Universidad de León

Facultad de Veterinaria

Departamento de Producción Animal

MAPEO FINO DE REGIONES GENÓMICAS PORTADORAS DE QTL CON INFLUENCIA SOBRE CARACTERES DE PRODUCCIÓN DE LECHE EN EL GANADO OVINO

FINE MAPPING OF GENOMIC REGIONS UNDERLYING QTL FOR MILK PRODUCTION TRAITS IN SHEEP

Elsa García Gámez

León, Septiembre de 2012

Las investigaciones de esta Memoria de Tesis Doctoral han sido financiadas por los proyectos AGL2009-07000 del Ministerio de Ciencia e Innovación (MICINN) y SheepMilkGenes (PERG02-GA-2007-224857) y 3SR (FP7-KBBE-2009-3) de la Comisión Europea.

La autora de esta Memoria ha sido beneficiaria de una beca de posgrado correspondiente al Programa de Formación de Profesorado Universitario (FPU) del Ministerio de Educación con referencia AP2008-01410.

AGRADECIMIENTOS

La elaboración de esta Tesis Doctoral ha sido posible gracias a la colaboración, directa o indirecta, de muchas personas a lo largo de estos años.

En primer lugar, quiero agradecer a mis directores de tesis, el Dr. Juan José Arranz Santos y la Dra. Beatriz Gutiérrez Gil por su confianza y apoyo durante este tiempo. Gracias por darme la oportunidad de trabajar en lo que me gusta, por vuestro apoyo en los momentos difíciles y por todas las oportunidades que me habéis brindado durante estos años.

A la Universidad de León y al Ministerio de Educación por proporcionarme los medios técnicos y económicos necesarios para la realización de este trabajo.

A todos los profesores del Departamento de Producción Animal que han contribuido durante mis estudios a mi formación científica y personal. A los primeros contactos que tuve con el Departamento de Producción Animal, el Dr. Fermín San Primitivo y el Dr. Luis Fernando de la Fuente, gracias porque vuestras clases durante la licenciatura me hicieron interesarme por este campo de la investigación. Un agradecimiento muy especial al Dr. Juan Pablo Sánchez, por su entusiasmo y su paciencia, por toda la ayuda que me ha prestado tanto desde León, como desde Lérida.

To Dr. James Kijas thank you for giving me the opportunity to collaborate with your research group in Brisbane and for sharing your knowledge with me. To Dr. Goutam Sahana thank you for your help during my internship at Aarhus University. Muchas gracias también al Dr. Antonio Reverter por contagiarme su entusiasmo, sus ganas de innovar. Gracias porque sin tu colaboaración mi estancia en CSIRO no hubiese sido tan productiva.

A mis compañer@s de Departamento por las horas de laboratorio, de café, las charlas y las confidencias. Porque estoy muy feliz por haber encontrado amigas de verdad en mi lugar de trabajo.

A todo el Personal del Departamento de Producción Animal, PDI y PAS, por su disponibilidad y por hacer más sencillo este camino.

A todos mis amig@s, l@s de León y l@s de Cantabria, los que están y los que ya no, por todos los momentos que hemos compartido. En especial a Tamara, por ser como eres y

estar siempre ahí. To all my friends around the world, in Brisbane, Denmark or wherever you are now, thank you for making me feel like home. A Marga, porque me aportas frescura, locura y a la vez sensatez, mil gracias por todo.

A mi familia, a todos, los que están cerca y los que están lejos. A mis padres, Enrique y Carmen, y a mis hermanos, Ana y Álvaro, por estar ahí siempre y por ayudar a que mis sueños se hagan realidad porque cada uno aportáis algo que hace la vida más fácil. Porque cuando he estado lejos, habéis sabido cómo hacer que me sintiera querida. También por todos los momentos que hemos compartido, los buenos y los no tan buenos, me alegra que todos podáis ver el final de este camino.

Y, por supuesto, a Javi, porque has traido una ilusión enorme a mi vida y a mi futuro. Por compartir estos momentos conmigo, por apoyarme y por estar ahí. Por no dejarme sola en los momentos más duros y por creer en mí con los ojos cerrados, te quiero.

MUCHÍSIMAS GRACIAS A TODOS.

ÍNDICE DE CONTENIDOS

ABREVIATURAS

ABCG2	Cassette de unión a ATP, familia G, miembro 2
ADN	Ácido desoxirribonucleico
ANCHE	Asociación Nacional de Criadores de Ganado Ovino Selecto de Raza Churra
ARN	Ácido ribonucleico
ARNm	Ácido ribonucleico mensajero
BTA	Cromosoma bovino
CI	Intervalo de confianza
cM	centimorgan
cm	centímetros
DGAT1	Diacil glicerol acil tranferasa 1
DYD	*Daughter yield deviations*
EBV	Valor genético estimado
EET	Encefalopatía Espongiforme Transmisible
FP	Porcentaje de grasa en leche
FN	*Florida Native*
F_{ST}	Medida de la diversidad genética
FxD	Población de retrocruzamiento East Friesian x Dorset
FY	Cantidad de grasa en leche
GAS	Selección Asistida por Genes
GCN	*Gulf Coast Native*
GEBVs	Valor genético genómico estimado
GHR	Receptor de la hormona de crecimiento
GS	Selección Genómica
GWAS	Estudio de asociación a nivel genómico
h^2	Heredabilidad
IA	Inseminación artificial
ISGC	Consorcio Internacional para la genómica ovina
Kb	Kilo bases, 1.000 pares de bases
Kg	Kilogramos
LA	Análisis de ligamiento
LALBA	Alfa-lactoalbúmina
LD	Desequilibrio de ligamiento

LDLA	Combinación de LD y LA
LN	*Louisiana Native*
MAF	Frecuencia del alelo menos frecuente
MAS	Selección Asistida por Marcadores
Mb	Megabases, 1.000.000 de pares de bases
Me	Número efectivo de marcadores por segmento cromosómico
MHC	Complejo Mayor de Histocompatibilidad
MY	Cantidad de leche
Ne	Tamaño efectivo de la población
OAR	Cromosoma ovino
pb	Pares de bases
PCR	Reacción en cadena de la polimerasa
PP	Porcentaje de proteína
PRL	Prolactina
PRNP	Proteína Priónica
PY	Cantidad de proteína en leche
QTL	*Locus/Loci* con influencia sobre un carácter cuantitativo
QTN	Mutación causal de un QTL
r^2	Medida del LD
SCS	Recuento de células somáticas
SNP	Polimorfismo de un solo nucleótido
SPP1	Osteopontina
SxL	Población de retrocruzamiento Sarda x Lacaune
YD	*Yield deviations*

INTRODUCCIÓN Y PLANTEAMIENTO

Durante los últimos años del siglo XX y comienzos del siglo XXI, la biotecnología genética ha experimentado un gran avance con el nacimiento de la genómica y la posibilidad de abordar el estudio de la estructura y funcionalidad de los genomas a nivel global. Así, desde mediados de los años 90, se han realizado numerosos proyectos para la detección de regiones con influencia sobre caracteres cuantitativos (QTL) en los animales domésticos. Las especies pioneras en estos estudios han sido el ganado porcino y el ganado vacuno, que presentan más de 300 publicaciones científicas y más de 5.000 QTL reportados en cada especie (http://www.animalgenome.org/cgi-bin/QTLdb/index). En el caso del ganado ovino, el número de proyectos realizados es más modesto, aunque cabe destacar varios proyectos de barridos genómicos que se han llevado a cabo en diferentes razas (Barillet et al., 2006; Gutiérrez-Gil et al., 2009; Raadsma et al., 2009a; Mateescu y Thonney, 2010a). La mayoría de estos estudios se han realizado utilizando técnicas de análisis de ligamiento en poblaciones segregantes y microsatélites como marcadores genéticos.

A partir de mediados de la primera década del siglo XXI surgen una serie de técnicas de secuenciación, denominadas de segunda generación (*Next Generation Sequencing*) o secuenciación masiva paralela, que han permitido el abordaje de la secuenciación de un genoma *de novo* en una escala temporal relativamente corta (2-5 años) y abaratar el coste de los proyectos de secuenciación en varios órdenes de magnitud (hasta más de 10.000 veces). Esto ha provocado que, en la mayoría de las especies domésticas, se hayan desarrollado proyectos, o estén en un estado muy avanzado, para el conocimiento de su genoma. También, se ha producido un desarrollo espectacular de las herramientas moleculares, derivadas de la gran información producida durante la secuenciación genómica, las más importantes los chips de SNPs, que nos permiten explorar el genoma en busca de regiones asociadas a caracteres de interés productivo. Una de las aplicaciones directas de estas herramientas en los animales domésticos es su utilización para el conocimiento de la arquitectura molecular y para incrementar la respuesta a la selección de los fenotipos de interés en Producción Animal.

De esta forma se ha pasado de los barridos genómicos utilizando marcadores de tipo microsatélite en familias segregantes a los barridos genómicos de alta densidad que utilizan chips de miles (50-1.000) de SNPs a un coste asequible. Esta herramienta genómica posibilita, además, el análisis de animales no relacionados genéticamente mediante los conocidos como análisis de asociación del genoma completo (GWAS, *Genome-wide Assocation Studies*).

Estos análisis se basan en el desequilibrio de ligamiento presente en las poblaciones de animales domésticos, evitando así la necesidad de establecer costosas poblaciones experimentales basadas en cruzamientos entre razas divergentes. Los resultados de los GWAS se caracterizan por la detección más precisa de regiones cromosómicas asociadas a los caracteres productivos. Aunque los resultados obtenidos han sido desiguales entre las distintas especies, se han producido, en general, grandes avances en la disección de la base genética de los caracteres productivos tanto simples como complejos.

Por otra parte, desde la perspectiva de la Producción Animal, la información derivada de estos estudios puede utilizarse en programas de selección asistida por marcadores (MAS, *Marker Assisted Selection*) para estimar los valores genéticos de los animales, sin necesidad de conocer la mutación causante del efecto. En el ganado ovino, la MAS solo se ha llevado a cabo mediante la selección de portadores de las mutaciones causantes de los efectos deseados en relación a fenotipos extremos de conformación de la canal y prolificidad, principalmente. En cambio, en el ganado vacuno lechero estos métodos de selección se están utilizando desde los años 2000 y 2003 en Francia y Alemania, cuando se crearon los consorcios para la aplicación de MAS (Boichard et al., 2006).

Asimismo, la selección de animales mejorantes ha comenzado a basarse en nuevos métodos como la Selección Genómica (GS, *Genomic Selection*). La GS se basa en la predicción de los valores genómicos de los animales utilizando toda la información disponible a lo largo del mismo en base a los miles de genotipos obtenidos con los chips de SNPs (Meuwissen et al., 2001). En el ganado vacuno lechero la GS se ha estado aplicando y mejorando durante los últimos años en países como Francia, Alemania, Estados Unidos, Nueva Zelanda, Australia y los Países Bajos (Hayes et al., 2009). El desarrollo de la GS ha conseguido aportar los mayores avances en mejora genética en el ganado vacuno de leche en los últimos 20 años, aunque todavía requiere de esfuerzos y mejoras para explotar todo su potencial en otras especies.

Dentro de las especies domésticas, el ganado ovino es una de las que presenta un proyecto avanzado de secuenciación del genoma completo. Este proyecto viene auspiciado por un esfuerzo internacional de diferentes grupos de investigación reunidos en el Consorcio Internacional para la Genómica Ovina (ISGC, *International Sheep Genomics Consortium*) y que está permitiendo la obtención de un genoma de alta calidad para la oveja. Este consorcio,

liderado por investigadores de Australia y Nueva Zelanda, se propuso como objetivo inicial el desarrollo de un chip de SNPs de media densidad en el ganado ovino (*OvineSNP50 BeadChip*), para poder utilizarlo en el análisis de la arquitectura genética de los caracteres heredables en esta especie. El desarrollo del chip se realizó en una etapa temprana del proyecto de secuenciación del genoma, cuando solo se disponía de un borrador obtenido por alineamiento de la secuencia genómica ovina con los genomas bovino, canino y humano (*Virtual Sheep Genome*; Dalrymple et al., 2007). Nuestro grupo de investigación pertenece al ISGC y colabora modestamente en algunas actividades del mismo, tales como el proyecto *SheepHapMap* planteado por el ISGC para el análisis de la diversidad, y estructura genética de la especie ovina a nivel mundial y para caracterizar el efecto que la adaptación al entorno doméstico y la selección artificial han tenido en esta especie (Kijas et al., 2012).

La actividad del grupo de investigación en el que se ha desarrollado la presente Tesis Doctoral ha sido históricamente la mejora genética del ganado ovino de leche, principalmente en la raza Churra. En una de las líneas de investigación desarrolladas en la última década, nuestro grupo ha sido uno de los más activos a nivel europeo en la búsqueda de QTL con influencia sobre caracteres relacionados con la producción de leche. De esta manera se han estudiado diferentes caracteres de producción de leche (Gutiérrez-Gil et al., 2009), morfología mamaria (Gutiérrez-Gil et al., 2008b), morfología corporal (Gutiérrez-Gil et al., 2011) y recuento de células somáticas, como indicador de resistencia a mastitis (Gutiérrez-Gil et al., 2007). Los resultados obtenidos nos permitieron detectar diversas regiones con clara evidencia de ser portadoras de genes que controlan la producción de leche en oveja.

El planteamiento de la presente Tesis Doctoral se realiza en un momento de transición entre la utilización de marcadores microsatélite para la localización de genes y los inicios de la utilización las herramientas genómicas en la oveja. Esta Tesis Doctoral comienza con un planteamiento de mapeo fino clásico utilizando marcadores microsatélite en una nueva población de raza Churra dentro de la actividad de uno de los proyectos desarrollados en ese momento por nuestro grupo de investigación (*SheepMilkGenes*). La disponibilidad del chip de SNPs desarrollado en los últimos años por el ISGC nos permitió adaptar el planteamiento inicial y, así, tanto la confirmación como el mapeo fino de algunos de los QTL detectados en análisis anteriores, se han complementado utilizando esta potente herramienta genómica. Estos últimos trabajos se han llevado a cabo dentro del marco de los proyectos desarrollados

por nuestro grupo de investigación en los últimos años: AGL2009-0900, financiado por el plan nacional y el proyecto Europeo 3SR (*Sustainable Solutions for Small Ruminants*).

Por lo tanto, el objetivo final de la presente Tesis Doctoral es el mapeo fino de algunas regiones detectadas previamente como QTL en la raza Churra. Como el número de regiones identificadas ha sido elevado, hemos seleccionado aquellas que influencian de una forma más evidente caracteres de riqueza de la leche, ya que no hay que olvidar que el destino final de la leche ovina es la producción de queso.

Los objetivos concretos que nos hemos planteado son los siguientes:

1- Confirmación de los QTL localizados en OAR20 y OAR3, que afectan al porcentaje de grasa y porcentaje de proteína, respectivamente, y que han mostrado anteriormente evidencia de segregación en la raza ovina Churra.

2- Evaluación de la utilidad del *OvineSNP50 BeadChip* como herramienta genómica en la raza Churra.

3- Utilización de dicha herramienta genómica para el análisis de la estructura del genoma de la raza Churra mediante el análisis de la extensión del desequilibrio de ligamiento en esta raza.

4- Mapeo fino de los QTL confirmados y análisis GWAS para la detección de nuevas regiones genómicas no detectadas anteriormente.

REVISIÓN BIBLIOGRÁFICA

1. IMPORTANCIA DE LA PRODUCCIÓN OVINA LECHERA

La explotación de razas ovinas lecheras se concentra principalmente en países de la Cuenca Mediterránea (Sur de Europa, Europa Central y Próximo Oriente) estando tradicionalmente ligada a la explotación de razas locales, bien adaptadas al medio y poco productivas. Actualmente, los sistemas de explotación varían desde extensivos a intensivos dependiendo de la raza y la importancia económica de la producción (Carta et al., 2009).

En los últimos años la demanda de productos de mayor calidad y seguridad alimentaria por parte de los consumidores exige que el sector ovino lechero deba adaptarse a este nuevo escenario. Esto ha incentivado el establecimiento de programas de mejora en muchas de las razas ovinas lecheras, cuyos objetivos han sido inicialmente los caracteres de producción (cantidad y composición de la leche), teniéndose en cuenta posteriormente y en algunas de las razas, otros caracteres de tipo funcional (caracteres morfológicos y de resistencia a enfermedades) (Barillet, 2007).

Además, existen aspectos culturales difícilmente cuantificables que influencian la producción de leche ovina (Ugarte et al., 2002). Esto ocurre con la explotación de razas locales altamente adaptadas al medio como es el caso de la raza Lacaune en la región de Rochefort (Francia), la raza Sarda en la isla de Cerdeña (Italia) o las razas Castellana, Churra, Manchega y Latxa en distintas regiones españolas. Los paisajes en estas regiones se han caracterizado por la presencia de rebaños ovinos y la explotación de estas razas ha constituido el sustento de la economía de dichas zonas.

La explotación de ganado ovino para producción de leche en otros países como Estados Unidos se encuentra en una fase inicial, aunque, dadas las características nutricionales de este producto, estos países parecen ser un nicho de mercados emergentes (BANR, 2008).

1.1. Producción de leche ovina en España

La industria lechera ovina española ocupa el décimo lugar en producción de leche entera de oveja a nivel mundial (FAOSTAT). En el año 2010, se produjeron en España 7.245 millones de litros de leche de los cuales 6.172 millones de litros (85%) fueron de leche bovina, 566 millones de litros (7,8%) de oveja y los restantes 507 millones de litros (7,2%) de

leche caprina. En la Figura 1 se representa la producción de leche de ganado ovino por Comunidad Autónoma en España en 2010. Dicha producción se concentra en dos regiones, Castilla y León, donde se produjeron 386 millones de litros (68%) y Castilla-La Mancha, con 131 millones de litros (23%) (MAGRAMA, 2010).

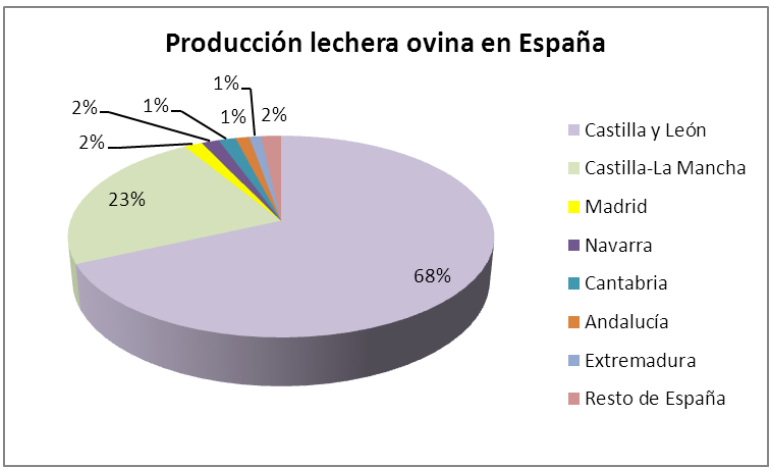


Figura 1 - Producción de leche ovina en España distribuida por provincias en el año 2010. (Fuente: MAGRAMA, 2010)

La explotación de ganado ovino lechero en España ha estado ligada, tradicionalmente, al aprovechamiento de los recursos naturales. En general, la producción de leche de ganado ovino se ha basado en la explotación de razas autóctonas muy bien adaptadas a sus respectivas zonas de origen. Este es el caso de la raza Manchega en Castilla-La Mancha, las razas Churra y Castellana en Castilla y León o las razas Latxa y Carranzana en el País Vasco y Navarra (Ugarte et al., 2002). En Castilla y León, las razas autóctonas, Churra y Castellana, están siendo desplazadas por razas foráneas más especializadas. El censo de animales de razas extranjeras más productivas, como Assaf, Awassi y Lacaune, ha aumentado en los últimos años, sustituyendo, en parte, la producción lechera de razas autóctonas. Actualmente, más del 45% de la producción de leche ovina procede de animales de razas foráneas o de sus cruces con razas autóctonas (Ugarte et al., 2001).

La importancia de la explotación de las razas locales radica no solo en la conservación del medio rural en las zonas donde han existido estas granjas tradicionalmente, sino en la elaboración de productos de alta calidad. A partir de la leche de estos animales se elaboran quesos comercializados con la etiqueta de calidad Denominación de Origen como son el "Queso Zamorano", con leche de las razas Churra o Castellana, el "Queso Manchego", de la raza Manchega, o el "Idiazábal", de las razas Latxa y Carranzana. Además, la venta de corderos bajo la etiqueta Indicación Geográfica Protegida "Lechazo de Castilla y León", de

las razas Churra, Castellana y Ojalada, o "Cordero Manchego", de la raza Manchega, proporciona razones para hacer un esfuerzo en la mejora de la producción lechera de las razas ovinas españolas.

1.2. La raza Churra

La raza ovina Churra, perteneciente al tronco Churro, es una oveja rústica de las más primitivas de la península ibérica. Forma parte del grupo de razas autóctonas más importantes de España, por su alta especialización en producción de leche y su elevado censo. La zona principal de explotación coincide con la Submeseta Norte del centro peninsular, y más concretamente con el valle del Duero, coincidiendo en gran medida con la Comunidad Autónoma de Castilla y León. Actualmente, convive en esta región con otras razas autóctonas, como Castellana, Merina y Ojalada, y con rebaños formados por animales con variable grado de mestizaje de la raza ovina Assaf (San Primitivo y De la Fuente, 2000).

En España, el número de ganaderías que se dedican a la explotación de ganado ovino de raza Churra en pureza es de 363, con un número medio de animales por explotación de 600, y estando 351 de ellas vinculadas a la Asociación Nacional de Criadores de Ganado Ovino Selecto de Raza Churra (ANCHE) (ARCA). La tendencia poblacional del censo de ganado Churro es estable, manteniéndose a lo largo de los últimos años. En el último censo, con fecha 31 de Diciembre de 2011, el número de cabezas de ganado de esta raza se situaba en torno a 155.000, siendo el 89% hembras reproductoras (ARCA).

Morfológicamente, se trata de una oveja de tamaño medio, longilínea, con aspecto de ganado rústico. La capa es de color blanco, con pigmentación centrífuga en negro que afecta a la porción terminal de las orejas, alrededor de los ojos, labios y hocico, región umbilical y zona distal de las extremidades. El vellón es de lana larga y basta. Presenta un cierto dimorfismo sexual en cuanto al tamaño de los animales: los machos alcanzan una altura a la cruz de 81 cm y un peso aproximado de 73 kg, mientras que las hembras miden aproximadamente 68 cm a la cruz y pesan alrededor de 66 kg (ARCA) (Figura 2).

Figura 2 - La raza Churra

La oveja Churra presenta una gran precocidad sexual, cubriéndose por primera vez entre los 10 y 14 meses de vida, si bien algunas cubriciones excepcionales pueden producirse a los cinco meses. El periodo de gestación es de, aproximadamente, 150-160 días. En algunos sistemas de explotación se producen tres partos cada dos años, concentrándose las parideras en tres periodos: noviembre-diciembre (31% del total), febrero-marzo (27%) y julio (10%), estando el restante 32% de los partos dispersos a lo largo del año (San Primitivo y De la Fuente, 2000). La prolificidad de estos animales es de 1,38 corderos por parto, aunque pueden existir diferencias entre rebaños. Las crías son amamantadas por la madre hasta el destete, entre los 25 y 30 días, para el sacrificio del cordero, con entre 9 y 11 kg de peso vivo.

La explotación de esta raza se basa en tres tipos de sistemas productivos: extensivo, semi-extensivo e intensivo. El primero es el más tradicional, los animales aprovechan los pastos comunales y suelen tener una paridera por año. El sistema semi-extensivo es el más común entre los criadores de raza Churra, en él los animales utilizan los pastos disponibles, pero también pasan una parte de su vida estabulados. Por último, existe un pequeño número de explotaciones basadas en un sistema intensivo de producción, con animales estabulados permanentemente (ARCA).

La raza ovina Churra presenta una doble aptitud productiva (leche-carne), aunque presenta una importante especialización lechera. La producción media por animal y lactación se puede cifrar en 119 Litros de leche a lo largo de los, aproximadamente, 120 días de duración de la misma. Respecto a la composición bromatológica de la leche, la raza Churra produce leche con valores medios de 6,7% de grasa y 5,7% de proteína (ARCA). Las cantidades de estos dos componentes de la leche se consideran importantes en la especie ovina ya que la leche se utiliza principalmente para la elaboración de queso y el precio de la misma está condicionado por parámetros de composición y calidad (Pirisi et al., 2007).

2. PROGRAMAS DE MEJORA GENÉTICA EN EL GANADO OVINO LECHERO

Los esquemas de mejora genética en el ganado ovino de leche a nivel mundial utilizan dos tipos de estrategias. Por un lado está el cruce de las razas locales, menos productivas, con animales de razas foráneas más productivos. Por otro tenemos la explotación en pureza de animales de razas autóctonas, utilizando la selección de reproductores mejorantes como herramienta para la mejora genética de la población. A continuación nos centraremos en esta segunda estrategia ya que es la elegida por ANCHE, asociación con la que el grupo de mejora genética en el que se ha desarrollado esta Tesis Doctoral colabora estrechamente.

El sistema de selección de reproductores más eficiente hasta el momento es el basado en un manejo piramidal de la población (Barillet, 1997). Los criadores del núcleo de selección se sitúan en el vértice de la pirámide. En este grupo de animales es en el que hay un control lechero oficial y de pedigrí, inseminación artificial o monta dirigida, y se estiman los valores genéticos de los animales. La mejora genética se transfiere después a través de los machos seleccionados al resto de la población. El tamaño óptimo del núcleo de selección, para maximizar las ganancias productivas, es entre el 10-20% del tamaño total de la población (Barillet, 1997).

Este proceso de valoración es costoso y se alarga en el tiempo ya que es necesario mantener a los candidatos a ser machos reproductores durante el tiempo que dure el periodo de valoración, desde el nacimiento hasta al menos la primera lactación de sus hijas. En el ganado ovino este periodo es variable desde un mínimo de 4 años en la raza Churra (San Primitivo, 1998) hasta los 5-7 años en la Lacaune (Duchemin et al., 2012).

2.1. Caracteres objeto de selección en el ganado ovino lechero

La cantidad de leche producida representa dos tercios del total de los beneficios obtenidos por el sector ovino lechero, así que dicho carácter es el más importante a la hora de diseñar un programa de selección (Carta et al., 2009). Dado que prácticamente toda la leche producida se utiliza para la elaboración de queso, los caracteres de cantidad de proteína y grasa en la leche son también de gran importancia. La correlación negativa entre ambos tipos de caracteres, cantidad de leche producida y contenidos, ha hecho que se introduzcan los porcentajes de proteína y grasa como objetivos en los programa de selección de algunas razas

(Carta et al., 2009). En las razas Lacaune y Manech se ha utilizado como criterio de selección una combinación lineal de los caracteres productivos (cantidades de proteína y grasa combinadas con los porcentajes de los mismos) (Barillet et al., 2001; Barillet et al., 2008). En el caso de la raza Churra, se ha incluido el porcentaje de proteína como criterio de selección (Gutiérrez-Gil et al., 2009).

La necesidad de adaptarse a la mayor competitividad y exigencias de los consumidores haría necesario incluir otros caracteres funcionales que ayudaran a incrementar el rendimiento de las explotaciones, la sanidad del producto, etc. En relación a los aspectos sanitarios de la leche, el carácter recuento de células somáticas constituye un buen indicador de la resistencia frente a mastitis (Shook y Schutz, 1994). Se estima que en ganado ovino de raza Churra la prevalencia de esta enfermedad es del 0,4% en casos clínicos, mientras que asciende hasta un 16-35% en los casos de mastitis subclínica e infecciones crónicas (Gutiérrez-Gil et al., 2007).

En España, se realiza un Control Lechero Oficial cuyo objetivo final es la evaluación genética de los reproductores de las especies de aptitud lechera para mejorar las producciones lácteas. Estos controles consisten en la comprobación sistemática, siguiendo las recomendaciones internacionales establecidas por el ICAR (*International Committee for Animal Recording*), de la cantidad de leche producida y de sus componentes, así como la recogida de otra información de validez para su incorporación en los esquemas de selección aprobados para las diferentes razas (Real Decreto 368/2005, de 8 de abril).

Además, en el ganado ovino, dentro de los programas de selección y en base a una decisión de la Comisión Europea (Commission Decision 2003/100/CE), se ha utilizado de manera sistemática la selección asistida por marcadores a favor de los alelos que confieren resistencia frente a una encefalopatía espongiforme transmisible (EET) que aparece en esta especie, el *scrapie*, conocida también con el nombre de tembladera. Esta selección se lleva a cabo mediante el genotipado de polimorfismos localizados en tres codones del gen *PRNP*, que codifica para la proteína priónica. Las diferentes combinaciones de esos polimorfismos determinan cinco haplotipos a nivel de los residuos aminoacídicos (ARR, ARQ, ARH, AHQ y VRQ), de los cuales el ARR es el que confiere una mayor resistencia frente a la enfermedad, mientras que el VRQ se considera asociado a la mayor sensibilidad frente a la misma.

Una de las consecuencias negativas de este tipo de método de selección es la reducción de variabilidad asociada a la selección en favor de un determinado genotipo.

Asimismo, un posible ligamiento o interacción del gen *PRNP* con otros *loci* asociados con caracteres de interés económico podría tener efectos negativos sobre algunos de los caracteres objeto de selección en relación con la producción de leche. Esta posibilidad ha sido estudiada por nuestro grupo de investigación, no habiéndose identificado asociación entre los genotipos del gen *PRNP* y caracteres de producción de leche (cantidad de leche y contenidos de grasa y proteína) en la raza Churra. Tampoco se detectó la presencia de ningún QTL con influencia sobre dichos caracteres a lo largo del cromosoma 13 (Álvarez et al., 2006). Estudios similares en otras razas ovinas lecheras han descartado, hasta el momento, asociaciones significativas entre los caracteres productivos o funcionales y los genotipos del gen *PRNP* en las razas estudiadas (Carta et al., 2009).

2.2. Programa de mejora genética de la raza ovina Churra

ANCHE diseñó el primer programa de selección para mejorar la producción láctea en el año 1984, basado en la prueba de machos por descendencia (De la Fuente et al., 1995). Inicialmente, el único carácter objeto de selección fue la cantidad de leche estandarizada a 120 días de lactación. La primera valoración genética de los machos se realizó en el año 1990 y el primer catálogo de sementales se publicó en 1991. En el año 1998 se incluye un nuevo carácter, el porcentaje de proteína, como indicador del rendimiento de la leche. A partir del año 2003, también se incluye como objetivo de selección la resistencia al scrapie a través del genotipado del gen *PRNP*, seleccionando a favor del haplotipo ARR (ARCA).

El actual programa de mejora genética de la raza para caracteres de producción de leche, aprobado en diciembre de 2010 por la Dirección General de Recursos Agrícolas y Ganaderos, contempla dos categorías diferenciadas en cuanto a los objetivos de selección (ANCHE). Se fijan como objetivos primarios la mejora de la cantidad y el rendimiento quesero de la leche. En ambos casos, la selección de los caracteres a través de los cuales se selecciona para estos objetivos, se ha hecho en función de la heredabilidad (h^2) de los mismos. La cantidad de leche se selecciona a través del carácter leche por lactación estandarizada, medida entre los 30 y 120 días de lactación (h^2 = 0,26 ± 0,02). Para el rendimiento o extracto quesero se considera el porcentaje de proteína en leche (h^2 = 0,28 ± 0,03), ya que este carácter presenta una heredabilidad sustancialmente superior al porcentaje de grasa (h^2 = 0,08 ± 0,03) (ANCHE).

Además, como hemos mencionado previamente, ANCHE lleva ejecutando desde el año 2003 el "Programa de selección de animales resistentes a las EETs". Dentro de este programa se ha genotipado toda la población inscrita en el libro genealógico, con el fin de realizar una selección que permita que en la población se vayan fijando los genotipos resistentes. El proceso de selección ha llevado a un aumento en las frecuencias de los haplotipos asociados con la resistencia, en especial el ARR, unido a un descenso en las frecuencias de los haplotipos relacionados con una mayor sensibilidad a la enfermedad (JJ Arranz, comunicación personal).

Como objetivos secundarios, desde el año 2003, se han incluido los caracteres de morfología mamaria y corporal, representando en total estos caracteres el 10% del índice de mérito total de los machos churros evaluados (ANCHE). El criterio de selección para mejorar la morfología mamaria se evalúa en función de cinco caracteres, profundidad e inserción de la ubre, verticalidad y longitud de los pezones y conformación general de la ubre. Para la morfología corporal los cinco caracteres que conforman el criterio de selección son estatura, aplomos de las patas posteriores, anchura de la grupa, inclinación de talones y apariencia general del animal. Las heredabilidades de estos caracteres varían desde 0,09 ± 0,02 para la inserción de la ubre hasta 0,26 ± 0,03 para la profundidad de la misma, lo que permite una selección eficiente de los animales de raza Churra para estos caracteres (De la Fuente et al., 2011).

Los animales que se encuentran dentro del programa de mejora genética, basado en la selección de sementales por descendencia, deben someterse a la toma de muestras y medida de los rendimientos productivos a través del Control Lechero Oficial. Éste lo lleva a cabo ANCHE a través de la colaboración con la Junta de Castilla y León y el Centro de Selección y Reproducción Animal de León. Posteriormente, la valoración de sementales se realiza en colaboración con el Grupo de Mejora Genética del Departamento de Producción Animal (Universidad de León).

Asimismo, existe otro Subprograma de Selección en la raza Churra para el aumento de la producción de lechazos aprobado en diciembre de 2010 por la Dirección General de Recursos Agrícolas y Ganaderos. El objetivo final de este programa es el de incrementar la rentabilidad de las explotaciones de raza Churra de no ordeño, donde el único producto es el lechazo. Actualmente, los caracteres objeto de selección se dividen en dos grupos, (i) la

prolificidad, y (ii) la aptitud maternal, evaluada como una combinación del crecimiento de la cría (edad a la que los corderos alcanzan un peso de 10 kg) y la valoración subjetiva del ganadero (ANCHE).

2.3. Selección Genómica

Una alternativa a los métodos tradicionales de mejora genética, propuesta por Meuwissen et al. (2001), es la Selección Genómica (GS, *Genomic Selection*) en la que la información proporcionada por marcadores distribuidos a lo largo del genoma, se utiliza para detectar los efectos de todas las regiones con influencia sobre caracteres productivos (QTL, *Quantitative Trait Loci*) existentes. Esta metodología se basa en el desequilibrio de ligamiento (LD) existente entre los marcadores y los QTL.

La implementación de la GS consiste básicamente en dos pasos, primero se estiman los efectos genéticos en cada uno de los segmentos cromosómicos en que se divide el genoma en una población de referencia y, posteriormente, se predice el valor genético genómico (GEBV, *genomic estimated breeding value*) de los animales candidatos a la selección. Para la estimación de los efectos necesitamos tanto el genotipo como el fenotipo de los animales, en el segundo paso predecimos los GEBVs a través del genotipo del animal utilizando la información obtenida en la población de referencia.

La mayor ventaja de la GS es la tremenda reducción del tiempo para obtener la valoración de los reproductores. Asimismo, la aplicación de la GS y la utilización de la información procedente de marcadores genéticos supone una mejora importante en la selección para caracteres difíciles de medir, caracteres con baja heredabilidad, aquellos que solo se miden en uno de los sexos o los que se obtienen post-mortem (Duchemin et al., 2012).

Para la primera fase de detección de los QTL se pueden utilizar diferentes metodologías, ya sean basadas en ligamiento, en LD o en una combinación de ambos (LDLA). Además hay una serie de factores que influyen en la exactitud de la GS: (i) el número de marcadores utilizados; (ii) el número de datos fenotípicos necesarios para estimar el efecto de los SNPs (o haplotipos); y (iii) la existencia de efectos no aditivos con influencia sobre el carácter. El número de marcadores utilizados y el LD existente entre ellos deben ser suficientes para permitir estimar el efecto de cada QTL a través de los genotipos (o haplotipos). Meuwissen et al. (2001) estimaron que el valor medio mínimo deseable de r^2,

medida del LD, debe ser de 0,2 ya que al disminuir el valor de r^2, disminuye también la exactitud en la estima de los efectos. La influencia del número de registros fenotípicos para estimar los efectos de los marcadores (o haplotipos) radica en que un aumento de los primeros mejora la precisión con que se determinan los segundos. Así, en ese mismo trabajo Meuwissen et al. (2001) sugirieron, utilizando datos simulados y diferentes metodologías, que para un carácter con una h^2 de 0,3, se requieren, al menos, 2.000 registros fenotípicos; si la h^2 es mayor, el número de registros necesarios será menor ya que el genotipo será un mejor predictor del fenotipo. Por último, los modelos de predicción de los valores genéticos, por definición, incluyen solo los efectos aditivos aunque en algunos casos existe la posibilidad de que un modelo en el que se incluyan efectos de dominancia o epistasia sean mejores predictores de los valores genéticos.

El éxito de la GS se basa en aprovechar el LD entre el marcador y el QTL, asumiendo que los efectos detectados serán igual en todas las poblaciones porque los marcadores están en LD completo con el QTL al que flanquean. El hecho de que la GS se base en la existencia de LD entre el marcador y el QTL y, además, en una fase cromosómica concreta, es decir, un genotipo (o haplotipo) está asociado al QTL produciendo un efecto mayor sobre el fenotipo que otro genotipo (o haplotipo), constituye una desventaja a la hora de aplicar los hallazgos de la GS en poblaciones diferentes a la de referencia. Esto ha sido estudiado en el ganado vacuno, donde ya ha comenzado a implementarse la GS (Hayes et al., 2009). Cuanto más divergentes son las razas o poblaciones, existe una menor probabilidad de que el marcador y el QTL estén en la misma fase en la población utilizada como referencia y en la que es objeto de selección. En un estudio sobre el LD y la persistencia de la fase en diferentes razas bovinas se llegó a la conclusión de que para poder aplicar la GS entre razas tan divergentes como Angus y Jersey, sería necesario disponer de 300.000 marcadores, asumiendo una distancia media marcador-QTL de 5 Kb (De Roos et al., 2008). Además, las poblaciones de referencia que se utilizan para estimar los efectos de los QTL deberían incluir animales de todas las razas objeto de selección para mejorar la exactitud de los GEBVs (Harris et al., 2008).

En el ganado ovino, como hemos comentado anteriormente, continúan utilizándose el testaje de machos y el manejo piramidal de la población como métodos de selección (Carta et al., 2009). La implementación de la GS en esta especie puede ofrecer nuevas oportunidades a la mejora genética ovina como la reducción de los costes derivados de la selección reduciendo el intervalo generacional, el aumento de la ganancia genética, la mejora en el control del

grado de consanguinidad en los rebaños y la posibilidad de inclusión de nuevos caracteres objeto de selección (Duchemin et al., 2012).

Estudios recientes realizados en la raza ovina de aptitud lechera Lacaune han demostrado que la utilización de la GS en esta especie mejora la predicción de los valores genéticos de los animales. El crecimiento de la población de referencia, debido al aumento del número de animales con genotipos y fenotipos, debe mejorar estas predicciones. Aun así, las ganancias que se esperan por el uso de esta metodología en el ganado ovino son menores que las obtenidas en el ganado vacuno hasta el momento debido a que los intervalos generacionales son más cortos y el gasto económico que supone mantener a los animales durante el periodo de testaje es mucho menor en esta especie (Duchemin et al., 2012).

Estudios preliminares llevados a cabo por nuestro grupo de investigación en la raza Churra, coinciden con los hallazgos descritos previamente. Se produce una mejora en la capacidad de predicción de los valores genéticos de los animales cuando se incluyen datos genómicos en el modelo de predicción utilizando el método *single-step* descrito por Aguilar et al. (2010) en lugar de únicamente la información procedente del pedigrí (Sánchez et al., 2012). Para el resto de metodologías testadas, *elastic net* (Friedman et al., 2010) y *marker assited selection* (Boichard et al., 2012), esta mejora en la precisión se produce para caracteres cuyos resultados de detección de QTL son constantes a lo largo de las diez réplicas realizadas (Sánchez et al., 2012).

3. TÉCNICAS DE DETECCIÓN DE GENES DE INTERÉS PRODUCTIVO

La mayoría de los caracteres de interés productivo en las especies ganaderas son cuantitativos, es decir, muestran una distribución fenotípica continua. A lo largo de la historia, se han propuesto dos modelos para explicar la herencia de este tipo de caracteres: el modelo infinitesimal por una parte, y por otra, un modelo en el que un número finito de *loci* influyen sobre el fenotipo. En el modelo infinitesimal se asume que los caracteres están genéticamente determinados por un número infinito de *loci* no ligados y con efecto aditivo, cada uno con un efecto muy pequeño (Fischer, 1918).

Sin embargo, la existencia de una cantidad limitada de material genético heredado y el descubrimiento de que existen alrededor de 25.000 genes o *loci* en el genoma humano (Pennisi, 2003), apoya la teoría de un número finito de *loci* que controlan a nivel genético la varianza fenotípica observada en los caracteres cuantitativos. Hayes y Goddard (2001) estudiaron la distribución de los efectos genéticos en caracteres cuantitativos en ganado vacuno de leche y ganado porcino, llegando a la conclusión de que hay pocos genes con gran efecto y otros muchos con efectos menores que controlan este tipo de caracteres, así como relaciones de interacción o epistasia entre ellos.

El interés por detectar esos genes o regiones portadoras de los mismos en el genoma, en especial los que tienen un efecto mayor sobre el fenotipo, y el potencial uso de esa información para seleccionar animales más productivos, determinó que a finales de los años 80 se comenzara a desarrollar la metodología, tanto estadística como molecular, que permitiría posteriormente llevar a cabo búsquedas sistemáticas de regiones del genoma con influencia sobre caracteres productivos o QTL. Se expone a continuación una breve descripción de las herramientas metodológicas, mapas genéticos, diseños experimentales y métodos de análisis desarrollados con este fin.

3.1. Marcadores genéticos

Actualmente, la detección de genes de interés en producción animal se realiza mediante el estudio de marcadores genéticos. Estos se definen como secuencias de ADN con una localización cromosómica identificable y que presentan una variación detectable entre los individuos de una población y un modo de herencia mendeliano. Las propiedades del

marcador genético ideal son: neutralidad fenotípica, elevado grado de polimorfismo, herencia mendeliana codominante, distribución uniforme a lo largo del genoma, detección sencilla, fácil interpretación genética, alta reproducibilidad y fácil automatización de su determinación (Arranz et al., 2006).

El rápido desarrollo de la Genética Molecular, sobre todo a partir del descubrimiento de la Reacción en Cadena de la Polimerasa (PCR, *Polymerase Chain Reaction*) por Mullis et al. en 1986, ha permitido la detección de diferentes tipos de marcadores genéticos que han sido utilizados en estudios de detección de QTL.

A continuación se describen brevemente los marcadores que han sido utilizados en los estudios realizados a lo largo del desarrollo de la presente Tesis Doctoral: repeticiones en tándem (microsatélites) y variaciones puntuales (SNPs).

3.1.1. Repeticiones en tándem: microsatélites

Las repeticiones en tándem (*Variable Number Tandem Repeats*, Jeffreys et al., 1985) son secuencias de ADN en las que un nucleótido o un fragmento nucleotídico de hasta miles de pares de bases se repite de manera consecutiva. Según la longitud de la repetición podemos observar ADN satélite, ADN minisatélite o microsatélites. Los últimos han sido descritos en todos los organismos vivos y comprenden repeticiones cortas en tándem, de una a seis pares de bases. Estos marcadores se encuentran, por lo general, en regiones no codificantes, presentan un alto nivel de polimorfismo y se distribuyen uniformemente a lo largo del genoma (Arranz et al., 2006).

Aproximadamente el 2,5% del genoma bovino (Bovine Sequencing and Analysis Consortium et al., 2009) y el 3% del genoma humano (Lander et al., 2001) están ocupados por secuencias tipo microsatélite. En el genoma bovino, la repetición dinucleotídica $(AC)_n$ constituye el 0,23% del genoma, mientras que el trinucleótido $(AGC)_n$ constituye el 0,13% (Bovine Sequencing and Analysis Consortium et al., 2009). En función de su estructura, los microsatélites se clasifican en tres grupos:

- Perfectos: un único fragmento repetido *n* veces. Ej: $(AC)_6$.
- Imperfectos: un solo fragmento repetido *n* veces donde hay nucleótidos intercalados entre las repeticiones. Ej: $(CA)_{10}AA(CA)_{12}$.
- Compuestos: distintos fragmentos repetidos en serie. Ej: $(GT)_2(TG)_{10}$.

Para el genotipado de este tipo de marcadores se diseñan cebadores o *primers* que contienen la secuencia flanqueante de la secuencia repetida para poder producir un elevado número de copias de la región que contiene las repeticiones mediante PCR. Posteriormente, se determinan los genotipos que portan los animales midiendo el tamaño de los fragmentos de ADN amplificados. La variación en el número de repeticiones crea diferentes alelos de tamaño aproximado, generalmente, entre 80 y 400 pares de bases.

Una gran ventaja de la PCR es la posibilidad de amplificar simultáneamente varios fragmentos de ADN mediante la técnica conocida como PCR-multiplex, utilizada por primera vez por Chamberlain et al. (1988). Para la amplificación conjunta de varios marcadores microsatélite hay que tener en cuenta que los *primers* deben hibridar a la misma temperatura y los fragmentos resultantes tienen que ser de tamaños diferentes o estar marcados con fluorocromos distintos para que se puedan diferenciar entre sí (Wallin et al., 2002).

Una de las técnicas de laboratorio más comúnmente utilizadas para el genotipado de microsatélites es la electroforesis capilar (Wenz et al., 1998). Para ello es necesario marcar uno de los *primers* con un fluorocromo y procesar las muestras en un secuenciador automático. Así, se han desarrollado protocolos de PCR-multiplex utilizando kits comerciales que permiten amplificar más de 15 microsatélites en una sola reacción de PCR para, por ejemplo, realizar pruebas de paternidad (Glowatzki-Mullis et al., 2007).

En el ganado ovino se han utilizado marcadores tipo microsatélite para la construcción de mapas genéticos (Maddox et al., 2001 entre otros), la realización de pruebas de paternidad (Glowatzki-Mullis et al., 2007), la detección de genes de interés productivo (por ejemplo Gutiérrez-Gil et al., 2009) y el estudio de la diversidad genética entre poblaciones (Uzun et al., 2006).

3.1.2. Variaciones puntuales o SNPs

Los SNPs son variaciones de un solo nucleótido en el ADN genómico. Representan el tipo de variación más frecuente en el genoma, siendo su frecuencia en el caso de la especie humana de un SNP cada 1.000 pares de bases (Arranz et al., 2006). Dentro de las variaciones puntuales o SNPs podemos encontrar dos tipos: las originadas por el cambio de un nucleótido (sustituciones nucleotídicas) y las producidas por la inserción o deleción de una base nucleotídica en la secuencia de ADN (inserciones/deleciones).

En ambos casos, los polimorfismos tipo SNP más interesantes, y menos frecuentes, son los que se producen en las secuencias codificantes de los genes. El primer caso, cambio de base, puede producir un cambio en el aminoácido al traducirse la secuencia a proteína u originarse un codón *stop* con lo que la traducción quedaría interrumpida en ese punto de la cadena de ADN/ARN. En el segundo caso, inserción o deleción de una base, cambia el marco de lectura durante la traducción, pudiendo formarse un polipéptido no funcional o, como en el caso anterior, un codón *stop* prematuro.

Los SNPs, al ser bi-alélicos, son aptos para el genotipado automatizado a gran escala. Además, su distribución uniforme a lo largo del genoma hace que puedan constituir una muestra representativa del mismo, sin necesidad de conocer la secuencia completa del genoma de la especie en estudio. Desde el año 2000, cuando se publicó la primera versión del genoma humano (International Human Genome Sequencing Consortium, 2001), la aparición de la tecnología de secuenciación masiva paralela ha permitido el desarrollo de numerosos proyectos de secuenciación de genomas de distintas especies animales. Como herramienta genómica directamente derivada de estos proyectos de secuenciación, y en base a las características de los marcadores SNPs, se han desarrollado los chips de SNPs. Además de la especie humana en la que el aumento de la densidad de los chips disponibles ha sido ingente (desde 10.000 SNPs en el año 2003, a chips de 4,3 millones de SNPs en la actualidad), los correspondientes Consorcios Internacionales han hecho grandes esfuerzos para el desarrollo de estos chips en casi todas las especies domésticas, vaca (50K, y 800K), cabra (50K), cerdo (60K), caballo (50K), perro (170K) y oveja (50K, y un próximo 700K en desarrollo). Dado que éste último, el *OvineSNP50 BeadChip*, ha sido una herramienta de gran importancia en el desarrollo de esta Tesis Doctoral, los detalles más importantes de su desarrollo se describen más adelante, en el apartado 3.3, tras la descripción del Proyecto de Secuenciación del Genoma Ovino.

3.2. Mapas de referencia

Además de disponer de marcadores genéticos, a la hora de detectar QTL es necesario conocer su localización en el genoma. A partir de los años 90, los marcadores tipo microsatélite se han utilizado para construir mapas genéticos o de ligamiento del genoma ovino en función de la frecuencia de recombinación entre marcadores adyacentes. También se han dedicado importantes esfuerzos al desarrollo de los mapas físicos, en los que mediante

técnicas citogenéticas se mapean los distintos genes y marcadores. En los últimos años, el desarrollo de un genoma virtual ovino (Dalrymple et al., 2007) y de una secuencia de referencia (ISGC et al., 2010) han permitido conocer la posición física de los marcadores, tanto microsatélite como SNP, a lo largo del genoma.

3.2.1. Mapas genéticos o de ligamiento

Los marcadores microsatélite han sido utilizados para el desarrollo de mapas genéticos de media y alta densidad en diferentes especies (Dib et al., 1996 en la especie humana; Ihara et al., 2004 en el ganado vacuno; Maddox et al., 2001 en el ganado ovino). En estos mapas genéticos o de ligamiento se estima la distancia entre marcadores en función de la frecuencia de recombinación entre ellos. La medida de la distancia se realiza en centimorgan (cM), unidad que equivale a un 1% de frecuencia de recombinación. Esta distancia genética se transforma en distancia física asumiendo que 1 cM equivale a 1 Megabase (Mb), relación media estimada en la especie humana y variable a lo largo del genoma (Yu et al., 2001).

En la década de los 90 se hicieron grandes esfuerzos para desarrollar un mapa de ligamiento informativo en el ganado ovino (Crawford et al., 1994; De Gortari et al., 1998). La utilización de rebaños formados por tres generaciones de familias de hermanos completos permitió la construcción de un mapa de ligamiento autosómico de baja densidad (Crawford et al., 1994) y otro para el cromosoma X (Galloway et al., 1996). La segunda generación de mapas comprendía marcadores con una distancia media entre ellos de 6 cM (De Gortari et al., 1998).

Posteriormente, en el año 2001 se publicó un nuevo mapa de media densidad que contenía 1.093 marcadores (Maddox et al., 2001). Este mapa cubría un total de 3.500 cM en autosomas y 132 cM en el mapa del cromosoma X, obtenido en hembras. La distancia media entre marcadores era de 3,4 cM en autosomas y 8,3 cM en el cromosoma X. Las diferentes actualizaciones del mapa genético ovino de referencia, obtenido genotipando animales pertenecientes al rebaño *International Mapping Flock*, se han ido publicando en la página web *Australian Sheep Gene Mapping Website* (http://rubens.its.unimelb.edu.au/~jillm/jill.htm).

La última versión del mapa genético ovino, v5.0 actualizada en Octubre de 2010, contiene 2.516 microsatélites distribuidos a lo largo de los 26 autosomas y el cromosoma X

ovino. Debido al diferente tamaño de los cromosomas el número de marcadores en cada uno varía desde más de 220 en los cromosomas 1, 2 y 3 (cromosomas metacéntricos) hasta una media de 55 en los cromosomas 20 a 26. Se incluyen en este mapa las posiciones de los mapas de ligamiento para el sexo masculino, el femenino y el promedio entre ambos.

Asimismo, en base a algunas poblaciones objeto de estudio por diferentes grupos de investigación, se han publicado diferentes versiones del mapa de ligamiento ovino. Este es el caso de las razas Soay (Beraldi et al., 2006) y Churra (Gutiérrez-Gil et al., 2008a) y de una población Awassi x Merino obtenida por retrocruzamiento (Raadsma et al., 2009b).

3.2.2. Mapas físicos

Los mapas físicos aportan información sobre la posición física de secuencias concretas de ADN en los cromosomas y las distancias entre *loci* se expresan en unidades de distancia en nucleótidos. Se han empleado diferentes tecnologías a lo largo de la historia para el desarrollo de los mapas físicos en el ganado ovino, revisadas por Broad et al. (1997) y Maddox y Cockett (2007), como son las técnicas basadas en hibridación *in situ* (por radiactividad o fluorescencia), el uso de células somáticas o híbridos por radiación y la construcción de clones y secuencias *contig* de regiones de interés.

El cariotipo diploide ovino está formado por 27 cromosomas (2n = 54), siendo tres de los autosomas metacéntricos (OAR1, OAR2 y OAR3), 23 pares telocéntricos (del OAR4 al OAR26), el cromosoma X acrocéntrico y el Y el más pequeño de los metacéntricos. En total, son 566 los *loci* localizados en el cariotipo ovino por técnicas citogenéticas, lo que muestra una distancia media entre ellos de 5,1 Mb (Goldammer et al., 2009).

En los últimos años, con el progreso de los proyectos de secuenciación de los genomas, la propia secuencia de referencia y la posición de marcadores y genes en la misma constituyen el mapa físico sobre el que basar posteriores análisis de asociación a nivel genómico. En esta Tesis Doctoral se ha usado como mapa físico la v2.0 del Genoma Ovino (http://www.livestockgenomics.csiro.au/cgi-bin/gbrowse/oarv2.0/). Dada la repercusión de la disponibilidad de esa secuencia de referencia en los avances que están teniendo y tendrán lugar en el campo de la genómica ovina, describimos a continuación las fases más importantes del desarrollo de esta secuencia de referencia, así como los detalles del desarrollo del *OvineSNP50 BeadChip.*

3.3. Proyecto de Secuenciación del Genoma Ovino y desarrollo del *OvineSNP50 BeadChip*

El Consorcio Internacional para la Genómica Ovina (ISGC, *International Sheep Genomics Consortium*) está desarrollando, desde 2002, estudios que tienen como objetivo último la secuenciación completa del genoma ovino. Este Consorcio ha desarrollado y coordinado el proyecto *SheepHapMap* (www.sheephapmap.org), iniciado en el año 2008. Dentro de este proyecto se incluyeron 3.004 animales pertenecientes a 71 razas ovinas procedentes de África, Asia, Sudamérica, Europa, Oriente Medio, Australasia, Estados Unidos y el Caribe. Los primeros resultados procedentes del análisis de dichas muestran han sido publicados recientemente (Raadsma, 2010; Kijas et al., 2012).

El desarrollo del proyecto de secuenciación del genoma ovino ha estado marcado por el escaso interés que esta especie, filogenéticamente cercana a la especie bovina, ha despertado en los proyectos de secuenciación de genomas completos. La utilización de un borrador del genoma bovino y técnicas de secuenciación de segunda generación sobre el genoma ovino, combinadas con el mapeo por radiación de híbridos (*radiation-hybrid mapping*), constituyeron el primer intento de obtener una secuencia completa del genoma ovino dando lugar al conocido como *Virtual Sheep Genome* (Dalrymple et al., 2007).

Las secuencias obtenidas de un macho de la raza Texel se analizaron mediante técnicas bioinformáticas, alineándolas con los genomas vacuno, canino y humano para ordenarlas y crear secuencias más largas por comparación entre especies. La especie bovina es la especie más cercana filogenéticamente al ganado ovino de la que se disponía del genoma completo (Btau2.0), pero dado el estado inicial del desarrollo de dicho genoma, se utilizaron además las secuencias del genoma del perro, siguiente especie en relación filogenética de la que se disponía del genoma completo (CanFam2.0), y las de la especie humana (Hinrichs et al., 2006), en la que la anotación génica era mucho más completa.

Por último, las secuencias se orientaron y ordenaron sobre el mapa genético ovino disponible en ese momento (versión 4.6 disponible en la página web *Australian Sheep Gene Mapping Website*). De esta manera se creó el *Virtual Sheep Genome*, que cubría el 76% del genoma de esta especie, y se dispuso, por primera vez, de una secuencia de referencia del genoma (Dalrymple et al., 2007). Además se utilizó información disponible sobre el grado de

conservación de los genomas y el mantenimiento de grandes grupos de sintenia entre especies para la anotación en este *Virtual Sheep Genome*.

La rápida evolución de técnicas de secuenciación a nivel genómico, unida al trabajo del ISGC cuyo objetivo último es la secuenciación completa del genoma ovino, han hecho posible, en un tiempo corto en relación a otros genomas desarrollados con anterioridad, pasar de un genoma virtual al borrador de un genoma real, realizado por alineamiento de fragmentos secuenciados en la oveja. Para la creación de una primera versión no-virtual del genoma ovino, se secuenció parcialmente el genoma de seis hembras pertenecientes a las razas Awassi, Merino Australiano, Poll Dorset, Romney, Scottish Blackface y Texel, respectivamente, utilizando una técnica de secuenciación de segunda generación, el 454-FLX. La secuenciación de las tres primeras razas se llevó a cabo en Baylor HGSC (USA) y la de las tres últimas en AgResearch (Nueva Zelanda). Como complemento a esta secuenciación, para mejorar la calidad de las secuencias obtenidas y detectar más polimorfismos, se añadió información obtenida mediante (i) la secuenciación de parte del genoma utilizando un Illumina Genome Analyser (GA) siguiendo el método descrito por Smith et al. (2008) y (ii) la creación de *paired end reads* de diferentes tamaños utilizando una combinación de técnicas de secuenciación, *Sanger* y de próxima generación (ISGC, http://www.sheephapmap.org/genseq.php).

Para las posteriores actualizaciones de este genoma ovino de referencia, además de los datos y técnicas descritas previamente, se secuenció el genoma de una hembra Texel, en Kunming Institute of Zoology y BGI Shenzhen (China) utilizando el Illumina GA; y el genoma de un macho Texel, previamente utilizado en la elaboración de *Virtual Sheep Genome*, en el Roslin Institute (Reino Unido) (ISGC et al., 2010). Los avances del proyecto de secuenciación del genoma ovino continúan. En breve se espera que se disponga en el dominio público la v3.0 de la secuencia de referencia, mientras que en el próximo año (2013) la v4.0 sea añadida a la base de datos ENSEMBL (http://www.ensembl.org/index.html), momento a partir del cual los mayores esfuerzos se centrarán en el subsiguiente proceso de anotación génica (ISGC et al., 2010).

La información procedente de la secuenciación del genoma de los primeros animales incluidos en este proyecto internacional permitió la identificación de miles de polimorfismos puntuales distribuidos a lo largo del genoma ovino que se utilizaron para el desarrollo del

OvineSNP50 BeadChip. Las posiciones físicas de los 590.000 polimorfismos detectados estaban referidas a la secuencia alineada del *Virtual Sheep Genome*. De ellos, los 270.000 heterocigotos en, al menos, dos animales se consideraron polimorfismos "de clase A". Los criterios en los que se basó la selección de SNPs que entraron a formar parte del chip comercial fueron los siguientes:

- distribución uniforme a lo largo del genoma.
- criterios de calidad en cuanto a la técnica de detección: las técnicas adicionales de secuenciación, los métodos *Sanger* e Illumina GA, se utilizaron para buscar más SNPs y estimar de manera más precisa la frecuencia del alelo menos frecuente (MAF, *Minor Allele Frequency*) de los mismos. Por tanto, los SNPs detectados o confirmados por estas dos técnicas se consideraron más probables que los detectados únicamente mediante la secuenciación con 454-FLX.
- se descartan los SNPs con MAF menor de 0,2.
- existe un parámetro técnico a la hora de elaborar el chip que se denomina *probe score* y que representa la probabilidad de que el genotipado funcione. Se desestimaron los SNPs con *probe score* menor de 0,8.
- la secuencia flanqueante (200 pb) del SNP, así como el oligonucleótido al que debe unirse el fragmento de ADN en el chip deben ser únicos, no estar repetidos en otras regiones del genoma.

El problema con este último criterio de calidad es que, en el momento de desarrollo del chip, la secuencia disponible era únicamente el 76% del genoma ovino. Siguiendo estos criterios, 54.241 SNPs "de clase A" fueron seleccionados para ser incluidos en el chip *OvineSNP50 BeadChip* comercializado por la empresa Illumina (Illumina Inc. San Diego, CA).

3.4. Diseños experimentales

Para la detección de QTL es necesaria la planificación de un diseño experimental adecuado para el tipo de análisis de asociación entre genotipos y fenotipos que se va a realizar. En general, existen dos opciones: crear una población específica para el estudio, población experimental, o utilizar una estructura ya existente, es decir, una población comercial.

3.4.1. Poblaciones experimentales

Este tipo de diseño experimental se basa en el establecimiento de una población mediante el cruce de dos razas o variedades que muestran diferencias fenotípicas claras para los caracteres de interés (Lynch y Walsh, 1998). Cruzando animales de estas dos líneas o razas divergentes, animales fundadores (F0), se crea la generación F1 en la que los individuos se consideran heterocigotos para los genes o *loci* con influencia sobre los caracteres de interés. Posteriormente, para estudiar la segregación de esos genes se pueden llevar a cabo dos tipos de cruzamientos:

- si se cruzan entre sí los animales de la F1 se crea una población F2 (o se sigue cruzando para crear una población F3, etc).
- si realizamos un cruzamiento entre animales F1 y animales fundadores, obtenemos una población de retrocruzamiento (backcross) de una o varias generaciones.

Todos los animales de la segunda (F1) y tercera generación se genotipan para los marcadores seleccionados, mientras que los datos fenotípicos se miden solo en la última generación.

La ventaja de utilizar este tipo de diseños experimentales radica en que los animales se encuentran en granjas experimentales donde el control ambiental y la medición de los datos fenotípicos exhaustivos son posibles y más sencillos que en las granjas comerciales. Además, el poder de estos experimentos se maximiza cuando las líneas F0 divergentes muestran alelos fijados o casi fijados para las mutaciones que controlan los fenotipos objeto de estudio. Como principal inconveniente cabe destacar los esfuerzos tanto económicos como de tiempo que son necesarios para el establecimiento de la población experimental. Además, los resultados obtenidos en estas poblaciones experimentales son en algunas ocasiones difíciles de implementar en poblaciones comerciales (Georges, 2007).

Este tipo de diseño experimental ha sido ampliamente utilizado en especies ganaderas como cerdos o pollos (Rothschild et al., 2007; Abasht et al., 2006). En el ganado vacuno se han utilizado cruces experimentales entre razas con aptitudes productivas divergentes (por ejemplo Esmailizadeh et al., 2011) o incluso entre especies distintas como son el ganado vacuno y el cebú (Kim et al., 2003). En el ganado ovino lechero, algunos de los estudios de búsqueda de QTL realizados se han basado en poblaciones experimentales. Así, una población de retrocruzamiento Sarda x Lacaune (Barillet et al., 2006; Carta et al., 2008), otra

Awassi x Merino (Raadsma et al., 2009a) y una tercera formada por retrocruzamiento de machos East Friesian y hembras Dorset (Mateescu y Thonney, 2010a) han sido objeto de estudio.

3.4.2. Poblaciones comerciales

Los diseños experimentales basados en poblaciones comerciales son los que tienen por objetivo detectar los *loci* responsables de la varianza fenotípica observada en las mismas poblaciones que se utilizan para la obtención de productos de interés comercial. La utilización de poblaciones comerciales para la búsqueda de QTL en especies ganaderas para producción lechera, como son el ganado vacuno y el ovino, se basa en el uso rutinario de la inseminación artificial (IA) en dichas poblaciones. De esta manera se aprovechan las estructuras familiares derivadas del uso de la IA consistentes en grandes familias en las que se puede estudiar la segregación de los marcadores genéticos y los QTL. Los principales diseños experimentales que se pueden aplicar a este tipo de poblaciones son el diseño hija y el diseño nieta.

En el diseño hija, propuesto por Neimann-Sörensen y Robertson (1961), los datos de producción se obtienen en las hijas y se genotipa toda la población para los marcadores genéticos, incluyendo a los machos. En el ganado vacuno lechero la IA ha sido utilizada rutinariamente durante bastantes años, por ello Weller et al. (1990) propusieron el diseño nieta, con un poder estadístico mayor para la detección de QTL. En este tipo de diseño experimental, se dispone de familias de medio-hermanos de toros evaluados, se mide la producción de leche en las hijas de esos toros, nietas de los machos fundadores, y se genotipan los machos fundadores y los padres de cada familia para la búsqueda de QTL. En este tipo de diseños se reduce de manera sustancial el número de animales que hay que genotipar y aumenta el poder estadístico para la detección de QTL con respecto al diseño hija. Otra de las ventajas del diseño nieta es la facilidad en la recogida de muestras, ya que los animales que se genotipan se encuentran en centros de selección y reproducción en vez de distribuidos por los diferentes rebaños. Sin embargo, en algunas poblaciones, como es el caso del ganado ovino, la estructura de la población impide utilizar el diseño nieta, así que en esta especie el diseño hija es el más utilizado (Arranz y Gutiérrez-Gil, *en prensa*).

