection could be compared with genetic differences between two breeds. In fact, Moradi *et al.*, (2012) found low differentiation between sheep populations, with F_{ST} average values of 0.024 and 0.027 for the Zel-Lori Bakhtiari and HapMap data sets, respectively. These results are in agreement with a previous study in which 23 domestic breeds and two wild sheep subspecies were considered and revealed that sheep breeds showed generally low differentiation (Kijas *et al.*, 2009).

Genetic relatedness and substructure within populations should be taken into account in order to avoid spurious associations in GWAS (Zenger *et al.*, 2007). The correction through population stratification (Lander and Schork, 1994) and the removal of outliers are alternatives to reduce false positive results in association studies. Results from population structure (Fig 1) showed outliers that were not considered for further association analysis. Besides no correction based on

population structure was applied, because in some instances, correcting for population structure could caused several markers and genes to lose significance (Wilson *et al.*, 2004).This loss of significance can have two main reasons: (1) this was a non functional polymorphism and the association was caused by population structure, or (2) the polymorphism is functionally related although the polymorphism distribution coincides with population structure (Wilson *et al.*, 2004). The second option would be the case of the divergent lines of this research, and correcting the association study could outcome in a false negative result and loss SNPs effect. Moreover F_{IS} index showed that there were not unbalanced allele frequencies within the lines due to population structure.

3.4.2 **Integrated bioinformatics: Association studies, Whole genome sequencing, RNA sequencing and databases.**

With respect to the SNP association results, the eight most highly significant SNP were identified for EPD FEC on OAR1, 2, 3, 4, 15, 20 and 25 (Table 1) and novel candidate genes were identified by the integrative approach (Table 4).

Table 4. Each associated SNP with their position are presented here. The nearest annotated gene in the ovine genome (Sheep Refseq) or in the Bovine genome (Bovine Refseq) to each SNP is reported. If there was expression (results from RNA-seq) but no annotation, BLAST was used to identify the genes. Positions and distances are in base pairs, and expression is exhibited as coverage in 1kbp of the target region .

OAR	SNP	SNP Position	Nearest Gene	Gene Position	Distance from the SNP	Sheep Refseq	Cattle Refseq	RNA-seq read Blast	Expression
1	OAR1_150528301	139331280	ubiquitin specific protease 25 (USP25)	139506254	174974	none	Refseq: XM_602857	NO	NO
	OAR1_213617854	198154177	somatostatin (SST)	198115131	39046	SHEEP_ENSP0000028641	Refseq: NM_173960	NO	124.1 (0-985)
2	s41345	102077706	prepronociceptin (PNOC)	102077706	within	SHEEP_ENSBTAP0000000695	Refseq: NM_174150	NO	NO
	OAR2_43175231	41543182	Disintegrin and metalloproteinase domain 28 (ADAM28)	41455282	87900	SHEEP_ENSBTAP0000004880	Refseq: XM_583562	NO	1.2 (0-124)
3	OAR3_227769239	209568058	poly (ADP-ribose) polymerase family, member 11 (PARP11)	209568058	within	none	Refseq: NM_001082452	NO	1.8 (0-20)
4	s54836	12189129	Ryruvate dehydrogenase kinase, isozyme 4 (PDK4)	12162678	26451	SHEEP_ENSBTAP0000003686	Refseq: NM_001101883	NO	106 (0-543)
13	OAR13_9727532	8764263	MACRO domain containing 2 (MACROD2)	9191022	426759	SHEEP_ENSBTAP0000004573	Refseq: NM_001105029	NO	NO
	s14792	32662299	Rho GTPase activating protein 12 (ARHGAP12)	32662299	within	SHEEP_ENSP00000345808	Refseq: NM_001102241	NO	6.9 (0-76)
15	OAR15_12066235	12121262	translation initiation factor 4E (EIF4E)	12166462	45200	none	none	XM_0040	3.2 (0-42)

								09683.1	
20	OAR20_38864697	35055068	SRY (sex determining region Y)-box 4 (SOX4)	35038455	16613	SHEEP_ENSP00000244745	Refseq: NM_001078128		1- 56.5 (0-333)
	OAR20_31086401	28138931	olfactory receptor Olr1690	28138931	within	SHEEP_ENSBTAP00000048959	Refseq: XM_001788659	NO	NO
25	OAR25_38239942_X	35219094	neuregulin 3 (NRG3)	35286936	67842	none	Refseq: XM_001788304	NO	NO

A QTL for FEC was previously reported by Marshall *et al.*, (2008) using a scan with microsatellite markers in a linkage analysis in Merino sheep. This QTL in OAR1 was confirmed by our results using the integrative approach. The associated genomic region determined by the first SNP (OAR1_213617854) was close to the somatostatin (SST) gene. Somatostatin is one of the most important genes in regulating animal growth and development, where it plays a negative role in animal growth regulation by inhibiting the release of growth hormone (GH) (Gao *et al.,* 2011). In fact, genetic correlations between body weight and indicator resistance traits were reported by Bishop *et al.*, (1996), Ciappesoni *et al.*, (2010). In addition, SST down-modulates a number of immune functions, among others lymphocyte proliferation, immunoglobulin production and the release of inflammatory cytokines such as interferon gamma (IFNG) (ten Bokum *et al.,* 2000). The second genomic region defined by the OAR1_150528301 associated SNP was close to the Ubiquitin-specific protease 25 (USP25) and was identified as a candidate gene (Table 4). It has been demonstrated in mice that USP25 is a negative regulator of Interleukin 17 (IL-17) which plays an important role in infection and autoimmunity (Zhong *et al.,* 2013).