La mayor ventaja de la utilización de poblaciones comerciales consiste en que los resultados obtenidos son directamente aplicables en programas de selección asistida por marcadores en la población estudiada. Como inconvenientes cabe destacar el mayor esfuerzo

que requiere la toma de muestras y la influencia de factores ambientales sobre los caracteres productivos, ya que las condiciones ambientales no son las mismas en todas las granjas mientras que en las poblaciones experimentales, aunque de forma costosa, estos factores son más fáciles de controlar (Arranz y Gutiérrez-Gil, *en prensa*).

En el ganado ovino el diseño nieta se ha utilizado en las razas francesas Lacaune y Manech para la búsqueda de QTL (Barillet et al., 2006), ya que en estas poblaciones los programas de mejora genética llevan bastante tiempo funcionando y la IA está implantada desde hace tiempo suficiente para proporcionar la estructura requerida para tal diseño. En cambio, en la raza Churra se han realizado varios barridos genómicos para la detección de QTL para caracteres lecheros y morfológicos bajo un diseño hija, más adecuado para la estructura poblacional de esta raza (Gutiérrez-Gil et al., 2007, 2008b, 2009, 2011).

3.5. Fenotipos

Los caracteres más estudiados en los proyectos de detección de QTL en el ganado lechero, tanto ovino como vacuno, son la cantidad de leche (MY, *milk yield*), las cantidades de proteína (PY, *protein yield*) y grasa en la leche (FY, *fat yield*), los porcentajes de proteína (PP, *protein percentage*) y grasa (FP, *fat percentage*) y el recuento de células somáticas (SCS, *somatic cell score*) (Arranz y Gutiérrez-Gil, *en prensa*).

En el ganado ovino de raza Churra, otros caracteres han sido valorados como posibles objetivos de selección y, por tanto, candidatos a la búsqueda de QTL con influencia sobre ellos (Othmane et al., 2002b). Estos son el contenido en caseínas, proteínas mayoritarias de la leche ovina, y el rendimiento quesero, pero su inclusión en los programas de mejora ha sido descartada debido a su difícil medición, en el caso del contenido en caseínas, o la baja heredabilidad del rendimiento quesero (h^2 = 0,08) (Othmane et al., 2002b).

Para la búsqueda de QTL se estudian como datos fenotípicos los valores brutos medidos en el animal, corregidos para una serie de factores ambientales que influencian dicho carácter. Estos factores ambientales han sido estudiados en algunas razas ovinas con el objetivo de determinar su influencia sobre los distintos fenotipos utilizados en la búsqueda de QTL, como es el caso de las razas East Friesian (Scharch et al., 2000; Kralickova et al., 2012), Churra (Othmane et al., 2002a), Rahmani y Chios (Allah et al., 2011), y un grupo de animales de las razas Tsigai, Improved Valachian y Lacaune (Oravcova et al., 2007). En la

raza Churra, utilizada en los trabajos presentados en esta Tesis Doctoral, los factores más importantes que presentan una influencia significativa sobre caracteres lecheros fueron el rebaño-día de control, fase de la lactación, edad de la oveja y número de corderos (Othmane et al., 2002a).

En el caso de poblaciones comerciales lecheras, las medidas fenotípicas se toman a través del Control Lechero Oficial y, dado que los valores obtenidos se utilizan en los programas de mejora genética, los datos fenotípicos que pueden utilizarse son los valores genéticos de los animales (EBV, *estimated breeding values*) o las desviaciones en producción del animal respecto de la media poblacional (YD, *yield deviations* o DYD, *daughter yield deviations*). Debido a que el número de controles de los animales es diferente, estos valores deben ser corregidos por la precisión, que aumenta con el número de mediciones (Thomsen et al., 2001).

En el cálculo de los EBV se tienen en cuenta las relaciones entre los animales (Israel y Weller, 1998), lo que hace que disminuya la variabilidad en los valores fenotípicos y que el efecto asignado al QTL sea menor (Arranz y Gutiérrez-Gil, *en prensa*), por lo que la utilización de estos valores dificulta la detección de QTL. Además, Thomsen et al. (2001) mostraron que el uso el EBV presenta una menor potencia estadística y, por tanto, recomendaron el uso de DYD. En el ganado ovino lechero este tipo de fenotipos han sido utilizados en un barrido genómico junto con datos de producción (Mateescu y Thonney, 2010a).

Las YD se estiman como las desviaciones de la media de producción de la población, corregidas para una serie de factores ambientales, variables en función de los que influyen sobre el carácter (VanRaden y Wiggans, 1991). En el caso de los diseños nieta, dado que los machos no tienen datos de producción lechera, se calculan las DYD que equivalen a los valores productivos de las hijas de cada uno de los machos de IA objeto de estudio. Las YD han sido utilizadas en los barridos genómicos realizados en la raza Churra (Gutiérrez-Gil et al., 2007, 2009) y las DYD en los análisis de las razas Lacaune y Manech, en las que se utilizan diseños nieta (Barillet et al., 2006).

Las variables dependientes a analizar en el modelo de detección de QTL o asociación pueden estimarse mediante la corrección de los factores ambientales a tener en cuenta en un paso previo o en el mismo paso en que se realiza el análisis de asociación o ligamiento. El uso

del análisis en dos pasos, con una pre-corrección de los fenotipos y posterior búsqueda de QTL, facilita de manera sustancial los cálculos. Sin embargo, hay que tener en cuenta, tal y como indican Crooks et al. (2009), que el uso de los fenotipos pre-corregidos parece sesgar la estimación del efecto del QTL, ya que debido a que parte de la varianza que podría ser debida al QTL se elimina en el primer paso, los efectos poligénicos se subestiman y se pierde potencia estadística para la detección de QTL. Desde otro punto de vista, el uso de esta aproximación más conservadora puede ser adecuada cuando se intenta disminuir el riesgo de detectar falsos positivos.

3.6. Metodología

La búsqueda de regiones del genoma con influencia sobre caracteres productivos se ha realizado desde dos enfoques o tipos de análisis. El primero, denominado estudio de genes candidatos, se basa en la secuenciación de genes que, debido a su función fisiológica conocida, podrían ser portadores de mutaciones con influencia sobre el carácter estudiado. Existen dos problemas en este tipo de análisis. En primer lugar, existe una limitación importante en los estudios basados en genes candidatos ya que requieren de conocimientos previos tanto a nivel fisiológico como bioquímico, funcional y de posición sobre los candidatos *a priori* (Zhu y Zhao, 2007). Además, normalmente, hay un gran número de genes que afectan a los caracteres productivos, por lo tanto, se requiere secuenciar un gran número de *loci* y estudiar la asociación de los polimorfismos detectados en muchos animales, lo que hace difícil el control de falsos positivos o error tipo I. También hay que tener en cuenta que la mutación causal puede encontrarse en regiones no codificantes o reguladoras del gen que no hayan sido secuenciadas o localizarse en un gen no estudiado, por no ser un candidato *a priori*.

Como alternativa al estudio de genes candidatos, y gracias al descubrimiento de los marcadores genéticos, se comenzó a mapear QTL por medio de barridos a nivel genómico. Esta técnica se basa en el estudio de la asociación entre marcadores genéticos distribuidos a lo largo del genoma y el carácter objeto de estudio. El avance en las técnicas moleculares, nos ha permitido multiplicar por mil el número de marcadores moleculares utilizados en los barridos genómicos y, así, aumentar el poder de detección de QTL y la resolución del mapeo. La detección de QTL mediante esta aproximación debe de confirmarse en una población o poblaciones independientes antes de centrar grandes esfuerzos al mapeo fino de la región en

cuestión. Estos estudios de mapeo fino posteriores tienen por objetivo identificar la mutación responsable del efecto inicialmente detectado, o mutación causal (QTN), o al menos identificar un polimorfismo en completo desequilibrio de ligamiento que pudiera utilizarse como marcador en un programa de selección asistida por marcadores.

Históricamente, se han utilizado dos métodos de análisis en la detección de QTL, uno basado en el *ligamiento* (LA, *Linkage Analysis*) y otro en la *asociación* (*Association Mapping*). Ambos se basan en la detección de desequilibrio de ligamiento o asociación entre el marcador genético y la mutación causal; en el primero se utiliza la información familiar y en el segundo se explota la información a nivel poblacional. Ambos métodos tienen sus ventajas e inconvenientes.

En el caso de los estudios basados en LA, el principal problema es la necesidad de un elevado número de animales por familia y de un diseño experimental adecuado (por ejemplo, familias de medio-hermanas); en caso contrario, el intervalo de confianza (CI) de cada QTL se extiende a lo largo de amplias regiones cromosómicas. Estos CI extensos hacen que los resultados obtenidos en la mayoría de los estudios de ligamiento no sean suficientemente informativos, es decir, se obtienen CI en los que suele haber cientos de genes candidatos (Georges, 2007). Por otra parte, es difícil utilizar estos resultados en selección asistida por marcadores, ya que si la asociación entre marcador y QTL no es suficientemente estrecha, no se mantiene entre poblaciones ni entre familias.

Desde el primer estudio realizado en vacuno de leche (Georges et al., 1995), este tipo de análisis, basado en LA, se ha utilizado de manera rutinaria en la detección de QTL en las diferentes especies domésticas, tal y como se recoge en la base de datos *Animal QTL Database* (http://www.animalgenome.org/cgi-bin/QTLdb/index). Algunos ejemplos de dichos estudios son:

- en ganado vacuno: caracteres lecheros (Khatkar et al., 2004), cárnicos (Esmailizadeh et al., 2011), reproductivos (Ashwell et al., 2004) o relacionados con la sanidad animal (Schulman et al., 2004)
- en ganado ovino: caracteres lecheros (Gutiérrez-Gil et al., 2009), de producción de lana (Purvis y Franklin, 2004), reproductivos (Mateescu y Thonney, 2010b) o relacionados con la sanidad animal (Gutiérrez-Gil et al., 2007)

- en ganado caprino: caracteres lecheros (Roldán et al., 2008) o de resistencia a parásitos (De la Chevrotière et al., 2012)
- en ganado porcino: caracteres de crecimiento y calidad de la carne (Liu et al., 2007), perfil de ácidos grasos en la carne (Clop et al., 2003) o reproductivos (Bidanel et al., 2008)
- en avicultura: caracteres de producción de huevos (Schreiweis et al., 2006), carne (Gao et al., 2009) o *foie gras* (Kileh-Wais et al., 2012).

La utilización de información poblacional, basada en desequilibrio de ligamiento (LD), a través de estudios de asociación, en contraposición a los diseños familiares, se ha propuesto como alternativa en el mapeo de QTL (Andersson, 2009). En este caso, debemos disponer de una densidad de marcadores tal que podamos detectar LD entre el marcador y la mutación causal a nivel poblacional. La utilización de este tipo de análisis ha estado ligada a los últimos avances que han tenido lugar en el campo de la Genómica, alcanzando un punto máximo con el desarrollo de los chips de SNPs. Así, se han realizado estudios de asociación a nivel genómico (GWAS, *Genome-Wide Association Studies*) en ganado vacuno lechero (Schopen et al., 2011) y de aptitud cárnica (Bolormaa et al., 2011), en distintas producciones avícolas (Gu et al., 2011; Liu et al., 2011), en cerdos (Onteru et al., 2012) y en ganado ovino (Becker et al., 2010; Mömke et al., 2011; Zhao et al., 2011, 2012).

El estudio previo de la extensión del LD a lo largo del genoma nos sirve para conocer la arquitectura molecular del genoma y realizar una estimación del número de marcadores que necesitamos para detectar la asociación existente entre una región del genoma y un carácter productivo. En una población en la que la extensión del LD es grande, se necesitará una menor densidad de marcadores; por el contrario si el LD es pequeño, muchos más marcadores serán necesarios para obtener el mismo poder en el experimento (Meadows et al., 2008).

En la misma línea, la eficiencia de la selección genómica se basa en el LD entre marcadores (Goddard, 2009; Daetwyler et al. 2010), ya que su objetivo es detectar los efectos que los genes tienen sobre los caracteres de producción a través del estudio de marcadores en LD con la/s mutaciones causales. Así, se han estimado el número de marcadores necesario para alcanzar una eficiencia adecuada tanto en selección genómica como en mapeo fino de QTL en función del tamaño efectivo de la población (*Ne*), el número efectivo de marcadores en LD (*Me*) y la longitud del genoma (Goddard, 2009; Daetwyler et al. 2010). Asumiendo un

Ne de 100 animales, número mínimo para asegurar una población viable en el tiempo (Meuwissen, 2009), un valor de *Me* de 20 marcadores, que representa el número de marcadores en cada segmento cromosómico independiente, y una longitud de 30 Morgans, se necesitarían 60.000 marcadores para obtener una potencia de estadística así como una precisión de mapeo adecuadas.

Los estudios de la extensión del LD en poblaciones de ganado vacuno y ovino, han revelado que ésta varía entre cromosomas e, incluso, existe cierto grado de LD entre marcadores situados en cromosomas diferentes (McRae et al., 2002; Farnir et al., 2002). En base a esto varios autores (Farnir et al., 2002; Meuwissen et al., 2001, 2002) han sugerido combinar la información de ligamiento con la proporcionada por el LD a nivel poblacional (LDLA) para la detección, y sobre todo, el mapeo fino de regiones portadoras de QTL. La información procedente de LD y LA se combina en la matriz de identidad por descendencia. Si tenemos, por ejemplo, un diseño hija y establecemos la fase cromosómica de los marcadores tendremos dos haplotipos por macho y los haplotipos paterno y materno de las hijas. En este caso, la información de LD poblacional estará contenida en los haplotipos del macho y los heredados de la madre en las hijas y los haplotipos paternos de la descendencia aportarán la información familiar (LA). Si el tamaño familiar es adecuado, es decir, si el número de haplotipos que aportan información sobre LD es representativo de la población, podemos afirmar que el diseño hija es adecuado para un análisis basado en LDLA (Hayes, 2008).

Este último método de análisis, aunque ha sido sugerido para la detección de QTL a nivel genómico, ha sido utilizado de manera más habitual para el mapeo fino de regiones cromosómicas donde previamente se han descrito QTL (Uleberg et al., 2005; Druet et al., 2008; Schulman et al., 2009; Olsen et al., 2010; Lee et al., 2011).

4. DETECCIÓN DE QTL EN GANADO OVINO LECHERO

En ganado ovino, se han llevado a cabo barridos genómicos para la detección de QTL en poblaciones con aptitudes productivas distintas. Se han estudiado caracteres lecheros (Barillet et al., 2006; Gutiérrez-Gil et al., 2009; Raadsma et al., 2009a; Mateescu y Thonney, 2010a), relacionados con la lactación (Jonas et al., 2011), de crecimiento y de características de la canal (Walling et al., 2004; Hadjipavlou y Bishop, 2009), de calidad de la carne (Karamichou et al., 2006; Cavanagh et al., 2010), reproductivos (Mateescu y Thonney, 2010b) y de morfología mamaria y corporal (Gutiérrez-Gil et al., 2008b, 2011). Los resultados obtenidos hasta el momento en producción de leche sugieren que muchos de los efectos detectados son específicos de población, poco significativos estadísticamente y se extienden a lo largo de amplios intervalos de confianza (Carta et al., 2009).

Hasta el momento, la metodología utilizada para la detección de QTL con influencia sobre caracteres de producción de leche en el ganado ovino se corresponde con un mapeo por intervalos, en el que se calcula la probabilidad de que un QTL esté en cada uno de los puntos del genoma, basándose en la información familiar o de ligamiento. En los diseños utilizados (diseño hija, diseño nieta o retrocruzamiento), el mapeo por intervalos no es igual de informativo en todo los puntos ya que no todas las familias serán heterocigotas para el QTL y los marcadores, por lo que la posición puede estar sesgada hacia el intervalo más informativo. Asimismo, el número de marcadores utilizados hasta el momento y su informatividad hacen que los resultados comprendan amplios intervalos de confianza, incluso cromosomas prácticamente completos.

Los resultados obtenidos en los barridos genómicos para caracteres de producción de leche realizados hasta el momento (Barillet et al., 2006; Gutiérrez-Gil et al., 2009; Raadsma et al., 2009a; Mateescu y Thonney, 2010a, entre otros) se resumen en la Figura 3 tomada de la revisión de Arranz y Gutiérrez-Gil (*en prensa*). Además, se han estudiado cromosomas candidatos a ser portadores de QTL con influencia sobre caracteres lecheros como es el caso de OAR1 (Calvo et al., 2004a; 2006), OAR2 (Calvo et al., 2004b), OAR3 (Singh et al., 2007), OAR6 (Díez-Tascón et al., 2001; Arnyasi et al., 2009), OAR9 (Rozen, 1999) OAR11 (García-Fernández et al., 2010b), OAR20 (Singh et al., 2007) y OAR22 (García-Fernández et al., 2010a).

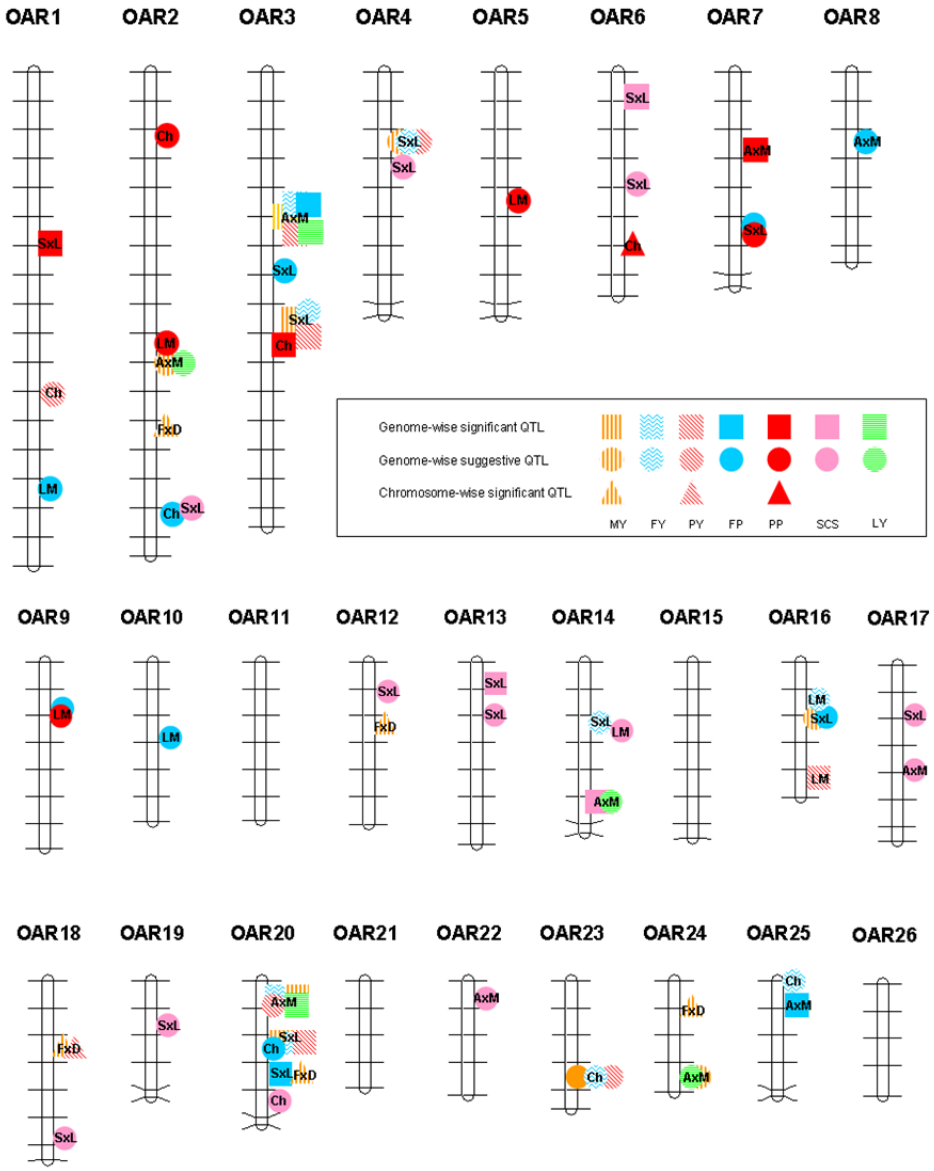


Figura 3 - Resumen de los resultados obtenidos hasta el momento en la búsqueda de QTL para caracteres de producción de leche (cantidad de leche, MY; cantidades de proteína y grasa, PY y FY; porcentajes de proteína y grasa, PP y FP; cantidad de lactosa, LY; recuento de células somáticas, SCS). Las poblaciones a las que se refiere cada una de las siglas son: AxM (Awassi x Merino), Ch (Churra), FxD (East Friesian x Dorset), LM (Lacaune y Manech) y SxL (Sarda x Lacaune). Los umbrales de significación a los que se refieren los resultados son: significativo a nivel genómico (*genome-wise significant QTL*), sugestivo a nivel genómico (*genome-wise suggestive QTL*) y significativo a nivel cromosómico (*chromosome-wise significant*). Imagen tomada de Arranz y Gutiérrez-Gil, *en prensa*.

En los barridos genómicos realizados en ganado ovino de leche, dentro del proyecto europeo *GeneSheepSafety* (QLK5CT20000656), se analizaron tres poblaciones: (i) una población experimental producida por retrocruzamiento Sarda x Lacaune (SxL); (ii) animales de las razas Lacaune y Manech bajo un diseño nieta; y (iii) una población comercial de raza Churra bajo un diseño hija (Barrillet et al., 2006). En la población experimental SxL, se

detectaron QTL en los cromosomas 1, 3, 4, 7, 14, 16, 17 y 20, significativos a nivel genómico; en las razas Lacaune y Manech los cromosomas 1, 2, 5, 9, 10, 14, 16, 17 fueron asociados a nivel genómico con varios de los caracteres estudiados; los resultados de la raza Churra se presentan más adelante y con mayor detalle ya que constituyen el punto de partida de esta Tesis Doctoral.

En la población experimental SxL, los QTL detectados en tres de los cromosomas, OAR3, OAR16 y OAR20, influían sobre las cantidades de leche, proteína y grasa con efectos en todos ellos del mismo signo y magnitudes similares, lo cual sugería que el mismo QTL influía sobre los tres caracteres en cada cromosoma (Barillet et al., 2006). Además, se observó una coincidencia entre los QTL detectados para porcentajes de proteína y grasa en los cromosomas OAR7 y OAR16 (Barillet et al., 2006).

En cuanto a los resultados obtenidos en las razas Lacaune y Manech cabe destacar la presencia de un QTL en OAR2 y otro en OAR9 con influencia sobre los porcentajes de grasa y proteína, aunque con efectos de magnitudes diferentes sobre ambos caracteres. El resto de QTL detectados a nivel genómico de significación mostraron asociación con un único carácter (Barillet et al., 2006).

Posteriormente, Raadsma et al. (2009a) estudiaron una población producida por retrocruzamiento de dos líneas con aptitudes productivas divergentes, una, Awassi y otra Merino. Se detectaron 24 QTL (13 significativos y 11 sugestivos a nivel cromosómico) en diez cromosomas (2, 3, 6, 7, 8, 9, 14, 20, 24 y 25). Los cromosomas 3 y 20 fueron los portadores de un mayor número de QTL para seis y cuatro de los caracteres estudiados, respectivamente. En ambos se detectaron asociaciones significativas con producción de leche, grasa, proteína y lactosa en los primeros 100 días de lactación. Estos dos cromosomas, OAR3 y OAR20, habían sido previamente descritos como portadores de QTL para los caracteres porcentaje de proteína en leche y persistencia de la lactación en la misma población Awassi x Merino (Singh et al., 2007).

En otra población producida por retrocruzamiento de machos East-Friesian, especializados en producción láctea, y hembras Dorset, seleccionadas para una mayor producción cárnica (FxD), se realizó otro barrido genómico (Mateescu y Thonney, 2010a). Basándose en el análisis de marcadores microsatélite, se detectaron regiones asociadas con producción lechera en los cromosomas 2, 12, 18, 20 y 24. La región de OAR20 con influencia

sobre el EBV para la cantidad de leche, coincide con la región descrita por Barillet et al (2006), lo que parece confirmar la importancia de este cromosoma como portador de genes que influyen sobre caracteres de interés en la producción de leche en ganado ovino.

A modo de resumen de los resultados previamente descritos, los cromosomas OAR3 y OAR20 parecen ser los portadores de un mayor número de QTL. En el OAR3, se han detectado tres regiones significativamente asociadas con caracteres de producción de leche en la región central del cromosoma en tres de las poblaciones objeto de estudio, SxL, Awassi x Merino, y Churra (Figura 3). En el caso del OAR20 se han descrito QTL a lo largo de todo el cromosoma en las poblaciones Awassi x Merino, SxL, Churra y FxD con influencia sobre las cantidades de leche, grasa, proteína, el porcentaje de grasa, la cantidad de lactosa y el recuento de células somáticas (Figura 3).

Por otra parte, los estudios de asociación en cromosomas candidatos también han permitido detectar otras regiones y genes con influencia sobre caracteres productivos. En la raza Manchega, estudios realizados en los cromosomas OAR1 y OAR2 revelaron la asociación entre polimorfismos detectados en los genes *FABP3* (*fatty acid binding protein 3*), *AMY* (*amylase*), *SLC27A3* (*solute carrier (fatty acids transporter) family 27, member 3*) y *ANXA9* (*annexin A9*) y diferentes caracteres de producción lechera, aunque la mayoría de dichas asociaciones fueron significativas a nivel intrafamiliar, no en el conjunto de la población estudiada (Calvo et al., 2004a, 2004b, 2006). También un estudio de asociación en OAR6 en tres poblaciones producidas por el cruzamiento de las razas Awassi y Merino mostró la existencia de asociación entre caracteres productivos (producción y composición de la leche y recuento de células somáticas) y los microsatélites utilizados en dicho estudio (Arnyasi et al., 2009).

4.1. Detección de QTL en la raza Churra

Los primeros estudios de QTL para caracteres lecheros en la raza Churra, se basaron en el análisis de cromosomas candidatos, con el estudio del cromosoma 6 (Díez-Tascón et al., 2001) y del cromosoma 9 (Rozen, 1999). Posteriormente, en esta misma raza, se llevó a cabo un barrido genómico en busca de regiones portadoras de QTL para diferentes caracteres de interés en producción de leche, tales como los clásicos caracteres de producción lechera (Gutiérrez-Gil et al., 2009), caracteres de morfología mamaria (Gutiérrez-Gil et al., 2008b), morfología corporal (Gutiérrez-Gil et al., 2011) y el recuento de células somáticas, como

parámetro relacionado con la resistencia a la mamitis (Gutiérrez-Gil et al., 2007). En todos estos estudios se utilizaron microsatélites como marcadores genéticos y análisis de ligamiento bajo un diseño-hija.

En el barrido genómico para caracteres de producción lechera (Gutiérrez-Gil et al., 2009) se estudiaron los caracteres MY, PY, FY, PP y FP. Únicamente se obtuvo un resultado significativo a nivel genómico en el cromosoma 3 para el carácter PP. El resto de resultados, significativos a nivel cromosómico, se resumen en la Tabla 1 (tomada de Gutiérrez-Gil et al., 2009). Así, se detectaron asociaciones significativas a nivel cromosómico en los cromosomas 1, 2, 20, 23 y 25 cuyos intervalos de confianza para la posición del QTL cubrían entre 39 cM (Kosambi, OAR23 para MY) y 274 cM (Kosambi, OAR1 para PY).

Nivel de significación	**OAR[1]**	**Carácter[2]**	**Posición cM[3]**	**Marcadores flanqueantes[4]**	**Familias segregantes[5]**
Significativo a nivel genómico	3	PP	186 [165–205]	**KD103**-OARVH34	4
Significativo a nivel cromosómico	1	PY	210 [1-274]	ILSTS004-CSSM4	2
	2	PP	25 [2-256]	**MCM147**	2
	2	FP	274 [48-292]	BMS356-OARFCB11	3
	20	FP	69 [21-116]	**OLADRBPS**	3
	23	MY	99 [76-115]	MCM136-URB031	2
	23	FY	100 [7-115]	MCM136-URB031	2
	23	PY	100 [20-115]	MCM136-URB031	-
	25	FY	1-2 [1-75]	**MCM200**-ILSTS060	2

Tabla 1 - Resumen de los resultados significativos del barrido genómico descrito por Gutiérrez-Gil et al, 2009. [1]OAR, cromosoma; [2]Carácter al que se refieren los resultados: cantidad de leche (MY), porcentajes de proteína y grasa (PP, FP, respectivamente) y cantidades de proteína y grasa (PY y FY, respectivamente); [3]Posición del valor máximo del test F en cM Kosambi, el valor entre corchetes indica el intervalo de confianza obtenido por *bootstrapping* (Visscher et al, 1996); [4]Marcadores flanqueantes de la posición del QTL, en negrita los marcadores que se encuentran a menos de 2 cM Kosambi del valor máximo del test F; [5]Número de familias segregantes detectadas.

Mientras que el CI del QTL de OAR3 presentó una longitud de 40 cM, los resultados obtenidos en el resto de los cromosomas, significativos a nivel cromosómico, muestran CI

muy amplios. Así por ejemplo, la longitud total del mapa de ligamiento de OAR1 es de 327.2 cM (Kosambi) y el CI del QTL identificado en este cromosoma abarca desde la posición 1 a 274 cM (Kosambi). Este es uno de los problemas más frecuentes de los barridos genómicos basados en el mapeo por intervalos y en mapas de microsatélites de baja-media densidad (Georges, 2007).

En cuanto al estudio de cromosomas candidatos, dentro del proyecto AGL2005-04321 del Ministerio de Ciencia e Innovación, se han realizado en la raza Churra otros estudios de detección de QTL en OAR11, cromosoma portador de los genes *ACACA* (*acetyl-CoA carboxylase alpha*) y *FASN* (*fatty acid synthase*) (García-Fernández et al., 2010b) y en OAR22 donde se localiza el gen *SCD* (*stearoyl-CoA desaturase*) (García-Fernández et al., 2010a). Ambos cromosomas se estudiaron, principalmente, como candidatos a ser portadores de QTL con influencia sobre la composición en ácidos grasos de la leche, pero, dada la disponibilidad fenotípica, también se estudiaron caracteres clásicos de producción lechera. Los análisis de asociación y ligamiento no revelaron ningún QTL para los caracteres de producción lechera (MY, PP y FP) en estos cromosomas.

5. DEL QTL AL QTN

Tanto en los estudios clásicos de detección de QTL, basados en ligamiento, como en los estudios de asociación, se han detectado multitud de regiones portadoras de genes con influencia sobre los caracteres estudiados (Georges, 2011). Sin embargo, son muy reducidos los casos en los que la verdadera mutación causal (QTN) ha sido identificada y su causalidad probada (Georges, 2007).

Para la confirmación del efecto sugerido para una mutación identificada como causal o QTN candidato, la estrategia seguida en el estudio de especies modelo, como por ejemplo el ratón, es la creación de animales *knock-out* (Ron y Weller, 2007). Con el objetivo de comprobar el efecto de un gen candidato sobre un fenotipo, se utilizan técnicas de ingeniería genética para crear animales en los que no se produzca la proteína codificada por dicho gen. Sin embargo, esta estrategia no es útil en las especies ganaderas debido a los largos intervalos generacionales, el alto coste de mantenimiento de cada animal y la dificultad para producir los animales *knock-out* (Ron y Weller, 2007).

Por tanto, en las especies ganaderas se han propuesto, como forma de evaluar la causalidad de una mutación, una serie de requisitos que el QTN putativo debería cumplir para ser considerado causante del efecto genético inicialmente detectado (Glazier et al., 2002; Rebbeck et al., 2004; Ron y Weller, 2007). A continuación se presentan, basándonos principalmente en el listado proporcionado por Ron y Weller (2007), los requisitos que debe cumplir una mutación para ser considerada el QTN:

1. El QTN está localizado dentro del CI definido por ligamiento.
2. El QTN está localizado en el CI estimado por desequilibrio de ligamiento.
3. La función del gen donde se localiza la mutación está relacionada con el fenotipo.
4. El patrón de expresión del gen se corresponde con su función, es decir, si el efecto del gen es positivo sobre el carácter, la expresión del mismo en animales más productivos es mayor.
5. Si hay organismos *knock-out*, el carácter objeto de estudio se encuentra afectado.
6. Los estudios funcionales muestran el mismo efecto sobre el fenotipo que el análisis de QTL.
7. Existe concordancia entre genotipos y fenotipos en la población base (estudiada).
8. Se identifica concordancia entre genotipos y fenotipos en otras poblaciones.

9. Los cambios en las frecuencias alélicas de la mutación están en concordancia con los criterios de selección seguidos.
10. El efecto del QTN se corresponde con el efecto observado en el análisis de ligamiento.
11. No hay otros polimorfismos que muestren resultados significativos cuando se incluye el QTN putativo en el modelo de análisis.

Para probar la causalidad de una mutación es necesario seguir una serie de pasos en los estudios de QTL. A continuación se presenta un ejemplo en cuatro pasos descrito por Glazier et al. (2002). En primer lugar, se debe obtener evidencia estadística de la asociación de una región del genoma con el carácter de interés. Para ello se pueden utilizar diferentes metodologías descritas en el apartado 3.6. En general, en los estudios basados en ligamiento los CI son amplios, entre 20 y 40 cM, por lo que se obtienen regiones donde puede haber entre 200 y 400 genes (Georges, 2007). Los nuevos estudios de asociación basados en chips de SNPs pueden mostrar intervalos más reducidos, pero normalmente son regiones con varios genes candidatos.

El segundo paso es el mapeo fino de las regiones candidatas con el objetivo de reducir el CI al máximo. Se trata de reducir el CI de manera que el número de genes candidatos nos permita la realización de estudios funcionales. Para ello es necesario conocer cuáles son los factores más importantes que influyen sobre la resolución de mapeo de QTL (Georges, 2007):

- Densidad de marcadores: en los análisis de búsqueda de QTL se trata de situar el QTL en el intervalo entre dos marcadores, cuanto menor sea la distancia entre estos, mayor será la resolución de mapeo.
- Número de meiosis: los cromosomas recombinantes son los únicos que proporcionan información útil de mapeo.
- Determinación del estatus del alelo del QTL: se refiere a la exactitud con la que se puede determinar el genotipo del QTL en un individuo a través de la información proporcionada por los marcadores.
- Naturaleza (o arquitectura molecular) del QTL: algunos QTL pueden mostrar los efectos de varios QTNs lo que puede hacer que los resultados sean poco claros y difíciles de interpretar.

Teniendo en cuenta estos factores debemos seleccionar la estrategia de mapeo fino más adecuada a cada situación. La estrategia más sencilla es el aumento de la densidad de

marcadores en la región de interés, pero esta aproximación ha ido perdiendo importancia con la utilización de herramientas moleculares de media densidad de marcadores y la disponibilidad de la secuencia completa de algunos genomas animales. El aumento de la del número de meiosis se obtiene utilizando poblaciones experimentales, como las descritas en el apartado 3.4.1, o aumentando el número de animales o familias en poblaciones comerciales. Además, se puede mejorar la resolución o precisión de mapeo combinando la información proporcionada por el ligamiento con información a nivel poblacional, LDLA. En este caso se suelen considerar haplotipos de diferentes longitudes en vez de marcadores individuales ya que la asociación de aquellos con los alelos del QTL es más estrecha (Georges, 2007). La determinación de los alelos del QTL que portan los animales ayuda a mejorar la detección del QTN. Esto se puede determinar cuando sabemos qué cromosoma del padre (o de la madre) ha heredado un descendiente (o un grupo de descendientes) en concreto, y cuál de los cromosomas determina que los animales que lo portan presenten un valor mayor (alelo *Q*) o menor (alelo *q*) para el carácter en estudio. Una vez definidos los cromosomas *Q* y *q* el test de concordancia (Seroussi, 2009) entre el genotipo de los polimorfismos identificados en la región candidata y el estatus del QTL en los animales estudiados es determinante a la hora de identificar posibles QTNs. Por último, para la disección de caracteres influenciados por un grupo de QTNs se pueden utilizar modelos multi-QTL (Georges, 2007).

Como tercer paso para la disección de un QTL, Glazier et al. (2002) propusieron el análisis de la secuencia dentro del CI. Dada la posible amplitud de dicho CI, la estrategia más utilizada es el estudio de los genes candidatos posicionales que a su vez resultan ser adecuados candidatos funcionales con respecto a la acción del QTL (Zhu y Zhao, 2007). Actualmente, la disponibilidad en muchas especies de la secuencia del genoma completo hace que la identificación de QTNs putativos localizados en el CI pueda ser directa, aunque algunas de las variantes nucleotídicas pueden no haber sido detectadas previamente en la especie estudiada. Además de cada mutación, es necesario conocer cómo la combinación de varias de ellas influyen sobre el fenotipo, ya que este puede ser el resultado de la combinación de más de un QTN (Glazier et al., 2002).

El último paso a seguir en esta estrategia de identificación de QTNs, es el análisis funcional de los QTNs candidatos. La evidencia más concluyente de la causalidad de una mutación es la demostración de que la presencia del QTN putativo modifica el fenotipo (Glazier et al., 2002). Esto se debe a que el estudio de polimorfismos en genes candidatos ha

llevado en algunas ocasiones a la obtención de resultados no replicables y a la detección de variantes alélicas supuestas causantes de fenotipos sin tener en cuenta aspectos funcionales (Georges, 2007). Las pruebas funcionales pueden basarse en la creación de un organismo *knock-out*, en una especie modelo, o en la creación de un transgénico en la especie objeto de estudio. La segunda sería la opción ideal, pero en el caso de especies ganaderas, no suele ser la elegida (Ron y Weller, 2007). Incluso si en el organismo *knock-out* el QTN putativo afecta el carácter cuantitativo, no quiere decir que exista una conexión funcional entre el gen y el QTN. La validación de un QTN se realiza mediante la demostración de que los valores del carácter son diferentes para los distintos alelos o existen diferencias en la funcionalidad de la proteína (Ron y Weller, 2007).

Los estudios funcionales realizados hasta el momento en relación a los QTNs candidatos para distintos efectos de QTL han demostrado que las pruebas funcionales son difíciles de interpretar y que no todos ellos cumplen todas las reglas descritas previamente (Ron y Weller, 2007). Así, por ejemplo, la mutación en el gen *DGAT1*, relacionado con la producción de leche en ganado vacuno (Grisart et al., 2002), no produce una alteración en los niveles de expresión de ARNm, sin embargo, dicha mutación sí modifica la actividad enzimática de la proteína, de acuerdo con el efecto fenotípico observado (Grisart et al., 2004).

En las especies ganaderas, son contados los casos en los que la detección de un QTL por ligamiento con influencia sobre uno o varios caracteres ha llevado a la determinación de un QTN. En el ganado vacuno lechero, se han descrito QTNs en los genes *DGAT1* (*di-acyl glicerol transferase 1*, Grisart et al., 2002), *ABCG2* (*ATP-binding cassette, sub-family G, member 2*, Olsen et al., 2007) y *GHR* (*growth hormone receptor*, Blott et al., 2003); en el ganado vacuno de carne el gen *NCAPG* (*non-SMC condensin I complex, subunit G*) ha mostrado tener efectos sobre el peso al nacimiento y el peso de la canal (Eberlein et al., 2009 y Setoguchi et al., 2009); en cerdos se ha asociado un QTN con el crecimiento muscular en el gen *IGF2* (*insulin-like growth factor 2*, Van Laere et al, 2003) y en el gen *RYR1* (*ryanodine receptor 1*, Stinckens et al., 2007), previamente asociado con la hipertermia maligna (Fujii et al., 1991); en el ganado ovino una mutación en el gen de la *miostatina* (*GDF8*) influye sobre la producción de carne (Clop et al., 2006); en esta especie, además, se han detectado mutaciones en diferentes razas ovinas que influyen sobre la tasa de ovulación en los genes *BMPR-1B* (*bone morphogenetic protein receptor, type IB*), *BMP15* (*bone morphogenetic protein 15*) y *GDF9* (*growth differentiation factor 9*) (Davis, 2004).

El desarrollo de nuevas herramientas moleculares y los estudios de asociación a nivel genómico o GWAS para la disección de caracteres simples, por ejemplo enfermedades mendelianas, han llevado a la determinación de las mutaciones causales en ganado ovino como la epidermólisis bullosa (Mömke et al., 2011), la microftalmia (Becker et al., 2010), el raquitismo (Zhao et al., 2011) o la condrodisplasia (Zhao et al., 2012). En cambio, el estudio de caracteres cuantitativos de interés económico en producción animal con la metodología del GWAS ha llevado a la identificación de un gran número de marcadores, principalmente SNPs, asociados a los caracteres en estudio aunque no a la detección y confirmación de los QTNs de manera directa. En el caso de la producción de leche en la especie bovina, se han estudiado caracteres de interés como la producción de leche (Jiang et al., 2010; Mai et al., 2010; Pryce et al., 2010; Meredith et al., 2012), el contenido de proteína leche (Schopen et al., 2011) o el perfil de ácidos grasos de la leche (Bouwman et al., 2011). En estos GWAS, en muchos casos, se ha llegado a la detección de asociaciones significativas en regiones portadoras de QTNs previamente descritos, como es el caso de las mutaciones en los genes *DGAT1* (Jiang et al., 2010; Mai et al., 2010; Pryce et al., 2010; Meredith et al., 2012), *GHR* y *ABCG2* (Jiang et al., 2010; Pryce et al., 2010).

Utilizando la metodología clásica de detección de QTL con análisis de ligamiento, en la que se empleaba una densidad baja de marcadores en un análisis inicial, el proceso completo desde la detección del QTL hasta la determinación del QTN se extendía a lo largo de, al menos, cuatro años. Sin embargo, los GWAS para caracteres cuantitativos sí han conseguido reducir los CIs de localización de los QTL a unas pocas Mb. Aprovechando las herramientas disponibles, como los *microarrays* de expresión y las secuencias del genoma completo, sería de esperar que el periodo desde la detección de un QTL hasta la determinación del QTN se reduzca sensiblemente (Ron y Weller, 2007).

Otro aspecto a tener en cuenta en la detección de QTNs son las diferencias interespecíficas. Debido a la próxima relación filogenética entre las especies bovina y ovina, y con el fin de validar los QTNs previamente descritos en ganado vacuno en la especia ovina, se han realizado análisis de asociación entre polimorfismos identificados en los genes causales y caracteres de producción de leche en la raza ovina Churra (García-Fernández et al., 2011). Los genes estudiados fueron *DGAT1,* con efecto sobre la producción de grasa y proteína en la leche (Grisart et al., 2002), *ABCG2* y *GHR,* con influencia sobre la producción y la composición de la leche (Olsen et al., 2007; Blott et al., 2003, respectivamente), y *SPP1*

(*osteopontin*), probablemente implicado en el proceso de la lactación (Schnabel et al. 2005). En dicho estudio ninguno de los polimorfismos descritos como causales en ganado vacuno se identificó en la población de raza Churra analizada. La búsqueda de variabilidad adicional en los genes analizados, reveló que ninguno de los polimorfismos detectados mostraba asociación significativa con los caracteres productivos estudiados tras una corrección de Bonferroni, aunque el gen *ABCG2* mostró ciertas asociaciones a nivel nominal con los caracteres MY, PP y FP, asociaciones que debieran ser confirmadas por otros estudios. Estos resultados han puesto de manifiesto que los genes causales en ganado vacuno parecen no tener un papel significativo en el control de los caracteres lecheros en ganado ovino, al menos en la raza Churra. Para explicar esto, los autores del trabajo propusieron dos hipótesis. Por un lado, parece que a pesar de la relación filogenética entre las especies bovina y ovina, la arquitectura genética de los caracteres de producción de leche es diferente en ambas. Alternativamente se sugirió que, probablemente, los procesos de selección que han tenido lugar en ambas especies hayan fijado o seleccionado diferentes mutaciones con efectos sobre los caracteres de producción (García-Fernández et al., 2011). Como sugieren estos autores, las diferencias observadas en el control genético de los caracteres de interés productivo entre ambas especies, hacen que cobren una mayor importancia los esfuerzos realizados por el ISGC para el desarrollo de herramientas genómicas en la especie ovina y que los avances no sólo se basen en las herramientas desarrolladas para la especie bovina.

5.1. Uso de la información genómica en los programas de selección

Los programas tradicionales de mejora genética, descritos en el apartado 2, se basan en la utilización de pedigrís y fenotipos. A esta información podemos añadir la contenida en el ADN del animal, ya sea de manera directa, en base a los QTNs detectados, o selección indirecta a través de marcadores en LD con el QTN. El primer caso se refiere a la Selección Asistida por Genes (GAS, *Gene Assisted Selection*) y el segundo a la Selección Asistida por Marcadores (MAS, *Marker Assisted Selection*). La selección de reproductores basada en MAS se realiza, en general, en dos pasos: (i) se estiman los efectos del marcador en la posición del QTL en una población de referencia; (ii) se calculan los valores genéticos de los animales candidatos a ser reproductores en función de la información molecular, proporcionada por los marcadores. En cambio la GAS se basa en la selección de los alelos del QTN con efecto favorable sobre el fenotipo.

Ambas metodologías de selección están siendo utilizadas en las especies ganaderas, siendo la GAS la preferida debido a su facilidad de aplicación y a que muestra ganancias genéticas de manera más directa (Dekkers, 2004). En el ganado ovino, los programas de mejora genética se basan, en general, en métodos tradicionales, como los descritos en el apartado 2.2, aunque existen ejemplos de MAS y GAS. Los caracteres objeto de selección para los que existen tests comerciales disponibles en ganado ovino son la resistencia a enfermedades, como el *scrapie* o inflamaciones de las pezuñas (*foot rot*), caracteres reproductivos, como los genotipos *Inverdale* o *Booroola*, y caracteres de producción cárnica como el fenotipo *Carwell* o la doble musculatura en Texel (Van der Werf, 2007). Únicamente la selección a favor de los alelos más resistentes al *scrapie* ha sido incluida de manera sistemática en los programas de selección en ganado ovino lechero (Carta et al., 2009).

La selección genómica se presenta como un paso más en la utilización de la información molecular en los esquemas de selección, ya que pretende incluir todos los efectos detectados sobre los caracteres productivos a lo largo del genoma en el mérito genético del animal, a través del estudio de marcadores en LD con los QTNs, sin necesidad de conocer las mutaciones causales. A pesar de que este método de selección se está implementando de manera satisfactoria en especies como el vacuno lechero (Hayes et al., 2009), el descubrimiento de los QTNs subyacentes puede ayudar a evitar algunas de las limitaciones del mismo tales como la necesidad de genotipar un elevado número de animales y de repetir el genotipado tras cierto número de generaciones (Taylor, 2012). Además, la detección de las mutaciones causales y su directa aplicación en los programas de mejora puede tener especial interés en el ganado ovino, ya que el limitado tamaño de las poblaciones objeto de selección dificulta el establecimiento de una población de referencia sobre la que estimar los valores genéticos.

RESULTADOS

García-Gámez E., Gutiérrez-Gil, B., García Fernandez, M., Sánchez, J.P., de la Fuente, L.F. San Primitivo, F., Arranz J.J. (2009). Confirmación de un QTL con efecto sobre caracteres de producción de leche en la raza ovina Churra. *XIII Jornadas sobre Producción Animal AIDA*. M. Joy, J. H. Calvo, C. Calvete, M. A. Latorre, I. Casasús, A. Bernués, B. Panea, A. Sanz, J. Balcells (Eds.). AIDA, Zaragoza.

García-Gámez E., Gutiérrez-Gil B., Sánchez J.P., Arranz J.J. (2012). Short Communication: Replication and refinement of a QTL influencing milk protein percentage on ovine chromosome 3. *Animal Genetics* 43(5), 636-641.

García-Gámez E., Gutiérrez-Gil B., Sánchez J.P., Bayón Y., De la Fuente L.F, San Primitivo F., Arranz J.J. (2012). Evaluación de la correspondencia entre los mapas genético (macho) y físico en el ganado ovino de la raza Churra. *XVI Reunión Nacional de Mejora Genética Animal.* 31 de mayo - 2 de junio 2012, Ciutadella de Menorca (España).

García-Gámez E., Sahana G., Gutiérrez-Gil B., Arranz J.J. (2012). Linkage disequilibrium and inbreeding estimation in Spanish Churra sheep. *BMC Genetics* 13:43.

García-Gámez E., Gutiérrez-Gil B., Sahana G., Sánchez J.P., Bayón Y., Arranz J.J. (2012). GWA analysis for milk production traits in dairy sheep and genetic support for a QTN influencing milk protein percentage in the *LALBA* gene. *PLoS ONE, In press,* doi: 10.1371/journal.pone.0047782.

CONFIRMACIÓN DE UN QTL CON EFECTO SOBRE CARACTERES DE PRODUCCION DE LECHE EN LA RAZA OVINA CHURRA

García-Gámez, E., Gutiérrez-Gil, B., García Fernandez, M., Sánchez, J.P., de la Fuente, L.F. San Primitivo, F. y Arranz J.J.

Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, 24071 León. E-mail: egarg@unileon.es

XIII Jornadas sobre Producción Animal AIDA. M. Joy, J. H. Calvo, C. Calvete, M. A. Latorre, I. Casasús, A. Bernués, B. Panea, A. Sanz, J. Balcells (Eds.). AIDA, Zaragoza.

INTRODUCCIÓN

Tradicionalmente, la producción de leche de ganado ovino se concentra en los países mediterráneos y va unida a la explotación de razas locales para la producción de queso de alta calidad. La mejora y el mantenimiento de estos sistemas de explotación son importantes, además de por una significación histórica y cultural, para el mantenimiento de estas poblaciones en áreas más desfavorecidas y la conservación de los ecosistemas. El principal factor que contribuye a este mantenimiento es la explotación de esas razas locales rentables. Para ello y desde hace años se han establecido programas de selección con objeto de mejorar sus producciones. Además, como complemento de los anteriores, se han realizado estudio con herramientas moleculares que intentan detectar genes con influencia sobre diferentes caracteres productivos.

Los primeras investigaciones de QTL en la raza Churra fueron hechos por nuestro grupo de investigación analizando cromosomas candidatos (Díez-Tascón et al., 2001).

Posteriormente se llevó a cabo un barrido genómico en busca de regiones que portaran QTL para diferentes caracteres: morfología mamaria y recuento de células somáticas como parámetro relacionado con la resistencia a la mamitis (Gutiérrez-Gil et al., 2007; Gutiérrez-Gil et al., 2008) y con caracteres de producción de leche (Gutiérrez-Gil et al., 2009). Con objeto de detectar la verdadera naturaleza de estos QTL, es necesaria la confirmación de estos efectos. Una de las formas de confirmar estos QTL es mediante el análisis de una muestra independiente de hijas de las familias que han mostrado evidencia de segregación, así como el análisis de nuevas familias. En el presente artículo presentamos el análisis de uno de los QTL detectado y la confirmación en una muestra independiente de familias analizadas.

MATERIAL Y MÉTODOS

En este trabajo se han utilizado dos conjuntos de datos, por un lado los producidos en el *genome scan* por Gutiérrez-Gil et al. (2007) y por otro las nuevas familias muestreadas para la confirmación. Los nuevos animales son un total de 841 medio hermanas distribuidas en 15 familias, pertenecientes a 16 rebaños del núcleo de selección de ANCHE. El tamaño medio de las familias es de 83 ovejas, variando entre 27 y 260 animales.

Para el estudio del cromosoma 20 se eligieron marcadores microsatélites utilizados en la construcción de un mapa de ligamiento en la población de raza Churra (Gutiérrez-Gil et al., 2008): BM1248, BM1905, BP34, DYA, INRA132, MCMA23 y OLADRB. Tras la obtención del ADN, se amplificaron los marcadores mediante PCR y se obtuvieron las variantes alélicas a partir de la electroforesis en un secuenciador *ABI3130 Genetic Analyzer*. La identificación alélica se llevó a cabo utilizando el software GeneMapper 4.0.

La elaboración del mapa de ligamiento se realizó mediante programa CRIMAP (Green et al., 1990) empleando las rutinas *build* y *chrompic* para la construcción del mapa y el control de múltiples recombinantes, respectivamente. Los caracteres utilizados en la detección de QTL han sido: cantidad de leche (MY), porcentaje de proteína (PP) y porcentaje de grasa (FP). Los caracteres empleados ha sido las "*yield deviations*" obtenidas al corregir los valores brutos para los factores fijos que mostraron significación. Posteriormente, el análisis para la confirmación del QTL se llevó a cabo con el programa GridQTL (Seaton et al., 2006) que utiliza un método de regresión con múltiples marcadores desarrollado por Knott et al. (1996). Los valores con significación chromosome-wise fueron determinados mediante 10.000 permutaciones.

RESULTADOS Y DISCUSIÓN

La figura 1 muestra el mapa obtenido tras el análisis de ligamiento en cada cM del cromosoma 20 ovino, en concordancia con el mapa previamente publicado (Maddox et al., 2001; Gutiérrez-Gil et al., 2008). Asimismo, se representan los valores del parámetro contenido de información (CI) a lo largo del mapa. La longitud del mapa fue de 89 cM (Kosambi). El contenido de información promedio fue del 70%, con un valor máximo de 88,52% en la posición del marcador DYA, y un mínimo de 49,49% en el intervalo [BM1905-MCMA23].

Además, se muestran en esta figura los perfiles para el test estadístico obtenidos en el análisis global de la población, para cada uno de los caracteres considerados en este trabajo. Se indican en ella los niveles de significación 5% y 10% *chromosome-wise* para el carácter porcentaje de grasa.

De acuerdo con este análisis aparece un QTL en este cromosoma para el carácter porcentaje de grasa, con valor máximo del estadístico en la posición 61 cM, con *p-value*

asociado de 0,0309 (*chromosome-wise*). Tras el análisis intrafamiliar (*within-family*), se obtuvieron un total de seis familias segregantes. De este modo, las nuevas familias añadidas confirmaron los resultados obtenidos a cerca de la presencia de un QTL para el parámetro porcentaje de grasa (FP).

En un estudio previo, Barillet et al. (2006) identificaron, en una población Sarda X Lacaune, este mismo QTL en la región del sistema mayor de histocompatibilidad (MHC), coincidente con la posición de nuestro QTL en el análisis "*across-family*". Además, un posible gen candidato, tanto posicional como funcional, como es la prolactina (PRL) se encuentra entre los marcadores OLADRB y BP34, así como otras proteínas relacionadas con la prolactina (PRLP1, PRLP3, PRLP4).

El siguiente paso sería un estudio de mapeo fino para comprender la arquitectura genética de esa región, con el fin de poner de manifiesto la mutación causante del efecto detectado. De todas formas, la localización cercana al efecto detectado del gen de la prolactina, señala a éste como un candidato idóneo, y posible responsable del efecto detectado. La detección de SNPs en este gen y la posible asociación mediante técnicas de desequilibrio de ligamiento, pueden ayudar a clarificar de forma más precisa la naturaleza de esta asociación y el papel de dicho gen en la misma.

AGRADECIMIENTOS

Este trabajo ha sido cofinanciado por el proyecto AGL2005-04321 del Ministerio de Ciencia e Innovación (MICINN) y del Proyecto GR43 de la Junta de Castilla y León para grupos de excelencia. Elsa García Gámez disfruta de un contrato para Investigadores jóvenes de la Junta de Castilla y León cofinanciado por el Fondo Social Europeo. Marta García Fernández es becaria FPI (MICINN). Beatriz Gutiérrez Gil disfruta de un Contrato del Programa Juan de la Cierva (MICINN).

REFERENCIAS BIBLIOGRÁFICAS

[1] Díez-Tascón, C., Bayón, Y., Arranz, J. J., De La Fuente, L. F., & San Primitivo F. 2001. J Dairy Res 68, 389-397.

[2] Green, P., Falls, K., & Crooks, S., 1990. Doc. for CRIMAP.

[3] Gutiérrez-Gil, B., Arranz, J. J., El-Zarei, M.F., Álvarez, L., Pedrosa, S., San Primitivo, F. & Bayón Y. 2008. J Anim Breed Genet 125, 201-204.

[4] Gutiérrez-Gil, B., El-Zarei, M.F., Alvarez, L., Bayón, Y., de la Fuente, L.F., San Primitivo, F. & Arranz, J.J., 2008. J Dairy Sci. 91(9):3672-81.

[5] Gutiérrez-Gil, B., El-Zarei, M.F., Alvarez, L., Bayón, Y., de la Fuente, L.F., San Primitivo, F. & Arranz J.J., 2009. Anim Genet. In press.

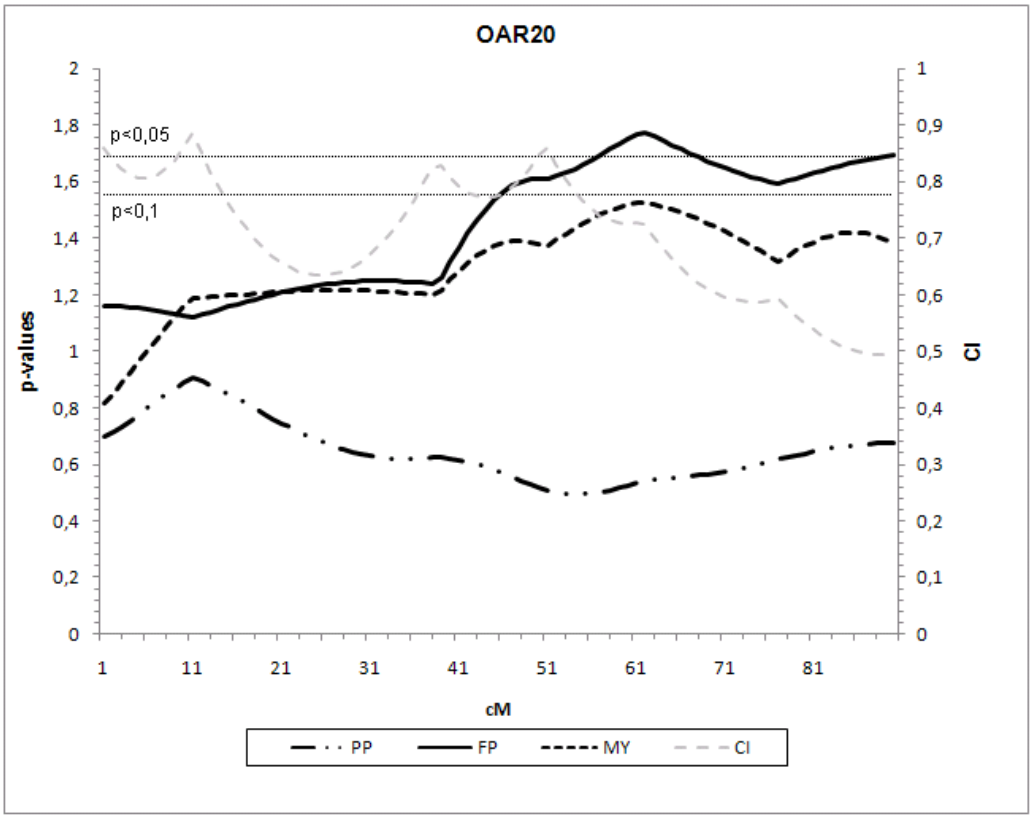
[6] Gutiérrez-Gil, B., El-Zarei, M. F., Bayón, Y., de la Fuente, L. F., San Primitivo, F. & Arranz J. J. 2007. J Dairy Sci 90, 422-426.

[7] Knott, S. A., J. M. Essen, & C. S. Haley, 1996. Theoret Appl Genet. 93: 71-80.

[8] Seaton, G., Hernandez, J., Grunchec, J.A., White, I., Allen, J., De Koning, D.J., Wei, W., Berry, D., Haley, C. & Knott, S. (2006) Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, August 13-18, 2006. Belo Horizonte, Brazil.

FIGURAS

Figura 1. Distribución de los valores de contenido de información (IC) y de significación del test estadístico expresado como $\log_{10}$ (1/P) para los caracteres de producción de leche (MY) Porcentaje de Proteína (PP) y porcentaje de grasa (FP) a lo largo del cromosoma OAR20 (análisis across-families).



CONFIRMATION OF QTL UNDERLYING DAIRY PRODUCTION TRAITS IN SPANISH CHURRA SHEEP

ABSTRACT: After a preliminary genome scan carried out in Churra sheep to detect QTL influencing milk production traits, additional families are now been analyzed to confirm some of the identified effects. We present here the confirmation of the QTL located in chromosome 20 for fat percentage. This meta-analysis was performed adding 15 new families to the ones previously analyzed. The total number of animals was 2.054, distributed in 25 families, one of them common in both studies. Phenotypic measurements studied included milk yield, protein percentage and fat percentage (FP). All the population was genotyped for seven microsatellite markers evenly distributed across the studied chromosome. Response variables used in the QTL analysis were yield deviations, estimated from the phenotypic data corrected for fixed environmental effects. For the QTL analysis, a multimarker regression method was implemented through the GridQTL software. Chromosome-wise critical values were calculated through 10.000 permutations. The average information content across the chromosome was 0.70. An across-family association analysis confirmed a region on this chromosome carrying the QTL for FP at the 5% chromosome-wise level (p-value = 0.0309). The within-family analysis revealed 6 families segregating for the QTL. This confirmation is the previous stage of a fine-mapping approach of this QTL.

Keywords: sheep, milk, QTL, fat percentage.

SHORT COMMUNICATION: REPLICATION AND REFINEMENT OF A QTL INFLUENCING MILK PROTEIN PERCENTAGE ON OVINE CHROMOSOME 3.

Running title: QTL replication for milk protein content on OAR3

García-Gámez E., Gutiérrez-Gil B., Sánchez J.P., Arranz J.J.

Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, 24071, León, Spain.

Animal Genetics 2012, 43(5): 636-641.

SUMMARY

A previous genome scan that was conducted in Spanish Churra sheep identified a significant quantitative trait locus (QTL) for milk protein percentage (PP) on chromosome 3 (OAR3) between markers KD103 and OARVH34. The aim of the present study was to replicate the results and refine the mapped position of this QTL. To do this, we genotyped 14 additional half-sib families of Spanish Churra sheep, including 1,661 ewes from 29 different flocks. This new population was genotyped for a total of 21 microsatellite markers mapping to OAR3. In addition to a classical linkage analysis (LA), a combined linkage disequilibrium and linkage analysis (LDLA) was performed with the aim of enhancing the resolution of QTL mapping. The LA performed in this population replicated the presence of a highly significant QTL for PP that was close to marker KD103 ($Pc < 0.001$; $Pexp < 0.001$). The phenotypic variance explained by the QTL was 2.74 %. Two segregating families for the target QTL were identified in this population with QTL effect estimates of 0.47 and 0.95 SD. The LDLA identified the same QTL with high statistical significance ($P = 9.184E-11$) and narrowed the CI to a 13 cM-long region. These results support the identification of the previously described OAR3 QTL that influences PP. Future research will be aimed at increasing the marker density across the refined CI and analysing candidate genes to identify the allelic variant or variants that underlie the studied genetic effect.

Keywords: dairy sheep, milk protein percentage, quantitative trait locus, linkage analysis, linkage disequilibrium.

Several QTLs have been identified using a whole genome scan approach in a commercial population of Spanish Churra sheep for milk traits in dairy sheep (Gutiérrez-Gil et al. 2009). A genome-wise significant QTL influencing milk protein percentage (PP) mapping to sheep chromosome 3 (OAR3) was identified in Churra sheep through a genome scan linkage analysis (LA) approach. The segregating population comprised 1,421 animals belonging to 11 half-sib families (Population A). Based on the linkage map that was constructed for this chromosome, which involved 11 microsatellite markers, the QTL peak was found between markers KD103 and OARVH34, and the estimated bootstrapping 95% confidence interval (CI) had a length of 40 cM (Gutiérrez-Gil et al. 2009).

The aim of the present study was to confirm and refine the mapped position of the QTL identified in OAR3 that has an effect on PP. Because we have used animals from the same population and the same study design, and following Igl et al. (2009), the term "replication" will be used to refer to the first of these objectives.

Hence, to replicate the target QTL, 14 additional half-sib families of Spanish Churra sheep, including a total of 1,661 animals from 29 different flocks (which will hereafter be referred to as Population B), were genotyped for a total of 21 microsatellite markers located on OAR3. Due to DNA degradation Population A could not be genotyped for the new markers selected. Based on the increased marker density that was reached in Population B (average marker distance between markers = 13.7 cM) and in addition to a classical LA, we performed a combined linkage disequilibrium and linkage analysis (LDLA) to narrow the CI of the QTL.