Two novel genomic regions on OAR2 were detected in this research for EPD FEC. From the integrative approach the gene ADAM28 (a disintegrin and metalloproteinase) was close to the OAR2_43175231 and OAR2_43304314 associated SNP and it is expressing (Table 4). It has been demonstrated that ADAM28 mediate leukocyte adhesion to endothelial cells (Shimoda *et al.,* 2007). On the other hand, Riggio *et al.,* (2013) detected QTLs for *Nematodirus* that are coincident with our results. As it was proposed by these authors, similar genetic mechanisms could be underlying the resistance either to *Strongyles* and *Nematodirus*. In fact, high genetic correlation between *Nematodirus* and *Strongyles* FEC was reported (Bishop *et al.,* 2004).

Many studies showed evidence of a QTL on OAR 3 near interferon gamma (IFNG) (Beh *et al.,* 2002; Davies *et al.,* 2006, Stear *et al.,* 2009) which appears to be the main candidate gene for resistance/susceptibility to GIP. Nevertheless, Silva *et al.,*

2012, reported that the IFNG region did not produce any particular evidence of linkage in their study, where associated genomic regions foraverages of last measurements for FEC (AVFEC) and blood packed cell volume at the start of the parasite challenge period PCVST were at position 260.0 on OAR 3 and were located more than 60 cM from IFNG. In addition, in more recent genome wide association studies, Salle´*et al.,* (2012) and Riggio *et al.,* (2013) reported a QTL on OAR 3 associated with worm burden and FEC, and Nematodirus FEC respectively. However no causative mutation was identified for these OAR3 QTL. Moreover, Riggio *et al.,* (2013) identified two SNPs on OAR 3 associated with immunoglobulin A (IgA), which were also significant for FEC. All of this is in agreement with our results. The associated SNP OAR3_227769239 is within the Poly (ADP-ribose) polymerase (PARP) PARP-11 gene, which is also showing expression (Table 4, Fig 3c). The PARP family has been shown to be involved in many cellular functions such as transcription regulation, telomere cohesion, mitotic spindle formation during cell division, inflammation, cell death, and DNA repair (AbdElmageed *et al.,* 2012). Recent works have highlighted the novel role of PARP-1 in regulating the intracellular trafficking of key cellular proteins such as p53 and nuclear factor-kappa B (NF-κB). Literature has revealed that ADP-ribosylation reactions may play important roles in cellular trafficking during inflammation, cell death, and DNA repair (AbdElmageed *et al.,* 2012). Additionally, Warrener *et al.,* (2012) have reported that PARPs are involved in interferon pathway activation and enhanced by IFNG. Shorts isoforms of PARPs, play key roles on induction of IRF3, interferon-α (IFN alfa), IFN beta (Hayakawa *et al.,* 2011), other cytokines after viral infection and interferon-γ (Warrener *et al.,* 2012).

Results from Salle *et al.* (2012), determined a 2-Mbp region between 70.2 and 72.2 Mbp on OAR13 which was significantly associated with FEC in every analysis they used. The same genomic region was detected in our analysis with the significant SNP OAR13_9727532. The closest gene to the associated marker was MACROD2 (Table 4), with no evident related biological function to EPD FEC described yet.

The region in OAR15 reported by Salle *et al.* (2012), for nematode resistance using a partial genome scan with microsatellite markers in a linkage analysis were also confirmed with the two significant SNPs in our research (Table 1). Using CSIRO sheep genome browser, (http://www.livestockgenomics. siro.au/cgi-bin/gbrowse/oarv2.0/), there appear to be no obvious candidate genes within this region. Nevertheless, between our two SNPs, RNA-seq results showed expression of a gene that was not present in the annotation (Figure 3b). However, Blast from the RNA-seq reads revealed the translation initiation factor 4E (EIF4E). This gene is involved in the translation of antigenic peptides that interact in the MHC class I pathway of the immune system (Apcher, *et al.,* 2011). Some of the regions previously reported by Salle *et al.,* 2012 for nematode resistance on OAR20 were also confirmed by our analysis. The significant SNP OAR20_31086401 reported here was close to olfactory receptor Olr1690 gene (Table 4).

3.4.3 **Conclusion**

Most of the genomic regions reported in this work have been previously described as associated to resistance/susceptibility to GIP. Nevertheless, the candidates genes identified were not reported before. Even if markers that were identified are likely to be only a subset of a larger number of polymorphisms affecting this trait, or are not a main causative mutation (if there is any), the results are promising and offer an interesting start point for future research. Comparison between studies is complex because of the different approaches used and the nature of this highly polygenic trait. Differences between populations, protocols of FEC records, among breeds, and so on, make these studies even harder to compare. However, the fact that several genomic regions consistently show association in different studies suggests that further research may shed light on this subject. Most studies have been focused on low density STR linkage analysis. Recently, some studies using higher density SNP panels have become more common. Nevertheless, the OvineSNP50 has still a low coverage over the ovine genome, therefore using next generation techniques and higher SNP density chips would help us to identify effective molecular markers.

Finally, a low density and cost effective SNP panel of QTL to EPD FEC would be very valuable for genetic improvement programs in sheep.

3.5 **ACKNOWLEDGMENTS**

This project was financed by the FCI program from the National Institute of Agriculture Research of Uruguay. We thank Alma Islas from the Animals Genetics department, UC Davis for the advices during the project.

3.6 **REFERENCES**

AbdElmageed ZY, Naura AS, Errami Y and Zerfaoui M. (2012). The poly(ADP-ribose) polymerases (PARPs): new roles in intracellular transport. Cellular signalling 24, 1–8.

Apcher S, Daskalogianni C, Lejeune F, Manoury B, Imhoos G, and Heslop L. (2011). Major source of antigenic peptides for the MHC class I pathway is produced during the pioneer round of mRNA translation, PNAS 108(28). doi:10.1073/pnas.1104104108.

Beh KJ, Hulme DJ, Callahan MJ, Leish Z, Lenane I, Windon RG and Maddox JF. (2002). A genome scan for quantitative trait loci affecting resistance to Trichostrongylus colubriformis in sheep.Animal Genetics 33, 97–106.

Benavides MV, Weimer TA, Borba MFS, Berne MEA and Sacco AMS (2002). Association between microsatellite markers of sheep chromosome 5 and faecal egg counts. Small Rumiant Research, 46: 97-105.

Bishop SC, Bairden K, McKellar QA, Park M, Stear MJ (1996). Genetic parameters for faecal egg count following mixed, natural, predominantly Ostertagtia circumcincta infection and relationships with live weight in young lambs. Animal Science. 63: 423–428.

Bishop SC, Jackson F, Coop RL, Stear MJ. (2004). Genetic parameters for resistance to nematode infections in Texel lambs and their utility in breeding programmes. Animal Science 78: 185–194.

Castells D, Nari A, Rizzo E, Mármol E and Acosta D. (1995). Efectos de los nematodos gastrointestinales sobre los parámetros productivos del ovino en la etapa de recría. Producción Ovina, (8): 17-32.

Castells D. (2002). Nuevo enfoque en el control parasitario de ovinos In: Resistencia genetica del ovino y su aplicacion en sistemas de control integrado de parasitos ; FAO Animal Production and HealthPaper (FAO) Castells Montes, D. / FAO, Rome (Italy). Direccion de Produccion y Sanidad Animal, p. 17-24.

Ciappesoni G, Gimeno D, Ravagnolo O. (2010). Genetic Relationships between Faecal Worm Egg Count and Production Traits in Merino Sheep of Uruguay. Proc. 9th World Congress Gen. Appl. Livest. Prod., Leipzig, Germany.

Davies G, Stear MJ, Benothman A, Kerr A, Mitchell S and Bishop SC. (2006). Quantitative trait loci associated with parasitic infection in Scottish blackface sheep. Heredity 96, 252–8.

Day-Williams AG and Zeggini E. (2011). The effect of next-generation sequencing technology on complex trait research. European journal of clinical investigation, 41(5), 561–7. doi:10.1111/j.1365-2362.2010.02437.x.

Ding C, Jin S. (2009). Single Nucleotide Polymorphisms. 578, 245-254. Humana Press: Totowa, NJ.

Evanno G, Regnaut S and Goudet J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study, Mol, Ecol,, 14:2611–2620.

Gao L, Zan LS, Wang HB, Hao RJ and Zhong X. (2011). Polymorphism of somatostatin gene and its association with growth traits in Chinese cattle. Genetics and molecular research : GMR, 10(2), 703–11. doi:10.4238/vol10-2gmr1029.

Hulme DJ, Blattman AN, Kinghorn BP, Woolaston RR, Gray GD and Beh KJ (1993). A search for associations between major histocompatibility complex restriction fragment length polymorphism bands and resistance to Haemonchus contortus infection in sheep. Anim. Genet., 24: 277-282.

Karlsson, LJE and Greeff JC. (2012). Genetic aspects of sheep parasitic diseases. Veterinary parasitology, 189(1), 104–12. doi:10.1016/j.vetpar.2012.03.039.