All the animals included in this work belonged to the Selection Nucleus of the National Association of Churra Breeders (ANCHE) and were bred by artificial insemination. The average family size in Population B was 119, ranging from 34 to 291 daughters per sire. It should be noted that some of the sires of Population B were related to the sires in the previously analysed Population A (Gutiérrez-Gil et al. 2009). The traits that were considered in the present work included the four milk production traits, milk yield (MY), milk protein percentage (PP), milk fat percentage (FP), and somatic cell score (SCS), that are routinely recorded in this breed through the official milk recording system.

Together with the markers that were previously analysed in Population A, additional markers were selected for genotyping in Population B to increase the marker density across the full-length chromosome, with a special focus on the region harbouring the targeted QTL. This selection was based on the Australian sheep linkage map v4.7 (http://rubens.its.unimelb.edu.au/~jillm/jill.htm). Animals were genotyped at the microsatellite loci using multiplexing and analysis of multi-loading combinations using GeneMapper® software (Applied Biosystems). After discarding markers BMS1350 and ILSTS042 due to difficulties in electropherogram interpretation, 21 markers were chosen for analysis in Population B (see Supplementary Table S1). Linkage maps were generated by multipoint linkage analysis using the BUILD option of CRIMAP (v. 2.4) (Green et al. 1990).

A classical regression LA was performed for Population B using the web-based GridQTL software (Seaton et al. 2006). The dependent variables used for this analysis were the yield deviations (YDs) of the traits under study, which were estimated as described by García-Férnandez et al. (2010). The information content (IC) along the marker map, the chromosome-wide significance thresholds and values (Pc-values) and the 95% confidence interval (95% CI) were calculated as previously described by García-Férnandez et al. (2010). Experiment-wide significance values (Pexp-values) were estimated using a Bonferroni correction that considered three independent traits that were predicted with a principal component analysis (R Development Core Team, 2008). The fraction of the phenotypic variance explained by the QTL was defined as the percent reduction in the residual variance that was due to the inclusion of the QTL in the model (adapted from Knott et al. 1996). The within-family analysis revealed which families were segregating for the QTL identified at the population level (those with $Pc < 0.05$). The presence of a second QTL was investigated using the two-QTL analysis option in GridQTL. The two-QTL model was accepted if there was a significant improvement over the best 1-QTL model at $Pc < 0.05$. This analysis was also performed in one of the segregating families because of the bimodal shape of the profile of the test statistic.

To refine the mapped position of the QTL identified for PP in the previous analysis, Population B was analysed for this trait using a combined LDLA procedure (Meuwissen & Goddard 2000) implemented in the corresponding module of the GridQTL platform (Hernández-Sánchez et al. 2009). The LDLA scan was conducted along a 49.7 cM interval (including 9 microsatellites) that was centred on the maximum LOD value observed in the LA for PP. In this scan, the analysis was repeated for each cM position by using the genotypes of the two flanking markers of the position tested. Haplotypes that were required for the analysis were calculated using the HapSim software (Montana 2005). Following Hernández-Sánchez et al. (2010) different values for the LDLA analysis parameters were used to model historical relationships. For many of these combinations, the analysis failed to converge on several points, and, therefore, the results described herein represent those with the highest level of convergence and which correspond to the parameters estimated for this experiment. Therefore, the effective sample size used in the reported LDLA was $Ne = 32$, whereas the number of discrete generations since population foundation until pedigree records began was $T = 7$. This was calculated taking into account 4 years as generation interval and considering

Churra selection scheme to begin 30 years ago. IBD probabilities were calculated using segregating information at the nearest pair of markers that bracketed each position tested. Following the methodology of Beraldi et al. (2007), Likelihood ratio (LR) test results were converted to the LOD scale. The CI for each QTL was determined by the LOD drop-off method (Lander & Botstein 1989). For the LDLA, we assumed the same significance thresholds as those used in the genome-wide association analyses in human pedigrees (LOD 3.3 ~ P < 0.0001; Lander & Kruglyak 1995).

The linkage map that was built for OAR3 based on the 21 microsatellite markers analysed in Population B (see Supplementary Table S1 for details) showed marker order and distances that were in agreement with the Australian linkage map v4.7 (http://rubens.its.unimelb.edu.au/~jillm/jill.htm) and that previously published for Churra sheep (Gutiérrez-Gil et al. 2008). The average information content (IC) across the studied chromosome was 0.69, which ranged from 0.53 at position 120 cM to 0.89 at position 191 cM (Figure 1a).

The across-family regression analysis revealed a highly significant QTL for PP (Pc = 0.0002; Pexp = 0.0006) with none of the other traits analysed showing significant results (Figure 1a). Detailed information for the QTL identified is given in Table 1. The maximum across-family F-value for this QTL was located at position 175 cM between markers CABB11 and KD103. The phenotypic variance explained by the QTL was 2.74 %. Despite evident grouping of bootstrap estimates around position 175 cM, the bootstrapping analysis estimated a large 95% CI for this QTL, which spanned from positions 109 to 281 cM. The individual within-family analyses revealed two segregating families (families 7 and 14) (Table 1). The estimated allele substitution effects for these two sire families (0.42 and 0.95 phenotypic SD) were slightly higher than those reported by Gutiérrez-Gil et al. (2009), which were approximately 0.3 SD. The statistical profile of family 14 showed a second peak at 258 cM which, according to a two-QTL analysis performed individually for this family, indicated the presence of a second chromosome-wise significant QTL for this half-sib group (Pc = 0.0075 for the two-QTL vs. one-QTL test). The two-QTL analysis performed at the whole population level was not significant (P = 0.193 for the two- vs. one-QTL test).

The LDLA scan performed for the PP trait was centred at position 175 cM, where the maximum F-value had been obtained in the LA and involved the interval flanked by markers

D469297 and OARVH34. The average distance between markers within this interval was 6.2 cM. The LDLA showed a plateau from positions 171 to 182 cM with a maximum LOD score at 178 cM (P = 9.184E-11), which is very close to marker KD103 (Figure 1b). The corresponding one-LOD drop-off CI was estimated to span a 13 cM-long interval (from positions 170 and 183 cM) flanked by markers CSAP17E and BL4. According to the Sheep Genome (v. 2.0), this interval includes the region between 133 Mbp and 148.5 Mbp of OAR3.

The first genome scan conducted in a commercial population of dairy sheep by our research group identified a genome-wide significant QTL for PP on OAR3 (Gutiérrez-Gil et al. 2009). Other QTLs for dairy traits in sheep have been reported on the same chromosome. A QTL for milk protein yield (PY) and PP was found in the first half of OAR3 (Raadsma et al. 2009; Singh et al. 2007). In addition, segregating QTLs for MY, milk fat yield (FY) and PY were identified in a Sarda x Lacaune backcross population and were close to marker BMC1009 (Barillet et al. 2006). According to the Australian sheep linkage map (http://rubens.its.unimelb.edu.au/~jillm/jill.htm) this marker maps close to marker CABB11 and is included within the 13 cM-long CI estimated by the LDLA described here. Further research would be required to assess the possible relationship of these QTL influencing yield traits and the one targeted here. In dairy cattle, Bennewitz et al. (2003) have reported a QTL for PP that maps to the orthologous region of our target QTL, on bovine chromosome 5 (BTA5).

The present study summarises the initial efforts to replicate and refine the most significant QTL detected in a previous genome scan in Churra sheep (Gutiérrez-Gil et al. 2009). The LA performed on Population B replicated the presence of a highly significant PP QTL close to marker KD103, with two out of the 14 newly analysed families segregating for this effect. The 95% CI that was estimated from this analysis involved an interval that was substantially larger (172 cM) than that estimated in the first report of this QTL (40 cM; Gutiérrez-Gil et al. 2009). This CI was also about the double of the resolving power that would be expected according to Darvasi & Soller´s (1997) formulae: 530/Nv where N = segregating population size (2 families x 119 average offspring here); v = the proportion of the variance explained by the QTL; 0.027 here).The identification of a second significant QTL for PP at the distal end of the chromosome for family 14 may be a plausible explanation for the long CI estimated in this analysis.

The joint analysis of different mapping experiments has been suggested to be a simple strategy for QTL confirmation and replication and to increase the experimental power of the independent experiments (Bennewitz et al. 2003). Because of the difficulties to analyse population A for the additional microsatellites analysed in population B, we performed a joint LA based on the nine markers that were commonly genotyped in these two populations (data not shown). The high level of statistical significance that the targeted QTL showed in this analysis ($Pc < 0.0001$), with 192 segregating families, may exemplify some of the potential benefits indicated by Bennewitz et al. (2003) and supports the replication of the QTL based on the previously described LA. The bootstrapping 95% CI estimated for this analysis spanned 46 cM, which according to the previous referred formulae, would be similar to the expected resolving power of the joint analysis design.

The LDLA performed on Population B allowed the rejection of the null hypothesis with much higher probability than that calculated in the initial LA and demonstrated a narrowing of the CI to a 13 cM-long region. As previously shown by other authors LDLA can potentially enhance the resolution of QTL mapping (Meuwissen et al. 2002; Hernandez-Sanchez et al. 2010)because LDLA takes into account that current pedigree founders are related among themselves relative to the ancestral generation of population founders and therefore using information of more generations. Although ideally LDLA should use a more saturated marker map than LA, Hernandez-Sánchez et al. (2010) showed that this is not a prerequisite.

To search for candidate genes in the refined CI, we used Ovine Assembly v. 2.0 (http://www.livestockgenomics.csiro.au/sheep/oar2.0.php). In addition to the candidate genes already suggested by Gutiérrez-Gil et al. (2009) (HDAC7, VDR and ENDOU), other positional candidates, such as IGFBP6 (insulin-like growth factor binding protein), AQP5 (aquaporin), ARF3 (ADP-ribosylation factor 3), LALBA (alpha-lactoalbumin) and IRAK4 (interleukin-1 receptor-associated kinase 4), are also included within the 133-148.5 Mbp of OAR3. Among these, LALBA appears to be a strong functional candidate for the QTL studied here. The protein encoded by this gene, α-LA, is a major whey protein found in milk that participates in lactose synthesis by modifying the substrate specificity of galactosyl-transferase (Ebner and Brodbeck, 1968). LALBA was studied in the 1990s as a potential tool for marker assisted selection in dairy species. In cattle, a single nucleotide polymorphism in the promoter of this gene has been reported to influence milk traits (Bleck & Bremel, 1993;

Schopen et al. 2011; Lundén & Lindersson, 1998). In dairy sheep a polymorphism has been described within the LALBA gene (Chiofalo & Micari, 1987), but to our knowledge there are not reporte studies on its influence on milk traits.

As a whole, the results presented here support the finding of a QTL on OAR3 that influences PP, which was originally reported by Gutiérrez-Gil et al. (2009). The initial replication of this QTL in an independent sample from the same population is an indispensable step before the results can be confirmed in other populations. This is especially true when the resource population is a commercial population, that can be used to refine the QTL position and attempt QTN identification. Because the trait influenced by our target QTL has a direct influence on cheese yield, the identification of the causal gene of the target QTL could be of great interest for the dairy sheep industry. However further research including high density SNPs screening would aim in the identification of the QTN or markers in strong LD that could have a practical implementation in selections programs.

ACKNOWLEDGEMENTS

This work was supported by a Marie Curie fellowship Reintegration grant from the European Commission (SHEEPMILKGENES: FP7-MC-ERG-224857) and the Spanish Ministry of Science (ProjectAGL2009-07000). Elsa García-Gámez is funded by a FPU contract from the Spanish Ministry of Education. Beatriz Gutiérrez-Gil is funded by the Juan de la Cierva Program of the Spanish Ministry of Science.

REFERENCES

[1] Barillet F., Arranz J.J., Carta A., Jacquiet P., Stear M. & Bishop S. (2006) Final Consolidated Report of the European Union Contract of Acronym genesheepsafety, QTLK5-CT-2000-00656, p. 145.

[2] Bennewitz J., Reinsch N., Grohs C., Levéziel H., Malafosse A., Thomsen H., Xu N., Looft C., Kühn C., Brockmann G.A., Schwerin M., Weimann C., Hiendleder S., Erhardt G., Medjugorac I., Russ I., Förster M., Brenig B., Reinhardt F., Reents R., Averdunk G., Blümel J., Boichard D. & Kalm E. (2003) Combined analysis of data from two granddaughter designs: A simple strategy for QTL confirmation and increasing experimental power in dairy cattle. Genetics Selection Evolution 35, 319-338.

[3] Beraldi D., McRae A.F., Gratten J., Slate J., Visscher P.M. & Pemberton J.M. (2007) Mapping quantitative trait loci underlying fitness-related traits in a free-living sheep population. Evolution 61, 1403-1416.

[4] Chiofalo L. & Micari P. (1987) Present knowledge of the variation of the milk proteins in the sheep population reared in Sicily. Experimental observations. Scienze e Tecnologie Lattiero Casearie 38, 104-114.

[5] Churchill G.A. & Doerge R.W. (1994) Empirical threshold values for quantitative trait mapping. Genetics 138, 963-971.

[6] Darvasi A. & Soller M. (1997) A simple method to calculate resolving power and confidence interval of QTL map location. Behavior genetics 27, 125-132.

[7] Ebner K. & Brodbeck U. (1968) Biological role of alpha-lactalbumin: a review. Journal of Dairy Science 51, 317-322.

[8] García-Fernández M., Gutiérrez-Gil B., García-Gámez E., Sánchez J.P. & Arranz J.J. (2010) Short communication: The identification of QTL that affect the fatty acid composition of milk on sheep chromosome 11. Animal Genetics 41, 324-328.

[9] Green P., Falls K. & Crooks S. (1990) Documentation for CRI-MAP, version 2.4. Washington University School of Medicine, St Louis.

[10] Gutiérrez-Gil B., Arranz J.J., El-Zarei M.F., Álvarez L., Pedrosa S., San Primitivo F. & Bayón Y. (2008) A male linkage map constructed for QTL mapping in Spanish Churra Sheep. Journal of Animal Breeding Genetics 125, 201-204.

[11] Gutiérrez-Gil B., El-Zarei M.F., Álvarez L., Bayón Y., De La Fuente L.F., San Primitivo F. & Arranz J.J. (2009) Quantitative trait loci underlying milk production traits in sheep. Animal Genetics 40, 423-434.

[12] Hernández-Sánchez J., Grunchec J. A. & Knott S. (2009) A web application to perform linkage disequilibrium and linkage analyses on a computational grid. Bioinformatics 25, 1377-1383.

[13] Hernández-Sánchez J., Chatzipli A., Beraldi D., Gratten J., Pilkington J.G. & Pemberton J.M. (2010) Mapping quantitative trait loci in a wild population using linkage and linkage disequilibrium analyses. Genetics Research 92, 273-281.

[14] Igl B.W., König I.R. & Ziegler A. (2009) What Do We Mean by 'Replication' and 'Validation' in Genome-Wide Association Studies? Human Heredity 67: 66-68

[15] Knott S.A., Essen J.M. & Haley C.S. (1996) Methods for multiple-marker mapping of quantitative trait loci in half-sib population. Theoretical and Applied Genetics 93, 71-80.

[16] Knott S.A., Marklund L., Haley C.S., Andersson K., Davies W., Ellegren H., Fredholm M., Hansson I., Hoyheim B., Lundström K., Moller M. & Andersson L. (1998) Multiple marker mapping of quantitative trait loci in across between out bred wild boar and large white pigs. Genetics 149, 1069-1080.

[17] Lander E.S. & Botstein D. (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121, 185-199.

[18] Lander E. & Kruglyak L. (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nature Genetics 11, 241-247.

[19] Lundén A. & Lindersson M. (1998) α-Lactalbumin polymorphism in relation to milk lactose. Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, Armidale, NSW, Australia.

[20] Meuwissen T.H. & Goddard M.E. (2000). Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. Genetics 155, 421-430.

[21] Meuwissen T.H., Karlsen A., Lien S., Olsaker I. & Goddard M.E. (2002) Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. Genetics 161, 373-379.

[22] Montana G. (2005) HapSim: a simulation tool for generating haplotype data with pre-specified allele frequencies and LD coefficients. Bioinformatics 21, 4309-4311.

[23] Raadsma H.W., Jonas E., McGill D., Hobbs M., Lam M.K. & Thomson P.C. (2009) Mapping quantitative trait loci (QTL) in sheep. II. Meta-assembly and identification of novel QTL for milk production traits in sheep. Genetics Selection Evolution 41, 45.

[24] R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

[25] Schopen G.C., Visker M.H., Koks P.D., Mullaart E., van Arendonk J.A. & Bovenhuis H. (2011) Whole-genome association study for milk protein composition in dairy cattle. Journal of Dairy Science 94, 3148-3158.

[26] Seaton G., Hernandez J., Grunchec J.A., White I., Allen J., De Koning D.J., Wei W., Berry D., Haley C. & Knott S. (2006) GridQTL: A Grid Portal for QTL Mapping of

Compute Intensive Datasets. Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Brazil.

[27] Singh M., Lam M., McGill D., Thomson P.C., Cavanagh J.A., Zenger K.R. & Raadsma H.W. (2007) High resolution mapping of quantitative trait loci on ovine chromosome 3 and 20 affecting protein yield and lactation persistency. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 17, 565-568.

[28] Visscher P.M., Thompson R. & Haley C.S. (1996) Confidence intervals in QTL mapping by bootstrapping. Genetics 143, 1013-1020.

TABLES

Table 1 Results from the regression linkage analysis that was performed on sheep chromosome 3 in the Churra population. . An across-family analysis of four traits identified a significant QTL for milk protein percentage (PP). The results of the within-family analysis for this trait are given for the two families identified as segregating.

ACROSS-FAMILY ANALYSIS					**WITHIN-FAMILY ANALYSIS**					
Trait	**Position[1] [Confidence[2] interval]**	**Flanking markers[3]**	**p_c-value[4]**	**% phenotypic variance[5]**	**Segregating families**	**Family size**	**Position[6] [Confidence[2,7] interval]**	**Flanking markers[8]**	**P_c[9]**	**Size effect[10], trait units (SD units)**
Protein Percentage	175 [109-281]	[**CABB11**-KD103]	0.0002	2.74	7	326	168 [61.5-219.5]	[**AGLA293** - CSAP17E]	0.003	- 0.182 ± 0.049 (0.47)
					14	82	174 [99-286]	[CSAP17E - **CABB11**]	0.029	- 0.369 ± 0.122 (0.95)

[1,6] Position (Kosambi cM) of the chromosome where the maximum F-statistic value was obtained in the across- and within-family analyses, respectively.

[2,7] For the protein percentage QTL identified at the across- and within-family levels. The 95% confidence interval, which was obtained by bootstrapping analysis (Visscher *et al.* 1996), is shown in square brackets (Kosambi cM).

[3, 8] Markers flanking the position of the maximum F-statistic in the across- and within-family analyses. Markers in bold caps are <1 cM from the maximum F-statistic.

[4, 9] Chromosome-wide *P*-values associated with the maximum F-statistic values of the across- and within-family analyses, which were calculated with a 10,000 permutation test (Churchill & Doerge 1994).

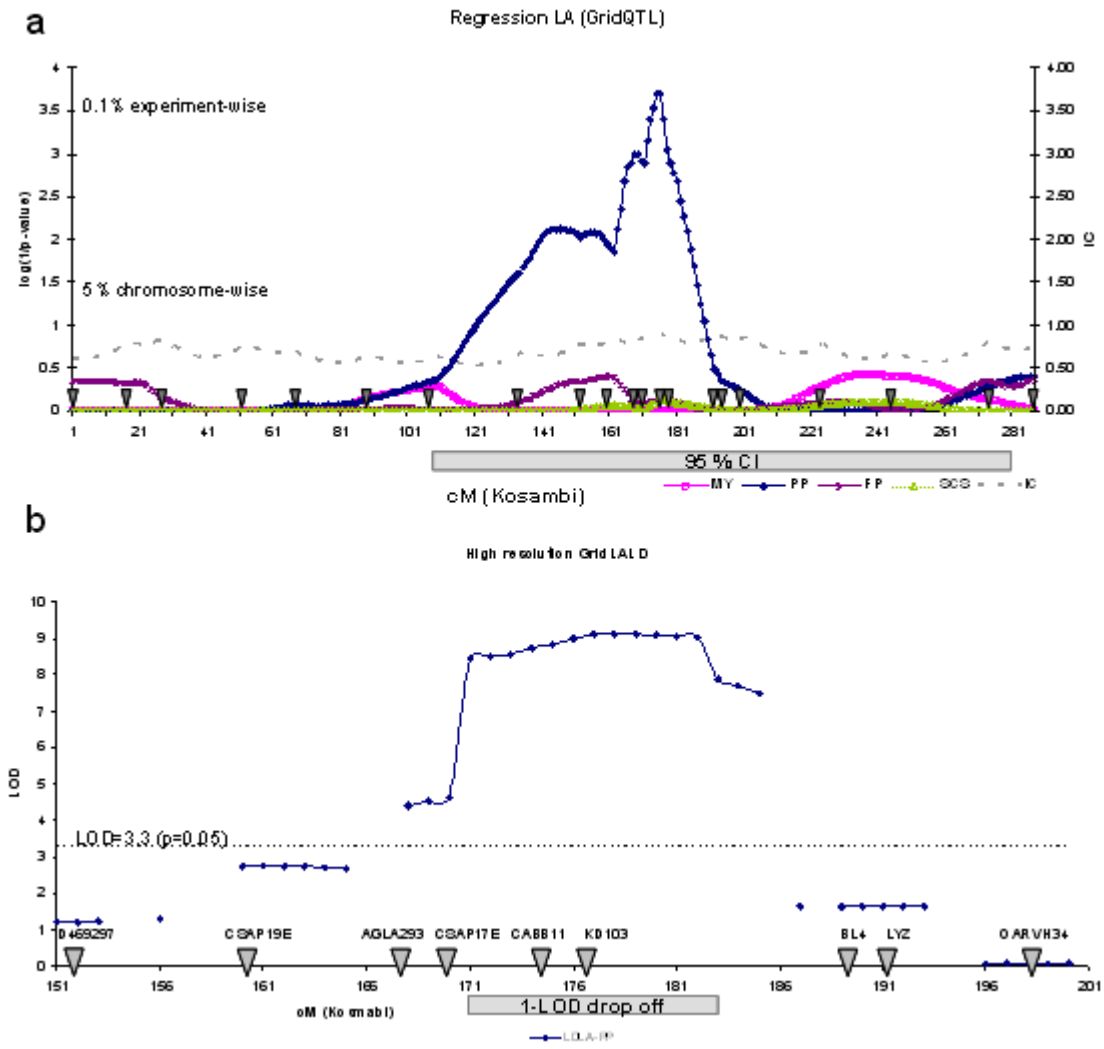
[5] Fraction of the phenotypic variance explained by the QTL, which was determined as the percentage of reduction in the residual variance due to the inclusion of the QTL in the model (adapted from Knott *et al*. 1998).

[10] Magnitude and standard error of the allelic substitution effect calculated for each segregating family, expressed in units of the trait (kg for milk yield and percentage points for composition traits) and in phenotypic SD units of the analysed traits (values in brackets).

FIGURES

Figure 1. a) Regression LA results. Statistical profiles obtained in the regression LA performed on chromosome 3 for the four traits analysed in the present study (milk yield, MY; FP, Fat percentage; PP, Protein percentage; SCS, Somatic cell score). The x-axis indicates the relative position on the linkage map (Kosambi cM), and the left y-axis represents the log (1/Pc). The horizontal lines indicate the significance thresholds that were considered (Pc < 0.05 and Pexp < 0.001). The information content that was obtained along the chromosome is represented on the right y-axis. Beginning at the centromeric end, the triangles on the x-axis indicate the relative positions of the markers that were analysed, including DK5391A, OARCP34, BMS460, BM8118, BM1831, INRA131, BM304, BM827, D469297, CSAP19E, AGLA293, CSAP17E, CABB11, KD103, BL4, LYZ, OARVH34, CSRD2111, BMS1248, MNS37A and MCMA13. The 95% CI, which was calculated by bootstrapping, is shown as a grey box at the bottom of the figure.

b) Results obtained in the LDLA The statistical profile obtained in the LDLA scan (1 cM step) performed on OAR3 for protein percentage is shown. This analysis was performed along a 49.7 cM-long interval that was centred on the QTL peak position identified by LA (175 cM) and flanked by markers D469297 and OARVH34. The LOD values are plotted against the positions of the linkage map. The maximum LOD value was located at position 178 cM (P = 9.184E-11). The dashed horizontal line indicates the significance threshold that was considered. The corresponding one-LOD drop-off CI is shown as a grey box at the bottom of the figure.


a
Regression LA (GridQTL)
0.1% experiment-wise
5% chromosome-wise
log(1/p-value)
95% CI
cM (Kosambi)
b
High resolution GridLALD
LOD
LOD=3.3 (p=0.05)
D469297
CSAP19E
AGLA293
CSAP17E
CABB11
KD103
BL4
LYZ
OARVH34
1-LOD drop off
cM (Kosambi)

SUPPLEMENTARY INFORMATION

Supplementary Table S1. Linkage map constructed for sheep chromosome 3 in this study for the analysis of Population B. Marker identities, order and positions (cM Kosambi) of markers are indicated for the two maps constructed with CRIMAP v2.0 (Green et al. 1990). The number of alleles and the number of heterozygous sires is indicated is also indicated for each locus analysed.

Chromosome	Marker	Population B		
		Position (cM)	Num Alleles	Num Het Sires
OAR3	*DK5391A*	0	7	6
	OARCP34	16.7	7	11
	BMS460	26.7	15	15
	BM8118	51.4	8	14
	BM1831	65.1	7	10
	INRA131	88.4	9	8
	BM304	108.8	12	8
	BM827	133.8	10	9
	D469297	151.2	8	11
	CSAP19E	160.8	9	11
	AGLA293	167.2	6	6
	CSAP17E	169.8	13	8
	CABB11	174.4	17	14
	KD103	177.2	11	11
	BL4	191	15	12
	LYZ	193	7	6
	OARVH34	200.8	11	11
	CSRD2111	223.1	16	13
	BMS1248	244.7	11	9
	MNS37A	272.6	19	12
	MCMA13	286.8	7	13

EVALUACIÓN DE LA CORRESPONDENCIA ENTRE LOS MAPAS GENÉTICO (MACHO) Y FÍSICO EN EL GANADO OVINO DE LA RAZA CHURRA

E. García-Gámez[1], B. Gutiérrez-Gil[1], J.P. Sánchez[2], Y. Bayón[1], L.F de la Fuente[1], F. San Primitivo[1], J.J. Arranz[1]

[1]Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, 24071 León. [2]IRTA Lleida. Av. Alcalde Rovira i Roure, 191 25198 Lleida. E-mail: jjarrs@unileon.es

XVI Reunión Nacional de Mejora Genética Animal. 31 de mayo - 2 de junio 2012, Ciutadella de Menorca (España).

INTRODUCCIÓN

El desarrollo de herramientas genómicas en el ganado ovino en los últimos años ha hecho posible la identificación de regiones asociadas a caracteres de interés en producción animal (Becker et al., 2010; García-Gámez et al., 2011; Zhao et al., 2011). La eficacia de estas herramientas está directamente relacionada con la calidad del alineamiento de la secuencia genómica de la que provienen. Debido a la naturaleza, fundamentalmente repetitiva, de los genomas animales y a las cortas longitudes de los fragmentos secuenciados por las tecnologías de secuenciación de segunda generación, no siempre la calidad de los mapas físicos derivados de los primeros alineamientos de los genomas presentan una elevada fiabilidad. Además, el conocimiento de cómo esta arquitectura física se traduce en la formación del fenotipo hace necesaria la elaboración de un mapa genético utilizado en análisis de asociación y de ligamiento. En este último caso, el gran número de meiosis informativas que deben ser analizadas para obtener un mapa de ligamiento de alta densidad, hace que se deban analizar pedigríes de un gran número de animales, difícilmente disponibles en un único experimento. En la mayoría de los casos lo que se hace es utilizar los datos medios obtenidos en los análisis del genoma humano (Yu et al., 2001) para convertir las distancias físicas en genéticas considerando que, como decía, a lo largo del genoma 1 cM es igual a 1 Mb. Esta medida es aproximada ya que se ha demostrado que existen regiones cromosómicas más susceptibles a sufrir recombinación (junglas de recombinación) y otras con muy poca tendencia a recombinar (desiertos de recombinación).

El objetivo de este trabajo es evaluar la versión 2 del alineamiento del genoma de la oveja mediante con la estimación de un mapa de ligamiento basado en las meiosis del sexo masculino en una población de ganado ovino de raza Churra.

MATERIAL Y MÉTODOS

Un población comercial de raza ovina Churra se ha genotipado para el chip de SNPs *Illumina OvineSNP50BeadChip*. Esta población está formada por 1.696 animales, distribuidos en 16 familias de medio-hermanas, con una media de 105 animales por familia.

El mapa físico, con las posiciones de los marcadores en pares de bases (bp), utilizado en este análisis se corresponde con la versión 2 del genoma de la oveja, disponible online (http://www.livestockgenomics.csiro.au/cgi-bin/gbrowse/oarv2.0/). En dicha versión del

genoma existen marcadores que pueden localizarse en diferentes regiones genómicas con una probabilidad muy similar (B. Dalrymple, comunicación personal).

El control de calidad de los genotipos se realizó, utilizando el software PLINK v1.06 (Purcell et al., 2007), a dos niveles: primero, un control por animal y, posteriormente, un control de calidad por marcador. En la primera fase, se eliminaron todos los animales con un porcentaje de genotipos inferior al 90 %. En la segunda, todos los SNPs con una tasa de genotipado inferior o igual a 95 %, frecuencia del alelo menos frecuente (MAF) inferior o igual a 0,01 y probabilidad de equilibrio Hardy-Weinberg menor de 0,00001, se eliminaron del análisis. Finalmente, todos los SNPs con localización en el genoma desconocida o que se localizan en cromosomas sexuales se excluyeron del análisis aunque hubiesen pasado del control de calidad.

Para la construcción del mapa de ligamiento, se utilizó la opción *fixed* del programa CRI-MAP (Green et al., 1990), v2.503 (proporcionada por JF Maddox). En este análisis, se asume como verdadero el orden de los marcadores, obtenido de la versión 2 del genoma ovino, y se calcula la distancia genética entre ellos en función de la frecuencia de recombinación en las familias estudiadas. Estos cálculos se llevaron a cabo tantas veces como fue necesario para minimizar los problemas de localización de algunos marcadores. En el caso de que un marcador mostrase un nivel de recombinación con los adyacentes mucho más elevado del esperado por la distancia física entre ellos existen cuatro opciones: (1) si dicho marcador no tiene otra posible localización en el genoma, el mapa se construye sin ese marcador, en caso de que la frecuencia de recombinación entre los marcadores restantes sea adecuada a su distancia, se elimina el marcador problemático y se asume como verdadero el nuevo orden (sin ese marcador); (2) si no tiene otra posible localización, y el mapa genético no mejora al eliminarlo, probamos eliminando los marcadores colindantes, si aun así no mejora, asumimos que puede haber un error, pero los siguientes cálculos se hacen asumiendo como correcto el mapa inicial (con el marcador problemático); (3) si el SNP tiene otra posible localización en el genoma, reconstruimos el mapa con la nueva ubicación y, si mejora, asumimos esta última como verdadera; (4) en caso de que el mapa no mejore, procedemos a las opciones 1 y 2, descritas previamente, eliminando el marcador del análisis. Las posiciones "definitivas" de los mapas físico (en Megabases, Mb) y de ligamiento (en centiMorgan, cM) se compararon y se ha estimado la correspondencia entre ambas variables tanto por cromosoma como a nivel genómico.

RESULTADOS Y DISCUSIÓN

En total, 15 animales se han eliminado del análisis por no cumplir el criterio de calidad (más del 90% de genotipos válidos), quedando un total de 1.681 animales cuyos genotipos se han utilizado en la construcción del mapa de ligamiento en oveja Churra. De los 54.241 SNPs de chip de Illumina, se eliminaron 6.431 SNPs en el control de calidad: 4.316 SNPs no pasaron el filtro por exhibir más de un 5% de genotipos no válidos; 1.540 SNPs mostraban un MAF por debajo del umbral (0,01); y 575 marcadores no estaban en equilibrio Hardy-Weinberg ($p < 0,00001$) en la población analizada. De los 47.810 SNPs restantes, se han utilizado para la construcción de la primera versión del mapa de ligamiento, los SNPs autosómicos con localización conocida en el genoma, es decir, 46.365.

En total se realizaron 6 iteraciones del mapa de ligamiento hasta obtener la versión "de trabajo". En el proceso de construcción del mapa 25 SNPs han sido eliminados del mapa de ligamiento y 114 marcadores han sido integrados en el mapa genético ocupado la segunda o tercera posición más verosímil en el mapa físico.

Al comprobar la relación entre los mapas físico y genético en la población de raza Churra estudiada, podemos observar que, como media, 1 Mb equivale a 1,85 cM. Este ratio (cM/Mb) varía a lo largo del genoma, desde 1,50 en OAR2 hasta 2,37 en el OAR20. En la Figura 1 se presentan de manera gráfica estos resultados. Al representar el genoma completo (Figura 1A), se observa claramente que el ratio observado es superior a 1 cM ~ 1 Mb. Las Figuras 1B y 1C muestran un zoom sobre los cromosomas con los ratios mayor y menor, respectivamente a lo largo del genoma. Así como en el cromosoma OAR2 (cM/Mb = 1,50), se puede observar una correspondencia entre el mapa físico (eje Y) y el mapa de ligamiento (eje X) constante sin regiones donde la recombinación sea muy elevada, en el cromosoma OAR20 (cM/Mb = 2,37) se observan diferencias importantes entre las distintas regiones del mismo.

Hay que tener en cuenta que las técnicas de secuenciación de segunda generación producen millones de lecturas de tamaño corto (35-500 bp) que se alinean a lo largo del genoma. Estas secuencias forman *contigs*, *scaffolds* y *super-scaffolds*, entre los que puede haber huecos y las distancias físicas estar sub- o sobrestimadas. En el caso del genoma ovino, aún en estado de borrador inicial, este fenómeno puede claramente afectar a la calidad del mapa físico. Además hay que tener en cuenta que la calidad del mapa de ligamiento obtenido está limitada por el reducido tamaño del pedigrí analizado. Las diferencias en la

correspondencia cM/Mb entre cromosomas ya se ha puesto de manifiesto en otras especies de mamíferos (Yu et al., 2001; Liu et al 2009) existiendo regiones donde la recombinación es mucho mayor que la media del genoma (selvas de recombinación) y otras donde se casi no se observan sucesos recombinatorios (desiertos de recombinación).

Los resultados obtenidos en nuestro análisis muestran que la relación cM/Mb es de 1,85. El refinamiento del mapa físico en futuras actualizaciones del genoma, puede ayudar a que este ratio se acerque más al ratio encontrado en las especies con un genoma mucho más elaborado 1 cM ~ 1 Mb. Por el momento, es necesario tener en cuenta, a la hora de interpretar resultados basados en esta versión del genoma, que se encuentra en un estado inicial de desarrollo y esto puede llevar a una incorrecta asociación de regiones genómicas con caracteres productivos.

AGRADECIMIENTOS

Este trabajo se ha realizado con financiación del proyecto AGL2009-07000 (Ministerio de Ciencia e Innovación) y "3SR" *Suistanable Solutions for Small Ruminants* (Comisión Europea). EGG es becaria FPU (Ministerio de Educación).

REFERENCIAS BIBLIOGRÁFICAS

[1] Becker D, Tetens J, Brunner A, *et al*. 2010. PLoS ONE. 5: e8689.

[2] García-Gámez E, Reverter A, Whan V, *et al*. 2011. PLoS ONE. 6: e21158.

[3] Green P, Falls K & Crooks S. 1990. Documentation for CRI-MAP, v2.4.

[4] Liu Y, Qin X, Song XZ, *et al*. 2009. BMC Genomics. 10:180

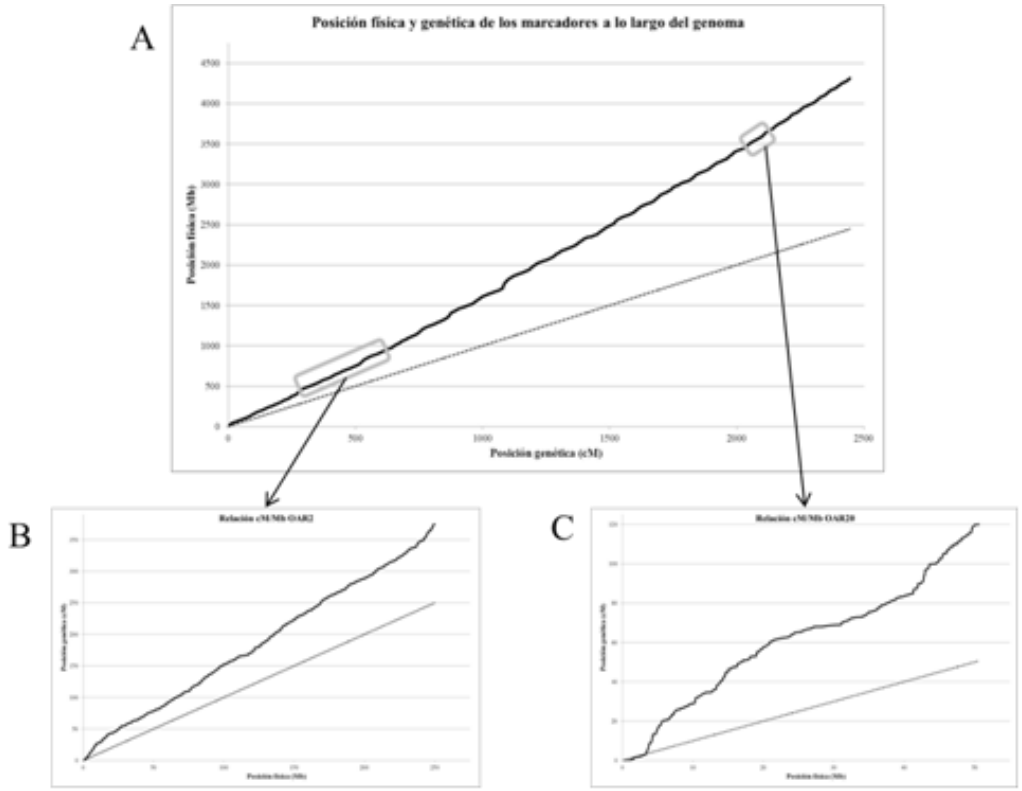
[5] Purcell S, Neale B, Todd-Brown K, *et al*. 2007. Am. J. Hum. Genet. 81.

[6] Yu A, Zhao C, Fan Y, *et al*. 2001. Nature. 409:951-3.

[7] Zhao X, Dittmer KE, Blair HT, *et al*. 2011. PLoS ONE, 6: e21739.

FIGURAS

Figura 1. Figura 1: Representación gráfica de la relación entre las posiciones físicas (en Megabases, Mb) y genéticas (en centiMorgan, cM) de los marcadores obtenidas en este estudio y la relación 1 cM ~ 1Mb (línea discontinua). (A) Relación entre posiciones de los mapas físicos y genéticos a lo largo de todo el genoma, se representan las posiciones en cM en el eje X y en Mb en el eje Y. (B) y (C) Posiciones físicas y genéticas para los marcadores situados en los cromosomas OAR2 (B) y OAR20 (C).



ASSESSMENT OF THE CORRELATION BETWEEN GENETIC (MALE) AND PHYSICAL MAPS IN SPANISH CHURRA SHEEP

ABSTRACT: To assess the correspondence between the physical map, derived from Ovine Assembly v2.0, and the genetic map in Spanish Churra sheep, we built a linkage map using the genotypes from the *Illumina Ovine SNP50BeadChip*. After a quality control per animal and per marker, 46,341 autosomal SNPs and 1,681 animals belonging to 16 half-sib families were used to estimate the equivalence between marker positions in cM and Mb. The results reported here show that the average recombination frequency in the 16 sires is 1.85 cM per Mb, higher than the average value in most mammals (1 cM ~ 1 Mb). Moreover, we have detected differences between chromosomes, with a minimum of 1.50 cM/Mb in OAR2 y a maximum of 2.37 cM/Mb in OAR20. These values should be used as a reference in studies where recombination information between markers is important, for example, in QTL detection or effective population size estimation based on molecular information. However, we have to take into account that in this work we have only analysed male meiosis and the number of animals is limited. It is also important that the Ovine Genome Assembly is still in a draft stage and future refinements will improve the sequence.

LINKAGE DISEQUILIBRIUM AND INBREEDING ESTIMATION IN SPANISH CHURRA SHEEP

García-Gámez E.[1,2], Sahana G. [2], Gutiérrez-Gil B. [1], Arranz J. J. [1 *]

* Corresponding author: Juan-José Arranz: jjarrs@unileon.es

[1]Dpto. Producción Animal, Universidad de León, 24071, León, Spain. [2]Department of Molecular Biology and Genetics, Aarhus University, Aarhus, Denmark

BMC Genetics 13:43.

ABSTRACT

Background: Genomic technologies, such as high-throughput genotyping based on SNP arrays, have great potential to decipher the genetic architecture of complex traits and provide background information concerning genome structure in domestic animals, including the extent of linkage disequilibrium (LD) and haplotype blocks. The objective of this study was to estimate LD, the population evolution (past effective population size) and the level of inbreeding in Spanish Churra sheep.

Results: A total of 43,784 SNPs distributed in the ovine autosomal genome was analyzed in 1,681 Churra ewes. LD was assessed by measuring r^2 between all pairs of loci. For SNPs up to 10 kb apart, the average r^2 was 0.329; for SNPs separated by 200–500 kb the average r^2 was 0.061. When SNPs are separated by more than 50 Mbp, the average r^2 is the same as between non-syntenic SNP pairs (0.003). The effective population size has decreased through time, faster from 1,000 to 100 years ago and slower since the selection scheme started (15-25 generations ago). In the last generation, four years ago, the effective population size was estimated to be 128 animals. Inbreeding coefficients, although differed depending on the estimation approaches, were generally low and showed the same trend, which indicates that since 2003, inbreeding has been slightly increasing in the studied resource population.

Conclusions: The extent of LD in Churra sheep persists over much more limited distances than reported in dairy cattle and seems to be similar to other ovine populations. Churra sheep show a wide genetic base, with a long-term viable effective population size that has been slightly decreasing since selection scheme began in 1986. The genomic dataset analyzed provided useful information for identifying low-level inbreeding in the sample, whereas based on the parameters reported here, a higher marker density than that analyzed here will be needed to successfully conduct accurate mapping of genes underlying production traits and genomic selection prediction in this sheep breed. Although the Ovine Assembly development is still in a draft stage and future refinements will provide a more accurate physical map that will improve LD estimations, this work is a first step towards the understanding of the genetic architecture in sheep.

BACKGROUND

The application of recently developed genomic technology, such as genome-wide SNP genotyping has great potential to increase our understanding of the genetic architecture of complex traits and to improve selection efficiency in domestic animals through genomic selection. However, the success of these approaches depends on the extent of the linkage disequilibrium (LD) across the genome, which may vary between populations. As an example, the extent of linkage disequilibrium serves to assess the number of markers required to associate genetic variation with economically important traits. A population with extensive LD will require a lower marker density; in contrast, if LD persists over short distances many more markers will be required to obtain the same power to detect association [1]; the same reasoning could be applied to genomic selection efficiency [2, 3]. Similarly, the signatures of genomic regions under positive selection can be identified by studying the haplotype block structure throughout the genome [4].

The linkage disequilibrium pattern can also provide insight into the evolutionary history of a population. The extension of LD in the genome could be used to infer ancestral effective population size (N_e) [5, 6, 7]. This is an important population parameter that helps to explain how populations evolved and can improve the understanding and modeling of the genetic architecture underlying complex traits [8].

Another aspect of interest while studying a commercial population under selection pressure is to study the level of inbreeding. Traditional estimation of the inbreeding coefficient based on pedigree data [9] is dependent on the completeness and accuracy of the available pedigree records. Currently, using the information provided by molecular markers (genome-wide SNP chip panels), we can estimate this coefficient with or without pedigree information [10]. Several methods have been described for this purpose [10, 11, 12, 13].

An increasing number of studies have analyzed LD features in livestock species, especially in cattle [4, 14, 15] but also in pigs [16], horses [17] and chicken [18]. In domestic sheep, LD studies based on microsatellite data [1, 19] found extended LD across the genome, although a marked variation between different breeds was reported [1]. Within the framework of the SheepHapMap project, the Illumina Ovine SNP50 BeadChip has been used to present a preliminary evaluation of LD in 74 diverse breeds [20]. A subset of informative SNPs from this chip has also been utilized in wild sheep to calculate the extent of LD and evaluate the

usefulness of this chip, which was developed for domestic sheep, for conducting genome-wide association studies in wild sheep populations [21].

The objective of this study was to characterize LD in a Spanish Churra sheep commercial population using data generated with the Illumina Ovine SNP50 BeadChip. This genomic tool is currently being used in this dairy sheep breed to fine-map previously reported QTLs [22] and to obtain a preliminary assessment of the genomic selection approach [23]. Thus, we have studied the genome structure (LD and haplotype blocks), population evolution (past effective population size) and the level of inbreeding present in this population, which will provide fundamental information on the genome organization of this Spanish sheep breed.

METHODS

Resource population and SNP genotyping

A commercial population of Spanish Churra sheep was analyzed in this research. Blood samples were collected from 1,710 Spanish Churra ewes belonging to 16 half-sib families and distributed across 20 different flocks. Semen straws were obtained for the 16 sires. The use of animals was performed in compliance with the guidelines approved by the University of Leon ethical commission.

DNA was extracted from blood and semen samples using standard protocols, as described in [24]. A control for Mendelian inheritance errors was performed at this stage using a panel of 18 microsatellite markers [25]. Finally, a total of 1,696 DNA samples with a concentration of 50 ng/µl and $A_{260/280}$ ratio above 1.8 were used for Illumina Ovine SNP50 BeadChip genotyping. Genotyping was performed commercially at AROS Applied Biotechnology AS (Aarhus, Denmark) and LABOGENA (Jouy-en-Josas, France). Quality control (QC) of the raw genotypes consisted of checking the GenCall Score (GCscore) using the GenomeStudio software (Illumina Inc. San Diego, CA). Genotypes with a GCscore < 0.15 were set as missing genotypes.

Quality control, marker order and genetic distances

The SNPs included in the Illumina Ovine SNP50 BeadChip were mapped using the Ovine Genome Assembly v2.0 [26]. The markers were filtered to exclude loci assigned to

unmapped contigs. Only the SNPs located on the sheep autosomes were considered in further analyses.

We performed QC of the genotypes in two stages, first implementing the control on a 'per-individual' basis prior to conducting QC on a 'per-marker' basis to maximize the number of markers that remained in the study [27, 28]. First, individuals were removed if they had more than 10 % missing genotypes. Secondly, the marker-QC included three steps: (i) control of call rate (≥ 0.95), (ii) minor allele frequency (MAF) (≥ 0.05) and (iii) correspondence with Hardy-Weinberg equilibrium (HWE) (p-value > 0.00001). For the markers that passed the previously mentioned QC, we performed a final QC using the available pedigree. Thus, the genotypes causing Mendelian inheritance inconsistencies were set to "missing" and afterwards inferred based on the sire's genotype and the population frequencies of the two possible alleles. This imputation process was done with an unpublished FORTRAN based program (VerifTyp 1.0; Boichard D and Druet T, personal communication), which performs 10 inference iterations where the base population frequencies are re-estimated at each step depending on the reconstruction of genotypes. A probability threshold was set to avoid over-representation of very frequent alleles.

The initial locus order between adjacent markers, which was based on the Sheep Genome Assembly v2.0, was assessed using the fixed option of a modified version of CRI-MAP [29], v2.503 (kindly provided by J. F. Maddox). The information derived from this control was used to mend some colocation problems, as some markers had more than one hit in the reference assembly (B. Dalrymple, personal communication). The resulting marker order and positions were used as the physical map to perform the LD analyses [Additional file 1].

Haplotype construction

The ideal scenario to measure the extent of LD within a population is to analyze "non-related" individuals. Our resource population of half-sib families had initially been selected to perform linkage-based QTL mapping studies using a daughter design [30], and therefore, the sampled individuals were related. To overcome this limitation, we attempted to obtain a representation of independent haplotypes of the population under study. With this purpose, we calculated chromosome phases taking into account the population pedigree structure using PHASEBOOK package [31]. Following the three-step approach described by the authors, we

first used the LinkPHASE 2.3 program (part of PHASEBOOK) to obtain partially phased genotypes using pedigree and linkage information (steps I and II). Then, DAGPHASE 1.1 (part of PHASEBOOK), in combination with BEAGLE 3.3 [32], was used to impute missing markers based on linkage disequilibrium (step III). For this analysis, we used i) DAGPHASE 1.1 option 1 to fill-in the missing base haplotypes at random, ii) 15 iterations using BEAGLE 3.3 to construct the optimal directed acyclic graph (DAG) and DAGPHASE 1.1 option 2 to sample the missing alleles of the base haplotypes according to a Hidden Markov Model (HMM) and iii) DAGPHASE 1.1 option 3 and the last DAG to calculate the haplotypes.

Linkage disequilibrium and effective population size

Reconstructed haplotypes were selected to not have an overrepresentation of the sires' haplotypes [15]. Sire haplotypes and maternal-inherited dam haplotypes were inserted into HAPLOVIEW v4.1 [33] to estimate LD statistics based on pairwise SNPs. For easy comparison of results with other reports, the two most commonly used statistics, D' [34] and r^2 [35], were computed for this study. For non-syntenic SNPs, a subset was used to estimate LD across the genome. This selection was based on a random representative sample of the SNPs analyzed in each chromosome (5 % of the SNPs used in the analysis). Both LD metrics (D' and r^2) were estimated for each non-syntenic pair. To assess how far LD extends, we average r^2 based on the SNP distance in 1-Mb intervals and calculated the half-length of r^2 [21]. This half-length is the distance at which LD decays to half of its maximum value [36].

HAPLOVIEW v4.1 was also used to define the haplotype blocks present in the genome. The method followed for block definition was previously described by Gabriel et al [37]. A pair of SNPs is defined to be in 'strong LD' if the one-sided upper 95 % confidence bound of D' is higher than 0.98 and if the lower bound is above 0.7. In contrast, 'strong evidence for historical recombination' is defined if the upper confidence bound on D' is less than 0.9 [37].

Past effective population size (N_e) was calculated for 11 time points. Based on the physical map used for the LD analysis, genetic distances between adjacent markers were calculated using three conversion rates: (i) considering a 1 cM ~ 1 Mb conversion rate for all the chromosomes, (ii) considering the specific cM/Mb ratio calculated for each chromosome by comparing the genetic and physical map in the CRI-MAP analysis and (iii) considering the average conversion rate estimated across the genome for all the chromosomes. Each genetic

distance (c, in Morgans) corresponds to a value of T generations in the past. This value was calculated as T = 1/(2c). The following formula was used to estimate N_e [8]:

$$r^2 = 1/(1+4N_ec) + 1/n,$$

where c is the distance in Morgans, and n is the chromosome sample size (number of haplotypes) used in the analysis. According to the genetic distances between markers, SNP pairs were stacked into bins of 1,000 pairs, and the average distance and r^2 estimated for each bin were used to calculate the N_e.

Inbreeding coefficients

Two different approaches were used to estimate the coefficient of inbreeding (F) within the Spanish Churra population. The pedigree-based F (F_{PED}) was estimated using the Relax2 software [38], based on the algorithm described by Meuwissen and Luo [39] using available pedigree records since 1978 (5,956 animals in total). Marker-based inbreeding coefficients were estimated using the GCTA software [12]. To calculate the marker-based F values, we used the population under study as the base population. Allele frequencies were estimated across the 1,681 animals, and the GCTA software was used to obtain F values. Three different metrics were obtained using the --ibc option of the program: a) based on the variance of the additive genotype (F_1), b) based on the excess of homozygosity (F_2) and c) based upon the correlation between uniting gametes (F_3) [12].

RESULTS

SNP distribution and frequencies

Out of a total of 54,241 SNPs genotyped in this study, 1,516 SNPs were unmapped, and 215 were located on sex chromosomes as per Ovine Genome Assembly v2.0. Thus, 52,510 SNPs mapped onto the 26 sheep autosomes were used in the described analyses.

Of the 1,696 genotyped animals, 15 individuals did not pass the QC. Thus, a total of 1,681 animals were used in the analyses. The number of markers removed during QC was 8,726 SNPs: 4,140 SNPs were deleted due to low call rate (< 0.95); 3,044 SNPs did not reach minimum MAF (< 0.05); and 1,542 markers were not in HWE ($P \leq 0.00001$). The total number of markers used in the analyses was 43,784 SNPs. The distribution of these SNPs per

chromosome is described in Table 1, ranging from 598 on OAR24 to 4,987 on OAR1. The average distance between SNPs was 55.74 kb, ranging from 51.47 kb in OAR9 to 69.71 kb in OAR24. The distribution of MAF across the chromosomes was similar, with a mean value of 0.288 (Table 1). Figure 1 represents the distribution of SNPs in MAF bins. Around 50 % of the SNPs had an MAF value over 0.3. The cM/Mb ratios calculated for each chromosome by comparing the genetic and physical maps in the CRI-MAP analysis ranged between 1.5 on chromosome 2 and 2.37 on chromosome 20 (Table 1). The average conversion rate estimated across the genome was 1.85 cM ~ 1 Mb.

Linkage disequilibrium and haplotype blocks

For the LD analysis, the total number of reconstructed 'non-related' haplotypes (chromosomes) was 1,692. A total of 42,381,374 syntenic pairs of SNPs was analyzed for all the autosomes. Average *D'* and r^2 values, pooled over autosomes in different categories of map distances, are presented in Table 2. The distribution of both *D'* and r^2 with respect to the physical distance separating loci is presented in Figure 2. As shown in Figure 2a and Table 2, there is a decline in r^2 with increasing physical distance between SNPs: for SNPs up to 10 kb apart, the average r^2 is 0.329; for SNPs separated by 200–500 kb, the average r^2 is 0.061. When SNPs are separated by more than 50 Mbp, the average r^2 (0.003) is the same as that found between non-syntenic SNP pairs (0.003). The distribution of *D'* is similar to that observed for r^2. The half-length of r^2 was 0.033, which corresponds to a distance between SNPs of 2.5 Mbp (Figure 2b). Small differences in average LD values were observed among chromosomes and they corresponded with differences in chromosome length (Table 1). Average r^2 per chromosome ranged from 0.006 on OAR1 to 0.015 for OAR20. These differences across the genome were lower when comparing values obtained at the different distance bins. For example, average r^2 values for SNPs up to 100 kb apart ranged from 0.126 on OAR22 to 0.180 on OAR10, which was the chromosome showing the highest level of LD between markers.

A summary of the distribution, size, number and SNPs involved in the haploblocks per chromosome are presented in Table 3. A total of 2,099 haploblocks spanning 56,726 kb (2.32 %) of the Churra autosomal genome were detected. The average block size was 27.03 kb, ranging from 0.04 kb (OAR14, 2 SNPs) to 1,263 kb (OAR2, 8 SNPs). In total, 4,780 SNPs (10.92 % of all SNPs used) formed blocks with a range of 2–13 SNPs per tract. The

chromosomes showing the longest and shortest haplotypic structures in the genome were OAR2 with 252 blocks spanning 11,149 kb and OAR26 with 22 blocks covering 367 kb. There was a region in OAR10, from 34.2 to 42.0 Mbp, with a high density of haplotype blocks. This region included 8 haploblocks involving the 61 % of haplotype block length of this chromosome (3,197 kb of the total of 83,632 Kb). The length of these blocks varied between 2 SNPs, for one of the haploblocks and 9 SNPs in three of the other haploblocks.

Past effective population size estimations

A graphical representation of the N_e values at each time point, from 250 to 1 generation ago, and for each of the three different cM/Mb conversion rates used to calculate genetic distance between markers is given in Figure 3. Taking into account a generation interval of 4 years, these results correspond to Churra populations 1,000 to 4 years ago, approximately. The results show that N_e has decreased through time, faster at 1,000 to 100 years ago and slower since selection scheme started (15-25 generations ago) (Figure 3). The effective population size in the last generation (4 years ago) averaged across three cM/Mb ratios is calculated to be 128 animals.

Inbreeding measurement

Available pedigree used in this analysis included 5,956 animals, which represents a pedigree depth of 6 generations. From the 11,912 expected parents, 41 % were missing data. The inbreeding coefficient calculated using this pedigree information, F_{PED}, was estimated to be null (no inbreeding) for 5,856 out of the 5,956 animals, with an average of 0.001. For the remaining 100 'inbred animals', the mean value was 0.064. To compare these results with marker-based inbreeding values, we extracted pedigree-based inbreeding coefficients for the 1,681 genotyped animals. For these animals, a mean value of 0.003 was obtained. Using molecular information, we did not obtain any 'non-inbred' animal ($F = 0$), but some results were negative. With the aim of comparing results between the marker-based and the pedigree-based methods, we transformed negative values to 0 and studied the differences in values across time. Average values for the positive estimates were 0.015 (F_1), 0.009 (F_2) and 0.005 (F_3). Figure 4 shows the inbreeding coefficients obtained for animals born each year since 2001. A peak for the values obtained using the F_1 method was found for year 2002. A more detailed study of the animals born that year showed more than half of these animals had an inbreeding coefficient higher than 0.0625, which is very high according to the results obtained

across the population. Although the estimates from various approaches were different, they all show the same trend and indicate that in the studied resource population inbreeding has increased slightly since 2003.

DISCUSSION

This study presents an analysis of the extent of LD in Spanish Churra sheep using 43,784 SNPs distributed across the autosomal genome, although the draft stage of the version of the Ovine Assembly it is based on should be taken into account. Future refinements and updates in the physical maps can lead to changes in the estimations reported here. To enable comparison with previous studies in sheep and other domestic species, we estimated two pairwise statistics: *D'* and r^2. *D'* values were higher than those estimated for r^2. This might be because rare alleles and unobserved haplotypes tend to inflate *D'* but not r^2 [1].

Comparing the level of LD obtained in different studies is difficult because of different sample sizes, LD measures, marker types, marker densities and recent and historical population demographics [4]. Previous reports in sheep based on microsatellite marker analysis have described LD as extensive (up to 20 cM) [1, 19], although its magnitude and significance was shown to vary markedly between different breeds [1]. The results reported for wild sheep [21] also showed LD extended over long distances (half-length r^2 of 4.6 Mb), which contrasts with the short extension of LD reported here for Churra sheep (half-length r^2 value 2.5 Mb). A recent assessment of LD based on the analysis of 51,446 SNPs in a sample of Sarda rams showed a similar level of LD than in Churra sheep, with an average r^2 value over 1,000 kb of 0.072 [40]. Compared with the results based on SNP genotyping described for cattle [4, 14, 15, 41], LD estimates between syntenic and nonsyntenic loci in Churra sheep was two times lower. Initial results from the analysis of 74 domestic sheep breeds with the Illumina Ovine SNP50 BeadChip [20] were in concordance with our findings, which suggests a relatively low level of LD in sheep and a substantially lower LD in sheep when compared with a wide range of cattle breeds, including dairy and beef cattle [42]. This analysis also showed Churra sheep as one of the breeds with a more remarkable decay of LD with the distance between markers when compared with other breeds [20]. Average r^2 between nonsyntenic SNP pairs provides an idea of the LD that can be expected by chance. None of the nonsyntenic SNP pairs tested showed a 'high' LD value ($r^2 > 0.25$).

Differences in LD between chromosomes have already been reported in Holstein cattle [4, 15]. These can be attributed to recombination rates varying between and within chromosomes, heterozygosity, genetic drift and effects of selection [4]. Our results for average LD within a chromosome are in concordance with the block structure across the genome. Chromosomes showing higher LD also have more and longer blocks than chromosomes with lower average LD. In Churra, 88 % of the blocks contained just two SNPs. Preliminary results from the SheepHapMap project also identified an overall limited genome coverage in haplotype blocks (of at least three SNPs) for domestic breeds with Churra showing the lowest coverage (0.8 %), whereas wild Soay sheep showed a large genome coverage (21.84 %) [20]. Compared with domestic sheep, the genomic distribution and coverage of the genome by haplotype blocks is higher in other species, such as cattle [4, 43], as expected according to the higher LD between markers reported in these species.

Also within the framework of the SheepHapMap project genomic regions containing signals of selection have been identified across a wide range of sheep breeds [44]. Higher homozygosity and LD is expected in regions that have undergone selection and are now fixed in the breed under study. Also, more and longer haplotype blocks are expected in those regions. Although there were haplotype blocks close to some of the regions related to selection, none of the high-Fst SNPs depicted by those authors [44] were involved in blocks in Spanish Churra sheep. For example, the region of OAR10 containing 8 haploblocks in Churra sheep is close to the *polled* locus (*RXFP2* gene), which is related to the presence of horns in sheep [44]. However, none of the SNPs linked to the polled phenotype were included in the Churra haploblocks. The longest haplotype block across the genome found in this study, which involved 8 SNPs and was located in OAR2 (111.9 – 113.1 Mbp), comprises the *HERC2* gene, which has already been related to pigmentation in cattle [45].

We also investigated the past effective population size (N_e) in the Churra sheep commercial population under study. First historical references from the existence of Spanish Churra sheep date from Middle Ages (thirteenth century) approximately 800 years ago [46]. Therefore, the time points chosen in this work were based on this historical information. The correlation between the results of the three different cM/Mb calculations was over 0.99. Major differences between the estimates based on the three different ratios are found at small distances, corresponding to more than 75 generations ago (Figure 3). Changes in the effective population size reflect past events that occurred in the corresponding populations. In Spanish

Churra, the N_e value has been descending through time until the selection scheme began. From that point on, no major changes are found. Effective population size estimated 50 generations ago in Churra (*Ne* = 467 animals) is in agreement with the observations reported within the framework of the SheepHapMap project, where most of the sheep breeds displayed high N_e, and only two populations showed a narrow genetic base comparable to the current N_e of domestic cattle breeds (N_e < 150) [44]. No other *Ne* estimations have been reported so far in sheep. High selection pressure and the use of artificial insemination are the main reasons for the low Ne values obtained in cattle [42]. To ensure an animal population is long-term viable, a threshold of *Ne* = 100 has been given [47]. Our results of current effective population size (*Ne* = 128) are above the threshold, but care should be taken on this regard to ensure that the effective population size is maintained.

The LD estimates reported in this work can serve to assess the utility of the Ovine SNP50 Beadchip to address fine-mapping studies in Churra sheep. In cattle, McKay et al. [48] showed that at a physical distance of 100 kb separating flanking SNP loci, the average r^2 was 0.15-0.2; considering a bovine genome length of 2.87 Gb, they concluded that 28,700 fully informative markers would be needed to saturate the cattle genome at an average resolution of 100 kb. Considering the lower value of LD estimates reported in this study, one can easily estimate that to obtain similar resolution in Churra sheep the marker density needs to be at least two times higher than the currently analyzed dataset. Hence, to implement genomic selection in this population with appropriate accuracy, a SNP array of higher density would be valuable. In this regard, following previous reported estimations [2, 3] we can estimate how many SNPs will be needed to accurately estimate breeding values in Churra sheep. Considering a marker density of 20 SNPs per genome effective segment, which represents each independent chromosome segment [3], a population with *Ne* of 128 animals and a genome of 30 Morgans, the SNP chip should include approximately 95,000 SNPs (assuming the same percentage of successful genotyping obtained in this study) to improve the accuracy of genetic breeding values estimation.

Pedigree-based inbreeding calculations rely on the completeness and accuracy of the available pedigree. The results reported herein based on the available Churra pedigree showed 94 % of the animals included in the analysis were ‘non-inbred’, although this is due to the lack of a deep pedigree. We obtained some negative values for inbreeding coefficients, which corresponded to animals with lower homozygosity than the average population. This could be

because we estimated the allele frequencies from the currently genotyped population instead of the base population. Correlation between the different methods to estimate inbreeding ranged from 0.27 (F_1 *vs.* F_2) to 0.83 (F_2 *vs.* F_3), with F_1 as the most different. In general, values calculated using pedigree information were lower than those obtained through marker analysis. The latter could be inflated because we assumed a homogeneous population [13], while there is a structure due to the experimental design of the linkage-based mapping studies for which the resource population had initially been selected. Comparing between the three marker-based methods, a different percentage of the analyzed animals showed an inbreeding value higher than the critical level (6.25 %, obtained when mating cousins, [49]). This proportion varied from 8.45 % (F_1) to 1.6 % (F_3), which is lower than that described for Finnsheep [49]. This percentage was very high when analyzing results from 2002 (method F_1), 50 % of the animals had an $F_1 > 0.0625$. Moreover, we were not able to compare between methods as most of the animals for F_2 and F_3 had negative values which might also explain the low correlation found between F_1 and F_2. In agreement with previous studies, the results presented here show that genomic data sets can provide useful information on a per sample basis in cases of complex genealogies or in the absence of genealogical data [13].