Kemper KE, Emery DL, Bishop SC, Oddy H, Hayes BJ, Dominik S, Henshall J M, *et al.* (2011). The distribution of SNP marker effects for faecal worm egg count in sheep, and the feasibility of using these markers to predict genetic merit for resistance to worm infections. Genetics Research, 93(03), 203–219..

Kijas JW, Townley D, Dalrymple BP, Heaton MP, Maddox JF, McGrath A, Wilson P, Ingersoll RG, McCulloch R, McWilliam S, Tang D, McEwan J, Cockett N, Oddy VH, Nicholas FW, and Raadsma H. (2009). A Genome Wide Survey of SNP Variation Reveals the Genetic Structure of Sheep Breeds. *PLoS*. 4, e4668.

Medrano JF, Aasen E, Sharrow L. (1990). DNA extraction from nucleated red blood cells. Biotechniques.;8(1):43

Moradi MH, Nejati-Javaremi A, Moradi-Shahrbabak M, Dodds KG y McEwan JC. (2012). Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. BMC genetics.13(1):10.

Lander ES and Schork NJ. (1994). Genetic dissection of complex traits. Science 265, 2037–2048.

2010. OvineSNP50 DNA Analysis Kit. En línea 5 de diciembre de 2012. http://www.illumina.com/products/ovinesnp50_dna_analysis_kit.ilmn

Patterson N, Price AL and Reich D. (2006). Population structure and eigenanalysis. PLoS Genetics. 2(12): e190.

Riggio V, Matika O, Pong-Wong R, Stear MJ and Bishop SC. (2013). Genome-wide association and regional heritability mapping to identify loci underlying variation in nematode resistance and body weight in Scottish Blackface lambs. Heredity, 110(5), 420–9. doi:10.1038/hdy.2012.90.

Pritchard JK, Stephens M, and Donnelly P. (2000). Inference of population structure using multilocus genotype data. Genetics, 155:945–959.

Salle´ G, Jacquiet P, Gruner L, Cortet J, Sauve´ C, Pre´vot F *et al.* (2012). A genome scan for QTL affecting resistance to Haemonchus contortus in sheep. J AnimSci 90:4690–4705.

Shimoda M, Hashimoto G, Mochizuki S, Ikeda E, Nagai N, Ishida S and Okada Y. (2007). Binding of ADAM28 to P-selectin glycoprotein ligand-1 enhances P-selectin-mediated leukocyte adhesion to endothelial cells. The Journal of biological chemistry, 282(35), 25864–74. doi:10.1074/jbc.M702414200.

Silva MVB, Sonstegard TS, Hanotte O, Mugambi JM, Garcia JF, Nagda S, Gibson JP, *et al.* (2012). Identification of quantitative trait loci affecting resistance to gastrointestinal parasites in a double backcross population of Red Maasai and Dorper sheep. Animal genetics, 43(1), 63–71. doi:10.1111/j.1365-2052.2011.02202.x.

Simpson HV. (2000). Pathophysiology of abomasal parasitism: is the host or parasite responsible? Vet. J. 160, 177–191.

Stear MJ, Boag B, Cattadori I, Murphy L. (2009). Genetic variation in resistance to mixed, predominantly Teladorsagia circumcincta nematode infections of sheep: from heritabilities to gene identification. Parasite Immunol 31: 274–282.

Hayakawa S,Shiratori S, Yamato H, *et al.* (2011). ZAPS is a potent stimulator of signaling mediated by the RNA helicase RIG-I during antiviral responses. Nature Immunology. 12, 37–44. (2011).

Sykes AR. (1994). Parasitism and production in farm animals. Anim. Prod., 59: 155-172.

ten Bokum AM, Hofland LJ, van Hagen PM. (2000). Somatostatin and somatostatin receptors in the immune system: a review. Eur Cytokine Netw. 2000 Jun;11(2):161-76.

Warrener P, Kim S, Williams SMG, Biery M, Gordon M. (2012). Synthetic lethality of PARP inhibition in BRCA-network disrupted tumor cells is associated with interferon pathway activation and enhanced by interferon-γ Apoptosis, Volume 17, Issue 7, pp 691-701 - Springer.

Weir BS and Cockerham CC. (1984). Estimating F-Statistics for the Analysis of Population Structure. Evolution. Vol. 38, No. 6, pp. 1358-1370.

Wilson LM, Whitt SR, Iba AM, Iv ESB. (2004). Dissection of Maize Kernel Composition and Starch Production by Candidate Gene Association. October 16, 2719-2733.

Zenger KR, Khatkar MS, Cavanagh JA, Hawken RJ and Raadsm HW. (2007). Genome-wide genetic diversity of Holstein Friesian cattle reveals new insights into Australian and global population variability, including impact of selection. *Animal Genetics* 38: 7–14.

Zhang H, Wang Z, Wang S, Li H. (2012). Progress of genome wide association study in domestic animals. Journal of animal science and biotechnology 3, 26.

Zhong B, Liu X, Wang X, Chang SH, Liu X, Wang A, Joseph M, *et al.* (2013). NIH Public Access, 13(11), 1110–1117. doi:10.1038/ni.2427. Negative regulation of IL-17-mediated signaling and inflammation by ubiquitin-specific protease 25.

4 CONCLUSIÓN

4.1 DISEÑO DE UN PANEL DE BAJA DENSIDAD DE SNP PARA VALIDACIÓN

Finalmente, se escogieron 4 regiones genómicas determinadas por SNP asociados en los cromosomas OAR2, OAR13, OAR15 y OAR20. Estas fueron evaluadas en detalle obteniendo 236 SNP de carácter posicional y funcional. Asimismo, se identificaron 15 SNP dentro de los genes candidatos USP25, SST, PARP11, SERHL2, NFAM1 y PDK4.

El total de 251 SNP candidatos para el carácter en estudio, fueron empleados para diseñar un panel de baja densidad, que en un futuro será evaluado en una población de validación.

Para cada uno de los SNP escogidos para el panel de validación, se diseñaron nuevas sondas (150pb hacia 5' y 3') para el nuevo genotipado. Asimismo, se identificó el SNP de interés entre paréntesis rectos y su alelo separado por una barra inclinada. En estas 150bp hacia 3' y 5', se indicó si existieron nuevos SNP además del escogido mediante el código de IUPAC.

4.2 CONCLUSIÓN Y CONSIDERACIONES FINALES

Las variantes encontradas mediante los estudios de asociación mostraron tener efectos individuales diferentes sobre la población en estudio, dependientes del modelo de asociación utilizado. Además combinando las tres tecnologías (genotipado masivo con el OvineSNP50, secuenciado de todo el DNA genómico y RNA-seq), las enormes regiones genómicas establecidas como asociadas presentan muchos marcadores, de los cuales solamente unos cientos debieron ser escogidos para el diseño de un panel de baja densidad que sea capaz de diagnosticar la

resistencia o susceptibilidad de la raza Corriedale a parásitos gastrointestinales. Sin embargo, el diseño de paneles de un número reducido de marcadores posicionales y funcionales obtenidos de la combinación de las tres tecnologías, que sean capaces de explicar gran porcentaje de la varianza fenotípica en una población de validación, sería una opción valiosa a investigar para que haya una mayor adopción de estas tecnologías de menor costo en los programa de mejora genética.

El principal problema de los estudios de asociación reside en las inconsistencias entre los resultados de estos estudios de asociación con todo el genoma para un mismo carácter, que puede atribuirse principalmente a aspectos tales como tamaño de la población, densidad de los marcadores (SNP), estructura genética de la población o la elección de los modelos estadísticos (Zhang *et al.*, 2012).

Para lograr la estimación precisa de los efectos de SNP en los caracteres de interés económico en un estudio de asociación, se requiere mayor tamaño poblacional y una mayor densidad de marcadores (Zhang *et al.*, 2012). Sin embargo, el costo de estas tecnologías sigue siendo prohibitivo para estudios con grandes cantidades de individuos.

Actualmente, los chips de SNP se aplican ampliamente para intentar mejorar la identificación de ATL para caracteres de interés en los animales domésticos. En comparación con esto, la secuenciación podría proporcionar casi toda la información acerca de las variaciones, incluyendo SNP, variación en el número de copias de DNA (CTV.) y deleciones / inserciones (indels), entre otras variantes, sobre la totalidad del genoma en cierta población.

Junto con la reducción en el precio de la secuenciación, es posible que individuos clave puedan ser secuenciados, donde estudios de asociación podrían llevarse a cabo empleando toda esa información. No obstante, los altos requerimientos informáticos y la formación de recursos humanos para el procesamiento de los datos son aspectos restrictivos para muchos grupos de investigación.

En un futuro, los estudios genómicos en animales domésticos deberían centrarse en la identificación de mutaciones causantes de rasgos económicamente importantes, por ejemplo las detectadas con RNA-seq. Estos hallazgos facilitarían la comprensión de la arquitectura genética de caracteres complejos en los animales domésticos y la mejora de las prácticas de los programas de mejoramiento.

5 BIBLIOGRAFÍA

AbdElmageed Z.Y., Naura A.S., Errami Y. and Zerfaoui M. 2012. The poly(ADP-ribose) polymerases (PARPs): new roles in intracellular transport. Cellular signalling (24): 1–8.

Alam M *et al.* 2011. A Whole Genome Association Study to Detect Single Nucleotide Polymorphisms for Body Conformation Traits in a Hanwoo Population. Asian-Aust. *Journal of Animal Science*. (24):322-329.

Apcher S., Daskalogianni C., Lejeune F., Manoury B., Imhoos G., y Heslop L. 2011. Major source of antigenic peptides for the MHC class I pathway is produced during the pioneer round of mRNA translation, PNAS 108(28). doi:10.1073/pnas.1104104108.

Beh K.J., Hulme D.J., Callahan M.J., Leish Z., Lenane I., Windon R.G. Y Maddox J.F. 2002. A genome scan for quantitative trait loci affecting resistance to Trichostrongyluscolubriformis in sheep.Animal Genetics (33): 97–106.