CONCLUSIONS

In the studied Churra sheep population, LD decayed with increasing genomic distance and the analysis yielded similar values than previously reported in other sheep populations. An estimation of genetic distance from physical position showed few differences between chromosomes, which did not affect the past effective population size results. Effective population size seems to have been decreasing until the recent selection scheme in Spanish Churra began. Marker information aided the estimation of the level of inbreeding present in the sample, although more accurate results could be obtained if we had a base population to estimate allele frequencies. In conclusion, the results reported herein are a first step toward understanding the genomic architecture of a domestic sheep breed and can be used to access the feasibility of direct future selection based on genomic data. The level of LD estimated in Churra sheep indicates that the present Illumina Ovine SNP50 BeadChip is not an optimum and the future availability of a high-density SNP-array or the use of next-generation sequencing methods will improve the performance of QTL fine-mapping studies and genomic selection accuracy in this population.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHOR'S CONTRIBUTIONS

EG-G performed the quality control, analyses and drafted the manuscript. BG-G and GS participated in the design and coordination of the study and helped to draft the manuscript. JJA conceived of the study and participated in its design and coordination. All authors have read and approved the final manuscript.

ACKNOWLEDGEMENTS

This work was supported by the Spanish Ministry of Science (Project AGL2009-07000) and by the European Commission by 3SR Project (Sustainable Solutions for Small Ruminants; http://www.3srbreeding.eu). Elsa García-Gámez is funded by an FPU contract from the Spanish Ministry of Education. The support and availability to the computing facilities of the Foundation of Supercomputing Center of Castile and León (FCSCL) (http://www.fcsc.es) is greatly acknowledged.

REFERENCES

[1] Meadows JR, Chan EK, Kijas JW: Linkage disequilibrium compared between five populations of domestic sheep. *BMC Genetics* 2008, 9: 61.

[2] Goddard M: Genomic selection: prediction of accuracy and maximization of long term response. *Genetica* 2009, 136: 245-252.

[3] Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA: The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 2010, 185(3):1021-1031.

[4] Qanbari S, Pimentel EC, Tetens J, Thaller G, Lichtner P, Sharifi AR, Simianer H: The pattern of linkage disequilibrium in German Holstein cattle. *Animal Genetics* 2010, 41(4):346-356.

[5] Sved JA: Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theoretical population biology* 1971, 2(2):125-141.

[6] Hill WG, Robertson A: Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* 1981, 38:226-231.

[7] Hayes BJ, Visscher PM, McPartlan HC, Goddard ME: Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Research* 2003, 13(4):635-643.

[8] Tenesa A, Navarro P, Hayes BJ, Duffy DL, Clarke GM, Goddard ME, Visscher PM: Recent human effective population size estimated from linkage disequilibrium. *Genome Research* 2007, 17(4): 520-526.

[9] Wright S: Coefficients of inbreeding and relationship. *American Naturalist* 1922, 56:330-338.

[10] Bouquet A, Sillanpaa MJ, Juga J: Estimating inbreeding using dense panels of biallelic markers and pedigree information. In *Book of Abstracts of the 62nd Annual Meeting of the European Association for Animal Production (EAAP): 29 August - 2 September; Stavanger, Norway*. Edited by: Wageningen Academic Publishers; 2011.

[11] Powel JE, Visscher PM, Goddard ME: Reconciling the analysis of IBD and IBS in complex trait studies. *Nature Reviews Genetics* 2010, 11: 800-805

[12] Yang J, Lee SH, Goddard ME, Visscher PM: GCTA: a tool for Genome-wide Complex Trait Analysis. *American Journal of Human Genetics* 2011, 88(1): 76-82.

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

[14] Khatkar MS, Nicholas FW, Collins AR, Zenger KR, Cavanagh JA, Barris W, Schnabel RD, Taylor JF, Raadsma HW: Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. *BMC Genomics* 2008, 9:187.

[15] Bohmanova J, Sargolzaei M, Schenkel FS: Characteristics of linkage disequilibrium in North American Holsteins. *BMC Genomics* 2010, 11:421.

[16] Uimari P, Tapio M: Extent of linkage disequilibrium and effective population size in Finnish Landrace and Finnish Yorkshire pig breeds. *Journal of Animal Science* 2011, 89(3):609-14.

[17] Corbin LJ, Blott SC, Swinburne JE, Vaudin M, Bishop SC, Woolliams JA: Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Animal Genetics* 2010, 41 Suppl 2:8-15.

[18] Rao YS, Liang Y, Xia MN, Shen X, Du YJ, Luo CG, Nie QH, Zeng H, Zhang XQ: Extent of linkage disequilibrium in wild and domestic chicken populations. *Hereditas* 2008, 145(5):251-7.

[19] McRae AF, McEwan JC, Dodds KG, Wilson T, Crawford AM, Slate J: Linkage disequilibrium in domestic sheep. *Genetics* 2002, 160(3): 1113-1122.

[20] Raadsma HW, International Sheep Genomics Consortium (ISGC): Linkage disequilibrium in the sheep genome: Findings from the ISGC HapMap iniciative. In *Proceedings of the XVIII Plant and Animal Genome (PAG), 9-13 January 2010, San Diego, CA, USA.* Available at URL http: Presentation

[21] Miller JM, Poissant J, Kijas JW, Coltman DW, International Sheep Genomics Consortium: A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. *Molecular Ecology Resources* 2010, 11(2): 314-322.

[22] García-Gámez E, Gutiérrez-Gil B, Sánchez JP, Arranz JJ: Replication and refinement of a QTL influencing milk protein percentage on ovine chromosome 3. *Animal Genetics* 2012, doi: 10.1111/j.1365-2052.2011.02294.x.

[23] Sánchez JP, García-Gámez E, Gutiérrez-Gil B., Arranz JJ: Preliminary evaluation of genomic selection procedures in the Churra dairy population. In *XIV Jornadas sobre Producción Animal 2011, Tomo II, Asociación Interprofesional de Desarrollo Agrario, 16-18 May 2011; Zaragoza (Spain).*

[24] García-Fernández M, Gutiérrez-Gil B, García-Gámez E, Arranz J-J: Genetic variability of the Stearoyl-CoA desaturase gene in sheep. *Molecular and Cellular Probes* 2009, 23: 107–111.

[25] Glowatzki-Mullis ML, Muntwyler J, Gaillard C: Cost-effective parentage verification with 17-plex PCR for goats and 19-plex PCR for sheep. *Animal Genetics* 2006, 38:81-91.

[26] The Ovine Genome Assembly v2.0 [http://www.livestockgenomics.csiro.au/sheep/oar2.0.php]

[27] Ziegler A: Genome-wide association studies: quality control and population-based measures. *Genetic Epidemiology* 2009, 33 Suppl 1:S45-50.

[28] Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT: Data quality control in genetic case-control association studies. *Nature Protocols* 2010, 5(9):1564-73.

[29] Green P, Falls K, Crooks S: Documentation for CRI-MAP, version 2.4. *Washington University School of Medicine*, St Louis 1990.

[30] Weller JI, Kashi Y, Soller M: Power of Daughter and Granddaughter Designs for Determining Linkage Between Marker Loci and Quantitative Trait Loci in Dairy Cattle. *Journal of Dairy Science* 1990, 73(9): 2525-2537

[31] Druet T, Georges M: A hidden markov model combining linkage and linkage disequilibrium information for haplotype reconstruction and quantitative trait locus fine mapping. *Genetics* 2010, 184(3): 789-798.

[32] Browning SR, Browning BL: Rapid and accurate haplotype phasing and missing data inference for whole genome association studies using localized haplotype clustering. *American Journal of Human Genetics* 2007, 81: 1084-1097.

[33] Barrett JC, Fry B, Maller J, Daly MJ: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005, 21(2): 263-265.

[34] Lewontin RC: The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. *Genetics* 1964, 49(1): 49-67.

[35] Hill WG, Robertson A: The effects of inbreeding at loci with heterozygote advantage. *Genetics* 1968, 60(3): 615-628.

[36] Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES: Linkage disequilibrium in the human genome. *Nature* 2001, 411(6834):199-204.

[37] Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D: The structure of haplotype blocks in the human genome. *Science* 2002, 296(5576):2225-2229.

[38] Strandén I, Vuori K: Relax2: pedigree analysis program. In *Proceedings of the 8th World Congress on Genetics Applied to Livestock Production (WCGALP), 13-18 August 2006; Brazil*.

[39] Meuwissen THE, Luo Z: Computing inbreeding coefficients in large populations. *Genetics Selection Evolution* 1992, 24: 305-313.

[40] Usai MG, Sechi T, Salaris S, Cubeddu T, Roggio T, Casu S, Carta A: Analysis of a representative sample of Sarda breed artificial insemination rams with the OvineSNP50 BeadChip. In *Proceedings of 37th Annual Meeting of International Committee for Animal Recording (ICAR), 31 May - 4 June 2010; Riga, Latvia.*

[41] de Roos AP, Hayes BJ, Spelman RJ, Goddard ME: Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics* 2008, 179(3):1503-12.

[42] Villa-Angulo R, Matukumalli LK, Gill CA, Choi J, Van Tassell CP, Grefenstette JJ: High-resolution haplotype block structure in the cattle genome. *BMC Genetics* 2009, 10:19.

[43] Kim ES, Kirkpatrick BW: Linkage disequilibrium in the North American Holstein population. *Animal Genetics* 2009, 40:279-288.

[44] Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, San Cristobal M, Servin B, McCulloch R, Whan V, Gietzen K, Paiva S, Barendse W, Ciani E, Raadsma H, McEwan J, Dalrymple B and other members of the International Sheep Genomics Consortium: Genome Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. *PLoS Biology* 2012, 10(2):e1001258.

[45] Han J, Kraft P, Nan H, Guo Q, Chen C, Qureshi A, Hankinson SE, Hu FB, Duffy DL, Zhao ZZ, Martin NG, Montgomery GW, Hayward NK, Thomas G, Hoover RN, Chanock S, Hunter DJ: A Genome-Wide Association Study Identifies Novel Alleles Associated with Hair Color and Skin Pigmentation. *PLoS Genetics* 2008, 4(5): e1000074.

[46] Sánchez Belda A, Sánchez Trujillano MC: *Razas Ovinas Españolas.* Publicaciones de Extensión Agraria: Ministerio de Agricultura, Pesca y Alimentación; 1986.

[47] Meuwissen T: Genetic management of small populations: A review. *Acta Agriculturae Scandinavica, Section A - Animal Science* 2009, 59: 71-79.

[48] McKay SD, Schnabel RD, Murdoch BM, Matukumalli LK, Aerts J, Coppieters W, Crews D, Dias Neto E, Gill CA, Gao C, Mannen H, Stothard P, Wang Z, Van Tassell CP, Williams JL, Taylor JF, Moore SS: Whole genome linkage disequilibrium maps in cattle. *BMC Genetics* 2007, 8: 74.

[49] Li MH, Strandén I, Kantanen J: Genetic diversity and pedigree analysis of the Finnsheep breed. *Journal of Animal Science* 2009, 87:1598-1605.

TABLES

Table 1 – A summary of statistics for the SNPs that passed quality control (QC).

Chromosome	NumSNP	Average Distance between SNPs (kb)	Average MAF	Average Heterozygosity	Rate cM/Mb	Average D'	Average r^2	Average r^2 (<100Kb)
1	4,987	55.38	0.287	0.352	1.55	0.127	0.006	0.151
2	4,676	53.45	0.284	0.345	1.5	0.138	0.008	0.171
3	4,164	53.78	0.288	0.351	1.64	0.132	0.008	0.164
4	2,246	52.76	0.287	0.352	1.75	0.151	0.010	0.163
5	1,978	54.26	0.290	0.349	1.71	0.161	0.012	0.157
6	2,190	53.22	0.291	0.354	2.04	0.166	0.013	0.166
7	1,887	52.84	0.293	0.354	1.72	0.144	0.009	0.139
8	1,717	52.91	0.282	0.346	1.62	0.151	0.011	0.170
9	1,836	51.47	0.283	0.346	1.71	0.148	0.009	0.150
10	1,523	54.86	0.294	0.356	1.61	0.155	0.012	0.180
11	963	64.12	0.292	0.355	2.12	0.154	0.012	0.151
12	1,456	54.36	0.296	0.358	1.64	0.148	0.011	0.145
13	1,402	59.14	0.283	0.347	1.85	0.150	0.009	0.155
14	959	64.89	0.277	0.344	2.15	0.162	0.010	0.142
15	1,374	58.59	0.291	0.356	1.54	0.139	0.010	0.163
16	1,312	54.33	0.282	0.346	1.86	0.158	0.011	0.144
17	1,178	61.54	0.291	0.354	1.88	0.158	0.011	0.160
18	1,192	56.35	0.292	0.358	1.83	0.152	0.010	0.146
19	1,032	58.76	0.279	0.344	2.05	0.158	0.009	0.147
20	910	55.46	0.301	0.361	2.37	0.170	0.015	0.145
21	724	66.68	0.277	0.339	1.94	0.164	0.010	0.138
22	942	53.60	0.284	0.353	2.06	0.158	0.010	0.126
23	934	66.93	0.296	0.358	1.82	0.167	0.013	0.149
24	598	69.71	0.303	0.366	2.22	0.152	0.013	0.153
25	846	52.05	0.295	0.355	2.02	0.173	0.014	0.134
26	758	57.91	0.281	0.347	2.05	0.169	0.014	0.148

Table 2 – Mean linkage disequilibrium among syntenic and nonsyntenic SNPs over different map distances. Average values for *D'* and r^2 pooled over autosomes in different categories; minimum, maximum and standard deviation are shown in this table. The same parameters for non-syntenic SNP pairs are also displayed.

		D'				r2			
Distance	NumPairs	Average	SD	Min	Max	Average	SD	Min	Max
< 10 Kb	1,814	0.763	0.312	0.001	1	0.329	0.325	0	1
10-20 Kb	4,218	0.694	0.322	0.001	1	0.256	0.282	0	1
20-40 Kb	13,963	0.601	0.329	0	1	0.191	0.240	0	1
40-60 Kb	16,364	0.543	0.323	0	1	0.152	0.207	0	1
60-100 Kb	32,426	0.487	0.312	0	1	0.120	0.175	0	1
100-200 Kb	80,273	0.422	0.288	0	1	0.086	0.133	0	1
200-500 Kb	239,021	0.364	0.261	0	1	0.061	0.098	0	1
500 Kb-1 Mb	395,047	0.328	0.243	0	1	0.049	0.078	0	0.987
1-2 Mb	781,115	0.301	0.228	0	1	0.040	0.066	0	0.898
2-5 Mb	2,288,329	0.255	0.203	0	1	0.028	0.048	0	0.953
5-10 Mb	3,644,862	0.203	0.170	0	1	0.017	0.030	0	0.747
10-20 Mb	6,661,473	0.158	0.140	0	1	0.009	0.016	0	0.661
20-50 Mb	15,118,487	0.120	0.114	0	1	0.005	0.008	0	0.530
> 50 Mb	13,103,982	0.106	0.106	0	1	0.003	0.006	0	0.229
Non-syntenic	2,257,088	0.107	0.106	0	1	0.004	0.006	0	0.169

Table 3 – Block structure per chromosome. The number of blocks, total, minimum and maximum block length, percentage of the sequence covered by blocks, number and percentage of SNPs in blocks on a per chromosome basis are reported in this table. Values obtained across the genome are also shown.

Chromosome	Number of blocks	Total block length (Kb)	% of Chromosome length in blocks	Min. block length (Kb)	Max. block length (Kb)	Number of SNPs in blocks	% of SNPs in blocks
1	257	5,440.86	1.97	1.74	265.59	569	11.41
2	252	11,148.79	4.46	3.92	1,262.79	634	13.56
3	200	5,321.47	2.38	2.79	385.36	464	11.14
4	103	3,273.46	2.76	2.89	228.94	244	10.86
5	86	2,023.86	1.88	0.07	325.55	191	9.66
6	115	2,705.20	2.31	3.72	225.92	257	11.74
7	95	1,753.56	1.75	4.06	158.77	204	10.81
8	88	2,256.32	2.49	2.62	262.30	198	11.53
9	109	2,886.37	3.04	3.66	472.64	249	13.56
10	79	5,231.93	6.26	2.46	1,075.88	210	13.79
11	41	779.45	1.25	5.76	177.74	88	9.14
12	80	1,154.64	1.46	1.95	182.06	165	11.33
13	71	1,725.72	2.08	3.75	175.94	161	11.48
14	34	648.62	1.04	0.04	120.50	76	7.92
15	62	1,716.25	2.12	2.13	248.80	140	10.19
16	52	980.30	1.37	2.46	134.49	113	8.61
17	42	1,225.16	1.69	2.99	231.71	103	8.74
18	50	854.06	1.25	5.08	152.75	107	8.98
19	51	978.44	1.61	5.73	316.69	105	10.17
20	33	660.86	1.31	6.33	155.29	72	7.91
21	34	521.40	1.08	2.84	116.56	73	10.08
22	48	870.01	1.72	2.63	200.82	104	11.04
23	39	1,009.87	1.61	6.71	328.76	85	9.10
24	19	453.21	1.09	4.38	233.82	40	6.69
25	37	738.85	1.68	6.76	205.07	82	9.69
26	22	366.91	0.83	0.92	128.69	46	6.07
All	2,099	56,725.56	2.32	0.04	1,262.79	4,780	10.92

FIGURAS

Figure 1 – The distribution of SNPs across the genome as a function of the minor allele frequency (MAF). The percentage of SNPs that passed quality control (QC) for different MAF bins is depicted here.

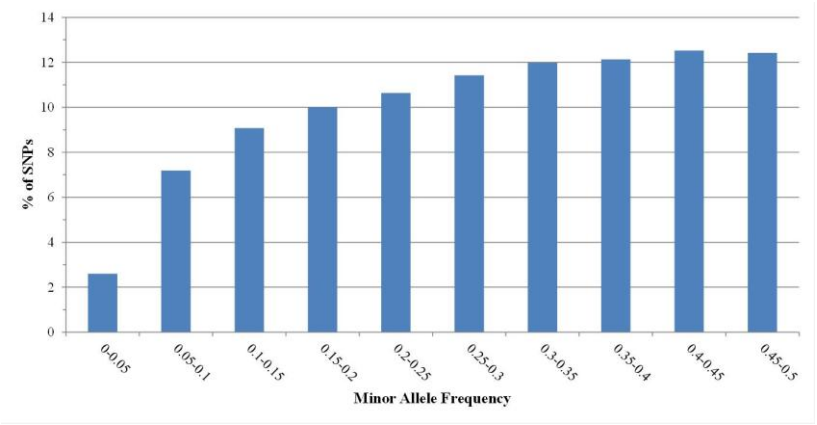


Figure 2 – Linkage disequilibrium (LD) across the genome as a function of genomic distance between markers is represented here for Spanish Churra sheep. **A**) Average LD, measured using two parameters, *D'* and r^2. SNP pairs were stacked according to their physical distance into 14 categories: < 10 kb, 10-20 kb, 20-40 kb, 40-60 kb, 60-100 kb, 200-500 kb, 0.5-1 Mb, 1-2 Mb, 2-5 Mb, 5-10 Mb, 10-20 Mb, 20-50 Mb or > 50 Mb; each point in the graph corresponds to one of these bins. **B**) Average LD (r^2) binning SNPs into 1 Mb-interval categories; half-length LD (r^2 = 0.033) approximate distance 2.5 Mb.

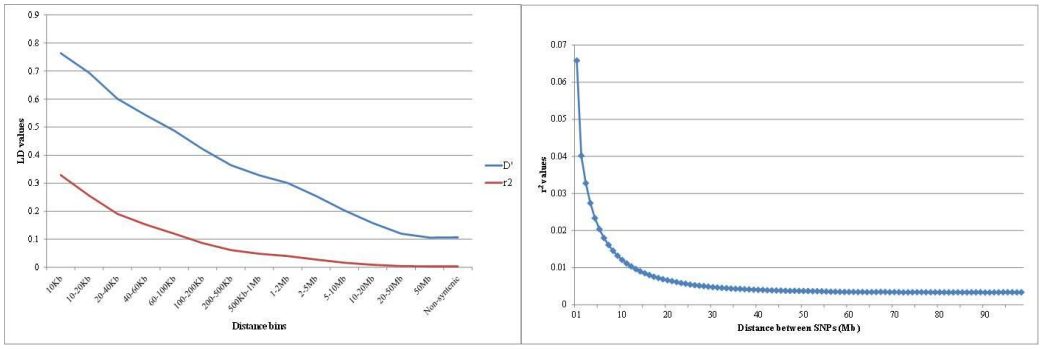


Figure 3 – Past effective population size (N_e) over the past generations based on linkage disequilibrium calculations from 26 autosomes. We estimated N_e from the average r^2 at different marker distances using three different cM/Mb conversion rates: a) 1 cM ~ 1 Mb (red), b) 1.85 cM ~ 1 Mb (blue) or c) individual rates for each chromosome, as indicated in Table 1 (green). Data points were based on at least 1,000 marker pairs.

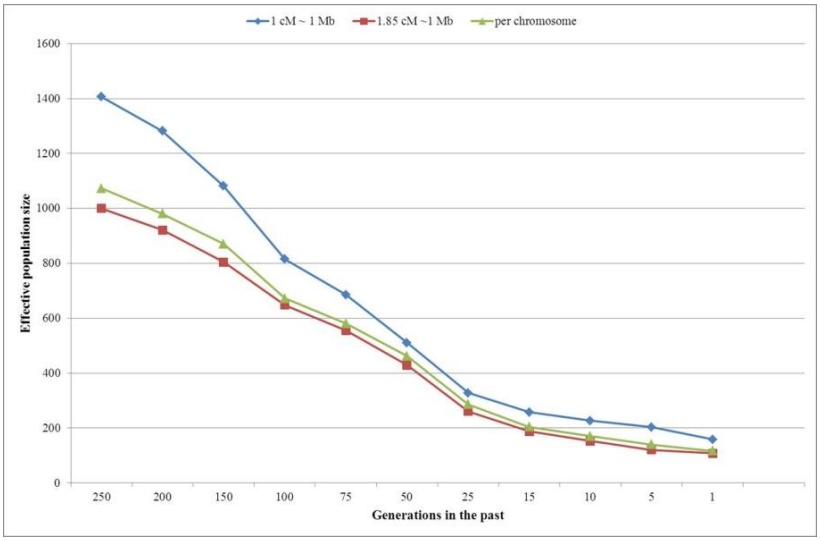
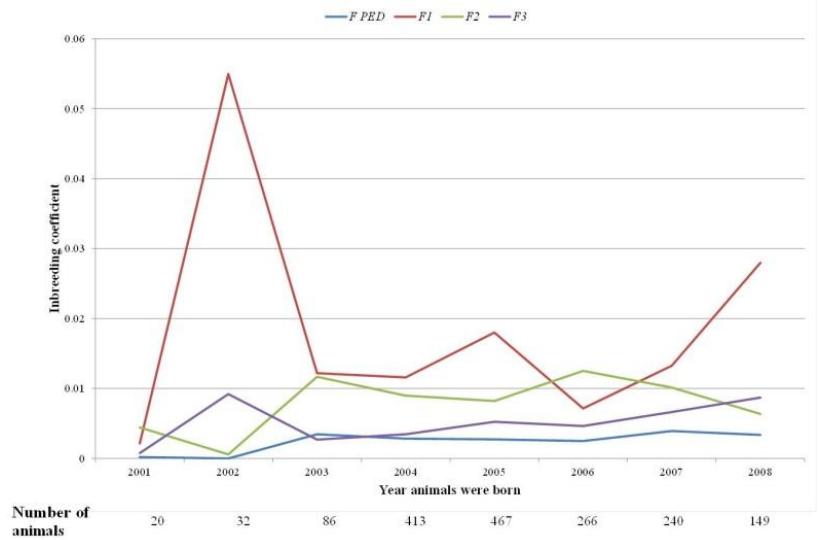


Figure 4 – Average inbreeding coefficients for animals born from 2001 to 2008. Average values of the positive inbreeding coefficients calculated using pedigree (F_{PED}) or marker information (F_1, F_2, F_3) are represented in this figure. The three marker-based metrics represented here are based on the variance of the additive genotype (F_1), the excess of homozygosity (F_2) and the correlation between uniting gametes (F_3). The number of animals used in each time point is listed below the X-axis.



GWA ANALYSIS FOR MILK PRODUCTION TRAITS IN DAIRY SHEEP AND GENETIC SUPPORT FOR A QTN INFLUENCING MILK PROTEIN PERCENTAGE IN THE *LALBA* GENE

Elsa García-Gámez[1], Beatriz Gutiérrez-Gil[1], Goutam Sahana[2], Juan-Pablo Sánchez[3], Yolanda Bayón[1], Juan-José Arranz[1*]

* Corresponding author: Juan-José Arranz: jjarrs@unileon.es

[1] Dpto. Producción Animal, Universidad de León, 24071, León, Spain. [2] Department of Molecular Biology and Genetics, Aarhus University, Aarhus, Denmark. [3] Mejora Genética Animal, IRTA, 25198, Lleida, Spain

PLoS ONE, *In press*, doi: 10.1371/journal.pone.0047782.

ABSTRACT

In this study, we used the *Illumina OvineSNP50 BeadChip* to conduct a genome-wide association (GWA) analysis for milk production traits in dairy sheep by analyzing a commercial population of Spanish Churra sheep. The studied population consisted of a total of 1,681 Churra ewes belonging to 16 half-sib families with available records for milk yield (MY), milk protein and fat yields (PY and FY) and milk protein and fat contents (PP and FP). The most significant association identified reached experiment-wise significance for PP and FP and was located on chromosome 3 (OAR3). These results confirm the population-level segregation of a previously reported QTL affecting PP and suggest that this QTL has a significant pleiotropic effect on FP. Further associations were detected at the chromosome-wise significance level on 14 other chromosomal regions. The marker on OAR3 showing the highest significant association was located at the third intron of the *alpha-lactalbumin (LALBA)* gene, which is a functional and positional candidate underlying this association. Sequencing this gene in the 16 Churra rams of the studied resource population identified additional polymorphisms. One out of the 31 polymorphisms identified was located within the coding gene sequence (*LALBA*_g.242T>C) and was predicted to cause an amino acid change in the protein (Val27Ala). Different approaches, including GWA analysis, a combined linkage and linkage disequilibrium study and a concordance test with the QTL segregating status of the sires, were utilized to assess the role of this mutation as a putative QTN for the genetic effects detected on OAR3. Our results strongly support the polymorphism *LALBA*_g.242T>C as the most likely causal mutation of the studied OAR3 QTL affecting PP and FP, although we cannot rule out the possibility that this SNP is in perfect linkage disequilibrium with the true causal polymorphism.

Keywords: Dairy sheep, GWAS, protein content, *LALBA*.

INTRODUCTION

Over the last five years, high-throughput SNP technologies have provided the opportunity to explore the genomes of livestock species to identify regions influencing traits of economic interest. Genome-wide association (GWA) studies are currently replacing traditional QTL linkage mapping analyses as more powerful gene detection tools. In cattle, many GWA studies based on SNP chips have been performed so far for milk production

[1,2,3], milk protein composition [4], milk fatty acid composition [5], growth [6,7] and reproductive traits [3,8,9,10,11]. Moreover, in sheep, the *Illumina OvineSNP50 BeadChip* has previously been used to accomplish medium-density marker genome scans to study categorical or disease-like traits and, in some cases, to identify candidate causative mutations for the traits under examination [12,13,14].

However, to date, no GWA analysis related to quantitative traits of economic interest in sheep has been reported, and only the results of classical QTL analysis based on microsatellite markers are available (http://www.animalgenome.org/cgi-bin/QTLdb/OA/index). Despite the close phylogenetic relationship between sheep and cattle, the few QTN previously identified in dairy cattle [15,16,17] appear not to have similar effects in dairy sheep, as indicated by analyses performed in Spanish Churra sheep [18]. In dairy cattle, genomic selection is now being successfully implemented based on the existing linkage disequilibrium (LD) between markers and QTL but without relying on the identification of the true causal mutations or QTN [19]. However, the identification of QTN could help to avoid some important limitations of the genomic selection approach, such as the need to genotype large numbers of animals and repeat genotyping after several generations [20]. Gene-assisted selection could be of special interest for applying genomic selection in sheep, in which the limited size of the populations hampers the establishment of training populations and the estimation of marker-QTL effects used to predict breeding values.

In dairy sheep, microsatellite-based whole genome-scans have identified several QTL influencing milk-related traits [21,22,23,24,25]. In Spanish Churra sheep, only one of the QTL identified for milk traits reached genome-wide significance. This QTL was located on chromosome 3 (OAR3) and influenced milk protein percentage [23]. After an initial increase of microsatellite marker density, some candidate genes, such as *HDAC7* (*Histone deacetylase 7*), *VDR* (*Vitamin D receptor*), *IGFBP6* (*insulin-like growth factor binding protein 6*) and *LALBA* (*alpha-lactalbumin*), were identified within the redefined 13-cM length confidence interval (CI) estimated for this QTL region [26].

In the present work, the *Illumina OvineSNP50 BeadChip* was used to perform a higher density GWA analysis to identify genomic regions influencing milk production traits in Churra sheep. The objectives of this study were, first, to identify, at the population level, novel regions associated with milk production traits that had not previously been identified

because of the lower resolution of microsatellite-based scans and/or the limitations of outbred linkage-based studies and, second, to replicate and fine-map QTL previously identified in Churra sheep.

MATERIALS AND METHODS

Resource population and phenotypic data

Blood samples from 1,696 Spanish Churra ewes were collected in order to extract DNA. The ewes were distributed in 16 half-sib families with an average size of 105 daughters per ram (ranging from 29 to 277 animals per half-sib family). The use of animals was performed in compliance with the guidelines approved by the University of Leon ethical commission.

The phenotypes included in the analysis were milk yield (MY), protein yield (PY), fat yield (FY), protein percentage (PP) and fat percentage (FP), which are collected routinely by the National Association of Churra Breeders (ANCHE) through the official milk recording process. The dependent variables used in the association analysis were the yield deviations (YDs) corresponding to the studied traits. The YDs were estimated as averages of the ewes' raw phenotypic records corrected for fixed environmental effects and the common environmental effect [27]. The calculation of the YDs was performed using multivariate animal repeatability models. The fixed effects used to calculate these YDs were the Herd-Test-Day effect, the birth order, the age of the ewe at parturition (as a covariate nested within birth order), the number of born lambs and the number of weeks of milk production of the ewe.

SNP genotyping, quality control and genetic maps

Genotyping for the *Illumina OvineSNP50 BeadChip* was performed commercially at AROS Applied Biotechnology AS (Aarhus, Denmark) and LABOGENA (Jouy-en-Josas, France). We applied the quality control (QC) criteria previously described by [28] on the raw genotypes: i) GenCall score for raw genotypes > 0.15; ii) known location of the marker in the ovine autosomes; iii) call rate per individual > 0.9; iv) call rate per SNP ≥ 0.95; v) minor allele frequency (MAF) ≥ 0.05; vi) correspondence with Hardy-Weinberg equilibrium (HWE) *P*-value > 0.00001. After applying these QC criteria, the VerifTyp software was used to

impute missing genotypes using pedigree and population information (Boichard D and Druet T, personal communication).

Initially, marker order and positions were based on the Ovine Genome Assembly v2.0 (http://www.livestockgenomics.csiro.au/sheep/oar2.0.php), assuming a conversion ratio of physical to genetic distances of 1 cM to 1 Mb. Next, a control of this initial genetic map was performed using a modified version of CRI-MAP [29], v2.503 (kindly provided by J. F. Maddox). The final genetic map used in the GWA analysis reported here was based on the physical map provided by [28], assuming the conversion ratio indicated above.

Genome-wide association analysis

For the GWA analysis, the following linear mixed model which includes the polygenic effect as a random effect and the genotypes at single SNP markers as fixed effects was applied to the data:

$$\boldsymbol{y} = \boldsymbol{Zu} + \boldsymbol{xb} + \boldsymbol{e} ,$$

where $\boldsymbol{y}$ is the vector of phenotypes (YDs) of the animal; $\boldsymbol{Z}$ is a matrix associating random additive polygenic effects to individuals; $\boldsymbol{u}$ is a vector containing the random polygenic effects; $\boldsymbol{x}$ is a vector with genotypic indicator (-1, 0, or 1) associating records to the marker effect; $\boldsymbol{b}$ is the allele substitution effect determined by the SNP genotype; and $\boldsymbol{e}$ is the random residual. The random variables $\boldsymbol{u}$ and $\boldsymbol{e}$ are assumed to be normally distributed. Specifically, $\boldsymbol{u}$ is normally distributed with $(\mathbf{0}, \boldsymbol{\sigma_g^2}\mathbf{A})$, where $\boldsymbol{\sigma_g^2}$ is the polygenic genetic variance and $\mathbf{A}$ is the additive relationship matrix derived from the pedigree. This association analysis was implemented by restricted maximum likelihood (REML) using the DMU software package (available at http://dmu.agrsci.dk), and the marker effect was tested using a Wald test against a null hypothesis of $\boldsymbol{b} = 0$.

In a later stage, to examine whether the two most significant SNPs were individually able to explain the QTL effect observed on OAR3 for PP and FP, the above linear mixed model was repeated with the addition of the genotypes of these 'top' SNPs as fixed effects in the model, and the analysis was repeated for the rest of the SNPs in this chromosome.

Significance thresholds for the GWA analyses were estimated using a Bonferroni correction for multiple testing. As not all of the markers tested for association were

independent because of LD, the Bonferroni correction for the total number of markers in the study is conservative. Therefore, based on the method described by [30], we calculated the number of independently analyzed markers tested for each chromosome and for the entire genome. For the whole genome, the total number of independent SNPs analyzed was 26,965. Hence, the 5 % genome-wise significance level corrected for multiple testing corresponded to a nominal *P*-value of 1.85×10^{-6}. Finally, the 5 % experiment-wise significance threshold was set by accounting for the three independent traits analyzed, which were determined by a principal component analysis performed in R (R Development Core Team, 2008). For simplicity, the statistical significance values given in the text refer to nominal *P*-values.

Candidate gene sequencing analysis

Sequence analysis was performed for the *LALBA* gene, which was identified as a strong candidate gene for the most significant associations identified in this study. Six primer pairs were designed for the sequence analysis of this gene in the 16 rams of the studied resource population based on the available reference sequences (GenBank Accession No. AB052168; http://www.livestockgenomics.csiro.au/cgi-bin/gbrowse/oarv2.0/). Following the sequencing protocol described by [31], we obtained the forward and reverse strand sequences of the complete *LALBA* gene (4 exons and 3 introns; 1,741 bp), a fraction of the promoter (656 bp) and the 3'UTR region (670 bp) (see Table S1 for the primer sequences and the amplification annealing temperatures for each amplicon). DNA sequence variants (DSVs) in the 16 Churra rams' sequences were identified using the SeqScape v2.5 software (Applied Biosystems, Foster City, CA). Among all the DSVs identified in this analysis, one of the SNPs located in Exon 1 of the *LALBA* gene (*LALBA*_g.242T>C) was genotyped across the entire population. This genotyping was performed by KBioScience using the fluorescence-based competitive allele-specific PCR (KASPar) assay (for details, see http://www.kbioscience.co.uk).

Combined linkage and linkage disequilibrium analysis (LDLA) in OAR3

For the chromosome showing the most significant associations in this study, OAR3, we performed a combined LDLA with the aim of refining the CI of the PP QTL detected on that chromosome [32]. For this purpose, all of the markers on this chromosome from the *Illumina OvineSNP50 BeadChip* and the commercially genotyped *LALBA* SNP were used to perform an LDLA using the QTLmap software [33]. The analysis was performed at every 0.1

cM. The significance thresholds were estimated through a total of 1,000 simulations. To estimate the CIs, the likelihood ratio test statistic (LRT) values were converted into LOD scores, and the LOD drop-off method, as described in [26], was used.

Concordance test performed on the QTL heterozygous sires

To assess the possible causality of the *LALBA* gene, all SNPs mapping on OAR3 were phased for the entire population using the PHASEBOOK software [34], following the protocol described by [28]. To identify the heterozygous sires for the OAR3 QTL, we performed a classical regression-based linkage analysis (LA) for the OAR3 markers included in the interval 70–200 Mb using the software GridQTL [35]. Following [36], we performed a fixed position analysis at the closest position to the *LALBA* gene (137 Mb). The families showing a significant QTL effect estimated at the 5 % nominal level (ABS(t) > 1.96) were considered to be heterozygous (*Qq*) at the QTL. The two phases of the segregating sires were assigned to one of the QTL alleles (*Q* and *q*) according to the sign of the estimated effect for the sire and the daughter's IBD status as obtained from GridQTL [35]. For the mutations showing concordance with the sires' QTL status, we then estimated the probability of concordance by chance between the QTL status and the SNP genotype. This probability was calculated as $pA^{(NQ)} \times pB^{(Nq)}$, where *pA* and *pB* are the allele frequencies of the SNP associated with *Q* and *q* QTL allele, respectively, and *NQ* and *Nq* are the number of concordant *Q* and *q* chromosomes [37].

RESULTS

After the QC performed on the raw genotypes, a total of 43,784 SNPs distributed on the 26 ovine autosomes were included in the GWA analysis. Descriptive data for the SNP chip, including the number of SNPs per chromosome, the distance between them, the average MAF and heterozygosity and LD calculations, have been previously reported by [28]. Here, we present the significant associations identified through the GWA analysis for the five milk production traits included in the study (MY, PP, FP, PY and FY). A summary of the discovered experiment- and chromosome-wise significant associations is presented in Table 1. Figure 1 depicts Manhattan plots across the whole genome for the five analyzed traits, with SNP associations represented as log(1/*P*-value) on the y-axis.

Experiment-wise significant associations

The most significant results from the GWA analysis were obtained for PP on OAR3. The significantly associated SNPs covered a 10.5 Mb-long region of the chromosome (from 130.0 to 140.5 Mbp), but the highest LRT values (*P*-value < $1.0x10^{-10}$) were obtained within a 30 Kb interval (from 137.29 to 137.32 Mbp). The SNP showing the most significant association was OAR3_147028849 (*P*-value = $3.78x10^{-26}$), which was located in the third intron of the *LALBA* gene. The same region of OAR3 also showed an experiment-wise significant association with FP (*P*-value = $1.8x10^{-10}$), in which marker OAR3_147028849 was also identified as the top SNP for this trait. Significant results for FP covered a narrower region (4.3 Mb; from 133.0 to 137.3 Mb) than for PP. The allelic substitution effect of this SNP was positive for both traits, although the magnitude was larger for PP (0.47 SD) than for FP (0.30 SD).

Chromosome-wise significant associations

SNPs on 14 chromosomes showed significant association with the studied traits at the 5 % chromosome-wise level. These chromosomes were OAR1, OAR2, OAR3, OAR4, OAR6, OAR7, OAR12, OAR14, OAR15, OAR16, OAR17, OAR20, OAR23 and OAR25. A summary of the significant SNPs associated with the studied traits is presented in Table 1. In general, most of the chromosome-wise significant results were single-point associated SNPs with the exception of a 20-Mb long region on OAR2, which showed significant associations with the three "yield" traits (MY, PY and FY). Significant SNPs in this linkage group were concentrated between 42.7 and 63.2 Mb, although most of them were located in the middle of that region, between 53.0 and 58.0 Mb. The central part of OAR20 exhibited significant associations with more than one trait, although there were also single point associations. At position 23.7 Mb, there was one marker significantly associated with FY, whereas in the region between 29.1 and 29.6 Mb, chromosome-wise significant associations were detected for PP and FP.

Candidate gene sequencing

The sequencing analysis was performed on the 16 rams for the LALBA gene, including all exons, introns, the promoter region and 3'UTR of the *LALBA* gene. A total of 3,067 bp were sequenced. The number of DSVs found in the sequenced region was 31, on average one polymorphism at every 98.9 bp. Two of the polymorphisms were one-base indel mutations, and 29 were single-base substitutions (Table S2). Only one of the SNPs identified

was located within a translated region of the gene, in Exon 1 (*LALBA*_g.242T>C; GenBank Accession No. AB052168). This mutation was predicted to cause an amino acid substitution of valine to alanine at position 27 of the protein sequence (Val27Ala).). The SNPs located in the non-coding region were distributed as follows, 13 mutations were located along the gene introns, including the two indels, whereas eight SNPs were found in the promoter and nine other SNPs in the 3'UTR, respectively.

Combined linkage disequilibrium and linkage analysis

The results from the LDLA performed in OAR3 for PP and FP are shown in Figure 2. For PP, the most significant result was obtained at 137.3 Mb (LRT = 140.2). According to the significance thresholds set through simulation (LRT = 78.5 for a chromosome-wise *P*-value of 0.0001), the analysis revealed a highly significant QTL located in the middle of a haplotype of 4 SNPs that includes the *LALBA* gene. The order and names of the markers included in this haplotype were *LALBA*_g.242T>C, OAR3_147028849, OAR3_147128672 and OAR3_147275963. Using the LOD drop-off method, the CI covered a 4 Kb-long region from 137.2 Mb to 137.6 Mb in OAR3.

The LDLA results for FP reached chromosome-wise significance, with the maximum LRT value obtained at 134.5 Mb (LRT = 84.6; *P*-value < 0.001). The LRT for FP at the location of the PP QTL (137.2 Mb), i.e., the location of LALBA gene, was 76.7 (*P*-value < 0.001).

Genetic support for the alpha-lactalbumin (LALBA) gene

The association analysis was repeated for OAR3 incorporating the genotyped SNP located in the *LALBA* exon, *LALBA*_g.242T>C, and the results for PP and FP are presented in Figure 3A. For both of the traits, the SNP *LALBA*_g.242T>C showed the most significant association (*P*-value for PP = $2.34x10^{-30}$; *P*-value for FP = $1.62x10^{-12}$). The allele substitution effect was higher for PP (0.46 SD) than for FP (0.3 SD). The second most significantly associated SNP for both PP and FP was also located in the intronic region of the *LALBA* gene that was identified as the top SNP in the initial GWA analysis, OAR3_147028849.

To test the ability of these two SNPs to explain variance in the studied traits, we individually included them in the mixed model equation as fixed factors. When we fixed marker *LALBA*_g.242T>C in the analysis, the significant associations previously identified on

OAR3 for PP and FP disappeared (Figures 3C and 3D). Significant associations were also removed when marker OAR3_147028849 was fixed in the FP analysis (Figure 3F). However, when this marker was fixed in the PP mixed model equation significant effects for markers *LALBA*_g.242T>C and OAR3_146832060 still remained (P values = $3.17x10^{-7}$ and $1.85x10^{-6}$, respectively). This latter SNP is located 23 Kb upstream from *LALBA* gene (Figure 3E). Using the information provided by the QTLMap LDLA approach, we examined the estimated haplotype effects for PP at the maximum LRT position. We observed that all haplotypes with a positive effect on PP were carrying the *LALBA*_g.242C allele at the *LALBA*_g.242T>C locus, whereas all haplotypes containing the *LALBA*_g.242T allele at that position showed a negative haplotype effect on PP (Table S3).

The LA performed at the fixed position of 137 Mb, the closest position to the *LALBA* gene, showed that three of the sires were segregating for the QTL (ABS(t) > 1.96). The concordance test performed using the phases estimated with PHASEBOOK showed that the three sires found to be segregating for the QTL in the LA analysis were heterozygous at the *LALBA*_g.242T>C locus but not at the OAR3_147028849 SNP. We assigned each of the sire's phases to the *Q* or *q* allele of the QTL according to the daughter's IBD and the QTL effect. All three sires' phases carrying the *Q* allele shared the *LALBA*_g.242C allele (Figure S1). Taking into account allelic frequencies at the *LALBA*_g.242T>C locus in the Churra ewes population (*LALBA*_g.242C allele = 0.71 and *LALBA*_g.242T allele = 0.29), the probability of this concordance by chance was estimated to be 0.0087. Within the haplotype block interval considered, from 136.8 to 138.1 Mb, there was another polymorphism, s19950, which showed concordance with the QTL status in the three segregating sires. On the basis of the marker allelic frequencies, the probability of the concordance by chance was 0.0155. In the initial GWA scan for the PP trait, the marker s19950 ranked 9th among the most significant markers (*P*-value = $1.3x10^{-9}$). The inclusion of this SNP as fixed effect in the model nonetheless yielded highly significant results on OAR3 for both PP (*P*-value = $3.35x10^{-22}$) and FP (*P*-value = $8.94x10^{-9}$). Additionally, among the sires that did not reach statistical significance to be considered segregating, there were four sires heterozygous for the *LALBA*_g.242T>C mutation, all of which shared a *LALBA*_g.242C allele at the parental haplotype associated with increased PP. The remaining nine sires of the population were homozygous for the *LALBA*_g.242C allele.

The concordance analysis was also performed for FP. Three sires segregated for the FP QTL at the *LALBA* gene (ABS(t) > 1.96). Two of them were coincident with the PP segregating sires, whereas the third, which was homozygous for the *LALBA*_g.242T>C polymorphism, was not segregating for the PP QTL.

DISCUSSION

In livestock species, GWA analyses have become a powerful strategy to identify DSVs affecting phenotypic variation. This study is, to our knowledge, the first reported GWA study for milk production traits in sheep using a high-throughput SNP array. The analyses reported here provide strong evidence for a highly significant region associated with PP and FP on OAR3 that was previously suggested in classical linkage analyses. Additionally, 14 other chromosome-wise significant associations were detected by the GWA study reported herein.

Among the QTL detected at the chromosome-wise significance level, two regions on OAR2 and OAR20 showed significant associations for more than one trait. In OAR2, we found significant associations for three of the traits under study: MY, PY and FY. Additional within-family linkage analyses showed the same sign for the estimated effects of these QTL, thereby supporting the existence of a pleiotropic QTL that increases milk yield without affecting milk composition. The LOD-drop off CI estimated by our analyses clearly located the QTL in the first third of OAR2 (42.7 to 63.2 Mb). QTLs influencing PP and FP have previously been described in the proximal end of OAR2 in Spanish Churra through a microsatellite-based genome-scan [23], although the estimated CIs from that study spanned most of the chromosome length. The orthologous bovine region corresponding to this OAR2 QTL, on BTA8, has also been associated with milk PY in dairy cattle [38]. According to the Ovine Genome Assembly v2.0, some positional candidate genes located close to the highest associations found in OAR2 (55.39 Mb), within the 48-60 Mb interval, are *TGFBR1* (transforming growth factor C beta receptor 1), *IGFBPL1* (Insulin-like growth factor binding protein-like 1), *CD72* (Cytokine 72), *STOML2* (stomatin (EPB72)-like 2), *GalNAc-T12* (UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 12), *TLN1* (talin 1), *PSAT1* (phosphoserine aminotransferase 1), *GCNT1* (glucosaminyl (N-acetyl) transferase 1 C core 2) and *RFK* (riboflavin kinase). Among them, the *IGFBPL1* gene might be considered a possible functional candidate because its coding protein has a domain that

resembles one found in the insulin-like growth factor binding proteins (IGFBPs) that is bound by insulin-like growth factors (IGFs) [39]. These IGFs have been suggested to be associated with milk performance in cattle [40].

Our analysis also detected a region on OAR20 that was associated at the chromosome-wise significance level with three of the studied traits: PP, FP and FY (at 23.7-29.6 Mb). This genomic region includes the major histocompatibility complex class-II genes (MHC-II). Notably, ovine QTL for FP [23] and for MY, FY, PY and FP [21] have been previously identified around the ovine MHC-II region in Spanish Churra sheep and a Sarda x Lacaune backcross population, respectively. Moreover, in the orthologous region of BTA23, QTLs influencing milk production traits have been described [41], with at least one gene located in the MHC-II region (*BoLA-DRB3*) showing a reported association with milk dairy traits [42].

The rest of the chromosome-wise significant associations reported in this paper were, in general, determined by a single, isolated SNP. Although some of these isolated associations could be spurious results or inaccurate chromosomal location mappings resulting from the draft stage of the ovine assembly, it is also possible that they indicate genuine genotype-phenotype associations that would require a more powerful design to be identified. This latter hypothesis is supported, for example, by the observation that the SNP associated with PP at the distal region of OAR6 (DU430803_572, 85.44 Mb) is located within the casein gene cluster (*CSN1S1*, *CSN1S2* and *CSN2*), which encodes the major proteins in sheep milk that determine the technological properties of milk during cheese manufacturing [4]. Genetic variants in these genes have been shown to influence milk casein content [43]. This region on OAR6 has been previously suggested to be associated with milk composition traits in sheep [21,43,44]. Another potential functional candidate identified for the chromosome-wise significant associations was found for the FY associated OAR25 peak, which is located in the region surrounding the *NRG3* gene (*Neuregulin 3*). In humans, this gene has been reported to be involved in mammary gland development [45].

The most significant association identified by our GWA analysis has previously been identified in Spanish Churra sheep through classical linkage analyses [23]. Using increased microsatellite markers and a LDLA approach, [26] replicated and redefined the QTL in the same population and postulated the *alpha-lactalbumin* gene (*LALBA*) as a strong functional and positional candidate affecting the traits. Alpha-lactalbumin is a major whey protein that

forms a subunit of the lactose synthase binary complex. Because lactose synthase is necessary for the production of lactose and the subsequent movement of water into the mammary secretory vesicles, this enzyme is critical in the lactational control and secretion of milk. Previous studies in *LALBA*-deficient mice have shown the influence of this enzyme on the protein and fat concentration in milk. Homozygous mutant mice produce highly viscous milk that is very rich in fat and protein and devoid of alpha-lactalbumin and lactose [46]. Polymorphisms in the *LALBA* gene were studied in the 1990s as possible markers related to milk production in dairy species. In cattle, one SNP located in the promoter of this gene (ULGR_SNP_U63109_1966) has been found to influence milk traits [4,47]. This polymorphism was one of the SNPs identified in our population (*LALBA*_g.82G>A), but it was homozygous for all the three sires segregating for the OAR3 QTL in our study. The ovine *LALBA* mRNA sequence reported by [48] (GenBank NM001009797) shows three nucleotide variants regarding the gene sequence used as reference here (GenBank AB052168), one of which corresponds to the coding mutation identified in our population, *LALBA*_g.242T>C. However, to our knowledge, there are no reported studies on the influence of the ovine *LALBA* polymorphisms on milk traits.

The candidacy initially suggested by [26] for the *LALBA* gene was strongly supported by the GWA study reported here. This GWA study identified marker OAR3_147028849, located in the third intron of this gene, as the SNP with the most highly significant association detected on OAR3 for PP and FP. The sequence analysis of the *LALBA* gene in our resource population identified 31 DSVs, one of which was located in the gene-coding region (*LALBA*_g.242T>C), and identified a non-synonymous mutation causing a Val27Ala substitution. To predict the possible structural changes that this amino acid change could cause in the LALBA protein, we used the software PolyPhen (http://genetics.bwh.harvard.edu/pph/index.html). The results indicated that the *LALBA*_g.242T>C mutation does not lead to any structural change in the protein. Moreover, the secondary structure for the protein sequence was identical at the mutation region independent of the LALBA_g.242T>C allele [49]. When examining whether the mutation was located in a conserved region of the protein across species, we observed that at the corresponding amino acid position, ruminants and mice have a valine residue, corresponding to the *LALBA*_g.242T allele, whereas other species, such as humans, horse, pig and macaque, have a leucine. The next amino acid of the protein sequence, Phe28Ser, also showed this

divergence pattern between species. On the basis of these results, we postulate that the *LALBA*_g.242T allele is the ancestral variant at that position, whereas the *LALBA*_g.242C allele appears to be a more recent mutation, potentially favored by classical selection in Churra sheep because of its favorable effect on milk protein content. Support for the *LALBA*_g.242T allele as the ancestral is also indicated by the large haplotype block shared by the *Q*-carrying chromosomes compared to the discontinuous haplotype blocks of the *q*-carrying chromosomes, as shown in Figure S1.

The results of the different analyses presented here for the *LALBA*_g.242T>C SNP, including the highly significant associations, the cancelation of the QTL effect when it is included as fixed effect in the linear mixed model, and the concordance test, provide strong statistical support that this SNP is the most explanatory QTN accounting for the milk protein and fat content QTL detected in OAR3 in Spanish Churra sheep. The sign of the allelic substitutions effects at the peak QTL position was identical for PP, FP, PY and FY, with slightly higher estimates (in SD units) for the protein quantity traits than for the fat traits (data not shown), whereas the calculated effect for MY was small and negative (-0.05 SD). On the basis of these observations, the high phenotypic correlation between milk content traits (r^2 = 0.49, based on our YD values) and the significance observed for the PP related QTL, we presume that the target QTL primarily affects PP, with a secondary pleiotropic effect on FP, as previously suggested by [23]. A possible hypothesis supporting this observation is that the Val27Ala substitution at *alpha-lactalbumin* would produce a reduction in the activity of the lactase synthase enzyme with a concomitant reduction in the synthesis of lactose. Because lactose is the major osmoticum in milk and the process of synthesis of lactose is responsible for drawing water into the milk, the reduction of this component would be expected to produce an alteration in the concentration of fat and protein. These two milk components are secreted independently by mammary epithelial cells. All of these observations agree with the concentration effect previously described in *LALBA*-deficient mice for both milk fat and milk protein [46] and support the implication of the *LALBA* gene for the OAR3 QTL effects reported here. Another study on transgenic mice has reported that only PP was significantly affected (*P*-value < 0.05) by the dilution effect due to high levels of expression of the *LALBA* gene, whereas minor differences were observed for the rest of milk components, lactose, cream (fat), and total solids [50]. These observations seem to be in agreement with our findings. In addition, the increased frequency of the favorable allele in the Churra population

(*f*(*LALBA*_g.242C) = 0.71, *f*(*LALBA*_g.242T) = 0.29) and in the rams (*f*(*LALBA*_g.242C) = 0.78, *f*(*LALBA*_g.242T) = 0.22) may indicate that the Churra selection scheme, for which protein content was one of the first traits to be included as a selection target, has already affected this locus by increasing the frequency of the favorable allele associated with increasing milk protein content, *LALBA*_g.242C. This mutation would be the first ovine-proposed QTN in relation to milk production traits.

It is not an easy task to conclusively prove the identification of a QTN. In addition to genetic analyses providing strong genetic support, identifying a QTN requires validation of the results in different populations and functional supporting assays. In their review, [51] provide a list of conditions for a candidate mutation to meet the burden of proof for QTN and discard possible false positives or neutral polymorphisms in near perfect LD with the causative mutation. In this work, we have demonstrated that the *LALBA*_g.242T>C mutation fulfills, for the PP trait, several of the genetic support-related conditions detailed in that list: (1) it is in the defined CI by linkage and linkage disequilibrium mapping; (2) the gene function is related to the trait; (3) there is concordance between the genotypes and the QTL status in the segregating sires (however, this concordance is not observed for the FP trait); (4) the allele frequency of the allele with favorable effect is higher, possibly because of intensive directional selection, and (5) the detected effect disappears when we fix the candidate SNP genotypes in the linear mixed model equation. Other points listed by [51], including the need to confirm the concordance of the results in additional sheep populations and the performance of functional assays complemented by physiological data will be required to finally prove the causality suggested herein for the *LALBA*_g.242T>C mutation. Moreover, there is a need to confirm whether the same physiological mechanism explains the effects detected on PP and FP, and why some of our analyses were not as conclusive for the FP trait as they are for the PP trait. On this regard, it should be taken into account that in dairy sheep, milk FP shows, in general, a lower heritability than milk PP [52,53]. Hence, it is possible that the lower heritability for this trait indicates that there are additional non-genetic factors that are not being controlled in the YD estimation model. However, we cannot currently discard this mutation to be in perfect LD with the genuine causal mutation or QTN of the OAR3 QTL reported here. In any case, the *LALBA*_g.242T>C genetic variant appears to be a useful marker for taking advantage of the commercial nature of the population that has served to identify this potential and could be directly used to assist Churra sheep breeders to make

informed decisions based on genomic information. The implementation in other dairy sheep populations will depend on the frequency of the variant, the rate of LD with the causal mutation, and the validation of the reported association.

ACKNOWLEDGEMENTS

This work was supported by the Spanish Ministry of Science (Project AGL2009-07000) and by the European Commission by 3SR Project (Sustainable Solutions for Small Ruminants; http://www.3srbreeding.eu). Elsa García-Gámez is funded by an FPU contract from the Spanish Ministry of Education. The support and availability to the computing facilities of the Foundation of Supercomputing Center of Castile and León (FCSCL) (http://www.fcsc.es) is greatly acknowledged.

AUTHORS' CONTRIBUTIONS

EG-G performed the quality control, analyses and drafted the manuscript. BG-G, JPS, GS and YB participated in the design and coordination of the study and helped to draft the manuscript. JJA conceived of the study and participated in its design and coordination. All authors have read and approved the final manuscript.

REFERENCES

[1] Jiang L, Liu J, Sun D, Ma P, Ding X, et al. (2010) Genome Wide Association Studies for Milk Production Traits in Chinese Holstein Population. PLoS ONE 5: e13661.

[2] Mai MD, Sahana G, Christiansen FB, Guldbrandtsen B. (2010) A genome-wide association study for milk production traits in Danish Jersey cattle using a 50K single nucleotide polymorphism chip. J Anim Sci 88: 3522–3528.

[3] Olsen HG, Hayes BJ, Kent MP, Nome T, Svendsen M, et al. (2011) Genome-wide association mapping in Norwegian Red cattle identifies quantitative trait loci for fertility and milk production on BTA12. Anim Genet 42: 466-474.

[4] Schopen GC, Visker MH, Koks PD, Mullaart E, van Arendonk JA, et al. (2011) Whole-genome association study for milk protein composition in dairy cattle. J Dairy Sci 94: 3148-3158.

[5] Bouwman AC, Bovenhuis H, Visker MH, van Arendonk JA. (2011) Genome-wide association of milk fatty acids in Dutch dairy cattle. BMC Genet 12: 43.

[6] Bolormaa S, Hayes BJ, Savin K, Hawken R, Barendse W, et al. (2011) Genome-wide association studies for feedlot and growth traits in cattle. J Anim Sci 89: 1684–1697.

[7] Pausch H, Flisikowski K, Jung S, Emmerling R, Edel C, et al. (2011) Genome-wide association study identifies two major loci affecting calving ease and growth-related traits in cattle. Genetics 187: 289-297.

[8] Olsen HG, Hayes BJ, Kent MP, Nome T, Svendsen M, et al. (2009) A genome wide association study for QTL affecting direct and maternal effects of stillbirth and dystocia in cattle. Anim Genet 41: 273-280.

[9] Sahana G, Guldbrandtsen B, Bendixen C, Lund MS. (2010) Genome-wide association mapping for female fertility traits in Danish and Swedish Holstein cattle. Anim Genet 41: 579-588.

[10] Sahana G, Guldbrandtsen B, Lund MS. (2011) Genome-wide association study for calving traits in Danish and Swedish Holstein cattle. J Dairy Sci 94: 479-486.

[11] Schulman NF, Sahana G, Iso-Touru T, McKay SD, Schnabel RD, et al. (2011) Mapping of fertility traits in Finnish Ayrshire by genome-wide association analysis. Anim Genet 42: 263-269.

[12] Becker D, Tetens J, Brunner A, Bürstel D, Ganter M, et al. (2010) Microphthalmia in Texel sheep is associated with a missense mutation in the Paired-Like Homeodomain 3 (PITX3) gene. PLoS ONE 5(1): e8689.

[13] García-Gámez E, Reverter A, Whan V, McWilliam SM, Arranz JJ, et al. (2011) Using regulatory and epistatic networks to extend the findings of a genome scan: identifying the gene drivers of pigmentation in merino sheep. PLoS ONE 6: e21158.

[14] Zhao X, Dittmer KE, Blair HT, Thompson KG, Rothschild MF, et al. (2011) A novel nonsense mutation in the DMP1 gene identified by a genome-wide association study is responsible for inherited rickets in Corriedale Sheep. PLoS ONE 6: e21739.

[15] Grisart B, Farnir F, Karim L, Mni M, Simon P, et al. (2002) Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Res 12: 222–231.

[16] Blott S, Kim JJ, Moisio S, Schmidt-Küntzel A, Cornet A, et al. (2003) Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. Genetics 163(1): 253-266.

[17] Olsen HG, Nilsen H, Hayes B, Berg PR, Svendsen M, et al. (2007) Genetic support for a quantitative trait nucleotide in the ABCG2 gene affecting milk composition of dairy cattle. BMC Genet 8: 32.

[18] García-Fernández M, Gutiérrez-Gil B, Sánchez JP, Morán JA, García-Gámez E, et al. (2011) The role of bovine causal genes underlying dairy traits in Spanish Churra sheep. Anim Genet 42: 415-420.

[19] Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. J Dairy Sci 92: 433-443.

[20] Taylor, J. (2012) Current Status of Genomic Selection in US Beef Cattle. XX Plant and Animal Genome Conference Sand Diego, USA, 14-18 January 2012. https://pag.confex.com/pag/xx/webprogram/Paper1662.html.

[21] Barillet F, Arranz JJ, Carta A. (2005) Mapping quantitative trait loci for milk production and genetic polymorphisms of milk proteins in dairy sheep. Genet Sel Evol 37 Suppl 1: S109-123.

[22] Gutiérrez-Gil B, El-Zarei MF, Bayón Y, Álvarez L, de la Fuente LF, et al. (2007) Short communication: detection of quantitative trait loci influencing somatic cell score in Spanish Churra sheep. J Dairy Sci 90: 422-426.

[23] Gutiérrez-Gil B, El-Zarei MF, Álvarez L, Bayón Y, De La Fuente LF, et al. (2009) Quantitative trait loci underlying milk production traits in sheep. Anim Genet 40: 423-434.

[24] Raadsma HW, Jonas E, McGill D, Hobbs M, Lam MK, et al. (2009) Mapping quantitative trait loci (QTL) in sheep. II. Meta-assembly and identification of novel QTL for milk production traits in sheep. Genet Sel Evol 41: 45.

[25] Mateescu RG, Thonney ML. (2010) Genetic mapping of quantitative trait loci for milk production in sheep. Anim Genet 41: 460-466.

[26] García-Gámez E, Gutiérrez-Gil B, Sánchez JP, Arranz JJ (2011). Short communication: Replication and refinement of a QTL influencing milk protein percentage on ovine chromosome 3. Anim Genet doi: 10.1111/j.1365-2052.2011.02294.x.

[27] VanRaden PM, Wiggans GR. (1991) Derivation, calculation, and use of national animal model information. J Dairy Sci 74: 2737-2746.

[28] García-Gámez E, Gutiérrez-Gil B, Sahana G, Arranz JJ. (2012) Linkage disequilibrium and inbreeding estimation in Spanish Churra sheep. BMC Genet 13: 43.

[29] Green P, Falls K, Crooks S. (1990) Documentation for CRI-MAP, version 2.4. Washington University School of Medicine, St Louis.

[30] Gao X, Becker LC, Becker DM, Starmer J, Province MA. (2009) Avoiding the high Bonferroni penalty in genome-wide association studies. Genet Epidemiol 34: 100-105.

[31] García-Fernández M, Gutiérrez-Gil B, García-Gámez E, Arranz JJ. (2009) Genetic variability of the Stearoyl-CoA desaturase gene in sheep. Mol Cell Probes 23: 107-111.

[32] Legarra A, Fernando RL. (2009) Linear models for joint association and linkage QTL mapping. Genet Sel Evol 41: 43.

[33] Filangi O, Moreno C, Gilbert H, Legarra A, Le Roy P, et al. (2010) QTLMap, a software for QTL detection in outbred populations. Proceedings of the 9th World Congress on Genetics Applied to Livestock Production ISBN 978-3-00-031608-1.

[34] Druet T, Georges M. (2010) A hidden markov model combining linkage and linkage disequilibrium information for haplotype reconstruction and quantitative trait locus fine mapping. Genetics 184: 789-798.

[35] Seaton G, Hernandez J, Grunchec JA, White I, Allen J, et al. (2006) GridQTL: A Grid Portal for QTL Mapping of Compute Intensive Datasets. Proceedings of the 8th World Congress on Genetics Applied to Livestock Production ISBN: 85-60088-00-8.

[36] Gutiérrez-Gil B, Wiener P, Williams JL, Haley CS. (2012) Investigation of the genetic architecture of a bone carcass weight QTL on BTA6. Anim Genet doi: 10.1111/j.1365-2052.2012.02322.x.

[37] Seroussi E. (2009) The concordance test emerges as a powerful tool for identifying quantitative trait nucleotides: lessons from BTA6 milk yield QTL. Anim Genet 40: 230-234.

[38] Daetwyler HD, Schenkel FS, Sargolzaei M, Robinson JA. (2008) A genome scan to detect quantitative trait loci for economically important traits in Holstein cattle using two methods and a dense single nucleotide polymorphism map. J Dairy Sci 91: 3225-3236.

[39] Gonda Y, Sakurai H, Hirata Y, Tabata H, Ajioka I, et al. (2007) Expression profiles of Insulin-like growth factor binding protein-like 1 in the developing mouse forebrain. Gene Expr Patterns 7:431-440.

[40] Mullen MP, Berry DP, Howard DJ, Diskin MG, Lynch CO, et al. (2011) Single nucleotide polymorphisms in the insulin-like growth factor 1 (IGF-1) gene are associated with performance in Holstein-Friesian dairy cattle. Front Genet 2: 3.

[41] Plante Y, Gibson JP, Nadesalingam J, Mehrabani-Yeganeh H, Lefebvre S, et al. (2001) Detection of quantitative trait loci affecting milk production traits on 10 chromosomes in Holstein cattle. J Dairy Sci 84: 1516-1524.

[42] Sharif S, Mallard BA, Wilkie BN, Sargeant JM, Scott HM, et al. (1998) Associations of the bovine major histocompatibility complex DRB3 (BoLA-DRB3) alleles with occurrence of disease and milk somatic cell score in Canadian dairy cattle. Anim Genet 29: 185-193.

[43] Moioli B, D'Andrea M, Pilla F. (2007) Candidate genes affecting sheep and goat milk quality. Small Rumin Res 68: 179–192.

[44] Diez-Tascón C, Bayón Y, Arranz JJ, De La Fuente F, San Primitivo F. (2001) Mapping quantitative trait loci for milk production traits on ovine chromosome 6. J Dairy Res 68: 389-397.

[45] Howard B, Ashworth A. (2006) Signalling pathways implicated in early mammary gland morphogenesis and breast cancer. PLoS Genet 2: e112.

[46] Stinnakre MG, Vilotte JL, Soulier S, Mercier JC. (1994) Creation and phenotypic analysis of alpha-lactalbumin-deficient mice. Proc Natl Acad Sci U S A 91: 6544-6548.

[47] Lundén A and Lindersson. (1998) α-Lactalbumin polymorphism in relation to milk lactose. Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, Volume 25.

[48] Gaye P, Hue-Delahaie D, Mercier JC, Soulier S, Vilotte JL, et al. (1987) Complete nucleotide sequence of ovine alpha-lactalbumin mRNA. Biochimie 69: 601-608.

[49] Zdobnov E.M. and Apweiler R. (2001) InterProScan - an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17: 847-848.

[50] Boston WS, Bleck GT, Conroy JC, Wheeler MB, Miller DJ. (2001) Short communication: effects of increased expression of alpha-lactalbumin in transgenic mice on milk yield and pup growth. J Dairy Sci 84:620-622.

[51] Ron M, Weller JI. (2007) From QTL to QTN identification in livestock--winning by points rather than knock-out: a review. Anim Genet 38: 429-439.

[52] Legarra A, Ugarte E. (2001) Genetic parameters of milk traits in Latxa dairy sheep. Anim Sci 73: 407-412.

[53] Othmane MH, Carriedo JA, San Primitivo F, De La Fuente LF. (2002) Genetic parameters for lactation traits of milking ewes: protein content and composition, fat, somatic cells and individual laboratory cheese yield. Genet Sel Evol 34: 581-596.

FIGURES

Figure 1. Result from the Genome-wide Association analysis based on the analysis of the *Illumina OvineSNP50 BeadChip*. For the five traits under study (milk yield, MY; protein percentage, PP; fat percentage, FP; protein yield, PY; and fat yield, FY) the log(1/*P*-value) are depicted here for all the 43,784 SNPs that passed the quality control.

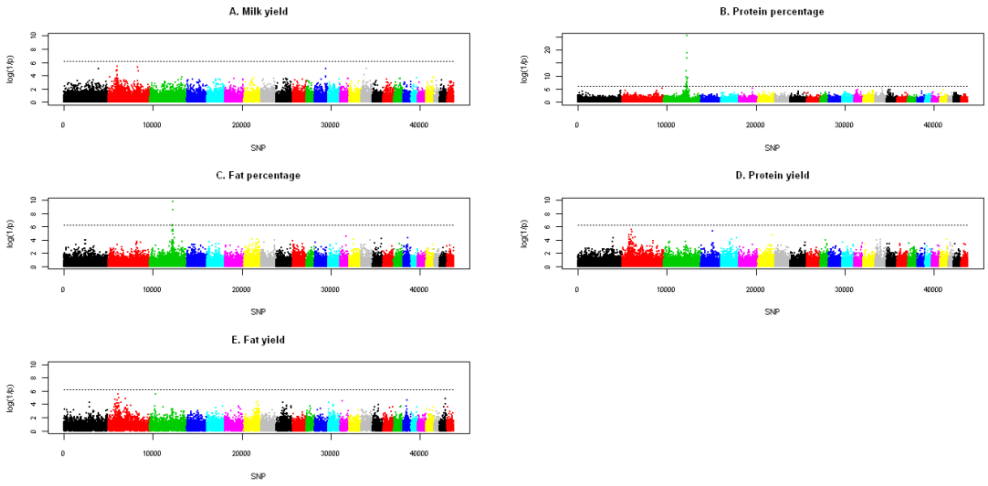


Figure 2. Statistical profiles obtained from the LDLA analyses in OAR3 for milk protein and fat percentages**.** Together with the LRT profiles obtained across the whole chromosome for protein percentage (PP)(A) and fat percentage (FP) (B), the detailed view of the profile around the *LALBA* gene region (from 1.3 to 1.4 Mb) is also provided for both traits, PP (C) and FP (D).

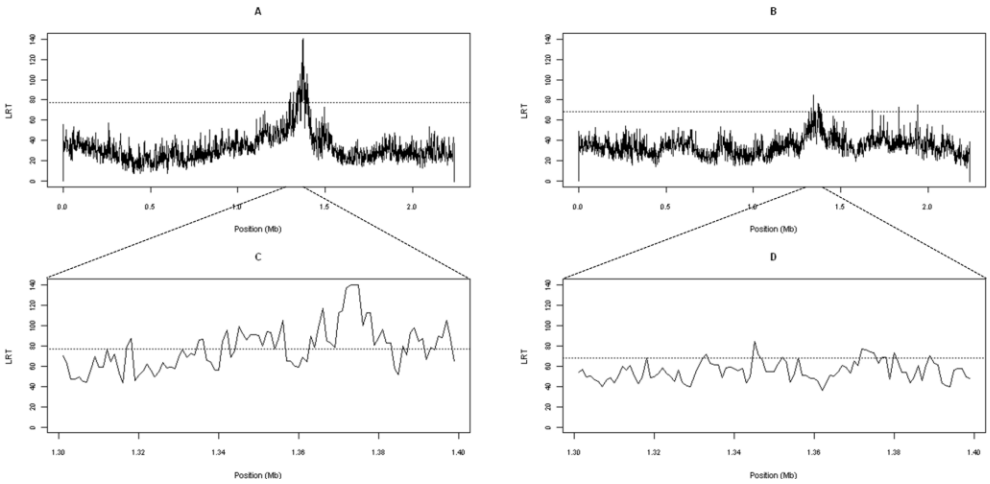
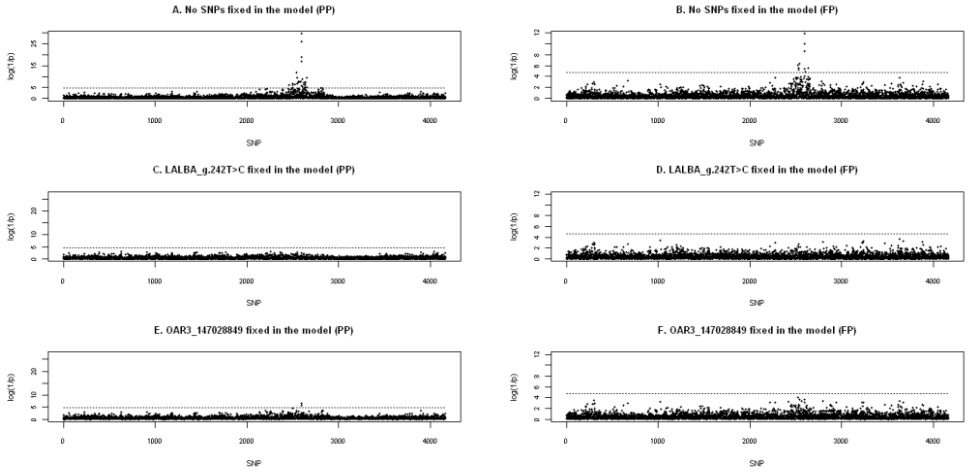


Figure 3. Association analyses for milk protein and fat percentages (PP and FP) including the *LALBA*_g.242T>C genotypes. Genotypes used in the analysis included SNPs from the *Illumina OvineSNP50 BeadChip* and the single SNP detected in the exonic sequence of the *LALBA* gene (*LALBA*_g.242T>C). The significance thresholds are set at the 5% chromosome-wise level (dashed lines). A and B) Log(1/*P*-value) profiles for the analysis including the pedigree information as a random polygenic effect and each SNPs' genotypes as a fixed effect in the mixed model for PP and FP, respectively. C and D) Results for the analysis where the genotypes of the *LALBA*_g.242T>C mutation were included as a fixed effect in the mixed model equation for PP and FP, respectively. E and F). Results of the analysis when the SNP OAR3_147028849 was included as a fixed effect in the mixed model equation for PP and FP, respectively.



TABLES

Table 1. Summary of the significant results obtained from the genome-wide association analysis reported herein.

Significance threshold	Chrom	SNP ID	Position (Mbp)	Trait	Allele substitution effect trait units (SD units)	*P*-value (Nominal)	*P*-value (Corrected)*
Experiment-wise significant	3	OAR3_147028849	137.3	PP	0.138 ± 0.013 (0.470)	$3.78x10^{-26}$	$9.24x10^{-23}$ ($2.77x10^{-22}$)
	3	OAR3_147028849	137.3	FP	0.169 ± 0.026 (0.297)	$1.80x10^{-10}$	$4.39x10^{-7}$ ($1.32x10^{-6}$)
Chromosome-wise significant	1	OAR1_233634722	216.9	MY	25.824 ± 5.815 (0.210)	$9.55x10^{-6}$	0.030
	2	OAR2_59276935	55.4	MY	26.476 ± 5.766 (0.215)	$4.74x10^{-6}$	0.013
	2	s25113	58.7	PY	-1.467 ± 0.321 (-0.193)	$5.27x10^{-6}$	0.015
	2	OAR2_67323597	63.2	FY	-2.249 ± 0.507 (-0.231)	$9.75x10^{-6}$	0.027
	2	OAR2_182893049	173.1	MY	21.526 ± 4.743 (0.175)	$6.10x10^{-6}$	0.017
	2	OAR2_251955154	239.6	PP	0.051 ± 0.011 (0.173)	$5.56x10^{-6}$	0.015
	3	s40946	35.0	FY	1.811 ± 0.387 (0.186)	$3.07x10^{-6}$	0.008
	4	OAR4_74926053	70.1	PY	-1.552 ± 0.337 (-0.204)	$4.46x10^{-6}$	0.006
	6	DU430803_572	85.4	PP	-0.044 ± 0.009 (-0.148)	$4.12x10^{-6}$	0.005
	7	OAR7_89700312	82.3	PY	-1.562 ± 0.363 (-0.205)	$1.78x10^{-5}$	0.021
	12	OAR12_74759500	68.2	MY	20.021 ± 4.513 (0.163)	$9.77x10^{-6}$	0.009
	14	s66781	13.3	PP	-0.050 ± 0.012 (-0.168)	$3.01x10^{-5}$	0.019
	14	OAR14_15268863	14.9	FY	1.664 ± 0.397 (0.171)	$2.87x10^{-5}$	0.018
	14	OAR14_28957918	27.6	PP	-0.041 ± 0.010 (-0.138)	$2.47x10^{-5}$	0.016
	14	s25830	41.6	FP	0.133 ± 0.032 (0.234)	$2.89x10^{-5}$	0.018
	15	s36641	69.3	PP	0.043 ± 0.010 (0.146)	$3.68x10^{-5}$	0.031
	16	s25440	32.5	MY	-22.771 ± 5.115 (-0.185)	$9.10x10^{-6}$	0.007
	16	OAR16_46325523	43.0	PP	-0.044 ± 0.010 (-0.148)	$6.80x10^{-6}$	0.006
	17	s42157	11.4	PP	-0.055 ± 0.013 (-0.185)	$2.66x10^{-5}$	0.019
	17	OAR17_23761428	21.4	PP	0.067 ± 0.016 (0.227)	$1.73x10^{-5}$	0.013
	17	OAR17_63857104	58.8	FP	-0.079 ± 0.020 (-0.139)	$6.03x10^{-5}$	0.044
	20	OAR20_25029391	23.7	FY	1.977 ± 0.465 (0.203)	$2.23x10^{-5}$	0.013
	20	s69570	29.1	FP	-0.139 ± 0.034 (-0.245)	$5.43x10^{-5}$	0.032
	20	OAR20_32868803	29.6	PP	-0.051 ± 0.013 (-0.173)	$5.51x10^{-5}$	0.032
	23	OAR23_28103191	27.1	PP	-0.042 ± 0.010 (-0.141)	$2.08x10^{-5}$	0.013
	23	s41936	50.8	PY	2.131 ± 0.536 (0.280)	$7.33x10^{-5}$	0.045
	25	s07823	35.3	FY	1.759 ± 0.446 (0.181)	$8.25x10^{-5}$	0.045

* Corrected *P*-values at the experiment-wise or chromosome-wise level are indicated for the associations reaching the corresponding 0.05 significance level. These values were obtained after applying a Bonferroni correction considering the number of independent markers analysed for each chromosome and the entire genome. The 0.05 experiment-wise significance threshold was set by correcting additionally for three independent traits analysed.

The most significant SNPs at each of the significant regions identified at the chromosome-wise and experiment-wise levels are indicated. For each of them, the location chromosome and position (in Megabase pairs), and the nominal and corrected *P*-values are provided. Also included in the table are the magnitude and standard error of the allele substitution effect in both trait units (mL, for yield traits and percentage points, for composition traits) and in phenotypic standard deviations (SD) units (in brackets).

SUPPORTING INFORMATION.

Figure S1. Haplotypes of the heterozygous sires for the PP QTL on OAR3 according to the regression analysis. The two QTL alleles, *Q* (increased protein content) and *q* (decreased protein content), were assigned to the respective haplotypes, as determined by half-sib family-based regression analysis performed with GridQTL. The genotypes of the 22 markers included in the 136.8-138.1 Mb interval were investigated to check the concordance with the estimated sire´s *QTL* status.

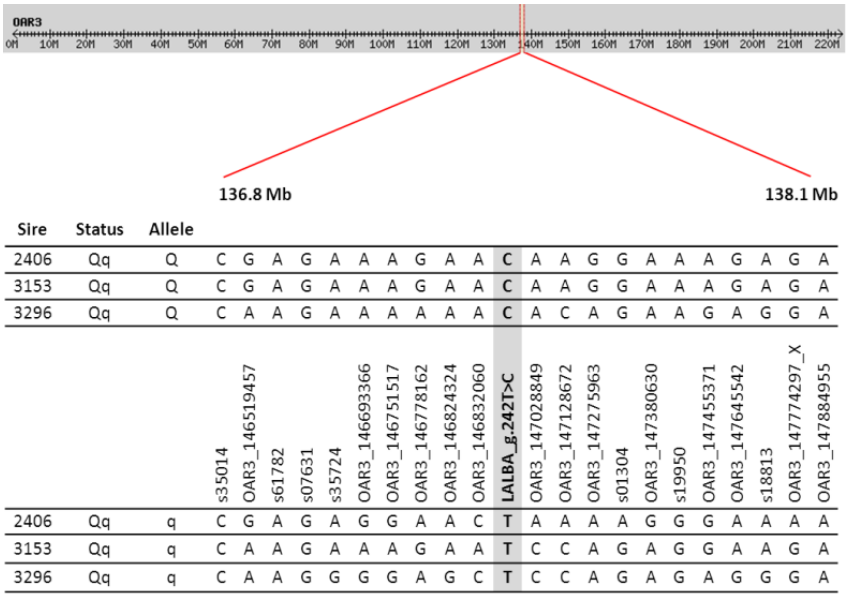


Table S1. Primers used in the amplification of the ovine *alpha-lactalbumin* (*LALBA*) gene. The size of the six amplified fragments and the melting temperature (T_m) used in the PCR amplifications are also indicated.

Fragment	*LALBA* region covered	Primer ID	Sequence (*5′* to *3′*)	Tm	Fragment size (pb)
Amplicon 1	Promoter region and complete 5'UTR	LALBA_ovis0_up	gggcagctagatactgtcatacac	60 ºC	693
		LALBA_ovis0_dn	accaggagcagagagacaaag		
Amplicon 2	Complete 5' UTR, exon 1 and intron 1	LALBA_ovis1_up	tttgggcaggtaacaattcc	60 ºC	743
		LALBA_ovis1_dn	tggaagagtccatattctgtgc		
Amplicon 3	Complete exon 2	LALBA_ovis2_up	ggttggagagccttttctg	58 ºC	599
		LALBA_ovis2_dn	tagaggcttatgcctgttgc		
Amplicon 4	Complete intron 2 and exon 3	LALBA_ovis3_up	gacctgagctgtttggctatc	60 ºC	595
		LALBA_ovis3_dn	tgttcaagtcctggtgaagg		
Amplicon 5	Complete intron 3 and exon 4	LALBA_ovis4_up	ccgttctctctatttcctggtc	58 ºC	764
		LALBA_ovis4_dn	cacgcatccctggagattag		
Amplicon 6	3' UTR	LALBA_ovis5_up	tgaacacctgctgtctttgc	58 ºC	873
		LALBA_ovis5_dn	tgttgctgttgttgttgctg		

Table S2. DNA sequence variants detected in the ovine *LALBA* gene through the sequencing analysis performed on the 16 Churra sires included in this study. Positions are referred to according to the GeneBank sequence (AB052168).

SNP ID	Fragment	Position in *LALBA* gene	Position according to GeneBank Acc. No. AB052168.1
*LALBA*_g.-421T>C	Amplicon 1	Promoter	-421
*LALBA*_g.-295A>G	Amplicon 1	Promoter	-295
*LALBA*_g.-275A>G	Amplicon 1	Promoter	-275
*LALBA*_g.-231C>G	Amplicon 1	Promoter	-231
*LALBA*_g.-212T>C	Amplicon 1	Promoter	-212
*LALBA*_g.-43G>T	Amplicon 2	Promoter	-43
*LALBA*_g.47G>A	Amplicon 2	Promoter	47
*LALBA*_g.82G>A	Amplicon 2	Promoter	82
*LALBA*_g.242T>C	Amplicon 2	exon 1	242
*LALBA*_g.298G>A	Amplicon 2	intron 1	298
*LALBA*_g.353delC	Amplicon 2	intron 1	353
*LALBA*_g.851insT	Amplicon 3	intron 2	851;1
*LALBA*_g.1005T>G	Amplicon 3	intron 2	1005
*LALBA*_g.1113G>A	Amplicon 3	intron 2	1113
*LALBA*_g.1196T>C	Amplicon 4	intron 2	1196
*LALBA*_g.1250C>T	Amplicon 4	intron 2	1250
*LALBA*_g.1397A>G	Amplicon 4	intron 3	1397
*LALBA*_g.1450T>A	Amplicon 4	intron 3	1450
*LALBA*_g.1477G>T	Amplicon 4	intron 3	1477
*LALBA*_g.1520C>T	Amplicon 4	intron 3	1520
*LALBA*_g.1664G>T	Amplicon 5	intron 3	1664
*LALBA*_g.1746T>C	Amplicon 5	intron 3	1746
*LALBA*_g.1977C>T	Amplicon 5	3' UTR	1977
*LALBA*_g.2020A>G	Amplicon 5	3' UTR	2020
*LALBA*_g.2023A>G	Amplicon 5	3' UTR	2023
*LALBA*_g.2058G>A	Amplicon 5	3' UTR	2058
*LALBA*_g.2223G>C	Amplicon 6	3' UTR	2223
*LALBA*_g.2241A>C	Amplicon 6	3' UTR	2241
*LALBA*_g.2277A>G	Amplicon 6	3' UTR	2277
*LALBA*_g.2422A>C	Amplicon 6	3' UTR	2422
*LALBA*_g.2485T>C	Amplicon 6	3' UTR	2485

Table S3. Description of the haplotypes found at the most significantly associated position in the LDLA QTLmap analysis for protein percentage (PP). The haplotypes include 4 SNPs: *LALBA*_g.242C>T, OAR3_147028849, OAR3_147128672 and OAR3_147275963. Haplotype frequency, estimated haplotype effect and the precision of the estimate are given in the table.

Haplotype	Frequency	Effect	Precision of the estimate
CAAA	0.14	0.033	0.224
CAAG	0.04	0.025	0.235
CACA	0.30	0.030	0.222
CACG	0.13	0.041	0.225
TAAA	0.01	-0.138	0.263
TACA	0.03	-0.081	0.235
TACG	0.01	-0.005	0.268
TCAA	0.01	-0.100	0.277
TCCA	0.31	-0.121	0.222
TCCG	0.01	-0.112	0.322

RESUMEN DE RESULTADOS Y DISCUSIÓN GENERAL

La búsqueda de regiones genómicas con influencia sobre caracteres productivos en las distintas especies ganaderas se ha beneficiado de los avances experimentados en los últimos años en el campo de la genómica. Los barridos genómicos realizados en los años 90 en el ganado vacuno, y comienzos del siglo XXI en el ovino, basados en marcadores microsatélite, proporcionaron un primer paso para la identificación de regiones genómicas que contienen las mutaciones directamente responsables de la variación fenotípica observada en caracteres de interés económico en estas especies. En los últimos años, las tecnologías de secuenciación de nueva generación, basadas en plataformas de secuenciación masiva paralela, han revolucionado la investigación genética, estando a día de hoy los genomas de varias especies de animales domésticos (pollo, perro, cerdo, vaca, oveja y caballo) parcial o completamente secuenciados (Bai et al., 2012). La variabilidad detectada a lo largo de estos genomas ha aportado la información necesaria para el desarrollo, en estas especies, de los chips de SNPs de primera generación (alrededor de 50.000 SNPs), estando los de segunda generación (~ 800.000 marcadores) en algunas especies ya disponibles (*BovineHD Genotyping BeadChip*) o en fase de desarrollo (*Ovine HD BeadChip*) (Taylor, 2012).

Esta memoria de Tesis Doctoral se encuadra, desde el punto de vista temporal, en un momento de transición de la investigación en el campo de la genética animal, entre los estudios clásicos basados en mapas de marcadores microsatélite de baja densidad y análisis de ligamiento, y lo que podríamos considerar la Era Genómica de las especies domésticas, en la que el análisis de una alta densidad de marcadores mediante las plataformas tipo chip proporciona la oportunidad de interrogar el genoma explotando directamente la información derivada del desequilibrio de ligamiento (análisis de asociación). Dicha transición, ha marcado las actividades del grupo de investigación en el que se ha realizado esta Tesis Doctoral, y, por tanto, el propio desarrollo de la misma.

En esta sección se presenta un resumen discutido de los resultados obtenidos en esta memoria de Tesis Doctoral, los cuales responden a los objetivos planteados en la sección inicial de la memoria.

Objetivo 1. Confirmación de los QTL localizados en OAR20 y OAR3, que afectan al porcentaje de grasa y porcentaje de proteína, respectivamente, y que han mostrado anteriormente evidencia de segregación en la raza ovina Churra.

Los resultados de un barrido genómico para la detección de QTL con influencia sobre caracteres de producción de leche realizado en una población comercial de ovejas de raza Churra son la base sobre la que se asientan las actividades realizadas en la fase inicial de esta memoria. Este proyecto previo, que consistió en el análisis de 181 marcadores microsatélite distribuidos homogéneamente a lo largo del genoma autosómico ovino, en 11 familias de medio-hermanas de raza Churra, permitió la identificación de nueve regiones cromosómicas que mostraron indicios de ser portadoras de un QTL. De ellas, únicamente un QTL localizado en OAR3, con efectos sobre el carácter PP, alcanzó una significación a nivel genómico (Gutiérrez-Gil et al., 2009).

Una de las principales limitaciones de este tipo de estudios clásicos de detección de QTL, basados en mapas de baja densidad, es la escasa precisión en el mapeo de las regiones identificadas como QTL. Las amplias regiones cromosómicas incluidas en los CI estimados, hacen necesario el posterior mapeo fino de las regiones identificadas como asociadas con caracteres de interés productivo. Este proceso de mapeo fino, con el objetivo de incrementar de manera sustancial la densidad de marcadores en la región de interés, resulta muy laborioso y costoso a nivel económico (Blott et al., 2003; Sellner et al., 2007). Por ello, antes de concentrar todo tipo de recursos y esfuerzos en una región genómica, es indispensable la confirmación de los efectos inicialmente detectados en una muestra independiente.

En base a esto, el primer objetivo de la presente Tesis Doctoral se planteó para confirmar algunos de los resultados previamente descritos por Gutiérrez-Gil et al. (2009) en la población de ganado ovino de raza Churra a la que nos referiremos como Población 1. Las regiones QTL inicialmente identificadas en esta población y seleccionadas para su confirmación, y posible mapeo fino posterior, fueron el QTL para PP localizado en OAR3 y un QTL con efectos sobre el carácter FP localizado en OAR20. El primero de ellos fue seleccionado por ser el resultado con mayor significación del barrido genómico presentado por Gutiérrez-Gil et al. (2009) y por el relativamente corto CI estimado por *bootstrapping* para esta región QTL (CI = 40 cM). La selección de la región localizada en OAR20, aunque

no presentó un apoyo estadístico destacable, se basó principalmente en que varios estudios realizados en otras poblaciones ovinas coincidían en la identificación de QTL con influencia sobre caracteres de producción de leche en la misma región cromosómica (Barillet et al., 2006).

Siguiendo la terminología inglesa y de acuerdo con el trabajo publicado por Igl et al. (2009), se definen dos términos diferentes para los estudios cuyo objetivo es la confirmación de resultados previamente descritos. En primer lugar, la replicación (*replication*) se refiere a estudios en los que se utiliza una muestra independiente procedente de la misma población animal. En la especie humana, se ha sugerido que estos estudios deben basarse en una muestra y un diseño experimental lo más parecidos posible a la muestra inicial (Chanock et al., 2007). Así, se pueden eliminar los falsos positivos debidos a la posible estructura poblacional no tenida en cuenta en el primer estudio. Por otra parte, el término validación (*validation*) hace referencia a la utilización de una población independiente para el confirmación de los resultados iniciales. En este caso las diferencias entre las muestras incluyen variaciones sistemáticas que puedan existir, por ejemplo, entre razas. Si los resultados son validados, éstos pueden ser más fácilmente generalizados que si únicamente son replicados (Igl et al., 2009).

En nuestro caso, el planteamiento de confirmación que se llevó a cabo en las dos regiones seleccionadas fue el de replicación, analizándose una nueva muestra de animales de la misma población comercial de raza Churra, en ambos casos siguiendo el diseño experimental utilizado por Gutiérrez-Gil et al. (2009) mediante el análisis de nuevas familias de medio hermanas pertenecientes al núcleo de selección de esta raza.

Una replicación inicial de los dos QTL objeto de estudio se llevó a cabo en una población de 800 ovejas distribuidas en 15 familias utilizada en los trabajos realizados por García-Fernández et al. (2010a, 2010b, 2011) para el estudio de caracteres relacionados con el perfil de ácidos grasos de la leche y a la que nos referiremos como Población 2. Dada la disponibilidad en esta población de los datos de producción lechera, se genotiparon en ella los mismos marcadores inicialmente analizados por Gutiérrez-Gil et al. (2009) en los cromosomas OAR3 y OAR20 y se planteó el mismo tipo de análisis que en el barrido genómico inicial basado en un clásico análisis de ligamiento siguiendo un diseño hija.

El primer trabajo presentado en esta Tesis Doctoral describe los resultados de confirmación obtenidos en relación al QTL para FP localizado en OAR20. En este caso se realizó un meta-análisis combinando los datos de las Poblaciones 1 y 2, analizándose un total de 2.120 animales distribuidos en 25 familias de medio-hermanas. El análisis de esta meta-población detectó en el cromosoma analizado, OAR20, un QTL significativo a nivel cromosómico para el carácter FP, con una localización muy cercana al previamente descrito por Gutiérrez-Gil et al. (2009), estando el pico del QTL localizado en la posición 61 cM, entre los marcadores BP34 y BM1905. El análisis intrafamiliar reveló un total de seis familias segregantes, siendo dos de ellas de las nuevas familias analizadas, otras tres pertenecientes a la Población 1 (Gutiérrez-Gil et al., 2009) y una última común a ambos estudios y que ya había sido identificada como segregante anteriormente. La identificación de nuevas familias segregantes para este QTL serviría como confirmación del mismo. Sin embargo, el CI estimado en este meta-análisis incluyó el cromosoma completo (0-89 cM), lo que puede deberse a la variable informatividad de los marcadores y a las diferencias observadas en cuanto a la localización del QTL en las familias segregantes (desde 21 a 88 cM), tal y como se observaba en los resultados descritos por Gutiérrez-Gil et al. (2009). Esta falta de mejora en la definición de la región QTL, a pesar de la identificación de nuevas familias segregantes, pone de manifiesto, una vez más, las limitaciones en cuanto a la precisión del mapeo de QTL realizados en poblaciones comerciales, en las que pueden existir importantes diferencias en cuanto a la informatividad de los marcadores entre las distintas familias.

Coincidiendo con la localización del QTL confirmado para el carácter FP en OAR20, se han descrito anteriormente varios QTL para caracteres de producción de leche en una población Sarda x Lacaune (Barillet et al., 2006). Así, se han descrito QTL para cantidad de leche, cantidad de grasa y de proteína en la región del complejo mayor de histocompatibilidad (*MHC-class II*). Otro QTL para el carácter porcentaje de grasa descrito en esta población experimental, se sitúa en una región cercana al marcador OARHH56, cercano al marcador BP34, y es por tanto coincidente con los resultados de nuestro análisis *across-family*. Además, en la región ortóloga bovina de BTA23 se han descrito varios QTL relacionados con la composición grasa de la leche en diferentes estudios basados en ligamiento con razas lecheras (Plante et al., 2001; Zhang et al., 1998). Un estudio de asociación realizado en animales de la raza Jersey danesa ha confirmado los efectos de este QTL también sobre al cantidad de grasa en leche (Mai et al., 2010).

La versión 2.0 del *Ovine Genome Assembly* se utilizó para la búsqueda de genes candidatos posicionales a lo largo de la región de interés del cromosoma 20. En esta región del OAR20 se localiza el gen que codifica para la prolactina (PRL), entre los marcadores OLADRB y BP34. La prolactina es una hormona esencial para el desarrollo mamario, la lactogénesis y la expresión de genes de proteínas de la leche (Horseman, 1999), lo que convierte el gen *PRL* en un posible candidato en relación al QTL estudiado en OAR20, tanto desde el punto de vista posicional como funcional. En ganado vacuno se han descrito varios polimorfismos en este gen. Además, Brym et al. (2005), Dybus et al. (2005) y Alipanah et al. (2007) realizaron estudios de asociación entre algunos de los polimorfismos detectados en la secuencia del gen *PRL* y caracteres lecheros identificando algunas asociaciones significativas.

En base a la identificación de este gen candidato en la región del QTL de OAR20, nuestro grupo de investigación planteó una aproximación alternativa al estudio de ligamiento, basada en el estudio de asociación entre polimorfismos detectados en este gen candidato y caracteres de producción de leche en el ganado ovino de raza Churra. Los análisis realizados en este sentido se basaron en el estudio de la Población 2, utilizándose los 15 machos de esta población para la identificación de los polimorfismos, mediante un análisis de secuenciación, y los datos productivos de sus hijas para realizar el correspondiente análisis de asociación (García-Gámez et al., 2010). Los resultados obtenidos mostraron cierta asociación entre dos de los polimorfismos identificados en el gen *PRL* y el carácter cantidad de leche (MY), aunque no se detectaron asociaciones significativas con los porcentajes de grasa y proteína en leche (García-Gámez et al., 2010).

Además del gen *PRL*, en esta misma región se encuentran otros genes que codifican para una serie de proteínas relacionadas con la prolactina (PRP1, PRP3, PRP5, PRP6, PRP8, PRP9, PRP12, PRP14). Estas proteínas son miembros del grupo de hormonas placentarias relacionadas directamente con la familia de la prolactina y la hormona del crecimiento (GH). Algunas de ellas se han descrito en ganado ovino, aunque estudios previos no han detectado función lactogénica para las mismas (Ushizawa et al., 2007). Por último, en esta misma región cromosómica, encontramos el gen que codifica para el lactógeno placentario (PL). La capacidad de esta hormona de unirse al receptor de la GH parece indicar la posibilidad de que tenga una función galactopoyética, que ha sido descrita en especies relacionadas con el ganado ovino, como es el ganado vacuno. Sin embargo, el tratamiento del tejido mamario de un grupo de ovejas en lactación con PL ovino, no mostró que esta proteína tuviera acción

galactopoyética ni somatogénica sobre las células de la glándula mamaria en el periodo de lactación de dichos animales (Basset et al., 1998). Este hallazgo parece descartar a esta hormona como un gen candidato funcional, al menos para un primer estudio de asociación.

Así pues, en la presente Tesis Doctoral se ha descrito la confirmación de la segregación en la oveja Churra del QTL previamente identificado por Gutiérrez-Gil et al. (2009) para porcentaje de grasa en OAR20. Si bien este resultado da validez al resultado previamente descrito, la falta de reducción del intervalo de confianza y de resultados significativos en el análisis del gen candidato estudiado en relación al carácter FP determinaron, en esta fase del proyecto, que no se planificaran esfuerzos adicionales en relación a esta región genómica inicialmente seleccionada. En este sentido, nos gustaría comentar que aunque se intentó incrementar la densidad de marcadores microsatélite en la segunda mitad de este cromosoma, dichos esfuerzos fueron infructuosos debido a la dificultad para genotipar de forma consistente y fiable los marcadores entonces disponibles para esa región según la v4.7 del mapa de ligamiento ovino disponible en la página web *Australian Sheep Gene Mapping Website*.

En el segundo trabajo de confirmación de resultados, en este caso el estudio del cromosoma OAR3 como portador de un QTL con efecto sobre el porcentaje de proteína, la estrategia seguida fue similar, aunque debido a los resultados positivos inicialmente obtenidos en la Población 2 para esta región (confirmación de un QTL altamente significativo en la misma región y para el mismo carácter, y el exitoso análisis de marcadores microsatélite adicionales) se planteó, para el estudio detallado de este QTL, el análisis en una nueva población a la que nos referiremos como Población 3. Dadas las limitaciones identificadas en las poblaciones anteriormente estudiadas (degradación de las muestras de ADN de la Población 1; y baja potencia estadística para la identificación y mapeo fino de QTL de la Población 2), el muestreo de esta nueva población se basó en la necesidad de contar con una población comercial que ofreciera una potencia estadística adecuada para la identificación de QTL y la realización de estudios de mapeo fino de las regiones más prometedoras identificadas. Así pues, la Población 3 incluyó un total de 1.696 ovejas repartidas en 16 familias de medio-hermanas y distribuidas en 29 rebaños del núcleo de selección de ANCHE. El número de hijas por familia varió entre 29 y 277 animales con un promedio de 105 medio-hermanas por familia.

Los primeros análisis realizados en la Población 3 tuvieron por objeto la confirmación del QTL identificado en OAR3 en relación al carácter PP. Para ello, se incrementó la densidad de marcadores microsatélite en la región de interés, pasando de 11 a 21 marcadores analizados en este cromosoma. A través de un análisis de ligamiento presentado en el segundo trabajo incluido en esta Tesis Doctoral se replicaron los resultados obtenidos por Gutiérrez-Gil et al. (2009), aunque, de nuevo, el análisis inicial basado en ligamiento clásico no permitió reducir el CI de localización del QTL. Con el objetivo de mejorar la precisión del mapeo, se realizó un análisis basado en la combinación de ligamiento y desequilibrio de ligamiento en la región del pico del QTL identificado en el análisis de ligamiento, incluyendo en total de nueve microsatélites que cubrían 50 cM con una distancia media entre ellos de 6,2 cM. Utilizando esta metodología de búsqueda de QTL se rechazó la hipótesis nula, de ausencia de QTL, con una probabilidad mucho mayor que la ofrecida con el LA y se consiguió reducir el CI a una región de 13 cM.

Siguiendo la línea de resultados obtenidos por otros autores, nuestros resultados indican que la aproximación basada en un análisis combinado LDLA mejora la resolución de mapeo de QTL (Meuwissen et al. 2002; Hernández-Sánchez et al., 2010), ya que en esta metodología se tienen en cuenta las relaciones genéticas existentes entre los cabezas de pedigrí. A pesar de que la densidad de marcadores en estudios basados en LDLA debería ser mayor que utilizando solo LA, Hernández-Sánchez et al. (2010) mostraron que esto no es un requisito en este tipo de análisis. Así, los resultados obtenidos en este trabajo muestran que la densidad de marcadores utilizada en este caso en la raza Churra parece ser suficiente para aprovechar las ventajas que ofrece esta metodología.

La búsqueda de genes candidatos posicionales en el intervalo redefinido de 13 cM en OAR3 mostró la presencia de varios genes, además de los descritos por Gutiérrez-Gil et al. (2009) (*HDAC7*, *VDR* y *ENDOU*). Entre ellos los genes *IGFBP6* (*insulin-like growth factor binding protein 6*), *AQP5* (*aquaporin 5*), *ARF3* (*ADP-ribosylation factor 3*), *LALBA* (lactalbumin, alpha-) e *IRAK4* (*interleukin-1receptor-associated kinase 4*) destacaron como los candidatos posicionales más relacionados con la producción de leche. Entre todos estos, el gen que codifica para la proteína alfa-lactoalbúmina (LALBA) parece ser un buen candidato funcional en relación al QTL objeto de estudio en este caso, ya que ésta constituye una de las proteínas séricas mayoritarias en la leche. En los años 1990 se estudiaron polimorfismos en este gen como posibles candidatos a ser utilizados en selección asistida por marcadores en las

especies productoras de leche. En el ganado vacuno, se ha descrito un polimorfismo asociado a caracteres de producción de leche (Lundén y Lindersson, 1998; Schopen et al. 2011), pero en el ovino, hasta el momento, a pesar de haberse detectado un polimorfismo en la región promotora de este gen (Chiofalo y Micari, 1987), no se han realizado estudios de asociación entre este gen y caracteres de producción de leche.

Los resultados obtenidos en el estudio de replicación y refinamiento del QTL localizado en OAR3 constituyen un primer paso para conocer la naturaleza de los efectos identificados sobre la producción de leche en la raza Churra en este cromosoma. El carácter porcentaje de proteína en producción ovina es importante ya que constituye uno de los objetivos de selección en esta raza dada su influencia sobre el rendimiento quesero de la leche. Para la aplicación de los resultados obtenidos en programas de selección sería necesario el estudio en profundidad de la región QTL redefinida para identificar la mutación causante del efecto e incluirla en la valoración de sementales.

Objetivo 2. Evaluación de la utilidad del *OvineSNP50 BeadChip* como herramienta genómica en la raza Churra.

Los dos trabajos descritos anteriormente constituyen la replicación, utilizando marcadores microsatélite y análisis de ligamiento, de algunos de los resultados descritos por Gutiérrez-Gil et al. (2009). Los avances en el proyecto de secuenciación del genoma ovino, a cargo del ISGC, permitieron que en el año 2009 estuviese disponible el *OvineSNP50 BeadChip*. La Población 3 anteriormente descrita, que incluía un total de 1.696 animales distribuidos en 16 familias de medio-hermanas, fue genotipada en su totalidad con este chip a través de servicios de genotipado de laboratorios externos. En el momento de obtención del conjunto completo de genotipos, estaba disponible la v2.0 del genoma de referencia ovino. Así, las posiciones de este borrador del genoma ovino correspondientes a los SNPs incluidos en el chip fueron utilizadas en los diferentes análisis realizados con estos genotipos.

Sobre este conjunto de datos se realizó un control de calidad tanto por individuo como por SNP y únicamente los animales y marcadores que pasaron dicho control se utilizaron en los análisis posteriores. El primer paso del control de calidad se realizó en base al parámetro *GenCall Score* (GCscore) utilizando el programa GenomeStudio (Illumina Inc. San Diego, CA). Este parámetro estima la calidad del genotipado y si los datos brutos obtenidos muestran

un valor de GCscore inferior a 0,15, se consideran como genotipos ausentes. El número total de marcadores obtenidos después de este paso fue de 54.241 SNPs para un total de 1.696 individuos. En el siguiente paso de este control de calidad, siguiendo las indicaciones de Anderson et al. (2010), se eliminan los animales con un porcentaje de marcadores genotipados inferior al 90%, eliminando así las muestras con ADN de baja calidad. Posteriormente, se utilizaron parámetros de calidad por SNP para eliminar los marcadores que mostraban un porcentaje de genotipos inferior al 95%, un MAF por debajo de 0,05 o una probabilidad de desviación del equilibrio Hardy-Weinberg inferior a 0,00001. Finalmente se retuvieron los genotipos de 1.681 animales y 43.784 SNPs para los posteriores análisis descritos en la presente Tesis Doctoral.

Como segundo objetivo de esta Tesis Doctoral y antes de utilizar el *OvineSNP50 BeadChip* en distintos tipos de análisis, se nos planteaba la necesidad de evaluar su aplicación en la raza objeto de estudio, en nuestro caso la raza Churra.

En primer lugar, con el objetivo de evaluar la calidad del mapa físico aportado por la secuencia de referencia v2.0, se planteó la comparación de éste con un mapa de ligamiento construido para la población de raza Churra objeto de estudio, en base a los genotipos obtenidos con el *OvineSNP50 BeadChip*. Para ello construimos un mapa de ligamiento basado en las meiosis de los machos de la Población 3 utilizando el orden de los marcadores de la versión 2.0 del genoma ovino. Debido a los requerimientos computacionales que supone crear un mapa de ligamiento de tal densidad sin información previa, lo que hicimos fue asumir el orden físico aportado por la v2.0 del genoma ovino de referencia y estimar, a partir de ahí, la frecuencia de recombinación entre marcadores adyacentes (opción *fixed* de CRI-MAP).

Los resultados obtenidos en este análisis mostraron que existen diferencias en cuanto a la frecuencia de recombinación entre marcadores a lo largo del genoma ovino. También se observaron diferencias en la correspondencia cM/Mb entre los distintos cromosomas, con un rango que varió entre 1,37 en el OAR2 y 2,50 en el OAR20. Estas diferencias tienen, fundamentalmente, tres causas: (i) calidad del mapa físico variable a lo largo del genoma ovino; (ii) tamaño del pedigrí estudiado; (iii) verdaderas diferencias en patrones de recombinación a lo largo del genoma.

Hay que tener en cuenta que las técnicas de secuenciación de segunda generación producen millones de lecturas de tamaño corto (35-500 bp) que se alinean a lo largo del genoma. Estas secuencias forman *contigs*, *scaffolds* y *super-scaffolds*, entre los que puede haber huecos y las distancias físicas estar sub- o sobrestimadas. En el caso del genoma ovino, aún en estado de borrador inicial, este fenómeno puede claramente afectar a la calidad del mapa físico. Además, la calidad del mapa de ligamiento obtenido también está limitada por el reducido tamaño del pedigrí analizado. En este sentido, cabe destacar el mapa de ligamiento que está siendo desarrollado por el grupo de JF Maddox, en Australia, en base a genotipos obtenidos con el *OvineSNP50 BeadChip* para un amplio número de las poblaciones ovinas incluidas dentro del proyecto *SheepHapMap*. Por otra parte, las diferencias en la correspondencia cM/Mb entre distintos cromosomas ya se han puesto de manifiesto en otras especies de mamíferos (Yu et al., 2001; Liu et al 2009) existiendo regiones donde la recombinación es mucho mayor que la media del genoma (selvas de recombinación) y otras donde casi no se observan sucesos recombinatorios (desiertos de recombinación).

Los resultados obtenidos en nuestro análisis muestran que la relación cM/Mb media a lo largo del genoma es de 1,85. El refinamiento del mapa físico en futuras actualizaciones del genoma, puede ayudar a que este ratio se acerque más al ratio 1 cM ~ 1 Mb, encontrado en las especies con un genoma de referencia mucho más elaborado. Por el momento, es necesario tener en cuenta, a la hora de interpretar resultados basados en esta versión del genoma, que la secuencia de referencia se encuentra en un estado inicial de desarrollo y esto puede llevar a una incorrecta asociación de regiones genómicas con caracteres productivos. Por todo ello, parece recomendable que siempre que se obtenga una asociación basada en un análisis de ligamiento con los marcadores del chip se compruebe que ésta no se encuentra en una región del genoma cuya calidad es dudosa.

Objetivo 3. Utilización del *OvineSNP50 BeadChip* para el análisis de la estructura del genoma de la raza Churra mediante el análisis de la extensión del desequilibrio de ligamiento en esta raza.

Para continuar con la valoración del *OvineSNP50 BeadChip* como herramienta para estudiar la arquitectura genética del genoma de la raza ovina Churra, nos planteamos determinar si el número de marcadores incluidos en este chip es suficiente para llevar a cabo análisis de mapeo fino y una posible puesta a punto de la Selección Genómica en la raza

Churra. Para ello utilizamos los genotipos del *OvineSNP50 BeadChip* para analizar la estructura de desequilibrio de ligamiento, LD, entre marcadores y los bloques haplotípicos a lo largo del genoma. Asimismo, utilizamos esta información sobre el LD para estimar el tamaño efectivo (*Ne*) de la población de Churra a lo largo de la historia y con los datos genotípicos calculamos los niveles de consanguinidad en esta raza.

Dadas las dificultades que entraña la comparación entre distintos estudios de cálculo de LD, en nuestro estudio utilizamos dos parámetros de medida del LD, r^2 y *D'*, con el fin de facilitar la posterior comparación de nuestros resultados con estudios previos descritos tanto en ganado ovino como en otras especies.

En el ganado ovino, los estudios iniciales valorando la extensión del LD a lo largo del genoma estuvieron basados en marcadores microsatélite y determinaron que el LD se extiende a lo largo de distancias tan amplias como 20 cM, existiendo diferencias entre razas (McRae et al., 2002; Meadows et al., 2008). En el caso de ovinos salvajes, también se comprobó que el LD es extenso (*half-length* r^2 = 4,6 Mb), al contrario que en nuestro estudio (*half-length* r^2 = 2,5 Mb). En cambio, posteriormente, estudios llevados a cabo en la raza Sarda, basados en el chip de SNPs, muestran un nivel similar de LD en esta raza con respecto a nuestros resultados para la raza Churra con valores de r^2 de 0,072 para SNPs separados hasta 1 Mb (Usai et al., 2010). Dentro del marco del proyecto *SheepHapMap* se han publicado recientemente los resultados de la estimación del LD en 74 razas ovinas (Raadsma, 2010). Estos resultados se acercan más a los resultados obtenidos en nuestro estudio, observándose un nivel bajo de la extensión del LD en la especie ovina, en contraste con los datos procedentes de diversas razas de ganado vacuno (Villa-Angulo et al., 2009). En ese trabajo, además, se pone de manifiesto que la raza Churra muestra un nivel de LD inferior al resto de razas ovinas estudiadas dentro del proyecto *SheepHapMap* (Raadsma, 2010).

Existen diferencias en cuanto al nivel de LD a lo largo del genoma que han sido descritas en otras especies ganaderas, como el ganado vacuno de raza Holstein (Bohmanova et al., 2010; Qanbari et al., 2010), que pueden ser atribuidas a variaciones en las frecuencias de recombinación a lo largo del genoma, diferencias en la heterocigosis, deriva genética o que pueden ser efecto de la presión de selección (Qanbari et al., 2010). En la raza ovina Churra, el patrón de LD a lo largo del genoma está relacionado con la estructura de bloques haplotípicos, es decir, los cromosomas con una extensión del LD más amplia muestran una mayor longitud

de los bloques haplotípicos que los cromosomas con LD menos extenso. En esta raza, el 88% de los bloques detectados a lo largo del genoma contenía únicamente dos marcadores. Los estudios previos dentro del marco del proyecto *SheepHapMap* mostraron un número limitado de bloques a lo largo del genoma con variaciones entre razas, desde un 0,8% en Churra hasta un máximo de 21,8% en la raza ovina Soay (Raadsma, 2010). Las diferencias en los valores de LD con otras especies, como el ganado vacuno, también se reflejan en la estructura haplotípica mostrando una mayor frecuencia de bloques en dicha especie (Kim y Kirkpatrick, 2009; Qanbari et al., 2010).

Dado que la presencia de bloques haplotípicos puede estar relacionada con la presión de selección sobre ciertos caracteres, resulta interesante comparar las regiones donde aparecen más bloques con regiones descritas como portadoras de huellas de selección en la especie ovina. Estas regiones han sido descritas por Kijas et al. (2012) estudiando las 74 razas incluidas en el proyecto *SheepHapMap*. A pesar de que varios de los bloques haplotípicos detectados en la raza Churra se encuentran en regiones próximas a las portadoras de huellas de selección descritas por Kijas et al. (2012), ninguno de los SNPs señalados por dichos autores formaba parte de los bloques de LD detectados en la raza Churra. Por ejemplo, la región con una mayor cantidad de bloques en nuestros análisis, situada en el OAR10, se localiza en las inmediaciones del *locus polled*, relacionado con la presencia y forma de los cuernos en ganado ovino (Johnston et al., 2011). Sin embargo, ninguno de los SNPs implicados en dicho carácter formó parte de los bloques en esa región. El bloque de mayor longitud identificado en Churra, de 1,2 Mb en el OAR2, incluye la secuencia del gen *HERC2*, relacionado con la pigmentación en la especie bovina (Han et al., 2008), aunque es una región en la que no se han detectado huellas de selección en el ganado ovino (Kijas et al., 2012).

Las estimaciones de la extensión del LD en el genoma de las diferentes especies pueden utilizarse para el cálculo del tamaño efectivo de las poblaciones a lo largo de la historia (Sved, 1971; Hill y Robertson, 1981; Hayes et al., 2003). Este parámetro aporta información sobre la evolución de la población a lo largo del tiempo y, además, nos muestra los efectos que la selección produce sobre el tamaño efectivo de la misma. Para ello es necesario transformar la distancia física entre marcadores (Mb) en distancia genética (cM), por lo que en nuestro trabajo realizamos tres estimaciones basadas en ratios cM/Mb diferentes: (i) el ratio comúnmente aceptado 1 cM ~ 1 Mb; (ii) el ratio medio obtenido en el

estudio previo en Churra 1,85 cM ~ 1 Mb; (iii) el ratio específico obtenido en dicho estudio para cada cromosoma.

Los primeros datos de la existencia de la raza Churra datan de la Edad Media (siglo XIII), aproximadamente hace 800 años (Sánchez Belda y Sánchez Trujillano, 1986). Así, para nuestros análisis se estimó el *Ne* de la población de Churra en base a esta información histórica. De acuerdo a nuestras estimaciones, el *Ne* en esta raza ha descendido a lo largo del tiempo hasta que comenzó el programa de selección en esta raza en el año 1984. A partir de ese momento, no se han producido grandes cambios en cuanto al *Ne*. El valor de *Ne* calculado para hace 50 generaciones está de acuerdo con los valores obtenidos en los análisis realizados dentro del proyecto *SheepHapMap* (Kijas et al., 2012).

El *Ne* mínimo para asegurar una población viable con una base genética suficientemente amplia fue estimado por Meuwissen (2009) en un valor de 100 animales. En el caso de la raza ovina Churra, el *Ne* estimado para la última generación, hace 4 años, fue de 128 animales, por encima del umbral de viabilidad (Meuwissen, 2009). Este valor de *Ne* estimado en Churra es mucho mayor cuando se compara con el estimado en la raza Holstein (*Ne* = 50), aunque hay que tener en cuenta que está cerca del umbral de viabilidad por lo que se debieran tomar ciertas precauciones para asegurar la diversidad genética dentro de esta raza.

En el trabajo centrado en el estudio del LD a lo largo del genoma de la raza ovina Churra también se estimó la consanguinidad de la población analizada utilizando la información procedente del pedigrí y la derivada de la información molecular. La estimación de este parámetro basándonos en el pedigrí depende de la profundidad y veracidad del mismo. En nuestro análisis utilizamos el pedigrí disponible desde 1987, que incluía un total de 5.956 animales, y los resultados obtenidos mostraron que únicamente el 6% de los animales tienen cierto grado de consanguinidad. En cambio, utilizando datos moleculares se obtuvieron niveles superiores de consanguinidad. Hay que tener en cuenta que los resultados basados en datos moleculares muestran valores negativos para algunos de los animales debido a la ausencia de una población base sobre la que calcular las frecuencias alélicas en las que se basa el coeficiente de parentesco. Al utilizar las frecuencias de los animales genotipados, existen algunos que tienen un nivel de homocigosis inferior a la media poblacional por lo que el resultado es un coeficiente negativo que no podemos interpretar.

Los mayores niveles de consanguinidad estimados en base a los datos moleculares, comparados con los calculados a través del pedigrí, podrían deberse, tal y como apuntaban Li et al. (2011), a que existe un sesgo que tiende a sobrestimarlos al asumir que no hay una estructura poblacional concreta en la población utilizada, cuando lo cierto es que el muestreo de los animales analizados se ha realizado siguiendo un diseño hija, destinado a realizar experimentos de detección de QTL. En este trabajo se estimaron los coeficientes de parentesco molecular mediante tres metodologías en base a la varianza del genotipo aditivo, en base al exceso de homocigosis, y en base a la correlación entre isogametos. La correlación entre los resultados obtenidos en base a las tres aproximaciones fue, como media, de 0,27. Existe un nivel crítico de consanguinidad, 6,25% obtenido al cruzar primos (Li et al., 2009b), que fue superado por un porcentaje de animales variable entre los diferentes métodos, pero en cualquier caso inferior al estimado en la raza Finnsheep (Li et al., 2009b). Aunque la comparación entre métodos se complica por la ausencia de una población base en la que estimar las frecuencias alélicas, los resultados descritos en este trabajo ponen de manifiesto que los métodos moleculares son capaces de calcular el coeficiente de parentesco entre animales en genealogías complejas o en ausencia de un pedigrí fiable.

Por último, otro de los objetivos de este trabajo, además de la caracterización de la estructura del genoma de la raza Churra, es la evaluación del chip de SNPs como herramienta para el mapeo fino de regiones con influencia sobre caracteres productivos y también para llevar a cabo predicciones del valor genético de los animales en base a la información molecular, lo que constituye las bases de la Selección Genómica. En ganado vacuno, basándose en la extensión media del LD, la densidad de marcadores, el tamaño efectivo de la población y la longitud total del genoma, Goddard (2009) y Daetwyler et al. (2010) determinaron el número de marcadores necesarios para obtener una buena precisión en la predicción de los valores genéticos en esta especie. Para ese mismo objetivo, en el caso de la raza Churra y dado el bajo nivel de LD observado, para un *Ne* de 128 animales, una densidad de 20 SNPs por segmento, la misma que la utilizada en el ganado vacuno, y una longitud del genoma de 30 Morgans, el número de SNPs necesarios serían en torno a los 95.000, teniendo en cuenta el porcentaje de SNPs eliminados en los distintos controles de calidad realizados. En base a estos resultados podríamos concluir que sería necesario duplicar la densidad ofrecida por el chip utilizado en esta Tesis Doctoral, para llevar a cabo de manera exitosa

tanto estudios de mapeo fino de regiones asociadas con caracteres de interés económico como para la utilización de esa información genómica en la predicción de valores genéticos.

Objetivo 4. Mapeo fino de los QTL confirmados y análisis GWAS para la detección de nuevas regiones genómicas no detectadas anteriormente.

A pesar de las conclusiones del trabajo anterior, según el cual el chip de SNPs disponible (*OvineSNP50 BeadChip*) no ofrece la densidad suficiente para la implementación eficiente de la GS en esta especie, hay que valorar el sustancial incremento de densidad de mapeo que esta herramienta ofrece a la comunidad científica, y a nuestro grupo de investigación, en comparación con la densidad de los mapas de microsatélites utilizados para los barridos genómicos que se llevaron a cabo en la raza Churra en la primera década del siglo XXI. En base a esto, el cuarto objetivo de esta Tesis Doctoral se ha centrado en la identificación de regiones portadoras de QTL para los caracteres de producción de leche en el genoma ovino mediante el uso de esta herramienta genómica. Además, la densidad de marcadores del chip nos ha ofrecido la posibilidad de realizar un análisis basado en asociación (GWAS, *Genome-Wide Association Study*). Así pues, con el objetivo general propuesto, nos proponemos, por una parte, identificar regiones que no hayan sido previamente identificadas por las limitaciones ya comentadas de los estudios de ligamiento basados en poblaciones comerciales y en mapas de baja densidad, y, por otra, la confirmación de regiones previamente detectadas en el barrido genómico descrito por Gutiérrez-Gil et al. (2009). Sería de esperar que, dado el aumento de la resolución del mapeo, este análisis sirviera también para el mapeo fino de algunas de las asociaciones detectadas.

El GWAS descrito en el quinto trabajo incluido en la presente memoria de Tesis Doctoral, se llevó a cabo en base al análisis de la Población 3 perteneciente al núcleo de selección de ANCHE, previamente utilizada en la replicación de resultados de OAR3, la construcción de un mapa de ligamiento y el estudio del LD a lo largo del genoma. Los resultados obtenidos en el análisis de asociación inicialmente realizado muestran una asociación significativa a nivel genómico entre una región del OAR3 y los caracteres PP y FP, además de 14 regiones significativamente asociadas a nivel cromosómico. A continuación describiremos brevemente las regiones significativas a nivel cromosómico para centrarnos posteriormente, y en mayor detalle, en el mapeo fino realizado en la región del OAR3.

Entre los resultados obtenidos a nivel de significación cromosómica, se seleccionaron las regiones con influencia sobre más de un carácter, OAR2 y OAR20, como las más interesantes para la búsqueda de genes candidatos. En el OAR2, se detectó un QTL con influencia sobre MY, PY y FY en el primer tercio del cromosoma, con resultados significativos localizados entre 42,7 a 63,2 Mb. Un QTL con influencia sobre PP y FP había sido descrito en esta región por Gutiérrez-Gil et al. (2009), aunque en ese caso el CI cubría prácticamente el cromosoma completo. La región ortóloga bovina, BTA8, ha sido descrita como portadora de un QTL con influencia sobre PY (Daetwyler et al., 2008).

La búsqueda de genes candidatos posicionales en esta región de OAR2, en base a la v2.0 del *Ovine Genome Assembly*, produjo múltiples resultados como *TGFBR1* (*transforming growth factor C beta receptor 1*), *IGFBPL1* (*Insulin-like growth factor binding protein-like 1*), *CD72* (*Cytokine 72*), *STOML2* (*stomatin (EPB72)-like 2*), *GalNAc-T12* (*UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 12*), *TLN1* (*talin 1*), *PSAT1* (*phosphoserine aminotransferase 1*), *GCNT1* (*glucosaminyl (N-acetyl) transferase 1 C core 2*) and *RFK* (*riboflavin kinase*). Entre estos genes, parece que el candidato funcional más interesante, desde nuestro punto de vista, podría ser el *IGFBPL1*, ya que la proteína que codifica contiene un dominio similar al existente en las proteínas de unión al factor de crecimiento similar a la insulina (IGFBP). A ese dominio se unen los factores de crecimiento tipo insulina (IGFs), un tipo de factores para los que en ganado vacuno se ha sugerido que influyen sobre la producción de leche (Mullen et al., 2011).

En el caso del OAR20, la región significativa detectada influye sobre los caracteres PP, FP y FY, cubriendo la región situada entre 23,7 y 29,6 Mb. En esta región se incluye el *MHC-II* (*complejo mayor de histocompatibilidad II*) y, en los alrededores del mismo, ya se había detectado un QTL con influencia sobre el carácter FP (Gutiérrez-Gil et al., 2009), confirmado en el primer trabajo de la presente Tesis Doctoral, y para los caracteres MY, FY, PY y FP en una población experimental de Sarda x Lacaune (Barillet et al., 2006). Además, la región ortóloga bovina, en BTA23, ha sido descrita como portadora de QTL para caracteres de producción de leche (Plante et al., 2001) y al menos un gen localizado en esta región, *BoLA-DRB3*, ha mostrado asociación con caracteres lecheros (Sharif et al., 1998). Sin embargo, en esta fase del estudio, parece que para entender la arquitectura genética de estas regiones de posible interés, localizadas en OAR2 y OAR20, se necesitarían estudios y análisis adicionales basados en el incremento de la densidad de marcadores principalmente.

Esperamos, que la ya próxima comercialización del chip ovino de alta densidad (*Ovine HD BeadChip*), permita confirmar y reducir a regiones más precisas las localizaciones de estos efectos genéticos identificados en este primer GWAS descrito en el ganado ovino de leche.

El resto de asociaciones identificadas a nivel cromosómico en este análisis inicial son SNPs aislados asociados con uno de los caracteres objeto de estudio. Cabe pensar, por un lado, que estas asociaciones aisladas podrían ser falsos positivos, debido al estado de "borrador" del genoma ovino y la posible confusión en la localización de algunos SNPs. Alternativamente, podría tratarse de asociaciones verdaderas entre genotipo y fenotipo que requieren de un diseño experimental más potente para ser detectadas. Apoyando esta segunda hipótesis está el ilustrativo caso del OAR6, en el que el SNP asociado con el carácter PP se localiza en la región de las caseínas (*CSN1S1*, *CSN1S2* y *CSN2*), proteínas mayoritarias en la leche. Se han descrito asociaciones entre polimorfismos en estos genes y el contenido en caseínas en leche (Moioli et al., 2007), así como la presencia de QTL con influencia sobre caracteres de producción de leche en esta región (Díez-Tascón et al., 2001; Barillet et al., 2006; Moioli et al., 2007). Otro plausible gen candidato cerca del SNP aislado asociado con uno de los caracteres analizados se identificó en relación a la asociación identificada para el carácter FY en el OAR25. En este caso, la región de interés incluye al gen *NRG3* (*Neuregulin 3*), relacionado con el desarrollo mamario en la especie humana (Howard y Ashworth, 2006).

En el GWAS descrito en este trabajo, los únicos resultados significativos a nivel genómico se localizaron en torno a la región de 137 Mb en el OAR3. Los caracteres influenciados por este QTL fueron PP y FP, aunque el primero mostró valores de probabilidad mucho más significativos (P = 3,78x10^{-26} para PP; P = 1,8x10^{-10} para FP). La influencia de esta región sobre PP en la raza Churra había sido previamente descrita por Gutiérrez-Gil et al. (2009) y los resultados se replicaron en una muestra independiente durante la presente Tesis Doctoral.

Para ambos caracteres significativos, el máximo valor del estadístico se identificó para un SNP del chip comercial, OAR3_147028849, localizado en el tercer intrón del gen que codifica para la proteína LALBA. Este gen había sido propuesto como candidato funcional y posicional para este QTL ya que, como se ha destacado previamente, la proteína LALBA es una de las mayoritarias del suero de la leche y forma parte de la lactosa sintasa. Este enzima es necesario para la producción de lactosa en leche, el componente que aporta la presión

osmótica que hace que el agua se desplace hacia los alveolos mamarios, por lo que su función es crítica en la producción y composición de la leche. Los ratones *knock-out* para este gen, producen una leche tan densa, con tan escasa cantidad de agua, que las crías no pueden mamar (Stinnakre et al., 1994).

En ganado vacuno, se ha descrito un polimorfismo (ULGR_SNP_U63109_1966) en la región promotora de este gen con influencia sobre los caracteres porcentaje de lactosa y cantidad de LALBA en leche en esta especie (Lundén y Lindersson, 1998; Schopen et al., 2011). Este mismo SNP se detectó en nuestra población, *LALBA*_g.82G>A, pero algunos de los machos segregantes para el QTL resultaron homocigotos para este SNP, por lo que quedó descartado como posible mutación causal y no fue genotipado en el conjunto de la población. Aunque en la oveja no se han publicado estudios previos describiendo en detalle polimorfismos en el gen *LALBA*, o valorando su posible relación con caracteres de producción de leche, la comparación *in silico* de las dos secuencias disponibles en GenBank para este gen ovino nos permitió la identificación de tres variaciones puntuales entre ambas. De las tres variaciones puntuales detectadas entre la secuencia de ARNm de Gaye et al. (1987) y la utilizada como referencia en nuestros análisis (GenBank AB052168), una de ellas se corresponde con el único polimorfismo detectado en la raza Churra en la región codificante de este gen (*LALBA*_g.242T>C).