Benavides M.V., Weimer T.A., Borba M.F.S., Berne M.E.A. y Sacco A.M.S. 2002. Association between microsatellite markers of sheep chromosome 5 and faecal egg counts. Small Rumiant Research, (46): 97-105.

Bentley D.R., Balasubramanian S., Swerdlow H.P., *et al.* 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature, (456):53-59.

Beuzen N.D., Stear M.J. y Chang K.C. 2000. Review: Molecular markers and their use in animal breeding. The Veterinary Journal. (160): 42–52.

Bishop S.C., Bairden K., McKellar Q.A., Park M., Stear M.J. 1996. Genetic parameters for faecal egg count following mixed, natural, predominantly Ostertagtia circumcincta infection and relationships with live weight in young lambs. Animal Science. (63): 423–428.

Bishop S.C., Jackson F., Coop R.L., Stear M.J. 2004. Genetic parameters for resistance to nematode infections in Texel lambs and their utility in breeding programmes. Animal Science (78): 185–194.

Camacho Vallejo M.E., Delgado Bermejo J.V. y Barba Capote C.J. 2000. Recursos genéticos de animales domésticos en España: situación actual y sistemas tradicionales de explotación. *Archivos de zootecnia.* 49, (187) 423-430.

Castells D. 2002. Nuevo enfoque en el control parasitario de ovinos In: Resistencia genetica del ovino y su aplicacion en sistemas de control integrado de parasitos ; FAO Animal Production and HealthPaper (FAO) Castells Montes, D. / FAO, Rome (Italy). Direccion de Produccion y Sanidad Animal, p. 17-24.

Castells D. y Gimeno D. 2011. Selection of Corriedale sheep for resistance or susceptibility to nematode infection in Uruguay. XXIII Congreso Internacional de la Asociación Mundial para el avance de la Parasitología Veterinaria (WAVVP). Buenos Aires, Argentina 2011.

Castells D., Nari A., Rizzo E., Mármol E. y Acosta D. 1995. Efectos de los nematodos gastrointestinales sobre los parámetros productivos del ovino en la etapa de recría. ProducciónOvina, (8): 17-32.

Ciappesoni G., Gimeno D., Ravagnolo O. 2010. Genetic Relationships between Faecal Worm Egg Count and Production Traits in Merino Sheep of Uruguay. Proc. 9th World Congress Gen. Appl. Livest. Prod., Leipzig, Germany.

Davies G., Stear M.J., Benothman A., Kerr A., Mitchell S. y Bishop S.C. 2006. Quantitative trait loci associated with parasitic infection in Scottish blackface sheep. Heredity (96), 252–8.

Day-Williams A.G. y Zeggini E. 2011. The effect of next-generation sequencing technology on complex trait research. European journal of clinical investigation, 41(5), 561–7.

Delgado J.V., Martínez A.M., Acosta A., Álvarez L.A., Armstrong E., Camacho E., *et al.* 2011. Genetic characterization of Latin-American Creole cattle using microsatellite markers. *Animal Genetics*. 43(1):2-10.

DI.CO.SE. (Dirección General De Servicios Ganaderos) 2011. Datos de la Declaracion Jurada de Dicose 2011 - Datos generales. Consultado Mayo 2011. Discponible en: www.mgap.gub.uy/DGSG/DICOSE/Informe2011/DJ2011_TotalNacional.pdf.

Ding C. y Jin S. 2009. Single Nucleotide Polymorphisms. Human Press: Totowa, NJ. (578): 245-254.

Evanno G., Regnaut., S y Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*. (14):2611–2620.

Fernández G. 2000. Situation of the local domestic genetic resources in Uruguay. *Archivos de Zootecnia*. (49): 333-340.

Flori L., Fritz S., Jaffrézic F., Boussaha M., Gut I., Heath S., Foulley J.L. y Gautier M. 2009. The genome response to artificial selection: a case study in dairy cattle. *PLoS One.* 4(8):e6595.

Gao L., Zan L.S., Wang H.B., Hao R.J. y Zhong X. 2011. Polymorphism of somatostatin gene and its association with growth traits in Chinese cattle. Genetics and molecular research : GMR, 10(2), 703–11.

Gutiérrez J.P. y Goyache F. 2005. A note on ENDOG: a computer program for analysing pedigree information. *Journal of Animal Breeding Genetics*. (122)**:** 357-360.

Gutiérrez-Gil B., García-Gámez E., Kijas J.W., San Primitivo F., de la Fuente L.F., Bayón Y., International Sheep Genomics Consortium (ISGC) y Arranz J.J. 2012. Preliminary analysis of selective sweeps in the Spanish Churra sheep. National Conference of Animal Genetic Improvement. Menorca (Spain), 04/31/2012.

Hayakawa S., Shiratori S., Yamato H., *et al.* 2011. ZAPS is a potent stimulator of signaling mediated by the RNA helicase RIG-I during antiviral responses. Nature Immunology. (12): 37–44.