En base a los resultados, y a su papel biológico, el gen *LALBA*, fue sometido a análisis adicionales con el fin de testar la posible causalidad con respecto al efecto genético sobre los caracteres FP y PP descrito en la oveja Churra. Con este fin, se inició un primer estudio destinado a la identificación de mutaciones en este gen. Para ello se secuenciaron los 16 machos de la Población 3 para obtener la secuencia completa de los cuatro exones del gen, los intrones, parte de la región promotora y parte de la región 3'UTR, secuenciándose en total 3.067 pb. Este análisis de secuenciación permitió la identificación de 31 polimorfismos. Únicamente uno de ellos, el SNP *LALBA*_g.242T>C, se localizó en la región codificante, en el primer exón, identificándose como responsable de un cambio aminoacídico en el vigésimo séptimo aminoácido de la proteína, una sustitución de valina por alanina (Val27Ala). La inclusión de las secuencias normal y mutada en el programa PolyPhen (http://genetics.bwh.harvard.edu/pph/index.html) para la predicción de los posibles cambios en la estructura proteica, no produjo resultados significativos, lo que indica que esta mutación es conservadora respecto a la actividad proteica. Además, se estudió el grado de conservación

entre especies de la región de la proteína donde se produce el cambio Val27Ala, observándose que el aminoácido mutado coincide con un residuo de valina en rumiantes y ratón, mientras que encontramos un residuo de leucina en otras especies como humanos, caballos, cerdos y macacos. Por esta razón, asumimos que el alelo *LALBA*_g.242T, que se corresponde con el residuo de valina, es la variante ancestral, mientras que el alelo *LALBA*_g.242C, responsable de la presencia de alanina en la proteína, es el mutado.

Los resultados incluidos en este trabajo muestran diferentes pruebas de la causalidad de la mutación *LALBA*_g.242T>C como la asociación significativa de los genotipos de la misma con los caracteres PP y FP, la concordancia observada entre los genotipos de los machos identificados como segregantes con el estatus del QTL para los mismos (machos *Qq)* y la cancelación de los resultados significativos en la región al añadir el genotipo de este SNP como factor fijo dentro del modelo de detección de QTL. El efecto de sustitución alélica, que sólo fue significativo para los dos caracteres anteriormente comentados, muestra el mismo signo, positivo, para los caracteres PP, FP, PY y FY, mientras que es negativo, aunque mínimo, sobre el carácter MY. Por lo tanto, el QTN descrito produce un aumento en la concentración de sólidos totales en la leche. Así, estos resultados en los que el efecto es mayor sobre PP que sobre FP, sugieren que el efecto de este gen es más importante sobre el primer carácter, como ya se había sugerido anteriormente (Gutiérrez-Gil et al., 2009). Apoyando la hipótesis de una diferente influencia de la proteína LALBA sobre los distintos componentes de la leche, Boston et al. (2001) detectaron, en ratones transgénicos que producían 5 a 15 veces más LALBA de lo normal, diferencias significativas únicamente en el porcentaje de proteína en leche ($P < 0,05$), siendo los porcentajes de grasa, lactosa y sólidos totales no significativamente diferentes entre los grupos transgénico y control. Además, los diferentes resultados obtenidos para los caracteres PP y FP pueden ser debidos a los valores de heredabilidad estimados para dichos caracteres en ganado ovino (Legarra y Ugarte, 2001; Othmane et al., 2002b), ya que una menor heredabilidad, como la obtenida para FP, pone de manifiesto la existencia de factores no genéticos que no han sido tenidos en cuenta en el modelo de estimación de los fenotipos.

Las frecuencias alélicas de este SNP muestran una mayor frecuencia en la población estudiada del alelo mutado, *LALBA*_g.242C, con respecto a la variante ancestral. Como se ha descrito en la Revisión Bibliográfica, el programa de mejora genética de la raza Churra incluye, desde 1998, el carácter PP como objetivo de selección, con un 20% de peso en el

valor genético de los machos (ANCHE). Por tanto el aumento en la frecuencia del alelo *LALBA*_g.242C, podría deberse a que el esquema de selección que se está implementando en esta raza, aunque basado en el modelo infinitesimal de la genética cuantitativa, es eficiente en la captación de las mutaciones segregantes que tienen efectos directos e importantes sobre los caracteres objeto de selección y favorece el aumento de la frecuencia del alelo de interés para las mismas.

Hay que tener en cuenta que este estudio, propone, por primera vez en el ganado ovino, un posible QTN responsable de un efecto genético sobre caracteres de producción de leche. La confirmación de la causalidad de una mutación requiere del cumplimiento de una serie de requisitos, descritos en la Revisión Bibliográfica de esta Tesis Doctoral. En este trabajo hemos demostrado que este QTN potencial: (i) se localiza dentro del CI estimado tanto por LA como por LD; (ii) la función del gen propuesto como causal se relaciona con el carácter; (iii) existe concordancia entre los genotipos de la mutación y el QTL en los machos segregantes; (iv) las frecuencias alélicas concuerdan con los objetivos de selección implementados en esta raza; (v) el efecto del QTL desaparece al incluir los genotipos de esta mutación en la ecuación de detección de QTL. Existen otra serie de requisitos, como la validación de los resultados en otras poblaciones o la realización de pruebas funcionales, que serán necesarios para probar de manera conclusiva la causalidad sugerida aquí para la mutación *LALBA*_g.242T>C.

A falta de pruebas funcionales que confirmen la verdadera naturaleza de los efectos detectados y la causalidad del SNP *LALBA*_g.242T>C, y sin poder descartar, totalmente, que esta mutación podría encontrarse en completo LD con el verdadero QTN, el trabajo realizado ofrece claras muestras de la importancia de este gen en el control genético de la producción lechera en el ganado ovino, al menos en la raza Churra. En esta raza, además, se podría plantear de manera directa, y tras valorar sus beneficios, el posible uso de esta mutación como marcador "ideal" en el programa de selección de la raza ovina Churra. El uso del mismo en otras razas ovinas dependerá, como es lógico, de la validación de la asociación aquí descrita en las mismas, y de las diferencias en LD, si las hubiera, con la verdadera mutación causal.

Los resultados presentados y discutidos en la presente Tesis Doctoral muestran la evolución que se está produciendo en el campo de la biotecnología genética ovina en estos primeros años del siglo XXI. Tomando como punto de partida los resultados previos descritos

por Gutiérrez-Gil et al. (2009), y utilizando una población independiente de la raza ovina Churra, hemos llevado a cabo un mapeo fino de las regiones previamente descritas como portadoras de QTL con influencia sobre la producción de leche. Los primeros estudios se basaron en marcadores microsatélite y, posteriormente, se aprovechó la aparición del *OvineSNP50 BeadChip* para mejorar la precisión de mapeo. El objetivo final de estos estudios de mapeo fino es encontrar la mutación causal o QTN para poder utilizarlo en los programas de mejora. En este caso, se han recopilado pruebas fehacientes de la directa influencia de la mutación *LALBA*_g.242T>C sobre el carácter PP en la raza Churra. La posibilidad de uso de esta información molecular en el programa de mejora de la raza Churra deberá ser valorada por el comité de seguimiento del plan de mejora genética de dicha raza.

TRABAJOS ADICIONALES

Kijas J.W., Miller J.E., Hadfield T., McCulloch R., Garcia-Gamez E., Porto Neto L.R., Cockett N. (2012). Tracking the emergence of a new breed using 49,034 SNP in sheep. *PLoS ONE* 7(7), e41508.

García-Gámez E., Reverter A., Whan V., McWilliam S.M., Arranz J.J., International Sheep Genomics Consortium, Kijas J. (2011). Using regulatory and epistatic networks to extend the findings of a genome scan: identifying the gene drivers of pigmentation in Merino sheep. *PLoS ONE* 6(6), e21158.

Durante el desarrollo de la presente Tesis Doctoral, además de los estudios realizados para completar los objetivos planteados incialmente, se han llevado a cabo dos Trabajos Adicionales. Dichos trabajos se encuadran dentro del proyecto *SheepHapMap*, en el que participa nuestro grupo de investigación, y han sido realizados durante una estancia de investigación en el grupo *Livestock Industries* perteneciente al *Commonwealth Scientific and Industrial Research Organisation* (CSIRO) en Australia encuadrada dentro del periodo formativo de la presente Tesis Doctoral.

La disponibilidad, a partir de 2008, del chip de SNPs ovino descrito en la Revisión Bibliográfica y utilizado para la consecución de algunos de los objetivos de esta Tesis Doctoral, además de permitir el estudio de caracteres cuantitativos relacionados con la producción ovina, ha permitido la disección de caracteres monogenéticos, tales como enfermedades mendelianas como la epidermólisis bullosa (Mömke et al., 2011) microftalmia (Becker et al., 2010), el raquitismo (Zhao et al., 2011) y la condrodisplasia (Zhao et al., 2012). El genotipado del chip en un total de 74 razas ovinas dentro del marco del proyecto *SheepHapMap* ha permitido caracterizar la diversidad genética de las poblaciones ovinas a nivel mundial. La información derivada de este proyecto ha permitido también estudiar la extensión del LD en el genoma ovino (Raadsma, 2010) e identificar huellas de selección en el genoma de las modernas razas ovinas (Kijas et al., 2012).

En el trabajo publicado por Kijas et al. (2012) se resumen los resultados de los análisis mencionados. Una de las conclusiones más importantes de este estudio es que los tamaños efectivos de las diferentes razas ovinas y, por tanto, la diversidad genética dentro de las mismas, se han mantenido en niveles mayores a los estimados en otras especies animales como el ganado vacuno o los perros. En base a estos resultados nos planteamos la necesidad de cuantificar cómo de diferentes, genéticamente, deben ser dos razas ovinas para ser consideradas razas separadas y no subpoblaciones de una misma raza. En relación a este tema, en el trabajo titulado *"Tracking the emergence of a new breed using 49,034 SNPs in sheep"*, se estudió el caso concreto de la raza americana *Gulf Coast Native* (GCN) en la que existen dos poblaciones separadas geográficamente y entre las que existen, además, ciertas diferencias morfológicas.

Esta raza ovina, GCN, tiene su origen en los rebaños llevados a América por exploradores procedentes de España alrededor del año 1500 y se considera que desciende de

la raza Churra. Sin embargo, la lana fina que muestran los animales de esta raza también sugiere cierta influencia de la raza Merina en su origen. Son animales de tamaño pequeño-mediano, de capa blanca aunque en algunas ocasiones pueden tener el vellón de color marrón más o menos oscuro. Dentro de la raza GCN, existen dos líneas o poblaciones distintas *Florida Native* (FN) *y Louisiana Native* (LN). Se ha sugerido que existen diferencias fenotípicas entre ambos grupos ya que podrían proceder de animales de regiones separadas. El grupo LN, se localiza en Arkansas, Luisiana, Missouri, Mississippi y Texas y, en general, son animales blancos con escasa pigmentación. La línea FN, localizada en Alabama, Florida y Georgia, está formada por animales que pueden presentar pigmentación marrón o negra en la cara y las patas (JE Miller, comunicación personal). Además, los animales pertenecientes al grupo LN parecen tener las patas más cortas y el cuerpo más grande que los incluidos dentro del grupo FN.

Con el objetivo de determinar si las dos subpoblaciones podrían considerarse razas diferentes se estudió la diversidad genética entre estos dos grupos, comparándolas además con un subconjunto de animales de 12 razas, procedentes de las 74 razas incluidas en el proyecto S*heepHapMap*. También se analizó la subestructura poblacional y las diferencias concretas que podemos encontrar en el genoma de las subpoblaciones LN y FN.

Por último, en el trabajo titulado *"Using regulatory and epistatic networks to extend the findings of a genome scan: identifying the gene drivers of pigmentation in Merino sheep"* se presenta la combinación de datos de expresión génica con un GWAS con el objetivo de facilitar la disección del fenotipo *piebald*, un tipo de pigmentación observado en la raza Merina. Desde un enfoque global en base a la biología de sistemas, que se fundamenta en que los organismos vivos funcionan como un conjunto, estando las diferentes rutas metabólicas relacionadas entre sí (Woelders et al., 2011), este estudio conjugó información de muy diversas fuentes para construir redes de interacción génica. Así, se analizaron conjuntamente los genotipos del *OvineSNP50 BeadChip* y datos de un *microarray* de expresión, usando como nexo de unión entre ambos los genes localizados cerca de un SNP genotipado y a su vez incluidos dentro del *microarray* de expresión. Además, se utilizó información relativa a la regulación génica a través de factores de transcripción para obtener una visión más global de las posibles interacciones génicas presentes en el organismo. Este trabajo representa un ejemplo en el que los resultados de un GWAS basado en los 49.034 SNPs del chip ovino se

combinan con datos expresión con el objetivo de llegar más allá en la interpretación de resultados de caracteres complejos.

Australia es el primer productor de lana a nivel mundial, con una media de producción de 500.000 Toneladas de lana por año entre los años 1992 y 2010 (FAOSTAT). La producción de lana ovina en Australia se basa en la explotación de animales cuya lana es apreciada por su blancura, el reducido diámetro y la suavidad de la fibra. La explotación de la raza Merina en pureza, constituye el 72% de las cabezas de ganado ovino del país (AWI). En los animales de raza Merina, se define el fenotipo *piebald* como la presencia de áreas de lana negra de distribución no simétrica en el vellón de lana blanca (Fleet y Smith, 1990). Los animales que presentan este fenotipo se convierten en inservibles para la producción de lana ya que a la hora de venderla, cuando aparecen fibras negras visibles entre el vellón de lana blanca, los precios de la misma se reducen de manera sustancial (Fleet et al., 2002). El estudio de los patrones de pigmentación en humanos y ratón ha identificado al menos 100 genes implicados en dichos patrones, por lo que muchos de los fenotipos relacionados con la pigmentación se consideran caracteres complejos (Seo et al., 2007). El fenotipo *piebald* ha sido estudiado en humanos (Thomas et al., 2004), perros (Karlsson et al., 2007) y ovejas (Brooker y Dooling, 1969). En el ganado ovino, estudios realizados hace 40 años comprobaron que la herencia del carácter no coincide con un modo mendeliano simple (Brooker y Dooling, 1969).

La aplicación de las teorías de biología de sistemas en organismos modelo se utiliza para estudiar los procesos biológicos que tienen lugar en los mismos. En la especie humana se han diseccionado caracteres complejos, casi siempre relacionados con enfermedades, por ejemplo problemas cardiovasculares (Diez et al., 2010), y en las especies ganaderas, como en el caso del ganado ovino, nos puede ayudar a explicar el funcionamiento de los organismos y las interacciones que se producen entre rutas metabólicas o entre genes con influencia sobre caracteres productivos (Woelders et al., 2011).

TRACKING THE EMERGENCE OF A NEW BREED USING 49,034 SNP IN SHEEP

James W Kijas[1], James E Miller[2], Tracy Hadfield[3], Russell McCulloch[1], **Elsa Garcia-Gamez**[1,4], Laercio R Porto Neto[1,5] and Noelle Cockett[3]

[1] Division of Livestock Industries, Commonwealth Scientific and Industrial Research Organisation, Brisbane, Queensland, Australia. [2] Department of Pathobiological Sciences, School of Veterinary Medicine and Department of Veterinary Science, Louisiana State University, Baton Rouge, Louisiana, USA. [3] Department of Animal, Dairy, and Veterinary Sciences, Utah State University, Logan, Utah, USA. [4] Departamento de Producción Animal, Universidad de León, León, Spain. [5] School of Veterinary Science, The University of Queensland, Gatton, Queensland, Australia.

PLoS ONE 7(7): e41508.

ABSTRACT

Domestic animals are unique in that they have been organised into managed populations called breeds. The strength of genetic divergence between breeds may vary dependent on the age of the breed, the scenario under which it emerged and the strength of reproductive isolation it has from other breeds. In this study, we investigated the Gulf Coast Native breed of sheep to determine if it contains lines of animals that are sufficiently divergent to be considered separate breeds. Allele sharing and principal component analysis (PCA) using nearly 50,000 SNP loci revealed a clear genetic division that corresponded with membership of either the Florida or Louisiana Native lines. Subsequent analysis aimed to determine if the strength of the divergence exceeded that found between recognised breed pairs. Genotypes from 14 breeds sampled from Europe and Asia were used to obtain estimates of pair-wise population divergence measured as F_{ST}. The divergence separating the Florida and Louisiana Native (F_{ST} = 6.2 %) was approximately 50% higher than the average divergence separating breeds developed within the same region of Europe (F_{ST} = 4.2 %). This strongly indicated that the two Gulf Coast Native lines are sufficiently different to be considered separate breeds. PCA using small SNP sets successfully distinguished between the Florida and Louisiana Native animals, suggesting that allele frequency differences have accumulated across the genome. This is consistent with a population history involving geographic separation and genetic drift. Suggestive evidence was detected for divergence at the *poll* locus on sheep chromosome 10, however drift at neutral markers has been the largest contributor to the genetic separation observed. These results document the emergence of populations that can be considered separate breeds, with practical consequences for bio-conservation priorities, animal registration and the establishment of separate breed societies.

INTRODUCTION

Sheep were first domesticated around 11,000 years ago for access to meat, before specialised breeds were subsequently developed to match animals to their captive environments. Today there are more than 1400 sheep breeds [1] containing considerable diversity in morphology, productive performance, size, shape and colour. The evolutionary history of breed development is of interest, and Menotti-Raymond and colleagues [2] describe four scenarios that may give rise to animal breeds. Firstly, 'natural' breeds are populations which became geographically dispersed following domestication and have subsequently undergone genetic drift and adaptation to their local environment. These are old breeds and often remain unmanaged, with the Soay and Gute breeds of sheep as good examples. Secondly, 'established' breeds are those that have undergone a managed process of human mediated selection towards a breed standard. Most economically relevant sheep breeds, such as the Merino, are considered to be established. A third scenario may give rise to 'mutation' breeds and involves the propagation of a specific desirable phenotype that distinguishes animals from their ancestors. This is most commonly focussed on pigmentation traits (e.g. Red Engadine, Swiss Black-Brown Mountain sheep) or horn type (e.g. Poll Dorset that have no horns and Jacob sheep which have either 4 or even 6 horns). The fourth scenario involves 'hybrid' breeds where deliberate crosses have been engineered between established breeds. The dual-purpose Perendale is an example that arose through the inter-breeding of Cheviot and Romney in New Zealand. Given this range of processes it is reasonable to anticipate the degree of genetic separation that distinguishes breeds will be highly variable. This has been demonstrated by previous investigations into the divergence between sheep breeds based on microsatellites [3,4], SNP [5] and the mitochondrial genome [6-8].

Knowledge describing the strength of genetic division that exists between breeds has a number of important applications. Firstly, the degree of divergence can be used to direct prioritisation of resources available for the conservation of biodiversity [9,10]. For example, where a high level of genetic differentiation is identified separating a phenotypically similar pair of breeds, high priority for conservation of animal genetic resources may be given to each. Conversely, phenotypically distinct 'mutation' breeds found to be genetically indistinguishable are less likely to individually attract a high priority for conservation. Beyond biodiversity, the relationship between breeds has practical implications for the delivery of emerging approaches to achieve genetic gain in livestock. Genomic selection is

currently being implemented to speed genetic gain through the forward prediction of phenotypic performance on the basis of genotypic data alone [11]. The success of across breed genomic prediction will, in part, be determined by the relatedness between breed pairs. It is also important to recognise that genetic division may exist within a breed. Population substructure, when undetected, has the potential to generate spurious associations in experiments seeking to identify disease or production genes [12,13]. The successful identification of population stratification, however, can be used to minimise inbreeding in closed populations [14].

Where substantial genetic separation has been identified between subpopulations within a breed, it becomes relevant to investigate if the divergence is sufficient to consider the subpopulations as separate breeds. This may trigger the establishment of separate breed societies to manage animal recording and accommodate divergent breeding objectives. It is important to note that the categorization of animals into breeds has a lot to do with non-genetic factors as diverse as human cultural identity, history and politics. This might explain why surprisingly little has been published concerning the minimum divergence required to declare subpopulations as separate breeds, given genetic distinction is not the sole determinant. While recognising non-genetic factors are important, the recent International Sheep Genomics Consortium's (ISGC) HapMap and Breed Diversity experiment offers the opportunity to explore this question at previously unattainable resolution through the use of 49,034 SNP [15]. We investigated the Gulf Coast Native (GCN), a breed adapted to the heat and semi tropical environment of the south-eastern part of the United States. The exact genetic origin of Gulf Coast sheep is unknown, however the Gulf Coast Native breed is considered to have descended from the Spanish Churra which were commonly brought to the Americas as early as the 1500s. The wool characteristics of modern Gulf Coast sheep indicate a contribution by Merino and/or British white faced breeds. Since its introduction the Gulf Coast Native breed has adapted to an environment characterised by high parasite loads, and a number of studies have investigated the genetic basis of their natural resistance to intestinal nematodes [16-19]. The majority of Gulf Coast animals lack wool on their faces, legs and bellies and most rams and some ewes carry horns, although box sexes may be polled. Size varies with rams weighing between 125 – 200 pounds and ewes 90 – 160 pounds. Importantly for this study, the breed contains two lines known as the Florida Native and the Louisiana Native. Anecdotal information suggests the two lines may have been founded by

separate importations of animals and that phenotypic differences exist. The Louisiana Native are found predominantly in Arkansas, Louisiana, Missouri, Mississippi and Texas and may have arrived with explorers from Latin America. They are almost totally white with only limited pigmentation. The Florida Native are found predominantly in Alabama, Florida and Georgia and likely derived from sheep arriving with settlers to the east coast of Florida. They are mostly polled and white, but black and brown body coloring is common on their face and legs (Miller, pers comm). The Louisiana Native tend to have shorter legs and a larger body than the Florida Native. Resource flocks were maintained at Louisiana State University and the University of Florida for several years, and the census size for each line numbers only in the thousands, across less than 50 flocks each.

In this study, we first evaluated the ability of a dense set of SNP markers to detect population substructure within the Gulf Coast Native breed while ignoring line membership. Further, we calibrated the strength of substructure detected against a series of population comparisons to evaluate if the lines could be considered sufficiently diverse to be declared separate breeds. In the process the results document the emergence of domestic breeds and provide insights into the mechanisms involved.

RESULTS

Strong Genetic Substructure Detected within the Gulf Coast Native

To examine the relationship between Gulf Coast Native animals and search for evidence of population substructure, the average proportion of alleles shared between individuals was calculated using 49,034 SNP. The resulting 2 x 2 allele sharing (A_S) matrix, when visualised as an ordered heat-map, revealed two distinct groups of animals within the breed (Figure 1). Animals within each group were, on average, more related to each other compared to a member of the second group. Inspection of the animals contained within each group revealed almost complete correspondence with membership of two established lines, the Florida Native and the Louisiana Native. The majority of Louisiana Natives formed a block displaying elevated allele sharing (lower right quadrant, Figure1) that was separate from the Florida Natives (upper left quadrant). The top right quadrant of Figure 1 shows allele sharing between animals from different lines. This indicated five Louisiana Natives that grouped within the Florida Native block retain elevated allele sharing with other Florida Natives. To further explore the relationship between animals, principal components analysis

(PCA) of allele sharing was performed to examine the clustering pattern within the population. Based on the full set of 49,034 SNP, animals from the two breed lines took non-overlapping positions in a plot of the two largest principal components (PCA1 and PCA2, Figure 2A). The Louisiana Natives had generally negative PCA1 values, while the Florida Natives had positive values. To test if the divergence detected between lines was the result of using a large number of markers, the analysis was repeated using a randomly selected set of 491 SNP (or 1% of the total). Individuals from the two lines again formed distinct clusters; however, the division between the groups was not complete when using a limited set of SNP (Figure2B). The five Louisiana Natives that appeared among the Florida Native in the ordered heat-map (Figure 1) took at intermediate position in the PCA plot where animals from the two lines were indistinguishable.

Are the Lines Sufficiently Divergent to be Considered Separate Breeds?

The degree of divergence observed between the Florida and Louisiana Natives was compared against the divergence that exists between populations recognised as separate breeds. The metric used was population F_{ST}, which provides a single value for the degree of differentiation by averaging across SNP. A collection of 12 populations with SNP50 genotypes were drawn from the International Sheep Genomics Consortium (ISGC) HapMap and Breed Diversity experiment [15]. The genetic diversity within each breed, measured as heterozygosity (H_e) and the proportion of polymorphic SNP (P_n), is given in Table S1. The 12 populations were used to calculate F_{ST} at four levels: a) selection lines within the same breed (e.g. Meat and Milk Lacaune); b) separate breeds developed in the same region of Europe (e.g. Merino and Rambouillet); c) separate breeds developed in different regions of Europe (e.g. Merino and Poll Dorset) and d) separate breeds developed in either Europe or Asia (e.g. Merino and Tibetan). SNP genotypes from a total of 921 animals were used in a single analysis to estimate the divergence between every combination of the populations. The breeds used, the number of animals and the F_{ST} separating each population pair are listed in Table 1. As expected, divergence increased for each of the four categories. Selection lines within breed had the lowest average population divergence (1.7 % from 2 comparisons, Table 2). Divergence increased to 4.2 % for closely related breeds that were developed in southern Europe (15 comparisons, Table 2). Average divergence was higher again between breeds from Southern and Northern Europe (11.4 %) and highest between breeds developed in Europe and Asia (13.6 %, Table 2). These values provided calibration points for the

interpretation of the genetic division between the Florida and Louisiana Native populations which had population F_{ST}, = 6.2 % (Tables 1 and 2). This is approximately three times higher than for selection lines within other breeds, and approximately 50% higher than the average divergence for breeds developed in Southern Europe.

The Genetic Basis of Differentiation Separating Gulf Coast Native Lines

Analysis was performed to identify the subset of SNP which contributed most to the observed divergence between the Florida and Louisiana Native. SNP were ranked using F_{ST} to identify the top 1% (491 SNP) and top 5% (2451 SNP) of markers. To ensure ranking on F_{ST} did indeed enrich for SNP explaining the divergence between lines, individuals were clustered using the top 5% and top 1% of F_{ST} ranked SNP (Figure 2C and D). This check confirmed that F_{ST} ranked SNP clearly delineated the lines much more strongly than a marker panel of the same size selected at random (Figure 2B and 2C). Plotting the genomic distribution of SNP with extreme F_{ST} revealed two chromosomes (OAR10 and OARX) contained an excess compared to the number expected based on chromosome size (Figure 3). Chromosome 10 contained 9.2% of the top 1% F_{ST} SNP which is more than double the proportion expected (3.8%). Chromosome 10 is of particular interest given it harbours the *Poll* locus responsible for the absence of horns [15,20] and matches to the anecdotal evidence suggesting a higher prevalence of polled animals within Florida Native compared with Louisiana Native (Miller, pers com). This finding prompted analysis of the *Poll* locus to compare haplotype frequencies present within the Australian Poll Merino, Merino, Florida Native and Louisiana Native. Genotypes at four SNP spanning 77 Kb at the *Poll* locus were phased into haplotypes, and their frequency is given in Table 3. Haplotype H1, known to be associated with selection for polled animals [15], was present at high frequency in both the Australia Poll Merino (0.57) and Florida Native (0.49). In contrast, haplotype H1 was at low frequency in the Merino (0.09) and Louisiana Native (0.19, Table 1), suggesting selection for the absence of horns has contributed to the divergence between the Florida and Louisiana Native populations.

Genetic Relationship Linking the Gulf Coast Native to Other Breeds

The population history of the Gulf Coast Native was explored by comparing them to a selection of four Spanish breeds, a meat type breed from the United Kingdom and two Asian breeds that served as outgroups. Allele sharing was calculated between 921 animals (Table

S1) and used to construct a PCA plot (Figure 4). Gulf Coast Native individuals were positioned separately from all other breeds (Figure 4) in a cluster near a group of Spanish breeds including the Rasa Aragonesa, Ojalada and Castellana. The GCN grouped away from the Poll Dorset (Northern European) and Asian derived animals.

DISCUSSION

Breeds are human constructs that assist in producing animals with a uniform phenotype and populations that have desirable attributes. Over the last few hundred years, the emergence of new breeds has most often occurred under the 'mutation' or 'hybrid' scenario where man intentionally develops new breeds based on phenotype [2]. Under these scenarios, the establishment of a new breed is not based on prior knowledge concerning genetic division or divergence within a population, although this is the long term consequence of erecting reproductive barriers within a population. The two remaining scenarios for breed development are emergence of 'natural' or 'established' breeds. If these events have occurred in the distant past, the expectation is that they will be characterised by genetic separation and division. In this framework, we sought to explore a specific case where a breed has been maintained over several hundred years where multiple lines of animals exist, but where no declaration has been made concerning their genetic eligibility to be termed separate breeds. To search for evidence for genetic division, a set of nearly 50,000 SNP markers were genotyped in the Florida and Louisiana Native lines, and the divergence separating them expressed as F_{ST}. To generate a type of calibration curve of F_{ST} values, other breeds were used that are expected to have increasing genetic separation. At one extreme pairs were examined using lines within the same breed while at the other extreme pairs were examined using breeds developed in Europe and Asia. These comparisons provided a guide to the percent values for F_{ST} to be expected when calculated using 50,000 SNP, and the values increased in a step-wise fashion for the four comparison types (Table 2). Importantly, the divergence observed between Florida and Louisiana Natives (F_{ST} = 6.2 %) was higher than the average value separating recognised breed pairs developed in Mediterranean Europe (average F_{ST} = 4.2%, Table 2). The Australian Merino and American Rambouillet are an example that contributed to the average value of Mediterranean derived breeds. Both were originally developed in southern Europe before being exported and subsequently adapted to production systems in Australia and the US, respectively. They have pair-wise population F_{ST} of 5.3 % which is lower than for the

Florida and Louisiana Native (Table 2). Our interpretation, therefore, is that on the basis of genetic data the two lines of Gulf Coast sheep can be considered as different breeds.

Analysis of the divergence between subpopulations suggests the emergence of separate breeds has occurred in a manner that most closely resembles the 'natural' scenario [2]. Specifically, there has been some degree of geographic separation (Florida versus Louisiana) followed by genetic drift within each separated population. This conclusion is based on the observation that PCA was able to reconstitute the genetic division between the Gulf Coast populations when performed using a random subset containing 1% of available SNP. This indicates allele frequency differences are present at the majority of SNP, rather than at a small number of markers in response to human mediated selection pressure. Suggestive evidence was detected for selection at the horn / poll locus; however, the difference in haplotype frequency was modest and did not contribute a meaningful amount to the total divergence observed. Furthermore, the divergence does not appear to have involved a strong founder effect whereby a small number of animals were used to create the two lines. Founder effects are accompanied by restricted genetic diversity; however, the proportion of SNP displaying polymorphism and expected heterozygosity observed within Gulf Coast animals was amongst the highest of any population tested (Table S1). In summary, the emergence of these new breeds is most consistent with the natural scenario, as opposed to the foundation of new breeds by either the 'mutation' or 'hybrid' scenarios.

MATERIALS AND METHODS

Animal Material

Gulf Coast Natives were sampled from a total of eleven breeders distributed across six states in the south-eastern region of the United States. Florida Natives (n = 35) samples were from six breeders distributed across Florida, Georgia and Texas, while the Louisiana Natives samples were collected from six breeders in Arkansas, Louisiana, Missouri and Texas. Five animals were sampled from a Florida flock of the 'cracker' line, and these are annotated separately from the Florida Native and Louisiana native individuals in Figure 2. Given sampling was performed across a number of different American States and within different flocks, the animals used can be considered representative of the two lines. Owners of all animals used in the study gave permission for the collection of blood prior to sampling. Animal handling procedures including the collection of blood samples was submitted to the

Louisiana Statue University Agricultural Center Institutional Animal Care and Use Committee (IACUC) and approved on June 16th 2005 under Protocol number AE AE 05-12. Details for all of the other breeds used have been described elsewhere [15].

Genotyping and Quality Control

All DNA samples were genotyped using the Illumina Ovine SNP50 BeadChip as part of the ISGC HapMap and Breed Diversity experiment [15]. SNP genotype calls were evaluated by a series of quality control filters to remove poor performing samples and 5207 SNP that failed to meet any of five quality criteria [15]. A total of 49,034 SNP remained and were used in this study.

Genetic Diversity and Analysis of Population Substructure

Within breed diversity was calculated as the proportion of polymorphic SNP (P_n) and gene diversity (H_e) from the full SNP dataset using PLINK v2.05 [21]. Allelic richness (A_r), which measures the normalised number of alleles and private allelic richness (pA_r), which gives a measure of population distinctiveness, were calculated using ADZE [22]. The average proportion of alleles shared between animals (A_s) was calculated as (IBS2+0.5*IBS1/N) where IBS1 and 2 are the number of loci that share either 1 or 2 alleles and N is the number of SNP pairs tested. These values were calculated using PLINK v2.05 [21] through use of commands - - cluster and - - distance-matrix. The resulting matrix of A_s values was used to generate an ordered heatmap and dendogram using R software package RColorBrewer. Principal components analysis (PCA) of A_s values was performed using smartpca implemented in Eigensoft [23]. Population divergence was calculated for each SNP as F_{ST} and as global F_{ST} using the methods of Nicholson and colleagues [24].

Analysis of the Poll Locus

SNP50 genotypes from four populations were used for the analysis of the *Poll* locus; the Florida Native, Louisiana Native, Poll Merino and Merino. The Poll Merino (98 animals with no horns) and Merino populations (88 animals with horns) were collected as unrelated industry sires in Australia as part of the ISGC HapMap experiment [15]. Haplotypes were constructed and their population frequencies estimated using Haploview [25] for four SNP

(*OAR10_29469450*, *OAR10_29511510*, *OAR10_29538398* and *OAR10_29546872*) which span the *Poll* locus on sheep chromosome 10 [15,18].

ACKNOWLEDGEMENTS

The International Sheep Genomics Consortium (ISCG) made the genotypes available to make this study possible.

REFERENCES

[1] Scherf BD (2000) World Watch List for Domestic Animals. Ed. 3. Food and Agriculture Organisation of the United Nations, Rome.

[2] Menotti-Raymond M, David VA, Pflueger SM, Lindblad-Toh K, Wade CM, et al. (2007) Patterns of molecular genetic variation among cat breeds. Genomics 91: 1-11.

[3] Lawson Handley LJ, Byrne K, Santucci F, Townsend S, Taylor M, et al. (2007) Genetic structure of European sheep breeds. Heredity 99: 620-631.

[4] Tapio M, Ozerov M, Tapio I, Toro MA, Marzanov N, et al. (2010) Microsatellite-based genetic diversity and population structure of domestic sheep in northern Eurasia. BMC Genet 11: 76.

[5] Kijas JW, Townley D, Dalrymple BP, Heaton MP, Maddox JF, et al. (2009) A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. PLoS One 4: e4668.

[6] Pereira F, Davis SJ, Pereira L, McEvoy B, Bradley DG, et al. (2006) Genetic signatures of a Mediterranean influence in Iberian Peninsula sheep husbandry. Mol Biol Evol (7): 1420-6.

[7] Pedrosa S, Arranz JJ, Brito N, Molina A, San Primitivo F, et al. (2007) Mitochondrial diversity and the origin of Iberian sheep. Genet Sel Evol 39: 91-103.

[8] Meadows JR, Li K, Kantanen J, Tapio M, Sipos W, et al. (2005) Mitochondrial sequence reveals high levels of gene flow between breeds of domestic sheep from Asia and Europe. J Hered 96: 494-501.

[9] Dalvit C, De Marchi M, Zanetti E, Cassandro M. (2009) Genetic variation and population structure of Italian native sheep breeds undergoing in situ conservation. J Anim Sci 87: 3837-44.

[10] Paiva SR, Facó O, Faria DA, Lacerda T, Barretto GB, et al. (2011) Molecular and pedigree analysis applied to conservation of animal genetic resources: the case of Brazilian Somali hair sheep. Trop Anim Health Prod 43: 1449-1457.

[11] Hayes B, Goddard M. (2010) Genome-wide association and genomic selection in animal breeding. Genome 53: 876-883.

[12] Yu K, Wang Z, Li Q, Wacholder S, Hunter DJ, et al. (2008) Population substructure and control selection in genome-wide association studies. PLoS One 3: e2551.

[13] Solovieff N, Hartley SW, Baldwin CT, Perls TT, Steinberg MH, et al. (2010) Clustering by genetic ancestry using genome-wide SNP data. BMC Genet 11: 108.

[14] Toro MA, Caballero A. (2005) Characterization and conservation of genetic diversity in subdivided populations. Philos Trans R Soc Lond B Biol Sci. 360): 1367-1378.

[15] Kijas J, Lenstra J, Hayes B, Boitard S, Porto Neto L, 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: e1001258.

[16] Bahirathan M, Miller JE, Barras SR. Kearney MT (1996) Susceptibility of Suffolk and Gulf Coast Native suckling lambs to naturally acquired strongylate nematode infections. Vet Parasitol 65: 259-268.

[17] Miller JE, Bahirathan M, Lemarie SL, Hembry FG, Kearney MT, et al. (1998) Epidemiology of gastrointestinal nematode parasitism in Suffolk and Gulf Coast Native sheep with special emphasis on relative susceptibility to *Haemonchus contortus* infection. Vet Parasitol 74: 55-74.

[18] Miller JE, Bishop SC, Cockett NE, McGraw RA (2006) Segregation of natural and experimental gastrointestinal nematode infection in F2 progeny of susceptible Suffolk and resistant Gulf Coast Native sheep and its usefulness in assessment of genetic variation. Vet Parasitology 140: 83-89.

[19] Shakya KP, Miller JE, Lomax LG, Burnett DD. (2011) Evaluation of immune response to artificial infections of *Haemonchus contortus* in Gulf Coast Native compared with Suffolk lambs. Vet Parasitology 181: 239-247.

[20] Johnston SE, McEwan JC, Pickering NK, Kijas JW, Beraldi D, et al. (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: 2555-2566.

[21] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559-575.

[22] Szpiech ZA, Jakobsson M, Rosenberg NA. (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. Bioinformatics. 24: 2498-2504.

[23] Patterson N, Price AL, Reich D. (2006) Population structure and eigenanalysis. PLoS Genet 2: e190.

[24] Nicholson G, Smith AV, Jonsson F, Gustafsson O, Stefansson K (2002) Assessing population differentiation and isolation from single nucleotide polymorphism data. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64: 695-715.

[25] Barrett JC, Fry B, Maller J, Daly MJ. (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21: 263-265.

Figure 1 Allele sharing heat map of the Gulf Coast Native. The proportion of alleles shared (A_s) between individuals was used to construct an ordered heat-map and dendogram. Cell colour represents the strength of allele sharing, where darker color indicates increasing allele sharing and relatedness between animals. As a guide, the lightest colors represent A_s values < 0.65 and the darkest colors A_s values > 0.80. Samples compared against themselves appear on the diagonal with the maximum A_s value of 1 and darkest colour. Two blocks were revealed that correspond to membership of either the Florida Native (indicated at right in green) or Louisiana Native (red).

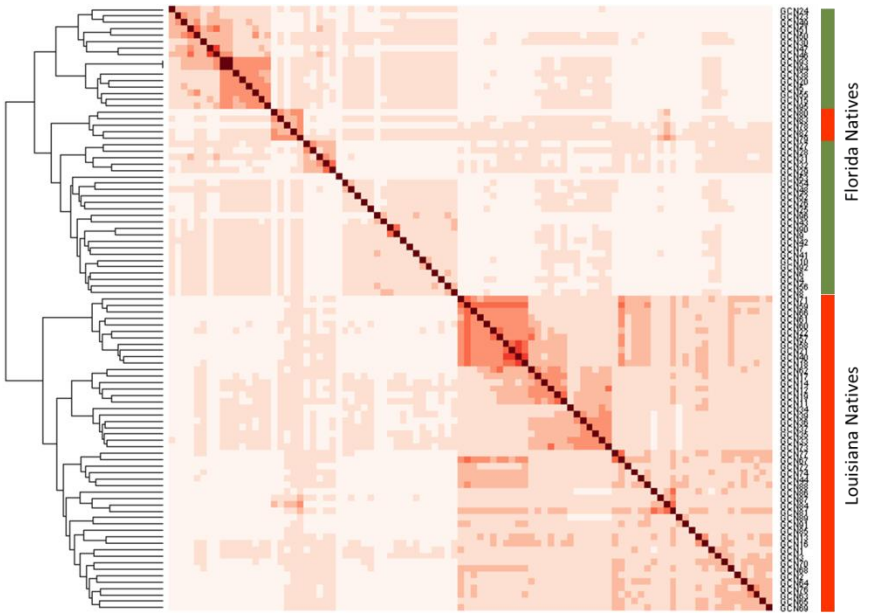


Figure 2. Clustering of animals based on principal component analysis of allele sharing. Individual animals were plotted for the first (PCA1) and second (PCA2) principal components using colours to distinguish the Louisiana Natives (red), Florida Natives (green) and Cracker (blue) lines. Four SNP sets were used to explore the effect of marker number and divergence on clustering: **(A)** all 49,034 SNP; **(B)** a random sample of 491 (1%) of SNP; **(C)** the top 1% of SNP ranked using F_{ST}; **(D)** the top 5% of SNP ranked using F_{ST}. Note that the scale for PCA1 differs between panels and that a larger proportion of the variation is captured by PCA1 using F_{ST} ranked SNP (C) compared with a random selection of markers (B).

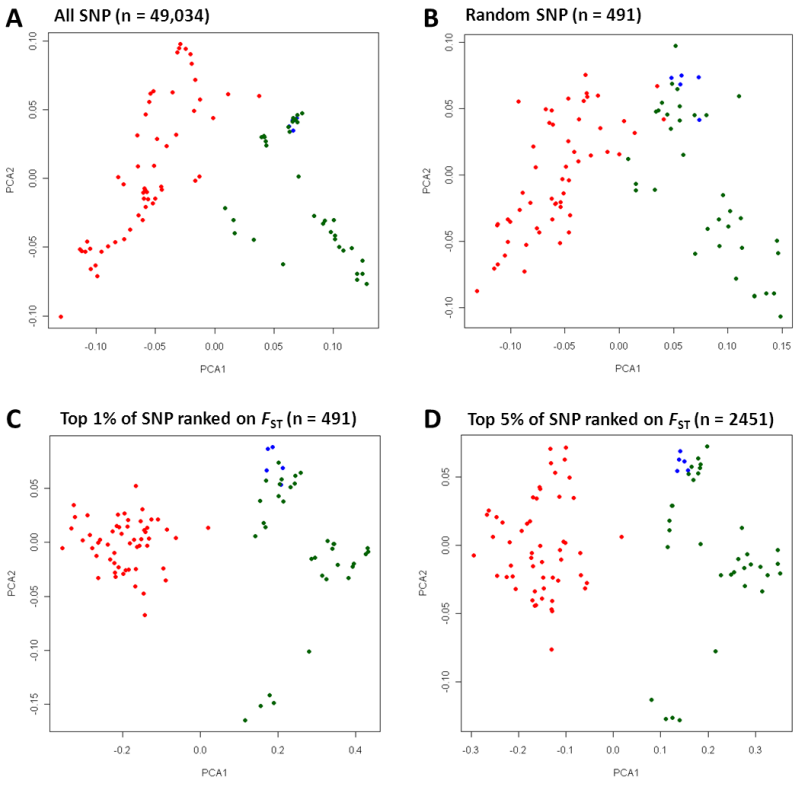


Figure 3. Distribution of divergent SNP across the sheep chromosome. The degree of divergence between the Florida and Louisiana Native populations was estimated for each SNP as F_{ST}. The genomic distribution of SNP with extreme values (either the top 1% or 5% of ranked values) is shown as a function of their chromosomal location. This revealed an over representation of high F_{ST} SNP on chromosome 10 when compared to the proportion expected.

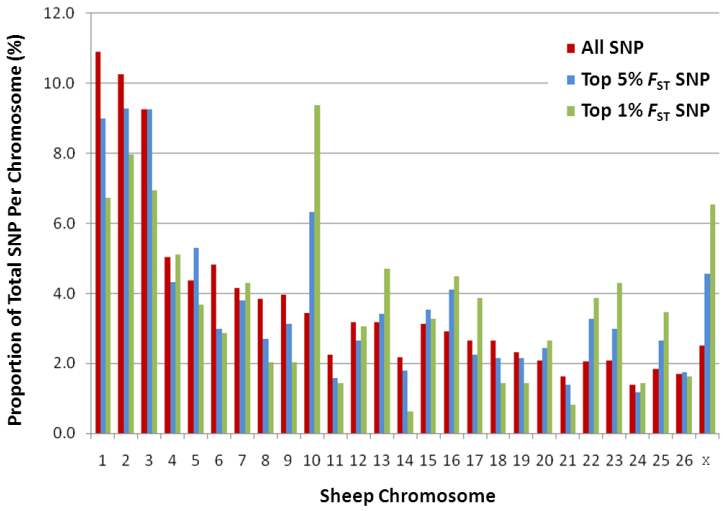


Figure 4. Genetic relationship between Gulf Coast Native and 12 other breeds. Individuals were clustered using PCA of allele sharing and plotted for PCA1 and 2 (**A**) and PCA1 and 6 (**B**). Individuals from different breeds are given using different colored symbols as follows: Gulf Coast Native (GCN gold), Merino (MER dark blue), Poll Dorset (APD dark green), Poll Merino (APM red), Castellana (CAS brown), Churra (CHU pink), Meat Lacaune (MEL light blue), Milk Lacaune (MIL black), Ojalada (OJA orange), Rambouillet (RAM purple), Rasa Aragonesa (RAS pale green), Sumatra (SUN violet) and Tibetan (TIB yellow). The two largest principal components (**A**) separated Asian (SUM, TIB) and Northern European animals (APD) from a cluster containing breeds developed in southern Europe (including RAM, MER and APM). Plotting PCA1 and PCA6 revealed the genetic division within GCN (**B**).

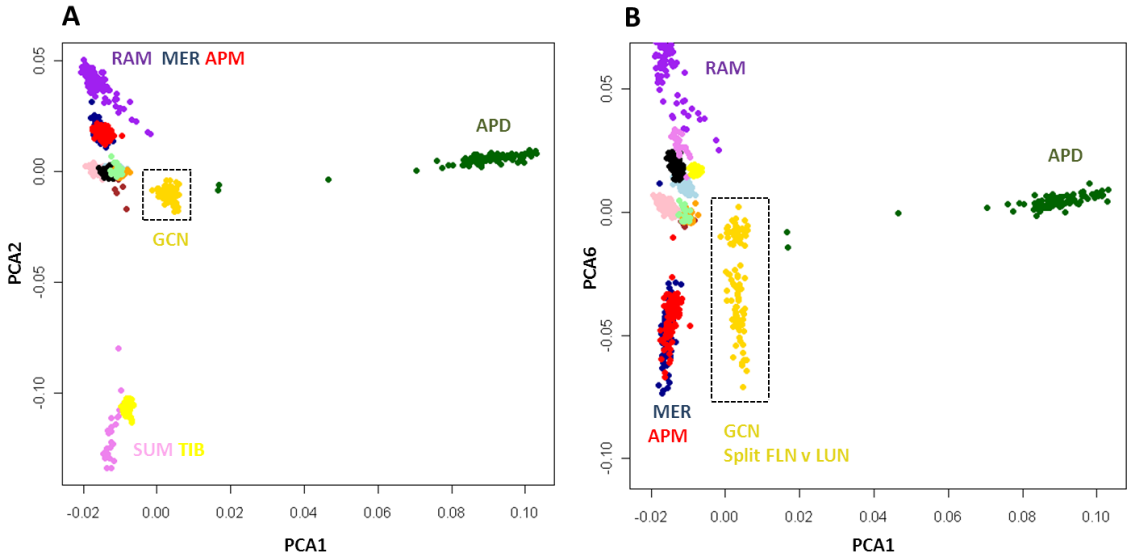


Table 1

Population divergence measured as F_{ST} (%)

Population	Code	Origin[1]	Animals	Pair-wise Population Divergence Measured as F_{ST} (SD) [2]													
				FLN	LUN	MER	APD	APM	CAS	CHU	MEL	MIL	OJA	RAM	RAS	SUM	TIB
Florida Native	FLN	SE	40	0	6.2	5.4	10.3	5.3	5.2	6.5	5.4	6.1	4.8	6.9	4.0	14.5	11.0
Louisiana Native	LUN	SE	54	0.8	0	7.5	12.4	7.4	7.3	8.8	7.5	8.3	7.0	9.0	6.1	16.9	13.2
Merino	MER	SE	88	0.7	1.0	0	11.4	1.0	4.0	5.4	4.5	5.2	3.7	5.3	2.7	14.4	11.1
Poll Dorset	APD	NE	108	1.3	1.6	1.5	0	11.2	11.1	12.4	11.1	12.0	10.7	13.0	10.0	20.6	16.9
Poll Merino	APM	SE	98	0.7	1.0	0.3	1.5	0	3.9	5.3	4.4	5.1	3.6	5.0	2.6	14.1	10.9
Castellana	CAS	SE	23	0.7	0.9	0.5	1.4	0.6	0	4.5	4.0	4.7	2.9	5.9	2.1	13.7	10.4
Churra	CHU	SE	120	1.0	1.1	0.7	1.6	0.8	0.7	0	5.3	6.0	4.0	7.2	3.5	15.3	12.0
Meat Lacaune	MEL	CE	78	0.7	0.9	0.6	1.4	0.6	0.6	0.8	0	2.3	3.7	6.4	2.7	14.5	11.2
Milk Lacaune	MIL	CE	103	0.7	1.0	0.6	1.5	0.7	0.6	0.9	0.3	0	4.4	7.1	3.3	15.2	11.8
Ojalada	OJA	SE	24	0.7	1.0	0.6	1.4	0.6	0.5	0.6	0.6	0.6	0	5.5	1.8	13.6	10.3
Rambouillet	RAM	SE	102	1.0	1.3	0.9	1.7	0.7	1.1	1.1	1.0	1.1	1.0	0	4.4	16.4	13.2
Rasa Aragonesa	RAS	SE	22	0.6	0.9	0.5	1.4	0.5	0.4	0.6	0.4	0.5	0.4	0.9	0	12.7	9.5
Sumatra	SUM	A	24	1.6	1.7	1.6	2.1	1.6	1.6	1.7	1.6	1.6	1.6	1.8	1.5	0	13.2
Tibetan	TIB	A	37	1.2	1.4	1.3	1.9	1.3	1.3	1.4	1.3	1.3	1.3	1.6	1.2	1.6	0

[1] The geographic region of breed development is given as SE (Southern Europe), NE (Northern Europe), (CE) Continental Europe or A (Asia).

[2] F_{ST} (given as a percentage) is given above the diagonal and its standard deviation (SD) for each combination is given below (x1000).

Table 2

Average F_{ST} for different population comparisons

Comparison	Populations Used	Number of Comparisons	Average_F_{ST}[1]
Selection lines within breed	MER,APM,MEL,MIL	2	1.7
Breed pairs of Mediterranean origin	MER,CAS,CHU,OJA,RAM,RAS	15	4.2
Breed pairs of Southern vs Northern European origin	APD,MER,CAS,CHU,OJA,RAM,RAS	6	11.4
Breed Pairs of Asian vs European Origin	SUM,TIB,APD,MER,CAS,CHU,OJA,RAM,RAS	14	13.6
Florida Native v Louisiana Native	FLN,LUN	1	6.2

[1] Average F_{ST} (percent) was calculated across comparisons.

Table 3. Haplotype frequencies at the Poll locus on sheep chromosome 10

Population	Animals	Haplotypes[1]	Haplotype Frequency [2]			
			H1	H2	H3	Other
Poll Merino	98	6	0.57	0.24	0.12	0.07
Merino	88	5	0.09	0.58	0.13	0.20
Florida Native	35	6	0.49	0.03	0.39	0.09
Louisiana Native	54	3	0.19	0.00	0.57	0.24

[1]Haplotypes were constructed using SNP genotypes at four loci (OAR10_29469450, OAR10_29511510, OAR10_29538398 and OAR10_29546872) that span 77 Kb at the Poll locus on sheep chromosome 10. The total number of haplotypes observed is given for each population.

[2] The most frequently observed haplotypes were labelled H1 – H3. Haplotype 1 (H1) comprised alleles GGAA at the four SNP 1 and was associated with selection for the absence of horns in the ISGC HapMap study [15]. Haplotypes H2 and H3 consisted of alleles AGGT and GGAT respectively. The combined frequency of all other haplotypes observed within each population is given as ‘other’.

SUPPORTING INFORMATION

Table S1. Basic indices of genetic diversity measured within breed. Breeds are listed with decreasing expected heterozygosity or gene diversity (H_e). Other measures include the proportion of SNPs displaying polymorphism (P_n); the inbreeding coefficient (F); allelic richness (pA_r). These results are taken from the ISGC HapMap and Breed Diversity Experiment [15].

Breed	H_e	P_n	F	A_r	pA_r
Gulf Coast Native	0.38	0.96	0.09	1.99	0.01
Rasa Aragonesa	0.38	0.95	0.04	1.98	0.01
Australian Industry Merino	0.37	0.96	0.10	1.99	0.00
Australian Poll Merino	0.37	0.96	0.09	1.99	0.01
Castellana	0.37	0.94	0.07	1.97	0.01
Meat Lacaune	0.37	0.96	0.10	1.99	0.01
Ojalada	0.37	0.95	0.06	1.98	0.00
Churra	0.36	0.96	0.11	1.99	0.00
Milk Lacaune	0.36	0.96	0.11	1.99	0.00
Rambouillet	0.36	0.96	0.14	1.99	0.00
Australian Poll Dorset	0.34	0.95	0.15	1.98	0.01
Tibetan	0.34	0.92	0.20	1.95	0.04
Sumatran	0.31	0.89	0.24	1.91	0.04

USING REGULATORY AND EPISTATIC NETWORKS TO EXTEND THE FINDINGS OF A GENOME SCAN: IDENTIFYING THE GENE DRIVERS OF PIGMENTATION IN MERINO SHEEP

Elsa García-Gámez[1,2], Antonio Reverter[1], Vicki Whan[1], Sean M. McWilliam[1], Juan José Arranz[2], International Sheep Genomics Consortium, James Kijas[1]*

[1] Livestock Industries, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Brisbane, Queensland, Australia. [2] Departamento de Producción Animal, Universidad de León, León, Spain.

PLoS ONE 6(6): e21158.

ABSTRACT

Extending genome wide association analysis by the inclusion of gene expression data may assist in the dissection of complex traits. We examined piebald, a pigmentation phenotype in both human and Merino sheep, by analysing multiple data types using a systems approach. First, a case control analysis of 49,034 *ovine* SNP was performed which confirmed a multigenic basis for the condition. We combined these results with gene expression data from five tissue types analysed with a skin-specific microarray. Promoter sequence analysis of differentially expressed genes allowed us to reverse-engineer a regulatory network. Likewise, by testing two-loci models derived from all pair-wise comparisons across piebald-associated SNP, we generated an epistatic network. At the intersection of both networks, we identified thirteen genes with insulin-like growth factor binding protein 7 (*IGFBP7*), platelet-derived growth factor alpha (*PDGFRA*) and the tetraspanin platelet activator *CD9* at the kernel of the intersection. Further, we report a number of differentially expressed genes in regions containing highly associated SNP including *ATRN*, *DOCK7*, *FGFR1OP*, *GLI3*, *SILV* and *TBX15*. The application of network theory facilitated co-analysis of genetic variation with gene expression, recapitulated aspects of the known molecular biology of skin pigmentation and provided insights into the transcription regulation and epistatic interactions involved in piebald Merino sheep.

INTRODUCTION

The population history and genetic structure of domestic animals offer advantages for the identification of the genetic drivers associated with phenotypic change. High throughput genotyping has lead to genome wide association studies (GWAS) which have successfully identified the genetic basis of monogenic disease in cattle [1], sheep [2] and dog [3]. Further, recent studies have identified complex traits such as skeletal morphology are under the control of a small number of genes of large effect [4]. For human and livestock populations however, the majority of traits are likely to be controlled by a larger number of genes which individually confer small effects [5,6]. Given that most tests for association proceed in a simple SNP-by-trait fashion using stringent significance levels to offset for multiple testing, the result is many of the genes that contribute to trait variation remain undetected.

To explore approaches which seek to incorporate GWAS into systems biology, we have merged SNP variation with analysis of differential gene expression to investigate the basis of a phenotypic trait in sheep. This was prompted by recent studies which exploit gene network theory and systems approaches to identify key genetic drivers [7–10]. In our case, we co-analysed variation in allele frequency, gene expression and transcription factor mediated gene regulation to incriminate genes which would otherwise have remained undetected using a single data type in isolation. To test our approach we selected piebald, which in humans is a leukoderma arising from disregulation of melanocyte development and migration often caused by mutations in the *KIT* gene [11]. In sheep however, piebald, is an economically important phenotype in Australian Merino sheep characterised by the presence of one or more asymmetric pigmented regions. Test matings indicate the condition is not consistent with a simple Mendelian mode of inheritance [12] and the location and extent of pigmentation in effected animals varies considerably, suggesting the coordinated action of multiple genes.

RESULTS

SNP Association and Gene Expression Confirm Piebald Has a Multigenic Basis

We collected DNA from 24 piebald Merinos characterised by the appearance of pigmentation spots (Figure 1A). To minimise unrelated genetic variability, we then selected 72 genetically similar but non-pigmented Merinos from a wider population sample using allele sharing calculated from 49,034 SNP. The resulting relationship matrix linking all 96

animals is shown in Figure 1B. Comparing allele frequency differences between piebald and non-pigmented animals revealed 226 loci were highly associated ($p<0.001$) and collectively distinguished piebald from non-pigmented animals (Figure 1C). The highest association ($p = 8.45610^{-27}$) was observed for SNP *s49104* located in the region containing *IGFBP7* (OAR 6 Mb 78.9), however the absence of a single and strong association peak confirmed a multigenic basis for *ovine* piebald (Figure S1).

We sought to interpret these genetic associations using gene expression obtained from five skin tissue types isolated from non-pigmented, piebald and also recessive black individuals known to be under the control of *Agouti* (Figure 1A, 2A). The five tissue types were white skin tissue from non-pigmented sheep (NOR); black skin tissue from a piebald animal (PBB); white skin tissue from a piebald animal (PBW); black skin tissue from a recessive black animal (RSB) and white skin tissue from the non-pigmented region of a recessive black animal (RSW). Seven contrasts between tissue types (named DE1–DE7, see Methods section) were examined using a microarray containing 3,685 unique skin-specific genes. Of these, 54 genes displayed differential expression (DE) in ≥4 contrasts and hierarchical cluster analysis revealed coexpression across tissue types (Figure S2). A set of 19 genes, including 11 keratin family members displayed coordinated down regulation in piebald tissue, again indicating no single gene alone appeared responsible for the trait.

Charting the proximity of SNP to the genomic location of genes revealed 17,223 SNP (35%) were either intragenic or within 2.5 Kb of a gene (Figure S3). Further, 1,935 genes present on the skin-specific microarray had a genotyped SNP within 1 Mb. This allowed us to search for genes displaying both DE and genetic association relating to piebald (Figure 2B). Analysis across all seven contrasts identified a total of 370 DE genes located within 1 Mb of a SNP (Figure S4). Of these, 287 had a SNP sufficiently close (>2.5 Kb) to be considered putatively cis-acting. On this basis, we identified 312 'piebald-associated genes' which displayed either (*i*) DE in one or more contrasts and a *cis*-SNP or (*ii*) the absence of DE but an associated SNP within the chromosomal region (<1 Mb) (Table S1). *IGFBP7* was both highly associated with piebald and down-regulated in piebald tissue (Table 1). Similarly, *ATRN*'s expression was up-regulated in piebald tissue and a putatively *cis*-acting SNP displayed strong association while *TMEM158*, *PDGFRA* and *PTK2* were not DE in any contrast, however co-localised SNP were highly associated (Table 1).

Regulatory and Epistatic Networks Identify the Gene Drivers of Pigmentation

A network was constructed to explore regulation of DE genes through the action of transcription factors (TF). First, promoter sequence analysis of each piebald-associated gene was performed to identify the complement of transcription factor binding sites (TFBS) associated with each. Then, a regulatory network was constructed where nodes represent genes. The presence of a TFBS created an edge linking a gene with the TF for which it contained a binding site. Gene expression was also used as input using the highest correlated contrasts (DE3, DE5 and DE6; Figure S5). The network was visualised using the Cytoscape software, as described in the methods. Fourteen TFs were present in the network (*MZF1*, *SF1*, *SMAD4*, *TEF*, *HBP1*, *MSX2*, *NF1*, *SIX3*, *DLX2*, *LEF1*, *NFAT5*, *HLF*, *IRX2* and *KLF6*), none of which displayed DE, but collectively linked 108 piebald-associated genes. Of these, three (*MZF1*, *SF1* and *SMAD4*) have a binding site in 20 or more piebald-associated genes (Figure 3), strongly suggests a pivotal role in regulation. Importantly, *SMAD4* inhibits *PAX3* [13] which is a key regulator of *MITF* that regulates the degree of black spotting in both dog [14] and cattle [6]. In our GWAS, SNP *OAR19_33605872* (located 310 Kb from *MITF*) shows suggestive association to piebald (P = 7.05E-03).

A second network was built to explore epistatic interactions between SNP and their closest genes. In this approach, all pair-wise combinations (25,425 pairs) of the 226 SNP found to be associated with piebald (Figure 1C) were tested for evidence of epistasis. A total of 645 significant pairs were identified (p<0.001) which served as input data for the epistatic network. To reduce network complexity, 'hubs' present in more than 20 significant pairs were identified (*ARHGAP15*, *BOC*, *C10H14orf37*, *CD9*, *EPHA4*, *IGFBP7*, *LOC785528* and *TMEM158*) and used to extract edges in a reduced network containing 121 nodes (genes) and 216 edges (Figure 3). This revealed *C10H14orf37* as the hub with the highest number of connections (61 different pairs). *IGFBP7* was the next most connected (32 pairs), ten of which were also significant for *PDGFRA*. Moreover, as a pair, *IGFBP7* and *PDGFRA* were able to distinguish between piebald and non-pigmented animals (P = 0.000053). Interestingly, *MAML2* and *PDGFRA* paired together suggesting a regulatory interaction however this could not be evaluated as *MAML2* is a non-DNA binding transcription co-activator.

The final analysis identified thirteen genes that were common to both the regulatory and epistatic networks. These formed the intersecting landscape displayed in Figure 3 (*CD9*,

EFNA5, FRY, IGFPB7, LARP7, MAML2, MHY10, PDGFRA, PTK2, PTPN18, SLC11A2, SP3 and *TMEM58*). Three transcription factors created all of the links from the regulatory network into the intersecting genes (*MZF1*, *TEF* and *SMAD4*) while an additional two transcription factors were present within the intersecting landscape via their connection to other genes made by epistatic interactions (*MAML2* and *SP3*).

DISCUSSION

The fundamental processes behind melanocyte differentiation and migration are well understood thanks largely to their relevance in the context of human melanoma and vitiligo [11]. Many cases involve disruption of a single gene, however analysis in this study failed to identify a gene of large effect via SNP based association testing. Therefore, the approach developed sought to investigate the genetic control of phenotypic variation through coanalysis of divergent data types. In our case, the three data types were genomic variation (SNP) which differed between case and control populations, gene expression differences derived from symptomatic versus asymptomatic tissues and transcription factor mediated gene regulation. We searched the resulting data for the coordinated regulation of genes via transcription factors (Figure 3, left side) and we reasoned that the coordinated action of multiple genes could be detected through analysis for epistatic interactions at the genetic level (Figure 3, right side). Using network theory, we then interrogated the results of both approaches and identified a small number of intersecting genes. This clearly revealed the molecular complexity underpinning the piebald phenotype in Merino sheep.

The current study has two technical limitations likely to have restricted our power to detect genes which contribute to piebald. Firstly, the microarray used to detect gene expression contained only a subset of sheep transcripts. This may not have been a critical limitation when investigating a pigmentation trait given the microarray was intentionally populated with 6,125 skin expressed sequence tags [15], however the possibility remains that biologically important genes were unavailable for testing. Secondly, the *ovine* SNP chip facilitated genotyping of 49,034 SNP however only 35% of these are located with 2.5 Kb of a gene. This meant many genes were not tagged by a physically associated genetic marker and recombination (between SNP and their nearest gene) is likely to have eroded the strength of association in many cases. In addition to these technical considerations, it is worthwhile noting the analytical approach likely influenced the genes identified. Network theory has long

recognised the powerful role that hubs play in networks, namely their ability to pervade most of the network topology and provide robustness [16]. We reduced network complexity by identification of hubs, but acknowledge this may have resulted in biologically relevant genes being missed where they do not act as hubs.

Despite these considerations, our approach has identified genes which would have remained hidden using either data type in isolation. In what follows, and as a proof of concept validating the biological relevance of our results, we describe how genes located within in the intersecting landscape of the joint network (Figure 3) are known to be involved in pigmentation and more specifically in vitiligo. CD9 is a cell surface protein member of the tetraspanin family mediating signal transduction events that play a role in the regulation of cell development, activation, growth and motility [17] including motility of epidermal keratinocytes [18,19]. The finding that it can trigger platelet activation [20] makes our finding of an epistatic SNP pair connecting *CD9* and *PDGFRA* highly relevant. Indeed, one of our main candidates underpinning the piebald phenotype is *PDGFRA* taking, along with *CD9* and *IGFBP7*, a prominent hub role in the landscape intersecting the regulatory and the epistatic networks. Xu et al. [21] speculate that the *PDGFRA* gene may be a candidate susceptibility gene of vitiligo. Furthermore, within the context of livestock species a genome-wide scan [22] found the gene *PDGFRA* to be under positive selection in Montbéliarde breed of cattle possibly underlying the primary importance of coat color patterns for herd-book registration. Another intersecting genes (Figure 3) is insulin-like growth factor-binding protein 7 (IGFBP7) which is a secreted protein and functions extracellularly. Hochberg et al. [23] first showed that IGFBP7 is underexpressed in psoriatic epidermis but is inducible by ultraviolet treatment. Since then, IGFBP7 has received particular attention due to its potential to induce apoptosis in human melanoma cell lines [24,25]. Of particular relevance is the fact that both *PDGFRA* and *IGFBP7* are located within a 3 Mb window of ovine chromosome 6 (OAR6), in a region that also contains the *KIT* gene to which piebald was first associated [26]. In this pioneering work, the authors did not exclude the involvement of other closely linked genes such as *PDGFRA*. In order to further shed light into this issue, we took a detailed look at the LD structure and Manhattan plot of the SNP association results on OAR6 from 75 to 80 Mb (Figure S6). The SNP associated to piebald (SNP ID: *OAR6_76377079*; $P<1.0E-3$) is 116.0 kb downstream from *PDGFRA* and 277.7 kb upstream from *KIT*. Hence, the association of *KIT* with piebald in sheep cannot be entirely disregarded.

Two additional genes found within the intersection between the regulatory and the epistatic networks are *PTK2* and *EFNA5*. Protein tyrosine kinase 2 (*PTK2*), also known as *FAK*, has been reported to stimulate melanocyte migration [27] and to induce vitiligo repigmentation in vitro [28]. *EFNA5* is a member of the ephrin family of proteins, components of cell signalling pathways involved in development [29]. Its role as a negative regulator of epidermal growth factor receptor (*EGFR*) has recently been established [30]. Equally relevant, mutations in *EGFR* have been reported to be associated with dark skin in mice [31], however the gene was not significantly associated in this study. We did find, however, that expression of *HBEGF* was 1.91-fold down regulated (P = 8.14E-3) in white compared to black samples (DE1 and DE7; Table S1) and a SNP in its coding region (*OAR5_53183086*) also displayed suggestive association (P = 0.0894).

Finally, we surveyed the Color Genes database [32] and found it to include a number of genes reported in our study as either being DE and/or harbouring a SNP associated with the piebald phenotype. These include: *ATRN*, *DOCK7*, *FGFR1OP*, *GLI3*, *SILV* and *TBX15* (Table 1). The molecular role of *ATRN* as a crosstalk between melanocortin-receptor signalling and immune function was first reported by [33]. Further, He et al. [34] established that *ATRN* is a transmembrane receptor for agouti protein. More recently, Seo et al. [35] reviewed the biology of epidermal and hair pigmentation in cattle and highlighted the likely role of *ATRN* in the switching between the synthesis of melanin components. A similar relationship with the agouti protein exists with TBX15, DOCK7 and GLI3. In the absence of TBX15, expression of agouti in mice is displaced dorsally [36]. Mice with mutations in the dedicator of cytokinesis 7 protein (DOCK7) have generalized hypopigmentation and white-spotting [37]. Matera et al. [38] showed that loss of GLI3 signalling disrupts melanoblast specification and that a mutation in *GLI3* causes increased hypopigmentation in mice.

In conclusion, the incorporation of experiments which utilise multiple data types in a systems biology context is an appropriate approach for understanding complex biological systems. Whole genome sequencing will soon become viable for livestock [39]. This will bring access to every SNP and structural variant differing between the genomes of case and control populations. When combined with methods to measure global transcription from target tissues, the approach described here holds enormous promise for the elucidation of the genetic drivers which underpin a spectrum of traits which, to this point, have remained intractable.

MATERIALS AND METHODS

Piebald Animal Resource, Genotyping and Testing for Association

Pigmented Merinos within Australian industry flocks had the location and extent of colour spots photographically recorded. Care was taken to avoid inclusion of cases which displayed symmetrical pigmentation which is diagnostic of the action of *Agouti* in sheep [40]. Piebald animals (n = 24) from 23 properties were selected for genotyping to reduce relatedness between cases. Genomic DNA was genotyped using the *Ovine* SNP50 BeadChip (http://www.illumina.com) before raw signal intensities were converted into genotype calls using Illumina's Genome Studio software. SNP were pruned using a series of quality filters to define a final set of 49,034 SNP. Breed matched control animals (non piebald) were selected from the 199 Australian Merinos genotyped as part of the International Sheep Genomics Consortium's HapMap project (www.sheephapmap.org). Pair-wise allele sharing was calculated between all animals (24 cases and 199 controls) from all SNP. The resulting allele sharing matrix was used in a step wise fashion to select 72 control animals which maximised allele sharing between cases and controls, thereby reducing genetic variability and substructure unrelated to piebald. The filtering schema proceeded as follows: for each piebald individual, three non-pigmented controls with the highest pair-wise allele sharing were selected. Once a normal individual was selected, it was not included in the search of the three most related individuals with the next piebald individual. We measured the statistical association of each of SNP with piebald as the difference between its average genotype across within the piebald sample (24 animals) minus its average genotype across the normal sheep (72 animals). SNPs were deemed to be significantly associated if the observed genotype difference was beyond three standard deviations from the mean and the chi-square test yielded a P-value<0.001.

Linking Gene Expression to GWAS

SNP was mapped with its nearest protein-coding gene using the available sheep genome sequence [39]. The final data set included 48,686 SNP that displayed polymorphism within the 96 cases and controls. Of these, 15,624 SNPs were intragenic, while a further 1,592 SNPs were located within 2.5 kb for either side of a gene's coding sequence and considered to be putatively cis-acting. A further 24,727 SNP were mapped beyond 20 kb of a gene and considered to be linked to another gene [41]. Figure S3 shows the empirical density

distribution of the 6,687 SNPs located from 1 base-pair to 20 kb distance from a gene. Finally, 1,741 SNPs were unmapped.

Study Design and Tissue Sampling for Gene Expression

We used the bovine/ovine skin gene expression microarray platform described previously [15]. In brief, 11,689 probes were printed in duplicate onto Corning UltraGAPS (Corning Inc., NY, USA) glass slides at a spacing of 210 mm. For skin biopsies, a wool staple sample was removed by close clipping a region of approximately 5 cm^2. Two skin trephines of 0.9 cm diameter were taken from the clipped region and immediately placed in RNA later (Ambion) for subsequent RNA extraction. Skin samples in RNA later were placed into -80ºC freezers for long-term storage. Total RNA was prepared from the skin samples using TRI Reagent in accordance with the manufacturer's recommendations (Sigma, St Louis, MO, USA).

The general experimental design for the microarray hybridisations is shown in Figure 2A. In total, 20 hybridizations were performed. The design took into account the limited animal material and was developed with an emphasis on the exploration of all possible contrasts of interest so that all samples were compared against each other in a series of hybridizations with alternate dye-swaps. Five skin samples, or experimental conditions, were explored and codified as follows:

NOR ('normal') = White sample from normal sheep;

PBW ('piebald white') = White sample from piebald sheep;

PBB ('piebald black') = Black sample from piebald sheep;

RSW ('recessive white') = White sample from Agouti recessive black sheep (sample taken from the inguinal, non-pigmented area); and

RSB ('recessive black') = Black sample from Agouti recessive black sheep.

We used the GenePix 4000A optical scanner and the GenePixPro 5.1 image analysis software (both from Molecular Devices, Sunnyvale, CA, USA) to quantify the gene expression level intensities. Two filtering criteria were applied during data acquisition. Firstly, probes with a signal to noise ratio less than two in all hybridisations were deemed

undetectable and removed from the analysis. Secondly, for genes represented by multiple probes the most abundant probe, averaged across all hybridisations, was used. This second criterion is based on the fact that abundant probes are better annotated and their intensity signals less prone to noise. After filtering, 300,480 gene expression intensity readings remained (half from each colour channel and 15,024 from each chip) from 3,685 unique skin-specific genes. Prior to normalization, signals were background corrected and base-2 log-transformed. The arithmetic mean and standard deviation (in brackets) for the red and green intensities were 7.04 (3.24) and 8.49 (2.19) respectively. The expression data from the entire set of 20 hybridisations was deposited in Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) and can be downloaded and can be accessed using accession number GSE24189.

Normalization of Gene Expression

Following previously described approaches [42] we fitted the following ANOVA mixed-effect model to normalize the gene expression data:

$$Yi_{jkftmn} = \mu + C_{ijk} + G_m + AG_{ijm} + DG_{km} + VG_{tm} + e_{ijktmn},$$

where $Y_{ijkftmn}$ represents the n-th background-adjusted, base-2 log-intensity from the m-th gene at the t-th sample variety (NOR, PBW, PBB, RSW, and RSB) taken from the i-th array, j-th printing block and k-th dye channel; μ is the overall mean; C represents a comparison group fixed effect defined as those intensity measurements that originate from the same array slide, printing block and dye channel; G represents the random gene effects with 3,685 levels; AG, DG, and VG are the random interaction effects of arrayxgene, dyexgene, and varietyxgene, respectively; and e is the random error term. Using standard stochastic assumptions, the effects of G, AG, DG, VG and e were assumed to follow a normal distribution with zero mean and between-gene, between-gene within-array, between-gene within-dye, between-gene within-variety and within-gene components of variance, respectively. Restricted maximum likelihood estimates of variance components and solutions to model effects were obtained using VCE6 software (ftp://ftp.tzv.fal.de/pub/vce6/). The solutions to the VG effect were used as the normalized mean expression of each gene in each of the five samples under scrutiny.

To contrast the expression of each gene across sample types, we explored the following seven measures of differential expression (DE):

DE1 ('black vs white within piebald') = PBB – PBW

DE2 ('piebald vs recessive within black') = PBB – RSB

DE3 ('piebald vs normal within white') = PBW – NOR

DE4 ('piebald vs recessive within white') = PBW – RSW

DE5 ('piebald vs others within white') = PBW – ½ (NOR+RSW)

DE6 ('piebald vs non-piebald') = ½ (PBW+PBB) – 1/3 (NOR+RSW+RSB)

DE7 ('black vs white') = ½ (PBB + RSB) – 1/3 (NOR+PBW+RSW)

Using a nominal P-value<0.01 from a two-tailed t-test statistic, genes were deemed to be DE if their normalized measure of differential expression across any of the 7 DE contrasts felt beyond 2.57 standard deviations. We used PermutMatrix [43] to perform hierarchical cluster analysis of gene expression across skin tissue samples.

Promoter sequence analysis

The bovine genome-wide promoter sequence database from Genomatix (http://www.genomatix.de/; ElDorado Btau 4, v-0709) was used in the absence of annotated promoter data for sheep. A total of 60,131 promoter sequences derived from 22,050 genes were downloaded. To ensure only high confidence promoters were selected we applied the concept of orthologous promoters [44] and retained only those promoters for which phylogenetically conserved sequences were documented in both the human and mouse genomes. This resulted in identification of 39,696 promoter sequences distributed over 13,623 genes. We subsequently applied a threshold of 1 (100% confidence) to core and matrix similarities [45] to identify a final set of 310,316 high confidence TFBS that were used for integration with the gene expression data.

Regulatory and Epistatic Networks

The gene regulatory network (Figure 3, left side) was constructed using two attribute types. First, output from the promoter sequence analysis was used to create edges linking genes and transcription factors (TFs). Specifically, edges were constructed only where analysis revealed a given gene contained a transcription factor binding site (TFBS) corresponding to the linked TF. Second, differential gene expression observed between piebald and non-piebald tissue was used. To obtain a global understanding of gene expression across the seven tissue contrasts (DE1–7), we first calculated the correlation (R-value) for each pairwise combination (all 21 combinations are plotted in Figure S5). This revealed the highest three correlations (R = 0.88, 0.82 and 0.79) were observed for combinations (DE3, DE5 and DE6) which in each case compared a piebald (only white or all piebald) versus a non-piebald sample (using the normal sample, all the white non-piebald or all the non-piebald, respectively). For each gene, we therefore used the average value (from DE3, DE5 and DE6) as input in the network to show genes as either over-expressed (red), under-expressed (green) or having unchanged expression (orange) in the piebald condition. Using the two attributes the network was joined, visualised and explored using the Cytoscape software [46]. For the epistatic network (Figure 3, right side) SNP pairs were used as input which displayed evidence for epistasis. Contingency tables comprising 9 rows (combinations of two loci genotypes) and 2 columns (containing genotype counts within either the piebald or non-piebald populations) were constructed for each of the 25,425 pair-wise combinations of 226 SNP significantly associated with the piebald condition. The distribution of observations in each table were tested (P<0.001 from a chi-square test with 8 degrees of freedom) to identify 645 significant pairs. A subset of these significant pairs may result from linkage disequilibrium (as opposed to epistasis) where SNP pairs are in close physical proximity (<100 Kb). To generate the epistatic network using this data, we used the closest gene to each SNP in a significant pair. As for the regulatory network, DE was incorporated from the average of DE3, DE5 and DE6 and Cytoscape was used to build and explore the network [46].

ACKNOWLEDGMENTS

The authors gratefully acknowledge Belinda Norris and Damien Turner for collection of the piebald animals and development of the expression array, as well as the Australian Sheep Industry CRC (Sheep CRC1) for making the resulting material and data available.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: JK AR. Performed the experiments: VW. Analyzed the data: EGG AR JK. Contributed reagents/materials/analysis tools: SMM JJA ISGC. Wrote the paper: EGG AR JK.

REFERENCES

[1] Charlier C, Coppieters W, Rollin F, Desmecht D, Agerholm JS, et al. (2008) Highly effective SNP-based association mapping and management of recessive defects in livestock. Nat Genet 40: 449–454.

[2] Becker D, Tetens J, Brunner A, Bu¨rstel D, Ganter M, et al. (2010) Microphthalmia in Texel sheep is associated with a missense mutation in the paired-like homeodomain 3 (PITX3) gene. PLoS One 5: e8689.

[3] Parker HG, Shearin AL, Ostrander EA (2010) Man's best friend becomes biology's best in show: genome analyses in the domestic dog. Annu Rev Genet 44: 309–36.

[4] Boyko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, et al. (2010) A simple genetic architecture underlies morphological variation in dogs. PLoS Biol 8: e1000451.

[5] Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, et al. (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42: 565–569.

[6] Hayes BJ, Pryce J, Chamberlain AJ, Bowman PJ, Goddard ME (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.

[7] Fortes MR, Reverter A, Zhang Y, Collis E, Nagaraj SH, et al. (2010) Association weight matrix for the genetic dissection of puberty in beef cattle. Proc Natl Acad Sci USA 107: 13642–13647.

[8] Peidis P, Giannakouros T, Burow ME, Williams RW, Scott RE (2010) Systems genetics analyses predict a transcription role for P2P-R: molecular confirmation that P2P-R is a transcriptional co-repressor. BMC Syst Biol 4: 14.

[9] Diez D, Wheelock AM, Goto S, Haeggstro¨m JZ, Paulsson-Berne G, et al. (2010) The use of network analyses for elucidating mechanisms in cardiovascular disease. Mol Biosyst 6: 289–304.

[10] Hecker M, Lambeck S, Toepfer S, van Someren E, Guthke R (2009) Gene regulatory network inference: data integration in dynamic models-a review. Biosystems 96: 86–103.

[11] Bondanza S, Bellini M, Roversi G, Raskovic D, Maurelli R, et al. (2006) Piebald trait: Implication of kit mutation on in vitro melanocyte survival and on the clinical application of cultured epidermal autografts. J Invest Dermatol 17: 676–685.

[12] Brooker MG, Dolling CHS (1969) Pigmentation of sheep: III. Piebald pattern in the Merino. Australian J Agricultural Research 20: 523–532.

[13] Yang, Li Y, Nishimura EK, Xin H, Zhou A, et al. (2008) Inhibition of PAX3 by TGF-b modulates melanocyte viability. Mol Cell 32: 554–563.

[14] Karlsson EK, Baranowska I, Wade CM, Salmon Hillbertz NH, Zody MC, et al. (2007) Efficient mapping of mendelian traits in dogs through genome-wide association. Nat Genet 39: 1321–1328.

[15] Smith WJ, Li Y, Ingham A, Collis E, McWilliam SM, et al. (2010) A genomicsinformed, SNP association study reveals FBLN1 and FABP4 as contributing to resistance to fleece rot in Australian Merino sheep. BMC Vet Res 6: 27.

[16] Barabasi A, Oltvai Z (2004) Network biology: Understanding the cell's functional organization. Nat Rev 5: 101–113.

[17] Charrin S, Le Naour F, Oualid M, Billard M, Faure G, et al. (2001) The major CD9 and CD81 molecular partner. Identification and characterization of the complexes. J Biol Chem 276: 14329–14337.

[18] Baudoux B, Castanares-Zapatero D, Leclercq-Smekens M, Berna N, Poumay Y (2000) The tetraspanin CD9 associates with the integrin alpha6beta4 in cultured human epidermal keratinocytes and is involved in cell motility. Eur J Cell Biol 79: 41–51.

[19] Peñas PF, García-Díez A, Sánchez-Madrid F, Yáñez-Mó M (2000) Tetraspanins are localized at motility-related structures and involved in normal human keratinocyte wound healing migration. J Invest Dermatol 14: 1126–1135.

[20] Lanza F, Wolf D, Fox CF, Kieffer N, Seyer JM, et al. (1991) cDNA cloning and expression of platelet p24/CD9. Evidence for a new family of multiple membrane-spanning proteins. J Biol Chem 266: 10638–10645.

[21] Xu S, Zhou Y, Yang S, Ren Y, Zhang C, et al. (2010) Platelet-derived growth factor receptor alpha gene mutations in vitiligo vulgaris. Acta Derm Venereol 90: 131–135.

[22] Flori L, Fritz S, Jaffre´zic F, Boussaha M, Gut I, et al. (2009) The genome response to artificial selection: a case study in dairy cattle. PLoS One 4: e6595.

[23] Hochberg M, Zeligson S, Amariglio N, Rechavi G, Ingber A, et al. (2007) Genomic-scale analysis of psoriatic skin reveals differentially expressed insulinlike growth factor-binding protein-7 after phototherapy. Br J Dermatol 156: 289–300.

[24] Wajapeyee N, Serra RW, Zhu X, Mahalingam M, Green MR (2008) Oncogenic BRAF induces senescence and apoptosis through pathways mediated by the secreted protein IGFBP7. Cell 132: 363–374.

[25] Scurr LL, Pupo GM, Becker TM, Lai K, Schrama D, et al. (2010) IGFBP7 is not required for B-RAF-induced melanocyte senescence. Cell 141: 717–727.

[26] Fleischman RA, Saltman DL, Stastny V, Zneimer S (1991) Deletion of the c-kit protooncogene in the human developmental defect piebald trait. Proc Natl Acad Sci USA 88: 10885–9.

[27] Wu CS, Lan CC, Chiou MH, Yu HS (2006) Basic fibroblast growth factor promotes melanocyte migration via increased expression of p125(FAK) on melanocytes. Acta Derm Venereol 86: 498–502.

[28] Lan CC, Wu CS, Chiou MH, Hsieh PC, Yu HS (2006) Low-energy helium-neon laser induces locomotion of the immature melanoblasts and promotes melanogenesis of the more differentiated melanoblasts: recapitulation of vitiligo repigmentation in vitro. J Invest Dermatol 126: 2119–2126.

[29] Boyd AW, Lackmann M (2001) Signals from Eph and ephrin proteins: a developmental tool kit. Sci STKE 2001: re20.

[30] Li JJ, Liu DP, Liu GT, Xie D (2009) EphrinA5 acts as a tumor suppressor in glioma by negative regulation of epidermal growth factor receptor. Oncogene 28: 1759–1768.

[31] Fitch KR, McGowan KA, van Raamsdonk CD, Fuchs H, Lee D, et al. (2003) Genetics of dark skin in mice. Genes Dev 17: 214–228.

[32] International Federation of Pigment Cell Societies Color Genes Database. Available: http://www.espcr.org/micemut/. Accessed 2010 May 15.

[33] Gunn TM, Miller KA, He L, Hyman RW, Davis RW, et al. (1999) The mouse mahogany locus encodes a transmembrane form of human attractin. Nature 398: 152–156.

[34] He L, Gunn TM, Bouley DM, Lu XY, Watson SJ, et al. (2001) A biochemical function for attractin in agouti-induced pigmentation and obesity. Nat Genet 27: 40–47.

[35] Seo K, Mohanty TR, Choi T, Hwang I (2007) Biology of epidermal and hair pigmentation in cattle: a mini-review. Vet Dermatol 18: 392–400.

[36] Candille SI, Van Raamsdonk CD, Chen C, Kuijper S, Chen-Tsai Y, et al. (2004) Dorsoventral patterning of the mouse coat by Tbx15. PLoS Biol 2: E3.

[37] Blasius AL, Brandl K, Crozat K, Xia Y, Khovananth K, et al. (2009) Mice with mutations of Dock7 have generalized hypopigmentation and white-spotting but show normal neurological function. Proc Natl Acad Sci USA 106: 2706–2711.

[38] Matera I, Watkins-Chow DE, Loftus SK, Hou L, Incao A, et al. (2008) A sensitized mutagenesis screen identifies Gli3 as a modifier of Sox10 neurocristopathy. Hum Mol Genet 17: 2118–2131.

[39] The International Sheep Genomics Consortium, Archibald AL, Cockett NE, Dalrymple BP, Faraut T, et al. (2010) The sheep genome reference sequence: a work in progress. Anim Genet 41: 449–453.

[40] Norris BJ, Whan VA (2008) A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep. Genome Res 18: 1282–1293.

[41] Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, et al. (2007) Genomewide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods 4: 651–657.

[42] Reverter A, Barris W, McWilliam S, Byrne KA, Wang YH, et al. (2005) Validation of alternative methods of data normalization in gene co-expression studies. Bioinformatics 21: 1112–20.

[43] Caraux G, Pinloche S (2005) PermutMatrix: a graphical environment to arrange gene expression profiles in optimal linear order. Bioinformatics 21: 1280–1281.

[44] Buske FA, Bode´n M, Bauer DC, Bailey TL (2010) Assigning roles to DNA regulatory motifs using comparative genomics. Bioinformatics 26: 860–866.

[45] Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, et al. (2005) MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics 21: 2933–2942.

[46] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13: 2498–504.

FIGURES

Figure 1. Genome wide association for piebald. (A) A piebald lamb is pictured with its non-pigmented mother (left hand side). The asymmetrical presentation of pigmentation (the lamb has one pigmented and one white forelimb) characterises this colour morph from Recessive Black (right hand side) which arises through action of the Agouti locus [39]. (B) The genetic relationship between 24 piebald animals and 72 non¬pigmented controls. Allele sharing, or genetic similarity, was calculated between each pair-wise combination of animal using 48,686 SNP. Increasing values of allele sharing are represented using darker colour. (C) Unsupervised hierarchical clustering of the 226 associated SNP (P,0.001) successfully distinguished all piebald animals as genetically distinct from the controls. Animals are arranged into columns and annotated above the matrix using either red (piebald) or black lines (controls). The 226 associated SNP are arranged into rows and each cell indicates the observed genotypic outcome as follows: homozygote (bright red), heterozygote (dark red), alternate homozygote (black).

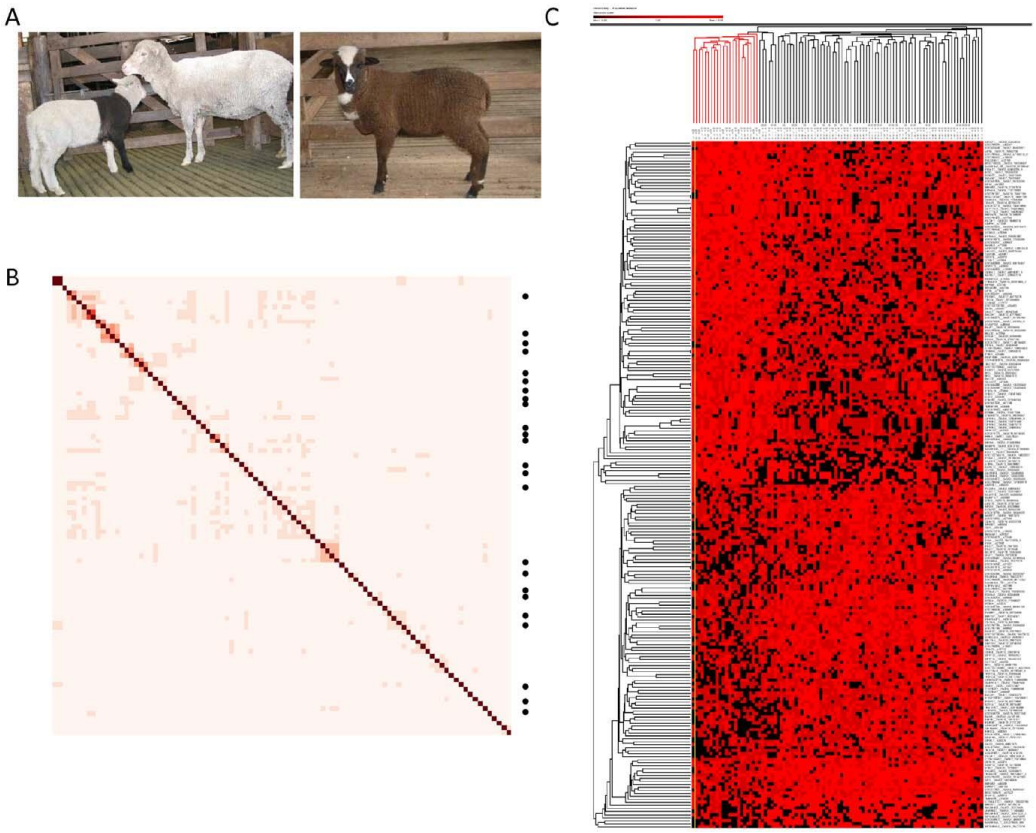


Figure 2. Gene expression relating to pigmentation. (A) Design of the microarray experiment, showing hybridizations (arrows) between five tissue types. Samples were labelled with either red (arrow head) or green dye (arrow tail) and the experiment was repeated in a dye swap using samples from independent biological replicates. A total of 20 hybridizations were performed to compare the following tissue types: white skin tissue from 4 pooled non-pigmented animals (NOR); black skin tissue from a piebald animal (PBB); white skin tissue from a piebald animal (PBW); black skin tissue from a recessive black animal (RSB) and white skin tissue sampled from the inguinal, non-pigmented area from a recessive black animal (RSW). Each sample intervened in four hybridizations (two labelled red and two labelled green). (B) Plot of SNP allele frequency difference between piebald and normal animals (Y-axis) and differential gene expression in contrast 3 (NOR versus PBW) (X-axis) for a set of 1,935 genes. Red symbols represent genes which were both differentially expressed (piebald versus normal) and have a SNP within 2.5 Kb. Black symbols represent genes which either displayed differential expression (P,0.01) or are located within 1 Mb of an associated SNP. The remaining green symbols represent genes which were neither differentially expressed nor located near associated SNP.

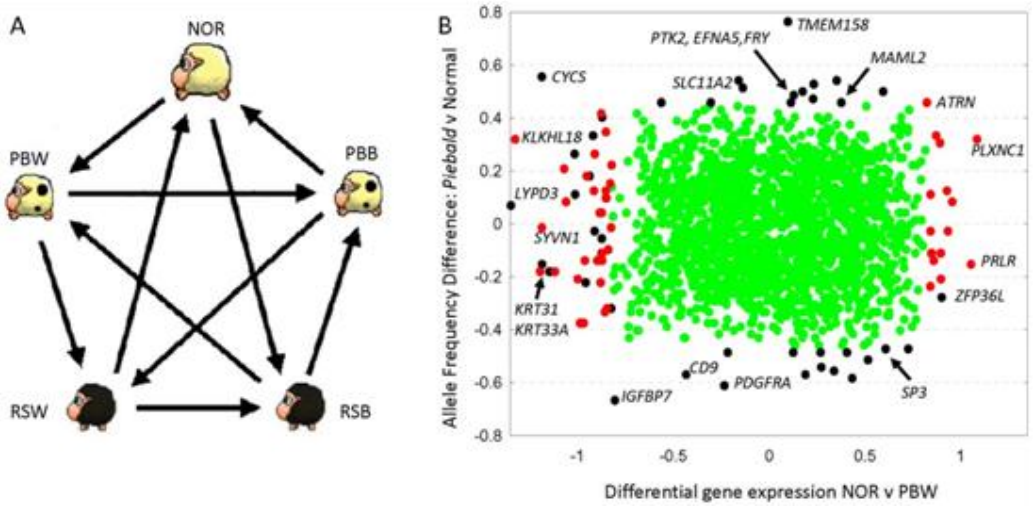
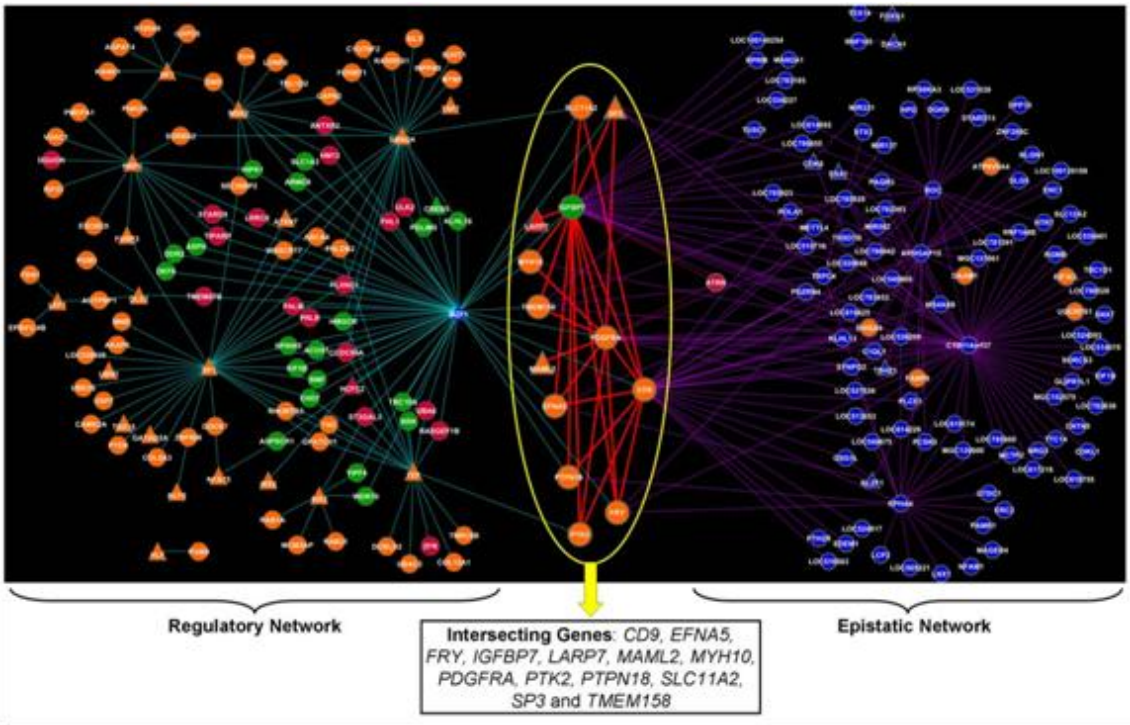


Figure 3. Regulatory and epistatic networks. Promoter sequence analysis of piebald-associated genes (Table S1) was performed to identify transcription factors (TF; represented by triangles) and build a regulatory network (left panel). Two-loci models for all pairwise comparisons among piebald-associated SNPs were fitted to develop an epistatic network (right panel). The overall landscape revealed the presence of thirteen genes located at the intersection of both networks: CD9, EFNA5, FRY, IGFBP7, LARP7, MAML2, MYH10, PDGFRA, PTK2, PTPN18, SLC11A2, SP3 and TMEM158. Gene color indicates over-expression (red), under-expression (green), unchanged expression (orange) and genes not represented on the microarray (blue) in highly correlated contrasts (DE3, DE5 and DE6). The network file (*.cys) is available on request from the authors to explore within Cytoscape [46].



TABLES

Table 1. A selection of key genes involved in piebald.

GeneID	Chr	Mb	Normalised Gene Expression [1]					SNP [2]	Allele Frequency [3]		P-value [4]	OR [4]
			NOR	PBB	PBW	RSB	RSW		Piebald	Non_Piebald		
ATRN	13	55.4	6.15	8.43	8.42	7.59	8.99	OAR13_55578882	0.67	0.44	5.96E-03	2.57
C1QL1	11	46.9	NA	NA	NA	NA	NA	s27884	0.19	0.5	1.47E-04	4.33
CD9	3	225.7	9.63	9.32	9.01	8.33	9.58	s65742	0.29	0.58	6.34E-04	3.3
DOCK7	1	38.6	2.47	4.18	4.36	4.67	4.18	OAR1_38581202	0.67	0.74	3.05E-01	1.45
EFNA5	5	112.9	8.44	9.15	8.91	8.59	9.15	s72651	0.58	0.79	4.40E-03	2.71
FGFR1OP	8	95.4	7.87	6.73	7.28	6.75	7.09	OAR8_95443730	0.9	0.81	1.80E-01	1.98
FRY	10	29.2	6.78	8.5	8.19	7.97	8.81	OAR10_29223007	0.69	0.42	1.54E-03	2.99
GLI3	4	84	5.85	7.48	7.58	6.81	8.39	OAR4_83976143	0.67	0.78	9.95E-02	1.82
IGFBP7	6	78.9	11.37	10.35	9.7	9.87	10.7	s49104	0.54	0.88	8.45E-07	5.92
KRT31	11	43.8	14.06	11.25	12.11	12.77	12.57	OAR11_43737226	0.71	0.67	6.55E-01	1.18
LARP7	6	15.7	6.29	7.99	7.71	6.51	8.76	s14758	0.46	0.67	7.80E-03	2.44
MAML2	15	13.4	7.05	7.57	8.31	7.34	8.6	s71238	0.88	0.65	2.62E-03	3.84
MYH10	11	29.3	5.36	5.77	5.97	6.43	7.54	s29973	0.46	0.74	2.74E-04	3.42
PDGFRA	6	76.2	6.12	5.33	6.22	5.99	6.1	OAR6_76377079	0.31	0.62	2.34E-04	3.56
PTK2	9	16.2	5.06	6.12	5.75	4.78	5.73	s22485	0.58	0.35	5.20E-03	2.55
PTPN18	2	121.1	6.99	7.19	7.87	7.54	8.69	s23889	0.71	0.46	2.68E-03	2.87
SHISA9	24	13.2	7.23	7.57	8.31	8.38	8.88	OAR24_13240468	0.9	0.63	4.28E-04	5.16
SILV	3	174.6	8.16	8.87	8.51	9.15	9.22	s59363	0.79	0.87	2.01E-01	1.73
SLC11A2	3	144.3	7.76	7.44	7.52	7.89	8.05	OAR3_144283427	0.88	0.64	2.03E-03	3.96
SP3	2	143.5	6.2	7.36	8.33	7.49	8.29	OAR2_143746835	0.54	0.78	1.63E-03	2.96
TBX15	1	101.9	4.38	3.67	4.96	5.29	4.53	OAR1_101890858	0.54	0.6	4.46E-01	1.29
TMEM158	19	57.1	6.18	6.95	6.76	6.41	7.77	s04445	0.79	0.41	4.57E-06	5.47

A selection of key genes differentially expressed (DE; $P < 0.01$) in at least one of the seven contrasts and/or near a SNP associated with the piebald conditions ($P < 0.001$). 1The mean normalised gene expression is given for each of five tissue types (NOR, PBB, PBW, RSB and RSW as define in Figure 2A). Values for piebald tissue (PBB or PBW) that are higher than in the normal tissue (NOR) indicate up regulation (eg ATRN). 2SNP with the highest association within a 1 Mb region centered on the gene. 3Allele frequencies are given in both the piebald and non-piebald animal populations. 4Both the P-value and associated odd ratio (OR) are given for SNP2. See Table S1 for full list of 312 key genes. Genes in bold type correspond to genes located at the intersection between the regulatory and the epistatic network (Figure 3).

SUPPORTING INFORMATION

Figure S1. Genome wide association results for piebald. The strength of association expressed as negative log P values (Y axis) are shown for 49034 SNP (X axis) arranged in genomic order from chromosome 1 (far left) to the X chromosome (far right).

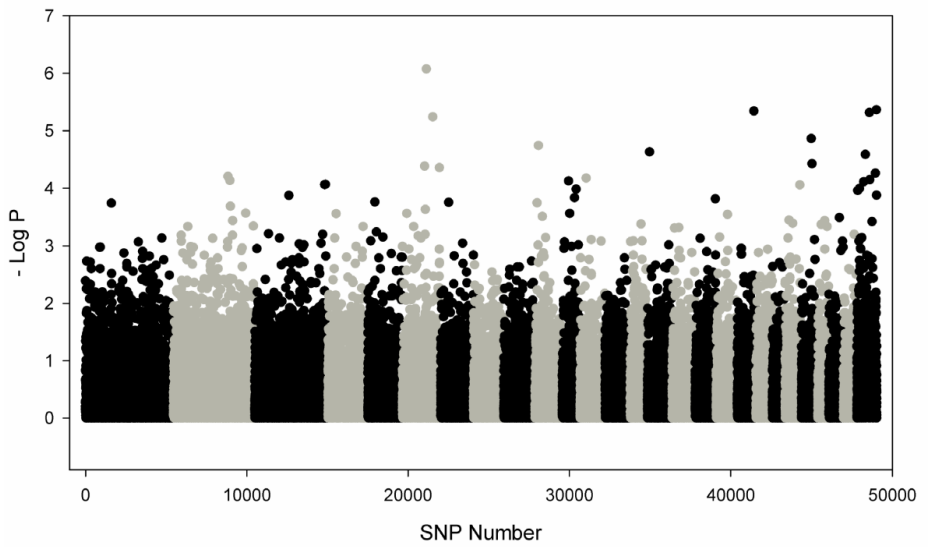


Figure S2. Co-regulation of 54 genes associated with pigmenta¬tion. Gene expression was examined using five tissue types as follows: white skin tissue from a non-pigmented animal (NOR); black skin tissue from a piebald animal (PBB); white skin tissue from a piebald animal (PBW); black skin tissue from a self color black animal (RSB) and white skin tissue from a self color black animal (RSW). The normalised mean expression (NME) within each tissue type is represented using color ranging from green (down regulation) through to red (up regulation). The 54 genes displayed were differentially expressed in at least 4 of the 7 tissue type contrasts examined (Figure 1 describes the 7 contrasts). Hierarchical clustering was performed to identify genes which are expressed in a coordinated way across tissue types, which revealed a set of 11 keratin genes strongly down regulated in piebald tissue.

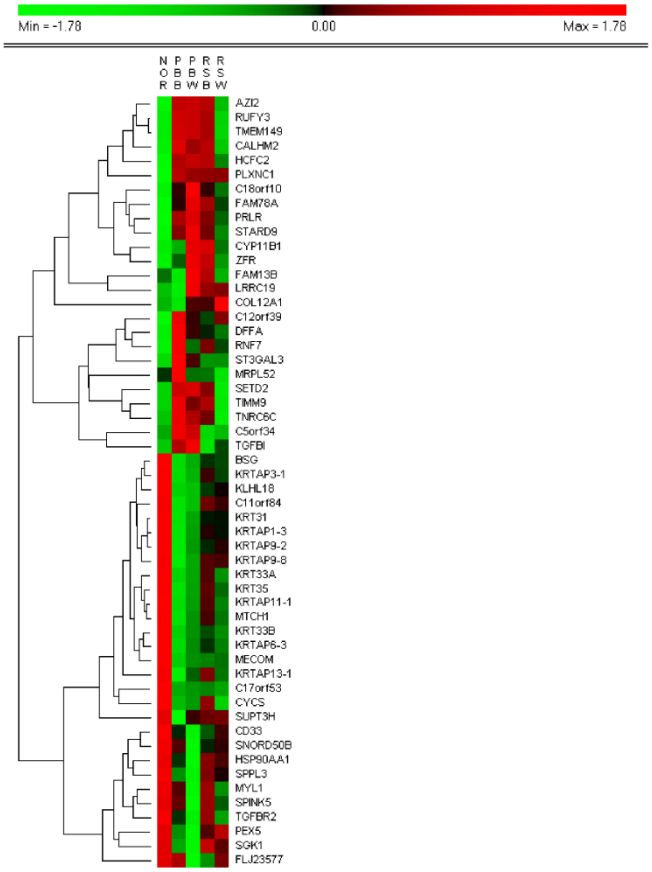


Figure S3. Physical proximity of SNP on the ovine SNP50 BeadChip to genes annotated in Ovine Genome Assembly v1.0 (https://www.biolives.csiro.au/cgi-bin/gbrowse/oar1.0/). A total of 47,275 SNP with known base pair location were examined. Of these, 15,624 SNP (or 33%) were intragenic while 24,921 (53%) were located greater than 20 Kb from the nearest gene. The distribution of the remaining 6,730 (14%) of SNP is shown where the SNP to gene distance was binned in increments of 0.2 Kb from 1 bp to 20 Kb. Importantly, the analysis reported 17,223 SNP were located with 2.5 Kb of the nearest gene which was our empirical threshold for defining SNP as potentially be *cis*-acting.

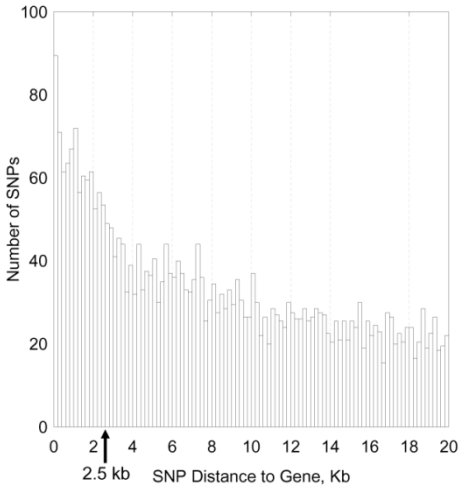


Figure S4. Plot of allele frequency difference versus gene expression for a set of 1,935 genes. For each of seven gene expression contrasts (termed DE1–DE7; refer to the materials and methods), the position of each symbol plots both the SNP allele frequency difference between piebald and non-piebald animals (Y axis) and differential gene expression (X axis). Red symbols represent genes which were both differentially expressed (piebald versus normal) and have a SNP within 2.5 Kb. Black symbols represent genes which either (i) displayed differential expression (p-value,0.05) or (ii) are located within 1 Mb of an associated SNP (p-value,0.01). The remaining green symbols represent genes which were neither differentially expressed nor located near associated SNP.

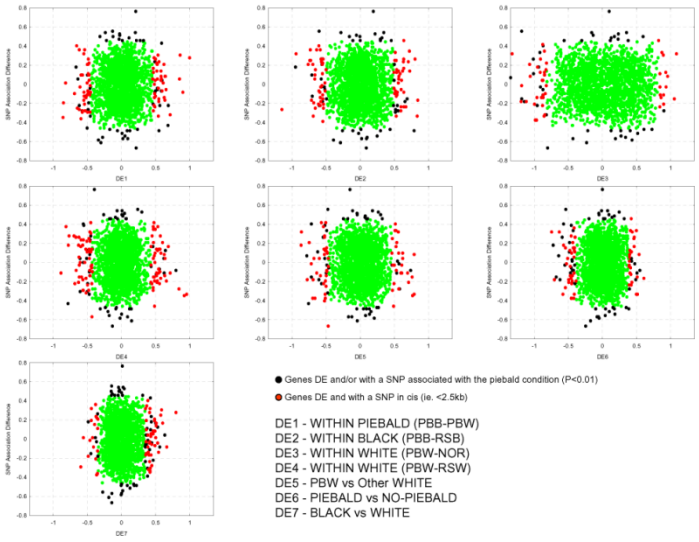


Figure S5. Correlation of differential expression for 1,935 genes. Differential gene expression observed within each of seven contrasts (termed DE1–DE7) were correlated in 21 pair-wise comparisons. The highest correlation (R = 0.88) was observed between DE3 and

DE5, both of which examined the difference between piebald and non-piebald tissues. The only difference between the two contrasts being inclusion of an additional tissue type (RSW) in DE5. Similarly, the second and third highest correlations (DE5 and DE6 R = 0.82; DE3 and DE6 R = 0.79) was also found between contrasts constructed between piebald and non-piebald tissue types. Three contrasts together (DE3, DE5 and DE6) were used to assign genes as either over-expressed (red), under-expressed (green) or having unchanged expression (orange) in the gene networks relating to piebald (Figure 3).

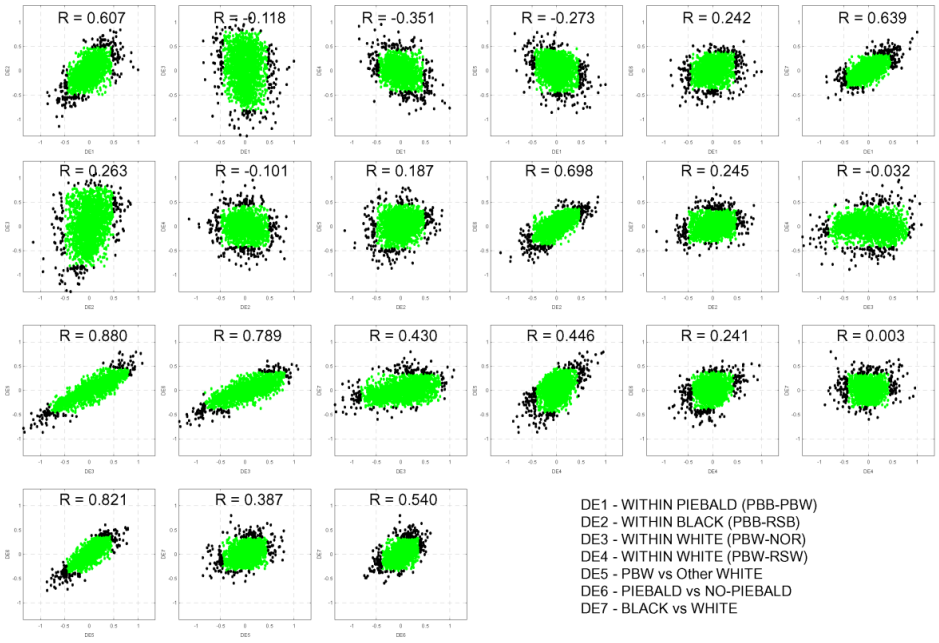
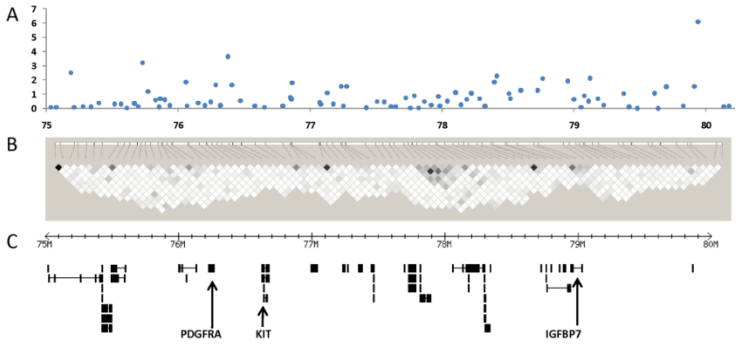


Figure S6. SNP association results for OAR6. (A) The strength of association between each SNP and piebald is given as negative log P values (X axis) across a 5 Mb region of sheep chromosome (OAR) 6 (X axis). This shows the position of the highest ranked SNP (s49104), genome wide, at Mb position 79.9. (B) Pair-wise linkage disequilibrium between SNP, measured as r 2, was calculated using all 96 animals (24 cases and 72 controls) and plotted as a heatmap in Haploview. This shows LD extends for only short distances across the region. The relative location of each annotated gene within the region is shown in (C).



RESUMEN DE RESULTADOS Y DISCUSIÓN DE LOS TRABAJOS ADICIONALES

En el primer Trabajo Adicional aquí presentado, con el objetivo de determinar una escala de medida de diversidad genética (F_{ST}) entre pares de razas, se seleccionaron las razas a comparar entre sí en función de las diferencias existentes *a priori* entre ellas. Por un lado, se cuantificó la diferencia entre líneas dentro de una raza, utilizando animales de raza Lacaune seleccionados para producción de leche y de carne, respectivamente, y dos subgrupos de Merinos australianos. Por el otro, se estimó el valor de F_{ST} entre razas tan alejadas entre sí como europeas y asiáticas. Los resultados obtenidos muestran un gradiente en la diferenciación entre grupos que varía desde un 0,017 entre líneas de una misma raza, hasta un 0,136 al comparar razas europeas y asiáticas. Los valores obtenidos dentro de la raza GCN, entre las líneas de Florida (FN) y Luisiana (LN), muestran un valor intermedio de F_{ST} (0,062), superior al obtenido al comparar razas perfectamente establecidas procedentes de la región Mediterránea, como Merino, Castellana, Churra, Ojalada, Rambouillet y Rasa Aragonesa con aptitudes productivas diferentes (F_{ST} media entre pares de razas 0,042).

Posteriormente, se valoró la subestructura poblacional dentro de la raza GCN, y su relación con el resto de razas incluidas en el estudio. Se observó que los animales de esta raza americana se agrupaban juntos y cerca de las razas españolas, en las que tiene su origen según la información disponible (*The American Livestock Breeders Organization Website*). Asimismo, el análisis de componentes principales, mostró que la población de GCN se subdivide en dos grupos, coincidentes con las subpoblaciones FN y LN, antes incluso de que el análisis sea capaz de diferenciar poblaciones perfectamente establecidas como las razas Rasa Aragonesa y Ojalada.

Con el objetivo de buscar posibles huellas de selección divergente entre estos dos grupos de animales, se analizaron los valores de F_{ST} por marcador, obteniendo como resultado que las mayores diferencias entre ambas se localizan en el cromosoma OAR10, cerca del *locus polled* relacionado con la presencia y forma de los cuernos en ovino (Johnston et al., 2011). Las diferencias detectadas entre las dos líneas de la raza GCN no están causadas por un efecto fundador, ya que esto implicaría una disminución en la diversidad genética en las subpoblaciones (Templeton, 1980) y, en este caso, el número de SNPs polimórficos y los valores de heterocigosis de los animales de la raza GCN muestran unos valores de diversidad

iguales o superiores a los de otras razas estudiadas. Los resultados presentados en este trabajo parecen indicar que el establecimiento de nuevas razas por separación geográfica, como en el caso de la GCN, sigue un procedimiento gradual debido a la deriva génica y, en menor medida, a la selección artificial.

Para concluir con la discusión de estos Trabajos Adicionales, se presentan a continuación los resultados del segundo trabajo incluido en este capítulo. La hipótesis inicial de que un gen mayor fuese el responsable del fenotipo *piebald*, como ocurre en la especie humana con algunos casos de vitíligo y melanoma (Bondanza et al., 2006), se descartó al no obtener resultados concluyentes en el análisis de asociación. Por tanto, utilizando como nexo de unión los genes asociados a marcadores incluidos tanto en el chip de SNPs como en el *microarray* de expresión, así como los potenciales factores de transcripción que puedan regular la expresión de dichos genes, se construyeron dos redes génicas, una denominada "de regulación" y otra "epistática", en base al tipo de interacciones entre genes que sirven como nexo de unión entre los mismos. Con el objetivo de poner de manifiesto los genes incluidos en ambas redes, éstas se combinaron entre sí, obteniéndose en la intersección de las mismas un total de trece genes mostrando asociación y/o expresión diferencial en alguno de los contrastes estudiados.

En este trabajo existen dos limitaciones técnicas importantes. Por una parte, el *microarray* de expresión fue diseñado para el estudio de genes que se expresan en la piel (Smith et al., 2010), pero podría no incluir algún gen biológicamente importante en la pigmentación, como los genes *KIT* y *MC1R*, relacionados con el color y patrón de la capa en caballos (Rieder, 2009). Por otra parte, a pesar de que el chip de SNPs contenía 49.034 marcadores, solo el 35% de estos se localizaban a menos de 2,5 Kb de un gen, según la v1.0 del Genoma Ovino de Referencia (http://www.livestockgenomics.csiro.au/cgi-bin/gbrowse/oar1.0/), disponible en el momento de realización de este análisis. También el estado de "borrador" inicial del genoma ovino podría haber influido sobre la veracidad de las asociaciones SNP-gen identificadas.

A pesar de estas limitaciones, los trece genes detectados en la intersección entre las dos redes objeto de estudio habían sido previamente asociados, de una u otra forma, con la pigmentación. El gen *CD9* codifica para una proteína de superficie celular que juega un papel importante en el desarrollo, activación, crecimiento y motilidad celular (Charrin et al., 2001),

incluyendo la motilidad de queratinocitos en la epidermis (Baudoux et al., 2000; Peñas et al., 2000). Este gen, además, puede activar las plaquetas (Lanza et al., 1991), por lo que la interacción detectada en este trabajo entre este gen y el *PDGFRA* (*platelet-derived growth factor receptor, alpha polypeptide*) parece relevante. De hecho, uno de los candidatos potenciales más interesantes detectados en este trabajo es el gen *PDFGRA*, y sus interacciones con los genes *CD9* e *IGFBP7* (*insulin-like growth factor binding protein 7*). En un trabajo publicado en 2010 por Xu et al. se especulaba que el gen *PDGFRA* era un candidato ideal en relación al vitíligo humano. Además, dentro del contexto de la producción animal se detectaron huellas de selección positiva en este gen en la raza bovina Montbéliarde, posiblemente relacionadas con el característico patrón de color de la capa en esta raza (Flori et al., 2009). El tercero de los genes que aparecen en la intersección entre las redes es el gen *IGFBP7*, que codifica para una proteína con funciones extracelulares. Esta proteína muestra una baja expresión en pieles con psoriasis, que aumenta mediante el tratamiento de esta enfermedad con radiación ultravioleta (Hochberg et al., 2007). Además, puede inducir la apoptosis de líneas celulares de melanoma, por lo que la proteína IGFBP7 está siendo estudiada como posible tratamiento de este tipo de cáncer en la especie humana (Wajapeyee et al., 2008; Scurr et al., 2010). Es importante destacar además, que los genes *PDGFRA* e *IGFBP7* se localizan a una distancia de 3 Mb en el cromosoma OAR6, en una región que contiene el gen *KIT*, asociado inicialmente con el fenotipo *piebald* en ovino (Fleischman et al., 1991). El estudio detallado de esta región puso de manifiesto que el SNP que mostró una mayor asociación con este fenotipo se localiza a 116,0 Kb del gen *PDGFRA* y a 277,7 Kb del gen *KIT*, por lo que la causalidad del segundo gen no puede ser descartada en base a los resultados de este análisis.

El gen que codifica para la proteín kinasa 2 (PTK2), también en la intersección entre las redes, estimula la migración de melanocitos (Wu et al., 2006) e induce la re-pigmentación *in vitro* de células portadoras de vitíligo (Lan et al., 2006). Otro de los genes interesantes, *EFNA5* (*ephrin A5*), es el precursor de una proteína miembro de la familia de las efrinas, implicadas en la señalización celular durante el desarrollo (Boyd y Lackmann, 2001). También se ha descrito un papel regulador de esta proteína, con influencia negativa, sobre el receptor del factor de crecimiento epidérmico (EGFR) (Li et al., 2009a), asociado a su vez con presencia de piel oscura en ratones (Fitch et al., 2003), aunque en principio este gen, *EGFR*, no mostró asociación significativa con el fenotipo en nuestro análisis. Sin embargo, sí

se identificó un bajo nivel de expresión en *HBEGF* (*heparin-binding EGF-like growth factor*), factor de crecimiento similar al EGF, en la piel blanca con respecto a la piel negra y un SNP en los alrededores de este gen presentando una asociación al nivel sugestivo ($P < 0,1$).

Finalmente, la búsqueda en la base de datos *Color Genes database* (IFPCS) mostró que varios de los genes diferencialmente expresados y/o con un SNP cercano asociado con el fenotipo estaban incluidos en la misma. Este conjunto de genes incluía *ATRN* (*atractin*), *DOCK7* (*dedicator of cytokinesis 7*), *FGFR1OP* (*FGFR1 oncogene partner*), *GLI3* (*GLI-Kruppel family member*), *SILV* (*premelanosome protein*) y *TBX15* (*T-box 15*). Entre ellos cabe destacar que el gen *ATRN* es un receptor para la proteína *Agouti* (He et al., 2001) y su función en la síntesis de los componentes de la melanina ha sido recientemente destacada en una revisión sobre la pigmentación de piel y pelo en vacuno (Seo et al., 2007). Tres de los genes restantes, *TBX15*, *DOCK7* y *GLI3*, también muestran una relación similar con la proteína *Agoutí* ya que la ausencia de la proteína TBX15 hace que la expresión de la proteína *Agoutí* se desplace dorsalmente (Candille et al., 2004); mutaciones en el gen *DOCK7* en ratones producen hipo-pigmentación y moteado blanco en la capa de estos animales (Blasius et al., 2009); así mismo la proteína GLI3 está implicada en la especificación de melanoblastos y mutaciones en el gen codificante de la misma causan hipo-pigmentación (Matera et al., 2008).

La conclusión principal de este trabajo, además de la complejidad de las interacciones génicas que producen el fenotipo *piebald* en ovino, es que la integración de diferentes conjuntos de datos utilizando las teorías basadas en la biología de sistemas, constituye un enfoque adecuado para el estudio de fenotipos complejos. Los modernos avances tecnológicos, en especial el acceso a la secuencia completa de los genomas, nos permitirán, utilizando metodología basada en la biología de sistemas, descifrar la arquitectura genética de caracteres complejos.

CONCLUSIONES

PRIMERA,

Los análisis de ligamiento realizados en una población de la raza Churra han permitido confirmar la presencia de dos QTL en los cromosomas ovinos 20 y 3, respectivamente. El primero de ellos afecta al porcentaje de grasa en la leche y se localiza en una amplia región de dicho cromosoma. Los análisis del cromosoma 3 han permitido, además de confirmar el QTL, refinar su localización a una región de 13 cM.

SEGUNDA,

Los análisis basados en el chip de SNPs *OvineSNP50 BeadChip*, disponible para el ganado ovino, han mostrado la utilidad de este tipo de herramientas que, aunque en fase inicial de desarrollo, pueden proporcionar información muy útil para el conocimiento de la arquitectura molecular del genoma. Así, tanto la información de los SNPs, como su localización basada en el genoma virtual ovino, han demostrado su validez para la realización de estudios genómicos en la raza Churra.

TERCERA,

Los cálculos realizados acerca de la estructura del desequilibrio de ligamiento en la raza Churra muestran que éste no es extenso a lo largo del genoma, puesto que el número de bloques haplotípicos en dicha raza es inferior al de otras razas ovinas. Además, se ha comprobado que la base genética de la población estudiada es amplia y que aunque el tamaño efectivo de la misma ha ido descendiendo a lo largo del tiempo, se mantiene un tamaño efectivo que hace que la población sea viable a largo plazo. Basándonos en dichas estimaciones, se necesitaría duplicar el número de marcadores incluidos en el chip disponible actualmente, para poder asegurar la idoneidad de esta herramienta en estudios de mapeo fino y programas de mejora basados en la selección genómica.

CUARTA,

El estudio de asociación a nivel genómico para los caracteres de producción de leche puso de manifiesto la presencia de 15 regiones asociadas significativamente con los fenotipos analizados. De ellas, las localizadas en los cromosomas 2 y 20 mostraron asociaciones significativas con varios caracteres, aunque se requiere de una mayor densidad de marcadores

para el mapeo fino de dichos QTL. Estos resultados vuelven a confirmar al cromosoma 20 como portador de QTL implicados en la producción láctea en la raza Churra. La región que alcanza un mayor nivel de significación en este análisis, se localiza en el cromosoma 3 e influencia los porcentajes de grasa y proteína de la leche. El SNP que mostró una mayor asociación con estos dos caracteres se localiza en el tercer intrón del gen que codifica para la alfa-lactoalbúmina.

QUINTA,

El análisis de secuenciación realizado durante el proceso de mapeo fino del QTL detectado en el cromosoma 3, nos ha permitido identificar 31 SNPs en el gen *LALBA*, candidato posicional y funcional. La recopilación de evidencias genéticas apoyando la posible causalidad de uno de los SNPs, localizado en la región codificante del gen (*LALBA*_g.242T>C), nos ha llevado a proponer esta mutación como QTN responsable del efecto sobre el porcentaje de proteína de la leche en la raza Churra.

CONCLUSIONS

FIRST,

The linkage analyses presented here confirmed the segregation of two QTL on sheep chromosomes 20 and 3, respectively, in a new population of Spanish Churra sheep. The former QTL influences milk fat percentage and its confidence interval covers a wide region of that chromosome. On chromosome 3, apart from confirming the QTL segregation, we were able to refine its location to a 13 cM-long interval.

SECOND,

The evaluation of the *OvineSNP50 BeadChip* performed in this PhD Thesis has shown the usefulness of this genomic tool to understand the molecular architecture of the sheep genome, even when it is still in an early stage of development. The informativity of the SNPs and their location, which was based on the Virtual Sheep Genome, seem to be appropriate to perform genomic studies in Spanish Churra sheep.

THIRD,

The study of the extent of linkage disequilibrium in Spanish Churra sheep showed that it does not expand over long distances as the number of haplotype blocks in this breed is lower than that estimated for other sheep populations. The studied population shows a wide genetic base and an effective population size which, in despite of having decreased through the time, still shows a long term viable size. Based on our calculations a higher-density ovine chip, with double number of SNPs than that currently available, will be required to improve its performance in QTL fine-mapping studies and breeding programs based on genomic selection.

FOURTH,

The genome-wide association analysis for milk production traits identified 15 genomic regions significantly associated with the phenotypes under study. Among them, the QTL on chromosomes 2 and 20 showed significant associations with more than one trait. However, a higher marker density is needed to fine-map these significant regions. The results observed on chromosome 20, confirmed again the importance of this chromosome in relation to milk production traits in Spanish Churra sheep. The most significantly associated region identified

in the whole study, located on chromosome 3, influences milk protein and fat percentages. The highest association for both traits was identified for a SNP located in the third intron of the gene encoding the alpha-lactalbumin protein.

FIFTH,

The sequencing analysis performed in order to fine-map the QTL previously described on chromosome 3 identified 31 SNPs within the *alpha-lactalbumin* gene, which was suggested as a strong positional and functional candidate. We have collected several genetic evidences supporting the causality of a mutation located in the coding region of this gene (*LALBA*_g.242T>C) based on which this SNP is proposed here as the putative QTN underlying the milk protein percentage effect described in Spanish Churra sheep.

RESUMEN

El planteamiento de la presente Tesis Doctoral se realiza en un momento de transición entre la utilización de marcadores microsatélite para la localización de genes de interés en las especies domésticas y los inicios de la utilización de las herramientas genómicas derivadas de los proyectos de secuenciación de los genomas de estas especies. El objetivo final de este trabajo es el mapeo fino de algunas regiones QTL detectadas previamente en la especie ovina, concretamente en una población comercial de la raza Churra. Para ello, en la primera etapa del desarrollo de esta Tesis Doctoral, se utilizó un planteamiento de confirmación y mapeo fino clásico basado en el genotipado de marcadores microsatélite en una nueva población de raza Churra. Posteriormente, la disponibilidad del chip de SNPs, *OvineSNP50 BeadChip*, desarrollado en los últimos años por el Consorcio Internacional para la Genómica Ovina hizo que tanto la confirmación como el mapeo fino de algunos de los QTL detectados previamente se hayan complementado utilizando esta potente herramienta genómica.

Los estudios iniciales, basados en el análisis de ligamiento con marcadores microsatélite en los cromosomas ovinos 3 y 20 nos permitieron confirmar los QTL previamente detectados en dichos cromosomas. Además, en relación al QTL del cromosoma 3, al incrementar la densidad de marcadores y utilizar de forma combinada la información familiar (ligamiento) con la obtenida a nivel poblacional (desequilibrio de ligamiento) se ha refinado la posición del QTL a un intervalo de 13 cM.

Como paso previo a la utilización del chip de SNPs ovino en posteriores análisis se nos planteaba la necesidad de evaluar su aplicación en la raza objeto de estudio, en nuestro caso la raza Churra. Para ello se elaboró un mapa de ligamiento, basado en el la v2.0 del genoma ovino de referencia, en una población formada por familias de medio-hermanas, siguiendo un diseño hija. Así se puso de manifiesto la existencia de diferencias a lo largo del genoma entre las distancias genéticas estimadas entre marcadores adyacentes y las que se obtendrían al convertir las distancias físicas asumiendo un ratio de 1 Mb ~ 1 cM. Estas diferencias tienen, fundamentalmente, tres causas: (i) la calidad del mapa físico variable a lo largo del genoma ovino; (ii) el tamaño del pedigrí estudiado; (iii) verdaderas diferencias en patrones de recombinación a lo largo del genoma. Así, parece recomendable que siempre que se obtenga una asociación basada en un análisis de ligamiento con los marcadores del chip se compruebe que ésta no se encuentra en una región del genoma cuya calidad es baja.