Hayes B. J., Pryce J., Chamberlain A. J., Bowman P. J. y Goddard M. E. 2010. Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. PLoS Genetics 6, e1001139.

Hayes B.J., Lien S., Nilsen H., Olsen H.G., Berg P., MacEachern S., Potter S., Meuwissen T.H.E. 2008. The origin of selection signatures on bovine chromosome 6. *Animal Genetics*. 39(2):105-111.

Hayes B, Goddard ME. 2001. The distribution of the effects of genes affecting quantitative traits in livestock. Genet Sel Evol. 33(3):209-29

Hulme D.J., Blattman A.N., Kinghorn B.P., Woolaston R.R., Gray G.D. y Beh K.J. 1993. A search for associations between major histocompatibility complex restriction fragment length polymorphism bands and resistance to Haemonchus contortus infection in sheep. Anim. Genet. (24): 277-282.

Johnston S.E., McEwan J.C., Pickering N.K., Kijas J.W., Beraldi D., Pilkington J.G., Pemberton J.M., Slate J. 2011. Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. Mol Ecol. 20(12):2555-66.

Karlsson L.J.E., y Greeff J.C. 2012. Genetic aspects of sheep parasitic diseases. Veterinary parasitology, 189(1), 104–12. doi:10.1016/j.vetpar.2012.03.039.

Kemper K., Emery D., Bishop S., Oddy H., Hayes B.J., Dominik S., Henshall J. y Goddard M. 2011. The distribution of SNP marker effects for faecal worm egg count in sheep, and the feasibility of using these markers to predict genetic merit for resistance to worm infections. *Genetics Research*. 93: 203-219.

Kijas J.W., Lenstra J.A., Hayes B., Boitard S., Porto Neto L.R., *et al.* 2009. Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. *PLoS Biology* 10(2).

Kijas J.W., Townley D., Dalrymple B.P., Heaton M.P., Maddox J.F., McGrath A., Wilson P., Ingersoll R.G., McCulloch R., McWilliam S., Tang D., McEwan J., Cockett N., Oddy V.H., Nicholas F.W., y Raadsma H. 2012. A Genome Wide Survey of SNP Variation Reveals the Genetic Structure of Sheep Breeds. *PLoS*. 4, e4668.

Lander E.S. y Schork N.J. 1994. Genetic dissection of complex traits. Science (265): 2037–2048.

Medrano J.F., Aasen E., Sharrow L. 1990. DNA extraction from nucleated red blood cells. Biotechniques.;8(1):43

Mendoza A. 1968. Raza Corriedale. En: Curso básico teórico práctico de lanares y lana. Comisión Honoraria de Mejoramiento de la Producción ovina. Ministerio de Ganadería y Agricultura, pp. 120-122.

Meuwissen, T.H.E. 2008. The origin of selection signatures on bovine chromosome 6. *Animal Genetics*. 39(2):105-111.

Miller J.M., Poissant, J., Kijas J.W. y Coltman D.W. 2011. A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. *Molecular EcologyResources*. (11): 314–322.

Montossi F., De Barbieri I., Ciappesoni G., Ravagnolo O., De Mattos D., Pérez Jones J., Fros A., Grattarola M., Mederos A. y Soares de Lima M. 2005. Núcleo fundacional de merino fino y superfino de la Unidad Experimental "Glencoe" - INIA Uruguay: una experiencia innovadora de mejoramiento genético asociativo y participativo. *Agrociencia*. IX-N°1. 609-616.

Moradi M.H., Nejati-Javaremi A., Moradi-Shahrbabak M., Dodds K.G. y McEwan J.C. 2012. Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. *BMC genetics*.13(1):10.

Morin R.D., Zhao Y., Prabhu A.L., Dhalla N., McDonald H., Pandoh P., Tam A., Zeng T., Hirst M., Marra M. 2010. Preparation and analysis of microRNA

libraries using the Illumina massively parallel sequencing technology. Methods Mol Biol (650):173–199.

Nari A., Cardozo H., Berdie J. 1977. Alza de lactación (Spring Rise) para nematodos gastrointestinales en ovinos. Primera comprobación en el Uruguay. Veterinaria 12 (65): 147-156.

OvineSNP50 Datasheet - Illumina. 2010. OvineSNP50 DNA Analysis Kit.Available: 12/05/2012 http://www.illumina.com/products/ovinesnp50_dna_analysis_kit.ilmn

Pareek C.S., Smoczynski R. y Tretyn A. 2011. Sequencing technologies and genome sequencing. *Journal of applied genetics* (52): 413-35.

Patterson N., Price A.L. y Reich D. (2006). Population structure and eigenanalysis. *PLoS Genetics*. 2(12): e190.

Perry B.D., McDermott J.J., Randolph T.F., Sones K.R., Thornton P.K. 2002. Investing in animal health research to alleviate poverty. ILRI (International Livestock Research Institute), Nairobi, Kenya, 148 pp.