El siguiente paso en el desarrollo de esta Tesis Doctoral fue el estudio de la extensión del desequilibrio de ligamiento en la raza ovina Churra. Los resultados obtenidos mostraron que dicha extensión es mucho menor que la descrita en el ganado vacuno y que también parece ser más reducida que en otras razas ovinas. La información procedente del análisis del desequilibrio de ligamiento entre marcadores se utilizó, además, para el cálculo del tamaño efectivo de la población de la oveja Churra a lo largo de la historia. Este tamaño ha ido descendiendo con el paso del tiempo, hasta mostrar un tamaño efectivo en la última generación de $Ne = 128$. Este valor es sustancialmente mayor que el estimado en el ganado vacuno de leche de raza Holstein ($Ne = 50$) y aunque no supone problemas a corto plazo, sería adecuado tomar ciertas precauciones para que se mantenga por encima del considerado como umbral de viabilidad a largo plazo ($Ne = 100$). En este estudio también se estimaron los coeficientes de parentesco a nivel molecular por diferentes métodos. Aunque la comparación entre métodos se complica por la ausencia de una población base en la que estimar las frecuencias alélicas, los resultados obtenidos pusieron de manifiesto que los métodos moleculares son capaces de calcular el coeficiente de parentesco entre animales en genealogías complejas o en ausencia de un pedigrí fiable. Por último, basándonos en los datos obtenidos se determinó que el número de SNPs necesarios para llevar a cabo de manera exitosa tanto estudios de mapeo fino de regiones asociadas con caracteres de interés económico, como para la utilización de esa información genómica en la predicción de valores genéticos en esta población ovina, sería en torno a los 95.000, el doble de los incluidos en el chip disponible actualmente.

A pesar de la última conclusión del trabajo anteriormente comentado y valorando el sustancial incremento de densidad de mapeo que el *OvineSNP50 BeadChip* ofrece comparado con los mapas de microsatélites utilizados en el anterior barrido genómico realizado en la raza Churra, la presente Tesis Doctoral presenta un estudio de asociación a nivel genómico para caracteres de producción de leche en esta raza. Los objetivos de este trabajo fueron (i) confirmar la segregación a nivel poblacional de QTL detectados en estudios de ligamiento previos, y si fuera posible refinar su posición y (ii) detectar nuevas regiones que no hayan sido previamente identificadas por las limitaciones de los estudios de ligamiento basados en poblaciones comerciales y en mapas de baja densidad. Este análisis de asociación identificó 15 regiones significativamente asociadas con los caracteres objeto de estudio, de las cuales únicamente los resultados del cromosoma 3 alcanzaron el nivel genómico de significación. De

entre las regiones asociadas a nivel de significación cromosómica, los cromosomas 2 y 20 mostraron tener influencia sobre más de un carácter, aunque parece que para entender la arquitectura genética de estas regiones de posible interés se necesitarían estudios y análisis adicionales basados, principalmente, en el incremento de la densidad de marcadores.

En el análisis de asociación realizado los únicos resultados significativos a nivel de significación genómico se localizaron en torno a la región de 137 Mb en el cromosoma 3. La influencia de esta región sobre el porcentaje de proteína en la raza Churra había sido previamente descrita y los resultados se replicaron inicialmente con marcadores microsatélite en un trabajo incluido en la presente Tesis Doctoral. Para los caracteres significativamente asociados, porcentajes de proteína y grasa, el máximo valor del test estadístico coincidió con un SNP del chip comercial localizado en el tercer intrón del gen que codifica para la alfa-lactoalbúmina (LALBA). En base a los resultados, y a su papel biológico, el gen *LALBA*, fue sometido a análisis adicionales con el fin de testar su posible causalidad con respecto al efecto genético descrito en la oveja Churra. A falta de pruebas funcionales que confirmen la verdadera naturaleza de los efectos detectados, y sin poder descartar totalmente que esta mutación se encuentre en completo LD con el verdadero QTN, el trabajo realizado da claras muestras de la relación directa de la mutación localizada en el primer exón del gen (*LALBA*_g.242T>C) con el control genético de los porcentajes de leche de grasa, y sobre todo de proteína, en el ganado ovino, al menos en la raza Churra. En cualquier caso, la naturaleza comercial de la población en la que se ha identificado este posible QTN haría posible la directa utilización del mismo como marcador en el programa de mejora genética de la raza Churra, poniendo así a disposición de los ganaderos de ANCHE la posibilidad de tomar decisiones de mejora en base a la información molecular.

SUMMARY

The development of this PhD Thesis has been influenced by the transition that has taken place in the last few years in the field of livestock genomics, from microsatellite-based genome scans to genome-wide association studies based on high-throughput technologies derived from the sequencing projects of the genome of domestic species. The final aim of this work was to fine-map genomic regions underlying QTL for milk production traits previously described in Spanish Churra sheep. For that purpose, firstly we used a classical replication and fine-mapping approach based on microsatellite linkage mapping with the purpose of confirming the presence of two of the previously detected QTL in this breed. Secondly, the confirmation and fine-mapping of previously detected QTL was completed using data from the commercial SNP chip, *OvineSNP50 BeadChip*, developed by the International Sheep Genomics Consortium.

Initial linkage studies on chromosomes 3 and 20 confirmed the QTL previously described in those chromosomes in a new Spanish Churra population. Moreover, the position of the QTL located on chromosome 3 was narrowed to a 13 cM-long region using an increased marker density and the combination of linkage and linkage disequilibrium information.

Before using the *OvineSNP50 BeadChip* in subsequent genetic studies, we evaluated its usefulness in the breed under study. Based on the physical order from the Ovine Genome Assembly v2.0 for the markers included in the chip, we built a linkage map using the genotypes from the half-sib families of the Spanish Churra population under study. We identified certain differences between the marker genetic distances obtained through this linkage analysis and those based on the physical distances of the sequence assuming a conversion ratio of 1 Mb ~ 1 cM. The divergence from this ratio varied across the genome which can be caused by (i) the quality of the physical map, (ii) the size of the pedigree under study, or (iii) true differences along the genome. Based on these observations, for every linkage association detected across the genome using the SNP chip, the quality of the reference sequence should be checked in order to avoid possible false positive due to a limited quality of the assembly.

Afterwards, the extent of the linkage disequilibrium was estimated in our resource commercial population of Spanish Churra sheep. The results of this study showed that the extent of linkage disequilibrium in Churra is much shorter than that reported in cattle

populations and it also appears to be shorter in other sheep populations. The analysis of the linkage disequilibrium in Churra sheep was also used to estimate the effective population size along the breed history. This size has been decreasing through the time showing for the last generation an estimated effective number of *Ne* = 128. This value is substantially higher than that reported in Holstein dairy cattle (*Ne* = 50). Although the current effective size in Churra does not represent a problem in the short term future, care should be taken to avoid crossing the value that is considered the long term viable population size threshold (*Ne* = 100). In this work we also calculated the inbreeding coefficients based on molecular information using three different methodologies. Although it is difficult to compare between methods because of the absence of a base population to use for the estimation of the allele frequencies, the molecular information seems to be able to calculate inbreeding when the pedigree is not accurate or in the absence of pedigree information. Finally, we estimated the number of SNPs needed to accurately perform fine-mapping studies and implement genomic selection in this sheep breed. Based on our calculations a higher-density ovine chip, with double number of SNPs than that currently available (about 95,000 SNP), would be required for these purposes.

Despite the last conclusion of the previously mentioned study and considering that the density offered by the *OvineSNP50 BeadChip* represents an important progress when compared with that of the microsatellite-based genome scan previously carried out in Churra sheep, the present PhD Thesis summarizes the results of a genome-wide association analysis for milk production traits performed on the basis of the chip genotypes. The main objectives of this work were (i) to confirm the presence of previously described QTL based on linkage analysis and, if possible, redefine their confidence interval and (ii) to detect new QTL segregating at the population level that were not detected previously based on the limitations shown by linkage and microsatellite-based studies. There were 15 genomic regions significantly associated with the phenotypes under study. Only the significant results detected on chromosome 3 reached the genome-wise significance level. Among the chromosome-wise significant results, chromosomes 2 and 20 were associated with more than one of the studied traits. However, it seems that further efforts, mainly focused in the increase of the marker density, will be needed to reveal the true nature of the effects detected in these two regions with regard to the Spanish Churra sheep population.

The genome-wise significant association described in this study was located at position 137 Mb on chromosome 3 and influenced milk protein and fat percentages. The

effect on milk protein percentage has already been described in Spanish Churra sheep and the results were replicated in the first stage of this PhD Thesis. The highest significant association for both traits was identified for a SNP located in the third intron of the gene encoding the alpha-lactalbumin protein (LALBA). Based on these results and the biological role known for this protein, additional studies were carried out in order to test the possible causality of the *LALBA* gene. Although functional assays complemented by physiological data will be required to finally prove the causality of the *LALBA*_g.242T>C mutation, this genetic variant is proposed here as the putative QTN underlying the milk protein percentage effect described in Spanish Churra sheep. However, we cannot discard this mutation to be in perfect LD with the genuine causal mutation of the OAR3 QTL reported here. In any case, the *LALBA*_g.242T>C genetic variant appears to be a useful marker for taking advantage of the commercial nature of the population that has served to identify this potential QTN and could be directly used to assist Churra sheep breeders to make informed decisions based on genomic information.

BIBLIOGRAFÍA

Abasht B, Dekkers JC, Lamont SJ. (2006). Review of quantitative trait loci identified in the chicken. *Poultry Science* 85, 2079-2096.

Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. (2010). Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science* 93, 743-752.

Alipanah M, Kalashnikova L, Rodionov G. (2007). Association of prolactin gene variants with milk production traits in Russian Red Pied cattle. *Iranian Journal of Biotechnology* 5, 158-161.

Allah MA, Abass SF, Allam FM. (2011). Factors affecting the milk yield and composition of Rahmani and Chios sheep. *International Journal of Livestock Production* 2, 24-30.

Álvarez L, Gutiérrez-Gil B, San Primitivo F, De la Fuente LF, Arranz JJ. (2006). Influence of Prion Protein genotypes on milk production traits in Spanish Churra sheep. *Journal of Dairy Science* 89, 1784-1791.

Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. (2010). Data quality control in genetic case-control association studies. *Nature Protocols* 5, 1564-1573.

Andersson L. (2009). Genome-wide association analysis in domestic animals: a powerful approach for genetic dissection of trait loci. *Genetica* 136, 341-349.

Arnyasi M, Komlósi I, Lien S, Czeglédi L, Nagy S, Jávor A. (2009). Searching for DNA markers for milk production and composition on chromosome 6 in sheep. *Journal of Animal Breeding and Genetics* 126, 142-147.

Arranz JJ, Gutiérrez-Gil B, Bayón Y, De la Fuente LF, San Primitivo F. (2006). Principios generales de la detección de genes de interés productivo. *OVIS* 101, 7-32.

Arranz JJ, Gutiérrez-Gil B. (*en prensa*). Detection of QTL underlying milk traits in sheep: an update. En: Milk Production – Advanced genetic traits, cellular mechanisms, animal management and health. ISBN: 978-953-51-0766-8.

Ashwell MS, Heyen DW, Sonstegard TS, Van Tassell CP, Da Y, VanRaden PM, Ron M, Weller JI, Lewin HA. (2004). Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle. *Journal of Dairy Science* 87, 468-475.

Bai Y, Sartor M, Cavalcoli J. (2012). Current status and future perspectives for sequencing livestock genomes. *Journal of Animal Science and Biotechnology* 3, 8.

Barillet F, Arranz JJ, Carta A, Jacquiet P, Stear M, Bishop S. (2006). Final Consolidated Report of the European Union Contract of Acronym genesheepsafety, QTLK5-CT-2000-00656, p. 145.

Barillet F, Astruc JM, Lagriffoul G, Aguerre X, Bonaïti B. (2008). Selecting milk composition and mastitis resistance by using a part lactation sampling design in French Manech red faced dairy sheep breed. *Proceedings of the 36th Biennial Session of the International Committee for Animal Recording (ICAR), Niagara Falls, NY*, pp: 129-136.

Barillet F, Marie C, Jacquin M, Lagriffoul G, Astruc JM. (2001). The French Lacaune dairy sheep breed: Use in France and abroad in the last 40 years. *Livestock Production Science* 71, 17–29.

Barillet F. (1997). Genetics of milk production. En: *The Genetics of sheep*, pp: 539-564. (Ed. L Piper y A Ruvinsky). ISBN 0-85199-200-5.

Barillet F. (2007). Genetic improvement for dairy production in sheep and goats. *Small Ruminant Research* 70, 60-75.

Bassett NS, Currie MJ, Breier BH, Klempt M, Min SH, McCutcheon SN, MacKenzie DDS, Gluckrnan PD. (1998). The effects of ovine placental lactogen and bovine growth hormone on hepatic and mammary gene expression in lactating sheep. *Growth Hormone and IGF Research* 8, 439-446.

Baudoux B, Castanares-Zapatero D, Leclercq-Smekens M, Berna N, Poumay Y. (2000). The tetraspanin CD9 associates with the integrin alpha6beta4 in cultured human

epidermal keratinocytes and is involved in cell motility. *European Journal of Cell Biology* 79, 41-51.

Becker D, Tetens J, Brunner A, Bürstel D, Ganter M, Kijas J; International Sheep Genomics Consortium, Drögemüller C. (2010). Microphthalmia in Texel sheep is associated with a missense mutation in the paired-like homeodomain 3 (PITX3) gene. *PLoS ONE* 5, e8689.

Beraldi D, McRae AF, Gratten J, Slate J, Visscher PM, Pemberton JM. (2006). Development of a linkage map and mapping of phenotypic polymorphisms in a free-living population of Soay sheep (Ovis aries). *Genetics* 173, 1521-1537.

Bidanel JP, Rosendo A, Iannuccelli N, Riquet J, Gilbert H, Caritez JC, Billon Y, Amigues Y, Prunier A, Milan D. (2008). Detection of quantitative trait loci for teat number and female reproductive traits in Meishan × Large White F2 pigs. *Animal: an International Journal of Animal Bioscience* 2, 813-820.

Blasius AL, Brandl K, Crozat K, Xia Y, Khovananth K, Krebs P, Smart NG, Zampolli A, Ruggeri ZM, Beutler BA. (2009). Mice with mutations of Dock7 have generalized hypopigmentation and white-spotting but show normal neurological function. *Proceedings of the National Academy of Sciences of the United States of America* 106, 2706-2711.

Blott S, Kim JJ, Moisio S, Schmidt-Küntzel A, Cornet A, Berzi P, Cambisano N, Ford C, Grisart B, Johnson D, Karim L, Simon P, Snell R, Spelman R, Wong J, Vilkki J, Georges M, Farnir F, Coppieters W. (2003). Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. *Genetics* 163, 253-266.

Board on Agriculture and Natural Resources (BANR). (2008). The US dairy sheep industry. En: *Changes in the Sheep Industry in the United States: Making the Transition from Tradition*, pp: 295-308. (Ed. Committee on the Economic Development and Current Status of the Sheep Industry in the United States, National Research Council). ISBN: 978-0-309-12161-3.

Bohmanova J, Sargolzaei M, Schenkel FS. (2010). Characteristics of linkage disequilibrium in North American Holsteins. *BMC Genomics* 11, 421.

Boichard D, Fritz S, Rossignol MN, Guillaume F, Colleau JJ, Druet T. (2006). Implementation of Marker Assisted Selection: practical lessons from dairy cattle. En: *Proceedings of the 8th World Congress on Genetics Applied to Livestock Production,* CD-ROM communication 22-11. Instituto Prociencia, Belo Horizonte, Brazil.

Boichard D, Guillaume F, Baur A, Croiseau P, Rossignol MN, Boscher MY, Druet T, Genestout L, Colleau JJ, Journaux L, Ducrocq V, Fritz S. (2012). Genomic Selection in French Dairy Cattle. *Animal Production Science* 52, 115-120.

Bolormaa S, Hayes BJ, Savin K, Hawken R, Barendse W, Arthur PF, Herd RM, Goddard ME. (2011). Genome-wide association studies for feedlot and growth traits in cattle. *Journal of Animal Science* 89, 1684-1697.

Bondanza S, Bellini M, Roversi G, Raskovic D, Maurelli R, Paionni E, Paterna P, Dellambra E, Larizza L, Guerra L. (2006). Piebald trait: implication of kit mutation on in vitro melanocyte survival and on the clinical application of cultured epidermal autografts. *The Journal of Investigative Dermatology* 127, 676-686.

Boston WS, Bleck GT, Conroy JC, Wheeler MB, Miller DJ. (2001). Short communication: effects of increased expression of alpha-lactalbumin in transgenic mice on milk yield and pup growth. *Journal of Dairy Science* 84, 620-622.

Bouwman AC, Bovenhuis H, Visker MH, van Arendonk JA. (2011). Genome-wide association of milk fatty acids in Dutch dairy cattle. *BMC Genetics* 12, 43.

Bovine Genome Sequencing and Analysis Consortium, Elsik, CG, Tellam, RL, Worley, KC, Gibbs, RA, Muzny, DM, Weinstock, GM, Adelson, DL, Eichler, EE et al. (2009). The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* 324, 522-528.

Boyd AW, Lackmann M. (2001). Signals from Eph and ephrin proteins: a developmental tool kit. *Science's STKE : signal transduction knowledge environment* 2001, re20.

Broad TE, Hayes H, Long SE. (1997). Cytogenetics: physical chromosome maps. En: *The Genetics of sheep*, pp: 241-296. (Ed. L Piper y A Ruvinsky). ISBN 0-85199-200-5.

Brooker MG, Dolling CHS. (1969). Pigmentation of sheep: III. Piebald pattern in the Merino. *Australian Journal of Agricultural Research* 20, 523–532.

Brym P, Kamiński S, Wójcik E. (2005). Nucleotide sequence polymorphism within exon 4 of the bovine prolactin gene and its associations with milk performance traits. *Journal of Applied Genetics* 46, 179-185.

Calvo JH, Marcos S, Beattie AE, Gonzalez C, Jurado JJ, Serrano M. (2004a). Ovine alpha-amylase genes: isolation, linkage mapping and association analysis with milk traits. A*nimal Genetics* 35, 329-332.

Calvo JH, Marcos S, Jurado JJ, Serrano M. (2004b). Association of the heart fatty acid binding protein (FABP3) gene with milk traits in Manchega breed sheep. *Animal Genetics* 35, 347-349.

Calvo JH, Martínez-Royo A, Beattie AE, Dodds KG, Marcos-Carcavilla A, Serrano M. (2006). Fine mapping of genes on sheep chromosome 1 and their association with milk traits. *Animal Genetics* 37, 205-210.

Candille SI, Van Raamsdonk CD, Chen C, Kuijper S, Chen-Tsai Y, Russ A, Meijlink F, Barsh GS. (2004). Dorsoventral patterning of the mouse coat by Tbx15. *PLoS Biology* 2, E3.

Carta A, Casu S, Salaris S. (2009). Invited review: Current state of genetic improvement in dairy sheep. *Journal of Dairy Science* 92, 5814-5833.

Carta A, Casu S, Usai MG, Addis M, Fiori M, Fraghì A, Miari S, Mura L, Piredda G, Schibler L, Sechi T, Elsen JM, Barillet F. (2008). Investigating the genetic component of fatty acid content in sheep milk. *Small Ruminant Research* 79, 22-28.

Cavanagh CR, Jonas E, Hobbs M, Thomson PC, Tammen I, Raadsma HW. (2010). Mapping Quantitative Trait Loci (QTL) in sheep. III. QTL for carcass composition traits derived from CT scans and aligned with a meta-assembly for sheep and cattle carcass QTL. *Genetics Selection Evolution* 42, 36.

Chamberlain JS, Gibbs RA, Ranier JE, Nguyen PN, Caskey CT. (1988). Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification. *Nucleic Acids Research* 16, 11141-11156.

Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, Hirschhorn JN, Abecasis G, Altshuler D, Bailey-Wilson JE, Brooks LD, Cardon LR, Daly M, Donnelly P, Fraumeni JF Jr, Freimer NB, Gerhard DS, Gunter C, Guttmacher AE, et al. (2007). Replicating genotype-phenotype associations. *Nature* 447, 655-660.

Charrin S, Le Naour F, Oualid M, Billard M, Faure G, Hanash SM, Boucheix C, Rubinstein E. (2001). The major CD9 and CD81 molecular partner. Identification and characterization of the complexes. *Journal of Biological Chemistry* 276, 14329-14337.

Chiofalo L, Micari P. (1987). Present knowledge of the variation of the milk proteins in the sheep population reared in Sicily. Experimental observations. *Scienze e Tecnologie Lattiero Casearie* 38, 104–114.

Clop A, Marcq F, Takeda H, Pirottin D, Tordoir X, Bibé B, Bouix J, Caiment F, Elsen JM, Eychenne F, Larzul C, Laville E, Meish F, Milenkovic D, Tobin J, Charlier C, Georges M. (2006). A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nature Genetics* 38, 813-818.

Clop A, Ovilo C, Perez-Enciso M, Cercos A, Tomas A, Fernandez A, Coll A, FolchJM, Barragan C, Diaz I, Oliver MA, Varona L, Silio L. (2003). Detection of QTL affecting fatty acid composition in the pig. *Mammalian Genome* 14, 650-656.

Crawford AM, Montgomery GW, Pierson CA, Brown T, Dodds KG, Sunden SL, Henry HM, Ede AJ, Swarbrick PA, Berryman T, Penty JM, Hill DF. (1994). Sheep linkage mapping: nineteen linkage groups derived from the analysis of paternal half-sib families. *Genetics* 137, 573-579.

Crooks L, Sahana G, de Koning DJ, Lund MS, Carlborg O. (2009). Comparison of analyses of the QTLMAS XII common dataset. II: genome-wide association and fine mapping. *BMC Proceedings* 3, Suppl 1, S2.

Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA. (2010). The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 185, 1021-1031.

Daetwyler HD, Schenkel FS, Sargolzaei M, Robinson JA. (2008). A genome scan to detect quantitative trait loci for economically important traits in Holstein cattle using two methods and a dense single nucleotide polymorphism map. *Journal of Dairy Science* 91, 3225-3236.

Dalrymple BP, Kirkness EF, Nefedov M, McWilliam S, Ratnakumar A, Barris W, Zhao S, Shetty J, Maddox JF, O'Grady M, Nicholas F, Crawford AM, Smith T, de Jong PJ, McEwan J, Oddy VH, Cockett NE; International Sheep Genomics Consortium. (2007). Using comparative genomics to reorder the human genome sequence into a virtual sheep genome. *Genome Biology* 8, R152.

Davis GH. (2006). Fecundity genes in sheep. *Animal Reproduction Science* 82-83, 247-253.

De Gortari MJ, Freking BA, Cuthbertson RP, Kappes SM, Keele JW, Stone RT, Leymaster KA, Dodds KG, Crawford AM, Beattie CW. (1998). A second-generation linkage map of the sheep genome. *Mammalian Genome* 9, 204-209.

De la Chevrotière C, C Bishop S, Arquet R, Bambou JC, Schibler L, Amigues Y, Moreno C, Mandonnet N. (2012). Detection of quantitative trait loci for resistance to gastrointestinal nematode infections in Creole goats. *Animal Genetics*, doi: 10.1111/j.1365-2052.2012.02341.x.

De la Fuente LF, Baro JA, San Primitivo F. (1995). Breeding programme for the Spanish Churra sheep breed. En: *Strategies for sheep and goat breeding = Stratégies pour l'amélioration génétique des ovins et caprins*, pp: 165-172. (Ed. Gabiña D., Zaragoza). ISSN 1022-1379.

De la Fuente LF, Gonzalo C, Sánchez JP, Rodríguez R, Carriedo JA, San Primitivo F. (2011). *Canadian Journal of Animal Science* 91, 585-591.

De Roos APW, Hayes BJ, Spelman RJ, Goddard ME. (2008). Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics* 179, 1503-1512.

Dekkers JC. (2004). Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *Journal of Animal Science* 82, E-Suppl, E313-328.

Dib C, Fauré S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissenbach J. (1996). A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380, 152-154.

Diez D, Wheelock AM, Goto S, Haeggström JZ, Paulsson-Berne G, Hansson GK, Hedin U, Gabrielsen A, Wheelock CE. (2010). The use of network analyses for elucidating mechanisms in cardiovascular disease. *Molecular Biosystems* 6, 289-304.

Diez-Tascón C, Bayón Y, Arranz JJ, De la Fuente F, San Primitivo F. (2001). Mapping quantitative trait loci for milk production traits on ovine chromosome 6. *Journal of Dairy Research* 68, 389-397.

Druet T, Fritz S, Boussaha M, Ben-Jemaa S, Guillaume F, Derbala D, Zelenika D, Lechner D, Charon C, Boichard D, Gut IG, Eggen A, Gautier M. (2008). Fine mapping of quantitative trait loci affecting female fertility in dairy cattle on BTA03 using a dense single-nucleotide polymorphism map. *Genetics* 178, 2227-2235.

Duchemin SI, Colombani C, Legarra A, Baloche G, Larroque H, Astruc JM, Barrillet F, Robert-Granié C, Manfredi E. (2012). Genomic selection in the French Lacaune dairy sheep breed. *Journal of Dairy Science* 95, 2723-2733.

Dybus A, Grzesiak W, Kamieniecki H, Szatkowska I, Sobek Z, Błaszczyk1 P, Czerniawska-Piątkowska E, Zych S, Muszyńska M. (2005). Association of genetic variants of bovine prolactin with milk production traits of Black-and-White and Jersey cattle. *Archiv für Tierzucht* 2, 149-156.

Eberlein A, Takasuga A, Setoguchi K, Pfuhl R, Flisikowski K, Fries R, Klopp N, Fürbass R, Weikard R, Kühn C. (2009). Dissection of genetic factors modulating fetal growth in cattle indicates a substantial role of the non-SMC condensin I complex, subunit G (NCAPG) gene. *Genetics* 183, 951-964.

Esmailizadeh AK, Morris CA, Cullen NG, Kruk ZA, Lines DS, Hickey SM, Dobbie PM, Bottema CDK, Pitchford WS. (2011). Genetic mapping of quantitative trait loci for meat quality and muscle metabolic traits in cattle. *Animal Genetics* 42, 592-599.

Farnir F, Grisart B, Coppieters W, Riquet J, Berzi P, Cambisano N, Karim L, Mni M, Moisio S, Simon P, Wagenaar D, Vilkki J, Georges M. (2002). Simultaneous mining of linkage and linkage disequilibrium to fine map quantitative trait loci in outbred half-sib pedigrees: revisiting the location of a quantitative trait locus with major effect on milk production on bovine chromosome 14. *Genetics* 161, 275-287.

Fischer RA. (1918). The correlation between relatives: the supposition of mendelian inheritance. *Transactions of the Royal Society of Edinburgh* 52, 399.

Fitch KR, McGowan KA, van Raamsdonk CD, Fuchs H, Lee D, Puech A, Hérault Y, Threadgill DW, Hrabé de Angelis M, Barsh GS. (2003). Genetics of dark skin in mice. *Genes and Development* 17, 214-228.

Fleet MR, Mahar TJ, Turk JA. (2002). Merino crossbreeding and objectionable sheep fibres: the problem and potential solution. *Wool Technology and Sheep Breeding* 50, 650-656.

Fleet MR, Smith DH. (1990). Pigmented Fibres in White-Skirted Fleece Wool from Piebald Merino Sheep. *Australian Journal of Agricultural Research* 41, 155-166.

Fleischman RA, Saltman DL, Stastny V, Zneimer S. (1991). Deletion of the c-kit protooncogene in the human developmental defect piebald trait. *Proceedings of the National Academy of Sciences of the United States of America* 88, 10885-10889.

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

FriedmanJ, Hastie T, Tibshirani R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33.

Fujii J, Otsu K, Zorzato F, de Leon S, Khanna VK, Weiler JE, O'Brien PJ, MacLennan DH. (1991). Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science* 253, 448-451.

Galloway SM, Hanrahan V, Dodds KG, Potts MD, Crawford AM, Hill DF. (1996). A linkage map of the ovine X chromosome. *Genome Research* 6, 667-677.

Gao Y, Du ZQ, Wei WH, Yu XJ, Deng XM, Feng CG, Fei J, Feng JD, Li N, Hu XX. (2009). Mapping quantitative trait loci regulating chicken body composition traits. *Animal Genetics* 40, 952-954.

García-Fernández M, Gutiérrez-Gil B, García-Gámez E, Sánchez JP, Arranz JJ. (2010a). Detection of quantitative trait loci affecting the milk fatty acid profile on sheep chromosome 22: Role of the stearoyl-CoA desaturase gene in Spanish Churra sheep. *Journal of Dairy Science* 93, 348-357.

García-Fernández M, Gutiérrez-Gil B, García-Gámez E, Sánchez JP, Arranz JJ. (2010b). The identification of QTL that affect the fatty acid composition of milk on sheep chromosome 11. *Animal Genetics* 41, 324-328.

García-Fernández M, Gutiérrez-Gil B, Sánchez JP, Morán JA, García-Gámez E, Alvarez L, Arranz JJ. (2011). The role of bovine causal genes underlying dairy traits in Spanish Churra sheep. *Animal Genetics* 42, 415-420.

García-Gámez E, García-Fernández M, Moran JA, Gutiérrez-Gil B, Sanchez JP, Arranz, JJ. (2010). Association analysis between Prolactin SNPs and milk production traits in Spanish Churra Sheep: preliminary results. En: *Proceedings of the 32nd Conference of the International Society of Animal Genetics, Edinburgh (Scotland)*, P3004.

Gaye P, Hue-Delahaie D, Mercier JC, Soulier S, Vilotte JL, Furet JP. (1987). Complete nucleotide sequence of ovine alpha-lactalbumin mRNA. *Biochimie* 69, 601-608.

Georges M. (2007). Mapping, fine mapping, and molecular dissection of quantitative trait Loci in domestic animals. *Annual Review of Genomics and Human Genetics* 8, 131-162.

Georges M. (2011). The long and winding road from correlation to causation. *Nature Genetics* 43, 180-181.

Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R, Pasquino AT, Sargeant LS, Sorensen A, Steele MR, Zhao X, Womack JE, Hoeschele I. (1995). Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics* 139, 907-920.

Glazier AM, Nadeau JH, Aitman TJ. (2002). Finding genes that underlie complex traits. *Science* 298, 2345-2349.

Glowatzki-Mullis ML, Muntwyler J, Gaillard C. (2007). Cost-effective parentage verification with 17-plex PCR for goats and 19-plex PCR for sheep. *Animal Genetics* 38, 86-88.

Goddard M. (2009). Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136, 245-257.

Goldammer T, Di Meo GP, Lühken G, Drögemüller C, Wu CH, Kijas J, Dalrymple BP, Nicholas FW, Maddox JF, Iannuzzi L, Cockett NE. (2009). Molecular cytogenetics and gene mapping in sheep (Ovis aries, 2n = 54). *Cytogenetics and Genome Research* 126, 63-76.

Grisart B, Coppieters W, Farnir F, Karim L, Ford C, Berzi P, Cambisano N, Mni M, Reid S, Simon P, Spelman R, Georges M, Snell R. (2002). Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Research* 12, 222-231.

Grisart B, Farnir F, Karim L, Cambisano N, Kim JJ, Kvasz A, Mni M, Simon P, Frère JM, Coppieters W, Georges M. (2004). Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. *Proceedings of the National Academy of Sciences of the United States of America* 101, 2398-2403.

Gu X, Feng C, Ma L, Song C, Wang Y, Da Y, Li H, Chen K, Ye S, Ge C, Hu X, Li N. (2011). Genome-wide association study of body weight in chicken F2 resource population. *PLoS One* 6, e21872.

Gutiérrez-Gil B, Álvarez L, De la Fuente LF, Sánchez JP, San Primitivo F, Arranz JJ. (2011). A genome scan for quantitative trait loci affecting body conformation traits in Spanish Churra dairy sheep. *Journal of Dairy Science* 94, 4119-4128.

Gutiérrez-Gil B, Arranz JJ, El-Zarei MF, Alvarez L, Pedrosa S, San Primitivo F, Bayón Y. (2008a). A male linkage map constructed for QTL mapping in Spanish Churra sheep.*Journal of Animal Breeding and Genetics* 125, 201-204.

Gutiérrez-Gil B, El-Zarei MF, Alvarez L, Bayón Y, De la Fuente LF, San Primitivo F, Arranz JJ. (2009). Quantitative trait loci underlying milk production traits in sheep. *Animal Genetics* 40, 423-434.

Gutiérrez-Gil B, El-Zarei MF, Alvarez L, Bayón Y, De la Fuente LF, San Primitivo F, Arranz JJ. (2008b). Quantitative trait loci underlying udder morphology traits in dairy sheep. *Journal of Dairy Science* 91, 3672-3681.

Gutiérrez-Gil B, El-Zarei MF, Bayón Y, Álvarez L, De la Fuente LF, San Primitivo F, Arranz JJ. (2007). Short communication: detection of quantitative trait loci influencing somatic cell score in Spanish Churra sheep. *Journal of Dairy Science* 90, 422-426.

Hadjipavlou G, Bishop SC. (2009). Age-dependent quantitative trait loci affecting growth traits in Scottish Blackface sheep. *Animal Genetics* 40, 165-175.

Han J, Kraft P, Nan H, Guo Q, Chen C, Qureshi A, Hankinson SE, Hu FB, Duffy DL, Zhao ZZ, Martin NG, Montgomery GW, Hayward NK, Thomas G, Hoover RN, Chanock S, Hunter DJ. (2008). A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genetics* 4, e1000074.

Harris BJ, Johnson DL, Spelman RJ. (2008). Genomic selection in New Zealand and the implications for national genetic evaluation. En: *Identification, breeding, production,*

health and recording of farm animals. Proceedings of the 36th ICAR biennial session, No. 13, Niagara Falls, USA, pp: 325-330. ISBN: 92-95014-09-X.

Hayes B, Goddard ME. (2001). The distribution of the effects of genes affecting quantitative traits in livestock. *Genetics Selection Evolution* 33, 209-229.

Hayes B. (2008). 'QTL mapping, MAS and Genomic Selection'. *Course Notes.*

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. (2009). Invited review: Genomic selection in dairy cattle: progress and challenges. *Journal of Dairy Science* 92, 433-443.

Hayes BJ, Visscher PM, McPartlan HC, Goddard ME. (2003). Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Research* 13, 635-643.

He L, Gunn TM, Bouley DM, Lu XY, Watson SJ, Schlossman SF, Duke-Cohan JS, Barsh GS. (2001). A biochemical function for attractin in agouti-induced pigmentation and obesity. *Nature Genetics* 27, 40-47.

Hernández-Sánchez J, Chatzipli A, Beraldi D, Gratten J, Pilkington JG, Pemberton JM. (2010). Mapping quantitative trait loci in a wild population using linkage and linkage disequilibrium analyses. *Genetics Research* 92, 273-281.

Hill WG, Robertson A. (1981). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* 38, 226-231.

Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, Furey TS, Harte RA, Hsu F, Hillman-Jackson J, Kuhn RM, Pedersen JS, Pohl A, Raney BJ, Rosenbloom KR, Siepel A, Smith KE, Sugnet CW, Sultan-Qurraie A, et al. (2006). The UCSC Genome Browser Database: update 2006. *Nucleic Acids Research* 34, D590-8.

Hochberg M, Zeligson S, Amariglio N, Rechavi G, Ingber A, Enk CD. (2007). Genomic-scale analysis of psoriatic skin reveals differentially expressed insulin¬like growth factor-binding protein-7 after phototherapy. *The British Journal of Dermatology* 156, 289-300.

Horseman ND. (1999). Prolactin and mammary gland development. *Journal of Mammary Gland Biology and Neoplasia* 4, 79-88.

Howard B, Ashworth A. (2006). Signalling pathways implicated in early mammary gland morphogenesis and breast cancer. *PLoS Genetics* 2, e112.

Igl BW, Konig IR, Ziegler A. (2009). What do we mean by 'replication' and 'validation' in genome-wide association studies? *Human Heredity* 67, 66-68.

Ihara N, Takasuga A, Mizoshita K, Takeda H, Sugimoto M, Mizoguchi Y, Hirano T, Itoh T, Watanabe T, Reed KM, Snelling WM, Kappes SM, Beattie CW, Bennett GL, Sugimoto Y. (2004). A comprehensive genetic map of the cattle genome based on 3802 microsatellites. *Genome Research* 14, 1987-1998.

International Human Genome Sequencing Consortium (IHGSC). (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860-921.

International Sheep Genomics Consortium, Archibald AL, Cockett NE, Dalrymple BP, Faraut T, Kijas JW, Maddox JF, McEwan JC, Hutton Oddy V, Raadsma HW, Wade C, Wang J, Wang W, Xun X. (2010). The sheep genome reference sequence: a work in progress. *Animal Genetics* 41, 449-453.

Israel C, Weller JI. (1998). Estimation of candidate gene effects in dairy cattle populations. *Journal of Dairy Science* 81, 1653-1662.

Jeffreys AJ, Wilson V, Thein SL. (1985). Hypervariable 'minisatellite' regions in human DNA. *Nature* 314, 67-73.

Jiang L, Liu J, Sun D, Ma P, Ding X, Yu Y, Zhang Q. (2010) Genome wide association studies for milk production traits in Chinese Holstein population. *PLoS ONE* 5, e13661.

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. *Molecular Ecology* 20, 2555-2566.

Jonas E, Thomson PC, Hall EJ, McGill D, Lam MK, Raadsma HW. (2011). Mapping quantitative trait loci (QTL) in sheep. IV. Analysis of lactation persistency and extended lactation traits in sheep. *Genetics Selection Evolution* 43, 22.

Karamichou E, Richardson RI, Nute GR, Gibson KP, Bishop SC. (2006). Genetic analyses and quantitative trait loci detection, using a partial genome scan, for intramuscular fatty acid composition in Scottish Blackface sheep. *Journal of Animal Science* 84, 3228-3238.

Karlsson EK, Baranowska I, Wade CM, Salmon Hillbertz NH, Zody MC, Anderson N, Biagi TM, Patterson N, Pielberg GR, Kulbokas EJ 3rd, Comstock KE, Keller ET, Mesirov JP, von Euler H, Kämpe O, Hedhammar A, Lander ES, et al.. (2007) Efficient mapping of mendelian traits in dogs through genome-wide association. *Nature Genetics* 39, 1321-1328.

Khatkar MS, Thomson PC, Tammen I, Raadsma HW. (2004). Quantitative trait loci mapping in dairy cattle: review and meta-analysis. *Genetics Selection Evolution* 36, 163-190.

Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, San Cristobal M, Servin B, McCulloch R, Whan V, Gietzen K, Paiva S, Barendse W, Ciani E, Raadsma H, McEwan J, Dalrymple B; International Sheep Genomics Consortium Members. (2012). Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biology* 10, e1001258.

Kileh-Wais M, Elsen JM, Vignal A, Feves K, Vignoles F, Fernandez X, Manse H, Davail S, Andrés JM, Bastianelli D, Bonnal L, Filangi O, Baéza E, Guéméné D, Genet C, Bernadet MD, Dubos F, Marie-Etancelin C. (2012). Detection of single and pleiotropic QTL controlling metabolism, meat and liver quality traits of the overfed inter-specific hybrid mule duck. *Journal of Animal Science, en prensa*.

Kim ES, Kirkpatrick BW. (2009). Linkage disequilibrium in the North American Holstein population. *Animal Genetics* 40, 279-288.

Kim JJ, Farnir F, Savell J, Taylor JF. (2003). Detection of quantitative trait loci for growth and beef carcass fatness traits in a cross between Bos taurus (Angus) and Bos indicus (Brahman) cattle. *Journal of Animal Science* 81, 1933-1942.

Kralickova S, Pokorna M, Kuchtik J, Filipcik R. (2012). Effect of parity and stage of lactation on milk yield, composition and quality of organic sheep milk. *Acta universitatis agriculturae et silviculturae mendelianae brunensis* 52, 71-78.

Lan CC, Wu CS, Chiou MH, Hsieh PC, Yu HS. (2006). Low-energy helium¬neon laser induces locomotion of the immature melanoblasts and promotes melanogenesis of the more differentiated melanoblasts: recapitulation of vitiligo repigmentation in vitro. *Journal of Investigative Dermatology* 126, 2119-2126.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860-921.

Lanza F, Wolf D, Fox CF, Kieffer N, Seyer JM, Fried VA, Coughlin SR, Phillips DR, Jennings LK. (1991). cDNA cloning and expression of platelet p24/CD9. Evidence for a new family of multiple membrane-spanning proteins. *Journal of Biological Chemistry* 266, 10638-10645.

Lee SH, van der Werf JH, Kim NK, Lee SH, Gondro C, Park EW, Oh SJ, Gibson JP, Thompson JM. (2011). QTL and gene expression analyses identify genes affecting carcass weight and marbling on BTA14 in Hanwoo (Korean Cattle). *Mammalian Genome* 22, 589-601.

Legarra A, Ugarte E. (2001). Genetic parameters of milk traits in Latxa dairy sheep. *Animal Science* 73, 407-412.

Li JJ, Liu DP, Liu GT, Xie D. (2009). EphrinA5 acts as a tumor suppressor in glioma by negative regulation of epidermal growth factor receptor. *Oncogene* 28, 1759-1768.

Li MH, Strandén I, Kantanen J. (2009). Genetic diversity and pedigree analysis of the Finnsheep breed. *Journal of Animal Science* 87, 1598-1605.

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

Liu G, Jennen DG, Tholen E, Juengst H, Kleinwächter T, Hölker M, Tesfaye D, Un G, Schreinemachers HJ, Murani E, Ponsuksili S, Kim JJ, Schellander K, Wimmers K. (2007). A genome scan reveals QTL for growth, fatness, leanness and meat quality in a Duroc-Pietrain resource population. *Animal Genetics* 38, 241-252.

Liu W, Li D, Liu J, Chen S, Qu L, Zheng J, Xu G, Yang N. (2011). A genome-wide SNP scan reveals novel loci for egg production and quality traits in white leghorn and brown-egg dwarf layers. *PLoS ONE* 6, e28600.

Liu Y, Qin X, Song XZ, Jiang H, Shen Y, Durbin KJ, Lien S, Kent MP, Sodeland M, Ren Y, Zhang L, Sodergren E, Havlak P, Worley KC, Weinstock GM, Gibbs RA. (2009). Bos taurus genome assembly. *BMC Genomics* 10, 180.

Lundén A, Lindersson M. (1998). α-lactalbumin polymorphism in relation to milk lactose. *Proceedings 6th World Congress on Genetics Applied to Livestock Production. Armidale NSW Australia* 25, pp: 47-50.

Lynch M, Walsh B. (1998). Genetics and analysis of quantitative traits. (Ed. Sinauer Associates Inc. Sunderland, Massachusetts, USA). ISBN: 0-87893-481-2.

Maddox JF, Cockett NE. (2007). An update on sheep and goat linkage maps and other genomics resources. *Small Ruminant Research* 70, 4-20.

Maddox JF, Davies KP, Crawford AM, Hulme DJ, Vaiman D, Cribiu EP, Freking BA, Beh KJ, Cockett NE, Kang N, Riffkin CD, Drinkwater R, Moore SS, et al. (2001). An enhanced linkage map of the sheep genome comprising more than 1000 loci. *Genome Research* 11, 1275-1289.

Mai MD, Sahana G, Christiansen FB, Guldbrandtsen B. (2010). A genome-wide association study for milk production traits in Danish Jersey cattle using a 50K single nucleotide polymorphism chip. *Journal of Animal Science* 88, 3522-3528.

Mateescu RG, Thonney ML. (2010a). Genetic mapping of quantitative trait loci for milk production in sheep. *Animal Genetics* 41, 460-466.

Mateescu RG, Thonney ML. (2010b). Genetic mapping of quantitative trait loci for aseasonal reproduction in sheep. *Animal Genetics* 41, 454-459.

Matera I, Watkins-Chow DE, Loftus SK, Hou L, Incao A, Silver DL, Rivas C, Elliott EC, Baxter LL, Pavan WJ. (2008). A sensitized mutagenesis screen identifies Gli3 as a modifier of Sox10 neurocristopathy. *Human Molecular Genetics* 17, 2118-2131.

McRae AF, McEwan JC, Dodds KG, Wilson T, Crawford AM, Slate J. (2002). Linkage disequilibrium in domestic sheep. *Genetics* 160, 1113-1122.

Meadows JR, Chan EK, Kijas JW. (2008). Linkage disequilibrium compared between five populations of domestic sheep. *BMC Genetics* 9, 61.

Meredith BK, Kearney FJ, Finlay EK, Bradley DG, Fahey AG, Berry DP, Lynn DJ. (2012). Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. *BMC Genetics* 13, 21.

Meuwissen TH, Hayes BJ, Goddard ME. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819-1829.

Meuwissen TH, Karlsen A, Lien S, Olsaker I, Goddard ME. (2002). Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. *Genetics* 161, 373-379.

Meuwissen TH. (2009). Genetic management of small populations: A review. *Acta Agriculturae Scandinavica, Section A - Animal Science* 59, 71-79.

Ministerio de Medio Ambiente, Medio Rural y Marino (MAGRAMA). (2010). Anuario de estadística Ministerio de Medio Ambiente y Medio Rural y Marino 2010. NIPO: 770-09-265-1.

Moioli B, D'Andrea L, Pilla F. (2007). Candidate genes affecting sheep and goat milk quality. *Small Ruminant Research* 68, 179-192.

Mömke S, Kerkmann A, Wöhlke A, Ostmeier M, Hewicker-Trautwein M, Ganter M, Kijas J; International Sheep Consortium, Distl O. (2011). A frameshift mutation within LAMC2 is responsible for Herlitz type junctional epidermolysis bullosa (HJEB) in black headed mutton sheep. *PLoS ONE* 6(5), e18943.

Mullen MP, Berry DP, Howard DJ, Diskin MG, Lynch CO, Giblin L, Kenny DA, Magee DA, Meade KG, Waters SM. (2011). Single Nucleotide Polymorphisms in the Insulin-Like Growth Factor 1 (IGF-1) gene are associated with performance in Holstein-Friesian dairy cattle. *Frontiers in Genetics* 2, 3.

Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. (1986). Specific enzymatic amplification of DNA in vitro; the polymerase chain reaction. *Cold Spring Harbor Symposium in Quantitative Biology* 51, 263-273.

Neimann-Sorensen A, Robertson A. (1961). The association between blood groups and 24 several production characteristics in three Danish cattle breeds. *Acta Agriculturae Scandinavica* 11, 163-196.

Olsen HG, Hayes BJ, Kent MP, Nome T, Svendsen M, Lien S. (2010). A genome wide association study for QTL affecting direct and maternal effects of stillbirth and dystocia in cattle. *Animal Genetics* 41, 273-280.

Olsen HG, Nilsen H, Hayes B, Berg PR, Svendsen M, Lien S, Meuwissen T. (2007). Genetic support for a quantitative trait nucleotide in the ABCG2 gene affecting milk composition of dairy cattle. *BMC Genetics* 21, 32.

Onteru SK, Fan B, Du ZQ, Garrick DJ, Stalder KJ, Rothschild MF. (2012). A whole-genome association study for pig reproductive traits. *Animal Genetics* 43, 18-26.

Oravcova M, Margetin M, Peskovicova D, Dano J, Milerski M, Hetenyi1 L, Polak P. (2007). Factors affecting ewe's milk fat and protein content and relationships between milk yield and milk components. *Czech Journal of Animal Science* 52, 189-198.

Othmane MH, Carriedo JA, San Primitivo F, De La Fuente LF. (2002a). Genetic parameters for lactation traits of milking ewes: protein content and composition, fat,

somatic cells and individual laboratory cheese yield. *Genetics Selection Evolution* 34, 581-596.

Othmane MH, De La Fuente LF, Carriedo JA, San Primitivo F. (2002b). Heritability and genetic correlations of test day milk yield and composition, individual laboratory cheese yield, and somatic cell count for dairy ewes. *Journal of Dairy Science* 85, 2692-2698.

Peñas PF, García-Díez A, Sánchez-Madrid F, Yáñez-Mó M. (2000). Tetraspanins are localized at motility-related structures and involved in normal human keratinocyte wound healing migration. *Journal of Investigative Dermatology* 14, 1126-1135.

Pennisi E. (2003). A low number wins the GeneSweep pool. *Science* 300, 1484.

Pirisi A, Lauret A, Dubeuf JP. (2007). Basic and incentive payments for goat and sheep milk in relation to quality. *Small Ruminant Research* 68, 167-178.

Plante Y, Gibson JP, Nadesalingam J, Mehrabani-Yeganeh H, Lefebvre S, Vandervoort G, Jansen GB. (2001). Detection of quantitative trait loci affecting milk production traits on 10 chromosomes in Holstein cattle. *Journal of Dairy Science* 84, 1516-1524.

Pryce JE, Bolormaa S, Chamberlain AJ, Bowman PJ, Savin K, Goddard ME, Hayes BJ. (2010). A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. *Journal of Dairy Science* 93, 3331-3345.

Purvis IW, Franklin IR. (2004). Major genes and QTL influencing wool production and quality: a review. *Genetics Selection Evolution* 37, Suppl 1, S97-107.

Qanbari S, Pimentel EC, Tetens J, Thaller G, Lichtner P, Sharifi AR, Simianer H. (2010). The pattern of linkage disequilibrium in German Holstein cattle. *Animal Genetics* 41, 346-356.

Raadsma HW, International Sheep Genomics Consortium (ISGC). (2010). Linkage disequilibrium in the sheep genome: Findings from the ISGC HapMap iniciative. En: *Proceedings of the XVIII Plant and Animal Genome (PAG),* San Diego, CA, USA.

Disponible *online*: http://www.sheephapmap.org/pag/cattle-sheepRAADSMApag2010.pdf.

Raadsma HW, Jonas E, McGill D, Hobbs M, Lam MK, Thomson PC. (2009a). Mapping quantitative trait loci (QTL) in sheep. II. Meta-assembly and identification of novel QTL for milk production traits in sheep. *Genetics Selection Evolution* 41, 45.

Raadsma HW, Thomson PC, Zenger KR, Cavanagh C, Lam MK, Jonas E, Jones M, Attard G, Palmer D, Nicholas FW. (2009b). Mapping quantitative trait loci (QTL) in sheep. I. A new male framework linkage map and QTL for growth rate and body weight. *Genetics Selection Evolution* 41, 34.

Rao YS, Liang Y, Xia MN, Shen X, Du YJ, Luo CG, Nie QH, Zeng H, Zhang XQ. (2008). Extent of linkage disequilibrium in wild and domestic chicken populations. *Hereditas* 145, 251-257.

Rebbeck TR, Spitz M, Wu X. (2004). Assessing the function of genetic variants in candidate gene association studies. *Nature Reviews. Genetics* 5, 589-597.

Rieder S. (2009). Molecular tests for coat colours in horses. *Journal of Animal Breeding and Genetics* 126, 415-424.

Roldán DL, Rabasa AE, Saldaño S, Holgado F, Poli MA, Cantet RJ. (2008). QTL detection for milk production traits in goats using a longitudinal model. *Journal of Animal Breeding and Genetics* 125, 187-193.

Ron M, Weller JI. (2007). From QTL to QTN identification in livestock--winning by points rather than knock-out: a review. *Animal Genetics* 38, 429-439.

Rothschild MF, Hu ZL, Jiang Z. (2007). Advances in QTL mapping in pigs. *International Journal of Biological Sciences* 3, 192-197.

Rozen F. (1999). Búsqueda de regiones del cromosoma 9 ovino con influencia sobre la producción láctea. *Tesis Doctoral.* Universidad de León.

San Primitivo F, De la Fuente LF. (2000). Situación actual de la oveja de raza Churra. *Archivos de zootecnia* 49, 185-186.

San Primitivo F. (1998). Situación de la mejora genética del ganado ovino lechero en España. *Producción Ovina y Caprina* XXIII: Ponencia 3. 37-46.

Sánchez Belda A, Sánchez Trujillano MC. (1986). Razas Ovinas Españolas. Publicaciones de Extensión Agraria: *Ministerio de Agricultura, Pesca y Alimentación* 1986.

Sánchez JP, García-Gámez E, Gutiérrez-Gil B, Arranz JJ. (2012). Marker assisted selection for milk production traits in Spanish Churra sheep. En: *Book of Abstracts of the 63rd Annual Meeting of the European Federation of Animal Science*, pp: 309. ISBN 978-90-8686-206-1.

Scharch C, Süß R, Fahr RD. (2000). Factors affecting milk traits and udder health in East Friesian milk sheep. En: *Proceedings of the 6th Great Lakes Dairy Sheep Symposium, Ontario, Canada*, pp: 1235-136.

Schnabel RD, Kim JJ, Ashwell MS, Sonstegard TS, Van Tassell CP, Connor EE, Taylor JF. (2005). Fine-mapping milk production quantitative trait loci on BTA6: analysis of the bovine osteopontin gene. *Proceedings of the National Academy of Sciences of the United States of America* 102, 6896-6901.

Schopen GC, Visker MH, Koks PD, Mullaart E, van Arendonk JA, Bovenhuis H. (2011). Whole-genome association study for milk protein composition in dairy cattle. *Journal of Dairy Science* 94, 3148-3158.

Schreiweis MA, Hester PY, Settar P, Moody DE. (2006). Identification of quantitative trait loci associated with egg quality, egg production, and body weight in an F2 resource population of chickens. *Animal Genetics* 37, 106-112.

Schulman NF, Sahana G, Iso-Touru T, Lund MS, Andersson-Eklund L, Viitala SM, Värv S, Viinalass H, Vilkki JH. (2009). Fine mapping of quantitative trait loci for mastitis resistance on bovine chromosome 11. *Animal Genetics* 40, 509-515.

Schulman NF, Viitala SM, de Koning DJ, Virta J, Mäki-Tanila A, Vilkki JH. (2004). Quantitative trait Loci for health traits in Finnish Ayrshire cattle. *Journal of Dairy Science* 87, 443-449.

Scurr LL, Pupo GM, Becker TM, Lai K, Schrama D, Haferkamp S, Irvine M, Scolyer RA, Mann GJ, Becker JC, Kefford RF, Rizos H. (2010). IGFBP7 is not required for B-RAF-induced melanocyte senescence. *Cell* 141, 717-727.

Sellner EM, Kim JW, McClure MC, Taylor KH, Schnabel RD, Taylor JF. (2007). Board-invited review: Applications of genomic information in livestock. *Journal of Animal Science* 85, 3148-3158.

Seo K, Mohanty TR, Choi T, Hwang I. (2007). Biology of epidermal and hair pigmentation in cattle: a mini-review. *Veterinary Dermatology* 18, 392-400.

Seroussi E. (2009). The concordance test emerges as a powerful tool for identifying quantitative trait nucleotides: lessons from BTA6 milk yield QTL. *Animal Genetics* 40, 230-234.

Setoguchi K, Furuta M, Hirano T, Nagao T, Watanabe T, Sugimoto Y, Takasuga A. (2009). Cross-breed comparisons identified a critical 591-kb region for bovine carcass weight QTL (CW-2) on chromosome 6 and the Ile-442-Met substitution in NCAPG as a positional candidate. *BMC Genetics* 10, 43.

Sharif S, Mallard BA, Wilkie BN, Sargeant JM, Scott HM, Dekkers JC, Leslie KE. (1998). Associations of the bovine major histocompatibility complex DRB3 (BoLA-DRB3) with production traits in Canadian dairy cattle. *Animal Genetics* 30, 157-160.

Shook GE, Schutz MM. (1994). Selection on somatic cell score to improve resistance to mastitis in the United States. *Journal of Dairy Science* 77, 648-658.

Singh M, Lam M, McGill D, Thomson PC, Cavanagh JA, Zenger KR, Raadsma HW. (2007). High resolution mapping of quantitative trait loci on ovine chromosome 3 and 20 affecting protein yield and lactation persistency. *Proceedings of the Association for the Advancement of Animal Breeding and Genetics* 17, 565-568.

Smith TP, Matukumalli LK, Sonstegard TS, Schnabel R, Taylor JF, Haudenschild CD, Lawley CT, Moore SS, Van Tassell CP. (2008). Generation of large numbers of SNP in cattle by coupling reduced genome representation with high throughput

sequencing. *Plant and Animal Genome Conference XVI (PAG*), San Diego, CA, USA, P90.

Smith WJ, Li Y, Ingham A, Collis E, McWilliam SM, Dixon TJ, Norris BJ, Mortimer SI, Moore RJ, Reverter A. (2010). A genomics-informed, SNP association study reveals FBLN1 and FABP4 as contributing to resistance to fleece rot in Australian Merino sheep. *BMC Veterinary Research* 6, 27.

Stinckens A, Van den Maagdenberg K, Luyten T, Georges M, De Smet S, Buys N. (2007). The RYR1 g.1843C>T mutation is associated with the effect of the IGF2 intron3-g.3072G>A mutation on muscle hypertrophy. *Animal Genetics* 38, 67-71.

Stinnakre MG, Vilotte JL, Soulier S, Mercier JC. (1994). Creation and phenotypic analysis of alpha-lactalbumin-deficient mice. *Proceedings of the National Academy of Sciences of the United States of America* 91, 6544-6548.

Sved JA. (1971). Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theoretical population biology* 2, 125-141.

Taylor J. (2012) Current Status of Genomic Selection in US Beef Cattle. *Plant and Animal Genome Conference XX (PAG)*, San Diego, USA, W140.

Templeton AR. (1981). The theory of speciation VIA the founder principle. *Genetics* 94, 1011-1038.

Thomas I, Kihiczak GG, Fox MD, Janniger CK, Schwartz RA. (2004). Piebaldism: an update. International Journal of Dermatology 43, 716-719.

Thomsen H, Reinsch N, Xu N, Looft C, Grupe S, Kühn C, Brockmann GA, Schwerin M, Leyhe-Horn B, Hiendleder S, Erhardt G, Medjugorac I, Russ I, Förster M, Brenig B, Reinhardt F, Reents R, Blümel J, Averdunk G, Kalm E. (2001). Comparison of estimated breeding values, daughter yield deviations and de-regressed proofs within a whole genome scan for QTL. *Journal of Animal Breeding and Genetics* 118, 357-370.

Ugarte E, Ruiz R, Gabiña D, Beltrán de Heredia I. (2001). Impact of high-yielding foreign breeds on the Spanish dairy sheep industry. *Livestock Production Science* 71, 3-10.

Ugarte E, Serrano M, De la Fuente LF, Pérez-Guzmán MD, Alfonso L, Gutiérrez JP. (2002). Situación actual de los programas de mejora genética en ovino de leche. *XI Reunión de Mejora Genética, Pamplona, España.*

Uleberg E, Widerøe IS, Grindflek E, Szyda J, Lien S, Meuwissen TH. (2005). Fine mapping of a QTL for intramuscular fat on porcine chromosome 6 using combined linkage and linkage disequilibrium mapping. *Journal of Animal Breeding and Genetics* 122, 1-6.

Usai MG, Sechi T, Salaris S, Cubeddu T, Roggio T, Casu S, Carta A. (2011). Analysis of a representative sample of Sarda breed artificial insemination rams with the OvineSNP50 BeadChip. En: *Farm animal breeding, identification, production recording and management. Proceedings of 37th Annual Meeting of International Committee for Animal Recording (ICAR), Riga, Latvia*, pp: 7-10. ISBN: 92-95014-10-3.

Ushizawa K, Takahashi T, Hosoe M, Ohkoshi K, Hashizume K. (2007). Expression and characterization of novel ovine orthologs of bovine placental prolactin-related proteins. *BMC Molecular Biology* 8, 95.

Uzun M, Gutiérrez-Gil B, Arranz JJ, San Primitivo F, Saatci M, Kaya M, Bayón Y. (2006). Genetic relationships among Turkish sheep. *Genetics Selection Evolution* 38, 513-524.

Van der Werf JH. (2007). Marker-assisted selection in sheep and goats. En: *Marker-assisted selection – Current status and future perspectives in crops, livestock, forestry and fish*, pp: 230-247. (Ed FAO). ISBN 978-92-5-105717-9.

Van Laere AS, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L, Archibald AL, Haley CS, Buys N, Tally M, Andersson G, Georges M, Andersson L. (2003). A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature* 425, 832-836.

Vanraden PM, Wiggans GR. (1991). Derivation, calculation, and use of national animal model information. *Journal of Dairy Science* 74, 2737-2746.

Villa-Angulo R, Matukumalli LK, Gill CA, Choi J, Van Tassell CP, Grefenstette JJ. (2009). High-resolution haplotype block structure in the cattle genome. *BMC Genetics* 10, 19.

Visscher PM, Thompson R, Haley CS. (1996). Confidence intervals in QTL mapping by bootstrapping. *Genetics* 143, 1013-1020.

Wajapeyee N, Serra RW, Zhu X, Mahalingam M, Green MR. (2008). Oncogenic BRAF induces senescence and apoptosis through pathways mediated by the secreted protein IGFBP7. *Cell* 132, 363-374.

Wallin JM, Holt CL, Lazaruk KD, Nguyen TH, Walsh PS. (2002). Constructing universal multiplex PCR systems for comparative genotyping. *Journal of Forensic Science* 47, 52-65.

Walling GA, Visscher PM, Wilson AD, McTeir BL, Simm G, Bishop SC. (2004). Mapping of quantitative trait loci for growth and carcass traits in commercial sheep populations. *Journal of Animal Science* 82, 2234-2245.

Weller JI, Kashi Y, Soller M. (1990). Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *Journal of Dairy Science* 73, 2525-2537.

Wenz H, Robertson JM, Menchen S, Oaks F, Demorest DM, Scheibler D, Rosenblum BB, Wike C, Gilbert DA, Efcavitch JW. (1998). High-precision genotyping by denaturing capillary electrophoresis. *Genome Research* 8, 69-80.

Woelders H, Te Pas MF, Bannink A, Veerkamp RF, Smits MA. (2011). Systems biology in animal sciences. *Animal* 5, 1036-1047.

Wu CS, Lan CC, Chiou MH, Yu HS. (2006). Basic fibroblast growth factor promotes melanocyte migration via increased expression of p125(FAK) on melanocytes. *Acta Dermato-Venereologica* 86, 498-502.

Xu S, Zhou Y, Yang S, Ren Y, Zhang C, Quan C, Gao M, He C, Chen H, Hhan J, Chen J, Liang Y, Yang J, Sun L, Yin X, Liu J, Zhang X. (2010). Platelet-derived growth

factor receptor alpha gene mutations in vitiligo vulgaris. *Acta Dermato-Venereologica* 90, 131-135.

Yu A, Zhao C, Fan Y, Jang W, Mungall AJ, Deloukas P, Olsen A, Doggett NA, Ghebranious N, Broman KW, Weber JL. (2001). Comparison of human genetic and sequence-based physical maps. *Nature* 409, 951-953.

Zhang Q, Boichard D, Hoeschele I, Ernst C, Eggen A, Murkve B, Pfister-Genskow M, Witte LA, Grignola FE, Uimari P, Thaller G, Bishop MD. (1998). Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree. *Genetics* 149, 1959-1973.

Zhao X, Dittmer KE, Blair HT, Thompson KG, Rothschild MF, Garrick DJ. (2011). A novel nonsense mutation in the DMP1 gene identified by a genome-wide association study is responsible for inherited rickets in Corriedale sheep. *PLoS ONE* 6, e21739.

Zhao X, Onteru SK, Piripi S, Thompson KG, Blair HT, Garrick DJ, Rothschild MF. (2012). In a shake of a lamb's tail: using genomics to unravel a cause of chondrodysplasia in Texel sheep. *Animal Genetics* 43, Suppl 1, 9-18.

Zhu M, Zhao S. (2007). Candidate Gene Identification Approach: Progress and Challenges. *International Journal of Biological Sciences* 3, 420-427.

PÁGINAS WEB CITADAS

ANCHE – Asociación de Criadores de Ganado Selecto de la raza Churra

http://www.anche.org/

Animal QTL database

http://www.animalgenome.org/cgi-bin/QTLdb/index

ARCA - Sistema Nacional de Información de Razas:

http://aplicaciones.magrama.es/arca-webapp/flujos.html?_flowId=anuncio-flow&_flowExecutionKey=e3s1

Australian Sheep Gene Mapping Website

http://rubens.its.unimelb.edu.au/~jillm/jill.htm

AWI - Australian Wool Industry

http://www.wool.com/

IFPCS - International Federation of Pigment Cell Societies Color Genes Database

http://www.espcr.org/micemut/

ISGC - International Sheep Genomics Consortium

http://www.sheephapmap.org/

MAGRAMA - Ministerio de Agricultura, Alimentación y Medio Ambiente

http://www.magrama.gob.es/es/

Ovine Genome Assembly v1.0

http://www.livestockgenomics.csiro.au/cgi-bin/gbrowse/oar1.0/

Ovine Genome Assembly v2.0

http://www.livestockgenomics.csiro.au/cgi-bin/gbrowse/oarv2.0/

FAOSTAT - Statistics Division of the Food and Agriculture Organization of the United Nations

http://faostat.fao.org/

The American Livestock Breeds Conservancy

http://www.albc-usa.org/cpl/gulfcoast.html

Virtual Sheep Genome

http://www.livestockgenomics.csiro.au/sheep/vsheep.php