Pritchard J.K., Stephens M., y Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics. (155):945–959.

Ramos M., Megens H.J., Crooijmans R.P.M., Schook L.B. y Groenen M. 2011. Identification of high utility SNPs for population assignment and traceability purposes in the pig using high-throughput sequencing. *Animal genetics*. 42(6):613-20.

Riggio V., Matika O., Pong-Wong R., Stear M.J. y Bishop S.C. 2013. Genome-wide association and regional heritability mapping to identify loci underlying variation in nematode resistance and body weight in Scottish Blackface lambs. Heredity, 110(5), 420–9. doi:10.1038/hdy.2012.90.

Salle´ G., Jacquiet P., Gruner L., Cortet J., Sauve´ C., Pre´vot F., *et al.* 2012. A genome scan for QTL affecting resistance to Haemonchus contortus in sheep. J Anim Sci 90:4690–4705.

Schaeffer L.R. 2006. Strategy for applying genome-wide selection in dairy cattle. Journal of Animal Breeding and Genetics. 123: 218–223.

Shimoda M., Hashimoto G., Mochizuki S., Ikeda E., Nagai N., Ishida S., y Okada Y. 2007. Binding of ADAM28 to P-selectin glycoprotein ligand-1 enhances P-selectin-mediated leukocyte adhesion to endothelial cells. The Journal of biological chemistry, 282(35), 25864–74. doi:10.1074/jbc.M702414200.

Silva M.V.B., Sonstegard T.S., Hanotte O., Mugambi J.M., Garcia J.F., Nagda S., Gibson J.P., *et al.* 2012. Identification of quantitative trait loci affecting resistance to gastrointestinal parasites in a double backcross population of Red Maasai and Dorper sheep. Animal genetics, 43(1), 63–71.

Simpson H.V. 2000. Pathophysiology of abomasal parasitism: is the host or parasite responsible? Vet. J. (160): 177–191.

Stear M.J., Boag B., Cattadori I., Murphy L. 2009. Genetic variation in resistance to mixed, predominantly Teladorsagia circumcincta nematode infections of sheep: from heritabilities to gene identification. Parasite Immunol (31): 274–282.

Sykes A.R. 1994. Parasitism and production in farm animals. Anim. Prod. (59): 155-172.

ten Bokum A.M., Hofland L.J., van Hagen P.M. 2000. Somatostatin and somatostatin receptors in the immune system: a review. Eur Cytokine Netw. 11(2):161-76.

Våge DI, Husdal M, Kent MP, Klemetsdal G, Boman IA. 2013. A missense mutation in growth differentiation factor 9 (GDF9) is strongly associated with litter size in sheep. BMC Genet. 2;14:1.

Wang W., Gerstein M., y Snyder M. 2010. “RNA-Seq : a revolutionary tool for t ranscriptomics.”. 10 (1): 57-63. doi:10.1038/nrg2484.RNA-Seq.

Warrener P., Kim S., Williams S.M.G., Biery M., Gordon M. 2012. Synthetic lethality of PARP inhibition in BRCA-network disrupted tumor cells is

associated with interferon pathway activation and enhanced by interferon-γ Apoptosis. 17 (7): 691-701.

Weir B.S. y Cockerham C.C. 1984. Estimating F-Statistics for the Analysis of Population Structure. *Evolution*. 38(6):1358-1370.

White S.N., Mousel M.R., Herrmann-Hoesing L.M., Reynolds J.O., Leymaster K.A., Neibergs H.L., Lewis G.S., Knowles D.P. 2012. Genome-wide association identifies multiple genomic regions associated with susceptibility to and control of ovine lentivirus. PLoS One.;7(10):e47829.

Wiggans G.R., VanRaden P.M., Bacheller L.R., Tooker M.E., Hutchison J.L., Cooper T.A. y Sonstegard T.S. 2010. Selection and management of DNA markers for use in genomic evaluation. *Journal of Dairy Science*. (93):2287–2292.

Wilson L.M., Whitt S.R., Iba A.M., Iv E.S.B. 2004. Dissection of Maize Kernel Composition and Starch Production by Candidate Gene Association. (16): 2719-2733.

Zeder M.A. 2008. Domestication and early agriculture in the Mediterranean Basin: origins, diffusion, and impact. PNAS. (105):11597–11604.

Zenger K.R., Khatkar M.S., Cavanagh J.A., Hawken R.J y Raadsm H.W. 2007. Genome-wide genetic diversity of Holstein Friesian cattle reveals new insights into Australian and global population variability, including impact of selection. *Animal Genetics* (38): 7–14.

Zhang H., Wang Z., Wang S., Li H. 2012. Progress of genome wide association study in domestic animals. Journal of animal science and biotechnology. 3, 26.

Zhong B., Liu X., Wang X., Chang S.H., Liu X., Wang A., Joseph M., *et al.* 2013. Negative regulation of IL-17-mediated signaling and inflammation by ubiquitin-specific protease 25. NIH Public Access, 13(11): 1110–1117.